Three Essays in the Economics of Higher Education

Paul David Cowell

Division of Economics

University of Stirling

A thesis submitted to the Division of Economics of the University of Stirling in partial fulfilment of the requirements for the degree of

Doctor of Philosophy

September 2017

Acknowledgements

I will forever be grateful to the Economics department at the University of Stirling for being my academic family since I arrived as an undergraduate in September 2007. To Bob Hart, who inspired me to pursue a career in academia, and who showed me how engaging a lecturer should be to students. To Frans de Vries, whose friendship, trust and guidance I have been the lucky recipient of since 2008. And to my supervisors, Mirko Moro and Tanya Wilson, who have given me all the personal and academic support I could have wished for. Thank you for your patience and your guidance.

This thesis has been supported generously by the Economic and Social Research Council, through a 1+3 studentship. To those who selected me to receive the award, I will eternally be grateful for giving the son of a joiner and a seamstress the chance to follow my academic dreams.

I am incredibly lucky to have had the support of my wife, Jen, throughout my PhD study. For all the long nights and early mornings, I apologise; and for all the cups of coffee and the hot meals, I thank you. You have been both the wind at my back and the comfort at the end of a long day. Your patience, kindness and motivation are part of the very foundations of this thesis.

To Anne, who guided me to choose Stirling, and for fanning the flame of learning.

And to my parents, Maureen and Dave. You have always given me encouragement without pressure, and sacrifice without regret. You made me feel like I could accomplish anything I wanted in life, and I could not have done this without your support throughout my education. Even throughout the toughest of times, your encouragement never wavered.

This is for you, Mam. I hope you are proud of your *canny bairn*.

Declaration

In accordance with the Regulations for Higher Degrees by Research, I hereby declare that the whole thesis now submitted for the candidature of Doctor of Philosophy is a result of my own research and independent work except where reference is made to published literature. I also hereby certify that the work embodied in this thesis has not already been submitted in any substance for any other degree and is not being concurrently submitted in candidature for any degree from any other institute of higher learning. I am responsible for any errors and omissions present in the thesis.

Candidate:

Paul David Cowell

Abstract

This thesis presents three empirical analyses in the economics of Higher Education within the United Kingdom.

The first analysis evaluates the impact of student funding reforms on participation and course choice, through the use of a difference-in-differences strategy with heterogeneous treatment effects. The results show that students who received the largest increase in study costs were less likely to move further away and also more likely to study a subject with lower graduate wage premia due to the significant reduction in the risk of investing in higher education. Students who received the largest increase in up-front financial support were more likely to attend a university further away.

The second question addresses whether undergraduate subject choice is affected by changes in the expected benefits and opportunity costs of investing in HE through variation in the labour market. Students who reside in areas of high unemployment are found to be less likely to choose subjects with the largest graduate wage and employment premia. This suggests that students may be afraid of failure in challenging labour markets and instead choose to study subjects with a greater chance of success. However, lower socioeconomic status students are more likely to study subjects with the highest graduate wage and employment premia. This suggests that the students who may be the most aware of the costs, are also the most aware of the benefits.

Finally, the third analysis investigates whether students who are socioeconomically disadvantaged incur a further penalty in terms of degree attainment. The results show that the most disadvantaged students outperform their advantaged counterparts. This may be due to pre-university attainment being an imperfect measure of ability in the most disadvantaged students, or that students who have

had to overcome the most challenges to attend university are better-equipped and more determined to succeed.

Table of Contents

List of Tables

<u>Chapter 3</u>

List of Figures

List of Abbreviations

ASHE	Annual Survey of Hours and Earnings
BIS	Department for Business, Innovation and Skills
DfES	Department for Education and Skills[1]
FE	Further Education
HE	Higher Education
HEI	Higher Education Institution
HESA	Higher Education Statistics Agency
JACS	Joint Academic Coding System
LFS	Labour Force Survey
ONS	Office for National Statistics
STEM	Science, Technology, Engineering and Maths
UCAS	Universities and Colleges Admissions Service

[1] DfES was replaced in 2007 by two new governmental bodies, the Department for Children, Schools and Families (DCSF) and the Department for Innovation, Universities and Skills (DIUS).

Introduction

This thesis presents three empirical analyses of Higher Education (HE) within the United Kingdom (UK), with an emphasis on estimating factors that can influence students' study choices and their degree outcomes. HE is defined as the final level of education that a student can undertake, after he or she completes compulsory schooling, and any pre-requisite education after school-leaving age (Further Education). Whilst there are a variety of different Higher Education Institutes (HEIs) in the UK, ranging from conservatoires (institutions that specialise in music) and specialised vocational providers; for most students, HE is obtained by going to a university.

The three analyses investigate two major themes: first, the question of how students at university respond in their study behaviour when the relative costs and benefits of acquiring HE are changing; and second, the question of whether socioeconomic factors have any deterministic effect on university academic achievement. Chapters 1 and 2 address the former, whilst Chapter 3 addresses the latter. If there is one central finding to these chapters, it is that the undergraduate student population is incredibly diverse and as such, any policy evaluation or empirical research design must be approached accordingly.

A Brief History of Higher Education in the UK

The research presented in this thesis focuses mainly on the period 2000-2008, which was a period of significant change for the higher education sector in the UK. However to place those analyses in context, it is useful to consider how the UK HE sector has changed in the period preceding. Historically, HE was provided free of tuition costs to students that attended, as funding was awarded from the public sector. This may have been sustainable when only 3% of 18-30 year olds participated in HE - as was the case in the UK in the 1950s - however student demand and participation in HE has increased significantly. As Wyness (2010) notes, the participation rate amongst young people had increased to 40% in 2000. This rising demand for, and participation in, HE and the lack of private direct contributions to the costs of the HE provision resulted in a significant lack of funding per student across universities in the UK. In response, the Dearing Report's recommendation of means-tested tuition fees were introduced for all new entrants to undergraduate study in September 1998, although contrary to the report they were introduced as an up-front cost rather than a graduate tax. However, although fees continued to be increased in line with inflation, this funding infrastructure did not provide sufficient increases in funding for universities.

Higher Education during the analysis period

In 2004, the introduction of the Higher Education Act (2004) brought about significant changes to the funding of HE in the UK. The up-front, flat-rate, means-tested tuition fees of £1,000 were replaced by variable, non-means tested tuition fees of up to £3,000 for all students independent of parental income, however they would be repaid on an income-contingent basis after graduation and the graduate's earnings were at least £15,000 at a rate of 9% of earnings above this threshold. This meant that students from lower income households who were previously exempt from paying towards their HE post-1998 would now be forced to pay the same as a student with the highest levels of parental income. Additionally, although the fees were designed to be variable, where a university could charge any amount up to a maximum of £3,000 per year, in practice almost every university charged the full amount in the first year. A further implication of this is that whilst the returns to an undergraduate degree may vary by subject, the pecuniary costs of acquiring HE did not. In addition to the fee increase, the amount of up-front financial support available to students was also increased. This was in the form of a non-means tested student loan (which, like the tuition fee loan, would be repayable through graduate earnings), and a means-tested, non-repayable maintenance grant for the poorest students. The maximum maintenance grant that a student could be awarded was £3,000, thus for the poorest students, the introduction of tuition fees was completely off-set by the new maintenance grant. However, as Chapter 1 identifies, students whose parental income was above the threshold for the full off-setting maintenance grant but who were still classed as low income may have experienced the largest relative increase in the costs of HE.

Recent developments in HE

In the 2009, the Browne Report outlined further student funding reforms, which increased the variable tuition fee cap from £3,375 per year in September 2011 to £9,000 per year in England from September 2012. As this represented a significant increase in the costs of tuition, the tuition fees charged varied by both institution and subject, such that higher quality institutions could charge the maximum fees possible, especially for subjects that they experience high demand for. As of September 2016, a further change to the funding of HE in the UK came into effect, which was to completely remove means-tested, non-repayable grants that students from the poorest families received to help fund the living costs of attending university.

In the context of introducing and increasing the tuition fees in the UK, there has been much emphasis on simultaneously improving the access to HE for students with disadvantaged backgrounds. Indeed, the report that detailed the case for the 2012 tuition fee increase stated that "everyone who has the potential should have the opportunity to benefit from higher education". (Browne, 2010: 24) However, to provide an opportunity to benefit from HE is not the same as ensuring the students (especially the poorest and the most socioeconomically disadvantaged students) do not make sub-optimal investment decisions due to a fear of debt, a fear of failure, or through cost-minimising behaviour such as a limiting of study choices by attending a university close to home. Since the introduction of the £9,000 per year fees in 2012, media and political debate has focused on the impact on student debt and student choices. Atherton et al. (2015) note that poorer students were on average 20% more likely to stay close to or live at home whilst attending university due to the increased costs of HE. This should not have been a surprise, as this thesis will show that students responded to the 2006 fee increase in exactly the same way. Furthermore, O'Leary (2016) reported that the increased fees in 2012 have led to decreased demand for undergraduate study of arts and humanities subjects (in particular, languages). However, this thesis will also show that when the relative costs of acquiring HE are directly increased (through tuition fees - Chapter 1) or indirectly increased (through labour market conditions - Chapter 2), this behaviour was also evident throughout the 2000-2008 period.

The Journey to Higher Education in the UK

Before proceeding to the empirical chapters, the final section of this introductory chapter presents a brief overview of the HE system in the UK, with particular reference to the transition from schooling and further education (FE) to HE. Throughout the UK, young people are required to remain in secondary education until the age of 16, which has been the case since 1972 when the minimum school leaving age was raised from the age of 15. Since 2013, students are required to enrol in some form of educational activity until the age of 18 which may take the form of an apprenticeship (a practical, work-based learning engagement where students learn a vocation and are paid a salary), a traineeship, or studying towards FE qualifications. If students decide to stay in full time education beyond the minimum school leaving age of 16, they can study towards their FE qualifications either at an FE college (which typically offers a range of academic and vocational subjects) or at schools that offer FE[2] (where the subjects offered are typically academic rather than vocational or practical). Given that students who attend university will usually be those who have pursued academic subjects at an FE college or by staying on full

[2] In England, Wales and Northern Ireland, these are known as Sixth Forms; a student will typically undertake two additional years of education before deciding to enter university. The first year is known as Lower Sixth and the final year at a sixth form is known as Upper Sixth. In Scotland, these two years are called S5 and S6, as they represent two further years at a Secondary school beyond the minimum school year (S4).

time at school or at a Sixth Form, students partially decide to attend university two years in advance of their eventual entry into HE.

In England, Wales and Northern Ireland, the educational system is almost identical except for minor subject differences such as compulsory Welsh language education in Wales until the age of 16. In these three countries of the UK, students will study towards General Certificates of Secondary Education (GCSEs) or equivalents during the final two academic years of compulsory schooling. The grades the students receive at GCSE level (and the subjects in which they receive them) will determine what subjects they are permitted to study at FE level between the ages of 16 and 18. Students who wish to attend university will typically study towards three or four Advanced Levels (almost always abbreviated to A-Levels) during these final two academic years, and an overall grade for their studied subject is awarded in the August of the student's last year of study.[3] These grades range from A (the highest) to E (the lowest), thus a student who obtains three A grades at A-Level are denoted AAA.[4] During their final year, students will apply to university through the Universities and Colleges and Admissions Service (UCAS) and receive a conditional offer, such that the results they receive in August will determine whether or not a student has met the conditions set by the university to attend in September. These conditions are usually that a student obtains a set of minimum A-Level grades (and potentially a minimum grade in a subject designated by the university).[5] If the student is successful in meeting those requirements, they can

[3] Since 2015, there have been changes to the A-Level format, whereby a student's overall grade is determined by an exam at the end of the two years of study. Furthermore, it used to be the case that an A-Level was comprised of two parts (reflecting the first and second years), namely the AS-Level and the A2-Level. Before 2015, students would typically choose four A-Levels to study in the first year of further education, and in the second year they would choose three of these four A-Levels to convert into a full A-Level by completing the 'A2' year. The period of analysis for this thesis however, does not contain these changes.
[4] In 2010, an additional grade was introduced of A*, however the change does not affect the period of analysis.
[5] Universities in the UK advertise their minimum requirement for entry onto a degree programme for a particular subject. Some universities may advertise a minimum grade for a particular subject, some may require simply a set of minimum grades from any of the student's chosen A-Levels, or the university may instead advertise a minimum UCAS tariff entry score. This tariff score is often used as A-Levels are not the only qualifications a prospective HE entrant can use to apply to university with, especially in the case of international students. The tariff score is computed by

confirm their attendance on an undergraduate degree course at university starting in September[6], which typically last for three years.

In Scotland, the process is largely the same, however the educational framework in Scotland is aligned to the Scottish Credit and Qualifications Framework (SCQF), where each progressive stage of education is considered a numerical level higher than the last.[7] Whilst the school-leaving age is still 16 in Scotland, there is no requirement to stay engaged in an educational activity past this age. Furthermore, students do not study GCSEs, but Higher Awards (shortened to Highers) which are the most common qualification used for entry into a Scottish university. This qualification, however, is not as advanced as the A-Level, and as such, degree programmes in Scotland last for four years. Students at Scottish schools and FE colleges do have the option to undertake Advanced Higher Awards (shortened to Advanced Highers), however this level of study is similar to the first year of the four year Scottish undergraduate degree programme in terms of difficulty. Students who do obtain Advanced Highers and who attend a Scottish university typically enter into Year 2 of the degree programme.[8]

Such fundamental structural differences mean that, even before considering the significant differences in student funding and tuition costs, direct comparisons between Scottish and English universities are complex. This is the fundamental reason why Scottish universities should not be used as a counterfactual for policy evaluation at English universities; and on the contrary, why Welsh universities can be used. This identification of a valid counterfactual case is used in Chapter 1 to

converting grades and qualifications into a numerical indicator of pre-university attainment, where the best grades on the highest qualification types score the largest numbers. A student's overall score is computed by aggregating their individual tariff points computed from their individual grades and qualifications. A table tariff points by grade and qualification can be found in Appendix C, Table C1.

[6] Or alternatively the student may elect to delay entry by one year, known as a gap year.

[7] For instance, a doctoral degree is the highest at level 12, a masters degree is level 11, and an undergraduate is level 10.

[8] For a complete overview of the SCQF framework, see http://www.scqf.org.uk/framework-diagram/Framework.htm

evaluate the impact of the 2006 funding reforms. Where policy evaluation through a natural experiment is not the research design, Scottish universities are included in the identification strategy, with attention paid particularly to start dates and degree programme lengths to correctly assign students and their influencing factors the appropriate time dimension. Chapter 1 therefore includes by definition only England and Welsh universities, whilst Chapters 2 and 3 are allowed to use observations from universities across the UK.

Chapter 1

Evaluating the Impact of the 2006 'Top-Up' Fees on Subject Choice & Geographic Mobility

"The social class gap among those entering higher education is unacceptably wide. Young people from professional backgrounds are over five times more likely to enter higher education than those from unskilled backgrounds."

(DfES: The Future of Higher Education, 2003)

1.1 Introduction

It remains a rational argument that an individual who benefits from education in the form of higher lifetime earnings should be expected to contribute towards the cost of acquiring that education. An important caveat is that those who are unable to afford the education at the outset should not be required to contribute to the costs until the benefits begin to realise, otherwise this will provide disincentives to participate in Higher Education (HE).

The right to an education is a fundamental element of most societies, founded on ethical, moral and legal considerations. Unlike basic schooling however, HE is an optional extension to an individual's investment in human capital. Whilst the provision of schooling in the UK – as in most developed economies – is free and universal, a student who enrols to study at the HE level must contribute to the costs of the HE provision. Following Becker's (1964) seminal work, the acquiring of HE is ultimately an investment decision for the individual, where the investment is undertaken if the costs (tuition costs, foregone earnings, living costs, relocation costs and psychic costs) are outweighed by the discounted net present value of the increase in lifetime earnings. However, since the returns to a degree not only allow higher lifetime earnings, but also allow upward social mobility, it is crucial that those who wish to, and have sufficient school attainment and ability,[9] are able to do so. Therefore, even if HE is not provided free of charge, there should be a right of access which particularly applies to students from lower income or otherwise disadvantaged backgrounds.

In the absence of means-tested grants, increases in or introductions of tuition fees have been widely found to cause a reduction in university enrolment. A significant body of research exists for tuition fees in the US, where there is both a large degree of heterogeneity in fee levels between states, as well as a reduced geographic mobility since fees are higher for out-of-state students. The collective consensus is that a $1,000 increase in fees reduces enrolment by around 2-5 percentage points (Hemelt & Marcotte, 2008; Kane, 1994; Heller, 1997; McPherson & Shapiro, 1991; Leslie & Brinkman, 1987). Dynarski (2003) also finds that a $1,000 increase in grants (the reverse of a fee rise) increases enrolment by 3.6 percentage points. In Europe, there is mixed evidence on the effect of tuition fees on participation. When fees were introduced in some German states in 2005-2007 at a rate of €500 per semester, Hübner (2012) and Dwenger et al. (2012) find that they reduced participation by 2-2.7 percentage points, in line with the US findings. However, other studies of the German tuition fee experiment (Bruckmeier & Wigger, 2014; Alecke et al., 2013) and evidence from tuition fees in the Netherlands (Huijsman et al., 1986; Canton & De Jong, 2005) find that there was no impact on enrolment.

In addition to the student's perspective, the universities' financial needs must also be satisfied for a student to obtain HE of a sufficient quality. In the UK, the university participation rate amongst 18-30 years olds has risen from around 5% in the early 1960s, to 40% by the year 2000, whilst university funding per HE student had reached its historic low. (Wyness, 2010).

[9] New entrants to HE are not constrained to school leavers, however they are the group for which a fair right of access is arguably the most important.

These two sentiments were the main motivation for the Higher Education Act 2004 which aimed to reform the provision of student support and the financing of HE provision, whilst encouraging widening participation amongst school leavers. In September 2006, tuition fees for new entrants to HE were increased from a means-tested £1,175 in the previous year to a flat fee of £3,000[10], but now the fees would be paid after graduation on an income-contingent basis, instead of being paid up-front under the old scheme. In addition to the tuition fee changes, the reform also introduced a large increase in up-front, non-repayable, means-tested maintenance grants to promote wider access to universities for lower-income background students.[11]

This chapter evaluates the impact of the 2006 funding reforms on the participation and choice behaviour of lower income background students to study at the undergraduate level. The contribution is that it is the first research to examine the policy change using the entire student population, using Welsh Higher Education Institutes[12] (HEIs) as the counterfactual in a difference-in-differences (DiD) approach. A further contribution is that unlike other research designs, the estimate of the treatment effect of the policy change in this analysis is allowed to vary across different types of students, depending on their subject and university choice, which relaxes the assumption that all student types and subjects were affected equally by the policy change. The evaluation is performed in a natural experiment setting, since the student support package available in England and Wales was identical, except for fees being increased in England, and being kept constant in Wales in September 2006.

The data used is the complete student population who left a UK HEI between the academic years 2005/06 and 2009/10. Their recorded start date, expected length of

---

[10] The new fees were widely known as 'top-up' fees, as they were introduced to top-up the funding shortfall for universities. Furthermore, although the fees were variable with a £3,000 cap, almost all universities chose to charge the full amount (Universities UK, 2009).

[11] See section 1.2.1 for a detailed overview of the changes in HE funding and student support, and section 1.2.2 for the 2006 policy change in particular.

[12] This term is used interchangeably throughout alongside 'universities'.

study, and actual length of study, are used to create an ex-ante population of new entrants between the academic years 2003/04 and 2007/08. Using a difference-in-differences (DiD) approach, the causal impact of the increased fees for English students is estimated, with a difference-in-difference-in-differences (DDD) approach to allow this estimate to be heterogeneous across different groups of students according to the subjects they chose to study and the distance between their parental home and their chosen university.

The results show that the policy reform had a significant impact on the participation and course choice of those students who were the most likely to suffer the largest net increase in tuition fees. These were students who, although classified as from a low income background, only received a partial or zero award of a non-repayable grant that was provided to offset the increase in tuition fees. The policy led to a 1.72 percentage point increase in the probability of observing such a student participating in HE overall, which suggests that at the aggregate level, the funding reforms (including the fee rise) actually increased participation of the students experiencing the largest net increases in course costs. When disaggregating the policy effect across subject types, the same classification of low income students were 1.77 percentage points less likely to study a subject with the highest graduate earnings premia, which initially is counter-intuitive.

However, when allowing for heterogeneous effects of the policy across subject types and whether a student decided to attend a university in the local area, the findings show that the policy led to a 2.87 percentage point increase in observing the lower income students most likely to experience the largest increase in course costs who are choosing to remain close to home whilst attending university, and choosing a subject with the lowest graduate earnings premia. This suggests that since the costs of attending university are repaid through income-contingent loans under the new scheme, students who do not receive the non-repayable grant to offset the fee increase are incentivised to engage in risk-free studying (since the amount a student repays is dependent on future earnings) at a local university and studying a subject that does not necessarily provide significant rewards in terms of graduate wage premia. These students were also observed to be less likely to attend a university far away, especially if the university was not an elite university[13], whereas low income students who received the offsetting non-repayable grants were on average 1.62 percentage points more likely to study at a university far from their registered address at the time of university application.

The rest of the paper is structured as follows: section 1.2 gives an overview of tuition fees and student support in the UK from 1998 until the 2006 reform and outlines the exogenous variation exploited; section 1.3 discusses the features of the data as well as the identification strategy; section 1.4.1 outlines the methodological approach of the paper with both the technical and empirical aspects, with section 1.4.2 addressing the implications of estimating a difference-in-difference regression with a binary outcome variable, with sections 1.4.3 and 1.4.4 discussing the DDD criteria of distance and subject groupings; section 1.5 presents the results of the analysis, with finally policy implications and conclusions in section 1.6.

1.2 Institutional Setting

1.2.1 Tuition Fees in the UK

Tuition fees were introduced in the UK as s response to the Dearing Report in 1997, where means-tested tuition fees of up £1,000 were introduced per year of a student's undergraduate degree for all new entrants to HE in September 1998.[14] The fees were to rise in line with inflation each year, which in practice resulted in yearly £25 increments for each cohort of new entrants. By 2005, the tuition fee level was £1,175. The rationale for means-testing the tuition fees was to try and ensure that only those that could afford to contribute to towards the cost of their degree did so,

[13] A university is classed as elite (for the purposes of this analysis) if it is listed as being a Russell Group university; a group of 24 (as of September 2017) highly prestigious universities in the UK. For a full list, see Appendix A, Table A1.
[14] See Table 1.1 for changes in tuition fees by parental income.

without providing disincentives to study for students from lower income backgrounds. In 2000, around 40% of HE students paid no fees, 20% paid partial fees, and 40% paid the full tuition fee (Greenaway & Haynes, 2003).

Table 1.1: Changes in Tuition Fees Levied by Parental Income (1992-2006)

Parental Income (£)	Sept 1992 Fee (£)	Sept 1998 Fee (£)	Sept 2004 Fee (£)	Sept 2006 Fee (£)
0 - 10,000	0	0	0	3000
20,000	0	373	0	3000
30,000	0	1172	980	3000
40,000	0	1172	1196	3000
50,000+	0	1172	1196	3000

Source: Dearden et al. (2011)

However, the income from those who paid fees arguably only increased revenue for universities marginally, and any policy that attempts to address widening participation should look beyond the basic cost of tuition; a common finding in the literature is that those from lower income backgrounds or from other under-represented or disadvantaged student groups have higher perceived costs of entering HE (Dunnett et al, 2012; Callender & Jackson, 2008).[15] Additionally, the 1998 reform did not seek to directly improve geographic mobility, as student bursaries were removed for students from lower-income backgrounds in 1999.[16] However, this reduction in grants coincided with maintenance loans being made available to students on an income-contingent basis in an attempt to leave students

[15] Callender & Jackson (2008) also find that not only do students from low income backgrounds perceive the costs as greater, they are more likely to see the cost of acquiring HE as a debt, whereas students from wealthier backgrounds see it as an investment.
[16] In September 1998, students could take out a maintenance loan to cover their costs of attending university (separate from the tuition costs), with lower-income students receiving a maximum annual (non-repayable) bursary of £810. In September 1999, the bursary amount was added to the maximum loan amount on a means-tested basis. See Table 1.2.

no worse off in terms of available support to attend university, but for lower income students their costs of attending university increased. Research by the Department for Business Innovation and Skills (BIS) using Labour Force Survey (LFS) data found that the 1998 reforms caused a 5.4 percentage point decrease in the participation rate amongst students from wealthy backgrounds, however there was no significant impact on the participation rate for students from low income backgrounds (BIS, 2010a).

1.2.2 The 2006 Reform

As it became apparent that the 1998 reforms had done little to improve funding for universities or access for students, a government report was published in January 2003 proposing universities could charge a ‘top-up’ fee up to a maximum of £3,000. The Higher Education Act received royal assent on 1st July 2004, and the increased fees would be active for new entrants in September 2006. This fee would be levied universally for all students on a non means-tested basis, however fees would be paid through the issuance of a tuition fee loan of the fee charged, which would be paid on an income-contingent basis when graduates were earning over £15,000 per year. To offset the increase in fees for the poorest students, means-tested maintenance grants of £2,700 per year were introduced, with an additional means-tested bursary of £300 to be provided by the university. Hence, for those students, the increase in fees from £0 to £3,000 was completely offset by the increase in grants and bursaries. The changes in student funding for England is summarised in Table 1.2 below.

Table 1.2: Changes in HE student funding for England (1998-2010)

	Sept 1998	Sept 1999	Sept 2000	Sept 2001	Sept 2002	Sept 2003	Sept 2004	Sept 2005	Sept 2006	Sept 2007	Sept 2008	Sept 2009	Sept 2010
Fees Payable Up Front (£)	1000	1025	1050	1075	1100	1125	1150	1175	0	0	0	0	0
Fees Payable in Graduate Earnings (£)	0	0	0	0	0	0	0	0	3000	3070	3145	3225	3290
Maximum Maintenance Loan available (£)	2735	3635	3725	3815	3905	4000	4095	4195	4405	4510	4625	4950	4950
Bursary available (£)	810	0	0	0	0	0	1000	1000	2700	2765	2835	2906	2906
Maximum Loan & Grant available (£)	3545	3635	3725	3815	3905	4000	5095	5195	5905	6045	6200	6405	6405

Sources:
SLC, Bolton (2014)
Notes:
All figures in nominal £ per academic year. Maximum Loan & Grant available: as household income increases, award of maintenance grant is reduced whilst amount of available loan increases to offset.

As seen in Table 1.2, in September 2006 the maximum combined loan and grant available for students is £5,905, of which up to £2,700 is available as a non-repayable bursary. It should also be noted that in September 2004, the Higher Education Grant of up to £1,000 was introduced to help support students from low income backgrounds. This was means-tested, and students whose household income was less than £15,200 received the full £1,000; whereas students whose annual household income was between £15,200 and £21,185 received a partial award on a decreasing basis.[17] However, to evaluate the impact of the funding reforms on students from lower income backgrounds, it is necessary to observe the

[17] SLC (2006) Higher Education Grants in England and Wales: Academic Year 2005/06 (Provisional).

changes in support by parental income.[18] Table 1.3 shows the gains/losses for a student from moving from the old funding system to the 2006 system, by parental income, which is also used to create Figure 1.1.

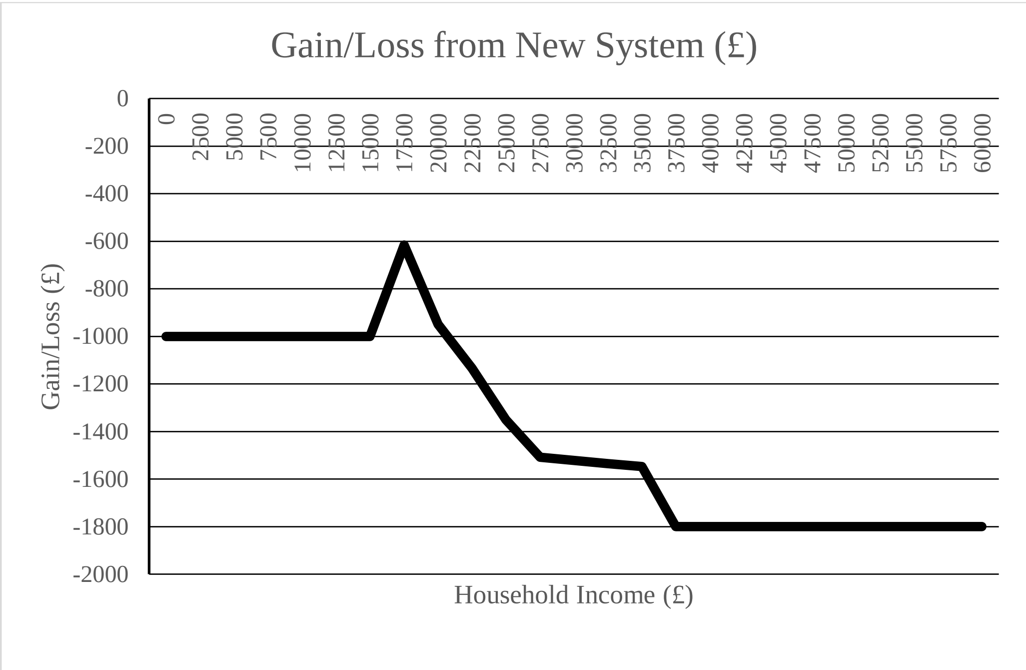
Table 1.3: Gains/Losses by Household Income in Transitioning from Old to New System

		Household Income			
		< £22,500	£25,000	£45,000	>£60,000
Old System in 2006	Fees	1200	1200	1200	1200
	Fee Subsidy	1200	1000	0	0
	HE Grant	1000	0	0	0
	NET	1000	- 200	- 1200	- 1200
New System (2006 Reform)	Fees	3000	3000	3000	3000
	Grant	2700	1449	0	0
	Bursary	300	0	0	0
	NET	0	- 1551	- 3000	- 3000
Change from Old-New		- 1000	- 1351	- 1800	- 1800
Change in Up Front, Non-Repayable Support		+ 2000	+ 1449	0	0
Increase in Costs (assuming no loan taken out)		+ 3000	+ 2800	+ 1800	+ 1800

Source: SLC, Dearden et al. (2008)
This table shows the difference in moving from the pre-2006 student finance system to the 2006 funding reform system. The net gain or loss is calculated by parental income groupings. Note that Dearden et al. include in their calculations a fee and maintenance loan subsidy, based on the fact that the loans taken out are at a zero interest rate. These subsidies are omitted here to clarify the student's decision making process when deciding to enter HE.

[18] The data used for the analysis in this chapter does not include specific values of parental income, however the net gains or losses calculated here are crucial for identifying students who suffer the greatest increases in costs whilst still being from a low income background.

Figure 1.1: Gains/Losses by Household Income in Transitioning from Old to New System at Representative Intervals



Focusing on the definitive aspects alone (i.e. the fees charged and the grants or bursaries awarded, and thus assuming that a student does not take out a loan[19]), Table 1.3 shows that the poorest students (parental or household income less than £15,000) suffer a financial loss in moving from the old system to the new. Although they receive £3,000 in non-repayable grants and bursaries, this only represents a £2,000 increase from the old system. At the same time, their tuition fees increase by £3,000, albeit not as an up-front cost. Students whose parental or household income is between £15,000 and £22,500 suffer less of a financial loss in switching systems, and students whose parental income was over £22,500 suffer an increasing financial penalty in the new system.

[19] This is different to the approach taken by Dearden et al. (2008) who include in their calculation the assumption that each student will take out the maximum maintenance loan possible.

Furthermore, for students whose parental income was over £17,500, their tuition fees are greater than their non-repayable grants and bursaries.[20] Hence, for these students whose parental income is above the threshold for the award of a full grant that offsets the fee increase, but is low enough to still be considered a student from a low income background, may be the most likely to be adversely affected in terms of an increase in the actual and perceived costs of HE. Conversely, since the fees are repaid on an income-contingent basis, the poorest students receive the largest increase in up-front non-means tested support despite suffering a negative financial penalty in moving from the old funding system to the new. This should allow for increased geographic mobility, which is especially important for low income students who reside in rural areas or who wish to study specialist subjects at university. These are the testable hypotheses that will be examined in the empirical analysis.

The majority of the existing research on the effects of the 2006 reform on participation finds that it had no effect on either participation overall, or on the participation of students from low income backgrounds in particular (Faggian, 2010; BIS, 2010a; BIS, 2010b; Wakeling & Jefferies, 2013; Universities UK, 2009). Due to the simultaneous increasing of fees and student grants, most approaches are unable to disentangle the opposing effects. Dearden et al (2011) use LFS data to separately estimate the effects of the increase in fees from £1200 to £3000, and the increase in student grants. Their estimates suggest that a £1000 increase in fees leads to a 3.9 percentage point decrease in HE participation, whilst a £1000 increase in student grants increases HE participation by 2.6 percentage points. As most approaches are estimating the effect of these concurrently, it may explain why the majority of research finds no significant overall impact of the reforms. An alternative approach of evaluating the effect of the reforms on a student's engagement with HE is university attrition (or 'drop-out') rates. Bradley

[20] For a comprehensive comparison between the old and new funding systems by household income, see Appendix 1, Table A2.

and Migali (2015) investigate this, and find that the reforms led to a fall in dropout rates, especially for the poorest students. However, the existing body of research either does not use a counterfactual outcome (Dearden et al, 2011; BIS, 2010a; Universities UK, 2009), and those that do use a difference-in-differences approach (Faggian, 2010; BIS, 2010b) perform only a simple, descriptive analysis, or use a counterfactual that is arguably too dissimilar.[21] Finally, none of the existing analyses allows for the reform to affect different groups of students heterogeneously within a regression difference-in-differences framework.

This research addresses these concerns by using the entire population of undergraduate students in England and Wales who began their studies in the academic years 2003/04 to 2007/08, which allows a regression DiD approach with several time periods. Furthermore, the estimate of the impact of the policy is allowed to vary across groups of students who may be substantially different (i.e. the funding reforms are not expected to affect students who attend a local university to study an arts subject in the same way they affect students who attend a university far away from their address at the time of application to study medicine).

[21] Faggian (2010) uses postgraduates as the counterfactual for the effect of the 2006 reforms on undergraduates in Northern Ireland. There is a significant argument to be made that postgraduates are not influenced in the same way that undergraduates are when choosing to study at university. Postgraduates are usually older, are generally not subject to the same financial constraints as school-leavers, and are overwhelmingly motivated by employment prospects.
BIS (2010b) uses Scotland as a control group for the impact of the reforms in England. There are again, many arguments to be made why students at Scottish universities are not sufficiently comparable to be a valid counterfactual for students at English HEIs. Firstly, the underlying educational structure differs significantly – in England and Wales, students study for GCSEs and A-Levels, whereas in Scotland, students study towards Standard Grades, Highers, and Advanced Highers. Consequently, undergraduate degree programmes in Scotland are typically four years' long, whereas they are three years' long in England and Wales. Finally, Scotland operated a graduate endowment scheme between the introduction of the fees in 1998 and the 2006 reforms, which removes the possibility of analysing a 'clean' policy break. As BIS (2010b) shows, the common trends assumption is violated.

1.2.3 The Experiment

The Welsh Assembly government was established by the Government of Wales Act 1998, and it obtained full devolved legislative powers in 2006 with the passing of the Government of Wales Act 2006, receiving royal assent on the 25th July 2006. Prior to the adoption of the Higher Education Act 2004 in England, the Welsh government delayed the adoption of the increased fees by one year – with the new fees becoming active in September 2007 instead of September 2006 as in England. Furthermore, even though the new fees were not adopted until 2007, the Welsh Assembly announced that they would introduce the same student support scheme as in England in 2006, in the form of grants and loans.[22] Additionally, the Welsh government announced in 2006 that when the increased fees were adopted in September 2007 (and for the foreseeable future), Welsh-domiciled students[23] would be given a non-means tested fee grant to fully offset the fee increase that happened in England[24]. Hence, the only difference between the policies in England and Wales in 2006 was that in England the fees were higher. The use of undergraduate students at Welsh universities as a control for the estimation of the effects of the policy change on undergraduate students at English universities is therefore reasonable, given the insulation from the fee increase for Welsh students. The use of a valid counterfactual in this empirical approach is a significant improvement over the use of postgraduate students compared to undergraduate students, which is the approach employed by Faggian (2010).

Normally, differential fees across a small spatial distribution with free movement of individuals would not allow a true estimate of the effect of the reform, since

[22] The Welsh government had already introduced the Assembly Learning Grant (ALG) for new entrants in the 2002/03 academic year, which made a £1,500 means-tested, non-repayable grant available for Welsh-domiciled students. When the Higher Education Grant (HEG) was introduced by England in 2004/05 for new entrants, the ALG simply topped the £1,000 award up by £500 for Welsh-domiciled students.

[23] A Welsh-domiciled student is any student who is a permanent and primary resident in the country of Wales at the time of applying to university, and has been for 5 consecutive years prior to entry into HE.

[24] This fee grant was payable directly to the university, so students could not use it to supplement their maintenance costs of attending university.

students could simply migrate from a high fee to a low fee area. However, English students (specifically, students from the rest of the UK who were not registered as living in Wales for 5 consecutive years previous to HE entry) would be liable to the same fee at Welsh HEIs as English HEIs. Conversely, the non-means tested fee grant awarded to Welsh-domiciled students from September 2007 onwards was valid for Welsh-domiciled students studying anywhere in the UK. Hence, there was no incentive for cross-border migration to avoid higher fees or to obtain more generous student support. The changes discussed in this section are summarised in Table 1.4 below.

Table 1.4: Changes in HE student funding for Wales (1998-2010)

	Sept 1998	Sept 1999	Sept 2000	Sept 2001	Sept 2002	Sept 2003	Sept 2004	Sept 2005	Sept 2006	Sept 2007	Sept 2008	Sept 2009	Sept 2010
Fees Payable Up Front	1000	1025	1050	1075	1100	1125	1150	1175	0	0	0	0	0
Fees Payable in Graduate Earnings	0	0	0	0	0	0	0	0	1200*	1225	1255	1285	3290
Maximum Maintenance Loan available	0	0	0	0	0	0	0	0	1200	3070	3145	3225	3290
Maximum Loan & Grant available	2735	3635	3725	3815	3905	4000	4095	4195	4405	4510	4625	4745	4745
Bursary available	810	0	0	0	1500	1500	1500	1500	2700	2765	2835	2906	5000

Sources:
SLC, Bolton (2014)
Notes:
All figures in nominal £ per academic year.
Maintenance loan - assuming not living at home, attending university outside London. Figures show maximum possible award. Note that this figure is reduced as parental income increases and also varies depending on award of means-tested grant.
Bursary - In September 2002, the Welsh assembly introduced a means-tested, non-repayable grant of £1,500. When the English system introduced maintenance grants in September 2004, the Welsh assembly grant (ALG) simply topped up the £1,000 award by £500.
*Fee of £1,200 only applied to Welsh HEIs, otherwise Welsh-domiciled students studying outside Wales in UK are liable to full 'English' fees.
Maximum Loan & Grant - as household income increases, award of maintenance grant is reduced whilst amount of available loan increases.

1.3 Data

1.3.1 HESA Data

The data used to analyse the 2006 funding reform comes from the Higher Education Statistics Agency (HESA), and it is a composite of three groups of student populations: the destinations of leavers population[25], the qualifiers population[26], and the student standard registration population.[27] The composite population therefore contains the record of all students who were registered at a Higher Education Institute (HEI) in the UK, and who left between the academic years 2005/06 and 2009/10. The period for analysis is such that there is a steady period before the reform in both the treatment and counterfactual groups, with an additional year after the policy's introduction to capture any re-adjustment behaviour. The advantage of using such data is the student population contains all students who left a HEI, including students who did not successfully complete their studies. Where the latter is true, a reason for an unsuccessful or early termination of studies is given. Students who are not included in the dataset are dormant[28], postdoctoral, visiting, exchange, writing-up students[29], or students on sabbatical.[30]

The dataset contains detailed information for each student observation, including course information (subject code, year of study, expected length of study, full or part time study, degree classification, level of study, end date, start date), university information (institution code, campus identifier, accommodation status), as well as student characteristics (age, gender, ethnicity, disability status, home postcode,

---

[25] Students who have left HE and responded to a follow-up survey with regards to their work or study situation, 6 months after graduation.
[26] All students who obtain any form of HE qualification.
[27] All students who are registered to study at an HEI.
[28] Students who have suspended their studies but are still registered at an HEI.
[29] Students who have completed their expected length of study, but who are registered at an HEI solely for writing their final dissertation or thesis – i.e. they are not actively engaged in studying.
[30] The original dataset contains more than 3.61 million student observations, however not all are relevant for the analyses in Chapters 1 to 3. Of the original dataset of 3.61 million, approximately 760,000 observations are retained for potential analysis, however this is before the specific identification strategies followed in the relevant chapters. See Appendix A, Table A3 for a detailed overview of the data cropping.

tariff score on entry, parental higher education indicator, socioeconomic status[31], new entrant to HE indicator, fee status, funding information).

As each student is recorded in the academic year they leave university for whatever reason, it is necessary to transform the population of leavers into a population of new entrants, which is easily performed using the student's start date, expected length of study, and actual time spent studying.[32] Each student in the population of new entrants is further assigned a calculated distance between their home postcode (recorded at the time of applying for university) and their eventual university attended. This is done by using eastings and northings to generate Euclidean distances in kilometres between all possible postcodes in the UK, and assigning the correct distance to a student depending on their home and university postcode.

One obvious limitation of this data is that there exists inherently a self-selection bias: students in the dataset are demonstrably those who have decided to participate in HE. Ideally it would be possible to record the participation decision and then subsequently observe students' choice of subject and university.[33] Hence, although it is not possible to identify a pure participation effect with the HESA data, it is possible to identify how the 2006 reform affected *how* students participated in HE, conditional on them having decided to participate.[34] However, one important advantage of the graduate data as opposed to applicant data is that it allows an overview of an entire student's engagement with higher education, including

[31] This code (1-9, with 1 = highest, 8 = lowest, and 9 = unknown) is derived from the occupation of the parent, guardian, or the student themselves (if neither parent nor guardian is applicable) with the highest income to the household.

[32] An alternative dataset considered was the population of new entrants, which would ensure maximum coverage of the new entrants in each academic year – in the leavers population, those students who take longer to complete their studies may not be captured. However, the advantages of the HESA data used in this analysis are that the completion status, degree classification, and time spent studying on a degree course are included for each student observation.

[33] Studies that have used such longitudinal data to capture the participation decision include Dynarski (2003), BIS (2010a) and Dearden et al. (2011).

[34] This participation decision would be captured using UCAS application data, but enquiries to link individuals' UCAS application information with their HESA student record were not fruitful. A further benefit of using the UCAS data is that students typically apply to more than one university, so in the HESA data this paper uses almost certainly observes students on courses and at universities which were the students' first choice, and which were the students' second, third or fourth choice.

attrition and award outcomes, and eventual length of study. The resulting student observation is arguably richer in information with regards to a student's entire engagement with higher education compared to simply observing students at their entry point.

1.3.2 Identification Strategy

Since the 2006 reforms were targeted at new entrants at the undergraduate level, only new entrants studying for an undergraduate degree between the academic years 2003/04 and 2007/08 are selected, both in the treatment group of English HEIs and the control group of Welsh HEIs. A more selective form of the treatment and control groups would be to only select English-domiciled students at English HEIs, and Welsh-domiciled students at Welsh HEIs. If students were able to migrate across borders to avoid the higher fees, this would invalidate the assumptions behind the difference-in-differences estimation.[35] Whilst this is possible using the home postcode of the student, it is not necessary for the purposes of this analysis. This is not only because the reforms provided no incentive for cross border migration, but additionally these migration flows did not change significantly over the period in question. This is a result that both Universities UK (2009)[36] and Wakeling & Jefferies (2013)[37] find. However, it is important to also note that the choice of study location depends not only on the cost of tuition, but also on the local living cost, the quality of the degree programmes and the universities themselves. If these migration assumptions are violated, however, it may question the validity of the results. Nevertheless, it is reasonable to expect that over a 5-year period there is not significant heterogeneous variation that would confound the analysis (especially in terms of living costs), however university and subject fixed effects are included in the analysis as controls.

[35] See section 1.4.1 for further discussion.
[36] See Universities UK (2009), Appendix 3: 'Cross-border flows: data for 2003/04 – 2006/07'.
[37] See Wakeling & Jefferies (2013), Figures 3 & 5.

A further distinction is made to select only students who were eligible to pay home fees (e.g. £3,000 in England and £1,200 in Wales in 2006), and since the analysis requires information on income backgrounds, only students with a valid socioeconomic code (SEC) are kept. Similarly, since the information is necessary to distinguish between sub-groups of students, only students without a missing or unknown home postcode, age, gender and fee eligibility are retained for analysis. As each student is given a SEC code (derived from the student's parental occupation that is the main contributor to the household income), it is possible to assign students to either a low or high income background. As is commonplace in the literature, students with SEC code 1-3 are denoted high income, whilst those with SEC code 4-8 are denoted low income. Observations that have SEC code 9 are dropped from the dataset, as this corresponds to 'Unknown'.[38]

However, given the nature of the reforms, students who are classified as low income but whose household income is too high to award the full offsetting meant-tested grant (household income equal to or greater than £17,500) are the most likely to be adversely affected by the tuition fee increase. Therefore, the low income grouping is further subdivided into an upper-low income background (SEC code 4-5) and a lower-low income background for those who are most likely to receive a full maintenance grant award (SEC code 6-8).[39] Since the upper-low income background students may be the most affected by the fee increase, it is rational to expect that this group of students will place more importance on graduate earnings and employability, whilst also the potential for observing some cost-reducing behaviour by choosing to live at or near the parental home whilst attending university to minimise maintenance costs. Students from the upper-low income

[38] A further extension to this analysis could be to investigate the impact of the funding reforms on those with an unknown socioeconomic code, however such an approach is not considered here.
[39] Ideally, the data would include information on parental or household income, so that students could be perfectly assigned to the upper-low and lower-low income categories, to reflect the lack of offsetting maintenance grants where household income is £17,500 or over. As such information is not available, the assignment of SEC codes 4-5 to the upper-low income category is checked for some robustness using the median earnings through the Annual Survey of Hours and Earnings (ASHE) by Standard Occupation Code 2000 (SOC 2000), from which the socioeconomic code is derived by HESA. See Appendix A, Table A4.

category who are unable or do not wish to stay in the local area to attend university will be the most adversely affected in financial terms by the reforms, whilst those that can live at home may be insulated from the increased costs. Hence, it is plausible that risk-averse behaviour in terms of subject and university choice will be observed in the former but not the latter. Nevertheless, although students are now assigned to one of three socioeconomic groups, the upper-low and the lower-low income categories are demonstrably still low-income. The benefit of splitting the low-income category into two is that it allows an identification of potentially opposing effects, whereby the upper-low income students are the most likely to bear the full increase of the tuition fees, and the lower-low income students are the least likely. Therefore, treating both groups as an average may lead to a loss of important information on how students responded to the policy change.

In terms of the HEIs, the HESA data contains the details of every student at every recognised HEI, which creates significant variation in the quality and academic focus of HEIs. To overcome this issue, students are only retained in England and Wales who are registered at a HEI that is not strictly vocational, such as arts colleges or drama schools. Specifically, the universities included for analysis are the Welsh and English universities that are identified by Gibbons and Vignoles (2009) as non-vocational, to remove HEIs such as arts and drama schools from the estimation.[40] To further classify HEIs, a distinction is made between elite and non-elite universities[41], by using the classification of the 24 members of the Russell Group.[42] The resulting dataset for analysis is approximately 690,000 observations, after

[40] See Gibbons & Vignoles (2009), Table A1; amended to include non-vocational Welsh, Scottish and Northern Irish universities as shown in Appendix A, Table A5.

[41] It is arguable that this definition of elite and non-elite universities may not fully capture any effects on the study choice behaviour of low income students, since low socioeconomic status students and minorities are typically under-represented in such universities. An alternative method of controlling for university quality could be to use an indicator for post-1992 institutions, or a continuous measure of university quality such as league table rankings. Such approaches are not taken in the analysis in these chapters for simplicity, but offer potential future avenues for research.

[42] See Appendix A, Table A1.

removing all student observations that are not from English or Welsh HEIs (approximately 70,000 of the cropped dataset outlined in Table A3.)

1.4 Methodology

1.4.1 Difference-in-Differences (DiD)

As has become increasingly popular in the field of policy evaluation, a difference-in-differences (DiD) analysis estimates the average treatment effect on the treated (ATET). To capture the ATET, it is necessary to observe the treatment group before and after the treatment, with and without the treatment. In essence, this would be to compare the observed outcome of the policy on English students with the participation behaviour of English students in the same time period in the absence of the policy. To solve this omitted variable problem, a counterfactual outcome is obtained from a control group that acts as the treatment group without treatment. It is therefore crucial that the common trends assumption holds, such that in the absence of treatment, the evolution of the treatment group is the same as the control group. In a regression setting (as shown in equation 1), a difference-in-differences (DiD) analysis is performed by regressing the dependent variable, y_{it} (the probability of observing a low income student i in country *c* at time *t*), on a country dummy, C_c (that takes the value 1 for the country in which the policy is active), a time dummy, T_t (which in the case of only two time periods takes the value 1 in the treatment period, and 0 in the pre-treatment period) and the interaction between the two (captured by the coefficient τ).

$$y_{i,c,t} = \alpha + \beta_1 T_t + \beta_2 C_c + \beta_3 X_{i,c,t} + \tau(G_c * T_t) + \varepsilon_{i,c,t} \tag{1}$$

This interaction, if the assumptions of OLS hold, correctly identifies the ATET:

$$\begin{aligned}\tau^{DID} &= E\left[y_{i,c,t=1} - y_{i,c,t=0} \middle| C_c = 1\right] - E\left[y_{i,c,t=1} - y_{i,c,t=0} \middle| C_c = 0\right] \\ &= E\left[\nabla y_{i,c} \middle| C_c = 1\right] - E\left[\nabla y_{i,c} \middle| C_c = 0\right] \\ &= (\beta + \tau) - (\beta)\end{aligned}$$

$$= \tau \tag{2}$$

As Imbens and Wooldridge (2008) show, this framework can be extended to multiple time periods and can include a full set of individual-specific covariates. As the dataset contains five cohorts of students, this will allow the inclusion of four time dummies which capture the year effects, assuming that they are country-invariant. Furthermore, the accounting for student heterogeneity (the term $X_{i,c,t}$ in equation 1) allows for a more precise calculation of the standard errors and therefore of the significance of the DiD estimates, and to account for any compositional changes in terms of the observed heterogeneity across the analysis period. (Imbens & Wooldridge, 2008)

A natural extension to this DiD model is the difference in the difference-in-differences (DDD) model, as shown in equation 3.

$$y_{i,j,c,t} = \alpha + \beta_1 T_t + \beta_2 C_c + \beta_3 J_j + \beta_4 X_{i,c,j,t} + \beta_5 (C_c * T_t) + \beta_6 \left(C_c * J_j\right) + \beta_7 \left(T_t * J_j\right) + \tau \left(C_c * J_j * T_t\right) + \varepsilon_{i,j,c,t} \tag{3}$$

Here, the treatment effect is disaggregated for treatment and control groups within groups, for example comparing students who study Science, Technology, Engineering and Maths (STEM) to those who do not (non-STEM). This is performed by creating a sub-group binary indicator, J_j (where J is equal to 1 if the subject studied by the student is classified as STEM and 0 otherwise), and including in the regression all pairwise interactions between the time, group and sub-group indicators (whose coefficients are β_5, β_6 and β_7), and the triple interaction between time, group and sub-group (τ). This triple interaction is the DDD estimator. If, within a treatment group, there exists a subgroup that are more likely to be affected by a policy than others, then the DDD is calculated to show the particular treatment effect of the policy on the sub-group. In this analysis, sub-groups of low income students will be identified by subject type (STEM and graduate earnings potential), and by whether they chose to study at a local university or not.[43]

1.4.2 Difference-in-Differences with a Binary Outcome Variable

In the context of the funding reforms of 2006, to obtain a true causal estimate it is necessary to observe students in 2006 at English HEIs with and without the higher fees, and calculate the difference in the differences over the period window. However, since the HESA data only allows the observation of a student at university conditional on them having decided to attend, the question of interest is whether the funding reform affected the probability of observing a low income student at university. Here, the dependent variable is whether a student is classified as being from a low income background, which either takes the value 1 if the student has the relevant SEC code, and 0 otherwise. This however means that the DiD analysis is performed on a binary outcome variable.

