UNIVERSITY OF SOUTHAMPTON

FACULTY OF HUMAN AND SOCIAL SCIENCES

Division of Social Statistics and Demography

THE MIGRATION PROCESSES OF STUDENTS INTO HIGHER EDUCATIONAL INSTITUTIONS IN THE UNITED KINGDOM

by

Neil Graeme Bailey

Thesis for the degree of Doctor of Philosophy

October 2015

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF HUMAN AND SOCIAL SCIENCES

Social Statistics and Demography

Thesis for the degree of Doctor of Philosophy

THE MIGRATION PROCESSES OF STUDENTS INTO HIGHER EDUCATIONAL INSTITUTIONS IN THE UNITED KINGDOM

Neil Graeme Bailey

The higher educational system in the UK plays a crucial role in the economic development of the country and significantly impacts on the future labour market outcomes for individuals. With participation rates in higher education continually increasing and the recent changes to student financing burdening the student with more of the costs of higher education, the decision of where and what to study has become increasingly important.

Despite this, there has been little work conducted to date that analyses in detail the migratory patterns of the large student population in the United Kingdom and their movements into Higher Education (HE). The overall aim of this thesis is to advance the current understanding of the student migration processes in the United Kingdom by considering three broad areas of enquiry and analysis; patterns and measurement of student migration, characteristics and correlations of student migration and lastly, future outcomes of student migration. This research uses data from the Higher Educational Statistics Agency (HESA) to provide a cross-sectional snapshot of the student migration situation in the UK.

The thesis puts forward a unique typology that is used to categorise and measure the different migration decisions that a person can undertake in order to attend a Higher Educational Institution (HEI). Using this typology, the results demonstrate that, the previously assumed traditional transition into higher education of migrating away from the parental home to study at a HEI is no longer the majority transition experienced by HE students in the UK. Secondly, a new spatial classification of student migration is created and the results show a clear difference in the migration outcomes of students from the South of the UK compared to the North, with the latter being less likely to migrate. Statistical modelling of the student migration process in the UK showed that migration into a HEI in the UK is not equal across ethnicity, socio-economic background and gender. Finally, the results regarding the impact of migrating in order to attend a HEI on the labour market outcomes after graduating were marginal. No clear causal impacts of the migration decision on the future labour market outcomes were identified.

Contents

List of tables

List of figures

DECLARATION OF AUTHORSHIP

I, Neil Graeme Bailey declare that the thesis entitled

The migration processes of students into higher educational institutions in the United Kingdom

and the work presented in the thesis are both my own, and have been generated by me as the result of my own original research. I confirm that:

- this work was done wholly or mainly while in candidature for a research degree at this University;
- where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
- where I have consulted the published work of others, this is always clearly attributed;
- where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
- I have acknowledged all main sources of help;
- where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
- none of this work has been published before submission

Signed:

Date: 26th October 2015

Acknowledgements

I am grateful to a number of organisations and individuals who have made the completion of this thesis possible due to their financial support. Firstly, I would like to acknowledge the Economic and Social Research Council and the University of Southampton Doctoral Training Centre for funding my MSc and PhD Studies through the 1+3 programme. Without this funding it would not have been possible for me to complete this thesis. Furthermore, I am grateful for the other financial support I received from the Department of Social Statistics and Demography, the Faculty of Human and Social Sciences, Dr Jakub Bijak and Professor James Raymer for contributing extra funds into my research, training and support grant. This extra funding was used to purchase data used in the thesis and enabled me to travel and present my work at numerous international conferences, at which I received some invaluable feedback from various academic peers which was an invaluable resource in helping me develop my work throughout the PhD process.

I am also grateful to many individuals who contributed towards this thesis through support, guidance, advice and encouragement. The majority of this came from the individuals that made up my supervisory team over the past four years. Dr Jakub Bijak, Dr Sylke Schnepf and Professor James Raymer, have all contributed over the long PhD process; from forming research topics and ideas at the start, providing technical and statistical guidance in the middle and providing feedback on written draft chapters of this thesis towards the end. Without all their help throughout the process I would not have been in the position to complete this thesis!

I must acknowledge how helpful HESA were in all my correspondence regarding data requests and their data has been a vital component to this PhD. I also need to acknowledge the use of the IRIDIS High Performance Computing Facility, and associated support services at the University of Southampton, in the completion of this research.

Lastly, I would like to thank all my friends, family and housemates who have supported me over the past four years. I am especially grateful to Sarah Walsh and Ken Bailey for their continued support throughout the process and for helping with proofreading sections towards then end.

Abbreviations

APEL	Accreditation of Prior Experimental Learning
APL	Accreditation of Prior Learning
CIA	Conditional Independence Assumption
CS	Common Support
FE	Further Education
GCE	General Certificate of Education
GCSE	General Certificate of Secondary Education
GPs	General Practitioners
HE	Higher Education
HEI	Higher Educational Institution
HEIPR	Higher Education Initial Participation Rate
HESA	Higher Education Statistics Agency
IB	International Baccalaureate
IIA	Independence of Irrelevant Alternative Assumption
IPS	International Passenger Survey
ISCED	International Standard of Certificate of Education
LA	Local Authority
MAUP	Modifiable Area Unit Problem
MNLM	Multinomial Logistic Regression Modelling
NHS	National Health Service (UK)
NHSCR	National Health Service Central Register
OECD	Organisation for Economic Co-operations and Development
ONS	Office for National Statistics

PRDS	General Practitioners Patient Register Data System
PSM	Propensity Score Matching
UCAS	Universities and Colleges Admissions Service
UK	United Kingdom
UN	United Nations
VCE	Vocational Certificates of Education
ZIP	Zero Inflated Poisson Model

1. Thesis Introduction

1.1 Student Migration: Introductory Remarks and Rationale

The Higher Education (HE) sector in the United Kingdom (UK) plays a crucial role in the country's development as a world leading economy and significantly impacts on individuals' future labour market outcomes (Dearden et al. 2004; O'Leary and Sloane 2005; BIS 2013a; Walker and Zhu 2013). Having a higher educational degree has become increasingly important for individuals aiming to enter the labour market in the UK. Over 50% of those in employment in the UK aged between 30 and 39 now have a higher educational degree, an increase of nearly 20% since 2003, and in testing economic conditions, graduates continue to experience better outcomes than non-graduates in both life time earnings and employability (Universities UK 2014).

There have also been significant changes in the UK regarding higher educational finances, in the way that students fund their studies and the way in which Higher Educational Institutions (HEIs) are funded. Individuals will now incur higher costs for studying at a HEI and for many potential students this will impact on the decision of where to study because, whilst they cannot control the cost of tuition fees, they can still maintain some control over their costs of living, accommodation, travel and so forth (Wilkins et al. 2013; Bachan 2014; Crawford and Jin 2014). Meanwhile, the current environment in which HEIs operate means they need to ensure they attract enough numbers of students to meet enrolment and budgetary targets. As a result, the decision of individuals about where to study has become increasingly important, not only to the individual themselves, but increasingly to the HEIs as well.

People have a choice about where they study and this thesis will examine how these choices differ across a range of factors and how these different choices impact on an individual's future life outcomes. The decision to migrate to enter HE and where to live is a complex one. People may migrate or not migrate for a variety of reasons, however, the concept of student debt and student financing has become increasingly prevalent in the decision making process over recent years (Reay et al. 2005; Callender and Jackson 2008; Reay et al. 2010; Bachan

2014). Accordingly research on these student migration patterns and the mobility of different population groups is highly important in the society of today. It will help inform individuals and HEIs in their decision-making. The study of student migration movements and motivations will also assist national and local governments in their higher education polices, and this thesis can inform private corporations in daily business and planning decisions, including the provision of services and facilities.

Previous research has examined the prevalence and spatial patterns of population movements by age, sex, ethnicity and employment status of the population, as well as a variety of other socio-economic characteristics (Patiniotis and Holdsworth 2005; Gibbons and Vignoles 2012; Holton and Riley 2013). However, research on the migration of students entering HE is one area within the wider literature of mobility and migration studies, which is conspicuous by its scarcity.

Demographers have long recognised that one of the major attributes affecting an individual's propensity to migrate is age, which acts as a proxy for events in the life course such as marriage or divorce (Courgeau 1985), and that persisting regularities appear in empirical age-specific migration schedules (Rogers and Castro 1981a; Cadwallader 1992; Raymer et al. 2006; Raymer et al. 2007; Dennett and Stillwell 2010; Wilson 2010). Internal migration statistics from the Office for National Statistics (2012) support these typical age-specific patterns, with young adults being the most likely to migrate. Around one in five people aged 18-19, and living in England and Wales in mid-2010 migrated to a different local authority (LA). This accounted for around six per cent of all the migration moves in that year, whilst another peak was observed amongst those aged 22. The Office for National Statistics (2012) internal migration report stated that the peaks in internal migration at young adult ages could largely be explained by moves to and from university or other HEIs. This is supported by evidence linking HE and the lifecourse of individuals which is strongly associated with age (Courgeau 1985; Pollard 2003). This observation was supported in the work conducted by Wilson (2010) in which an extension to the model migration schedule in an international context was proposed to account for highly age-concentrated migration patterns that were strongly related to entry and exit from HE.

The significance of analysing students as a separate group in migration research was first recognised in the Tuckman (1970) seminal article on the determinants of college student migration in the United States. Yet despite acknowledgement of the importance of students within the wider migration literature (Long 1988; Skeldon 1997; Salt 2001; Skeldon 2012), to date, there are only limited studies of the internal migration of university students in the UK. This appears neglectful in a landscape where inter-regional migrants dominate the make-up of the majority of UK student populations (Smith 2002; Smith and Jons 2015). The general lack of research into inter-regional student flows can be largely attributed to the lack of robust migration datasets that accurately measure student migration flows (Champion and Coombes 2007). However, with the combination of acquired student population data from Higher Educational Statistics Agency (HESA) and the development of a new student migration typology, this thesis aims to supply three substantive analysis chapters that will begin to fill this void in the literature.

The motivation for focusing on student migration behaviour originated from the increasing multidimensionality and complexity of the higher educational landscape in UK in recent decades. The HE sector in the UK has been extremely policy relevant and has received a high media profile for many years. Tony Blair's 'Education, Education, Education' speech in the 1996 Labour Party conference and the resultant polices that encouraged increased and widened participation in HE was a significant factor in the growth of interest within the higher educational sector. The policies included those to ensure equal access into HE, with the aim that anyone with the ability and desire to study at a HEI would be given the opportunity to do so regardless of their socio-economic or ethnic background. More recently, the impact of raising tuition fees (HMSO 2011) received considerable media attention and the impact these policy changes will have on access to HE are of serious relevance to policy makers. These increased fees prompted violent protests on the streets of London and inflicted deep damage on the reputation of the Liberal Democrats, who went into the 2009 General Election pledging to vote against a fee rise. The number of university applicants in England dropped by 8.8% after the tuition fee increase compared with the tally before the fee increase in 2010 (Vasagar 2012), which was the first drop in applications since the large restructuring of the HE sector and the Tony Blair government.

Another important policy that is closely related to the aspects under investigation within this thesis is the 'Widening Participation Policy' (HEFCE 2013), which states that anyone with the ability and desire to go to university should have the opportunity to do so, whatever their geographical location or socio-economic background. It is, therefore, important to analyse the spatial patterns of student migration to investigate whether government policy for equal access to HE for all is actually achieving its aims in terms of spatial and individual level differences in student migration outcomes. With the increasing costs of HE being borne by the student, this thesis also provides analysis into whether or not migrating to attend a HEI is economically beneficial in regards to an individual's future labour market outcomes. This is extremely policy relevant as well as beneficial to future potential students to aid their decision making process around where to study.

These examples of the high profile and policy relevance of the higher educational sector in the UK provided the motivation to investigate this research area in more detail.

In order to become a student and study at a HEI an individual has to make three important decisions, all of which are impacted by many influencing factors. The first major decision is whether to participate in HE or not. The second decision - dependent on the first decision being positive towards participation – relates to which institution to attend and the third on which course to study and thereby some form of student migration decision process is undertaken.

In layman's terms, once a person has decided they want to enter HE, in order for a person to become a student at a HEI they have three basic options. The first being, if there is a higher educational facility in close proximity, the person could attempt to study there. The second would be for the person to identify an institution they wish to attend elsewhere in the country and apply to study there and consequently migrate to this place to live and study if accepted. The final choice would be to identify an institution they wish to attend elsewhere in the country and apply to study there and if accepted travel to this place in order to study but do not migrate.

However, in reality, the decision process that an individual must go through in order to attend a HEI in the UK is much more complex and there are many important facets that impact on the transition into HE.

This thesis is concerned with quantifying the different migration movements of students entering into HEIs in the UK, the characteristics that are correlated with these movements, and how the different movements impact on labour market outcomes after graduating.

In the UK HE system there has been a long held assumption that the majority of young people experience a traditional transition into HE in which young people move away from their parental home in order to attend a HEI a relatively large distance away from their parents (Chatterton 1999; Patiniotis and Holdsworth 2005; Chatterton 2010). Thereby acquiring the personal and cultural capital, in the transition from youth to adulthood, as well as academic credentials and skills to enter the labour market (Smith and Sage 2014; Smith and Jons 2015). Students have also been traditionally perceived as a highly mobile part of society, and a large amount of population turnover has been associated with student areas, as young people move to study at their desired HEI. This significant amount of population turnover associated with student areas has been shown to have a profound effect on university towns and cities (Duke-Williams 2009). Therefore, it is also important to quantify, at a local and national level, the patterns of student migration that contribute to these large scale population changes and how they redefine the areas in which students decide to settle.

It is important for each HEI to be able to measure the type of migration movement their students undertake in order to attend their HEI. This information can then be used to focus the HEI's recruitment strategies (Beech 2014b, a) in order to continue to address the needs of the types of students that the institution is already attracting, as well as allowing institutions to target those types of students they are not currently attracting but would like to. Local housing and service providers will also benefit from information about the types of students that are attracted to their areas. Those areas with large numbers of student migrants would need more services and available housing than those areas which host mainly local students. As a result, an in-depth analysis of the student migration patterns – whether they migrate, where they are migrating from and where they are migrating to – is expected to benefit not only the HEIs themselves, but local and national government, planning authorities, the national population statistic offices, as well as other entities, such as transport providers and retailers in the surrounding economy.

Students have a large impact on the local economy, the services provided within an area, and the types of housing that are made available. However, the specific impact on a student area often depends on the type of transition that the student has made in order to attend an HEI. With the vast and significant changes already observed in the HE sector, one motivation of this thesis is to examine if these traditional transitions are still the dominant experience of students in the UK or, if like the sector itself, the migration choices experienced by the majority of students in this analysis are different to those reported in the previous literature.

Alongside the large structural and policy changes observed in the higher educational sector in recent decades, there has been progressively increasing enrolment into HE in the UK. The numbers of people entering HE is now at a record high. The number of students attending HEIs in the UK has steadily expanded over the last half century, while over the past 15 years student numbers have risen sharply from 1.7 million in the academic year 1995/96 to 2.6 million in 2010/11 (Higher Education Statistics Agency 2012b). The latest figures indicate that the Higher Education Initial Participation Rate (HEIPR)[1] for the 2011/12 academic year for English domiciled students was at a record high of 49% (BIS 2013b). This indicates that almost half of all 17 year olds that lived in England at the start of the 2011/12 academic year will participate in HE by the age of thirty, assuming the current age-specific participation rates.

With enrolment rates generally increasing, it is imperative that there is a good evidence-based understanding of the patterns of students' migration behaviour and the student characteristics that are associated with these observations. The migration decision process experienced often differs widely between individual students, affecting the distance students choose to relocate and

[1] The Higher Education Initial Participation Rate (HEIPR) is calculated individually for the four constituent countries of the United Kingdom. A description of how the measure is calculated and links to the most updated statistical releases are available at: https://www.gov.uk/government/collections/statistics-on-higher-education-initial-participation-rates

where they choose to relocate to. The decision to migrate in order to attend a HEI will be influenced by many overarching factors. These factors include ethnicity, socio-economic status, parental background and educational achievement. Many previous studies have indicated that inequalities exist for a variety of reasons across the life course, for example; differences between ethnic and socio-economic groups in their educational attainment, access to HE and future career earnings (Blanden and Gregg 2004; Blanden and Machin 2004; Blanden et al. 2010). However, despite anecdotal evidence of motivations and the relationship to the spatial patterns of migration, there has been little attention paid to the spatial migration patterns of HE students attending HEIs in the UK and how these differ between certain social demographic groupings.

This thesis contributes to filling the current gap in knowledge about HE related migration. It does so by examining information on the intensity, spatial patterns and differences between personal characteristics and social groupings in HE related migration to institutions within the UK. It also analyses how HE related migration can impact on individuals' future labour market outcomes. The key contributions, value added and originality of the thesis are explained in more detail below, where the aims and scope of the thesis are introduced.

1.2 Thesis Aims, Research Questions and Scope

The overall aim of this thesis is to advance the current understanding of the student migration process in the United Kingdom. To achieve this aim the thesis is split into three broad areas of enquiry and analysis; patterns and measurement of student migration, characteristics and correlations of student migration and future outcomes of student migration. This research uses data from the Higher Educational Statistics Agency (HESA) to provide a cross-sectional snapshot of the student migration situation in the UK for the 2010/11 and 2011/12 academic years, while the analysis is conducted on three geographic levels, UK Counties, UK Local Authorities and UK Government Office Regions (GORs).

The aim of the first substantive analysis chapter is to identify and accurately measure the different types of student migration taking place across the United Kingdom. In order to capture the true complexity of student migration in the UK a new innovative typology of student migration is proposed. The typology is required in order to take advantage of the three locational variables within the dataset to depict the different possible types of student migration an individual could undertake. Without this unique and innovative typology the analysis of the student migration data would lack the necessary detail to isolate the different types of student migration.

With the use of the typology the chapter aims to provide a descriptive overview of the current situation of student migration across the UK. The student migration typology is then used in the creation of an area classification of student migration for the UK. The reason for creating this student migration area classification is to add statistical robustness to the creation of the typology by running cluster analysis techniques to analyse the typology groupings in a spatial context. The student migration area classification will then also be used with the aim of analysing any spatial patterns in student migration across the UK.

The aim of the second substantive analysis chapter is to build upon the first chapter by analysing how the student migration outcomes (from the typology) are impacted by the individuals' personal level characteristics. Therefore, gaining an in-depth understanding of how the student migration processes of people entering HEIs in the UK are impacted by a student's characteristics, the course they studied and the institute they attended.

This chapter uses a variety of statistical techniques including linear, logistic and multinomial regression. Analysis of this data using these techniques has not been conducted previously. The aim is to provide further insights into what factors may be impacting on the different migration transitions experienced by people entering into HE and how these may differ between social and ethnic groups or spatially across the country.

The aim of the final analysis chapter is to progress from looking at the overall pattern of student migration and how this correlates to different personal characteristics to investigate how this student migration decision impacts on the individual in later life. The main aim is to identify the true economic value of migrating in order to attend a HEI. No previous research has estimated the impact of the student migration outcome on the future economic outcomes of the student. As a result, the aim is to identify if there is any economic benefit for migrating in order to attend a HEI.

The research aims set out above lead to the following specific research questions to be answered throughout this thesis:

1. How best can we examine the student migration process in the UK?
2. How well does the student migration typology perform in measuring the UK student migration process?
3. How were students attending HEIs distributed across the UK during the 2010/11 academic year?
4. How was the student population classified into the different student migration categories?
5. Can this typology be used to create a Student Migration Area Classification for the UK?
6. How did student migration patterns and trends differ across the geographical areas of the UK?
7. How does a student's social background, ethnicity or gender impact on the migration outcomes experienced in order to attend a HEI?
8. How do the course studied and institution attended impact on the student's migration outcome?
9. How does student migration into higher education impact on the future employment status after graduation?
10. How does student migration into higher education impact on the first wage achieved after graduation?

The scope of this project focuses on the HE system of the United Kingdom for the academic years 2010/11 and 2011/12. These academic years were the most recent years available from the data providers at the start of the research project and the aim is to provide the most up-to-date analysis possible.

The study area of the United Kingdom is chosen because of the unique HE system in place in the country. The UK has around 160 higher educational institutions (HEIs) in a country that is relatively small in size on a global scale. As a result, students have a large amount of choice regarding where they chose to study and for the majority of people in the UK there will be several HEIs in a relatively short distance from their location of residence. The geographical distribution of the 160 HEIs in the UK is mapped in Figure 1-1, and this clearly illustrates the wide dispersal of HEIs across the whole UK.

The geography of the UK is split into many different types of administrative areas that differ in their size and structural composition. As a result of these differences in geographical areas a phenomenon known as the Modifiable Areal Unit Problem (MAUP) arises. The MAUP was first identified by Gehlke and Biehl (1934) while Openshaw (1984) provides a comprehensive review on the early research on the subject area. The MAUP is simply defined by Fotheringham and Wong (1991) as ‘the sensitivity of analytical results to the definition of units for which data are collected’. Wrigley et al. (1996) as read in Bell et al. (2002) state that there are two main aspects of the MAUP that are traditionally recognised; those of scale and those of zonation. Those of scale occur because an area may be divided into geographies with differing numbers of special units, while those of zonation occur because an area may be divided into the same number of units in a variety of ways.

Figure 1-1: Map of All Higher Educational Institutions in the United Kingdom

Source: University of Wolverhampton (nd)

Throughout the thesis three administrative levels of geography are used for the analysis; Local Authority (LA), County level and Regions. Local Authority (LA) is a generic term for any level of local government in the UK. In geographic terms LAs therefore include English counties, non-metropolitan districts, metropolitan districts, unitary authorities and London boroughs; Welsh unitary authorities; Scottish council areas; and Northern Irish district council areas (Office for National Statistics 2011b). The dataset from HESA contains 408 LAs in the UK; 328 in England, 32 in Scotland, 22 in Wales, and 26 in Northern Ireland.

Counties were formerly administrative units across the whole UK. Due to various administrative restructurings the only administrative areas still referred to as counties are the non-metropolitan (shire) counties of England. The English metropolitan counties, although no longer administrative units, are also used for statistical purposes (Office for National Statistics 2011b). The dataset from HESA contains 94 Counties in the UK; 47 in England, 8 in Wales, 11 in Scotland, 26 in Northern Ireland, Isle of Man and Channel Islands (same as LA). For reference, a labelled map of the UK counties can be found in Appendix A – Figure A-1.

Regions, formally known as Government Office Regions (GORs) are the highest tier of sub-national division used by the official statistic services. However, in 2011 it was decided to shift away from using this level of geography and provide a more local focus. Despite this there is still a requirement to maintain this regional level geography for statistical purposes within Eurostat (NUTS 1 regions) and therefore statistics are still available at this level. The United Kingdom is split in 12 regions; nine in England and Wales, Scotland and Northern Ireland being an individual region in their own entity. For reference, a labelled map of the UK regions can be found in Appendix A – Figure A-2.

Later in the thesis (Chapter 3 and 4) some analysis will look at the North-South differences in the UK. The North and South regions have been defined using the region level geography. The South West, South East, East of England and London regions are classified as being South Regions, while the rest are categorised as being Northern Regions (See: Dorling (2007, 2011, 2012); Dorling and Thomas (2011) and Thomas and Dorling (2011) for discussions on the North-South Divide in the UK).

1.3 Thesis Structure

The remainder of this thesis is organised into 5 chapters.

An overview of the research context that underpins the research and analysis presented in this thesis is provided in Chapter 2. The structure of the education system in the UK is introduced which is then followed by a detailed explanation of the UK HE system, including a brief history and its current structure. The second section introduces the various conceptual and methodological issues within the study of migration, with a latter focus primarily on the migration of students. This chapter is then concluded with the creation and discussion of the overarching conceptual framework used throughout the analysis presented within this thesis.

The first analytical chapter of this thesis, 'Towards a typology of student migration: Incorporating a student migration area classification for the United Kingdom' is presented in Chapter 3. It identifies that measuring student migration is a complex process in which many differing types of movement can be undertaken. As a result, in order to accurately define and measure all the different types of student migration this chapter proposes a typology that depicts the complexity of the student migration decision process. The chapter then develops on this typology by using it to illustrate the geographic patterns and recent trends of student migration in the UK. Finally, by using the student migration typology to accurately measure the student migration within an area, the data are used to create an area classification of student migration to analyse how patterns of student migration differ spatially across the UK.

The second analytical chapter of this thesis, 'Migration Choices of Students Entering Higher Education in the UK: What are the Impacts of Personal Characteristics, Institution Attended and Course Studied?' is then presented in Chapter 4. This chapter builds on the findings presented in Chapter 3 and shifts the focus of the analysis towards the differences in the student migration decision process as a result of each student's personal characteristics. The chapter uses a variety of statistical techniques to model three different outcome variables, all of which measure student migration in some form.

The final analytical chapter, ‘The value of gaining a higher educational degree in the UK: Does migration matter?’ is presented in Chapter 5. This chapter analyses the graduate’s labour market outcomes six months after graduating and how these are impacted by many factors including the migration decision experienced when entering HE. This chapter uses a unique linked dataset and a combination of statistical techniques in order to answer the policy relevant question: is migrating to attend a HEI worth it?

Finally, Chapter 6, ‘Thesis discussion and final conclusions’ presents the final overarching summary of the thesis. Within the chapter the main substantive contributions are highlights, ideas for further work expressed and limitations of the work acknowledged.

2. Analysing Student Migration in the United Kingdom: Conceptualisation and Previous Research

The principle aim of the current chapter is to set out the theoretical, practical and definitional terms of reference that will be relevant for the rest of this thesis. The underlying concepts that underpin the student migration process are outlined and explained to give a foundation for the analysis presented hereafter. An overview of previous relevant research is produced to give a general understanding of the issues that are involved in this study, the prior work that has already been conducted and to establish the gaps in the literature that this thesis aims to fill. A more detailed review of previous work is provided in the specific analysis chapters later in this thesis. This chapter is needed to provide a clear and succinct background to the field of study under investigation in this thesis while also motivating the analysis conducted in the subsequent chapters.

The concept of student migration involves the decision to participate in HE and the student migration decision process. As a result these will be explained in turn. Analysing the student migration decision involves evaluating factors that impact on people that migrate and those that do not. Therefore, to get a clear understanding of all the processes involved an evaluation of the UK HE system and how these impact on an individual's decision process will be provided. Subsequently, defining what constitutes a student migrant will become an important concept that shapes the first sections of the chapter. Once the concept of defining a student migrant has been discussed the focus shifts to the factors that may impact on the student migration decision process. This involves analysing what micro level individual characteristics will be evaluated and previous research introduced to show what factors have already been proven to impact on the HE experience and which areas need to be analysed further. Finally all the concepts, factors and previous literature are condensed and summarised in a clear and succinct conceptual framework of the student migration decision process.

2.1 The Higher Education System in the United Kingdom

2.1.1 Overview of the UK Education System

Higher education (HE) is just one sector of a wider education system in the UK. The education system directly influences an individual's progression through their life course and an individual's probability of participating in HE. To have a clear understanding of the pathways an individual must take and the challenges they must overcome in order to participate in HE, it is necessary to have a clear understanding of the education system in the UK as a whole.

Across the UK there are five stages of education; early years, primary, secondary, Further Education (FE) and Higher Education (HE). Education is compulsory for all children between the ages of 5 (4 in Northern Ireland) and 16 across the UK. However, the minimum age of school leaving has recently been increased in England where an individual must partake in some form of FE until their 18th birthday if they were born after 1st September 1997 (United Kingdom Government 2014), while FE is not compulsory in the other constituent countries of the UK. FE covers non-advanced education which can be taken at further (including tertiary) education colleges and HEIs.

The fifth stage, HE, is the sector of the education system being analysed throughout this thesis. HE is the study beyond assessment level of GCE (General Certificate of Education) A levels and their equivalent which, for most full-time students, takes place in universities, HEIs or colleges (United Kingdom Government 2012). The overall hierarchy of the education system in the UK is visually presented in Figure 2-1.

Figure 2-1: Flow Chart of Education System in the United Kingdom

Source: United Kingdom Government (2012)

The progression of an individual through the UK education system (as illustrated in Figure 2-1) has several stages that will impact on the likelihood of that individual participating in HE in the future and the student migration decision if that individual enters HE. Such stages include the choice of which primary and secondary school to attend, the subjects the individual choses to study at certain stages of the education system and more importantly the level of attainment achieved by the individual at stages of assessment.

The type (private or public) and choice of school attended is known to impact on educational attainment, probability of participation in HE and later-life outcomes (Ryan and Sibieta 2010). This shows that choices made even at the very early stages of the education system will impact on the probability of participating in HE. There are also key decisions an individual must make throughout the education system that will affect future progression to HE and employment. In year 9 of secondary school students usually choose what they want to study at GCSE level. This can then often affect their options of what they can study post-16 at the FE level which will then directly impact on what degree programs they can apply for at universities (Russell Group 2011). This directly impacts on the migration decision of students entering HE because some courses are only available at certain institutions.

One of the key processes of the education system that impacts on future participation in HE and the student migration process is the measurement of attainment. Throughout the education system a student's attainment is assessed by some form of assessment or examination. There are two key stages that have a direct impact on whether or not an individual will participate in HE and if they do these will also impact on the quality and reputation of the institution and course that the student is able to attend and therefore the student migration decision.

For a student to be able to attend a FE institution, most FE and Sixth Form institutions have entry requirements which are evaluated by the attainment of the student in their GSCE's or equivalent[2]. Performance in GCSE exams or

[2] GCSE stands for General Certificate of Secondary Education. GCSE examinations are taken by most pupils at the end of compulsory school education (year 11) in England, Wales and Northern Ireland and assesses all subjects graded from A* to G (with U being "ungraded")

equivalent in year 11 can also affect future options for study at university. For example, many medical courses will expect students to have very good grades (A and A*) in English, maths and science, and for degrees in business and psychology a grade B in maths is often needed. So GCSEs or equivalent are the first assessments of attainment that directly impact on HE.

The next stages of assessment of attainment are crucial for HE participation and are directly related to the university admissions process. In year 12 most students will take AS level exams. Universities will look at performance in these to see if they are broadly in line with the grades predicted at A-Level. The final entry decision to a HEI is often based on the individual attaining a certain level of qualification in their GCE A-Levels or equivalent.

The overall progression through the stages of the education system is a prerequisite to being able to participate in HE in the UK and many factors impact on this progression throughout a student's childhood and young adolescence. Once an individual has navigated the first four stages of the UK education system they will then have the choice of attempting to participate in HE. If they chose to participate in HE the individual is then exposed to a student migration decision.

The importance of a student's level of attainment at the different levels of assessment prior to entry to HE is crucial to the process being examined in this thesis. In later analysis chapters, differences in individuals' attainment levels will need to be carefully considered. In Chapter 4 in order to accurately evaluate what factors impact on the migration decision experienced the individual's attainment level needs to be considered in the evaluation process. Also in Chapter 5, when evaluating the value of the student migration decision the individual's ability will again be a very important factor to consider.

The focus of this section will now examine the higher educational system itself in more detail.

2.1.2 The Structure of Higher Education in the UK

The HE system and its many institutions have varied histories which are reflected in major differences seen today in universities missions, legal status, constitutional arrangements and organisation. In general, over the last 50 years HE in the UK has undergone significant changes and has been transformed from an elitist system to a mass participation system with differing types of higher educational courses now on offer (Reay et al. 2005). This mass participation system has been shown by the marked increase in participation rates from around six per cent in the early 1960's to just under 50% according the most recent published figures (BIS 2013b).

One of the major driving contributors towards this significant shift in participation was the Further and Higher Education Act 1992 (HMSO 1992) which enabled former polytechnics to gain degree-awarding powers and to use the word 'university' in their title. This saw a significant growth in the HE industry and significantly changed the numbers and types of students participating in HE.

Despite the rapid expansion of the HE sector during the 1990's, public funding for HEIs fell by around 25 per cent, putting considerable pressure on universities and colleges alike. In response to these pressures, in May 1996, the National Committee of Inquiry into HE was established, by agreement between the main political parties, to make recommendations on how the purposes, shape, size and funding of HE, including support for students, should develop to meet the needs of the UK over the next 20 years. The committee, chaired by Sir Ron Dearing, reported in July 1997 (Dearing 1997). Key themes and recommendations of the report included an increase and widening of participation, mainly through two-year courses of HE provided in colleges of further education and the implementation of measures to improve standards in teaching and to ensure the comparability of qualifications. The committee also made a number of recommendations concerning the funding of HE, including a proposal that full-time students in HE should pay some of the costs of their tuition.

As a result, the Teaching and Higher Education Act 1998 (HMSO 1998) introduced measures to change financial support for students, including tuition fees to be paid by all except the poorest students from academic year 1998-99, the replacement of the maintenance grant for living expenses with loans from academic year 1999-2000, the availability of a supplementary hardship loan of £250 a year, and bursaries for students entering teacher training or health and social care courses. This was later followed by the

Higher Education Act 2004 (HMSO 2004) which allowed HEIs in the UK to charge variable tuition fees of up to £3,000 per year, rising only with inflation. The most recent development came after the publication of The Browne Review (Browne 2010), which suggested the removal of the cap on tuition fees and as a result The Education Act 2011 (HMSO 2011) was passed and gave publicly subsidised HEIs the right to charge up to £9,000 a year for their annual tuition for UK and EU domiciled students and considerably more for students from overseas.

The students under investigation in this research were enrolled at universities before the most recent tuition fee increase and were part of the cohort of students that had to pay tuition fees up to £3,000 per year, rising only with inflation. These costs of tuition differ between institutions and the funding support to students was means tested on their parental or own income. The cost of tuition therefore acts as an intervening obstacle in the student migration decision process. A student's socio-economic background will influence the impact of tuition fees while also impacting on the level of financial support available, while the impact of fees and the availability of student finance may influence participation, the choice of institution and therefore the student migration outcome. As a result the impact of cost of tuition and variable financial support on the analysis in later chapters should be considered.

2.1.3 Higher Education Institutions (HEIs)

One of the most influential factors that impact on the student migration process is the HEIs themselves. There were around 2.5 million students registered at the UK's 160 HEIs in the 2011/12 academic year (Universities UK 2013). The great majority of these HEIs were classified as government-dependent private institutions. They are autonomous, independent organisations, with their own legal identities and powers, both academically and managerially. They are not owned by the state, although they are dependent to a greater or lesser degree on state funding.

The publicly funded HE sector is very diverse, encompassing HEIs varying in size, history, mission and subject mix. HEIs which do not have university title include small, specialist institutions of art and design, drama, music and agriculture. Those with the power to award taught degrees but which do not meet the numerical criteria for university title (having at least 4,000 full time equivalent HE students, of whom at least 3,000 are registered on degree level courses) are entitled to apply to use the title ‘university college’. However, not all choose to do so. The use of titles other than ‘university’ and ‘university college’ is not controlled by law.

All universities offer research opportunities, as well as education in a wide range of taught subjects, although the balance between these activities varies between institutions. The balance between the types of qualifications offered and the subject mix also varies, within as well as across these categories. Institutions established as universities prior to 1992, such as Ancient, Red Brick and Plate Glass Institutions, typically focus on traditional academic courses at bachelor’s degree level and above, although many also provide a range of professionally accredited degree courses, such as medical studies, engineering and accountancy. The ‘new’ or ‘post-1992’ universities, often former polytechnics or teacher training colleges, typically offer a wider range of vocational courses, some of which may be below bachelor’s degree level.

Universities in the UK range greatly in their types and reputations and HEI can be split into five broad groups (see Section 4.3.3) ranging from the Ancient Universities such as University of Oxford and University of Cambridge, to Red Brick Institutions, Plate Glass Institutions and the Post-1992 and Recently Created Universities. Universities have also formed their own collaborative groupings or ‘mission groups’ based on their shared interests. The most well-known is the Russell Group which is an association of 20 major research-intensive universities of the UK. The group is so-called because it traditionally met at the Russell Hotel, London. Million+ is a university think-tank that works to help solve complex problems in HE and to ensure that policy reflects the potential of the UK's world-class university system. This group mainly comprises post-1992 universities.

The choice of HEI by the student plays an integral role in the migration decision process. The student will choose a HEI on a number of interlinking factors, however the actual location of the chosen HEI will have the most direct link to whether a migration takes place or not. It can be seen in Figure 1-1 that the UK is a unique country with many HEIs spread all over the country with most students having a wide array of choice of HEIs within relatively short travelling times. This wide network of available HEIs will play a key role in all the subsequent chapters when trying to analyse why some students migrate and some don't. Throughout the further analysis the location of available HEIs should always be considered when analysing the student migration decision.

HEI choice by a student will also be linked to the reputation of the HEI as well as the expected grades needed for the admissions process to the HEI which will discussed in more detail in the following sub-section. The choice of HEI is also interlinked to the course chosen by the student as well as the student's attainment, while also being strongly linked back to the student's individual characteristics and the intervening obstacles. The HEI therefore plays an integral role in the migration decision chosen and despite the rhetoric of widening participation by the UK government it is clear that the different types of HEIs attract very different proportions of student types and non-traditional students, draw upon students from varying catchment areas and provide very different social and academic experiences (Sutton Trust 2000; Crozier et al. 2008; Clayton et al. 2009).

2.1.4 The admissions process

In general there is no central control over admissions criteria across the UK. HEIs determine their own admissions policies and the entry requirements for each programme, which are set out in the institution's prospectuses.

For a few subject areas, there is a greater degree of central control. Undergraduate medical and dental courses are subject to quotas to ensure that the number of medical and dental students required to meet national needs is delivered. Nursing and midwifery degree provision is largely funded by the health authorities, which contract with institutions for the delivery of specified numbers of trainee nurses and midwives. In England, the Training and Development Agency for Schools sets intake targets for initial teacher training for those wanting to work in primary and secondary schools. Similar arrangements exist in Wales and Northern Ireland.

In all cases, prospective students apply for a specific programme and the minimum admissions requirements for each programme is determined by the individual institution. Many courses require some or all of the qualifications for entry to be in specific subjects or range of subjects and at specific grades. Again, these requirements are set out in the institution's undergraduate prospectus.

For full-time first cycle programmes at ISCED (International Standard Classification of Education) level 5 (e.g. bachelor's degrees), the minimum entry requirement is two or three GCE A-level passes, as well as a minimum number of General Certificate of Secondary Education (GCSE) passes at grade C or above. These remain the most common form of entry qualification held by young entrants to HEIs in the UK.

A wide range of other qualifications are acceptable for entry. They include GCE A-levels in applied subjects (formerly Vocational Certificates of Education, VCEs), Edexcel BTEC National Qualifications and the International Baccalaureate. In Wales, a Welsh Baccalaureate qualification is available in several schools and colleges; the advanced qualification is also acceptable for entry to HEIs.

Access courses provide another route, particularly for mature entrants. These programmes were originally designed for students over the age of 21 without formal qualifications but, since 2003-04, the lower age limit has been 19. Some access courses provide guaranteed entry to specific undergraduate courses on successful completion. Most HEIs also welcome applications from mature candidates who have had appropriate experience but may lack formal qualifications. Many HEIs give credit for prior study and informal learning acquired through work or other experiences; such as Accreditation of Prior Learning (APL) or Accreditation of Prior Experiential Learning (APEL).

In 2002, the Universities and Colleges Admissions Service (UCAS) introduced a points scoring system for expressing entry requirements. The 'UCAS tariff'' establishes agreed comparability between different types of qualifications in the whole of the UK, including GCE A-levels, some vocational qualifications, the Welsh Baccalaureate, the International Baccalaureate, and Scottish and Irish qualifications. However, HEIs are not obliged to express their entry requirements in these terms. An applicant who meets the published minimum admission requirements for a particular programme may be offered a place, but this is not guaranteed. Entry is competitive, and there are wide variations between institutions and programmes in terms of the competition for places.

It must therefore be considered that when the migration process is undertaken by a student entering a HEI, the student has to meet the admissions criteria in order to attend a specific HEI and as a result this will also play a significant role in the decision making process of that student and the overall migration outcome.

Another important process in the UK admissions process that will have a direct impact on the migration decision experienced by a student is the ‘clearing’ process. The clearing process is in place for those students that do not acquire the required grades to be admitted to their university of choice. The clearing process is a way of matching students without a university place with universities with vacancies. This process will directly impact on the student migration decision as a clearing place will often be in a different location to that originally chosen.

At this point it is relevant to highlight the insightful piece of work by Hoare and Johnston (2011) which examined the impact of the application and allocation processes of places at UK HEIs. They stated that although students can apply to any course and any university irrespective of where they live or their ability each HEI has the autonomy to accept or reject them. Acceptance is usually based on the applicant’s academic record and as a result Hoare and Johnston (2011) state that the allocation and admissions processes of HEIs and especially elite universities are likely to disadvantage the lower attaining students from disadvantaged backgrounds.

The admissions process is therefore a vital component of the student migration process and should be included in the future analysis where possible. However, the impact of the admissions process on the student migration decision will be hard to quantify.

2.2 The Study of Migration: Concepts, Methods and Theory

The study of human populations has seen the rise and development of a set of concepts and techniques collectively referred to as ‘demography’ (Rees 2009) and along with fertility and mortality, migration is one of the three key components of population change. The main facet under investigation in this thesis is the migration of students into HE, and therefore it is necessary to have a clear understanding of the concepts and methodological issues in the general field of migration studies.

Migration has widespread consequences, both for the individual involved and for the society within which it takes place and is both the cause and consequence of social change. As a result, policymakers have become increasingly aware of the role of migration in the context of economic growth, social well-being, political representation and urban change (Cadwallader 1992). The demographic and socio-economic composition of areas are determined by migration flows and a clear understanding of internal migration processes is crucial for anyone trying to analyse the general structure and process of demographic change within areas (Cadwallader 1992). Unlike births and deaths, which benefit from being relatively easy to define and measure, migration can often defy definition.

Researchers in the field of migration have long recognised the multiple issues and problems that beset the analysis and interpretation of population mobility (Bell 2004). Human mobility involves the movement of people between an origin and a destination (Belfield and Morris 1999) and is comprised of international and internal migration flows (Rowland 2003). However, human mobility can also occur in a much wider more complex such as residential mobility and the flows of students to university.

International migration, immigration and emigration, is the movement of people across international borders. In comparison, internal migration, in-migration and out-migration, is the movement of people within the international borders of a given country but across some form of administrative internal boundary. However, this is where the simplicity of migration concepts stop and multiple issues arise.

The main issue within the study of migration is that there is no set definition of what constitutes a migration and there are several broad groups in which differences derive. The main focus within this thesis refers to the internal migration of students within the UK, although in the Chapter 3 international student migration into the UK is measured and therefore a set definition and understanding of both concepts are required.

2.2.1 Who is a migrant?

Some problems and differences in migration studies are arising due to the concept of time. Questions and dissimilarities arise when one is distinguishing between a permanent or temporary migration. How long a time period must a person be in a different place to be defined as a permanent migrant? The answer differs depending on the country being investigated and the data source being used for the analysis (Bell et al. 2002; Bell 2004).

Despite there being no set definition on the length of stay that constitutes a migration, the United Nations advised the following two definitions which distinguish between long-term and sort-term migrants when referring to international migration (United Nations 1998):

> *Long-term migrant: "A person who moves to a country other than that of his or her usual residence for a period of at least a year (12 months), so that the country of destination effectively becomes his or her new country of usual residence. From the perspective of the country of departure the person will be a long-term emigrant and from that of the country of arrival the person will be a long-term immigrant".*
>
> *Short-term migrant: "A person who moves to a country other than that of his or her usual residence for a period of at least 3 months but less than a year (12 months) except in cases where the movement to that country is for purposes of recreation, holiday, visits to friends and relatives, business, medical treatment or religious pilgrimage. For purposes of international migration statistics, the country of usual residence of short-term migrants is considered to be the country of destination during the period they spend in it."*

However, it is common for definitions of permanent migration to differ from the advised United Nations definition. In the Australian context permanent migration is classified as a change in the person's usual residence for a period of at least six months, while temporary migration being for a period of less than six months (Bell and Ward 2000; Rowland 2003; Bell 2004). Therefore, what duration of move is of interest to the researcher? Is one interested in a migration if it is only temporary, does the migrant make significant impacts on their place of temporary residence to warrant the migration classification? Rees (1977) defined a migration:

> "As a permanent change of usual residence and therefore does not define temporary moves as a migration",

But no classification of what constituted permanent in Rees (1977) definition was made. A more detailed definition by Rees et al. (2009) stated that (p.64):

> *"Migration is the event of transfer from one residential location to another by a person who is termed a migrant. In this context, an event is an activity that takes place over a short period of time and a transfer involves travel over some distance from one location to another."*

This second definition raises another issue. What constitutes a migration in terms of the distance one must relocate to be classified as a migrant? Should an individual be classified as a migrant if they migrate only a short distance, say within a suburb or within a city? A common assumption of population mobility analysis is that it is only the relatively long distance moves which have a significant and disruptive effect on an individual's life that should be classified as a migration (Long 1988; Rowland 2003).

Short distance or local moves, are viewed as only affecting daily habits (Long 1988) and subsequently short distance moves are often not classified as a migration by many scholars, yet there is no set definition of what distance constitutes a short or long distance move. In the Rees et al. (2009) definition above, distance is mentioned but no set size of distance was stated as to what constitutes a migration. In the definition of migration proposed by Lee (1966), he defines migration as:

> "A *permanent or semi-permanent change of resident of any distance, even if that move is only a few meters*"

Therefore, it is clear that there is no definitive precedent set in the literature as to what distance constitutes a migration.

In contrast to these arguments, Boyle (2009) explains that the definition of migration used in human geography is often influenced more by the data resources available and the capability of defining migration from those resources, as opposed to theoretically guided principles. As a result a migration is often recorded if an administrative boundary has been crossed rather than a certain distance being travelled.

Another important distinction to be made refers to the way in which migration data has been recorded. The two most common forms of data measurement capture migration as either a transition or as an event (Bell et al. 2002). Transition data identify migrants by comparing their place of usual residence at the time of enumeration with that at a specified earlier date. If the usual residence is different between the two time points then a migration has occurred within the enumeration period. However, transition data have several limitations, the most serious being the failure to capture multiple and return moves within the enumeration period. Another is that transition data miss migrants who are born or die within the measurement period.

In contrast to transition data are event data. Event data attempt to record every move that was made by an individual and each move recorded as a single migration event. Therefore, event data should, in practice, include multiple and return migrations as well as moves by new-borns and those immediately before death.

Transition and event data are associated with how migrants and migrations are recorded. Transition data are associated with migrant stock data. Migrant stocks refer to the number of migrants in a given area at a certain time point. In contrast event data are associated with migrant flow data. Migrant flows report the number of migrations from one area to another within a given time period, therefore the number of migration events between place *A* and place *B* between time *t* and time *t+n*.

So in this thesis how is migration defined in light of the issues mentioned above? The study involves both internal and international migration and therefore a migration is recorded when an administrative boundary is crossed. In the case of this analysis the administrative boundaries consist of the international border of the UK, counties and local authorities (See Section 1.2). This was driven by the available data consistent with the issues discussed previously by Boyle (2009), and as a result there is no set distance of what constitutes a migrant in this analysis, it just requires that a boundary has been crossed. The data in this analysis consist of transition data and migrants are recorded as stocks, however this will be explained in more detail in Chapter 3.

2.2.2 Migration Literature and Theories

The concept of the migration process sums up a complex set of factors and interactions that all influence a migration decision. Migration is a process which affects every dimension of social existence and which develops its own complex dynamics (Castles and Miller 2009). Research on migration is therefore intrinsically interdisciplinary: sociology, political science, history, economics, geography, demography, psychology, cultural studies and law are all relevant (Brettell and Hollifield 2007; Castles and Miller 2009). Each of these disciplines have their place in understanding the complexity of migration and as interest in migration has grown in recent years this has led to the proliferation and interaction of theoretical approaches to understanding the concepts of migration. A detailed review of migration theory is not necessary here (See: (Massey et al. (1993); Arango (2000); Brettell and Hollifield (2007)), however some basic concepts and their links to student migration will be discussed below.

Lee (1966) developed on Ravenstein's (1885, 1889) laws of migration to create the push-pull framework of migration. Lee (1966) stated that a migration decision was ultimately influenced by 4 main factors; (1) origin factors [push factors] (2) destination factors [pull factors] (3) intervening obstacles between the origin and destination and (4) personal characteristics or Attributes. Therefore, an individual makes a migration decision by deciding whether (after considering the four main factors) migrating would prove beneficial. A large number of theories related to migration are borrowed from other disciplines and focus on the economic and labour market factors that instigate migration (Massey et al. 1993). However, as discussed by King (2002) the importance of new migratory circumstances that involve movements for non-economic or partly economic reasons need further consideration. Courgeau (1985) and King (2002) state that migrations linked to the life course such as student and retiree migrations both have potential for further investigation and future expansion in the area of migration theory. While Van Mol and Timmerman (2014) discuss that one of the groups of intra-European migrants that have long been ignored in migration studies is international students. Due to the rise in the number of mobile students in Europe, student mobility should now be considered as an integral part of the 'new map of European migration' as proposed by King (2002).

The major conceptual and methodological issues that are commonly experienced in the study of migration have been outlined above. The majority of the migration concepts and theories focus on factors other than students and rarely take into account the factors influencing the student migration process. This thesis focuses on the migration of students and therefore it is necessary now to build upon the basic migration terminology and concepts to specify how the above issues directly relate to the concept of student migration and as a result in the following section the idea that student migration requires a separate conceptual base to migration in general will be proposed.

2.3 Student Migration: Conceptual Issues and Previous Research

There is a growing base of literature on HE related migration and in particular on the migration of HE students. Discussions on the emerging geographies of education have highlighted the importance of students as agents of change in various geographical contexts (Smith and Holt 2007; Smith 2009; Holton and Riley 2013; Smith and Jons 2015). However, despite the growth in research on tertiary education, surprisingly to date, comprehensive analyses of student related migration in the UK have been limited [see Duke-Williams (2009)]. As a result there are inconsistencies in key conceptual issues.

As explained in Section 2.2, there are several issues with regards to defining and measuring migration and a migrant. This is no different with regards to defining and measuring who is a student migrant and what constitutes student migration. It can be argued these issues are even more pronounced when defining and measuring student migration. Aspects that are particularly disjointed include the definition of a migration involving HE students, whether HE mobility can be classified as a temporary or a permanent migration, the impact that the intention of the student can have on the type of migration experienced and the function that short distance moves have in student residential mobility.

Participation in HE has expanded substantiality in the UK over the past half-century (Chowdry et al. 2013) with around 2.6 million students registered as attending a HEI in the 2010/11 academic year (Higher Education Statistics Agency 2012b). This expansion in student numbers was primarily driven by government policy and the major restructuring of HE in the UK in the early 1990s, which saw the emergence of the 'post 1992' University – those former polytechnics and colleges given university status through the 1992 Further and Higher Education Act (HMSO 1992; Christie 2007; Holton and Riley 2013). These changes were implemented under the premise of building a workforce capable of sustaining the shift towards a knowledge economy, which saw incentives introduced to increase the number of school leavers entering HE to 50 per cent (Munro et al. 2009). The expansion in HE was also underpinned by the desire to bring a more diverse set of 'non-traditional students'[3] into universities as a means to resolve problems of social exclusion and poverty within the UK (Christie 2007). This strategy was born out of Tony Blair's 'education, education, education' speech at his party's labour conference of 1996 (Holton and Riley 2013). There was a particular policy drive to improve the representation of previously underrepresented groups in the student population, such as people from lower socio-economic backgrounds and ethnic minority groups (Chowdry et al. 2008). The desired expansion of non-traditional student populations in HE remains a major policy issue within the UK today. The 'Widening Participation Policy' (HEFCE 2013) states that anyone with the ability and who wants to go to university should have the chance to do so, whatever their economic or social background.

2.3.1 Conceptual Issues: Who is a 'student migrant'?

With this expansion in HE participation and the changing dynamics of the student population, the construction of the term 'student' is changing. The previously accepted one-dimensional definition of the 'student' is being replaced with a range of possible definitions that reflect much greater diversity and uncertainty within this population group (Ozga and Sukhnandan 1998; Leathwood and O'Connell 2003; Morley 2003).

In recent years a significant shift away from the traditional notion of a student being a young white man from an upper-class or middle-class background studying as far away from the parental home possible has been observed

[3] Non-traditional students refer to students who would not, in previous generations, have been expected to attend university. These included students brought up in working-class families, ethnic minorities and mature students.

(Leathwood and O'Connell 2003). This has been the result, in part, of campaigns to further the participation of previously excluded groups, not least of which has resulted in the dramatic increase in the proportion of women studying in HE (Leathwood and O'Connell 2003). Meanwhile there has also been an emerging discussion of whether it is still common practice for students to study away from home. Ideologies have been shifting over time, with ever increasing numbers of students residing locally (Holdsworth 2006; Christie 2007; Holdsworth 2009; Holton and Riley 2013).

With this changing identity of the 'student' in HE it is crucial that institutions and researchers alike move away from the one-dimensional view of the industry and the students that attend. This has seen the construction of the concept of 'the new student' that has very different experiences of HE and spend much less time on campus (Leese 2010).

One of the major results of this shift away from the traditional notion of what constitutes being a 'student' is the difficulty in having one definition of a 'student' and therefore a 'student migrant' or 'student migration'. An important feature in the analysis and understanding of HE related student migration is being able to clearly define what constitutes 'student migration'. Therefore this section now attempts to state how these 'student' related terms are defined in this thesis.

In the UK, the academic year traditionally starts around late September or early October and finishes around mid-June. While each HEI will have slight variations, it is common that the academic year is split into 2 semesters and includes 3 terms; autumn, spring and summer. In the most commonly adopted format, a university term consists of around 10 or 11 weeks in which the students are expected to attend contact teaching hours throughout the week. Therefore, it is expected that a student be present at university for around 30 weeks per year.

This raises the question as to where the student should be recorded as living and this can vary between individual students as a result of their actions when not attending university teaching hours. For example, a student (A) may return home each weekend during term-time and only stay in their term-time accommodation for the minimum of time possible. Therefore, it could be argued that student (A) should not be classed as a migrant to the place they study as they spend very little time there. In contrast, student (B) migrates to their location of study and stays even out of term-time. Therefore, it is clear that student (B) had migrated to their place of study and should be recorded so accordingly. However, previous research has shown that the amount of time spent at the term-time address and the location of that term-time address is greatly influenced by many factors (Christie 2007; Faggian and McCann 2009; Holdsworth 2009; McNay 2012; Cotton et al. 2013; Khambhaita and Bhopal 2013) and making a generalisation for all students will mask these differences.

In terms of standard migration terminology, it is argued that a person should be registered as living in the location where they spend the majority of their time. Therefore, if the student lives in the same geographical area of the HEI for all of the 30 teaching and examination weeks there place of usual residence should technically be recorded using their term-time address as they spend more than half of the weeks in a year in that location. This links back to the concept of place of usual residence as used by the official statistics offices. This is defined as;

> "The housing unit or collective living quarters at which the person usually lives, i.e. sleeps, keeps his/her clothes and other belongings etc. It is the residence from which a person generally goes to work or if a student, attends school/college/university" (Office for National Statistics 2009a).

Therefore, from the ONS definition of usual residence, it suggests a student should be recorded as living at the place they attend a HEI. So if this is a different location to where they lived before then they should be classed as a student migrant. So do the official UK internal migration estimates record a student migration in this way?

At present, there is no compulsory system to record movements of the population within the UK. To estimate this information a combination of three administrative data sources are used; namely the Patient Register Data Service (PRDS), the National Health Service Central Register (NHSCR) and the Higher Education Statistics Agency (HESA) (Office for National Statistics 2011c; Raymer et al. 2012). A person is recorded as migrating in the official UK internal migration estimates when they change their place of usual residence, defined as above, and when this change of residence is detected in one of the three aforementioned data sources. Measuring internal migration in this fashion is not perfect as the data sources were not designed for this purpose and as a result there is often miss reporting of the amount of internal migration occurring in the UK. For a more detailed discussion on how migration statistics are recorded and measured in the UK see Lomax et al. (2011) and Raymer et al. (2012).

As a result, students in the UK are recorded as making a migration in official internal migration estimates if their term-time address was in a different geographical area to their domicile. Although it must be considered, that due to the way internal migration is measured in the UK, as mentioned above, for this move to appear in the official statistics then the student had to have re-registered with their GP in the location of their term-time address or provided adequate information to HESA for their migration event to be recorded.

Therefore, in the official statistics, if students change their location for term-time purposes and register this move then they are classified as an internal migrant regardless of the amount of time they spend at the term-time address or whether this was their place of usual residence. As a result, in order to seek consistency with the official national statistics office, in terms of this research a student is also classified as making a migration if their term-time address is in a different administrative geographical area than their domicile address.

Another argument in the migration literature is the distinction between permanent and temporary migration. How long must a person be in one place to be classified as a permanent migrant and should temporary migration be classified as a migration at all? These arguments are clearly an issue in the study of student migration. This relates back to the distinction between migration and residential mobility (Cadwallader 1992), where it is argued that a migration should not only imply a move that involves a change in location but also probably a change in employment (or educational institution) and a departure from previous social groups (Dennett 2010).

It could be argued that all student migration is actually just student movement as the movement made by students is only temporary and that students make very few permanent ties to the area in which they reside during term-time. Again, a student may be technically recorded as a migrant but only stay for the 10 week term period and then return to their domicile. Does this mean this type of student has made a departure from their previous social groups? Does this type of student make any permanent ties or impact on the area in which they reside for term-time? In contrast, a student may migrate every academic year and only return home for the summer months, or not at all. So should this student be classified as a permanent migrant, have they made permanent ties with the area they have migrated to?

This debate was also visible with regards to international student migrants and whether intentional students should be included in the net migration figures for a country. HMSO (2012) discussed how they aimed to introduce an annual limit on the number of non-EU economic migrants admitted to the UK. It has been stressed that the UK Government includes overseas students in the policy and this has been raised as a major concern by the HE sector. However, the OECD (2013) does not include overseas students in the net migration figure until they exceed 36 months in a country. This contradiction in definitions between large global organisations shows the discontinuity in the student migration terminology. Throughout this research, if an international student attended a HEI in the UK, this was classified as a permanent migration without considering the length of stay.

On reflection, it is hard to generalise for all students what should constitute a migration. If the student makes a large impact on the area in which they reside during term-time then they should be considered as resident in this location and if this is a different location to their domicile then they should be classified as a student migrant. So this raises the question on how to classify a student that makes an impact on a new area but only during term-time. Local authorities will argue that if a student is using services such as housing, public transport, medical facilities such as the GP and dentist and are spending money in the local economy then these students should have their usual residence registered as their term-time address.

2.3.2 Previous Research on Student Migration

There has been a growing amount of research into the patterns and issues around student migration in recent decades which has been associated with the increase in student numbers in the UK. Research became increasing interested in both the student movements as well as the implications of burgeoning numbers of students on university towns and cities (Holton and Riley 2013).

Duke-Williams (2009) produced one of the more detailed analyses of student movements to date. The study used 2001 census data to examine the migration flows associated with areas with high concentrations of students. Holdsworth (2009) used UCAS (University and Colleges Admissions Service) admissions data to analyse the pattern of moving away from the parental home in order to pursue university education. These studies found that a large amount of population turnover in small areas was associated with student movement and that the levels of in and out migration were not evenly distributed across the country.

These studies on student migration within the UK triggered further studies on the impact of these students on the areas in which they reside. Students have been found to be a highly mobile part of society and have an evident impact on the locations they migrate to as a consequence of their expressive lifestyles and consumption practices (Chatterton 1999, 2000). Students ensure the viability of some retail businesses and help to increase the range of goods, services and attractions available to the community. The average annual spend by a full time English domiciled students on living costs excluding rent in 2010 was £6,496, which, on average, breaks down into the following expenditure categories: £1,724 on food, £1,828 on personal items (e.g. clothes, toiletries, mobile phones, CDs, magazines and cigarettes), £1,154 on entertainment, £1,343 on travel, £279 on household goods. This equated to a total expenditure of around £7.9 billion a year (Universities UK 2010). Students also contribute to the local community in many ways such as volunteering to help older people, young children and people with disabilities, while universities provide cultural activities for the community such as art galleries, concerts and theatres.

During the period 2003–04 to 2011–12, the total number of students enrolled at HEIs in the UK increased by almost 300,000, or 13.5% (Universities UK 2013). With this significant increase in student numbers attending universities it was not surprising that the student population increased faster than the HEIs ability to accommodate them (Hubbard 2009). There was therefore, a large unmet need of bed space in university settlements for these students to reside in close proximity to the HEI they were attending. This triggered a reliance on the private rental sector to accommodate a large number of students. A survey in 2007 suggested that 46% of all students in the UK lived in private homes of multiple occupation (HMO) (Savills Research 2007; Hubbard 2009). As a result of this rapidly growing student population and the number of HMOs in close proximity to university campuses, the term 'studentification' was derived to encapsulate the growing concentration of students in student areas/distinct enclaves of university settlements and how this triggered significant levels of urban change (Smith 2002, 2005; Smith and Holt 2007; Smith 2008; Hubbard 2009; Smith 2009).

Smith (2005: 73) introduces the conceptual meaning of studentification as:

> 'the distinct social, cultural, economic and physical transformations within university towns, which are associated with the seasonal in-migration of higher education students', with the transformation of properties from single-family properties to HMOs for higher educational students.

As a result this triggered the replacement of permanent or semi-permanent groups of residents within an area with temporary student groups. This results in a distinctive change in the local class and household structure, with a distinct change in services and the creation of 'student ghettos' (Hubbard 2008; Kinton et al. 2014).

An example of the impact studentification can have on an area was provided by Harris and McVeigh (2002) as read in Smith (2005: 74):

> Pubs have been converted into theme bars, which often shut during summer months when students have returned to their homes. Fast-food takeaways and off-licences selling cheap alcohol dominate the shopping streets. Schools have seen their class sizes plummet as families move out of the area … House prices have also rocketed as landlords have created a property boom and now people wishing to move house but stay in the area have found themselves priced out the market

It is therefore clear that in some university towns and cities student in-migration has a profound impact on the area. However, how long can such rapid student growth and large scale studentifications continue for?

In a recent commentary, Chatterton (2010) suggested that the student population had probably reached a historical peak and plateau due to the financial constraints of government spending cuts, the increase in tuition fees and the impact of the 2008-09 recession. It was suggested that the process of studentification may also go into reverse as a result of this plateau in student numbers, the changing migration transitions students are undertaking and the increase in choice of student housing available to student migrants.

This leads onto a new term first developed by Smith (2008) to describe the aftermath of studentification – Destudentification:

> "The reduction of a student population in a neighbourhood which leads to social (for example, population loss), cultural (for example, closure of retail and other services), economic (for example, devaluation of property prices) and physical (for example, abandonment of housing) decline" (Smith 2008: 2552).

However, to understand the process of destudentification further, a more in-depth analysis of the causes, conditions and catalysts is required (Kinton et al. 2014). After further analysis the work in progress by Kinton et al. (2014) suggests there are two main factors driving the process of destudentification. The first being the evolution in student lifestyle choices in which students are becoming far more demanding in terms of the quality of accommodation they wish to reside in – students no longer reside in substandard housing as they now have the option to reside in high quality dedicated student housing (Hubbard 2009; Chatterton 2010; Sage et al. 2012b; Kinton et al. 2014). The second is the oversupply of student accommodation in university towns and cities. Where large scale private projects to build purpose built student accommodation and the aftermath of large scale studentification has resulted in the supply of accommodation outweighing demand (Kinton et al. 2014). As a result, the areas of lower quality and poorer reputation are no longer finding tenants as students head into new halls of residence in the city centre which can lead to abandonment of student style HMOs in the student enclaves (Munro and Livingston 2011).

This previous research has shown that the impact of rising and declining numbers of students attending HEIs in the UK is extremely significant and the level of this impact changes depending on the migration transition experienced in order to attend a HEI. Studies have also suggested that there are signs that the student migration patterns seem to be changing over time. It is therefore imperative to have a clear understanding of all student movements and the factors that impact on them. An in-depth knowledge of these student migration transitions and the factors affecting them will be beneficial to the HEIs, local and national government, planning authorities, as well other entities, such as retailers and transport providers. Therefore, in this thesis any student who relocated to an area to attend a HEI is defined as a student migrant, in order to get the highest level of detail of the student migration taking place in the UK.

It must be noted that student migration is a unique type of migration in which standard concepts of migration may not hold and hence there is no clear, unambiguous definition of student migration that corresponds with the conventional definitions of migration. By classifying all student relocations, as student migrations, some trends may be missed as a result of the issues mentioned above. Despite this, due the perceived significant level of impact students have on student areas as discussed above it was deemed necessary to classify all students in a new term-time location as undertaking some form of student migration.

With the concepts of migration and student migration discussed in detail the focus now shifts towards the factors that impact upon the student migration decision process.

2.4 The Impact of Socio-Demographic Characteristics on Student Migration

A substantial body of previous research has examined a wide range of factors that are interlinked with the student migration process. The overriding themes of this pre-existing research focused on the differences and inequalities observed in attainment, HE participation and the migration and housing choices of students. In order to have a clear conceptual understanding of the student migration process, a substantive review of all the previous studies that have investigated the differences in the higher educational system is necessary.

A significant amount of previous research focused on the inequalities of access and participation in HE in the UK, and this remains a major policy issue to this day (Department for Education and Skills 2003, 2006; HEFCE 2013). Although this thesis does not analyse differences in those that participate in HE or not, it is of great policy interest to investigate whether these inequalities and differences were still visible within those students that do participate in HE by analysing inequalities in the migration transitions and distance travelled to attend HEIs.

It was highlighted earlier in this chapter how important an individual's level of attainment and achievement are in the student migration outcome. A longstanding body of literature across the developed world and the UK focuses on educational inequality and the relationship between the educational attainment of children at pre-higher educational levels and their social and ethnic origins (Galindo-Rueda et al. 2004; Goodman et al. 2010; Gregg and Macmillan 2010; Sullivan et al. 2011; Bukodi and Goldthorpe 2012; Jerrim 2012). The previous research shows that at lower levels of education, children are greatly influenced by the socio-economic and ethnic background of their parents, with those from low income families and ethnic minority groups faring the worst in overall educational attainment.

There have been persistent inequalities in HE participation rates among school leavers from poor neighbourhoods and those from rich neighbourhoods. Analysis based on data for the period 1994 to 2000 showed that young people living in the most advantaged 20% of areas are five to six times more likely to enter HE than those living in the least advantaged 20% of areas (HEFCE 2005). These patterns of inequality in participation in HE are rooted in divisions which emerge earlier in the education system and are related to issues such as parental background and school type, not only the student's social class (Halsey et al. 1997; Brantlinger 2003; Christie 2007).

In the book entitled 'Degrees of Choice' by Reay et al. (2005) it is claimed that choices in HE in regards to where and what to study are greatly impacted by an individual's social class, ethnicity and gender. The previous research analysing the impact of these three personal characteristics on the participation and migration process to HE are discussed in turn below.

Socio-economic Background

Recent statistics from UK Universities and Colleges Admissions Service (UCAS) show that around 25% of students accepted to university were from the four lowest socio-economic groups (Reay et al. 2010). This is compared to 35% of the total population being from these four lowest groups (Hill 2005), so this suggests participation of lower social backgrounds is not equal to those from higher social classes.

Machin and Vignoles (2004) investigated the links between HE and family background by analysing the experiences of two cohorts of individuals born in 1958 and 1970. The findings indicated that educational inequality increased between the two cohorts and that the expansion in HE during this period benefitted children from richer families rather than the most able. Blanden and Machin (2004) also investigated the links between family background and HE by studying the temporal shifts in participation and attainment across parental income groups for children going to university in the 1970s, 1980s and 1990s. Their key finding was similar to that of Machin and Vignoles (2004) in that they found that the HE expansion was not equally distributed across people from richer and poorer backgrounds.

Further research by Galindo-Rueda et al. (2004) investigated whether the socio-economic gap in HE participation had widened over time and if this gap emerged on entry to university or earlier in the education system. They did this in two ways, firstly by looking at samples of school leavers at different time periods and analysing how the likelihood of them going to university differed as a result of the socio-economic status of the student's neighbourhood. Secondly, they used more detailed individual level data; to model determinates of participation in HE, focusing on changes in the relationship between family background and participation over time. The main findings of the study indicated that actual growth in participation amongst poorer students had been remarkably high but the gap between the rich and poor widened during the 1990s. They did however indicate that much of the class difference in HE participation seems to reflect inequalities at earlier stages of the education system. Therefore, despite decades of policy designed to widen participation, it appears from the majority of research that social inequality within HE in the UK increased during the 1980s, 1990s and early 2000s. A recent and detailed piece of research into widening participation in HE was the study conducted by the Institute for Fiscal Studies (Chowdry et al. 2008, 2010, 2013), which used a unique individual-level administrative dataset that provided information on a particular cohort of state school pupils as they progressed through the education system. The report found that students from materially deprived backgrounds were much less likely to participate in HE at age 18 or 19 than students from less deprived backgrounds. They also found that the socio-economic differences observed in HE – including at high status institutions – arise as a result of substantial socio-economic differences in educational achievement earlier in life.

Another important aspect in discussions regarding fair and widening access to HE, is the levels of unfair and unequal selectivity seen from differing HEIs. Selectivity refers to what level of educational attainment is typically required to gain admission to a HEI. The UCAS tariff system (as explained in section 2.4.4) gives a summary measure of educational attainment used in order to gain entry into to HE, and an Office For Fair Access (OFFA) report (Harris 2010) classified institutions according to whether their entrants have, on average, higher tariff, medium tariff or lower tariff scores from their entry qualifications. Harris (2010:95) found that the relative participation of advantaged and disadvantaged young people in individual institutions varied widely and was associated with the tariff group that the institution was in. In lower tariff institutions disadvantaged young people typically have only slightly lower participation rates than advantaged young people. For some lower tariff institutions the participation rates of disadvantaged young people are higher than for advantaged young people, up to twice as high for a few institutions. However, it was found that disadvantaged young people were much less likely to enter higher tariff institutions than advantaged young people, in some cases as much as 15 times less likely. This clearly shows that individuals from less advantaged background are still significantly under-represented in those institutions demanding the highest entry requirements and therefore access within HE in the UK is still not equal (Harris 2010; OFFA 2014).

The situation of social-economic inequality in HE in the UK is well summarised by Field (2003:30):

> "Socio-economic inequality is startlingly persistent at every stage of the higher education system. Although the flows into the system from all classes have increased, they have increased unequally for different groups of people … Rather than widening participation the growth over the past two decades has increased inequality. The new students have come from the transformation of higher education into a mass experience of the middle class".

It is, therefore, of great interest to see if these inequalities with regards to participation, as observed in these previous studies, are present within the student migration outcomes.

It has been found that differences associated with social background are present within the HE sector beyond the issue of participation. Students from disadvantaged social background groups show differentials in the process of leaving the parental home to attend HEIs (Kerckhoff and Macrae 1992; Belfield and Morris 1999; Jones 2002). For example, research shows that working-class students more often attend local post-1992 university institutions that offer relatively low entry requirements, reducing the financial implications of moving away and providing culturally and geographically familiar learning environments (Ball et al. 2000; Reay et al. 2001; Clayton et al. 2009). Patiniotis and Holdsworth (2005) conducted a qualitative study to analyse why recent trends have shown that more students were choosing to stay at home and not migrate in order to attend a HEIs. The authors found that the decision to migrate was strongly linked to socio-economic class with an association between those choosing not to migrate being from poorer backgrounds, with these students stating they chose to stay home for 'financial reasons'. Holdsworth (2009) analysed admissions data to HE and found a trend towards more localised study, while the data also indicated that those that still chose to migrate and the process of student migration had become an elite practice mostly undertaken by those from richer and higher socio-economic backgrounds. These findings were also supported in research conducted by Christie (2007) which found marked differences in participation and student migration trends between those in the most advantaged areas of the UK and those from the least advantaged areas.

These previous pieces of research clearly show that an individual's social class impacts on the student migration process experienced and therefore must be a clear and distinct component in the student migration analysis conducted within this thesis.

Ethnicity

There is now a considerable body of research on the issues of minority ethnic groups and their access and participation in HE. There is however, much less work conducted on minority ethnic students experiences of HE and almost no work on how minority ethnic students chose the HEI they attend and the student migration transition they experience (Ball et al. 2002).

Previous research has indicated that there are substantial differences in participation rates in HE across the different ethnic groups (Modood and Shiner 1994; Dearing 1997; Tomlinson 2001; Khambhaita and Bhopal 2013). Chowdry et al. (2008) found that ethnic minority students were significantly more likely to participate in HE than their White British peers, while Ball et al. (2002) argued that the differences amongst ethnic minority students cannot be fully understood without reference to their social class background, which as discussed before, has strong associations with the student migration process.

It is well established that candidates from black and minority ethnic groups go to university in good numbers, but we also know that candidates from some minority groups tend to be concentrated in less prestigious institutions (Noden et al. 2014). Shiner and Modood (2002) and Chowdry et al. (2008) both found that there were large institutional bias with regards to ethnicity and that there were large socio-economic and ethnic gaps in the likelihood of attending high status HEIs within the UK. Access to high status institutions is important for several reasons, not least because it is likely to affect candidates' subsequent destinations and their ability to access elite professions (Noden et al. 2014).

There have also been several pieces of research that investigated the relationship between ethnicity and the transition into university. Khambhaita and Bhopal (2013) investigated the link between ethnicity and term-time accommodation status, this therefore has a direct link to the student migration process. Khambhaita and Bhopal (2013) found that female students from Indian, Pakistani and Bangladeshi ethnic groups were all more likely to stay at home and not migrate for HE relative to their white counterparts. These findings supported those of Faggian et al. (2006) who stated that all non-white UK students were much less likely to migrate to attend university compared to their white peers, while McNay (2012) found that black and minority students were more likely to study closer to home than white students.

Further work by Faggian et al. (2007b) modelled the decision to migrate for university and the subsequent decision to migrate for employment. Their work confirmed previous findings that those with more human capital and from higher socio-economic groups were more likely to migrate, while the study also indicated that these patterns differed by gender and ethnicity. The work conducted by Finney and Simpson (2008b), Simpson and Finney (2009) and Finney (2011) must also be noted as they identified ethnic differences in migration flows for the population as a whole and identified student migration trends within their studies. However, a focus solely on ethnic differences within student migration patterns was lacking.

Gender

There is also a substantial and growing body of literature that analysed the gender and ethnic differences in the students' experiences and attainment in HE in the UK and many other countries across the world. The reversal of gender inequalities has been found to be well established in OECD (Organisation for Economic Co-operations and Development) countries including the UK. More women than men are now entering HE, irrespective of age or type of HE except at the doctoral level although current trends suggest that women will outnumber males at this level within a few years (Vincent-Lancrin 2008). In terms of attainment, the gender gap has grown even further in favour of women. In 2005, OECD countries awarded 57% of their degrees on average to women and this is expected to increase in future years (Vincent-Lancrin 2008). A study by Cotton et al. (2013) also found a clear gender gap within UK HEIs but also investigated if an ethnic attainment gap existed. Cotton et al. (2013) found that there was a clear ethnic attainment gap within UK HEIs with ethnic minority students having lower completion rates to their white counterparts. Therefore, with these gender and ethnic gaps already noted in previous research with regards to attainment and participation, the current study will build on this by investigating if these gaps are present with regards to the migration transitions experienced by students and the distances migrated in order to study at UK HEIs.

Despite the value of the previous research reviewed above being unquestioned, there is a sense that a focus on the actual patterns of student migration and the different student migration processes as a result of the student's socio-economic, ethnic or gender group has been lacking. These previous studies were also largely qualitative and survey based, and often focused on specific case study areas. In contrast, this thesis aims to supplement this research by providing a solely quantitative analysis of the whole UK student population.

The discussions in the preceding paragraphs of this chapter have highlighted the vast amounts of prior research in topics related to the study under investigation in this thesis. Despite this vast amount of prior research, a truly quantitative analysis of socio-economic, ethnic and gender differences on the specific migration transitions experienced by students is missing. The current thesis therefore builds upon this prior research by applying solely quantitative techniques on population data that has not previously been conducted. As a result, this thesis supports and critiques the findings of the previous research by using differing techniques and data sources and subsequently comparing there results to the findings produced herein.

2.5 Conceptualisation of the student migration decision process

Throughout the current chapter, overviews of the concepts that underpin the student migration decision process have been presented. The final section of this chapter aims to summarise these theories, influencing factors and concepts into a clear and succinct conceptual framework. In order to conceptualise the student migration decision process in a comprehensive but succinct manner, it is necessary to review the factors that may influence this decision and how they are all interlinked within a conceptual framework.

At this stage it must again be stressed that the outcome in question in this thesis is the student migration decision process and future labour market outcomes for those individuals. Those students who decided not the study in HE are therefore not subject to the student migration process and are not of interest within this thesis. However, it must be noted that the factors that impact on the student migration decision process may be similar to the factors that impact on the decision whether to participate in HE or not. Therefore, although the conceptual framework does not include the student decision on whether or not to participate in HE, as this decision is integral to the whole process of student migration, this conceptual framework could also be applied to the higher educational participation decision.

The conceptual framework proposed within this section is illustrated in the conceptual framework diagram Figure 2-2. The conceptual framework is presented as a 'cone' shaped decision diagram with influencing factors presented in four groups, that range from domicile macro factors to factors relating to individual choice. The concept of the 'cone' diagram represents the level of control the individual has on the factors within the student migration decision process. The concept is that each individual student migration decision will be influence by the factors represented in the diagram. Factors higher up the 'cone' diagram are influencing factors in which individuals have very little control over (macro factors), while the level of the individuals control on the factors affecting the student migration decision increase the lower down the 'cone' diagram you progress.

The framework includes all the factors that have been identified and deemed by the author to impact on the student migration decision process. Even though a large number of these factors are not explicitly analysed in this thesis, it was viewed necessary to include them within the conceptual framework of the study to highlight the potential biases these non-observable factors could have on results presented later in the thesis. Non-observed factors are especially important when evaluating the results presented in Chapters 4 and 5, as non-observable factors can create bias in the results and also result in the underlying assumptions of the methodologies being violated. The variables that are explicitly included within the analysis presented later in this thesis are shaded green in Figure 2-2.

A number of different academic disciplines, including economics, geography, sociology and demography, amongst others, have contributed to research on student migration. Yet when analysing why students migrate, it appears that answers are nearly always based on models that point out various factors, at different levels, that are seen as encouraging or impeding migration. These models are often referred to as push-and-pull model factors or drivers and barriers to migration (Mazzarol and Soutar 2002; Li and Bray 2007; Rodriguez Gonzalez and Mesanza 2011; Carlson 2013). The majority of the previous work focuses on modelling international student migration and although the focus of this thesis is primarily on internal student migration, a large amount of insight can be gained from analysing the previous studies.

In a paper evaluating existing theorisations on the student mobility, Findlay (2011) sees the previously outlined models as 'demand-side' theorisations and criticises them for using 'simple behavioural models of the choices made by students' without 'recognizing the importance of the cultural, social and economic contexts within which "decisions" are taken' (Findlay 2011:164-165). It is therefore imperative that the conceptualisation of the student migration decision proposed in this section takes into account the complexity of student migration that Findlay (2011) observed in his research.

The conceptualisation of the student migration process presented below takes aspects of several pieces of previous work and adapts their concepts to match those put forward in this thesis. Firstly, as previously discussed in Section 2.2.2, the theory of migration proposed by Lee (1966) in which he introduced the push-pull framework of migration, he suggested that each individual migration decision will involve the individual weighing up the positives and negatives of making a migration (or not). In the case of this research we build upon the framework suggested by Lee (1966) by adapting the concepts to be specific to the migration decision process into HE, in which interlinking factors affecting a student's migration decision are conceptually visualised.

However, the choices individuals make with regards to HE are far more complex than simply weighing up positives and negatives or push and pull factors of migrating or not migrating. Reay et al. (2005) state that the area of HE choice is both under-researched and under-theorised and they draw on Bourdieu's concepts of habitus, field and cultural capital to conceptualise the complex choices at play in the area of HE (Bourdieu 1977, 1984, 1985, 1986). These complexities are constructed into a clear student-choice model put forward by Perna (2006). This student-choice model is grounded in the economic theory of human-capital and the sociologists' constructs of habitus and social and cultural capital. This model views students HE decisions as being shaped by 'layers of context' such as a student's habitus, school and community context and the HE context (Perna 2006:116). A clear and concise description of these 'layers of context' is provided by Salisbury et al. (2010:617-618) and each layer is summarised below.

Human capital refers to the productive capacities (knowledge, skills and abilities) possessed by an individual. An investment in HE, as a form of human capital, can enhance an individual's productive capacities as well as their future earnings and occupational status (Becker 1993; Paulsen 2001).

Financial capital represents an individual's actual or perceived financial resources, such as income or financial support. This financial capital is especially important in the aspect of HE decision-making when students compare the monetary costs and benefits of participation and relocating (Paulsen 2001; Perna and Titus 2005).

Social capital describes the access to networks, support systems and information resources that might inform or constrain the range and type of options, procedures and opportunities available to migrate in order to participate in HE (Coleman 1988; Massey et al. 2003; Perna and Titus 2005)

Cultural capital refers to class-based cultural knowledge, norms, activities, skills, and values - typically derived from one's parents - such as those related to the acquisition of educational credentials and occupational status (Bourdieu and Passerson 1977; McDonough 1997; Massey et al. 2003; Perna 2006)

Finally, each of these four forms of capital is consistently related to, and clearly shaped by, each individual's habitus. Habitus refers to an enduring, social-class-based set of beliefs, values, perceptions, attitudes, and aspirations an individual acquires through their early home, community and school environments that serve to frame and constrain the choices they make in their life (Bourdieu and Passerson 1977).

Recent research has demonstrated that indicators of habitus and measures of each of these four forms of capital influence HE student decision-making and are situated with a series of overlapping educational, familial and societal contexts (Perna 2000; Paulsen and St John 2002; Perna and Titus 2005; Perna 2006)

The remainder of this chapter will propose a conceptualisation of the student migration process that considers Lee's (1966) push-pull framework, the complexity of habitus, field and social capital in the individual's choices in entering HE (Reay et al. 2005) and all factors that have been identified in the previous research presented in this chapter thus far.

The proposed conceptual framework of the student migration decision process has been visually depicted in a 'cone' diagram (Figure 2-2) that shows a simplification of factors impacting on the student migration decision process. In the proposed conceptual framework the determinants differ from Lee's (1966) model, in the fact that the origin and destination effects are not explicitly illustrated. This is due to the fact that the origin and destination factors in the student migration concept are better illustrated in terms of domicile and institution effects. Lee (1966) also stated that it is not so much the actual factors at origin and destination that results in migration but personal sensitivities and awareness of conditions elsewhere that also enter the equation, which are strongly linked with individual characteristics rather than the origin and destination effects themselves. The intervening obstacles and individual characteristics (micro-level factors) are explicitly expressed in the conceptual framework shown herewith. The interlinking factors influencing the student migration decision process of a student into a HEI were split into four main categories: Individual (micro) factors, Domicile (macro) factors, Individual Choice and Intervening Obstacles. As shown in Figure 2-2, there were many different variables within these four categories that all contribute in some fashion towards the student migration decision. As mentioned earlier, the perceived level of individual control on these factors varies as one progress through the conceptual diagram, and all of these factors are explained in turn below.

The first of the categories depicted in Figure 2-2 are those on the domicile (macro) factors. This category is included in the conceptual framework due to the links between these variables and the individual and individual choice factors that appear lower down in the framework. The geographic area in which a student was brought up and went to school can influence the future migration outcome in several ways. The area where an individual was brought up may impact on the quality of schooling available to that individual, for example, if the individual was brought up in an inner city area with high levels of deprivation then it is highly likely that the quality of schooling available in that area was not of a high standard. This can have a direct link on the individuals' future achievement and therefore their future higher educational experience. Other domicile factors that will impact on the student migration decision will be the proximity to a HEI. If a student is from a domicile location that is not close to a HEI then if a student wants to study at a HEI there only option would be to migrate, commute or distance learn. This category is placed at the top of the decision diagram as it is perceived that the individual student has very little control over these macro level factors, but nevertheless, these factors will likely impact on the student migration decision in some fashion. These macro level factors are often a construct of the area in which the individual grew up and was educated, all of which were likely to have been controlled and influenced by their parents choices and location of employment as opposed to a direct decision made by the individual student themselves.

The next group category in the conceptual framework is the major grouping of variables referring to individual (micro) level factors. This group of factors are more closely linked to the individual themselves, but are located towards the top of the diagram as again the level of individual control over these factors is low. Individuals have no control over there age, ethnicity and gender, while it can also be argued that socio-economic status is very much set by the individual's parents, although this often depends how this variable is measured within the dataset. This group of determining factors, although mostly out of the individual's control, will have many interlinking facets across the student migration decision process. These individual factors have a direct relationship with the decision to participate in HE and the student migration outcome, as well as being interlinked to all of the other factors within the conceptual framework.

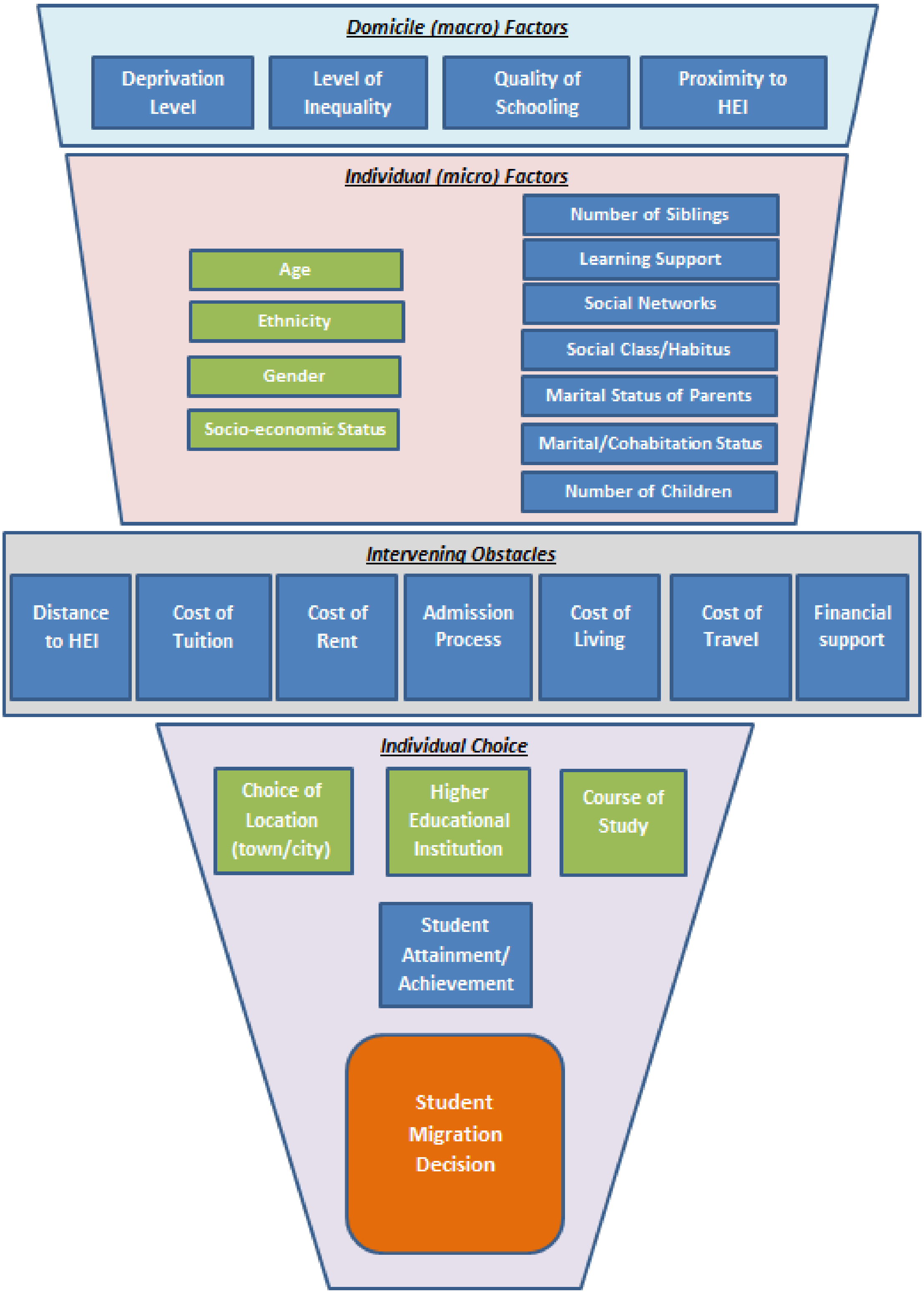
The first of the micro level factors in the framework to be discussed refers to the age of the individual. Previous studies have related the impact of age (Rogers et al. 1977; Rogers and Castro 1981b; Wilson 2010) and an individual's stage in the lifecourse (Courgeau 1985) on the propensity to migrate. As previously mentioned, internal migration statistics for the UK suggest that the peaks in migration numbers at ages 19 and 22 were highly associated with migration to and from HEIs (Office for National Statistics 2013a). The number of students aged under 30 increased by 388,000 between

2003–04 and 2011–12, while the number aged 30 and over decreased by 79,000, while in 2011-12 students under the age of 30 made up 73.7% of the total student population (Universities UK 2013; Higher Education Statistics Agency 2014b). Age is highly correlated with many of the other factors impacting on the student migration decision. Many studies in demography link an individual's age to their levels of fertility and marital/cohabiting status. While an individual's age is also likely to impact on factors within the intervening obstacles category which in turn will impact on the migration decision process. It is therefore clear that an individual's age is likely to impact on whether they decide to participate in HE and if they do what type of institution they will attend and the student migration decision they make.

Ethnicity is another extremely important individual level factor that will impact on the student migration process. As discussed previously in Section 2.4, previous research has indicated that an individual's ethnicity has been found to impact directly on participation rates in HE and the migration decision an individual makes if they do participate in HE (Finney and Simpson 2008a; Simpson and Finney 2009; Finney 2011; Smith and Jons 2015). The choice of HEI is also seen to be directly influenced by ethnicity with certain institutions still having very low participation rates of ethnic minority groups.

Gender has also been seen to impact on certain inequalities in HE and therefore will impact on the student migration process. Certain courses have high proportions of one sex, for example, engineering courses have a high male dominance whereas Education degree have a female dominance (Universities UK 2013). Therefore, an individual's gender is likely to impact on the courses they apply for and resultantly this will impact on the institution they attend and the migration transition they experience.

Figure 2-2: Conceptualisation diagram of the factors influencing student participation in higher education and the student migration outcome



Source: Authors own creation

The next individual level variable refers to the student's social background by using the student's socio—economic status and parental education. Again, this variable will directly impact on the student migration outcome as well as being interlinked to many other variables within the framework. Many previous studies have linked socio-economic status of students and the probability of them migrating in order to attend a HEI (Patiniotis and Holdsworth 2005; Christie 2007; Holdsworth 2009). Reay et al. (2005) discussed how the choice of higher education institution is a choice of lifestyle and a matter of taste in which social class is a key aspect. As a result the choice of institution can be seen as a choice of social-class matching, therefore individuals tend to choose an institution in which they feel comfortable and one that matches there social habitus and social classification.

Social-economic status also has an association with financial support, as the financial support offered to a student is means tested (United Kingdom Government 2013) and therefore students from families with lower incomes are entitled to more government funded financial support than students from families with higher incomes. The socio-economic status of the student is also interlinked to the impact of costs of tuition, rent and travel as these issues will be less of an issue to those students from more advantaged backgrounds compared to those from less advantaged backgrounds. The individual's socio-economic status is also be interlinked to the deprivation and inequality levels on the domicile (macro) scale, which in turn will impact on the quality of schooling, learning support and therefore will play some role in the level of student attainment.

The four individual level factors discussed above are therefore explicitly examined throughout this thesis to investigate how they impact on the student migration decision and future labour market outcomes. However, there are several other interlinked individual level factors that will not be examined that may also impact on the student migration process. For example, those students that have children and/or are married may be less likely to migrate in order to attend a HEI as they might be more tied to the domicile area compared to a single student with no children. The level of learning support provided to an individual may directly impact on the individual's attainment which in turn impacts on the probability of entering HE and the types of institutions they can attend. An individual's social networks can also play a key role in the student migration process. Some individuals might want to remain local to stay in the same social networks while in contrast other individuals will choose to migrate to attend a university with the intention of creating new social networks. Again, these choices by the individual will be interlinked back to the factors already discussed, such as social class or ethnic group.

The next grouping of factors within the conceptual framework is the intervening obstacles category which derives directly from Lee's (1966) push-pull framework. In the case of student migration the same principles are also relevant. The obstacles that impact on the migration decisions of students may also result in a student deciding not to participate in HE at all, however, these obstacles also have a major impact on the decision of where and what to study. This intervening obstacles grouping is situated in the middle part of the diagram as the obstacles will be influenced by factors above it, in which the individual has little control over and also influenced by factors below it in which individuals have much greater control. For example, the obstacle of 'Cost of Living' will be directly influenced by the individual choice of where to study, but the impact level of this obstacle will also be determined by other factors such as age and socio-economic status in which an individual has less or no control over.

The first intervening obstacle in the HE process to be discussed is the cost of tuition. As mentioned in Section 2.1.2, the structure of HE has changed in the UK in recent years and this has seen the introduction of tuition fees in order to attend at a HEI. Different institutions charge differing levels of fees and therefore this will have a direct impact on the choice of institution. This is also heavily interlinked to the socio-economic status of an individual, as well as other intervening obstacles such as cost of travel, rent and living and access to financial support. The distance to HEI, cost of travel, rent and living may also impact on the migration decision and are all very much interlinked with each other and the individual characteristics. Individuals from poorer backgrounds are more likely to be influenced by where they can afford to study whereas these issues may be of less of an obstacle to those student from financially better off backgrounds. The high cost of living in certain areas may also directly impact on the type of students they attract or the student migration transitions students experience in order to study there. For example the cost of living in London is 17% higher than in Edinburgh, and 23% higher than in

Manchester and the majority of this is attributable to the very high price of housing in the UK capital (Mayor for London 2003). As a result, this could impact on the decision process of individuals when deciding on the institution to attend and the location of where to reside when studying.

Arguably the most important category in the conceptual framework of the student migration decision process is the individual's choice. As in any voluntary migration decision, the decision on whether to migrate or not is primarily made by the individual as a balanced decision weighing up the positives and negatives of migrating or staying in the domicile location (Lee 1966). In the student migration process the individual student has several key choices in which they have ultimate control. However, the decision is greatly influence by all the factors above it in the framework and as discussed already in this section.

The first choice the student has to make is whether to participate in HE or not. This decision is influenced my many interlinking factors as discussed previously including social class, status, ethnicity, attainment etc. However, this thesis is not examining this decision and assumes that an individual has chosen to study at a HEI. The next major decision once an individual has decided to participate in HE is what HEI and courses to apply for. The choice of institution and course are crucial to whether or not a migration occurs and these are directly linked to the intervening obstacle of the higher education admissions process and these differ per institution.

Student's attainment level will also influence on what HEI the student will apply for in terms of that HEIs reputation and courses available. This link works in both directions. Although student achievement is linked to other factors such as quality of schooling, deprivation, socio-economic background and so forth, overall the level of student achievement is determined by the amount of effort and the ability of that individual student. As a result, it can be said that if an individual has an aspiration to study at a highly reputable HEI with high entry requirements, that this may act as an incentive for that student to highly achieve.

The choice of HEI may also be affected by the students social habitus, for example, an individual might not want to attend a prestigious institution that has little or no association with people from the social class group in which they associate themselves to belong to. Another important influencing factor on the choice of HEI is the choice of location. An individual might have several institutions in mind that offer similar courses but an individual has a certain desire to study in or to avoid certain locations. For example an individual might want to study in a large city rather than a remote campus based institution, while another student might want to avoid London. The student choice of HEI and location will also have a direct link with the migration outcome and the intervening obstacles. If the student only choses local HEIs then there is no possibility that a migration will occur while the opposite is true if the student choses all non-local HEIs. Also, the choice of HEI by the student will directly influence the intervening obstacles as HEI choice is directly linked with the distance to the HEI, cost of travel, tuition cost and cost of rent.

It is important to reiterate that every student migration decision undertaken by an individual entering HE is one that involves a complex combination of factors discussed above that differ on an individual basis. Each individual may order the importance of each of these influencing factors differently. In their choice of where and what to study some individuals may prioritise certain factors in the conceptual framework whereas other students may not even consider some of the factors discussed at all. The purpose of this framework is to consider as many of the possible factors that could influence an individual's student migration decision process in a comprehensive but succinct fashion. However, generalising for all student migration decisions is a complex task and each individual decision may be slightly different to the next.

3. Towards a typology of student migration: Incorporating a student migration area classification for the United Kingdom

3.1 Introduction

The preceding review chapter set the scene for the remainder of the thesis, both conceptually and theoretically, through providing an overview of the student migration landscape by defining concepts, reviewing previous literature and assessing the availability of data. It was evident from the previous chapter that there is a lack of previous research that analyses the migration patterns of students entering into the HE system in the UK in great detail. The principle aim of this thesis is to provide a thorough and detailed analysis of the patterns and influencing factors of student migration behaviour in the UK and how this will impact on individuals in later life. Therefore, in order to carry out the analysis for the remainder of this thesis, a well-designed mechanism of measuring student migration is required, as well as a sound grounding of the most up to date situation of the student population and migration patterns in the UK. These aspects will be adhered to here in the first substantive analysis chapter of the thesis.

As discussed within Chapter 2, there has been a significant increase in student participation in the UK in recent decades and students can have a profound impact on the locations that they reside. As a result, it is important to have an understanding of the geographical distribution of these students across the UK to be able to establish which areas of the country are impacted by these large student populations. Therefore, the first aim of this chapter is to provide a generic and descriptive overview of the student population in the UK at varying levels of geographic detail. This will provide a basic overview of the size of the student population in the UK and how it is dispersed across the country.

It was also highlighted in previous chapters that the impact students have upon an area in which they study will depend on the type of student migration transition they experience in order to attend the HEI. This is a result of the differing demands certain types of students place upon an area as a result of differing living arrangements and pressure they place on local services etc. which will be explained in greater detail later in the chapter.

However, to date, there has been no attempt to accurately categorise the different types of student migration an individual can experience in order to attend a HEI. As a result, there are no set definitions or groupings of students by their migration outcomes in the literature and it is therefore difficult to identify areas in which certain types of student groups are prominent or under represented.

The second aim of this chapter will look to address these problems by proposing a new and innovative typology of student migration. The typology will provide a well-grounded basis of how students should be categorised by their differing transitions into HE, and this can then be used to accurately measure student migration across areas and the results be used to help inform policy makers, HEIs and new potential students on a wide variety of topical areas. This typology will then be used to provide a descriptive overview of the current situation of student migration across the UK. It will be possible to highlight areas in which certain student groups are dominant or under-represented.

The typology is required in order to take advantage of the three locational variables within the dataset to depict the different possible types of student migration each individual could possibly undertake. Without this unique and innovative typology the analysis of the student migration data would lack the necessary detail to isolate the different types of student migration and would not capture the true complexity of the different movements that are occurring within the student population.

A cluster analysis will then be conducted with the aim to create a student migration area classification of the UK. This cluster analysis will add statistical robustness to the student migration typology by running clustering techniques on variables created from the student migration typology. These cluster results can then be used to analyse whether the variables created from the student migration typology accurately represent the student migration patterns across the UK. Clustering and area classification techniques are useful tools to summarise a large amount of data on a spatial scale and enabling the users to identify certain spatial patterns. Therefore, by using the student migration area classification, the final aim of the chapter refers back to the policy implications of equal access to HE across the UK by exploring if there are any spatial differences in the student migration decision and access to HEIs. No previous work has analysed how the student migration outcomes of individuals may differ as a result of geographical location across the country. If a statement is to be made about spatially equal access across the HE system in the UK then geographical patterns of student migration need to be analysed and the area classification technique will provide new summary information at the local authority level which will indicate the main type of student migration that is occurring within that area.

In the following sections of this chapter the data that is used is introduced first and subsequently the geographical distribution of the UK student population is described. The proposed typology is then defined, which offers categorisation of the different student migration transitions. Finally, the area classification of local authorities according to their student migration patterns is used to identify any spatial patterns of student migration across the UK.

3.2 HESA Student Record Data

The dataset used in this chapter is the Higher Education Statistics Agency (HESA) Student Record Data (Higher Education Statistics Agency 2012b). HESA is the official agency for the collection, analysis and dissemination of quantitative information about higher education. It was set up by agreement between the relevant government departments, the higher education funding councils and the universities and colleges in 1993, following the White Paper "Higher Education: a new framework", which called for more coherence in HE statistics, and the 1992 Higher and Further Education Acts, which established an integrated higher education system throughout the United Kingdom (Higher Education Statistics Agency 2012c).

The HESA Student Record is collected in respect of all students registered at a reporting HEI in the United Kingdom, which follow courses that lead to the award of a qualification or institutional credit, excluding those registered as studying wholly overseas. The record excludes students studying overseas for the entire duration of their course, even when they are formally registered at a UK-based HE institution. Students studying overseas by distance learning are similarly excluded; unless they are funded by a UK HE funding body. It was also deemed necessary, for the purpose of the analysis in this thesis, to remove those students registered as studying at the 'Open University'. The 'Open University' is dedicated to part-time distance learning and it was decided these students were not of interest to this study.

It must be noted that this dataset consists of 'population data' as every student in the UK is recorded in the dataset. Therefore the data is not derived from a survey and as a result standard statistical practices regarding the analysis of survey data are not required within the analysis using this data source.

The subset of the student record dataset used in this chapter contains three locational variables:

- *Domicile of Student*: The students place of permanent residence prior to undertaking a course at a Higher Educational Institution. This data is provided in the form of postcodes for UK domiciled students or country codes for internationally domiciled students. Where no data is supplied about the student's domicile, fee eligibility is used to assign student to either UK region unknown or Non-UK unknown.

- *Term-time Address of Student:* The student's term-time address postcode at some point during the reporting year. This field is required for all students except those studying by distance learning and those on placements. Although completion of the field is compulsory for all students, 'Unknown' values are acceptable. It will be possible to provide only the outward part of the postcode if this is all that is known, although that is not expected to arise commonly.

- *Institution Address:* The allocation of the HEI to a geographical region is done by reference to the administrative centre of that HEI.

These variables are aggregated by HESA and made available at two different levels of geography; Region and LA (see Section 1.2.2), while the dataset also includes international students that are recorded as having a 'Non-UK' domicile. The geographical level of counties is also used within this chapter. This geographical level has been aggregated by the authors from the reported LA variables and the county assigned to it dependent on the county in which the LA fell inside.

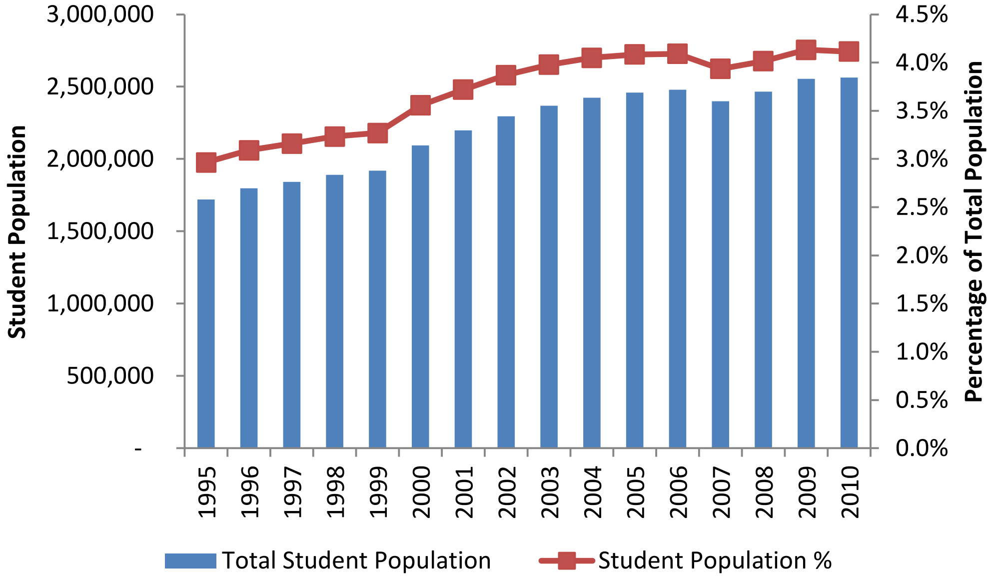
This dataset contains information on migrant stocks as opposed to flows of migrants. As explained in Section 2.2, migration data can either be recorded as transitions or events. This data measures the transitions of student migration by recording the student at their term-time address and asking the student the location of their domicile. As a result, the data used in this chapter is student migrant stock data at the start of the reported academic year within a given geographical area, as opposed to the flow of students from one area to another within a given time period.

As also discussed in Section 2.2, there are many issues with regards to defining migration and these often differ between studies. With regards to defining migration in this chapter there were a couple of restraints as a result of the available data. There was no capability to report migrations with regards to ‘any change in address’ that would have corresponded with Lee’s (1966) and Rees’s (1977) definitions of migration. For this research the ideal definition of recording a migration would have been if there was any change of address. However, due to the data availability it was only possible to record a migration if an administrative geographical boundary is crossed.

3.3 The Student Population in the United Kingdom

The extract of student record dataset obtained from HESA for the analysis in this chapter contains geographical and characteristic variables on just over 9 million students attending one of the 160 HEIs across the UK between 2007/8 to 2010/11. In the 2010/11 academic year, there were 2,562,100 student attending a HEI, which accounted for 4.1% of the total 2010 mid-year population (Office for National Statistics 2011d). The majority of students in the HESA dataset (63%) were aged between the ages of 18 and 24 and out of all people aged 18 to 24 years in the United Kingdom in 2010, 24% were attending a HEI (Office for National Statistics 2011d; Higher Education Statistics Agency 2012b). The total student population and corresponding percentage of the UK population between 1995 and 2010 are shown in Figure 3-1.

Figure 3-1: Student population in the UK 1995 - 2010

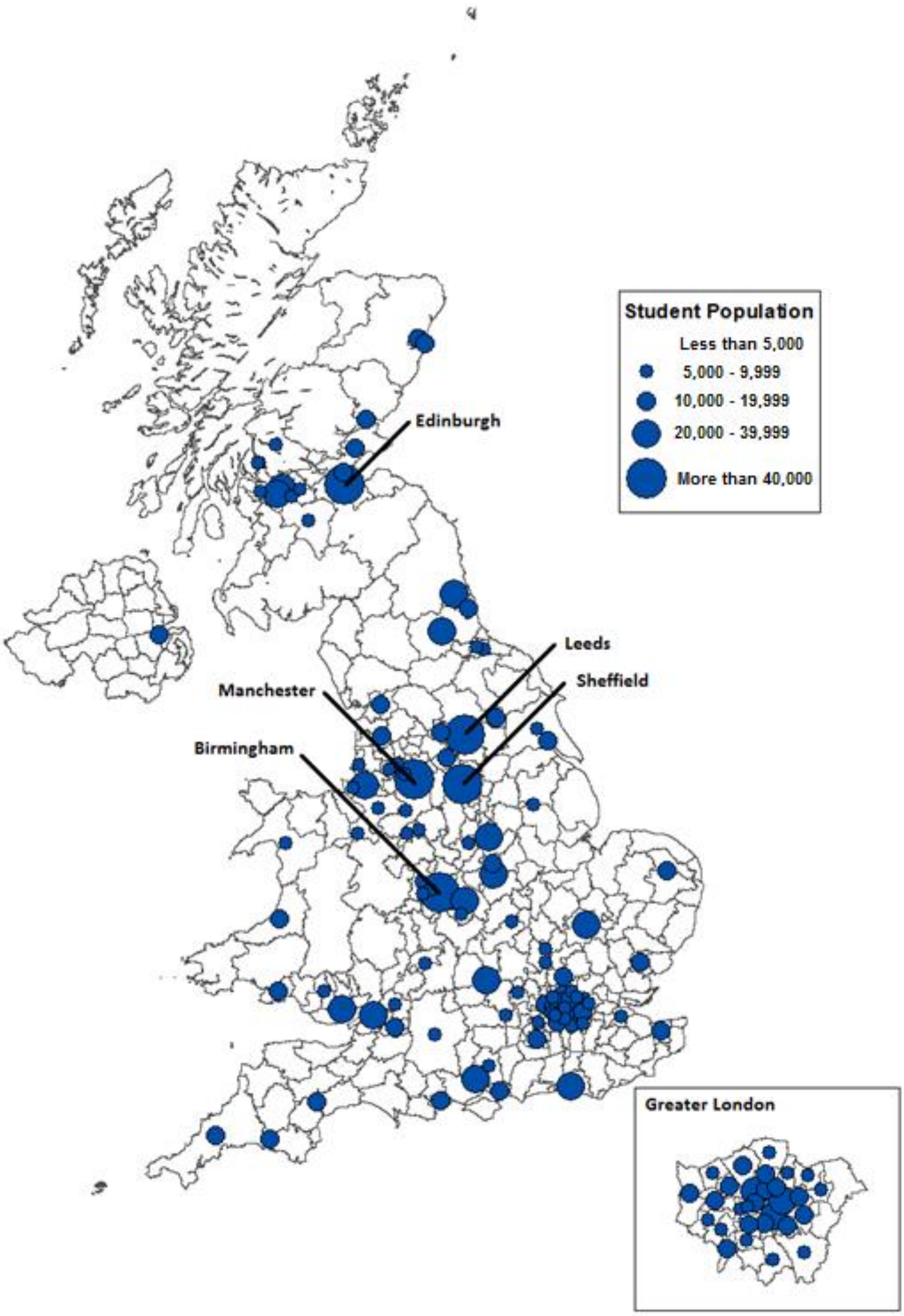


Source: Higher Education Statistics Agency (2012b); Office for National Statistics (2011d)

With the ever-increasing number of students in the UK over recent decades and the fact that students are representing a larger percentage of total population than ever before, it is important to have an understanding of how these students are geographically distributed across the UK. The geographical locations of students that attended a HEI in the UK in the academic year 2010/11 are shown in Figure 3-2. Term-time address by LA is used to identify settlements by the number of term-time resident students. It is possible to identify towns and cities within the UK that are heavily influenced by the HEIs located within them and the large numbers of students that reside in the LA. There are 5 LAs – Birmingham, Edinburgh, Manchester, Sheffield and Leeds – with over 40,000 student residents and these LAs are labelled in Figure 3-2.

However, the LAs highlighted in Figure 3-2 are LAs with large overall populations. It is often more intuitive to observe the percentages of total population that were term-time resident students rather than just using crude counts of students. The percentage of students in relation to the total population of a LA is mapped in Figure 3-3. The two LAs with the highest percentages of the total population that are term-time resident students are the well-known and renowned university settlements of Oxford (20.7% students) and Cambridge (17.4% students).

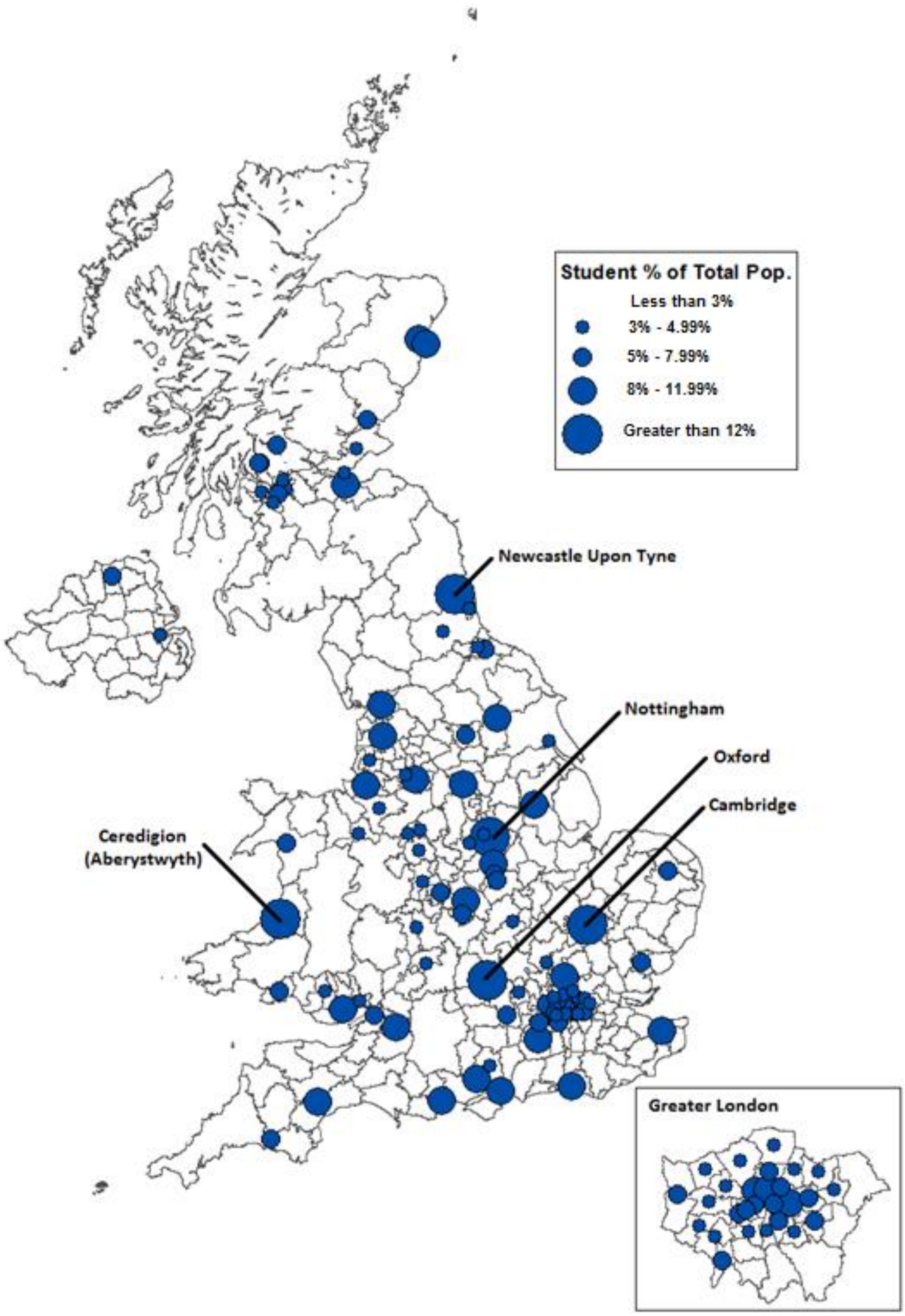
Figure 3-2: Student population by Local Authority



Source: Higher Education Statistics Agency (2012b)

Note: Term-time Address Variable used to record students geographical location. 'Open University' students not included.

Figure 3-3: Student population as a percentage of total population by Local Authority



Source: Higher Education Statistics Agency (2012b), Office for National Statistics (2011d)
Note: Term-time address variable used to record students geographical location. Open University students not included.

Table 3-1: 2010/11 UK student population at the County and Local Authority Level

	County				Local Authority			
Rank	Name	Student Pop.	Name	Student Pop. (%)	Name	Student Pop.	Name	Student Pop. (%)
1st	Inner London	192,169	Lothian	13.9	Leeds	58,104	Oxford	20.7
2nd	Outer London	162,739	South Glamorgan	8.6	Birmingham	56,890	Cambridge	17.4
3rd	West Midlands	114,652	Hereford & Worcester	6.8	Manchester	55,488	Ceredigion	14.0
4th	Greater Manchester	104,151	Inner London	6.2	Sheffield	51,252	Nottingham	12.9
5th	West Yorkshire	91,650	Oxfordshire	6.2	Edinburgh	42,524	Newcastle	12.3
6th	Strathclyde	82,796	Coleraine	5.6	Glasgow	40,546	Exeter	12.2
7th	Hampshire	69,753	Tyne And Wear	5.2	Nottingham	39,630	Southampton	11.8
8th	South Yorkshire	61,034	West Glamorgan	5.2	Cardiff	37,453	Canterbury	11.5
9th	Tyne And Wear	58,239	Leicestershire	5.1	Newcastle	35,856	Welwyn Hatfield	11.3
10th	Nottinghamshire	53,666	Belfast	5.0	Liverpool	34,623	Manchester	11.1

Source:Higher Education Statistics Agency (2012b), Office for National Statistics (2011d)

The top ten counties and LAs for total student population and percentage of the total population that were term-time resident students are shown in Table 3-1, these values supplement the data mapped in Figure 3-2 and Figure 3-3. Manchester, Newcastle and Nottingham appear in the top 10 for both student numbers and student percentages. Therefore, it is clear these LAs are greatly influenced by student populations, as students represent a large percentage of the area's population but also amount for a large crude number of the total population.

The county with the highest percentage of students was Lothian, the county that contains Edinburgh - Scotland's Capital City – and the county with the second highest percentage of students was South Glamorgan, the county containing Cardiff – the Capital of Wales. The county of Inner London was 4th and Belfast 10th. Therefore, when considering the number of students as a percentage of the total population, the four capital cities of the four constituent countries in the United Kingdom were all prominent in the top ten.

However, many cities that were in the top ten for total student population in the UK were not in the top ten when considering the percentage of students from the total population. This shows the importance of considering the percentage of students instead of just the crude counts as these can be exaggerated by large total populations. Counties with relatively small populations such as Herefordshire and Worcestershire, Coleraine, West Glamorgan and Dyfed, have large percentages of students and as a result were highlighted as 'student dominant' locations whereas they were not highlighted when considering just crude numbers. Counties containing larger HEIs, such as Oxfordshire and Avon, were also more pronounced as ‘student dominant’ areas when reviewing the percentages of students as opposed to just the crude numbers. The only counties in the top ten for both crude numbers and percentage of students were Inner London and Tyne and Wyre.

A very high percentage of students within a LA can have a profound impact on the area’s economy, housing market and demographic composition. However, it is important to investigate the number of students that migrated to these areas to study compared to the number that were already resident in the area. This is because people migrating from different areas will have a much greater impact on an areas composition and resources, while putting extra demand on an areas housing market compared to those who originate from an area, are familiar with the surroundings and local customs and are likely to already have accommodation. Therefore, the focus of the chapter will now shift to identifying the different types of student movements, measuring these student migration types within the UK and analysing any spatial differences across the country.

3.4 Typology of Student Migration

The importance and underlying complexity of accurately measuring student migration has been set out in previous chapters. Despite there being a clear understanding that student migration movements are a key and policy relevant area at the local and national level, as to date, no framework has been put forward that categorises the different migratory moves an individual can experience when entering HE. Therefore, a typology of student migration is proposed herewith that will enable the accurate and detailed definition and measurement of the different migration movement’s student’s experience when entering into HE.

Migration as a process involves at least three key variables: the migrant, the origin and the destination (Dennett and Stillwell 2010). With reference to the analysis of student migration within this chapter; the migrant refers to a student, the origin refers to the domicile of that student and the destination refers to the location of term-time address of the student or the location of the Higher Educational Institution (HEI) attended. To measure the migration transitions of people into a HEI using the HESA Student Record Data there are two possible approaches; a simple approach where the two destination variables are analysed inter-changeably or a more detailed and robust approach where a typology is created that uses all three location variables simultaneously.

To truly understand the complexity of student migration, the three variables need to be analysed simultaneously, and in order to do this, an innovative typology of student migration is proposed. This typology ideally needs to be sufficiently versatile so it can be applicable at different geographical levels of analysis, so it can be implemented on data from any country that possesses the minimum required variables and so that it can easily be applied to a variety of populations other than students, such as graduates or retirees.

It is important to be able to accurately categorise and measure the different types of transition a student can experience when entering HE. Student migration contributes to large scale population changes and as a result can redefine areas in which students decide to settle. It is also important for the individual HEIs to understand the type of transition students make in order to attend their institution. This information can then be used to focus the HEIs recruitment strategies (Beech 2014b, a) in order to maintain the types of students that institutions are already attracting, as well as allowing institutions to target those types of students they are not currently, but interested in attracting. Local housing and service providers would also benefit from an understanding of the types of students that are attracted to their areas. Those areas with large numbers of student migrants would need more services and available housing than those areas which host mainly local students. The service needs of student migrants would also be a lot more seasonal and differential between term and non-term time compared to local students who do not predominately vacate the area out of term dates.

As a result, a better understanding of the decision process experienced by young people in their student migration choices is likely to benefit many governmental and private sectors industries with regards to planning, service provision and policy creation and evaluation. Much of the previous research lacks a focus on the actual patterns of student migration and the different student migration processes. The most comprehensive study to date on student migration in the UK is now rather outdated, as it used 2001 census data (Duke-Williams 2009). The study by Belfield and Morris (1999) also examined the patterns of undergraduate movements to study at HEIs and is the only example of previous research that created some form of typology of student migration. The authors identified four categories to map 'graduate' migration; those students who moved to attend HEI and then stayed in or moved away from that region after graduation; and those who stayed in their region to attend HEI and stayed in or moved away from the region after graduation. However, this research was conducted using a survey and was only analysed at the regional level. The typology also categorised student mobility as a combination of the transition undertaken towards the HEI and the mobility experienced after graduation. Measuring the sequential mobility patters of students in a four category system greatly simplifies the student migration process which is much more inherently complex. As a result, the Belfield and Morris (1999) typology did not capture the full complexity of the possible transitions towards a HEI as suggested in the typology put forward in the current chapter. Also, the Belfield and Morris (1999) research is now rather out-dated and does not take into account the considerable increase in student numbers following the Labour Party's manifesto pledge to increase participation in HE to 50% of 18-24 year olds by 2010 (Blair 1999).

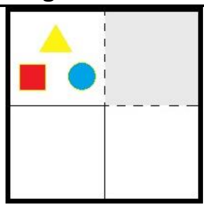
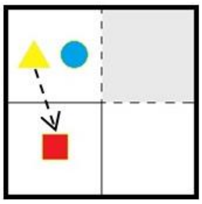
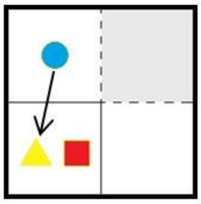
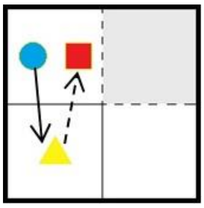
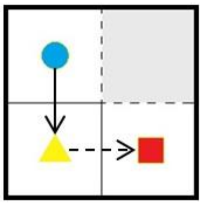
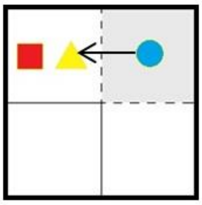
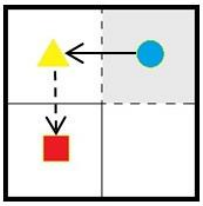
It can also be noted that the majority of the previous research in the area of student migration has been largely qualitative, survey-based and often focused on specific case study areas. This current study aims to supplement previous research, set out in Chapter 2, by providing a solely quantitative analysis of the whole UK student population. The aim is to develop a typology of student migration that can be used to conduct this quantitative analysis and as a result this will develop upon the Belfield and Morris (1999) typology by proposing a new unique typology that encompasses the complexity of the student migration decision process and the possible transitions a student could make when entering the HE system in the UK.

As previously mentioned, when analysing student migration patterns it is possible to produce some basic statistics by interpreting the destination variables of term-time address and institution address independently. However, analysing these variables independently raises the problem of which destination variable actually provides the desired and most accurate answers and results in redundant data from the unused variable. For example, when comparing domicile and term-time address, 43% of all students in 2010/11 had a term-time address in a different LA to their domicile, suggesting that under half of all students in 2010/11 migrated to attend a HEI. However, when institute address was used instead of term-time address as the destination variable, this value increases significantly to 88%. This shows that a lot more students attended a HEI in a different LA to their domicile than those students who lived in a different LA to their domicile during term-time.

So which variable is most suitable and which variable best reports student migration? The answer to this question is ambiguous. This example highlights the deficiency in analysing the student migration patterns in this way and illustrates that this simple technique would not provide the full picture of the student migration transitions taking place. It also raises the point that by analysing the destination variables separately, there are several key questions that cannot be answered with any certainty. These include how many students actually migrated to attend a HEI, how many stayed in the same area and commuted or how many stayed in the same area and attend a local HEI? If any of these more complex and relevant questions are to be answered with any degree of certainty, all three locational variables need to analysed simultaneously. This can be done best by creating a typology that combines all three location variables in unison. The creation of this typology allows the researcher to accurately capture the actual transitions a student may experience in order to attend a HEI and results in all the available data being utilised.

The typology is derived by categorising an individual student as a result of the geographical location of all three locational variables. Due to the different possible combinations of these variables, eight main categories of student movements have been created, as illustrated in Table 3-2.

Table 3-2: Typology of Student Migration: Categories, Descriptions and Diagrams

Student Category	Description	Diagram
1 – Local Student	- No migration to attend a HEI = Domicile in same geographical area as term-time address. - Domicile, term-time address and institution address in the same geographical area, suggests student did not migrate to attend a HEI as they attend a local HEI.	
2 – Commuter or Distance Learner	- No migration to attend a HEI = Domicile in same geographical area as term-time address. - Institution address is in different geographical area as domicile and term-time address which suggests the student commutes to the HEI or is distance learning.	
3 – Internal Student Migrant	- Migrates to attend a HEI = Domicile in different geographical area as term-time address. - Term-time address is in same geographical area as institution address, therefore the student is assumed to live close to the institution. - UK Domicile – Internal Migrant	
4 – Migrant Commuter or Distance Learner attending local HEI	- Migrates to attend a HEI = Domicile in different geographical area as term-time address. - Domicile and institution address in same geographical area but term-time address is different, therefore they migrate away to live and commute back to the same area as the domicile to attend the HEI or they distance learn.	
5 – Internal Migrant Commuter or Distance Learner	- Migrates to attend a HEI = Domicile in different geographical area as term-time address. - All three addresses are in different geographical areas, therefore they migrate to different term-time address but still commute to the HEI or distance learns. - UK Domicile – Internal Migrant	
6 – International Student Migrant	- Internationally Migrates to attend a HEI = Domicile outside of UK and term-time address in the UK. - Term-time address is in same geographical area as institution address, therefore the student is assumed to live close to the institution. - Non-UK Domicile – International Migrant	
7 – International Migrant Commuter or Distance Learner	- Internationally Migrates to attend a HEI = Domicile outside of UK and term-time address in the UK. - Term-time address in different geographic area as institution address and therefore commute to the HEI or distance learn - Non-UK Domicile – International Migrant	
8 – Unknown	- Domicile or Term-time address recorded as unknown	

Notes: = Internal Geographical Area Border = International Geographical Area Border = Domicile = Term-time address = Institute Address = Migration = Commute

The eight categories were designed so that each category is mutually exclusive and that every student fits into only one category. If the geographical location for one or more of the variables is different to the other(s), then a geographical boundary has been crossed and some sort of movement has occurred. Three different types of movement are identified within the student migration typology, depending on which variables were in different geographical areas. These were; International Migration, Internal Migration and Commuting. The eight student categories have been ordered in terms of the perceived size/distance of the movement recorded, where category one experiences no movement and category seven experiences the largest movement.

This typology is vital to gain a better understanding of student migration because all three variables are interlinked, with each variable reflecting a student decision process. This typology is unique and no such typology has been created to accurately categorise and measure the student migration movements. This typology is also unique in that it can easily be used at differing geographical levels and can be used to measure the migratory movements of population groups other than students.

The domicile and term-time addresses reflect the student's decision to relocate, study locally or commute/distance learn. This migration decision is directly affected by a number of factors. One factor being the distance of the nearest HEI to the students domicile (refer back to Figure 1-1), for example there is no possibility a student could be a local student if there is no HEI in their area. Other possible factors that influence this decision are those depicted in Figure 2-2 that include unobservable variables such as the student's financial situation or the level of parental support. The choice of institution is also influenced by the attainment level of the student, the location of the desired HEI, the student's views on the quality of the HEI and courses available. Without this typology to observe the three location variables simultaneously, these unique interactions would be lost and the dimensionality of the research would be significantly reduced.

The numbers and percentages of students in each student category for the academic years 2007/08 to 2010/11by County and LA are shown in Table 3-3.

Table 3-3: Total number and percentage of UK student population by student category and academic year at the County and Local Authority levels

Student Category	County								Local Authority							
	2007/08		2008/09		2009/10		2010/11		2007/08		2008/09		2009/10		2010/11	
	Total	%	Total	%	Total	%	Total	%	Total	%	Total	%	Total	%	Total	%
1 – Local Student	553,215	25.5	595,703	26.5	615,659	26.4	600,967	25.6	202,081	9.3	213,118	9.5	216,628	9.3	210,999	9.0
2 – Commuter or Distance Learner	438,636	20.2	462,857	20.6	480,079	20.6	485,218	20.7	729,339	33.7	777,323	34.6	808,215	34.6	804,443	34.3
3 – Internal Student Migrant	512,187	23.6	577,303	25.7	614,659	26.3	636,128	27.1	486,103	22.4	547,834	24.3	586,218	25.1	604,226	25.8
4 – Migrant Commuter or Distance Learner attending local HEI	12,700	0.6	12,984	0.6	13,317	0.6	13,689	0.6	9,437	0.4	9,596	0.4	9,668	0.4	9,901	0.4
5 – Internal Migrant Commuter or Distance Learner	41,397	1.9	42,955	1.9	45,823	2.0	49,034	2.1	131,175	6.1	143,931	6.4	148,808	6.4	155,467	6.6
6 – International Student Migrant	223,459	10.3	268,232	11.9	302,765	13	323,654	13.8	179,359	8.3	217,215	9.7	246,515	10.6	265,442	11.3
7 – International Migrant Commuter or Distance Learner	37,328	1.7	40,036	1.8	46,154	2.0	49,495	2.1	81,428	3.8	91,053	4.1	102,404	4.4	107,707	4.6
8 – Unknown	348,460	16.1	250,401	11.1	216,978	9.3	186,621	8.0	348,460	16.1	250,401	11.1	216,978	9.3	186,621	8.0

Source: Higher Education Statistics Agency (2012b)

Note: Students registered with the Open University excluded.

The distributions of all students across the student categories remained constant over the four-year period, except for the increase of those in the 'Internal Student Migrant' and 'International Student Migrant' categories. The increase in the number of 'Internal Student Migrant' and 'International Student Migrant' categories can be largely attributed to the decline in the numbers recorded in the 'Unknown' category. This decline in domicile and term-time address being recorded as 'Unknown' only resulted in increases in internal and international student migrants, which implies that a large percentage of the students still recorded as unknown also belong to these two categories. However, this cannot be said with any statistical certainty.

Due to the distributions of the student categories remaining constant over the time period in which the data were available, the remaining analysis will focus on the most recent academic year available, 2010/11. Each of the student migration categories illustrated in Table 3-2 are explained in turn below.

3.4.1 Local Students

The first student migration category, 'Local Student', refers to a student that has all three location variables in the same geographical area and therefore travelled no distance and attended one of the local HEIs. There has been growing interest in this group of students in recent years. Several studies have investigated the changing patterns of student types - in particular the increasing numbers of local students - and how this could be related to increases in the overall participation rates in HE, increases in tuition fees, increasing costs of living and changing social-economic composition of university students (Patiniotis and Holdsworth 2005; Allinson 2006; Faggian et al. 2006; Christie 2007; Holdsworth 2009; Khambhaita and Bhopal 2013).

There is, however, no capability within this dataset to distinguish between those local students that remain in the parental home and those that do not. All that can be interpreted from the data is that the individuals domicile and term-time address is in the same geographical area, it cannot be measured if this is the same address or a different address within the same area. Previous research by Patiniotis and Holdsworth (2005) found that 22.7% of students that responded to their questionnaire remained in the parental home while studying at a HEI, while according to the 2001 census, just over a quarter of all students were still living in the parental home in 2001 (Munro et al. 2009). However, it is not known whether these students in these pieces of research represent local students as no information was provided regarding the location of their HEI so it cannot be said if they attended a local HEI or commuted to a HEI in a different LA.

In 2010/11, 600,967 (25.6%) of all students were classified as a local student at the county level of geography, which tells us that just over 1 in 4 of all students attended a HEI in the same county as their domicile and term-time address. However, when the LA level of geography was used then only 210,999 (9.0%) of all students in 2010/11 were classified as a local student. When using the LA level geography, the area in which a student must remain to be categorised as 'local' is much smaller than when using the county level geography, hence the decline in numbers. A student may be categorised as a local student when using the county level geography but might be attending a HEI, that although is within the same county, is quite a large distance away. Therefore, the choice of geography is important when analysing local students and the results can differ greatly as a consequence of this choice.

Table 3-4: Analysing the Student Migration Typology Categories - Numbers and Percentages of Local Students – Top 10 Counties and LA

	County				LA			
Rank	**Name**	**Count**	**Name**	**%**	**Name**	**Count**	**Name**	**%**
1st	Inner London	57,108	Suffolk	67.8	Birmingham	13,412	Highland	32.8
2nd	Strathclyde	54,446	Strathclyde	62.6	Glasgow	13,314	Bradford	26.7
3rd	West Midlands	45,498	Cleveland	43.5	Leeds	9,128	Cornwall	20.1
4th	Greater Manchester	36,992	Hereford & Worcester	37.9	Sheffield	8,798	Bolton	19.7
5th	Outer London	34,956	Highland	37.6	Edinburgh	8,365	Swansea	19.1
6th	West Yorkshire	31,321	Kent	36.3	Liverpool	6,694	Glasgow	19.0
7th	Lancashire	19,963	Northamptonshire	34.1	Cardiff	5,816	Ipswich	18.8
8th	Tyne And Wear	19,781	Staffordshire	33.9	Manchester	5,705	Birmingham	17.5
9th	Hampshire	15,944	Humberside	33.7	Leicester	4,962	Kirklees	17.1
10th	Merseyside	15,303	Greater Manchester	33.3	Aberdeen	4,524	Sunderland	17.0

Source: Higher Education Statistics Agency (2012b)

The counties and LAs with the largest numbers and highest percentages of local students are shown in Table 3-4. The county of Inner London had the largest number of local students with 57,108 (21.2%). This was not surprising given Inner London's total population size and the large number of HEI's (30) in the county. Due to the large number of HEI's in Inner London, there was a large variety of institutions and courses that a student could choose from, which results in less of a need to migrate in order to find a specific course. It must also be noted, that although a student will be classified as a local student within the county of Inner London, the actual journey times may be very long and a large number of these students may be better recorded as a student commuter, however it was not possible to make this distinction from the data.

The county of Strathclyde had a high percentage of local students as well as a large crude number of local students. 62.6% of all students in Strathclyde were classified as local students. One of the reasons for this large percentage of people staying in Strathclyde to attend a HEI may be a result of the Scottish Government subsidising tuition fees for Scottish students attending a Scottish HEI and as a result providing an incentive to students to remain local. There are also 6 HEIs within Strathclyde offering a wide range of choice to those students who want to stay local. This trend may also be explained by the research that shows Scottish students routinely live in the parental home and attend a local university (Munro et al. 2009). This perception is backed up by another previous study that found Scottish students to be less mobile than their British counterparts, with between 85% and 90% of all undergraduate places at Scottish HEIs being filled by Scottish Domiciled students (Scottish Executive 2005; Christie 2007). Glasgow LA also had a large crude number of local students as well as a high percentage of local students. Glasgow LA is part of the Strathclyde County, and therefore the same explanations still hold here.

Birmingham was the LA with the largest number of 'Local Students' with 13,412, while also having a high percentage of local students, 17.5%, and ranked 8th. Again, this is not surprising given Birmingham's large total population and hosting 7 HEI's, which also results in less of a need to migrate in order to find a HEI or specific course. Another potential cause of the high percentages of local students in Birmingham could be the ethnic composition of the LA population. Previous work by Khambhaita and Bhopal (2013) found that non-white students were more likely to stay in the parental/guardian home during their first year at university and as a result there is a potential link between the number of local students and the ethnic composition of a LA. Work by Faggian et al. (2006) also found that non-white UK students were much less likely to migrate to attend university. According to the 2011 census, the LA of Birmingham had a very high non-white population, accounting for 42% of the total population compared to 15% for the whole of England (Office for National Statistics 2013b). This high number of non-white people in the population could be correlated to the high percentages of local students recorded in Birmingham. The LA of Bradford also had a very high percentage of local students and similarly to Birmingham also had a very high non-white population (32.6%) according to the 2011 census (Office for National Statistics 2013b).

It is also important to remember that not all counties or LAs host a HEI - only 98 of the 408 UK LAs host a HEI. As a result, students domiciled in those areas without a HEI had no possibility of being a local student and this will partly explain the low percentage of all students that were local students at the LA level. The Local Student category is also particularly vulnerable to the Modifiable Area Unit Problem (MAUP) as discussed previously, as the geographical level being used greatly impacts on the size of an area a student can study in and remain being categorised as 'local'. For example a local student on the region level may not actually be studying in a HEI in close proximity as a region area is vast, in comparison at the LA level the areas are much smaller and a 'local' student would have to be studying in a HEI within a much smaller distance range.

When only those students that lived in a domicile with a HEI were analysed (i.e. those that were at risk of being a local student, the percentage of those that were local students increased from only 9% to 27% at the LA level.

3.4.2 Commuter or Distance Learner

The second category, 'Commuter or Distance Learner', refers to a student whose domicile and term-time addresses are in the same geographical area but the HEI they attend is in a different geographical area. These students made a movement across a geographical boundary in order to attend a HEI, but no boundary was crossed between domicile and term-time address. Therefore, the student does not migrate to attend a HEI but has to regularly commute, are distance learners[4] or combine the two in order to attend the course at their HEI. As a result, the social, economic and cultural impact of these students follow closely those of local students closely as, in both cases, no migration took place, however there is a need to ensure that these students are provided with adequate transport links and financial support to enable these students to commute to the HEI they are attending. A limitation of this category comes as a result of the restraints of the data used in this research. As a result, there was no information recorded on the frequency of the students travel to the HEI and there was no way to distinguish between a commuting student or someone who was distance learning.

When studying this category of students there are two main areas of interest; what are the locations of the HEIs that attract a large amount of commuting or distance learning students and what are the locations that these students reside in during term-time.

In 2010/11, 485,218 (20.7%) of all students were classified as commuting or distance learning at the county level of geography. This figure rises significantly to 804,443 (34.3%) when analysing the LA level of geography. This again emphasizes the importance of the level of geography used to run the analysis and the impact of the choice of geographical level will be reviewed in more detail later in the chapter.

First, the geographical areas hosting HEIs with large numbers of commuting or distance learning students are investigated, with the top 10 counties and LAs by the institute location, shown in Table 3-5.

Table 3-5: Analysing the Student Migration Typology Categories - Numbers and Percentages of Commuter or Distance Learners – Top 10 Counties and LA – by Institute Address

	County				LA			
Rank	**Name**	**Count**	**Name**	**%**	**Name**	**Count**	**Name**	**%**
1st	Inner London	78,863	Shropshire	59.9	Camden	33,274	West Lancashire	80.8
2nd	Outer London	34,611	Highland	53.1	Westminster	29,133	Renfrewshire	73.9

[4] "A distance learner can be simply defined as a person who partakes in a system of education delivery in which the majority of learning takes place with the learner and the teacher separated by space and/or time, the gap between the two being bridged by technology. A distance learner is one who experiences the majority (80+%) of their learning off-campus at a distance from the teacher and consequently has limited face-to face interaction with their teachers and peers" (Tynan 2010).

3rd	Lancashire	30,720	Belfast	52.3	Glasgow City	28,497	Ipswich	67.9
4th	West Midlands	25,788	Coleraine	50.0	West Lancashire	23,122	Southwark	65.8
5th	Greater Manchester	14,774	Lancashire	41.5	Manchester	20,734	Middlesbrough	65.6
6th	Avon	14,477	Cumbria	39.9	Preston	18,283	Wolverhampton	62.7
7th	West Yorkshire	14,288	Gwent	38.2	Middlesbrough	18,012	Telford and Wrekin	61.5
8th	Belfast	14,053	Buckinghamshire	38.1	Southwark	16,909	Chelmsford	59.9
9th	Coleraine	13,035	Cleveland	35.7	Birmingham	16,632	Worcester	59.0
10th	Essex	11,461	Mid Glamorgan	34.4	Greenwich	16,514	Newcastle-under-Lyme	57.3

Source: Higher Education Statistics Agency (2012b)

23.4% of commuting or distance learning students attended a HEI within the two counties that represent London, while around 5.5% of all commuting or distance learning students attended a HEI in London and lived in the neighbouring counties of Berkshire, Essex, Hertfordshire, Kent and Surrey. When you look at the LA level of geography, 3 of the top 10 LAs (Camden, Westminster and Southwark) were also in London. This strongly suggests that these students did not migrate to London and commuted to take advantage of the cheaper housing costs of living outside of London itself in conjunction with the good transport links into London and the financial benefits of remaining in the parental home. London has very good transport links and offers a wide choice of institutions and courses making the option of commuting to London a viable option for many students from surrounding areas.

The LA of West Lancashire had the highest percentage of commuting or distance learning students (81%), the majority of these resided in the neighbouring LAs of Sefton, Liverpool or Wigan. Edge Hill University is located in West Lancashire and attracts a large amount of students from neighbouring LAs. The transport links in this area are good and the distances travelled across LAs to Edge Hill University are quite short which may be the cause of these surprisingly high numbers. Renfrewshire LA also had a very high percentage of commuting or distance learning students; this may be a result of a large number of students studying at the 'West of Scotland University' and commuting from the nearby LA of Glasgow.

The county of Shropshire and LA of Telford & Wrekin (LA of Telford & Wrekin lies within the county of Shropshire) had very high percentages of commuters or distance learners. This is the location of Harper Adams University College which is one of the leading agricultural universities in the UK and attracts its students from a wide variety of areas. However, a large percentage of these commute rather than residing within the area, this may be a result of the HEI not being located in a large settlement that has the services that attract young people to reside in the area.

The areas in which commuting or distance learning students live during term-time are also of interest. These areas are likely to have strong inter-linking relationships with the areas in which the students attend a HEI and these interactions are important due to the necessary transport corridors between the two locations. The top 10 sending counties and LAs of commuters or distance learners are shown in Table 3-6.

9 out of the top 10 counties were from Northern Ireland (NI). This can be explained by the location of HEIs in NI (Figure 1-1). In NI, the counties of Belfast and Coleraine have HEIs and therefore the system of HE is very different to that of the rest of the UK. As a result, students from other counties in Northern Ireland have to migrate or commute to these HEIs to participate in HE. However, it appears from the results of the data that large percentages of students in Northern Ireland chose to commute to attend a HEI in Northern Ireland rather than migrating to either Belfast or Coleraine.

Table 3-6: Analysing the Student Migration Typology Categories - Numbers and Percentages of Commuter or Distance Learners – Top 10 Counties and LA – by Domicile/Term-time Address

	County				LA			
Rank	**Name**	**Count**	**Name**	**%**	**Name**	**Count**	**Name**	**%**
1st	Outer London	72,024	Newtownabbey	62.1	Birmingham	9,412	North Lanarkshire	84.1
2nd	Inner London	26,651	Carrickfergus	60.4	South Lanarkshire	7,657	West Dunbartonshire	82.6
3rd	Greater Manchester	16,956	Castlereagh	59.3	Cornwall	7,556	South Lanarkshire	78.2
4th	Merseyside	14,385	Antrim	58.1	Brent	7,516	Inverclyde	76.5
5th	Cheshire	11,829	Lisburn	58.0	County Durham	7,498	East Renfrewshire	76.4
6th	Kent	10,589	Magherafelt	56.5	Lambeth	7,443	East Dunbartonshire	75.0
7th	West Midlands	10,134	Craigavon	56.2	North Lanarkshire	7,046	Stoke-on-Trent	74.9
8th	Lancashire	9,666	Durham	56.0	Croydon	6,855	South Tyneside	73.9
9th	Essex	9,595	Ballymena	55.1	Redbridge	6,852	East Ayrshire	73.6
10th	Berkshire	9,568	Derry	54.8	Ealing	6,776	Redcar and Cleveland	73.4

Source: Higher Education Statistics Agency (2012b)

The largest flows of commuting or distance learning students between the domicile/term-time address LA and the institute address LA were from Cornwall LA to Plymouth LA (5,517 students), from South Lanarkshire to

Glasgow City (4,442) and from North Lanarkshire to Glasgow City (4,010). All of these flows were over relatively short distances and show the interconnectedness of these LAs for the short-term/daily movement of students from term-time address to place of study.

3.4.3 Internal Student Migrant

The third category, 'Internal Student Migrant', refers to a UK-domiciled student who migrates across a geographical boundary to attend a HEI and their term-time address is in the same geographical area as the institution attended. This group of students have been of interest to researchers for a significant period of time, as it has long been perceived that a large proportion of students attending HE in the UK move away from their parental home to take up their studies (Nicholson and Wasoff 1989; Boyle et al. 1998; Belfield and Morris 1999; Allinson 2006). Other areas of research related to internal student migrants is the link between student migration and the transition to adulthood, as well as the studies of studentification and shifting housing market needs of areas (Smith and Holt 2007; Holdsworth 2009; Sage et al. 2012a, b). Links have also been made between the traditional pattern of students leaving home to go to university and factors such as socio-economic class, ethnicity and parental education (Holdsworth 2009; Khambhaita and Bhopal 2013), although this is discussed in much greater detail in Chapter 4.

In the 2010/11 academic year, 636,128 (27.1%) students were categorised as being an internal student migrant at the county level of geography, while the number of internal student migrant at the LA was relatively similar at 604,226 (25.8%). This indicates that the choice of geography does not significantly impact on the overall numbers and percentages of internal student migrants. This suggests that those that chose to migrate in order to attend a HEI do so over a large enough distance that ensures they not only change the LA that they reside but also the county as well.

The top 10 counties and LA with regards to crude numbers and percentages of internal student migrant by term-time and institute address (destination variable) are shown in Table 3-7.

The county of West Yorkshire (39,617) and the LA of Leeds (35,597) were resident to the largest crude stocks of internal student migrants. It has already been noted (in Section 3.3) that both the county of West Yorkshire and the LA of Leeds had a very large student populations and a large percentage of those students migrated into the area in order live and study at a HEI - 52.8% of Leeds LA students were internal student migrants.

Table 3-7: Analysing the Student Migration Typology Categories - Numbers and Percentages of Internal Student Migrants – Top 10 Counties and LA – by Term-time/Institution Address (Destination)

	County				LA			
Rank	**Name**	**Count**	**Name**	**%**	**Name**	**Count**	**Name**	**%**
1st	West Yorkshire	39,617	Cornwall	69.2	Leeds	35,597	Cornwall	69.2
2nd	Inner London	38,744	Durham	57.2	Manchester	29,960	Cotswolds	63.4
3rd	Hampshire	32,758	Cambridgeshire	54.1	Sheffield	29,167	Ceredigion	59.4
4th	West Midlands	32,363	North Yorkshire	50.1	Nottingham	25,969	Charnwood	57.6
5th	Greater Manchester	31,958	Lincolnshire	48.1	Birmingham	20,849	County Durham	57.4
6th	South Yorkshire	28,263	Dorset	48.0	Newcastle upon Tyne	20,496	Lancaster	54.2
7th	Nottinghamshire	27,994	Hampshire	47.1	Oxford	18,535	Cambridge	53.0
8th	Avon	24,256	Berkshire	44.9	Southampton	17,981	Leeds	52.8
9th	Leicestershire	23,319	Nottinghamshire	44.4	Liverpool	17,854	Lincoln	52.4
10th	Tyne And Wear	22,087	Fife Region	44.0	Cardiff	17,136	York	52.2

Source: Higher Education Statistics Agency (2012b)

Inner London also had a very high number of Internal Student Migrants, however, this only represented around 14.4% of all students attending HEIs in Inner London. The county of Hampshire had the 3rd highest number of internal student migrants, accounting for 47.1% of all students, while the LA of Southampton, which falls within the county of Hampshire, also had a large internal student migrant population of 17,981 (48.6%).

When focusing on the crude number of internal student migrants all of the top 10 counties and LAs were areas with large overall student populations due to issue discussed in Section 3.3. However, the percentage of all students that were internal student migrants were analysed a different picture appears. The LAs of Cornwall and Cotswolds had very high percentages of internal student migrants, although these LAs actually only represented a very small number of students. The counties of Durham (57.2%) and Cambridgeshire (54.1%) and the encompassed LAs of Cambridge (53.0%) and County Durham (57.4%) had very high percentages of internal student migrants. These percentages were over two times higher than the UK average. Cambridge LA also had a very low number of local students (2.7%), this value was one of the lowest throughout the UK, suggesting that very few students from Cambridge LA got the chance to study at the prestigious local institution, The University of Cambridge. Also, the lack of neighbouring LAs in the top flows to the LA of Cambridge suggests that students are attracted to this LA over large distances. A similar trend was visible with the LA of County Durham where only 4.6% of students were local students and the top four sending LAs (Leeds, Wiltshire, Cheshire East and Richmond upon Thames) of internal student migrants to County Durham were not neighbouring counties but were quite large distances away.

3.4.4 Internal Migrant Commuter or Distance Learner

The 'Internal Migrant Commuter Distance Learner' category refers to a UK-domiciled student who migrates across a geographical boundary to attend a HEI but unlike a student migrant, the institution address is also in a different geographical area to the term-time address. This group of students are of interest because whilst they migrate away from an area to reside during term-time and therefore have the housing and service needs of a student migrant, the area in which these services were required was not in the same geographical area to their HEI. Therefore, these students need to have housing services in one area and good transport links to the HEI they attend and, as a result, these students need to be considered separately in any analysis.

In the 2010/11 academic year 45,823 (2.0%) students were classified as an internal migrant commuter or distance learner at the county level. This figure increases 3 fold to 155,467 (6.6%) when categorised using the LA level data. This increase is mainly caused by those students who were classified as an internal student migrant at the county level but at the LA level the term-time address and institution address are not in the same LA but are in the same county.

When the term-time address and institution location of students within this category are compared simultaneously, some interesting patterns appear in the results. These can be easily explained once the geographical areas and institutional locations were explored in greater detail. The largest number of this type of students for county term-time address was Outer London while the corresponding area for county institution address was Inner London. Therefore, it can be said that these students migrated away from their domicile county to attend a HEI in Inner London but actually resided during term-time in Outer London and commuted to Inner London to attend a HEI. This was probably a result of the very high housing and renting costs within Inner London and many of the HEIs in Inner London are located in very central locations where housing is simply not available. Also, a large number of institutionally owned housing, such as halls of residence, are located in the outer areas of London due to cheaper land prices and hence a large number of students studying in London fall within this category.

When the lower geographical level of LA is used then a similar trend was also apparent. A large number of LAs in regards to numbers and percentages of students in the two tables were LAs within London. Just under half of students attending a HEI in the City of London were internal migrant commuters or distance learners, while the top two LAs for the number of internal migrant commuters were Westminster and Camden.

The neighbouring counties of Warwickshire and West Midlands also have large numbers of students in this category indicating that a large number of students migrated to live in Warwickshire but attended a HEI in the neighbouring West Midlands. Also, Bournemouth LA had very high numbers of this student category with regards to term-time address while neighbouring Poole LA had very high numbers of this student category registered as attending a HEI in the LA. This can be explained by Bournemouth University actually having its institute address in Poole LA and a large amount of students attending this university actually residing in halls of residence and privately rented housing in the neighbouring Bournemouth LA. It must be noted that in recent years the two settlements of Poole and Bournemouth have grown in size and there is no longer a clear divide between the two settlements and the distance commuted by students in this category between these two LAs are very small. This explanation was also valid for the LAs of Wokingham and Reading that appear in the top 5 for percentage of student migrant commuters. Again, the University of Reading is actually located in the Wokingham LA although a lot of students resided within the city of Reading itself and travelled the short distance to the university campus. The LA of Warwick had a large number and percentage of internal student migrants commuters with term-time address in the LA, this was caused as a result of the University of Warwick being located in the LA of Coventry and a large amount of students lived in the town of Warwick and travelled the short distance to the University of Warwick campus which is located just over the LA boarder between the Warwick and Coventry LAs. The LA of Nottingham also had a large amount of students from this category that attend a HEI in Nottingham but migrated to reside in a different LA. Again, a large amount of these students migrated to the neighbouring LAs of Rushcliffe (27.4%) and Broxtowe (17.1%) which were very short distances away from the campuses of the two HEIs within Nottingham LA.

This internal migrant commuter or distance learner category is a prime example of how the level of geography used in the analysis had a significant impact on the classification of a student migration transition. There have been several examples shown here which illustrate how certain students may have resided in a different LA to the HEI but in reality they lived just on the other side of an administrative boundary and the distances involved were very small. Here the user of these findings must be aware of these issues and the possible impact of the modifiable area unit problem on such student classifications (Bell et al. 2002).

3.4.5 International Students

The sixth and seventh categories refer to students that immigrate from outside of the UK in order to attend a HEI in the UK. These categories are identical to the third and fifth categories, except that these students were from Non-UK domiciles. In the 2010/11 academic year, 373,149 (15.9%) of all students in the UK were international students with a domicile from outside the UK. Like the internal students, the international student migrants are split into two categories: one for those whose term-time and institute address are in the same geographical area and one for those whose term-time and institute address are in different areas and therefore they either commute or are distance learners.

International Student Migrant

In the 2010/11 academic year 323,654 (13.8%) students were international student migrants at the county level compared to 265,442 (11.3%) at the LA level. The entire decline in the number at the LA level was a result of those students being classified as an international student migrant commuter or distance learner at LA level. The top 10 counties and LAs for numbers and percentage of international student migrants are shown in Table 3-8. The counties and LAs with very large numbers of international student migrants were all areas with large HEIs, with high international reputations. The county of Lothian and LA of Edinburgh (Edinburgh falls within county of Lothian) had very high numbers and percentages of international students, Inner London also attracts a large number and percentage of international student migrants, while the well-established university cities of Oxford and Cambridge (as well as Oxfordshire and Cambridgeshire) also appear in Table 3-8. This suggests that international students are attracted to the large cities and well established HEIs throughout the UK. When you consider the LA of Cambridge, 56.3% of students were internal student migrants and 28.0% were international student migrants, therefore over 81% of all students studying and living in Cambridge LA migrated to the LA to do so – this was by far the highest percentage for a LA in the whole UK.

Table 3-8: Analysing the Student Migration Typology Categories - Numbers and Percentages of International Student Migrants – Top 10 Counties and LA – by Institute Address

	County				LA			
Rank	**Name**	**Count**	**Name**	**%**	**Name**	**Count**	**Name**	**%**
1st	Inner London	52,273	Fife Region	34.4	Edinburgh, City of	11,827	Aylesbury Vale	39.3
2nd	West Midlands	24,853	Cambridgeshire	29.0	Birmingham	11,391	Central Bedfordshire	35.0
3rd	Greater Manchester	15,816	Clwyd	28.6	Manchester	11,047	Fife	34.4
4th	Outer London	15,478	Bedfordshire	24.0	Sheffield	10,496	Wrexham	28.5
5th	Lothian	12,577	Surrey	20.9	Coventry	10,275	Cambridge	28.0
6th	Tyne And Wear	12,025	Lothian	20.5	Oxford	8,576	Guildford	25.5
7th	West Yorkshire	11,677	Grampian Region	20.1	Glasgow City	8,367	Colchester	24.0
8th	Hampshire	10,956	Oxfordshire	19.7	Newcastle upon Tyne	8,069	Edinburgh, City of	21.2
9th	South Yorkshire	10,555	Inner London	19.4	Nottingham	7,972	Runnymede	20.7
10th	Strathclyde Region	9,463	Tyne And Wear	17.3	Leeds	7,046	Exeter	20.3

Source: Higher Education Statistics Agency (2012b)

Other LAs in Table 3-8 stand out for having very high percentages of International Student Migrants that might not have been expected to do so. Aylesbury Vale, home of University of Buckingham, had an international student migrant population of 39.3%, the highest of all LAs. 35% of students from Central Bedfordshire LA, home of Cranfield University, were international student migrants, while 25.5% of the student population in the LA of Guildford, University of Surrey, were international student migrants. This may be a result of these LAs being in close proximity to London and offering many courses that were attractive to international students, but they can benefit from not having to live in the London where living costs are much higher than other parts of the UK.