Whilst a nonlinear approach could be used in this instance, nonlinear DiD regressions are challenging due to the interpretation and consistency problems of interactions in nonlinear models. This is a recurring issue in the policy evaluation literature, but was first highlighted by Ai and Norton (2003). The coefficient on the interaction term can be of incorrect significance, magnitude or even sign, and the recommendation is to calculate the cross difference to obtain the true estimate of the treatment effect. Furthermore, as Karaca-Mandic et al. (2012) show, the estimate of the ATET of the policy would depend on where in the distribution the treatment effect is calculated, and even without an interaction term there can be an erroneously significant treatment effect for a nonlinear DiD estimator.

Unlike Dietrich and Gerner (2012) who estimate the DiD on the probability of school leavers to enrol at a university using a nonlinear probit, this paper's approach is to use the Linear Probability Model (LPM) to estimate the DiD, as the interaction

[43] See section 1.4.3 for full details of the DDD categories.

between the group and time variable correctly identifies the DiD estimate. However, since the LPM can return probabilities that are outside the [0,1] bound, and the error term is inherently heteroscedastic, the LPM estimates are supplemented with a nonlinear estimation of all DiD and DDD specifications using probit. Despite the concerns raised by Ai and Norton (2003), the coefficient on the interaction term in a nonlinear DiD (and DDD) model returns the correct sign and significance (Puhani, 2008). Hence, as a robustness check, the sign and significance of the nonlinear estimates is found to be consistent with the LPM estimates.[44]

1.4.3 Distance & Subject Groupings

Instead of assuming that the treatment effect of the 2006 reform was constant across all low income background students, the impact of the increased fees is investigated by subject studied[45], and by distance between a student's home and their university attended. This allows the policy effect to vary across groups of subjects and students with potentially different study motivations. It is reasonable to expect that the funding reforms influenced the study behaviour of low income students studying subjects with high graduate wage premia and/or higher entrance requirements, as opposed to subjects with lower graduate wage premia and/or lower entrance requirements. Furthermore, distinguishing between students that do and do not attend a local university facilitates an evaluation of whether the policy changes affected low income students who attended a local university in a different way than those who attended a university far from their home at the time of application. The latter of these questions allows a direct test of whether the lower-low income students (who were the most likely to receive the full non-repayable maintenance grant) experienced greater geographic mobility as a result of the inclusion of the increased up-front financial support, and consequently, if this increased geographic mobility was heterogeneous across subject types.

[44] See Appendix A, Table A7.
[45] See Appendix A, Table A6 for subject groupings.

Every possible degree subject is assigned to one of 19 broad subject groupings (letters A-X) called the Joint Academic Coding System (JACS), and since each student observation has a valid JACS code, it is possible to use these subject groupings to allow the treatment effect of the policy to differ by classifications. One popular class of subject grouping is STEM and is defined as JACS codes A-K[46], whilst non-STEM subjects are defined as K-X.[47] However, in using JACS codes to classify STEM and non-STEM subjects, it leads to the inclusion of subjects that may not fit the definition of STEM (e.g. sports science with code C600, psychology with code C800), hence the classification may not be wholly accurate. Nevertheless, since STEM graduates typically have a higher than average graduate wage premium, and are less likely to be unemployed (Universities UK, 2010), students who are motivated by graduate job prospects may choose STEM subjects rather than non-STEM at the margin.[48]

Since degrees in different subjects offer differing individual rates of return (IRR), a more direct measure of grouping subjects by graduate earnings prospects is the classification constructed in this paper of High Future Income (HFI) and non-HFI subjects. A subject is classified as HFI if the IRR for both males and females is above the mean[49] IRR for undergraduate degrees, as calculated by BIS (2011),[50] and non-HFI subjects are those whose IRRs are below the average IRR for both males and females. This classification is also done using the JACS codes[51], hence the same issue may arise in classifying a broad group of subjects as all providing higher or lower than average IRRs when it is reasonable to expect at least some

[46] See http://www.publications.parliament.uk/pa/ld201213/ldselect/ldsctech/37/3705.htm
[47] There are two common definitions of STEM: broad and narrow. Unless otherwise specified, STEM refers to the broad definition (i.e. including Medicine, Subjects allied to Medicine, Architecture, and Agricultural Science).
[48] Universities UK (2010) also note however, that variations of graduate job prospects and earnings differ within the broad STEM grouping – further highlighting the imperfect nature of this classification.
[49] The graduate wage premium is universally found to vary by gender, for example Education carries higher lifetime earnings premium only for women, whereas the opposite is true for Social Studies.
[50] BIS (2011), Figure 13: ‘Individual rates of return to undergraduate degrees’.
[51] See Appendix A, Table A6 for a full list of JACS codes by HFI and non-HFI status.

variation within the 19 alphabetical groups.[52] To reduce the effect of incorrectly classifying subjects as either HFI or non-HFI, and to further provide contrast between the two groups of subjects, JACS codes whose IRRs were not above or below the average IRR for both males and females were omitted from the classification. Nevertheless, the same rationale applies for HFI subjects as STEM subjects – students who put emphasis on graduate job prospects may choose a HFI subject rather than a non-HFI subject at the margin.

The treatment effect is also allowed to vary across types of students according to the distance from their parental home at the time of applying to university and their eventual HEI attended. This is to identify those students who have a large degree of geographic mobility in pursuing their university education, and those students who choose to remain close to home whilst attending university. Both Kelchtermans and Verboven (2010), and Gibbons and Vignoles (2012) find that although distance to a university does not affect the participation decision itself, it does affect the choice of subject and the choice of university. Denzler and Wolter (2011) also find that this is only true for lower income students, since students from the higher income families are not constrained by the costs of moving to a university far away in the same way as students from lower socioeconomic groups. Furthermore, the fear of debt or an increase in the perceived costs of attending university is likely to negatively affect the willingness to attend a university far from the parental home for low income students (Callender and Jackson, 2008). For the purpose of the analysis, and taking the relatively smaller geographic size of Wales into account, a student is classified as having remained in the local area if the difference between their parental home and university postcode is less than 20km in Euclidean space, since it is reasonable to expect that such a distance is easily commutable.[53] Furthermore, a student is defined as having attended a

[52] For example, in the definition that will follow, Economics (code L100) is not classed as an HFI subject for the purposes of this analysis.
[53] The 2011 UK Census identifies the average commuter distance – albeit for commuters to full and part time work – as being 15.0km in England and Wales, and 13.4km in the previous census (2001). The slightly greater distance choice of 20km as being the threshold for a local student accounts for the possibility of students not being required to travel every day of the week, nor at

university far away if their point-to-point distance between parental and university postcode is more than 80km.[54] These classifications, in conjunction with a student's accommodation information, allows for even more detailed distinction between students who are living at home, and those who are living in university halls or in private sector accommodation. Table 1.5a below shows the summary statistics for the treatment group (English HEIs) and the control group (Welsh HEIs), both before and after the introduction of the policy. Table 1.5b provides a t-test comparison of the means of the treatment and control groups in the pre-policy period. As can be seen from column 3 of Table 1.5b, students at Welsh HEIs do differ from English HEIs in terms of statistical significance before the funding change was enacted. This does not invalidate the approach of using Wales as a counterfactual however, since the underlying assumption of the difference-in-differences strategy used in this analysis is not that the control group *is* identical to the treatment group, but that the evolution over time at Welsh HEIs is provides the counterfactual outcome for English HEIs. Hence, although the two groups are statistically different, this difference is assumed to be the same in the treatment period, in the absence of treatment. As discussed in the proceeding section, the more specific assumption of time-varying but country-invariant trends is outlined.

peak times, which reduces the financial burden of travel costs. As seen from Table 1.5a and Appendix A, Figure A1, around one fifth of the student population in England is classified as attending a local student, and slightly less in Wales (due to Wales' smaller size compared to England, and relatively more rural demography). For full details of the commuter distance results from the 2011 UK Census, see: http://webarchive.nationalarchives.gov.uk/20160107181447/http://www.ons.gov.uk/ons/dcp171776_357812.pdf

[54] This is the same classification as used by Frenette (2004). Alternative specifications that were considered included a distance of 50km, as used by Mitze et al. (2013), however this is in the context of distance to a fee border given regional variation in study fees; and the approach by Denzler and Wolter (2011) to use commuting time instead of distance, however this was impractical given the geographic extent of the analysis.

Table 1.5a: Summary Statistics of Treatment and Control Groups

	(1) Total	(2) Treat (England)	(3) Control (Wales)	(4) Treat (England) [Pre-Policy]	(5) Control (Wales) [Pre-Policy]	(6) Treat (England) [Post-Policy]	(7) Control (Wales) [Post-Policy]
Low Income =1	0.29	0.29	0.30	0.28	0.29	0.30	0.31
(SEC Code 4-8)	(0.453)	(0.453)	(0.459)	(0.447)	(0.454)	(0.459)	(0.464)
Upper-Low Income =1	0.12	0.12	0.13	0.11	0.13	0.12	0.13
(SEC Code 4-5)	(0.322)	(0.320)	(0.338)	(0.318)	(0.336)	(0.323)	(0.341)
Lower-Low Income	0.17	0.17	0.17	0.16	0.16	0.18	0.18
(SEC Code 6-8)	(0.377)	(0.377)	(0.375)	(0.369)	(0.367)	(0.387)	(0.385)
England	0.93	-	-	-	-	-	-
	(0.249)	-	-	-	-	-	-
Wales	0.07	-	-	-	-	-	-
	(0.249)	-	-	-	-	-	-
Age	19.35	19.35	19.34	19.34	19.34	19.37	19.34
	(3.696)	(3.693)	(3.748)	(3.695)	(3.791)	(3.690)	(3.696)
Educated Parent	0.07	0.08	0.00	0.00	0.00	0.17	0.00
	(0.260)	(0.268)	(0)	(0.0545)	(0)	(0.377)	(0)
White	0.83	0.82	0.95	0.83	0.95	0.81	0.95
	(0.376)	(0.383)	(0.222)	(0.378)	(0.223)	(0.389)	(0.222)
Female	0.57	0.57	0.58	0.57	0.59	0.57	0.58
	(0.495)	(0.496)	(0.493)	(0.496)	(0.492)	(0.495)	(0.494)
British	0.95	0.95	0.95	0.95	0.92	0.96	0.99
	(0.213)	(0.212)	(0.216)	(0.221)	(0.266)	(0.201)	(0.121)
Disabled	0.09	0.09	0.09	0.08	0.09	0.09	0.09
	(0.281)	(0.281)	(0.288)	(0.278)	(0.287)	(0.285)	(0.289)
Local	0.21	0.22	0.18	0.21	0.17	0.22	0.20
	(0.410)	(0.411)	(0.385)	(0.407)	(0.374)	(0.417)	(0.398)
Live at Home	0.19	0.19	0.16	0.18	0.15	0.21	0.17
	(0.394)	(0.396)	(0.364)	(0.386)	(0.353)	(0.408)	(0.376)
STEM	0.22	0.22	0.20	0.24	0.22	0.19	0.18
	(0.412)	(0.413)	(0.401)	(0.428)	(0.412)	(0.392)	(0.386)
HFI	0.31	0.31	0.29	0.33	0.30	0.29	0.28
	(0.464)	(0.464)	(0.454)	(0.470)	(0.459)	(0.456)	(0.447)
Russell Group	0.36	0.36	0.29	0.39	0.31	0.32	0.27
	(0.479)	(0.481)	(0.455)	(0.488)	(0.464)	(0.468)	(0.441)

N	691470	645535	45935	358185	25240	287355	20695

Notes:

Educated Parent is a dummy variable that takes the value 1 if the student is known to have a parent with university level education, and 0 otherwise; White, Female, British and Disabled are dummy variables whose values are equal to 1 if the student's ethnicity, gender, nationality and self-reported disability status is that of the relevant category, and 0 otherwise. Local is a dummy variable that takes the value 1 if the Euclidean distance between the student's domiciled postcode at the time of applying to higher education and the student's eventual university that they attend is less than 20km. Live at Home is a dummy variable that takes the value 1 if the student's term time accommodation is denoted as living at a parental or family home, and 0 otherwise. STEM and HFI are both dummy variables that are equal to 1 (and 0 otherwise) if the student's subject that they study is classified as either STEM or HFI according to the classification discussed in this section (1.4.3) and in Appendix A, Table A6. Russell Group is a dummy variable that takes the value 1 if the university the student attends is classified as an elite university, as defined by Appendix A, Table A1. Raw numbers have been rounded to the nearest multiple of 5 to comply with HESA's Standard Rounding Methodology.

Table 1.5b: Comparison of Treatment and Control Groups Means

	(1) **Treat** (England) [Pre-Policy]	(2) **Control** (Wales) [Pre-Policy]	(3) **Difference** (t-test of means)
Low Income =1	0.28	0.29	-0.0138***
(SEC Code 4-8)	(0.447)	(0.454)	
Upper-Low Income =1	0.11	0.13	-0.0158***
(SEC Code 4-5)	(0.318)	(0.336)	
Lower-Low Income	0.16	0.16	0.00202
(SEC Code 6-8)	(0.369)	(0.367)	
Age	19.34	19.34	0.00376
	(3.695)	(3.791)	
Educated Parent	0.00	0.00	0.00298***
	(0.0545)	(0)	
White	0.83	0.95	-0.121***
	(0.378)	(0.223)	
Female	0.57	0.59	-0.0208***
	(0.496)	(0.492)	
British	0.95	0.92	0.0254***

	(0.221)	(0.266)	
Disabled	0.08	0.09	-0.00631***
	(0.278)	(0.287)	
Local	0.21	0.17	0.0411***
	(0.407)	(0.374)	
Live at Home	0.18	0.15	0.0354***
	(0.386)	(0.353)	
STEM	0.24	0.22	0.0252***
	(0.428)	(0.412)	
HFI	0.33	0.30	0.0540***
	(0.470)	(0.459)	
Russell Group	0.39	0.31	0.0771***
	(0.488)	(0.464)	
N	358185	25240	

Notes:

Columns (1) and (2) in Table 1.5b above are the same as columns (4) and (5) in Table 1.5a. Column (3) gives the t-tested difference in the means between the treatment and control groups before the funding policy change with the corresponding statistical difference in the two group means being denoted as stars. * p<0.05, ** p<0.01, *** p<0.001

1.4.4 Empirical Model

Given the data and the parameters of the natural experiment, the DiD model is estimated using OLS of the following:

$$Y_{i,c,u,t} = \lambda_c D_c + \sum_{t=2004}^{T} \psi_t D_t + \sum_{k=1}^{K} \beta_k X_{i,c,u,t} + \sum_{u=1}^{U} a_u A_u + \delta_1 (D_p * D_c) + \varepsilon_{i,c,u,t} \tag{4}$$

Where $Y_{i,c,u,t}$ denotes the low income background status of student *i* in country *c* at university *u* in time period *t*, D_c is a country dummy that takes the value of 1 for students in England, and D_t are a full set of year dummies.[55] The DiD estimate is the coefficient δ_1 on the interaction between the country indicator (D_c) and an additional period indicator (D_p) that takes the value 1 if the policy is active in the wave year, and 0 otherwise.[56] Furthermore, a number (K) of individual controls ($X_{i,c,u,t}$) that measure student characteristics are added to take into account of any heterogeneity in the trends of the observable student characteristics between students at English and Welsh HEIs. These controls are gender, disability status, ethnicity, whether a student has a parent with university education, nationality, age, mode of study, and whether the student is from the local area (as previously defined). Also included are university fixed effects (A_u) to capture any heterogeneity between English and Welsh universities over time.[57] Furthermore, since the error term is likely to violate the independent and identically distributed (iid) assumption (as errors are likely to be correlated within universities, and as the LPM inherently creates heteroscedasticity), the errors ($\varepsilon_{i,c,u,t}$) are clustered at the university level.

Since the coefficient on the interaction term δ_1 is the coefficient of interest, as it is the difference-in-differences estimator which estimates the ATET, it is these coefficients that are reported in the following tables. Finally, it should be noted that the common trends assumption made in this specification is not that the overall trends are common across England and Wales, but the individual year effects are. Thus the trend is allowed to vary by year, but across the two countries they are assumed common.

[55] For all estimations, the base year is 2003.
[56] This allows the difference-in-differences to be computed as an average of the post-policy periods, or separately for the 2006/07 and 2007/08 cohort. See the opening paragraph of section 1.5 for a full discussion.
[57] As can be seen from Appendix A, Table A7, the inclusion of department fixed effects instead of university fixed effects - which would further capture any heterogeneity within universities - does not have a significant impact on the estimates.

To perform the DDD specifications, the existing interaction term is expanded with a third interaction between a sub-group of the treatment group, as shown by the following equation:

$$Y_{i,c,u,j,t} = \lambda_c D_c + \sum_{t=2004}^{T} \psi_t D_t + \sum_{k=1}^{K} \beta_k X_{i,c,u,t} + \sum_{u=1}^{U} a_u A_u + \theta_j J_j + \delta_1 \left(D_p * D_c\right) + \delta_2 \left(D_p * J_j\right) + \delta_3 \left(J_j * D_c\right) + \delta_4 \left(D_p * D_c * J_j\right) + \varepsilon_{i,c,u,j,t} \tag{5}$$

The original DiD equation (4) is extended to a DDD estimation with the addition of a sub-group indicator, J_j (which is equal to 1 if the student observation is in the sub-group of interest[58] and 0 otherwise), all pairwise interactions of the sub-group, time period and country indicators (δ_1, δ_2 and δ_3), and the DDD estimator δ_4. In this analysis, distinctions are made between low income students who study STEM and HFI subjects (Table 1.7), and low income students who remained in the local area (Table 1.8). This approach is then combined (by estimating the Local DDD separately by subject groups) to examine whether there is a significant effect to be found when considering the subject choice of those students who can be classified as being local, those who are living at home whilst attending university, and those who are attending a university far from their registered home address at the time of application (Tables 1.8-1.10). The coefficients reported in the DDD tables are therefore the coefficient δ_4 in equation 5 from each DDD regression.

1.5 Results

For each specification of the DiD and DDD analysis, the estimation is performed across all universities and across those universities with the Russell Group universities excluded, to investigate whether the policy impacted students studying

[58] Either subject sub-groups (STEM==1, HFI==1) or distance sub-groups (Local==1, Live at Home==1, Live Far Away==1).

at non-elite institutions, who tend to have a higher proportion of lower socioeconomic status students.[59] Furthermore, each DiD and DDD estimation is performed across three different treatment periods: the 2006 period (where the country dummy is interacted with the 2006 year dummy), the 2007 period (where the 2007 year dummy is used in the interaction) and the average treatment period (where the country dummy is interacted with a post policy dummy). All the coefficients presented in Tables 1.6-1.10 (and their corresponding clustered standard errors) are the DiD estimates from a separate DiD regression.

In terms of the 2006 treatment effect, although the students who started in September 2006 were the first cohort of students subject to the new funding reforms, many would have already committed themselves to the decision to attend university by their subject choice at school. As students decide which subjects to study two years before they attend university[60], the subjects students choose to study prior to HE entry can define which subjects a student is eligible to study at university through subject-specific entrance requirements[61] (e.g. pre-university mathematics qualifications are normally required to study an undergraduate mathematics degree). Hence in 2006, the effects can be considered a partial adjustment to the policy in this regard. The only choice these students may have had was which university to study at, given their choice of degree subject is limited to their subjects chosen at school or college. In 2007 however, the treatment effect can be thought of in terms of full adjustment, as students who started in September 2007 had a chance to better delay or change any decision to attend university, including the decision to change which subjects to study pre-university. Finally, each specification of the DiD and DDD is performed across three classifications of

[59] See, for example: https://www.suttontrust.com/newsarchive/access-highly-selective-universities-stalls/

[60] See Chapter 2, Table 2.1 for a complete timeline of a typical new entrant to HE with regards to subject choice; also see the Introduction to this thesis for discussion of the transition from further education to higher education in general.

[61] In the UK context, students in England and Wales typically study 3 or 4 A-Level subjects to gain entry into undergraduate study (whereas in Scotland, students study Highers or Advanced Highers). See the Introduction to this thesis for a detailed explanation.

low income students, in accordance with the identification strategy: low income students (SEC code 4-8), upper-low income students (SEC code 4-5), and lower-low income students (SEC code 6-8).

Table 1.6 shows the results from the OLS estimation of equation 4, where each coefficient is the difference in differences in the probability of observing a low income student. Panel A uses the broader definition of low income students (SEC code 4-8), whilst Panels B and C show the DiD estimates in the probability of observing an upper-low and lower-low income student respectively. In each panel, the DiD is estimated separately for the average of the treatment period, and for the partial (the 2006/07 cohort) and full (the 2007/08 cohort) adjustment effects[62]. Columns 1 and 3 uses university fixed effects (dummies for each university) to capture unobserved heterogeneity and hence account for omitted bias, whilst Columns 2 and 4 use department fixed effects (dummies for each department at each university). Comparing the two types of fixed effects, there is no significant difference across the type of fixed effects used; hence for computational ease (and as shown in equation 4, university fixed effects are used henceforth.

Panel A shows that the increased tuition fees and funding reforms actually increased the probability of observing a low income student by 1.66 percentage points, conditional on them having applied to university. This represents a 5.98% increase in the probability of observing a low income student, relative to the pre-reform mean of 27.78%. However, as seen from Panel B, this increase in probability is significant only for the upper-low income category of students (SEC code 4-5; the most likely to be adversely affected by the increased fees), with an increase in probability of 1.72 percentage points which represents an increase of 14.96% relative to the pre-reform mean of 11.50%. The implication is that when estimating the DiD averaged across all subjects, and not distinguishing between whether a student is local to their university, the policy reforms appear to have had a positive

[62] For ease, the estimates for students who began their studies in the 2006/07 academic year is simply named ‘2006’, and the 2007/08 cohort is named ‘2007’ Tables 1.6-1.10.

impact on the probability of observing an upper-low income group student once full adjustment took place in September 2007. This is somewhat counter-intuitive, as given the structure of the policy introduced, the upper-low income students should have been the most adversely affected by the increased fees, hence it would have been reasonable to expect the DiD estimate to be negative, especially for non-elite universities. Furthermore, the policy reforms appear not to have affected the conditional probability of observing a lower-low income student, which supports the existing research as outlined in section 1.2.2.

Table 1.6: DiD Estimate of the 2006 Policy Reform – Baseline

				All Universities		Non-Elite Universities	
				(1)	(2)	(3)	(4)
Panel A							
SEC Code	4-8						
		Treatment Period	Average	0.0117	0.0129	0.00642	0.00626
				(0.00939)	(0.00922)	(0.00647)	(0.00632)
			2006	-0.00251	-0.00235	-0.00371	-0.00442
				(0.00554)	(0.00560)	(0.00497)	(0.00480)
			2007	0.0166**	0.0179**	0.0110*	0.0113*
				(0.00837)	(0.00797)	(0.00627)	(0.00596)
Panel B							
SEC Code	4-5	Treatment Period	Average	0.0103**	0.00986*	0.00859	0.00725
				(0.00484)	(0.00506)	(0.00648)	(0.00666)
			2006	-0.00539	-0.00581	-0.00757	-0.00864
				(0.00573)	(0.00593)	(0.00709)	(0.00715)
			2007	0.0172**	0.0171**	0.0168*	0.0160
				(0.00715)	(0.00717)	(0.0100)	(0.0101)
Panel C							
SEC Code	6-8	Treatment Period	Average	0.00141	0.00301	-0.00217	-0.000990
				(0.00987)	(0.00951)	(0.0104)	(0.0103)

	2006	0.00288	0.00346	0.00386	0.00422
		(0.00334)	(0.00348)	(0.00413)	(0.00447)
	2007	-0.000618	0.000865	-0.00584	-0.00470
		(0.0125)	(0.0122)	(0.0147)	(0.0145)
Fixed Effects		University	Department	University	Department
Controls		yes	yes	yes	yes
Obs (SEC Code 4-8)		691,475	691,475	444,395	444,395
Obs (SEC Code 4-5)		691,475	691,475	444,395	444,395
Obs (SEC Code 6-8)		691,475	691,475	444,395	444,395

This table reports the results from the DiD estimation as outlined by equation (4), estimated across all universities (columns 1-2) and then across non-elite universities (columns 3-4). Elite status is denoted by membership of the Russell Group of universities, as seen in Appendix A, Table A1. The coefficients reported are the coefficients on the interaction between the treatment group indicator (English HEIs) and the treatment period indicator, and show the change in the probability of observing a low income student studying as a result of the 2006 funding reforms. To allow for the estimation of the effect of the policy averaged across the post-policy period and to capture the partial and full adjustment effects to the policy change, the DiD is estimated separately for the average of the post-policy cohorts, and for the 2006/07 and 2007/08 cohorts. There are therefore three rows of DiD estimates for each column, where each coefficient relates to the specific treatment period. There are three panels (A, B and C) to allow an estimation of the DiD across the broad low income background category (Panel A, SEC Code 4-8), and disaggregated across the upper-low income background category of students (Panel B, SEC Code 4-5) and the lower-low income background category of students (Panel C, SEC Code 6-8). University fixed effects are included for columns 1 and 3, whilst department fixed effects (20 departments per university, as created by the 20 broad JACS codes) are used in columns 2 and 4. Controls are included for observable student heterogeneity, which are: gender, ethnicity, age, disability status, and nationality. Raw numbers have been rounded to the nearest multiple of 5 to comply with HESA’s Standard Rounding Methodology. Standard errors are shown in parentheses and are clustered at the HEI level. *** p<0.01, ** p<0.05, * p<0.1

Table 1.7 attempts to investigate this finding by disaggregating the treatment effect into subject groupings using the DDD approach by estimating equation number 5. Using the triple difference approach by subject grouping, Panel B shows that the initial effect identified in Table 1.6 is driven by upper-low income students studying non-High Future Income subjects. The implication is therefore that not only are students more likely to be observed at university from the upper-low income category, the subjects that this effect is found for may not offer the best investment in terms of graduate employability. Column 4 shows that there is a decrease of 1.77 percentage points on the conditional probability of observing an upper-low income student studying High Future Income subjects compared to non-HFI subjects, once full adjustment has taken place. Furthermore, this effect is larger in magnitude and statistical significance at non-Russell Group universities. Once again, the counter-intuitive nature of the disaggregation of the DiD estimate by subject is that whilst students from the upper-low income category are more likely to be studying as a result of the policy reforms, they are less likely to be studying subjects that offer the highest graduate earnings premia. The story therefore cannot be that students from the upper-low income category are more likely to attend university, and as a consequence of the increased costs are choosing to study a subject with the highest rewards upon graduation in terms of earnings.

Table 1.7: DiD Estimate of the 2006 Policy Reform – DDD by Subject

				All Universities		Non-Elite Universities	
				STEM	HFI	STEM	HFI
				(1)	(2)	(3)	(4)
Panel A							
SEC Code	4-8						
		Treatment Period	Average	-0.000528	-0.00660	-0.000813	-0.0143
				(0.00999)	(0.0131)	(0.0144)	(0.0184)
			2006	0.00900	0.0184	-0.0220	0.0147
				(0.0208)	(0.0112)	(0.0148)	(0.0168)
			2007	-0.00841	-0.0238**	0.0163	-0.0298**
				(0.0168)	(0.00964)	(0.0145)	(0.0120)
Panel B							
SEC Code	4-5	Treatment Period	Average	-0.00274	-0.000767	-0.0108	-0.00414
				(0.00646)	(0.00623)	(0.00695)	(0.0103)
			2006	0.00348	0.0134*	-0.0161	0.0146
				(0.0143)	(0.00783)	(0.0118)	(0.0107)
			2007	-0.00649	-0.0124*	-0.000520	-0.0177**
				(0.00815)	(0.00637)	(0.00852)	(0.00784)
Panel C							
SEC Code	6-8	Treatment Period	Average	0.00221	-0.00583	0.00996	-0.0101

		(0.0112)	(0.00855)	(0.0149)	(0.0120)
	2006	0.00552	0.00499	-0.00588	0.000121
		(0.00961)	(0.00941)	(0.0107)	(0.0133)
	2007	-0.00191	-0.0113*	0.0168	-0.0122
		(0.0176)	(0.00598)	(0.0204)	(0.00791)
Fixed Effects		University	University	University	University
Controls		yes	yes	yes	yes
Obs (SEC Code 4-8)		691,475	691,475	444,395	444,395
Obs (SEC Code 4-5)		691,475	691,475	444,395	444,395
Obs (SEC Code 6-8)		691,475	691,475	444,395	444,395

This table reports the results from the DDD estimation as outlined by equation (5), estimated across all universities (columns 1-2) and then across non-elite universities (columns 3-4). Elite status is denoted by membership of the Russell Group of universities, as seen in Appendix A, Table A1. The coefficients reported are the coefficients on the triple interaction between the treatment group indicator (English HEIs), the subject group indicator and the treatment period indicator, and show the change in the probability of observing a low income student studying either a STEM or HFI subject compared to non-STEM or non-HFI, as a result of the 2006 funding reforms. Columns 1 and 3 therefore show the impact of the funding reforms on low income students studying STEM subjects compared to non-STEM subjects, and columns 2 and 4 show the impact of the funding reforms on low income students studying HFI subjects compared to non-HFI subjects. All relevant cross-products were included between the treatment group indicator, treatment period indicator and the subject group indicator, but are not shown here in the interests of concision. To allow for the estimation of the effect of the policy averaged across the post-policy period and to capture the partial and full adjustment effects to the policy change, the DiD is estimated separately for the average of the post-policy cohorts, and for the 2006/07 and 2007/08 cohorts. There are therefore three rows of DiD estimates for each column, where each coefficient relates to the specific treatment period. There are three panels (A, B and C) to allow an estimation of the DiD across the broad low income background category (Panel A, SEC Code 4-8), and disaggregated across the upper-low income background category of students (Panel B, SEC Code 4-5) and the lower-low income background category of students (Panel C, SEC Code 6-8). University fixed effects are included for all specifications, as are controls for observable student heterogeneity. Namely, these individual controls are: gender, ethnicity, age, disability status, and nationality. Raw numbers have been rounded to the nearest multiple of 5 to comply with HESA's Standard Rounding Methodology. Standard errors are shown in parentheses and are clustered at the HEI level. *** p<0.01, ** p<0.05, * p<0.1

To disaggregate the policy effect even further, in Table 1.8 a further distinction is made for students who stay in the local area (within 20km), and the DDD is estimated as an average across all subjects (columns 1 and 6), and then across STEM/non-STEM subjects (columns 2-3, 7-8) and HFI subjects (columns 4-5, 9-10). Whilst it may have been counter-intuitive so far for the policy to have a positive impact on students most likely to be adversely affected, the distinction between local and non-local students gives a more logical explanation. Columns 1 and 6 show that the increase in probability of observing an upper-low income student can be attributed to those upper-low income students who remained in the local area, although the coefficients are marginally insignificant at the 5% level. However, as columns 5 and 10 highlight, when this DDD estimation for local students is allowed to be heterogeneous across subject types, the increase in probability of observing an upper-low income student appears only to be true for non-HFI subjects. Furthermore, as found in Table 1.7, these effects are slightly larger in magnitude for non-Russell Group HE institutions. Thus, by not taking into account heterogeneous treatment effects between students who chose to stay in the local area and those who do not (Tables 1.6-1.7), the results are initially somewhat counter-intuitive. However, given that by staying in the local area, the costs (excluding tuition fees) of attending university can be reduced through parental support and living at home, and given that tuition fees are now repayable through earnings above a certain level upon graduation, the decision to attend university if staying in the local area has less risk involved than before the policy (where tuition fees were paid up front with no loans available).

The disaggregation of treatment effects across subject types also shows which types of subjects this increase in local upper-low income students originates. Whilst the finding that the increase in upper-low income students studying non-HFI subjects appeared surprising at first, by looking at the local DDD results by subject type the finding in Table 1.7 is the result of an increase in local upper-low income students studying non-HFI subjects. Given the lack of risk involved as a result of the 2006 policy changes, students from the upper-low income category who wanted to study non-HFI subjects now find it easier and less risky to do so.

Table 1.8: DiD Estimate of the 2006 Policy Reform – Local DDD by Subject

					All Universities					Non-Elite Universities				
					Local					Local				
					All	STEM	Non-STEM	HFI	Non-HFI	All	STEM	Non-STEM	HFI	Non-HFI
					(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Panel A														
	SEC Code	4-8	Treat	Average	-0.00537	-0.00494	-0.00522	0.000129	-0.00793	-0.0103	-0.0140	-0.00891	-0.00951	-0.00778
			Period		(0.0132)	(0.0177)	(0.0150)	(0.0195)	(0.0141)	(0.0184)	(0.0258)	(0.0197)	(0.0262)	(0.0180)
				2006	-0.00694	-0.0209	-0.00212	-0.00452	-0.00975	-0.0125	-0.0426***	-0.00601	-0.0126	-0.0133
					(0.0117)	(0.0186)	(0.0159)	(0.0180)	(0.0114)	(0.0164)	(0.0157)	(0.0195)	(0.0270)	(0.0144)
				2007	-0.00191	0.0104	-0.00571	0.00342	-0.00285	-0.00289	0.0147	-0.00639	-0.000685	0.000357
					(0.0103)	(0.0175)	(0.0106)	(0.0188)	(0.0132)	(0.0133)	(0.0227)	(0.0128)	(0.0248)	(0.0156)
Panel B														
	SEC Code	4-5	Treat	Average	0.00507	0.00791	0.00389	-0.00105	0.00845	0.0101	0.0196	0.00791	-0.000981	0.0158
			Period		(0.0117)	(0.0163)	(0.0127)	(0.00924)	(0.0141)	(0.0137)	(0.0203)	(0.0146)	(0.0113)	(0.0153)
				2006	-0.0170	-0.0164	-0.0175	-0.0140	-0.0203	-0.0129	-0.0100	-0.0142	-0.0164	-0.0131
					(0.0109)	(0.0133)	(0.0130)	(0.0132)	(0.0135)	(0.0140)	(0.0160)	(0.0160)	(0.0181)	(0.0158)
				2007	0.0185*	0.0211	0.0177	0.00903	0.0258**	0.0212*	0.0285	0.0198	0.0114	0.0287**
					(0.0101)	(0.0173)	(0.0114)	(0.00925)	(0.0130)	(0.0115)	(0.0227)	(0.0123)	(0.0112)	(0.0144)
Panel C														
	SEC Code	6-8	Treat	Average	-0.0104	-0.0128	-0.00911	0.00118	-0.0164	-0.0203	-0.0336*	-0.0168	-0.00853	-0.0236

Period		(0.0186)	(0.0156)	(0.0212)	(0.0185)	(0.0229)	(0.0235)	(0.0197)	(0.0256)	(0.0258)	(0.0271)
	2006	0.0100	-0.00445	0.0153	0.00951	0.0106	0.000349	-0.0325	0.00817	0.00386	-0.000174
		(0.0166)	(0.0179)	(0.0191)	(0.0203)	(0.0193)	(0.0213)	(0.0207)	(0.0233)	(0.0297)	(0.0222)
	2007	-0.0204*	-0.0107	-0.0234*	-0.00561	-0.0286	-0.0241	-0.0138	-0.0262	-0.0120	-0.0283
		(0.0120)	(0.0126)	(0.0139)	(0.0143)	(0.0183)	(0.0146)	(0.0196)	(0.0164)	(0.0177)	(0.0216)
Fixed Effects		University	University	University	University	University	University	University	University	University	University
Controls		yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
Obs (SEC Code 4-8)		691,475	150,250	541,220	216,505	474,970	444,395	67,640	376,760	130,240	314,155
Obs (SEC Code 4-5)		691,475	150,250	541,220	216,505	474,970	444,395	67,640	376,760	130,240	314,155
Obs (SEC Code 6-8)		691,475	150,250	541,220	216,505	474,970	444,395	67,640	376,760	130,240	314,155

This table reports the results from the DDD estimation as outlined by equation (5), estimated across all universities (columns 1-5) and then across non-elite universities (columns 6-10). Elite status is denoted by membership of the Russell Group of universities, as seen in Appendix A, Table A1. The coefficients reported are the coefficients on the triple interaction between the treatment group indicator (English HEIs), the indicator of being a local student and the treatment period indicator, and show the change in the probability of observing a low income student as a result of the 2006 funding reforms studying at a local university compared to a non-local university, where local is defined as being within 20km Euclidean distance from the student's domiciled postcode at the time of application to HE. Column 1 shows the DDD impact of the funding reforms for low income students studying at a local university across all subjects, compared to non-local universities across all subjects. Column 6 estimates the same as column 1, except the DDD estimation is only performed for non-elite universities. Columns 2 and 3 perform the same estimation as column 1, but for STEM and non-STEM subjects separately. Columns 7 and 8 also show the DDD for STEM and non-STEM subjects separately, but only at non-elite universities. The DDD estimates in these columns therefore show the impact of the funding reforms on the probability of observing a low income student who attends a local university compared to a non-local university who studies a STEM or a non-STEM subject. Columns 4 and 5, and columns 9 and 10 show the results for the same approach for HFI and non-HFI subjects at elite and non-elite universities respectively. All relevant cross-products were included between the treatment group indicator, treatment period indicator and the local indicator, but are not shown here in the interests of concision. To allow for the estimation of the effect of the policy averaged across the post-policy period and to capture the partial and full adjustment effects to the policy change, the DiD is estimated separately for the average of the post-policy cohorts, and for the 2006/07 and 2007/08 cohorts. There are therefore three rows of DiD estimates for each column, where each coefficient relates to the specific treatment period. There are three panels (A, B and C) to allow an estimation of the DiD across the broad low income background category (Panel A, SEC Code 4-8), and disaggregated across the upper-low income background category of students (Panel B, SEC Code 4-5) and the lower-low income background category of students (Panel C, SEC Code 6-8). University fixed effects are included for all specifications, as are controls for observable student heterogeneity. Namely, these individual controls are: gender, ethnicity, age, disability status, and nationality. Raw numbers have been rounded to the nearest multiple of 5 to comply with HESA's Standard Rounding Methodology. Standard errors are shown in parentheses and are clustered at the HEI level. *** p<0.01, ** p<0.05, * p<0.1

Table 1.9 shows that by disaggregating the local classification even further into those who live at home in the local area and those who live away from home but in the local area, similar results to Table 1.8 are found. Those who are living at home in the local area are the most likely to insulate themselves from the increase in fees (by effectively not requiring a maintenance loan to fund living costs, given they are likely to be less eligible for a means-tested maintenance grant), whilst also able to repay tuition fees on an income-contingent basis. Hence, such students should display the least cautious behaviour in terms of subject choice, as there is even less risk involved than a student simply living locally. Columns 1 and 5 confirm this hypothesis, with an increase in the probability of 2.85 percentage points across all subjects, and an increase of 3.61 percentage points for those students studying non-HFI subjects.

By looking specifically at students who chose to live at home, it is also possible to examine the effect of the policy on geographic mobility of the lower-low income students. Since the policy introduced both tuition fee loans and maintenance grants, the poorest students who previously only had the option of student loans now have greater access to universities further away from home, since maintenance grants effectively fund living costs. In column 1 and 6 of Table 1.9, there is a negative and significant effect of the policy reform on the conditional probability of observing a lower-low income student living at home whilst attending university across all subjects overall and across most subject categories.

Table 1.9: DiD Estimate of the 2006 Policy Reform – Live at Home DDD by Subject

				All Universities					Non-Elite Universities				
				Live at Home					Live at Home				
				All	STEM	Non-STEM	HFI	Non-HFI	All	STEM	Non-STEM	HFI	Non-HFI
				(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Panel A													
SEC Code	4-8	Treat Period	Average	-0.00595	0.00374	-0.00946	0.00789	-0.0129	-0.0110	-0.0277	-0.00756	-0.0102	-0.00839
				(0.0155)	(0.0356)	(0.0161)	(0.0245)	(0.0177)	(0.0191)	(0.0450)	(0.0186)	(0.0258)	(0.0185)
			2006	0.00411	0.00265	0.00413	0.0216	-0.00627	0.00570	0.00320	0.00630	0.0294	-0.00647
				(0.00941)	(0.0275)	(0.0132)	(0.0160)	(0.0105)	(0.0115)	(0.0369)	(0.0139)	(0.0195)	(0.0121)
			2007	-0.0111	0.00228	-0.0155	-0.00833	-0.0112	-0.0182	-0.0364	-0.0145	-0.0352	-0.00542
				(0.0150)	(0.0302)	(0.0134)	(0.0312)	(0.0164)	(0.0176)	(0.0385)	(0.0158)	(0.0314)	(0.0160)
Panel B													
SEC Code	4-5	Treat Period	Average	0.00625	0.0103	0.00432	-0.00349	0.0124	0.00663	0.0171	0.00359	-0.00896	0.0153
				(0.0140)	(0.0219)	(0.0157)	(0.0139)	(0.0160)	(0.0157)	(0.0290)	(0.0167)	(0.0127)	(0.0175)
			2006	-0.0268*	-0.0262	-0.0277*	-0.0292	-0.0255	-0.0235	-0.0248	-0.0240	-0.0372**	-0.0174
				(0.0139)	(0.0207)	(0.0162)	(0.0188)	(0.0159)	(0.0162)	(0.0248)	(0.0185)	(0.0181)	(0.0166)
			2007	0.0285**	0.0319	0.0272	0.0185	0.0361**	0.0263*	0.0387	0.0232	0.0193	0.0324*
				(0.0141)	(0.0265)	(0.0170)	(0.0121)	(0.0179)	(0.0149)	(0.0346)	(0.0164)	(0.0140)	(0.0183)
Panel C													
SEC Code	6-8	Treat	Average	-0.0122	-0.00657	-0.0138	0.0114	-0.0253	-0.0176	-0.0449	-0.0112	-0.00127	-0.0237

Period	(0.0184)	(0.0293)	(0.0196)	(0.0202)	(0.0225)	(0.0219)	(0.0342)	(0.0213)	(0.0243)	(0.0245)
2006	0.0309*	0.0288	0.0318	0.0507**	0.0192	0.0292	0.0280	0.0303	0.0665***	0.0109
	(0.0157)	(0.0204)	(0.0194)	(0.0247)	(0.0186)	(0.0183)	(0.0263)	(0.0217)	(0.0217)	(0.0205)
2007	-0.0396***	-0.0296	-0.0427***	-0.0268	-0.0472*	-0.0445***	-0.0750**	-0.0378***	-0.0545**	-0.0378
	(0.0134)	(0.0335)	(0.0144)	(0.0252)	(0.0239)	(0.0151)	(0.0331)	(0.0142)	(0.0209)	(0.0229)
Fixed Effects	University	University	University	University	University	University	University	University	University	University
Controls	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
Obs (SEC Code 4-8)	691,475	150,250	541,220	216,505	474,970	444,395	67,640	376,760	130,240	314,155
Obs (SEC Code 4-5)	691,475	150,250	541,220	216,505	474,970	444,395	67,640	376,760	130,240	314,155
Obs (SEC Code 6-8)	691,475	150,250	541,220	216,505	474,970	444,395	67,640	376,760	130,240	314,155

This table reports the results from the DDD estimation as outlined by equation (5), estimated across all universities (columns 1-5) and then across non-elite universities (columns 6-10). Elite status is denoted by membership of the Russell Group of universities, as seen in Appendix A, Table A1. The coefficients reported are the coefficients on the triple interaction between the treatment group indicator (English HEIs), the indicator of being a student living at home and the treatment period indicator, and show the change in the probability of observing a low income student as a result of the 2006 funding reforms studying at university whilst living at home compared to students at university not living at home, where living at home is one of the categories of term time accommodation status, where alternative categories of term time accommodation are being in private or university rented accommodation. Column 1 shows the DDD impact of the funding reforms for low income students studying whilst living at home across all subjects, compared to students not living at home across all subjects. Column 6 estimates the same as column 1, except the DDD estimation is only performed for non-elite universities. Columns 2 and 3 perform the same estimation as column 1, but for STEM and non-STEM subjects separately. Columns 7 and 8 also show the DDD for STEM and non-STEM subjects separately, but only at non-elite universities. The DDD estimates in these columns therefore show the impact of the funding reforms on the probability of observing a low income student who is living at home whilst attending university compared to not living at home, who studies a STEM or a non-STEM subject. Columns 4 and 5, and columns 9 and 10 show the results for the same approach for HFI and non-HFI subjects at elite and non-elite universities respectively. All relevant cross-products were included between the treatment group indicator, treatment period indicator and the live at home indicator, but are not shown here in the interests of concision To allow for the estimation of the effect of the policy averaged across the post-policy period and to capture the partial and full adjustment effects to the policy change, the DiD is estimated separately for the average of the post-policy cohorts, and for the 2006/07 and 2007/08 cohorts. There are therefore three rows of DiD estimates for each column, where each coefficient relates to the specific treatment period. There are three panels (A, B and C) to allow an estimation of the DiD across the broad low income background category (Panel A, SEC Code 4-8), and disaggregated across the upper-low income background category of students (Panel B, SEC Code 4-5) and the lower-low income background category of students (Panel C, SEC Code 6-8). University fixed effects are included for all specifications, as are controls for observable student heterogeneity. Namely, these individual controls are: gender, ethnicity, age, disability status, and nationality. Raw numbers have been rounded to the nearest multiple of 5 to comply with HESA's Standard Rounding Methodology. Standard errors are shown in parentheses and are clustered at the HEI level. *** p<0.01, ** p<0.05, * p<0.1

Whilst the upper-low income students who lived at home were more likely to experience the least risk in attending university, the upper-low income students who moved away for university were likely to experience the most risk. This can be seen in Panel B of Table 1.10, where Column 6 shows that there is a decrease of 1.82 percentage points overall for these students studying at non-Russell Group universities. Furthermore, when disaggregating this effect across subject types, the only subjects where there is a significant (and negative) effect on the conditional probability of observing a student are those subjects that arguably do not offer the best investment in terms of graduate employment or earnings. Specifically, the negative probability effects for the upper-low income students are for studying non-HFI subjects (-1.85 percentage points, which is greater for non-Russell Group universities at -2.44 percentage points), as well as for non-STEM subjects (-1.80 percentage points, only at non-Russell Group universities). The implication is that for the students who were the most likely to exhibit cautious behaviour in terms of subject and university choice (by choosing a subject and/or a university with the higher graduate reward) as a result of the 2006 policy, there is evidence of careful subject and university choice. However, this is conditional on the student having decided to participate in higher education.

Table 1.10: DiD Estimate of the 2006 Policy Reform – Live Far Away DDD by Subject

				All Universities					Non-Elite Universities				
				Live Far Away					Live Far Away				
				All	STEM	Non-STEM	HFI	Non-HFI	All	STEM	Non-STEM	HFI	Non-HFI
				(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Panel A													
SEC Code	4-8	Treat	Average	0.000163	0.00993	-0.00264	-0.00927	0.00467	0.00543	0.0232	0.00150	-0.0155	0.0114
		Period		(0.00749)	(0.0150)	(0.00704)	(0.0170)	(0.00958)	(0.00902)	(0.0246)	(0.00762)	(0.0207)	(0.0105)
			2006	0.00464	0.0520*	-0.00919	0.00210	0.00818	0.0189	0.0875**	0.00536	0.0102	0.0234
				(0.0155)	(0.0297)	(0.0164)	(0.0153)	(0.0166)	(0.0187)	(0.0345)	(0.0187)	(0.0227)	(0.0179)
			2007	-0.00241	-0.0300*	0.00565	-0.0121	0.000688	-0.00769	-0.0405*	-0.00170	-0.0263	-0.00350
				(0.00958)	(0.0175)	(0.0109)	(0.0160)	(0.0139)	(0.0116)	(0.0218)	(0.0122)	(0.0190)	(0.0170)
Panel B													
SEC Code	4-5	Treat	Average	-0.00686	0.00686	-0.00989	0.00321	-0.0120*	-0.0107	0.00177	-0.0126	-0.00772	-0.0130
		Period		(0.00764)	(0.00842)	(0.00816)	(0.0141)	(0.00714)	(0.00862)	(0.00726)	(0.00967)	(0.0173)	(0.00884)
			2006	0.00451	0.0232*	-0.000361	0.00521	0.00610	0.00804	0.0272*	0.00496	0.00160	0.0124
				(0.00737)	(0.0128)	(0.00852)	(0.0122)	(0.00960)	(0.00960)	(0.0158)	(0.0105)	(0.0177)	(0.0110)
			2007	-0.0109	-0.01000	-0.0106	0.000562	-0.0185**	-0.0182***	-0.0181	-0.0180***	-0.00957	-0.0244***
				(0.00725)	(0.0139)	(0.00739)	(0.00877)	(0.00723)	(0.00600)	(0.0167)	(0.00676)	(0.00820)	(0.00687)
Panel C													
SEC Code	6-8	Treat	Average	0.00702	0.00308	0.00725	-0.0125	0.0167	0.0162*	0.0214	0.0141	-0.00776	0.0244**

		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Period		(0.00883)	(0.0172)	(0.00874)	(0.0133)	(0.0109)	(0.00883)	(0.0247)	(0.00890)	(0.0191)	(0.0114)
	2006	0.000134	0.0288	-0.00883	-0.00310	0.00208	0.0108	0.0603**	0.000399	0.00855	0.0110
		(0.0104)	(0.0230)	(0.0106)	(0.0133)	(0.0113)	(0.0120)	(0.0275)	(0.0119)	(0.0183)	(0.0130)
	2007	0.00851	-0.0200	0.0162	-0.0127	0.0192	0.0105	-0.0224	0.0163	-0.0168	0.0209
		(0.0105)	(0.0126)	(0.0125)	(0.0115)	(0.0154)	(0.0129)	(0.0211)	(0.0145)	(0.0163)	(0.0181)
Fixed Effects		University	University	University	University	University	University	University	University	University	University
Controls		yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
Obs (SEC Code 4-8)		691,475	150,250	541,220	216,505	474,970	444,395	67,640	376,760	130,240	314,155
Obs (SEC Code 4-5)		691,475	150,250	541,220	216,505	474,970	444,395	67,640	376,760	130,240	314,155
Obs (SEC Code 6-8)		691,475	150,250	541,220	216,505	474,970	444,395	67,640	376,760	130,240	314,155

This table reports the results from the DDD estimation as outlined by equation (5), estimated across all universities (columns 1-5) and then across non-elite universities (columns 6-10). Elite status is denoted by membership of the Russell Group of universities, as seen in Appendix A, Table A1. The coefficients reported are the coefficients on the triple interaction between the treatment group indicator (English HEIs), the indicator of being a student attending a university far away and the treatment period indicator, and show the change in the probability of observing a low income student as a result of the 2006 funding reforms studying at a university far away compared to a university not defined as far away, where attending a university far away is defined as being in excess of 80km Euclidean distance from the student's domiciled postcode at the time of application to HE. Column 1 shows the DDD impact of the funding reforms for low income students studying at a university far away across all subjects, compared to universities not defined as far away across all subjects. Column 6 estimates the same as column 1, except the DDD estimation is only performed for non-elite universities. Columns 2 and 3 perform the same estimation as column 1, but for STEM and non-STEM subjects separately. Columns 7 and 8 also show the DDD for STEM and non-STEM subjects separately, but only at non-elite universities. The DDD estimates in these columns therefore show the impact of the funding reforms on the probability of observing a low income student who attends a university far away compared to a university not defined as far away, who studies a STEM or a non-STEM subject. Columns 4 and 5, and columns 9 and 10 show the results for the same approach for HFI and non-HFI subjects at elite and non-elite universities respectively. All relevant cross-products were included between the treatment group indicator, treatment period indicator and the far away group indicator, but are not shown here in the interests of concision To allow for the estimation of the effect of the policy averaged across the post-policy period and to capture the partial and full adjustment effects to the policy change, the DiD is estimated separately for the average of the post-policy cohorts, and for the 2006/07 and 2007/08 cohorts. There are therefore three rows of DiD estimates for each column, where each coefficient relates to the specific treatment period. There are three panels (A, B and C) to allow an estimation of the DiD across the broad low income background category (Panel A, SEC Code 4-8), and disaggregated across the upper-low income background category of students (Panel B, SEC Code 4-5) and the lower-low income background category of students (Panel C, SEC Code 6-8). University fixed effects are included for all specifications, as are controls for observable student heterogeneity. Namely, these individual controls are: gender, ethnicity, age, disability status, and nationality. Raw numbers have been rounded to the nearest multiple of 5 to comply with HESA's Standard Rounding Methodology. Standard errors are shown in parentheses and are clustered at the HEI level. *** p<0.01, ** p<0.05, * p<0.1

1.6 Conclusions & Discussion

To evaluate whether the introduction of the 2006 tuition fee policy affected the participation and behaviour of students in England, a difference-indifferences approach was adopted to estimate the average treatment effect of the treatment on the treated. Using Wales as a control group, where Welsh-domiciled students were exempt from the policy in 2006 and insulated from the policy from 2007 thereafter, HESA data was used in a DiD and DDD estimation framework.