International Student Migrant Commuter or Distance Learner

In the 2010/11 academic year 49,495 (2.1%) students were registered in this category at the county level, while the figure increases to 107,707 (4.6%) at the LA level. This category represented a relatively small number of all students in the United Kingdom and these students follow very similar spatial patterns to their corresponding internal migrant commuter students. At the county level 23.5% of all the students in this category were attending a HEI within the county of Inner London, while when both London counties were counted together 38.7% of all the international student migrant commuters or distance learners were attending a HEI within London.

3.4.6 Unknowns

The final category contains any student in which any of the locational variables (domicile, term-time and institute address) were recorded as unknown. As previously mentioned, all the data within this dataset is provided to HESA from the HEIs themselves and as a result the institute address for all students in the dataset was provided and therefore had no students recorded as unknown. However, for the variables of domicile and term-time address, certain HEIs were better than others at collecting this data from their students.

The numbers of students being categorised in the unknown category have been decreasing steadily from 348,460 (16.1%) in 2007/8 to 186,621 (8.0%) in 2010/11, which shows the HEIs are improving in collecting this data. However, 8% of the total dataset was still a high percentage of the total dataset and therefore a more detailed analysis of where the unknowns were occurring is beneficial.

The counties of Tayside, Avon and Coleraine all had very high percentages of unknowns with around one in five students having either their domicile or term-time address recorded as unknown. The University of Dundee, in Tayside County had 4,208 students recorded as unknown; The University of West of England, in Avon had 8,569 students record as unknown, while The University of Ulster in Colerain had 5,196 unknown students. While the LAs of Bexley, South Gloucestershire, Dundee City and Coleraine also had very high percentages of unknown students.

3.5 Differences in Student Migration Categorisation as a result of geographical level of analysis

When analysing the location of students using there term-time or institution address, the geographical level of analysis will only affect the geographic scale of the results. Therefore, when the purpose of the analysis is to locate where large numbers or percentages of students reside or study, then the geographical level of analysis to be used will depend on the level of spatial detail required in the research.

As with any level of geographical aggregation it is possible to miss important patterns that are occurring within the spatial unit. When using the county level geography it is possible to locate counties with large student populations, however varying patters within a county can be hidden. When analysing at the LA level geography, it was possible to locate LAs within counties that were highly influenced by student populations on a much smaller scale, although this caused a significant increase in the number of spatial units that needed to be analysed. As the number of spatial units are increased then so do the number of potential flows and interactions between areas, therefore, it is necessary that a balance be found between the level of geographical detail and the complexity of the analysis undertaken.

The choice of geographical level has a much more pronounced impact when the data is used to categorise students by their migration transitions in order to attend a HEI by using the student migration typology. One of the pros of creating a typology to measure the student migration is that it can be applied to a number of different spatial levels of analysis. However, as a result, the results produced when using the typology proposed within this chapter are highly sensitive to the level of geography used for the analysis. The following section will illustrate how the classifications of students into the student categories vary as a result of the chosen level of geographical detail by comparing the classifications of students for the county and LA levels of geographical aggregation.

The geography of the UK is split into many different types of administrative areas that differ in their size and structural composition (See Section 1.2). As a result of these differences in geographical areas, a phenomenon known as the Modifiable Areal Unit Problem (MAUP) arises. The MAUP was first identified by Gehlke and Biehl (1934) while Openshaw (1984) provides a comprehensive review on the early research on the subject area. Fotheringham and Wong (1991) simply define the MAUP as ‘the sensitivity of analytical results to the definition of units for which data are collected’. Wrigley et al. (1996) as read in Bell et al. (2002) state that there are two main aspects of the MAUP that are traditionally recognised; those of scale and those of zonation. Those of scale occur because an area may be divided into geographies with differing numbers of special units, while those of zonation occur because an area may be divided into the same number of units in a variety of ways.

The issue at play in this analysis refers to issues referring to those of scale. This chapter uses two differing levels of geography to divide the UK, Counties and LAs. Issues arise because these two levels of geography differ greatly in scale. The geographical level of counties divides the UK into 94 spatial entities whereas there are 408 LAs. Issues also arise because all of the LAs in the UK can be allocated into one of the counties and as a result when the level of analysis changes then the classifications of certain students can also change in turn.

To evaluate how the MAUP impacts on the student migration classifications, as created by the topology proposed in this chapter, it is necessary to refer back to Table 3-3. By comparing the numbers and percentages of the student migration categories for the county and LA levels, clear differences can be observed as a result of the geographical level used. One example of this is the significant increase in the numbers of commuting or distance learning students when using the LA level geography compared to the county level. There are also a much lower number of local students when using LA compared to county and there is a slight decrease in the number of internal and international student migrants at the LA level compared to the county level.

It is clear that the categorisation of students into the student migration categories using the typology was affect by MAUP. To gain a better understanding of how the categorisations changed as a result of the level of geography used, the student migration categories for county and LA were cross-tabulated as shown in Table 3-9.

Table 3-9: Analysing the effect of geographic scale on the Student Migration Typology Categories - Cross-tabulation of Student Categories by County and Local Authority

	Local Authority								
County	**Local Student**	**Commuter/Distance Learner**	**Internal Migrant**	**Migrant Commuter/Distance learner attending local HEI**	**Internal Migrant Commuter/Distance Learner**	**Int. Migrant**	**Int. Migrant Commuter/Distance Learner**	**Unknown**	**Total**
Local Student	210,999	331,726	40,689	5,187	12,366	*	*	*	600,967
Commuter/Distance Learner	*	472,717	*	*	12,501	*	*	*	485,218
Internal Migrant	*	*	563,537	*	72,591	*	*	*	636,128
Migrant Commuter/Distance learner attending local HEI	*	*	*	4,714	8,975	*	*	*	13,689
Internal Migrant Commuter/Distance Learner	*	*	*	*	49,034	*	*	*	49,034
Int. Migrant	*	*	*	*	*	265,442	58,212	*	323,654
Int. Migrant Commuter/Distance Learner	*	*	*	*	*	*	49,495	*	49,495
Unknown	*	*	*	*	*	*	*	186,621	186,621
Total	210,999	804,443	604,226	9,901	155,467	265,442	107,707	186,621	2,344,806

Source: Higher Education Statistics Agency (2012b)

Note: * = Structural Zero – Impossible Combination

31,902 students were no longer classified as an internal student migrant as a result of using the LA level of geography instead of county is caused by the smaller geographical size of the LA areas. This results in students that previously had their term-time and institute address in the same county now having these addresses in different LAs, which re-classifies them as internal migrant commuters/distance learners. For example, the largest amount of this change was where 9,777 students migrated to the county of Dorset and were classed as a student migrant, however when using the LA level, these 9,777 students had migrated to a term-address in Bournemouth LA but their institution address was in neighbouring Poole LA (as previously explained), so they were re-classified as an internal migrant commuters/distance learners. This decline in internal student migrants was offset by the 40,689 students that were classified as a local student but when the LA level was used the domicile was in a different LA to term-time and institute address and was therefore re-classified as an internal student migrant.

The largest change as a result of the different geographical level used was the drop in the number of local students and the increase in commuting/distance learning students when using LA. This is relatively intuitive as the geographical size of a LA is much smaller than a county and therefore the likelihood of all three variables being in the same LA is much lower than for the county level. 331,726 students were re-classified from a local student using counties to a commuter/distance learner using LA, which explains the large shifts in the two categories. Again the largest numbers of these changes are from those students in Scotland that commute into Glasgow City LA as mentioned previously. In the county level analysis these students would have been classified as local students within Strathclyde, however when LA was used a large number of students were re-classified due to domicile and institute addresses being in different LAs.

3.6 Local Authority Student Migration Area Classification

Analysing the student migration patterns of the whole UK at small geographical levels is complex. Unlike single-zone geographical data, migration data involve both origin and destination geographies and when considering the 408 LAs in the UK, this represents a possible 166,464 internal migration interactions. As a result, this makes any national level analysis extremely difficult to summarise or comprehend.

It is possible to analyse the spatial patterns of student migration across the UK by using the typology of student migration proposed above to create variables to measure the different levels of student migration and inputting these into a cluster analysis to create a student migration area classification. Area classifications group together geographic areas into clusters according to key characteristics common to the population in that grouping (Dennett and Stillwell 2011). Thus using an area classification will significantly decreases the complexity of analysing all the possible origin-destination flows in a full migration system.

There are a number of general purpose area classifications at differing spatial scales that exist in the UK's commercial and non-commercial sectors (See: Vickers et al. (2003), Vickers and Rees (2006, 2007), Dennett and Stillwell (2011)). The Office for National Statistics (ONS) have made general area classifications based on the national censuses, the most recent is called the 2011 Area Classifications (Office for National Statistics 2011a). However, the characteristics of migrant populations differ noticeably from those of the general population and as such general purpose area classifications built on data from total populations are not the best for analysing migration flows (Duke-Williams 2010; Dennett and Stillwell 2011). In response to this, the CIDER Migration Classification produced by Dennett and Stillwell (2011) provided a framework for analysing internal migration flows in Britain using a migration specific area classification. This migration classification is the only one of its kind produced for the UK and as a result the classification proposed within this section will be very similar to Dennett and Stillwell's but will only analyse student migration trends as opposed to all internal migration.

3.6.1 Methodology

There are a vast variety of clustering techniques that can be applied to finding groups within data and therefore can be used to find clusters within a vast amount of areas (See: Everitt et al. (2001), Bailey et al. (2000), (Vickers and Rees (2006), 2007); Dennett and Stillwell (2011)). However, as per the methodology used by Dennett and Stillwell (2011), the methodology proposed in this chapter will use the k-means algorithm to cluster the LAs into the student migration classifications.

The objects clustered in this analysis were the 408 LAs of the UK. These areas were clustered into groups of similar characteristics based on 13 variables that indicate the level of student migration associated with each LA. These variables are all measures of the different student types within the student migration typology proposed earlier, and are explained in Table 3-10.

Table 3-10: List of Variables used in the Student Migration Area Classification Cluster Analysis

Variable Name	Associated Typology of Student Migration Category	Description
Local Students	1 - Local Students	The number of Local Students in a LA
Commuters/Distance Learners - IN	2 - Commuter or Distance Learner	The number of students that attend a HEI in the LA but commute or distance learn from their domicile LA
Commuters/Distance Learners - OUT	2 - Commuter or Distance Learner	The number of students that live in their domicile LA but study in a different LA and therefore commute or distance learn
Internal Student Migrants - IN	3 - Internal Student Migrant	The number of students that attend a HEI and live in the LA and migrated in order to do so
Internal Student Migrants - OUT	3 - Internal Student Migrant	The number of students that migrated away from the LA to live and attend a HEI in a different LA
Local Migrant Commuter - TERM	4 - Migrant Commuter or Distance Learner attending local HEI	The number of students that live in the LA but attend a HEI in their domicile LA
Local Migrant Commuter - DOM/INST	4 - Migrant Commuter or Distance Learner attending local HEI	The number of students the attend a HEI in the LA which is also their domicile LA but live in a different LA
Internal Migrant Commuter - DOM	5 - Internal Migrant Commuter or Distance Learner	The number of students that migrated away from the LA to live in a different LA and then commute or distance learn to HEI in another LA
Internal Migrant Commuter - TERM	5 - Internal Migrant Commuter or Distance Learner	The number of students that have migrated away from the domicile to live in the LA but commute or distance learn to a HEI in a different LA
Internal Migrant Commuter - INST	5 - Internal Migrant Commuter or Distance Learner	The number of students that attend a HEI in the LA but commute or distance learn from a different LA that is not their domicile
International Student Migrant	6 - International Student Migrants	The number of international students that attend a HEI and live in the LA and migrated in order to do so
International Student Commuter - TERM	7 - International Migrant Commuter or Distance Learner	The number of international students that have migrated away from the domicile to live in the LA but commute or distance learn to a HEI in a different LA
International Student Commuter - INST	7 - International Migrant Commuter or Distance Learner	The number of international students that attend a HEI in the LA but commute or distance learn from a different LA that is not their domicile

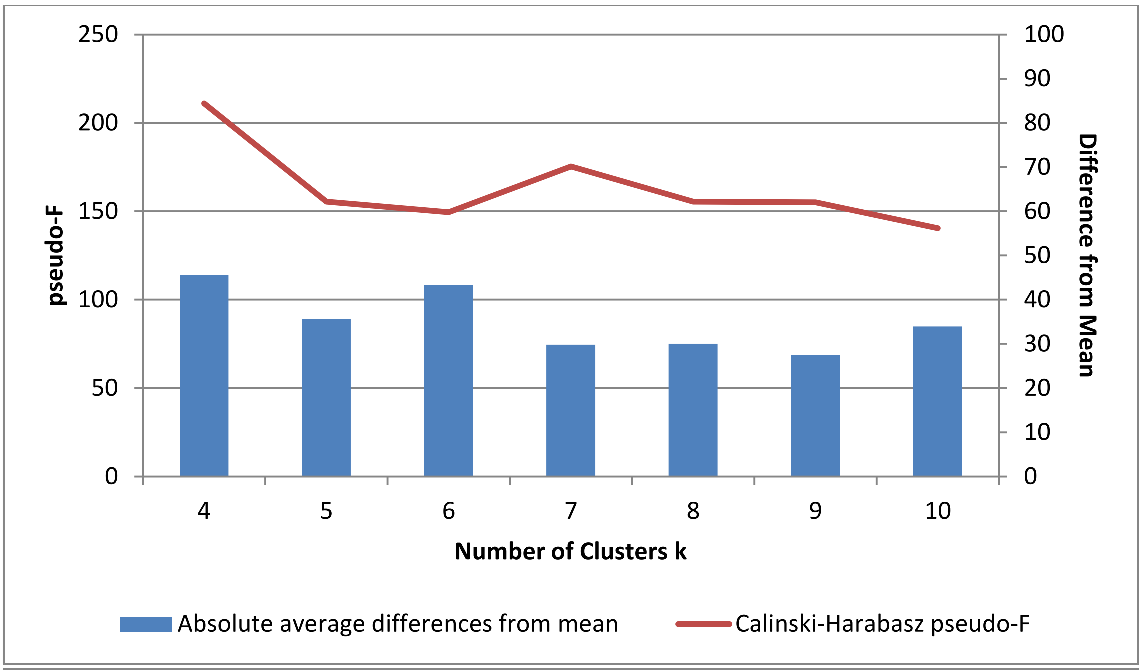
From the seven student migration categories 13 variables can be identified which directly relate to the origin and destination(s) of the certain student migration categories. For example, a ‘Local Student’ only impacts on one geographical area as the domicile, term-time address and institution are all in one areas and therefore is represented by only one variable. In contrast an 'Internal Migrant Commuter or Distance Learner' impacts on three locations as the domicile, term-time address and institution are all in different areas, therefore this category is represented by three variables. When you sum all the areas impacted by the 7 categories of student migration a total of 13 variables of student migration are created.

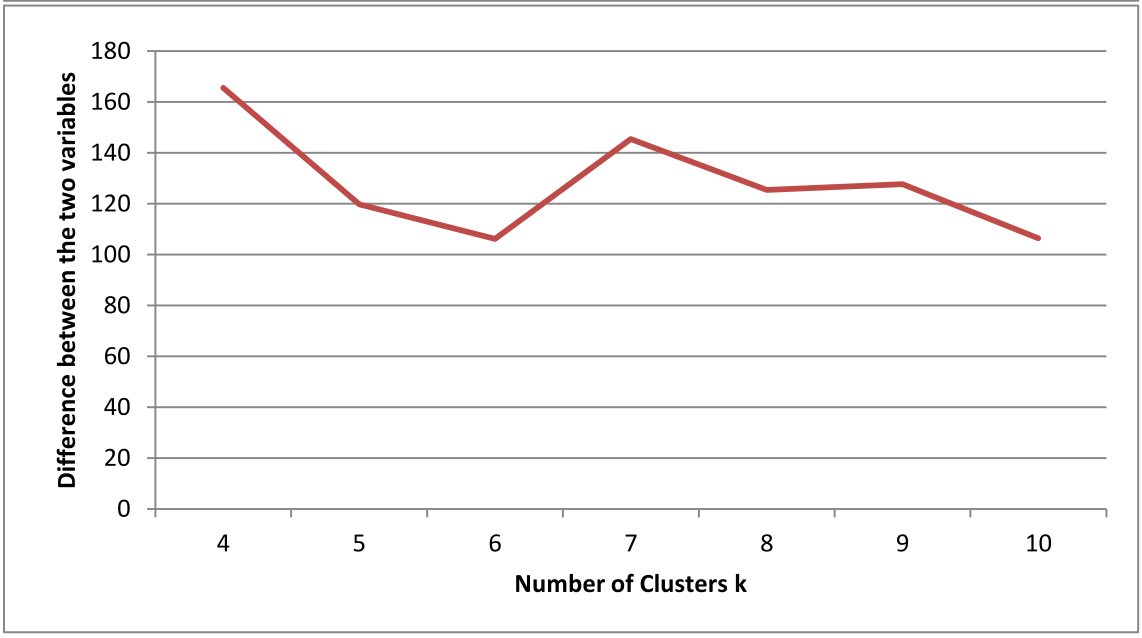
All variables included in the analysis are represented as rates, using a LA student turnover variable as the denominator. Rates are used instead of raw counts to ensure that the clusters are not biased by the differing sizes of student population associated with LAs with large overall populations. Each of the variables are also standardised into Z-Scores to account for some variables having varying standard deviations and as a result standardisation prevents the cluster analysis being dominated by certain variables.

The major decision when conducting a cluster analysis is the number of clusters to be included in the classification. As with many elements of classification building, the literature offers no definitive answer for deciding the most appropriate value of clusters (Dennett and Stillwell 2011). Everitt et al. (2011) discussed the problem of determining the number of clusters and described several stopping rules, one being the Calinski and Harabasz (1974) pseudo-F index, where large values indicate distinct clustering. By contrast, Dennett and Stillwell (2011) state that having evenly sized clusters is the most desirable outcome and, therefore, the smaller the values in the absolute difference in cluster size from the mean cluster size, the better the solution. However, it should be remembered that the cluster classification has to make intuitive sense in explaining the real world scenario regardless of the statistical tests of fit.

The Calinski-Harabasz pseudo-F values and absolute average differences from the mean cluster size are displayed in Figure 3-4. The best number of *k* clusters would be chosen as the best combination of a high Calinski-Harbasz value and a low absolute average difference from mean (the highest difference between the two indicators shows the best fit). From the indicator values shown in Figure 3-4, the two best number of k clusters are either 4 (difference of 165.5) or 7 (difference of 145.5). 7 clusters were chosen as 4 clusters was too small a number to clearly illustrate the complexity of the student migration movements occurring in the UK. And therefore, the final classification was made up of the *k7-means* clustering algorithm.

Figure 3-4: Creating the Student Migration Area Classification - Absolute average differences from mean cluster size and the Calinski-Harabasz pseudo-F values for clusters k 4 to 10





Source: Higher Education Statistics Agency (2012b)

3.6.2 Results: The Student Migration Area Classification

The final seven groups of the student migration area classification are now presented in turn and the numerical breakdown of the 408 LAs into the 7 area classifications are shown in Table 3-11.

Table 3-11: Breakdown of the number of Local Authorities into each of the Student Migration Area Classifications

Cluster Name	Number of LAs in Cluster	Percentage (%)
1 - Large University Settlement	37	9.1
2 - Medium University Settlement	19	4.7
3 - Commuting/Distance Learning HEIs	36	8.8
4 - Migrant Commuting Student Settlements	37	9.1
5 - Special Scenario Areas	13	3.2
6 - Sending LAs - Commuters	134	32.8
7 - Sending LAs - Student Migrants	132	32.3

There appears to be a large unbalance in the allocation of LAs into the different cluster groupings shown in Table 3-11, however this shows that the clustering is working well. Only 98 of the 408 LAs have a HEI and therefore the vast majority of HEIs cannot be recorded into one of top three categories.

Two maps illustrating the 408 LAs defined by their Student Migration Area Classification are shown in Figure 3-5 and Figure 3-6. Figure 3-5 shows a map of the UK with the size of the LA area on the map representing the physical geographical area of the LA. Figure 3-6 shows a cartogram of the UK created in ArcGIS in which area represents total population size (Gastner and Newman 2004). Each LA area is transformed to represent its population size and therefore this draws attention to areas of high population rather than in the conventional map in which large areas dominate even though these areas are sparsely populated (Dorling and Thomas 2011). Both Figure 3-5 and Figure 3-6 show the same Student Migration Area classification and each LA is coloured the same in both maps.

Each cluster will now be discussed in detail to best summarise the student migration profile that each cluster represents. A detailed breakdown of how each of the 7 groupings corresponds with the 13 variables shown in Table 3-10 can be found in Appendix B Table B-1, and a list of all LAs within each of the 7 area classification groups can also be found in Appendix B.

Cluster 1: Large University Settlements

Cluster 1 represents LAs that accommodate large HEIs. These HEIs attract large numbers of students to study at their institutions and this has a profound effect on the settlements themselves. This cluster is defined by having very high numbers of Student In-Migrants, both internal and international, as well as having high levels of local students and medium levels of students commuting from outside the LA to study at one of the LAs HEIs.

In Dennett and Stillwell (2011) CIDER Internal Migration Classification, one of the clusters was defined as Student Towns and Cities. The Large University Settlement category in this analysis appears to group LAs in a very similar fashion. This indicates that the student migration area classification is grouping LAs in a meaningful fashion.

Cambridge, Fife, Lancaster and Durham are the LAs best represented by this cluster. All are home to large HEIs with very high in-flows of internal and international students. In Cambridge, 81% of all students studying at a HEI in the LA were internal or international student migrants. Compared to the national average for all LAs of 37%, this value is extremely high. There were a few LAs that had weaker associations with the cluster definition; for example, Colchester and Welwyn Hatfield had high levels of commuting students but were classified in this cluster as a result of relatively high numbers of international student migrants.

The LAs in this student area classification closely match those found to be in the Top 10 LAs for Internal Student Migrants and International Student Migrants as shown earlier in Section 3.4.

Towards a typology of student migration:

Figure 3-5: The student migration area classification – Map of the United Kingdom Local Authorities

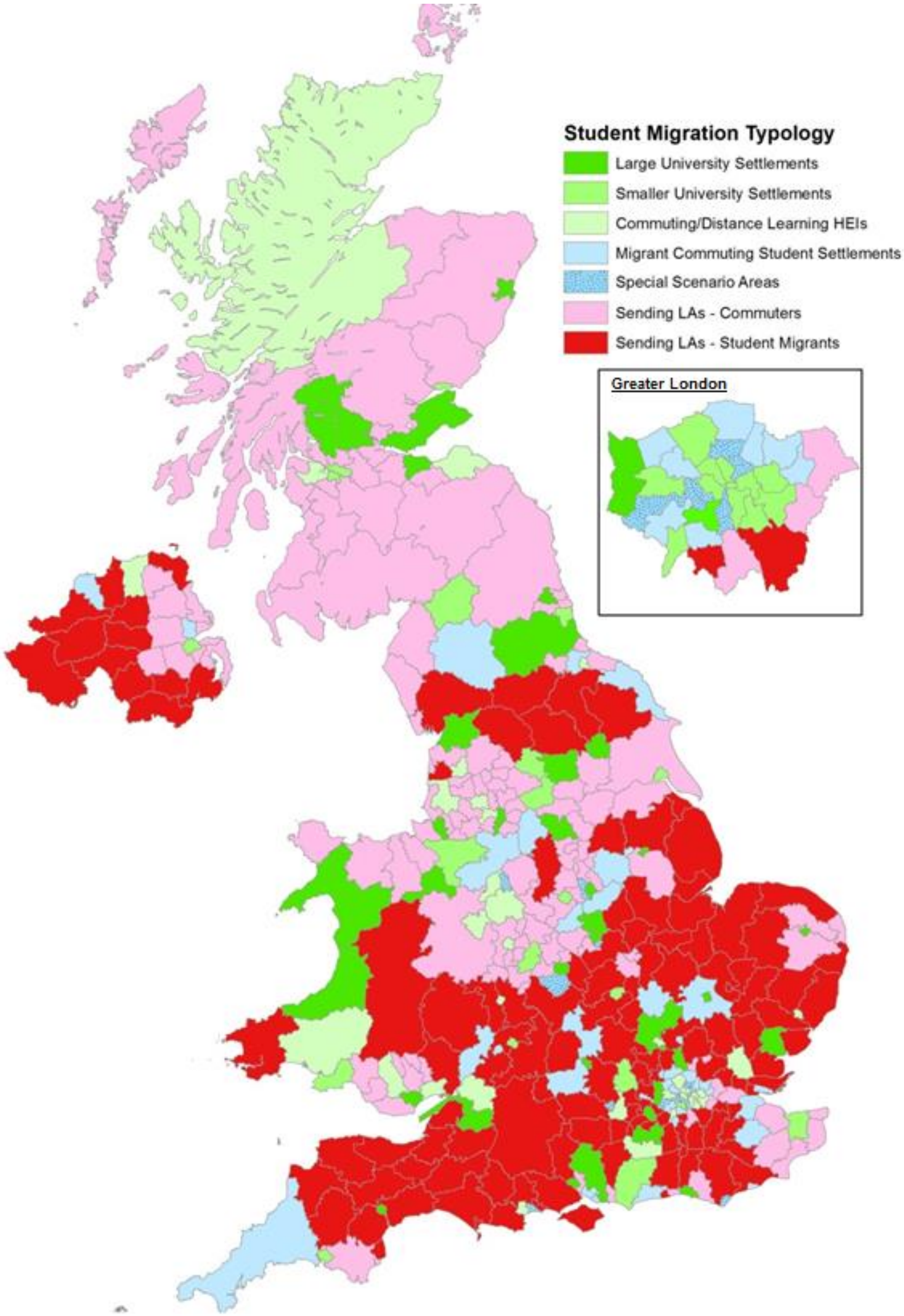
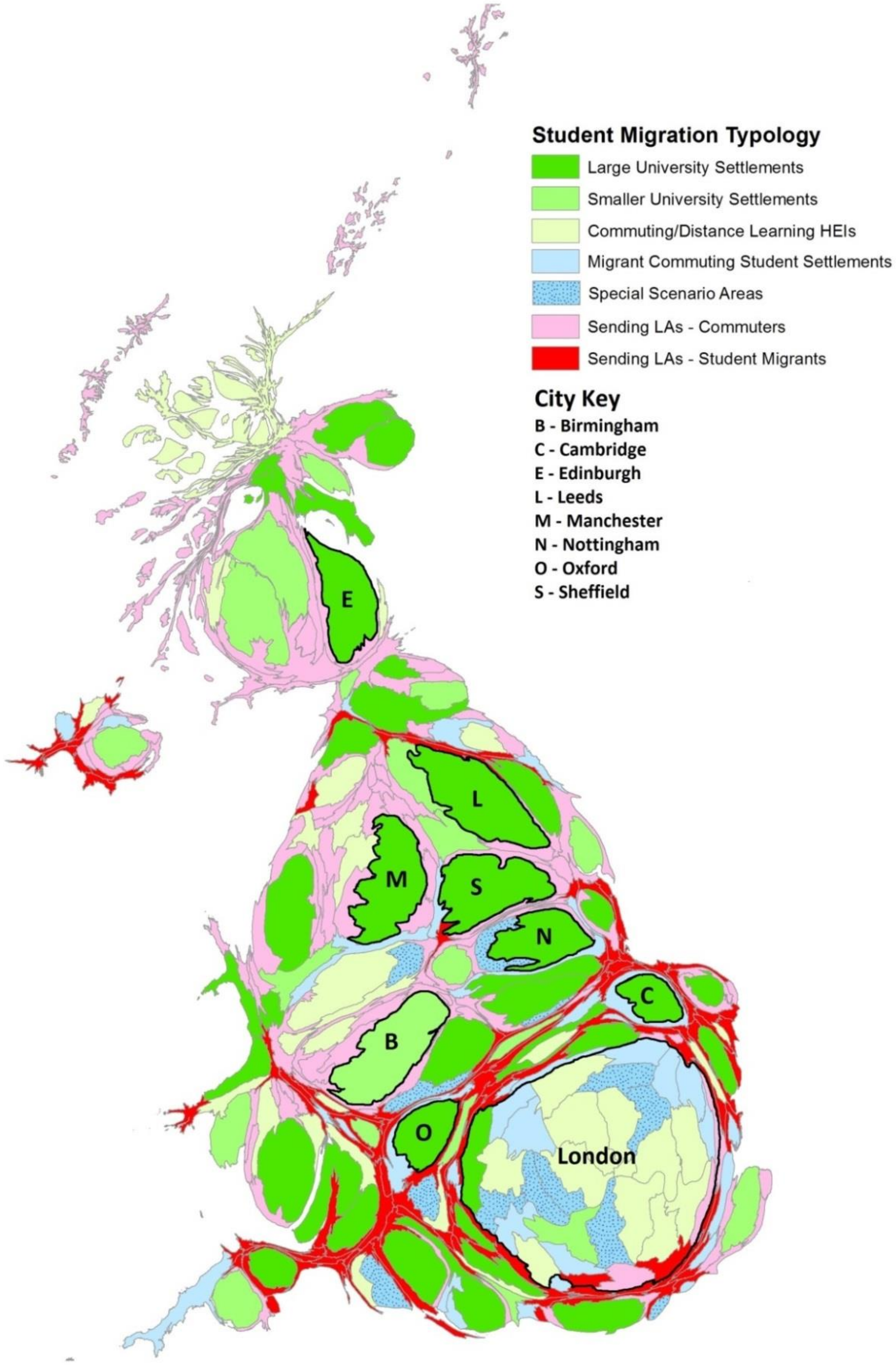


Figure 3-6: The student migration area classification – Cartogram of the United Kingdom Local Authorities



Cluster 2: Smaller University Settlements

Cluster 2 is similar to cluster 1, with the exception of having a stronger relationship with the number of local and in-commuting students and a weaker association with the number of internal and international in-migrants. The LAs in cluster 2 have a much higher representation of local students than cluster 1 and as mentioned previously in Section 3.4.1, previous research suggests that this can be linked to the ethnic composition of an area and the differing migration behaviour of students from differing ethnic backgrounds. It can also be noted that LAs in cluster 2 attract smaller numbers of internal and international migrants to study at the LA's HEIs, this may suggest that the HEIs in these LAs have less of a 'pull' effect attracting students to migrate to these LAs in order to study. LAs such as Birmingham and Glasgow represent this cluster well; these LAs have large local student populations as well as a large number of commuting students that reside in neighbouring LAs. As discussed in Section 3.4.1 and 3.4.2, these LAs were found to have strong associations with the percentages of local and commuting/distance learning students.

Cluster 3: Commuting/Distance Learning HEIs

This cluster represents LAs hosting HEIs that attract students to study at their institution but large proportions of the student population do not reside in the LA and therefore either commute or are distance learning. This cluster has a very strong association with the in-commuter variable as well as the internal migrant in-commuter variable.

The majority of these LAs are located within Greater London, where students study at universities located in the inner zones of the city but cannot afford to live in the same area (Mayor for London 2003) and therefore commute from LAs further out of the city. Just under half of students attending a HEI in the City of London were internal migrant commuters or distance learners, while the top two LAs for the number of internal migrant commuters were studying in the LAs of Westminster and Camden.

Outside of London, West Lancashire had the highest percentage of commuters or distance learners (81%), the majority of these commuters or distance learners resided in the neighbouring LAs of Sefton, Liverpool or Wigan. Edge Hill University is located in West Lancashire and attracts a large number of students from neighbouring LAs, however the transport links in this area are good and the distances travelled across LAs to Edge Hill University are quite short, which may be the cause of these surprisingly high numbers.

There are also a number of LAs in this cluster that are strongly correlated with a LA in cluster 5 due to special scenarios, often attributed to the modifiable area unit problem of where an administrative boundary is drawn (Bell et al. 2002). For example the University of Reading is located on the boarder of Wokingham and Reading, and as a result Wokingham is grouped in cluster 3 and Reading in cluster 5. However, if the boundary between these LAs was drawn in a different place this scenario would not occur.

Cluster 4: Migrant Commuter Student Settlements

This cluster groups LAs that attract student in-migrants but these students study in a differing (often neighbouring) LA. Therefore, this cluster has low values for all variables except for commuters-out and migrant-commuters term time location. LAs that have relatively high numbers of these students in comparison to other LAs and are therefore classified in this cluster, actually have relatively low numbers of these students in the LA. It could therefore be argued that the other student groups have a larger impact on the LA than the internal migrant commuting students. However, due to these LAs having high numbers of migrant commuting students in comparison to all the other LAs, this is why they are clustered into this student migration category.

The LAs of Rushcliffe (neighbouring Charnwood [Loughbourgh University] and Nottingham), Arun (neighbouring Chichester and in close proximity to Brighton) and Epsom & Ewell (neighbouring Kingston-upon-Thames) fit this cluster well as they have very high numbers of internal migrant commuters. However, the overall small number of internal migrant commuting students in the UK (only 6%) makes this cluster hard to define accurately.

Cluster 5: Special Scenario Areas

This cluster picked up a very small amount of LAs and is very similar to cluster 4 in its characteristics. However, for all the LAs in this grouping there is an obvious explanation as to why these LAs had been grouped together. All LAs in this group had very large numbers of internal migrant commuters and most can be explained by the modifiable area unit problem (Bell et al. 2002).

We will illustrate the example of the special scenarios with two case studies: Bournemouth University and the University of Reading. Bournemouth University is actually located in the LA of Poole and therefore a large majority of its students are recorded as migrant commuters because they reside in a term-time address in Bournemouth and travel the very short distance to the university just over the border in Poole. The same scenario is present in Reading where the university and a large amount of its halls of residences are in Wokingham. However a large number of students chose to reside in neighbouring Reading where the rent is considerably cheaper.

Cluster 6 and Cluster 7: Sending LAs – Commuters and Student Migrants

These two clusters group together a large number of LAs, of which the vast majority have no HEI located there. Cluster 6 groups together LAs in which the majority of students from these LAs do not migrate away but commuted or were distance learning at a HEI in a different LA. Whereas, cluster 7 groups LAs in which the majority of students did migrate away from the LA in order to study at a HEI in a different LA and were therefore student migrants.

These two clusters make up the large majority of LAs in the UK and this can be seen in Figure 4 where the map is dominated by these two categories. However, it must be remembered that these LAs are often sparsely populated and therefore do not account for a large proportion of the total population. This is why the overall pattern of Figure 3-6 is vastly different to that of Figure 3-5, because Figure 3-6 is weighted to represent an area's population instead of geographical area.

There are a number of LAs in cluster 4 that could easily have been grouped in cluster 6. However, they have been grouped into cluster 4 because they had slightly higher numbers of internal migrant commuters than those LAs in cluster 6 but overall they show the characteristics of a sending LA.

3.6.3 Spatial Patterns in the area classification

There are some clear observable patterns in the two maps (Figure 3-5 and Figure 3-6) that visually display the student migration area classification. The first three clusters clearly pick out the LAs that are home to larger towns and cities and are the location of many HEIs. As a result, these areas had a net gain of students, and these gains are greater depending on the size and prestige of the institution as well as the size and reputation of the town or city. In contrast, the smaller LAs and those without a HEI experienced a net loss of students to areas with institutions.

The areas of net loss were clustered into two distinct groups; those areas where the majority of people migrated away in order to attend a HEI (as shown in red) and those areas where the majority of people stayed living in the LA but commuted or were distance learning at a HEI (shown in pink). When observing Figure 3-5, a clear spatial difference is visible between LAs that were dominated by those who migrated away and those that were dominated by commuting or distance learning students. This divide clearly shows a North-South difference in the student migration transitions within the UK (See: Dorling (2007, 2011, 2012); Dorling and Thomas (2011) and Thomas and Dorling (2011), for a further discussion of the demographic North-South divide in the UK). Those LAs in the South were mainly sending student migrants (red) and those from the North were mainly sending student commuters (pink).

This finding has prompted some further analysis to investigate the student movement differences between the North and the South in the UK (a map showing the definition of the North and South Region in the UK is shown in Appendix A Figure A-2). Of all those students domiciled in a LA in the North only 6.6% migrated to the South to study at a HEI. This is in vast contrast to moves in the opposite direction where 18.4% of students that were domiciled in a LA from the South migrated to study in a LA in the North. This North-South difference suggests that those students from the South are migrating more than their Northern counterparts in order to attend HEIs, whereas those from the North show a much greater association with commuting to study or distance learning.

This raises the question as to whether this pattern is down to differences in student choice or if other factors are at play such as the unobservable factors highlighted in Figure 2-2, for example, differences in financial support in order to afford to live and study away from the domicile location. This differential in north-south student migration trends can also be discussed within the topic of economic push and pull factors (Lee 1966) that influence migration. These high flows of students from the South to the North could be attributable to them being ‘pulled’ by the economic advantages that can be gained from studying in the North compared to the South. Studies have shown the cost of living is significantly lower in the North compared to the South - £3,000 per year more expensive in the South compared to the North (Poulter 2014; Office for National Statistics 2016) - with the main driver of this being housing and rental prices. Therefore, students may be encouraged to migrate to the North to study to take advantage of the lower costs of living as these economic pulls may out-weigh the intervening obstacles of the cost of travel and relocating in order to study and the economic pull comparison of staying and studying in the South. This explanation may also hold for why fewer students from the North migrate to study in the South. The increased living costs in the South compared to remaining in the North, coupled with the intervening travel and relocation costs result in a negative economic impact of studying in the South for Northern domiciled students and hence very few students from the North migrated to the South. This does raise questions over the geographical distribution of top rated HEIs and whether or not access to top HEIs is equal across the UK.

A Joseph Rowntree Foundation Report indicated that this student migration from South to North was one factor that was contributing to offsetting the uneven growth of cities across the UK and was helping tackle city decline and in essence offsetting the growing North-South Divide (Pike et al. 2016). The report found the 10 of the poorest cities in the UK were in the North but two key core Northern cities, Manchester and Newcastle, were excluded from this list as they host high-level service facilities and both hold strong attractions for students from both the North and the South (idem). Universities have played an important part in supporting economic recovery in some Northern cities through the growth of knowledge-economy jobs and student population numbers (idem).

These results suggest that the migration patterns of students are not equal across the UK. Students from LAs in the South appear to be more likely to be student migrants than their peers from LAs in the North. With the UK government promoting widening participation and equal access to HE, government policies should be adapted in an attempt to eradicate this North-South divide in student migration practices. There is a rising cost of HE in the UK and, for equal access to be achieved financial support should be in place to help those students with the ability to migrate in order to attend a HEI of their choice and of the reputation that matches their academic ability, regardless of where in the country they originate from.

3.7 Chapter Summary

The complexity in this study derives from the need to interact three different location variables - domicile, term-time address and institution address - in order to accurately categorise and measure the possible migration transitions of people entering HE in the UK. Therefore within this chapter a unique typology that categorises all students in the UK into one of eight mutually exclusive and exhaustive student migration categories was designed. This typology was then used to create 13 variables to measure student migration, which were subsequently incorporated into a cluster analysis that created a new student migration area classification by local authority. The area classification grouped all 408 UK LAs into 7 area classifications categories which were characterised by the student migration characteristics of that area. This enabled the spatial analysis of complex student migration patterns for the whole of the UK in a succinct and manageable fashion.

Using the typology of student migration at the LA level the data indicate that around 25% of all students were internal student migrants, 9% were local students, while 35% stayed in the LA but studied elsewhere, suggesting they were distance learners or commuters. When international students and unknowns were removed from the data, the typology shows that 57% of all UK domiciled students, remained living in the same domicile LA and therefore did not migrate to attend a HEI. This shows that when the migration patterns of students are accurately measured, the previously assumed notion that the traditional transition into HE of migrating away from the parental home to study at a HEI in a far-away location (Chatterton 1999; Patiniotis and Holdsworth 2005; Chatterton 2010), is actually not the majority transition experienced by HE students in the UK. Only 38% of all UK domiciled students followed the traditional pattern and migrated away from their domicile and resided close to the HEI they attended. The results also indicate that a large percentage of UK domiciled students now commute or distance learn as their term-time address was not in the same area as their HEI.

These findings could be the result of the major increase in student numbers and the increased participation of 'non-traditional' students in HE (Chatterton 1999). The rising cost of tuition and living could also be associated with this changing model of student migration transitions currently being observed in the UK. The result of fewer students undertaking the more traditional migratory transition away from the parental home towards university will have an impact on the assumed transitions to adulthood and life-course transition theory in the UK. While this shift in the number of student migrants and the number of local students raises questions about widening participation policy; for example, are students able to attend the HEIs their ability levels entitle them to attend or are students simply attending the HEI in the closest proximity to restrict the overall cost of obtaining a higher educational degree? This research could certainly be linked to the increasing costs of tuition and living in the UK and it demonstrates a need for the government and HEIs to adapt polices, taking into account these observed migration (or non-migration) transitions into HE as opposed to the previously assumed situation of vast student migration.

These results will also be of interest to local planners and housing authorities, because the previously assumed to be true patterns of student migration have been shown to not actually be the case. Lower proportions of students are migrating and therefore the question arises are less students requiring student housing? A new body of research by Kinton et al. (2014) discusses the impact of 'Destudentification' where reductions of a student population in a neighbourhood that was previously renowned for its high student numbers leads to the 'Social, cultural, economic and physical decline' (Smith 2008: 2552). The results from this work show student migrants were not the norm in 2010/11 in the UK and therefore the process of 'Destudentification' may well be seen across a large number of the LAs identified in the student migration area classification as now attracting more commuting students rather than student migrants.

The student migration area classification also highlighted a distinct difference between the North and South regions of the UK in regards to the student migration patterns. Students from Southern LAs were much more likely to migrate than their Northern peers, who tended to commute or distance learn instead of migrating to attend a HEI. This overlaps with the economic North-

South divide literature discussion in the UK and can also be linked to the economic benefit (economic ‘Pull’) from living in the North compared to the South regarding the cost of living and the lower rental prices. This again is an important finding in reference to the equal opportunities and access to HE, as well as the widening participation policy of the UK government. If there are such clear spatial differences then these need to be addressed in the government HE policies, as well as the HEIs recruitment strategies.

This study could be extended by analysing an extensive longitudinal dataset that held detailed population data on all students for a longer period of time in a way in which the impact of increasing tuition fees could be accurately investigated. This chapter only provides a cross-sectional analysis of the most recently available time period, while a longitudinal analysis would enable the analysis of changing student migration trends over time. Another limitation of this study is the inability to answer why these geographical differences are present and why students experience difference student migration transitions. As a result, further work which will analyse how a student’s characteristics impact on the migration transition experienced is conducted in Chapter 4. Further extensions to this work could include a qualitative evaluation of the opinions of students from the North and South regions to determine whether students from the North commute out of personal choice or because of other influencing factors that were not overserved.

The key methodological contribution of this chapter is the creation of a typology that accurately records the complexity of the student migration process and the creation of a new and unique student migration area classification. The possibility for the typology proposed in this chapter to be used on any given level of geographical data and be adapted and used on any dataset that contains the locational details of students or any other relevant sub-population is another benefit. When doing so, it is important to remember the necessity to be able to distinguish between the different types of student movements, migrations and motivations as demonstrated in this chapter. It has also been highlighted how extremely difficult it is to analyse the complex spatial patterns of a whole migration system due to the large amount of possible interactions between the origins and destinations. This chapter has illustrated that when there are three locations to be considered in a migration system, such as in the student migration system in the UK, the complexity increases again. The creation of the student migration area classification within this chapter has shown the value of using such area classifications for analysing whole system processes. As a result, clear spatial differences in the UK student migration behaviour have been identified which would not have been possible without the creation of the typology or the student migration area classification.

4. Migration Choices of Students Entering Higher Education in the UK: What are the Impacts of Personal Characteristics, Institution Attended and Course Studied?

4.1 Introduction

In Chapter 3, a typology of student migration was created that enabled the accurate measurement of student migration. This was then used to analyse the main patterns of student migration for the UK and how these differed spatially across the UK. The focus of the analysis now shits to focus on the individual student and how the student migration patterns differ on the individual level. Therefore, the aim of this chapter is to gain an in-depth understanding of how student migration transitions of people entering HEIs in the UK are impacted by a student's characteristics, the course they study and the institute they attend.

In Chapter 2, it was discussed that in recent decades the student population in the UK has been steadily increasing and now stands at a record high. This expansion in student numbers was driven by government policy and the major restructuring of HE in the United Kingdom in the early 1990s (HMSO 1992; Christie 2007). These changes were implemented under the premise of building a workforce capable of sustaining the shift towards a knowledge economy and promoting economic growth. However, the expansion of HE was not primarily driven by the desire to increase the numbers of people participating in HE. It was also underpinned by the desire of government and policy makers to bring a more diverse set of non-traditional students into universities as a means to counter problems of social exclusion and poverty within the UK (Christie 2007).

This expansion in student numbers also leads to the question of where all these extra students have gone to take up their studies (Allinson 2006) and do all these extra student experience similar student migration transitions. It has been well documented, that in the past in the UK, a long move away from the parental home to take up your HE studies was seen as part of the university experience and an ideal transition into adulthood for a large percentage of students, especially those from white, high social class, backgrounds (Nicholson and Wasoff 1989; Boyle et al. 1998; Belfield and Morris 1999; Chatterton 1999; Allinson 2006; Gibbons and Vignoles 2012). However, it was found in Chapter 3, that in the UK these traditional moves away in order to attend a HEI were not the majority transition for students in the 2010/11 academic year at the local authority level and the group in the majority were those students who did not migrate but travelled to study at their HEI or were distance learning.

The expansion of student numbers and representation of non-traditional students in the HE population in the UK and the change in the majority patterns of student migration found in Chapter 3, raises the question of whether the two trends are interlinked. It was also found in Chapter 3 that there were significant geographical differences in student migration patterns. Students originating from the South were found to be much more migratory than their counterparts from the North. Therefore, if student migration patterns differ as a result of geographic location, this raises further questions regarding if these differences are purely the result of geography or if more detailed and interlinked factors at the individual level are at play. These factors therefore provide the rationale to further investigate student migration patterns in the UK by investigating the individual level impact of a student's personal characteristics on the student migration outcome.

Thus far, little quantitative research has analysed the differences in the student migration transition experienced as a result of an individual's personal characteristics, the course they study and the institute they attend. Therefore, the analysis in this chapter will investigate if any patterns or associations within the data occur and whether or not students of differing characteristics do experience differing student migration outcomes.

The analysis presented in this chapter is conducted by implementing three different techniques of evaluating a student's migration transition. The use of three different techniques is a reflection on the complexity of the concept of student migration and that the student migration can be measured in different ways. A migration could be seen as a simple binary do they migrate, yes or no. Migration experience could also be measured by the distance they move and so forth. The concepts and issues regarding the complexity of understanding and measuring migration were discussed in detail in Chapter 2. Another reason for using multiple techniques for analysing the student migration is to add statistical robustness to the findings, as the range of techniques will allow for checking whether the findings were a result of the measure used or if findings were consistent across techniques this would add confidence to those findings.

The first technique uses a logistic model to evaluate what variables impact on the probability of whether a student migrated or stayed local to attend a HEI. The second technique uses distance travelled by the student in a Tobit regression model to investigate how distance is associated with the variables within the data. The final technique uses the typology put forward in Chapter 3 that categorised all students that attended a HEI in the UK into one of eight categories which illustrates the type of transition the student experienced in-order to attend a HEI. This typology is then used as the dependent variable in a multinomial logistic regression model to evaluate what variables impact on a student's migration category.

These three techniques are used and reported in this order as each technique advances the information used to understand the type of migration transition undertaken to attend a HEI. Three techniques are used to show the commonalities within the methods and the data as well as to corroborate the outcomes of three different dependent variables. This analysis is then used to investigate if inequalities or dissimilarities are present in the migration patterns of students into HEIs in the UK by answering the following main research questions:

1. How does a student's social background, ethnicity or gender impact on the migration outcomes experienced in order to attend a HEI?
2. How does the course studied or institute attended impact on the student's migration outcome?

The findings to these research questions will be of great policy interest to government and policy makers but also to the HEIs who need to been seen as providing equal opportunities and equal access to all applicants and meeting university entry quotas.

The remainder of this chapter takes the following structure. An exploration of the previous research conducted in the subject area of student migration and social inequalities is conducted in Section 4.2. A description of the data used in the analysis and an argument for further in-depth investigation are put forward in Section 4.3. In Section 4.4, the results of the preliminary analysis are presented and provide the evidence that supports the need to use statistical modelling techniques in the analysis. The methodologies of the analysis are then explained in detail in Section 4.5 while the results of the different methodologies are presented in Section 4.6. Finally, the chapter and its findings are concluded in Section **Error! Reference source not found.**.

4.2 Migration into Higher Education: What do we already know and why do we need to know more?

An overview of the previous research and a conceptualisation of the student migration process in the UK as a whole were provided in Chapter 2. It was shown that the student migration process is impacted on and affected by many differing but inter-related factors. Some of these factors have been highlighted to be important in previous qualitative and quantitative studies, while some factors were conspicuous in their absence within previous studies. The overriding themes of this pre-existing research focused on the differences and inequalities observed in attainment, HE participation and the migration and housing choices of students in the UK and a general review of this research can be found in Section 2.4.

In short, the majority of previous research indicated there were significant differences in participation rates in HE as a result of an individual's socio-economic background, ethnicity and gender. However, the aim of this chapter is to focus solely on what factors impact on the student migration outcome and not HE participation. Therefore, the current chapter builds on the previously mentioned research by analysing if the findings of inequalities that were visible in participation trends are also visible in the migration transitions and distance travelled to attend a HEI. This will be conducted by applying a series of statistical models to try and capture and measure the impacts of as many of the observable variables depicted in Figure 2-2 that can be acquired from the available data on the student migration outcome of the individuals.

There were also a number of previous studies that did analyse differences in the migration transition in a qualitative fashion which were also explained earlier in Section 2.4. Patiniotis and Holdsworth (2005) found that the decision to migrate was strongly linked to socio-economic class. Holdsworth (2009) found that student migration had become an elite practise mostly undertaken by those from richer and higher socio-economic backgrounds. Khambhaita and Bhopal (2013) found that female students from Indian, Pakistani and Bangladeshi ethnic groups were all more likely to stay at home and not migrate for HE relative to their white counterparts. Faggian et al. (2006) found that all non-white UK students were much less likely to migrate to attend university compared to their white peers. Further work by Faggian et al. (2007b) confirmed the previous findings, that those with more human capital and from higher socio-economic groups were more likely to migrate. And an investigation by McNay (2012) found that black and minority students were more likely to study closer to home than their white student peers.

The analysis of student the distance travelled by student migrants is another area of previous research that is related to the current study. Gibbons and Vignoles (2012) provided new quantitative evidence on the impact of a student's characteristics, the distance from a HEI and there likelihood of participating in HE. Geographical distance is a potential barrier to students thinking of going to university, because of the direct monetary, informational and psychological costs involved in relocating or commuting a long way from the family home. These costs are often said to have an important influence on university choices, particularly for low income families, ethnic groups with cultural incentives to stay at home, or others for whom distance related costs could be particularly high. It is for these reasons that such intervening obstacles are important in the student migration decision process and why they were included in the conceptual framework in Section 2.5, Figure 2-2 and in the analysis presented later in this chapter. Students with different backgrounds and abilities choose different types and qualities of universities, and the spatial distribution of both university types and student characteristics is not uniform across the country.

Gibbons and Vignoles (2012) used a composite linked dataset from a number of administrative sources. The core dataset was the 2002 Pupil Level Annual Census (PLASC), which contains details on all pupils in England sitting their age-16 exams (when compulsory education ends). This dataset was then linked to the HESA Student Record to enable to researchers to investigate the subsequent decision of each pupil to enrol (or not) into HE and then subsequently calculate the distance between the school and the HE attended (if attended). Gibbons and Vignoles (2012) found that university intakes are, on average, skewed towards those students whose parents live relatively close-by to the HEI. This in itself was likely to explain the apparent over-representation of some ethnic minority groups in inner-urban universities. Some ethnic groups – especially Bangladeshi and Pakistani women – appear to be considerably more sensitive than others to distance, and possible reasons for this have been documented elsewhere. Students from lower socio-economic backgrounds differ too in the sensitivity of their choices to distance, with the sensitivity increasing as income and occupational status decreases. At the same time, students from low income/status backgrounds have a low probability of attending high research-quality institutions relative to their equally qualified peers from better-off backgrounds.

The work by Gibbons and Vignoles (2012) is extremely valuable and analyses the same facets under investigation in the current chapter (migration distance to HE). However, the study uses pupils aged 16 in 2002 and therefore the findings do not encompass the recent significant changes in the HE sector regarding expansion in numbers and the results are now quite dated. This study also focused on the impact on HE participation and the types of HEI attended, as opposed to the study within this chapter which focuses on the impact on the student migration outcome. Although very similar these slight differences in the focus of the studies impact on what findings can be taken from the different pieces of research. It will therefore be of great interest to see if the results presented later in this chapter support or contradict the findings of Gibbons and Vignoles (2012) with regards to the differences between students from different backgrounds and the distance they migrate in order to attend a HEI.

The discussions in the preceding paragraphs of this section and in Chapter 2 have highlighted the prior research in the field of student migration. However, a truly quantitative analysis of socio-economic, ethnic and gender differences on the specific migration transitions experienced by students was lacking. These previous studies were largely qualitative and survey based, and often focused on specific case study areas. In contrast, the current study aims to supplement this research by providing a solely quantitative analysis of the whole UK student population by using population data that has not been analysed for this purpose before. As a result, the current study will either support or critique the findings of the previous research that was conducted by using differing techniques and data sources and by comparing there results to the findings produced here.

The following section of this chapter introduces these datasets in more detail, with the methodology and results sections to follow.

4.3 Data

The data source used in this chapter is the Higher Education Statistics Agency (HESA) Student Record Data, as described in detail and used in the analysis within Chapter 3. However, for the purpose of the analysis undertaken in this chapter, students that migrated from overseas to attend a HEI in the UK are omitted. The aim of this chapter is to investigate how a student's background and ethnicity impact on the migration decision of that student attending a HEI in the UK. Due to the fact that all international students have migrated in order to study in the UK and due to the poor information collected on international students regarding their socio-economic background and ethnicity, it was best to omit this group from the analysis.

After removing the International and Open University students, the remaining data consists of 1,797,492 students that were enrolled at a UK HEI in the 2011/12 academic year (Higher Education Statistics Agency 2014b). The data consists of three main groups of variables. Variables on the student's locations represent the outcome variables of interest used in this study. While variables' representing a student's individual characteristics and information regarding the higher educational system are employed as the explanatory variables. These three main groups of variables are explained in turn below.

The dataset and the variables were chosen with the intention of capturing as many of the factors that influence the student migration decision as possible. An attempt to conceptualise all the possible influencing factors and there interlinked relationships were illustrated in Figure 2-2. Only those variables within the dashed rectangle on the left hand-side of the diagram are investigated and analysed in this chapter. The reason for this was the availability of data. The HESA Student Record Data is unique in that it provides population data on every student enrolled in HE in the UK and contains a large set of covariates to be analysed. However, as can be seen from Figure 2-2, there are also many other factors that contribute to the outcome of interest, student migration, that were not captured or measured in the data source and were therefore not included in the analysis. As a consequence of this, when analysing the results of this research later in the chapter, it must be considered that some of the findings and patterns may be attributable to some factors that are present in Figure 2-2, but are unobservable variables in the analysis.

4.3.1 Students' locational variables

The variables on the student's locations were used and explained in detail in Chapter 3. The variables and the proposed typology from Chapter 3 are also used in this analysis chapter because they can be used to measure and record the outcome of interest, student migration. By analysing the students domicile, term-time address and institution address simultaneously, an accurate understanding of the student migration undertaken for each individual is recorded. It can be seen from Figure 2-2, that the student migration decision is the key outcome of interest of this study and that there are many interlinking factors that influence this outcome. In this chapter the outcome of interest, student migration, is measured using three different outcome variables, all of which depict student migration in a slightly different fashion. Each of these three outcome variables are now discussed in turn.

Internal Migration

The first outcome variable is a simple binary indicator depicting whether the student had migrated internally in order to attend a HEI or not. An internal migration is recorded if a student's term-time address was in a different LA to their domicile address. Therefore, from the Student Migration Typology proposed in Chapter 3, students in categories 3, 4 or 5 are classed as internal migrants, while those in 1 or 2 are not. This indicator allows for a simple comparison of how the student characteristics differ between student migrants and non-student migrants.

Distance

The second outcome variable is a continuous numeric variable that measures how far the student travelled in kilometres to attend a HEI. This was calculated using ArcGIS software, in which the distance is calculated between the LA centroid for the three locational variables; domicile, term-time and institutional address.

Due to the different combinations of the three locational variables, as explained in the typology of student migration (Chapter 3), three different measures of distance were calculated;

- *Migration Distance*: the distance between domicile and term-time address – the two variables represent where a student resides and therefore the transition between these two variables is deemed to represent a migration as this movement is deemed to be relatively permanent (duration of the university term).

- *Commuting Distance*: the distance between term-time address and institute address – the two variables represent the distance travelled from term-time residence to place of study, therefore it is assumed this distance is travelled on a day by day basis and does not represent a migration.

- *Total Distance*: the sum of Migration and Commuting distance – this represents the total distance a student has travelled to gain access to higher education.

The analysis in this chapter focuses on the Total Distance variable. The outcome of interest is how far a student travels in order to attend a HEI and therefore total distance can be used to capture all movements made by an individual. As a result, a student's total distance is 0km if no movement is made and the student attends their local HEI. In contrast, all students that do not attend their local HEI have a distance figure greater than 0km, representing the distance they commute, migrate or a combination of both.

Student Migration Classification

The third outcome variable is a categorical variable that categorises the type of student migration decision taken by the student. As illustrated in Chapter 3, the migration outcome is often more complex than that illustrated by a binary migration; yes or no. As a result, the factors that influence an individual's probability of being in one of the different types of student migration is analysed using the same typology put forward in Chapter 3.

In the current chapter the ‘International’, ‘Migrant Commuter or Distance Learner attending a local HEI’ and ‘Unknown’ categories are omitted. The international student category is omitted because international students are omitted from this chapter completely as explained earlier. The migrant commuter or distance learner attending a local HEI were also omitted as they represent a very small number of the total student population (0.4%). The unknown category is also omitted as estimating what characteristics influence a student being in the unknown category is not relevant to the research or to any public policy, as students are categorised here as a result of data recording errors often at the institution not the individual level.

4.3.2 Students’ individual characteristic variables

The HESA Student record data contains micro level characteristic variables which are used in this chapter to explore any associations with the three outcome variables discussed above. The following sub-section explains how these variables are recorded by HESA and how (if any) manipulation or merging was conducted by the author.

Ethnicity

Students domiciled in England, Wales, Scotland, Northern Ireland, Guernsey, Jersey and the Isle of Man were required to report their ethnic origin for the HESA student record. The coding frame adopted by HESA is that recommended by ONS for UK-wide data collection and uses the following ethnic category groupings:

- **White** includes White and Irish Traveller.
- **Black** includes Black or Black British - Caribbean, Black or Black British - African, and other Black background.
- **Asian** includes Asian or Asian British - Indian, Asian or Asian British - Pakistani, Asian or Asian British - Bangladeshi, Chinese, and other Asian background.

- **Other** (including mixed) includes mixed - White and Black Caribbean, mixed - White and Black African, mixed - White and Asian, other mixed background, and other ethnic background.
- **Unknown** includes not known and information refused.

As discussed in Chapter 2, many previous studies have found links between an individual's ethnicity and many factors closely related to student migration. Ethnicity is clearly one of the key influencing factors in the student migration process as shown in Figure 2-2 and as a result, it is imperative to include ethnicity into the analysis.

The ethnic groups provided by HESA in the dataset are very broad and differences within these groups will be masked as a result. Because ethnic differences are not the sole interest of the research in this chapter or this thesis, investigating differences between these broad ethnic groupings will provide satisfactory findings, as this will provide a useful insight into the overall ethnic patterns within student migration. For a more detailed understanding of the ethnic differences in student migration behaviour in the UK, this work could be further extended by looking into data that break down ethnicity into smaller and more defined groupings. However, these more detailed ethnic breakdowns were not available to use for the research conducted within this thesis.

Social Background

It has been discussed in Chapter 2, that a longstanding body of literature exists across the developed world and the UK that indicates that an individual's social economic status impacts on many factors that are inter-related to the student migration decision. Social economic status has been found to impact on the choice of institution to attend as individuals tend to apply to institutions that match their personal background. It has also been found the social economic status will impact on the level of financial support which will also influence on a student's ability and decision to migrate to study or not.

The participation in HE of individuals from lower social economic backgrounds is also a key political and policy relevant topic in the UK at present. The widening participation act was implemented to encourage equal access into HE based on ability regardless of your social economic background. Despite these policies being in place, the hypothesis here is that those students who were from the highest social-economic classification would have the most advantageous background in terms of financial support and encouragement to attend more prestigious HEIs and therefore be more likely to make a migration to attend HEI.

Previous research has also indicated that chance of attending university and the attitudes towards which HEIs to attend are greatly influence by the individual's parents' education. Therefore, parents' education is likely to indirectly impact upon the student migration process. A study by the Institute of Education (2012) investigated the influence of parental and family background on HE choices. They found the students whose parents also held a higher educational degree were five times more likely to go to university than those whose parents had no HE qualifications. The report found that even after taking into account the pupils academic achievements, those with university educated parents were still twice as likely to attend HE than their peers with no qualifications. It is therefore evident that students whose parents attended HE were influencing the decisions of students entering HE and as a result, information on the parents education shall be included in this analysis. The hypothesis here is that those students whose parents attended HE would be more advantaged in regards to attending HE and more likely to receive the support and encouragement to migrate if needed in order to attend the HEI of their choice.

HESA provided both a Socio-Economic Classification variable and a Parental Education variable in the HESA Student Record. The Socio-Economic Classification (SEC) variable records the socio-economic background of students aged 21 and over at the start of their course, or for students under 21, the SEC of their parent, step-parent or guardian who earns the most is recorded. The SEC is based on occupations, and if the parent or guardian is retired the SEC is based on their most recent occupation. The SEC classification is provided by HESA in the standard eight class NE-SEC format as recommended by the Office for National Statistics (2010b), with the addition of two categories, 'unknown' and 'not classified'.

HESA also provided a parental education variable that recorded information about whether an entrant's parents have HE qualifications. This field splits the students into 5 categories; Yes, No, Don't Know, Information Refuses and Unknown.

It would have been preferred to use these two variables independently in the research; however, both variables have a large amount of non-response. In the SEC variable 43% of students are classified as 'Not Classified' or 'Unknown', while in the parental education question 34% of students are classified as either 'don't know', 'information refused' or 'Unknown'. As a result of this large non-response in the variables, merging the two variables enabled more information about an individual's social background to be gained.

The merging process involved running a series of robustness checks on all possible logical combinations of the two variables and the best performing combination was chosen. The SEC variable was redefined from eight categories to four, following the Office for National Statistics (2010b) guidelines for three categories with the addition of an unknown category. The parental education variable was also redefined into three categories; yes, no and unknown. After redefining the two variables the merging process took place. The process of how the two original variables were merged is explained in detail in Appendix C. After the merging process the following five Social Background categories were created and are used in the analysis:

- Most Advantaged
- Advantaged
- Less Advantaged
- Least Advantaged
- Unknown

As a result of the merging process the number of students recorded in the unknown category is 21%. Comparing this to 34% and 42% unknown when the variables were included separately, this shows that from merging the two variables information about individuals have been gained and the new variable has significantly less unknowns which improves the data quality and the subsequent quality of the later analysis.

The Social Background variable indicates how advantageous the student's personal background was in supporting a student to enter HE and experiencing the perceived traditional process of migrating away from the parental home to attend a HEI. The hypothesis being that those students who are from the highest social-economic classification and whose parents attended a HEI would have the most advantageous and supportive background in terms of financial support and encouragement to attend more prestigious HEIs and therefore be more likely to make a migration to attend HEI. In contrast those students who were from the lowest socio-economic groups and parents did not attend a HEI would have the least advantageous background as a result of a lack of financial support and parental encouragement to participate in HE and less likely to migrate to attend a HEI.

Age

The age of the student was recorded as at 31st August in the reporting period, therefore in this study the age of the student on 31st August 2011. Age is provided by HESA by categorising each student into one of six categories: 17 years and under, 18-20 years, 21-24 years, 25-29 years, 30 years and over, and Age Unknown.

Gender

The specification for student gender falls within the scope of the Aligned Data Definitions adopted by the Information Standards Board (ISB) for education, skills and children's services (escs). Gender is spilt into Male, Female and Indeterminate. Indeterminate gender means unable to be classified neither as either male or female, and intended to identify students who are intersex and not trans-gender nor as a proxy for not-known. However, the indeterminate field has only a few observations and therefore is left out of the analysis in this chapter.

Number of years in HE

This field indicates the number of years that the student had been enrolled in a HE course or programme leading to the student's qualification aim. This number did not restart if the intended subject or class had changed and as a result the number of years in HE could be different from the number of years on a course, if the student had changed course or retaken a year. The number of years in HE variable is recorded continuously until the open ended top category of 6 years and above.

Level of Study

Level of study is taken from the course aim of the student and classifies a student as either Undergraduate or Postgraduate – more details can be found in Appendix C.

4.3.3 Higher Educational Indicators

For the 2011/12 academic year, every student is recorded as being registered at one of 160 HEIs in the UK. It would be possible to use this detailed information for each institution in the analysis, however, interpretation of an indicator of such size would not bring much intuitive detail to the analysis.

As a result, the institution variable is used to create a categorical indicator variable that categorises all HEIs into six groups. The differences between the HEIs in the UK was discussed briefly in Section 2.1.3, however in the following sub-section, more detail on how these differences can be interpreted and used to create a variable to analyse the differences between the HEIs is provided.

University Categorisation

It is possible to create a categorical variable that groups all 160 HEIs into 6 categories based on when the HEI was founded. There are 6 different time periods that universities in the UK were formed and these are highly correlated with the ranking and perceived prestige of the institutions. The number and percentage of institutions and the percentage of the student population associated with each of the university categories are displayed in Table 4-1 and a list of all the HEIs in each of the Institution Categories is provided in Appendix C. An explanation of how each of the student categories is specified is explained below.

Table 4-1: Categorical Variable of Higher Educational Institutions - The number of HEIs and Percentage of the 201/12 Student Population

Category	Number of HEIs	Percentage of all HEIs (%)	Percentage of the 2011/12 student population (%)
1 – Ancient Universities	7	4.4	6.26
2 – Red Brick or Civic Universities	36	22.5	25.88
3 - Plate Glass or 1960s Universities	20	12.5	14.81
4 – Post-1992 Universities	42	26.3	38.74
5 - Recently Created Universities	19	11.8	8.29
6 – Other	36	22.5	6.02
Total	160	-	-

Source: Higher Education Statistics Agency (2014b)

Ancient Universities

These HEIs were founded during the 'Middle Ages' and the 'Renaissance' periods and generally refer to institutions formed before the 19th century and are some of the oldest universities in the world. Owing to their sheer age and continuous academic and scientific output, all of the ancient universities are highly reputable and are often found in the top segments of the British university rankings. As a result, these universities often have extremely high entry requirement tariffs and as explained in Chapter 2, previous research has shown there to still be high levels of inequality with regards to students from less advantaged backgrounds attending high tariff institutions such as Ancient Universities.

Only 7 out of the 160 HEIs were classified as an Ancient University and the grouping accounted for 6% of the total 2011/12 student population.

Red Brick or Civic Universities

Red Brick or Civic Universities were founded in the nineteenth century, the majority within major industrial cities of the UK. The creation was seen as a reflection of the increasing need for the university level of study of technical, science, design and engineering subject. This group was then extended to include those institutions granted a charter between 1900 and 1963, known as Civic Universities. Similar to the Ancient Universities, the Red Brick or Civic

Universities are often highly reputable and highly placed in the university rankings and also have high entry requirements.

36 of the 160 HEIs were classified as a Red Brick or Civic University and just over a quarter of the student population attend one of these institutions.

Plate Glass or 1960s Universities

Plate Glass Universities refers to any university founded between 1963 and 1992 in the era of the Robbins Report on Higher education (HMSO 1963) which recommended the immediate expansion of the higher educational sector in the UK and the resultant creation of these Plate Glass universities. This group of universities are of slightly lower prestige as the previous two groups of universities and are often found in the upper middle places of the university rankings. Institutions within this group often have middle tariff entry requirements.

20 of the 160 HEIs were classified as a Plate Glass university and around 15% of the 2011/12 student population attended one of these institutions.

Post-1992 Universities

Post-1992 universities are a specific group of universities that relate to any of the former polytechnics, central institutions or colleges of HE that were given university status by John Major's government in 1992 through the Further and Higher Education Act 1992 (HMSO 1992). The most visible impact of the Higher Education Act 1992 was the reclassification of 35 polytechnic colleges into universities allowing them to award their own degrees.

Post-1992 universities are formed from a less traditional background and often offer non-traditional degree courses that are not offered at institutions in the earlier university groupings. Institutions in this grouping therefore also attract larger numbers of non-traditional students which can be linked back to the widening participation policies and often these institutions require medium to low traffic scores.

These universities were the creation of a large restructuring of HE which aimed to increase the number of people entering HE and providing more institutions for people to attend. As a result, these institutions make up the largest grouping out of the 160 HEIs with 42 being classified as a post-1992 university. This group of institutions also has the largest share of the student population with 39%.

Recently Created Universities

This category is for any university created since 2005. These were often formerly further education, teacher training colleges or other specialist colleges and have only been granted university status since 2005.

Other Universities

This category refers to 36 of the 160 HEIs that are classified as HESA as a HEI but do not fit into any of the previous five categories. These HEIs include independent universities, unique institutions (i.e. post-graduate only), colleges of Higher Education, Music, Drama and Art School.

Although this other category accounts for quite a large amount of the HEIs in the UK (22.5%), these institutions are all very small and only account for 6% of the total 2011/12 student population.

Course Studied

HESA provided 19 subject areas in terms of their JACS codes to report which subject area the students degree course was best defined. However, for the analysis in this study a smaller number of subject areas were required and therefore a new variable (sub2) was created in which each student is categorised into one of the following groups to represent the subject area in which they study:

- Medicine
- Science or Engineering
- Agricultural or Veterinary
- Social or Human Sciences
- Business or Law
- Humanities
- Combined

A detailed explanation of how the new variable for course studied was created can be found in Appendix C.

4.4 Preliminary Analysis

This section provides a preliminary analysis into the explanatory variables set out above and whether they have an association with the three outcome variables used in this analysis. By analysing the three outcome variables in turn by cross tabulating them by each of the explanatory variables on a simple bivariate level, enables the identification of any patterns in student migration behaviour. This therefore provides the rationale to investigate the relationship between the explanatory variables and the three outcome variables further by applying more complex multivariate techniques that analyse these relationships simultaneously.

As discussed above in the data section (Section 4.3), three different outcome variables are investigated in this chapter, each of which are discussed in turn below.

4.4.1 Internal Migration

The first outcome variable simply distinguishes between students that migrate in attend a HEI and those that do not. The percentages of all students that were migrants and non-migrants have been cross-tabulated against five of the major explanatory variables as shown in Table 4-2. This enables the identification of any association that is present when these explanatory variables are analysed individually. The purpose for this is to assess whether further and more detailed analysis is worthwhile.

One of the aims of this chapter is to investigate how ethnic groups were represented in the HE population and to investigate what impact ethnicity has on the migration outcome. The total non-international student population in the UK is still heavily dominated by the White ethnic group in the 2011/12 academic year with nearly 4 out of 5 students falling within this category. However, when this is compared to the ethnic composition of the total population from the 2011 census, 87.1% of the total population were in the White ethnic group (Office for National Statistics et al. 2011). Therefore, it can be said that the total non-international student population is more ethnically diverse than the total population of the UK.

Table 4-2: Percentage of Students that were Internal Migrants by Individual Characteristic Explanatory Variables

2011/12		Internal Migration - Yes (%)	Internal Migration - No (%)	All Students (%)
Ethnicity	White	47.4	52.6	79.1
	Black	30.1	69.9	6.0
	Asian	32.0	68.0	9.0
	Other	45.5	54.5	4.0
	Unknown	27.9	72.1	1.9
Social Background	Most Advantaged	71.1	28.9	17.5
	Advantaged	57.8	42.2	17.8
	Less Advantaged	45.9	54.1	22.8
	Least Advantaged	32.6	67.4	20.6
	Unknown	21.8	78.2	21.3
Gender	Male	49.6	50.4	42.5
	Female	40.8	59.2	57.5
Institution Category	Ancient	67.1	32.9	6.3
	Red Brick	58.4	41.6	25.9
	Plate Glass	47.5	52.5	14.8
	Post 1992	35.2	64.8	38.7
	Recent	34.5	65.5	8.3
	Other	42.6	57.4	6.0
Course	Medicine	30.5	69.5	17.2
	Science/ Engineering	51.8	48.2	24.2
	Agr/ Vet	52.4	47.6	1.2
	Social/ Human	36.3	63.7	20.3
	Business/ Law	39.6	60.4	14.8
	Humanities	60.0	40.0	21.2
	Combined	15.2	84.8	1.1
Total		44.6	55.4	-

Total Population Size – 1,797,492

Source: Higher Education Statistics Agency (2014b)

So do the migration patterns differ across the ethnic groups? It appears that there are large differences between the percentage of White students that migrated to attend a HEI compared to those in the Black and Asian groups. The percentage of White students that migrate to attend a HEI is by far the highest and is much higher than the Black and Asian groups. Therefore, by simply analysing the ethnic differences in the binary internal migration outcome, it is clear that there are marked ethnic differences in the migration experiences of students attending HEIs.

Another aim of this chapter is to investigate how students from different social backgrounds are represented in the HE population and what impact social background has on the migration outcome. The total population of students is divided relatively equal across the five categories of student social background, although there are significant differences by social background in the migration outcomes.

The difference between the most advantaged and least advantaged social background groups with regards to the percentage of those students that migrated is extremely large. Over 70% of students in the most advantaged group migrated in order to attend a HEI, while this number dropped significantly to 33% for those in the least advantaged group.

As suggested in previous research (e.g. Cotton et al. 2013), there is also a prominent gender gap in the participation of students in the 2011/12 academic year with 57.5% of the non-international student population being female. However, the more interesting patterns are visible in the migration outcomes. Despite having much greater numbers of students, female students are much less likely to migrate in order to attend a HEI than their male peers.

The final variables investigated are those recording the institution that the student attended and the course that the student studied. With regards to the institution category a clear pattern is visible. Those students attending the more prestigious and reputable ancient, red brick and plate glass institutions have much higher percentages of students that migrated in order to do so. In contrast the newer and more recent institutions have a student population with much fewer migrants.

4.4.2 Distance Travelled

With clear associations between the explanatory variables and whether or not a student migrated, next it is investigated if these trends were mirrored in the distance travelled by students to attend a HEI. The mean distance in kilometres travelled by students in each of the five explanatory variable categories are shown in Table 4-3.

Table 4-3: Mean Distance Travelled by Individual Characteristic Explanatory Variables

2011/12		Mean Total Distance (km)
Ethnicity	White	96.8
	Black	59.7
	Asian	61.9
	Other	85.1
	Unknown	79.4
Social Background	Most Advantaged	126.7
	Advantaged	102.9
	Less Advantaged	89.4
	Least Advantaged	68.8
	Unknown	73.2
Gender	Male	99.7
	Female	83.9
Institution Category	Ancient	157.7
	Red Brick	108.9
	Plate Glass	94.1
	Post 1992	72.5
	Recent	77.8
	Other	92.8
Course	Medicine	72.3
	Science/ Engineering	98.1
	Agr/ Vet	140.2
	Social/ Human	79.8
	Business/ Law	83.1
	Humanities	111.4
	Combined	60.5

Total Population Size – 1,797,492

Source: Higher Education Statistics Agency (2014b)

The mean total distance travelled by students also differs as a result of their ethnic group. White students on average travel around 35km more than their Black and Asian peers. Therefore, this also supports the case that migration patterns to HEI are impacted by the students' ethnic background.

A similar trend is also visible with regards to mean distance travelled by social background. Those in the most advantageous group travel on average 57.9 km more than those students with the least advantageous social backgrounds. Again, males are shown to travel on average 15km further than their female peers.

The patterns are also very similar when analysing institution category. There appears to be a linear association between university category and distance travelled. As the university category declines in age and prestige so does the mean distance travelled. This suggests that students are willing to travel further distances to attend a higher ranked and more prestigious institution. There are also noticeable differences between the course studied and the mean distance travelled, with some courses being associated with greater average distances travelled than others. For example, students who study an agricultural or veterinary subject travel the furthest distances on average. This trend in larger distances can be linked to the fact that very few of the 160 HEIs in the UK offer agricultural or veterinary courses and those that do are more often than not located in quite remote/rural areas.

4.4.3 Student Migration Category

The final outcome of interest is the type of student migrant, as categorised using the typology proposed in Chapter 3. Again, the percentages of all students that are in the different student migration categories have been cross-tabulated against five of the major explanatory variables as shown in Table 4-4.

Clear differences between the ethnic groups appear. There are significantly more white internal student migrants than their black and Asian peers, while the percentage of black and Asian commuters or distance leaners are higher than their white counterparts. This suggests that white students are still taking traditional migrational moves away from the domicile to attend a HEI. In comparison many more Asian and black students are living in the same area as the domicile, presumably the parental home, and either studying at a local HEI or commuting in order to study.

There are also clear differences as a result of the student's social background when looking at the different types of student migrants. The most advantageous group have over double the percentage of internal student migrants than the least advantaged group, while the least advantaged group have over double the percentage of local students than the most advantaged group.

Table 4-4: Student Migration Category by Individual Characteristic Explanatory Variables

2011/12		Local Students (%)	Commuter /Distance Learner (%)	Internal Student Migrant (%)	Migrant Commuter/Distance Learner attended local HEI (%)	Internal Migrant Commuter/Distance Learner (%)
Ethnicity	White	10.3	42.3	37.7	0.5	9.2
	Black	13.2	56.7	19.8	0.6	9.7
	Asian	16.4	51.7	23.6	0.4	8.0
	Other	11.8	42.7	34.0	0.6	10.9
	Unknown	15.7	56.4	19.1	0.8	8.0
Social Background	Most Advantaged	5.4	23.4	60.4	0.3	10.5
	Advantaged	8.2	34.0	47.6	0.4	9.8
	Less Advantaged	11.0	43.1	35.1	0.6	10.2
	Least Advantaged	13.9	53.5	24.1	0.6	7.9
	Unknown	15.9	62.3	13.4	0.8	7.6
Gender	Male	10.4	40.0	39.5	0.5	9.6
	Female	11.8	47.4	31.4	0.6	8.8
Institution Category	Ancient	13.1	19.8	63.5	0.5	3.1
	Red Brick	8.7	32.9	46.8	0.5	11.1
	Plate Glass	9.6	42.9	39.4	0.5	7.6
	Post 1992	12.9	51.9	25.5	0.6	9.0
	Recent	11.0	54.5	26.8	0.5	7.2
	Other	11.1	46.3	28.1	0.5	14.0
Course	Medicine	12.3	57.2	20.8	0.7	9.0
	Science/ Engineering	10.2	38.0	42.5	0.5	8.8
	Agr/ Vet	4.9	42.7	31.8	0.4	20.2
	Social/ Human	12.9	50.8	29.0	0.6	6.7
	Business/ Law	12.6	47.8	30.7	0.6	8.4
	Humanities	8.5	31.5	47.6	0.4	12.0
	Combined	25.5	59.3	11.1	0.5	3.7
Total		11.2	44.3	34.9	0.5	9.1

Total Population Size – 1,797,492

Source: Higher Education Statistics Agency (2014b)

There are also gender differences in the types of student migration experienced. Both sexes had similar percentages of those that are categorised local students, migrant commuter/distance learner attended local HEI and internal migrant commuter/distance learner. Although, there are differences between the sexes in the percentages categorised in commuter/distance learners and internal student migrants. Males have significantly higher percentages of internal student migrants than their female counterparts, while women have much higher percentages of commuter/distance learners than men.

Some interesting and differing patterns arise when analysing institution category and the student migration category. Ancient universities have the highest percentage of internal student migrants as would be expected from the results of the previous outcome categories. However, Ancient universities also have the highest percentage of local students which was the opposite finding to what was expected. From the previous outcome variables it was predicted that the newer and less prestigious institutions would have had the highest percentages of local students. While, the newer/more recent institutions do have significantly higher percentages of commuting students compared to any of the other categories as expected.

In summary, when analysed unconditionally independent of any other explanatory variables, there appears to be significant differences between the majority of the explanatory variables and all three of the migration outcomes.

The ethnic groups have clear associations in all three of the migration outcomes variables. These preliminary results support findings from the literature that suggest ethnicity plays an important role in the migration decision process (Modood and Shiner 1994; Dearing 1997; Tomlinson 2001; Khambhaita and Bhopal 2013). Clear differences in the migration outcomes experienced by the students are also strongly associated with the different types of student social background. Students with the more advantaged backgrounds were more likely to migrate and migrate further distances than those in the lesser advantaged groups. While there are also clear associations between the outcome variables and gender, institution and course studied. This therefore provides strong evidence to support further and more detailed analysis of these variables later on in the chapter.

4.5 Methodology

In the previous section it was shown that when analysed unconditionally independently from all other explanatory variables, there appears to be some form of association between all the explanatory variables and the three outcome variables used to illustrate student migration (or non-migration). As a result, this encouraged further investigation into how these variables would impact on the migration outcomes when the explanatory variables are analysed simultaneously.

The three outcome variables used in the analysis are quite different in their format and as a result they require different methods to analyse the student migration outcomes against the explanatory variables simultaneously. This section explains each of these methods in turn and the results of these methodologies are presented in Section 4.6.

4.5.1 Logistic Regression Modelling

The first of the migration outcome variables that is analysed is the binary outcome variable that depicts whether a student had migrated or not in order to attend a HEI. A value of 0 (a failure) is recorded for no migration and a value of 1 (a success) if a migration has occurred. For binary outcomes, the most commonly used model is the logistic regression model (Long and Freese 2006; Agresti 2013). Logistic regression models analyse how each explanatory variables affect the probability of the event occurring, for this analysis, the probability of a student migrating to attend a HEI.

For the binary response variable of migration Y and the multiple explanatory variables x_p the rearranged logistic regression model to calculate the predicted probabilities for $\pi(x) = P(Y = 1)$, the predicted probability of a student making a migration to attend a HEI, at values $x = (x_1, \dots , x_p)$ of p predictors is (Agresti 2013, p18):

$$\pi(x) = \frac{e^{(\alpha+\beta_1 x_1+\beta_2 x_2+\cdots+\beta_p x_p)}}{1+e^{(\alpha+\beta_1 x_1+\beta_2 x_2+\cdots+\beta_p x_p)}} \tag{4.1}$$

Where the parameter β_j refers to the effect of x_j on the log odds that $Y = 1$, adjusting for the other x_k . Therefore, $e^{(\beta_j)}$ is the multiplicative effect on the odds of a student migration of a 1-unit increase in x_j when the other variables levels of x_k are left constant.

As previously mentioned, there are many variables available to analyse their impact on the probability of a student migrating to attend HEI. However, which combinations of these variables that best predict the outcome are unclear. There is no set criterion on how to select the 'best' fitting model. Assessing which combination of variables for the final model involves both the analysis of the statistical fit of individual observations and the evaluation of scalar measures of fit for the model as a whole (Long and Freese 2006).

How to select the final model often involves a trade-off between the best statistical fitting model and the model that includes the combination of variables that makes most theoretical sense towards answering the research questions. It is also the case that within the field of social sciences that researchers may not necessarily be interested in finding the best statistically fitted model, as researchers can be equally interested in those variables that are not significant as those that are.

As a result, a model selection process is undertaken to find a model that has a good statistical fit compared to other models tested but also only includes variables that could be justified and linked back to the factors identified as impacting on the student migration process in the contextual framework presented in Chapter 2 (Figure 2-2). The process started with a simple model with the most important variables (ethnicity, social background and gender). Then the numbers of variables in the model were sequentially increased as well as including the three interaction variables between ethnicity, background and gender. With the additional of variables in a sequential manner each new model is nested within the previous and therefore the impact of adding each set of variables can be statistically tested in turn. Each time a number of scalar measures of fit of the models are recorded. The measures of fit are created using the *fitstat* command designed and recommended by Long and Freese (2006). This command produces several measures of fit which are explained in turn below and the numeric values for these measures for each model in the model selection process are reported in Table D-1 (Appendix D).

Log-likelihood: This is the function of the parameters of the statistical model. The likelihood of a set of parameter values, θ, given outcomes x, is equal to the probability of those observed outcomes given those parameter values. Maximum likelihood iterations are calculated by computing the log likelihood of the model with all parameters but the intercept constrained to zero, referred to as $L(M_{Intercept})$. The log likelihood on convergence, referred to as $L(M_{Full})$ is then reported and is the value reported in Table D-1. The smaller the value of $L(M_{Full})$ the better the fit of the model, however, this statistic is more commonly used in the Likelihood Ratio Test in which the difference in log-likelihood scores between nested models are tested against a critical value dependent on the number of degrees of freedom, although this statistic is not reported in the analysis presented in this chapter (Long and Freese 2006; Argesti 2013).

Pseudo-R^2s: The squared multiple correlation, R^2 is also used to assess the goodness of fit of a statistical model as it represents the proportion of variation in the criterion that is explained by the predictor variables as shown in Equation 4.2 (Cohen et al. 2002; Long and Freese 2006):

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{N}(y_i - \bar{y}_i)^2} = 1 - \left\{\frac{L(M_{Intercept})}{L(M_{Full})}\right\}^{2/N} \tag{4.2}$$

AIC (Akaike's Information Criterion): Akaike (1973) information criterion can be used to compare nested and non-nested models and is defined by Equation 4.3:

$$AIC = \frac{\{-2\ln\hat{L}(M_k) + 2P_k\}}{N} \tag{4.3}$$

Where $\hat{L}(M_k)$ is the likelihood of the model and P_k is the number of parameters in the model. With all else being equal, the model with the smaller AIC is considered the better fitting model (Long and Freese 2006).

BIC (The Bayesian Information Criterion): this method was proposed by Raftery (1996) to again compare nested and non-nested models. When the model M_k has the deviance $D(M_k)$ then the BIC is defined as shown in Equation 4.4:

$$BIC_k = D(M_k) - df_k \ln N \tag{4.4}$$

In Equation 4.4 df_k is the degree of freedom associated with the deviance.

The choice of which model was preferred was based primarily on the observed differences in the BIC values of the nested models. Raftery (1996) as read in Long and Freese (2006) suggested guidelines for the strength of evidence favouring model$_{i+1}$ against model$_i$ on the difference in the BIC score:

Absolute Difference	*Evidence*
0-2	Weak
2-6	Medium
6-10	Strong
>10	Very Strong

The model selection process (in which the statistical evidence supporting the process undertaken is shown in Table D-1) resulted in the selection of model 10 to be final model in which the results will be formulated from. This model included all the variables available from the student record data and the three interaction terms. The full model had the best statistical fit and all variables including the interaction terms have theoretical groundings in the literature that shows they are likely to impact on the student migration transition.

Each of the nested models had difference in BIC values that suggested very strong evidence that the models with the additional variables were favoured. However, it must be considered here that when evaluating the strength of the different models used in the analysis presented in this chapter that the sample size was very large with just fewer than 1.8 million observations. It must also be considered here that the data used in the current analysis is population data and is not derived from a sample or survey and as a result our interpretation of significance levels and best fitting models must be adapted accordingly.

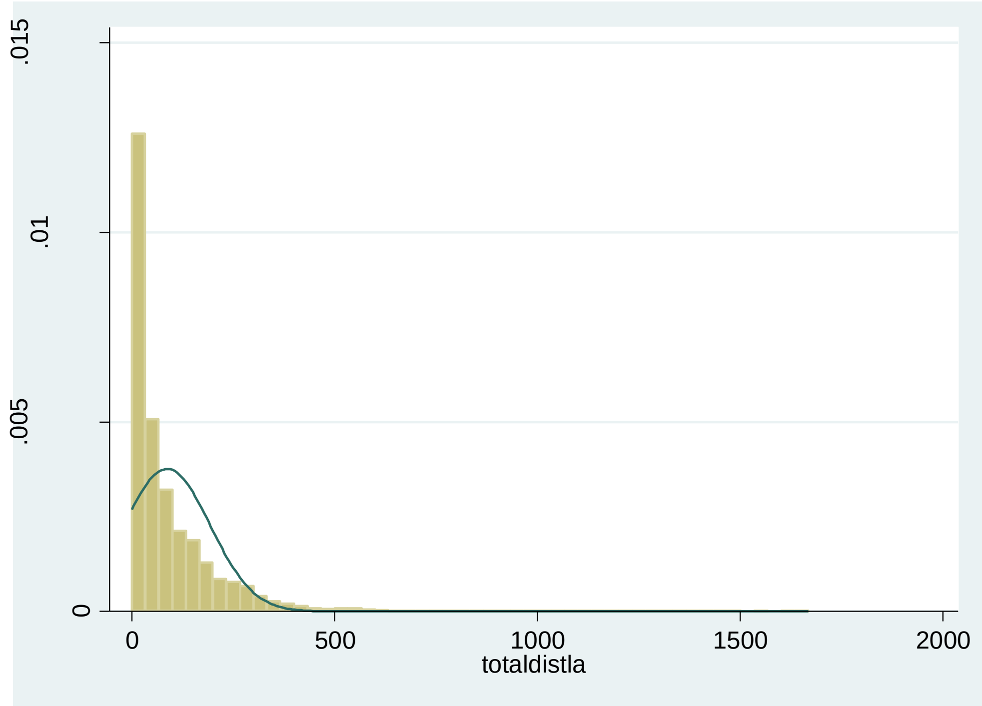
In standard statistical and sampling techniques a researcher often has a sample of n units, say s, drawn from the population of N units by a specified stochastic procedure often referred to as the survey design. The probability p, of any particular sample outcome s is determined by the survey design (Hartley and Sielken 1975). However, when the data being analysed contains all of the N units of the population, when interpretation and inferences are being made then it can be said the statistical inferences being made are actually referring to a hypothetical super-population (Cochran 1939; Hartley and Sielken 1975; Dorfman and Valliant 2005). The data used in this analysis is population data as the student record data is an administrative census of all students enrolled in HE in the UK. As a result, the significance levels are therefore referring to a hypothetical super-population. The associations between the explanatory variables and the outcome variables in reference to the real life student population are of primary interest. As a result, the model's measures of fit, the significance levels and interpretation should be made with care as a result of this super-population being invoked.

The results for the final chosen logistic regression model of migration can be found in Section 4.6.1.

4.5.2 Tobit Regression

The second migration outcome variable is the variable of distance travelled by the student in order to attend a HEI as explained in Section 4.3. The outcome variable is represented as a numerical value of distance in kilometres on a continuous scale. For continuous outcome variables the common method to estimate linear relationships with multiple explanatory variables is multiple linear regression (Greene 1993). However, one of the criterions for linear regression is that the outcome variable is normally distributed. The distribution of the distance outcome variable is shown in Figure 4-1 and is clearly not normally distributed but was heavily clustered and censored at zero km. These individuals clustered at zero are local students that do not travel any distance to attend a HEI. Therefore, the data are censored at zero as it is impossible to travel a negative distance.

Figure 4-1: Histogram of Total Migration Distance (km)



Source: Higher Education Statistics Agency (2014b)

Note: 'totaldistla' refers to the total distance measured in km for each individual student in the 2011/12 academic year. Non-UK Domiciled students were removed.

Taking into account the distribution of distance was not normally distributed and was clustered and censored at zero km, a different methodology other than multiple linear regression is needed to be used or a transformation of the outcome variable be performed.

The Tobit Model

The Tobit model is a statistical model originally proposed by Tobin (1958) that was designed to estimate the linear relationship between variables when the outcome variable has a number of its values clustered at a limiting values, usually zero (McDonald and Moffitt 1980). The Tobit technique uses all observations, both those at the limit and those above it, to estimate a regression line. As a result of the model taking into account the clustering in the outcome variable, it is preferred, in general, over alternative techniques that estimate the regression line only with observations above the limit (McDonald and Moffitt 1980).

As shown in Figure 4-1, the outcome variable of total distance is an example of an outcome variable that is clustered around zero. This clustering around zero is a result of those students that do not make a migration or commute in order to attend a HEI and are therefore classified as a local student (Section 3.4). Due to the structure of the outcome variable in this example the Tobit model is a good methodology than can analyse the impact of the explanatory variables on the distance travelled in order to attend a HEI.

In the Tobit model used here y_i refers to the total distance travelled by the student in kilometres and several explanatory variables x_p . The structural equation for the Tobit model is shown in Equation 4.5 (Long 1997).

$$y_i^* = x_{1i}\beta_1 + x_{2i}\beta_2 + \cdots + x_{pi}\beta_p + \varepsilon_i \tag{4.5}$$

In Equation 4.5, $\varepsilon_i \sim N(0, \sigma^2)$ the x's are observed for all cases. y^* is a latent variable that is observed for values greater the τ and is censored for values less than or equal to τ. The observed y was defined by the measurement Equation 4.6.

$$y_i = \begin{cases} y_i^* & if \quad y_i^* > \tau \\ \tau_y & if \quad y_i^* \leq \tau \end{cases} \tag{4.6}$$

As previously mentioned, the data in this analysis are censored at zero therefore $\tau = 0$, with this in mind and combining Equations 4.5 and 4.6 the final Tobit model is shown in Equation 4.7.

$$y_i = \begin{cases} y_i^* = x_{1i}\beta_1 + x_{2i}\beta_2 + \cdots + x_{pi}\beta_p + \varepsilon_i & if \quad y_i^* > 0 \\ 0 & if \quad y_i^* \leq 0 \end{cases} \tag{4.7}$$

The model selection process undertaken for the Tobit model is identical to that for the logistic regression methodology (Section 4.5.1) and the measure of fit statistics and the output from model selection process are shown in Table C-2 (Appendix C). Model 10 is selected to be final model in which the results can be found in Section 4.6.2. This model includes all the variables available from the student record data and the three interaction terms. The full model has the best statistical fit and all variables including the interaction terms have theoretical groundings in the literature that shows they are likely to impact on the distance travelled by the student.

Alternative Methodologies

The Tobit methodology is chosen as the most appropriate methodology to conduct the analysis of distance migrated. However, there are other methodologies that deal with data that have large or unusual amounts of zeros that could also have been used in the analysis.

The Hurdle Regression Model (HRM) combines a binary model to predict values of zero and a zero-truncated Poisson model to then predict the non-zero counts (Mullahy 1986; Cameron and Trivedi 1998). Another method is zero-inflated models, introduced by Lambert (1992), which change the mean structure to allow zeros to be generated by two distinct processes, compared with one process in the hurdle model. The zero-inflated model assumes that there are two latent groups. An individual in the 'Always Zero' group has an outcome of 0 with a probability if 1, whereas an individual in the 'Not Always Zero' groups might have a zero count, but there is a non-zero probability that they have a positive count. This process is developed in three steps: (1) model membership into the latent groups; (2) model counts for those in the 'Always Zero' group and (3) compute observed probabilities as a mixture of the probabilities for the two groups (Long and Freese 2006).

However, both the hurdle model and zero-inflated count models are designed to model count data. In this analysis the data do not represent count data but distance on a continuous scale with a clustering at zero. Despite this zero-inflate Poisson (ZIP) modelling is could still would work on the data used here. Therefore, ZIP modelling is also conducted in the same manner as for the Tobit models and the same model selection process is used. The result of these ZIP models will be used to check the robustness of the findings produced using the Tobit methodology.

4.5.3 Multinomial Logistic Regression Modelling

The third and final migration outcome variable is the variable of student migration category that was proposed in Section 3.4, in which student migration was categorised into eight different groups. However, as explained in Section 4.3.1 international students and students with locations unknown are removed from the analysis in this chapter which leaves the remaining five categories of student migration.

In this section of the analysis, to ease the interpretation of the final model, the student category 'Migrant Commuter/Distance Learner attending local HEI (%)' is dropped from the analysis. Therefore the outcome variable has only four student migration categories. An interpretation of four outcome groups instead of five is much simpler and the group dropped from the analysis only made up 0.5% of the total student population being analysed.

The Multinomial logistic regression modelling (MNLM) is perhaps the most commonly used regression model for nominal outcomes. The model is easy to estimate, and interpretation is straightforward, albeit complicated due to the large number of parameters involved. (Cheng and Long 2007). The MNLM is suitable where the response variable is nominal and has three or more categories that are unordered. The basic principle of MNLM is the prediction of the probability of membership to each group of the outcome variable as a result of the observed explanatory variables. Therefore, for this section of the analysis, the MNLM is used to predict the probability of a student being in one of the student migration categories given their explanatory characteristics. In predicting the probabilities the response categories are simultaneously compared to a reference category. MNLM models the log of probability ratio; the log of probability of response in one category compared to the probability of the reference category. The set-up of these models, where the outcome variable has four categories, as in this analysis, is shown in Equations 4.8 (Agresti 2013):

$$\log\left(\frac{\pi_1}{\pi_4}\right) = \alpha_1 + x_1\beta_1 \quad \log\left(\frac{\pi_2}{\pi_4}\right) = \alpha_2 + x_2\beta_2 \quad \log\left(\frac{\pi_3}{\pi_4}\right) = \alpha_3 + x_3\beta_3 \tag{4.8}$$

In Equation 4.8, π_1 is the response category 1, π_2 is the response category 2, π_3 is the response category 3 and π_4 is the response category 4 (reference category), α_i the intercept, x_i a vector of the explanatory variables and β_i the coefficients.

The reference category used in this analysis is 'Local Student' (4). Regression equations are set up for Commuter/Distance Learner (1), Internal Student Migrant (2) and Internal Migrant Commuter/Distance Learner (3). In order to ease interpretation, results from the logit equations are used to calculate the predicted probabilities of being in a student category by transforming the equations into the predicted probabilities as shown in Equations 4.9 to 4.12 (Agresti 2013):

Probability of the reference category (Local Student):

$$P_4 = \frac{1}{1+e^{\alpha_1+\beta_1 x}+e^{\alpha_2+\beta_2 x}+e^{\alpha_3+\beta_3 x}} \tag{4.9}$$

Probability of category 1 (Commuter/Distance Learner):

$$P_1 = \frac{e^{\alpha_1+\beta_1 x}}{1+e^{\alpha_1+\beta_1 x}+e^{\alpha_2+\beta_2 x}+e^{\alpha_3+\beta_3 x}} \tag{4.10}$$

Probability of category 2 (Internal Student Migrant):

$$P_2 = \frac{e^{\alpha_2+\beta_2 x}}{1+e^{\alpha_1+\beta_1 x}+e^{\alpha_2+\beta_2 x}+e^{\alpha_3+\beta_3 x}} \tag{4.11}$$

Probability of category 3 (Migrant Commuter/Distance Learner):

$$P_3 = \frac{e^{\alpha_3+\beta_3 x}}{1+e^{\alpha_1+\beta_1 x}+e^{\alpha_2+\beta_2 x}+e^{\alpha_3+\beta_3 x}} \tag{4.12}$$

As is the case in all types of regression modelling, the MNLM methodology relies on a number of assumptions to be met in order for the methodology to predict the probabilities accurately and without bias. The MNLM assumes the data are case specific and the collinearity is relatively low. There is one assumption that causes concern to many researchers which is the assumption of the independence of irrelevant alternatives (IIA) and is implicit to the MNLM (Cheng and Long 2007). The IIA means that, with all else being equal, a person’s choice between two alternative outcomes is unaffected by what other choices are available. If the MNLM is being used to model the choices of individual’s (as is the case in this chapter), it may in some situations, where the IIA is violated, impose too much constraint on the relative preferences between the alternative choices. However, this point is only important and needs to be taken into account if the analysis aims to predict how an individual's choices would change if one of the categories/options of choice were to be removed or if another option were to be added.

In the analysis conducted within this chapter, there are four types of student migration options available to the students. As a result the student migration choice that is being modelled using MNLM models how a student's personal characteristics impact on their decision to be in one of these student migration categories. As explained in Section 3.4, these student categories are exhaustive and a UK domiciled student can only be classified into these categories. No categories will ever become unavailable and no more can be created. Therefore, the MNLM in this analysis is simply modelling the characteristics of students with the choice of the four student categories. At no stage will the MNLM be used to assess the impact or influence of the student characteristics as a result of one of the categories being removed or more categories being added to the model. Therefore the issue of IIA for this analysis is null and void.

The model selection process undertaken here is identical to that of the previous two methods and the measure of fit statistics and model selection process are shown in Table D-3 (Appendix D). Due to the relatively small value added to the overall fit of the model by including the interaction terms and due to the interaction terms making a large number of the other variables insignificant, the best model for interpretation in this analysis is the model with all explanatory variables but no interaction terms – Model 4.

4.6 What factors impact on the Student Migration Outcomes?

The following section brings together the previous sections of the chapter, which have reviewed relevant past research, introduced the data to be analysed and explained the methodologies, to answer the research questions proposed in the introduction: 1) How does a student's social background, ethnicity or gender impact on the migration outcomes experienced in order to attend a HEI? 2) How does the course studied or institute attended impact on the student's migration outcome?

The student migration outcome is measured using three differing variables each with varying levels of complexity and by comparing the findings using these three different techniques, the robustness of the overall results can be assessed. The remainder of the section presents the results of each of these three outcome variables in turn.

4.6.1 Internal Migration

Out of the 1,797,492 non-international students enrolled at a HEI in the United Kingdom in the 2011/12 academic year, 44.6% of them migrated across a LA boundary to do so (Higher Education Statistics Agency 2014b). It has already been shown that when considered individually, all the explanatory variables analysed seemed to have some form of association with the student migration outcome. In the current sub-section, the method and final chosen model as explained in Section 4.5.1 are used to explain what factors impact on the probability of a student migrating to attend a HEI by analysing the impacts of the explanatory variables simultaneously. The coefficients (β), standard errors, 95% confidence intervals and the odds of a being a student migrant (e^{β}) of the final logistic regression model are shown in Table 4-5.

One of the aims of this chapter is to analyse whether a student's background or ethnicity impacted on the migration outcomes experienced in order to attend a HEI. As a result, an interaction terms between these variables is included in the model and the outcomes were significant at the 1% level (with the exception of ethnicity unknown and Asian*High, which were insignificant at the 10% level). These two variables are also interacted with gender, to see if there were any significant differences between ethnicity and background as a result of the student's gender. Again, these two new interaction terms were significant at the 1% level (with the exception of Other*Female and Unknown*Female, which were insignificant at the 10% level).

It is important to note that when interpreting the figures shown in Table 4-5, that those variables that were involved in the significant interaction terms, their main effect terms cannot be interpreted individually since the individual main effects of interacted variables cannot be isolated. Therefore, interpretation regarding the ethnicity, background and gender variables should be in terms of their interactions in order to make appropriate conclusions.

It is however important to remember that the high significance levels and measures of fit found in the result may be a product of having population data and as a result the statistical inferences being made are technically referring to a super-population that has been invoked within this analysis (Cochran 1939; Hartley and Sielken 1975; Dorfman and Valliant 2005).

Table 4-5: Multiple logistic regression results of the association between student migration and the student characteristic variables

VARIABLES	Coefficient (β)	P-Value	SE	95% Confidence Interval	e(β)
Constant	1.783	0.000	(0.0136)	1.757 - 1.810	5.948
Ethnicity					
White[a]					
Black	-0.641	0.000	(0.0211)	-0.682 - -0.600	0.527
Asian	-0.971	0.000	(0.0174)	-1.005 - -0.937	0.379
Other (Including Mixed Race)	-0.397	0.000	(0.0223)	-0.441 - -0.354	0.672
Unknown	-0.0932	0.105	(0.0575)	-0.206 - 0.0194	0.911
Social Background					
Most Advantaged[a]					
Advantaged	-0.277	0.000	(0.00932)	-0.295 - -0.258	0.758
Less Advantaged	-0.485	0.000	(0.00893)	-0.503 - -0.468	0.616
Least Advantaged	-0.717	0.000	(0.00953)	-0.736 - -0.698	0.488
Unknown	-0.945	0.000	(0.0101)	-0.965 - -0.925	0.389
Gender					
Male[a]					
Female	-0.145	0.000	(0.00886)	-0.162 - -0.127	0.865
Subject					
Medicine[a]					
Science/Engineering	0.181	0.000	(0.00623)	0.168 - 0.193	1.198
Argicultural/Veterinary	0.540	0.000	(0.0173)	0.506 - 0.574	1.716
Social/Human	-0.0200	0.002	(0.00646)	-0.0326 - -0.00733	0.980
Business/Law	-0.00134	0.845	(0.00683)	-0.0147 - 0.0121	0.999
Humanities	0.464	0.000	(0.00633)	0.451 - 0.476	1.590
Combined	-0.746	0.000	(0.0239)	-0.793 - -0.699	0.474
Institution Category					
Ancient[a]					
Red Brick	-0.337	0.000	(0.00903)	-0.355 - -0.319	0.714
Plate Glass	-0.853	0.000	(0.00955)	-0.872 - -0.835	0.426
Post 1992	-1.213	0.000	(0.00880)	-1.230 - -1.195	0.297
Recent University	-1.289	0.000	(0.0103)	-1.309 - -1.269	0.276
Other	-1.030	0.000	(0.0111)	-1.052 - -1.009	0.357
Age					
17 years and under	-1.002	0.000	(0.0216)	-1.044 - -0.960	0.367
18-20 years[a]					
21-24 years	-0.748	0.000	(0.00487)	-0.758 - -0.738	0.473
25-29 years	-1.614	0.000	(0.00724)	-1.628 - -1.600	0.199
30 years and over	-2.729	0.000	(0.00727)	-2.743 - -2.714	0.065
Age unknown	-3.127	0.000	(0.183)	-3.485 - -2.769	0.044
Number of Years in HE					
1[a]					
2	0.174	0.000	(0.00446)	0.166 - 0.183	1.190
3	0.394	0.000	(0.00500)	0.384 - 0.404	1.483
4	0.710	0.000	(0.00778)	0.695 - 0.725	2.034
5	0.771	0.000	(0.0144)	0.742 - 0.799	2.162
6 or more	1.071	0.000	(0.0212)	1.030 - 1.113	2.918
Unknown	0.722	0.000	(0.0943)	0.537 - 0.907	2.059
Level of Study					
Post-Graduate[a]					
Under-Graduate	-0.158	0.000	(0.00641)	-0.171 - -0.146	0.854

VARIABLES	Coefficient (β)	P-Value	SE	95% Confidence Interval	e(β)
Domicile					
North[a]					
South	0.520	0.000	(0.00375)	0.512 - 0.527	1.682
Interaction Terms					
Ethnicity * S.Background					
White*Most Advantaged[a]					
Black*Advantaged	0.0627	0.024	(0.0278)	0.00816 - 0.117	1.065
Black*Less Advantaged	0.124	0.000	(0.0244)	0.0764 - 0.172	1.132
Black*Least Advantaged	0.115	0.000	(0.0252)	0.0658 - 0.165	1.122
Black*Unknown	0.302	0.000	(0.0270)	0.249 - 0.355	1.353
Asian*Advantaged	-0.0234	0.286	(0.0220)	-0.0665 - 0.0196	0.977
Asian*Less Advantaged	-0.0574	0.005	(0.0203)	-0.0972 - -0.0175	0.944
Asian*Least Advantaged	-0.0835	0.000	(0.0200)	-0.123 - -0.0442	0.920
Asian*Unknown	0.399	0.000	(0.0226)	0.355 - 0.444	1.490
Other*Advantaged	0.113	0.000	(0.0290)	0.0557 - 0.169	1.120
Other*Less Advantaged	0.0718	0.007	(0.0265)	0.0199 - 0.124	1.074
Other*Least Advantaged	0.0326	0.245	(0.0281)	-0.0224 - 0.0876	1.033
Other*Unknown	0.271	0.000	(0.0299)	0.213 - 0.330	1.311
Unknown*Advantaged	-0.0130	0.854	(0.0706)	-0.151 - 0.125	0.987
Unknown*Less.Adv	-0.133	0.041	(0.0653)	-0.261 - -0.00550	0.875
Unknown*Least.Adv	-0.195	0.004	(0.0684)	-0.329 - -0.0612	0.823
Unknown*Unknown	-0.183	0.002	(0.0598)	-0.300 - -0.0657	0.833
S.Background*Gender					
V. Advantaged*Male[a]					
Advantaged*Female	-0.0568	0.000	(0.0119)	-0.0801 - -0.0335	0.945
Less Advantaged*Female	-0.0993	0.000	(0.0113)	-0.121 - -0.0772	0.905
Least Advantaged*Female	-0.173	0.000	(0.0118)	-0.197 - -0.150	0.841
Unknown*Female	-0.138	0.000	(0.0124)	-0.162 - -0.114	0.871
Ethnicity*Gender					
White*Male[a]					
Black*Female	0.0948	0.000	(0.0162)	0.0631 - 0.126	1.099
Asian*Female	0.0698	0.000	(0.0125)	0.0453 - 0.0943	1.072
Other*Female	0.00344	0.847	(0.0178)	-0.0315 - 0.0384	1.003
Unknown*Female	0.00881	0.768	(0.0298)	-0.0496 - 0.0672	1.009
Observations	1,797,942				
R-Squared	0.252				

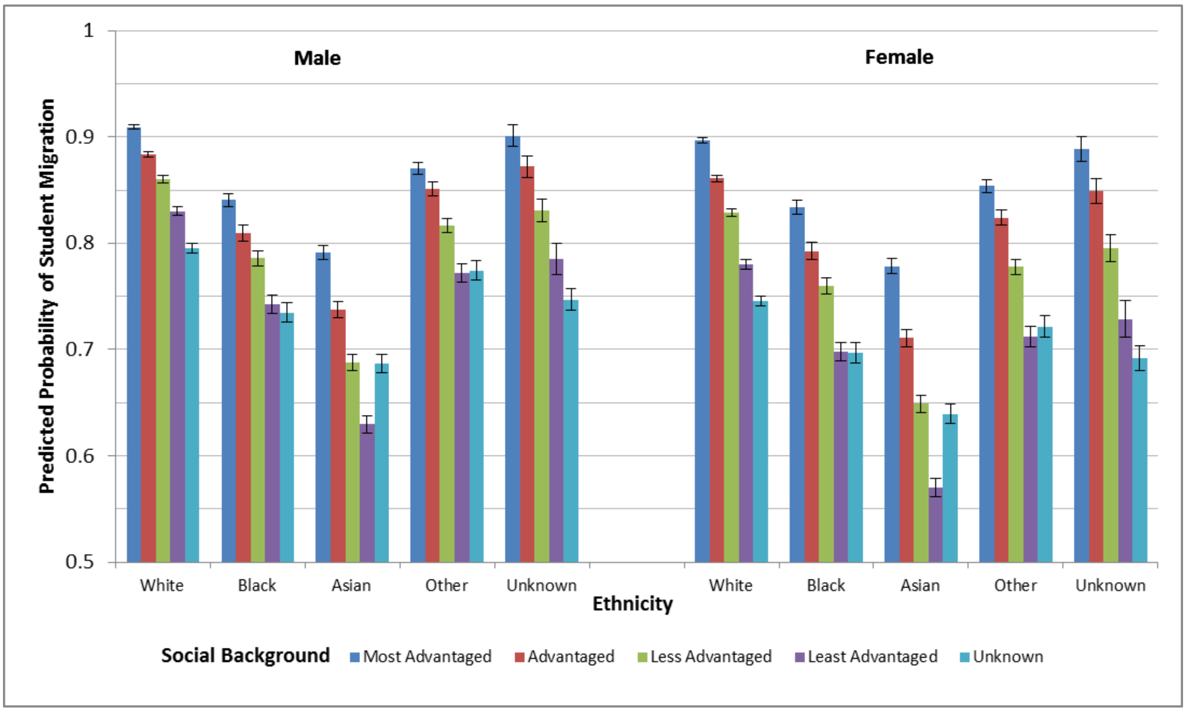
Standard errors in parentheses
Notation “0.000” refers to P-Values smaller the 5 x 10-4
[a] Denotes Reference Category

Source: Higher Education Statistics Agency (2014b)

Note: International students are omitted

As a result of the interaction terms in the model, the easiest way of interpreting the results of ethnicity, social background and gender is by calculating their predicted probability values of being a student migrant. The predicted probabilities are calculated by substituting the coefficients in Table 4-5 into the regression Equation 4.1. The results for social background, ethnicity and gender are graphically displayed in Figure 4-2 and Figure 4-3. Figure 4-2 and Figure 4-3 graph the same predicted probabilities; however Figure 4-2 uses social background as the focus variable, whereas Figure 4-3 switches the focus to ethnicity.

Figure 4-2: Logistic Regress Model - Predicted Probabilities of Student Migration by gender and ethnicity, for different social backgrounds.

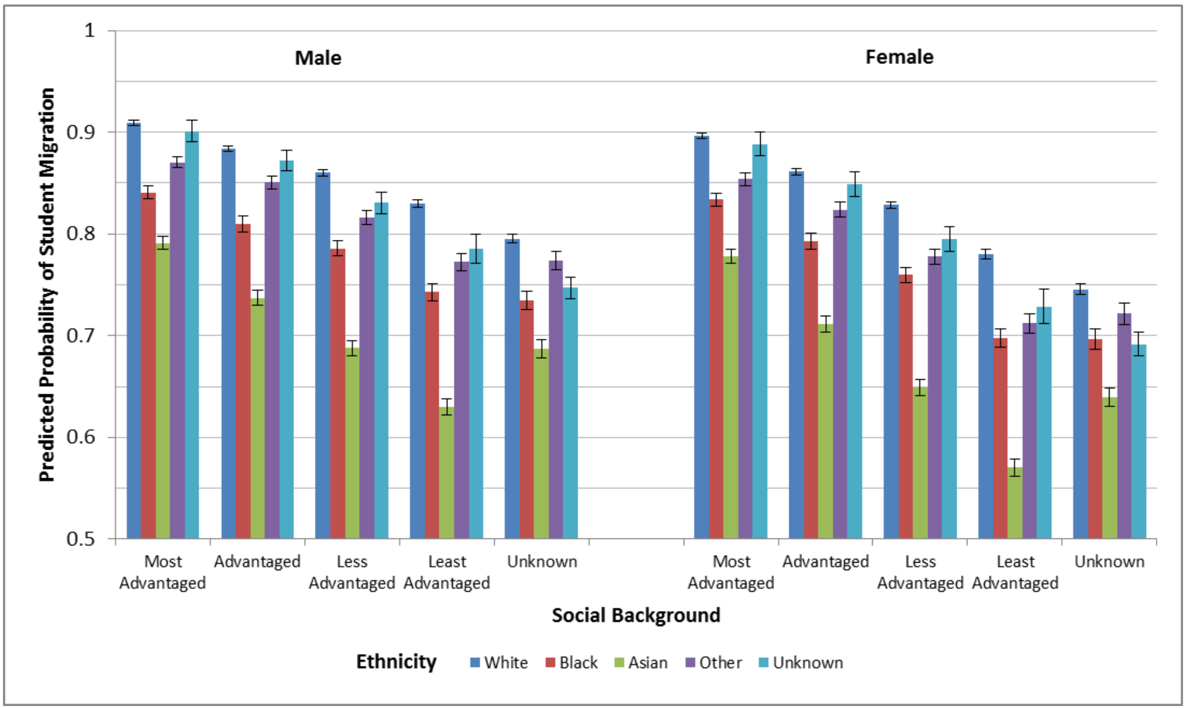


Source: Higher Education Statistics Agency (2014b)
Note: The predicted probabilities assume all other variables in the model were set to the reference category. Error Bars represent 95% CIs.

When studying the patterns in the two figures, it is important to note that the differences between the variables in the graphs are not parallel. This indicates that the interaction terms in the model are significant and without the interaction terms these differences would not have been visible.

In terms of ethnicity, social background and gender the predicted probabilities of student migration vary significantly. Fixing the remaining variables in the model at the reference category, it is clear to see that a White, most advantaged social background male is the type of student that has the highest predicted probability of making a student migration. In contrast, an Asian, least advantaged social background female has the lowest predicted probability of making a student migration.

Figure 4-3: Logistic Regress Model - Predicted Probabilities of Student Migration by gender and social background, for different ethnic groups.



Source: Higher Education Statistics Agency (2014b)
Note: The predicted probabilities assume all other variables in the model were set to the reference category. Error Bars represent 95% CIs.

It is also equally important to consider the impact of the student's social background on the probability of making a student migration. Those in the most advantaged group have the highest predicted probabilities than all other social backgrounds and this is the case for all ethnicities and both genders. Those students in the advantaged background group have the second highest predicted probabilities and the less advantaged group the third highest. Again, this is the case for all ethnicities and both genders. There is also a linear trend to the relationship between how advantageous the students social background and their predicted probability to migrate. As social background advantageousness declines so does the predicted probability of making a student migration and this is clearly visible within the figures.

However, there is variation in these trends with regards to the least advantaged and unknown social background groups, as the order of the predicted probabilities changes as a result of ethnicity and gender. The least advantaged social background group has the second lowest predicted probabilities for White, Black and Unknown ethnicity males but the lowest probability for Asian and Other. For females the trend is similar, however, Black, least advantaged social background females also fall below the unknown social background predicted probability unlike their male counterparts.

Focusing on the differences between ethnicities it is clear that those students from the Asian ethnicity group were the least likely to migrate in order to attend a HEI irrespective of their social background or gender. Asian females however are even less likely to migrate than their male counterparts. In contrast, White students have the highest predicted probabilities irrespective of social background, while White males are again more likely to migrate than their White female counterparts. This trend supports the findings of Khambhaita and Bhopal (2013) who found that Asian female students were much more likely than White students to stay living in the parental/guardian home during the first year at university. In the current study a student staying in the parental/guardian home during the first year at university would be recorded as not making a migration and with the predicted probability of not migrating being highest for Asian females from the least advantaged social background it appears the findings here mirror those reported by Khambhaita and Bhopal (2013).

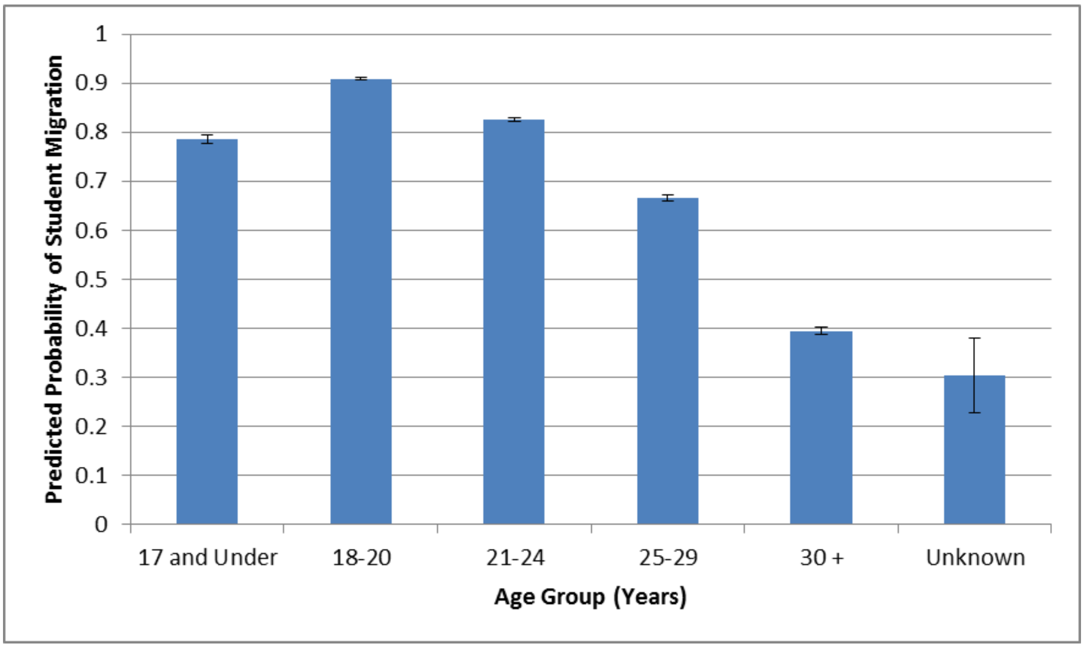
These results show that the social background of the student appeared to play a significant role in the likelihood of a student migrating in order to attend a HEI. This was suggested in the preliminary analysis and has been confirmed here when all other variables are considered simultaneously. These patterns tend to support the previous research regarding the tradition in the UK for HE students to migrate away from their parental home in order to study at a HEI. These findings are especially apparent for those from more traditional backgrounds, such as those from higher social classes and with supportive parents. There also appears to be an interaction between Asian students and the least advantaged social background group. Students of Asian ethnicity in the least advantaged groups were unlikely to migrate in order to study. As a result of these findings further and more in-depth analysis, which could include detailed qualitative research, would allow for more detailed and policy relevant findings to be made. This further analysis would hopefully find the reasons why students from this social background and ethnic group were found to be significantly less likely to migrate to attend a HEI than peers from other social background and ethnic groups. This is important as this quantitative study can only analyse and identify patterns in the data but can only speculate as to why these findings are occurring.

The remainder of this section focuses on the second research question by analysing the probability of being a student migrant when a student's characteristics (other than ethnicity, social background and gender), course studied institution attended and geographical location of domicile are all controlled for. These remaining variables were not involved in any interaction terms and therefore their main effects can be interpreted individually. All the remaining control variables in the model are significant at the 1% level.

When considering the impact on the probability of migration as a result of a student's age, clear differences are visible. In comparison to the 18-20 years (the reference category) age group, all of the remaining age groups have odds of a success less than one, suggesting that students in these age groups were less likely to migrate than those aged 18-20 years. The decline in the predicted probabilities by age is clearly illustrated in Figure 4-4.

Another interesting trend was the probability of migration by year of student as shown in Figure 4-5. The students least likely to migrate are those in their first of study, while the probability of migration increases sequentially with each year of study. This may be caused be influenced by large number of ethnic minority students that tend to remain in the parental home, especially in the first year of study (Khambhaita and Bhopal 2013). The subsequent increase in the probability of migration as people progress through university may also be a result of students deciding to migrate after the initial decision to remain in the parental home and commute.

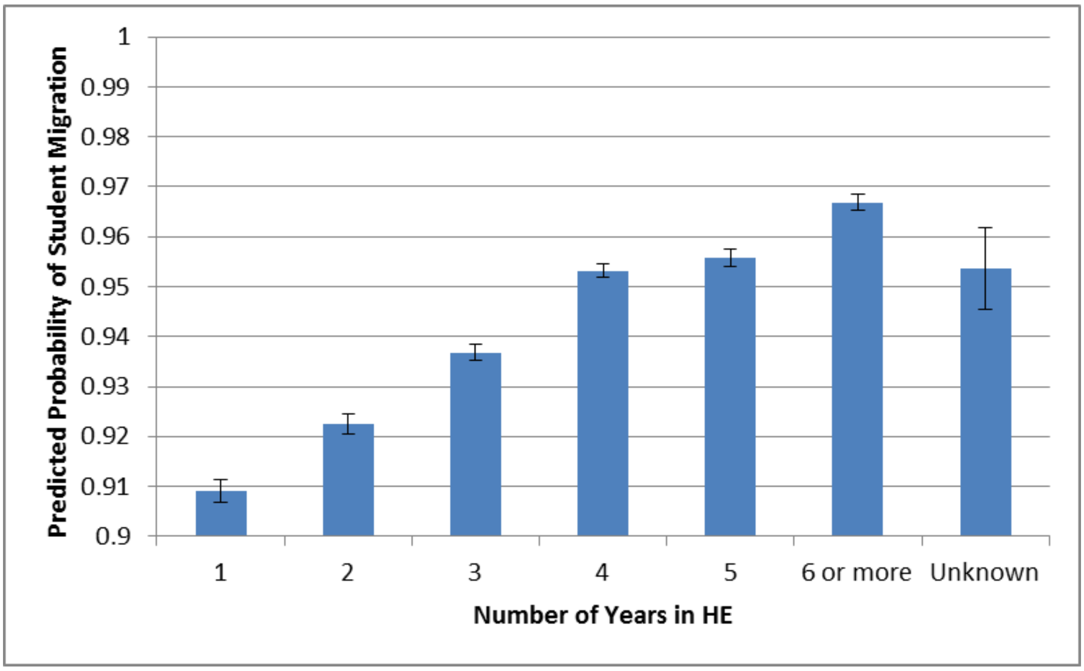
Figure 4-4: Logistic Regress Model - Predicted Probabilities of Student Migration by age group.



Source: Higher Education Statistics Agency (2014b)

Note: The predicted probabilities assume all other variables in the model were set to the reference category. Error Bars represent 95% CIs.

Figure 4-5: Logistic Regress Model - Predicted Probabilities of Student Migration by Number of Years in HE.



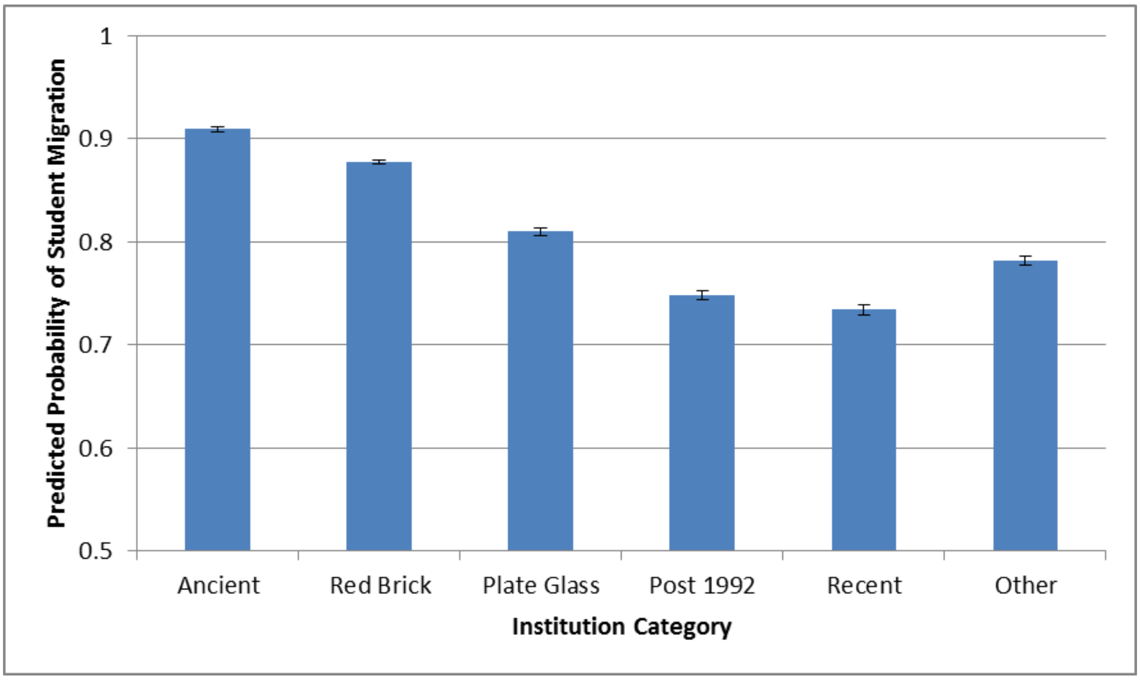
Source: Higher Education Statistics Agency (2014b)

Note: The predicted probabilities assume all other variables in the model were set to the reference category. Error Bars represent 95% CIs.

When considering the level of study of the student the data shows that, while holding all other variables constant, postgraduate students are more likely to migrate than undergraduates. However, although the variable is significant at the 1% level the difference in the predicted probabilities between the two groups are quite small.

There is also a clear pattern in the association between institution attended and the predicted probability to migrate in order to attend that HEI. The predicted probabilities by institution category, as shown in Figure 4-6, show a clear association between the two variables. As the prestige and reputation of the institution categories decrease so do the predicted probabilities of migrating in order to attend them. This suggests that a student is more likely to migrate in order to attend a higher reputable university. Whereas if the student is attending a more recent post 1992 institution then the student is much less likely to have migrated in order to do so.

Figure 4-6: Logistic Regress Model - Predicted Probabilities of Student Migration by Institution Category.



Source: Higher Education Statistics Agency (2014b)

Note: The predicted probabilities assume all other variables in the model were set to the reference category. Error Bars represent 95% CIs.

There were also some observed differences between courses studied and the predicted probability of migration. Those students that studied Agricultural or Veterinary courses and those studying Humanities have a much greater probability of migration than those who study Medicine (the reference category). Social or Human Science and Business or Law are very similar to the reference category, while students that study Combined degrees are much less likely to migrate in order to attend their HEIs.

In Chapter 3, clear geographical differences were identified in the number of students who were student migrants or student commuters depending on whether they were domiciled from a Northern or Southern region. As a result, an indicator variable is included into the model to see if any difference between the predicted probabilities of being a student migrant occurred as a result of a student being domiciled in a Southern or Northern region. The exponential of the coefficient for students from a Southern domicile region, as shown in Table 4-2, shows that students from the south were 1.68 times more likely to be a student migrant compared to those students from a Northern domicile region. This finding supports the conclusion made in Chapter 3 that there are clear spatial differences in the patterns of student migration as shown from evaluating data using the unique student migration typology and student migration area classification in Figure 3-5 and Figure 3-6.

4.6.2 Distance Travelled

Every student in the HESA student record data has the total distance they have travelled in order to attend a HEI measured using ArcGIS software. The process in which this was conducted was explained in Section 4.3.1. The distance was recorded by calculating the number of kilometres the student migrated or commuted in order to attend a HEI by measuring the distance between the centroid locations of the students domicile, term-time address and institution location. These distances vary from 0km for those students who study at their local institution to a maximum of 1723km, while the mean distance travelled across all the non-international students (including those that travelled 0km) is 90.6km.

In the current sub-section, a Tobit Model (Section 4.5.2) is used to analyse the effects of ten explanatory variables and three interaction terms on the predicted total distance travelled for a student to attend a HEI, and the results of this model are shown in Table 4-6.

Table 4-6: Tobit Model results of the association between total distance travelled to attend a HEI and the students' characteristic variables

VARIABLES	Coefficient (β)	Sig	SE	95% Confidence Interval
Constant	184.2	0.000	(0.592)	183.0 - 185.4
Ethnicity				
White[a]				
Black	-38.12	0.000	(1.036)	-40.16 - -36.09
Asian	-45.03	0.000	(0.880)	-46.76 - -43.31
Other (Including Mixed Race)	-18.26	0.000	(1.072)	-20.36 - -16.16
Unknown	-11.26	0.000	(2.554)	-16.26 - -6.252
Social Background				
Most Advantaged[a]				
Advantaged	-12.98	0.000	(0.433)	-13.83 - -12.14
Less Advantaged	-19.16	0.000	(0.420)	-19.98 - -18.34
Least Advantaged	-28.96	0.000	(0.449)	-29.84 - -28.08
Unknown	-23.39	0.000	(0.456)	-24.29 - -22.50
Gender				
Male[a]				
Female	-5.989	0.000	(0.404)	-6.781 - -5.198
Subject				
Medicine[a]				
Science/Engineering	2.176	0.000	(0.283)	1.622 - 2.731
Argicultural/Veterinary	51.73	0.000	(0.781)	50.20 - 53.26
Social/Human	-5.418	0.000	(0.281)	-5.969 - -4.867
Business/Law	-0.459	0.136	(0.308)	-1.063 - 0.144
Humanities	11.16	0.000	(0.288)	10.59 - 11.72
Combined	-28.63	0.000	(0.873)	-30.34 - -26.91
Institution Category				
Ancient[a]				
Red Brick	-44.51	0.000	(0.395)	-45.29 - -43.74
Plate Glass	-54.89	0.000	(0.424)	-55.72 - -54.06
Post 1992	-71.43	0.000	(0.386)	-72.19 - -70.68
Recent University	-69.95	0.000	(0.454)	-70.84 - -69.06
Other	-59.11	0.000	(0.496)	-60.08 - -58.13
Age				
17 years and under	-23.66	0.000	(1.068)	-25.76 - -21.57
18-20 years[a]				
21-24 years	-14.19	0.000	(0.242)	-14.66 - -13.72
25-29 years	-41.11	0.000	(0.335)	-41.76 - -40.45
30 years and over	-47.58	0.000	(0.275)	-48.12 - -47.04
Age unknown	-31.15	0.000	(4.663)	-40.29 - -22.01
Number of Years in HE				
1[a]				
2	-0.589	0.004	(0.205)	-0.991 - -0.188
3	1.477	0.000	(0.231)	1.025 - 1.930
4	11.04	0.000	(0.366)	10.32 - 11.76
5	8.118	0.000	(0.670)	6.804 - 9.432
6 or more	11.50	0.000	(0.963)	9.616 - 13.39
Unknown	-16.97	0.000	(4.480)	-25.75 - -8.193
Level of Study				
Post-Graduate[a]				
Under-Graduate	-14.44	0.000	(0.266)	-14.96 - -13.92
Domicile				
North[a]				
South	33.47	0.000	(0.172)	33.14 - 33.81
Interaction Terms				
Ethnicity * S.Background				
White*Most Advantaged[a]				
Black*Advantaged	-0.372	0.785	(1.364)	-3.045 - 2.301

VARIABLES	Coefficient (β)	Sig	SE	95% Confidence Interval
Black*Less Advantaged	1.570	0.181	(1.175)	-0.732 - 3.873
Black*Least Advantaged	2.455	0.039	(1.187)	0.128 - 4.782
Black*Unknown	3.439	0.005	(1.211)	1.065 - 5.813
Asian*Advantaged	-1.285	0.253	(1.124)	-3.488 - 0.918
Asian*Less Advantaged	-1.572	0.124	(1.023)	-3.576 - 0.432
Asian*Least Advantaged	-2.173	0.023	(0.992)	-4.118 - -0.228
Asian*Unknown	8.924	0.000	(1.061)	6.845 - 11.00
Other*Advantaged	1.347	0.342	(1.416)	-1.429 - 4.123
Other*Less Advantaged	-3.049	0.018	(1.290)	-5.577 - -0.521
Other*Least Advantaged	-10.08	0.000	(1.353)	-12.73 - -7.430
Other*Unknown	-7.680	0.000	(1.379)	-10.38 - -4.978
Unknown*Advantaged	11.84	0.000	(3.159)	5.646 - 18.03
Unknown*Less.Adv	-2.147	0.469	(2.968)	-7.963 - 3.669
Unknown*Least.Adv	-2.419	0.425	(3.029)	-8.357 - 3.519
Unknown*Unknown	1.723	0.511	(2.623)	-3.418 - 6.863
S.Background*Gender				
V. Advantaged*Male[a]				
Advantaged*Female	-3.962	0.000	(0.559)	-5.059 - -2.866
Less Advantaged*Female	-6.569	0.000	(0.531)	-7.610 - -5.527
Least Advantaged*Female	-7.254	0.000	(0.551)	-8.333 - -6.174
Unknown*Female	-10.02	0.000	(0.545)	-11.09 - -8.955
Ethnicity*Gender				
White*Male[a]				
Black*Female	4.573	0.000	(0.726)	3.151 - 5.996
Asian*Female	3.094	0.000	(0.595)	1.927 - 4.260
Other*Female	0.356	0.678	(0.860)	-1.328 - 2.041
Unknown*Female	1.509	0.221	(1.232)	-0.906 - 3.924
Observations	1,797,492			
ML (Cox-Snell) R2	0.127			

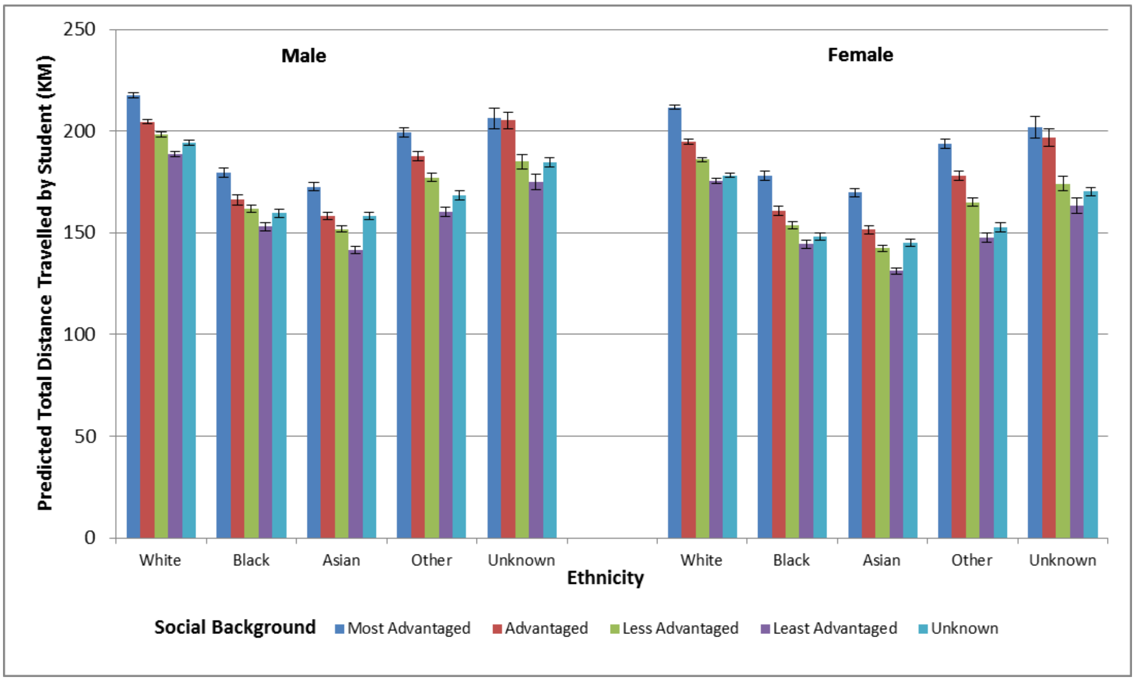
Standard errors in parentheses
Notation “0.000” refers to P-Values smaller the 5 x 10^{-4}
[a] Denotes Reference Category

Source: Higher Education Statistics Agency (2014b)
Note: International students are omitted

As in the logistic regression methodology; the ethnicity, social background and gender variables are interacted with each other in order to answer one of the main research questions of the study. Again, due to these three variables being involved in interaction terms the findings for these three variables need to be interpreted together. The predicted total distance travelled by a student by ethnicity, background and gender are shown in Figure 4-7 and Figure 4-8.

The overall results seemed very similar to those produced when investigating the probability of student migration. Those students predicted to travel the largest distances are the same group that had the highest probability of migration, White, most advantaged social background Males. Similarly, the group predicted to travel the shortest distances were Asian, least advantaged social background Females.

Figure 4-7: Tobit Regression Model - Predicted Total Distance Travelled by gender and ethnicity, for different social backgrounds.

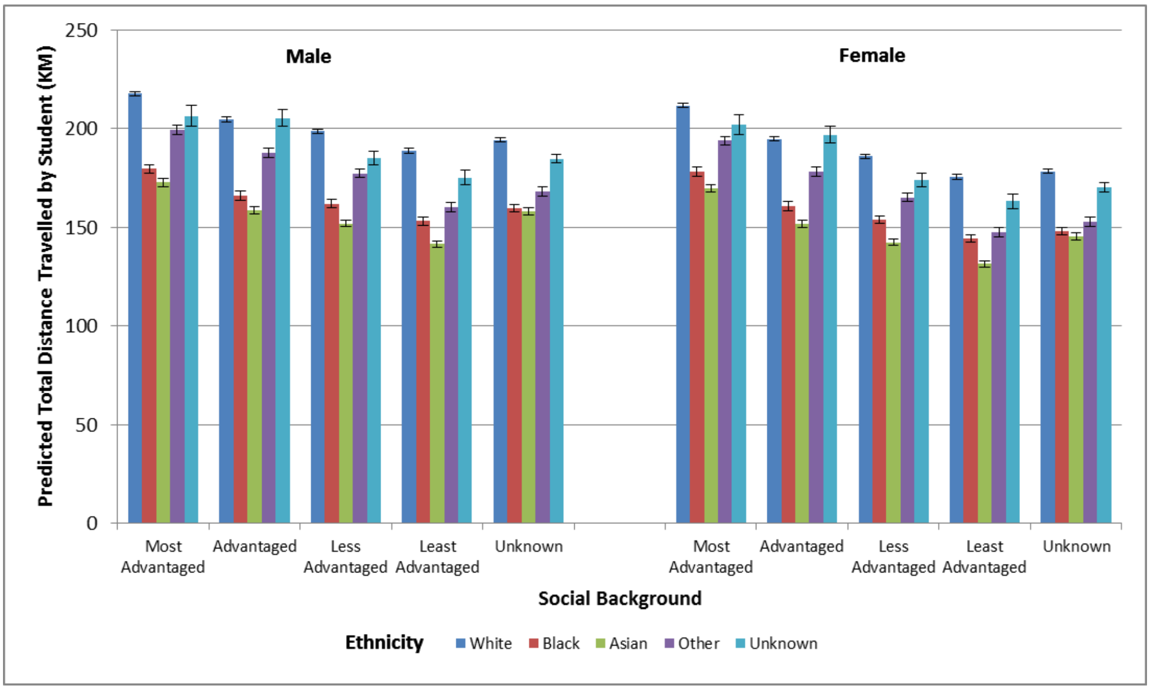


Source: Higher Education Statistics Agency (2014b)

Note: The predicted total distances travelled assume all other variables in the model were set to the reference category. Error Bars represent 95% CIs.

When considering the impact of a student's social background on the predicted distance travelled to attend a HEI, the patterns are again very similar to those observed in the probability of migration results (Note: the similarities between Figure 4-2 and Figure 4-7). The one noticeable difference as a result of social background between the probability to migrate and the predicted total distance travelled was for the unknown social background category. The predicted total distances for students with social background unknown were much higher than expected and much higher when compared to their predicted probability of migration. This may indicate some form of non-reporting bias in the data. Because the data is population data, this technically is not non-response but there seems to be some association between the migration outcome and those students that there social background was recorded as unknown.

Figure 4-8: Tobit Regression Model - Predicted Total Distance Travelled by gender and social background, for different ethnic groups.



Source: Higher Education Statistics Agency (2014b)

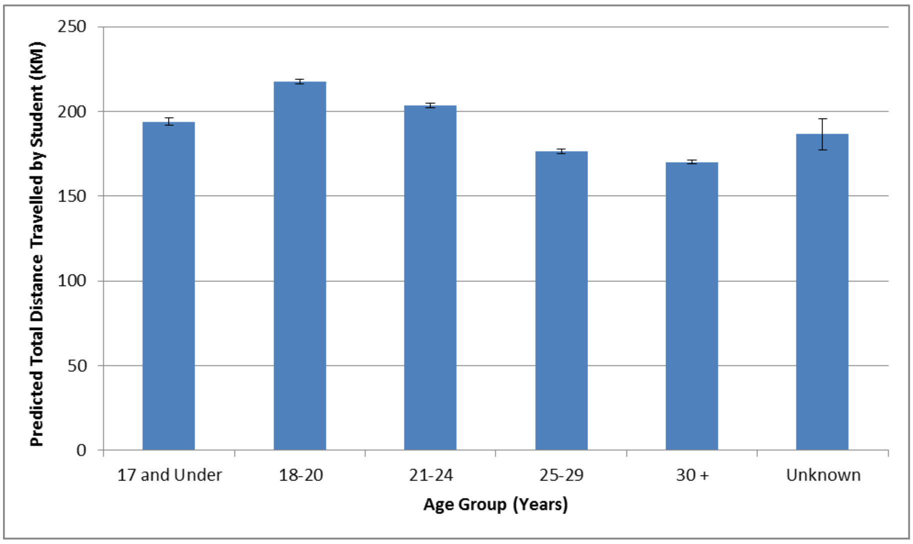
Note: The predicted total distances travelled assume all other variables in the model were set to the reference category. Error Bars represent 95% CIs.

Shifting the focus to ethnicity, the trends are again very similar with regards to predicted distance as they were with predicted probability of migration (again note the similarities between Figure 4-3 and Figure 4-8). Those students of Asian ethnicity are predicated to travel the shortest distances and the difference between the Asian and Other ethnic groups are quite large. Again, female students are predicted to travel shorter distances than their male counterparts. The White group are predicted to travel the furthest distances across all social backgrounds with the exception of the advantaged social background group.

The remaining variables in the model are not involved in any interaction terms and as a result their main effects can be interpreted individually. All the remaining control variables in the model are again significant at the 1% level.

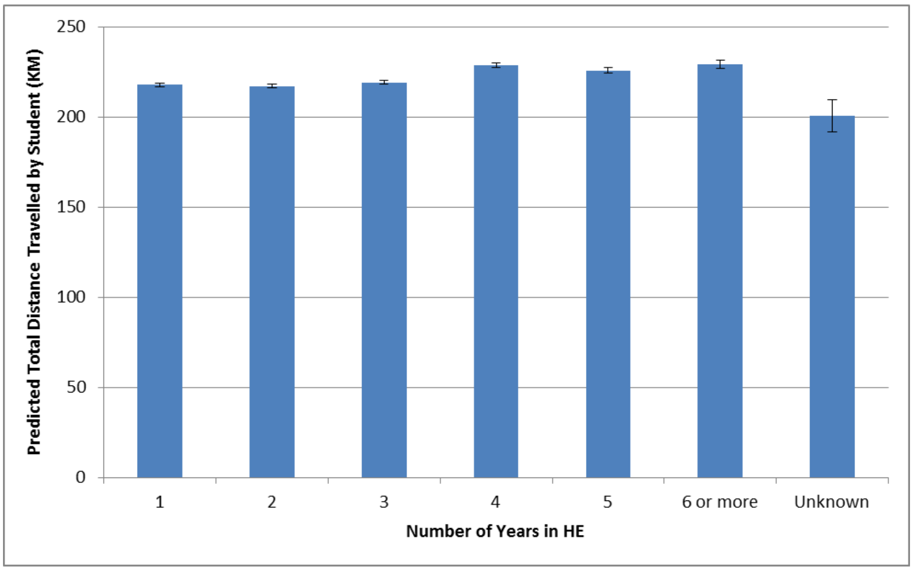
The direction and strength of the associations between the remaining variables and the predicted distance travelled again mirrored those found in the logistic regression model results. The effect of age (Figure 4-9) on distance is the same as observed on the probability of migration (Figure 4-4), the average predicted distance travelled by a student declines as age increased. While the effect of the number of years in HE (Figure 4-10) has the same direction as observed on the probability of migration (Figure 4-5) the differences in the average distance travelled do not change significantly between the year groups as shown by the overlap in the 95% confidence intervals.

Figure 4-9: Tobit Regression Model - Predicted Total Distance Travelled by Age Group



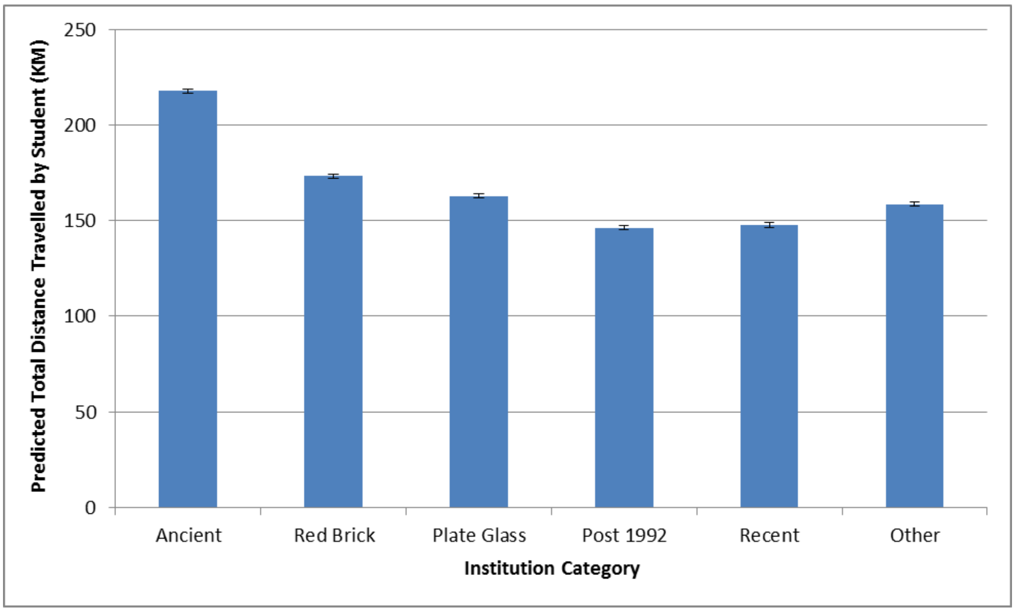
Source: Higher Education Statistics Agency (2014b)
Note: The predicted total distances travelled assume all other variables in the model were set to the reference category. Error Bars represent 95% CIs.

Figure 4-10: Tobit Regression Model - Predicted Total Distance Travelled by Number of Years in HE



Source: Higher Education Statistics Agency (2014b)
Note: The predicted total distances travelled assume all other variables in the model were set to the reference category. Error Bars represent 95% CIs.

Figure 4-11: Tobit Regression Model - Predicted Total Distance Travelled by Institution Category



Source: Higher Education Statistics Agency (2014b)
Note: The predicted total distances travelled assume all other variables in the model were set to the reference category. Error Bars represent 95% CIs.

There are significant differences between the predicted distances travelled and the instituion category of the HEI attended as shown in Figure 4-11. As seen in the probability to migrate, the predicted distance travelled are highest for the higher more prestigous instituion categories and the declines as reputation decreases. The ancient institutions have the highest prdcited distance. This is not surprising given the relatively small number of these institutions but the very high reputaiton and prestigure associated with these HEIs. As a result, it is not surprising that on average students were willing to travel further distances to attend the anicent universities in comparison to the more recent and less reputable post 1992 or recent HEIs.

As mentioned in Section 4.5.2, it is possible to run a number of different statistical techniques to analyse what factors impact on the migraiton distance and one such method is Zero Inflated Poisson (ZIP) Models.

ZIP models are different to the Tobit model as they predicted the probability of a distance being 0km and the predicted distance for those students that did not have a distance of 0km seperatley. The modelling techinque conducted using the ZIP method was the same as for the Tobit model and the final ZIP model had 10 explanatory variables and three interaction terms as shown in Table 4-7.

Table 4-7: Zero Inflated Poisson Model results of the association between total distance travelled to attend a HEI and the students' characteristic variables

	Inflate (prob. of 0Km)		Total Distance	
VARIABLES	**Coefficient (β)**	**Sig**	**Coefficient (β)**	**Sig**
Constant	-2.744	0.000	5.419	0.000
Ethnicity				
White[a]				
Black	0.620	0.000	-0.241	0.000
Asian	0.757	0.000	-0.308	0.000
Other (Including Mixed Race)	0.395	0.000	-0.104	0.000
Unknown	0.412	0.000	-0.058	0.000
Social Background				
Most Advantaged[a]				
Advantaged	0.296	0.000	-0.070	0.000
Less Advantaged	0.501	0.000	-0.097	0.000
Least Advantaged	0.556	0.000	-0.179	0.000
Unknown	0.552	0.000	-0.115	0.000
Gender				
Male[a]				
Female	0.112	0.000	-0.036	0.000
Subject				
Medicine[a]				
Science/Engineering	0.208	0.000	0.072	0.000
Argicultural/Veterinary	-0.690	0.000	0.403	0.000
Social/Human	0.234	0.000	-0.009	0.000
Business/Law	0.270	0.000	0.059	0.000
Humanities	0.145	0.000	0.149	0.000
Combined	0.783	0.000	-0.068	0.000
Institution Category				
Ancient[a]				
Red Brick	-0.448	0.000	-0.374	0.000
Plate Glass	-0.453	0.000	-0.470	0.000
Post 1992	-0.210	0.000	-0.625	0.000
Recent University	-0.274	0.000	-0.609	0.000
Other	-0.235	0.000	-0.477	0.000
Age				
17 years and under	0.465	0.000	-0.179	0.000
18-20 years[a]				
21-24 years	0.547	0.000	-0.084	0.000
25-29 years	1.071	0.000	-0.288	0.000
30 years and over	0.991	0.000	-0.424	0.000
Age unknown	0.626	0.000	-0.236	0.000
Number of Years in HE				
1[a]				
2	-0.004	0.463	-0.006	0.000
3	-0.037	0.000	0.012	0.000
4	-0.184	0.000	0.091	0.000
5	-0.118	0.000	0.085	0.000
6 or more	-0.042	0.088	0.147	0.000
Unknown	0.242	0.045	-0.067	0.000
Level of Study				
Post-Graduate[a]				
Under-Graduate	0.064	0.000	-0.188	0.000
Domicile				
North[a]				
South	-1.184	0.000	0.182	0.000
Interaction Terms				
Ethnicity * S.Background				
White*Most Advantaged[a]				

	Inflate (prob. of 0Km)		Total Distance	
VARIABLES	**Coefficient (β)**	**Sig**	**Coefficient (β)**	**Sig**
Black*Advantaged	-0.063	0.165	-0.080	0.000
Black*Less Advantaged	-0.212	0.000	-0.135	0.000
Black*Least Advantaged	-0.085	0.028	-0.158	0.000
Black*Unknown	-0.121	0.002	-0.126	0.000
Asian*Advantaged	0.008	0.820	-0.067	0.000
Asian*Less Advantaged	-0.049	0.120	-0.110	0.000
Asian*Least Advantaged	0.084	0.006	-0.138	0.000
Asian*Unknown	-0.201	0.000	-0.014	0.000
Other*Advantaged	-0.012	0.811	-0.003	0.011
Other*Less Advantaged	-0.092	0.041	-0.070	0.000
Other*Least Advantaged	0.153	0.001	-0.141	0.000
Other*Unknown	0.102	0.020	-0.099	0.000
Unknown*Advantaged	-0.344	0.002	0.060	0.000
Unknown*Less.Adv	-0.018	0.851	-0.006	0.014
Unknown*Least.Adv	-0.148	0.140	-0.055	0.000
Unknown*Unknown	-0.088	0.338	0.023	0.000
S.Background*Gender				
V. Advantaged*Male[a]				
Advantaged*Female	-0.015	0.469	-0.043	0.000
Less Advantaged*Female	-0.013	0.484	-0.077	0.000
Least Advantaged*Female	-0.015	0.407	-0.122	0.000
Unknown*Female	-0.030	0.099	-0.143	0.000
Ethnicity*Gender				
White*Male[a]				
Black*Female	-0.036	0.068	-0.011	0.000
Asian*Female	0.074	0.000	0.012	0.000
Other*Female	0.043	0.083	-0.007	0.000
Unknown*Female	-0.040	0.193	0.003	0.003
Observations		1,797,492		
ML (Cox-Snell) R2				

Source: Higher Education Statistics Agency (2014b)

Notation “0.000” refers to P-Values smaller the 5 x 10-4

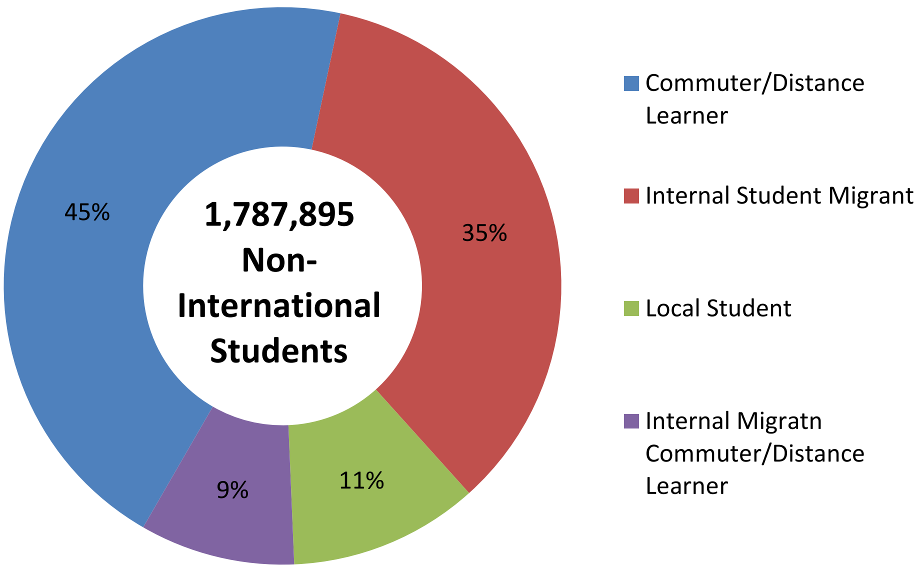
Note: International students are omitted

The over-riding findings using the ZIP model show confounding results that suggest similar overal results as shown using the logisitic and tobit modelling. The first section of the ZIP models the probabilty of a student have a distance of 0km or non-0km, and these results show very similar results to those produced using the logistic model to predict the probabiltiy of a student migrating. The second secontion of the ZIP models the impact of the explanatory factors on the predicted distance travelled for those indivduals that have a non-0km distance. Again, these results produce very similar results to those produced usign the Tobit model. However, as disscussed in Section 4.5.2, the ZIP method is desgined for modelling count data and the data used here is continous data measuring distance. Therefore, although the results produce similar findings, the result produced using to Tobit model are demmed to be better suited to the data type and are preffered over the ZIP models outputs.