The results suggest that the probability of observing an upper-low income student (which was the group expected to be the most detrimentally affected by the policy) actually increased in the treatment period. However, this can be traced to an increase in probability of observing a student in that particular income grouping remaining in the local area. Furthermore, when combining this with subject choice, it appears that the coupling of the tuition fee loans up front, alongside the guarantee that students only start repaying these loans when they can afford to do so, may have removed the risk involved in investing in a university degree, leading to an increase in the probability of observing an upper-low income student staying in the local area who chose a non-High Future Income subject. For those low income students for whom a decrease in the probability of observing them attend university, the magnitude is similar to the findings in the US and German literature.

In the context of estimating treatment effects, this chapter highlights the importance of considering the heterogeneous manner in which a policy may affect certain sub-groups of individuals. Without the DDD approach, the average DiD findings may have been misleading. Additionally, this chapter also highlights the importance of considering the policy response lag when estimating treatment effects. The majority of the DiD and DDD estimates for the 2006/07 cohort were insignificant, and if the policy impact was evaluated only on that basis, the funding reforms would have appeared to have little effect on low income student participation behaviour. Given the structure of pre-university schooling and the subject-specific pre-requisites for some courses, the impact of the policy reforms on the full adjustment cohort

(2007/08) – and the lack of impact on the partial adjustment cohort – is arguably rather reasonable to expect.

The use of a valid control group – in this case, Welsh universities – means that the estimates of the policy impact were obtained comparing students within the same higher educational framework, subject to the same macroeconomic influences and who also attended university with an almost identical structure of schooling. Where previous studies have estimated the impact of the 2006 funding reforms using arguably quite different groups of students, or where no counterfactual outcome was used at all, this may be the fundamental reason why the empirical literature suggests the reforms had little impact. Once a valid control group is identified, and when the research design acknowledges the significant heterogeneity in the student population (both in terms of how and where they study), a more representative impact of the policy may be obtained.

Chapter 2

Are Local Labour Markets Important in Determining Subject Choice?

"Conditional on changes in lifetime expected earnings, recessions encourage women to enter male-dominated fields, and students of both genders [to] pursue more difficult [subjects], such as [Science, Technology, Engineering and Mathematics] (STEM) fields."

(Blom et al., 2015)

"Over a third (37%) of graduates regret going to university…and 49% believe they could have got to where they are in life without the benefit of a degree."

(Collinson, 2016)

2.1 Introduction

As students approach the end of their school education, they must decide whether to apply to university and study towards a first degree (or similar higher education qualification[63]), or whether to enter the jobs market and search for non-graduate employment.[64] The above quotes highlight that if students do enter higher education (HE), then the returns on their investment in the form of higher wage premia and employment probability are not certain. Moreover, since the returns to the investment decision tend to vary by subject studied, students may undertake an investment that does not offer them significant advantages in their labour market,

[63] Students may elect to obtain vocational or professional qualifications for employment reasons, which are not considered here.

[64] There is a third option – namely, to delay the investment decision by taking a 'gap year'. For the purposes of this analysis, students who elect to delay the investment decision are not of interest. Moreover, if students do elect to take a gap year, they will either enter the non-graduate jobs market, or not be looking for employment.

and as such the costs of acquiring the HE will be greater than the benefits. The relative importance and magnitude of the benefits of HE may also vary depending on the labour market conditions in which the student makes the decision, and the conditions in which the student expects to graduate in. This changing of the benefits and costs to obtaining a university education depending on the subject studied, labour market conditions (and expectations thereof), and whether students are responsive to those changes, are the focus of this chapter.

This investment in HE (and schooling, more generally) has been well-researched in the context of labour and education economics, most famously by Becker (1964) who formalised the acquisition of education as an investment decision. From a purely theoretical view, students will decide to undertake HE if the discounted present value of the future flow of benefits are greater than the costs of acquisition, conditional on the assumptions of perfect information and non-binding credit constraints. The benefits to HE are higher earnings, and a lower likelihood of unemployment; whilst the costs to acquiring the degree are both pecuniary (tuition costs, books, additional living expenses whilst studying, foregone earnings in the acquisition period) and non-pecuniary (effort required to study, psychic costs of being away from home).

However, the investment decision in practice is made under uncertainty, which may mean that students undertake HE when the eventual costs will outweigh the benefits, and vice versa. The uncertainty in the benefits originate from the graduate wage premium, and the probability of obtaining graduate employment; both of which vary by the choice of subject and university.[65] If the decision is made by a rational, forward-looking individual, the student will choose the subject and university that maximises the expected returns to HE, given the subject-specific

[65] The uncertainty is further complicated by the individual-specific graduate wage premium is not necessarily the average graduate wage premia for each subject, but rather a probabilistic draw from a distribution of potential graduate wage premia.

likelihood of employment post-graduation and the average subject-specific graduate wage premium, conditional on ability.

This chapter seeks to analyse this uncertain investment decision in the context of the changing costs and benefits to an undergraduate degree with respect to variation in the local labour market conditions. A significant body of research exists which estimates the impact of labour markets on participation in HE, but not to any great extent in terms of subject choice. For instance, although unemployment has been robustly found to be a positive determinant of participation in HE, this aim of this research is to evaluate what subjects those students are likely to choose. Furthermore, it is plausible that students form expectations of the future labour market given its current trends, which implies that choosing a degree subject with a higher probability of graduate employment may be relatively more important.

The results show that, contrary to what might be expected in terms of maximising the benefits of an investment, students who experience high or rising levels of unemployment are less likely to study a subject with the highest returns with respect to graduate wage and employment probability premia. This suggests that students who experience increasingly challenging local labour markets respond by choosing subjects with lower rewards, but with a higher probability of degree success. Students may therefore be exhibiting behaviour that points toward a fear of failure when the relative importance of obtaining HE. However, the results also show that students who are classified as lower socioeconomic status students are more likely to choose a subject with the highest graduate premia and the greatest probability of employment. The implication is therefore that the students who are the most aware of the costs are the most aware of the benefits of HE, especially in the context of social mobility.

The remainder of this chapter is structured as follows: section 2.1.1 provides an overview of the human capital investment decision hitherto discussed, with a formal presentation of the problem under uncertainty in section 2.1.2. Section 2.1.3

discusses how unemployment may affect the decision with reference to the formal problem. Section 2.2.1 proposes a summary of the determinants of subject choice, since that is the outcome of interest, with discussion on how the labour market may influence subject choice in section 2.2.2. Section 2.2.3 thus proposes the research hypotheses which are estimated using student and labour market data discussed in section 2.3 and using the methodology discussed in section 2.4. The results are presented in section 2.5, and conclusions and policy implications are discussed in section 2.6.

2.1.1 Human Capital Theory

At the end of compulsory education, further or higher education is an investment decision, and as such the investment should rationally be undertaken if the net present value is greater than the net present cost, over the student's life-cycle. If a student decides to acquire HE, they receive the benefit of additional units of human capital that allow higher earnings to be obtained.[66] The empirical estimates of that benefit – the graduate wage premium – vary significantly, from a 17% return to an undergraduate degree on average (Blundell et al, 2000), to 8-10% (Harmon et al, 2000).

Much of the variation in calculating the average return to HE can be explained by the numerous dimensions along which the returns are heterogeneous. Ultimately the benefit to HE depends on the subject studied[67], the university attended[68], eventual degree class[69], the student's own characteristics (such as gender[70], ethnicity[71]), and the occupation in which the graduate eventually works[72]. Nevertheless, in addition to an increase in life-cycle earnings, an investment in HE

[66] Much debate surrounds whether the returns to a degree are from increased productivity, from signalling of higher quality in an asymmetric information jobs market (Spence, 1979), or a mixture of both. See Chevalier et al. (2004) for a synthesis of the debate.
[67] e.g. Chevalier (2011).
[68] e.g. Chevalier (2014); Hussain et al. (2009).
[69] e.g. Di Pietro (2010); Naylor et al. (2015).
[70] e.g. Britton et al. (2016); Sloane & O'Leary (2004).
[71] e.g. Britton et al. (2016).
[72] e.g. for STEM subjects, see Greenwood et al. (2011).

also increases the probability of employment[73], decreases the probability of spells of future unemployment[74], and has been shown to have indirect benefits that have spill-over effects to the wider society, such as improved health outcomes and a reduction in the incidence of crime.[75]

However in order to acquire the education that attains these benefits, the student must incur both pecuniary and non-pecuniary costs. The most direct of these costs is the cost of the tuition itself, which may or may not vary across subjects.[76] Additionally, a student will incur indirect costs whilst attending university, including accommodation, study materials, living expenses and travel; although the effect of these may be mitigated by students living close to or at home for at least part of their degree studies. Furthermore, although not incurred directly as a cost, the foregone earnings that the student is not able to earn due to full-time studying are the opportunity costs of HE, and can vary depending on the individual student and their local labour market characteristics, as well as the length of degree study. Non-pecuniary costs of the acquisition of HE are generally psychic costs (the effort expended by the student in order to obtain the degree, the potential of not being able to see family), which again can vary according to the student's individual characteristics.

Whether the student decides to invest in HE depends on whether the perceived net present value of the future flow of benefits over the student's life-cycle outweigh the net present value of all incurred costs.

[73] See Walker & Zhu (2013), Figure 6.
[74] Walker & Zhu (2013).
[75] For an overview of the empirical literature on the social returns to higher education, see for example Toutkoushian & Paulsen (2016).
[76] At the graduate level and for part-time students of both undergraduate and postgraduate levels of study, it is common for tuition fees to vary by course and university in the UK. At the full-time undergraduate level in the UK, the tuition costs vary by university but not by subject. See the Introduction of this thesis for a comprehensive overview, and section 1.2 for a history of tuition fees in the UK until September 2012.

2.1.2 A Model of Human Capital Acquisition

To formalise this relationship, and to allow the explicit interaction of labour markets with the benefits and costs of obtaining a degree[77], we can categorise the decision as a net present value calculation, similar to Becker (1964) and Koch (1972). It is assumed that the student has perfect information of the parameters, time periods are denoted by *t*, and for simplicity the investment duration is 1 period, and if it is undertaken it is taken in period 0. For further simplicity, the graduate premium is assumed to be homogenous across subjects, but this assumption will be relaxed once the model is constructed. Furthermore, all costs and benefits of the human capital decision are assumed to be independent of student heterogeneity.[78]

$$H^* = \sum_{t=1}^{T} \frac{W_t^G - W_t^N}{(1+r)^t} - (C_0 + F_0) \tag{6}$$

where:

$$\left.\begin{aligned} W_t^G &= \rho^G \cdot w_t^G \\ W_t^N &= \rho^N \cdot w_t^N \end{aligned}\right\} \forall\, t \geq 1$$
$$F_0 = \rho^N \cdot w_0^N \} \; t = 0 \tag{7}$$

The net present value of the investment in higher education is captured by H^*. If the student decides to acquire higher education in period 0, they are able to earn W_t^G in period 1 and thereafter, which is the graduate wage. If the student decides not to acquire the degree, the student can only obtain the non-graduate wage, W_t^N in all periods. Note that this is the same wage that is foregone in period 0 if the

[77] This chapter is an empirical investigation, rather than an empirical test of a theoretical model. The model proposed however, is useful in giving a framework to the interaction of perceived present and future labour market conditions to the investment decision in higher education.

[78] It has been found that the returns to higher education are heterogeneous across gender, income background and other familial characteristics (see section 2.2.1), but for simplicity the actual benefits and costs are assumed to be identical to every potential student, independent of student or course/university characteristics. This is arguably more realistic with respect to the costs of the acquisition of human capital, since tuition fees in the period of study did not vary according to university or subject. Nevertheless, it is an oversimplification of the non-course costs, such as relocation and costs of living, which are assumed to exhibit significant variation.

student does invest in higher education, and this opportunity cost is captured by F_0. The student's direct costs (both pecuniary and non-pecuniary) are also incurred in just period 0[79], and are denoted by C_0, and the student discounts the future at rate r.

However, as this is a decision made under uncertainty, there are two measures of employment probability included, which are, ρ^N (the probability of obtaining employment as a non-graduate) and ρ^G (the probability of obtaining employment as a graduate). By definition, both probabilities are independently within of the interval [0,1], and they are assumed to be time-invariant. As it is not guaranteed whether the student is in employment irrespective of the investment decision, the benefit of acquiring the investment is not only the graduate wage premium (since it assumed that $W_t^G > W_t^N \;\forall\; t$), but also the increased probability of obtaining employment (since it is also assumed that $\rho^G > \rho^N$). Following Universities UK (2010), graduates experience an increased probability of obtaining non-graduate employment, thus if the investment decision is taken it implies that there may be a further diverging in the gap between ρ^G and ρ^N as a result of the increased difficulty for non-graduates to obtain non-graduate employment, especially in the context of business cycles. The average probability of obtaining a non-graduate job is effectively the inverse of the youth unemployment rate in the student's local area, whilst the average probability of obtaining non-graduate employment in future periods may be determined by both local and national labour market trends.

The discount rate is central to the decision of whether or not to invest in HE, as it determines the relative value of future benefits to current incurred costs to the individual student. Oreopoulos (2006) finds that the discount rate for school-leavers is significantly higher, which renders the decision to invest atypical since the potential benefits of a degree may not be fully appreciated by the prospective student, relative to the costs of doing so. Nevertheless, given a student's individual characteristics (discount rate, psychic costs), labour market characteristics (the

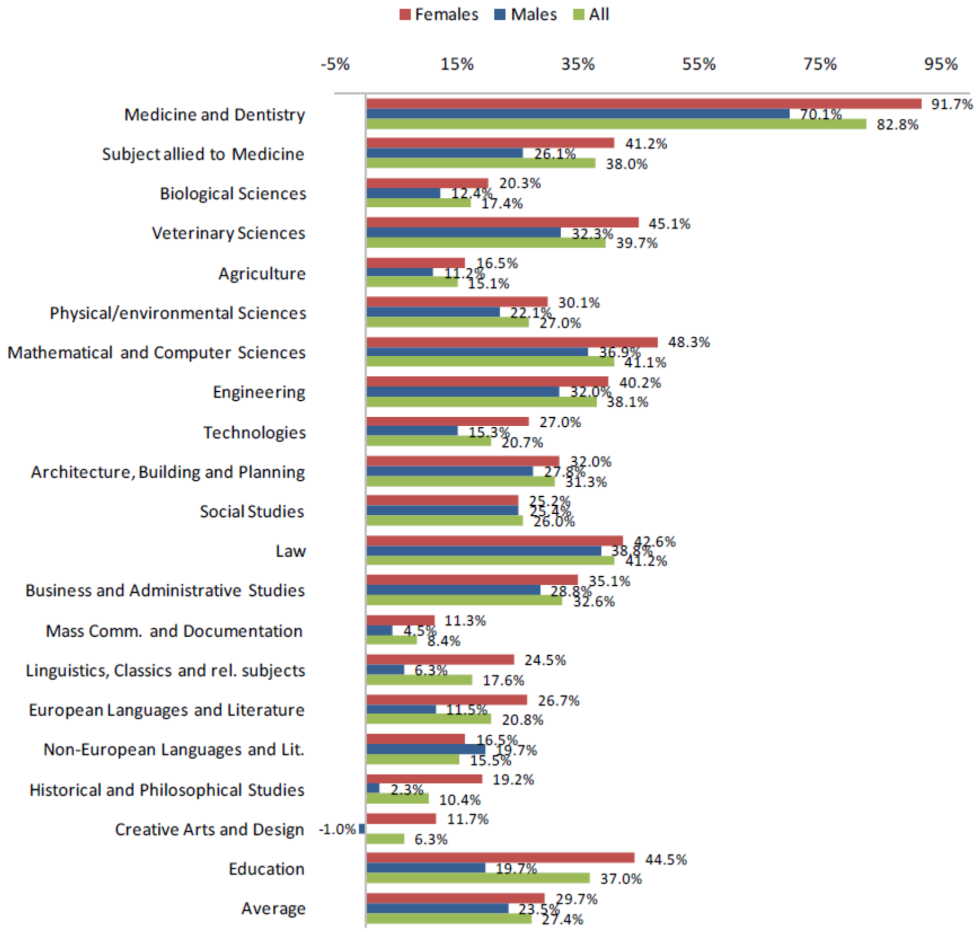
[79] This is a possible oversimplification, since the psychic costs may be incurred in the post-investment period, too.

average wage and employment probability premia), and the direct costs of the higher education, the student will decide to invest if $H^{*} \geq 0$, otherwise the student will decide against investing ($H^{*} < 0$).

However, the assumption that the benefits to a degree are constant across subjects is a gross oversimplification. The above model does not account for the large and significant variation in the graduate wage premium, and the heterogeneity in the probability of graduate employment, according to subject studied. As Walker and Zhu (2011) show that even when controlling for the net present value of the tuition fees paid, students who undertake Law, Business or Economics degrees experience a significantly higher lifetime earnings profile than Social Sciences, Arts and Humanities.[80] This reflects the findings of the variation in the returns to an undergraduate degree by subject, as shown by BIS (2011) which can be seen in Figure 2.1.

[80] In particular, see Walker & Zhu (2011), Figures 1 & 2, p.6 for a visual representation.

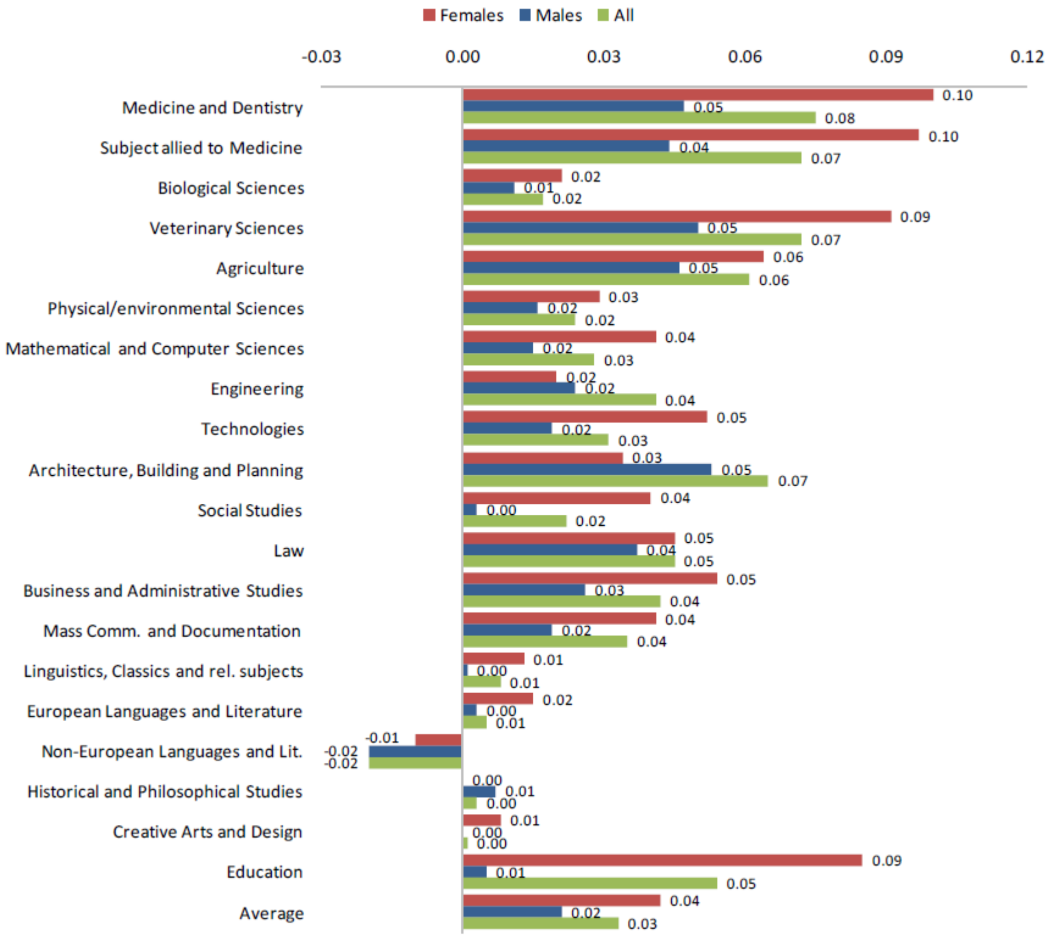
Figure 2.1: Returns to Undergraduate Degree by Subject



Source: BIS (2011), Figure 4

The oversimplification of the homogeneity of the benefits of a degree are also not likely to be limited to the graduate wage premium, but also the probability of graduate employment. Some subjects, such as STEM and Business, may offer a higher probability of employment beyond that of school-leavers qualifications, compared to Arts and Humanities. This was also documented by BIS (2011) and is shown in Figure 2.2 below.

Figure 2.2: Employment Premia, by Undergraduate Degree Subject



Source: BIS (2011), Figure 8

Further oversimplifications of the model outlined in this section include the possibility of a non-graduate obtaining a graduate wage or a graduate obtaining a non-graduate wage – both of which are non-zero events – and the lack of an explicit channel for subject choice. The aim of this chapter is empirical rather than theoretical, therefore a complete and comprehensive human capital acquisition model which accounts for the heterogeneity in the costs and benefits of HE is not the objective. Nevertheless, the model provides a useful theoretical framework upon which to ask the question: given there are costs and benefits to HE, how would the labour market affect the decision?

2.1.3 How Unemployment May Affect Higher Education Decisions

Perhaps the most direct impact of unemployment on the investment decision is the opportunity cost, especially when considering the youth unemployment rate in the local area[81]. If the student does decide to acquire HE, there is a cost of the foregone earnings that could have been earned through non-graduate employment after schooling.[82] However, those foregone earnings are not a certainty, and thus the size of the opportunity cost of HE participation is dependent on (1) the probability of employment, and (2) the non-graduate wage. If there is higher youth unemployment concurrently with the decision, this effectively reduces the size of the opportunity costs, since it is less likely the student would be able to obtain the non-graduate employment if they decide to not attend university. There may also be a further negative impact on the size of the opportunity cost if the youth unemployment is so substantial that it puts downward pressure on wages, although that is unlikely given the majority of earnings for school-leavers in the UK are at or around the national minimum wage.[83]

Nevertheless, if larger youth unemployment in the local area does reduce the size of the opportunity cost of HE, this should mean there is a positive relationship between the local youth unemployment rate and the participation rate of HE. Thus, the participation rate may be counter-cyclical with respect to the business cycle. Bell and Blanchflower (2011) note that increases in participation by young people

[81] It is assumed that at least initially, there is a significant degree of geographic immobility for a school-leavers.
[82] We also assume that students whilst studying are not engaged in part-time employment, but even if this assumption is relaxed, it would likely only reduce the size of the opportunity costs rather than remove them altogether – since it is highly unlikely that the part time employment would equate the earnings from full-time employment in the counterfactual case.
[83] See, for example, the Annual Survey of Hours and Earnings: 2016 provisional results, Figure 14. Available at: https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/earningsandworkinghours/bulletins/annualsurveyofhoursandearnings/2016provisionalresults#earnings-by-age-group

in HE in response to an adverse labour market may be due to a lack of other alternatives, or in an attempt to be insulated from the inclement conditions. Moreover, a significant body of research exists on the impact of unemployment on college (US) and university (UK, Europe) participation. Pissarides (1981) was one of the first to explicitly include unemployment as a determinant of the destinations of school leavers, and found that varying measures of unemployment did have a positive relationship with the likelihood of studying beyond the compulsory school-leaving age. Similar effects were found by Rice (1987), Whitfield and Wilson (1991), Bennett et al. (1992). In the US, Betts and McFarland (1995) showed that a 1 percentage point increase in the unemployment rate led to a 4.5% increase in the probability of enrolment into college, particularly for full-time students, whilst Fredriksson (1997) found similar effects in the Swedish HE sector, and Di Pietro (2004) in Italy.

Fernández and Shioji (2000) specifically focused on the effect of unemployment on HE incentives, rather than estimating the demands for HE in a more general context. They posit two potential effects of unemployment: an investment effect (where unemployment lowers the opportunity cost of investing in HE), and a wealth effect where higher unemployment may dissuade HE participation through financial constraints on the student's household. They find that both effects are present in their study, and the investment effect was also found to be significant by Card and Lemieux[84] (2001), Dellas and Sakellaris (2003), Dellas and Koubi (2003), Long (2004, 2014), Arkes (2010) and Clark (2011). Hillman and Orians (2013) show that for community college enrolments in the US, a 1 percentage point increase in the US increases enrolment demand by 1-3%.

However, the positive relationship between unemployment and HE participation has not been estimated as consistent across all students. Black and Sufi (2002) and

[84] The authors also note the importance of the finding by Oreopoulos (2006) that effects of unemployment on participation may be exacerbated by the relatively high discount rate of school-leavers.

Dellas and Sakellaris (2013) both find the effect of unemployment on participation to be quantitatively larger for students from higher socioeconomic status students, whilst Boffy-Ramirez (2016) finds that the effect of unemployment on participation is concentrated to students at the lower end of the ability distribution. Furthermore, Johnson (2013) finds that for graduate school students, only females respond to unemployment, and Kahn (2010) shows that it is national unemployment rather than local unemployment that graduate students respond to. This is likely due to the decision to acquire further HE being fundamentally different from acquiring college/higher education. Long's study (2014) of how college enrolment in the US was affected by the Great Recession showed that the positive relationship did indeed exist, but mainly for part-time enrolment. Thus, since it is likely that the opportunity cost of acquiring HE varied by the individual's labour market characteristics, it is unsurprising that the effect of unemployment on the investment decision is heterogeneous along the same dimensions.

In addition to affecting the likelihood of participation, labour market trends have also been evaluated in terms of their impact on the likelihood of completion, conditional on participation. It is likely that if the acquisition period of HE is lengthy, then higher unemployment will mean that the opportunity costs of remaining in HE are low, compared to lower unemployment. Thus at the margin, students may be less likely to drop-out. The effect of unemployment on the rate of student attrition has attracted relatively less attention, but Stratton et al (2007) show that there is an inverse relationship between the unemployment rate and probability that a student drops out of HE, but only for full time students. Di Pietro (2006) finds the same relationship in Italy, which had the highest drop-out rate in the OECD in the 1990s due to the low costs of tuition, and the non-selective admissions policy.

2.2 Undergraduate Degree Subject Choice

As the returns to an investment in HE vary significantly by subject (Walker & Zhu, 2010), the choice of subject studied should be a significant factor in determining whether or not to participate in HE. If higher unemployment affects the investment decision through a lowering of the costs and, as Fernández and Shioji (2000) remark, the expectation of higher future unemployment reduces the benefits of HE, then it may also be the case that the choice of subject is affected by labour market factors. In order to consider the possible influences on subject choice of the labour market, the other determinants of subject choice must be examined.

2.2.1 What Determines Subject Choice

The first major studies to evaluate what determines subject choice in HE were Freeman (1971) and Koch (1972), who formally proposed that the choice depended on the utility (Freeman) or the individual rates of return (Koch). Thus, in accordance with human capital theory, students choose the subjects that maximise their returns to investment in HE; evidence of which is shown by Koch, where subject choice is partly determined by changes in the individual rates of return to subjects. In support of this argument, in the UK, Pitcher and Purcell (1998) surveyed final-year undergraduates and over 50% of students stated that career prospects were one of the main reasons behind course choice. In the UK context of ever-increasing tuition fees – some of which will exceed £9,000 in September 2017 – employment factors will only become more important in the HE investment decision, especially given Pitcher and Purcell's finding when direct tuition costs had not yet been introduced. Chevalier (2011) finds that the returns to some subjects in terms of graduate earnings compared to others are more than double after controlling for student heterogeneity, which given that UK tuition fees are constant across subjects within a university, implies significant variation in the rate of return to individual investment choices. In addition, the risk of graduate unemployment is also found to be significantly heterogeneous across subjects (Universities UK, 2010), although it is still found to be lower than non-graduates.[85]

However, the choice of subject is a multi-faceted decision, and is likely to be made from a mixture of socioeconomic factors, ability, preferences and the wider course

[85] See for example Smith et al. (2000).

characteristics, rather than just the financial benefits to the subject degree. This idea of non-monetary factors influencing course choice was estimated by Cebula and Lopes (1982), who extend Koch's analysis to include a larger array of subject choice determinants. In addition to the earnings motivation (wage differentials and graduate earnings growth were the most important factors for subject choice), the authors highlighted the importance of non-monetary influences such as the probability of successful attainment and the characteristics of the eventual job. Furthermore, as different subjects will have different choice motivations, Skatova and Ferguson (2014) identify four major categories of motivations, and attribute the choice of engineering to career and employment reasons, whilst the choice of arts and humanities subjects were ascribed to a degree of 'loafing' – where students choose a subject in order to enter HE with greater ease.

Aside from interest in the subject, employment reasons are likely to be the most reported factor in course choice (Montmarquette et al., 2002), and one fundamental element is that of income expectations. In human capital theory, prospective students are assumed to have perfect knowledge of the monetary benefits to the investment in HE, but in practice, graduate earnings are estimated with a degree of uncertainty. Whilst wage signals have been found to be important for course choice (Wales, 2010), the expectations of graduate earnings and employment prospects, and hence the returns to individual degree subjects, are not always made using all of the information available to the prospective students (Universities UK, 2010). Furthermore, even if all of the information is used by the prospective student, the size of the income expectations and the benefits in terms of graduate employment depends on the manner in which the expectation is formed.

Firstly, the high discount rate for school-leavers estimated by Oreopoulos (2006) was found to be even higher for students from low income backgrounds (Oosterbeek & Van Ophem, 2000). The implication is that students from poorer families may underestimate the potential returns to HE, which is one possible explanation why Davies et al (2013) find that lower income students are less likely to choose high wage premia subjects. This supports the earlier finding that income expectations can be over, under or well-estimated, and can vary by socioeconomic background, as shown by Smith and Powell (1990). The authors find that not having a parent with university level education may overestimate the returns to a degree, conditional on family income. Delaney et al (2011) estimate that income expectations vary significantly by socioeconomic status, even when controlling for student characteristics such as ability, demographics and personality, and suggest that if students from lower income backgrounds are underestimating the returns, it may be capturing some expected wage or hiring discrimination in the labour market. This would imply that although the returns to a degree vary by subject, students expect that there is additional variation in the returns by their own individual socioeconomic factors.

One further finding from Montmarquette et al. (2002) is that students from less affluent backgrounds are likely to choose less risky subjects in terms of difficulty and drop-out, which may in part explain the finding by Davies et al. (2013) that lower income students are less likely to choose high wage premia subjects. If the most rewarding subjects in terms of graduate earnings and employment prospects are also the most risky in terms of difficulty (both of application and completion), then any systematic degree of risk or disappointment aversion from lower socioeconomic status students may imply that socioeconomic advantages have educational advantages. Anderberg and Cerrone (2016) show there may be disappointment aversion for lower ability students, which emphasises the supposition by Duru and Mingat (1979) that there is a trade-off between the risk of failure and the higher returns to certain degrees. The extent to which students choose less risky subjects may also be exacerbated if the students are debt-averse (Callender & Jackson, 2008) which may occur even through higher perceived costs of obtaining the HE qualification. Furthermore, Pennell and West (2005) note that in response to debt aversion or higher perceived costs, students from lower socioeconomic backgrounds are more likely to attend university closer to their parental home, which may mean that the range of course choices available to the student are limited.

Arguably, one of the groups of subjects with the highest collective reward in terms of earnings and employment probability premia is that of Science, Technology, Engineering and Mathematics (STEM) subjects, and these subjects have seen some of the highest increases in student uptake between 2003-2012.[86] Graduates of STEM subjects are significantly less likely to be unemployed (Universities UK, 2010; Harrison, 2012), and earn significantly more[87] (Sloane & O'Leary, 2004; Mellors-Bourne et al., 2011), compared to other degree subjects. Choice of a STEM subject varies by socioeconomic background, and ethnic minorities have been shown to be more likely to study a STEM subject, conditional on family background and attainment (Codiroli McMaster, 2017). However, despite the seemingly large rewards to studying STEM subjects, not all STEM graduates choose to work in STEM jobs (approximately 1/3 in 2011 according to Bosworth et al. (2013)), and income expectations are not a significant determinant of choosing a STEM subject (Mellors-Bourne et al., 2011).

Ultimately, the determinants of subject choice are numerous and vary according to the individual subject and student. The relatively consistent finding of risk aversion to choosing the most rewarding subjects by the most disadvantaged students however, implies that students who could benefit the most from social mobility may inherently be less likely to choose the subject that allows it. Nevertheless, there is a significant body of evidence to suggest that students do respond to the changing benefits and costs of a degree, albeit in a heterogeneous manner. This heterogeneity in forming income expectations, responding to wage signals, and the socioeconomic differences in choosing degree subjects requires a heterogeneous approach to evaluating the impact of the changing of the benefits and costs to acquiring HE with respect to the labour market.

[86] See Universities UK (2014), Table 3, p.15.
[87] This is especially true of graduates of mathematics (Greenwood et al, 2011).

2.2.2 Labour Markets and Subject Choice

The outlining of how the labour market may affect the decision to invest in HE was under the simplifying assumption that the differences in probability of employment and the graduate wage premium were homogenous across subject types. This particular question has had limited attention in research output, although Universities UK (2010) note that there is some descriptive evidence to suggest that the Great Recession had an impact on the subject choice of undergraduates in the UK. Sectors that were the most hit during the recession (construction and finance) saw a corresponding decline in the demand for and uptake of subjects related to those industries.

The report also notes that if students expect higher unemployment in the future, students may choose subjects with higher probabilities of employment (for example, vocational or public sector orientated subjects), although no evidence of such behaviour is presented. One major caveat to re-optimising subject choice in the context of increasing unemployment is that it is difficult to switch subjects with full flexibility, given that the requirements for entry into some subjects at university may be unobtainable for students depending what a student has decided to study at school pre-university.[88] However, there may still be a degree of flexibility available for a student to switch their course choice in response to higher expectations of future unemployment amongst subjects of similar subject requirements, or those without subject-specific requirements altogether. A further constraint on choosing a degree subject that maximises the benefits is if the adverse economic conditions incentivise the student to reduce the direct costs of acquiring the degree by living

[88] In the UK context, this is determined by the student's choice of 3-4 A-Level subjects at school or college, which are chosen roughly two years prior to university entry. Further information on the timeline of the application process in the UK for school-leavers can be found in Table 2.1 and the Introduction to this thesis. A further dimension of university entry is that students do not require A-Levels or equivalent school/college qualifications, but instead students can be admitted on the basis of practical or work experience. However, as the analysis is performed on students who have entered university at school-leaving age, this is not a consideration for the empirical approach.

close to or at home. If this is the case, the set of possible course choices is likely to be reduced, suggesting that although living close to or at home minimises the potential costs of HE, it may also minimise the potential benefits.[89]

The only study that attempts to model university subject choice explicitly as a function of previous subject choice, student and labour market characteristics is Wales (2010), who finds that the youth unemployment rate is a positive determinant for the choice of six subjects: mathematics, engineering, history, medicine and dentistry, veterinary sciences, and languages. Since four of these are STEM subjects, it suggests that higher youth unemployment may influence students to choose subjects with higher relative returns, although the magnitude of the effects are relatively small. Adult unemployment is found not to be a significant determinant of subject choice, although there is evidence that industry-specific trends in employment are important for signalling to the student.

2.2.3 Research Hypotheses

Given the possible labour market effects on the HE decision, two major hypotheses are that:

1. Higher local youth unemployment reduces the opportunity cost in the investment decision through a lowering of the probability of obtaining non-graduate employment (the direct effect) and by the potential downward pressure on wages (indirect). Hence, at the margin, it is likely that students decide to enrol in HE.

    a. Students who enrol due to a lowering of the opportunity cost may not choose subjects for employment or wage premium reasons.

[89] Universities UK (2010) finds no evidence of students living close to home in response to adverse economic conditions or financial concerns, unlike Pennell and West (2005) and the findings in Chapter 1 of this thesis.

2. Positive and significant growth rates of youth and adult unemployment give signals to the student about the future condition of the labour market, and hence may improve the student's perception of the benefits (through a lowering of the discount rate, or a more accurate estimate of the changes in the non-graduate probability of being employed). Hence, at the margin, it is likely that more students decide to enrol.

    a. Students who enrol due to a lowering of their discount rate or a perception of a worsening labour market in the future may choose subjects to maximise the returns in graduate wage premia and employment probabilities.

If these hypotheses are correctly supposed, the local youth unemployment rate should be a positive determinant of students participating in HE, but not necessarily for subjects that offer the greatest benefits in the labour market. Additionally, significant growth rates of local unemployment (both youth and adult) should be positive determinants of participating in HE, but specifically for subjects that offer the greatest labour market advantages. Fernández and Shioji's (2000) wealth effect, should not be significant in the UK context, if the funding schemes during the period of interest truly render HE accessible, independent of household income. Ultimately, it is reasonable to expect that significant current (or expectations of future) unemployment at the local level may incentivise students to switch into more rewarding degree subjects at the margin. On the contrary, significant or rising youth unemployment reduces the opportunity cost of acquiring higher education, and students that switch into higher education as a result of the lowering of the opportunity cost may choose less rewarding degree subjects if there is a 'weathering the storm' effect.

2.3 Data

To examine the relationship between local labour markets and subject choice, a combination of student and labour market data is used.

2.3.1 Higher Education Statistics Agency (HESA) Data

The student data used is the 2005/06 to 2009/10 destinations and leavers of individuals from higher education compiled by HESA, as outlined in Chapter 1 of this thesis (section 1.3.1). The individual-level data allows for the estimation of the determinants of subject choice as outlined in section 2.2.1, including socioeconomic factors (income background, ethnicity, parental education level), student characteristics (age, tariff score, gender) and course characteristics (subject studied, mode of attendance, university attended, length of study). Furthermore, since each student observation has the postcode within which the student was domiciled at the time of applying to university through the Universities and Colleges Admissions Service (UCAS), the mapping of local area characteristics to the students at the time of applying is possible. As the outcome of interest is the subject studied by the student[90], the student's specific UCAS course code will be used to create groups of subjects (see section 2.4.3).

2.3.2 Labour Market Data

A variety of different labour market variables will be used to capture the various possible effects of interest as proposed in sections 2.1.3 and 2.2.2. Firstly, to estimate the effect of youth unemployment on the opportunity cost of investing in HE, the youth unemployment rate is calculated at the local level using data from the Quarterly Labour Force Survey (QLFS) for 1999-2003 and from the Annual

[90] This will be estimated with a degree of imprecision, since it is not known whether the student changed their subject between entering and leaving higher education. All that is known is the eventual subject of the degree awarded. In the case of combined degrees where two or more subjects are studied, the subject indicated is that with the highest weighting in the student's degree structure or the subject that appears first in the degree programme (for example, a student studying Politics, Philosophy and Economics would be recorded as a Politics student).

Population Survey (APS) for 2004-2010.[91] In both cases, the youth unemployment rate is defined as the ratio of unemployed 16-24 year olds to the population of 16-24 year olds that are in the labour force, averaged across the 12 month period ending in December each year. Additionally, the adult unemployment rate at ‘prime’ age (25-49 year olds) is calculated from the same data sources over the same periods, to capture any potential effect of the adult unemployment rate on subject choice.

However, it is more likely that whilst the concurrent youth unemployment rate will affect the opportunity cost of HE, it is the expectation of the conditions of the labour market in the future that will affect the perceived benefits to the investment. To attempt to estimate this, the year-on-year growth rate of both the youth and adult unemployment rates are calculated using 1 period differences in the local observations.

An alternative approach to using local unemployment data would be to use trends in graduate employment, graduate earnings, or graduate vacancies. This may be considered the most useful if it is believed that students when deciding which subject to study for their degree investment are using graduate-level trends. However, this is not the approach taken in this analysis for two reasons: firstly, since it is reasonable to expect the headline unemployment rate at the local level is the most relevant for students to respond to through media and everyday life; and secondly, graduate employment data by subject and by time is not widely available.

2.3.3 Spatial & Time Matching

To accurately estimate the impact of local unemployment on HE decisions, the unemployment data is matched to the student data using the student’s UK domiciled postcode at the time of the application.[92] The unemployment data is measured at

[91] The annual population survey effectively replaced the local area QLFS in providing labour market data at a local level in 2004.

[92] A further avenue of research could be to investigate the possible role of students choosing to move to a particular university in preparation of finding employment in the local area. In order to

the 2nd level of the Nomenclature of Territorial Units for Statistics (NUTS 2, 2010) level, which comprises of 37 regional areas of the UK.[93] Since the classification underwent structural changes within the analysis period (2003, 2006), the NUTS 2 (2010) regional data is created using the lower level of the hierarchical classification which did not change over the time period (NUTS 3, 2010) to ensure consistent measures of regional unemployment are collected. Each student is assigned to their local NUTS 2 region using their domiciled postcode, and the unemployment observations are matched by NUTS 2 region. For each NUTS 2 region, there are 11 years of observations of the unemployment rate, and 10 years of observations of the growth rate of the unemployment rate (due to the raw calculation of the year-on-year change). This is a similar approach to Wales (2010), except the measures of locality used in that study are much smaller (Local Area Districts).[94]

There is significant variation in the regional differences in the unemployment rates and growth rates: Inner London has on average around 20% youth unemployment across the period, compared to Surrey, and East and West Sussex with less than 10% youth unemployment. In addition, some regions show an increasing trend of youth and adult unemployment (Derbyshire, Cumbria) whilst others show a decline over the period (North Eastern Scotland, Outer London).[95] However, the spatial dimension alone does not sufficiently match the labour market conditions to the investment decision – the time element is also important, as shown in Table 2.1.

---

facilitate such an investigation, it would be necessary to have information on the location of the student pre-university, at-university and post-university once in graduate employment. Whilst the data used in this analysis contains the two former, it does not contain the latter, and therefore is not possible.

[93] See Appendix B, Table B1.

[94] Whilst using this classification of local areas was considered for this analysis, the additional cohorts of students included in this dataset require a consistent measure of unemployment from 2002 to 2006 for students starting 2003/04 to 2007/08. A further source of richness of the data is this analysis includes the entire population of students who entered higher education, whereas students included in Wales (2010) are those that respond to the Destinations and Leavers of Higher Education (DLHE) survey, which is typically performed six months after a student graduates.

[95] See Appendix B, Figures B1-B6.

Table 2.1: An Example Timeline of UK Undergraduate Study (of a Student Commencing Their Studies in Year '*t*'

Year	Month	Event
t-2	June	Student leaves compulsory secondary school; decides whether to study A-Levels[96], where to study (college / sixth form) and which A-Level subjects to study.
	September	Student begins 2 year A-Level study.
t-1	June	Finalise A-Level subject choices.[97]
	September	Student begins application process to university through Universities and Colleges Admissions Service (UCAS); may apply to up to 6 course choices at any UK university.[98]
t	January	Deadline for UCAS application submission; course and university choice finalised.
	June	Receive and respond to offers of study from university admissions.
	August	Receive exam results and confirm attendance on preferred course.
	September	Begin degree studies.

In the context of the UK higher education system, a student will partially decide to go to university by deciding whether to study A-Levels at age 16 at college or sixth form, and they will also decide to some extent which subjects they may wish to study, since subject-specific A-Levels (or equivalent) are often required for university admission. For a student who wishes to start in year *t*, their application to university begins in September *t*-1, and must be completed by early January in year t. Thus, the majority of the decision of whether to and what to study takes place nine months previous to entry. Therefore, for a student starting in year *t*, it is the local labour market conditions observed in September-December in year *t*-1 that determine the opportunity cost of HE. In addition, it is the expectations of future

[96] As outlined in the Introduction, these are the pre-university qualifications a student will obtain by studying at school, college or sixth form in England and Wales. In Scotland, students study for similar pre-university qualifications called Highers or Advanced Highers.
[97] Typically, students choose a reduced number of the A-Level subjects they started (AS-Level) to continue on to the full A-Level (A2). The subjects chosen to study in the final year of college or sixth form to a large extent will determine the available course choices at university in terms of entrance requirements.
[98] Since 2007, this has been reduced to 5.

unemployment at that same time that help to determine the expectations of the future labour market that they may face either as a graduate or non-graduate. With respect to matching student observations to labour market data, a student who is observed to have started in year *t* is matched with the 12 month average local unemployment rates ending in year *t*-1. Whilst this means that students are matched with the 12 month average unemployment rate in the previous calendar year – and students decide what subjects to study in September *t*-2, this will be captured by the growth in the local unemployment rate from *t*-2 to *t*-1. The matching procedure is therefore also performed with the growth rates of the local unemployment rates, as shown in Table 2.2 below.

Table 2.2: Matching of Unemployment Data to Student Entrants

Student Enters HE	**Student Begins Application to HE**	**Local Unemployment Matched (12 month average)**	**Growth in Local Unemployment Matched (percentage point change)**
September 2003	September 2002	January to December 2002	(January to December 2002) – (January to December 2001)
September 2004	September 2003	January to December 2003	(January to December 2003) – (January to December 2002)
September 2005	September 2004	January to December 2004	(January to December 2004) – (January to December 2003)
September 2006	September 2005	January to December 2005	(January to December 2005) – (January to December 2004)
September 2007	September 2006	January to December 2006	(January to December 2006) – (January to December 2005)

One final comment on the unemployment data is that of a potential simultaneity problem. As Bell and Blanchflower (2011) note, an increase in the participation in HE may have two potential effects. Increased participation may increase the youth unemployment rate further, since the remaining pool of those aged 16-24 and not in further education may become less skilled. Alternatively, since there has been an increase in the participation in HE, this across a longer time dimension will make those aged 16-24 more skilled, and may therefore reduce the youth unemployment rate. The result is that, as Clark (2011) also notes, such simultaneity may cause a downward bias in the estimates on the effect of unemployment on HE investments. However, since this analysis concentrates on the subject choice effects of labour market trends, and that the data ensures that those observed are conditional on having decided to participate in HE, such bias should not be as problematic as looking at the impact on participation directly. Furthermore, by examining the impact of local adult unemployment and expectations of future local adult unemployment on subject choice, negates this simultaneity bias.

2.3.4 Identification Strategy

As the research question focuses on local labour markets in the UK, only students with a valid UK postcode are kept from the complete population of leavers, which removes students who have entered HE as non-domiciled within the UK from the analysis. Further restrictions ensure that only students with a valid (i.e. non-missing) socioeconomic code are kept[99], in order to control for socioeconomic factors in the determination of subject choice. Also excluded are any exchange students (approximately 2.5% of the student observations), any students who are not studying an undergraduate degree for the first time, and any students studying at arts, drama or music Higher Education Institutes (HEIs). Students who attended a university in Northern Ireland are removed due to significant missing observations of labour market data. In a further attempt to only include students who are making

[99] Approximately one third of undergraduate students do not have a valid socioeconomic code, but this is predominantly those who did not apply to university as a school leaver.

the investment as a school-leaver, only students with a valid tariff score[100] (similar to Wales, 2010), who are eligible to pay home tuition fees[101] (explicitly excluding foreign students) and those aged 17-18 at the time of entry to university are retained in the dataset. Part-time students are also excluded in order to restrict the analysis of the unemployment effect only on those who incur the full opportunity cost of HE.[102]

2.4 Methodology

Ideally, to fully estimate on the demand for particular subjects with respect to variation in the local labour market, it would be necessary to capture changes in the demand for HE through applications data. This would allow the estimation of the student, socioeconomic and regional characteristics on the decision not to invest in HE, rather than the conditional probabilities we are able to obtain with the observations for those who have self-selected into HE. Given that the HESA data is ex-post, we only observe the subject that a student studies, conditional on them having applied, accepted an offer of study, and subsequently been accepted on the basis of pre-university attainment. What we can estimate therefore is a compositional effect: given that students have decided to enrol, what is the relationship between the local labour market and the subjects they choose?

2.4.1 The Linear Probability Model

As discussed in detail in the proceeding section, each student's specific course code is grouped into one of three main categories: STEM, LEM and OTHER. Given that all the student observations have decided to participate in HE, the relationship of

[100] A universal measure of pre-university attainment that is calculated by UCAS' own application structure. See section 3.4.3 for a full description of how UCAS tariff scores are computed.
[101] Note that this does not exclude students who received full or partial tuition fee waivers or bursaries, but that the students were eligible to pay the home fees as domiciled within the UK for at least 3 years prior to entry.
[102] From the cropped dataset outlined in Appendix A, Table A3, approximately 330,000 observations are dropped which are students not aged 17-18, and a further 225,000 observations are dropped where the student's tariff score is unknown, missing, or unclassified. The resulting dataset is approximately 200,000 students.

interest is what determines which subject group a student belongs to, given the explanatory variables. Thus, there will be three dependent binary variables for each subject category. One approach would be to use a logit or probit model instead of OLS, however the coefficients are not immediately quantifiable, especially when multiple interactions are included. If OLS is used, and the error term is assumed to be distributed normally, the estimation of a limited dependent variable becomes the Linear Probability Model (LPM). Although the coefficients are easily interpreted as they represent marginal effects directly, the lack of a restrictive condition on the error term means that predicted probabilities can exceed the [0, 1] boundary and thus cause inherent problems of heteroscedasticity, resulting in potentially biased standard errors.

However, in in the interests of comparability with the estimates from the preceding chapter, the linear probability model will be used to perform the analysis, with a logit specification which re-estimates all the analysis with a non-linear error term distribution to as a robustness check to confirm the sign and significance of the coefficients reported in the main tables.[103] The LPM approach will also allow a degree of comparison with Wales' (2010) study, which is the closest study to this analysis. Clustered standard errors will also be used in the LPM estimations to address the heteroscedasticity issue, especially since the error term (irrespective of regression method) may not be independent across observations due to certain groups of students and subjects being more likely to be observed at certain universities.

2.4.2 Subject Groupings

Using each student's specific 4-digit Joint Academic Coding System (JACS) identifier, 20 broad subject areas are defined. These 20 subject areas are then assigned to one of three possible subject groupings, to reflect three different sets of

[103] A summary of these results are presented in Tables 2.11, 2.12, 2.13 and 2.14, where the average marginal effects are also given following a multinomial logit regression. The complete set of results are available on request from the author.

subjects.[104] The first grouping used is STEM: Science, Technology, Engineering and Mathematics subjects.[105] As noted in section 2.2.1, these subjects have some of the highest entry requirements (see Table 2), but they also offer arguably the greatest benefits in terms of a high graduate wage premium and a low risk of graduate unemployment for those students who are willing and able to obtain a STEM degree. For students who are seeking to maximise their return on their investment in higher education, an investment in a STEM subject would offer a good chance of doing so, despite the lack of employment motivation noted by Mellors-Bourne et al. (2011).

However, given that STEM subjects may be prohibitively difficult to obtain offers of study for, especially since the timeline of A-Level prerequisite subjects has already been decided by the student in advance of observing any changes in the labour market, it may be unlikely that students are able to switch into a STEM subject easily. As Britton et al. (2016) and Walker and Zhu (2010) also identify, there exists a group of subjects that still offer high graduate wage premia, but are relatively more accessible to students in terms of prerequisite subjects and grades. These subjects are henceforth defined as LEM: Law, Economics and Management. Finally, the remainder of the subjects – which are predominantly arts and humanities courses – are defined as OTHER. Although not constant across all subjects, on average these subjects will offer a lower return on the investment decision, especially compared to LEM and STEM subjects in the graduate labour market.

The resulting dataset comprises approximately 200,000 students who were domiciled in either England, Scotland or Wales prior to the beginning of their undergraduate degree studies, which they began between September 2003 and

[104] See Appendix B, Table B2.
[105] As noted in Section 1.4.3, the STEM classification is created using the JACS 2-digit code to ensure that individual non-STEM degree courses are not incorrectly included in the STEM definition.

September 2007, of which a summary of the key variables can be found below in Table 2.3.