4.6.3 Student Migration Type

The final outcome variable used to measure the student migration choice is the typology of student migration proposed in Chapter 3. The typology of student migration categorised every student registered at a UK HEI into one of eight categories that depicted the type of migration the student experienced in order to attend a HEI. As previously mentioned, the analysis presented in this chapter only uses four of the eight categories of student migration. The distributions of the student sub-population between these four categories are illustrated in Figure 4-12.

Figure 4-12: Breakdown of Student Population by Student Migration Category



Source: Higher Education Statistics Agency (2014b)

Note: ‘International’, ‘Migrant Commuter or Distance Learner attending a local HEI’ and ‘Unknown’ student migration categories are omitted

The results in this sub-section develop on the binary analysis of migration from the first outcome variable. The four categories of student internal migration are directly linked to the binary indicator of migration. Local student and commuter/distance learners do not migrate and were classified as not migrating in the first outcome variable; while in contrast, the internal student migrant and internal migrant commute/distance learner categories do migrate and were recorded in the migration yes category. The analysis in this sub-section acknowledges and takes into the extra complexity behind the phenomenon of student migration which was brought to the fore in Chapter 3.

A multinomial logistic regression model is used to analyse the probability of a student being in one of the four student migration categories and the coefficients of the multinomial logistic model are shown in Table 4-8. The final chosen multinomial logistic regression model includes no interaction terms and as a result the main effects of all the variables in the model can be interpreted individually.

The predicted probabilities represent the probability of being in each of the student categories when holding the other variables within the model constant at the reference category. Therefore, when looking at the differences between variables one should look at the probability of being in each category independently to the other categories; this is also a function of the IIA assumption.

The impact of the student's ethnic group on the predicted probabilities of being in each of the student migration categories is illustrated in Figure 4-13. There are some clear differences in the predicted probabilities as a result of the students' ethnicity and these differences again illustrate similar findings to the previous two outcome variables.

The predicted probability of being a local student is significantly higher for Asian students than any other of the ethnic groups. This finding echo's the results found using the previously mentioned methods and supports the findings of previous research that suggests that Asian students were more likely to remain in the parental home during HE than the other ethnic groups (Khambhaita and Bhopal 2013). Asian students are also the most likely to be a commuter/distance learner, which again supports the idea that Asian students tend to remain in the parental home while studying at a HEI. In contrast, the White ethnic group have the lowest predicted probabilities of being a commuter/distance learner and a local student. Again, these results support the findings found using the previous two outcome variables.

In contrast to the local and commuter/distance learner students are those students who made an internal migration to attend a HEI. The White ethnic group have the highest predicted probabilities of being an internal student migrant while the Asian group have the lowest predicted probabilities. Again, these findings support those presented earlier where White students had the highest probability of making a migration in order to attend a HEI.

Table 4-8: Multinomial logistic regression results of the association between student migration categories and student characteristic variables

VARIABLES	Commuter		Migrant		Migrant Commuter	
	Coefficient (β)	Sig	Coefficient (β)	Sig	Coefficient (β)	Sig
Constant	0.719	0.000	2.907	0.000	-0.121	0.000
Ethnicity						
White[a]						
Black	-0.344	0.000	-0.882	0.000	-0.517	0.000
Asian	-0.441	0.000	-1.361	0.000	-1.009	0.000
Other (Including Mixed Race)	-0.356	0.000	-0.663	0.000	-0.441	0.000
Unknown	-0.223	0.000	-0.494	0.000	-0.316	0.000
Social Background						
Most Advantaged[a]						
Advantaged	-0.0384	0.001	-0.322	0.000	-0.317	0.000
Less Advantaged	-0.0938	0.000	-0.644	0.000	-0.463	0.000
Least Advantaged	-0.0814	0.000	-0.921	0.000	-0.723	0.000
Unknown	0.00952	0.367	-1.029	0.000	-0.702	0.000
Gender						
Male[a]						
Female	-0.0204	0.000	-0.259	0.000	-0.212	0.000
Subject						
Medicine[a]						
Science/Engineering	-0.280	0.000	0.0966	0.000	-0.359	0.000
Agricultural/Veterinary	0.532	0.000	0.773	0.000	1.430	0.000
Social/Human	-0.236	0.000	-0.0588	0.000	-0.552	0.000
Business/Law	-0.265	0.000	-0.109	0.000	-0.428	0.000
Humanities	-0.342	0.000	0.293	0.000	0.0438	0.000
Combined	-0.544	0.000	-1.089	0.000	-1.126	0.000
Institution Category						
Ancient[a]						
Red Brick	0.826	0.000	-0.0114	0.378	1.520	0.000
Plate Glass	1.079	0.000	-0.318	0.000	1.133	0.000
Post 1992	0.971	0.000	-0.921	0.000	1.131	0.000
Recent University	1.046	0.000	-0.837	0.000	0.868	0.000
Other	0.924	0.000	-0.872	0.000	1.455	0.000
Age						
17 years and under	0.0872	0.003	-1.048	0.000	-0.845	0.000
18-20 years[a]						
21-24 years	-0.187	0.000	-0.954	0.000	-0.555	0.000
25-29 years	-0.395	0.000	-2.376	0.000	-1.043	0.000
30 years and over	-0.137	0.000	-3.730	0.000	-1.729	0.000
Age unknown	0.318	0.012	-4.015	0.000	-1.636	0.000
Number of Years in HE						
1[a]						
2	-0.0507	0.000	0.0841	0.000	0.214	0.000
3	-0.108	0.000	0.245	0.000	0.369	0.000
4	-0.0932	0.000	0.584	0.000	0.624	0.000
5	-0.134	0.000	0.525	0.000	0.796	0.000
6 or more	-0.239	0.000	0.595	0.000	1.046	0.000
Unknown	-0.341	0.010	0.129	0.386	1.154	0.000
Level of Study						
Post-Graduate[a]						
Under-Graduate	-0.0831	0.000	0.00470	0.633	-0.546	0.000
Domicile						
South[a]						
North	1.077	0.000	1.326	0.000	1.660	0.000
Observations			1,787,895			
ML (Cox-Snell) R2			0.1843			

Standard errors in parentheses

[a] Denotes Reference Category

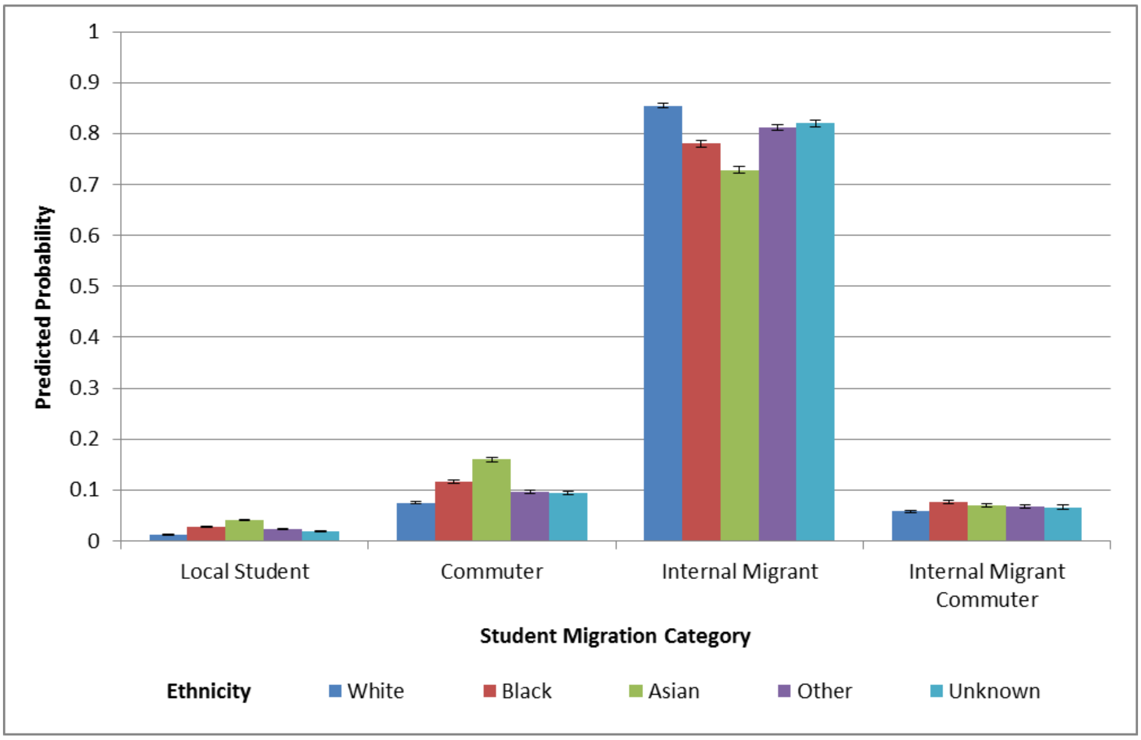
Notation “0.000” refers to P-Values smaller the 5 x 10-4

The reference category for the response variable was ‘Local Student’

Source: Higher Education Statistics Agency (2014b)

Note: ‘International’, ‘Migrant Commuter or Distance Learner attending a local HEI’ and ‘Unknown’ student migration categories are omitted

Figure 4-13: Multinomial Regression Model - Predicted probabilities of being in one of the Student Migration Categories by ethnic group



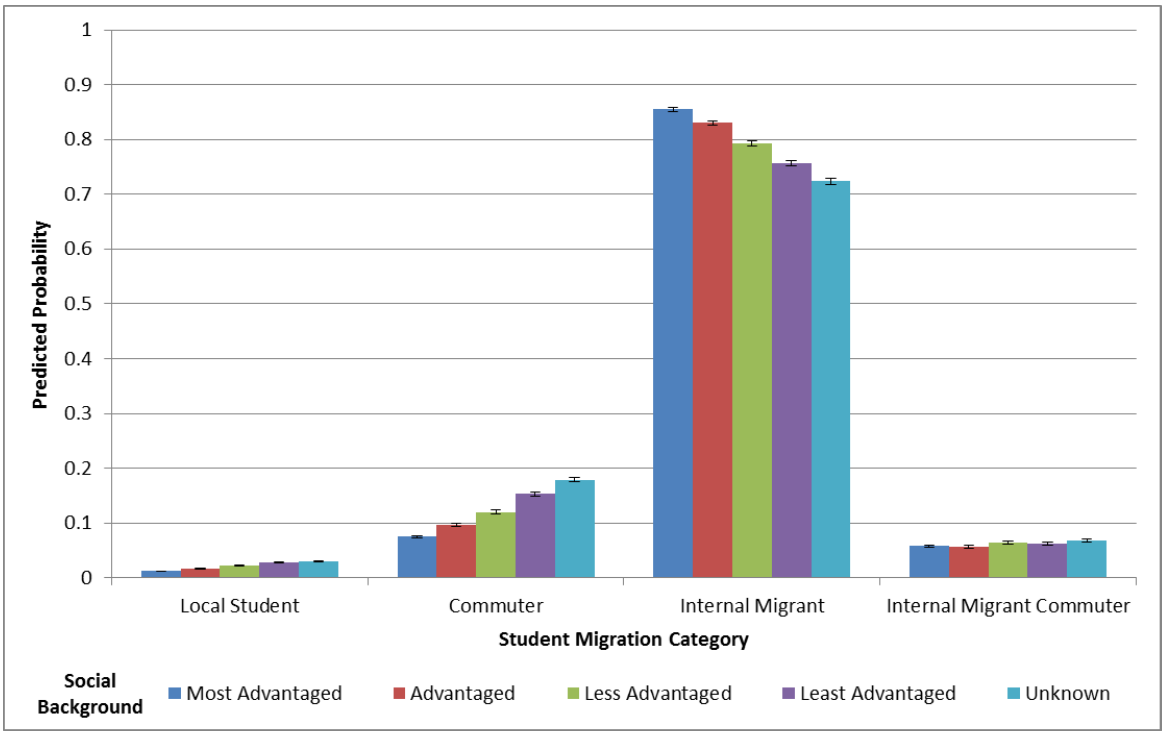
Source: Higher Education Statistics Agency (2014b)

Note: The predicted probabilities assume all other variables in the model were set to the reference category. Error Bars represent 95% CIs.

However, a new pattern arises within the migrant commuter group that has not been highlighted within the previous outcome variables. The Black ethnic group have the highest predicted probability of all the ethnic groups of being a student migrant commuter, closely followed by the Other ethnic group. A large percentage of these migrant commuter students were studying in London HEIs and they are recorded as crossing LA boundaries but still remain with the area of Greater London itself. This can also be linked back to the large Black ethnic group population in London that is not found in other LAs across the UK (Office for National Statistics 2013b).

There are also clear differences in the predicted probabilities of being in each of the student migration categories as a result of the student's social background. The predicted probabilities of the different migration categories by social background are shown in Figure 4-14.

Figure 4-14: Multinomial Regression Model - Predicted probabilities of being in one of the Student Migration Categories by social background



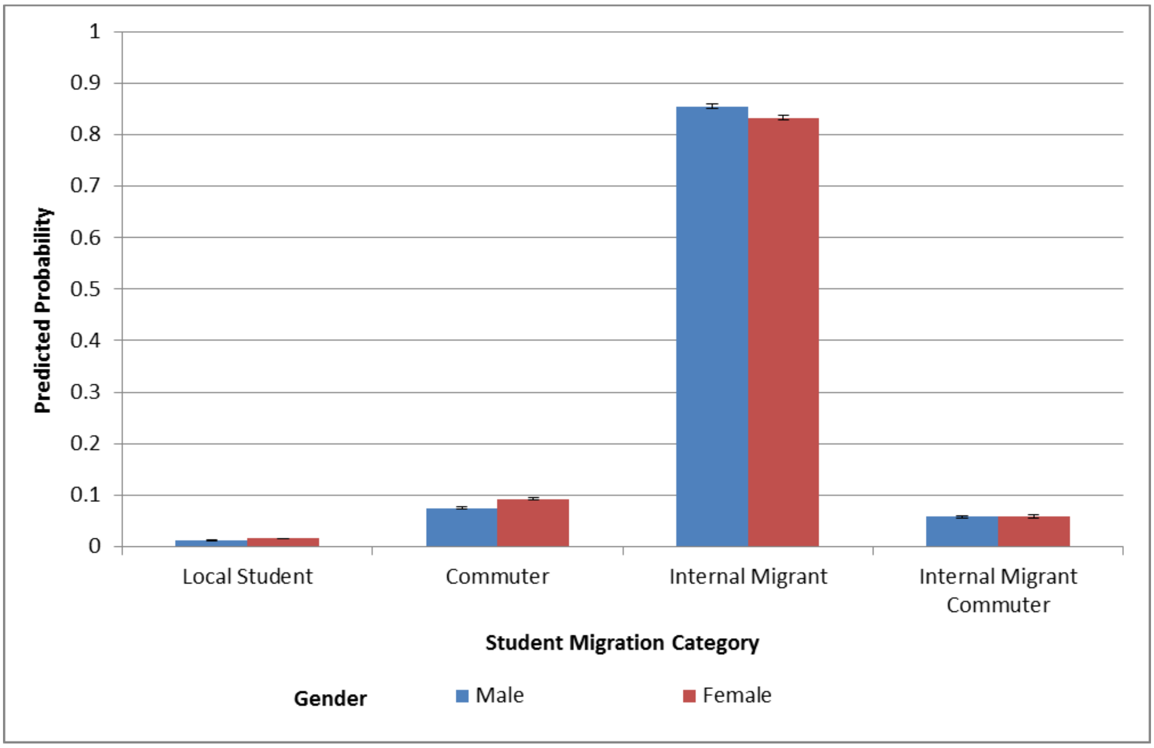
Source: Higher Education Statistics Agency (2014b)

Note: The predicted probabilities assume all other variables in the model were set to the reference category. Error Bars represent 95% CIs.

The predicted probabilities of being a commuter/distance learner, a local student and an internal student migrant appear to have a linear relationship with social background. The predicted probabilities for being a commuter/distance learner and a local student increase as social background advantageousness decreases, while the opposite trend is apparent for being an internal student migrant. These results again support the findings produced using the other outcome variables. These results were not surprising given that the creation of the social background variable was done in such a fashion that this variable was designed to show how advantageous an individual socio-economic variable would be towards making a student migration and the results here support this hypothesis. The only findings that stood out in the analysis, that did not when analysing the other methods, was the very high predicted probability of being a commuter/distance learner if your social background is unknown. However, there is no logical explanation from the literature or any previous research to explain why this is the case.

In the previous two methodologies there were also observed differences in the migration outcomes as a result of the student's gender. The predicted probabilities of the four migration categories by gender are shown in Figure 4-15. The differences in the predicted probabilities as an impact of gender are minimal, especially when compared to the differences by ethnicity and social background. Females have a higher predicted probability of being a local student than their male counterparts but the size of the difference is marginal. A similar but reversed pattern is seen for migrant commuters but again the difference between the probabilities is very small and insignificant. The differences between the genders for internal student migrants and commuter/distance learners are larger in size but again the differences are not substantial. Males are more likely than females to be internal student migrants and for commuter/distance learners the trend is reversed.

Figure 4-15: Multinomial Regression Model - Predicted probabilities of being in one of the Student Migration Categories by gender



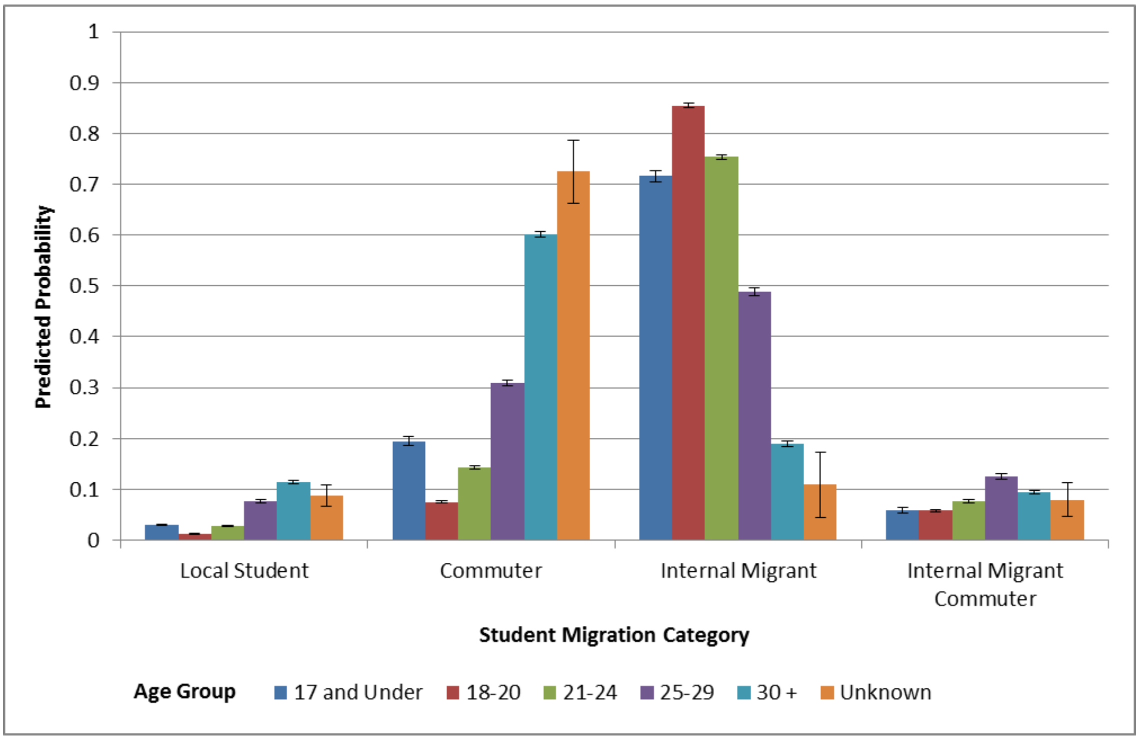
Source: Higher Education Statistics Agency (2014b)

Note: The predicted probabilities assume all other variables in the model were set to the reference category. Error Bars represent 95% CIs.

The only major differences by age were in the internal student migrant and commuter/distance learner groups, as illustrated in Figure 4-16. Those aged

18-20 years are the most likely to be internal student migrants and least likely to be local students. The predicted probability of being a commuter/distance learner increases with age. These findings were not surprising as previous empirical evidence has shown how migration intensity is interlinked with age (Wilson 2010) and as age increases so does the likelihood that the individual will have stronger ties to an area such as owning a house or having children and therefore the probability of that individual migrating to study would decrease.

Figure 4-16: Multinomial Regression Model - Predicted probabilities of being in one of the Student Migration Categories by age group

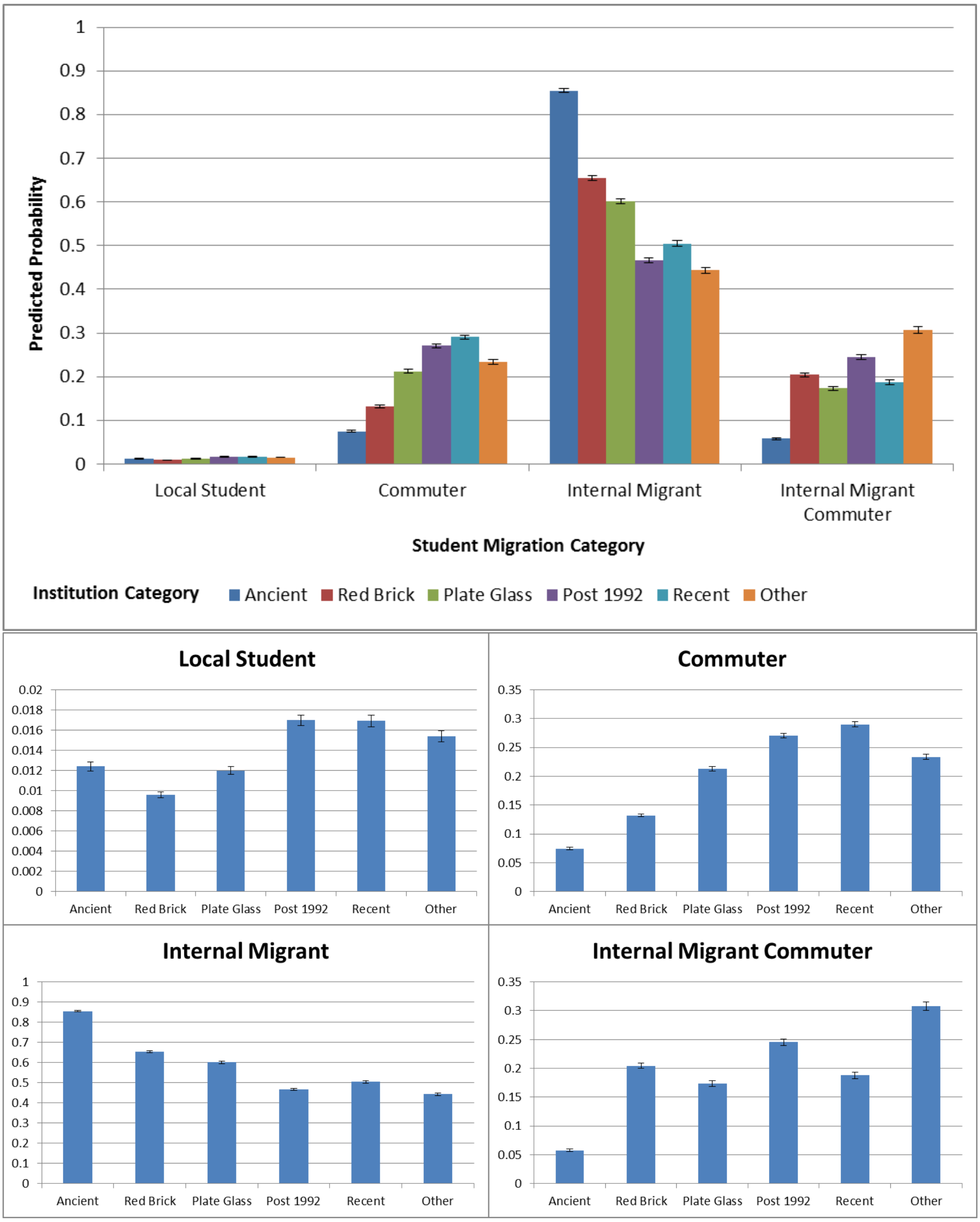


Source: Higher Education Statistics Agency (2014b)

Note: The predicted probabilities assume all other variables in the model were set to the reference category. Error Bars represent 95% CIs.

Finally, the impact of the category of institution attended on the predicted probability of a student being in one of the four student migration categories is shown in Figure 4-17. Those attending an ancient institution have the highest probability of being a student migrant. It is also clear that the probability of being a student migrant declines with the decline in institution category. The inverse pattern is visible with regards to student commuters with students attending post 1992 and recent HEIs having the highest predicted probabilities.

Figure 4-17: Multinomial Regression Model - Predicted probabilities of being in one of the Student Migration Categories by Institution Category



Source: Higher Education Statistics Agency (2014b)

Note: The predicted probabilities assume all other variables in the model were set to the reference category. Error Bars represent 95% CIs.

The patterns associated with being a local student are less clear. Due to the predicted probabilities being calculated with the other response variables in model constrained to the reference categories, the predicted probability of being a local student is very low (this is due to the reference category being set to white and most advantageous and from the previous results it has been shown these groups have the lowest probability of being of a local student). However, when the differences between the probabilities of being a local student as a result of the institution category are more closely examined, clear statistical differences are observed. The post 1992, recent and other HEIs have the highest predicted probability of being a local student and these probabilities decline for plate glass and red brick HEIs. These patterns are expected from previous results and make theoretical and practical sense. However, the predicted probability of being a local student for those attending an ancient HEI rises again to an unexpectedly high predicted probability compared to the other higher reputable and prestigious HEIs. This may be a result of those students being in the position of having an ancient HEI in their domicile LA and as a result they may perceive no benefit of migrating to attend a HEI if they have a highly prestigious and reputable institution in their domicile LA.

Overall, the results from the multinomial analysis provide the same conclusive findings as the previous methods. The overall relationship between ethnicity and social background are very similar irrespective of the methodology and outcome variable used, with those from white and most advantageous social backgrounds most likely to be an internal student migrant, while those from the least advantageous social backgrounds and non-white ethnic groups are much more likely to be local students or commuters/distance learners.

4.7 Chapter Summary

This chapter has provided an in-depth analysis of how student migration choices of people entering into HEIs in the UK were impacted by a student's characteristics, the course they studied and the institution attended. This was conducted by analysing a detailed HESA dataset of population data that had not been previously analysed for this purpose.

It has been recognised that the migration choices of people entering into HE is of great policy interest to HEIs as well as government and non-government organisations. This is a result of the impact students have on the locations that they reside as well as other factors such as equality in access to higher education, widening participation, increases in tuition fees and the changing patterns in student migration illustrated in the previous chapter.

The migration choices of individual students have been shown to vary as a result of many over-arching interlinked contributing factors as shown in the conceptual framework earlier in the thesis (Figure 2-2). However, this analysis presented in this chapter aimed to investigate if any general themes or trends were apparent in the data and if any patterns in the migration outcome experienced by a student were impacted by their social background, ethnic group or gender.

The preliminary analysis on the explanatory variables found evidence that supported the view that these explanatory variables did explain some of the differences in the migration outcomes of students. Further analysis was then conducted to answer the aforementioned research questions. The three outcome variables used in the analysis were quite different in their format and as a result they required different methods to model the outcomes against the different explanatory variables simultaneously. However, the findings from this chapter indicate that despite the complexities and different techniques available to measure and quantify student migration, the three outcome variables used in this analysis illustrate very similar results.

The main findings indicate that ethnicity, social background and gender all have a significant impact on the student migration experience in order to attend a HEI in the UK. The most concurrent finding across the three techniques was the group most likely to migrate, travel the furthest distances and be internal student migrant were students from the White ethnic group, most advantageous social background and were male. In contrast, the group of students least likely to migrate, travel the shortest distances and be local students were from the Asian ethnic group, least advantageous social background and were female.

The results produced within this chapter are very similar to the main findings reported by Gibbons and Vignoles (2012) who found that some ethnic groups – especially Bangladeshi and Pakistani women – appear to be considerably more sensitive than others to distance regarding student migration. The results presented in this chapter found that Asian women were predicted to travel the shortest distance to attend a HEI of all ethnic and gender combinations. This shows that the results from the two studies portray a very similar picture despite the differing sources, methods and time periods. Furthermore, Gibbons and Vignoles (2012) also found that students from lower socio-economic backgrounds also differed with regards to their sensitivity to student migration distance, with the sensitivity increasing as income and occupational status decreases. Again the results presented within this chapter regarding the students social background showed the same relationship between predicted total distance and social background with the least advantageous background groups predicted to travel the shortest distances.

The results presented throughout this chapter have analysed the trends between the available variables within the dataset. However, the underlying factors that influence the student migration decision process are plentiful. In Chapter 2, the many interlinking factors that influence a student migration decision were discussed in detail (Figure 2-2). It must be noted here that one of the limitations of this study is that it was not possible to quantify or to take into account in this analysis several of the factors, as identified in Chapter 2, that influence the student migration decision and as a result it is likely that the results will be impacted by unobservable variable bias.

One of the key influencing variables that was not taken into account in this analysis was the impact of a student's achievement level prior to HE on the student migration outcome. This was not included as the variable was not available in the dataset. The work by Gibbons and Vignoles (2012, p.109) found that the variations in distances travelled to HEIs was not caused by the 'heterogeneity in prior achievement across groups' but it should be noted that 'students with lower achievement scores are less likely to travel far' to attend a HEI. Student achievement is therefore directly linked to the HE admissions process and will often influence the student in their choice of course and HEI. There have been previous studies that have linked the achievement level of students to their socio-economic status, ethnicity and the level of schooling at earlier stages of the education system. The findings of this analysis concluded that the student migration outcome was influenced by the student's ethnicity, social background and gender. However, due to the inability of this study to disseminate the results by factors not included in the models, these findings might be a by-product of other influencing factors that cannot be identified within this study, such as student achievement and levels of deprivation.

The research conducted in this thesis could be extended further by analysing the patterns and differences between the migration and commuting distance variables independently. It has been discussed in previous chapters that there were clear geographical differences between those students that migrate compared to those that commute. As a result, further and more detailed analysis that investigates these differences in distance migrated and distance commuted by the student characteristics would also provide interesting and policy relevant research findings.

Further extensions of this work could include changing the focus of the outcome variables, for example, within the dataset it is possible to see how social background, ethnicity and gender impact on the institution attended or course studies instead of the focus here on the migration transition experienced. Further extensions could also include obtaining linked data that allow for the same analysis to be conducted but with the addition of controlling for a student's prior attainment level. This work could also be extended by conducting a qualitative study to find more in-depth reasoning behind the observed differences between the sub-groups in the student population as the current quantitative study can only illustrate that such differences exist but do not provide any indication as to why these differentials were so apparent.

The findings from this chapter indicate that despite the complexities and different techniques available to measure and quantify student migration, the outcome variables used in this analysis all illustrate the same substantive findings. All techniques undertaken here have suggested there to be substantial differences between the ethnic and social background groups as well as significant gender differences in the patterns of migration into HE in the UK. The use of three different techniques and the cofounding results provide statistical evidence that support these findings of ethic, social background and gender differences in the migration decision process of students and that access to HE is still not equal across the social spectrum in the United Kingdom. These finding should be of great concern to policy makers, government officials and anyone involved in the running and management of the UK HE sector.

5. The value of gaining a higher educational degree in the UK: Does migration matter?

5.1 Introduction

The aim of this chapter is to examine the economic value of migrating for higher educational purposes in the UK. The decision of where to study is a choice, and the choice to migrate to attend a HEI is associated with added economic costs in comparison to not migrating and studying at a local HEI.

The aim is to evaluate the future economic value of this choice by analysing differences between those that migrated and those that did not, on their labour market outcomes six months after graduating. By analysing graduates' employment status and first salary, the economic value of the choice between staying at home when entering HE or migrating away and studying at a HEI further away can be estimated. With the increasing monetary cost of HE in the UK and the increasing costs associated with any migration, the key policy relevance to this study is whether there is any future economic benefit when entering the labour force for those that migrated to attend a HEI in comparison to those that did not. Therefore, informing the decision process of future students on whether it is beneficial to migrate or study locally.

In previous chapters it has been shown that the student migration decision is influenced by many factors. In the previous literature, it was argued that an individual's experience of HE, in regards to their participation, institution attended, course studied and migration process experienced, has been found to be associated with variety of factors. Some of these factors occur well before entering HE but some are still visible throughout the higher educational sector. The migration outcome experienced by an individual differs as a result of their domicile location, with a clear difference in migration patterns of those originating from the North compared to the South of the UK as shown in Chapter 3. While it was shown in Chapter 4, that the type of student migration experienced and the distance travelled by students varied greatly as a result of their social-economic background, ethnicity and gender.

A substantial amount of previous research has been conducted across a variety of developed countries that has assessed the value of obtaining a HE degree.

This research has become increasingly important in recent years in the UK, as the monetary cost of obtaining a degree has been increasing for a substantial period of time. However, there has been no previous research that has estimated the impact of migration to obtain a HE degree on the future economic outcomes of the students. Therefore, no previous studies have been able to estimate the economic value of migrating in order to attend a HEI.

A large number of policies have been designed and implemented to encourage equal access to HE. In previous chapters, it has been shown that the student migration experience differs between individuals according to their socio-demographic characteristics and spatially across the UK. It is therefore of great policy interest to see if there are any differences in future labour market outcomes of graduates as a result of these previously mentioned differences in the migration outcome when entering HE in the UK. This could enable the evaluation of whether differences in the student migration decision result in visible disparity between socio-demographic groups later in the life course or if, at this end stage of the education system, the differences have been eradicated. The findings will also aim to give a definitive evaluation of whether or not it is economically beneficial after graduating for a student to migrate to attend a HEI or remain at home and study locally.

This chapter, therefore, attempts to evaluate whether observed differences in student migration patterns, as a result of the multiple factors mentioned in previous chapters, impact on individuals after graduating. This is conducted by applying a variety of statistical techniques to a combination of two HESA datasets that have not previously been analysed in this fashion and by answering the following main research questions; ‘How does student migration into higher education impact on the future employment status after graduation?’ and ‘How does student migration into higher education impact on the first wage achieved after graduation?’

The remainder of the chapter takes the following structure. Firstly, an overview of previous studies investigating the value of gaining a higher educational degree in the UK is provided. This is followed by introducing the dataset and the methodologies that are used throughout this chapter. Finally, the results of the statistical analysis are presented and concluding comments, limitations and recommendations are summarised.

5.2 What do we know about the benefits of migrating in order to obtain a higher educational degree?

The decision to migrate is rarely a simple individual action in which a person decides to relocate, but one that takes into account multiple interrelated factors. As illustrated in Chapter 2, the process of student migration is no exception and it has been shown that the decision to migrate in order to attend a HEI is a complex decision process and is impacted by many factors as depicted in Figure 2-2.

One area that is identified as being an important factor in influencing the student migration decision is the financial cost that would be burdened on the individual and how this would impact people from different socio-economic backgrounds and social class groupings in varying levels. In this chapter the principle aim is to investigate how the migration decision experienced into HE impacts on an individual entering the labour market after graduating with regards to employment and earnings. In order to do this it is necessary to have a clear understanding of the theoretical perspectives behind the economics of migration, as well as setting out a substantive review of all the previous research that has evaluated the value of obtaining a HE degree and the student migration decision experienced in order to do so.

The neoclassical theory of migration is widely used in the context of migration research and is taken from the discipline of economics. The neoclassical framework corresponds to the laws of migration set out by Ravenstein (1885) and Lee (1966) which emphasise the importance of economic drivers in migration. The neoclassical framework remains a dominant strand of economics and has played a considerable role in migration studies (Castles and Miller 2009).

At a macro-level, this theory was developed to explain the migration of workers within the structuring context of economic development (Todaro 1969), where migration was caused by geographical differences in the supply of and demand for labour (Massey et al. 1993). It was also argued that further to differences in availability of employment, wage differentials across geographical space would also contribute to migration flows (Arango 2000).

On the micro-level, the theory was adapted to include personal considerations. These included the cost of transport, adapting to a new area, the upheaval of changing social networks, and therefore the micro-level theory is more suited to the concepts being observed in this chapter. According to Borjas (1989), individual migrants evaluate the costs and benefits of migration and make a rational decision about the destination of their migration based on which location will provide the highest benefit, in terms of their lifetime earnings.

These macro and micro-level concepts of the neoclassical migration theory can be easily related to the process of student migration. Student migration will be caused by differentials in the availability of HE spaces in geographical areas. For example, if you live in an area with no HEI your migration is triggered by the availability of HE elsewhere. It can also be indirectly linked to the concept of wage differentials, where student migration can be triggered by the financial benefit through increased wages from studying in HE in general or through the perceived benefit of studying at a certain HEI. However, as illustrated in Figure 2-2, on the micro-level, the decision to migrate will involve many personal considerations which will impact on the decision to migrate or not when entering a HEI.

Therefore, according to basic neoclassical theory of migration, one would assume that people decide to migrate in order to gain a financial benefit. However, can this be said for student migration, are student migration decisions purely financial? Results presented in previous chapters would suggest that a purely economic theory of student migration would not take into account all the factors at play. Gaining a financial benefit from migrating or not migrating in order to attend a HEI will be important but it is likely that it may not be the only or most important factor in the student migration decision. Other factors such as the transition in adulthood and leaving the parental home, the cultural and social capital gain from migrating to a new environment, the desire to attend a certain institution or the desire to move to a certain settlement may also be influencing factors in the students choice to migrate or not (Smith and Sage 2014; Smith and Jons 2015).

The analysis presented within in this chapter focuses solely on the financial value of migrating. Any evaluation of the non-economic, social and cultural values of migrating is beyond the scope of this research and is a possible further extension but would require different data sources and possibly a qualitative element in order to be successful.

Previous literature on migration modelling shows some widely observed overarching findings. It has been found that better educated people tend to be generally more mobile than less educated people (Sjaastad 1962; Schwartz 1976). Furthermore, the amount of previous migration is positively correlated to the amount of subsequent mobility (DaVanzo 1976, 1983). While the likelihood of migration depends on the economic attractiveness of the destination in comparison to the origin (Faggian and McCann 2009). These findings are important to keep in mind when analysing the value of migrating for HE purposes. However, this previous research did not focus on the value of migrating for HE purposes and therefore this has been identified as a clear gap in the literature that this chapter aims to fill.

The pieces of previous research with the most relevance to the work conducted in this chapter were those that identified the role of the individual characteristics of students and the role of particular universities in the migration behaviour of students and graduates. Faggian et al (2006, 2007a) analysed the sequential migration behaviour of students and graduates by studying the migration of individuals at the start and end of their HE studies. All individuals were categorised into one of five categories of sequential migration. The previous studies found that ethnicity, gender, levels of human capital and local economic conditions all impacted on the sequential mobility trends experienced by students and graduates (Faggian et al. 2006, 2007a, b; Faggian and McCann 2008, 2009; Pemberton et al. 2013).

However, there was no analysis that analysed how individuals from the different categories of sequential student mobility differed in regards to their future labour market outcomes. Furthermore, these previous studies did not evaluate the impact or the value of the migration into HE. Therefore, although this work is extremely valuable, the impact of the migration into HE on future labour market outcomes cannot be identified, as it was always analysed alongside the migration decision experienced after graduating. In contrast, the analysis conducted in this chapter does isolate the migration decision into HE and therefore will aim to evaluate the economic impact of this migration.

The value of gaining a higher educational degree in the UK:

It is important to analyse the value of migrating for HE purposes on future earnings and employment status for a variety of reasons. The findings in the previous chapters have shown that the migration choices to HEIs in the UK are not equal between different groups within the population and therefore it is necessary to investigate whether these differences impact on the student's future after graduation. The substantial expansion of HE participation in the UK and in Europe has allowed a substantial proportion of young people to obtain a university degree. As a result, there has become significantly more competition for graduate jobs and being a university graduate no longer provides the perceived guarantee of immediate employment after graduation. The large expansion in the number of people obtaining a degree has also coincided with a period of transition in the UK in regards to the financial structuring of the HE system. The has resulted in a marked increase in the interest and the amount of research conducted to evaluate the perceived benefit to an individual of gaining a higher educational degree for future employment prospects and future career earnings (Brynin 2012; Walker and Zhu 2013).

As discussed in Chapter 2, there have been significant changes in tuition fee policy in recent years in the UK. Before 1998, the cost of a university degree was entirely supported by the government. Since then, students have been asked to pay part of the cost of HE with the introduction and later increasing of tuition fees (Sa 2014). The policy of seeking to expand the HE sector in the UK was primarily driven by the desire to increase the skilled workforce of the population. Although it was envisaged that young people in the UK were set to benefit from this expansion, they can also be seen to be exposed to greater risk.

The expansion in the number of people gaining degrees has blurred the boundary between graduate and non-graduate work and thus altered the risk environment associated with the benefits of HE. Knowing what is a graduate job is surely an important factor in the decision process of whether to participate in HE or not. It is unlikely that many young people calculate the economic value of education relative to an expected career. They are likely instead to have a notion of a 'good' job, which would partially be based on some idea of expected pay (Brynin 2012).

As the costs of HE are becoming increasingly shouldered by the students themselves, there is more risk of financial loss involved with undertaking HE. As a result, a large amount of research has been conducted to help inform people wanting to consider the positives and negatives of undertaking a higher educational degree.

Many studies have found that the economic returns associated with higher educational qualifications are substantial; however, the actual level of this impact differs between the studies conducted. Universities UK (2007) found that an individual with an undergraduate qualification would earn an average of £160,000 more in the labour market over their lifetime compared to an individual with A-Levels. Universities UK (2007) also stated that there was significant variation in the amount of lifetime benefit as a result of the degree subject, qualification type and the age of attaining the qualification. The results also indicated that men from lower socio-economic groupings and families with relatively lower family income do particularly well from attaining HE qualifications. In contrast, women do relatively well irrespective of their family background or circumstances. In a more recent study, Walker and Zhu (2013) estimate substantial effects of gaining a higher educational degree on the net value of lifecycle incomes. The likely impact of having a degree relative to not having a degree on lifecycle earnings was found to be 28% for men (approx. £168k) and 53% (approx. £252k) for women.

These previous studies have all looked at the impact of gaining a degree on future life-time earnings as well as the probability of being employed. All studies found differences as a result of the course studies and the quality of the institution attended. In previous chapters, it has been discussed that the migration experience into HE has a direct link with the institution attended and course studied and it was also found that there were significant differences in those migration experiences as a result of ethnicity, socio-economical background and gender. However, despite these links between the future life prospects after graduating and the institution attended and the corresponding links with the migration decision process, there is an absence of previous research that has directly compared the differences in the migration decision process to attend a HEI and the individuals employability and salary after graduation.

The question that arises is whether we observe these differences in the benefit of gaining a higher educational degree as a direct result of the migration transition experienced or any other observable characteristics of the individual. Unlike the previous research, the analysis presented in this chapter aims to fill the gap in the literature by attempting to isolating the impact of the migration transition into HE on future labour market outcomes, rather than analysing the sequential mobility patterns and the overall impact of gaining a HEI degree. The analysis presented in this chapter uses a unique combination of two HESA datasets, which up to now, have not previously been analysed in this fashion and therefore this analysis provides a clear and policy relevant piece of new research. The following section of this chapter introduces these datasets in more detail, with the methodology and results sections to follow.

5.3 Data

This chapter uses two linked datasets provided by the Higher Education Statistics Agency (HESA); the HESA Student Record (Higher Education Statistics Agency 2014b), as used in previous chapters of the thesis (for a description see Section 3.2 and 4.3), and the HESA Destinations of Leavers from Higher Education (DLHE) survey (Higher Education Statistics Agency 2014a).

The DLHE data can be linked to the HESA Student Record Data to provide information on the leavers' student experience and personal characteristics. As mentioned previously, the HESA Student Record contains population data on every student registered at a UK HEI. In contrast to this, the DLHE is a survey of graduates and therefore those individuals that responded to the DLHE survey were then retrospectively linked back to their Student Record Data observations. The student record variables used in this chapter are identical to those used in Chapter 4 and therefore will not be explained again here. The remainder of this section will review the coverage and response rate of the DLHE survey, while introducing the new variables that will be used in the remainder of the chapter. This section will also analyse how representative the DLHE sample is in comparison to the total student population by comparing the sample population characteristics with those in the Student Record Data.

5.3.1 DLHE Coverage and Response

The DLHE provides first phase information about patterns of employment and further study six months after the completion of a higher educational degree. Therefore, the DLHE target population contains all students reported to HESA as obtaining relevant higher education qualifications and whose study was full-time or part-time.

In 2011/12, 411,005 UK and other EU domiciled students provided information about their destinations from a possible 567,390, while a further 28,920 explicitly declined to give information, giving an overall response rate for UK and EU domiciled qualifiers of 77.4% (Higher Education Statistics Agency 2013). As was the scenario in Chapter 4, the analysis in this chapter only analysed UK-domiciled students. 385,640 UK domiciled leavers provided their information to the DLHE in 2011/12, a response rate of 77.0% (Higher Education Statistics Agency 2013), and these respondents are those that are analysed throughout the remainder of this chapter.

5.3.2 Representativeness of DLHE compared to Student Record Data

It is necessary to investigate how representative the 385,640 UK domiciled survey respondents for the 2011/12 DLHE were compared to all UK domiciled students that graduated in the same time period. The response rate to the survey is good but in order to make accurate conclusions it is necessary to have a clear understanding of the potential non-response bias in the survey. Potential bias from non-response represents a major threat to the validity of findings from survey based analysis (Micklewright et al. 2011).

To investigate the general representativeness of the DLHE compared to the total student population it is necessary to investigate the Unit Non-response. However, this is not a simple task, due to the way the data is merged between the two datasets. The population data tell us the true real composition of the student population by the different the characteristic variables. From the DLHE, the compositions of those students that responded can easily be identified. The problem however arises when trying to identify the student characteristics of the non-responders.

From the linked data it is simple to identify the students that responded to the DLHE and those that did not and because of the linked nature of the dataset information on the non-responders characteristics can be drawn. However, those that did not respond to the DLHE in the student population data include students that were graduating and therefore were actual non-responders to the survey, but also include all other students that were not finishing and were therefore not targeted by the survey. There is no way to accurately identify individuals in the linked dataset that were actual non-responders or were just continuing students. As a result of this a proxy for non-responders was necessary to enable some form of rough analysis of the non-response, although this is noted is being a far from perfect solution.

The proxy used to analyse the non-response was to limit the analysis of responders and non-responders to students in the dataset that were third year undergraduate students. The majority of undergraduate degrees in the UK last three years and therefore this was the closest way to examine those that were expected to be finishing and a target of the DLHE. However, of the third year undergraduate students in the student record, DLHE responders only represent 44.7%. Comparing this value to the known 77.0% response rate, it is clear that the non-response value used below contains a large amount of students that were not finishing and weren't actual non-responders.

The ethnic composition of the population data compared to the DLHE is shown in Table 5-1. It can be seen that the differences between the student population data and those who respond to the DLHE are quite small. The DLHE has around a 3% high representation of White students compared to the total student's population and an under-representation of the Black Ethnic Group. However, the difference between the DLHE responders and the proxy for non-responders for ethnicity is much larger and there is a very strong statistical difference[5] between the observed and non-observed values. White students were significantly over-represented in the DLHE compared their non-white peers.

[5] Z-Score calculated by using the pooled sample proportion in the denominator obtained by combining the two samples given by: $p = (n_1p_1 + n_2p_2)/(n_1 + n_2)$

Similar to ethnicity, there were small observed differences between the population values and the DLHE responders with regards to the students Background as shown in Table 5-2. All background groups were over represented compared to the population values because of the significantly smaller number of DLHE responders in the unknown social background category. However, when directly comparing the responders and non-responders statistically significant differences appear. The most advantaged and advantaged groups were slightly over-represented in the DLHE, while the less and least advantaged groups were slightly under-represented. However, the difference for the less advantaged group was not statistically significant.

The differences by gender are shown in Table 5-3, again the differences were statistically significant with females being very slightly over-represented in the DLHE.

In conclusion there appears to be significant unit non-response bias within the DLHE survey. When simply comparing the DLHE with the total student population only very small differences are apparent. However, when these differences are statistically tested against the non-responders the differences are more visible.

This could be a result of selection bias from the universities where they are not encouraging students that may not have achieved what they desired from their time at the HEI. A significant amount of the advertisement that encourages leavers to fill in the DLHE are targeted at surveying those who are high achievers. This may be influencing what is being observed here with regards to the unit non-response by ethnicity, social background and gender, as those groups within these variables that have been found to underachieve and are less likely to migrate were under represented in the DLHE. However, as mentioned previously, this non-response bias wasn't able to be identified with any certainty and was measured using a proxy for non-responders and therefore this should be considered when evaluating the representativeness of the DLHE survey. It must be remembered that the official reported response rate was very high and these levels of non-response bias may be a production of not being able to identify specifically who was a non-responder.

Table 5-1: Analysis of the Representativeness of the DLHE compared to the true student population as measured from Student Record Data – Ethnicity Variable

Ethnicity	Total Student Population (%)	DLHE - Responders (%)	DLHE – Non-Responders (%)	Z-Score (diff between responders and non-responders	P-Value
White	79.1	82.2	77.5	35.35	0.000
Black	6.0	4.5	6.9	-31.02	0.000
Asian	9.0	8.5	9.8	-12.73	0.000
Other	4.0	3.7	4.3	-8.67	0.000
Unknown	1.9	1.1	1.6	-12.43	0.000
N	1,797,492	200,538	162,127	-	-

Source: Higher Education Statistics Agency (2014a, b)

Table 5-2: Analysis of the Representativeness of the DLHE compared to the true student population as measured from Student Record Data – Student Background Variable

Student Background	Total Student Population (%)	DLHE - Responders (%)	DLHE – Non-Responders (%)	Z-Score (diff between responders and non-responders	P-Value
Most Advantaged	17.5	23.66	21.59	14.84	0.000
Advantaged	17.8	23.48	20.51	21.52	0.000
Less Advantaged	22.8	24.11	24.34	-1.61	0.054
Least Advantaged	20.6	20.25	21.30	-7.74	0.000
Unknown	21.3	8.5	12.26	-36.60	0.000
N	1,797,492	200,538	162,127	-	-

Source: Higher Education Statistics Agency (2014a, b)

Table 5-3: Analysis of the Representativeness of the DLHE compared to the true student population as measured from Student Record Data – Gender Variable

Gender	Total Student Population (%)	DLHE - Responders (%)	DLHE – Non-Responders (%)	Z-Score (diff between responders and non-responders	P-Value
Male	42.5	41.03	43.54	-15.21	0.000
Female	57.5	58.97	56.46	15.21	0.000
N	1,797,492	200,538	162,127	-	-

Source: Higher Education Statistics Agency (2014a, b)

5.3.3 DLHE variables

On top of the variables used in previous chapters that were derived from the Student Record Data, the analysis in this chapter also uses four labour market indicator variables that have been derived from the DLHE questionnaire.

Economic Activity

In the DLHE survey respondents are able to report what they are doing six months after graduating in relation to both employment and study. They are able to report up to eight individual activities, of which one must be indicated to be the 'most important'. The responses to this question are then used by HESA to derive a category for publication that reflects the range of activities undertaken. The eight derived categories are as follows (Higher Education Statistics Agency 2012a):

- *Full-time work* includes those who indicated their most important activity was working full-time, and who's other activity did not include either full-time or part-time further study, training or research, and those who were due to start a job in the next month.
- *Part-time work* includes those who indicated their most important activity was working part-time, and whose other activities did not include either full-time or part-time further study, training or research. It also includes those where the most important activity was due to start a job in the next month and other activities included working part-time but not working full-time, or engaged in full-time further study, training or research.
- *Primarily in work and also studying* includes those who indicated their most important activity was working full-time or part-time, and whose other activities included full-time or part-time study, training or research.
- *Primarily studying and also in work* includes those who indicated their most important activity was full-time or part-time study, training or research, and whose other activities included working full-time or part-time.
- *Full-time study* includes those who indicated their most important activity was full-time further study, training or research, and whose other activities did not include working full-time or part-time. It also includes those where the most important activity was due to start a job

in the next month, and an additional activity included full-time further study, training or research, provided that working full-time was not also reported as an activity.

- *Part-time study* includes those who indicated their most important activity was part-time further study, training or research, and whose other activities did not include working full-time or part-time.
- *Due to start work* includes those who indicated in their most important activity that they were due to start a job in the next month, but neither working full-time, working part-time, or further study was reported as an activity.
- *Unemployed* includes those who indicated in their most important activity that they were unemployed and looking for work.
- *Other* includes those whose most important activity was either taking time out in order to travel, or something else.

This economic activity variable was recoded into a binary variable from the original eight categories provided by HESA. It was decided to code all those that were unemployed as 1 and those that were employed as 0. Because this chapter is interested in measuring the labour market outcomes of graduates, it was decided that those graduates that were still in further education be removed from this economic activity variable. Those individuals have been removed because they have not yet finished their education and therefore are not taking part in the labour market and are not of interest for this analysis.

After all non-UK domiciled graduates and those still in education were removed from the dataset there was data on the economic activity of 324,711 graduates six months after graduating. Therefore, the item non-response can be said to be 15.8%. Item non-response analysis has been conducted to analyse if there were any differences between those that responded and did not response to the employment question by the main characteristics variables and can be found in Appendix G (Table G-1 to G-3).

As seen in the unit non-response there seemed to be a significant association between non-response to this question and ethnicity, social background and gender. Those from the white ethnicity group were the best responders while those from Asian and other ethnicity groups were those groups with the largest non-response. In contrast to unit non-response, the item non-response for unemployment had the largest non-response from the most advantaged and advantaged groups, while the least advantaged group had the lowest non-response. It is unclear whether these individuals, who responded to the DLHE survey, didn't respond to the question because they were unemployed or if they decided to refuse to answer for another reason.

Salary

This variable describes the annual salary to the nearest thousand pounds before tax. It is collected for all leavers who indicated an activity of either full-time or part-time work, regardless of whether it was classed as their most important activity.

The variable was recorded to the nearest thousand pounds for all values between £6,000 and £90,000, and with two opens ended categorical responses for 'Less than £5,000' and 'Greater than £90,000'. It was decided to remove all respondents from these two open ended categories because; firstly, it is not clear what actual salary value each individual was earning in these categories and therefore it was unclear how to handle these groups in the analysis, secondly these two groups represent the tails within the distribution of the salary variable and thirdly these opened groups only represent less than 3% of those that responded and therefore removing them will not have a profound impact on the overall results.

Of the 385,640 leavers in the 2011/12 DLHE, 176,616 provided information about their salary 6 months after graduation and after removing those recorded in the open ended categories the dataset contained salary information on 171,581 graduates. Those students that were recorded as unemployed however did not have a salary recorded, so to calculate the non-response for salary, one is really interested in how many employed students did not respond to the salary question. Therefore of the 324,711 students that responded to the employment question 283,797 were employed. Therefore, of those that responded and was employed, only 60.5% responded to the question on salary. Again, item non-response analysis has been conducted to analyse if there were any differences between those that responded and did not response to the salary question by the main characteristics variables and can be found in Appendix G (Table G-4 to G-6).

As seen in the unit non-response and the item non-response for unemployment, there seemed to be a significant association between item non-response to this question on first salary and ethnicity, social background and gender. The same trend appeared with ethnicity with whites being over represented at the cost of an under-representation of the Asian group. And similar with the social background group with the more advantaged groups with the higher item non-response. These item non-responders, responded to the employment question and the DLHE in general but did not respond to the question regarding salary. There was no indication as to why they did not respond, however, questions in surveys regarding salary are often less well responded than other questions within a survey.

Standard Industrial Classification (SIC)

The SIC is used to classifying business establishments and other statistical units by the type of economic activity in which they are engaged in. The classification provides a framework for the collection, tabulation, presentation and analysis of data. HESA coincides with the Office for National Statistics SIC classification and is formed of 21 broad sections of industry (Office for National Statistics 2009b).

As this variable is only relevant to those in employment, this variable is only use in the analysis on salary.

Standard Occupational Classification (SOC)

The Standard Occupational Classification (SOC) is a common classification of occupational information for the United Kingdom. Within the context of the classification jobs are classified in terms of their skill level and skill content. It is used for career information to labour market entrants, job matching by employment agencies and the development of government labour market policies. HESA provide SOC information that follows the structure set out by the Office for National Statistics (Office for National Statistics 2010a) and consists of 9 major groups.

As was the case with the SIC variable, SOC is only relevant to those in employment therefore this variable is only use in the analysis on salary.

Location of Employment

This describes the location of the HE leaver's place of work. Data is supplied to HESA in the form of postcodes (for employment in the UK, Guernsey, Jersey and the Isle of Man) or country codes. Postcodes are mapped to counties, unitary authorities, Government Office Regions and UK countries using the Office for National Statistics Postcode Directory (ONSPD). Countries are mapped to geographical regions, informed by the National Statistics Country Classification 2006 grouping of countries (Higher Education Statistics Agency 2012a).

In the analysis in this chapter Government Office Regions (GORs) were used to indicate where in the UK the student was employed (http://www.ons.gov.uk/ons/guide-method/geography/beginner-s-guide/administrative/england/government-office-regions/index.html). Again, as place of employment is only relevant to those in employment this variable is only used in the analysis of salary.

5.4 Methodology

The aim of this chapter is to identify and measure what factors impact on graduates labour market outcomes six months after graduation. However, a particular interest is placed on measuring the impact of migration into HE. In previous sections, it has been highlighted that identifying the economic impact of student migration on the individual's labour market outcomes six months after graduating is of great policy interest. With the increasing costs of HE and living costs in general, there is a growing interest to knowing if migrating to attend a HEI, which in itself endures further expense to the individual, has any economic benefit in the future. In order to identify the value of migrating, the causal impact of migrating on the future labour market outcomes needs to be estimated.

Migration is measured using a binary indicator, as used for the first outcome variable in Chapter 4, where a migrant is record as 1 and a non-migrant as 0. A migration is recorded if a student relocates over a LA boundary in order to attend a HEI. For the purpose of the analysis in the current chapter, a migration is interpreted as a 'treatment' on the individual. However, it is important to stress that migration is an individual's choice and is not a forced upon or compulsory treatment.

The labour market outcomes of each individual student are measured using two variables from the DLHE as explained in Section 5.3. First, a binary variable adapted from the economic activity variable of unemployment for all graduates and second, a continuous variable of salary for all those graduates that were in employment.

The remainder of this section first introduces the issues one may encounter when estimating the causal counterfactual effect by providing a general overview of the evaluation problem and policy evaluation methods. Thereafter, two methodologies for estimating the causal effect are introduced, evaluated and critiqued; regression and propensity score analysis (PSM).

5.4.1 Overview of Policy Evaluation Methods

The problem of measuring the effect of migration on future labour market outcomes falls neatly into the evaluation problem literature, where the aim is to quantify the causal impact of a treatment on an outcome of interest (Heckman et al. 1999; Blundell et al. 2005). As previously mentioned, the treatment under investigation here is the decision of whether or not the student migrates in order to attend a HEI.

To describe the problem in this analysis student migration needs to be considered as a binary variable, $D_i = \{0, 1\}$ and the labour market outcome variable (unemployment or salary) as denoted by Y_i. The analysis aims to identify the difference between Y_{1i} and Y_{0i}, which can be said to be the causal effect of migrating to attend a HEI for individual i on the labour market outcome (Angrist and Pischke 2009). However, to get the true causal impact of migration, unobserved counterfactual information needs to be estimated. This is shown in the model of potential outcomes which is also known as the Neyman-Fisher-Cox-Roy-Quandt-Rubin model (Heckman et al. 1999; Sianesi 2012). The potential outcomes model shows that the observed outcome Y_i can be written as:

$$Y_i = \begin{cases} Y_{1i} & if\ D_i = 1 \\ Y_{0i} & if\ D_i = 0 \end{cases}$$

$$= Y_{0i} + (Y_{1i} - Y_{0i})D_i$$

$$\equiv Y_{0i} + \beta_i D_i$$

$$Treatment\ Effect\ \beta_i \equiv Y_{1i} - Y_{0i}$$

(5.1)

This notation is useful because $Y_{1i} - Y_{0i}$ is the causal effect of migration for an individual. However, it is only possible to observe $Y_{1i}\ or\ \ Y_{0i}$ for any one unit and not possible to observe both, as is required. This is because one individual either migrates or does not. It is, therefore, impossible to measure the impact of both treatments for one individual as one individual is only ever exposed to one treatment at one time point. 'Counterfactual' relates to something that has not happened and therefore in measuring the true causal effect the counterfactual measurement of a treatment that has not occurred on a given outcome variable is required.

In summary, it is impossible to observe the counterfactual due to the 'Fundamental Problem of Causal Inference' which states that it is impossible to observe the outcomes of the same unit in both treatment conditions at the same time (Holland 1986).

In order to learn about the causal effect of migration one needs to compare the labour market outcomes of those who were and were not migrants. The average labour market outcome conditional on migrant status is formally linked to the average causal effect by the following equations (Angrist and Pischke 2009: 19):

$$E[Y_i|D_i = 1] - E[Y_i|D_i = 0]$$
$$= \underbrace{E[Y_{1i}|D_i = 1] - E[Y_{0i}|D_i = 1]}_{\text{Average treatment on the treated}} + \underbrace{E[Y_{0i}|D_i = 1] - E[Y_{0i}|D_i = 0]}_{\text{Selection Bias}}$$

(5.2)

While the term,

$$ATT = E[Y_{1i}|D_i = 1] - E[Y_{0i}|D_i = 1] = E[Y_{1i} - Y_{0i}|D_i = 1]$$

(5.3)

is the Average Treatment Effect of the Treated (ATT). This term captures the averages between the labour market outcomes of the migrants, $E[Y_{1i}|D_i = 1]$, and what would have happened to them if they had not migrated, $E[Y_{0i}|D_i = 0]$. There is, however, an additional term in Equation 5.2 known as the selection bias. The selection bias is the difference in the average Y_{0i} between those who were and were not migrants (Angrist and Pischke 2009). As seen in previous chapters there are many interlinking factors that impact on whether a student was a migrant or not, such as ethnicity, social background and gender. Therefore it is highly likely that the selection bias will be non-zero and as a result this needs to be accounted and controlled for in the methodology used.

In general, the value that a researcher is most interested in measuring is an estimate of the mean impact of a treatment for the whole population. This is obtained by averaging the impact of the treatment across all the individuals in the population. This is known as the Average Treatment Effect or ATE (Sianesi 2012):

$$ATE = E(Y_{1i} - Y_{0i}) \equiv E(\beta_i) \tag{5.4}$$

The ATE is useful to evaluate what is expected on the outcome if individuals in the population were randomly assigned to treatment, as random assignment of treatment D_i eliminates selection bias (Angrist and Pischke 2009). However, as this study is an observational study and has not been conducted using random allocation to the migrant and non-migrant groups, it can be said that selection bias will be present in the analysis and depending on the method used the ATE estimate is therefore likely to contain bias.

The goal in most empirical economic research is to overcome this selection bias and be able to make concluding remarks on the causal effect of the treatment. As a result, methodologies have been created to estimate the causal effect by estimating the missing data that are not observed and ensuring selection bias is eradicated. These methods are known as policy evaluation techniques. The different methods available apply differing assumptions to calculate the counterfactual and they also differ in the way in which the causal effect is estimated and presented in terms of either ATE or ATT.

All policy evaluation methods rely on the Stable Unit Treatment Value Assumption (SUTVA) being met. SUTVA is the assumption that the model's representation of outcomes is adequate, that is that the observed outcome for an individual exposed to treatment depends only on the individual and not on what treatments other individuals receive, nor on the mechanism assigning treatment to individuals and that whether the individual participates only depends on the individual (Blundell et al. 2001; Heckman 2005). In the analysis, it must be assumed that the decision of the individual student to migrate or not was their own and was not dependent on the decision of other individuals in the dataset. Because the dataset consist of population data, the vast majority of individuals in the dataset will have no association at all with other individuals in the population. Those individuals that may be affected by other individuals in the dataset in their migration decision would only make up a very small proportion of the dataset and due to the way the data is collected there is no way of testing if this assumption is met.

Another assumption that is relevant to all policy evaluation methods that only account for observable variables within the method is the conditional independence assumption (CIA). The CIA states that, ensuring a given a set of observable covariates, X, are not affected by the treatment, the potential outcomes of Y are independent of the treatment assignment D (Khandker et al. 2009: 55):

$$(Y_{0i}, Y_{1i}) \perp D_i | X_i \tag{5.5}$$

Equation (5.5) states that the potential outcome is independent of the treatment status, given X. Therefore after controlling for X the treatment assignment is 'as good as random'.

This assumption is also called unconfoundness (Rosenbaum and Rubin 1983), and it implies that the decision to migrate is based entirely on observed characteristics. CIA is a strong assumption and is not a directly testable criterion. If unobservable characteristics determine the treatment allocation (decision to migrate), CIA will be violated (Khandker et al. 2009). The CIA is crucial for correctly identifying the impact of the feature under examination, since it ensures that, although selection biased treated and untreated groups differ, these differences may be accounted for in order to reduce the selection bias. This therefore allows the untreated units to be used to construct the counterfactual for the treatment group (Heinrich et al. 2010).

As previously mentioned, the HESA datasets used in this analysis contain a rich selection of covariates that include information on the student's socio-economic and demographic background. As a result, it is hoped that all relevant differences between graduates have been captured within the variables covered in the dataset. However, in Chapter 2, variables were identified as being important factors in the student migration process but are not observed in the dataset. The variables include a student's ability, educational attainment and motivation, and these variables cannot be directly measured using the datasets acquired. It was discussed previously that an individual's ability, attainment and motivation are important factors that can impact on the student migration process and their labour market chances. Therefore, they are important factors that need to be considered when evaluating the causal impact of student migration. As a result, the type of HEI attended is used as a proxy in an attempt to eradicate unobservable variable bias on the models. Type of HEI attended can be used to proxy for the unobservable variables due to the varying levels of admission tariffs required to enter institutions of differing types and the varying levels of motivation required to attend differing institutions (as discussed in Section 2.1.4). This approach was decided to be the best option available to proxy and control for an individual's attainment, ability and motivation level in the analysis and thus limited the impact of the CIA being violated and limiting the impact of unobservable variable bias on the results and conclusions. It is understood that this is a limitation of the study and that HEI category is an important variable in its own right. It must be noted that the inclusion of HEI category is not designed to control for institution attended but primarily as a proxy for attainment, ability and motivation.

To identify causal effects the CIA/unconfoundness alone is not enough. Another assumption that needs to be met is the common support or overlap condition (Khandker et al. 2009: 56):

$$0 < P(D_i = 1|X_i) < 1 \qquad (5.6)$$

The common support assumption implies that for each value of X, there is a positive probability of being both a migrant and a non-migrant. Therefore, the probability of receiving treatment for each possible value of the vector X is strictly within the unit interval: as is the probability of not receiving treatment.

This assumption of common support ensures that there is sufficient overlap in the characteristics of treated and untreated units to be able to make causal statements about the impact of the treatment on the outcome (Baum 2013).

When the assumptions of CIA and common support are satisfied, the treatment assignment is said to be strongly ignorable (Rosenbaum and Rubin 1983), and causal inferences can be made.

After discussing the concept of policy evaluation and the underlying assumptions the focus now shifts onto the specific methods that will be used in this analysis. Two methods are used for estimating the counterfactual effect of the observed factors for both economic activity and the salary outcome variables: regression analysis and propensity score matching (PSM). Further details of how these methods have been specified and there underlying assumptions are provided in the remainder of this sub-section.

5.4.2 Regression Analysis

The first methodology to be used when analysing both labour market outcome variables is regression analysis. Due to the differing format of the two outcome variables two different forms of regression analysis will be conducted: Linear and Logistic.

Economic Activity

The labour market variable for economic activity is a binary outcome that depicted whether a graduate (who was not enrolled in further study) was unemployed. A value of 1 was recorded for those graduates unemployed and 0 for those employed. For binary outcomes, the most commonly used model is the logistic regression model (Long and Freese 2006; Agresti 2013). This method was used and explained in detail in Chapter 4 (Section 4.5.1), and the methodology undertaken here is identical except the change in outcome variable used.

In order to find the combination of variables that had the best statistical fit to the data and made the most theoretical sense a modelling procedure was conducted and the Scalar Measures of Fit are shown in Table E-1 (Appendix E). The final chosen logistic model was model 5 and contained all explanatory variables available except Standard Occupational Categorisation (SOC), Standard Industrial Categorisation (SIC), Salary and place of employment. An interaction term between gender and social background was also included. SOC, SIC, Salary and place of employment were omitted as these variables were only recorded for those in employment and therefore did not provide any benefit to modelling the impact on employment. It was also decided to not include all the interaction terms (as in Chapter 4) in the final model as when there addition was tested they did not provide enough statistical improvement to the model to make the increased complexity in the interpretation worthwhile. The addition of these additional interaction terms actually made the statistical fit of the model worse hence they were not included in the models for this part of the analysis.

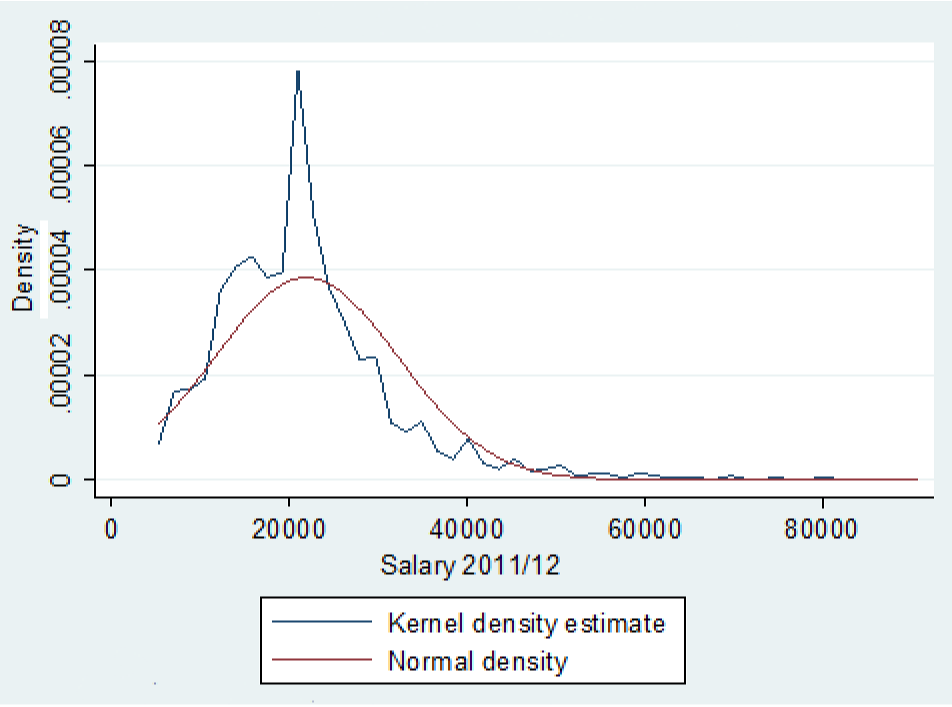
Salary

The labour market outcome of salary is a continuous variable and therefore an ordinary least square (OLS) regression model allows us to study the relationship between the dependent variable (salary) and several independent variables taken from the Student Record Data and the DLHE. For the continuous response variable of salary y_i and the multiple explanatory variables x_p , the OLS linear regression model takes the form as shown in Equation (5.7) (Greene 1993).

$$y_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} + \varepsilon_i \qquad (5.7)$$

There are a set of assumptions that accompany the OLS regression model, these include linearity and normality. However, the outcome variable of salary used in this chapter does not conform to these assumptions. The salary variable is positively skewed as shown in Figure 5-1and therefore violates the standard assumptions of classical OLS regression.

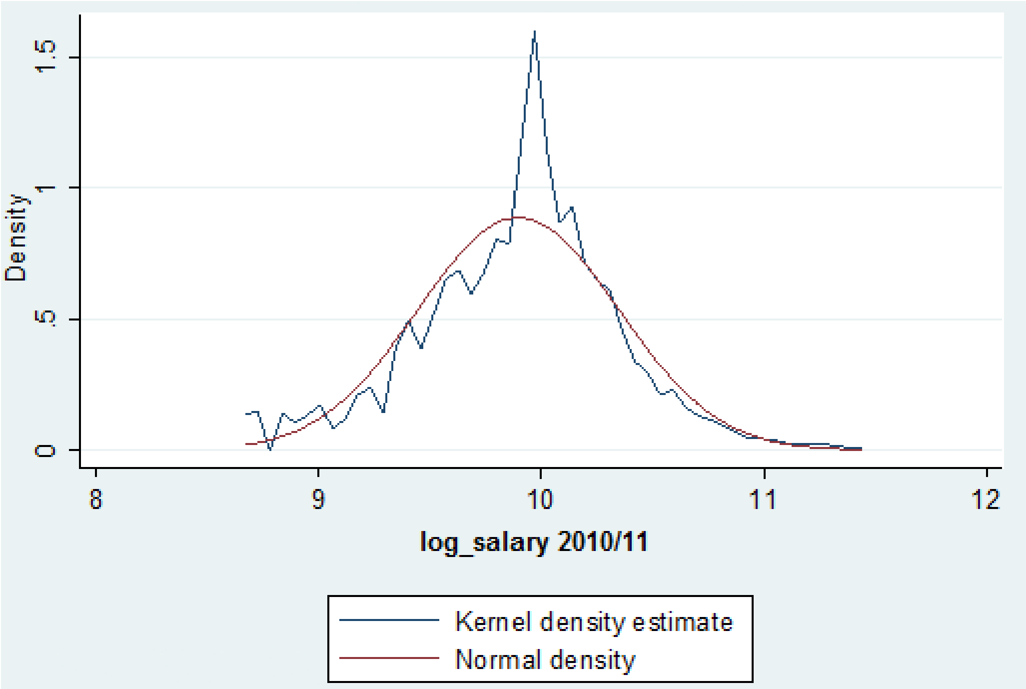
Figure 5-1: Kernel Density Estimate and Normal Density Curve of Salary



Source: Higher Statistics Agency (2014a)

Taking a logarithmic transformation of the highly skewed salary variable is a convenient means of making the outcome variable more approximately normal and therefore not violating the assumptions of the multiple linear regression model. The distribution of the new transformed variable log_salary is shown in Figure 5-2. The distribution of the transformation of the salary variable as shown in Figure 5-2 is still not normally distributed. However the transformation has transformed the variable into a more symmetric shape and is therefore much closer to meeting the normality assumption than it was prior to the log transformation.

Figure 5-2: Kernel Density Estimate and Normal Density Curve of log_salary



Source: Higher Statistics Agency (2014a)

The literal interpretation of the OLS regression model as expressed in Equation 5.7 will still hold when variables have been logarithmically transformed. However, it often makes more sense to interpret the results as percentage changes rather than in log-units and therefore the model specification changes slightly (Benoit 2011).

With the outcome variable of salary being logarithmically transformed the model technically becomes a Log-linear model for a continuous outcome variable and takes the form shown in Equation 5.8 (Benoit 2011).

$$\ln(y_i) = \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} + \varepsilon_i \qquad (5.8)$$

Therefore, to be able to compute the expected value of y_i by a one-unit change in x then the exponent of the coefficient (β) is needed, $e^{\widehat{\beta}}$.

As with the logistic regression model, in order to find the combination of variables that had the best statistical fit to the data and made the most theoretical sense a modelling procedure was conducted and the Scalar Measures of Fit are shown in Table E-2 (Appendix E). The final chosen logistic model was model 8 and contained all explanatory variables available and an interaction term between gender and social background. Again, the statistical gains of including further interaction terms into the model were negative and therefore the modelling process concluded at model 8.

Regression Modelling for Counterfactual Analysis

Regression analysis is a useful tool for the study of causal impacts when the analysis is conducted on data from experiments in which treatment is randomly assigned (Angrist and Pischke 2009). However, when observational data is used and allocation to treatment is not random, the regression modelling technique needs to meet the assumptions of policy evaluation techniques, as well as the underlying assumptions of its own methodology.

Regression modelling, if specified correctly (meets the regression assumptions and CIA) can be an efficient estimator of causal effects. The regression models produce the ATE in their output and therefore conclusions can be made for the whole population, not just those who receive the treatment.

However, regression modelling should not be used to make causal inferences for non-random observational data because the methodology often fails in the assumption of common support. Regression models that are not saturated in X_i may violate common support since covariate cells without both treated and control observations can end up contributing to the estimates by extrapolation (Angrist and Pischke 2009). As a result of not having the common support the regression model is not matching individuals from the treatment and the control for all values of X_i and this will bias the results. The lack of common support results in the inability to make casual inferences as the research is unable to achieve ignorability (Rosenbaum and Rubin 1983).

5.4.3 Propensity Score Matching

As discussed previously, the greatest challenge in evaluating any intervention, programme or treatment is obtaining a credible estimate of the counterfactual: what would have happened to the participants of a treatment/programme if they had not participated (Heinrich et al. 2010). It was shown in the previous sub-section that regression techniques rarely meet all the requirements needed to make a causal inference.

Matching techniques on the other hand have become a popular methodological approach when attempting to estimate a casual treatment effect. It has been widely applied when evaluating labour market policies (Heckman et al. 1997; Dehejia and Wahba 1999), as well as in many other diverse fields of study. It applies for all situations where there is a treatment that can be measured, a group of treated individuals and a group of untreated individuals (Caliendo and Kopeinig 2008).

In this chapter, the aim is to evaluate the economic value of migrating in order to attend a HEI in the UK. Therefore, the treatment under investigation here is migration and the dataset contains students that were 'treated' (migrated in order to attend a HEI) and individuals that were 'not treated' (did not migrate to attend a HEI). This chapter focuses on a specific non-experimental matching methodology known as propensity-score matching (PSM).

PSM uses information on all individual units regardless of if they received the treatment or not. Propensity score matching (PSM) constructs a statistical comparison group that is based on a model of the probability of participating in the treatment, using observed characteristics. Participants are then matched on the basis of this probability, or propensity score, to nonparticipants. By comparing how outcomes differ for the treated units relative to the observationally similar non-treated units, it is possible to estimate the effect of the treatment (Khandker et al. 2009).

One of the critical issues in implementing PSM is defining clearly and being able to justify what similar means with regards to how treated and non-treated units are matched. If the matching process is to successfully mitigate potential bias, this has to be conducted by considering a full range of covariates across which the treatment and comparison units may differ. This concept is pivotal to PSM and refers to the assumptions of the methodology that need to be met in order to make the desired causal inference.

The validity of PSM depends on two conditions that were discussed previously: conditional independence and sizable common support or overlap in propensity scores across the participant and nonparticipant samples (Khandker et al. 2009). If the CIA holds and if there is sizable overlap in $P(X)$ accross participants and nonparticipants then PSM can be used to estimate the ATT. The ATT can be specified as the mean difference in Y over the common support, weighting the comparison units by the propensity score distribution of the participants, which can be specified as follows (Khandker et al. 2009: 56):

$$ATT_{PSM} = E_{P(X)|D=1}\{E[Y_{1i}|D = 1, P(X)] - E[Y_{0i}|D = 0, P(X)]\} \quad (5.9)$$

In recent years, facilitated in part by improvements in computing capacity and associated algorithms, PSM approaches that directly match treated with non-treated who have similar characteristics have replaced regression as one of the preferred methods for estimating the counterfactual effect (Heinrich et al. 2010). Although being very similar to regression in terms of the CIA, PSM goes beyond regression in terms of estimating the casual impact because observations not in common support are excluded from the analysis and therefore PSM matches individuals and compares like with like unlike regression modelling. Another benefit of PSM over regression is that PSM is a non-parametric technique and therefore the parametric assumptions that apply to regression analysis do not apply to PSM. As a result of being non-parametric PSM also avoids potential misspecification of $E(Y_{0i}|X)$ and allows for arbitrary heterogeneity in the causal effects $E(Y_{0i}|X)$ (Grilli and Rampichini 2011).

A limitation of the PSM is that the method only produces the ATT and therefore the focus of the results are restricted to only those who have migrated and we cannot infer to the whole population. The PSM also requires a very high sample size (Sianesi 2012). However, despite these limitations it was deemed appropriate to run the PSM techniques to try and measure a more accurate causal effect of migrating in order to attend a HEI.

It must be noted that when comparing the results of the regression analysis and the PSM it is not expected that the results will be similar. This is because of the different assumptions of the two techniques and that regression produces the ATE and PSM produces the ATT. As a result the methods are looking at different populations and therefore the results are not directly comparable.

Estimating the Propensity Score

To conduct PSM it is necessary to calculate propensity scores of receiving the treatment for those that received the treatment and those that did not. In this study it is therefore necessary to calculate propensity scores that show the probability of migrating given a set of covariates X. This is conducted by running a probit regression conditioning on the same rich set of covariates that have been used previously throughout the thesis.

Unlike the previous methods used in this thesis, the model selection and choice of variables for PSM requires a much more thoughtful and thorough process. This is due to the matching strategy within PSM building upon the CIA assumption discussed previously, which requires that the outcome variable in the matching must be independent of treatment (migration) conditional on the propensity score (Caliendo and Kopeinig 2008).

Implementing matching therefore requires choosing a set of variables X that credibly satisfy the CIA. Heckman et al. (1997) and Dehejia and Wahba (1999) have shown that omitting important variables can seriously increase bias in the results. Only variables that influence the participation (migration) decision and the outcome variable simultaneously are included in the probit model. The better and more informative the data and choice of variables, the easier it is to credibly justify the CIA and the matching procedure.

10 different probit models with the same combination of variables and interactions nested within each model were simulated as was conducted for the modelling procedures in the Chapter 4. From these 10 models, three models were chosen for employment status and four models (with the addition of a model including employment region) are chosen for salary. These different probit models are used to construct the propensity scores of an individual migrating to attend a HEI in the UK. Model 1 includes just basic individual characteristic variables from the student record data which were shown in chapter 4 to significantly impact on the migration choice of the individual. These variables were also seen to be significant and included in the regression models for economic activity and salary used later in this chapter. Model 1 represents the most basic model and therefore the results for this model are likely to have the highest levels of selection bias. As a result further models are needed to test if any variables are missing from model 1 and are therefore causing bias in the results and violating the CIA.

Model 2 includes the addition of the institution category variable. As previously mentioned this variable is used as a proxy for an individual’s attainment, ability and motivation. This is a slight limitation to the study but these unobservable characteristics need to be included in the analysis in some fashion in an attempt to remove omitted variable bias as a result of not observing an individual's ability or attainment. This use of institution category as a proxy is also aimed to ensure that the CIA is met. The inclusion of institution category however was not intended to control for the institution itself as this in itself plays a very important role in the student migration decision process. Despite this, when institution category is included in the model to act as a proxy it does therefore mean that the model does control of HEI attended and this is a limitation. HEI choice plays a key role in the decision to migrate and the outcomes after graduation due to large variety of HEIs within the UK. Controlling for the type of HEI therefore may have a significant impact on the ability to truly measure the impact of the migration decision as the results show the impact of the migration decision conditional on the HEI attended.

Model 3 represents the full model and includes all variables available and interactions between gender, background and ethnicity. While Model 4 which is only conducted for the salary analysis is nested within Model 2 but includes the addition of the place of employment variable. This variable was deemed important to be included as differences in peoples salary may be directly influenced by the location of their job, for example wages in London are always higher than elsewhere in the UK as a result of the London wage weighting[6].

The model outputs for the four probit models are shown in Table 5-4.

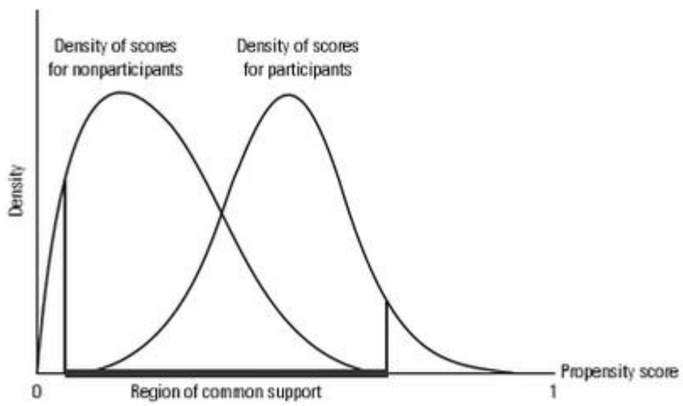
The PSM is implemented using the four different probit models to see how the choice of matching variables impact on the findings. The choice of model impacts on the amount of bias in the results and also impacts on whether the CIA has been met or violated. Also the levels of CS also differ between the models. Those propensity scores where individuals from both treatments are present are within common support and those that do not have individuals from both treatments are outside common support and therefore are omitted from the PSM analysis. The desire is to have all individuals within common

[6] The **London Weighting** is an allowance paid to people who work in London's public sector. Its original purpose was to compensate London workers for the extra costs that they incurred in relation to public sector employees elsewhere in the country.

support as this will result in no data being lost from the analysis as a result of common support not being met.

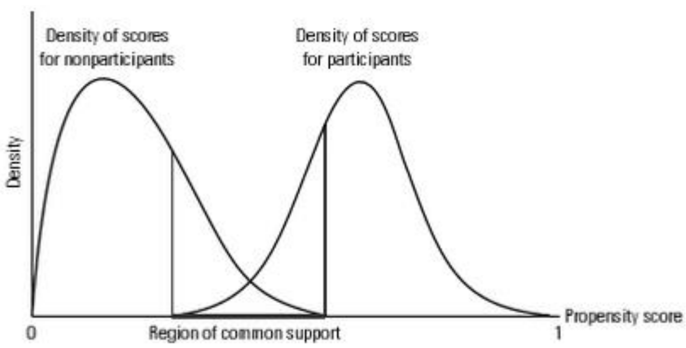
Graphs illustrating the common support assumption can be found in Appendix F. A graph is produced for each probit model and common support is met if there is good overlap between untreated and treated individuals across the propensity scores. Figure 5-3 shows an illustration of good common support where for the majority of individuals in both treatment groups overlap, while Figure 5-4 illustrates an example of weak common support where individuals only overlap in the middle of region of the propensity score and as a result many do not satisfy common support.

Figure 5-3: Example of Good Common Support



Source: Khandker et al. (2009: 57)

Figure 5-4: Example of Weak Common Support



Source: Khandker et al. (2009: 57)

The density of individuals from both the migrant and non-migrant groups to illustrate the common support is shown in Figures F-1 to F-4 in Appendix F. It can be seen that strong common support is met in all four probit models. Model 4 has the strongest common support and smother density lines that the other models. In general the non-smooth and distorted look to the density lines are attributable to the explanatory variables X_i being categorical.

Out of the four probit models, model 4 has the best statistical fit and all variables in the model had a significant association with the propensity to migrate to attend a HEI. Model 3 had a very similar statistical fit to model 2 but several of the interaction variables terms in the model were not significant.

At this point it must be noted that data being analysed herewith in Chapter 5 refer to survey data as opposed to the population data that was analysed in Chapter 4. Therefore, the super-population invoked within Chapter 4 with regards to statistical interpretation of the models statistical measure of fit, is not invoked within this chapter. Therefore, standard statistical and sampling techniques apply within Chapter 5 and the P-Values and measures of fit can be interpreted accordingly (Hartley and Sielken 1975; Agresti 2013).

Table 5-4: Probit Models of being a migrant for Propensity Score Matching Analysis

	Model 1			Model 2			Model 3			Model 4		
VARIABLES	**Coef. (β)**	**Sig**	**SE**	**Coef. (β)**	**Sig**	**SE**	**Coef. (β)**	**Sig**	**SE**	**Coef. (β)**	**Sig**	**SE**
Constant	0.949	0.000	(0.0122)	1.271	0.000	(0.0157)	1.271	0.000	(0.0169)	1.495	0.000	(0.0199)
Ethnicity												
White[a]												
Black	-0.192	0.000	(0.0107)	-0.135	0.000	(0.0108)	-0.257	0.000	(0.0309)	-0.223	0.000	(0.0136)
Asian	-0.410	0.000	(0.00794)	-0.428	0.000	(0.00806)	-0.551	0.000	(0.0244)	-0.452	0.000	(0.0100)
Other	-0.0475	0.000	(0.0116)	-0.0617	0.000	(0.0118)	-0.146	0.000	(0.0315)	-0.113	0.000	(0.0145)
Unknown	0.124	0.000	(0.0168)	0.0783	0.000	(0.0171)	0.0475	0.512	(0.0724)	0.0559	0.000	(0.0200)
Background												
Most Advantaged[a]												
Advantaged	-0.258	0.000	(0.00740)	-0.203	0.000	(0.00755)	-0.208	0.000	(0.0119)	-0.207	0.000	(0.00900)
Less Advantaged	-0.444	0.000	(0.00711)	-0.363	0.000	(0.00726)	-0.351	0.000	(0.0114)	-0.359	0.000	(0.00864)
Least Advantaged	-0.664	0.000	(0.00745)	-0.543	0.000	(0.00762)	-0.513	0.000	(0.0122)	-0.524	0.000	(0.00903)
Unknown	-0.566	0.000	(0.00804)	-0.517	0.000	(0.00818)	-0.532	0.000	(0.0122)	-0.474	0.000	(0.00963)
Gender												
Male[a]												
Female	-0.152	0.000	(0.00454)	-0.135	0.000	(0.00461)	-0.106	0.000	(0.0114)	-0.139	0.000	(0.00541)
Subject												
Medicine[a]												
Science/Engineering	0.0365	0.000	(0.00774)	0.0978	0.000	(0.00789)	0.0960	0.000	(0.00790)	0.0918	0.000	(0.00894)
Agricultural/Veterinary	0.264	0.000	(0.0209)	0.354	0.000	(0.0212)	0.352	0.000	(0.0212)	0.333	0.000	(0.0252)
Social/Human	-0.142	0.000	(0.00764)	-0.0694	0.000	(0.00781)	-0.0686	0.000	(0.00781)	-0.105	0.000	(0.00864)
Business/Law	-0.115	0.000	(0.00830)	-0.0001	0.985	(0.00848)	-0.001	0.898	(0.00849)	-0.001	0.994	(0.00958)
Humanities	0.236	0.000	(0.00782)	0.309	0.000	(0.00801)	0.306	0.000	(0.00802)	0.300	0.000	(0.00919)
Combined	0.454	0.000	(0.0417)	0.454	0.000	(0.0417)	0.455	0.000	(0.0417)	0.531	0.000	(0.0534)
Age												
17 years and under	-1.469	0.000	(0.0923)	-1.314	0.000	(0.0929)	-1.316	0.000	(0.0929)	-1.313	0.000	(0.197)
18-20 years[a]												
21-24 years	-0.135	0.000	(0.00549)	-0.140	0.000	(0.00557)	-0.139	0.000	(0.00558)	-0.160	0.000	(0.00678)
25-29 years	-0.627	0.000	(0.00839)	-0.574	0.000	(0.00851)	-0.573	0.000	(0.00851)	-0.585	0.000	(0.00989)
30 years and over	-1.097	0.000	(0.00768)	-0.996	0.000	(0.00780)	-0.995	0.000	(0.00781)	-0.983	0.000	(0.00914)
Age unknown	-1.550	0.000	(0.287)	-1.384	0.000	(0.289)	-1.371	0.000	(0.289)	-1.425	0.000	(0.298)
Level of Study												
Post-Graduate[a]												
Under-Graduate	-0.126	0.000	(0.00632)	0.0179	0.000	(0.00653)	0.0171	0.009	(0.00653)	0.0716	0.000	(0.00734)
Institution Category												
Ancient[a]												
Red Brick				-0.280	0.000	(0.0111)	-0.279	0.000	(0.0111)	-0.343	0.000	(0.0139)
Plate Glass				-0.496	0.000	(0.0117)	-0.495	0.000	(0.0117)	-0.544	0.000	(0.0145)
New University				-0.814	0.000	(0.0108)	-0.814	0.000	(0.0108)	-0.871	0.000	(0.0135)

	Model 1			Model 2			Model 3			Model 4		
VARIABLES	Coef. (β)	Sig	SE	Coef. (β)	Sig	SE	Coef. (β)	Sig	SE	Coef. (β)	Sig	SE
Recent University				-0.872	0.000	(0.0126)	-0.872	0.000	(0.0127)	-0.965	0.000	(0.0156)
Other				-0.785	0.000	(0.0136)	-0.785	0.000	(0.0136)	-0.876	0.000	(0.0166)
Region of Employment												
North East										-0.477	0.000	(0.0148)
North West										-0.320	0.000	(0.0107)
Yorkshire and Humber										-0.189	0.000	(0.0118)
East Midlands										-0.155	0.000	(0.0128)
West Midlands										-0.277	0.000	(0.0151)
East of England										0.0138	0.219	(0.0112)
London[a]												
South East										-0.002	0.779	(0.00940)
South West										-0.013	0.253	(0.0114)
Northern Ireland										-0.676	0.000	(0.0167)
Scotland										-0.497	0.000	(0.00940)
Wales										-0.345	0.000	(0.0140)
Non-UK										-0.013	0.270	(0.0124)
Interaction Terms												
Ethnicity*Background												
White*Most Advantaged[a]												
Black*Advantaged							0.0575	0.152	(0.0402)			
Black*Less Advantaged							0.0932	0.008	(0.0351)			
Black*Least Advantag.							0.128	0.000	(0.0359)			
Black*Unknown							0.185	0.000	(0.0361)			
Asian*Advantaged							0.0392	0.191	(0.0299)			
Asian*Less Advantaged							0.0904	0.001	(0.0281)			
Asian*Least Advantaged							0.0652	0.017	(0.0274)			
Asian*Unknown							0.311	0.000	(0.0289)			
Other*Advantaged							0.116	0.004	(0.0398)			
Other*Less Advantaged							0.0556	0.129	(0.0366)			
Other*Least Advantag.							0.0955	0.014	(0.0387)			
Other*Unknown							0.143	0.000	(0.0390)			
Unknown*Advantaged							0.00259	0.977	(0.0893)			
Unknown*Less.Adv							-0.0464	0.570	(0.0817)			
Unknown*Least.Adv							-0.0273	0.744	(0.0834)			
Unknown*Unknown							0.0387	0.605	(0.0747)			
Background*Gender												
V. Advantaged*Male[a]												
Advantaged*Female							-0.0073	0.630	(0.0152)			
Less Advant.*Female							-0.0406	0.005	(0.0144)			
Least Advant.*Female							-0.0697	0.000	(0.0151)			
Unknown*Female							-0.0430	0.004	(0.0149)			
Ethnicity*Gender												
White*Male[a]												
Black*Female							0.0339	0.119	(0.0217)			
Asian*Female							0.0273	0.086	(0.0159)			
Other*Female							0.00416	0.861	(0.0238)			
Unknown*Female							0.0641	0.061	(0.0342)			
Pseudo R2	0.1356			0.1621			0.1626			0.1760		
BIC	70770			84602			84542			324906		
Change in BIC	-			13800	(v.v.strong +ive)		-60	(v. strong -ive)		114352	(v.v.strong +ive)	
Observations	377,920			377,920			377,920					

Standard errors in parentheses
Notation "0.000" refers to P-Values smaller the 5 x 10^{-4}
[a] Denotes Reference Category

Source: Higher Education Statistics Agency (2014a, b)

Matching Estimators

The PSM estimator in its general from was shown in Equation 5.9, however, all matching algorithms differ in the way they use the propensity score to match comparison units with treated units. There are many factors to be considered when choosing which matching algorithm to use, these include; matching with or without replacement, the choice of proximity, whether and how to weight cases in the analysis and the number of comparison units matched to each treatment unit. There are many different PSM estimators that can be used in the PSM methodology, each of which requires different choices to be made when they are used. The technical details of each method will not be discussed here in depth, for a discussion see Imbens (2004). In this analysis two matching algorithms are used; Caliper and Kernel matching.

The most straightforward matching method is Nearest Neighbour (NN) matching. NN matching matches an individual from the comparison group with a matching partner in the treatment group that is closest in terms of the propensity score. However NN matching faces the risk of bad matches if the closest neighbour is far away (large difference in the two propensity scores).

The risk of incorrect matching can be avoided by imposing a tolerance level to the maximum propensity score distance between the matched individuals in the treatment and non-treatment groups. This tolerance level is known as a caliper and therefore caliper matching is an extension to NN matching with the addition of the set propensity range (Khandker et al. 2009). Applying caliper matching means that those individuals from the non-migrant group are matched with a partner from the migrant group that lie within the caliper range and is closest in terms of propensity score. One drawback of caliper matching is that it is difficult to know what tolerance level is reasonable and provides the best quality of matching (Smith and Todd 2005). In the analysis conducted in this chapter the caliper was set to 0.01.

Kernel matching was also conducted in this analysis. Kernel matching is a very different technique compared to NN and caliper techniques. Unlike NN and caliper, which do not use all observations from the comparison group to construct the counterfactual, kernel matching is a non-parametric method which uses weighted averages of all individuals in the control group to construct the counterfactual outcomes. One major advantage of kernel matching is the lower variance of the estimates due to the large amount of information used in the matching. One major downside to kernel matching is the large computational requirements and these methods can take a long time to complete. Also some observations used may be poor matches hence the proper imposition of the common-support condition is of major importance for kernel matching (Caliendo and Kopeinig 2008; Heinrich et al. 2010).

Implementation and Bootstrapping

After running the probit models, creating the propensity scores and choosing the matching algorithm to be used, the next stage is to run the matching techniques on the dataset. In this analysis the statistical-analysis software STATA13 was used to implement the PSM using a program developed by Leuven and Sianesi (2003) called PSMATCH2.

It is not possible to interpret the results of the PSM estimation properly without estimating the standard errors, which provide an indication of the importance of the sampling error in the generated estimates. However, testing the significance of treatment effects and computing their standard errors is not straightforward. The problem is that the estimated variance of the treatment effect should also include the variance due to the estimation of the propensity score and the imputation of the common-support and as a result these steps add variation beyond the normal sampling variation. Conventionally, standard errors of PSM estimates are obtained using bootstrap methods (Lechner 2002; Caliendo and Kopeinig 2008; Heinrich et al. 2010) where the estimated standard error is the standard deviation of the estimate across multiple replications. Despite Imbens (2004) stating that there is little evidence to justify bootstrapping, it is widely used in other studies (Black et al. 2004; Sianesi 2004) and therefore is applied here. The standard errors reported in the output of the PSM in Section 5.5 have been estimated using bootstrapping. For the caliper matching bootstrapping was ran for 500 replications. However, due to the extremely computationally expensive, data and time-consuming nature of kernel matching and bootstrapping it was only possible to run 50 replications on the kernel matching algorithms.

5.5 Results

The previous sections of this chapter have outlined the research aims and the major rationale behind them. A clear gap in the literature was identified and the methodological approach was introduced showing how these aims would be achieved. This section presents the results of the regression and PSM analysis introduced previously. First, the individual's economic activity six months after graduating will be analysed. The focus then switches to only those graduates that were employed sixth months after graduating by analysing their first reported salary.

5.5.1 Impact of Student Migration Outcome on Economic Activity Six Months after Graduation

Economic activity is measured by a binary variable that categorises a graduate who is not in education and either employed or unemployed six months after graduation. After all non-UK domiciled graduates and those still in education were removed from the dataset there was data on the economic activity of 324,711 graduates, of which 87.4% were employed and 12.6% were unemployed.

In Chapter 4 it has been found that a student's ethnicity, social background, gender and institution attended significantly impacted on the migration outcome experienced when entering into the HE system in the UK. It was therefore justified to investigate if these covariates impacted on the employment status of students after graduation to see if these visible differences throughout the education system in the UK are still present when entering the labour market.

In a recent report by the Equality Challenge Unit (2014), a large amount of inequality in employment status sixth months after graduating is attributed to ethnicity. Using the same data source as used in this chapter (DLHE) the report states that the percentage of Black and minority graduates who were unemployed six months after graduating was 10.8% compared to only 5.2% for white graduates. Despite the report using different breakdowns of ethnicity and a deferent definition of unemployment (used the unemployment category vs all other individuals and did not removed those still in HE), similar patterns could be observed from our data. The breakdown of our economic activity variable by ethnicity is shown in Table 5-5.

There is a significant association between the employment status of students six months after graduation and their ethnic grouping at the 1% level as shown by the Chi-Squared test result. Clear differences were visible as a result of the graduates' ethnicity with black students having double the percentage unemployed compared to their white counterparts. Graduates from the Asian ethnic group also had much higher percentages unemployed that White graduates.

Table 5-5: Economic Activity six months after graduation by Ethnicity

	White (%)	Black (%)	Asian (%)	Other (%)	Unknown (%)	All (%)
Employed	88.77	77.99	80.80	81.68	85.43	87.35
Unemployed	11.23	22.01	19.20	18.32	14.57	12.65

Observations – 324,711

Pearson chi^2 (4df) = 3000.0 Pr = 0.000

Source: Higher Education Statistics Agency (2014a, b) and authors own calculations.

In Chapter 4, it was also been shown that a student's social background impacted on their migration outcome, however there seems to be less of an association between social background and employment after graduating as shown in Table 5-6. The Chi-Squared test shows the association between social background and employment was less than with ethnicity but the difference was still significant at the 1% level.

Table 5-6: Economic Activity six months after graduation by Social Background

	Most Advantaged (%)	Advantaged (%)	Less Advantaged (%)	Least Advantaged (%)	Unknown (%)	All (%)
Employed	86.15	87.19	86.98	86.84	89.21	87.35
Unemployed	13.85	12.81	13.02	13.16	10.79	12.65

Observations – 324,711

Pearson chi^2 (5df) = 315.7679 Pr = 0.000

Source: Higher Education Statistics Agency (2014a, b) and authors own calculations.

Once a student had graduated it seems that their student background became a less significant factor in determining the employment status as the differences between the social background groupings with regards to employment status were relatively small.

There was a slight difference in unemployment six months after graduating as a result of gender with 14.64% of males being unemployed compared to only 11.26% of females.

There was also a small but interesting relationship with institution category as shown in Table 5-7. The association as shown by the Chi-Squared is small but still significant at the 1% level.

Table 5-7: Economic Activity six months after graduation by Institution Category

	Ancient (%)	Red Brick (%)	Plate Glass (%)	Post 1992 (%)	Recent (%)	Other (%)	All (%)
Employed	86.76	87.39	87.15	87.24	88.31	87.46	87.35
Unemployed	13.24	12.61	12.85	12.76	11.69	12.54	12.65

Observations – 324,711

Pearson chi^2 (5df) = 33.7563 Pr = 0.000

Source: Higher Education Statistics Agency (2014a, b) and authors own calculations.

Although the differences in unemployment between the categories are relatively small the trend is of a surprising direction. Those graduates from Ancient Universities, such as University of Oxford and University of Cambridge, which are renowned as the most prestigious have the highest amount of unemployment. In contrast, those students from the recent universities created post-1992, mostly less prestigious former polytechnic institutions, had the lowest percentage of unemployment. However, this is likely to be a result of the subjects being studied and many courses offered by these HEIs having direct links to certain professions and employment opportunities after graduating.

It has been shown that when simply analysed individually, the covariates that were found in Chapter 4 to impact on the migration outcome into HE, had differing levels of impact on the students employment status after graduating. Therefore, can differences in migration outcomes be seen to directly impact on the economic activity of students if the covariates seem to impact on both the migration outcome and the economic activity?

By simply comparing the percentage of student migrants that were unemployed (13.29%) with the percentage of student non-migrants that were unemployed (12.05%), there appeared to be very little impact as a result of the student migration outcome experienced. However, by simply observing the percentages between two groups does not take into account other attributable factors that have already been shown in previous and the current chapters to impact on both employment status and the migration outcome. The next step of the analysis will therefore investigate how the detailed set of covariates impact on the economic activity of a student six months after graduating when the variables are analysed simultaneously using a logistic regression model.

The modelling procedure for the logistic model for unemployment was explained in Section 5.4.2 and a detailed explanation and the measures of fit are shown in Table E-1. The final model outputs for the chosen model (model 5) are shown in Table 5-8 and the nested model outputs for models 1 to 6 are shown in Table 5-9.

Table 5-8: Logistic Regression of Employment – Model 5: All covariates and an interaction between Gender and Background

VARIABLES	Coefficient (β)	P-Value	SE	e(β)
Constant	-3.087	0.000	(0.0426)	0.0456
Migrate LA				
Yes[a]				
No	0.0167	0.170	(0.0122)	1.017
Ethnicity				
White[a]				
Black	0.902	0.000	(0.0220)	2.465
Asian	0.745	0.000	(0.0178)	2.106
Other (Including Mixed Race)	0.561	0.000	(0.0259)	1.752
Unknown	0.361	0.000	(0.0399)	1.435
Background				
Most Advantaged[a]				
Advantaged	-0.0392	0.122	(0.0254)	0.962
Less Advantaged	-0.0516	0.037	(0.0247)	0.950
Least Advantaged	-0.0440	0.092	(0.0261)	0.957
Unknown	-0.124	0.000	(0.0274)	0.884
Gender				
Male[a]				
Female	-0.328	0.000	(0.0261)	0.721
Subject				
Medicine[a]				
Science/Engineering	1.117	0.000	(0.0233)	3.055
Argicultural/Veterinary	1.204	0.000	(0.0520)	3.333
Social/Human	0.679	0.000	(0.0237)	1.973
Business/Law	0.928	0.000	(0.0245)	2.529
Humanities	1.380	0.000	(0.0233)	3.975
Combined	1.611	0.000	(0.0802)	5.009
Institution Category				
Ancient[a]				
Red Brick	-0.0526	0.047	(0.0265)	0.949
Plate Glass	-0.136	0.000	(0.0282)	0.873
Post 1992	-0.162	0.000	(0.0261)	0.851
Recent University	-0.240	0.000	(0.0308)	0.786
Other	-0.307	0.000	(0.0333)	0.736
Age				
17 years and under	0.724	0.002	(0.237)	2.064
18-20 years[a]				
21-24 years	-0.103	0.000	(0.0134)	0.902
25-29 years	-0.0982	0.000	(0.0222)	0.906
30 years and over	0.0893	0.000	(0.0190)	1.093
Age unknown	-1.195	0.244	(1.025)	0.303
Level of Study				
Post-Graduate[a]				
Under-Graduate	0.500	0.000	(0.0171)	1.649
Domicile				
South[a]				
North	-0.0431	0.000	(0.0111)	0.958
Interaction: Background*Gender				
Most Advantaged*Male[a]				
Advantaged*Female	0.0150		(0.0361)	1.015
Less Advantaged*Female	0.194	0.000	(0.0341)	1.214
Least Advantaged*Female	0.209	0.000	(0.0351)	1.233
Unknown*Female	0.298	0.000	(0.0359)	1.347
Observations	324,711			
R-Squared				

Standard errors in parentheses

Notation "0.000" refers to P-Values smaller the 5×10^{-4}

[a] Denotes Reference Category

Table 5-9: Nested Logistic Regression of Employment - Outputs for Models 1 to 6

Model	M1		M2		M3		M4		M5 (Final Model)		M6	
Variables	**Coef.(β)**	**pval**	**Coef.(β)**	**pval**	**Coef.(β)**	**pval**	**Coef.(β)**	**pval**	**Coef.(β)**	**pval**	**Coef.(β)**	**pval**
Constant	-1.875	0.000	-3.185	0.000	-3.157	0.000	-3.118	0.000	-3.087	0.000	-3.137	0.000
Migrate (LA Level)												
Yes[a]												
No	-0.113	0.000	0.013	0.309	0.018	0.142	0.020	0.110	0.017	0.170	0.018	0.137
Ethnicity												
White[a]												
Black			0.915	0.000	0.900	0.000	0.617	0.000	0.902	0.000	0.772	0.000
Asian			0.749	0.000	0.742	0.000	0.473	0.000	0.745	0.000	0.636	0.000
Other			0.569	0.000	0.559	0.000	0.440	0.000	0.561	0.000	0.553	0.000
Unknown			0.361	0.000	0.356	0.000	0.0481	0.765	0.361	0.000	0.278	0.000
Background												
Most Advantaged[a]												
Advantaged			-0.032	0.079	-0.032	0.081	-0.058	0.004	-0.039	0.122	-0.031	0.079
Less Advantaged			0.047	0.008	0.047	0.007	0.007	0.711	-0.052	0.037	0.047	0.007
Least Advantaged			0.061	0.001	0.064	0.001	-0.001	0.952	-0.044	0.092	0.064	0.001
Unknown			0.0247	0.222	0.0290	0.151	-0.048	0.033	-0.124	0.000	0.028	0.167
Gender												
Male[a]												
Female			-0.178	0.000	-0.178	0.000	-0.178	0.000	-0.328	0.000	-0.221	0.000
Subject												
Medicine[a]			1.110	0.000	1.108	0.000	1.108	0.000	1.117	0.000	1.109	0.000
Science/Engineering			1.195	0.000	1.193	0.000	1.191	0.000	1.204	0.000	1.198	0.000
Argicultural/Veterinary			0.682	0.000	0.679	0.000	0.680	0.000	0.679	0.000	0.681	0.000
Social/Human			0.922	0.000	0.920	0.000	0.920	0.000	0.928	0.000	0.923	0.000
Business/Law			1.374	0.000	1.369	0.000	1.367	0.000	1.380	0.000	1.372	0.000
Humanities			1.607	0.000	1.605	0.000	1.604	0.000	1.611	0.000	1.607	0.000
Combined			1.110	0.000	1.108	0.000	1.108	0.000	1.117	0.000	1.109	0.000
Institution Category												
Ancient[a]												
Red Brick			-0.046	0.080	-0.052	0.049	-0.050	0.058	-0.052	0.047	-0.052	0.049
Plate Glass			-0.130	0.000	-0.136	0.000	-0.131	0.000	-0.136	0.000	-0.136	0.000
Post 1992			-0.157	0.000	-0.161	0.000	-0.159	0.000	-0.162	0.000	-0.162	0.000
Recent University			-0.231	0.000	-0.241	0.000	-0.237	0.000	-0.240	0.000	-0.240	0.000
Other			-0.299	0.000	-0.309	0.000	-0.305	0.000	-0.307	0.000	-0.309	0.000
Age												
17 years and under			0.718	0.002	0.730	0.002	0.727	0.002	0.724	0.002	0.727	0.002
18-20 years[a]												
21-24 years			-0.100	0.000	-0.101	0.000	-0.101	0.000	-0.103	0.000	-0.100	0.000
25-29 years			-0.097	0.000	-0.098	0.000	-0.096	0.000	-0.098	0.000	-0.097	0.000
30 years and over			0.095	0.000	0.092	0.000	0.095	0.000	0.089	0.000	0.095	0.000
Age unknown			-1.193	0.245	-1.203	0.241	-1.181	0.250	-1.195	0.244	-1.208	0.239
Level of Study												
Post-Graduate[a]												
Under-Graduate			0.498	0.000	0.501	0.000	0.499	0.000	0.500	0.000	0.500	0.0000
Domicile												
South[a]												
North					-0.044	0.000	-0.043	0.000	-0.043	0.000	-0.043	0.000
Interaction: Ethnicity*Background												
White*Most Advantaged[a]							0.150	0.080				
Black*Advantaged							0.241	0.001				
Black*Less Advantaged							0.371	0.000				
Black*Least Advantaged							0.484	0.000				
Black*Unknown							0.216	0.002				
Asian*Advantaged							0.261	0.000				
Asian*Less Advantaged							0.330	0.000				
Asian*Least Advantaged							0.393	0.000				
Asian*Unknown							-0.007	0.937				
Other*Advantaged							0.151	0.057				
Other*Less Advantaged							0.214	0.009				
Other*Least Advantaged							0.213	0.013				
Other*Unknown							0.505	0.010				
Unknown*Advantaged							0.395	0.033				
Unknown*Less.Adv							0.394	0.037				
Unknown*Least.Adv							0.273	0.111				
Unknown*Unknown												
Interaction: BackgroundGender												

Model	M1		M2		M3		M4		M5 (Final Model)		M6	
Variables	**Coef.(β)**	**pval**	**Coef.(β)**	**pval**	**Coef.(β)**	**pval**	**Coef.(β)**	**pval**	**Coef.(β)**	**pval**	**Coef.(β)**	**pval**
Most Advantaged*Male[a]												
Advantaged*Female									0.015	0.677		
Less Advantaged*Female									0.194	0.000		
Least Advantaged*Female									0.209	0.000		
Unknown*Female									0.298	0.000		
Ethnicity*Gender												
White*Male[a]												
Black*Female											0.224	0.000
Asian*Female											0.206	0.000
Other*Female											0.0134	0.796
Unknown*Female											0.160	0.044
Observations	324,711		324,711		324,711		324,711		324,711		324,711	
R-Squared	0.0005		0.0408		0.0409		0.0413		0.0413		0.0413	
BIC	101.538		9723.437		9726.077		9627.262		9783.713		9734.959	
Difference in BIC	-		9621.899		2.640		-98.815		57.636		8.882	
Evidence	-		V.V. Strong		Medium		V. Strong Negative (Nested in M3)		V. Strong (Nested in M3)		Strong (Nested in M3)	

Notation "0.000" refers to P-Values smaller the 5 x 10^{-4}

Source: Higher Education Statistics Agency (2014a, b)

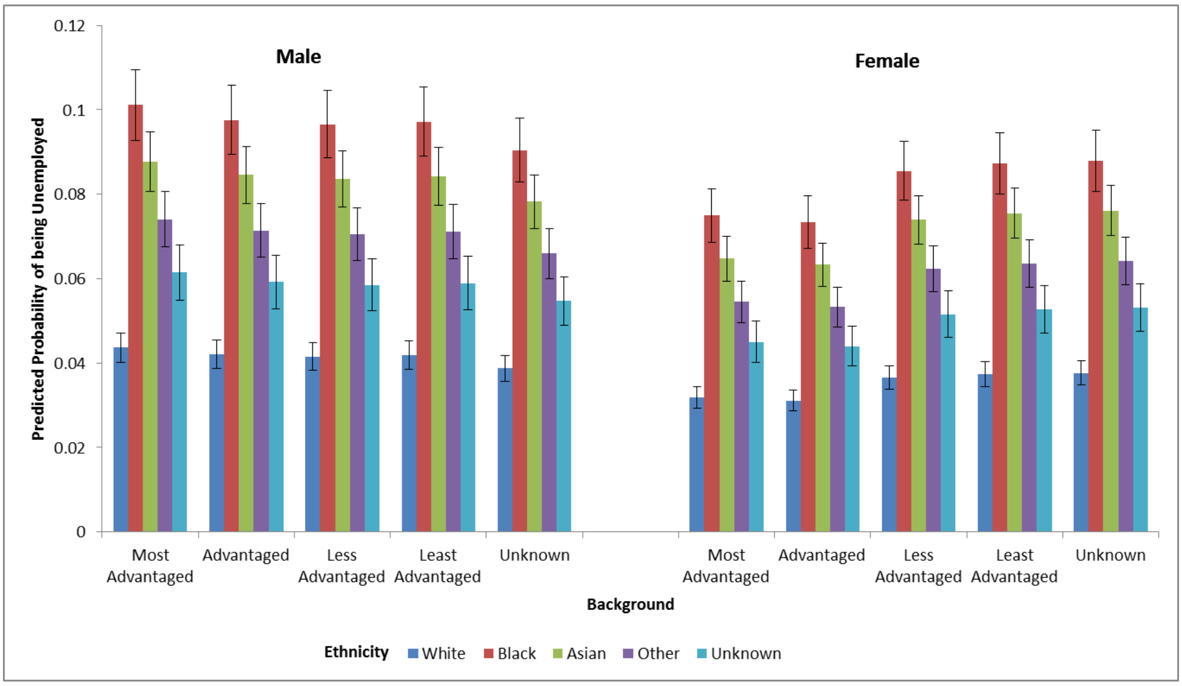
When interpreting the output shown in Table 5-8, it is important to note that the main effects of the variables involved in the significant interaction terms cannot be interpreted individually because the individual main effects of interacted variables cannot be isolated. Therefore, interpretation regarding the background and gender variables should be in terms of their interactions in order to make appropriate conclusions.

From evaluating the output and predicted probabilities from the logistic regression model of the covariates in the dataset, it is clear that some variables had a significant relationship with employment activity after graduation. When simply looking at the differences in the percentage unemployed for migrants compared to non-migrants there seemed to be only a marginal impact of migration and this was confirmed in the regression modelling. The variable migrate which depicts whether a student migrated at the local authority level or not to attend a HEI was insignificant in the final chosen regression model (model 5). Migrate was insignificant in all the models except Model 1, as shown in Table 5-9. Model 1 was the only model where Migrate was the only variable, therefore this shows when other covariates are considered the effect of migration on unemployment became insignificant.

All the other variables in the model were significant at the 5% level (or better) with the exception of some categories of student background. Therefore, unemployment six months after graduating is impacted by several factors but the migration decision into university was insignificant. From this evidence it suggests migrating to attend a HEI provides no future value with regards to employment status after graduating.

In order to evaluate the predicted impact of the other variables in the model of unemployment, the predicted probabilities of being unemployed are calculated by substituting the coefficients in Table 5-8 into the regression Equation 5.1 and have been graphed and explained below. The predicted probabilities and 95% confidence intervals by ethnicity, background and gender are shown in Figure 5-5.

Figure 5-5: Predicted Probabilities of Unemployment six months after graduation by Gender, Ethnicity and Social Background



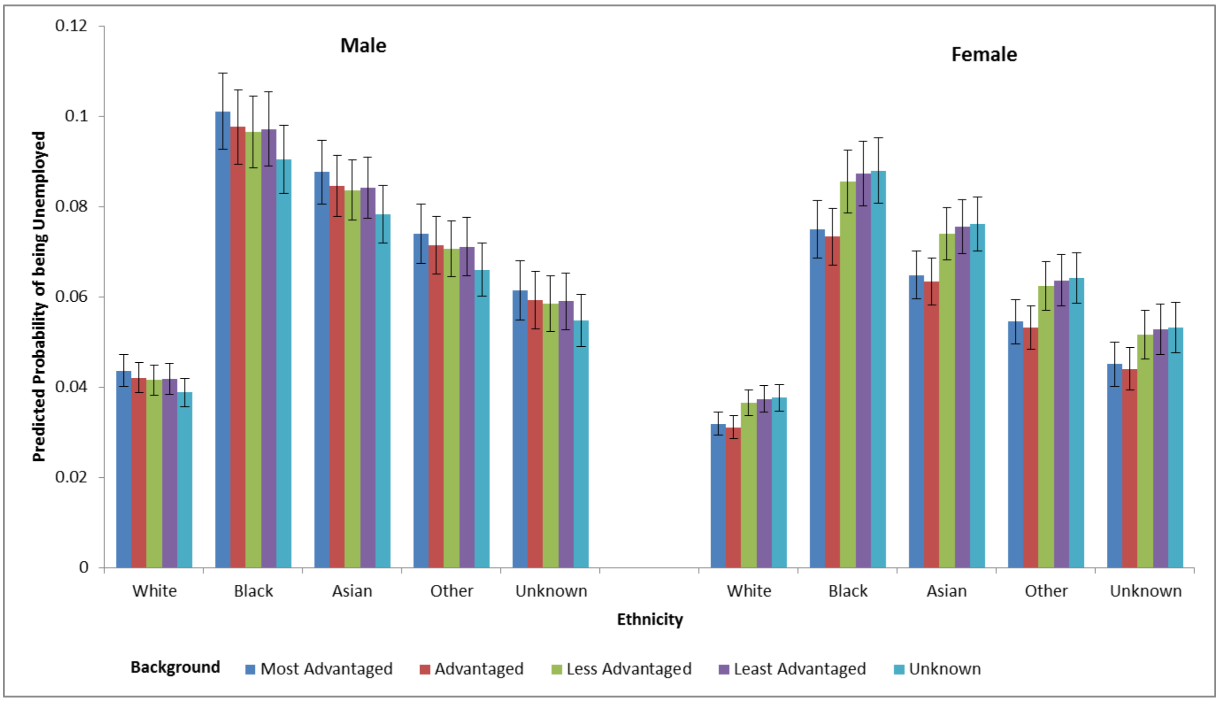
Source: Higher Education Statistics Agency (2014a, b)

Note: The predicted probabilities assume the remaining variables were set to the reference category.

There is a clear statistically significant difference in the probability of being unemployed between white and non-white graduates. White graduates for all background groups and both genders had the lowest predicted probabilities of being unemployed. In contrast the black ethnic group had the highest predicted probability of unemployment. These results mirror those described in the basic descriptive analysis (Table 5-5) and in the previous research which indicate large ethnic differences in employment even in those with the highest educational qualifications.

The effect of a student's social background can also be observed from Figure 5-5 but for an easier interpretation Figure 5-6 has been rearranged to highlight the differences by social background. Due to the significant interaction between gender and social background the trends appear to not be the same for both sexes. Males from more advantageous backgrounds appear more likely to be unemployed than males from less advantageous backgrounds, while the inverse pattern is seen for females. However, when the 95% confidence intervals are observed there is significant overlap between the predicated probabilities of unemployment between the background groupings for both sexes and therefore it can be concluded that there is no significant difference in the predicted probabilities of unemployment as a result of the students social background.

Figure 5-6: Predicted Probabilities of Unemployment six months after graduation by Gender, Social Background and Ethnicity



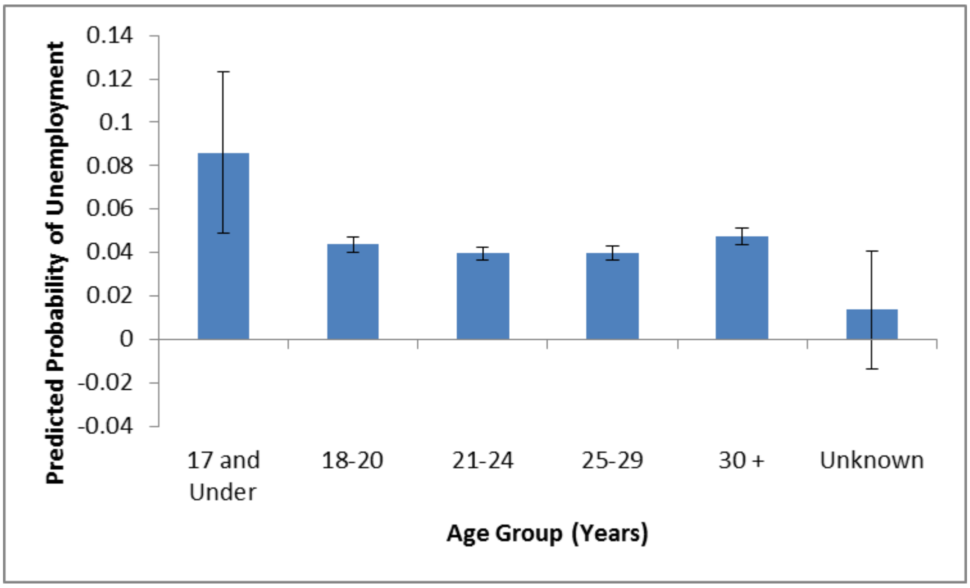
Source: Higher Education Statistics Agency (2014a, b)

Note: The predicted probabilities assume the remaining variables were set to the reference category.

The predicted probabilities of being unemployed by age group are shown in Figure 5-7. There was no significant difference in the probability of unemployment according to age. The 17 and under and Unknown groups make up very small proportions of those that are leaving HE and this therefore explains the differences in the findings for these groups and the large 95% confidence intervals.

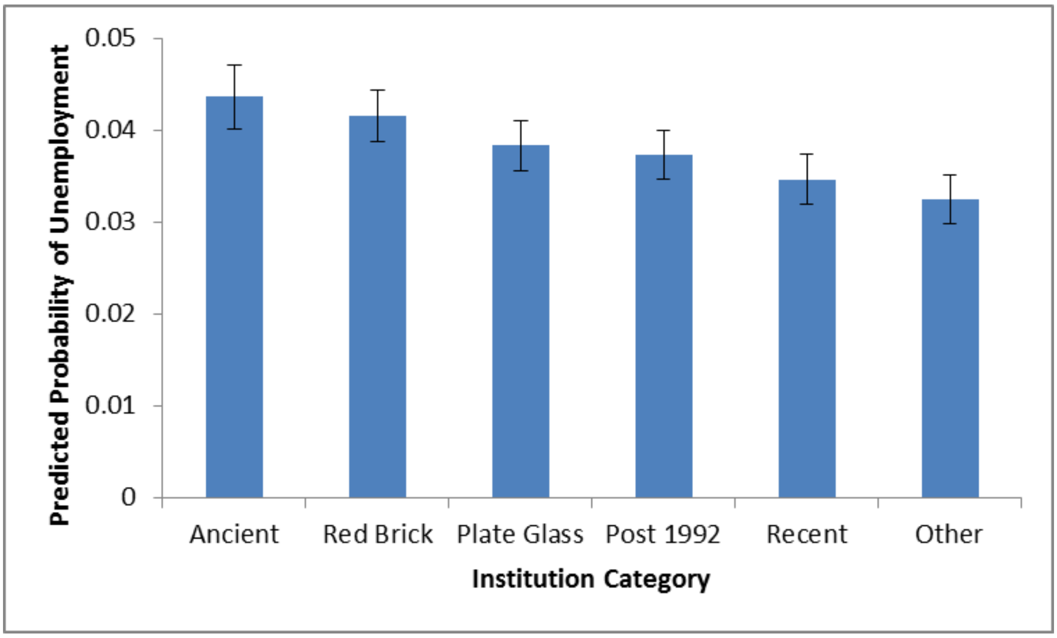
The predicted probabilities by institution category of the HEI attended by the student are shown in Figure 5-8. The association here shows that the graduates from the more prestigious universities have a higher predicted probability of being unemployed six months after graduating than those students that graduated from newer more recent institutions. These differences in predicted unemployment were statistically significant for institutions at the ends of the categorisation with students from Ancient Universities statistically more likely to be unemployed than students from recent universities at the 5% level.

Figure 5-7: Predicted Probabilities of Unemployment six months after graduation by Age Group



Source: Higher Education Statistics Agency (2014a, b)
Note: The predicted probabilities assume the remaining variables were set to the reference category.

Figure 5-8: Predicted Probabilities of Unemployment six months after graduation by Institution Category



Source: Higher Education Statistics Agency (2014a, b)

Note: The predicted probabilities assume the remaining variables were set to the reference category.

By evaluating the regression modelling output the results indicate that when all the covariates are considered simultaneously only ethnicity and institution attended significantly impacted on the employment status of students six months after graduating. One of the aims of this chapter is to evaluate the value of migrating into HE on the student's future labour market outcomes, however, the basic statistics and the regression modelling indicated that the migration outcome experienced by the student had no significant association with the employment activity six months after graduating.

As discussed in Section 5.4, in order to evaluate the causal impact of migration accurately, regression techniques do not suffice in isolating the causal impact as we are often interested in a counterfactual that cannot be observed. We therefore conducted a series of PSM techniques in order to effectively evaluate the causal impact of migration into HE on economic activity six months after graduation.

Table 5-10: The impact of migration on Employment – Regression and PSM results

Outcome - Unemployment		Migrated	SE	P-Value
Model 1	Logistic Regression	0.0116	(0.0119)	0.327
	PSM Caliper	0.0492	(0.0294)	0.094
	PSM Kernel	0.0043	(0.0014)	0.002
Model 2	Logistic Regression	-0.0123	(0.0121)	0.309
	PSM Caliper	0.0193	(0.0146)	0.186
	PSM Kernel	0.0013	(0.0015)	0.368
Model 3	Logistic Regression	-0.0132	(0.0121)	0.274
	PSM Caliper	0.0197	(0.0146)	0.178
	PSM Kernel	0.0016	(0.0015)	0.301
Model 4	Logistic Regression	Not applicable to Unemployment Analysis		
	PSM Caliper			
	PSM Kernel			

Standard errors for PSM are Bootstrapped; 50bs for Caliper and 500bs for Kernel
Notation "0.000" refers to P-Values smaller the 5 x 10^{-4}
The model outputs for the logistic regression models used here can be found in Table E-3

Source: Higher Education Statistics Agency (2014a, b)

The difference in the predicted probability of being unemployed for those that migrated compared to those that did not have been calculated using logistic regression, PSM with caliper matching and PSM with kernel matching. These techniques were used on three different model specifications (as explained in Table 5-4) and the results of the three techniques and the three models are shown in Table 5-10.

It can be seen from Table 5-10 that across the three models and three techniques used to analyse the effect of migration into HE on unemployment only the PSM techniques for Model 1 produced significant results at the 10% level. The PSM caliper predicted that students who migrated to attend a HEI were 4.9% more likely to be unemployed than non-migrants at the 10% level. The PSM kernel predicted the same directional finding however with a much smaller impact, with migrants predicted to be 0.4% more likely to be unemployed, however the PSM kernel was significant at the 1% level. Model 1, however, was only the basic model and didn't control for institution category or any interactions between the variables. Because of this there may well be bias in the result due to the CIA assumption not being met as a result of other variables not included in Model 1 impacting on the level of unemployment.

Models 2 and 3 do include more variables and include institution category which is being used as a proxy for ability and attainment. Therefore it is more likely that the CIA assumption holds for Model 2 and 3. However, the predicted differences for Models 2 and 3 were not significant at the 15% level and therefore indicted there was no significant impact on migration once controlling for institution category.

From analysing the factors effecting a graduate's economic activity six months after graduating it has been found that only a student's ethnicity had a statistically significant impact on unemployment. It was shown by using regression and matching techniques that the migration transition experienced entering HE in the UK had no significant impact regarding the individuals employment status after achieving a HE degree. Graduate employment research has evidenced that students who leave home develop self-sufficiency sooner and therefore have potentially better job prospects than those who stay with family (The Complete University Guide 2014), however our research shows no support towards this claim. These findings were shown in all three models showing that when the matching techniques were conducted conditional on institution category or not, there was found to be no difference in future employment status as a result of the migration choice undertaken when entering HE.

5.5.2 Impact of Student Migration Outcome on Salary Activity Six Months after Graduation

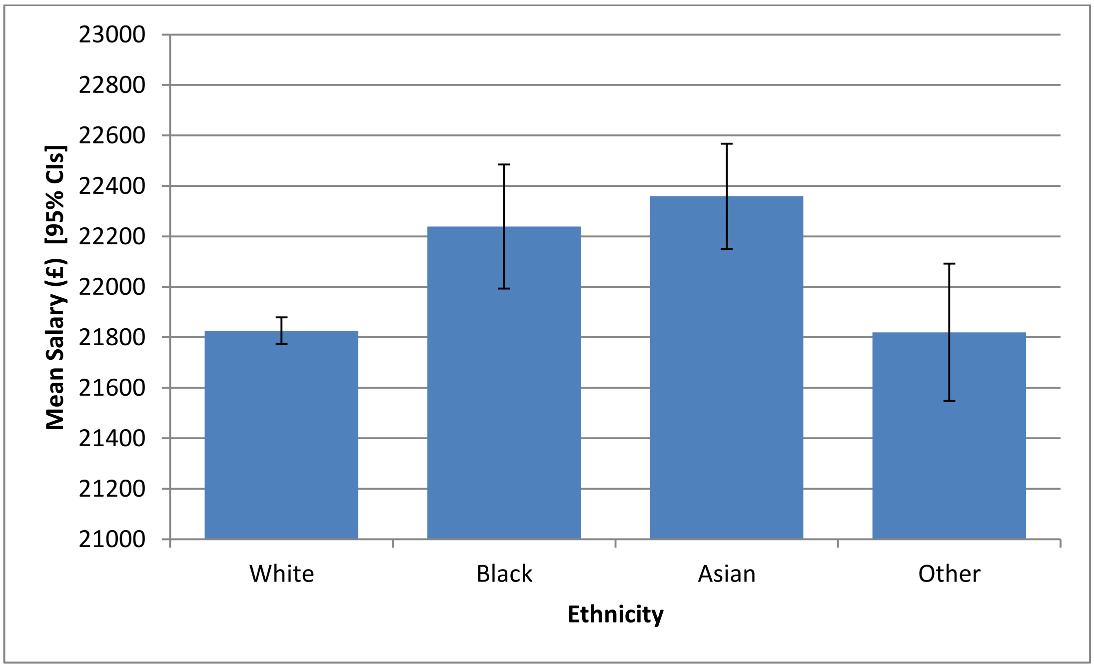
After observing the factors that impact on a graduate's employment status, it was of interest to investigate if there were any differences associated with the covariates available and the amount a graduate earned six months after graduating, conditional on them being employed. This will enable us to conclude if any of the inequalities observed throughout the education system in the UK can be found to have transferred into the labour market outcomes and therefore impact on an individual's future life success.

Salary is measured using a continuous variable that recorded a graduate's salary six months after graduation as long as the graduate was not in further study and was employed. The dataset contained information on 171,581 graduate's salary six months after graduation. The mean salary was £21,951

(SE - 10332) and the distribution of the salary variable can be seen in Figure 5-1. However, as per all income distributions, due to their skewed nature, care must be taken when interpreting simple sample means. It must also be stated that by simply analysing the sample means of salary from the DLHE, the item non-response biases, as discussed in detail in Section 5.3.3, are not taken into account and care should be used when interpreting these values. Later in the chapter, salary is analysed in more complexity using regression and propensity score matching techniques, which provide more intuitive findings as opposed to simply comparing sample means of salary.

As per economic activity, a logical place to start analysing differences in salary was to investigate if there was any association between first salary and ethnicity, background and gender. The difference in mean salary by ethnicity is shown in Figure 5-9. It can be seen here that there were no large scale differences in mean salary as a result of a graduate's ethnicity.

Figure 5-9: Mean salary six months after graduation by Ethnicity

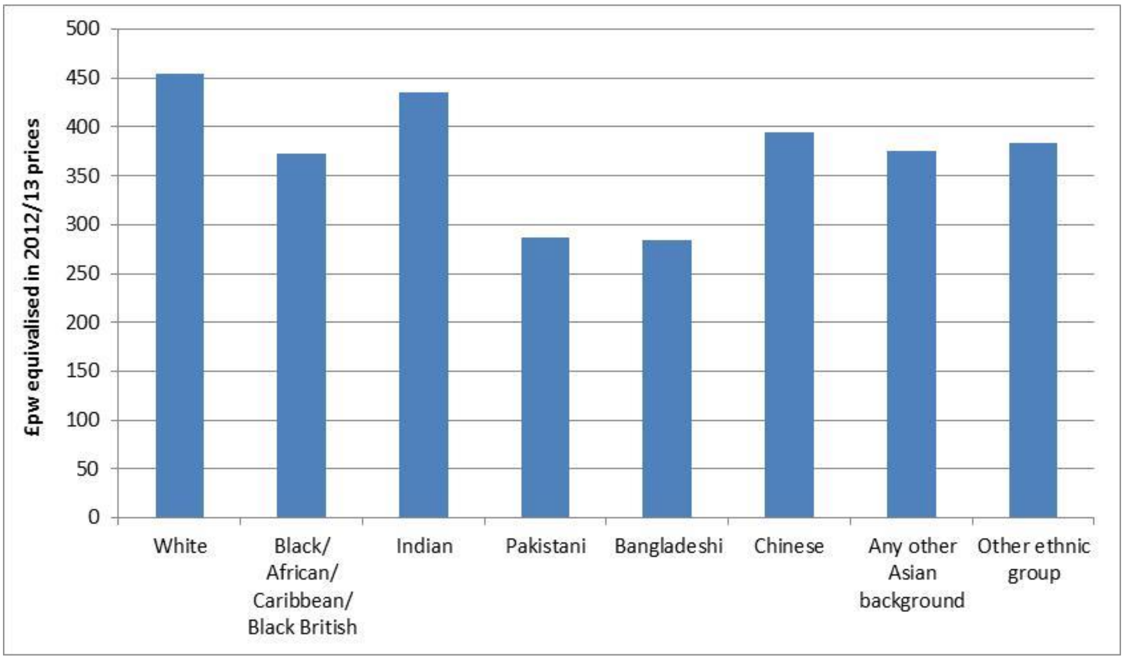


Source: Higher Education Statistics Agency (2014a)

Although the differences in actual wage levels between the ethnic groups are quite small, it can be seen from the 95% confidence intervals in Figure 5-9 that the differences between the white and non-white groups were statistically significant at the 5% level as the confidence intervals do not overlap. The findings shown from comparing the sample mean salary's by ethnicity showed a counter-intuitive trend with white graduates having the lowest mean salary of all the ethnic groups. It was shown earlier that the white ethnic group had a higher percentage of graduates employed than the non-white groups. However, it is shown here that although more white graduates were employed, of those that were employed, white graduates had a slightly lower mean salary that there non-white counterparts.

These counter-intuitive findings show a contrast to the trends observed for the total population of the UK which found that there were substantial ethnic inequalities with regards to earnings, with Black and Asian ethnic groups in general earning significantly less than those from the White ethnic group (Hills et al. 2010; Platt 2011). Data from the Department for Work and Pensions (2015) Family Resources Survey also show that the median weekly equivalised net disposable household income is highest for white ethnic groups as shown in Figure 5-10. This raises questions as to why the mean salary data for graduates from the DLHE show this opposite trend to other data sources for the whole population. Is there a real difference in graduate earnings by ethnicity compared to the whole population or are these findings a caused by another contributing factor?

Figure 5-10: Median weekly equivalised net disposable household income for all individuals, by ethnicity of household reference person, in average 2012/13 prices, United Kingdom

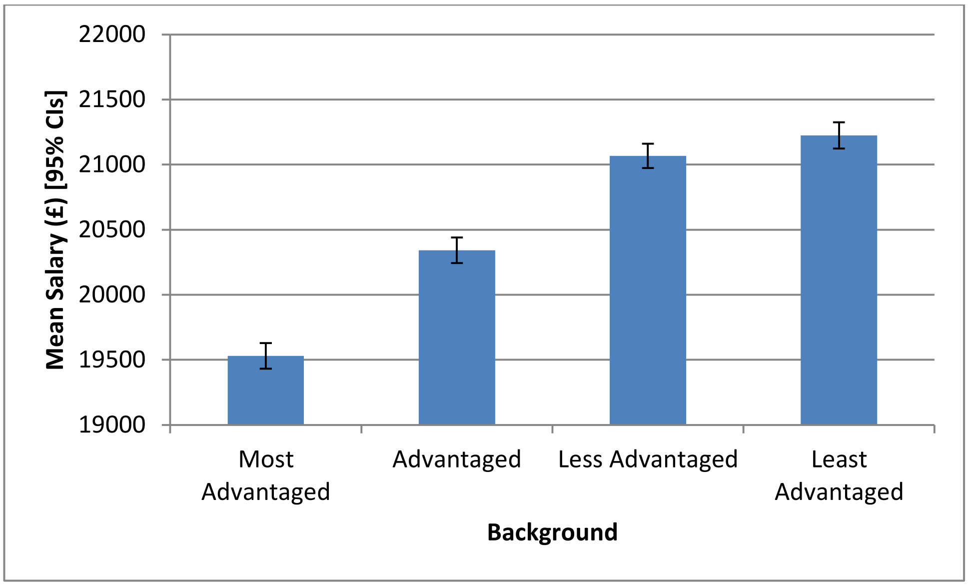


Source: Department for Work and Pensions (2015)

These counter-initiative findings from the mean salary from the DLHE are likely to have been skewed as a result of the item and unit non-response observed in the DLHE as discussed in Section 5.3.3. As previously discussed, Black and Asian groups were under-represented in the DLHE as a whole and it might be the case that those that did respond were those in higher paid jobs while those that did not respond were either unemployed or in a low paid job in which they did not want to report to the DLHE. There was also a large amount of item non-response within the salary question (Table G-3) in which a high number of white responders to the survey did not respond to the question regarding salary. If these white item non-responder were in high paying salaries this would explain the skewed mean salary findings within the DLHE. However, there is no way for this to be proved but it is highly likely that the non-response biases are a key contributing factor here and it is therefore important to consider this when interpreting simple sample means and provides rational to explore these difference in more detail later in the chapter.

The mean salary for graduates by social background is shown in Figure 5-11. As for ethnicity, there are slight but statistically significant (at the 5% level) differences in mean salary by social background. However, these differences were also in the opposite direction than one might expect. Mean earnings are highest for those from the least advantaged background and lowest for those from the most advantaged backgrounds.

Figure 5-11: Mean salary six months after graduation by Social Background



Source: Higher Education Statistics Agency (2014a, b)

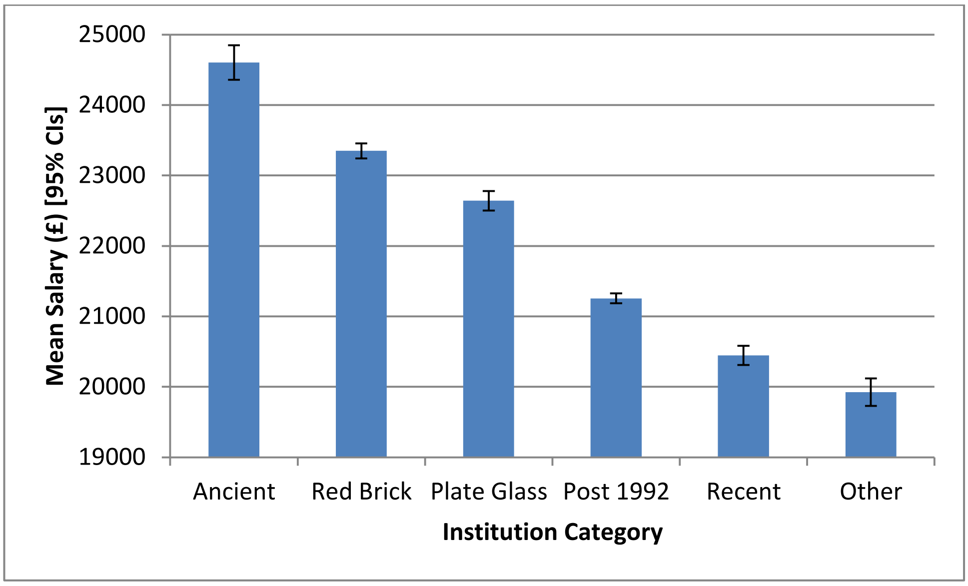
Again, the issues highlighted above regarding the impact of non-response in the DLHE on ethnicity also hold herewith regarding social background. It was not possible to find any data sources to directly compare the findings of mean salary by the background variable used in this study due to how the background variable in this dataset is constructed. However, a number of studies have shown there to be a clear link between income and socio-economic status for the population as a whole (Pickett and Wilkinson 2010; Payne 2013; Savage et al. 2013), that clearly show that the trends shown in Figure 5-11 are in the opposite direction to what the previous studies suggest. One would expect that those graduates from the most advantageous background to earn the highest salaries, although the sample means from the DLHE suggest the otherwise.

It is highly likely that these sample means are being skewed by high levels of item non-response within the DLHE, where large numbers of most advantaged and advantages survey responders did not respond to the question regarding salary (Table G-4). However, as mentioned previously, there is no way for this to be proved but it is highly likely that the non-response biases are a key contributing factor towards these counter-intuitive findings regarding mean first salary.

There was also a clear gender gap by mean salary, with males having a mean salary £2383 a year higher than females, that shows that males were earning on average 11.34% more six months after graduating than there female peers. This gender gap was significant at the 1% level.

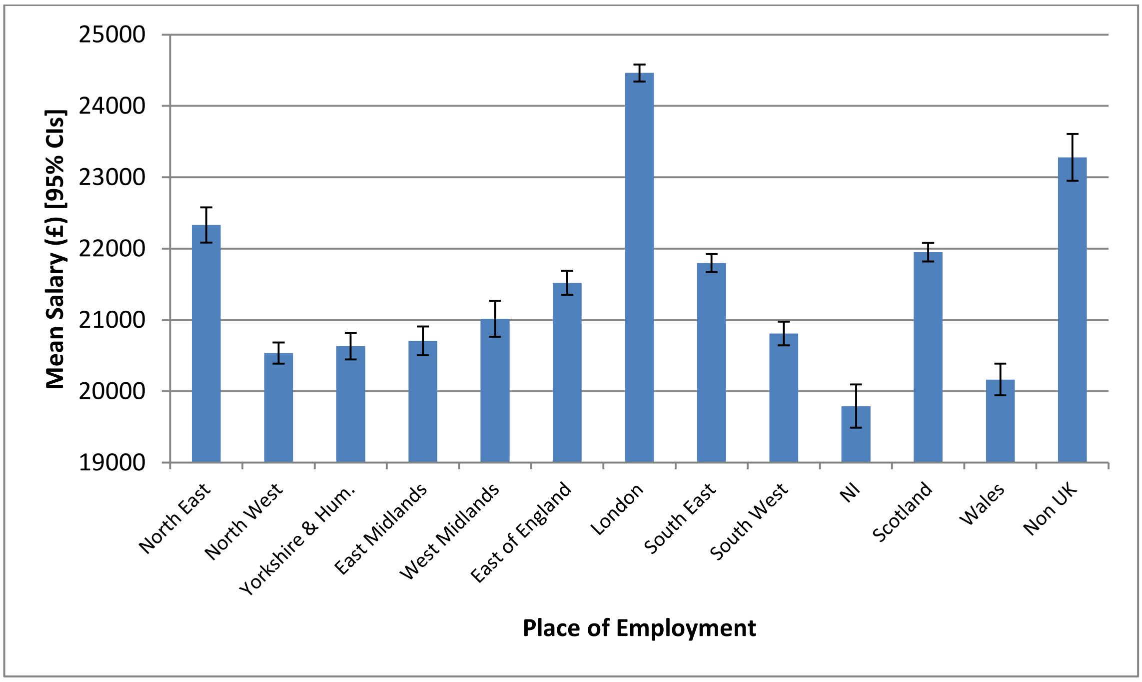
There were also quite substantial and significant differences in the mean salary according to the type of institution that the graduate attended, as shown in Figure 5-12. It was shown previously that the institution category had the inverse association with employment status than expected with the more reputable universities being associated with higher levels of unemployment compared to the newer less reputable institutions. However, when analysing salary the association is opposite to that of employment. For those graduates that are employed those from the ancient universities earn significantly more than any other institution group and there is a liner trend of mean salary across the categories, as reputation and prestige of the institution decreases, so does the mean first salary. This indicates that employers place value in where the student gained their degree and this may indicate value for migrating in order to attend a higher reputable institution as the data indicates students from these institutions earn more on average than lower ranking institutions.

Figure 5-12: Mean salary six months after graduation by Institution Category



Source: Higher Education Statistics Agency (2014a, b)

Figure 5-13: Mean salary six months after graduation by Place of Employment



Source: Higher Education Statistics Agency (2014a, b)

Differences in mean salary by place of employment are shown in Figure 5-13. Clear differences can be seen in the mean salary with London having the highest mean salary. As mentioned earlier this was not surprising seeing as salaries in London are slightly inflated using the London Wage Weighting. As a result of these statistically significant differences in wages associated with location, it would sensible to control for these differences later in the analysis.

When observing the mean salary for those that migrated compared to those that did not, there seemed to be the opposite in terms of value of migrating. Those that migrated to attend a HEI earned on average £2104.77 a year less than those students that did not migrate. However, simply observing the means between the migrants and non-migrants does not take into account other attributes that have already been proven to impact on salary and the migration outcome. Also differences in mean salary by gender, institution, ethnicity etc. might be attributable to interlinked differences. For example, the type of courses and industries frequented by males may differ to those frequented by females. This could be a driving factor in the simple observed differences in mean salary by gender. As a result of this the next step of the analysis therefore investigates how the detailed set of covariates impact on first salary after graduating when the variables are analysed simultaneously using a multiple regression model.

As explained in Section 5.4.2, in order to be in accordance with the assumptions of a multiple linear regression model, the natural logarithm transformation of salary was modelled as opposed to the highly skewed original salary variable. From the modelling procedure shown in Table E-2 the final chosen multiple linear regression model for log_salary was Model 8. This model provides the best statistical fit to the data as tested in the modelling process and all variables (with the exception of Domicile North/South) in the model added statistical significance while making theoretical sense. This model included all the variables available from the Student Record Data and the DLHE as well as an interaction term between Gender and Background. The regression output of the final model (Model 8) is shown in Table 5-11 and the nested model outputs for models 1 to 9 are shown in Table 5-12. It must be remember when interpreting the results that the model is using the natural logarithm transformation as explained in Section 5.4 and the output must be interpreted accordingly.