Table 2.3: Summary Statistics

	(1) Total	(2) STEM	(3) LEM	(4) Other
England	82.66%	80.37%	81.05%	84.15%
Scotland	10.99%	14.20%	13.17%	8.94%
Wales	6.35%	5.44%	5.78%	6.91%
Student Characteristics:				
Russell Group =1	0.40	0.57	0.35	0.34
	(0.490)	(0.495)	(0.477)	(0.474)
Oxbridge =1	0.04	0.05	0.03	0.04
	(0.199)	(0.220)	(0.167)	(0.198)
Age	17.96	17.95	17.95	17.97
	(0.200)	(0.228)	(0.227)	(0.176)
Tariff Score	341.42	364.34	342.26	333.63
	(118.0)	(126.3)	(119.8)	(113.4)
Female =1	0.58	0.42	0.51	0.67
	(0.494)	(0.493)	(0.500)	(0.470)
White =1	0.85	0.80	0.75	0.89
	(0.361)	(0.400)	(0.431)	(0.308)
Disabled =1	0.07	0.07	0.05	0.07
	(0.251)	(0.252)	(0.222)	(0.260)
British =1	0.97	0.96	0.95	0.97
	(0.181)	(0.189)	(0.210)	(0.167)
Low Income =1	0.27	0.26	0.29	0.27
(SEC Code 4-8)	(0.443)	(0.436)	(0.452)	(0.443)
Educated Parent =1	0.07	0.06	0.07	0.07
	(0.250)	(0.231)	(0.251)	(0.258)
No Educated Parent =1	0.07	0.05	0.08	0.07
	(0.252)	(0.221)	(0.276)	(0.255)
Local =1	0.19	0.20	0.24	0.18
(<20km)	(0.394)	(0.398)	(0.426)	(0.380)
Live at Home =1	0.19	0.18	0.24	0.19

	(0.396)	(0.381)	(0.428)	(0.390)
Local Labour Market Characteristics:				
Local Youth Unemployment	12.79	12.78	13.14	12.69
(16-24)	(3.487)	(3.489)	(3.525)	(3.468)
Local Adult Unemployment	3.79	3.82	3.91	3.74
(25-49)	(1.157)	(1.161)	(1.197)	(1.139)
Youth Unemployment Growth	0.05	0.04	0.05	0.05
(1 year)	(0.161)	(0.156)	(0.160)	(0.164)
Adult Unemployment Growth	0.02	0.01	0.02	0.03
(1 year)	(0.162)	(0.162)	(0.159)	(0.163)
Aggregate Unemployment	4.98	5.00	5.13	4.93
(16-49)	(1.442)	(1.451)	(1.476)	(1.423)
Aggregate Unemployment	0.08	0.05	0.08	0.10
Growth (1 year)	(0.595)	(0.590)	(0.596)	(0.596)
N	198,280	48,240	36,345	113,695
	(100%)	(24.33%)	(18.33%)	(57.34%)

Mean coefficients; standard errors in parentheses. The table shows the characteristics of the student data, by subject classification. For each subject group (and across all subject groups), the summary of student and labour market characteristics are shown. Raw numbers have been rounded to the nearest multiple of 5 to comply with HESA's Standard Rounding Methodology.
Low Income is a binary variable that takes the value 1 if the student's socioeconomic code is 4-8.
Educated Parent is a binary variable that takes the value 1 if the student declares that they have at least 1 parent with university education. No Educated Parent is a binary variable that takes the value 1 if the student declares that neither parent has university education. This variable was not compulsory in all waves of the data, and students were given the option of responding 'Unsure'.
White is a binary variable that takes the value 1 if the student's ethnicity is coded as being white. Black & Asian is a binary variable that takes the value 1 if the student's ethnicity is coded as being either black or Asian. Female and British are binary variables that take the value 1 if the student's gender and nationality are declared as female and British respectively.
Disabled is a binary variable that takes the value 1 if the student has declared that they have a known disability.
Local is a binary variable that takes the value 1 if the student attends a university within 20km in Euclidean space from their domiciled postcode where the student resided at the time of making the application through UCAS.
Live at Home is a binary variable that takes the value 1 if the student is known to be living at home whilst undertaking their university study.
Russell Group is a binary variable that takes the value 1 if the university the student attended is classified as a Russell Group university, as in Appendix A, Table A1. Furthermore, Oxbridge is a binary variable that takes the value 1 if the university the student attended was either the University of Oxford or the University of Cambridge.

2.4.3 Empirical Model

The model estimated across the three subject groupings is performed with OLS, and in Table 2.4 the determinants of choice for each subject grouping are estimated before including any labour market measures. This is done both to establish a baseline of subject choice depending on socioeconomic factors, and also to see whether the influence of those factors remains when adding local labour market trends and any possible interactions. Hence, for STEM, LEM and OTHER subject groups, the model estimated initially is:

$$y_{i,j,u,t} = \delta_0 + \delta_1 \mathrm{T}_i + \sum_{k=1}^{K} \beta_k X_{i,j,u,t} + \sum_{t=2004}^{T} \psi_t D_t + \sum_{n=2}^{N} \psi_j N_n + \sum_{u=2}^{U} a_u A_u + \varepsilon_{i,j,u,t} \tag{8}$$

where $y_{i,j,u,t}$ is a binary indicator which equals one if student *i* from local area *j* studies at university *u* within academic year *t*, a subject classified as being in the subject group of interest, and 0 otherwise; T_i is the student's tariff score on entry which measures the student's ability before attending university; and X_k is a set of categorical student controls (ethnicity, disability status, gender, low/high socioeconomic status, attending a local university within 20km Euclidean distance). Also included are a set of time fixed effects (D_t), regional fixed effects at the NUTS 1 (2010) level (N_n) and a full set of university fixed effects (A_u). Finally, the error term ($\varepsilon_{i,j,u,t}$) is clustered at the university level to account for the likely violation of the independent error term assumption.

Once the model specification has been performed for all three subject groups, measures of unemployment are added to the model for Tables 2.5-2.7.[106] Depending

[106] One potential consequence of including more than one measure of unemployment in the same specification is that of multicollinearity between the explanatory variables. Although this would not bias the OLS estimates, it may result in the standard errors being overestimated. To ensure the

on the specification used, these unemployment measures matched with student *i* who begins their studies in time *t* are either the 12 month average in t-1, or the growth in local unemployment from t-2 to t-1. However, given that the benefits and costs may vary according to socioeconomic factors (section 2.2.1), the effect of unemployment on subject choice is also allowed to be heterogeneous across students with different socioeconomic characteristics. Specifically, the effect of unemployment is allowed to be heterogeneous across gender and socioeconomic status.[107] To do this, a pairwise interaction between the unemployment measures and the categorical socioeconomic variables of female and low income are included in Tables 2.8-2.10. The regression model therefore becomes:

$$y_{i,j,u,t} = \delta_0 + \delta_1 \mathrm{T}_i + \sum_{k=1}^{K} \beta_k X_{i,j,u,t} + \sum_{t=2004}^{T} \psi_t D_t + \sum_{n=2}^{N} \psi_j N_n + \sum_{u=2}^{U} a_u A_u + \delta_2 U_i + \delta_3 (X_1 * U_i) + \varepsilon_{i,j,u,t} \tag{9}$$

Where δ_2 measures the impact of the local unemployment measure on the probability of student studying a particular subject group, and δ_3 captures any heterogeneity in the effect of δ_2 across the socioeconomic characteristics of female and low income students.

2.5 Results

The results from the model specification analysis for the STEM, LEM and OTHER subject groupings are shown in Table 2.4, where the determinants of subject choice

unemployment measures did not exhibit collinearity, the Variance Inflation Factor (VIF) was computed for all specifications in Tables 2.5 to 2.7, and all of the explanatory variables showed a vif significantly less than the accepted threshold value of 10. Only when including aggregate and adult unemployment measures in the same specification does the VIF exceed 10, however this combination of explanatory variables does not appear in this chapter's analysis.

[107] An additional source of heterogeneity was that of students who decided to attend a local university compared to those who moved further away. There was no significant heterogeneity in the distance of a student's university from their home, and the effect of local unemployment on subject choice.

as suggested by the empirical body of research are included, subject to the data used. These initial analyses establish a baseline of subject choice before estimating the impact of the labour market variables. Once established, the unemployment measures are included in Tables 2.5-2.7, which only show the unemployment coefficients. Finally, to see how the effects of unemployment on subject choice may differ by socioeconomic factors, Tables 2.8-2.10 show the interaction between socioeconomic variables, and with the unemployment rates, to establish any apparent heterogeneity. Tables 2.11-2.14 show the average marginal effects from a multinomial logit specification to help quantify and compare the aforementioned.

In establishing the determinants of subject choice, the majority of the explanatory variables are the expected sign and significance. For STEM subjects, having a higher tariff score makes it more likely that a student chooses a STEM subject[108], however this effect becomes statistically insignificant once controlling for regional, time and university fixed effects. What is perhaps surprising is that students from a low income background are more likely to be observed studying a STEM subject.[109] Also surprising is that if a student is attending a local university, they are more likely to be studying a STEM subject, which casts doubt on the idea of a limited geographical radius for university choice negatively impacting on the availability of possible courses. There is also a positive effect for disabled students on the probability of choosing STEM subjects, where having a self-reported disability should not have any significant impact on course choice. Finally in column (3), there is an additional negative effect for females who are also from a financially poorer background on the probability of choosing STEM subjects, which was also found by Codiroli McMaster (2017). Whilst low income was found to increase the probability of studying a STEM subject, for females this is significantly negative, which does not reduce the magnitude of being female. Females from low income backgrounds are therefore approximately 14 percentage points in total less likely to

[108] Whilst being female or white makes it less likely. As evidenced in Column (3) of Table 2.11, females are 12 percentage points less likely to study a STEM subject.
[109] As seen in Column (3) of Table 2.11, low income background students are 2.01 percentage points more likely to be studying a STEM subject.

study a STEM subject. This result also increases the magnitude of the positive effect for low income on the probability of STEM subjects being studied overall from 1.06 percentage points to 3.78 percentage points. The other interaction between low income and local is also significant, but negative. This suggests that, although local students are more likely to study STEM subjects, students who attend a local university and who are also low income are less likely to study STEM subjects.

For LEM subjects, students who are female, white or disabled are less likely to be studying LEM subjects conditional on participation, and students who are local and have a higher tariff score are more likely.[110] There is no apparent heterogeneity in the effect of income background, and no effect overall of a student's socioeconomic status on the probability of choosing LEM. Finally, for the OTHER subjects, as expected, students with a lower tariff score are more likely to be studying a non-STEM and non-LEM subject. Also as expected, female and white students are more likely to be studying what are predominantly arts and humanity degrees.[111] Surprising however, is that low income students are less likely to be studying OTHER subjects, which is contrary to the finding of disappointment or loss aversion by choosing less risky subjects by Montmarquette et al. (2002). Further surprising is that local students are less likely to be studying OTHER subjects, which is contrary to the ideas proposed by Callender and Jackson (2008) and Pennell and West (2005). Furthermore, as this is in part a compositional analysis, the findings for low income and local students in STEM are of the opposite sign but of similar magnitude of the coefficients on the same explanatory variables in the OTHER subject grouping.

[110] Only once regional, year and university fixed effects are included in the regression does tariff become significant. Also see Column (4) of Table 2.11 for the average marginal effects.
[111] For the average marginal effects from the multinomial logistical specification, see Column (5) of Table 2.11.

Table 2.4: Baseline Model of Subject Choice

	(1) STEM	(2) STEM	(3) STEM	(4) LEM	(5) LEM	(6) LEM	(7) OTHER	(8) OTHER	(9) OTHER
Tariff	0.00718***	-0.00190*	-0.00189*	0.00133	0.00438***	0.00440***	-0.00852***	-0.00250**	-0.00250**
	(0.00112)	(0.000968)	(0.000968)	(0.00106)	(0.000977)	(0.00100)	(0.00123)	(0.00104)	(0.00104)
Female	-0.140***	-0.122***	-0.115***	-0.0618***	-0.0626***	-0.0635***	0.202***	0.185***	0.179***
	(0.00664)	(0.00608)	(0.00585)	(0.00548)	(0.00483)	(0.00501)	(0.00747)	(0.00695)	(0.00716)
White	-0.0593***	-0.0427***	-0.0428***	-0.153***	-0.152***	-0.153***	0.212***	0.196***	0.196***
	(0.0153)	(0.00857)	(0.00854)	(0.0145)	(0.0107)	(0.0104)	(0.0163)	(0.0125)	(0.0125)
British	-0.0127	0.00925	0.00910	-0.0263**	-0.0186	-0.0169	0.0390***	0.00768	0.00778
	(0.00934)	(0.0164)	(0.0164)	(0.0118)	(0.0155)	(0.0157)	(0.0114)	(0.0305)	(0.0305)
Low Income	0.0106***	0.0196***	0.0378***	0.00504*	0.00114	-0.00293	-0.0156***	-0.0206***	-0.0349***
	(0.00325)	(0.00297)	(0.00573)	(0.00291)	(0.00289)	(0.00493)	(0.00346)	(0.00293)	(0.00497)
Local	0.0189**	0.0162***	0.0207***	0.0308***	0.0302***	0.0268***	-0.0497***	-0.0461***	-0.0474***
	(0.00922)	(0.00519)	(0.00538)	(0.00870)	(0.00560)	(0.00615)	(0.0118)	(0.00512)	(0.00565)
Disabled	0.00930*	0.0138***	0.0139***	-0.0394***	-0.0385***	-0.0383***	0.0301***	0.0245***	0.0245***
	(0.00483)	(0.00410)	(0.00411)	(0.00381)	(0.00416)	(0.00418)	(0.00525)	(0.00513)	(0.00513)
Low Income x Female			-0.0253***			0.00314			0.0222***
			(0.00560)			(0.00528)			(0.00571)
Low Income x Local			-0.0126**			0.00886			0.00377
			(0.00623)			(0.00644)			(0.00708)
Constant	0.213***	0.198***	0.193***	0.363***	0.448***	0.449***	0.424***	0.355***	0.358***
	(0.0212)	(0.0278)	(0.0276)	(0.0274)	(0.0236)	(0.0235)	(0.0278)	(0.0363)	(0.0363)
Observations	198,280	198,280	198,280	198,280	198,280	198,280	198,280	198,280	198,280

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
R-squared	0.043	0.106	0.106	0.028	0.061	0.061	0.077	0.122	0.122
Year FE	✗	✓	✓	✗	✓	✓	✗	✓	✓
University FE	✗	✓	✓	✗	✓	✓	✗	✓	✓
Region FE (NUTS 1 2010)	✗	✓	✓	✗	✓	✓	✗	✓	✓
Clustered SE	✓	✓	✓	✓	✓	✓	✓	✓	✓

Standard errors shown in the parentheses are clustered at the institutional level across all 9 specifications. All regressions are performed using OLS to estimate equation 8. The dependent variable is the binary subject group indicator, which takes the value 1 for STEM in specifications (1)-(3) and 0 otherwise; 1 for LEM in specifications (4)-(6) and 0 otherwise; and 1 for OTHER in specifications (7)-(9) and 0 otherwise. For each subject grouping, the same 3 models are estimated - models (1), (4) and (7) include socioeconomic controls, and specifications (2), (5) and 8) include a full set of year, institutional and regional fixed effects. Specifications (3), (6) and (9) further allow the effect of low income to be heterogeneous across gender and geographical proximity by including interactions with the low income indicator. Raw numbers have been rounded to the nearest multiple of 5 to comply with HESA's Standard Rounding Methodology. *** p<0.01, ** p<0.05, * p<0.1

Tables 2.5-2.7 show the addition of various measures of unemployment in the subject groupings, using specifications 3, 6 and 9 from Table 2.4 for the STEM, LEM and OTHER subjects respectively. Tables 2.12-2.14 provide the average marginal effects from a multinomial logit specification of Tables 2.5-2.7 to provide robustness and quantifiable comparison. The effect of local unemployment rate amongst young people (16-24) represents changes in the opportunity cost of attending university, and for those students at the margin who do enrol into higher education, it was hypothesised that choosing a subject with the highest returns in graduate wage and employment premia would be less important for those students. From Table 2.5, there is no significant effect of local youth unemployment on STEM subjects, which may reflect the difficulty in switching into STEM subjects with respect to required subjects to be studied at school or college. Since LEM and OTHER subjects may be easier to change courses, any changes in the opportunity cost may affect these subjects more so than STEM. Tables 2.6 and 2.7 show a significant impact of local youth unemployment and course choice, but the sign is opposite to those which were hypothesised. In column (9) of both tables, once all the disaggregated unemployment measures are included, experiencing higher local youth unemployment in the previous 12 months has a positive effect on choosing a LEM subject, and a negative effect on the probability of choosing an OTHER subject. However, the effect is quantitatively small; a 1 percentage point increase in the youth unemployment rate increases the probability of a LEM subject being studied by 0.442 percentage points, or by 2.4% relative to the mean of the number of students studying LEM subjects. This would support a hypothesis to the contrary of what was proposed in Section 2.2.3, whereby course choice is actually more important for those motivated by changes in the opportunity cost of higher education. Although the opportunity cost has fallen for students who experience higher local youth unemployment, the worsening outside options in the labour market may at the margin as a signal to study a subject with greater labour market benefits. This suggests that changes in the opportunity cost are relatively insignificant with respect to subject choice, compared to changes in the future benefits of a degree subject.

Changes in the expected benefits to a degree are also partially captured by changes in the growth rates of adult and youth unemployment, such that if a student observes a worsening youth, or especially adult[112], labour market, this may help form expectations of the labour market that the student will be faced with upon graduation. Thus, if the expectations are for higher unemployment in the future, it is reasonable to expect that students will maximise their returns to HE by choosing a subject with higher graduate wage and employment premia. However, this is not what is observed in the data. In Tables 2.5 and 2.6, there are significantly negative coefficients on the growth rates of youth and aggregate unemployment, which imply that students who reside in an area of high or rising unemployment are less likely to choose STEM or LEM subjects; the subjects with the highest likely returns.[113]

Instead, as seen from Table 2.7, students who reside in areas with positive growth rates in youth and aggregate unemployment are more likely to study OTHER subjects.[114] This may suggest that students are engaging in some degree of loss or disappointment aversion; if students are aware that being a graduate carries a lower risk of unemployment, they may choose subjects where they are less likely to drop-out and more likely to obtain a high degree classification. Alternatively, it may be that students who select into university due to worsening labour market conditions are engaging in 'weathering the storm', although this is unlikely given that students' decision to participate is already partially made through the studying of A-Levels two years prior to entry. Irrespectively, these findings imply that current labour

[112] Since degree study takes typically 3-4 years, and that higher education is expected to benefit individuals throughout their working career, the prospective adult labour market should be the most pertinent to the student.
[113] Table 2.13, Column (3) shows that a 1 unit increase in the growth rate of local youth unemployment suggests that a student is 2.13 percentage points less likely to study a LEM subject.
[114] Table 2.14, Column (6) shows that a 1 unit increase in the growth rate of local aggregate unemployment suggests that a student is 0.98 percentage points more likely to be studying a subject in the OTHER category.

market conditions and the formation of future labour market conditions have different mechanisms for students in their subsequent course choice behaviour.

Table 2.5: STEM Subjects and Local Unemployment

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	STEM	STEM	STEM	STEM	STEM	STEM	STEM	STEM	STEM
Local Youth Unemployment (16-24yr)	-0.00109*						-0.000812		-0.00105
	(0.000635)						(0.000775)		(0.00110)
Local Adult Unemployment (25-49yr)		-0.00231						-0.00258	0.000586
		(0.00167)						(0.00180)	(0.00244)
Growth in Local Youth Unemployment (16-24yr)			-0.00928				-0.00559		-0.00244
			(0.00624)				(0.00741)		(0.00902)
Growth in Local Adult Unemployment (25-49yr)				-0.00879				-0.00530	-0.00922
				(0.00851)				(0.00886)	(0.00962)
Aggregate Local Unemployment (16-49yr)					-0.00305**				
					(0.00142)				
Growth in Local Aggregate Unemployment (16-49yr)						-0.00539**			
						(0.00216)			
Observations	181,090	187,825	158,720	178,995	196,355	196,355	158,720	178,995	158,720
R-squared	0.104	0.105	0.105	0.105	0.106	0.106	0.105	0.105	0.105
Year FE	✓	✓	✓	✓	✓	✓	✓	✓	✓
University FE	✓	✓	✓	✓	✓	✓	✓	✓	✓
Region FE (NUTS 1 2010)	✓	✓	✓	✓	✓	✓	✓	✓	✓
Clustered SE	✓	✓	✓	✓	✓	✓	✓	✓	✓

Standard errors shown in the parentheses are clustered at the institutional level across all 9 specifications. All regressions are performed using OLS to estimate equation 9. The dependent variable is the binary subject group indicator, which takes the value 1 for STEM and 0 otherwise. All 9 specifications are estimating column (3) from Table 2.4 with various indicators of unemployment. Specifications (1)-(6) estimate the unemployment variables separately, where specifications (7)-(8) include the level of local unemployment and the growth in local unemployment respectively. Specification (9) includes all measures of youth and adult local unemployment. The socioeconomic controls as included in specification (3) of Table 2.4 are included, but the results are supressed here to focus on the effect of the unemployment variables. The regression results including the socioeconomic controls can be found in Appendix B, Table B3. Raw numbers have been rounded to the nearest multiple of 5 to comply with HESA's Standard Rounding Methodology. *** p<0.01, ** p<0.05, * p<0.1

Table 2.6: LEM Subjects and Local Unemployment

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	LEM	LEM	LEM	LEM	LEM	LEM	LEM	LEM	LEM
Local Youth Unemployment (16-24yr)	0.000220						0.000363		0.00442***
	(0.000789)						(0.000979)		(0.00139)
Local Adult Unemployment (25-49yr)		-0.00160						-0.00229	-0.0135***
		(0.00200)						(0.00235)	(0.00307)
Growth in Local Youth Unemployment (16-24yr)			-0.0216***				-0.0233***		-0.0467***
			(0.00679)				(0.00797)		(0.0104)
Growth in Local Adult Unemployment (25-49yr)				-0.00421				-0.00112	0.0250**
				(0.00743)				(0.00854)	(0.0107)
Aggregate Local Unemployment (16-49yr)					0.000436				
					(0.00181)				
Growth in Local Aggregate Unemployment (16-49yr)						-0.00417**			
						(0.00205)			
Observations	181,090	187,825	158,720	178,995	196,355	196,355	158,720	178,995	158,720
R-squared	0.063	0.063	0.064	0.063	0.061	0.061	0.064	0.063	0.064
Year FE	✓	✓	✓	✓	✓	✓	✓	✓	✓
University FE	✓	✓	✓	✓	✓	✓	✓	✓	✓
Region FE (NUTS 1 2010)	✓	✓	✓	✓	✓	✓	✓	✓	✓
Clustered SE	✓	✓	✓	✓	✓	✓	✓	✓	✓

Standard errors shown in the parentheses are clustered at the institutional level across all 9 specifications. All regressions are performed using OLS to estimate equation 9. The dependent variable is the binary subject group indicator, which takes the value 1 for LEM and 0 otherwise. All 9 specifications are estimating column (6) from Table 2.4 with various indicators of unemployment. Specifications (1)-(6) estimate the unemployment variables separately, where specifications (7)-(8) include the level of local unemployment and the growth in local unemployment respectively. Specification (9) includes all measures of youth and adult local unemployment. The socioeconomic controls as included in specification (6) of Table 2.4 are included, but the results are supressed here to focus on the effect of the unemployment variables. The regression results including the socioeconomic controls can be found in Appendix B, Table B4. Raw numbers have been rounded to the nearest multiple of 5 to comply with HESA’s Standard Rounding Methodology. *** p<0.01, ** p<0.05, * p<0.1

Table 2.7: OTHER Subjects and Local Unemployment

	(1) OTHER	(2) OTHER	(3) OTHER	(4) OTHER	(5) OTHER	(6) OTHER	(7) OTHER	(8) OTHER	(9) OTHER
Local Youth Unemployment (16-24yr)	0.000871						0.000448		-0.00337**
	(0.000650)						(0.000814)		(0.00132)
Local Adult Unemployment (25-49yr)		0.00390**						0.00486**	0.0129***
		(0.00183)						(0.00212)	(0.00337)
Growth in Local Youth Unemployment (16-24yr)			0.0309***				0.0289***		0.0492***
			(0.00830)				(0.00934)		(0.0115)
Growth in Local Adult Unemployment (25-49yr)				0.0130				0.00641	-0.0157
				(0.00906)				(0.00976)	(0.0122)
Aggregate Local Unemployment (16-49yr)					0.00262*				
					(0.00148)				
Growth in Local Aggregate Unemployment (16-49yr)						0.00956***			
						(0.00241)			
Observations	181,090	187,825	158,720	178,995	196,355	196,355	158,720	178,995	158,720
R-squared	0.122	0.123	0.124	0.124	0.122	0.122	0.124	0.124	0.124
Year FE	✓	✓	✓	✓	✓	✓	✓	✓	✓
University FE	✓	✓	✓	✓	✓	✓	✓	✓	✓
Region FE (NUTS 1 2010)	✓	✓	✓	✓	✓	✓	✓	✓	✓
Clustered SE	✓	✓	✓	✓	✓	✓	✓	✓	✓

Standard errors shown in the parentheses are clustered at the institutional level across all 9 specifications. All regressions are performed using OLS to estimate equation 9. The dependent variable is the binary subject group indicator, which takes the value 1 for OTHER and 0 otherwise. All 9 specifications are estimating column (9) from Table 2.4 with various indicators of unemployment. Specifications (1)-(6) estimate the unemployment variables separately, where specifications (7)-(8) include the level of local unemployment and the growth in local unemployment respectively. Specification (9) includes all measures of youth and adult local unemployment. The socioeconomic controls as included in specification (9) of Table 2.4 are included, but the results are supressed here to focus on the effect of the unemployment variables. The regression results including the socioeconomic controls can be found in Appendix B, Table B5. Raw numbers have been rounded to the nearest multiple of 5 to comply with HESA's Standard Rounding Methodology. *** p<0.01, ** p<0.05, * p<0.1

These effects of unemployment on course choice however are assuming that the effect of unemployment is homogenous across all types of students. Since it has been discussed that different students may have different psychic costs, and different expectations of the benefits of HE, the effect of unemployment may also differ according to the same factors. Tables 2.8-2.10 present the findings of unemployment interacted with socioeconomic indicators to investigate whether there is any heterogeneity in the labour market influences of course choice. These tables use specifications 3, 6 and 9 from Table 2.4 with the inclusion of some socioeconomic indicator interactions between income background, gender and geographic proximity and the unemployment measures.

When allowing the unemployment effect to be heterogeneous across individuals, an added layer of student behaviour is revealed: female students who experience significant growth in aggregate unemployment are more likely to study a STEM subject, and less likely to study a LEM subject. This finding opposes the lower likelihood of a female studying a STEM subject by Codiroli McMaster (2017) and the result from Table 2.4. Moreover, the coefficient on female remains highly significant and negative, as does the interaction between female and low income on the probability of studying a STEM subject, even when the interaction between female and the growth in local aggregate unemployment is included in columns (8) and (9) in Table 2.5. Further to this, the quantitative effect of growth in local aggregate unemployment becomes greater, such that a 1 percentage point increase in the growth of local unemployment reduces the probability of a STEM subject being studied by 1.34 percentage points (a fall of 5.7% relative to the mean of STEM subjects studied), compared to 0.534 percentage points without the interactions included. This suggests that although females are less likely to study STEM subjects, and disadvantaged females are even less likely to study STEM subjects, significant growth in adult unemployment in the student's local area may incentivise females particularly to study STEM subjects as they offer some of the highest graduate wage and employment premia. Furthermore, as this is a compositional story, the effects for females studying STEM in response to changes in the growth of local aggregate unemployment is the opposite of what is observed in columns (8) and (9) of Table 2.9. This implies that the females who do study STEM in response to changes in aggregate local unemployment are switching from LEM subjects, rather than arts and humanities.

There are no apparent heterogeneous effects of unemployment on OTHER subjects, although, similar to the effect for females discussed above, students who are from a low income background and who experience higher aggregate unemployment in the local area are more likely to study a STEM subject. The coefficients on the interaction term in columns (5) and (6) of Table 2.8 are significant and positive, however the effect of being from a low income background overall becomes marginally statistically insignificant. This implies that the positive effect of poorer students choosing STEM subjects is driven by those students who experience higher aggregate unemployment in the local area. Similar to the effect for females, the results suggest that students that are the most disadvantaged in the labour market and who experience worsening local labour market conditions are more likely to choose to study a subject with higher graduate wage and employment benefits.

Table 2.8: STEM Subjects and Unemployment (Socioeconomic Heterogeneity)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	STEM	STEM	STEM	STEM	STEM	STEM	STEM	STEM	STEM
Female	-0.110***	-0.110***	-0.123***	-0.115***	-0.115***	-0.132***	-0.115***	-0.115***	-0.118***
	(0.00581)	(0.00581)	(0.0129)	(0.00582)	(0.00582)	(0.0127)	(0.00582)	(0.00582)	(0.00599)
Low Income	0.0363***	0.0169	0.0176	0.0376***	0.0172*	0.0181*	0.0374***	0.0374***	0.0372***
	(0.00562)	(0.0117)	(0.0117)	(0.00562)	(0.00999)	(0.00994)	(0.00562)	(0.00562)	(0.00561)
Low Income x Female	-0.0256***	-0.0257***	-0.0262***	-0.0258***	-0.0259***	-0.0266***	-0.0258***	-0.0258***	-0.0257***
	(0.00567)	(0.00567)	(0.00557)	(0.00554)	(0.00554)	(0.00545)	(0.00553)	(0.00553)	(0.00553)
Local Youth Unemployment (16-24yr)	-0.00108*	-0.00153**	-0.00212**						
	(0.000632)	(0.000596)	(0.000845)						
Low Income x Local Youth Unemployment (16-24yr)		0.00152*	0.00149*						
		(0.000871)	(0.000872)						
Female x Local Youth Unemployment (16-24yr)			0.00102						
			(0.000849)						
Aggregate Local Unemployment (16-49yr)				-0.00302**	-0.00424***	-0.00624***			
				(0.00142)	(0.00137)	(0.00192)			
Low Income x Aggregate Local Unemployment (16-49yr)					0.00414**	0.00402**			
					(0.00184)	(0.00184)			
Female x Aggregate Local Unemployment (16-49yr)						0.00341*			
						(0.00202)			

Growth in Local Aggregate Unemployment (16-49yr)							-0.00534**	-0.00541**	-0.0134***
							(0.00216)	(0.00223)	(0.00330)
Low Income x Growth in Aggregate Local Unemployment (16-49yr)								0.000235	-0.0000732
								(0.00368)	(0.00366)
Female x Growth in Aggregate Local Unemployment (16-49yr)									0.0137***
									(0.00395)
Observations	181,090	181,090	181,090	196,355	196,355	196,355	196,355	196,355	196,355
R-squared	0.104	0.104	0.104	0.106	0.106	0.106	0.106	0.106	0.107
Year FE	✓	✓	✓	✓	✓	✓	✓	✓	✓
University FE	✓	✓	✓	✓	✓	✓	✓	✓	✓
Region FE (NUTS 1 2010)	✓	✓	✓	✓	✓	✓	✓	✓	✓
Clustered SE	✓	✓	✓	✓	✓	✓	✓	✓	✓

Standard errors shown in the parentheses are clustered at the institutional level across all 9 specifications. All regressions are performed using OLS to estimate equation 9. The dependent variable is the binary subject group indicator, which takes the value 1 for STEM and 0 otherwise. All 9 specifications are estimating column (3) from Table 2.4 with various indicators of unemployment, and various interactions of the unemployment measures with the socioeconomic controls. Specifications (1)-(3) estimate the effect of local youth unemployment on the probability of a STEM subject being studied, whilst specifications (4)-(6) and (7)-(9) estimate the effect of aggregate and the growth of aggregate local unemployment respectively. Interactions are included between the unemployment variable and the socioeconomic indicators of low income and female, to allow the unemployment effect to be heterogeneous across these binary categories. The socioeconomic controls as included in specification (3) of Table 2.4 are included, but the results are supressed here (except those controls that are being interacted with the unemployment measures) to focus on the effect of the unemployment variables. The regression results including the socioeconomic controls can be found in Appendix B, Table B6. Raw numbers have been rounded to the nearest multiple of 5 to comply with HESA's Standard Rounding Methodology. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 2.9: LEM Subjects and Unemployment (Socioeconomic Heterogeneity)

	(1) LEM	(2) LEM	(3) LEM	(4) LEM	(5) LEM	(6) LEM	(7) LEM	(8) LEM	(9) LEM
Female	-0.0653***	-0.0653***	-0.0580***	-0.0637***	-0.0637***	-0.0571***	-0.0637***	-0.0637***	-0.0615***
	(0.00512)	(0.00512)	(0.0105)	(0.00501)	(0.00501)	(0.00941)	(0.00500)	(0.00500)	(0.00516)
Low Income	-0.000734	0.0131	0.0127	-0.00197	0.00998	0.00960	-0.00199	-0.00193	-0.00184
	(0.00529)	(0.0107)	(0.0107)	(0.00497)	(0.00922)	(0.00927)	(0.00497)	(0.00502)	(0.00500)
Low Income x Female	0.00293	0.00298	0.00326	0.00285	0.00290	0.00316	0.00289	0.00289	0.00274
	(0.00539)	(0.00538)	(0.00532)	(0.00533)	(0.00533)	(0.00528)	(0.00533)	(0.00533)	(0.00533)
Local Youth Unemployment (16-24yr)	0.000216	0.000534	0.000868						
	(0.000788)	(0.000795)	(0.000829)						
Low Income x Local Youth Unemployment (16-24yr)		-0.00109	-0.00107						
		(0.000746)	(0.000748)						
Female x Local Youth Unemployment (16-24yr)			-0.000570						
			(0.000691)						
Aggregate Local Unemployment (16-49yr)				0.000420	0.00114	0.00191			
				(0.00181)	(0.00190)	(0.00196)			
Low Income x Aggregate Local Unemployment (16-49yr)					-0.00242	-0.00238			
					(0.00170)	(0.00171)			
Female x Aggregate Local Unemployment (16-49yr)						-0.00131			
						(0.00158)			

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Growth in Local Aggregate Unemployment (16-49yr)							-0.00418**	-0.00410*	0.00237
							(0.00205)	(0.00207)	(0.00326)
Low Income x Growth in Aggregate Local Unemployment (16-49yr)								-0.000295	-0.0000474
								(0.00356)	(0.00354)
Female x Growth in Aggregate Local Unemployment (16-49yr)									-0.0110***
									(0.00408)
Observations	181,090	181,090	181,090	196,355	196,355	196,355	196,355	196,355	196,355
R-squared	0.063	0.063	0.063	0.061	0.061	0.061	0.061	0.061	0.061
Year FE	✓	✓	✓	✓	✓	✓	✓	✓	✓
University FE	✓	✓	✓	✓	✓	✓	✓	✓	✓
Region FE (NUTS 1 2010)	✓	✓	✓	✓	✓	✓	✓	✓	✓
Clustered SE	✓	✓	✓	✓	✓	✓	✓	✓	✓

Standard errors shown in the parentheses are clustered at the institutional level across all 9 specifications. All regressions are performed using OLS to estimate equation 9. The dependent variable is the binary subject group indicator, which takes the value 1 for LEM and 0 otherwise. All 9 specifications are estimating column (6) from Table 2.4 with various indicators of unemployment, and various interactions of the unemployment measures with the socioeconomic controls. Specifications (1)-(3) estimate the effect of local youth unemployment on the probability of a LEM subject being studied, whilst specifications (4)-(6) and (7)-(9) estimate the effect of aggregate and the growth of aggregate local unemployment respectively. Interactions are included between the unemployment variable and the socioeconomic indicators of low income and female, to allow the unemployment effect to be heterogeneous across these binary categories. The socioeconomic controls as included in specification (6) of Table 2.4 are included, but the results are supressed here (except those controls that are being interacted with the unemployment measures) to focus on the effect of the unemployment variables. The regression results including the socioeconomic controls can be found in Appendix B, Table B7. Raw numbers have been rounded to the nearest multiple of 5 to comply with HESA's Standard Rounding Methodology. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 2.10: OTHER Subjects and Unemployment (Socioeconomic Heterogeneity)

	(1) OTHER	(2) OTHER	(3) OTHER	(4) OTHER	(5) OTHER	(6) OTHER	(7) OTHER	(8) OTHER	(9) OTHER
Female	0.175***	0.175***	0.181***	0.179***	0.179***	0.189***	0.179***	0.179***	0.179***
	(0.00686)	(0.00686)	(0.0143)	(0.00712)	(0.00712)	(0.0142)	(0.00712)	(0.00711)	(0.00725)
Low Income	-0.0355***	-0.0300**	-0.0303**	-0.0356***	-0.0271**	-0.0277**	-0.0354***	-0.0354***	-0.0354***
	(0.00525)	(0.0128)	(0.0128)	(0.00501)	(0.0111)	(0.0111)	(0.00501)	(0.00506)	(0.00506)
Low Income x Female	0.0227***	0.0227***	0.0229***	0.0230***	0.0230***	0.0234***	0.0230***	0.0230***	0.0229***
	(0.00595)	(0.00595)	(0.00591)	(0.00575)	(0.00575)	(0.00570)	(0.00575)	(0.00575)	(0.00575)
Local Youth Unemployment (16-24yr)	0.000863	0.000991	0.00125						
	(0.000651)	(0.000763)	(0.000975)						
Low Income x Local Youth Unemployment (16-24yr)		-0.000437	-0.000423						
		(0.000892)	(0.000892)						
Female x Local Youth Unemployment (16-24yr)			-0.000451						
			(0.000914)						
Aggregate Local Unemployment (16-49yr)				0.00260*	0.00310*	0.00433*			
				(0.00148)	(0.00175)	(0.00225)			
Low Income x Aggregate Local Unemployment (16-49yr)					-0.00172	-0.00165			
					(0.00201)	(0.00201)			
Female x Aggregate Local Unemployment (16-49yr)						-0.00209			

						(0.00209)			
Growth in Local Aggregate Unemployment (16-49yr)							0.00952***	0.00951***	0.0111***
							(0.00241)	(0.00253)	(0.00367)
Low Income x Growth in Aggregate Local Unemployment (16-49yr)								0.0000604	0.000121
								(0.00460)	(0.00460)
Female x Growth in Aggregate Local Unemployment (16-49yr)									-0.00268
									(0.00444)
Observations	181,090	181,090	181,090	196,355	196,355	196,355	196,355	196,355	196,355
R-squared	0.122	0.122	0.122	0.122	0.122	0.122	0.122	0.122	0.122
Year FE	✓	✓	✓	✓	✓	✓	✓	✓	✓
University FE	✓	✓	✓	✓	✓	✓	✓	✓	✓
Region FE (NUTS 1 2010)	✓	✓	✓	✓	✓	✓	✓	✓	✓
Clustered SE	✓	✓	✓	✓	✓	✓	✓	✓	✓

Standard errors shown in the parentheses are clustered at the institutional level across all 9 specifications. All regressions are performed using OLS to estimate equation 9. The dependent variable is the binary subject group indicator, which takes the value 1 for OTHER and 0 otherwise. All 9 specifications are estimating column (9) from Table 2.4 with various indicators of unemployment, and various interactions of the unemployment measures with the socioeconomic controls. Specifications (1)-(3) estimate the effect of local youth unemployment on the probability of an OTHER subject being studied, whilst specifications (4)-(6) and (7)-(9) estimate the effect of aggregate and the growth of aggregate local unemployment respectively. Interactions are included between the unemployment variable and the socioeconomic indicators of low income and female, to allow the unemployment effect to be heterogeneous across these binary categories. The socioeconomic controls as included in specification (9) of Table 2.4 are included, but the results are supressed here (except those controls that are being interacted with the unemployment measures) to focus on the effect of the unemployment variables. The regression results including the socioeconomic controls can be found in Appendix B, Table B8. Raw numbers have been rounded to the nearest multiple of 5 to comply with HESA's Standard Rounding Methodology. *** p<0.01, ** p<0.05, * p<0.1

Table 2.11: Baseline Model of Subject Choice (Multinomial Logit & Marginal Effects)

	Log-odds		Average Marginal Effects		
	(1)	(2)	(3)	(4)	(5)
	STEM	LEM	STEM	LEM	OTHER
Tariff	-0.00706	0.0285***	-0.00215**	0.00454***	-0.00240**
	(0.00686)	(0.00655)	(0.000953)	(0.000964)	(0.00107)
Female	-1.010***	-0.685***	-0.120***	-0.0633***	0.183***
	(0.0359)	(0.0348)	(0.00476)	(0.00487)	(0.00617)
White	-0.659***	-1.070***	-0.0422***	-0.153***	0.195***
	(0.0646)	(0.0523)	(0.00812)	(0.00788)	(0.0113)
British	0.0284	-0.0874	0.00755	-0.0145	0.00700
	(0.152)	(0.131)	(0.0159)	(0.0146)	(0.0293)
Low Income	0.154***	0.0476***	0.0201***	0.000774	-0.0209***
	(0.0191)	(0.0179)	(0.00277)	(0.00269)	(0.00294)
Local	0.192***	0.248***	0.0169***	0.0302***	-0.0471***
	(0.0354)	(0.0317)	(0.00526)	(0.00518)	(0.00525)
Disabled	0.0290	-0.276***	0.0147***	-0.0397***	0.0250***
	(0.0299)	(0.0315)	(0.00423)	(0.00394)	(0.00496)
Constant	-1.118***	0.396**			
	(0.214)	(0.164)			
Observations	198,280	198,280			
R-squared	0.106	0.061			
Year FE	✓	✓	✓	✓	✓
University FE	✓	✓	✓	✓	✓
Region FE (NUTS 1 2010)	✓	✓	✓	✓	✓
Clustered SE	✓	✓	✓	✓	✓

Standard errors shown in the parentheses are clustered at the institutional level. Raw numbers have been rounded to the nearest multiple of 5 to comply with HESA's Standard Rounding Methodology. Columns (1) to (5) from the output of one single regression estimation, which is the multinomial logit specification of the OLS model estimated in Table 2.4. Specifically, the dependent variable is the categorical subject choice indicator (Y=1(STEM), =2(LEM), =3(OTHER)) and as such the estimates presented in Table 2.11 are equivalent to the model estimated in Table 2.4, Columns (2), (5) and (8) where the interactions are not included between the socioeconomic indicator explanatory variables, but do include a full set of year, institutional and regional fixed effects. Columns (1) and (2) report the log odds where the base category is OTHER, whilst columns (3) to (5) report the average marginal effects of the explanatory variables on the probability that the subject category is equal to STEM (Column 3), LEM (Column 4) or OTHER (Column 5). *** p<0.01, ** p<0.05, * p<0.1

Table 2.12: STEM Subjects and Local Unemployment (Marginal Effects)

	(1) STEM	(2) STEM	(3) STEM	(4) STEM	(5) STEM	(6) STEM
Local Youth Unemployment (16-24yr)	-0.00116*					
	(0.000595)					
Local Adult Unemployment (25-49yr)		-0.00235				
		(0.00159)				
Growth in Local Youth Unemployment (16-24yr)			-0.0100			
			(0.00674)			
Growth in Local Adult Unemployment (25-49yr)				-0.00915		
				(0.00841)		
Aggregate Local Unemployment (16-49yr)					-0.00305**	
					(0.00131)	
Growth in Local Aggregate Unemployment (16-49yr)						-0.00571***
						(0.00210)
Observations	181,090	187,825	158,720	178,995	196,355	196,355
Year FE	✓	✓	✓	✓	✓	✓
University FE	✓	✓	✓	✓	✓	✓
Region FE (NUTS 1 2010)	✓	✓	✓	✓	✓	✓
Clustered SE	✓	✓	✓	✓	✓	✓

Standard errors shown in the parentheses are clustered at the institutional level across all 6 specifications. All regressions are performed using a multinomial logit specification to estimate equation 9. The dependent variable is the categorical subject group indicator, which takes the value 1 for STEM, 2 for LEM and 3 for OTHER. All 6 specifications are estimating column (2) from Table 2.4 with various indicators of unemployment. Specifications (1)-(6) estimate the unemployment variables separately, and correspond to columns (1)-(6) of Table 2.5. The coefficients shown are the average marginal effects of the unemployment measures on the probability of the categorical subject outcome variable being equal to STEM. The socioeconomic controls as included in specification (2) of Table 2.4 are included, but the results are supressed here to focus on the effect of the unemployment variables. The regression results including the socioeconomic controls are available on request from the author. Raw numbers have been rounded to the nearest multiple of 5 to comply with HESA's Standard Rounding Methodology. *** $p<0.01$, ** $p<0.05$, * $p<0.1$*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 2.13: LEM Subjects and Local Unemployment (Marginal Effects)

	(1) LEM	(2) LEM	(3) LEM	(4) LEM	(5) LEM	(6) LEM
Local Youth Unemployment (16-24yr)	0.000199					
	(0.000741)					
Local Adult Unemployment (25-49yr)		-0.00150				
		(0.00187)				
Growth in Local Youth Unemployment (16-24yr)			-0.0213***			
			(0.00700)			
Growth in Local Adult Unemployment (25-49yr)				-0.00449		
				(0.00767)		
Aggregate Local Unemployment (16-49yr)					0.000428	
					(0.00167)	
Growth in Local Aggregate Unemployment (16-49yr)						-0.00412**
						(0.00204)
Observations	181,090	187,825	158,720	178,995	196,355	196,355
Year FE	✓	✓	✓	✓	✓	✓
University FE	✓	✓	✓	✓	✓	✓
Region FE (NUTS 1 2010)	✓	✓	✓	✓	✓	✓
Clustered SE	✓	✓	✓	✓	✓	✓

Standard errors shown in the parentheses are clustered at the institutional level across all 6 specifications. All regressions are performed using a multinomial logit specification to estimate equation 9. The dependent variable is the categorical subject group indicator, which takes the value 1 for STEM, 2 for LEM and 3 for OTHER. All 6 specifications are estimating column (5) from Table 2.4 with various indicators of unemployment. Specifications (1)-(6) estimate the unemployment variables separately, and correspond to columns (1)-(6) of Table 2.6. The coefficients shown are the average marginal effects of the unemployment measures on the probability of the categorical subject outcome variable being equal to LEM. The socioeconomic controls as included in specification (5) of Table 2.4 are included, but the results are supressed here to focus on the effect of the unemployment variables. The regression results including the socioeconomic controls are available on request from the author. Raw numbers have been rounded to the nearest multiple of 5 to comply with HESA's Standard Rounding Methodology. *** p<0.01, ** p<0.05, * p<0.1*** p<0.01, ** p<0.05, * p<0.1

Table 2.14: OTHER Subjects and Local Unemployment (Marginal Effects)

	(1) OTHER	(2) OTHER	(3) OTHER	(4) OTHER	(5) OTHER	(6) OTHER
Local Youth Unemployment (16-24yr)	0.000965					
	(0.000626)					
Local Adult Unemployment (25-49yr)		0.00386**				
		(0.00179)				
Growth in Local Youth Unemployment (16-24yr)			0.0313***			
			(0.00843)			
Growth in Local Adult Unemployment (25-49yr)				0.0136		
				(0.00913)		
Aggregate Local Unemployment (16-49yr)					0.00263*	
					(0.00143)	
Growth in Local Aggregate Unemployment (16-49yr)						0.00982***
						(0.00238)
Observations	181,090	187,825	158,720	178,995	196,355	196,355
Year FE	✓	✓	✓	✓	✓	✓
University FE	✓	✓	✓	✓	✓	✓
Region FE (NUTS 1 2010)	✓	✓	✓	✓	✓	✓
Clustered SE	✓	✓	✓	✓	✓	✓

Standard errors shown in the parentheses are clustered at the institutional level across all 6 specifications. All regressions are performed using a multinomial logit specification to estimate equation 9. The dependent variable is the categorical subject group indicator, which takes the value 1 for STEM, 2 for LEM and 3 for OTHER. All 6 specifications are estimating column (8) from Table 2.4 with various indicators of unemployment. Specifications (1)-(6) estimate the unemployment variables separately, and correspond to columns (1)-(6) of Table 2.5. The coefficients shown are the average marginal effects of the unemployment measures on the probability of the categorical subject outcome variable being equal to OTHER. The socioeconomic controls as included in specification (8) of Table 2.4 are included, but the results are supressed here to focus on the effect of the unemployment variables. The regression results including the socioeconomic controls are available on request from the author. Raw numbers have been rounded to the nearest multiple of 5 to comply with HESA's Standard Rounding Methodology. *** p<0.01, ** p<0.05, * p<0.1*** p<0.01, ** p<0.05, * p<0.1

2.6 Conclusions & Discussion

This chapter has attempted to establish whether labour market influences have any effect on the subject choice of school-leavers in the UK. A linear estimation approach was adopted for comparative and computational ease over non-linear models, and specifications of subject choice were created for three main subject groups, both with and without the effect of unemployment. The effect of unemployment was then allowed to be heterogeneous across different types of students. The results show that unemployment is an important factor for establishing how students decide what subject to study at university, and this excludes the impact on the demand for a subject which is not possible to estimate with the data used. Two channels were identified for affecting the decision of which subject to study: the changing of the opportunity costs of HE, and the changing of the expected benefits. These were hypothesised to be the contemporaneous local youth unemployment rate averaged over the 12 month period prior to the student's starting date, and the growth in local unemployment rates respectively.

Whilst both of these effects were present in the data and significant in affecting the probability of observing a student studying a particular subject type, the sign of these effects was of the opposite direction than that was hypothesised. The higher was the youth unemployment rate in the local area, the higher the probability of observing a student studying a LEM subject, which implies that students who experience a lowering of the opportunity cost of studying are more likely to choose a subject with higher potential rewards in terms of wages and employment probability. This contradicted the original hypothesis, however it implies that students may engage in more forward-looking investment behaviour than originally proposed.

Furthermore, students who were either female or the most likely to be socioeconomically disadvantaged were both more likely to choose STEM subjects, conditional on having decided to participate in HE. This further proposes evidence that certain students may be aware of their future disadvantages in the labour market (as Delaney et al. (2011) propose), and that this may be an attempt to compensate for future labour market disadvantages, especially in the context of rising unemployment. Given that low income students may be the most aware of the costs of education, they may also be the most aware of the benefits of choosing a degree subject with larger graduate wage and employment premia relative to their peers.

However, this analysis is performed under the simplifying assumption that it is the average local unemployment rate that students respond to, and signals from either the average local unemployment rate for low-skilled workers or high-skilled workers do not have an additional, heterogeneous effect. This simplifying assumption may be somewhat problematic if it is the case that students from low-skilled backgrounds respond more to the low-skilled labour market trends rather than the average, as it would imply that study choices are motivated by the skill-heterogeneity in the local unemployment rate. Whilst it is not the focus of this chapter, the local labour market signals that university entrants from low-skilled, low-income backgrounds respond to in terms of study choices presents an important avenue for further research. The findings are also conditional on the student having participated in higher education, since the analysis is performed on graduate data.

Nevertheless, the results suggest that some students – particularly those who experience significantly high growth in aggregate unemployment and who are not from a disadvantaged background – are more likely to study a subject with lower graduate employment and wage premia, but that may have a higher chance of success and higher degree attainment. This points towards a potential of students engaging in risk aversion: an attempt to increase the certainty of obtaining a degree but with less rewards. Thus, what is observed is arguably evidence in support of the hypothesis of Callender and Jackson (2008). The findings of this research also represent an important contribution to the empirical literature, with a longer time period of analysis than Wales (2010), and the strategy of comparing subject types which allows the impact of local unemployment to be directly comparable between high and low reward subjects in economic terms.

Chapter 3

'Disadvantaged' Students and University Academic Performance

"Young people from the top socio-economic quintile group remain ... 3.7 percentage points more likely to graduate with a first or a 2:1 than young people from the bottom socio-economic quintile group."

(Crawford, 2014a: 20)

"In other words, those students from private schools…were much less likely to achieve at least a 2:1 compared to equivalently qualified state school pupils, all else equal."

(Lasselle et al., 2014: 308)

3.1 Introduction

These seemingly contradictory quotes serve to highlight the complex processes that determine a student's undergraduate academic performance at university. Whilst Crawford (2014a) highlights the penalty incurred by lower socioeconomic status (SES) students in terms of degree attainment, Lasselle et al. (2014) suggest that students from private schools[115] are less likely to graduate with a 2:1 or above. Although 33% of private school students received some means-tested assistance

[115] In the UK, primary (aged 4-11) and secondary education (aged 11-18) is provided free by the state, or parents may elect to enrol their child in an independent (also known as a private) school where parents must pay a tuition fee per school term (there are normally 2 terms to each academic year). Private schools in 2011 accounted for roughly 6.5% of the UK school-age population, approximately 625,000 pupils (Independent School Council Annual Census 2011). Some students are eligible for discounted or free tuition at private schools through means-tested family income and/or child performance on entrance examinations. In 2017, the average term tuition fee for secondary education was approximately £4,800, and for primary education the per term tuition cost was approximately £4,200. If the private schools are boarding schools where pupils reside at the school during term time, the average costs are considerably higher - £11,000 and £8,000 per term respectively. (Independent School Council Annual Census 2017).

with school fees[116], the majority of students who attend private school are inherently more likely to be from a family of higher socioeconomic status and income background. The aim of this chapter is to empirically test whether students with socioeconomic disadvantages are more, less or no more likely to obtain the best degree outcomes, all other things being equal.

Whilst access to university study is crucial for social mobility, the extent of success of a student's engagement with higher education (HE) is not solely determined by participation alone. Previous research has found that, 1) students from poorer backgrounds are inherently less-inclined to participate in undergraduate study[117], and 2) even if they do participate, risk aversion may lead to sub-optimal course choices at university[118], and 3) moreover students from poorer backgrounds are more likely to drop-out of university study[119]; however there is little attention paid to the effects of student's socioeconomic backgrounds on university performance. The work of Lasselle et al. (2014) is arguably the closest to the aims of this research, as they investigate whether overall school performance is useful for predicting a student's university attainment. In researching this question, they employ a non-linear regression model with a binary dependent variable of degree success, explained by pre-university attainment, and socioeconomic controls. This research significantly improves upon the existing literature by using a much more comprehensive research design and by using a rich dataset for analysis.

One of the major improvements over the work of Lasselle et al. (2014) is that this analysis will be performed for using data containing the entire population of students who started their undergraduate studies during the academic years 2003/04 and 2007/08. Consequently, not only does this offer two additional cohorts of data compared to Lasselle et al (2014), but the effects of socioeconomic disadvantage on university academic attainment are estimated across all Higher Education

[116] Independent Schools Council Annual Report 2017, Figures 26 & 28.
[117] e.g. Chowdry et al. (2013).
[118] See section 2.2.1 of this thesis for further discussion.
[119] e.g. Powdthavee & Vignoles (2009); Johnes & McNabb (2004).

Institutions (HEIs) throughout the United Kingdom (UK). This is opposed to the work of Lasselle et al. (2014) who focus only on three cohorts totalling 1,300 students from one institution – the University of St. Andrews. Furthermore, the data that is used in this analysis contains detailed socioeconomic indicators such as disability status, gender, age, ethnicity, whether the student has a parent or guardian with university education. The only socioeconomic data used in Lasselle et al. (2014) are age and gender. Finally, this research design controls for significant variation in degree attainment by subject by using subject fixed effects with 20 subject categories, whereas Lasselle et al. (2014) only control for faculty fixed effects with 3 categories.

The results show that students that experience socioeconomic disadvantages before entering HE are less likely to obtain a good degree at university, controlling for observable determinants of degree success and conditional on attainment pre-university. The results further show that this is predominantly the case, and larger in magnitude for students who enter HE with the lowest levels of pre-university attainment. Furthermore, students with the most significant socioeconomic disadvantages are found to outperform their peers in terms of university attainment. This suggests that for those students who are the most disadvantaged, their level of pre-university attainment is an imperfect indicator of true ability at university. The results of this research therefore find evidence strongly in favour of the use of contextual admissions to HE, where the decision to offer a prospective student entry onto a course depends on their school grades as well as their socioeconomic characteristics.

The chapter is structured as follows: section 3.1 outlines the determinants of degree outcomes, as well as presenting hypotheses for why disadvantaged may affect a student's university attainment; section 3.2 discusses the institutional aspects of measuring widening participation and to what extent degree outcomes matter for social mobility; section 3.3 presents an overview of the data and identification strategy; section 3.4 outlines the methodological considerations and approaches;

section 3.5 presents the results, and section 3.6 concludes with some discussion of the results in the context of policy implications.

3.1.1 Factors Affecting Degree Outcomes

Before presenting an overview of how socioeconomic factors may affect academic performance at university, it is useful to consider perhaps the more obvious factors that may affect degree attainment. It is a relatively robust finding that academic performance at school or college is a significant predictor of academic performance at university. It therefore follows, that all things being equal, students with higher pre-university attainment should be expected to obtain better degree outcomes. A range of studies have either researched the effect of pre-university attainment on degree outcomes directly, or as part of an investigation of other determinants of degree success. In particular, Smith and Naylor (2001a) use UK data from the 1993 cohort of university leavers and find that a 1-grade increase in the A-Level[120] grade of a student increases the probability of obtaining a 'good'[121] degree (2:1 or above) by 9 percentage points. Similarly positive effects of pre-university scores on degree outcomes are found in a descriptive study by HEFCE (2014), where 80% of students who obtain AAB[122] at A-Level obtain a 2:1 or above, compared to only 50% who enter university with CCC or below.

In addition to this finding being confirmed in studies estimating the determinants of degree outcomes (Ardila, 2001; Naylor and Smith, 2004; Koh & Koh, 1999), both McNabb et al. (2002) and Barrow et al. (2009) find pre-university education scores as a positive determinant of higher degree outcomes. Naylor and Smith (2004) in particular find that for Economics graduates in the UK who commenced

[120] An outline of the UK higher education system and entrance qualifications can be found in the Introduction to this thesis. In this chapter, see also section 3.2.

[121] A 'good' degree is widely acknowledged as being an upper second class honours or first. See section 3.2 of this chapter.

[122] During the period of analysis, A-Level grades ranged from A to E, where A=80% or above, B=70%, C=60%, D=50%, E=40%. For international comparisons, the A-Level is similar to the International Baccalaureate (IB) and Advanced Placement (AP) in the United States. As previously outlined, students in England and Wales typically apply to university having completed 3 full A-Level qualifications, hence the three letter grade reporting (e.g. AAB, CCC).

their study between 1984/85 and 1992/93, a 1 unit grade increase in a student's A-Level results increases the probability of obtaining a good degree by 2 percentage points, approximately a 4% increase relative to the mean.[123] A similar finding is reported by Powdthavee and Vignoles (2009) in estimating the effect of socioeconomic background on university attrition rates. However, despite this seemingly robust finding, there is some evidence (Boyle et al., 2002; Gist et al., 1996; Bartlett et al., 1993) to suggest that at the individual subject and course level, pre-university attainment in certain school or college subjects does not significantly influence the degree outcome in a particular subject.

Gender is also robustly found as a determinant of degree outcomes, where females are more likely to obtain a 2:1 or above (Barrow et al., 2009; Smith & Naylor, 2001a; HEFCE, 2014). However, McNabb et al. (2002) find that whilst females are more likely to obtain a 2:1 or above, females are less likely than their male counterparts to receive a 1st class degree. Gender may influence degree outcomes through the choice of subjects, gender-specific differences in other determinants of degree outcomes, or through biological differences. The latter is explored by Mellanby et al. (2000) who suggest the gender gap in degree outcomes may be explained by biological differences in educational and emotional intelligence. However, this may be further confounded by socioeconomic differences if female students are expected to be interested in certain subjects, and this expectation becomes self-fulfilling.

Age and ethnicity are also found to be significant in modelling degree outcomes. It is consistently found that ethnic minorities perform worse than 'White British' students ceteris paribus (Connor et al., 2004; Barrow et al., 2009; HEFCE, 2010; 2015b), even when allowing for the lower pre-university attainment of ethnic minorities (Broecke & Nicholls, 2007). Age, however, has mixed evidence in the

[123] As will be discussed in section 3.4.3, although this analysis uses tariff scores to measure pre-university attainment, they will be rescaled such that a 1 unit increase in the tariff score is equal to a 1 grade change at A-Level.

empirical literature. Whilst some research identifies older students as being more likely to obtain better degree outcomes (HEFCE, 2015b; Smithers & Griffin, 1986; Osborne et al., 1997), other studies show that – similar to the effect of pre-university attainment – there may be heterogeneous effects of age. Both Bartlett et al. (1993) and Koh and Koh (1999) for example find that older students are less likely to obtain a 2:1 or above for accountancy degrees, which gives additional evidence to the existence of subject-specific differences in the determinants of degree outcomes.

Two further factors that have been found to affect degree outcomes are income background and school quality. HEFCE (2014; 2015a) show that, conditional on pre-university attainment, students from lower income backgrounds or from regions with the lowest HE participation obtain significantly poorer degree outcomes.[124] This effect is also found based on parental occupation (Blundell et al., 2000; McNabb et al., 2002), where in particular Smith and Naylor (2001a) estimate that 'poorer' students are 13-15 percentage points less likely to obtain a 2:1 or above. Income background is potentially correlated with school quality, and almost certainly correlated with whether a student attends an independent or private school. Nevertheless, there is little evidence of school quality and type on degree outcomes (Eide & Showalter, 1998; Dearden et al., 2002) and conditional on students' pre-university attainment, students from independent and private schools are in fact less likely to obtain a 2:1 or above compared to state school students (HEFCE, 2014; Crawford, 2014b). As Smith and Naylor (2001a) note after finding the same effect, two students who have the same pre-university attainment but from state and private schools are likely to be drawn from different ability distributions.

Finally, considerations should also be paid to modes of study (part time students are less likely to get a 2:1 or above (HEFCE, 2015b)), disability status (disabled students are less likely to get a 2:1 or above (HEFCE, 2013; 2015b)), and the choice

[124] See Murphy et al. (2017) for a historical overview of this relationship in England.

of subjects and universities (Crawford, 2014a) which may all have a deterministic impact on a student's degree outcome. Ultimately the determinants of degree outcomes contain a significant amount of overlap, which has implications for empirical estimation.

3.1.2 Why a Disadvantaged Background May (Not) Affect Degree Outcomes

Two students, A and B, are about to enter higher education in the same year and have the same level of pre-university attainment,[125] who choose to go to the same university to study the same degree course, and are the same age, gender, ethnicity and nationality. The research question of interest is: if student A has a disadvantaged background relative to student B, does this have a zero, a positive, or a negative effect on student A's degree outcome, all else being equal? A student from a potentially disadvantaged background will be hereafter classed as a student from the lower end of the socioeconomic classification[126], whilst an additional indicator for being potentially disadvantaged that will be used is a student who does not have a parent or guardian with university-level education.[127]

It is inherently assumed that coming from a disadvantaged background should not have a further penalty on, or benefit for, degree outcomes. However this may not be the case, especially if students from disadvantaged backgrounds do not have the same access to high quality institutions as higher SES students.[128] Nevertheless, whilst students who come from disadvantaged backgrounds tend to have a higher concentration of the characteristics outlined in the previous section that have a negative impact on university attainment, if these compounded effects are taken

---

[125] Ideally we would like to be able to say the exact same grades in the same subject using the same pre-university qualification, however this is not possible in the data.
[126] See section 3.3.2 for a full identification.
[127] Parental education level is a widely-used measure of socioeconomic status, due to the correlation between education level and household income (see for example, the effect of schooling on earnings; Devereux and Hart, 2010), and due to the intergenerational transmission of human capital (e.g. Chevalier, 2003; Delaney et al., 2011). Thus a student who is both classed as being low SES and is known to not have a parent with university level education is potentially the most socioeconomically disadvantaged.
[128] In particular, the Universities of Oxford and Cambridge have been criticised for failing to admit sufficient students from the poorest households. (See Adams (2017) for an example).

into account, simply coming from a poorer household in income terms should not have any deterministic effect on that student's degree classification, all else being equal. If this is not the case, then the returns to a university degree not only depend on the subject, the university, and the classification, but also on the social and familial characteristics of the student who obtains the degree. University would not therefore be a 'levelling of the playing field'[129] with respect to the potential for academic achievement, conditional on ability.

Alternatively, it may be the case that – using the two exemplar students above – student A has attained the same pre-school qualifications and grades, and is attending the same university degree course, having overcome more barriers to study than student B due to the disadvantaged background of the student. Such barriers may include financial hardship and lower historical rates of HE participation (both familial and spatially). The students that do overcome those barriers may be more likely to succeed in their degree studies, conditional on attainment, since the value of the returns to the degree relative to their familial socioeconomic characteristics may elicit more effort and a determination to succeed. This may especially be the case if for those students, degree success is perceived to be fundamental to social mobility. Ultimately, for a low SES student to achieve a better degree outcome as a high SES student conditional on the same pre-university attainment, they need to be either: 1) of higher ability such that previous attainment is not a true reflection of ability and thus is a poor predictor of future attainment; or 2) the low SES student must have worked harder for the same grade, such that ability may be comparable, but unobserved components related to ability such as motivation and determination are higher for the low SES student.

If there is a positive effect between socioeconomic background alone and degree outcomes, we should expect students from poorer backgrounds to be more likely to

[129] If universities provide sufficient support, student funding schemes are such that poorer students are not financially constrained in attending university, and students are proactive in seeking support where necessary, university attainment should be a reflection on ability and effort, rather than socioeconomic disadvantage.

attain a 2:1 or 1st in the degree studies, after controlling for all other determinants of degree classification. This scenario would support the finding that students from worse-performing schools are more likely to graduate with a good degree than those from better-performing schools, conditional on attainment[130]; which may reflect that students from disadvantaged backgrounds hold more potential than students from better-performing or private/independent schools (Crawford, 2014b). This also alludes to the effect of coming from a private school compared to a state school in Smith and Naylor (2001a). If disadvantaged students are more likely to obtain a 2:1 or above conditional on attainment and controlling for other factors, this would add weight to the argument in favour of contextual admissions policies[131] as noted by Crawford (2014b), and as used in practice as of September 2017 by the Universities of Oxford, Cambridge, St. Andrews, Southampton, Manchester, Edinburgh, Strathclyde, Newcastle, Sheffield and Warwick.[132]

The final scenario is that those barriers experienced by students pre-university continue to affect the student whilst studying for their degree. In the case of financial hardship, this may present a barrier to the student in practical terms, by making the purchase of course materials such as textbooks or paying for university accommodation more difficult, which may impact on university attainment. Furthermore, since students who are from the lower socioeconomic distribution are more likely to be financially constrained or affected by poor financial well-being whilst at university (Pennell & West, 2005; Callender, 2003), students may engage

[130] This finding by Crawford (2014a), which was also confirmed in HEFCE (2014), is further supported by the finding by both authors show evidence of students from worse-performing schools are less likely to drop out than better-performing schools.

[131] A contextual admissions policy is whereby a student's application to a university depends on prior academic achievement, a personal statement, as well as socioeconomic factors, school characteristics and regional indicators. A university's admissions service may use this contextual information to offer a place of study with lower academic entry requirements, to target students who would benefit from university as part of widening access. For more information, see: https://www.ucas.com/advisers/guides-and-resources/adviser-news/news/contextualised-admissions-and-what-it-means-your-students

[132] It is unclear precisely which universities operate a contextual admissions policy and to what extent, but the practice is becoming more common. With the exception of Imperial College London and Lancaster University, all of the top ten universities (as ranked by The Guardian University Guide 2018) declare the use of contextual admissions.

in part time work whilst at university which although may relieve financial pressure may constrain the amount of time or effort available for studying towards assessments. This may be further exacerbated by the extent of any risk aversion to engaging in student loans, since students from poorer backgrounds are more likely to be more averse to debt (Pennell & West, 2005). However, the potentially negative effect of coming from a disadvantaged background on degree outcomes is not necessarily limited to financial causes. As HEFCE (2015a) remarks, the issues and challenges that university students face are complex and varied, thus any degree attainment must require some amount of coping skills, both academically and practically. If students from disadvantaged backgrounds have fewer of these coping skills[133], then this may inhibit the potential for degree success. De Vries and Rentfrow (2016) find that across various personality measures, 'better'[134] personality traits are more likely to be found in students from the higher socioeconomic groups. Furthermore, since 'better' personality traits are linked to better degree outcomes (Parker et al., 2003; Lenton, 2014), this indicates that students from lower socioeconomic groups may be inhibited in their studies relative to higher socioeconomic groups, due to personality traits. This finding is also shown by Poropat (2009) who finds that especially conscientiousness has a positive effect on educational outcomes. If such a student encounters difficulties in their studies, this may compound the likelihood for degree success, since students from the higher socioeconomic distribution are relatively more likely to seek help and advice from academic staff (McNabb et al., 2002).

3.1.3 Other Confounding Issues

Whilst socioeconomic factors may affect degree outcomes, it is unlikely that it is the sole causal effect on a student's engagement with HE. One significant

[133] An alternative hypothesis is that students from lower socioeconomic backgrounds may possess equally good or even greater coping skills than their higher socioeconomic background peers through greater determination and work ethic, given that lower SES students may have encountered more obstacles to overcome in obtaining the necessary grades for university entry. This would therefore mitigate the effect of socioeconomic disadvantage on degree outcomes.

[134] These were found to be extroversion, conscientiousness and openness.

relationship is that between socioeconomic background and student attrition. Powdthavee and Vignoles (2009) find that the drop-out rate is significantly higher for students from lower income backgrounds, even when controlling for other possible influences on the rate of attrition. Johnes and McNabb (2004) also find this effect on the likelihood of drop out, which is strictly monotonic throughout the socioeconomic distribution. Although Crawford (2014a) also finds this effect, it is suggested that this effect may be somewhat explained by differences in human capital. Nevertheless, controlling for the rate of attrition is necessary in the identification of the effect of interest, especially since Smith and Naylor (2001a) propose that this link between socioeconomic background and drop out explains at least in part their finding of socioeconomic background negatively affecting degree classification[135].

Controlling for student attrition is relatively simple, but controlling for the effect of socioeconomic background on the probability of applying to university is more complex. It is a robust finding that students from poorer or otherwise disadvantaged backgrounds are significantly less likely to go to university (Blanden & MacMillan, 2014; Galindo-Rueda et al., 2004; Gayle et al., 2002); with the poorest quintile of students approximately 45 percentage points less likely to attend university than the richest quintile (Chowdry et al., 2013). Further to this, there is also a wide body of evidence to show a significant link between socioeconomic status and subject choice, both at the pre-university and degree subject level (Davies et al., 2008; Anders, 2012; Codiroli McMaster, 2017). Hence it is important to consider that any investigation into the impact of socioeconomic background on degree outcomes may also have these factors influencing the observed outcomes.

[135] As the outcome variable is a student's final degree classification, and the only degree classifications kept for analysis are 'pass' grades, drop-outs are inherently excluded from this analysis.

3.2 Institutional Setting

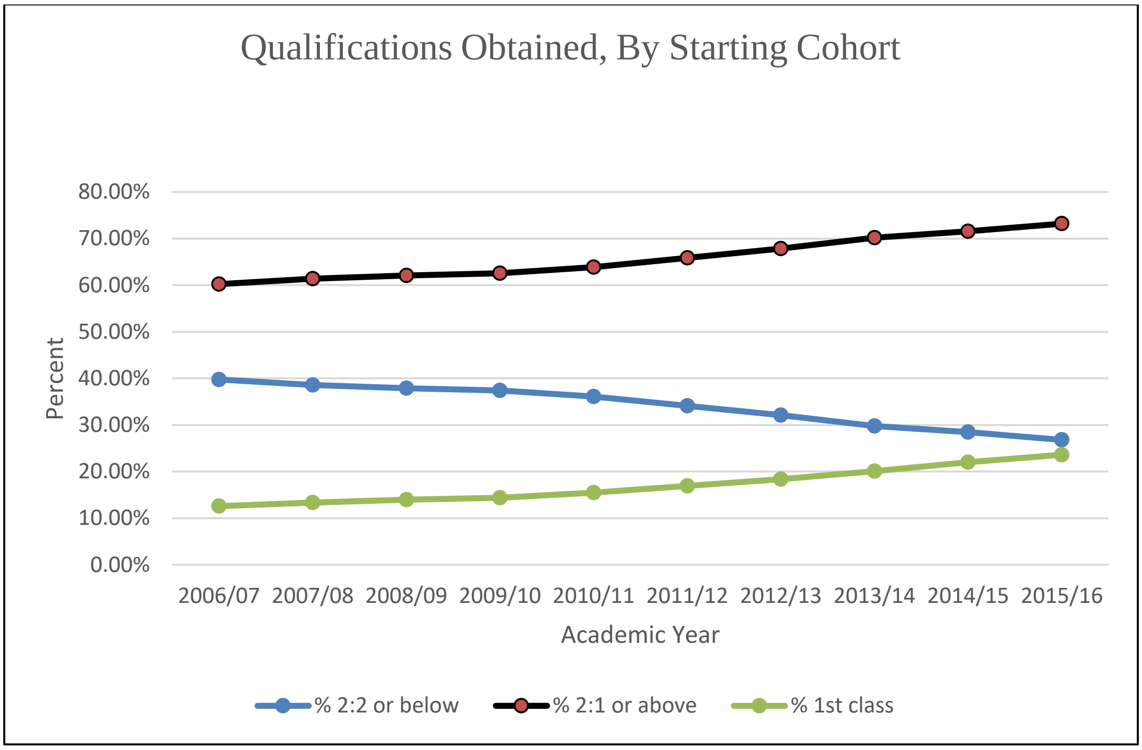
Traditional entry into undergraduate university study in the UK occurs at age 18 where school-leavers (from colleges, schools or 'sixth form[136]') apply through the Universities and Colleges Admissions Service (UCAS) system in January to up to five different courses at any university they wish. Universities receive these applications and decide, based on obtained and projected grades as well as references and a personal statement, whether to offer a place of study in September of the same year. Once attending a university, students will spend between 3-4 years typically to complete their studies, the latter 2 years being counted towards the student's overall degree classification. Whilst each individual university may use a slightly different metric or algorithm for calculating overall degree classification, at the undergraduate level a score of 70% overall is a 1st class degree, 60-69% a 2:1, 50-59% is a 2:2, and 40-49% is a 3rd class.

3.2.1 The Importance of Degree Class

As a higher degree class reflects better academic achievement, it is signal of a quality of graduate, and thus positively correlated with graduate earnings as expected (eg. Naylor et al., 2007). Moreover, obtaining a 1st class degree increases the probability of working in a high wage job by 13 percentage points (Feng & Graetz, 2013), as well as carrying a 9-13% increase in earnings compared to a 2:1 classification on average (Battu et al, 1999).

[136] Sixth form refers to the final stages of secondary education which is most commonly takes place in a traditional schooling structure and lasts two years. It is also known as Years 12 and 13 or Lower and Upper Sixth (in England, Northern Ireland and Wales). In Scotland, Sixth Form is usually known as Sixth Year, or S6, which only lasts one year.

Figure 3.1: Qualifications Obtained, by Starting Cohort



Source: HESA

As seen from Figure 3.1[137], there has been a continual and consistent increased in the proportion of ‘good ‘degrees being awarded, especially driven by the increase in 1st class degrees. As a result of the increasing proportion of good degrees awarded, and a significant increase in HE participation, in 2012 over 75% of graduate vacancies required at least a 2:1, compared to 52% in 2004.[138]

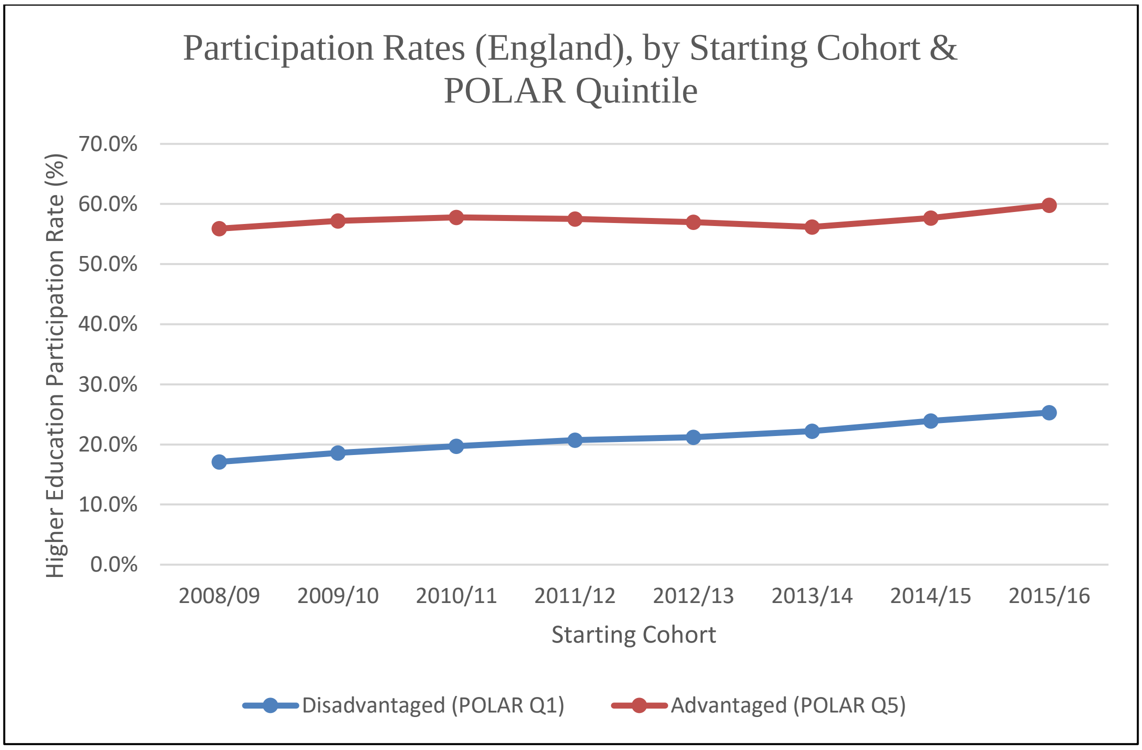
3.2.2 Widening Participation

Whilst overall participation in HE has steadily increased, this increase must also be the case for disadvantaged students to allow social mobility. Figure 3.2 shows that participation by students from the lowest quintile of participation areas (POLAR

[137] Note that the percentage of 1st degrees awarded is a subset of the percentage of 2:1 or above, hence the sum of all three categories do not sum to 100%.
[138] See the Introduction to this thesis for further discussion on the trends of the participation in higher education in the UK.

Q1[139]) has increased from around 17% to over 25%, which compared to the relatively stable participation rate for the most advantaged students (Q5), indicates the participation gap is reducing. Specifically the gap has reduced from 40 percentage points to 35 percentage points, which is a 17.5% decrease in 7 years.

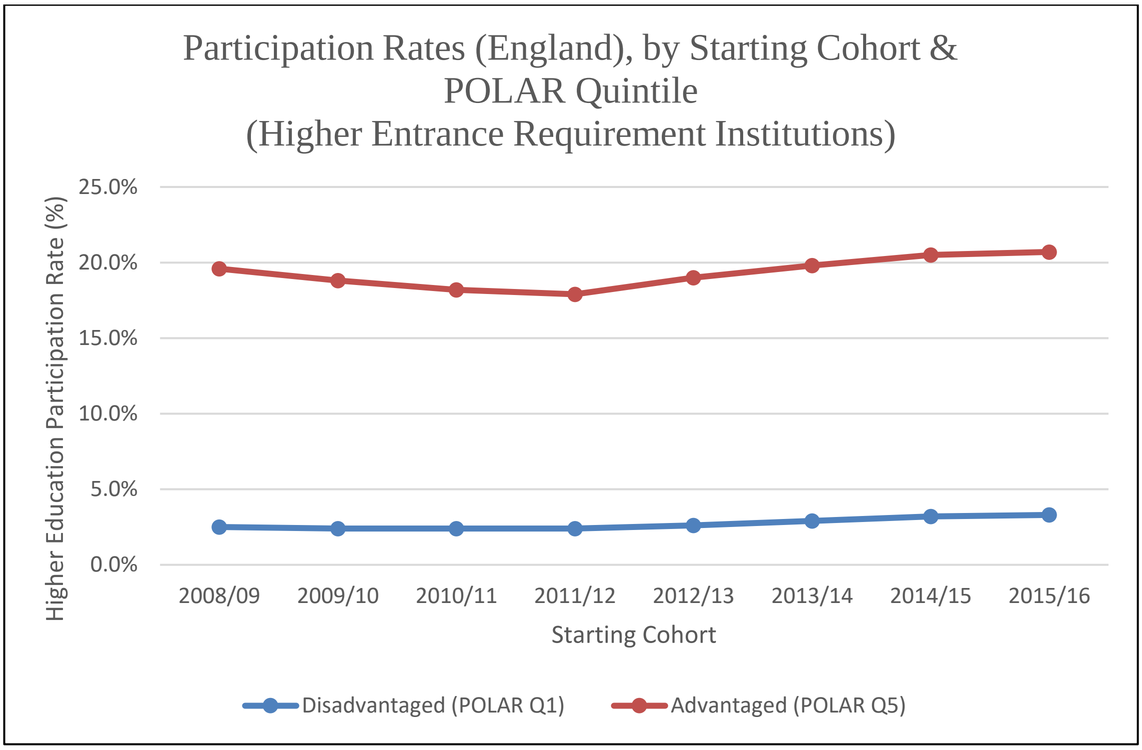
Figure 3.2: Participation Rate in Higher Education (England), by Cohort & POLAR Quintile



Source: BIS, UCAS

139 POLAR is a classification metric for determining whether a student lives in an advantaged area (where there is high rates of participation in higher education) or a disadvantaged area (where participation in higher education is low). The classification is performed at the student's local area or ward, and each ward is designated a number ranging from the lowest quintile of participation (POLAR Q1) to the highest quintile of participation (POLAR Q5). For further details, see: http://www.hefce.ac.uk/analysis/yp/POLAR/

Figure 3.3: Participation Rate in Higher Education (England), by Cohort & Quintile for Higher Tariff Institutions[140]

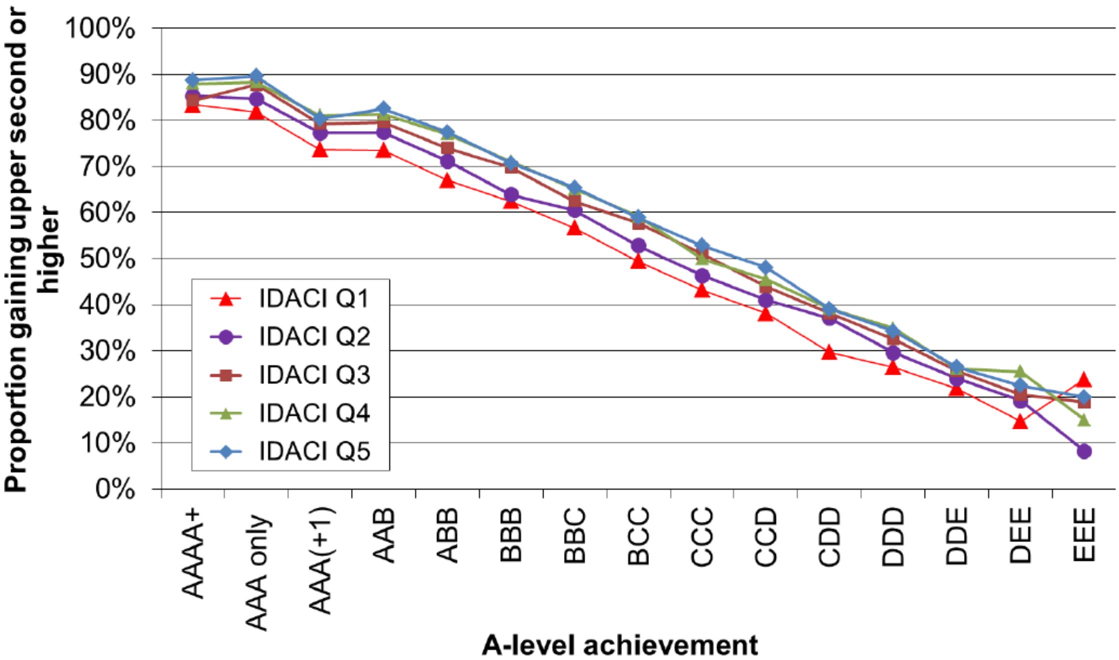


However, as already noted, the returns to a degree will also depend on the subject studied and the university attended, and as seen from Figure 3.3, there has been no significant increase in the participation rate for students from the most disadvantaged quintile who are attending the more elite universities – here denoted by institutions with a higher than average entry requirement (in terms of pre-university schooling grades) across all subjects. This suggests that although the participation gap may be closing, it may not be reflected in terms of returns to a degree and hence social mobility. Finally, as HEFCE (2014) show in Figure 3.4, even when students are participating in HE, conditional on pre-university entry grades, students from the more disadvantaged areas do consistently worse than the

[140] These are the institutions as having an average tariff entry requirement greater than the average. For full specification, see Widening participation in higher education: 2016, Table B; available at: https://www.gov.uk/government/statistics/widening-participation-in-higher-education-2016

most advantaged quintile in terms of degree success, apart from those at the lowest part of the entry grades distribution. It is therefore of significant interest to estimate the impact of a disadvantaged background on degree success, conditional on other factors that may determine degree outcomes, to evaluate whether the effect in Figure 3.4 exists with an econometric specification of degree attainment.

Figure 3.4: Qualifications Obtained, by IDACI[141] Quintile (2007/08 Cohort)



Source: HEFCE, 2014: 29

[141] The Income Deprivation Affecting Children Index (IDACI) is a measure of socioeconomic depravity at the local level. This measure is not specifically targeted at applications to higher education, unlike POLAR. The IDACI score reflects the proportion of children aged 15 or below who are income deprived, and a score is assigned in the [0,1] interval. The upper quintile (Q5) is the least income deprived, and the lower quintile (Q1) is the most income deprived. However, neither IDACI nor POLAR data is available for the HESA individual-level data.

3.3 Data

3.3.1 HESA Data

The data used is the 2005/06 to 2009/10 destinations and leavers HESA dataset that is outlined previously in Chapter 1. The same definitions of low income according to socioeconomic codes created from parental occupation classification are used here, as well as the set of student characteristics that can be used to control for the determinants of degree outcomes outlined in section 3.1.1. In addition, where a UK domiciled student applied via UCAS for their first undergraduate degree, their tariff score is known and is used to control for students with similar levels of pre-university academic attainment. The classification of a student's socioeconomic status and whether a student is known to not have a parent with university education[142], will be used as measures of a disadvantaged background. A student's degree classification is also known, and will be used to create either an ordered or binary indicator of degree success.

3.3.2 Identification Strategy

The original dataset has over 3.6 million observations, but those include postgraduate and part-time students, and student engagements with missing observations for key variables. Since observations for tariff score are only for those students who apply through UCAS for their first degree, only students who are studying for their first undergraduate degree, with a known tariff score at a non-vocational[143] university are selected. As the outcome variable is degree classification, only those students with a pass grade are selected. In practice, this is

[142] An alternative strategy that could be used is to capture additional socioeconomic disadvantage using a student's ethnicity, and thus compare students classified as higher SES and whose ethnicity is White, with students classified as lower SES and whose ethnicity is Black or Asian (potentially the most disadvantaged). This approach would lead to a significantly increased sample size (N= approximately 285,000), however the mechanisms for ethnicity affecting degree outcomes are potentially much more complex and confounded compared to using parental education as an indicator of potential socioeconomic disadvantage. Nevertheless, similar effects are found when using ethnicity as an additional SES measure as those reported here using parental education. Results are available on request.

[143] This is the same specification of universities as used in Chapter 1, in accordance with the universities classified as non-vocational by Gibbons & Vignoles (2009). See Appendix A, Table A5.

dropping all observations with an unclassified or a ‘Further Education’ classification to their university award.[144] Furthermore, since socioeconomic status is required, only students with a known socioeconomic code are kept for analysis, and additionally only students who study full time are retained. Whilst the variable ‘Educated Parent’ – to denote whether a student has a parent with previous university education – exists for all students, it is self-reported by the student on application to UCAS, and hence is subject to missing or unknown observations for some students. Therefore, two indicator variables are created, to denote whether a student either is known to have a parent with university education, or is known to not have a parent with university education. Finally, to concentrate the analysis on UK-domiciled school leavers who began their studies between September 2002 and September 2007, no exchange student observations are kept, and only students who are aged between 17 and 19 are kept.[145] The final dataset contains almost 70,000[146] students, and Table 3.1 below presents the summary statistics for the total sample created by the identification strategy outlined above, and by measures of a disadvantaged background. As expected, the mean tariff score is significantly higher amongst those students from either higher socioeconomic backgrounds, or with parental university education. Additionally, both the ordered degree class variable and the degree success indicator show that the most advantaged students tend to perform better in terms of overall university attainment.

[144] Although medical degrees are traditionally unclassified, they are excluded from the analysis in the interest of consistency of identifying degree attainment within the normal academic degree classification structure.

[145] From the cropped dataset outlined in Appendix A, Table A3, approximately 430,000 observations are dropped due to missing, unknown or unclassified tariff scores. In addition, approximately 50,000 observations are further removed due to the student’s degree classification being below a 3rd or unclassified, and a further 25,000 students are removed if they are aged 20 or above at the beginning of their degree studies. Finally, since the student’s educated parent status is required for the analysis, only students known to either have or known to either not have a parent or guardian with higher educations are kept – approximately 190,000 students are removed from the analysis.

[146] Whilst the size of the dataset is not as large as it would have been without the missing observations for the parental education variable, the 70,000 student observations used in this analysis is still a significant improvement over the 1,200 students used by Lasselle et al. (2014).

Table 3.1: Summary Statistics

	(1) Total	(2) Low Income	(3) High Income	(4) No Educated Parent	(5) Educated Parent
Ordered Degree Class	2.81	2.73	2.84	2.76	2.86
	(0.693)	(0.705)	(0.686)	(0.695)	(0.688)
Degree Success = 1	0.71	0.66	0.73	0.68	0.74
	(0.452)	(0.474)	(0.442)	(0.466)	(0.437)
Tariff Score	349.79	316.30	361.69	330.28	366.61
	(118.5)	(114.7)	(117.6)	(115.7)	(118.4)
Age	18.30	18.29	18.30	18.28	18.32
	(0.475)	(0.473)	(0.476)	(0.465)	(0.484)
Low Income = 1	0.26	-	-	0.42	0.13
	(0.440)	-	-	(0.494)	(0.332)
Educated Parent = 1	0.54	0.26	0.64	-	-
	(0.499)	(0.438)	(0.481)	-	-
No Educated Parent = 1	0.46	0.74	0.36	-	-
	(0.499)	(0.438)	(0.481)	-	-
White = 1	0.83	0.75	0.86	0.82	0.83
	(0.376)	(0.433)	(0.349)	(0.380)	(0.372)
Female = 1	0.59	0.63	0.58	0.61	0.57
	(0.492)	(0.484)	(0.494)	(0.488)	(0.495)
British = 1	0.97	0.96	0.97	0.97	0.97
	(0.173)	(0.193)	(0.165)	(0.161)	(0.183)
Disabled = 1	0.08	0.07	0.08	0.07	0.09
	(0.270)	(0.253)	(0.276)	(0.251)	(0.285)
Local = 1	0.16	0.25	0.13	0.21	0.12
	(0.368)	(0.433)	(0.337)	(0.406)	(0.327)
Live at Home = 1	0.20	0.31	0.17	0.26	0.16
	(0.404)	(0.464)	(0.372)	(0.439)	(0.363)
STEM = 1	0.18	0.17	0.18	0.17	0.19
	(0.384)	(0.379)	(0.386)	(0.374)	(0.392)
Russell Group = 1	0.41	0.27	0.46	0.33	0.48
	(0.491)	(0.446)	(0.498)	(0.469)	(0.500)
Oxbridge = 1	0.03	0.01	0.03	0.01	0.04
	(0.163)	(0.104)	(0.179)	(0.0945)	(0.203)
N	67045	17580	49465	31050	35995

Mean coefficients; standard errors in parentheses. Raw numbers have been rounded to the nearest multiple of 5 to comply with HESA's Standard Rounding Methodology.
Ordered Degree Class is ordered such that higher values correspond to better outcomes. Hence, 1=3rd, 2=2:2, 3=2:1, and 4=1st. Degree Success is a binary variable that takes the value 1 if the student's degree classification was either 2:1 or 1st.
Low Income is a binary variable that takes the value 1 if the student's socioeconomic code is 4-8.
Educated Parent is a binary variable that takes the value 1 if the student declares that they have at least 1 parent with university education. No Educated Parent is a binary variable that takes the value 1 if the student declares that neither parent has university education.
White is a binary variable that takes the value 1 if the student's ethnicity is coded as being white.
Female and British are binary variables that take the value 1 if the student's gender and nationality are declared as female and British respectively.
Disabled is a binary variable that takes the value 1 if the student has declared that they have a known disability.
Local is a binary variable that takes the value 1 if the student attends a university within 20km in Euclidean space from their domiciled postcode where the student resided at the time of making the application through UCAS.
Live at Home is a binary variable that takes the value 1 if the student is known to be living at home whilst undertaking their university study.
STEM is a binary variable that takes the value 1 if the subject studied is classified as being a STEM subject in accordance with the classification table in Appendix B, Table B2.
Russell Group is a binary variable that takes the value 1 if the university the student attended is classified as a Russell Group university (see Appendix A, Table A1). Furthermore, Oxbridge is a binary variable that takes the value 1 if the university the student attended was either the University of Oxford or the University of Cambridge.

3.4 Methodology

As the outcome variable is discrete, a combination of nonlinear estimation strategies are considered, including ordered and binary models. Since a student's degree classification has a potential of four, ordered outcomes[147], it is ordered nonlinear estimation that is first considered. A binary method however, is computationally simpler and allows for a direct comparison between 'success/fail' scenarios, thus the approach and the justification for its adoption and appropriate use is also outlined.

Irrespective of whether degree outcomes are examined from a success/fail approach, or whether they are examined in their original ordered classifications, we are interested in estimating the probability that the observed outcome (y_i) takes a particular discrete value, given a vector of explanatory variables:

[147] Specifically the analysis considers only the degree outcomes of: 1st, 2:1, 2:2, and 3rd class. Unknown or unclassified degrees are omitted, as are those students who did not complete their degree.

$$\Pr(y_i = 1|x_i) = [.] \tag{10}$$

where:

$y_i \in \{0, 1, 2, 3, 4\}$ in the case of ordered degree outcomes, or:
$y_i \in \{0, 1\}$ in the case of using the success/fail approach

This could be calculated using OLS, which with a binary dependent variable becomes the linear probability model (LPM):

$$\Pr(y_i = 1|x_i) = x_i'\beta \tag{11}$$

which is the case as:

$$\Pr(y_i = 1|x_i) = x_i'\beta \tag{12}$$

$$\Pr(y_i = 0|x_i) = 1 - (x_i'\beta) \tag{13}$$

hence since:

$$E[y_i|x_i] = x_i'\beta \tag{14}$$

$$y_i = E[y_i|x_i] + (y - E[y_i|x_i]) = x_i'\beta + \varepsilon \tag{15}$$

However, the LPM does not constrain the predicted probabilities $(x_i'\beta)$ to the interval [0,1], hence predicted probabilities violate the assumptions that:

$$\lim_{x_i'\beta \to +\infty} \Pr(y_i = 1|x_i) = 1 \tag{16}$$

$$\lim_{x_i'\beta \to -\infty} \Pr(y_i = 1|x_i) = 0 \tag{17}$$

Thus, to correctly estimate either the binary or the ordered model where the predicted probabilities are constrained to the [0,1] interval, (following Greene, 2003) we can instead assume that:

$$\Pr(y_i|x_i) = F(x_i'\beta) \qquad (18)$$

where F is some nonlinear function that satisfies the boundary condition. If F is specifically assumed to be the logistic cumulative density function (Λ), then we obtain the logit model:

$$\Pr(y_i|x_i) = \frac{e^{x_i^i\beta}}{1+e^{x_i^i\beta}} = \Lambda(x_i'\beta) \qquad (19)$$

where the log likelihood of the logit model is:

$$\ell(\beta) = \sum_{y_i=1} \ln \Lambda(x_i'\beta) + \sum_{y_i=0} \ln\left(1 - \Lambda(x_i'\beta)\right) \qquad (20)$$

Alternatively, if we assumed a normal distribution of the error term instead of the logistic, and hence use the normal cumulative density function then we obtain the probit model:

$$\Pr(y_i|x_i) = \int_{-\infty}^{x_i'\beta} \phi(t)dt = \Phi(x_i'\beta) \qquad (21)$$

where the log likelihood of the probit model is:

$$\ell(\beta) = \sum_{y_i=1} \ln \Phi(x_i'\beta) + \sum_{y_i=0} \ln\left(1 - \Phi(x_i'\beta)\right) \qquad (22)$$

In either case, the globally concave log likelihood function is maximised using Maximum Likelihood (ML) estimation, which sets the score function of the log likelihood function equal to zero, and iteratively converges on the estimate of β, $\hat{\beta}$

that makes the observations in the data the most likely to be observed. The process of maximising the likelihood is akin to that of minimising the residual sum of squares in OLS.

Moreover, whilst we observe the discrete outcomes in either the binary or ordered case, it is likely that these outcomes are themselves a function of some underlying, continuous, but unobservable regression. The dependent variable can therefore be split into an observed (y_i) and a latent, unobserved variable (y_i*) that determines what category of the dependent variable each observation is. In the general case, if:

$$y_i^* = x_i'\beta + \varepsilon \qquad \text{we observe:} \qquad y_i = \begin{cases} 1 \;\; if \;\; y_i^* \geq 0 \\ 0 \;\; if \;\; y_i^* < 0 \end{cases} \tag{23}$$

3.4.1 Ordered Logit

If the discrete outcomes are not dichotomous, estimation of logit or probit is usually extended into multinomial regression design, but as the degree outcomes are ordinal there is a relationship between the dependent variable's categories. Such ordinal information would not be taken into account by a multinomial model (i.e. a 2:2 vs 2:1 outcome comparison would be treated the same as a 2:2 vs 1st outcome), hence an ordered model must be used. We assume that there is still a latent outcome variable which is unobserved, but now:

$$y_i = \begin{cases} 0 \;\; if \;\; y_i^* \leq \alpha_1 \\ 1 \;\; if \; \alpha_1 < \; y_i^* \leq \alpha_2 \\ 2 \;\; if \; \alpha_2 < \; y_i^* \leq \alpha_3 \\ \qquad \vdots \\ J \;\; if \;\; \alpha_{J-1} < y_i^* \end{cases} \tag{24}$$

Specifically with the ordered degree outcomes:

$$y_i = \begin{cases} 3rd \;\; if \;\; y_i^* \leq \alpha_1 \\ 2{:}\,2 \;\; if \;\; \alpha_1 < \; y_i^* \leq \alpha_2 \\ 2{:}\,1 \;\; if \;\; \alpha_2 < \; y_i^* \leq \alpha_3 \\ 1st \;\; if \;\; \alpha_3 < y_i^* \end{cases} \qquad (25)$$

where the J-1 threshold parameters (α) are estimated using ML in conjunction with obtaining $\hat{\beta}$. Assuming the logit form of the cumulative probability density function, following Wooldridge (2011) we can estimate each of the probabilities that the degree classification will take one of the possible categories:

$$\begin{aligned} Pr(y_i = 3rd|x_i) &= Pr(y_i^* \leq \alpha_1|x_i) = Pr(x_i'\beta + \varepsilon \leq \alpha_1|x_i) = \Lambda(\alpha_1 - x_i'\beta) \\ Pr(y_i = 2{:}\,2|x_i) &= Pr(\alpha_1 < \; y_i^* \leq \alpha_2|x_i) = \Lambda(\alpha_2 - x_i'\beta) - \Lambda(\alpha_1 - x_i'\beta) \\ Pr(y_i = 2{:}\,1|x_i) &= Pr(\alpha_2 < \; y_i^* \leq \alpha_3|x_i) = \Lambda(\alpha_3 - x_i'\beta) - \Lambda(\alpha_2 - x_i'\beta) \\ Pr(y_i = 1st|x_i) &= Pr(y_i^* > \alpha_3|x_i) = 1 - \Lambda(\alpha_3 - x_i'\beta) \end{aligned} \qquad (26)$$

As there are four possible categories of the dependent variable, there are 3 threshold parameters to be estimated. In the binary case considered in section 3.4.2, J=1 and hence we arrive back at the binary logit model. The ordered model is estimated using the ML method, where the log likelihood function to be maximised is:

$$\begin{aligned} \ell(\alpha, \beta) = 1[y_i = 3rd\,]log[\Lambda(\alpha_1 - x_i'\beta)] \\ + 1[y_i = 2{:}\,2\,]log[\Lambda(\alpha_2 - x_i'\beta) - \Lambda(\alpha_1 - x_i'\beta)] \\ + 1[y_i = 2{:}\,1\,]log[\Lambda(\alpha_3 - x_i'\beta) - \Lambda(\alpha_2 - x_i'\beta)] \\ + 1[y_i = 1st\,]log[1 - \Lambda(\alpha_3 - x_i'\beta)] \end{aligned} \qquad (27)$$

which is similar to the approach used by Barrow et al. (2009). However, just like the binary model considered in the proceeding section, although the assumption of a particular probability distribution of the error term negates the boundary assumption violation, the coefficients reported by any logit or probit model (whether binary, multinomial or ordered) are not readily interpretive. Initially, the coefficients estimated by the ordered logit model are not marginal effects, but ordered log-odds. Whilst the sign and significance of the coefficient can be interpreted, a positive significant coefficient only indicates that the explanatory variable has a significant and positive effect on the dependent variable, averaged across all possible outcome categories in an ordinal fashion. Thus, the actual marginal effect of x_i on y_i is not only non-constant across the distribution of individual observations (unlike OLS), but it also non-constant across the ordered outcomes. Where there are 4 ordered categories of the outcome variable, there will be a marginal effect for each category. Hence, the marginal effect of x_i on y_i where $y_i = j$ is given by:

$$\frac{\partial Pr[y_i=j]}{\partial x_i} = \{F'(\alpha_{j-1} - x_i'\beta) - F'(\alpha_j - x_i'\beta)\}\beta \qquad (28)$$

Where marginal effects are reported for each outcome category, they are the average marginal effects (AMEs), where the marginal effect is computed as above for all individuals within an outcome category, and averaged. Alternatives to using AMEs are the marginal effects at a representative value (MERs) and marginal effects at the mean (MEMs). Arguably AMEs may be more representative of the population of observations within a category however, since MEMs and MERs do not take into account each individual's marginal effects across the entire distribution.

3.4.2 Binary Estimation of an Ordered Variable

The ordered model outlined in the previous section may be converted into a binary logit model if only 1 threshold parameter is required – i.e. the discrete dependent variable is binary. This may be useful if the outcome variable itself is at least to some extent inherently dichotomous, which may be the case given that graduate employment vacancies and schemes regularly require at least a 2:1 at degree level. Dichotomisation of a continuous variable however generally leads to a loss of statistical inference and allows for the possibility of missed effects in the aggregation process (Royston et al, 2006). Nevertheless, there may be some situations where it is applicable due to the nature of the research design, and especially when the dependent variable is discontinuous. In the case of degree outcomes, it has already been outlined that the 4 ordered discrete outcomes are considered, so any process of dichotomisation would be to collapse 4 categories into 2 (success/fail). In addition to the contextual arguments for considering a degree success as 2:1 or above, and a fail as 2:2 and below, it is common to consider degree outcomes in such a fashion (e.g. Lasselle et al, 2014; Crawford, 2014a; Smith and Naylor, 2001a).

Whilst there is some loss of inference by classifying 2:1 and 1st as the same category, there is already a degree of lost inference and imprecision in the ordered outcomes themselves. As a generalised example, if Student A receives 69% as their overall degree score they are awarded a 2:1 classification, and if Student B receives 71% they are awarded a 1st class degree. However, since the classification is known and the overall score is not, the difference between students A and B with respect to their degree outcomes are the same as Student C who receives 61% and Student D who receives 95% respectively. If the exact degree scores were known (and continuous), it would be possible in theory to implement a regression discontinuity design at the degree classification thresholds. The design would need to be fuzzy to allow discretionary borderline cases, but such a design was used by Feng and Graetz (2013) using university-specific data on degree classifications and scores at the London School of Economics and Political Science.