Table 5-11: Multiple Linear Regression of First Salary – Model 8: All covariates and an interaction between Gender and Background

Variables	Coefficient (β)	P-Value	e(β)
Constant	10.347	0.000	31,163
Migrate (LA Level)			
Yes[a]			
No	-0.023	0.000	0.977
Ethnicity			
White[a]			
Black	-0.07	0.000	0.932
Asian	-0.028	0.000	0.972
Other	-0.02	0.000	0.980
Unknown	0.016	0.013	1.016
Background			
Most Advantaged[a]			
Advantaged	0.007	0.124	1.007
Less Advantaged	-0.009	0.021	0.991
Least Advantaged	-0.005	0.227	0.995
Unknown	0.054	0.000	1.055
Gender			
Male[a]			
Female	-0.045	0.000	0.956
Subject			
Medicine[a]			
Science/Engineering	-0.074	0.000	0.929
Argicultural/Veterinary	-0.037	0.000	0.964
Social/Human	-0.069	0.000	0.933
Business/Law	-0.006	0.083	0.994
Humanities	-0.162	0.000	0.850
Combined	-0.081	0.000	0.922
Institution Category			
Ancient[a]			
Red Brick	-0.028	0.000	0.972
Plate Glass	-0.044	0.000	0.957
Post 1992	-0.066	0.000	0.936
Recent University	-0.072	0.000	0.931
Other	-0.08	0.000	0.923
Age			
17 years and under	-0.013	0.832	0.987
18-20 years[a]			
21-24 years	0.048	0.000	1.049
25-29 years	0.135	0.000	1.145
30 years and over	0.263	0.000	1.301
Age unknown	0.554	0.000	1.740
Level of Study			
Post-Graduate[a]			
Under-Graduate	-0.139	0.000	0.870
Domicile			
South[a]			
North	-0.002	0.405	0.998
SIC			
Managers, Directors and Senior Officials[a]			
Professional occupations	-0.102	0.000	0.903
Associate professional and tech. occ…	-0.232	0.000	0.793
Administrative and secretarial occup…	-0.428	0.000	0.652
Skilled trades occupations	-0.343	0.000	0.710
Personal service occupations	-0.558	0.000	0.572
Sales and customer service occupations	-0.584	0.000	0.558
Process, plant and machine operatives	-0.519	0.000	0.595
Elementary occupations	-0.61	0.000	0.543
SOC			

Agriculture, forestry and fishing[a]			
Mining and quarrying	0.425	0.000	1.530
Manufacturing	0.212	0.000	1.236
Electricity, gas, steam…	0.304	0.000	1.355
Water supply, sewerage, waste…	0.241	0.000	1.273
Construction	0.168	0.000	1.183
Wholesale and retail trade; repair…	-0.033	0.065	0.968
Transport and storage	0.187	0.000	1.206
Accommodation and food …	-0.072	0.000	0.931
Information and communication	0.095	0.000	1.100
Financial and insurance activities	0.233	0.000	1.262
Real estate activities	0.057	0.003	1.059
Professional, scientific and tech…	0.067	0.000	1.069
Administrative and support serv…	0.061	0.001	1.063
Public administration and def…	0.172	0.000	1.188
Education	0.016	0.383	1.016
Human health and social work activities	0.09	0.000	1.094
Arts, entertainment and recreation	-0.096	0.000	0.908
Other service activities	-0.044	0.025	0.957
Activities of households as employ…	0.074	0.038	1.077
Activities of extraterritorial…	0.24	0.000	1.271
Place of Employment			
North East	-0.156	0.000	0.856
North West	-0.172	0.000	0.842
Yorkshire and the Humber	-0.166	0.000	0.847
East Midlands	-0.165	0.000	0.848
West Midlands	-0.175	0.000	0.839
East of England	-0.132	0.000	0.876
London[a]			
South East	-0.123	0.000	0.884
South West	-0.177	0.000	0.838
Northern Ireland	-0.238	0.000	0.788
Scotland	-0.158	0.000	0.854
Wales	-0.212	0.000	0.809
Non-UK	-0.139	0.000	0.870
Interaction: Background*Gender			
Most Advantaged*Male[a]			
Advantaged*Female	-0.007	0.221	0.993
Less Advantaged*Female	-0.023	0.000	0.977
Least Advantaged*Female	-0.042	0.000	0.959
Unknown*Female	-0.064	0.000	0.938
Observations		169,887	
R-Squared		0.478	

Notation “0.000” refers to P-Values smaller the 5 x 10^{-4}

Source: Higher Education Statistics Agency (2014a, b)

Table 5-12: Nested Multiple Linear Regression of First Salary - Outputs for Models 1 to 9

Model	**M1**		**M2**		**M3**		**M4**		**M5**		**M6**		**M7**		**M8 (Final Model)**		**M9**	
Variables	**Coef.(β)**	**pval**	**Coef.(β)**	**pval**	**Coef.(β)**	**pval**	**Coef.(β)**	**pval**	**Coef.(β)**	**pval**	**Coef.(β)**	**pval**	**Coef.(β)**	**pval**	**Coef.(β)**	**pval**	**Coef.(β)**	**pval**
Constant	9.849	0.000	10.243	0.000	10.285	0.000	10.346	0.000	10.267	0.000	10.364	0.000	10.361	0.000	10.347	0.000	10.364	0.000
Migrate (LA Level)																		
Yes[a]																		
No	0.089	0.000	-0.036	0.000	-0.028	0.000	-0.025	0.000	-0.027	0.000	-0.023	0.000	-0.023	0.000	-0.023	0.000	-0.023	0.000
Ethnicity																		
White[a]																		
Black			-0.044	0.000	-0.066	0.000	-0.022	0.000	-0.025	0.000	-0.069	0.000	-0.040	0.001	-0.070	0.000	-0.077	0.000
Asian			-0.005	0.205	-0.015	0.000	0.002	0.530	0.002	0.600	-0.027	0.000	0.013	0.171	-0.028	0.000	-0.018	0.000
Other			0.008	0.156	-0.007	0.211	0.005	0.304	0.007	0.115	-0.020	0.000	-0.001	0.957	-0.020	0.000	-0.027	0.000
Unknown			0.023	0.002	0.017	0.024	0.026	0.000	0.028	0.000	0.017	0.008	-0.002	0.926	0.016	0.013	0.045	0.000
Background																		
Most Advantaged[a]																		
Advantaged			0.000	0.925	0.000	0.907	0.004	0.194	0.001	0.660	0.003	0.316	0.006	0.058	0.007	0.124	0.003	0.324
Less Advantaged			-0.037	0.000	-0.037	0.000	-0.022	0.000	-0.023	0.000	-0.023	0.000	-0.020	0.000	-0.009	0.021	-0.023	0.000
Least Advantaged			-0.060	0.000	-0.057	0.000	-0.033	0.000	-0.033	0.000	-0.030	0.000	-0.023	0.000	-0.005	0.227	-0.030	0.000
Unknown			-0.009	0.010	-0.003	0.383	0.011	0.000	0.009	0.003	0.017	0.000	0.021	0.000	0.054	0.000	0.017	0.000
Gender																		
Male[a]																		
Female			-0.111	0.000	-0.110	0.000	-0.087	0.000	-0.074	0.000	-0.074	0.000	-0.075	0.000	-0.045	0.000	-0.073	0.000
Subject																		
Medicine[a]																		
Science/Engineering			-0.148	0.000	-0.151	0.000	-0.054	0.000	-0.074	0.000	-0.073	0.000	-0.073	0.000	-0.074	0.000	-0.073	0.000
Argicultural/Veterinary			-0.190	0.000	-0.194	0.000	-0.060	0.000	-0.052	0.000	-0.036	0.000	-0.035	0.000	-0.037	0.000	-0.036	0.000
Social/Human			-0.146	0.000	-0.150	0.000	-0.077	0.000	-0.064	0.000	-0.070	0.000	-0.070	0.000	-0.069	0.000	-0.070	0.000
Business/Law			-0.095	0.000	-0.097	0.000	0.013	0.000	-0.001	0.818	-0.006	0.113	-0.006	0.132	-0.006	0.083	-0.006	0.116
Humanities			-0.331	0.000	-0.338	0.000	-0.167	0.000	-0.148	0.000	-0.161	0.000	-0.160	0.000	-0.162	0.000	-0.160	0.000
Combined			-0.207	0.000	-0.213	0.000	-0.067	0.000	-0.070	0.000	-0.080	0.000	-0.079	0.000	-0.081	0.000	-0.079	0.000
Institution Category																		
Ancient[a]																		
Red Brick			-0.063	0.000	-0.073	0.000	-0.046	0.000	-0.037	0.000	-0.029	0.000	-0.029	0.000	-0.028	0.000	-0.028	0.000
Plate Glass			-0.087	0.000	-0.097	0.000	-0.065	0.000	-0.058	0.000	-0.044	0.000	-0.044	0.000	-0.044	0.000	-0.044	0.000
Post 1992			-0.152	0.000	-0.158	0.000	-0.100	0.000	-0.084	0.000	-0.066	0.000	-0.066	0.000	-0.066	0.000	-0.066	0.000
Recent University			-0.151	0.000	-0.165	0.000	-0.116	0.000	-0.089	0.000	-0.072	0.000	-0.072	0.000	-0.072	0.000	-0.072	0.000
Other			-0.146	0.000	-0.160	0.000	-0.112	0.000	-0.086	0.000	-0.080	0.000	-0.080	0.000	-0.080	0.000	-0.080	0.000
Age																		
17 years and under			-0.136	0.066	-0.113	0.125	0.017	0.799	-0.010	0.875	-0.013	0.834	-0.013	0.834	-0.013	0.832	-0.013	0.835
18-20 years[a]																		
21-24 years			0.113	0.000	0.111	0.000	0.052	0.000	0.049	0.000	0.048	0.000	0.047	0.000	0.048	0.000	0.047	0.000
25-29 years			0.243	0.000	0.239	0.000	0.147	0.000	0.138	0.000	0.135	0.000	0.135	0.000	0.135	0.000	0.135	0.000
30 years and over			0.388	0.000	0.384	0.000	0.275	0.000	0.261	0.000	0.262	0.000	0.262	0.000	0.263	0.000	0.262	0.000
Age unknown			0.703	0.000	0.689	0.000	0.508	0.000	0.532	0.000	0.555	0.000	0.555	0.000	0.554	0.000	0.555	0.000
Level of Study																		
Post-Graduate[a]																		

Model	M1		M2		M3		M4		M5		M6		M7		M8 (Final Model)		M9	
Variables	**Coef.(β)**	**pval**	**Coef.(β)**	**pval**	**Coef.(β)**	**pval**	**Coef.(β)**	**pval**	**Coef.(β)**	**pval**	**Coef.(β)**	**pval**	**Coef.(β)**	**pval**	**Coef.(β)**	**pval**	**Coef.(β)**	**pval**
Under-Graduate			-0.200	0.000	-0.196	0.000	-0.126	0.000	-0.146	0.000	-0.139	0.000	-0.139	0.000	-0.139	0.000	-0.139	0.000
Domicile																		
South[a]																		
North					-0.065	0.000	-0.059	0.000	-0.062	0.000	-0.002	0.415	-0.002	0.445	-0.002	0.405	-0.002	0.420
SIC																		
Managers, Directors and Senior Officials[a]																		
Professional occupations							-0.090	0.000	-0.101	0.000	-0.102	0.000	-0.102	0.000	-0.102	0.000	-0.102	0.000
Associate professional and tech. occ…							-0.203	0.000	-0.227	0.000	-0.233	0.000	-0.233	0.000	-0.232	0.000	-0.233	0.000
Administrative and secretarial occup…							-0.401	0.000	-0.432	0.000	-0.428	0.000	-0.428	0.000	-0.428	0.000	-0.428	0.000
Skilled trades occupations							-0.382	0.000	-0.351	0.000	-0.343	0.000	-0.343	0.000	-0.343	0.000	-0.343	0.000
Personal service occupations							-0.578	0.000	-0.563	0.000	-0.559	0.000	-0.559	0.000	-0.558	0.000	-0.559	0.000
Sales and customer service occupations							-0.639	0.000	-0.594	0.000	-0.584	0.000	-0.584	0.000	-0.584	0.000	-0.584	0.000
Process, plant and machine operatives							-0.496	0.000	-0.528	0.000	-0.519	0.000	-0.519	0.000	-0.519	0.000	-0.519	0.000
Elementary occupations							-0.710	0.000	-0.621	0.000	-0.611	0.000	-0.611	0.000	-0.610	0.000	-0.611	0.000
SOC																		
Agriculture, forestry and fishing[a]																		
Mining and quarrying									0.443	0.000	0.426	0.000	0.425	0.000	0.425	0.000	0.426	0.000
Manufacturing									0.215	0.000	0.213	0.000	0.213	0.000	0.212	0.000	0.213	0.000
Electricity, gas, steam…									0.307	0.000	0.305	0.000	0.304	0.000	0.304	0.000	0.305	0.000
Water supply, sewerage, waste…									0.244	0.000	0.244	0.000	0.243	0.000	0.241	0.000	0.244	0.000
Construction									0.190	0.000	0.170	0.000	0.169	0.000	0.168	0.000	0.170	0.000
Wholesale and retail trade; repair…									-0.014	0.440	-0.032	0.074	-0.032	0.071	-0.033	0.065	-0.032	0.074
Transport and storage									0.216	0.000	0.188	0.000	0.188	0.000	0.187	0.000	0.188	0.000
Accommodation and food …									-0.057	0.002	-0.071	0.000	-0.071	0.000	-0.072	0.000	-0.071	0.000
Information and communication									0.130	0.000	0.095	0.000	0.095	0.000	0.095	0.000	0.095	0.000
Financial and insurance activities									0.260	0.000	0.233	0.000	0.232	0.000	0.233	0.000	0.233	0.000
Real estate activities									0.087	0.000	0.058	0.003	0.058	0.003	0.057	0.003	0.058	0.003
Professional, scientific and tech…									0.096	0.000	0.068	0.000	0.068	0.000	0.067	0.000	0.068	0.000
Administrative and support serv…									0.085	0.000	0.062	0.001	0.061	0.001	0.061	0.001	0.062	0.001
Public administration and def…									0.183	0.000	0.174	0.000	0.173	0.000	0.172	0.000	0.173	0.000
Education									0.027	0.133	0.017	0.356	0.016	0.366	0.016	0.383	0.017	0.355
Human health and social work activities									0.102	0.000	0.090	0.000	0.090	0.000	0.090	0.000	0.090	0.000
Arts, entertainment and recreation									-0.072	0.000	-0.095	0.000	-0.095	0.000	-0.096	0.000	-0.095	0.000
Other service activities									-0.022	0.256	-0.042	0.031	-0.042	0.031	-0.044	0.025	-0.042	0.031
Activities of households as employ…									0.123	0.001	0.075	0.037	0.074	0.039	0.074	0.038	0.075	0.036
Activities of extraterritorial…									0.255	0.000	0.246	0.000	0.247	0.000	0.240	0.000	0.246	0.000
Place of Employment																		
North East											-0.156	0.000	-0.157	0.000	-0.156	0.000	-0.156	0.000
North West											-0.172	0.000	-0.172	0.000	-0.172	0.000	-0.172	0.000
Yorkshire and the Humber											-0.165	0.000	-0.166	0.000	-0.166	0.000	-0.165	0.000
East Midlands											-0.166	0.000	-0.166	0.000	-0.165	0.000	-0.165	0.000
West Midlands											-0.175	0.000	-0.175	0.000	-0.175	0.000	-0.175	0.000
East of England											-0.133	0.000	-0.133	0.000	-0.132	0.000	-0.133	0.000
London[a]																		
South East											-0.123	0.000	-0.123	0.000	-0.123	0.000	-0.123	0.000
South West											-0.177	0.000	-0.177	0.000	-0.177	0.000	-0.177	0.000
Northern Ireland											-0.238	0.000	-0.239	0.000	-0.238	0.000	-0.238	0.000

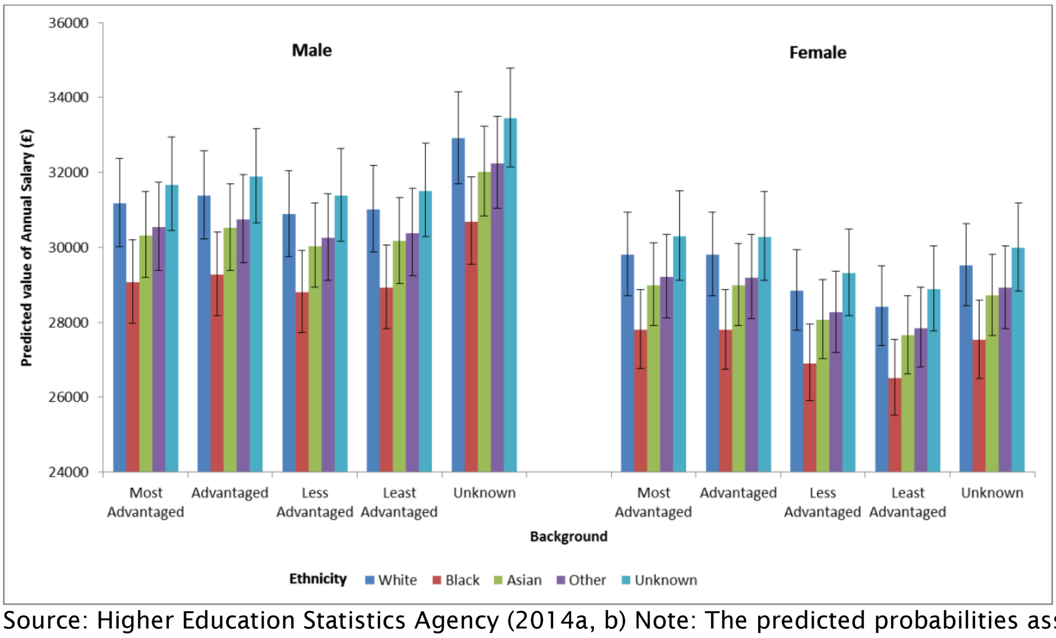
Model	M1		M2		M3		M4		M5		M6		M7		M8 (Final Model)		M9	
Variables	**Coef.(β)**	**pval**	**Coef.(β)**	**pval**	**Coef.(β)**	**pval**	**Coef.(β)**	**pval**	**Coef.(β)**	**pval**	**Coef.(β)**	**pval**	**Coef.(β)**	**pval**	**Coef.(β)**	**pval**	**Coef.(β)**	**pval**
Scotland											-0.158	0.000	-0.158	0.000	-0.158	0.000	-0.158	0.000
Wales											-0.212	0.000	-0.212	0.000	-0.212	0.000	-0.212	0.000
Non-UK											-0.138	0.000	-0.139	0.000	-0.139	0.000	-0.138	0.000
Interaction: Ethnicity*Background																		
White*Most Advantaged[a]																		
Black*Advantaged													-0.032	0.051				
Black*Less Advantaged													-0.018	0.211				
Black*Least Advantaged													-0.044	0.002				
Black*Unknown													-0.041	0.004				
Asian*Advantaged													-0.033	0.007				
Asian*Less Advantaged													-0.036	0.002				
Asian*Least Advantaged													-0.064	0.000				
Asian*Unknown													-0.038	0.001				
Other*Advantaged													-0.006	0.708				
Other*Less Advantaged													-0.017	0.249				
Other*Least Advantaged													-0.026	0.092				
Other*Unknown													-0.045	0.003				
Unknown*Advantaged													0.014	0.683				
Unknown*Less.Adv													0.026	0.393				
Unknown*Least.Adv													-0.001	0.969				
Unknown*Unknown													0.025	0.355				
Interaction: Background*Gender																		
Most Advantaged*Male[a]																		
Advantaged*Female															-0.007	0.221		
Less Advantaged*Female															-0.023	0.000		
Least Advantaged*Female															-0.042	0.000		
Unknown*Female															-0.064	0.000		
Ethnicity*Gender																		
White*Male[a]																		
Black*Female																	0.013	0.134
Asian*Female																	-0.017	0.007
Other*Female																	0.011	0.250
Unknown*Female																	-0.053	0.000
R-Squared	0.010		0.265		0.270		0.429		0.462		0.478		0.478		0.478		0.478	
BIC	1662.516		52125.995		53265.605		94960.546		104645.915		109318.847		109187.364		109476.028		109299.369	
Difference in BIC	-		50463.479		1139.611		41694.941		9685.369		4672.932		-131.483		157.181		-19.478	
Evidence	-		V.V. Strong		V.V. Strong		V.V. Strong		V.V. Strong		V.V. Strong		V.V. Strong Negative (with M6)		V.V. Strong (with M6)		V.V. Strong Negative (with M6)	

Notation “0.000” refers to P-Values smaller the 5 x 10^{-4}

From evaluating the output from the final chosen multiple regression model (model 8) of the covariates in the dataset as shown in Table 5-11, it is clear that there are significant associations with first salary after graduation, which supports some of the patterns observed when analysing the basic descriptive statistics. The migration variable is significant in predicting first salary and shows that those migrating had a higher predicted salary than those did not. From observing how the coefficients changed across the different models, as shown in Table 5-12, it can be seen that the migration variable had a positive impact on predicted salary for all models expect model 1. When covariates are added into the regression model, the negative impact of migration on future salary, changes to a positive trend. Therefore, the inclusion of one of the basic characteristic variables that are added in model 2 can be attributable to this conditional change in the relationship between migration and first salary.

The predicted salaries are calculated by substituting the coefficients in Table 5-11 into the regression Equation 5.2 and then the exponential of the coefficients were taken to take into account the log transformation. These predicted values of salary have been calculated and explained below. The predicted probabilities and 95% confidence intervals by ethnicity, background and gender are shown in Figure 5-14.

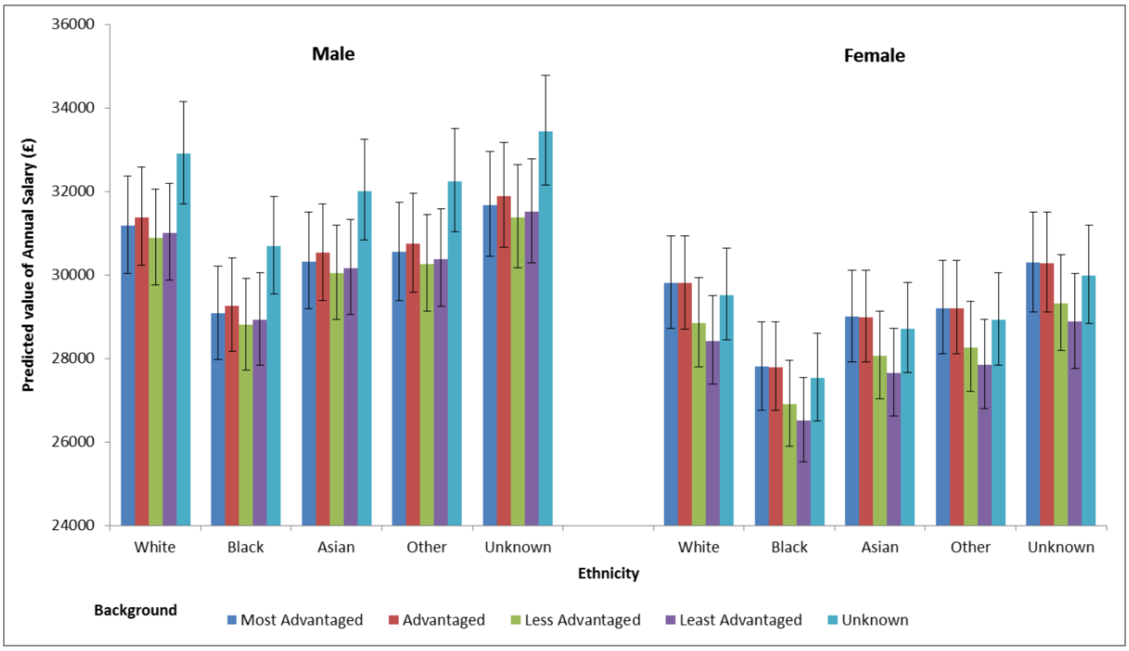
Figure 5-14: Predicted value of Annual Salary (£) six months after graduation by Gender, Ethnicity and Social Background



Source: Higher Education Statistics Agency (2014a, b) Note: The predicted probabilities assume the remaining variables were set to the reference category.

There is an observable difference in salary as a result of background, ethnicity and gender, with graduates from the black ethnic group having the lowest predicted salary for all social backgrounds and both genders. The groups with the highest salaries are those where ethnicity and social background are unknown; however these represent very small proportions of the sample. When the unknowns are ignored, male Asian graduates have the highest predicted salaries. However, the significant amount of overlap between the 95% confidence intervals indicate that these predicted differences between ethnic groups in first salary were not statistically different when all other covariates in the model were controlled for. A similar finding is found when focusing on social background as shown in Figure 5-15, where although there appears to be differences as a result of social background in the predicted first salary, these finding are not statistically significant at the 5% level.

Figure 5-15: Predicted value of Annual Salary (£) six months after graduation by Gender, Social Background and Ethnicity

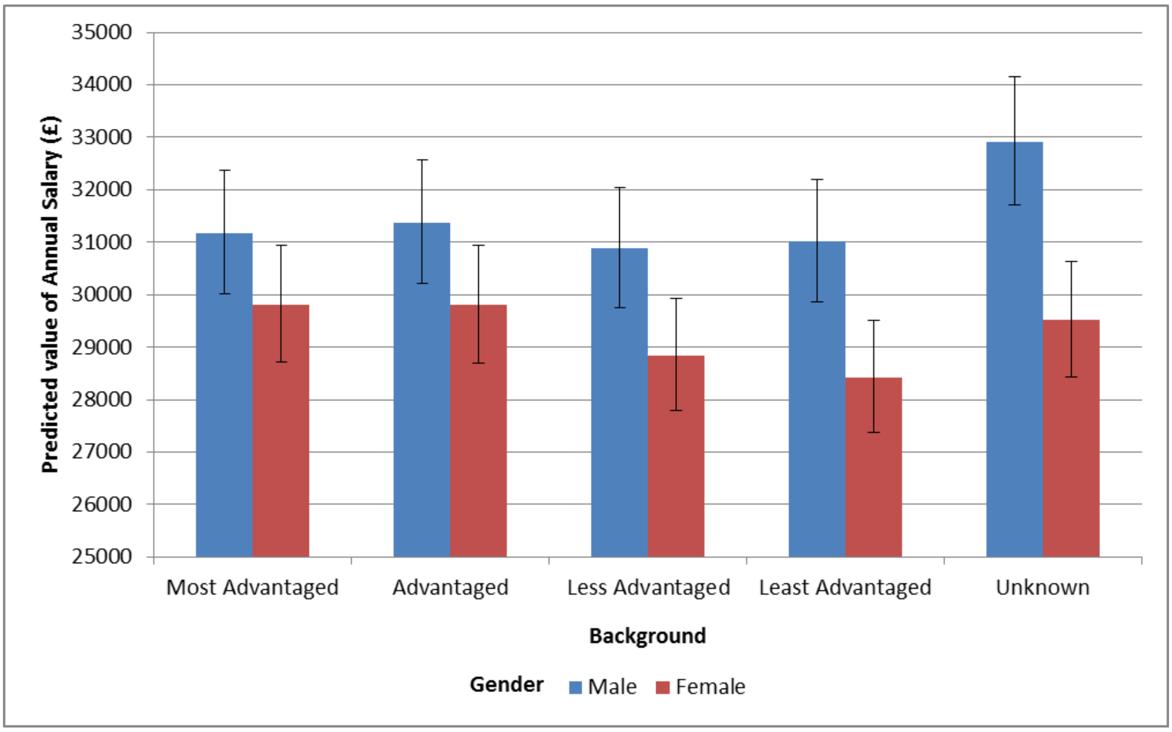


Source: Higher Education Statistics Agency (2014a, b)

Note: The predicted probabilities assume the remaining variables were set to the reference category.

In the descriptive statistics there is a substantial difference in mean first salary by gender, with males having higher salaries than females. Looking at the gender gap from the regression output of model 5, the significant interaction with social background means the findings need to be interpreted simultaneously. Males have consistently higher predicted mean first salaries than females across all ethnic and background groups, as Figure 5-14 and Figure 5-15 show. However, are these gender gaps still significant? Due to the interaction between social background and gender, the significance of the gender gap in mean first wage is determined by the social background variable, as shown in Figure 5-16. Across all social background classes males have a higher mean first wage than females, however, as outlined in more detail below, this difference is only significant at the 5% level for those in the least advantaged and unknown social background categories.

Figure 5-16: Predicted value of Annual Salary (£) six months after graduation by Gender and Social Background



Source: Higher Education Statistics Agency (2014a, b)

Note: The predicted probabilities assume the remaining variables were set to the reference category.

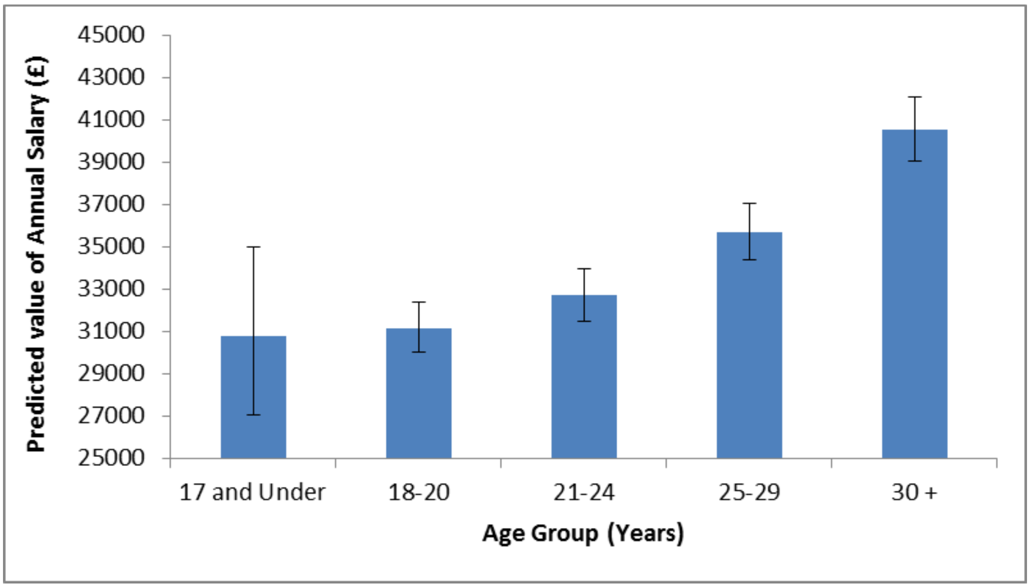
When holding all other variables in the model at the reference category, the predicted first salary six months after graduating is £1365.96 a year higher for males than females in the most advantaged social background category, however, this difference was not significant at the 5% level. While the gender gap in the least advantaged social group was £2588.24 higher for males and was significant at the 5% level.

This shows that, by simply controlling for gender, the observed gender gap in wages can be explained by other variables, such as gender difference in the type of industry employed within or the type of courses studied or qualifications gained and, in particular, social background. On further investigation, it appears that when all these extra variables are controlled for in the regression model, the gender gap in mean first salary is no longer significant at the 5% level for all social backgrounds, with the exception of those in the least advantaged and unknown groups. This indicates that gender differences in graduate earnings are more often a by-product of the different choices of jobs and degrees and not just a gender issue. The exception being those graduates from less advantageous social background, when it appears that males do still earn significantly more than females – demonstrating that there is still an issue regarding gender equality for all which should be addressed.

The predicted first salaries by age are shown in Figure 5-17, there is an increase in salary as age increases and these differences are statistically significant. The predicted first salaries by institution category are shown in Figure 5-18. Predicted salary is highest for those in the highest reputable universities and declines with each category. This was in contrast to the findings with employment status. Therefore, those students from the most reputable institutions were more likely to be unemployed, but if they were unemployed they were found to be earning significantly higher wages. It may be that graduates from the higher reputable institutions have higher salary expectations and are therefore applying for higher paid jobs which are harder to be successful in application, this may explain the higher wages of those employed but the higher probability of being unemployed.

The value of gaining a higher educational degree in the UK:

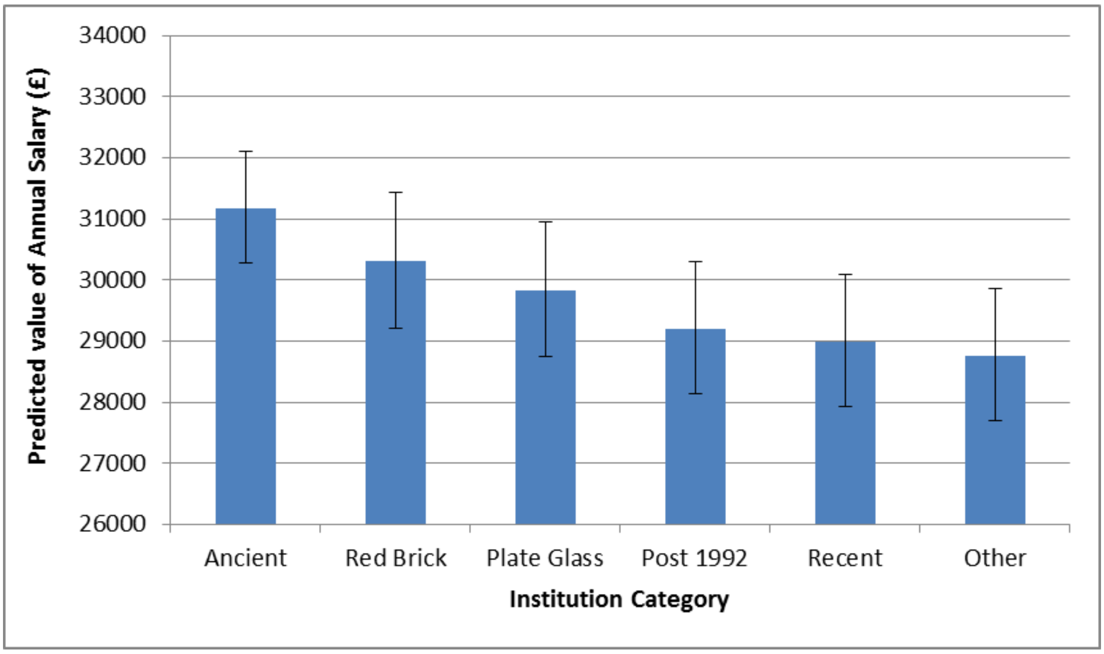
Figure 5-17: Predicted value of Annual Salary (£) six months after graduation by Age



Source: Higher Education Statistics Agency (2014a, b)

Note: The predicted probabilities assume the remaining variables were set to the reference category.

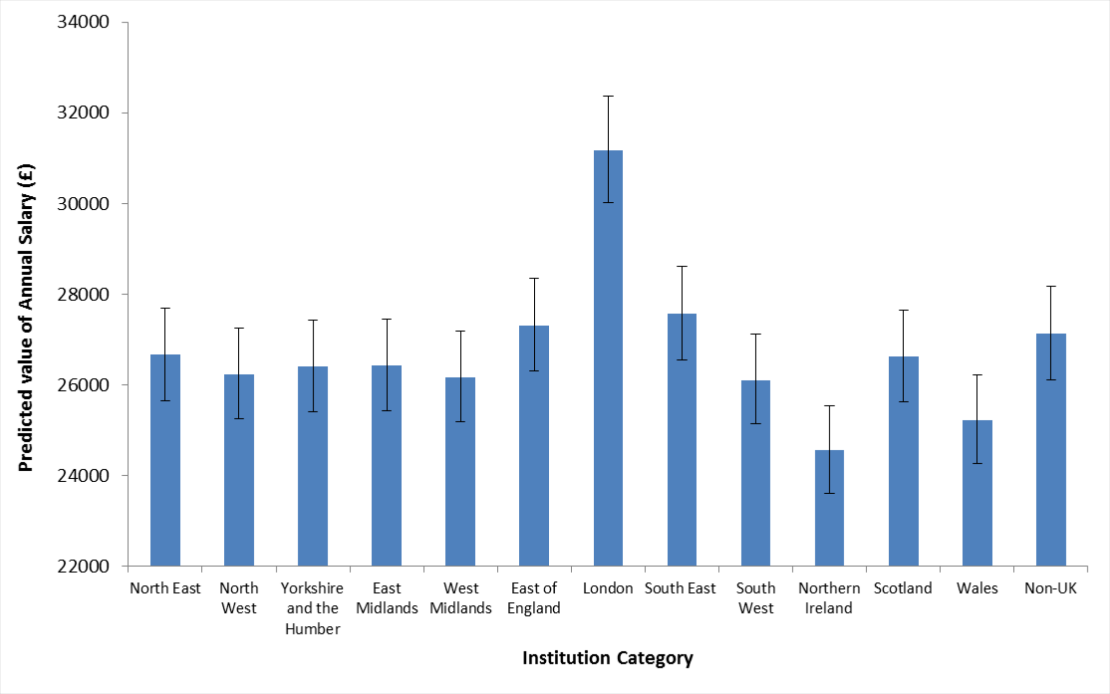
Figure 5-18: Predicted value of Annual Salary (£) six months after graduation by Institution Category



Source: Higher Education Statistics Agency (2014a, b)

Note: The predicted probabilities assume the remaining variables were set to the reference category.

Figure 5-19: Predicted value of Annual Salary (£) six months after graduation by Place of Employment



Source: Higher Education Statistics Agency (2014a, b)

Note: The predicted probabilities assume the remaining variables were set to the reference category.

The predictied mean first salary by place of employment is shown in Figure 5-19. This is a statistically signifcant variable in all of the models in Table 5-12 in which it was included and its inclusion also vastly improved the statistical measures of fit of the model in general. Overall the area of employment made no difference to mean first salary at the 5% level expect for London which had significantly higher wages for graduates.

It has been shown that the migration variable was signifcant in predicting first salary. From the regression model 5 the predicted difference in annual salary was £527.93 higher for a migrant compared to a non-migrant. This shows a completely opposite association between salary and migration that was shown when simply observing the difference in mean salary for migrants and non-migrants. However, as disscussed previously, regression modelling techniques do not accuraltey estimate the causual impact of migration on salary and as a result further analysis was required to determine to casual conterfactual impact on first salary for migrants compared to non-migrants.

PSM techniques are used here in order to evaluate the possible causal impact of migration on salary six months after graduation. The predicted difference in salary for those that migrated compared to those that did not have been calculated using multiple regression, PSM with Caliper Matching and PSM with Kernel matching.

These techniques were used on four different model specifications (as explained in Table 5-4) and the results of the three techniques and the four models are shown in Table 5-13. It can be seen from Table 5-13 that the size and significance level of the difference between migrants and non-migrants salaries is heavily determined by the choice of model – and therefore which student characteristics were controlled for and which individuals were matched upon) and the techniques used.

Table 5-13: The impact of migration on Salary – Regression and PSM results

Salary		Migrated	SE	Sig.
Model 1	**Multiple Regression**	1382.24	0.002	0.000
	PSM Caliper	716.95	579.69	0.216
	PSM Kernel	903.65	43.97	0.000
Model 2	**Multiple Regression**	994.89	0.002	0.000
	PSM Caliper	918.05	302.02	0.002
	PSM Kernel	615.08	49.16	0.000
Model 3	**Multiple Regression**	978.57	0.002	0.000
	PSM Caliper	908.48	302.29	0.003
	PSM Kernel	619.12	48.93	0.000
Model 4	**Multiple Regression**	709.00	0.002	0.000
	PSM Caliper	151.86	122.54	0.215
	PSM Kernel	256.81	59.42	0.000

Standard errors for PSM are Bootstrapped; 50bs for Caliper and 500bs for Kernel
Notation "0.000" refers to P-Values smaller the 5×10^{-4}

Standard Errors for Multiple Regression refer to the SE for log_salary whereas the estimate is the exponential of the coefficient multiplied by the constant minus the constant to show the difference in salary by a unit change in migration
The model outputs for the multiple regression models used here can be found in Table E-4

Source: Higher Education Statistics Agency (2014a, b)

When analysing model 1 (model 1 only models and matches graduates on their prior characteristics) the PSM with caliper results show no significant difference in mean first salary as a result of migration. The regression and PSM with kernel results, however, did show highly significant differences indicating that migrants were earning between £903-£1382 more than non-migrants six months after graduation depending on the method used. However, model 1 did not take into account the student's ability or attainment level. These variables have been shown in the previous results to significantly impact on several factors including the migration decision and the mean predicted salary and therefore should be included in the analysis in order to satisfy the CIA.

Model 2 was identical to model 1 but with the inclusion of institution category. Institution category is also being used as a proxy for attainment and it is therefore desirable to include institution category into the model in a best attempt to satisfy to CIA assumption of PSM. After modelling and matching graduates according to their prior characteristics and the institution attended the predicted difference in annual salary as a result of the migration decision dropped to between £615-£918. When using model 2 all three methods found significant differences in mean first salary between the migrant and non-migrant groups at the 1% level.

Model 3 included all variables in the dataset plus three interaction terms between; ethnicity and gender, social background and gender and ethnicity and social background. This model is classified as the full model and has been classified as so throughout the thesis. As a result of using all variables available the aim was that this model will eradicate as much of the bias as possible and satisfy the CIA. When using model 3 the predicted difference in salary between migrants and non-migrants is between £619-£908, these values were extremely similar to that of model 2. When using model 3 both PSM matching techniques and the regression modelling also found statistically significant results at the 1% level as was the case in model 3.

It was shown in the descriptive statistics and regression analysis that the location of employment had a significant impact on the mean first salary. Therefore, this variable needed to be added to the analysis in order to control for these differences and for the CIA to be satisfied. Due to the results of Model 2 and 3 being near identical it was decided that the place of employment variable be added to Model 2[7] to create Model 4. When using

[7] Decided not to add it to Model 3 because Model 3 contained complex interactions which impacted very little on the results (as shown in the near identical results between model 2 and 3)

model 4 the predicted difference in salary between migrants and non-migrants once prior characteristics, ability and attainment (through proxy) and location of employment had been controlled for was significantly smaller than found using the other models. The predicted difference was £256.81 when using the PSM Kernel method that was significant at the 1% level. The PSM caliper method did not produce a significant difference, while the regression analysis predicted a significant difference of £709. However, as mentioned previously due to the increased assumptions associated with regression analysis one needs to be careful when making conclusions of the causal impact of migration when using regression.

It can there be said that the PSM analysis indicates that there is a statistically significant difference in the first annual salary achieved by graduates six months after graduating. The size of this difference depends on the method of matching used and what variables to match graduates on. Using the result of model 4 it can be concluded that the casual impact of migrating to attend a HEI on first wages, when ability, attainment and location of employment have been controlled for, is only around £256.81 a year. Therefore, it can be said that the results regarding the impact of migrating in order to attend a HEI on first salary after graduating were marginal.

5.6 Chapter Summary

Migrating in order to attend a HEI in the UK is no longer the decision of choice for the majority of UK domiciled students. Large numbers of students are now deciding to study locally, commute or partake in distance learning. But do these students that do not migrate lose out in the future labour market compared to their peers that migrate? There has been little work to date that has analysed how this migration decision into HE impacts on the student's post graduate labour market outcomes.

This topic has become increasingly relevant in recent years as a result of the changing costs of HE in the UK. The rising cost of gaining a degree is increasingly burdening the individual student and there have also been large increases in the cost of living and travel. Not migrating to attend a HEI could reduce this financial burden on individual students, but at what cost to their future employability and earning potential? Therefore, an understanding of the pros and cons of migrating in order to study is becoming ever more important.

Attempting to address this need, this chapter uses a unique combination of HESA Student Record Data and Destinations of Leavers from Higher Education Survey to examine whether the student migration decision into HE does impact on the students future labour market outcomes. The chapter considers both employment status and salary six months after graduating. In particular, this work aimed to answer the following policy relevant question: what is the future economic value of migrating in order to attend a HEI?

Many interesting patterns were identified concerning both the future employment status and mean first salary of graduates. Students who attended an Ancient Institution -with the highest reputations - were found to be the most likely to be unemployed sixth months after graduating. However, students from these most reputable institutions were found to have the highest annual salaries.

Significant differences in employment status and salary by ethnicity were also found. Non-white students were significantly more likely to be unemployed than white graduates. However, of those graduates that were employed, ethnicity did not impact on salary.

Regarding mean first salary, there appeared to be a large gender wage gap favouring males. However, this gender driven wage gap was only significant for those in the less advantageous social background groups. No significant differences were observed in employment status or first annual salary by the student's social background. This suggests a student's social background was no longer influential by the time the student had graduated from HE.

To answer the question of the value of migrating, it was necessary to investigate whether the differences in labour market outcomes could be directly attributed to the migration outcome. In order to do this, a combination of regression analysis and PSM were conducted to attempt to estimate the counterfactual casual impact of migration on both employment status and first salary.

Employment status was found not to be impacted by migration, as no significant difference in unemployment between migrants or non-migrants was found. There was, however, a statistically significant impact of migration on first salary after graduating. The exact size of this difference depended on the model selected and matching technique used, however it was concluded that the size of this impact was in the region of £257 per year more for migrants compared to non-migrants.

Although this difference in salary in favour of migrants was statistically significant, can a difference in salary of £257 per year be described as a substantial difference? Yes, it can be said that migrants on average earn more than non-migrants six months after graduating, but does this represent value? Migrating in order to attend a HEI will incur higher costs for the student than studying and living locally, especially if non-migrant students remain in the parental home. This salary difference does not appear that large compared to the savings made by not migrating, however it is important to consider that all pay increases in the future will be based on an increment from this starting salary and could be larger in the future. However, these findings only isolate the impact of migration into HE six months after graduation and, therefore, no conclusion can be made from this analysis regarding how migration will impact on career lifetime earnings.

Limitations of this work include those regarding the data and the underlying assumptions of the regression and PSM methods. The dataset that was sourced for this analysis did not include all the desired variables that would have identified and observed all the possible factors that influence on the student migration decision, as set out in Figure 2-2, and the future labour market outcomes. Throughout this chapter the institution attended by the student was used as a proxy for the individual’s attainment as institutions differ regarding their entry requirements. In an ideal scenario, data on the individuals attainment levels prior to entering HE and the classification of degree achieved would have been used in the modelling and matching procedures. The lack of this data in the matching technique might have resulted in the observed difference in salary for migrants and non-migrants actually being attributable to differences in ability and attainment of the student that was not be captured using the proxy of institution category. As a result of this, it is possible that the CIA assumption of the regression and PSM method may be violated and this is a known limitation of these findings.

Possible extensions of this work include tackling the limitations aforementioned by obtaining linked data that includes an individual's school level attainment as well as the two datasets used throughout this chapter. Further analysis could also be conducted to extend the findings by conducting the PSM analysis by ethnicity and social background. The work could also follow the students up 3 years after graduating by using the HESA destination of leavers longitudinal study. This would enable the analysis of the more long-term impact of migrating on future labour market outcomes and acting as a sensitivity analysis for the models and techniques used in the current analysis.

To summarise, the most significant and policy-relevant findings within this chapter were threefold. First, the large significant difference in unemployment by ethnicity is concerning. For this to be evident even for those individuals with the highest levels of education should be a real policy concern for those involved in equality and equal access in the labour market for the UK. Second, gender inequality in first wages after graduation for those in the least advantageous social backgrounds should also be addressed by policy makers. Eradicating gender difference for those in the most advantaged groups is progress, but those in a less advantageous position should not be left behind.

Finally, the results regarding the impact of migrating in order to attend a HEI on the labour market outcomes after graduating were marginal. There appears to be little value in migrating to attend a HEI in the UK, however, due to limitations within the data it was difficult to estimate the true impact of the student migration decision on the labour market outcomes. With regards to employment, the migration transition experienced entering HE in the UK had no statistically significant impact regarding the individuals employment status after achieving a HE degree. Therefore, it can be concluded, the student migration decision does not impact on employability after graduating, so it cannot be stated whether migrating to study 'pays off' or not. With regards to first salary after graduating, the results have shown that there is a marginally small but statistically significant benefit to migrating. However, the overall economic benefits of migrating to attend a HEI were extremely limited. It therefore appears that migration does not provide any additional value with regards to future economic gain solely as a result of the migration decision. Due to the additional costs associated with migrating to study, it could be argued that students would actually benefit economically from studying at their local HEI if possible, however this analysis does not look at any of the potential non-economic benefits that could be gained from migrating to study at a HEI. As a result, it is important for schools, colleges and UK Government to convey the benefits of studying locally and migrating to potential new HEI students. The results show that the impact of the student migration decision are marginal, therefore, there are arguments in favour of both migrating and studying locally, both of which should be highlighted and clearly communicated to potential new students. However, organisations that are advising potential new HEI students must also consider other benefits of migrating to study beyond those that are solely economic.

6. The Student Migration Decision: Key Findings and Contributions

6.1 Introduction

The work in the thesis has addressed the aims set out in the introduction: to advance the current understanding of student migration in the United Kingdom. In completing this research there have been a number of specific achievements: creation of a new typology to accurately measure student migration decisions; production of a new student migration based area classification of local authority districts; application of a number of statistical models has shed new light on the factors impacting the student migration decision in the UK; and, for the first time, propensity score matching analysis has been used to try to evaluate the true value of student migration on future labour market outcomes after graduation.

This chapter concludes the thesis by summarising the main research findings and contributions. Section 6.2 addresses the specific objectives of each chapter as laid out in Chapter 1; Section 6.3 outlines some of the overall limitations of the thesis whilst Section 6.4 suggests possible further extensions to the research.

6.2 Summary of research, findings and contributions

The beginning of this thesis identified that despite the policy relevance, media interest and growing participation in the HE sector in the UK, very little work has been conducted to analyse the migration patterns of individuals entering HE in detail. It was shown through an extensive review of the literature that the student population in the UK has a significant impact on the areas in which they reside (Duke-Williams 2009) and that the level of this impact depends on the student migration experienced . However, studies of student migration were limited and, like the study of migration as a whole, the subject area is inherent with complexities in definitions, theories, measurement and data quality.

In previous studies, student migration was often measured and analysed without taking into account the true underlying complexity of the student migration process. As a result, a new typology of student migration was required to accurately measure the complexity of student migration. This typology needed to encompass all the possible movements a student could make when entering the HE system. Once this typology was created, it was used as the methodological base for the remainder of the analysis throughout this thesis. This is because the typology made it possible to accurately measure the student movements that were occurring across the UK.

This thesis has a clear underlying structure. First, a method that accurately measured student migration needed to be designed. Once this method was produced, it was used to investigate student migration patterns on the national scale. Then the focus switched from the national scale to identify how student migrations differed spatially across the UK. After analysing the spatial student migration patterns the focus then progressed to the individual level, to explore how a student's characteristics impacted on the student migration experienced. And finally the thesis analysed the economic impact of the student migration experience on the individual's future labour market outcomes.

The main findings and contributions of the thesis will now be discussed in turn.

The aim of Chapter 3 was to create a typology that best captured the complexity of student migration. This typology was then used to provide a detailed explanation of the student migration patterns for the UK as a whole and to identify if there were any spatial differences within the UK by creating a student migration areas classification. As a result two methodological contributions were provided within Chapter 3.

Researchers in the field of migration have long recognised the multiple issues and problems that beset the capture and accurate measurement of population mobility (Bell 2004). Measuring student migration - especially in the UK - is well known to be one of the hardest types of human mobility to accurately measure and record, primarily because of the way that official migration estimates are recorded (Office for National Statistics 2011c; Raymer et al. 2012). This is a result of a combination of definitional issues (what is defined as a migration for a student? Where should a student be reported as being resident?), poor quality of data (doctors registers) and low quality reporting from students themselves on their mobility (for example, not reporting to doctors when they migrate)(Lomax et al. 2011; Raymer et al. 2012).

However, the HESA Student Record Data provides locational information on every student enrolled in HE in the country. As a result this dataset provided the opportunity to capture and measure the migration outcomes of all students in the UK regardless of whether they had registered with their new doctor or whether they were picked up in the official migration estimates.

To truly understand the complexity of student migration, three locational variables from the HESA student record data (domicile, term-time and institution address) needed to be analysed simultaneously and, in order to do this, an innovative typology of student migration was proposed. The creation of the student migration typology provided the tool that accurately measured student migration throughout this thesis. This typology has the capability to be adapted, so it can be applied at varying levels of geography, and for any country that has the required data. The typology expanded on the work by Belfield and Morris (1999) which was the only past research to date to proposed any form of student migration typology. The typology proposed in chapter 3 extended the complexity with the typology proposed by Belfield and Morris (idem) and solely focus on the migration into HE as opposed to the migration into and then subsequently out of HE. The new typology has now enabled to truly quantitative and detailed examination and measurement of the factors impacting on the migration decisions of students when entering HE in the UK.

This typology was then used in the creation of a new student migration based area classification of local authority districts, which has never been produced for UK LAs before. The student migration area classification developed and progressed the migration area classification research conducted by Dennett and Stillwell (2011) as a result of solely focusing on student migration rather than all migration as a whole. The migration area classification produced by Dennett and Stillwell (idem) included one categorisation of a student migration area, this work expanded on this classification to create an area classification solely investigating student migration to provide added detail and a more complex understanding of the geographical difference in UK student migration decisions. The creation of this student area classification allowed for the detailed and complex analysis of the whole UK student migration system, which would not previously have been possible without the creation of the student migration typology. The student migration area classification output clusters closely aligned with the student migration groups within the typology, therefore providing statistical robustness to the creation of the student migration typology.

Using these two unique and innovative methodological outputs, two major substantive policy findings can be inferred from the results. Firstly, using the typology of student migration, it was identified that the previously assumed traditional transition to HE of migrating away from the parental home to study at a HEI in a far-away location (Chatterton 1999; Patiniotis and Holdsworth 2005; Chatterton 2010), was no longer the majority transition experienced by HE students in the UK in the 2010/11 academic year. Only 38% of all UK domiciled students followed the traditional pattern and migrated away from their domicile and resided close to the HEI they attended. The typology also indicated that a large percentage of UK domiciled students now commuted or were distance learning, as their term-time address was not in the same area as their HEI.

This raises questions over the academic debate regarding student migration and its impact on local areas, both areas that send student migrants and those that receive them. Duke-Williams (2009) discussed that students can have a profound effect on university towns and cities, as well as having a wider impact for acting as a process of relocating people around the country. However, if student migrants are no longer the dominant type of student, this raises questions on whether students still have such a profound effect in the UK. These findings can also be linked into the debates around studentification (Smith 2005, 2009), saturation in student populations (Duke-Williams 2009) and the prospect of destudentification (Kinton et al. 2014). Is the finding of student migrants no longer being the norm a response to the saturation of student populations and new students being unable to become student migrants? Is this move away from student migration a result of other factors such as increases in tuition fees and as a result is a conscious choice of the student but as a result has triggered this new observable trend of destudentification. This finding within chapter 3 expands the debate of the true impact of students in local areas within the UK and raises the question of whether there has been a shift away from the traditional decision of migrating away to study at a HEI and students becoming more static. If students are more static in general this will surely impact on the debates regarding studentification, destudentification and the overall geographies of student migration in the UK (Duke-Williams 2009; Smith 2009; Kinton et al. 2014).

Secondly, the student migration area classification highlighted a distinct difference between the North and South regions of the UK with regard to their student migration patterns. Students from the Southern Region migrated more than their Northern peers, who tended to be local students, commute or were distance learning instead. These findings overlap and extend the academic discussion around the economic North-South divide in the UK literature (Dorling 2007, 2011) and can be related to the differentials in living and housing costs between the two areas. Without the creation of the student migration typology or the student migration area classification it would not have been possible to identify these findings.

Chapter 4 set out to use the typology of student migration created in Chapter 3 but with the focus shifted to the individual level factors that impacted on the student migration outcome. The aim of Chapter 4 was to gain an in-depth understanding of how student migration transitions of people entering HEIs in the UK were correlated to the student's characteristics, the course they studied, and the institute they attended.

It was recognised that the migration choices of people entering into HE is of great policy interest to HEIs, as well as to government and non-government organisations. This is a result of the impact students make on the locations that they reside in, as well as other factors such as equality in access to HE, widening participation, increases in tuition fees and the change in the dominant migration behaviour of students as illustrated in Chapter 3.

This analysis was conducted using a number of well-known statistical techniques on a detailed dataset of population data that had not previously been analysed for this purpose. The three outcome variables used in the analysis were quite different in their format and as a result they required different methods to model the outcomes against the different explanatory variables simultaneously. The methods used were a logistic regression, a Tobit regression and a multinomial regression. Despite the wide range of different techniques used throughout the chapter, the overall substantive findings were the same for all three outcome variables and all methods used.

The main substantive contribution of Chapter 4 found that ethnicity, social background and gender all had a significant impact on the student migration experience in order to attend a HEI in the UK. The most consistent finding across the three techniques was that the group most likely to migrate, travel the furthest distances and be an internal student migrant were those students from the White ethnic group, most advantageous social background and were male. In contrast, the group of students least likely to migrate, travel the shortest distances and be local students were from the Asian ethnic group, least advantageous social background and were female. The use of three different techniques and the similar results provide statistical evidence that support these findings of ethnic, social background and gender differences in the migration outcomes of students, therefore demonstrating that access to HE was still not equal across the social spectrum in the United Kingdom.

These findings within Chapter 4 extend the pre-existing academic debate in a number of areas. The ethnic differences in student migration identified with this thesis supplement the work conducted by Finney and Simpson (2008b), Simpson and Finney (2009) and Finney (2011) in which ethnic differences in all migration flows were identified and discussed in detail. The work within this study highlights that these findings are mirrored when the focus of the analysis is solely on student migration. The research findings within this chapter also extend the debate regarding previous work which investigated the higher educational experiences of students from ethnic minority backgrounds (Patiniotis and Holdsworth 2005; Faggian and McCann 2008; Holdsworth 2009; McNay 2012; Khambhaita and Bhopal 2013), by highlighting that the migration experiences of students entering HE were not observed to be equal. Additionally the work within Chapter 4 show similar findings and emphasise the issues brought to light in the research presented by Gibbons and Vignoles (2012) which highlighted the sensitivity to distance migrated experienced by Bangladeshi and Pakistani women. The results of modelling distance presented in Chapter 4 also found that non-White women were predicated to migrate the shortest distances and this is inter-linked to the proximity of ethnic minority populations to inner-city HEI locations.

The aim of Chapter 5 was to examine the economic value of migrating for higher educational purposes in the UK. No previous research has estimated the impact of the student migration decision on the future economic outcomes of the students. As a result, no previous studies have been able identify the true economic value of migrating in order to attend a HEI.

A large number of policies have been designed and implemented to encourage equal access to HE across the UK. In Chapters 3 and 4 it has been shown that the student migration decision is influenced by many factors and that the student migration patterns differ spatially across the UK and as a result of individual level characteristics. Therefore, chapters 3 and 4 outlined that the student migration experience was not equal across the UK or across individuals. It is of great policy interest to see if there are any differences in future labour market outcomes of graduates as a result of these unequal differences in the migration decision when entering HE in the UK. This research enabled evaluation to find out whether differences in the student migration decision resulted in visible inequality later in the life course, or if inequality was eradicated at this end stage of the education system. The findings also enabled a definitive evaluation of whether or not it was economically beneficial for a student to migrate to attend a HEI or remain at home and study locally.

There were three significant and policy-relevant findings within Chapter 5. Large scale differences in unemployment by ethnicity were identified with Non-white students being significantly more likely to be unemployed than white graduates. For this to be evident within individuals with the highest levels of education should be a real policy concern for those involved in equality and equal access in the labour market for the UK. Secondly, the analysis identified gender inequality in first wages for those in the least advantageous social background groups where on average, women earned less than men. This should also be a real concern to policy makers and campaigners for gender equality in the workplace. These findings extend the debate raised from previous studies which found that ethnicity, gender and levels of human capital all impacted on sequential mobility trends by students and graduates (Faggian et al. 2006, 2007a, b; Faggian and McCann 2008, 2009; Pemberton et al. 2013), and the results from the research presented in Chapter 5 appear to confound these findings.

Finally, the value of migrating to attend a HEI in the UK was marginal. With regards to employment, the impact of the migration decision was not statistically significant. Migrating to attend a HEI did not improve or decline the probability of finding a job and so - from an employment perspective - it cannot be concluded whether migration pays off. With regards to first salary after graduating, the results showed that there was a small but statistically significant benefit of migrating to attend a HEI. However, the overall economic benefits were extremely marginal. It therefore appears that student migration does not significantly impact on future labour market outcomes. The results show that the student migration decision on its own does not provide any particular value with regards to future economic gain and students may benefit economically from studying at their local HEI if possible, because they won't have to factor in the costs of migrating in order to study. There are arguments in favour of both migrating and studying locally, both of which should be highlighted and clearly communicated to potential new students and this should be taken on board by policy makers, HEIs and any person involved in providing advice to people considering the HE options. As there was no prior literature that examined the explicit impact of the student migration decision into HE on the labour market outcomes after graduating, the research presented within this chapter would ideally prompt further analysis and research in the subject area with the aim to truly understand and measure the economic impact of the student migration decision.

6.3 Limitations and Critiques

This thesis has quantified and examined the factors impacting student migration in the United Kingdom and the impact this migration outcome has on the individual in later life. In spite of the innovative contributions provided within this thesis, there were some inevitable limitations within the study and issues with the methodology which, whilst noted in the relevant thesis chapters, are explicitly acknowledged here.

There were several general limitations that apply to the analysis throughout the thesis as a whole. Firstly, the thesis has provided a cross-sectional analysis of the student population at two time points, the academic year of 2010/11 for Chapter 3 and the academic year of 2011/12 for Chapters 4 and 5. This therefore only allowed for the analysis of the student migration process at one cross-sectional point in time. To gain a more in-depth understanding of how the student migration process has changed over time, the analysis would need to be extended to study the migration processes over a prolonged time period. This would add a longitudinal aspect to the study. However, due to limitations in the availability of the data source used within this study it was not possible to acquire enough of a time series of the dataset to avoid this limitation.

Another general comment on the study was its quantitative nature, which can be seen as both a strength and a weakness. The quantitative nature allowed for the whole student population to be analysed and for population level patterns and associations to be identified. However, the quantitative nature of the study meant it was only possible to examine the migration patterns. It was not possible to investigate in detail why these patterns and associations were occurring. Therefore it was impossible to add to the understanding of why these migration patterns were occurring or gain any information on the mechanisms driving the different student migration patterns. A more in depth understanding of the underlying mechanisms and factors driving these different migration choices by individuals could have been investigated with the addition of some follow up qualitative interviews, surveys or focus groups. However, the Student Population Data provided from HESA was anonymised and therefore it would not have been possible to identify individuals from the data to conduct follow up qualitative research. Any qualitative aspect of the research would have needed to be a standalone chapter and it would not have been possible to link any of the quantitative results to any of the potential qualitative findings.

The final comment regarding the data used within the analysis is that, whilst the HESA student record data provided unique population level data, the level of detail within some of the variables and the lack of some variables in the dataset resulted in some limitations within the study. For example, there was a lack of detail in the socio-economic background and ethnicity variables which caused problems within Chapters 4 and 5. Issues regarding the assumptions of the methods used and the need to merge variables to gain data were challenge that had to be confronted throughout the thesis. Also, the dataset did not contain data on many of the variables set out in Figure 2-2, as such, it is important to remember – when interpreting the results throughout the thesis - that the non-observable variables in the methods may impact on the findings.

The main limitation regarding the available data variables was the lack of a variable that could be used to directly measure an individual's attainment level. As seen in Figure 2-2, the attainment level of the student plays a key role in the student migration process as it is directly linked with the admissions process as well as being closely linked with many individual (micro) level variables. Ideally, if it were possible, it would have been desirable to acquire a three-way linked dataset between the HESA Student Population Data, the HESA follow-up survey and the National Pupil Database. Within the thesis linked data between the two HESA sources was used in Chapter 5. However, it was not possible to obtain the three-way linked dataset from the data provider at the time the research was being conducted. If it had been possible this thesis could have included the individuals' attainment levels with the analysis conducted within Chapter 4, which would have added an extra level of robustness to the findings.

Another limitation as a result of the available data was the level of detail in the term-time address variable. It was not possible to derive from this variable what type of accommodation the student was living in. For example, if the student was still in the parental home, university provided accommodation, private renting etc. This extra information within the variable would have provided much greater detail within the student migration typology. It would also have shed more light on the housing situations of students, especially those student that were 'local students', as in the current typology it is impossible to distinguish between those that study at the local university and stay in the parental home and those that study locally but move into their own accommodation. These two groups of students are very different in terms of the housing demands they place on an area and therefore this distinction would provide more useful information to policy makers and planners alike. A recommendation to HESA and the HEIs who collect this data would be to collect information on the type of accommodation the student is residing at the term-time address. This would make it possible to identify those students still residing at the parental home.

There were also more chapter-specific limitations. In Chapter 3, a student migration area classification was created and it must be noted that the process of building an area classification is fraught with problems as explained in detail by Dennett and Stillwell (2011). At each stage of the area classification process, careful decisions were made. However, often an argument could be made for an alternative decision to have been taken at many stages of the process. That said, the clustering algorithm used and the distance measure chosen were all made for the best reasons at the time of creation in the opinion of the author.

In Chapter 5, an important limitation of the work relates to the strong underlying assumptions of the regression and PSM techniques used within the analysis. This is again linked back to the lack of a student attainment variable as mentioned previously. It is therefore a known limitation of the work that the CIA assumption of the PSM may be violated.

6.4 Recommendations for Future Work

The research presented within this thesis has the potential to be enhanced and taken forward in a number of directions. Some of these potential extensions could result in the limitations discussed above being addressed, or further work could be done to support or critique the findings of this thesis using alternative data sources. Additionally, further work could build on the prior research by increasing the complexity of the current study. As per the limitations, the recommendations for future work can be split into more general recommendations associated to the whole thesis and more chapter specific extensions that have been mentioned previously but are highlighted again here.

The first general recommendation for further work would be to source a time series of data from HESA and conduct the analysis presented within this thesis over a longer time period. This would enable the analysis to observe if and how the student migration patterns observed in this thesis have changed over time and would also enable further investigation into the impact of certain higher educational polices and tuition fee increases.

The second recommendation would be to source Census 2011 data from the Office for National Statistics (2015) on students and migration. From the 2011 Census, students can be identified from the question about the individual's occupation and migration information is collected from the question about what the individual's address was one year ago. In addition, data is also collected on an individual's out of term address. By using this Census data, many extensions to the current study could be conducted. The work and findings in Chapter 3 could be validated - or critiqued - by using the census data to compare student migrant stocks and flows as reported using the HESA student record. Moreover, the analysis of student migrants' characteristics in Chapter 4 could be extended using census data because, the information collected on an individual in the census will be of much greater detail (more variables and variables in greater detail) than those provided in the HESA student record data.

The analysis in this thesis focused on the migration decision of students for UK HEIs. A further extension to this research would be to conduct an international comparison to see how the migration patterns of students differ across the globe. Many arguments can be made for making cross-national comparisons of migration analysis (Bell et al. 2002; Bell and Muhidin 2009). Measures for individual countries become more meaningful when placed in a comparative context and much more can be learnt from comparisons around the links between migration research and public policy (Bell et al. 2002; Bell and Muhidin 2009). Therefore, extending this project to analyse student migration across the globe would be a logical and worthwhile progression.

Another possible extension would be to conduct some qualitative research as an extension to the work in Chapters 3 and 4. This would provide more information as to why the patterns that were observed in the quantitative analysis were occurring. In turn, this would provide more details and possible solutions to policy makers within the higher educational sector.

Finally, throughout Chapters 4 and 5 a clear imitation to the research was the inability – because of the data available - to measure a student's attainment level prior to entering HE. To address this, the work in Chapters 4 and 5 could be extended by sourcing a linked dataset that connects the school pupil census or national pupil data, the HESA student record and the DLHE. At the time of writing, there were discussions in place to make this three-way linked dataset available in the near future via the Administrative Data Research Centre England (ADRC-E). If this linked data does become available, and this extension takes place, it would significantly improve the robustness of the analysis presented within this thesis.

6.5 Concluding Remark

This thesis has explored and enhanced the research of student migration in the United Kingdom at a time where there is a continuing need to increase understanding in the higher educational section. There is of course research still to be conducted, new datasets to explore and techniques to be improved and further study to be done as new patterns of student migration evolve over time. However, it is hoped that the novel contributions and substantive findings provided within this thesis will enable researchers wishing to continue studying student migration, in the UK or internationally, to do so. Researchers can use this thesis to inform and guide their future work and can take advantage of the new tools created to better understand the complexity of student migration.

List of References

Agresti, A. 2013. Categorical Data Analysis. 3rd Edition. New Jersey: John Wiley & Sons.

Akaike, H. 1973. "Information theory and an extension of the maximum likelihood principle." In Second Internaitonal Symposium on Information Theory, eds. B. Petrov and F. Csaki. Budapest: Akademiai Kiado.

Allinson, J. 2006. "Over-educated, over-exuberant and over here? The impact of students on cities." Planning Practice & Research 21(1):79-94.

Angrist, J.D. and J. Pischke. 2009. Mostly Harmless Econometrics. Oxford: Princeton University Press.

Arango, J. 2000. "Explaining migration: a critical view." International Social Science Journal 52(3):283-+.

Bachan, R. 2014. "Students' expectations of debt in UK higher education." Studies in Higher Education 39(5):848-873.

Bailey, S., J. Charlton, G. Dollamore and J. Fitzpatrick. 2000. "Families, groups and clusters of local and health authorities: revised for authorities in 1999." Population Trends 99:37-52.

Ball, S.J., M. Maguire and S. Macrae. 2000. Choice, pathways and transitions post-16: New youth, new economies in the global city. London: Falmer.

Ball, S.J., D. Reay and M. David. 2002. "'Ethnic Choosing': minority ethnic students, social class and higher education choice." Race, Ethnicity and Education 5(4):333-357.

Baum, C. 2013. "Propensity Score Matching, Regression Discontinuity and Limited Dependent Variables." In Applied Econometrics: Boston College.

Becker, G.S. 1993. Human capital: A theoretical and empirical analysis, with a special reference to education. Chicago: Univeristy of Chicago Press.

Beech, S.E. 2014a. "International student mobility: The role of social networks." Social and Cultural Geography.

Beech, S.E. 2014b. "Why place matters: Imaginative geography and international student mobility." Area 46(2):170-177.

Belfield, C and Z. Morris. 1999. "Regional Migration to and from Higher Education Institutions: Scale, Determinants and Outcomes." Higher Educational Quarterly 53(3):240-263.

Bell, M. 2004. "Measuring Temporary Mobility: Dimensions and Issues." St Lucia: Queensland Centre for Population Research, School of Geography, Planning and Architecture, The University of Queensland.

Bell, M., M. Blake, P. Boyle, O. Duke-Williams, P. Rees, J. Stillwell and G. Hugo. 2002. "Cross-national comparison of internal migration: issues and measures."

Journal of the Royal Statistical Society Series a-Statistics in Society 165:435-464.

Bell, M. and S. Muhidin. 2009. "Cross-National Comparisons of Internal Migration." In Human Development Research Paper 3009/30. New York: United Nations.

Bell, M. and G. Ward. 2000. "Comparing temporary mobility with permanent migration." Tourism Geographies 2(1):87-107.

Benoit, K. 2011. "Linear Regression Models with Logarithmic Transformations." London School of Economics: http://www.kenbenoit.net/courses/ME104/logmodels2.pdf.

BIS. 2013a. "The Benefits of Higher Education Participation for Individuals and Society: key findings and reports "The Quadrants"." In BIS RESEARCH PAPER NO. 146. London: Department for Business, Innovation and Skills.

BIS. 2013b. "Participation rates in higher education: academic years 2006/2007 – 2011/2012 (provisional)." ed. Matthew Bollington. London: Department for Business, Innovation and Skills.

Black, D., J. Smith and A. Jefferey. 2004. "How robust is the evidence on the effects of college quality? Evidence from matching." Journal of Econometrics 121(1-2):99-124.

Blair, T. 1999. "Tony Blair's Speech in Full." http://news.bbc.co.uk/1/hi/uk_politics/460009.stm: BBC.

Blanden, J. and P. Gregg. 2004. "Family income and educational attainment: A review of approaches and evidence for Britain." Oxford Review of Economic Policy 20(2):245-263.

Blanden, J., K. Hansen and S. Machin. 2010. "The Economic Cost of Growing Up Poor: Estimating the GDP Loss Associated with Child Poverty." Fiscal Studies 31(3):38-73.

Blanden, J. and S. Machin. 2004. "Educational inequality and the expansion of UK higher education." Scottish Journal of Political Economy 51(2):230-249.

Blundell, R., L. Dearden and B. Sianesi. 2001. "Estimating the Returns to Education: Models, Methods and Results." London School of Economics.

Blundell, R., L. Dearden and B. Sianesi. 2005. "Evaluating the effect of education on earnings: models, methods and results from the National Child Development Survey." Journal of Royal Statistics Society A 168(3):473-512.

Borjas, G.J. 1989. "Economic theory and international migration." International Migration Review, Special Silver Anniversary Issue 23(3):457-485.

Bourdieu, P. 1977. "Cultural Reproduction and Social Reproduction." In Power and Ideology in Education, eds. J. Karabel and A.H. Halsey. New York: Oxford University Press.

Bourdieu, P. 1984. Distinction. London: Routledge and Kegan Paul.

Bourdieu, P. 1985. "The genesis of the concepts of Habitus and of Field." Sociocriticism 2:11-24.

Bourdieu, P. 1986. "The Forms of Capital." In Handbook of Theory and Research for the Sociology of Education, ed. J.G. Robinson. New York: Greenwood Press.

Bourdieu, P. and J.C. Passerson. 1977. Reproduction in education, society, and culture. Beverly Hills, CA: Sage.

Boyle, P. 2009. "Migration." In International Encyclopedia of Human Geography, eds. R. Kitchin and N. Thrift. Oxford: Elsevier.

Boyle, P., K. H. Halfacree and V. Robinson. 1998. Exploring Contemporary Migration. London: Longman.

Brantlinger, E. 2003. Dividing Classes: How the Middle Class Negotiates and Rationalises School Advantage. London: Routlege Falmer.

Brettell, C.B. and J.F. Hollifield. 2007. Migration Theory: Talking Across Disiplines. 2nd Edition. London: Routeledge.

Browne, J. 2010. "Securing a sustainable future for higher education: An independent review of higher education funding & student finance." http://www.bis.gov.uk/assets/biscore/corporate/docs/s/10-1208-securing-sustainable-higher-education-browne-report.pdf.

Brynin, M. 2012. "Individual Choice and Risk: The Case of Higher Education." Sociology 0(0):1-17.

Bukodi, E. and J.H. Goldthorpe. 2012. "Decomposing 'social origins': The effects of parents' class, status and educational on the educational attainment of their children." In Barnett Papers in Social Research: Department of Social Policy and Intervention - University of Oxford.

Cadwallader, M. 1992. Migration and Residential Mobility - Macro and Micro Approaches. Madisob: The University of Wisconsin Press.

Caliendo, M. and S. Kopeinig. 2008. "Some practical guidance for the implementation of Propensity Score Matching." Journal of Economic Surveys 22(1):31-72.

Calinski, T. and J. Harabasz. 1974. "A dendrite method for cluster analysis." Communications in Statistics 3:1-27.

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

Cameron, A.C. and P.K. Trivedi. 1998. Regression Analysis of Count Data. New York Cambridge University Press.

Carlson, S. 2013. "Becoming a Mobile Student – a Processual Perspective on German Degree Student Mobility." Population, Space and Place 19:168-180.

List of References

Castles, S. and M.J. Miller. 2009. The age of migration: International population movements in the modern world. 4th Edition. Basingstoke: Palgrave Macmillan.

Champion, T and M. Coombes. 2007. "Using the 2000 Census to study human capital movements affecting Britain's large cities: insights and issues." Journal of the Royal Statistical Society A 170:447-467.

Chatterton, P. 1999. "University students and city centres - the formation of exclusive geographies - The case of Bristol, UK." Geoforum 30(2):117-133.

Chatterton, P. 2000. "The cultural role of universities in the community: revisiting the university - community debate." Environment and Planning A 32:165-181.

Chatterton, P. 2010. "Commentary: the student city: an ongoing story of neoliberalism, gentrification, and commodification." Environment and Planning A 42(3):509-514.

Cheng, S. and J. Long. 2007. "Testing for IIA in the Multinomial Logit Model." Sociological Methods Research 35(4):583-600.

Chowdry, H., C. Crawford, L. Dearden, A. Goodman and A. Vignoles. 2008. "Widening participation in higher education: analysis using linked administrative data." Institute for Fiscal Studies.

Chowdry, H., C. Crawford, L. Dearden, A. Goodman and A. Vignoles. 2010. "Widening Participation in Higher Education: Analysis using Linked Adminstrative Data." In IFS Working Paper W10/04: Institute for Fiscal Studies.

Chowdry, H., C. Crawford, L. Dearden, A. Goodman and A. Vignoles. 2013. "Widening participation in higher education: analysis using linked administrative data." Journal of the Royal Statistical Society 176(2):431-457.

Christie, H. 2007. "Higher education and spatial (im)mobility: nontraditional students and living at home." Environment and Planning A 39(10):2445-2463.

Clayton, J., G. Crozier and D. Reay. 2009. "Home and away: risk, familiarty and the multiple geographies of the higher education experience." International Studies in Sociology of Education 19(3-4):157-174.

Cochran, W.G. 1939. "The Use of the Analysis of Variance in Enumeration by Sampling." Journal of the American Statistical Association 34(207):492-510.

Cohen, J., P. Cohen, S.G. West and L.S. Aiken. 2002. Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences (3rd ed.). London: Routledge.

Coleman, J.S. 1988. "Social capital in the creation of human capital." American Journal of Sociology 94:95-120.

Cotton, D., R. George and M. Joyner. 2013. "The gender and ethnicity attainment gap research project." In PedRIO paper no.2. Pedagogic Research Institute and Observatory, Plymouth University, Plymouth.

Courgeau, D. 1985. " Interaction between spatial mobility, family and career life cycle: A French survey." Sociological Review 1(2):139-162.

Crawford, C. and W. Jin. 2014. "Payback time? Student debt and loan repayments: what will the 2012 reforms mean for graduates?" In IFS Report R93. London: Institute of Fiscal Studies.

Crozier, G., D. Reay and J. Clayton. 2008. "Different strokes for different folks: Diverse students in diverse institutions – experiences of higher education." Research Papers in Higher Education 23:167-177.

DaVanzo, J. 1976. "Differences between Return and Non-return Migration: An econometric analysis." International Migration Review 10:13-27.

DaVanzo, J. 1983. "Repeat migration in the United States: Who moves back and who moves on?" Review of Economics and Statistics 65:552-559.

Dearden, L., L. McGranahan and B. Sianesi. 2004. "Returns to Education for the Marginal Learner: Evidence from the BCS70." CEE Discussion Papers: SLE.

Dearing, R. 1997. "Higher Education in the Learning Society." In Report of the National Committee of Inquiry into Higher Education. Norwich: HMSO.

Dehejia, R.H. and S. Wahba. 1999. "Causal effects in nonexperimental studies: reevaluating the evaluation of training programs." Journal of American Statistical Association 94(448):1053-1062.

Dennett, A. 2010. "Understanding Internal Migration in Britian at the start of the 21st Century." In School of Geography: The Univeristy of Leeds.

Dennett, A. and J. Stillwell. 2010. "Internal Migration Patterns by Age and Sex at the Start of the 21st Century." In Technologies for Migration and Commuting Analysis, eds. J. Stillwell, O. Duke-Williams and A. Dennett. New York: Business Science Reference.

Dennett, A. and J. Stillwell. 2011. "A new area classification for understanding internal migration in Britain." Population Trends 145:1-26.

Department for Education and Skills. 2003. "Widening Participation in Higher Education." London: DfES.

Department for Education and Skills. 2006. "Widening Participation in Higher Education." London: DfES.

Department for Work and Pensions. 2015. "Family Resources Survey - United Kindom 2013/14." https://www.gov.uk/government/collections/family-resources-survey--2.

Dorfman, A.H. and R. Valliant. 2005. Superpopulation Models in Survey Sampling.

Dorling, D. 2007. "The North-South Divide - Where is the line?". Scoail And Spatial Inequalities: The Univeristy of Sheffield.

Dorling, D. 2011. So you think you know about Britain? The surprising truth about modern Britain. London: Constable & Robinson Ltd.

Dorling, D., J. Rigby, B. Wheeler, D. Ballas, B. Thomas, E. Fahmy, D. Gordon and R. Lupton. 2007. "Poverty, wealth and place in Britain, 1968 to 2005." Bristol: The Policy Press.

Dorling, D. and B. Thomas. 2011. Bankrupt Britain: An atlas of social change. Bristol: The Policy Press.

Duke-Williams, O. 2009. "The geographies of student migration in the UK." Environment and Planning A 41(8):1826-1848.

Duke-Williams, O. 2010. "Mapping the geodemographic classifications of migrants' origins and destinations." Journal of Maps:360-369.

Equality Challenge Unit. 2014. "Equality in higher education: statistical report 2014." http://www.ecu.ac.uk/publications/equality-higher-education-statistical-report-2014/.

Everitt, B.S., S. Landau and M. Leese. 2001. Cluster Analysis. 4th Edition. London: Arnold.

Everitt, B.S., S. Landau, M. Leese and Stahl. D. 2011. Cluster Analysis. 5th Edition. Chichester: Wiley.

Faggian, A. and P. McCann. 2008. "Universities, agglomerations and graduate human capital mobility." Tijdschrift voor Economische Sociale Geografie 100:461-471.

Faggian, A. and P. McCann. 2009. "Human capital, graduate migration and innovation in British regions." Cambridge Journal of Economics 33:317-333.

Faggian, A., P. McCann and S. Sheppard. 2006. "An analysis of ethnic differences in UK graduate migration behaviour." Annals of Regional Science 40(2):461-471.

Faggian, A., P. McCann and S. Sheppard. 2007a. "Human capital, higher migration: An analysis of Scottish and Welsh students." Urban Studies 44(13):2511-2528.

Faggian, A., P. McCann and S. Sheppard. 2007b. "Some evidence that women are more mobile than men: Gender differences in UK graduate migration behaviour." Journal of Regional Science 47(3):517-539.

Field, J. 2003. "Getting real: evidence on access and achievement?" In Higher Education and the Life Course, eds. M. Slowey and D. Watson. Maidenhead: Open University Press.

Findlay, A. 2011. "An Assessment of Supply and Demand-side Theorizations of International Student Mobility." International Migration 49(2):162-190.

Finney, N. 2011. "Understanding ethnic differences in the migration of young adults within Britain from a lifecourse perspective." Transactions of the Institute of British Geographers 36(3):455-470.

Finney, N. and L. Simpson. 2008a. "Internal migration and ethnic groups: Evidence for Britain from the 2001 census." Population Space and Place 14(2):63-83.

Finney, N. and L. Simpson. 2008b. "Internal Migration and Ethnic Groups: Evidence for Britain from the 2001 Census." Population, Space and Place 14:63-83.

Fotheringham, A.S. and D.W.S. Wong. 1991. "The modifiable areal unit problem in multivariate statistical analysis." Environment and Planning A 23:1025-1044.

Galindo-Rueda, F., O. Marcenaro-Gutierrez and A. Vignoles. 2004. "The widening socio-economic gap in UK Higher Education." National Institute Economic Review 190.

Gastner, M.T. and M.E.J. Newman. 2004. "Diffusion-based methods for producing density equlizing maps." Proceedings of the National Academy of Sciences of the United States of America 101:7499-7504.

Gehlke, C.E. and K. Biehl. 1934. "Certain effects of grouping upon the size of the correlation coefficient in census tract material." Journal of American Statistical Association Supplement 29:169-170.

Gibbons, S. and A. Vignoles. 2012. "Geography, choice and participation in higher education in England." Regional Science and Urban Economics 42(1-2):98-113.

Goodman, A., P. Gregg, A. Chowdry, C. Crawford, L. Dearden, R. Joyce, L. SIbieta, K. Sylva and E. Washbrook. 2010. "Poorer children's educational attainment: How important are attitudes and behaviour?", eds. A. Goodman and P. Gregg. York: Joseph Rowntree Foundation.

Greene, W.H. 1993. Econometric Analysis. 2nd Edition. New York: Macmillan.

Gregg, P. and L. Macmillan. 2010. "Family income, education and cognitive ability in the next generation: exploring income gradients in education and test scores for current cohorts of youth." Longitudinal and Life Course Studies 1(3):259-280.

Grilli, L. and C. Rampichini. 2011. "Propensity scores for the estimation of average treatment effects in observational studies - Training lessons on causal inference." Dipartamento di Statistica “Giuseppe Parenti” Univeriti di Firenze.

Halsey, A.H., H. Lander, O. Brown and A.S. Wells. 1997. Education: Culture, Economy, Society. Oxford: Oxford University Press.

Harris, M. 2010. "What more can be done to widen access to highly selective universities? ." Bristol: Office for Fair Access.

Harris, P. and T. McVeigh. 2002. "Student takeover alarms cities." In The Observer.

Hartley, H.O. and R.L. Jr. Sielken. 1975. "A "Super-Populaiton Viewpoint" for Finite Population Sampling." Biometrics 31(2):411-422.

Heckman, J. 2005. "The Scientific Model of Causality." Sociological Methodology 35(1):1-97.

Heckman, J., H. Ichimura and P. Todd. 1997. "Matching as an econometric evaluation estimator: evidence from evaluating a job training programme." Review of Economic Studies 64(4):605-654.

Heckman, J., R. LaLonde and J. Smith. 1999. "The economics and econometrics of active labor market programs." In Handbook of Labour Economics, eds. O. Ashenfelter and D. Card. Amsterdam: Elsevier.

HEFCE. 2005. "Young participation in higher education." Bristol: Higher Education Funding Council for England.

HEFCE. 2013. "Widening Participation Policy." London: Higher Education Funding Council for England.

Heinrich, C., A. Maffioli and G. Vazquez. 2010. "A Primer for Applyign Propensity-Score Matching." Inter-American Development Bank: Impact-Evaluation Guidelines.