Finally, a binary approach to the outcome variable is employed since, not only is it computationally simpler, it would create more direct comparisons of the marginal effects between the explanatory variables of interest across the tariff score groupings. However, the ordered model will be conducted as a first stage to check whether the initial marginal effects from the ordered logit are seemingly dichotomous. If that is the case, then we should expect to see the sign, significance and magnitude to be consistent across the two ordered categories within both the success and the fail collapsed categories.

As with the ordered model, estimation output of the logit model returns the log-odds as coefficients. Since in the binary model it is the marginal effect of x_i on y_i which is of interest, then the marginal effect (for a continuous regressor) is:

$$\frac{\partial Pr[y_i=1]}{\partial x_i} = \Lambda(x_i'\beta)[1 - \Lambda(x_i'\beta)]\beta \qquad (29)$$

and in the case of a binary explanatory variable (d):

$$\frac{\partial Pr[y_i = 1]}{\partial x_i} = Pr\big[y_i = 1|\overline{x}_{(d)}, d = 1\big] - Pr\big[y_i = 1|\overline{x}_{(d)}, d = 0\big] \qquad (30)$$

where $\overline{x}_{(d)}$ are the means of all other variables included in the model, since the marginal effects with respect to a small change with a continuous regressor cannot be used with a binary indicator such as gender.

3.4.3 Subject & Tariff Groupings

Since the effect of interest is that of students' socioeconomic background on degree outcomes conditional on pre-university attainment, students' tariff scores are used to group students together into groups of similar pre-university academic achievement. Any student applying to university through the UCAS has their pre-university academic qualifications and grades converted into a tariff score. Although subject to recent revisions, during the period of analysis a student's tariff score was calculated as follows[148]:

[148] In the interests of concision, only AS and A-Levels are listed here. A comprehensive guide to calculating a tariff score from other pre-university qualifications (such as BTECs, International Baccalaureates) can be found in Appendix 3, Table C1.

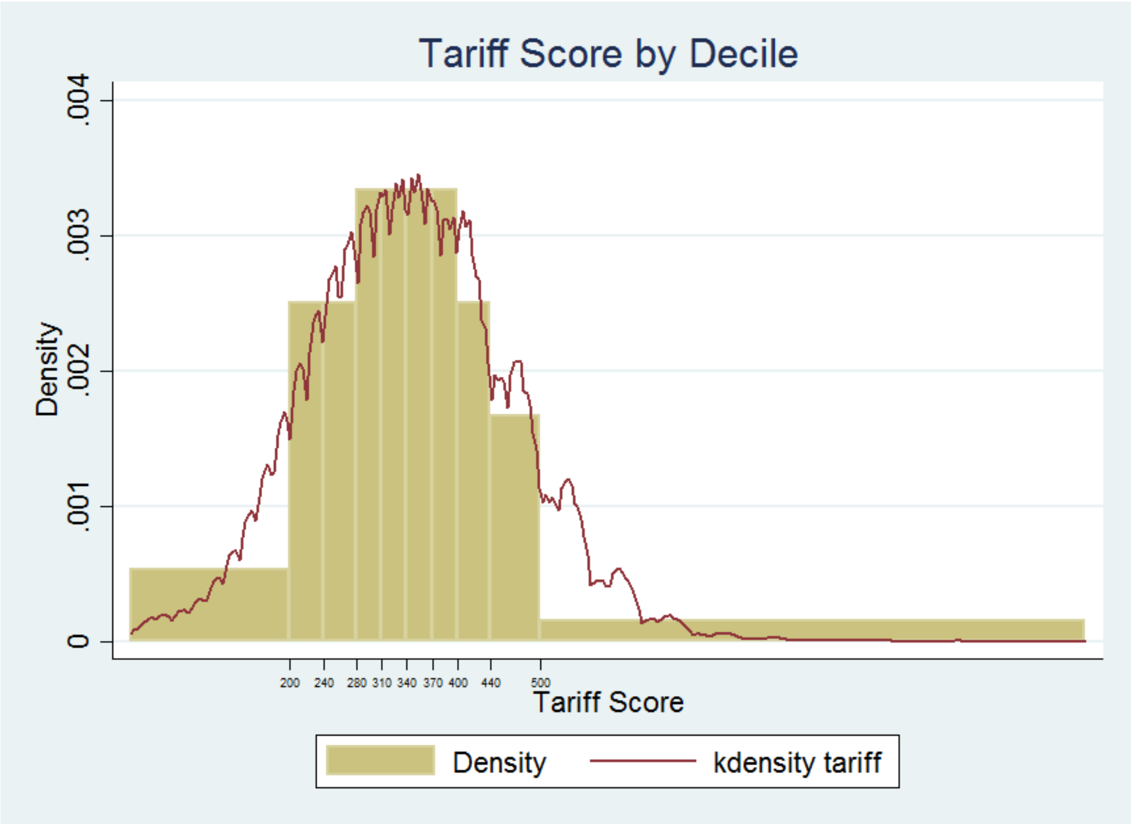
Table 3.2: Tariff Scores by Grade and Qualification

GCE A-Level	GCE AS-Level	Scottish Advanced Higher	Scottish Higher	Tariff Score
A		A		**120**
B		B		**100**
C		C		**80**
		D	A	**72**
D	A		B	**60**
	B			**50**
			C	**48**
			D	**42**
E	C			**40**
	D			**30**
	E			**20**

So a student who applies to university with 3 A-Levels (AAB) and 1 AS-Level (D) would have a total tariff score of 370. However, a student may also attain the same tariff score by applying to university with 3 A-Levels (AAC) and 1 AS-Level (B). Thus, since the exact grades are unknown within the dataset, the measure of pre-university attainment is to some degree imperfect. Further complications of using a student's tariff score is that the subjects studied (in addition to the exact qualifications and grades) are unknown. Nevertheless, within certain bandwidths of tariff score, it is reasonable to expect pre-university attainment to be relatively comparable. One further aspect of using tariff scores is, due to the nature of the calculation, only students who have applied to university through UCAS have a valid tariff score observation. However since the research question is aimed at the performance of school-leavers, and all school-leavers must apply through UCAS for undergraduate study, observations with missing tariff scores are not relevant.

The specific 10 tariff score groupings are created by taking the decile frequencies of the tariff score across all subjects, as seen in Figure 3.5.

Figure 3.5: Histogram and Kernel Density of Tariff Score with Variable Bin Widths



As the tariff scores are not perfectly continuous (individual grade scores are predominantly multiples of 10), the decile groups do not contain an equal number of observations, which is also seen by the discontinuous kernel density plot. This method of creating sub-populations of students conditional on similar levels of pre-university attainment has been used extensively by the Higher Education Funding Council for England (HEFCE), in various forms. Where the exact grades of the students' qualifications were unknown, students were split into 6 tariff score groups (HEFCE, 2013; 2015b) and where students' grades were known, students were split into 10 tariff groups (HEFCE, 2014). The latter of these approaches however, is not directly comparable to this grouping design, since the tariff scores used in this dataset are calculated from all of the student's applicable qualifications, as opposed to the best 3 A-Levels. The result is that the tariff score bandwidths used in this design are slightly inflated due to the incorporation of all of the students' grades. Nevertheless, across the tariff score groupings there is almost a perfectly monotonic increase in the proportion of high income students moving from the lowest tariff score group to the highest (Table 3.3), as well as a similarly monotonic increase in the proportion of students obtaining a higher degree class (Table 3.4). Lower socioeconomic status students are therefore under-represented in the higher bandwidths of pre-university attainment compared to lower pre-university attainment, and from a descriptive standpoint students with lower tariff scores are less likely to obtain a good degree. Table 3.4 also shows a further potential benefit of considering a binary indicator of degree success, in that as students enter HE with particularly higher levels of pre-university attainment (e.g. a tariff scores in excess of 440), the proportion of students in the highest bandwidths obtaining a 2:1 falls, simply as the highest achieving students are more likely to obtain a 1st class degree. By dichotomising the dependent variable, the marginal effects calculated will be consistently estimated between two distinct outcomes, rather than across four outcomes that may exhibit some overlap.

Table 3.3: Tariff Score, by High and Low Income (All Subjects)

Tariff Score	SEC 1-3 (High Income)	SEC 4-8 (Low Income)	Total
0 – 199	3,565 *(60.9%)*	2,290 *(39.1%)*	5,890 *(100.0%)*
200 – 239	3,100 *(63.2%)*	1,805 *(36.8%)*	4,910 *(100.0%)*
240 – 279	4,510 *(66.2%)*	2,300 *(33.8%)*	6,810 *(100.0%)*
280 – 309	4,630 *(69.4%)*	2,045 *(30.6%)*	6,675 *(100.0%)*
310 – 339	4,175 *(71.6%)*	1,650 *(28.4%)*	5,825 *(100.0%)*
340 – 369	5,565 *(75.2%)*	1,835 *(24.8%)*	7,405 *(100.0%)*
370 – 399	4,835 *(77.7%)*	1,385 *(22.3%)*	6,225 *(100.0%)*
400 – 439	6,335 *(79.5%)*	1,635 *(20.5%)*	7,970 *(100.0%)*
440 – 499	6,655 *(81.4%)*	1,520 *(18.6%)*	8,175 *(100.0%)*
500 +	6,090 *(84.7%)*	1,100 *(15.3%)*	7,195 *(100.0%)*
Total	49,465 *(73.8%)*	17,580 *(26.2%)*	67,045 *(100.0%)*

Raw numbers have been rounded to the nearest multiple of 5 to comply with HESA's Standard Rounding Methodology.

Table 3.4: Tariff Score, by Ordered Degree Outcomes (All Subjects)

Tariff Score	3rd	2:2	2:1	1st	Total
0 – 199	550	2,695	2,350	265[149]	5,860
	(9.4%)	*(46.0%)*	*(40.1%)*	*(4.6%)*	*(100.0%)*
200 – 239	325	2,180	2,155	250	4,910
	(6.6%)	*(44.4%)*	*(43.9%)*	*(5.1%)*	*(100.0%)*
240 – 279	300	2,590	3,475	445	6,810
	(4.4%)	*(38.0%)*	*(51.0%)*	*(6.53%)*	*(100.0%)*
280 – 309	230	2,100	3,775	570	6,675
	(3.4%)	*(31.5%)*	*(56.6%)*	*(8.6%)*	*(100.0%)*
310 – 339	170	1,650	3,485	515	5,825
	(2.9%)	*(28.3%)*	*(59.9%)*	*(8.9%)*	*(100.0%)*
340 – 369	155	1,680	4,690	885	7,405
	(2.1%)	*(22.7%)*	*(63.3%)*	*(11.9%)*	*(100.0%)*
370 – 399	105	1,210	4,075	830	6,225
	(1.7%)	*(19.5%)*	*(65.5%)*	*(13.4%)*	*(100.0%)*
400 – 439	135	1,135	5,305	1,400	7,970
	(1.7%)	*(14.2%)*	*(66.5%)*	*(17.6%)*	*(100.0%)*
440 – 499	120	1,050	5,315	1,685	8,175
	(1.5%)	*(12.9%)*	*(65.1%)*	*(20.6%)*	*(100.0%)*
500 +	100	695	4,430	1,965	7,195
	(1.4%)	*(9.7%)*	*(61.6%)*	*(27.3%)*	*(100.0%)*
Total	2,185	16,985	39,055	8,815	67,045
	(3.3%)	*(25.3%)*	*(58.3%)*	*(13.2%)*	*(100.0%)*

Raw numbers have been rounded to the nearest multiple of 5 to comply with HESA's Standard Rounding Methodology.

[149] The further benefit of dichotomising the ordered degree outcomes variable is that it negates the small number of observations at the top and bottom of the distribution of good and bad outcomes. The relatively small number of observations for those who obtained a 1st class degree but whose tariff score was less than 199 (N=267) is significantly smaller than those who obtained a good degree within the same tariff band (N=2,615). This will be particularly important for STEM subjects, where the number of observations will be even smaller given the same outcome and tariff parameters.

Whilst the effect of student background on degree outcomes is estimated conditional on pre-university attainment for all subjects, it is further allowed to be heterogeneous across different subjects. Using the student's subject code (as classified by the Joint Academic Coding System, JACS), students are grouped into either STEM or non-STEM subjects.[150] This is a similar approach to the six subject groups identified by Smith and Naylor (2001a), however by using only STEM and non-STEM, the subjects that are arguably the most rewarding in terms of rates of return (and have the higher tariff entry requirements on average) can be directly compared with the subjects that are arguably the least rewarding.[151] Therefore the effect of coming from a disadvantaged background on degree outcomes is allowed to be heterogeneous between students who choose to study high and low reward subjects, which have the higher and lower entry requirements on average respectively.

Finally, although the raw tariff score is retained and used to assign students to bandwidths, the tariff score when included in an estimation as an explanatory variable will be rescaled by a factor of 20. This is so that, given how the tariff scores are constructed, a 1 unit change in the rescaled tariff score is equal to a 1 grade change in a student's A-Level subject result or equivalent.[152]

3.4.4 Empirical Models

The model to be estimated in the first instance is the ordered logit model, to establish model specification and to evaluate whether a binary success/fail approach is plausible. Hence, Table 3.5 reports the ordered log-odds from the following ordered logit model:

[150] The same considerations that were highlighted in section 1.4.3 are applicable here in creating a STEM indicator using JACS codes.
[151] e.g. Greenwood et al. (2011).
[152] As seen from Table 3.2 and Table C1 in Appendix C, a grade B at A-Level is worth 100 tariff points, whereas a grade A is worth 120 tariff points. Thus, the rescaled tariff score would reflect a change in the tariff points from 5 to 6.

$$y_{i,j,u,t} = \delta_0 + \delta_1 \mathrm{T}_i + \delta_2 S_i + \delta_3 P_i + \delta_4 (P_i * S_i) + \sum_{k=1}^{K} \beta_k X_{i,j,u,t} + \sum_{t=2004}^{T} \psi_t D_t + \sum_{j=2}^{J} \psi_j J_j + \sum_{u=2}^{U} a_u A_u + \varepsilon_{i,j,u,t} \tag{31}$$

where $y_{i,j,u,t}$ is one of 4 possible degree outcomes[153] for student *i* who studies subject *j* at university *u* in a given academic year (t); T_i is the tariff score associated with individual i's pre-university academic attainment, S_i is a dummy indicator denoting whether the student is classified as low income from their socioeconomic code, and P_i is also a dummy indicator denoting whether the student is known to not have a parent or guardian with university education. An interaction between the low income indicator and this parental education indicator is used to capture any additional effect of a student being classified as both from a lower income background and who is known to not have a parent or guardian with university education, which is captured by the coefficient δ_4 . Also included are a set of binary student controls ($X_{i,j,u,t}$; gender, nationality, disability status and whether the student is attending a local university to their domiciled address)[154], a complete set of time dummies (D_t) to capture year effects, as well as university fixed effects (A_u) to account for unobserved heterogeneity across universities. Finally, to capture the heterogeneity in degree outcomes across subjects due to differences in difficulty and material, subject fixed effects are included at the JACS 1 digit level. This approach addresses all the possible determinants of degree outcomes listed in section 3.1.1, given the limitations of the data. This strategy is also an improvement on the specification of Lasselle et al. (2014), since subject and year fixed effects are used, rather than just faculty fixed effects; in addition to the significantly larger number of observations due to the estimation being performed across all

[153] The convention is for the outcome variable is ordered such that '1' relates to the 'worst' outcome, and higher categories relate to better outcomes. As such, the ordered degree variable is coded such that 1=3rd class, 2=2:2, 3=2:1 and 4=1st class.
[154] Specifically, British==1, Female==1, White==1, Disabled==1 and Local==1; or 0 otherwise.

universities in the UK. As such, the error term ($\varepsilon_{i,t}$) is clustered at the institutional level, to account for the likely violation of independence of the errors between student observations.

Following on from the baseline model, the tariff variable (T_i) is then omitted from the empirical estimation, and instead the model that is now estimated is:

$$y_{i,j,u,t,m} = \delta_0 + \delta_2 S_i + \delta_3 P_i + \delta_4 (P_i * S_i) + \sum_{k=1}^{K} \beta_k X_{i,j,u,t,m} + \sum_{t=2004}^{T} \psi_t D_t + \sum_{j=2}^{J} \psi_j J_j + \sum_{u=2}^{U} a_u A_u + \varepsilon_{i,j,u,t,m} \tag{32}$$

which is performed using the constructed sample of all students (*i*), who studies a particular subject (*j*) at a university (*u*) across all cohorts of students by academic year (*t*), within a tariff bandwidth group (*m*) across all subjects. Instead of controlling for pre-university attainment, this approach allows the estimates to be conditional on the pre-university attainment being relatively similar. The estimates from these tables are presented in Tables 3.7a (coefficients are log odds) and 3.7b (coefficients are marginal effects), after the evaluation of the marginal effects from the ordered model and thus whether the ordered approach can be dichotomised (Table 3.6). Irrespectively, the tariff bandwidth model in Table 3.7 is performed in the same manner for two further sub-populations of students: students studying Science, Technology, Engineering and Mathematics[155] (STEM) subjects (Table 3.8a and Table 2.8b) and non-STEM subjects (Table 3.9a and Table 3.9b). Table 3.10 shows the predicted probabilities of attaining degree success within each bandwidth and across the subject groupings, with the explanatory variables set at

[155] As defined in the preceding two chapters, using the classification as outlined in Appendix B, Table B2.

their mean values. This can be used to give context to the marginal effects found in the preceding tables.

3.5 Results

Table 3.5 presents the model specification and the additive effects of the student controls on ordered degree outcomes. As expected from the empirical body of research, students from lower socioeconomic backgrounds are less likely to obtain good degree, however this effect is only statistically significant when accounting for subject, year and university fixed effects, and allowing for the additional effect for students from lower socioeconomic backgrounds and not having a parent with university education. Students who do not have a parent with university education and students who declare themselves as being disabled also have a significantly lower probability of obtaining higher degree outcomes, and in both cases these effects are larger in magnitude and statistical significance when controlling for fixed effects. Conversely, and still in line with empirical findings, students who are white, British, and female are all significantly more likely to obtain higher degree outcomes. Moreover, the significantly positive effect of tariff score on the probability of obtaining a higher degree outcome is robust even when controlling for subject, year and university fixed effects, and the full set of student control characteristics, which in this analysis conclusively finds that pre-university attainment is a significant determinant of degree attainment.

Furthermore, one of the indicators of socioeconomic disadvantage – not having a parent with university education – is also robust to controlling for student heterogeneity, which suggests that disadvantaged students are less likely to do well in their university studies, when controlling for pre-university attainment and other student characteristics. Finally, although a disadvantaged background appears to reduce the probability of degree success on average, students who are the most disadvantaged (who are from a low income background and who do not have a parent with university education) are significantly more likely to obtain a better degree outcome. This interaction coefficient, in ordered log odds, gives the difference between the log-odds ratio comparing students from low and high income backgrounds, with and without a parent with university education. This may suggest that those students who encounter the most obstacles in attending university may have higher potential to succeed compared to their counterparts, ceteris paribus.[156]

[156] As discussed in section 3.1.3, any estimate of the effect of a disadvantaged background on degree outcomes using data that has a self-selection bias (since we only observe those students who do participate in university education) must also be set in the context that such obstacles may prevent students from participating, and hence being observed in the data.

Table 3.5: Determinants of Degree Success (Ordered Logit Regression)

	(1)	(2)	(3)	(4)	(5)	(6)
Tariff	0.119***	0.118***	0.117***	0.117***	0.120***	0.120***
	(0.00562)	(0.00537)	(0.00535)	(0.00534)	(0.00523)	(0.00523)
Low Income =1		-0.0158	-0.0172	-0.0179	-0.0260	-0.109***
		(0.0270)	(0.0262)	(0.0262)	(0.0261)	(0.0419)
No Educated Parent =1		-0.0982***	-0.101***	-0.103***	-0.0798***	-0.109***
		(0.0279)	(0.0269)	(0.0269)	(0.0267)	(0.0294)
White =1		0.354***	0.374***	0.376***	0.494***	0.495***
		(0.0387)	(0.0382)	(0.0383)	(0.0413)	(0.0413)
Female =1		0.211***	0.212***	0.211***	0.205***	0.205***
		(0.0201)	(0.0203)	(0.0203)	(0.0202)	(0.0202)
British =1		0.193***	0.222***	0.223***	0.254***	0.254***
		(0.0638)	(0.0651)	(0.0649)	(0.0617)	(0.0618)
Local =1			0.0713	0.0704	-0.0187	-0.0194
			(0.0520)	(0.0519)	(0.0444)	(0.0444)
Disabled =1				-0.0791**	-0.0946***	-0.0950***
				(0.0326)	(0.0341)	(0.0341)
Low Income * No Edc Parent						0.127***
						(0.0437)
Observations	67,045	67,045	66,480	66,480	66,480	66,480
Subject FE	×	✓	✓	✓	✓	✓
Year FE	×	×	✓	✓	✓	✓
University FE	×	×	×	×	✓	✓
Cluster SE	✓	✓	✓	✓	✓	✓

Standard errors shown in the parentheses are clustered at the institutional level across all 6 specifications. All regressions are performed using ordered logit to estimate equation 31. The dependent variable is the ordered degree outcome, where 1=3rd class degree, 2=2:2, 3=2:1 and 4=1st. Coefficients reported are ordered log odds and therefore report the effect of the explanatory variable averaged across the 4 possible outcomes of degree classification. The number of observations from model specification (3) is lower due to some missing observations of a valid UK postcode which makes the Local binary indicator neither 1 nor zero. Specifications (2)-(6) include subject fixed effects at the JACS 1 digit code level, specifications (3)-(6) include year effects, and specifications (5) and (6) include university fixed effects. Raw numbers have been rounded to the nearest multiple of 5 to comply with HESA's Standard Rounding Methodology. *** p<0.01, ** p<0.05, * p<0.1

Table 3.6 shows the average marginal effects of the explanatory variables on the ordered outcomes using the final model specification (column 6) from Table 3.5, with four AMEs for each degree classification. Firstly, there is a clear case for dichotomising the outcome variable, since the predictors that are positive and significant for predicting a student obtaining a 2:1 or 1st are negative and significant for predicting a student obtaining a 2:2 or 3rd. Whilst there is some variation in the magnitude of the AMEs between the outcomes, the sign and significance are consistent within the 'good' and 'bad' outcomes. Hence, the analysis can be converted into a binary estimation method, as outlined in section 3.4.2. Before moving to the bandwidth results, the marginal effects in Table 3.6 give quantifiable effects on the predicted probabilities of the outcome variable categories. If a student's tariff score is increased by 1 unit (i.e. by 1 A-Level grade or 20 tariff points), they are 0.369 percentage points less likely to obtain a 3rd class degree, whilst they are 0.912 percentage points and 1.27 percentage points more likely to obtain a good degree (2:1 and 1st respectively). Similarly, a student coming from a low income background means they are 0.343 percentage points more likely to get a 3rd class degree, and 1.66 percentage points more likely to get a 2:2, and 1.13 percentage points less likely to get a 1st class degree. These effects are now evaluated for students with similar pre-university attainment, instead of controlling for it across all students, by assigning students to one of the ten tariff bandwidths outlined in section 3.4.3 and omitting the tariff variable from the empirical model as shown in equation 32.

Table 3.6: Average Marginal Effects for Ordered Degree Outcomes

	(1) Degree Class: 3	(2) Degree Class: 2:2	(3) Degree Class: 2:1	(4) Degree Class: 1
Tariff	-0.00369***	-0.0181***	0.00912***	0.0127***
	(0.000271)	(0.000831)	(0.000332)	(0.000640)
Low Income =1	0.00343**	0.0166***	-0.00872**	-0.0113***
	(0.00141)	(0.00639)	(0.00347)	(0.00432)
No Educated Parent =1	0.00335***	0.0165***	-0.00837***	-0.0114***
	(0.000921)	(0.00445)	(0.00233)	(0.00302)
White =1	-0.0174***	-0.0778***	0.0486***	0.0466***
	(0.00208)	(0.00648)	(0.00487)	(0.00341)
Female =1	-0.00639***	-0.0311***	0.0161***	0.0213***
	(0.000816)	(0.00299)	(0.00167)	(0.00205)
British =1	-0.00869***	-0.0393***	0.0233***	0.0248***
	(0.00254)	(0.00962)	(0.00650)	(0.00561)
Local =1	0.000600	0.00293	-0.00149	-0.00204
	(0.00138)	(0.00671)	(0.00346)	(0.00464)
Disabled =1	0.00303***	0.0145***	-0.00772***	-0.00976***
	(0.00114)	(0.00525)	(0.00296)	(0.00342)
Low Income * No Educated Parent	-0.00380***	-0.0189***	0.00891***	0.0137***
	(0.00134)	(0.00640)	(0.00281)	(0.00491)
Observations	66,480	66,480	66,480	66,480

The coefficients reported are the average marginal effects (AMEs) of the effect of the explanatory variable for each ordered outcome. The AMEs are calculated after estimating specification (6) from Table 3.5, which estimates equation 31 using an ordered logit regression. The AMEs show (for example) the additive effect of a 1 unit increase in the tariff score in the probability of obtaining a 2:1 of 0.912 percentage points (in the case of continuous explanatory variables); and that being female compared to being male increases the probability of obtaining a 1st class degree by 2.13 percentage points (in the case of indicator explanatory variables), holding all other things constant. Raw numbers have been rounded to the nearest multiple of 5 to comply with HESA's Standard Rounding Methodology. *** p<0.01, ** p<0.05, * p<0.1

Table 3.7a gives the binary logit results across the tariff bandwidths as outlined in section 3.4.3, averaged across all subjects. The coefficients are log-odds, and average marginal effects for the variables of interest (income background and parental income) on degree success are presented in Table 3.7b. First considering the control variables in Table 3.7a, the positive effect of being female on degree success is statistically significant across all levels of pre-university attainment, and as students enter HE with higher pre-university grades, the effect on degree success of being female increases in magnitude. Furthermore, whilst the effects of being female or white are significant, the effect of being disabled or British are now statistically insignificant in determining degree success, conditional on similar levels of pre-university attainment.

Having controlled for other student characteristics, Table 3.7b shows that the average marginal effect of being from a low income background has a negative impact on the probability of obtaining a good degree, particularly for those with higher or lower than average levels of pre-university attainment. For students who entered HE with the lowest pre-university attainment (the least likely to obtain degree success, as seen from Table 3.4), being classified as lower socioeconomic status has an additional negative effect on the probability of obtaining a good degree. The average marginal negative effect of low socioeconomic status on the probability of degree success across all subjects is 13.8 percentage points, which is a 31.4% decrease in the probability of degree success associated with being from a low income background, relative to the mean probability of degree success of 44% (for the lowest tariff bandwidth) as shown in Table 3.10. Even for students with pre-university attainment towards the middle of the distribution of tariff scores, entering HE as a student from a lower income background has a negative impact on the probability of obtaining a good degree. For students who obtained between 240-279 tariff points, the probability of degree success is 11.3 percentage points lower for low socioeconomic status students across all subjects, which is a 19.5% decrease relative to the mean probability of degree success for that tariff bandwidth of 57.9%

For students who attained higher than average pre-university grades, if they were known to not have a parent with university education, this carried a similar penalty of entering HE as a socioeconomically disadvantaged student. However, the magnitude of this effect is not as large as the effect of lower socioeconomic status, such as students in the highest tariff bandwidth experiencing on average a 4.73 percentage point decrease in the probability of degree success for students without a parent with university education, although this is only a 6.6% decrease relative to the mean probability of degree success of 74.5%. The effect of being classified as a lower socioeconomic status student however is also decreasing in magnitude as the level of pre-university attainment is increasing. Students who scored between 440-499 tariff points are still less likely to obtain a good degree (conditional on completion) due to coming from a low income background, however this fall of 5.56 percentage points represents only a 6.3% decrease in the probability of degree success, relative to the mean of 87.9%. The magnitude of the effect decreasing as pre-university attainment increases may indicate that disadvantaged students with higher tariff scores are more able to cope with university study.

Surprisingly though, students who are both from a low income background and who are known to not have a parent with university education are more likely to obtain a good degree, both at higher and lower than average tariff bands. Students who have the lowest level of pre-university attainment (less than 200 tariff points) and are from both low income backgrounds and who don't have a university-level educated parent are 10.7 percentage points more likely to attain degree success. Given the predicted probability of success within this tariff bandwidth across all subjects is 44% (Table 3.10), this is a 24.3% increase in the probability of degree success. Similar effects are also found for the 240-279 tariff bandwidth, and the second highest tariff bandwidth, albeit but with quantitatively lower effect. This may mean that although disadvantaged students are less likely to succeed in obtaining a good degree, students that are the most disadvantaged – who potentially have entered HE by overcoming the most obstacles – are more likely to succeed.

An argument contrary to this story however is the result from tariff bandwidth 6 (which is centred on the median tariff score of 350), which shows that for students whose pre-university attainment is in the middle of the distribution, to have one indicator of socioeconomic disadvantage has a positive effect on degree success, and the most disadvantaged students are less likely to graduate from university with a good degree. However, since Tables 3.7a and 3.7b are estimating the marginal effects of a disadvantaged background across all subjects, Tables 3.8 and 3.9 estimate the same model of degree success as Table 3.6, but for STEM and non-STEM subjects respectively. It may be that the effect of a disadvantaged background in column 6 of Table 3.6 may be confined to students who study some subjects and not others.

Table 3.7a: Logit Model, by Tariff Score Bandwidth (All Subjects)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Tariff Score	0-199	200-239	240-279	280-309	310-339	340-369	370-399	400-439	440-499	500+
Disabled =1	-0.0857	-0.202	0.0670	-0.0158	-0.182	0.00129	0.0497	0.0214	-0.0570	-0.0945
	(0.109)	(0.148)	(0.107)	(0.105)	(0.133)	(0.0977)	(0.127)	(0.139)	(0.172)	(0.213)
White =1	0.536***	0.346***	0.417***	0.603***	0.524***	0.461***	0.496***	0.598***	0.522***	0.490***
	(0.111)	(0.0985)	(0.106)	(0.117)	(0.114)	(0.109)	(0.127)	(0.107)	(0.178)	(0.172)
Female =1	0.193***	0.195**	0.212***	0.190***	0.345***	0.399***	0.353***	0.539***	0.493***	0.571***
	(0.0700)	(0.0760)	(0.0630)	(0.0711)	(0.0679)	(0.0731)	(0.0820)	(0.0891)	(0.0802)	(0.112)
British =1	0.139	0.238	0.280*	0.203	0.0670	0.253	0.00676	0.315*	0.465*	0.0393
	(0.140)	(0.225)	(0.148)	(0.206)	(0.179)	(0.192)	(0.231)	(0.187)	(0.254)	(0.290)
Local =1	-0.0728	0.0263	-0.100	0.0866	0.0523	-0.0406	-0.116	-0.218	-0.0226	0.110
	(0.0940)	(0.104)	(0.0777)	(0.107)	(0.108)	(0.0847)	(0.151)	(0.134)	(0.162)	(0.288)
Low Income =1	-0.623***	-0.144	-0.489***	0.0823	0.229*	0.302***	-0.0714	-0.190	-0.445***	-0.0417
	(0.134)	(0.150)	(0.121)	(0.128)	(0.131)	(0.114)	(0.121)	(0.150)	(0.154)	(0.161)
No Educated Parent =1	-0.0430	0.0593	-0.217***	0.116	-0.107	-0.0147	-0.244***	-0.0656	-0.290***	-0.490***
	(0.0876)	(0.0833)	(0.0697)	(0.0815)	(0.0812)	(0.0740)	(0.0842)	(0.100)	(0.0853)	(0.154)
Low Income * No Educated Parent	0.482***	0.141	0.721***	-0.156	-0.173	-0.425***	0.280*	0.00734	0.679***	0.271
	(0.149)	(0.176)	(0.141)	(0.177)	(0.173)	(0.160)	(0.168)	(0.201)	(0.203)	(0.249)
Constant	-0.730***	-0.517**	0.0786	0.318	0.874***	0.554	-0.140	1.097	-0.0862	1.227***
	(0.281)	(0.241)	(2.248)	(0.542)	(0.319)	(0.589)	(0.413)	(0.938)	(0.677)	(0.344)
Observations	5,790	4,855	6,750	6,580	5,730	7,300	6,120	7,795	7,975	6,875
Subject FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
University FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

Standard errors shown in the parentheses are clustered at the institutional level across all 10 bandwidths. All regressions are performed using logit to estimate equation 32 across all subjects, which is the same as model specification 6 in Table 3.5 but without the inclusion of tariff scores. The unscaled tariff score is used to assign students to groups of similar pre-university attainment. The dependent variable is the binary degree success outcome, which takes the value 1 if the student obtained a 2:1 or 1st. The coefficients reported are the log odds of the effect of the explanatory variables on the probability of degree success, conditional on a similar level of pre-university attainment (within a particular tariff score bandwidth), and not marginal effects. Raw numbers have been rounded to the nearest multiple of 5 to comply with HESA's Standard Rounding Methodology. ***p<0.01, ** p<0.05, * p<0.1

Table 3.7b: Average Marginal Effects from Logit Model, by Tariff Score Bandwidth (All Subjects)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Tariff Score	0-199	200-239	240-279	280-309	310-339	340-369	370-399	400-439	440-499	500+
Low Income =1	-0.138***	-0.0338	-0.113***	0.0178	0.0459*	0.0515***	-0.0114	-0.0245	-0.0556***	-0.00387
	(0.0288)	(0.0350)	(0.0274)	(0.0275)	(0.0257)	(0.0185)	(0.0195)	(0.0199)	(0.0210)	(0.0151)
No Educated Parent =1	-0.00967	0.0139	-0.0501***	0.0251	-0.0219	-0.00261	-0.0388***	-0.00819	-0.0339***	-0.0473***
	(0.0197)	(0.0196)	(0.0160)	(0.0176)	(0.0166)	(0.0131)	(0.0135)	(0.0126)	(0.0100)	(0.0154)
Low Income * No Educated Parent	0.107***	0.0331	0.159***	-0.0342	-0.0358	-0.0799**	0.0420*	0.000911	0.0669***	0.0232
	(0.0321)	(0.0412)	(0.0285)	(0.0391)	(0.0364)	(0.0314)	(0.0238)	(0.0249)	(0.0170)	(0.0198)
Observations	5,790	4,855	6,750	6,580	5,730	7,300	6,120	7,795	7,975	6,875

Standard errors in parentheses. The coefficients reported are the average marginal effect of the independent variable on the probability that the dependent variable (Degree Success) is equal to 1, conditional on similar pre-university attainment (within a given tariff score bandwidth). The marginal effects are calculated following the estimation of specification 6 from Table 3.5. Raw numbers have been rounded to the nearest multiple of 5 to comply with HESA's Standard Rounding Methodology. *** p<0.01, ** p<0.05, *p<0.1

Table 3.8a once again shows that the effect of being white or female is consistently positive and significant across almost[157] all tariff bandwidths for students studying STEM subjects. The impact of being disabled on the probability of degree success is now significant for STEM subjects (where previously it was insignificant across all subjects), but only for students with lower pre-university attainment. Therefore, for students entering higher education to study a STEM subject with the lowest pre-university grades, to be disabled carries a statistically significant negative penalty in the probability of degree success compared to non-disabled students, conditional on similar pre-university attainment. It is also the case that students who are in the lowest tariff bandwidth and who study a STEM subject at a local university, are less likely to graduate with a good degree compared to those who attend a university further away.

Table 3.8b shows that, although there is evidence that a student from a disadvantaged background (using low income) within a higher tariff bandwidth are less likely to obtain degree success, and that the most disadvantaged are more likely to obtain degree success, the results averaged across all subjects within the lower tariff bandwidths are statistically insignificant. This may reflect that STEM degree courses are more difficult to secure places for in terms of entry requirements – both the higher tariff score needed, and the complexity of the A-Level subjects. This, coupled with the fact that students from lower socioeconomic groups tend to be under-represented in STEM subjects[158] may explain the lack of significant findings for the effect of disadvantaged backgrounds on degree success for STEM degrees. This therefore may be evidence that students either cannot gain significant access to studying STEM subjects due to the higher grade entrance requirements, or that students (assuming they could study STEM subjects) are self-selecting out of this

[157] The smaller sample size for STEM subjects compared to Non-STEM subjects may cause a lack of statistical power and an increase in the standard error. Therefore, it may be statistical power that is masking any effect for students studying STEM, especially when the distribution of pre-university attainment for STEM students has significantly fewer observations in the tails of the tariff distribution.

[158] See Codiroli Mcmaster (2017).

subject group. However, similar to the marginal effects calculated across all subjects, once again there is the finding that students whose tariff scores is at or near the middle of the distribution of pre-university attainment are more likely to succeed in terms of their degree classification if they come from a low income background (column 5).

Nevertheless, students who are studying a STEM subject who are the most disadvantaged and who have higher-than average pre-university attainment are more likely to obtain a good STEM degree compared to students from the higher socioeconomic status categories and/or who have a parent with university education. Conditional on pre-university attainment, the most disadvantaged students are 13.4 and 16.4 percentage points more likely to obtain a 2:1 or above in their degree classification in tariff bandwidths 370-399 and 440-499 respectively. That is an increase of 17.4% and 19.3% relative to the means as shown in Table 3.10.

Table 3.8a: Logit Model, by Tariff Score Bandwidth (STEM Subjects)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Tariff Score	0-199	200-239	240-279	280-309	310-339	340-369	370-399	400-439	440-499	500+
Disabled =1	-0.544**	-0.648**	-0.0150	0.170	0.110	-0.512*	-0.0534	-0.0663	-0.184	-0.234
	(0.232)	(0.316)	(0.279)	(0.296)	(0.267)	(0.306)	(0.347)	(0.333)	(0.213)	(0.332)
White =1	0.805***	0.360	1.191***	0.649**	0.783***	0.569**	0.750***	0.867***	0.314	0.295
	(0.310)	(0.271)	(0.240)	(0.256)	(0.259)	(0.224)	(0.238)	(0.199)	(0.276)	(0.218)
Female =1	0.616***	0.526**	0.349*	0.262	0.438**	0.380**	0.470***	0.506***	0.512***	0.495***
	(0.153)	(0.242)	(0.181)	(0.195)	(0.190)	(0.163)	(0.166)	(0.145)	(0.137)	(0.150)
British =1	0.320	-0.725	-0.116	-0.332	0.254	0.536	0.224	1.069**	0.393	-0.0684
	(0.286)	(0.451)	(0.314)	(0.427)	(0.356)	(0.350)	(0.501)	(0.434)	(0.440)	(0.417)
Local =1	-0.464**	-0.0712	-0.120	0.375	-0.129	0.141	-0.177	-0.362	-0.106	0.341
	(0.189)	(0.216)	(0.178)	(0.273)	(0.262)	(0.227)	(0.243)	(0.245)	(0.174)	(0.283)
Low Income =1	-0.487*	0.238	-0.456	0.0270	0.670***	0.355	-0.208	-0.123	-0.890***	-0.0746
	(0.271)	(0.424)	(0.291)	(0.320)	(0.237)	(0.281)	(0.311)	(0.293)	(0.274)	(0.299)
No Educated Parent =1	-0.0219	0.251	-0.0701	-0.0777	-0.305	0.0462	-0.266	0.257	-0.225	-0.580**
	(0.252)	(0.285)	(0.177)	(0.172)	(0.243)	(0.204)	(0.185)	(0.159)	(0.173)	(0.238)
Low Income * No Educated Parent	0.361	-0.469	0.880**	0.333	-0.152	-0.276	0.761*	0.0271	1.154***	0.345
	(0.342)	(0.537)	(0.374)	(0.409)	(0.390)	(0.350)	(0.399)	(0.309)	(0.371)	(0.514)
Constant	-0.826*	0.548	-0.749	0.845	-1.070**	0.786	-0.710	0.102	-1.701	1.957***
	(0.471)	(0.615)	(1.734)	(0.832)	(0.507)	(1.245)	(0.670)	(1.817)	(1.549)	(0.476)
Observations	780	670	980	945	875	1,190	1,050	1,420	1,695	1,755
Subject FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
University FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

Standard errors shown in the parentheses are clustered at the institutional level across all 10 bandwidths. All regressions are performed using logit to estimate equation 32 across STEM subjects, which is the same as model specification 6 in Table 3.5 but without the inclusion of tariff scores. The unscaled tariff score is used to assign students to groups of similar pre-university attainment. The dependent variable is the binary degree success outcome, which takes the value 1 if the student obtained a 2:1 or 1st. The coefficients reported are the log odds of the effect of the explanatory variables on the probability of degree success, conditional on a similar level of pre-university attainment (within a particular tariff score bandwidth), and not marginal effects. Raw numbers have been rounded to the nearest multiple of 5 to comply with HESA's Standard Rounding Methodology. ***p<0.01, ** p<0.05, * p<0.1

Table 3.8b: Average Marginal Effects from Logit Model, by Tariff Score Bandwidth (STEM Subjects)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Tariff Score	0-199	200-239	240-279	280-309	310-339	340-369	370-399	400-439	440-499	500+
Low Income =1	-0.105*	0.0505	-0.0962	0.00561	0.130***	0.0699	-0.0405	-0.0223	-0.168***	-0.0118
	(0.0572)	(0.0890)	(0.0597)	(0.0665)	(0.0437)	(0.0534)	(0.0611)	(0.0537)	(0.0546)	(0.0478)
No Educated Parent =1	-0.00472	0.0534	-0.0149	-0.0162	-0.0611	0.00933	-0.0513	0.0454	-0.0392	-0.0938**
	(0.0545)	(0.0602)	(0.0375)	(0.0358)	(0.0483)	(0.0412)	(0.0357)	(0.0280)	(0.0305)	(0.0394)
Low Income * No Educated Parent	0.0771	-0.0993	0.180**	0.0680	-0.0306	-0.0569	0.134**	0.00482	0.164***	0.0504
	(0.0717)	(0.112)	(0.0702)	(0.0814)	(0.0789)	(0.0730)	(0.0625)	(0.0549)	(0.0411)	(0.0695)
Observations	780	670	980	945	875	1,190	1,050	1,420	1,695	1,755

Standard errors in parentheses. The coefficients reported are the average marginal effect of the independent variable on the probability that the dependent variable (Degree Success) is equal to 1 for STEM subjects, conditional on similar pre-university attainment (within a given tariff score bandwidth). The marginal effects are calculated following the estimation of specification 6 from Table 3.5. Raw numbers have been rounded to the nearest multiple of 5 to comply with HESA's Standard Rounding Methodology. *** p<0.01, ** p<0.05, *p<0.1

If the majority of the findings averaged across all subjects do not originate from students taking STEM subjects, then it must be the case that they originate from students taking non-STEM subjects. Tables 3.9a and 3.9b present the bandwidth results from the non-STEM subpopulation of students, and once again being white or female is a significantly positive predictor of degree success across most tariff bandwidths.

In evaluating the marginal effects, the same pattern exhibited in Table 3.7b is evident. Students from a low income background or who are known to not have a parent with university education are significantly less likely to obtain degree success conditional on pre-university attainment, at the both the upper and lower end of the tariff distribution. The finding that students who are from the most disadvantaged backgrounds are more likely to obtain a 2:1 or above, conditional on pre-university attainment is evident for non-STEM subjects, especially for students whose pre-university attainment was relatively low. The marginal effects for columns 1 and 3 are almost identical when estimated across all subjects (Table 3.7b) and across non-STEM subjects (Table 3.9b). This suggests that students from the lower tariff bandwidths who study STEM subject do not incur the same penalty of a disadvantaged background, possibly due to self-selection out of studying STEM subjects, or because students who have decided (and are able) to study a STEM subject with a lower tariff score are more able and more determined to succeed in their degree studies. Furthermore, the results suggests that for students studying non-STEM subjects with higher tariff scores, coming from a low income background is not significant in determining degree success; but the effect of not having a parent with university education is significant. Additionally, the positive effect of potentially being the most socioeconomically disadvantaged only applies for the lower bandwidths of non-STEM subjects.

Ultimately, the results lend weight to the argument that tariff scores are an imperfect predictor of ability for students from disadvantaged backgrounds, such that true ability is higher than indicated ability. This may especially be true for students who experience the most socioeconomic disadvantages, such as being from a low income household and where neither parent has a university level education. In that case, it is reasonable to expect that students who are disadvantaged are less likely to obtain a good degree conditional on pre-university attainment, especially at lower levels of indicated ability. However, where a student experiences significant disadvantage, the indicated tariff score may be such a poor predictor for university attainment and true ability that the most disadvantaged students outperform their peers, holding all things constant. This may especially be true of non-STEM subjects, where disadvantaged students with lower tariff scores are both less likely and less able to study them. This is the pattern that is observed in Tables 3.7-3.9.

For students whose pre-university attainment lies in the middle of the distribution in terms of tariff score, both Tables 3.8b and 3.9b show that coming from a lower income background has a positive effect on the probability of degree success. This may be the case if the tariff score is a good predictor of ability for these students, such that the underperformance relative to ability for the lower tariff score students does not apply. Thus, conditional on pre-university attainment, students who obtain average grades and are from a low income background may be more determined to succeed than their peers. Students who are the most disadvantaged though, and whose tariff score may be a good predictor or ability, may be less likely to succeed due to financial constraints or the lack of coping skills.

Table 3.9a: Logit Model, by Tariff Score Bandwidth (Non-STEM Subjects)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Tariff Score	0-199	200-239	240-279	280-309	310-339	340-369	370-399	400-439	440-499	500+
Disabled =1	-0.0434	-0.173	0.0640	-0.0411	-0.283*	0.122	0.0450	0.0633	-0.0406	0.00458
	(0.130)	(0.163)	(0.128)	(0.109)	(0.150)	(0.101)	(0.143)	(0.156)	(0.220)	(0.256)
White =1	0.525***	0.343***	0.281***	0.594***	0.496***	0.453***	0.415***	0.469***	0.617***	0.675***
	(0.109)	(0.110)	(0.108)	(0.122)	(0.118)	(0.117)	(0.140)	(0.129)	(0.166)	(0.194)
Female =1	0.148*	0.145*	0.201***	0.187**	0.348***	0.399***	0.322***	0.505***	0.454***	0.592***
	(0.0795)	(0.0783)	(0.0696)	(0.0731)	(0.0722)	(0.0809)	(0.101)	(0.117)	(0.0966)	(0.156)
British =1	0.125	0.445**	0.363**	0.386*	-0.0243	0.226	-0.0445	0.0450	0.565*	0.404
	(0.182)	(0.218)	(0.167)	(0.228)	(0.236)	(0.255)	(0.274)	(0.297)	(0.331)	(0.391)
Local =1	0.00794	0.0450	-0.0877	0.0616	0.111	-0.0778	-0.110	-0.105	0.0201	-0.0841
	(0.101)	(0.125)	(0.0820)	(0.104)	(0.118)	(0.0892)	(0.167)	(0.144)	(0.210)	(0.362)
Low Income =1	-0.629***	-0.179	-0.501***	0.0857	0.182	0.289**	0.0129	-0.234	-0.218	0.000248
	(0.153)	(0.155)	(0.127)	(0.139)	(0.148)	(0.131)	(0.160)	(0.176)	(0.238)	(0.218)
No Educated Parent =1	-0.0467	0.0532	-0.243***	0.131	-0.0891	-0.0293	-0.248**	-0.187	-0.292***	-0.458**
	(0.0839)	(0.0923)	(0.0738)	(0.0844)	(0.0933)	(0.0812)	(0.0990)	(0.126)	(0.0861)	(0.192)
Low Income * No Educated Parent	0.499***	0.207	0.711***	-0.205	-0.178	-0.433**	0.121	0.0341	0.461*	0.172
	(0.170)	(0.179)	(0.146)	(0.183)	(0.193)	(0.173)	(0.198)	(0.237)	(0.275)	(0.255)
Constant	-0.766**	-0.710***	-0.810	-0.513	0.402	-0.722*	-0.176	0.529	0.0753	0.637
	(0.359)	(0.274)	(1.173)	(0.546)	(0.337)	(0.399)	(0.497)	(0.556)	(0.488)	(0.454)
Observations	4,970	4,160	5,750	5,585	4,815	6,035	5,010	6,300	6,160	5,060
Subject FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
University FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

Standard errors shown in the parentheses are clustered at the institutional level across all 10 bandwidths. All regressions are performed using logit to estimate equation 32 across non-STEM subjects, which is the same as model specification 6 in Table 3.5 but without the inclusion of tariff scores. The unscaled tariff score is used to assign students to groups of similar pre-university attainment. The dependent variable is the binary degree success outcome, which takes the value 1 if the student obtained a 2:1 or 1st. The coefficients reported are the log odds of the effect of the explanatory variables on the probability of degree success, conditional on a similar level of pre-university attainment (within a particular tariff score bandwidth), and not marginal effects. Raw numbers have been rounded to the nearest multiple of 5 to comply with HESA's Standard Rounding Methodology. ***p<0.01, ** p<0.05, * p<0.1

Table 3.9b: Average Marginal Effects from Logit Model, by Tariff Score Bandwidth (Non-STEM Subjects)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Tariff Score	0-199	200-239	240-279	280-309	310-339	340-369	370-399	400-439	440-499	500+
Low Income =1	-0.138***	-0.0416	-0.115***	0.0183	0.0357	0.0471**	0.00190	-0.0267	-0.0222	1.69e-05
	(0.0325)	(0.0359)	(0.0287)	(0.0296)	(0.0285)	(0.0205)	(0.0236)	(0.0209)	(0.0255)	(0.0149)
No Educated Parent =1	-0.0104	0.0124	-0.0561***	0.0283	-0.0178	-0.00498	-0.0371**	-0.0207	-0.0290***	-0.0331**
	(0.0187)	(0.0216)	(0.0168)	(0.0181)	(0.0187)	(0.0138)	(0.0149)	(0.0140)	(0.00870)	(0.0145)
Low Income * No Educated Parent	0.110***	0.0482	0.156***	-0.0447	-0.0362	-0.0784**	0.0174	0.00369	0.0400*	0.0112
	(0.0364)	(0.0414)	(0.0294)	(0.0403)	(0.0399)	(0.0330)	(0.0279)	(0.0255)	(0.0211)	(0.0157)
Observations	4,970	4,160	5,750	5,585	4,815	6,035	5,010	6,300	6,160	5,060

Standard errors in parentheses. The coefficients reported are the average marginal effect of the independent variable on the probability that the dependent variable (Degree Success) is equal to 1 for non-STEM subjects, conditional on similar pre-university attainment (within a given tariff score bandwidth). The marginal effects are calculated following the estimation of specification 6 from Table 3.5. Raw numbers have been rounded to the nearest multiple of 5 to comply with HESA's Standard Rounding Methodology. *** p<0.01, ** p<0.05, *p<0.1

Table 3.10: Predicted Probabilities of Degree Success, by Tariff Bandwidth & Subject Group

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Tariff Score	0-199	200-239	240-279	280-309	310-339	340-369	370-399	400-439	440-499	500+	ALL
All	0.440***	0.492***	0.579***	0.657***	0.696***	0.765***	0.803***	0.863***	0.879***	0.915***	0.745***
	(0.000322)	(0.000858)	(0.000322)	(0.000887)	(0.000858)	(0.00107)	(0.00185)	(0.00258)	(0.00224)	(0.00310)	(0.000880)
STEM	0.461***	0.509***	0.524***	0.634***	0.631***	0.672***	0.704***	0.748***	0.768***	0.814***	0.691***
	(0.00197)	(0.00114)	(0.00258)	(0.00208)	(0.00197)	(0.00246)	(0.00310)	(0.00408)	(0.00290)	(0.00513)	(0.00143)
Non-STEM	0.432***	0.489***	0.582***	0.662***	0.710***	0.780***	0.822***	0.885***	0.900***	0.941***	0.761***
	(0.000466)	(0.000130)	(0.000378)	(0.000910)	(0.000960)	(0.00101)	(0.00157)	(0.00198)	(0.00195)	(0.00258)	(0.000837)
N (All)	5,745	4,830	6,420	6,525	5,690	7,225	6,060	7,720	7,860	6,815	64,895
N (STEM)	780	670	670	945	875	1,190	1,050	1,420	1,695	1,755	11,050
N (Non-STEM)	4,970	4,160	5,750	5,585	4,815	6,035	5,010	6,300	6,160	5,060	53,845

Standard errors in parentheses. The coefficients reported are the base probabilities that the dependent variable (Degree Success) is equal to one, when all explanatory variables are set at their mean values within each tariff bandwidth and subject group. Raw numbers have been rounded to the nearest multiple of 5 to comply with HESA's Standard Rounding Methodology. *** p<0.01, ** p<0.05, * p<0.1

3.6 Conclusions & Discussion

This chapter has evaluated whether coming from a disadvantaged background, after controlling for other factors, has any effect on degree outcomes. A mixture of ordered and binary nonlinear models were used to measure the determinants of degree success, including differing measures of a disadvantaged student. The research design and the data used are a significant improvement on the existing studies estimating the impact of socioeconomic disadvantage on the probability of obtaining a good degree. In particular, the estimation of a model of degree success using data rich with socioeconomic indicators, the use of individual subject fixed effects, the estimation of the model separately for different subject groupings, and by using student data from every university in the UK, all help contribute to a significant addition to the empirical literature.

Ultimately, whilst coming from a lower income background or not having a parent with prior university experience has a significantly negative impact on the probability of degree success, this is mainly true for students with low levels of pre-university attainment, and predominantly for non-STEM subjects. However, students that are the most disadvantaged are, on the contrary, more likely to obtain degree success, across all subject groupings, especially for higher and lower than average levels of pre-academic attainment.

This may suggest that students who are the most disadvantaged are not able to fulfil their potential at school or college, but when studying at university that potential for higher attainment is shown in the higher probability of degree success. Alternatively, it may be that students who are the most disadvantaged, conditional on attainment, place a higher value on the degree and its returns, and therefore are more determined to succeed than their counterparts who have not had to overcome such disadvantages. In both scenarios but particularly the former, this would support the growing utilisation of contextual admissions policies since pre-university attainment for the most disadvantaged pupils may not be a true reflection of true academic ability at the HE level. The implication is therefore, although pre-university attainment is an important and positive determinant in a student’s university performance, prior attainment can also be a reflection of disadvantage, rather than a true predictor of ability.

What cannot be argued, however, is that university study is not a level playing field: the extent to which a student is considered socioeconomically disadvantaged has a statistically and quantitatively significant impact on their likelihood of obtaining a degree success. To refer back to the example in section 3.1.2, student A (the disadvantaged) may outperform student B, or he/she may underperform; it depends both on the extent to which they are disadvantaged, and whether their tariff scores are likely to be a good or poor predictor of true ability. Given that the findings of this chapter are also conditional on the student having decided to participate in higher education, it may also be that these findings underestimate the true potential in the most socioeconomically disadvantaged students, as these are the students who are empirically found to be less likely to attend university in the first place.

Concluding Remarks

This thesis has presented three empirical analyses that address two central questions in higher education (HE), namely: what factors affect student participation behaviour, and whether a student's socioeconomic background affects a student's chances of academic success at university. What is clear from all three analyses is that students who participate in HE are extremely diverse in their backgrounds, their motivations and their choices. If policy-makers and researchers do not acknowledge the extent of heterogeneity that exists, and design policies and empirical strategies accordingly, there may be unintended consequences.

At an empirical level, if policies are evaluated at an aggregate level across all subjects and all students (BIS 2010a; BIS 2010b; Universities UK, 2009), the impact may be insignificant. Even if empirical strategies do acknowledge heterogeneity, a control group may not be used (Dearden et al., 2011; BIS, 2010a) which makes the estimation of the impact of the policy difficult to interpret, given there is no counterfactual outcome. Furthermore, if a natural experiment setting is used, the treatment and control groups may not be sufficiently comparable (e.g. Faggian, 2010) thus potentially affecting the validity of the estimates of the impact of a policy. When a comprehensive acknowledgement of student heterogeneity is made, along with allowing a policy to impact different groups of students in different ways across different subjects and different modes of study, the singular estimate of a policy introduction is disaggregated.

Even when there is no policy evaluation, the diversity of the student population must be taken into account, as Chapters 2 and 3 show. For example, although students who are disadvantaged may have lower academic attainment in terms of degree classification (HEFCE, 2014, 2015a; Blundell et al., 2000; McNabb et al., 2002; Smith & Naylor, 2001a) there may be a different effect for those students that experience the most significant socioeconomic disadvantages. Furthermore, although it was found in Chapter 2 that students who experience high or rising levels of local unemployment were more likely to choose a subject with the lowest graduate wage and employment premia, the opposite effect was found for students from the lower socioeconomic classifications. Thus, local unemployment does not affect subject choice homogenously across student backgrounds and arguably therefore across differing perceptions of the relative costs and benefits to a university degree.

From a policy perspective, the results from these empirical studies show that student heterogeneity is significant and unless addressed, a policy change may have negative impacts on the most disadvantaged students. When the funding reforms of 2006 were introduced, the increase in the means-tested, non-repayable maintenance grant allowed students from the lowest socioeconomic distribution to experience greater geographic mobility. The result is that the poorest students are no longer confined through financial concerns to attend university in the local area, which may have significantly impacted the range of suitable courses and universities. When the tuition fees were increased to £9,000 per year in 2012, and when the means-tested, non-repayable maintenance grants were removed in favour of maintenance loans (which would further add to student debt) in 2016, this has almost certainly had a detrimental impact on geographic mobility. Chapter 1's finding that students who experienced the largest increase in costs but still classified as low income were more likely to attend a local university and live at home should be a cautionary note. If the fear of debt does constrain students' choices, as Callender and Jackson (2008) propose, then increasing those fears through higher tuition fees and taking out loans to fund geographic mobility may lead to students from low income families choosing subjects that do not represent an optimal investment, or not participating in HE at all. This is echoed in the finding in Chapter 2, where students with higher local unemployment are more likely to study subjects with lower graduate returns. The policy implication here, is that some students living in areas of high or worsening unemployment may benefit from additional support in choosing degree subjects that offer the best investment, rather than choosing subjects that are the least likeliest to fail.

A further policy perspective can also be presented in the context of student attainment. The findings of the Chapter 3 strongly support the existing empirical literature (HEFCE, 2014, 2015a; Blundell et al., 2000; McNabb et al., 2002; Smith & Naylor, 2001a) that, conditional on pre-university attainment and student characteristics, some students perform worse in their degree outcomes if they are from a disadvantaged background. Thus, students from low HE participation areas, low income families, or those who are the first to enter HE from their immediate family should be offered additional support in adjusting to university life and study. However, there is also strong evidence in support of contextual admission policies, since students who are the most socioeconomically disadvantaged outperform their peers in terms of degree classification, all else being equal. Since this conditional pre-university attainment is assigned by an index of school grades (the UCAS tariff score), it may be the case that for students with low levels of pre-university attainment, the measure of ability is significantly underestimating true ability for the most disadvantaged students. Additionally, the most disadvantaged students who have high levels of pre-university attainment may, holding all else equal, outperform their peers due to a greater ability and desire to succeed at university since they have potentially had to overcome socioeconomic obstacles.

One final note that should be made is that the results presented in these chapters have attempted to focus on specific research questions, and in doing so the empirical approach has been to control for all possible, observable confounding factors. Some of these controls – such as university quality and subject-specific effects – may indeed be outcomes, particularly in chapters 1 and 2. It therefore offers further avenues of potential research to instead consider these conditioning factors as outcomes.

Appendices

Appendix A: Appendix to Chapter 1

Table A1: HEI Members of the Russell Group

University	***Year Joined Russell Group***
Imperial College London	1994
London School of Economics	1994
University of Birmingham	1994
University of Bristol	1994
University of Cambridge	1994
University of Edinburgh	1994
University of Glasgow	1994
University of Leeds	1994
University of Liverpool	1994
University of Manchester	1994
University of Newcastle-Upon-Tyne	1994
University of Nottingham	1994
University of Oxford	1994
University of Sheffield	1994
University of Southampton	1994
University College London	1994
University of Warwick	1994
Cardiff University	1998
King’s College, London	1998
Queen’s University Belfast	2006
Durham University	2012
Queen Mary University of London	2012
University of Exeter	2012
York University	2012

Table A2: A Comparison of the Old and New Funding Systems by Household Income

Old System in 2006					**New 2006 Reform**					**Gain/Loss**
Income (£)	Fees (£)	Fee Subsidy (£)	HE Grant (£)	Net (£)	Income (£)	Fees (£)	2006 Grant (£)	Bursary (£)	Net (£)	(£)
0	-1200	1200	1000	1000	0	-3000	2700	300	0	-1000
2500	-1200	1200	1000	1000	2500	-3000	2700	300	0	-1000
5000	-1200	1200	1000	1000	5000	-3000	2700	300	0	-1000
7500	-1200	1200	1000	1000	7500	-3000	2700	300	0	-1000
10000	-1200	1200	1000	1000	10000	-3000	2700	300	0	-1000
12500	-1200	1200	1000	1000	12500	-3000	2700	300	0	-1000
15000	-1200	1200	1000	1000	15000	-3000	2700	300	0	-1000
17500	-1200	1200	616	616	17500	-3000	2700	300	0	-616
20000	-1200	1200	232	232	20000	-3000	2283	0	-717	-949
22500	-1200	1200	0	0	22500	-3000	1866	0	-1134	-1134
25000	-1200	1000	0	-200	25000	-3000	1449	0	-1551	-1351
27500	-1200	750	0	-450	27500	-3000	1041	0	-1959	-1509
30000	-1200	500	0	-700	30000	-3000	778	0	-2222	-1522
32500	-1200	250	0	-950	32500	-3000	515	0	-2485	-1535

35000	-1200	0	0	-1200	35000	-3000	252	0	-2748	-1548
37500	-1200	0	0	-1200	37500	-3000	0	0	-3000	-1800
40000	-1200	0	0	-1200	40000	-3000	0	0	-3000	-1800
42500	-1200	0	0	-1200	42500	-3000	0	0	-3000	-1800
45000	-1200	0	0	-1200	45000	-3000	0	0	-3000	-1800
47500	-1200	0	0	-1200	47500	-3000	0	0	-3000	-1800
50000	-1200	0	0	-1200	50000	-3000	0	0	-3000	-1800
52500	-1200	0	0	-1200	52500	-3000	0	0	-3000	-1800
55000	-1200	0	0	-1200	55000	-3000	0	0	-3000	-1800
57500	-1200	0	0	-1200	57500	-3000	0	0	-3000	-1800
60000	-1200	0	0	-1200	60000	-3000	0	0	-3000	-1800

Table A3: Data Cropping of the Original HESA Data

Observations	Condition	Observations Dropped	Observations Remaining
3,610,180	Remove observations from The Open University	118,650	3,491,530
3,491,530	Remove exchange students	115,900	3,375,630
3,375,630	Remove students whose gender is recorded as unknown	250	3,375,385
3,375,385	Keeping only undergraduate students registered for their first degree	1,649,990	1,725,400
1,725,400	Keeping only students who pay home fees	182,385	1,543,010
1,543,010	Removing students with unknown or missing socioeconomic code	513,350	1,029,660
1,029,660	Removing observations from HEIs not included in Table A5 (i.e. vocational HEIs)	59,520	970,140
970,140	Removing observations with missing age	20	970,120
970,120	Keeping only full time students	115,420	854,700
854,700	Removing observations with unknown or Non-UK postcodes at the time of application	19,110	835,590
835,590	Keeping only observations who commenced studies in the academic years 2003/04 to 2007/08	77,105	758,485

Note:

all the above figures ending in 0, 1 and 2 are rounded to 0, and all other figures are rounded to the nearest 5. This is to comply with HESA's rounding strategy to protect student anonymity. As a result, discrepancies are present.

Table A4: Median Gross Earnings by SOC 2000 and Socioeconomic Status

SOC 2000 Code[159]	**SOC 2000 Broad Category**	**Median Annual Gross Earnings (£)**[160]
1	Managers and senior officials	£32, 928
2	Professional occupations	£31, 402
3	Associate professional and technical occupations	£24, 371
4	Administrative and secretarial occupations	£15, 063
5	Skilled trades occupations	£21, 207
6	Personal service occupations	£11, 066
7	Sales and customer service occupations	£8, 623
8	Process, plant and machine operatives	£19,330

[159] The SOC 2000 classification is not mapped to the socioeconomic code perfectly as it is a complex hierarchical process, however this table provides an indication. For instance, the majority of the occupations coded 8 (Process, plant and machine operatives) are coded as being socioeconomic code 6. For full details of the mapping, see http://webarchive.nationalarchives.gov.uk/20160106042025/http://www.ons.gov.uk/ons/guide-method/classifications/current-standard-classifications/soc2010/soc2010-volume-3-ns-sec--rebased-on-soc2010--user-manual/section-13--deriving-the-ns-sec--full--reduced-and-simplified-methods.pdf

[160] Data from the 2006 release of the Annual Survey of Hours and Earnings, Table 2.7a: “Annual pay - Gross (£) - For all employee jobs: United Kingdom, 2006”. Figures reported are the median gross earnings for each broad SOC 2000 category.

Table A5: List of HEIs (Gibbons-Vignoles Augmented)

England

Anglia Ruskin University	University of Central Lancashire
Aston University	University of Chester
Bath Spa University	University of Chichester
Bournemouth University	University of Derby
Brunel University	University of Durham
Buckinghamshire Chilterns University	University of East Anglia
Canterbury Christ Church University	University of East London
City University	University of Essex
College of St Mark and St John	University of Exeter
Coventry University	University of Gloucestershire
De Montfort University	University of Greenwich
Edge Hill College of Higher Education	University of Hertfordshire
Goldsmiths College	University of Huddersfield
Harper Adams University College	University of Hull
Imperial College of Science, Technology	University of Keele
King's College London	University of Kent
Kingston University	University of Lancaster
Leeds Metropolitan University	University of Leeds
Liverpool Hope University	University of Leicester
Liverpool John Moores University	University of Lincoln
London Metropolitan University	University of Liverpool
London School of Economics and Political Science	University of Luton (University of Bedfordshire)
London South Bank University	University of Manchester
Loughborough University	University of Newcastle-upon-Tyne
Manchester Metropolitan University	University of Northampton
Middlesex University	University of Northumbria at Newcastle
Newman College of HE	University of Nottingham

Nottingham Trent University
Oxford Brookes University
Queen Mary and Westfield College
Roehampton University
Royal Holloway and Bedford New College
Sheffield Hallam University
Southampton Solent University
St Mary's College
Staffordshire University
Thames Valley University
Trinity and All Saints College
University College Falmouth
University College London
University of Bath
University of Birmingham
University of Bolton
University of Bradford
University of Brighton
University of Bristol
University of Cambridge
University of Central England in Birmingham
University of Oxford
University of Plymouth
University of Portsmouth
University of Reading
University of Salford
University of Sheffield
University of Southampton
University of Sunderland
University of Surrey
University of Sussex
University of Teesside
University of Warwick
University of Westminster
University of Winchester
University of Wolverhampton
University of Worcester
University of York
University of the West of England, Bristol
York St John College

Wales

Aberystwyth University
Bangor University
Cardiff University
Glyndwr University
Swansea Metropolitan University
Swansea University
Trinity University College
University of Glamorgan
University of Wales Institute, Cardiff
University of Wales, Lampeter
University of Wales, Newport

Scotland

Edinburgh Napier University	University of Dundee
Glasgow Caledonian University	University of Edinburgh
Heriot-Watt University	University of Glasgow
Queen Margaret University, Edinburgh	University of Strathclyde
Robert Gordon University	University of St Andrews
University of Aberdeen	University of Stirling
University of Abertay, Dundee	University of the West of Scotland

Table A6: Subject Groupings for DD and DDD

JACS Code & Subject	STEM Indicator	HFI/non-HFI
A Medicine & Dentistry	STEM	HFI
B Allied to Medicine	STEM	HFI
C Biological Sciences	STEM	non-HFI
D Veterinary Sciences & Agriculture	STEM	-
F Physical Sciences	STEM	-
G Mathematical & Computer Sciences	STEM	HFI
H Engineering	STEM	HFI
J Technologies	STEM	non-HFI
K Architecture, Building & Planning	STEM	-
L Social Studies	non-STEM	-
M Law	non-STEM	HFI
N Business & Administrative Studies	non-STEM	HFI
P Mass Communication	non-STEM	non-HFI
Q Linguistics & Classics	non-STEM	non-HFI
R European Languages & Literature	non-STEM	non-HFI
T non-European Languages & Lit	non-STEM	non-HFI
V Historical and Philosophical Studies	non-STEM	non-HFI
W Creative Arts & Design	non-STEM	non-HFI
X Education	non-STEM	-

STEM: STEM subjects are classified as those identified as being Broad STEM subjects, following http://www.publications.parliament.uk/pa/ld201213/ldselect/ldsctech/37/3705.htm
The STEM classification is performed using 2-digits JACS codes, to distinguish between STEM and non-STEM courses within a 1-digit code. See section 1.4.3 for a full discussion.
HFI: HFI classification calculated using individual rates of return to degree subjects as calculated by BIS (2011), Figure 13. Missing HFI classification is due to subject having close to average returns, so neither High nor Low future income.

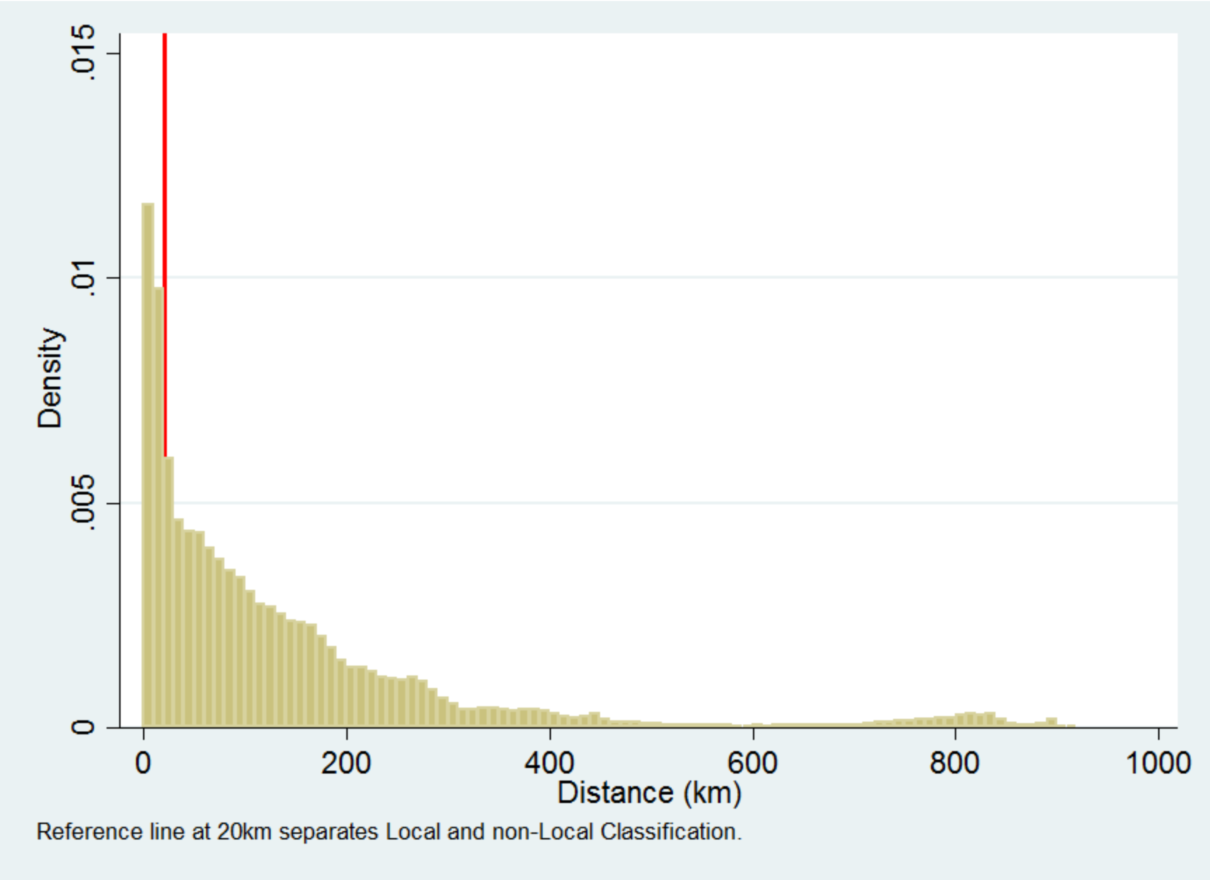
Table A7: Consistency of Baseline DiD Estimates across Estimation Type

				All Universities			Non-Elite Universities		
				Probit	Logit	LPM	Probit	Logit	LPM
				(1)	(2)	(3)	(4)	(5)	(6)
Panel A									
SEC Code	4-8								
		Treatment Period	Average	0.0340	0.0576	0.0117	0.0169	0.0290	0.00642
				(0.0262)	(0.0422)	(0.00939)	(0.0183)	(0.0301)	(0.00647)
			2006	-0.00484	-0.00849	-0.00251	-0.00964	-0.0165	-0.00371
				(0.0160)	(0.0261)	(0.00554)	(0.0137)	(0.0224)	(0.00497)
			2007	0.0459**	0.0773**	0.0166**	0.0285	0.0485*	0.0110*
				(0.0229)	(0.0368)	(0.00837)	(0.0177)	(0.0292)	(0.00627)
Panel B									
SEC Code	4-5	Treatment Period	Average	0.0494**	0.0954**	0.0103**	0.0386	0.0746	0.00859
				(0.0223)	(0.0423)	(0.00484)	(0.0282)	(0.0528)	(0.00648)
			2006	-0.0233	-0.0476	-0.00539	-0.0330	-0.0653	-0.00757
				(0.0264)	(0.0499)	(0.00573)	(0.0307)	(0.0564)	(0.00709)
			2007	0.0808**	0.157**	0.0172**	0.0751*	0.146*	0.0168*
				(0.0316)	(0.0619)	(0.00715)	(0.0424)	(0.0812)	(0.0100)
Panel C									
SEC Code	6-8	Treatment Period	Average	-0.00250	0.00219	0.00141	-0.0194	-0.0242	-0.00217

		(1)	(2)	(3)	(4)	(5)	(6)
		(0.0436)	(0.0707)	(0.00987)	(0.0512)	(0.0827)	(0.0104)
	2006	0.0137	0.0253	0.00288	0.0157	0.0270	0.00386
		(0.0125)	(0.0225)	(0.00334)	(0.0144)	(0.0254)	(0.00413)
	2007	-0.0140	-0.0174	-0.000618	-0.0356	-0.0497	-0.00584
		(0.0550)	(0.0888)	(0.0125)	(0.0672)	(0.108)	(0.0147)
Fixed Effects		University	University	University	University	University	University
Controls		yes	yes	yes	yes	yes	yes
Obs (SEC Code 4-8)		691,475	691,475	691,475	444,395	444,395	444,395
Obs (SEC Code 4-5)		691,475	691,475	691,475	444,395	444,395	444,395
Obs (SEC Code 6-8)		691,475	691,475	691,475	444,395	444,395	444,395

This table reports the results from the DiD estimation as outlined by equation (4), estimated across all universities (columns 1-3) and then across non-elite universities (columns 4-6). Elite status is denoted by membership of the Russell Group of universities, as seen in Appendix A, Table A1. This table establishes the consistency of the DiD estimates to linear and non-linear regression methods, where columns 1 and 4 estimate equation (4) using probit, columns 2 and 5 estimates equation (4) using logit, and columns 3 and 6 estimate equation (4) using OLS. Columns 3 and 6 are therefore identical to columns 1 and 3 in Table 1.6. The coefficients reported are the coefficients on the interaction between the treatment group indicator (English HEIs) and the treatment period indicator, and show the change in the probability of observing a low income student studying as a result of the 2006 funding reforms. To allow for the estimation of the effect of the policy averaged across the post-policy period and to capture the partial and full adjustment effects to the policy change, the DiD is estimated separately for the average of the post-policy cohorts, and for the 2006/07 and 2007/08 cohorts. There are therefore three rows of DiD estimates for each column, where each coefficient relates to the specific treatment period. There are three panels (A, B and C) to allow an estimation of the DiD across the broad low income background category (Panel A, SEC Code 4-8), and disaggregated across the upper-low income background category of students (Panel B, SEC Code 4-5) and the lower-low income background category of students (Panel C, SEC Code 6-8). University fixed effects are included for all specifications, as are controls for observable student heterogeneity. Namely, these individual controls are: gender, ethnicity, age, disability status, and nationality. Standard errors are shown in parentheses and are clustered at the HEI level. Raw numbers have been rounded to the nearest multiple of 5 to comply with HESA’s Standard Rounding Methodology. *** p<0.01, ** p<0.05, * p<0.1

Figure A1: Distance Histogram and Local Threshold



Appendix B: Appendix to Chapter 2

Table B1: NUTS 2 (2010) Coding and Labels

NUTS 1 (2010) Code	**Data Code**	**NUTS 2 (2010) Code**	**Data Code**	**Label**
UKC	1	UKC1	1	Tees Valley & Durham
UKC	1	UKC2	2	Northumberland & Tyne and Wear
UKD	2	UKD1	3	Cumbria
UKD	2	UKD3	4	Greater Manchester
UKD	2	UKD4	5	Lancashire
UKD	2	UKD6	6	Cheshire
UKD	2	UKD7	7	Merseyside
UKE	3	UKE1	8	East Yorkshire & Northern Lincolnshire
UKE	3	UKE2	9	North Yorkshire
UKE	3	UKE3	10	South Yorkshire
UKE	3	UKE4	11	West Yorkshire
UKF	4	UKF1	12	Derbyshire & Nottinghamshire
UKF	4	UKF2	13	Leicestershire, Rutland & Northamptonshire
UKF	4	UKF3	14	Lincolnshire
UKG	5	UKG1	15	Herefordshire, Worcestershire & Warwickshire
UKG	5	UKG2	16	Shropshire & Staffordshire
UKG	5	UKG3	17	West Midlands
UKH	6	UKH1	18	East Anglia
UKH	6	UKH2	19	Bedfordshire & Hertfordshire
UKH	6	UKH3	20	Essex
UKI	7	UKI1	21	Inner London

UKI	7	UKI2	22	Outer London
UKJ	8	UKJ1	23	Berkshire, Buckinghamshire & Oxfordshire
UKJ	8	UKJ2	24	Surrey, East and West Sussex
UKJ	8	UKJ3	25	Hampshire & Isle of Wight
UKJ	8	UKJ4	26	Kent
UKK	9	UKK1	27	Gloucestershire, Wiltshire & Bristol/Bath Area
UKK	9	UKK2	28	Dorset & Somerset
UKK	9	UKK3	29	Cornwall & Isles of Scilly
UKK	9	UKK4	30	Devon
UKL	10	UKL1	31	West Wales & The Valleys
UKL	10	UKL2	32	East Wales
UKM	11	UKM2	33	Eastern Scotland
UKM	11	UKM3	34	South Western Scotland
UKM	11	UKM5	35	North Eastern Scotland
UKM	11	UKM6	36	Highlands & Islands
UKN	12	UKN0	37	Northern Ireland

Table B2: Subject Groupings by JACS Code

JACS Code	Subject Area	STEM [161]	LEM [162]	Other [163]
A	Medicine & dentistry	Y		
B	Subjects allied to medicine	Y/N		Y/N
C	Biological sciences	Y/N		Y/N
D	Veterinary science, agriculture & related subjects	Y/N		Y/N
F	Physical sciences	Y/N		Y/N
G	Mathematical & computer sciences	Y		
H	Engineering	Y		
J	Technologies	Y/N		Y/N
K	Architecture, building & planning			Y
L	Social studies		Y/N	Y/N
M	Law		Y	
N	Business & administrative studies		Y	
P	Mass communications & documentation			Y
Q	Linguistics, classics & related subjects			Y
R	European languages, literature & related subjects			Y
T	Eastern, Asiatic, African, American and Australasian languages, literature & related subjects			Y
V	Historical & philosophical studies			Y
W	Creative arts & design			Y
X	Education			Y
Y	Combined			Y

[161] STEM subjects are classified as STEM through the JACS coding system using the 2-digit JACS code, where only Broad STEM subjects are defined as STEM, as given by http://www.publications.parliament.uk/pa/ld201213/ldselect/ldsctech/37/3705.htm
[162] In line with the Britton et al. (2016) definition of LEM subjects, only Economics (JACS 2 digit code 'L1') is classified as LEM within the social studies (L) classification in addition to JACS codes M and N.
[163] All remaining subjects from classifications B, C, D and F that are not classified as STEM, and all remaining subjects within the L classification that are not Economics (L1) are classified as 'Other'.

Table B3: Full Regression Results from Table 2.5

	(1) STEM	(2) STEM	(3) STEM	(4) STEM	(5) STEM	(6) STEM	(7) STEM	(8) STEM	(9) STEM
Tariff	-0.00216**	-0.00210**	-0.00228**	-0.00212**	-0.00195**	-0.00195**	-0.00228**	-0.00213**	-0.00228**
	(0.000963)	(0.000972)	(0.000965)	(0.000969)	(0.000971)	(0.000971)	(0.000965)	(0.000968)	(0.000965)
Female	-0.117***	-0.119***	-0.116***	-0.119***	-0.122***	-0.122***	-0.116***	-0.119***	-0.116***
	(0.00598)	(0.00599)	(0.00585)	(0.00585)	(0.00603)	(0.00603)	(0.00585)	(0.00585)	(0.00585)
White	-0.0419***	-0.0426***	-0.0405***	-0.0424***	-0.0435***	-0.0427***	-0.0409***	-0.0429***	-0.0408***
	(0.00859)	(0.00849)	(0.00856)	(0.00859)	(0.00849)	(0.00857)	(0.00851)	(0.00850)	(0.00849)
British	0.00699	0.00975	0.00447	0.0108	0.0101	0.0102	0.00427	0.0105	0.00425
	(0.0153)	(0.0164)	(0.0150)	(0.0167)	(0.0167)	(0.0165)	(0.0151)	(0.0167)	(0.0151)
Low Income	0.0181***	0.0184***	0.0189***	0.0184***	0.0192***	0.0190***	0.0190***	0.0185***	0.0190***
	(0.00294)	(0.00293)	(0.00292)	(0.00296)	(0.00292)	(0.00293)	(0.00291)	(0.00294)	(0.00291)
Local	0.0149***	0.0154***	0.0158***	0.0163***	0.0172***	0.0163***	0.0162***	0.0167***	0.0162***
	(0.00527)	(0.00518)	(0.00525)	(0.00508)	(0.00514)	(0.00519)	(0.00522)	(0.00506)	(0.00521)
Disabled	0.0118***	0.0127***	0.0130***	0.0144***	0.0139***	0.0139***	0.0130***	0.0144***	0.0131***
	(0.00426)	(0.00427)	(0.00469)	(0.00439)	(0.00407)	(0.00408)	(0.00468)	(0.00438)	(0.00468)
Local Youth Unemployment (16-24yr)	-0.00109*						-0.000812		-0.00105
	(0.000635)						(0.000775)		(0.00110)
Local Adult Unemployment (25-49yr)		-0.00231						-0.00258	0.000586
		(0.00167)						(0.00180)	(0.00244)
Growth in Local Youth Unemployment (16-24yr)			-0.00928				-0.00559		-0.00244
			(0.00624)				(0.00741)		(0.00902)
Growth in Local Adult Unemployment (25-49yr)				-0.00879				-0.00530	-0.00922
				(0.00851)				(0.00886)	(0.00962)
Aggregate Local Unemployment (16-49yr)					-0.00305**				

					(0.00142)				
Growth in Local Aggregate Unemployment (16-49yr)						-0.00539**			
						(0.00216)			
Constant	0.217***	0.213***	0.206***	0.203***	0.221***	0.201***	0.218***	0.216***	0.220***
	(0.0266)	(0.0284)	(0.0251)	(0.0276)	(0.0286)	(0.0277)	(0.0270)	(0.0290)	(0.0273)
Observations	181,090	187,825	158,720	178,995	196,355	196,355	158,720	178,995	158,720
R-squared	0.104	0.105	0.105	0.105	0.106	0.106	0.105	0.105	0.105
Year FE	✓	✓	✓	✓	✓	✓	✓	✓	✓
University FE	✓	✓	✓	✓	✓	✓	✓	✓	✓
Region FE (NUTS 1 2010)	✓	✓	✓	✓	✓	✓	✓	✓	✓
Clustered SE	✓	✓	✓	✓	✓	✓	✓	✓	✓

Standard errors shown in the parentheses are clustered at the institutional level across all 9 specifications. All regressions are performed using OLS to estimate equation 9. The dependent variable is the binary subject group indicator, which takes the value 1 for STEM and 0 otherwise. All 9 specifications are estimating column (3) from Table 2.4 with various indicators of unemployment. Specifications (1)-(6) estimate the unemployment variables separately, where specifications (7)-(8) include the level of local unemployment and the growth in local unemployment respectively. Specification (9) includes all measures of youth and adult local unemployment. Raw numbers have been rounded to the nearest multiple of 5 to comply with HESA's Standard Rounding Methodology. *** p<0.01, ** p<0.05, * p<0.1

<u>Table B4: Full Regression Results from Table 2.6</u>

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	LEM	LEM	LEM	LEM	LEM	LEM	LEM	LEM	LEM
Tariff	0.00444***	0.00442***	0.00432***	0.00437***	0.00439***	0.00439***	0.00432***	0.00437***	0.00428***
	(0.00102)	(0.00101)	(0.00104)	(0.00102)	(0.000997)	(0.000998)	(0.00104)	(0.00102)	(0.00104)
Female	-0.0645***	-0.0647***	-0.0651***	-0.0650***	-0.0629***	-0.0629***	-0.0651***	-0.0650***	-0.0650***
	(0.00494)	(0.00489)	(0.00499)	(0.00493)	(0.00477)	(0.00477)	(0.00499)	(0.00493)	(0.00498)
White	-0.155***	-0.155***	-0.155***	-0.155***	-0.153***	-0.153***	-0.155***	-0.155***	-0.155***
	(0.0105)	(0.0105)	(0.0109)	(0.0107)	(0.0104)	(0.0104)	(0.0109)	(0.0107)	(0.0109)
British	-0.0177	-0.0182	-0.0211	-0.0216	-0.0180	-0.0183	-0.0210	-0.0218	-0.0219
	(0.0172)	(0.0170)	(0.0149)	(0.0157)	(0.0165)	(0.0165)	(0.0149)	(0.0157)	(0.0149)
Low Income	0.00254	0.00190	0.00240	0.00153	0.00157	0.00158	0.00236	0.00163	0.00255
	(0.00307)	(0.00297)	(0.00336)	(0.00306)	(0.00285)	(0.00285)	(0.00336)	(0.00305)	(0.00335)
Local	0.0278***	0.0293***	0.0248***	0.0280***	0.0295***	0.0297***	0.0247***	0.0284***	0.0249***
	(0.00556)	(0.00555)	(0.00575)	(0.00564)	(0.00558)	(0.00565)	(0.00569)	(0.00557)	(0.00561)
Disabled	-0.0386***	-0.0380***	-0.0380***	-0.0385***	-0.0387***	-0.0387***	-0.0380***	-0.0385***	-0.0379***
	(0.00446)	(0.00432)	(0.00470)	(0.00440)	(0.00420)	(0.00420)	(0.00470)	(0.00439)	(0.00469)
Local Youth Unemployment (16-24yr)	0.000220						0.000363		0.00442***
	(0.000789)						(0.000979)		(0.00139)
Local Adult Unemployment (25-49yr)		-0.00160						-0.00229	-0.0135***
		(0.00200)						(0.00235)	(0.00307)
Growth in Local Youth Unemployment (16-24yr)			-0.0216***				-0.0233***		-0.0467***
			(0.00679)				(0.00797)		(0.0104)
Growth in Local Adult Unemployment (25-49yr)				-0.00421				-0.00112	0.0250**
				(0.00743)				(0.00854)	(0.0107)
Aggregate Local Unemployment (16-49yr)					0.000436				

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
					(0.00181)				
Growth in Local Aggregate Unemployment (16-49yr)						-0.00417**			
						(0.00205)			
Constant	0.448***	0.459***	0.462***	0.456***	0.445***	0.449***	0.456***	0.468***	0.463***
	(0.0259)	(0.0253)	(0.0244)	(0.0241)	(0.0242)	(0.0242)	(0.0270)	(0.0251)	(0.0269)
Observations	181,090	187,825	158,720	178,995	196,355	196,355	158,720	178,995	158,720
R-squared	0.063	0.063	0.064	0.063	0.061	0.061	0.064	0.063	0.064
Year FE	✓	✓	✓	✓	✓	✓	✓	✓	✓
University FE	✓	✓	✓	✓	✓	✓	✓	✓	✓
Region FE (NUTS 1 2010)	✓	✓	✓	✓	✓	✓	✓	✓	✓
Clustered SE	✓	✓	✓	✓	✓	✓	✓	✓	✓

Standard errors shown in the parentheses are clustered at the institutional level across all 9 specifications. All regressions are performed using OLS to estimate equation 9. The dependent variable is the binary subject group indicator, which takes the value 1 for LEM and 0 otherwise. All 9 specifications are estimating column (6) from Table 2.4 with various indicators of unemployment. Specifications (1)-(6) estimate the unemployment variables separately, where specifications (7)-(8) include the level of local unemployment and the growth in local unemployment respectively. Specification (9) includes all measures of youth and adult local unemployment. Raw numbers have been rounded to the nearest multiple of 5 to comply with HESA's Standard Rounding Methodology.
*** p<0.01, ** p<0.05, * p<0.1

Table B5: Full Regression Results from Table 2.7

	(1) OTHER	(2) OTHER	(3) OTHER	(4) OTHER	(5) OTHER	(6) OTHER	(7) OTHER	(8) OTHER	(9) OTHER
Tariff	-0.00227**	-0.00232**	-0.00204*	-0.00225**	-0.00245**	-0.00244**	-0.00204*	-0.00224**	-0.00200*
	(0.00103)	(0.00104)	(0.00104)	(0.00105)	(0.00103)	(0.00103)	(0.00104)	(0.00105)	(0.00104)
Female	0.181***	0.184***	0.181***	0.184***	0.185***	0.185***	0.181***	0.184***	0.181***
	(0.00669)	(0.00687)	(0.00652)	(0.00684)	(0.00688)	(0.00688)	(0.00652)	(0.00683)	(0.00651)
White	0.197***	0.197***	0.195***	0.197***	0.197***	0.196***	0.196***	0.198***	0.196***
	(0.0127)	(0.0126)	(0.0128)	(0.0126)	(0.0125)	(0.0125)	(0.0129)	(0.0126)	(0.0128)
British	0.0107	0.00847	0.0167	0.0108	0.00792	0.00803	0.0168	0.0113	0.0176
	(0.0309)	(0.0319)	(0.0282)	(0.0308)	(0.0316)	(0.0314)	(0.0283)	(0.0309)	(0.0283)
Low Income	-0.0207***	-0.0203***	-0.0213***	-0.0199***	-0.0208***	-0.0206***	-0.0214***	-0.0202***	-0.0215***
	(0.00314)	(0.00311)	(0.00332)	(0.00313)	(0.00294)	(0.00293)	(0.00332)	(0.00314)	(0.00333)
Local	-0.0428***	-0.0447***	-0.0406***	-0.0444***	-0.0467***	-0.0460***	-0.0408***	-0.0451***	-0.0411***
	(0.00518)	(0.00516)	(0.00510)	(0.00522)	(0.00519)	(0.00516)	(0.00511)	(0.00523)	(0.00510)
Disabled	0.0268***	0.0253***	0.0250***	0.0241***	0.0248***	0.0247***	0.0250***	0.0240***	0.0248***
	(0.00532)	(0.00535)	(0.00577)	(0.00543)	(0.00516)	(0.00516)	(0.00577)	(0.00542)	(0.00576)
Local Youth Unemployment (16-24yr)	0.000871						0.000448		-0.00337**
	(0.000650)						(0.000814)		(0.00132)
Local Adult Unemployment (25-49yr)		0.00390**						0.00486**	0.0129***
		(0.00183)						(0.00212)	(0.00337)
Growth in Local Youth Unemployment (16-24yr)			0.0309***				0.0289***		0.0492***
			(0.00830)				(0.00934)		(0.0115)
Growth in Local Adult Unemployment (25-49yr)				0.0130				0.00641	-0.0157
				(0.00906)				(0.00976)	(0.0122)
Aggregate Local Unemployment (16-49yr)					0.00262*				

					(0.00148)				
Growth in Local Aggregate Unemployment (16-49yr)						0.00956***			
						(0.00241)			
Constant	0.334***	0.329***	0.332***	0.341***	0.334***	0.349***	0.325***	0.315***	0.317***
	(0.0389)	(0.0406)	(0.0358)	(0.0368)	(0.0394)	(0.0371)	(0.0398)	(0.0412)	(0.0400)
Observations	181,090	187,825	158,720	178,995	196,355	196,355	158,720	178,995	158,720
R-squared	0.122	0.123	0.124	0.124	0.122	0.122	0.124	0.124	0.124
Year FE	✓	✓	✓	✓	✓	✓	✓	✓	✓
University FE	✓	✓	✓	✓	✓	✓	✓	✓	✓
Region FE (NUTS 1 2010)	✓	✓	✓	✓	✓	✓	✓	✓	✓
Clustered SE	✓	✓	✓	✓	✓	✓	✓	✓	✓

Standard errors shown in the parentheses are clustered at the institutional level across all 9 specifications. All regressions are performed using OLS to estimate equation 9. The dependent variable is the binary subject group indicator, which takes the value 1 for OTHER and 0 otherwise. All 9 specifications are estimating column (9) from Table 2.4 with various indicators of unemployment. Specifications (1)-(6) estimate the unemployment variables separately, where specifications (7)-(8) include the level of local unemployment and the growth in local unemployment respectively. Specification (9) includes all measures of youth and adult local unemployment. Raw numbers have been rounded to the nearest multiple of 5 to comply with HESA's Standard Rounding Methodology. *** p<0.01, ** p<0.05, * p<0.1

Table B6: Full Regression Results from Table 2.8

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	STEM	STEM	STEM	STEM	STEM	STEM	STEM	STEM	STEM
Tariff	-0.00216**	-0.00216**	-0.00215**	-0.00194**	-0.00194**	-0.00194**	-0.00195**	-0.00195**	-0.00195**
	(0.000963)	(0.000963)	(0.000963)	(0.000971)	(0.000971)	(0.000971)	(0.000972)	(0.000972)	(0.000972)
Female	-0.110***	-0.110***	-0.123***	-0.115***	-0.115***	-0.132***	-0.115***	-0.115***	-0.118***
	(0.00581)	(0.00581)	(0.0129)	(0.00582)	(0.00582)	(0.0127)	(0.00582)	(0.00582)	(0.00599)
White	-0.0420***	-0.0419***	-0.0419***	-0.0435***	-0.0434***	-0.0433***	-0.0428***	-0.0428***	-0.0429***
	(0.00857)	(0.00856)	(0.00856)	(0.00846)	(0.00845)	(0.00846)	(0.00854)	(0.00854)	(0.00854)
British	0.00687	0.00705	0.00706	0.00993	0.0101	0.0101	0.0101	0.0101	0.0102
	(0.0153)	(0.0153)	(0.0153)	(0.0167)	(0.0167)	(0.0167)	(0.0165)	(0.0165)	(0.0166)
Low Income	0.0363***	0.0169	0.0176	0.0376***	0.0172*	0.0181*	0.0374***	0.0374***	0.0372***
	(0.00562)	(0.0117)	(0.0117)	(0.00562)	(0.00999)	(0.00994)	(0.00562)	(0.00562)	(0.00561)
Local	0.0188***	0.0197***	0.0196***	0.0213***	0.0225***	0.0224***	0.0205***	0.0205***	0.0205***
	(0.00539)	(0.00542)	(0.00542)	(0.00531)	(0.00530)	(0.00530)	(0.00536)	(0.00537)	(0.00537)
Disabled	0.0118***	0.0119***	0.0118***	0.0140***	0.0140***	0.0140***	0.0140***	0.0140***	0.0139***
	(0.00427)	(0.00427)	(0.00427)	(0.00408)	(0.00408)	(0.00408)	(0.00409)	(0.00409)	(0.00409)
Low Income x Female	-0.0256***	-0.0257***	-0.0262***	-0.0258***	-0.0259***	-0.0266***	-0.0258***	-0.0258***	-0.0257***
	(0.00567)	(0.00567)	(0.00557)	(0.00554)	(0.00554)	(0.00545)	(0.00553)	(0.00553)	(0.00553)
Local Youth Unemployment (16-24yr)	-0.00108*	-0.00153**	-0.00212**						
	(0.000632)	(0.000596)	(0.000845)						
Low Income x Local Youth Unemployment (16-24yr)		0.00152*	0.00149*						
		(0.000871)	(0.000872)						
Female x Local Youth Unemployment (16-24yr)			0.00102						
			(0.000849)						
Aggregate Local Unemployment (16-49yr)				-0.00302**	-0.00424***	-0.00624***			
				(0.00142)	(0.00137)	(0.00192)			

Low Income x Aggregate Local Unemployment (16-49yr)					0.00414**	0.00402**			
					(0.00184)	(0.00184)			
Female x Aggregate Local Unemployment (16-49yr)						0.00341*			
						(0.00202)			
Growth in Local Aggregate Unemployment (16-49yr)							-0.00534**	-0.00541**	-0.0134***
							(0.00216)	(0.00223)	(0.00330)
Low Income x Growth in Aggregate Local Unemployment (16-49yr)								0.000235	-0.0000732
								(0.00368)	(0.00366)
Female x Growth in Aggregate Local Unemployment (16-49yr)									0.0137***
									(0.00395)
Constant	0.213***	0.218***	0.226***	0.216***	0.222***	0.232***	0.197***	0.197***	0.199***
	(0.0265)	(0.0271)	(0.0277)	(0.0284)	(0.0291)	(0.0294)	(0.0276)	(0.0276)	(0.0276)
Observations	181,090	181,090	181,090	196,355	196,355	196,355	196,355	196,355	196,355
R-squared	0.104	0.104	0.104	0.106	0.106	0.106	0.106	0.106	0.107
Year FE	✓	✓	✓	✓	✓	✓	✓	✓	✓
University FE	✓	✓	✓	✓	✓	✓	✓	✓	✓
Region FE (NUTS 1 2010)	✓	✓	✓	✓	✓	✓	✓	✓	✓
Clustered SE	✓	✓	✓	✓	✓	✓	✓	✓	✓

Standard errors shown in the parentheses are clustered at the institutional level across all 9 specifications. All regressions are performed using OLS to estimate equation 9. The dependent variable is the binary subject group indicator, which takes the value 1 for STEM and 0 otherwise. All 9 specifications are estimating column (3) from Table 2.4 with various indicators of unemployment, and various interactions of the unemployment measures with the socioeconomic controls. Specifications (1)-(3) estimate the effect of local youth unemployment on the probability of a STEM subject being studied, whilst specifications (4)-(6) and (7)-(9) estimate the effect of aggregate and the growth of aggregate local unemployment respectively. Interactions are included between the unemployment variable and the socioeconomic indicators of low income and female, to allow the unemployment effect to be heterogeneous across these binary categories. Raw numbers have been rounded to the nearest multiple of 5 to comply with HESA's Standard Rounding Methodology. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table B7: Full Regression Results from Table 2.9

	(1) LEM	(2) LEM	(3) LEM	(4) LEM	(5) LEM	(6) LEM	(7) LEM	(8) LEM	(9) LEM
Tariff	0.00444***	0.00443***	0.00443***	0.00439***	0.00439***	0.00439***	0.00439***	0.00439***	0.00439***
	(0.00101)	(0.00101)	(0.00101)	(0.000997)	(0.000997)	(0.000997)	(0.000998)	(0.000998)	(0.000998)
Female	-0.0653***	-0.0653***	-0.0580***	-0.0637***	-0.0637***	-0.0571***	-0.0637***	-0.0637***	-0.0615***
	(0.00512)	(0.00512)	(0.0105)	(0.00501)	(0.00501)	(0.00941)	(0.00500)	(0.00500)	(0.00516)
White	-0.155***	-0.155***	-0.155***	-0.153***	-0.153***	-0.153***	-0.153***	-0.153***	-0.153***
	(0.0105)	(0.0105)	(0.0105)	(0.0104)	(0.0104)	(0.0104)	(0.0104)	(0.0104)	(0.0104)
British	-0.0177	-0.0178	-0.0178	-0.0180	-0.0181	-0.0181	-0.0182	-0.0182	-0.0183
	(0.0172)	(0.0172)	(0.0172)	(0.0165)	(0.0165)	(0.0165)	(0.0165)	(0.0165)	(0.0165)
Low Income	-0.000734	0.0131	0.0127	-0.00197	0.00998	0.00960	-0.00199	-0.00193	-0.00184
	(0.00529)	(0.0107)	(0.0107)	(0.00497)	(0.00922)	(0.00927)	(0.00497)	(0.00502)	(0.00500)
Local	0.0256***	0.0250***	0.0250***	0.0268***	0.0261***	0.0261***	0.0269***	0.0269***	0.0269***
	(0.00599)	(0.00604)	(0.00604)	(0.00612)	(0.00618)	(0.00618)	(0.00618)	(0.00617)	(0.00617)
Disabled	-0.0386***	-0.0386***	-0.0386***	-0.0387***	-0.0387***	-0.0387***	-0.0387***	-0.0387***	-0.0386***
	(0.00446)	(0.00446)	(0.00446)	(0.00419)	(0.00420)	(0.00419)	(0.00420)	(0.00420)	(0.00420)
Low Income x Female	0.00293	0.00298	0.00326	0.00285	0.00290	0.00316	0.00289	0.00289	0.00274
	(0.00539)	(0.00538)	(0.00532)	(0.00533)	(0.00533)	(0.00528)	(0.00533)	(0.00533)	(0.00533)
Local Youth Unemployment (16-24yr)	0.000216	0.000534	0.000868						
	(0.000788)	(0.000795)	(0.000829)						
Low Income x Local Youth Unemployment (16-24yr)		-0.00109	-0.00107						
		(0.000746)	(0.000748)						
Female x Local Youth Unemployment (16-24yr)			-0.000570						
			(0.000691)						
Aggregate Local Unemployment (16-49yr)				0.000420	0.00114	0.00191			
				(0.00181)	(0.00190)	(0.00196)			

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Low Income x Aggregate Local Unemployment (16-49yr)					-0.00242	-0.00238			
					(0.00170)	(0.00171)			
Female x Aggregate Local Unemployment (16-49yr)						-0.00131			
						(0.00158)			
Growth in Local Aggregate Unemployment (16-49yr)							-0.00418**	-0.00410*	0.00237
							(0.00205)	(0.00207)	(0.00326)
Low Income x Growth in Aggregate Local Unemployment (16-49yr)								-0.000295	-4.74e-05
								(0.00356)	(0.00354)
Female x Growth in Aggregate Local Unemployment (16-49yr)									-0.0110***
									(0.00408)
Constant	0.449***	0.445***	0.441***	0.446***	0.443***	0.439***	0.450***	0.450***	0.449***
	(0.0258)	(0.0257)	(0.0271)	(0.0241)	(0.0243)	(0.0252)	(0.0241)	(0.0241)	(0.0241)
Observations	181,090	181,090	181,090	196,355	196,355	196,355	196,355	196,355	196,355
R-squared	0.063	0.063	0.063	0.061	0.061	0.061	0.061	0.061	0.061
Year FE	✓	✓	✓	✓	✓	✓	✓	✓	✓
University FE	✓	✓	✓	✓	✓	✓	✓	✓	✓
Region FE (NUTS 1 2010)	✓	✓	✓	✓	✓	✓	✓	✓	✓
Clustered SE	✓	✓	✓	✓	✓	✓	✓	✓	✓

Standard errors shown in the parentheses are clustered at the institutional level across all 9 specifications. All regressions are performed using OLS to estimate equation 9. The dependent variable is the binary subject group indicator, which takes the value 1 for LEM and 0 otherwise. All 9 specifications are estimating column (6) from Table 2.4 with various indicators of unemployment, and various interactions of the unemployment measures with the socioeconomic controls. Specifications (1)-(3) estimate the effect of local youth unemployment on the probability of a LEM subject being studied, whilst specifications (4)-(6) and (7)-(9) estimate the effect of aggregate and the growth of aggregate local unemployment respectively. Interactions are included between the unemployment variable and the socioeconomic indicators of low income and female, to allow the unemployment effect to be heterogeneous across these binary categories. Raw numbers have been rounded to the nearest multiple of 5 to comply with HESA's Standard Rounding Methodology. *** p<0.01, ** p<0.05, * p<0.1

Table B8: Full Regression Results from Table 2.10

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	OTHER	OTHER	OTHER	OTHER	OTHER	OTHER	OTHER	OTHER	OTHER
Tariff	-0.00228**	-0.00228**	-0.00228**	-0.00245**	-0.00245**	-0.00245**	-0.00244**	-0.00244**	-0.00244**
	(0.00103)	(0.00103)	(0.00103)	(0.00103)	(0.00103)	(0.00103)	(0.00103)	(0.00103)	(0.00103)
Female	0.175***	0.175***	0.181***	0.179***	0.179***	0.189***	0.179***	0.179***	0.179***
	(0.00686)	(0.00686)	(0.0143)	(0.00712)	(0.00712)	(0.0142)	(0.00712)	(0.00711)	(0.00725)
White	0.197***	0.197***	0.197***	0.196***	0.196***	0.196***	0.196***	0.196***	0.196***
	(0.0127)	(0.0127)	(0.0127)	(0.0125)	(0.0125)	(0.0125)	(0.0125)	(0.0125)	(0.0125)
British	0.0108	0.0108	0.0108	0.00803	0.00795	0.00796	0.00814	0.00814	0.00812
	(0.0309)	(0.0310)	(0.0310)	(0.0316)	(0.0316)	(0.0316)	(0.0314)	(0.0314)	(0.0314)
Low Income	-0.0355***	-0.0300**	-0.0303**	-0.0356***	-0.0271**	-0.0277**	-0.0354***	-0.0354***	-0.0354***
	(0.00525)	(0.0128)	(0.0128)	(0.00501)	(0.0111)	(0.0111)	(0.00501)	(0.00506)	(0.00506)
Local	-0.0444***	-0.0446***	-0.0446***	-0.0480***	-0.0485***	-0.0485***	-0.0474***	-0.0474***	-0.0474***
	(0.00570)	(0.00572)	(0.00572)	(0.00572)	(0.00576)	(0.00576)	(0.00569)	(0.00569)	(0.00569)
Disabled	0.0267***	0.0267***	0.0267***	0.0247***	0.0247***	0.0247***	0.0247***	0.0247***	0.0247***
	(0.00533)	(0.00532)	(0.00533)	(0.00517)	(0.00516)	(0.00517)	(0.00517)	(0.00517)	(0.00517)
Low Income x Female	0.0227***	0.0227***	0.0229***	0.0230***	0.0230***	0.0234***	0.0230***	0.0230***	0.0229***
	(0.00595)	(0.00595)	(0.00591)	(0.00575)	(0.00575)	(0.00570)	(0.00575)	(0.00575)	(0.00575)
Local Youth Unemployment (16-24yr)	0.000863	0.000991	0.00125						
	(0.000651)	(0.000763)	(0.000975)						
Low Income x Local Youth Unemployment (16-24yr)		-0.000437	-0.000423						
		(0.000892)	(0.000892)						
Female x Local Youth Unemployment (16-24yr)			-0.000451						
			(0.000914)						
Aggregate Local Unemployment (16-49yr)				0.00260*	0.00310*	0.00433*			
				(0.00148)	(0.00175)	(0.00225)			

Low Income x Aggregate Local Unemployment (16-49yr)					-0.00172	-0.00165			
					(0.00201)	(0.00201)			
Female x Aggregate Local Unemployment (16-49yr)						-0.00209			
						(0.00209)			
Growth in Local Aggregate Unemployment (16-49yr)							0.00952***	0.00951***	0.0111***
							(0.00241)	(0.00253)	(0.00367)
Low Income x Growth in Aggregate Local Unemployment (16-49yr)								0.0000604	0.000121
								(0.00460)	(0.00460)
Female x Growth in Aggregate Local Unemployment (16-49yr)									-0.00268
									(0.00444)
Constant	0.338***	0.336***	0.333***	0.338***	0.335***	0.329***	0.353***	0.353***	0.353***
	(0.0389)	(0.0389)	(0.0402)	(0.0394)	(0.0397)	(0.0406)	(0.0370)	(0.0370)	(0.0370)
Observations	181,090	181,090	181,090	196,355	196,355	196,355	196,355	196,355	196,355
R-squared	0.122	0.122	0.122	0.122	0.122	0.122	0.122	0.122	0.122
Year FE	✓	✓	✓	✓	✓	✓	✓	✓	✓
University FE	✓	✓	✓	✓	✓	✓	✓	✓	✓
Region FE (NUTS 1 2010)	✓	✓	✓	✓	✓	✓	✓	✓	✓
Clustered SE	✓	✓	✓	✓	✓	✓	✓	✓	✓

Standard errors shown in the parentheses are clustered at the institutional level across all 9 specifications. All regressions are performed using OLS to estimate equation 9. The dependent variable is the binary subject group indicator, which takes the value 1 for OTHER and 0 otherwise. All 9 specifications are estimating column (9) from Table 2.4 with various indicators of unemployment, and various interactions of the unemployment measures with the socioeconomic controls. Specifications (1)-(3) estimate the effect of local youth unemployment on the probability of an OTHER subject being studied, whilst specifications (4)-(6) and (7)-(9) estimate the effect of aggregate and the growth of aggregate local unemployment respectively. Interactions are included between the unemployment variable and the socioeconomic indicators of low income and female, to allow the unemployment effect to be heterogeneous across these binary categories. Raw numbers have been rounded to the nearest multiple of 5 to comply with HESA's Standard Rounding Methodology. *** p<0.01, ** p<0.05, * p<0.1

Figure B1: Histogram of Youth Unemployment

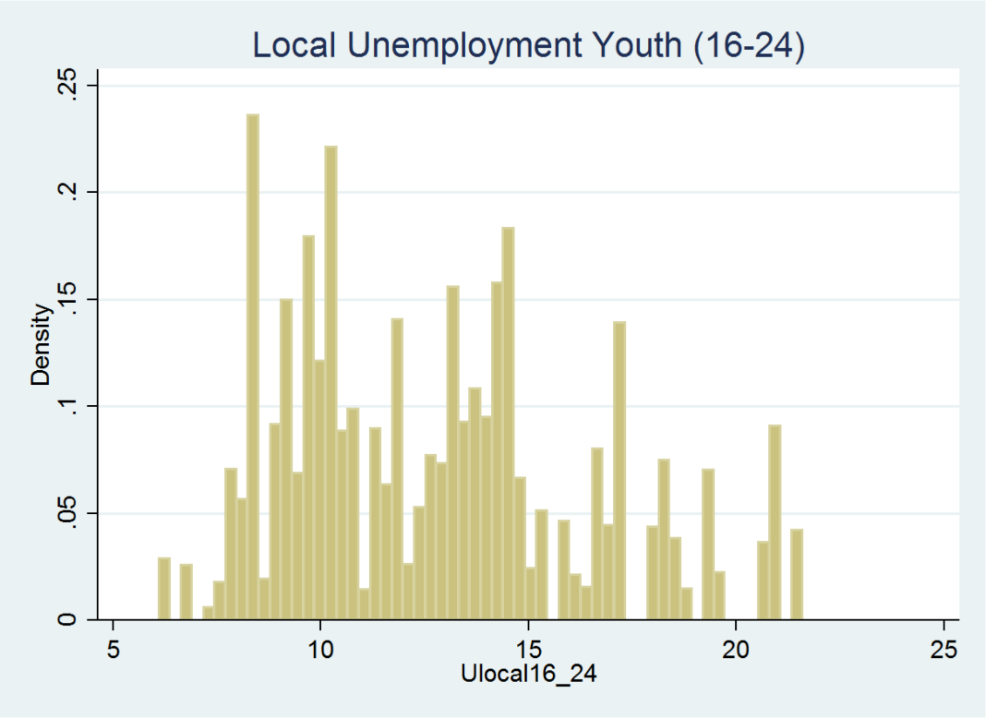


Figure B2: Histogram of Youth Unemployment Growth

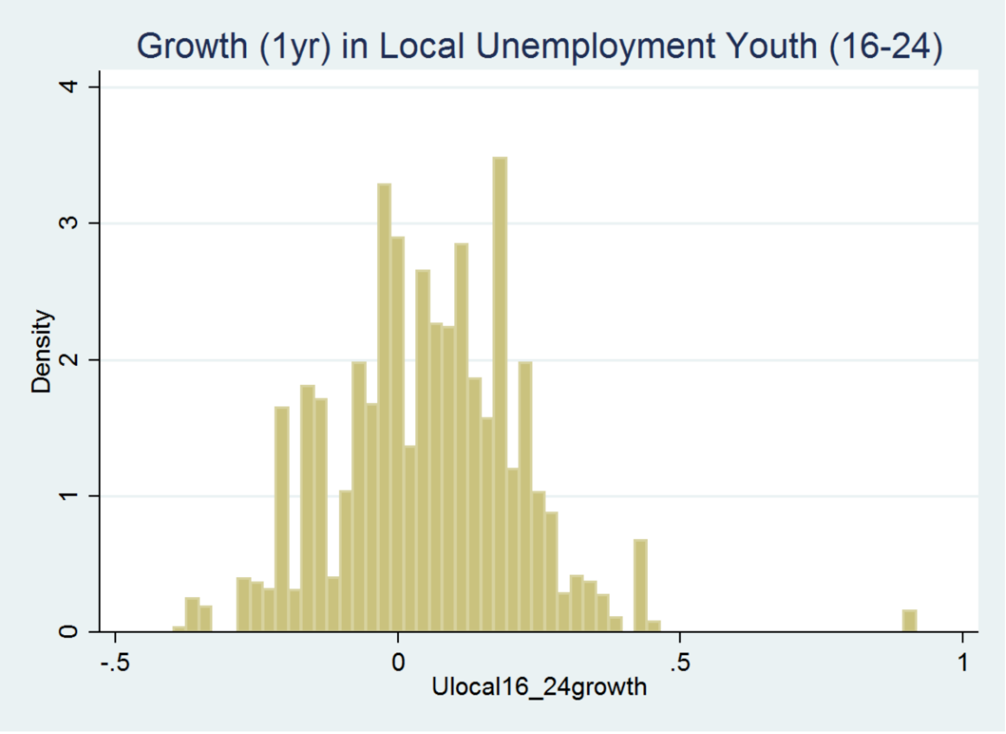


Figure B3: Histogram of Adult Unemployment

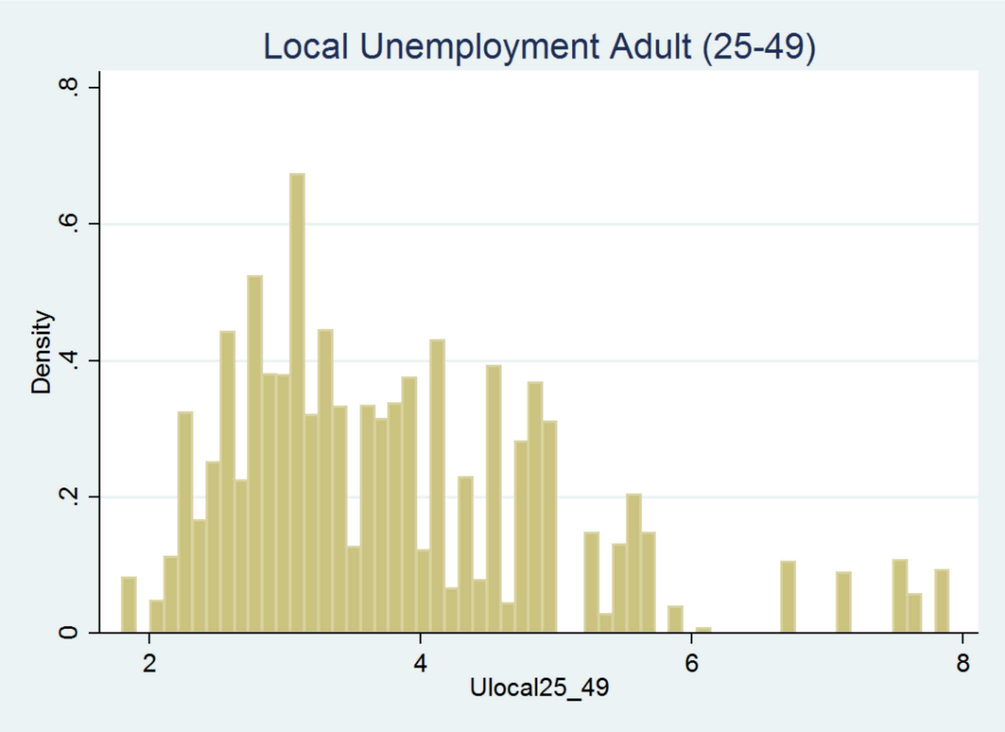


Figure B4: Histogram of Adult Unemployment Growth

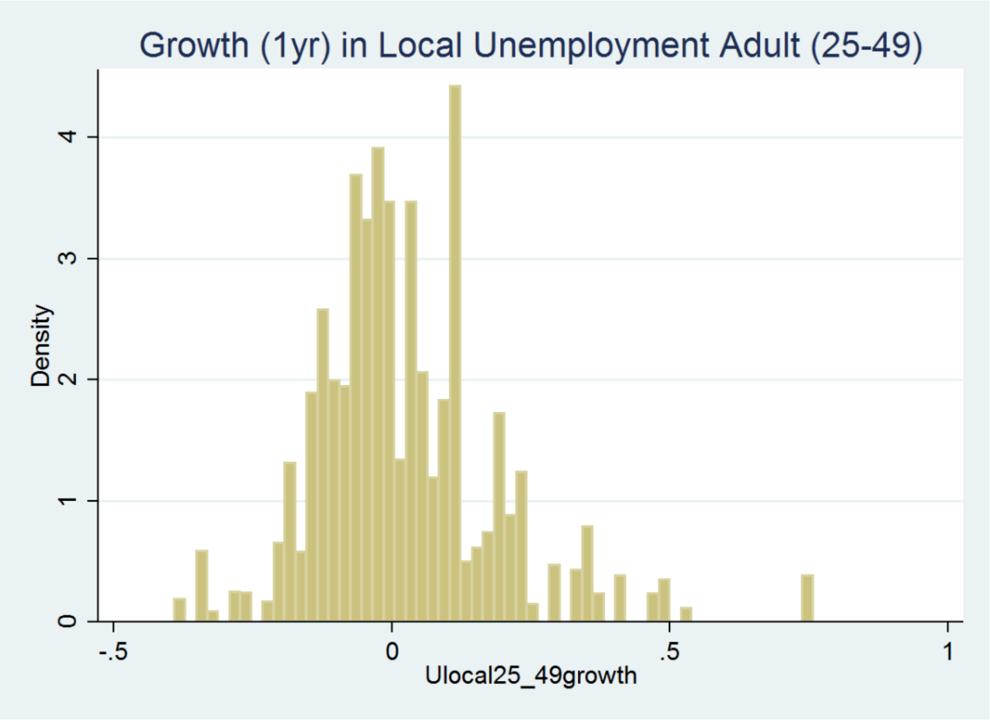


Figure B5: Youth Unemployment by NUTS 2 (2010) and Start Year

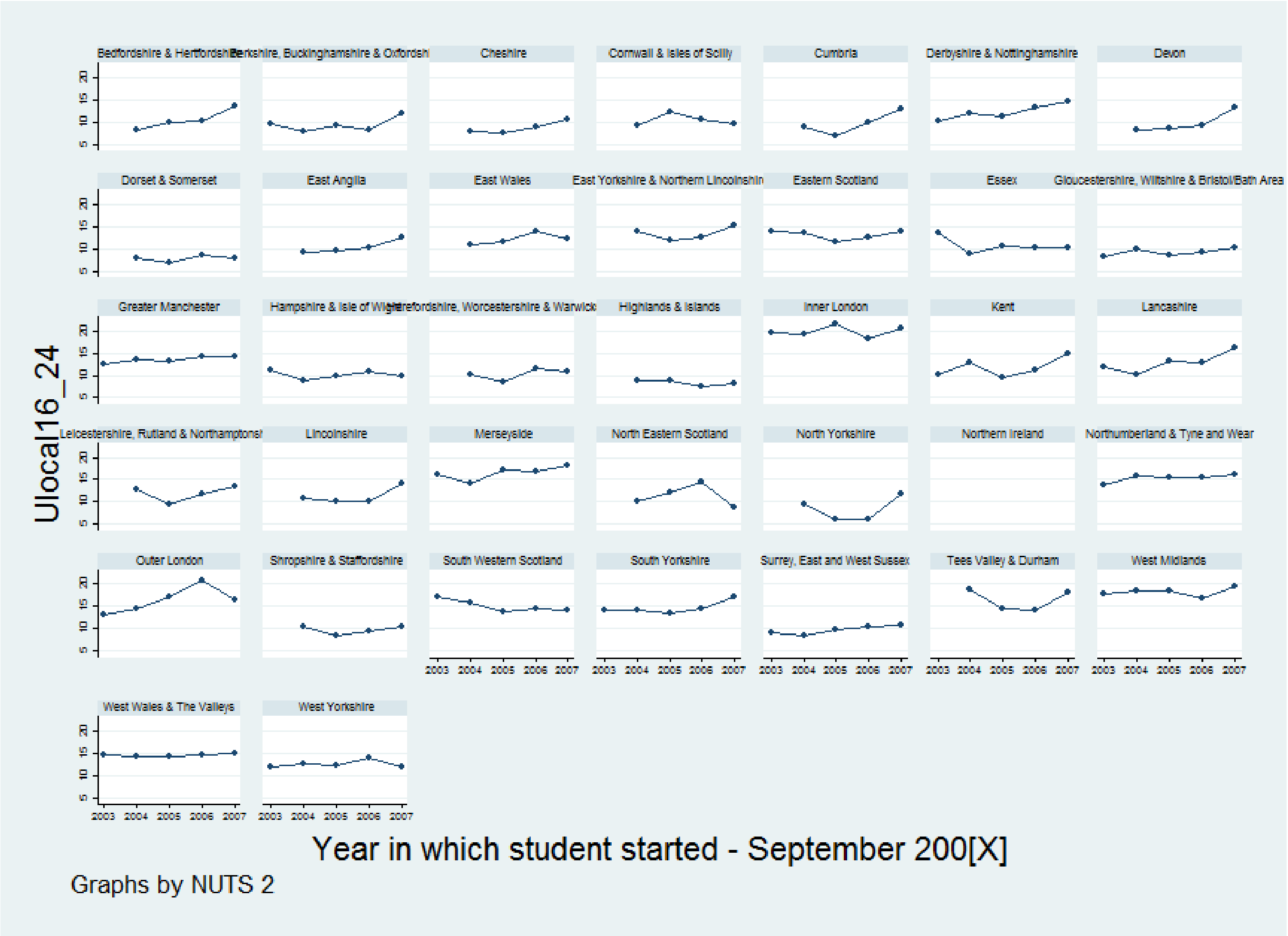
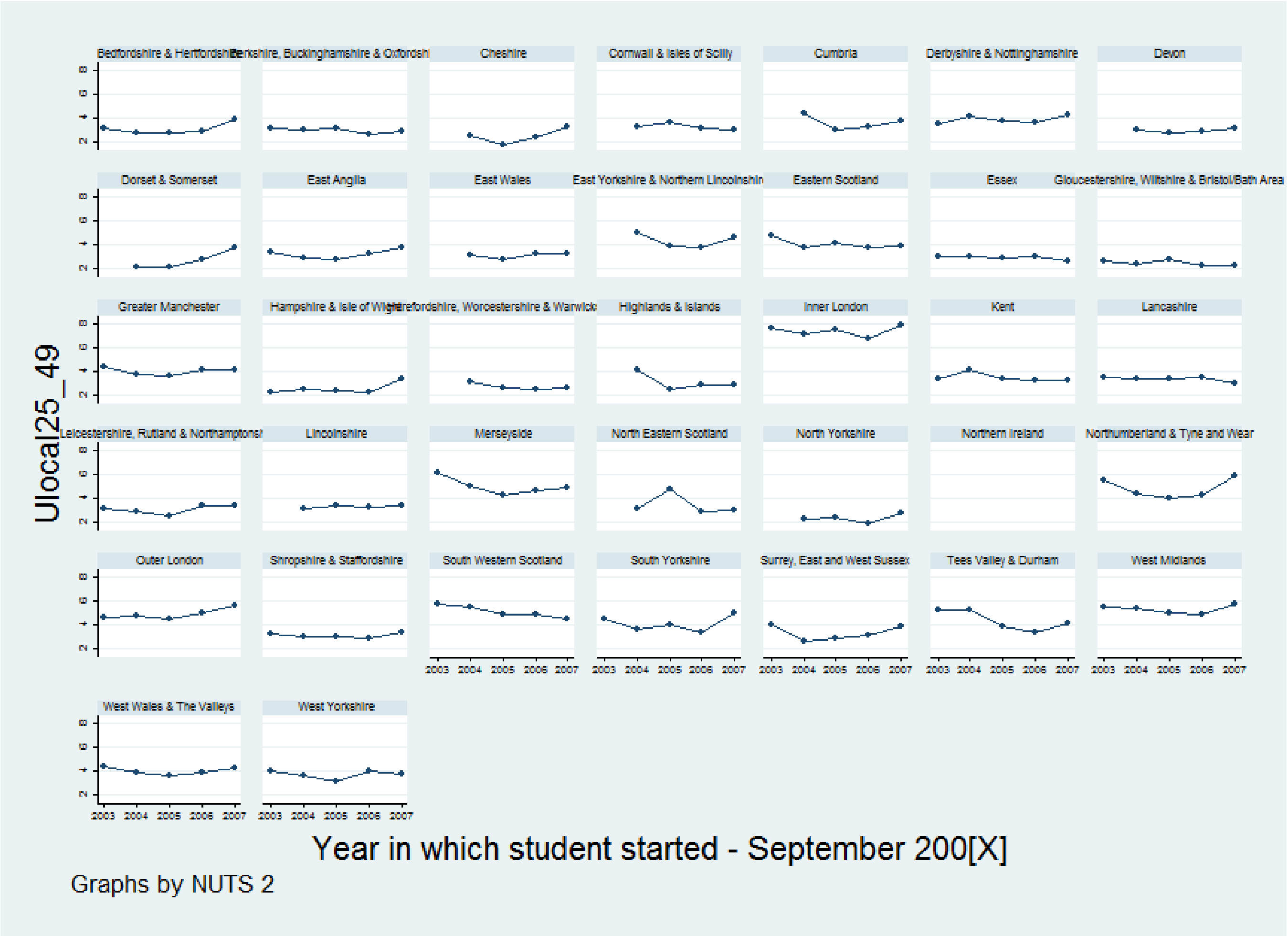


Figure B6: Adult Unemployment by NUTS 2 (2010) and Start Year



Appendix C: Appendix to Chapter 3

Table C1: UCAS Tariff Scores, by Qualification & Grade

GCE/VCE Qualifications			BTEC Nationals[1]			OCR Nationals[2]			CACHE Diploma[3]		Diploma in Foundation Studies (Art and Design)[4]	Points	Irish Leaving Cert[5]		Scottish Qualifications			
GCE AS/ AS VCE	GCE A level/ AVCE	AVCE Double Award	Award	Certificate	Diploma	Certificate	Diploma	Extended Diploma	Theory	Practical			Higher	Ordinary	Advanced Higher	Higher	Int 2	Standard Grade
					DDD			**D1**				**360**						
					DDM			**D2/M1**				**320**						
											Distinction	**285**						
					DMM			**M2**				**280**						
		AA		**DD**	**MMM**		**D**	**M3**	**A**			**240**						
											Merit	**225**						
		AB										**220**						
		BB		**DM**	**MMP**		**M1**	**P1**	**B**			**200**						
		BC										**180**						
											Pass	**165**						
		CC		**MM**	**MPP**		**M2/P1**	**P2**	**C**			**160**						
		CD										**140**						
	A	**DD**	**D**	**MP**	**PPP**	**D**	**P2**	**P3**	**D**	**A**		**120**			**A**			

	B	**DE**								**B**		**100**			**B**			
												90	**A1**					
	C	**EE**	**M**	**PP**		**M**	**P3**		**E**	**C**		**80**			**C**			
												77	**A2**					
												72			**D**	**A**		
												71	**B1**					
												64	**B2**					
A	**D**									**D**		**60**				**B**		
												58	**B3**					
												52	**C1**					
B												**50**						
												48				**C**		
												45	**C2**					
												42				**D**	**A**	
C	**E**		**P**			**P**				**E**		**40**						
												39	**C3**	**A1**				
												38						**Band 1**
												35					**B**	
												33	**D1**					
D												**30**						
												28					**C**	**Band 2**
												26	**D2**	**A2**				

E												**20**	**D3**	**B1**				
												17						
												14		**B2**				
												13						
												10						
												7		**B3**				

Welsh Bacc Core[6]	**Advanced Extension Awards**[7]	**Core Skills**[8]	**Key Skills**[9]	**Free standing Maths**[10]	**IFS CeFS**[11]	**Points**	**Music Examinations**[12]					
							Practical			**Theory**		
							Grade 6	Grade 7	Grade 8	Grade 6	Grade 7	Grade 8
Pass						**120**						
						75			**D**			
						70			**M**			
					A	**60**		**D**				
						55		**M**	**P**			
					B	**50**						
						45	**D**					
	Distinction				**C**	**40**	**M**	**P**				
			Level 4		**D**	**30**						**D**
						25	**P**					**M**
	Merit	**Higher**	**Level 3**	**A**	**E**	**20**					**D**	**P**
				B		**17**						
						15				**D**	**M**	
				C		**13**						
		Int 2	**Level 2**	**D**		**10**				**M**	**P**	
				E		**7**						
						5				**P**		

Source: UCAS

[1] The points shown are for the newly specified BTEC National Award, Certificate and Diploma
[2] The points for the OCR Nationals come into effect for entry to higher education in **2007** onwards
[3] Covers the CACHE Diploma in Child Care and Education
[4] The points for the Diploma in Foundation Studies (Art and Design) come into effect for entry to higher education in **2006** onwards
[5] The points shown for the Irish Leaving Certificate Higher and Ordinary levels, come into effect for entry to higher education in **2006** onwards
[6] Points for the Core of the Advanced Welsh Baccalaureate Qualification
[7] Points for Advanced Extension Awards are over and above those gained from the A level grade and come into effect for entry to higher education in **2006**
[8] Covers the five Scottish Core Skills – Communication, Information Technology, Numeracy, Problem Solving & Working with Others
[9] Covers the three main three Key Skill subjects – Application of Number, Communication and Information Technology with the three Wider Key Skills coming into the Tariff for **2007** entry
[10] Covers free-standing Mathematics qualifications – Additional Maths, Using and Applying Statistics, Working with Algebraic and Graphical Techniques, Modelling with Calculus
[11] The points shown are for the revised Institute of Financial Services Certificate in Financial Studies (CeFS) taught from September '03
[12] The points shown are for ABRSM, Guildhall, LCMM, Rockschool and Trinity music examinations, grades 6, 7 and 8.

Bibliography

Adams, R. (2017). *Oxford and Cambridge 'need to improve access for disadvantaged students'*. Available at: https://www.theguardian.com/education/2017/jun/29/oxford-cambridge-improve-access-disadvantaged-students

Ai, C., & Norton, E. C. (2003). Interaction terms in logit and probit models. *Economic Letters*, 80, 123-129

Alecke, B., Burgard, C., & Mitze, T. (2013). The effect of tuition fees on student enrollment and location choice – interregional migration, border effects and gender differences. *Ruhr Economic Paper*, No. 404

Anderberg, D. & Cerrone, C. (2016). Investment in education under disappointment aversion. *Max Planck Instutute Collective Goods Preprint*, No. 2016/16

Anders, J. (2012). The link between household income, university applications and university attendance. Fiscal Studies, 33, 185-210

Ardila, A. (2001). Predictors of university academic performance in Colombia. *International Journal of Educational Research*, 35, 411-417

Arkes, J. (2010). Using unemployment rates as instruments to estimate returns to schooling. *Southern Economic Journal*, 76(3), 711-722

Artess, J., McCulloch, A., & Mok, P. (2014). Learning from Futuretrack: studying and living at home. *BIS Research Paper*, No. 167, Retrieved from: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/287699/bis-14-642-learning-from-futuretrack-studying-and-living-at-home-bis-research-paper-167.pdf

Atherton, G., Jones, S., & Hall, A. (2015). Does cost matter? Students' understanding of the higher education finance system and how cost affects their decisions. *National Education Opportunities Network Report*

Barrow, M., Reilly, B. & Woodfield, R. (2009). The determinants of undergraduate degree performance: How important is gender? *British Educational Research Journal*, 35(4), 575-597

Bartlett, S., Peel, M. J., & Pendlebury, M. (1993). From fresher to finalist: a three year analysis of student performance on an accounting degree programme. *Accounting Education*, 2(2), 111-122

Battu, H., Belfield, C. R., & Sloane, P. J. (1999). Over-education among graduates: a cohort view. *Education Economics*, 7, 21-38

Becker, G. S. (1964). Human capital: a theoretical and empirical analysis, with special reference to education (3rd ed.). Chicago: The University of Chicago Press.

Bell, D. N. F. & Blanchflower, D. G. (2011). Young people and the great recession. *Oxford Review of Economic Policy*, 27(2), 241-267

Bennett, R., Glennerster, H., & Nevison, D. (1992). Investing in skill: to stay on or not to stay on? *Oxford Review of Economic Policy*, 8(2), 130-145

Betts, J. R. & McFarland, L. L. (1995). Safe port in a storm: the impact of labor market conditions on community college enrollments. *The Journal of Human Resources*, 30(4), 741-765

BIS (2010a). The impact of higher education finance on university participation in the UK. *BIS Research Paper*, No. 11

BIS (2010b). The impact of the 2006-07 HE finance reforms on HE participation. *BIS Research Paper*, No. 13

BIS (2011). The returns to higher education qualifications. *BIS Research Paper*, No. 45

Black, S. E. & Sufi, A. (2002). Who goes to college? Differential enrollment by race and family background. *National Bureau of Economic Research Working Paper*, No. 9310

Blom, E., Cadena, B. C., & Keys, B. J. (2015). Investment over the business cycle: insights from college major choice. *IZA Discussion Paper*, No. 9167

Blundell, R., Dearden, L., Goodman, A., & Reed, H. (2000). The returns to higher education in Britain: evidence from a British cohort. *The Economic Journal*, 110, 82-99

Boffy-Ramirez, E. (2016). The heterogeneous impacts of business cycles on educational attainment. *IZA Discussion Paper*, No. 10167

Bolton, P. (2014). Value of student maintenance support. *House of Commons Library,* SN/SG/916

Boyle, R., Carter, J., & Clark, M. (2002). What makes them succeed? Entry, progression and graduation in Computer Science. *Journal of Further and Higher Education*, 26(1), 3-18

Bradley, S., & Migali, G. (2015). The effect of a tuition fee reform on the risk of drop out from university in the UK. *Lancaster University Management School Economics Working Paper Series*, No. 2015/016

Britton, J., Dearden, L., Shephard, N. & Vignoles, A. (2016). How English domiciled graduate earnings vary with gender, institution attended, subject and economic background. *Institute for Fiscal Studies Working Paper*, No. W16/06

Broecke, S. & Nicholls, T. (2007). Ethnicity and degree attainment. *Department of Education and Skills Research Report*, No. RW92

Browne Review of Higher Education. (2010). Securing a sustainable future for higher education: an independent review of higher education funding and student finance. Retrieved from: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/422565/bis-10-1208-securing-sustainable-higher-education-browne-report.pdf

Bruckmeier, K., & Wigger, B. U. (2012). The effects of tuition fees on transition from high school to university in Germany. *Economics of Education Review*, 41(1), 14-23

Callender, C. (2003). Student financial support in higher education: access and exclusion, in Tight, M. (eds.) *Access and Exclusion: International Perspectives on Higher Education Research*, Elsevier Science, London

Callender, C. & Jackson, J. (2008). Does fear of debt constrain choice of university and subject of study? *Studies in Higher Education*, 33(4), 405-429

Canton, E., & De Jong, F. (2005). The demand for higher education in the Netherlands, 1950-1999. *Economics of Education Review*, 24, 651-663

Card, D. & Lemieux, T. (2001). Dropout and enrolment trends in the post-war period: what went wrong in the 1970s? In J. Gruber (ed.), *Risky Behavior among Youths: An Economic Analysis*. Chicago: University of Chicago Press.

Cebula, R. & Lopes, J. (1982). Determinants of student choice of undergraduate major field. *American Educational Research Journal*, 19(2), 303-312

Chevalier, A. (2003). Parental education and child's education: a natural experiment. *IZA Discussion Paper*, No. 1153

Chevalier, A. (2007). Education, occupation and career expectations: determinants of the gender pay gap for UK graduates. *Oxford Bulletin of Economics and Statistics*, 69(6), 819-842

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

Chevalier, A. (2014). Does Higher Education quality matter in the UK? *IZA Discussion Paper*, No. 8363

Chevalier, A., Harmon, C., Walker, I., & Zhu, Y. (2004). Does education raise productivity, or just reflect it? *The Economic Journal*, 114, 499-517

Chowdry, H., Crawford, C., Dearden, L., Goodman, A., & Vignoles, A. (2013). Widening participation in higher education: analysis using linked administrative data. *Journal of the Royal Statistical Society: Series A*, 176, 431-457

Clark, D. (2011). Do recessions keep students in school? The impact of youth unemployment on enrolment in post-compulsory education in England. *Economica*, 78, 523-545

Codiroli Mcmaster, N. (2017). Who studies STEM subjects at A level and degree in England? An investigation into the intersections between students' family background, gender and ethnicity in determining choice. *British Educational Research Journal*, doi:10.1002/berj.3270

Collinson, P. (2016). More than a third of UK graduates regret going to university. *The Guardian*, 10 August. Available at: https://www.theguardian.com/education/2016/aug/10/more-than-a-third-of-uk-graduates-regret-attending-university (Accessed on 26 September 2017)

Connor, C., Tyers, T., Modood, J., & Hillage, J. (2004). Why the difference? A closer look at higher education minority ethnic students and graduates. *Department of Education and Skills Research Report,* No. RR552

Crawford, C. (2014a). Socio-economic differences in university outcomes in the UK: drop-out, degree completion and degree class. *Institute for Fiscal Studies Working Paper*, No. W14/31

Crawford, C. (2014b). The link between secondary school characteristics and university participation and outcomes. *The Centre for Analysis of Youth Transititions (CAYT) Research Report*. Retrieved from: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/317276/RR353_-_The_link_between_secondary_school_characteristics_and_university_participation_and_outcomes_FINAL.pdf

Davies, P., Mangan, J., Hughes, A., & Slack, K. (2013). Labour market motivation and undergraduates' choice of degree subject. *British Educational Research Journal*, 39(2), 361-382

Davies, P., Telhaj, S., Hutton, D., Adnett, N., & Coe, R. (2008). Socioeconomic background, gender and subject choice in secondary schooling. *Educational Research*, 50(3), 235-248

De Vries, R. & Rentfrow, J. (2016). A winning personality: the effects of background on personality and earnings. *The Sutton Trust*. Retrieved from: http://www.suttontrust.com/wp-content/uploads/2016/01/Winning-Personality-FINAL.pdf

Dearden, L., Ferri, J., & Meghir, C. (2002). The effect of school quality on educational attainment and wages. *The Review of Economics and Statistics*, 84(1), 1-20

Dearden, L., Fitzsimons, E., Goodman, A., & Kaplan, G. (2008). Higher education funding reforms in England: the distributional effects and the shifting balance of costs. *The Economic Journal*, 118(526), F100-F125

Dearden, L., Fitzsimons, E., & Wyness, G. (2011). The impact of tuition fees and support on university participation in the UK. *Centre for the Economics of Education Discussion Paper*, No. 126

Delaney, L., Harmon, C., & Redmond, C. (2011). Parental education, grade attainment and earnings expectations among university students. *Economics of Education Review*, 30, 1136-1152

Dellas, H. & Koubi, V. (2003). Business cycles and schooling. *European Journal of Political Economy*, 19, 843-859

Dellas, H. & Sakellaris, P. (2003). On the cyclicality of schooling: theory and evidence. *Oxford Economic Papers*, 55, 148-172

Denzler, S. & Wolter, S. C. (2011). Too far to go? Does distance determine study choices? *IZA Discussion Paper*, No. 5712

Devereux, P. J. & Hart, R. A. (2010). Forced to be rich? Returns to compulsory schooling in Britain. *The Economic Journal*, 120, 1345-1364

DfES (2003). The future of higher education. Available at: http://webarchive.nationalarchives.gov.uk/20080205154241/http://www.dcsf.gov.uk/hegateway/strategy/hestrategy/exec.shtml (Accessed on 26 September 2017)

Di Pietro, G. (2004). The determinants of university dropout in Italy: a bivariate probability model with sample selection. *Applied Economics Letters*, 11(3), 187-191

Di Pietro, G. (2006). Regional labour market conditions and university dropout rates: evidence from Italy. *Regional Studies*, 40(6), 617-630

Di Pietro, G. (2010). The impact of degree class on the first destinations of graduates: a regression discontinuity approach. *IZA Discussion Paper*, No. 4836

Dickson, M. & Smith, S. (2011). What determines the return to education: an extra year or a hurdle cleared? *Economics of Education Review*, 30, 1167-1176

Dietrich, H., & Gerner, H. (2012). The effects of tuition fees on the decision for higher education: evidence from a German policy experiment. *Economics Bulletin*, 32(3), 2407-2413

Dunnett, A., Moorhouse, J., Walsh, C., & Barry, C. (2012) Choosing a university: a conjoint analysis of the impact of higher fees on students applying for university in 2012. *Tertiary Education and Management*, 18(3), 199-220

Duru, M. & Mingat, A. (1979). Comportement des bacheliers: modèle des choix de disciplines. *Consommation*, 3(4), 245-262

Dwenger, N., Storck, J., & Wrohlich, K. (2012). Do tuition fees affect the mobility of university applicants? Evidence from a natural experiment. *Economics of Education Review*, 31(1), 155-167

Dynarski, S. M. (2003). Does aid matter? Measuring the effect of student aid on college attendance and completion. *American Economic Review*, 93(1), 279-288

Eide, E. & Showalter, M. W. (1998). The effect of school quality on student performance: a quantile regression approach. *Economics Letters*, 58, 345-350

Faggian, A. (2010). The effect of tuition fees on higher education participation in Northern Ireland: an interim report. *DELNI Research Report*, accessed at: http://www.delni.gov.uk/effect-of-variable-tuition-fees-interim-report.pdf

Feng, A. & Graetz, G. (2013). A question of degree: the effects of degree class on labor market outcomes. *CEP Discussion Paper*, No. 1221

Fernández, R. M. & Shioji, E. (2000). Human capital investment in the presence of unemployment: application to university enrolment in Spain. *University of Oxford Department of Economics Working Paper*, No. 66

Fredriksson, P. (1997). Economic incentives and the demand for higher education. *Scandinavian Journal of Economics*, 99(1), 129-142

Freeman, R. (1971). The market for college-trained manpower. Cambridge, MA: Harvard University Press.

Frenette, C. (2004). Access to college and university: does distance to school matter? *Canadian Public Policy / Analyse De Politiques*, 30(4), 427-443

Galindo-Rueda, F., Marcenaro-Gutierrez, O., & Vignoles, A. (2004). The widening socio-economic gap in UK higher education. *National Institute Economic Review*, 190, 70-82

Gayle, V., Berridge, D., & Davies, R. (2002). Young people's entry into higher education: quantifying influential factors. *Oxford Review of Education*, 28, 5-20

Gibbons, S., & Vignoles, A. (2009). Geography, choice and participation in higher education in England. *Regional Science and Urban Economics*, 42, 98-113

Gist, W. E., Goedde, H., & Ward, B. H. (1996). The influence of mathematical skills and other factors on minority student performance in principles of accounting. *Issues in Accounting Education*, 11(1), 49-60

Greenaway, D., & Haynes, M. (2003) Funding higher education in the UK: the role of fees and loans. *The Economic Journal*, 113, F150-F166

Greene, W. H. (2010). Testing hypotheses about interaction terms in nonlinear models. *Economics Letters*, 107, 291-296

Greene, W. H. (2003). Econometric Analysis. Upper Saddle River, N.J., Prentice Hall.

Greenwood, C., Harrison, M., & Vignoles, A. (2011). The labour market value of STEM qualifications and occupations. *Institute of Education / The Royal Academy of Engineering*. Retrieved from: http://www.raeng.org.uk/publications/reports/the_labour_market_value_of_stem

Harmon, C., Oosterbeek, H., & Walker, I. (2000). The returns to education: a review of the evidence, issues and deficiencies in the literature. *Centre for the Economics of Education*, London School of Economics and Political Science.

Harrison, M. (2012). Jobs and growth: the importance of engineering skills to the UK economy. The *Royal Academy of Engineering*. Retrieved from: http://www.raeng.org.uk/publications/reports/jobs-and-growth

Harrison, N. (2011). Have the changes introduced by the 2004 Higher Education Act made higher education admissions in England wider and fairer? *Journal of Education Policy*, 26(3), 449-468

Heckman, J. J., Ichimura, H., & Smith, J. (1998). Characterizing selection bias using experimental data. *Econometrica*, 66(5), 1017-1098

HEFCE (2010). Student ethnicity: profile and progression of entrants to full-time, first degree study. *Higher Education Funding Council for England*. Retrieved from: http://www.hefce.ac.uk/media/hefce1/pubs/hefce/2010/1013/10_13.pdf

HEFCE (2013). Higher education and beyond: outcomes from full-time first degree study. *Higher Education Funding Council for England*. Retrieved from: http://www.hefce.ac.uk/media/hefce/content/pubs/2013/201315/Higher%20education%20and%20beyond%20Outcomes%20from%20full-time%20first%20degree%20study.pdf

HEFCE (2014). Differences in degree outcomes: Key findings. *Higher Education Funding Council for England*. Retrieved from: http://www.hefce.ac.uk/media/hefce/content/pubs/2014/201403/HEFCE2014_03.pdf

HEFCE (2015a). Causes of differences in student outcomes. *Higher Education Funding Council for England*. Retrieved from: http://www.hefce.ac.uk/media/HEFCE,2014/Content/Pubs/Independentresearch/2015/Causes,of,differences,in,student,outcomes/HEFCE2015_diffout.pdf

HEFCE (2015b). Differences in degree outcomes: the effect of subject and student characteristics. *Higher Education Funding Council for England*. Retrieved from: http://www.hefce.ac.uk/media/HEFCE,2014/Content/Pubs/2015/201521/HEFCE2015_21.pdf

Heller, D. E., (1997). Student price response in higher education: An update to Leslie and Brinkman. *Journal of Higher Education*, 68(6), 624-659

Hemelt, S. W., & Marcotte, D. E. (2008). Rising tuition and enrollment in public higher education. *IZA Discussion Paper*, No. *3827*

Hillman, N. W. & Orians, E. L. (2013). Community colleges and labor market conditions: how does enrollment demand change relative to unemployment rates? *Research in Higher Education*, 54, 765-780

Hübner, M. (2012). Do tuition fees affect enrollment behavior? Evidence from a 'natural experiment' in Germany. *Economics of Education Review*, 31(6), 949-960

Huijsman, R., Kloek, T., Kodde, D. A., & Ritzen, J. M. M. (1986). An empirical analysis of college enrollment in the Netherlands. *De Economist*, 134(2), 181-90

Hussain, I., McNally, S., & Telhaj, S. (2009). University quality and graduate wages in the UK. *Centre for the Economics of Education Discussion Paper*, No. 99

Imbens, G. M., & Wooldridge, J. M. (2008). Recent developments in the econometrics of program evaluation. *NBER Working Paper*, No. *14251*

Johnes, G. & McNabb, R. (2004). Never give up on the good times: student attrition in the UK. *Oxford Bulletin of Economics and Statistics*, 66(1), 23-47

Johnson, M. T. (2013). The impact of business cycle fluctuations on graduate school enrollment. *Economics of Education Review*, 34, 122-134

Kahn, L. B. (2010). The long-term labor market consequences of graduating from college in a bad economy. *Labour Economics*, 17(2), 303-316

Kane, T., (1994). College entry by blacks since 1970: The role of college costs, family background, and the returns to education. *Journal of Political Economy,* 105, 878-911

Karaca-Mandiac, P., Norton, E., C., & Dowd, B. (2012). Interaction terms in nonlinear models. *Health Services Research*, 47, 255-274

Kelchtermans, S., & Verboven, F. (2010). Participation and study decision in a public system of higher education. *Journal of Applied Econometrics*, 25, 355–391

Koch, J. V. (1972). Student choice of undergraduate major field of study and private internal rates of return. *Industrial and Labor Relations Review*, 26(1), 680-685

Koh, M. Y. & Koh, H. C. (1999). The determinants of performance in an accountancy degree programme. *Accounting Education*, 8(1), 13-29

Laselle, L., McDougall-Bagnall, J., & Smith, I. (2014). School grades, school context and university degree performance: Evidence from an elite Scottish institution. *Oxford Review of Education*, 40(3), 293-314

Lenton, P. (2014). Personality characteristics, educational attainment and wages: an economic analysis using the British Cohort Study. *Sheffield Economic Research Paper Series*, No. 2014011

Leslie, L. L., & Brinkman, P. T. (1987). Student price response in higher education. *Journal of Higher Education*, 58, 181-204

Long, B. T. (2004). How have college decisions changed over time? An application of the conditional logistic choice model. *Journal of Econometrics*, 121, 271-296

Long, B. T. (2014). The financial crisis and college enrollment: how have students and their families responded? In J. Brown and C. Hoxby (Eds.), *How the financial crisis and the great recession affected higher education*. Forthcoming from University of Chicago Press.

McGuigan, M., McNally, S., & Wyness, G. (2014). Student awareness of costs and benefits of educational decisions: effects of an information campaign and media exposure. *Department of Quantitative Social Science Working Paper*, No. *14-19*

McNabb, R., Pal, S., & Sloane, P. (2002). Gender differences in educational attainment: the case of university students in England and Wales. *Economica*, 69, 481-503

McPherson, M. S., & Shapiro, M. O. (1991). Does student aid affect college enrollment? New evidence on a persistent controversy. *American Economic Review*, 81(1), 309-318

Mellanby, J., Maxtin, M., & O'Doherty, J. (2000). The 'gender gap' in final examination results at Oxford University. *British Journal of Psychology,* 91, 377-390

Mellors-Bourne, R., Connor, H., & Jackson, C. (2011). STEM graduates in non STEM jobs. *Department for Business, Innovation and Skills Research Paper*, No. 30

Mitze, T., Burgard, C., & Alecke, B. (2013). The tuition fee ‘shock’: analysing the response of first-year students to a spatially discontinuous policy change in Germany. *Papers in Regional Science*, 94(2), 385-419

Montmarquette, C., Cannings, K., & Mahseredjian, S. (2002). How do young people choose college majors? *Economics of Education Review*, 21, 543-556

Murphy, R., Scott-Clayton, J, & Wyness, G. (2017). The End of Free College in England: Implications for Quality, Enrolments, and Equity. *NBER Working Paper*, No. 23888

Naylor, R., Smith. J., & McKnight, A. (2000). Sheer class? Returns to educational performance: evidence from UK graduates’ first destination labour market outcomes. *University of Warwick Economics Research Paper*, No. 786

Naylor, R. & Smith, J. (2004). Determinants of educational success in higher education, in: G. Johnes & J. Johnes (Eds) *International Handbook on the Economics of Education* (Cheltenham, UK, Edward Elgar), 415-461

O’Leary, J. (2016). Are arts and humanities still worth pursuing? *The Times*, 27 September. Available at: https://www.thetimes.co.uk/article/are-arts-and-humanities-still-worth-pursuing-6dlzprm6g (Accessed at 26 September 2017)

Oosterbeek, H. & Van Ophem, H. (2000). Schooling choices: preferences, discount rates and rates of return. *Empirical Economics*, 25(1), 15-34

Oreopoulos, P. (2006). Estimating average and local average treatment effects of education when compulsory schooling laws really matter. *American Economic Review*, 96(1), 152-175

Osborne, M., Leopold, J., & Ferrie, A. (1997). Does access work? The relative performance of access students at a Scottish university. *Higher Education*, 33(2), 155-176

Parker, J. D. A., Summerfeldt, L. J., Hogan, M. J., & Majeski, S. A. (2003). Emotional intelligence and academic success: examining the transition from high school to university. *Personality and Individual Differences*, 36, 163-172

Pennell, H. & West, A. (2005). The impact of increased fees on participation in higher education in England. *Higher Education Quarterly*, 59(2), 127-137

Pissarides, C. A. (1981). Staying-on at school in England and Wales. *Economica*, 48, 345-363

Pitcher, J & Purcell, K. (1998). Diverse expectations and access to opportunities. *Higher Education Quarterly*, 52(2), 179-203

Poropat, A. E. (2009). A meta-analysis of the five-factor model of personality and academic performance. *Psychological Bulletin*, 135, 322-338

Powdthavee, N. & Vignoles, A. F. (2009). The socio-economic gap in university drop-out. *The B. E. Journal of Economic Analysis and Policy*, 9, 19-39

Puhani, P. A. (2012). The treatment effect, the cross difference, and the interaction term in nonlinear "difference-in-differences" models. *Economics Letters*, 115, 85-87

Rice, P G. (1987). The demand for post-compulsory education in the UK and the effects of educational maintenance allowances. *Economica*, 54, 465-475

Royston, P., Altman, D. G., & Sauerbrei, W. (2006). Dichotomizing continuous predictors in multiple regression: a bad idea. *Statistics in Medicine*, 25(1), 127-141

Skatova, A. & Ferguson, E. (2014). Why do different people choose different university degrees? Motivation and the choice of degree. *Frontiers in Psychology*, 5, 1244

SLC (2006). Higher Education Grants in England and Wales: Academic Year 2005/06 (Provisional), accessed at: http://www.slc.co.uk/media/5510/slcsfr022006.pdf

Sloane, P. J. & O'Leary, N. C. (2004). The return to a university education in Great Britain. *IZA Discussion Paper*, No. 1199

Smith, H. L., & Powell, B. (1990). Great expectations: variations in income expectations among college seniors. *Sociology of Education*, 63(3), 194-207

Smith, J., McKnight, A., & Naylor, R. (2000). Graduate employability: policy and performance in higher education in the UK. *The Economic Journal*, 110, F382-F411

Smith, J. & Naylor, R. (2001a). Determinants of degree performance in UK universities: a statistical analysis of the 1993 student cohort. *Oxford Bulletin of Economics and Statistics*, 63(1), 29-60

Smith, J. & Naylor, R. (2001b). Dropping out of university: a statistical analysis of the probability of withdrawal for UK university students. *Journal of the Royal Statistical Society: Series A*, 164, 389-405

Smithers, A. & Griffin, A. (1986). Mature students at university: entry, experience and outcomes. *Studies in Higher Education*, 11(3), 257-268

Stratton, L S., O'Toole, D. M., & Wetzel, J. N. (2007). Are the factors affecting dropout behavior related to initial enrollment intensity for college undergraduates? *Research in Higher Education*, 48(4), 453-485

Toutkoushian, R., K. & Paulsen, M., B. (2016). Private and social returns to higher education. In: Economics of Higher Education. Springer, Dordrecht, pp. 93-147

Universities UK. (2009). Variable tuition fees in England: assessing their impact on students and higher education institutions. *A fourth report.*

Universities UK. (2010). Changes in student choices and graduate employment.

Universities UK. (2014). Patterns and trends in UK higher education.

Wakeling, P., & Jefferies, K. (2013). The effect of tuition fees on student mobility: the UK and Ireland as a natural experiment. *British Educational Research Journal*, 39(3), 491-513

Wales, P. (2010). Geography or economics? A micro-level analysis of the determinants of degree choice in the context of regional economic disparities in the UK. *Spatial Economics Research Centre Discussion Paper*, No. 56

Walker, I. & Zhu, Y. (2011). Differences by degree: evidence of the net financial rates of return to undergraduate study in England and Wales. *Economics of Education Review*, 30, 1177-1186

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

Whitfield, K. & Wilson, R. A. (1991). Staying on in full-time education: the educational participation rate of 16-year olds. *Economica*, 58, 391-404

Wooldridge, J.M. (2011). Econometric analysis of cross section and panel data. Cambridge, Mass, MIT.

Wyness, G. (2010). Policy changes in UK higher education funding, 1963-2009. *Department of Quantitative Social Science Working Paper*, No. 10-15