Higher Education Statistics Agency. 2012a. "Destinations of leavers 2011/12-2012/13: Definitional support." https://www.hesa.ac.uk/content/view/2889/.

Higher Education Statistics Agency. 2012b. "HESA Student Record Data - 2007/08 to 2010/11."

Higher Education Statistics Agency. 2012c. "Higher Education Statistics Authority - About Us - Overview."

Higher Education Statistics Agency. 2013. "DLHE Introduction 2011/12." https://www.hesa.ac.uk/intros/dlheintro1112.

Higher Education Statistics Agency. 2014a. "HESA Destination of Leavers survey 2011/12."

Higher Education Statistics Agency. 2014b. "HESA Student Record Data - 2011/12."

Hill, C. 2005. "A picture of the United Kingdom using the National Statstics Socio-economic Classification." Population Trends 125:7-14.

Hills, J., M. Brewer, S. Jenkins, R. Lister, R. Lupton, S. Machin, C. Mills, T. Modood, T. Rees and S. Riddell. 2010. "An anatomy of economic inequality in the UK: Report of the National Equality Panel." London: Government Equalities Office/Centre for Analysis of Social Exclusion.

HMSO. 1963. "Higher education: report of the Committee appointed by the Prime Minister under the Chairmanship of Lord Robbins 1961-63, Cmnd. 2154, London: HMSO."

HMSO. 1992. "Further and Higher Education Act." London.

HMSO. 1998. "Teaching and Higher Education Act." London.

HMSO. 2004. "Higher Education Act." London.

HMSO. 2011. "Education Act 2011."

HMSO. 2012. "Immigration policy - Overseas Students and Net Migration - Business, Innovation and Skills Committee." http://www.publications.parliament.uk/pa/cm201213/cmselect/cmbis/425/42505.htm#n32.

Hoare, A.G. and R. Johnston. 2011. "Widening participation through admissions policy - a British case study of school and university performance." Higher Education Quarterly 36:21-41.

Holdsworth, C. 2009. "'Going away to uni': mobility, modernity, and independence of English higher education students." Environment and Planning A 41:1849-1864.

Holdsworth, C. 2006. "Don't you think you're missing out, living at home? Student experiences and residential transitions." The Sociological Review 54:495-519.

Holland, P. 1986. "Statistics and Causal Inference." Journal of American Statistical Association 81(396):945-960.

Holton, M and M Riley. 2013. "Student Geographies: Exploring the Diverse Geographies of Students and Higher Education." Geography Compass 7:61-74.

Hubbard, P. 2008. "Regulating the social impacts of studentification: a Loughborough case study." Environment and Planning A 40(323-341).

Hubbard, P. 2009. "Geographies of studentifciation and purpose-built student accommodation: leading separate lives?" Environment and Planning A 41:609-629.

Imbens, G. 2004. "Nonparametric estimation of average treatment effects under exogeneity: a review." Review of Economics and Statistics 86(1):4-29.

Institute of Education. 2012. "Chance Of Attending University Still Influenced By Parents' Education." TO BE COMPLETED http://www.huffingtonpost.co.uk/2012/12/14/parents-still-influence-university-chance_n_2298821.html.

Jerrim, J. 2012. "The Socio-Economic Gradient in Teenagers' Reading Skills: How Does England Compare with Other Countries?" Fiscal Studies 33(2):159-184.

Jones, G. 2002. "TheYouth Divide: Dividing Paths to Adulthood." York: Joseph Rowntree Foundation.

Kerckhoff, A.C. and J. Macrae. 1992. "Leaving the parental home in Great Britain: a comparative perspective." The Sociological Quarterly 33:281-301.

Khambhaita, P. and K. Bhopal. 2013. "Home or away? The significance of ethnicity, class and attainment in the housing choices of female univeristy students." Race, Ethnicity and Education:1-32.

Khandker, S.B., G.B. Koolwal and H. Samad. 2009. Handbook on Impact Evaluation : Quantitative Methods and Practices. Washington D.C.: World Bank Publications.

King, R. 2002. "A New Map of European Migration " International Journal of Population Geography 8(2):89-106.

Kinton, C., D. Smith and J. Harrison. 2014. "Destudentification: A new expression of urban abandonment and decline." Loughborough University.

Lambert, D. 1992. "Zero-inflated Poission regression with an application to defects in manufactoring." Technometrics 34:1-14.

Leathwood, C. and P. O'Connell. 2003. "It's a struggle': the construction of the 'new student' in higher education." Journal of Education Policy 18(6):597-615.

Lechner, M. 2002. "Some practical issues in the evaluation of heterogenous labour market programmes by matching methods." Journal of the Royal Statistical Society, A 165(59-82).

Lee, E. 1966. "A theory of migration." Demography 3(1):47-57.

Leese, M. 2010. "Bridging the gap: supporting student transitions into higher education." Journal of Further and Higher Education 34(2):239-251.

Leuven, M. and B. Sianesi. 2003. PSMATCH2: Stata Module to Perform Full Mahalanobis and Propensity-Score Matching, Common Support Graphing, and Covariate Imbalance Testing. Newton, MA, United States: Boston College Department of Economics.

Li, M. and M. Bray. 2007. "Cross-border flows of students for higher education: Push–pull factors and motivations of mainland Chinese students in Hong Kong and Macau." Higher Education 53:791-818.

Lomax, N., J. Stillwell, P. Norman and P. Rees. 2011. "Data Sources and Estimation Methods for Sub-national Migration in the United Kingdom." Working paper 11/4, School of Geography, University of Leeds.

Long, J. 1997. Regression Models for Categorical and Limited Dependent Variables. Thousand Oaks, CA: Sage Publications.

Long, J. and J. Freese. 2006. Regression Models for Categorical Dependent Variables Using STATA. Texas: Stata Press.

Long, L. 1988. Migration and Residential Mobility in the United States. New York: Russel Sage Foundation.

Machin, S. and A. Vignoles. 2004. "Educational inequality: The widening socio-economic gap." Fiscal Studies 25(2):107-128.

Massey, D. S., J. Arango, G. Hugo, A. Kouaouci, A. Pellegrino and J. E. Taylor. 1993. "Theories of International Migration - a Review and Appraisal." Population and Development Review 19(3):431-466.

Massey, D.S., C.Z. Charles, G.F. Lundy and M.J. Fischer. 2003. The source of the river: The social origins of freshmen at America's selective colleges and universities. Princeton, NJ.: Princeton Univeristy Press.

Mayor for London. 2003. "Cost of Living Comparisons Between London, Edinburgh and Manchester." http://www.london.gov.uk/mayor/economic_unit/case_for_london/costofliving_report.pdf.

Mazzarol, T. and G.N. Soutar. 2002. "“Push-pull” factors influencing international student destination choice." International Journal of Educational Management 16(2):82-90.

McDonald, J.F. and R.A. Moffitt. 1980. "The Uses of Tobit Analysis." The Review of Economics and Statistics 62(2):318-321.

McDonough, P. 1997. Choosing colleges: How social class and schools structure opportunity. Albany, NY: SUNY Press.

McNay, I. 2012. "The new decade: a watershed in UK higher education?" Higher education review 45(1):5-29.

Micklewright, J., S.V. Schnepf and C. Skinner. 2011. "Non-response biases in surveys of school children: the case of the English PISA samples." Institute of Education, University of London, University of Southampton.

Modood, T. and M. Shiner. 1994. Ethnic Minorities and Higher Education: Why are There Differential Rates of Entry? London: Policy Studies Institute.

Morley, L. 2003. "Reconstructing students as consumers: power and assimilation?" In Higher Education and the Lifecourse, eds. M. Slowey and D. Watson. Maidenhead: Society for Research into Higher Education and Open University.

Mullahy, J. 1986. "Specification and testing of some modified count data models." Journal of Econometrics 33(341-365).

Munro, M. and M. Livingston. 2011. "Student Impacts on Urban Neighbourhoods: Policy Approaches, Discourses and Dilemmas." Urban Studies 49(8):1903-1923.

Munro, M., I. Turok and M. Livingston. 2009. "Students in cities: a preliminary analysis of their patterns and effects." Environment and Planning A 41(8):1805-1825.

Nicholson, L. and F. Wasoff. 1989. Students Experience of Private Rented Housing in Edinburgh. Edinburgh: University of Edinburgh

Noden, P., M. Shiner and T. Modood. 2014. "Black and Minority Ethnic Access to Higher Education: A Reassessment." In Nuffield Foundation: London School of Economics.

O'Leary, NC and PJ Sloane. 2005. "The returns to a university education in Great Britain." National Institute Economic Review 93(1):75-89.

OECD. 2013. "International Migration Outlook." OECD Publishing: http://dx.doi.org/10.1787.migr_outlook-2013-en.

OFFA. 2014. "Trends in young participation by student background and selectivity of institution." Bristol: Office for Fair Access.

Office for National Statistics. 2009a. "Final Population Definitions for the 2011 Census." 2011 Census Programme.

Office for National Statistics. 2009b. "UK Standard Industrial Classification 2007 (UK SIC 2007)." ed. Lindsay Prosser. http://www.ons.gov.uk/ons/guide-method/classifications/current-standard-classifications/standard-industrial-classification/index.html.

Office for National Statistics. 2010a. "Standard Occupational Classification 2010: Volume 1 Structure and descriptions of unit groups." http://www.ons.gov.uk/ons/guide-method/classifications/current-standard-classifications/soc2010/soc2010-volume-1-structure-and-descriptions-of-unit-groups/index.html.

Office for National Statistics. 2010b. "Standard Occupational Classification 2010: Volume 3 The National Statistics Socio-economic Classification: (Rebased on the SOC2010) User Manual." In Palgrave Macmillan. Titchfield: Office for National Statistics.

Office for National Statistics. 2011a. "2011 Area Classifications." Titchfield: Office for National Statistics.

Office for National Statistics. 2011b. "About National Statistics and ONS." http://www.statistics.gov.uk/geography/admin_geog.asp.

Office for National Statistics. 2011c. "Internal Migration Estimates - Methodology." Titchfield: Office for National Statistics.

Office for National Statistics. 2011d. "Mid-2010 Population Estimates: Selected age groups for local authorities in the United Kingdom; estimated resident population." Titchfield: Office for National Statistics.

Office for National Statistics. 2012. "Internal Migration by Local Authorities in England and Wales, Year Ending June 2011." Titchfield: Office for National Statistics.

Office for National Statistics. 2013a. "Internal Migration by Local Authority in England and Wales, Year Ending June 2012.". Titchfield: ONS.

Office for National Statistics. 2013b. "Neighbourhood Statistics." ed. ONS. Titchfield.

Office for National Statistics. 2015. "Welcome to the 2011 Census for England and Wales." http://www.ons.gov.uk/ons/guide-method/census/2011/index.html.

Office for National Statistics. 2016. "House Price Index, November 2015." Titchfield: Office for National Statistics.

Office for National Statistics, Northern Ireland Statisics and Research Agency and National Records of Scotland. 2011. "2011 Census: Ethnic group, local authorities in the United Kingdom."

Openshaw, S. 1984. Concepts and Techniques in Modern Geography, Number 38. The Modifiable Areal Unit Problem. Norwich: Geo Books.

Ozga, J. and L. Sukhnandan. 1998. "Undergraduate non-completion: developing an explanatory model." Higher Education Quarterly 52(3):316-333.

Patiniotis, J. and C. Holdsworth. 2005. "'Seize That Chance!' Leaving Home and Transitions to Higher Education." Journal of Youth Studies 8(1):81-95.

Paulsen, M.B. 2001. "The economics of human capital and investment in higher education." In The finance of higher education; Theory, research, policy & practice, eds. M.B. Paulsen and J.C. Smart. New York: Agathon Press.

Paulsen, M.B. and E.P. St John. 2002. "Social class and college costs: Examining the financial nexus between college choice and persistence." The Journal of Higher Education 73:189-239.

Payne, G. 2013. "Models of Contemporary Social Class: the Great British Class Survey." Methodological Innovations Online 8(1).

Pemberton, J., S. Jewell, A. Faggian and Z. King. 2013. "Higher education as a portfolio investment: students' choices about studying, term time employmentm leisrue and loans." Oxford Economic Papers 65:268-292.

Perna, L.W. 2000. "Differences in the decision to enroll in college among African-American, Hispanics, and Whites." Journal of Higher Education 71:117-141.

Perna, L.W. 2006. "Understanding the relationship between information about college prices and financial aid and students' college-related behaviors." American Behavioural Scientist 49:1620-1635.

Perna, L.W. and M.A. Titus. 2005. "The relationship between parental involvement as social capital and college enrollment: An examination of racial/ethnic group differences." Journal of Higher Education 76:485-518.

Pickett, K. and R. Wilkinson. 2010. The Spirit Level: Why Equality is Better for Everyone. London: Penguin.

Pike, A., D. MacKinnon, M. Coombes, T. Champion, D. E. Bradley, A. Cumbers, L. Robson and C. Wymer. 2016. "Uneven growth: tackling city decline." In Inspiring Social Change: Josheph Rowntree Foundaiton.

Platt, L. 2011. "Inequality within Ethnic Groups." Joseph Rowntree Foundation Paper: Poverty and Ethnicity.

Pollard, A. 2003. "Learning through life - higher education and the lifecourse of individuals." In Higher Education and the Lifecourse, eds. M. Slowey and D. Watson. Maidenhead: The Society fro Research into Higher Education and The Open Univeristy.

Poulter, S. 2014. "It really is cheaper up North: How it's £3,000 a year more expensive to live in the South (and the cost of houses is the reason why)." In Daily Mail. 11th June 2014 - http://www.dailymail.co.uk/news/article-2654562/It-really-cheaper-North-How-3-000-year-expensive-live-South-cost-houses-reason-why.html.

Raftery, A.E. 1996. "Bayesian model selection in social research." In Sociological Methodology, ed. P.V. Marsden. Oxford: Blackwell.

Ravenstein, E. 1885. "The Laws of Migration." Journal of the Statistical Society of London 48(2):167-235.

Ravenstein, E. 1889. "The Laws of Migration." Journal of the Royal Statistical Society 52(2):241-305.

Raymer, J., G. Abel and P. W. F. Smith. 2007. "Combining census and registration data to estimate detailed elderly migration flows in England and Wales." Journal of the Royal Statistical Society Series a-Statistics in Society 170:891-908.

Raymer, J., A. Bonaguidi and A. Valentini. 2006. "Describing and Projecting the Age and Spatial Structures of Interregional Migration in Italy." Population, Space and Place 12:371-388.

Raymer, J., P. Rees, A. Blake, P. Boden, J. Brown, G. Disney, N. Lomax, P. Norman and J. Stillwell. 2012. "Conceptual Framework for UK Population and Migration Statistics." Titchfield: Office for National Statistics.

Reay, D., G. Crozier and J. Clayton. 2010. "'Fitting in' or 'standing out': working-class students in UK higher education." British Educational Research Journal 36(1):107-124.

Reay, D., M. David and S. Ball. 2005. Degrees of Choice: Social Class, Race and Gender in Higher Education. Stoke-on-Trent: Trenrham Books.

Reay, D., J. Davies, M. David and S. Ball. 2001. "Choices of degree or degrees of choice? Social class, race and the higher education choice process." Sociology 35:855-874.

Rees, P. 1977. "The measurement of migration from census and other sources." Environment and Planning 9:257-280.

Rees, P. 2009. "Demography." In International Encyclopedia of Human Geography, eds. R. Kitchin and N. RThrift. Oxford: Elsevier.

Rees, P., J. Stillwell, P. Boden and A. Dennett. 2009. "A review of migration statistics literature." UK Statistics Authority.

Rodriguez Gonzalez, C. and R.B. Mesanza. 2011. "The determinants of international student mobility flows: an empirical study on the Erasmus programme." Higher Education 62(413-430).

Rogers, A. and L. Castro. 1981a. "Model Migration Schedules." In Research Report-81-30: International Institute for Applied System Analysis, Laxenburg.

Rogers, A. and L.J. Castro. 1981b. "Model Migration Schedules." In Research Report RR-81-30. Laxenburg: International Institute for Applied Systems Analysis.

Rogers, A., R. Raquillet and L. Castro. 1977. "Model Migration Schedules and their Applications." International Institute for Applied Systems Analysis.

Rosenbaum, P.R. and D.B. Rubin. 1983. "The Central Role of the Propensity Score in Observational Studies for Causal Effects." Biometrika 70:41-55.

Rowland, D.T. 2003. Demographic Methods and Concepts. Oxford: Oxford University Press.

Russell Group. 2011. "Informed Choices: A Russell Group guide to making decisions about post-16 education." London: www.russellgroup.ac.uk.

Ryan, C. and L. Sibieta. 2010. "Private schooling in the UK and Australia." Institute for Fiscal Studies.

Sa, F. 2014. "The Effect of Tuition Fees on University Applications: Evidence from the UK." IZA Discussion Paper Series 8364.

Sage, J., D. Smith and P. Hubbard. 2012a. "The Diverse Geographies of Studentification: Living Alongside People Not Like Us." Housing Studies 27(8):1057-1078.

Sage, J., D. Smith and P. Hubbard. 2012b. "The Rapidity of Studentification and Population Change: There Goes the (Student)hood." Population Space and Place 18(5):597-613.

Salisbury, M.H., M.B. Pulsen and E.T. Pascarella. 2010. "To See the World or Stay at Home: Applying an Integrated Student Choice Model to Explore the Gender Gap in the Intent to Study Abroad." Research in Higher Education 51:615-640.

Salt, J. 2001. "The business of international migration." In International Migration into the 21st Century: Essays in Honour of Reginald Appleyard, ed. M. Siddique. Cheltenham: Edward Elgar.

Savage, M., F. Devine, N. Cunningham, M. Taylor, Y. Li, J. Hjellbrekke, B. Le Roux, S. Friedman and A. Miles. 2013. "A New Model of Social Class? Findings from the BBC's Great British Class Survey Experiment." Sociology 0(0).

Savills Research. 2007. "Student housing report." London: FPD Savills.

Schwartz, A. 1976. "Migration, Age and Education." Journal of Political Economy 24:701-720S.

Scottish Executive. 2005. "Cross Border Student Flows: Higher Education Tuition Fee Levels." Edinburgh.

Shiner, M. and T. Modood. 2002. "Help or Hinderance? Higher Education and the Route to Ethnic Equality." British Journal of Sociology of Education 23(2):209-232.

Sianesi, B. 2004. "An evaluation of the Active Labour Market Programmes in Sweden." The Review of Economics and Statistics 86(1):133-155.

Sianesi, B. 2012. A course on Policy Evaluation Methods: London.

Simpson, L. and N. Finney. 2009. "Spatial Patterns of Internal Migration: Evidence for Ethnic Groups in Britain." Population Space and Place 15(1):37-56.

Sjaastad, L. 1962. "The Costs and Returns of Human Migration." Journal of Political Economy 70:80-93.

Skeldon, R. 1997. Migration and Development: A Global Perspective. London: Longman.

Skeldon, R. 2012. "Migration Transitions Revisited: Their Continued Relevance for the Development of Migration Theory." Population Space and Place 18:154-166.

Smith, D. 2002. "Patterns and processes of studentification in Leeds." Regional Review 12(1):14-16.

Smith, D. 2005. "Studentification: the gentrification factory?" In Gentrification in a Global Context, eds. R. Atkinson and G. Bridge. London: Routledge.

Smith, D. 2008. "The Politics of Studentification and '(Un) balanced' Urban Populations: Lessons for Gentrification and Sustainable Communities?" Urban Studies 45(12):2541-2564.

Smith, D. 2009. "Student geographies, urban restructuring and the expansion of higher education." Environment and Planning A 41(1795-1804).

Smith, D. and H. Jons. 2015. "Education and Internal Migration." In Internal Migration: Geographical Perspectives and Processes, eds. D. Smith, N. Finney, K. H. Halfacree and N. Walford.

Smith, D. P. and L. Holt. 2007. "Studentification and 'apprentice' gentrifiers within Britain's provincial towns and cities: extending the meaning of gentrification." Environment and Planning A 39(1):142-161.

Smith, D. and J. Sage. 2014. "The regional migration of young adults in England and Wales (2002-2008): a 'conveyor-belt' of population redistribution?" Children's Geographies.

Smith, J. and P. Todd. 2005. "Does matching overcome LaLonde's critique of nonexperimental estimators?" Journal of Econometrics 125(1-2):305-353.

Sullivan, A., A. Heath and C. Rothon. 2011. "Equalisation or inflation? Social class and gender differentials in England and Wales." Oxford Review of Education 37(2):215-240.

Sutton Trust. 2000. "Entry to leading universities." London: Sage.

The Complete University Guide. 2014. "The Location - Where do you want to go?". http://www.thecompleteuniversityguide.co.uk/universities/choosing-the-right-university/the-location/.

Thomas, B. and D. Dorling. 2007. Identity in Britain: A cradle-to-grave atlas. Bristol: The Policy Press.

Tobin, J. 1958. "Estimation of Relationships for Limited Dependent Variables." Econometrica 26:24-36.

Todaro, M.P. 1969. "A Model of Labor Migration and Urban Unemployment in Less Developed Countries." The American Economic Review 59(1):138-148.

Tomlinson, S. 2001. "Some success, could do better: education and race 1976-2000." In Education, Reform and the State: Twenty-Five Years of Politics, Policy and Practice, eds. R. Phillips and J. Furlong. London: Routledge.

Tuckman, H.P. 1970. "Determinants of College Student Migration." Southern Economic Journal 37(2):184-189.

United Kingdom Government. 2012. "Education System in the UK."

United Kingdom Government. 2013. "Student Finance." https://www.gov.uk/student-finance/overview: GOV.UK.

United Kingdom Government. 2014. "School leaving age."

United Nations. 1998. "Recommendations on Statistics of International Migration Revision 1." Statistical Papers Series M, Department of Economic and Social Affairs Statistics Division 58(1).

Universities UK. 2007. "The economic benefits of a degree." In Research Report: Price Waterhouse Coopers.

Universities UK. 2010. "Universities: engaging with local communities." www.universitiesuk.ac.uk.

Universities UK. 2013. "Patterns and Trends in UK Higher Education." In In Focus: HIGHER EDUCATION: A DIVERSE AND CHANGING SECTOR: Universities UK and HESA.

Universities UK. 2014. "Patterns and Trends in UK Higher Education 2014." http://www.universitiesuk.ac.uk/highereducation/Documents/2014/PatternsAndTrendsInUKHigherEducation2014.pdf.

University of Wolverhampton. nd. "http://www.scit.wlv.ac.uk/ukinfo/."

Van Mol, C. and C Timmerman. 2014. "Should I Stay or Should I Go? An Analysis of the Determinants of Intra-European Student Mobility." Population, Space and Place 20:465-479.

Vasagar, Jeevan. 2012. "Tuition fees increase led to 15,000 fewer applicants." In The Guardian.

Vickers, D. and P. Rees. 2006. "Introducing the area classification of output areas." Population Trends 125:15-29.

Vickers, D. and P. Rees. 2007. "Creating the UK National Statistics 2001 output area classification." Journal of the Royal Statistical Society Series a-Statistics in Society 170(2):379-403.

Vickers, D., P. Rees and M. Birkin. 2003. "A new classification of UK local authorities using 2001 Census key statistics." University of Leeds.

Vincent-Lancrin, S. 2008. "The Reversal of Gender Inequalities in Higher Education: An On-going Trend." In Higher Education to 2030 - Volume 1: Demography, ed. OECD.

Walker, I. and Y. Zhu. 2013. "The impact of university degrees on the lifecycle of earnings: some further analysis." In BIS Research paper number 112, ed. Inovation and Skills Department for Business.

Wilkins, S., F. Shams and J. Huisman. 2013. "The decision-making and changing behavioural dynamics of potential higher education students: the impacts of increasing tuition fees in England." Educational Studies 39(2):125-141.

Wilson, T. 2010. "Model migration schedules incorporating student migration peaks." Demographic Research 23:191-222.

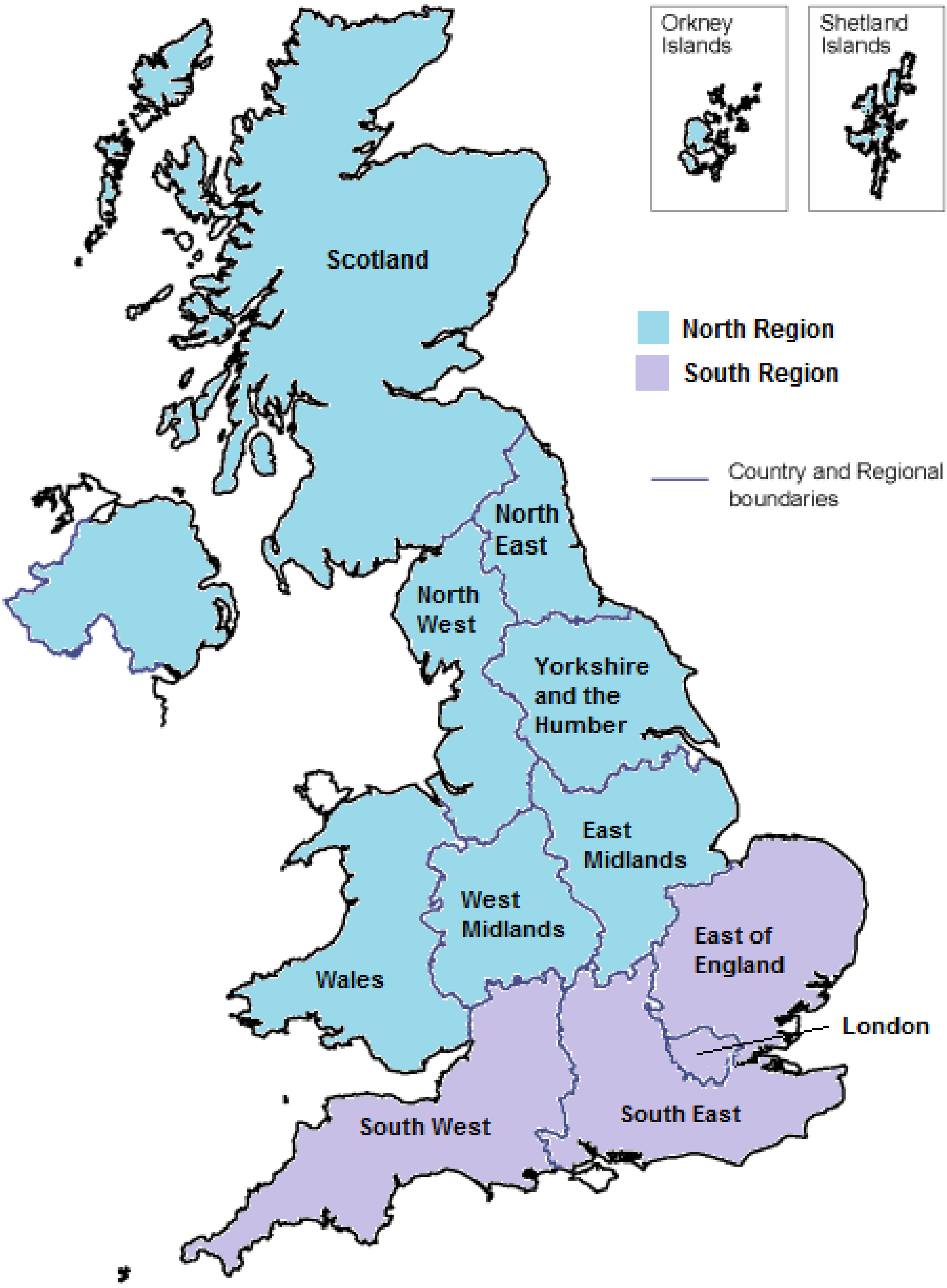
Wrigley, N, T Holt, D Steel and M Tranmer. 1996. "Analysing, modelling, and resolving the ecological fallacy." In Spatial Analysis, Modelling in a GIS Environment, eds. P Longley and M Batty. Cambridge: GeoInformation International.

Appendix A – Reference Maps of United Kingdom Geography

Figure A-1: Map of United Kingdom Counties

Source: Authors own creation

Figure A-2: Map of Regions in the United Kingdom



Source: Authors own creation

Appendix B – Student Migration Area Classification Outputs

Table B-1: Student Migration Area Classification Groupings definitions by student migration typology variables

Variable Name	Cluster Groups						
	Large University Settlement	Medium University Settlement	Commuting/Distance Learning HEIs	Migrant Commuting Student Settlements	Special Scenario Areas	Sending LAs - Commuters	Sending LAs - Student Migrants
Local Stud.	High	Very High	High	Low	Low	Low	Low
Commuters - IN	Medium	High	Very High	Low	Low	Low	Low
Commuters - OUT	Very Low	Very Low	Very Low	Medium	Low	High	Medium
Internal Student Migrants - IN	Very High	High	Medium	Low	Low	Low	Low
Internal Student Migrants – OUT	Very Low	Very Low	Very Low	Low	Low	Medium	High
Local Migrant Commuter - TERM	Low	Low	Low	Medium	Medium	Very High	Low
Local Migrant Commuter - DOM/INST	High	Very High	High	Low	Low	Low	Low
Internal Migrant Commuter - DOM	Very Low	Very Low	Very Low	Medium	Low	Low	High
Internal Migrant Commuter - TERM	Low	Low	Low	High	Very High	Low	Low
Internal Migrant Commuter - INST	Medium	Medium	Very High	Low	Low	Low	Low
Int. Student Migrant	Very High	High	Medium	Low	Low	Low	Low
Int. Student Commuter - TERM	Low	Low	Medium	Medium	Very High	Low	Low
Int. Student Commuter - INST	Medium	Medium	Very High	Low	Low	Low	Low

List of Local Authorities by Student Migration Area Classification

Cluster 1: Large University Settlements

- Aberdeen City
- Bath and North East Somerset
- Brighton and Hove
- Bristol, City of
- Cambridge
- Cardiff
- Central Bedfordshire
- Ceredigion
- Charnwood
- Colchester

- County Durham
- Coventry
- Edinburgh, City of
- Exeter
- Fife
- Guildford
- Gwynedd
- Hillingdon
- Lancaster
- Leeds
- Leicester
- Lincoln
- Liverpool
- Manchester
- Newcastle upon Tyne
- Norwich
- Nottingham
- Oxford
- Portsmouth
- Runnymede
- Sheffield
- Southampton
- Stirling
- Welwyn Hatfield
- Winchester
- Wrexham
- York

Cluster 2: Smaller University Settlements

- Belfast
- Birmingham
- Bradford
- Canterbury
- Carlisle
- Cheltenham
- Cheshire West and Chester
- Chichester
- Derby
- Dundee City
- Glasgow City
- Kingston upon Hull, City of
- Kirklees
- Northampton
- Plymouth
- Sunderland
- Swansea
- Wandsworth
- Wycombe

Cluster 3: Commuting/Distance Learning HEIs

- Barnet
- Bolton
- Camden
- Carmarthenshire
- Chelmsford
- City of London
- Coleraine
- Ealing
- East Lothian
- Greenwich
- Highland
- Ipswich
- Islington
- Kingston upon Thames
- Lewisham
- Luton
- Middlesbrough
- Newcastle-under-Lyme
- Newham
- Newport
- Poole
- Preston
- Renfrewshire
- Rhondda Cynon Taf
- Salford
- South Gloucestershire
- Southwark
- Stafford
- Telford and Wrekin
- Tower Hamlets
- Waverley
- West Lancashire
- Westminster
- Wokingham
- Wolverhampton
- Worcester

Cluster 4: Migrant Commuter Student Settlements

- Arun
- Barking and Dagenham
- Bedford
- Brent
- Cherwell
- Cheshire East
- Cornwall
- Derry
- Eden
- Enfield

- Epsom and Ewell
- Fareham
- Forest of Dean
- Gloucester
- Gosport
- Harrow
- Hastings
- Hertsmere
- High Peak
- Maidstone
- Medway
- Merton
- Newark and Sherwood
- Newtownabbey
-
- North West Leicestershire
- Redbridge
- Richmond upon Thames
- Rushcliffe
- Scarborough
- Slough
- South Cambridgeshire
- Southend-on-Sea
- Stevenage
- Stockton-on-Tees
- Vale of White Horse
- Waltham Forest
- Watford

<u>Cluster 5: Special Scenario Areas</u>

- Bournemouth
- Broxtowe
- Eastbourne
- Hackney
- Hammersmith and Fulham
- Haringey
- Hounslow
- Kensington and Chelsea
- Lambeth
- Oadby and Wigston
- Reading
- Stoke-on-Trent
- Warwick

<u>Cluster 6: Sending LAs – Commuters</u>

- Aberdeenshire
- Adur
- Allerdale
- Amber Valley
- Angus
- Antrim
- Ards
- Argyll & Bute
- Ashfield
- Ashford
- Ballymena
- Ballymoney
- Barnsley
- Barrow-in-Furness
- Bexley
- Blaby
- Blackburn with Darwen
- Blackpool
- Blaenau Gwent
- Bolsover
- Bridgend
- Broadland
- Bromsgrove
- Broxbourne
- Burnley
- Bury
- Caerphilly
- Calderdale
- Cannock Chase
- Carrickfergus
- Castlereagh
- Chesterfield
- Chorley
- Clackmannanshire
- Conwy
- Copeland
- Corby
- Craigavon
- Croydon
- Darlington
- Denbighshire
- Doncaster
- Dover
- Dudley
- Dumfries & Galloway
- East Ayrshire
- East Dunbartonshire
- East Renfrewshire
- East Riding of Yorkshire
- East Staffordshire
- Eastleigh
- Eilean Siar
- Erewash

- Falkirk
- Flintshire
- Gateshead
- Gedling
- Great Yarmouth
- Halton
- Harlow
- Hartlepool
- Havant
- Havering
- Hinckley and Bosworth
- Hyndburn
- Inverclyde
- Isle of Anglesey
- Kettering
- Knowsley
- Larne
- Lewes
- Lichfield
- Lisburn
- Mansfield
- Merthyr Tydfil
- Midlothian
- Moray
- Neath Port Talbot
- North Ayrshire
- North Down
- North East Derbyshire
- North Kesteven
- North Lanarkshire
- North Lincolnshire
- North Tyneside
- North Warwickshire
- Northumberland
- Nuneaton and Bedworth
- Oldham
- Orkney Islands
- Pendle
- Perth & Kinross
- Redcar and Cleveland
- Redditch
- Ribble Valley
- Rochdale
- Rossendale
- Rotherham
- Sandwell
- Scottish Borders
- Sefton
- Selby
- Shepway
- Shetland Islands
- Shropshire
- Solihull
- South Ayrshire
- South Derbyshire
- South Hams
- South Lanarkshire
- South Norfolk
- South Ribble
- South Staffordshire
- South Tyneside
- St. Helens
- Staffordshire Moorlands
- Stockport
- Swale
- Tameside
- Tamworth
- Thanet
- The Vale of Glamorgan
- Thurrock
- Torfaen
- Trafford
- Wakefield
- Walsall
- Warrington
- West Dunbartonshire
- West Lothian
- Wigan
- Wirral
- Wyre
- Wyre Forest

Cluster 7: Sending LAs - Student Migrants

- Armagh
- Aylesbury Vale
- Babergh
- Banbridge
- Basildon
- Basingstoke and Deane
- Bassetlaw
- Boston
- Bracknell Forest
- Braintree
- Breckland
- Brentwood
- Bromley
- Castle Point
- Channel Islands
- Chiltern
- Christchurch
- Cookstown
- Cotswold
- Craven

- Crawley
- Dacorum
- Dartford
- Daventry
- Derbyshire Dales
- Down
- Dungannon
- East Cambridgeshire
- East Devon
- East Dorset
- East Hampshire
- East Hertfordshire
- East Lindsey
- East Northamptonshire
- Elmbridge
- Epping Forest
- Fenland
- Fermanagh
- Forest Heath
- Fylde
- Gravesham
- Hambleton
- Harborough
- Harrogate
- Hart
- Herefordshire, County of
- Horsham
- Huntingdonshire
- Isle of Man
- Isle of Wight
- Isles of Scilly
- King's Lynn and West Norfolk
- Limavady
- Magherafelt
- Maldon
- Malvern Hills
- Melton
- Mendip
- Mid Devon
- Mid Suffolk
- Mid Sussex
- Milton Keynes
- Mole Valley
- Monmouthshire
- Moyle
- New Forest
- Newry and Mourne
- North Devon
- North Dorset
- North East Lincolnshire
- North Hertfordshire
- North Norfolk
- North Somerset
- Omagh
- Pembrokeshire
- Peterborough
- Powys
- Purbeck
- Reigate and Banstead
- Richmondshire
- Rochford
- Rother
- Rugby
- Rushmoor
- Rutland
- Ryedale
- Sedgemoor
- Sevenoaks
- South Bucks
- South Holland
- South Kesteven
- South Lakeland
- South Northamptonshire
- South Oxfordshire
- South Somerset
- Spelthorne
- St Albans
- St Edmundsbury
- Strabane
- Stratford-on-Avon
- Stroud
- Suffolk Coastal
- Surrey Heath
- Sutton
- Swindon
- Tandridge
- Taunton Deane
- Teignbridge
- Tendring
- Test Valley
- Tewkesbury
- Three Rivers
- Tonbridge and Malling
- Torbay
- Torridge
- Tunbridge Wells
- Uttlesford
- Waveney
- Wealden
- Wellingborough
- West Berkshire
- West Devon
- West Dorset
- West Lindsey
- West Oxfordshire
- West Somerset

- Weymouth and Portland
- Wiltshire
- Windsor and Maidenhead
- Woking
- Worthing
- Wychavon

Appendix C – Creation of Variables

Throughout the thesis a number of variables have been created by manipulating and adapting the variables provided directly from the datasets sourced to conduct this research. A detailed explanation of how these new variables were created and the rationale behind their creation is provided in this appendix section.

Background Variable

The variable used in the analysis labelled Social *Background* was created by merging the original *Parental Background* and *Socio-Economic Classification* variables from the HESA dataset. The process that was undertaken in this merge is explained below.

The first step was to re-categorise the original *Parental Background* (pared) variable which records information about whether an entrant's parents have HE qualifications. This field is only compulsory for entrants through UCAS at institutions in England and Scotland and splits the students into 5 categories; Yes, No, Don't Know, Information Refuses and Unknown. It was decided to merge the final 3 categories into 1 category called Unknown, and the new variable created was labelled pared2 (please refer to Table C-1)

The second step was to re-categorise the original Socio-Economic Classification (sec) variable which recorded the socio-economic background of students aged 21 and over at the start of their course, or for students under 21 the socio-economic background of their parent, step-parent or guardian who earns the most. It is based on occupation, and if the parent or guardian is retired or unemployed, this is based on their most recent occupation. The 'sec' variable categorises a student into 10 categories (please refer to Table C-2), one of which was called 'Not Classified'. 'Not classified' includes the 3 categories: 'Students', 'occupations not stated or inadequately described' and 'not classifiable for other reasons'. It was decided to create a new variable (sec2) that merged these 10 categories into 4 broader categories as shown in

Table . This re-categorisation followed the ONS guidelines reducing the classification from 10 categories to 3 plus an unknown field (Office for National Statistics 2010b)

The final step was to decide what combination of the pared2 and sec2 variables would produce the best background variable. Eight different combinations of the two variables were tested by modelling the different combinations on how well they predicted migration of a student and the combination that was chosen is shown in Table C-3 and Table C-4.

Table C-1: Reformatting of 'pared' variable into 'pared2' variable

	pared2			
pared	**Yes**	**No**	**Unknown**	**Total**
Yes	622,964			622,964
No		561,130		561,130
Don't Know			193,564	193,564
Information Refused			220,142	220,142
Unknown			199,692	199,692
Total	622,964	561,130	613,398	1,797,492

Table C-2: Reformatting of Sec variable into sec2 variable

sec	sec2				
	Higher	Intermediate	Low	Not Classified/Unknown	Total
Higher Managerial	236,394				236,394
Lower Managerial	314,898				314,898
Intermediate		142,224			142,224
Small Employers		75,851			75,851
Lower Supervisory		49,413			49,413
Semi-routine			143,930		143,930
Routine			63,016		63,016
Never Worked			3,345		3,345
Not Classified				360,422	360,422
Unknown				407,999	407,999
Total	551,292	267,488	210,291	768,421	1,797,492

Table C-3: Classification of new social background variable based on pared2 and sec2 classification

pared2	sec2			
	Higher	Intermediate	Low	Not Classified/Unknown
Yes	Most Advantaged	Advantaged	Less Advantaged	Less Advantaged
No	Advantaged	Less Advantaged	Least Advanced	Least Advantaged
Unknown/Don't Know	Advantaged	Less Advantaged	Least Advantaged	Unknown

Table C-4: Classification of new social background variable with student numbers

pared2	sec2			
	Higher	**Intermediate**	**Low**	**Not Classified/Unknown**
Yes	314,255	93,354	48,182	177,179
No	126,543	117,041	108,609	207,937
Unknown/Don't Know	110,494	66,093	53,500	383,311

Level of Study

Level of study is taken from the course aim of the student. HESA classifies courses according to a framework which aligns with the framework for HE qualifications in England, Wales and Northern Ireland (FHEQ), the Scottish Credit and Qualifications Framework (SCQF) (of which the framework for qualifications of HE institutions in Scotland is a constituent part) and the International Standard Classification of Education (ISCED) and Bologna frameworks. Details are available at www.hesa.ac.uk/C11051/a/COURSEAIM. It includes level M for taught masters degrees, and level H for honours degrees.

Postgraduate courses are those leading to higher degrees, diplomas and certificates (including Postgraduate Certificate in Education (PGCE at level M) (unless shown separately) and professional qualifications) which usually require a first degree as an entry qualification (i.e. already qualified at level H).

Undergraduate courses are programmes of study at level H, I, J and C including, but not limited to, first degrees (including eligibility to register to practice with a health or social care or veterinary statutory regulatory body), first degrees with Qualified Teacher Status (QTS)/registration with a General Teaching Council (GTC), postgraduate bachelors degrees at level H, enhanced first degrees (including those leading towards obtaining eligibility to register to practice with a health or social care or veterinary statutory regulatory body), first degrees obtained concurrently with a diploma and intercalated first degrees, Professional Graduate Certificate in Education (PGCE at level H), foundation degrees, diplomas in HE (including those leading towards obtaining eligibility to register to practice with a health or social care or veterinary statutory regulatory body), Higher National Diploma (HND), Higher National Certificate (HNC), Diploma of Higher Education (DipHE), Certificate of Higher Education (CertHE), foundation courses at higher education level, National Qualifications Framework (NQF) levels 4 and 5, post-degree diplomas and certificates at undergraduate level (including those in Teaching in the Lifelong Learning Sector), professional qualifications at undergraduate level and other undergraduate diplomas and certificates including post-registration health and social care courses. Entrants to these programmes of study do not usually require a higher education qualification.

Course Studied

The subject of study of a student in higher education is recorded by HESA using a system known as JACS2 codes in which the initial letter identifies the subject group, for example F for physical sciences. A full listing of the JACS2 can be found at www.hesa.ac.uk/jacs2.

HESA has defined 19 subject areas in terms of JACS2 codes for reporting information broken down by subject to present a useful broad-brush picture. The subject areas do not overlap, and cover the entire range of JACS2 principal subjects.

For the purpose of this study an even more refined number of subject areas were required and as a result the original 19 subject areas defined by HESA were merged into 7 groups to depict the subject area of study in a new variable called sub2.

The original HESA subject areas, the corresponding JACS2 codes and the new sub2 groupings are shown in Table C-5.

Table C-5: HESA Subject areas, JACS2 Codes and sub2 grouping

HESA - Subject areas	JACS2 code	sub2 Group
Medicine & dentistry	A	Medicine
Subjects allied to medicine	B	Medicine
Biological sciences	C	Science or Engineering
Veterinary science	D1/2	Agricultural or Veterinary
Agriculture & related subjects	D0/3/4/5/6/7/9	Agricultural or Veterinary
Physical sciences	F	Science or Engineering
Mathematical sciences	G00/01/1/2/3/90/91	Social or Human Sciences
Computer science	G02/4/5/6/7/92	Science or Engineering
Engineering & technology	H, J	Science or Engineering
Architecture, building & planning	K	Science or Engineering
Social studies	L	Social or Human Sciences
Law	M	Business or Law
Business & administrative studies	N	Business or Law
Mass communications & documentation	P	Humanities
Languages	Q, R, T	Humanities
Historical & philosophical studies	V	Humanities
Creative arts & design	W	Humanities
Education	X	Social or Human Sciences
Combined	Y	Combined

Institution Category

The 160 Higher Institution Categories have been split into 6 categories depending on the year in which they were founded. A description of these is provided in Section 4.3.3. Below is a list of all 160 HEIs split into the 6 Institution Categories.

Ancient Universities

- The University of Cambridge
- University of Durham
- The University of Oxford
- The University of Edinburgh
- The University of Glasgow
- The University of Aberdeen
- The University of St Andrews

Red Brick or Civic Universities

- The University of Birmingham
- The University of Bristol
- The University of Exeter
- The University of Hull
- The University of Leeds
- The University of Leicester
- The University of Liverpool
- Birkbeck College
- Goldsmiths College
- Imperial College of Science, Technology
- Institute of Education
- King's College London
- London Business School
- London School of Economics and Politica
- London School of Hygiene and Tropical M
- Queen Mary and Westfield College
- Royal Holloway and Bedford New College
- The Royal Veterinary College
- St George's Hospital Medical School
- The School of Oriental and African Stud
- The School of Pharmacy
- University College London
- University of London (Institutes and ac
- The University of Newcastle-upon-Tyne
- The University of Nottingham
- The University of Reading
- The University of Sheffield
- The University of Southampton
- The University of Dundee
- Aberystwyth University
- Bangor University
- Cardiff University
- Swansea University
- The Queen's University of Belfast
- The Institute of Cancer Research
- The University of Manchester

Plate Glass or 1960s Universities

- Aston University
- The University of Bath
- The University of Bradford
- Brunel University
- The City University
- The University of East Anglia
- The University of Essex
- The University of Keele
- The University of Kent
- The University of Lancaster
- Loughborough University
- The University of Salford
- The University of Surrey
- The University of Sussex
- The University of Warwick
- The University of York
- The University of Strathclyde
- Heriot-Watt University
- The University of Stirling
- University of Ulster

Post-1992 Universities

- University of Bedfordshire
- The University of Northampton
- University of Cumbria
- The University of Worcester
- Anglia Ruskin University

- Bath Spa University
- The University of Bolton
- Bournemouth University
- The University of Brighton
- Birmingham City University
- The University of Central Lancashire
- University of Derby
- The University of East London
- University of Hertfordshire
- The University of Huddersfield
- Kingston University
- Leeds Metropolitan University
- Liverpool John Moores University
- The Manchester Metropolitan University
- Middlesex University
- De Montfort University
- The University of Northumbria at Newcas
- The Nottingham Trent University
- Oxford Brookes University
- The University of Plymouth
- The University of Portsmouth
- Sheffield Hallam University
- London South Bank University
- Staffordshire University
- The University of Sunderland
- Teesside University
- The University of West London
- University of the West of England, Bris
- The University of Westminster
- The University of Wolverhampton
- Cardiff Metropolitan University
- University of Glamorgan
- University of Abertay Dundee
- The Robert Gordon University
- The University of the West of Scotland
- Glasgow Caledonian University
- Edinburgh Napier University

Recently Created Universities

- University of Chester
- Canterbury Christ Church University
- York St John University
- Edge Hill University
- University College Falmouth
- Harper Adams University College
- The University of Winchester
- Newman University College
- Roehampton University
- Southampton Solent University
- Leeds Trinity University College
- University of Gloucestershire
- The University of Greenwich
- The University of Chichester
- The University of Wales, Newport
- Glyndwr University
- Swansea Metropolitan University
- Queen Margaret University, Edinburgh
- University of Wales Trinity Saint David

Other Universities

- Cranfield University
- Royal College of Art
- Bishop Grosseteste University College L

- Buckinghamshire New University
- Central School of Speech and Drama
- University College Plymouth St Mark and
- University of the Arts, London
- Ravensbourne
- Rose Bruford College
- Royal Academy of Music
- Royal College of Music
- Royal Northern College of Music
- St Mary's University College, Twickenha
- Trinity Laban Conservatoire of Music an
- Coventry University
- The University of Lincoln
- Glasgow School of Art
- Royal Conservatoire of Scotland
- Scottish Agricultural College
- Writtle College
- Norwich University College of the Arts
- Stranmillis University College
- St Mary's University College
- Royal Agricultural College
- University of the Highlands and Islands
- The Arts University College at Bournemo
- Conservatoire for Dance and Drama
- University College Birmingham
- Courtauld Institute of Art
- The University of Buckingham
- Heythrop College
- University for the Creative Arts
- Guildhall School of Music and Drama
- The Liverpool Institute for Performing
- University Campus Suffolk
- Leeds College of Art

Appendix D – Modelling Selection Process for Chapter 4

Logistic Modelling Process

Table D-1: Scalar measures of fit for logistic modelling process for migration

Model	Variables	LL	R^2	AIC	BIC	Difference in BIC	Evidence
1	Ethnicity Background Gender	-1108969	0.1022	2217959.092	2218083.111		
2	M1 + Subject Institution Category	-1046663	0.1527	2093369.123	2093629.563	124453.548	V.V. Strong for M2 over M1 (Positive)
3	M2 + Age Year of Student Level of Student	-935324	0.2428	1870715.040	1871124.303	222505.260	V.V. Strong for M3 over M2 (Positive)
4	M3 + Domicile North/South	-924472	0.2516	1849015.108	1849449.175	21675.128	V.V. Strong for M4 over M3 (Positive)
Adding Interaction Terms – So models below nested in model 4							
5	M4 + EthnicityGender	-924450	0.2516	1848979.157	1849462.831	-13.656 (with M4)	Positive for M4 over M5 (Negative)
6	M4 + BackgroundGender	-924349	0.2517	1848777.404	1849261.078	188.097 (with M4)	V. Strong for M6 over M4 (Positive)
7	M4 + EthnicityBackground	-924043	0.2519	1848188.302	1848820.799	628.375 (with M4)	V.V. Strong for M7 over M4 (Positive)
Model 7 added most value – below testing adding 2 interactions to model 7							
8	M7 + BackgroundGender	-923918	0.2520	1847947.761	1848629.866	190.933 (with M7)	V. Strong for M8 over M7 (Positive)
9	M7 + EthnicityGender	-924018	0.2519	1848146.767	1848828.871	-8.072 (with M7)	V. Strong for M7 over M9 (Negative)
Model 8 added most value – finally model 10 tests adding 3 interaction terms into the model – nested in model 8							
10	M7 + BackgroundGender	-923888	0.2520	1847895.031	1848626.743	3.123	Positive for M10 over M8 (Positive)

Tobit Modelling Process

Table D-2: Scalar measure of fit for Tobit modelling process for distance

Model	Variables	LL	R^2	ML (Cox-Snell) R^2	AIC	BIC	Difference in BIC	Evidence
1	Ethnicity Background Gender	-9993495	0.0050	0.054	19990000	19990000		
2	M1 + Subject Institution Category	-9957519	0.0086	0.091	1.992e+07	1.992e+07	71794.483	V.V. Strong for M2 over M1 (Positive)
3	M2 + Age Year of Student Level of Student	-9940691	0.0102	0.108	1.988e+07	1.988e+07	33482.807	V.V. Strong for M3 over M2 (Positive)
4	M3 + Domicile North/South	-9921919	0.0121	0.127	1.984e+07	1.984e+07	37529.396	V.V. Strong for M4 over M3 (Positive)
Adding Interaction Terms – So models below nested in model 4								
5	M4 + EthnicityGender	-9921897	0.0121	0.127	1.984e+07	1.984e+07	-12.021	V. Strong for M4 over M5 (Negative)
6	M4 + BackgroundGender	-9921734	0.0121	0.127	1.984e+07	1.984e+07	312.396	V. Strong for M6 over M4 (Positive)
7	M4 + EthnicityBackground	-9921739	0.0121	0.127	1.984e+07	1.984e+07	130.815	V. Strong for M7 over M4 (Positive)
Model 6 added most value – below testing adding 2 interactions to model 6								
8	M6 + EthnicityBackground	-9921557	0.0121	0.127	1.984e+07	1.984e+07	123.989	V. Strong for M8 over M6 (Positive)
9	M6 + EthnicityGender	-9921716	0.0121	0.127	1.984e+07	1.984e+07	-194.122	V. Strong for M6 over M9 (Negative)
Only Model 8 added value – finally model 10 tests adding 3 interaction terms into the model – nested in model 8								
10	M8 + BackgroundGender	-9921526	0.0121	0.127	1.984e+07	1.984e+07	4.469	Positive for M10 over M8 (Positive)

Multinomial Logistic Modelling

Table D-3: Scalar measure of fit for multinomial logistic modelling process for student migration category

Model	Variables	LL	R^2	ML (Cox-Snell) R^2	AIC	BIC	Difference in BIC	Evidence
1	Ethnicity Background Gender	-1992228	0.0659	0.145	3984517.380	3984889.276		
2	M1 + Subject Institution Category	-1908817	0.1050	0.222	3817760.841	3818541.824	166347.453	V.V. Strong for M2 over M1 (Positive)
3	M2 + Age Year of Student Level of Student	-1770494	0.1699	0.333	3541187.420	3542414.678	276127.145	V.V. Strong for M3 over M2 (Positive)
4	M3 + Domicile North/South	-1739612	0.1843	0.356	3479434.418	3480736.055	61678.623	V.V. Strong for M4 over M3 (Positive)
Adding Interaction Terms – So models below nested in model 4								
5	M4 + EthnicityGender	-1739547	0.1844	0.356	3479329.079	3480779.475	-43.420	V.Strong for M4 over M5 (Negative)
6	M4 + BackgroundGender	-1739425	0.1844	0.356	3479083.964	3480534.360	201.695	V.Strong for M6 over M4 (Positive)
7	M4 + EthnicityBackground	-1739005	0.1846	0.356	3478316.986	3480213.658	522.397	V.Strong for M7 over M4 (Positive)
Model 7 added most value – below testing adding 2 interactions to model 7								
8	M7 + EthnicityBackground	-1738815	0.1847	0.356	3477961.066	3480006.497	207.162	V.Strong for M8 over M7 (Positive)
9	M7 + EthnicityGender	-1738940	0.1846	0.356	3478210.239	3480255.669	-42.011	V.Strong for M7 over M9 (Negative)
Only Model 8 added value – finally model 10 tests adding 3 interaction terms into the model – nested in model 8								
10	M8 + BackgroundGender	-1738745	0.1847	0.356	3477845.310	3480039.499	-33.003	Strong for M8 over M10 (Negative)

Appendix E – Modelling Selection Process for Chapter 5

The first regression model in Chapter 5 was a Logistic Regression of Economic Activity. In order to find the combination of variables that had the best statistical fit to the data and made the most theoretical sense the following modelling procedure was conducted – the Scalar Measures of Fit are Show in Table E-1.

Table E-1: Scalar measure of fit measures for Logistic Regression Model of Economic Activity

Model	Variables	LL	R^2	AIC	BIC	Difference in BIC	Evidence
1	Migrate LA (Yes/No)	-123219	0.0005	0.759	101.538		
2	M1 + Characteristic Variables Ethnicity Background Subject Institution Cat. Gender Age Level of Student	-118243	0.0408	0.729	9723.437	9621.899	V.V. Strong for M2 over M1 (positive)
3	M2 + North/South Domicile	-118235	0.0409	0.728	9726.077	2.640	Medium for M3 over M2 (positive)
Adding Interaction Terms – So models below nested in Model 3							
4	M3 + Interaction: EthnicityBackground	-118183	0.0413	0.728	9627.262	-98.815	V. Strong for M3 over M4 (negative)
5	M3 + Interaction: BackgroundGender	-118181	0.0413	0.728	9783.713	57.636	V. Strong for M5 over M3 (positive)
6	M3 + Interaction: EthnicityGender	-118205	0.0411	0.728	9734.959	8.882	Strong for M6 over M3 (positive)
Model 5 added most value – below testing adding 2 interactions to Model 5							
7	M6 + Interaction: EthnicityGender	-118159	0.0415	0.728	9777.591	-6.122	Strong for M5 over M7 (negative)
8	M6 + Interaction: EthnicityBackground	-118130	0.0417	0.728	9683	-100.434	V. String for M5 over M8 (negative)
Models 7 and 8 did not add any value so Modelling Process Stopped Here – Model 5 was the best statistically fitting model of Economic Status							

Model	Variables	LL	R^2	AIC	BIC	Difference in BIC	Evidence
9	Full Model: All three interactions	-118132	0.0417	0.728	9627	-156	V. Strong for M5 over M9 (negative)

The second regression model in Chapter 5 was a Multiple Linear Regression of Log_Salary. Again, in order to find the combination of variables that had the best statistical fit to the data and made the most theoretical sense the following modelling procedure was conducted – the Scalar Measures of Fit are Show in Table E-2. This was the same modelling procedure as conducted for the Economic Activity with the addition of adding the SIC and SOC variables. These were not used in the modelling of Economic Activity as those unemployed will not have a SIC or SOC categorisation. However when modelling salary everyone in the model are employed and therefor will have a SIC and SOC classification and these variables may impact on the salary earned.

Table E-2: Scalar measure of fit measures for Multiple Linear Regression of Log_Salary

Model	Variables	LL	R^2	AIC	BIC	Difference in BIC	Evidence
1	Migrate LA (Yes/No)	-104550	0.010	1.228	1662.516		
2	M1 + Characteristic Variables Ethnicity Background Subject Institution Cat. Gender Age Level of Student	-79162	0.265	0.930	52125.995	50463.479	V.V Strong for M2 over M1 (positive)
3	M2 + North/South Domicile	-78586	0.270	0.923	53265.605	1139.611	V.V Strong for M3 over M2 (positive)
4	M3 + SOC Variable	-57648	0.429	0.678	94960.546	41694.941	V.V Strong for M4 over M3 (positive)
5	M4 + SIC Variable	-52385	0.462	0.618	104645.915	9685.369	V.V Strong for M5 over M4 (positive)
6	M5 + Work GOR	-49792	0.478	0.588	109318.847	4672.932	V.V Strong for M6 over M5 (positive)
Adding Interaction Terms – So models below nested in Model 6							
7	M6 + Interaction: Ethnicity*Background	-49761	0.478	0.588	109187.364	-131.483	V. Strong for M6 over M7 (negative)
8	M6 + Interaction: Background*Gender	-49689	0.478	0.587	109476.028	157.181	V. Strong for M8 over M6 (positive)
9	M6 + Interaction: Ethnicity*Gender	-49777	0.478	0.588	109299.369	-19.478	V. Strong for M6 over M9 (negative)
Model 8 added most value – below testing adding 2 interactions to Model 8							
10	M8 + Interaction: Ethnicity*Gender	-49679	0.478	0.587	109446.896	-29.132	V. Strong for M8 over M10 (negative)
11	M8 + Interaction: Ethnicity*Background	-49657	0.478	0.587	109347.340	-128.688	V.V. Strong for M8 over M11 (negative)
Models 10 and 11 did not add any value so Modelling Process Stopped Here – Model 8 was the best statistically fitting model of Log_Salary							

Table E-3: Logistic Regression outputs for Unemployment of PSM Models 1 to 3

Unemployment	M1		M2		M3	
Variables	**Coef. (β)**	**P-Val**	**Coef. (β)**	**P-Val**	**Coef. (β)**	**P-Val**
Constant	-3.246	0	-3.172	0	-3.052	0
Migrate						
Yes	0.0116	0.327	-0.0123	0.309	-0.0132	0.274
No[a]						
Ethnicity						
White[a]						
Black	0.903	0	0.915	0	0.533	0
Asian	0.755	0	0.749	0	0.396	0
Other	0.571	0	0.569	0	0.451	0
Unknown	0.375	0	0.361	0	-0.00243	0.988
Background						
Most Advantaged[a]						
Advantaged	-0.0447	0.0134	-0.0319	0.0787	-0.0647	0.0148
Less Advantaged	0.0278	0.110	0.0468	0.00751	-0.0839	0.00127
Least Advantaged	0.0364	0.0457	0.0616	0.000809	-0.0949	0.000665
Unknown	0.00957	0.634	0.0247	0.222	-0.194	0
Gender						
Male[a]						
Female	-0.180	0	-0.178	0	-0.348	0
Subject						
Medicine[a]						
Science/Engineering	1.092	0	1.110	0	1.118	0
Agricultural/Veterinary	1.161	0	1.195	0	1.209	0
Social/Human	0.661	0	0.682	0	0.684	0
Business/Law	0.892	0	0.922	0	0.931	0
Humanities	1.341	0	1.374	0	1.383	0
Combined	1.601	0	1.607	0	1.615	0
Age						
17 years and under	0.612	0.00925	0.718	0.00238	0.708	0.00281
18-20 years[a]						
21-24 years	-0.0997	0	-0.100	0	-0.100	0
25-29 years	-0.108	1.14e-06	-0.0968	1.29e-05	-0.0932	2.73e-05
30 years and over	0.0764	5.22e-05	0.0946	6.30e-07	0.0969	3.67e-07
Age unknown	-1.280	0.212	-1.193	0.245	-1.167	0.255
Level of Study						
Post-Graduate[a]						
Under-Graduate	0.463	0	0.498	0	0.494	0
Institution Category						
Ancient[a]						
Red Brick			-0.0463	0.0803	-0.0450	0.0897
Plate Glass			-0.130	4.14e-06	-0.126	8.32e-06
New University			-0.157	1.47e-09	-0.155	2.46e-09
Recent University			-0.231	0	-0.226	0
Other			-0.299	0	-0.295	0
Interaction Terms						
Ethnicity*Background						
White*Most Advantaged[a]						
Black*Advantaged					0.149	0.0822
Black*Less Advantaged					0.217	0.00203
Black*Least Advantag.					-0.00256	0.977
Black*Unknown					0.508	0.00985
Asian*Advantaged					0.230	0.00196
Asian*Less Advantaged					0.259	7.30e-05
Asian*Least Advantaged					0.149	0.0602
Asian*Unknown					0.405	0.0291
Other*Advantaged					0.356	1.72e-06
Other*Less Advantaged					0.320	4.36e-07
Other*Least Advantag.					0.212	0.0102
Other*Unknown					0.402	0.0328
Unknown*Advantaged					0.469	4.48e-10
Unknown*Less.Adv					0.391	5.78e-09
Unknown*Least.Adv					0.209	0.0149
Unknown*Unknown					0.291	0.0899
Background*Gender						
V. Advantaged*Male[a]						
Advantaged*Female					0.0126	0.727
Less Advant.*Female					0.180	1.39e-07
Least Advant.*Female					0.182	2.56e-07
Unknown*Female					0.275	0
Ethnicity*Gender						
White*Male[a]						
Black*Female					0.192	9.77e-06
Asian*Female					0.173	6.13e-07
Other*Female					0.00382	0.941
Unknown*Female					0.0944	0.238

Table E-4: Multiple Regression outputs for Log_Salary of PSM Models 1 to 3

Log_Salary	M1 Coef. (β)	P-Val	M2 Coef. (β)	P-Val	M3 Coef. (β)	P-Val	M4 Coef. (β)	P-Val
Constant	10.12	0	10.21	0	10.19	0	10.35	0
Migrate								
Yes	0.0540	0	0.0361	0	0.0361	0	0.0225	0
No[a]								
Ethnicity								
White[a]								
Black	-0.0497	0	-0.0444	0	-0.0811	5.20e-08	-0.118	0
Asian	0.00260	0.489	-0.00475	0.205	0.0706	2.01e-09	-0.0506	0
Other	0.0113	0.0381	0.00769	0.156	-9.53e-06	0.999	-0.0401	0
Unknown	0.0309	5.47e-05	0.0231	0.00242	0.0449	0.154	0.00345	0.645
Background								
Most Advantaged[a]								
Advantaged	-0.00793	0.0167	0.000310	0.925	0.00181	0.729	0.00282	0.385
Less Advantaged	-0.0492	0	-0.0374	0	-0.0272	6.33e-08	-0.0350	0
Least Advantaged	-0.0768	0	-0.0596	0	-0.0297	2.76e-08	-0.0518	0
Unknown	-0.0143	5.69e-05	-0.00904	0.0103	0.0304	9.10e-09	0.00848	0.0149
Gender								
Male[a]								
Female	-0.114	0	-0.111	0	-0.0901	0	-0.109	0
Subject								
Medicine[a]								
Science/Engineering	-0.155	0	-0.148	0	-0.148	0	-0.155	0
Agricultural/Veterinary	-0.198	0	-0.190	0	-0.190	0	-0.179	0
Social/Human	-0.155	0	-0.146	0	-0.145	0	-0.158	0
Business/Law	-0.112	0	-0.0950	0	-0.0949	0	-0.109	0
Humanities	-0.340	0	-0.331	0	-0.331	0	-0.357	0
Combined	-0.213	0	-0.207	0	-0.206	0	-0.228	0
Age								
17 years and under	-0.151	0.0424	-0.136	0.0661	-0.136	0.0673	-0.118	0.104
18-20 years[a]								
21-24 years	0.113	0	0.113	0	0.113	0	0.108	0
25-29 years	0.237	0	0.243	0	0.242	0	0.236	0
30 years and over	0.375	0	0.388	0	0.387	0	0.385	0
Age unknown	0.670	4.72e-10	0.703	0	0.701	5.14e-11	0.713	0
Level of Study								
Post-Graduate[a]								
Under-Graduate	-0.222	0	-0.200	0	-0.200	0	-0.187	0
Institution Category								
Ancient[a]								
Red Brick			-0.0634	0	-0.0631	0	-0.0582	0
Plate Glass			-0.0875	0	-0.0874	0	-0.0750	0
New University			-0.152	0	-0.151	0	-0.132	0
Recent University			-0.151	0	-0.152	0	-0.140	0
Other			-0.146	0	-0.147	0	-0.149	0
Region of Employment								
North East							-0.173	0
North West							-0.203	0
Yorkshire and Humber							-0.195	0
East Midlands							-0.189	0
West Midlands							-0.191	0
East of England							-0.150	0
London[a]								
South East							-0.133	0
South West							-0.199	0
Northern Ireland							-0.301	0
Scotland							-0.169	0
Wales							-0.243	0
Non-UK							-0.140	0
Interaction Terms								
Ethnicity*Background								
White*Most Advantaged[a]								
Black*Advantaged					-0.00782	0.683		

Log_Salary	M1 Coef. (β)	M1 P-Val	M2 Coef. (β)	M2 P-Val	M3 Coef. (β)	M3 P-Val	M4 Coef. (β)	M4 P-Val
Black*Less Advantaged					-0.0562	0.000104		
Black*Least Advantag.					0.00882	0.631		
Black*Unknown					-0.0304	0.442		
Asian*Advantaged					0.0183	0.270		
Asian*Less Advantaged					-0.0682	4.08e-07		
Asian*Least Advantaged					0.0125	0.466		
Asian*Unknown					0.00572	0.874		
Other*Advantaged					-0.0206	0.220		
Other*Less Advantaged					-0.111	0		
Other*Least Advantag.					-0.0174	0.336		
Other*Unknown					-0.0380	0.297		
Unknown*Advantaged					-0.00147	0.930		
Unknown*Less.Adv					-0.0828	1.43e-09		
Unknown*Least.Adv					-0.0290	0.104		
Unknown*Unknown					-0.00349	0.914		
Background*Gender								
V. Advantaged*Male[a]								
Advantaged*Female					0.00218	0.742		
Less Advant.*Female					-0.0144	0.0219		
Least Advant.*Female					-0.0341	1.48e-07		
Unknown*Female					-0.0576	0		
Ethnicity*Gender								
White*Male[a]								
Black*Female					0.0614	7.20e-10		
Asian*Female					-0.00310	0.676		
Other*Female					0.0205	0.0640		
Unknown*Female					-0.0272	0.0742		

Appendix F – Test of common support for probit PSM models

Figure F-1: Density plot to test for common support – Probit PSM Model 1

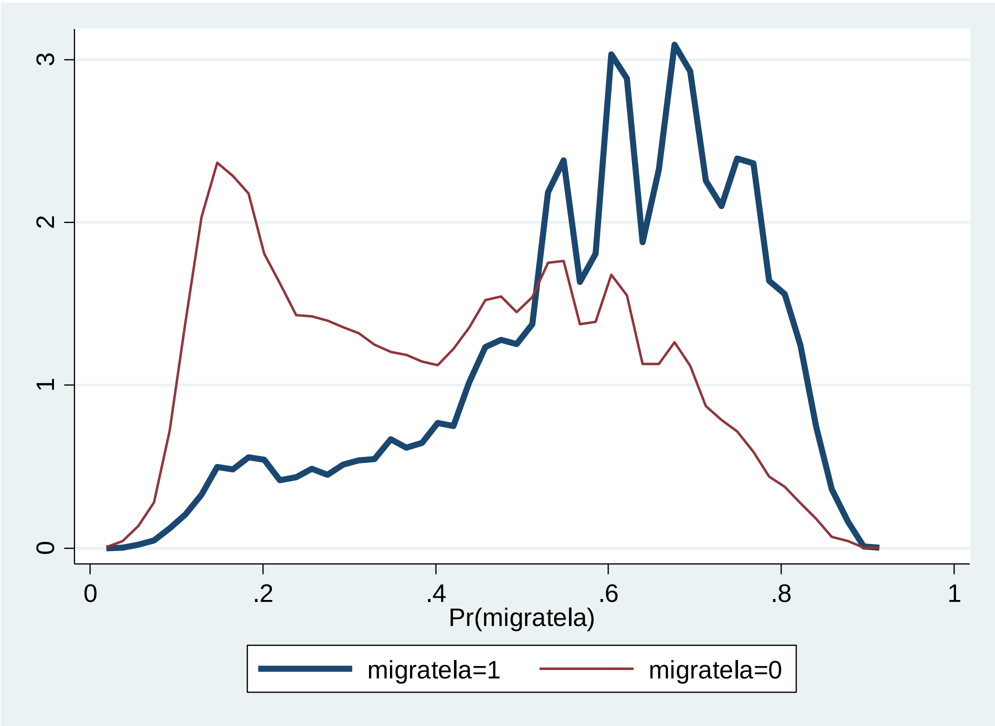


Figure F-2: Density plot to test for common support – Probit PSM Model 2

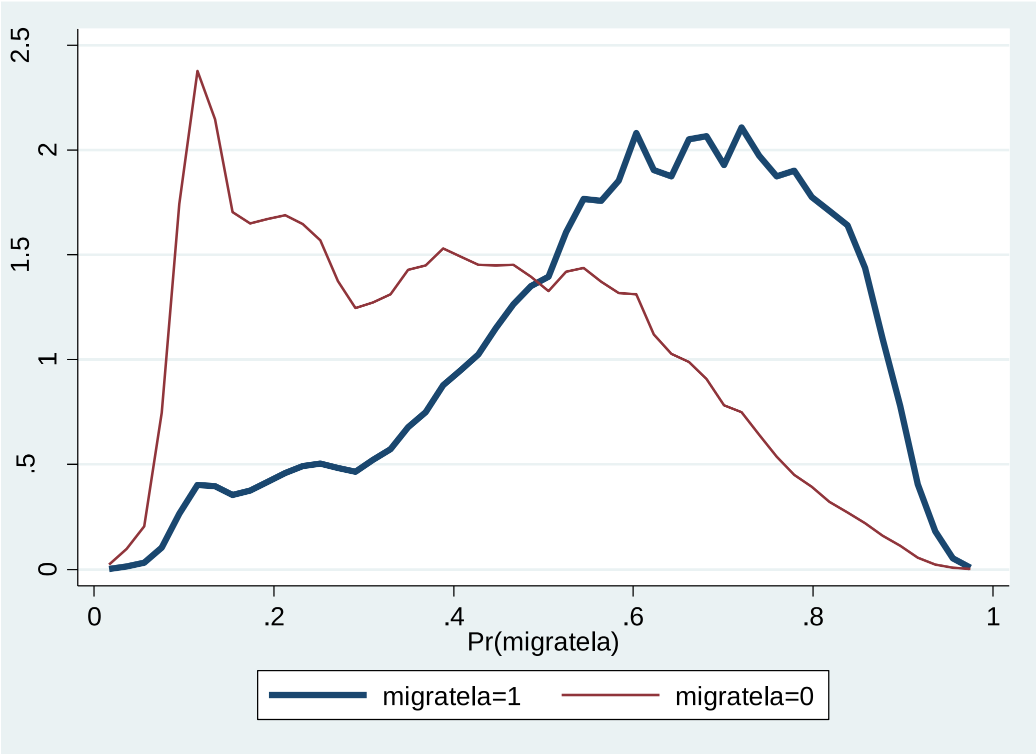


Figure F-3: Density plot to test for common support – Probit PSM Model 3

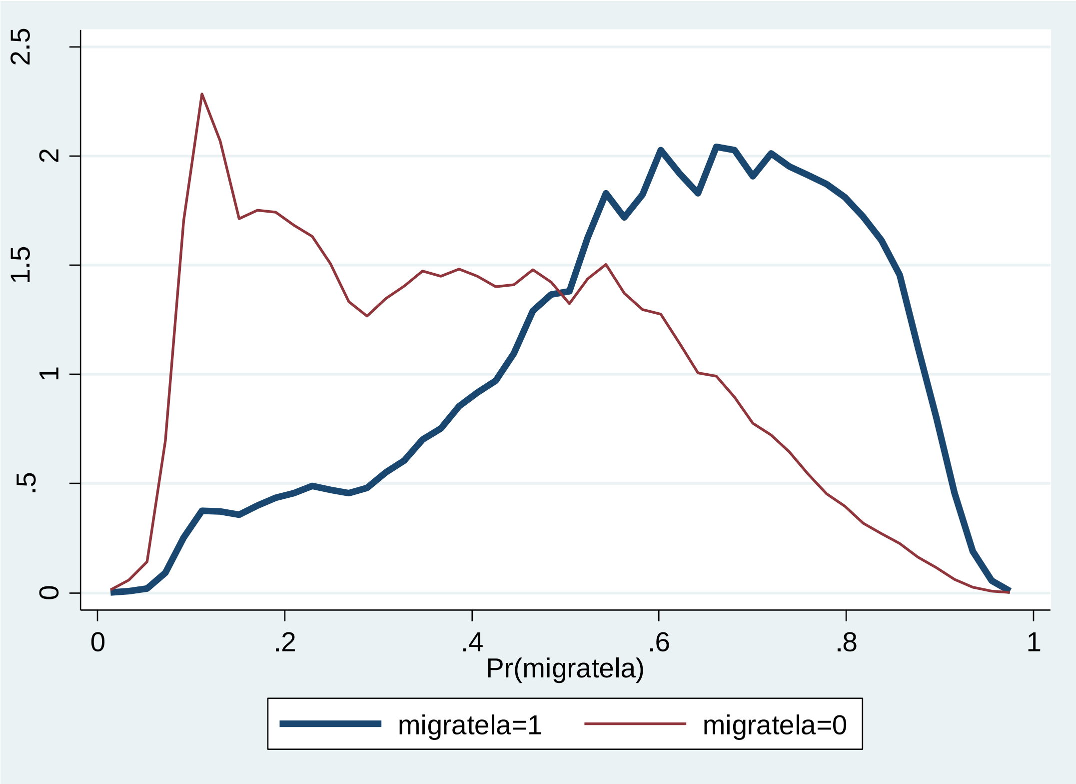
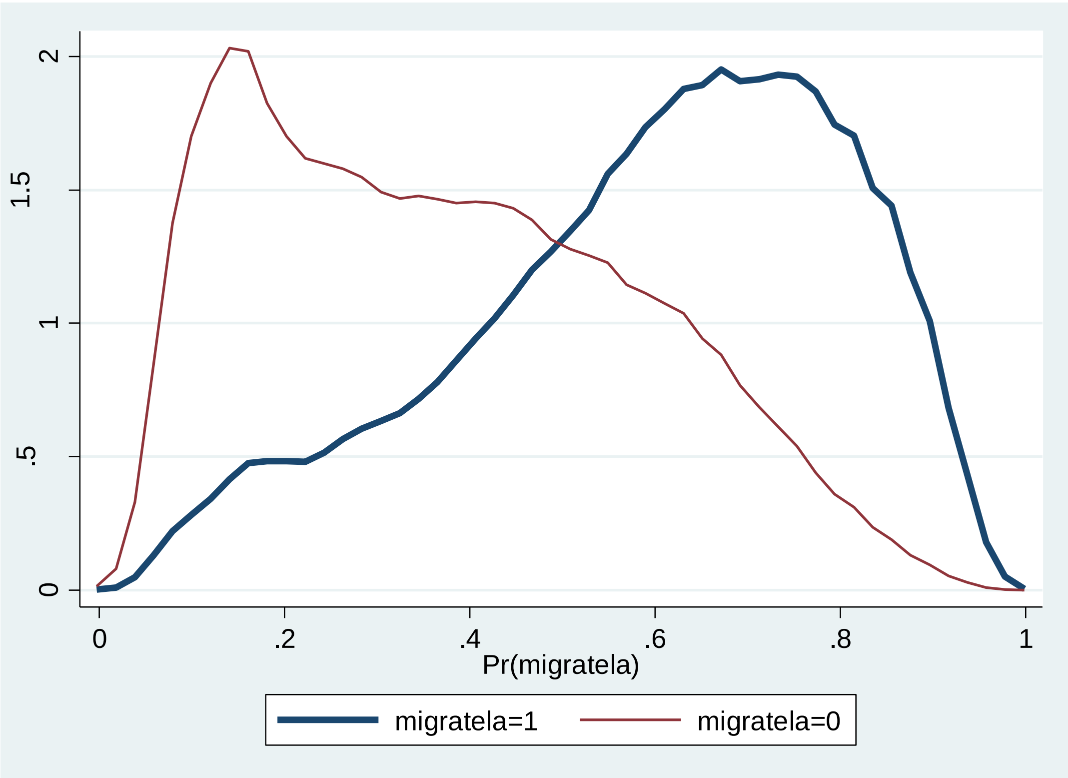


Figure F-4: Density plot to test for common support – Probit PSM Model 4



Appendix G – Non-Response Analysis of the DLHE

Table G-1: Analysis of Item Non-Response in Unemployment Variables with Ethnicity within the DLHE Survey

Ethnicity	DLHE– Unemployment - Item Responders (%)	DLHE – Unemployment – Item Non-Responders (%)	Z-Score (diff between responders and non-responders	P-Value
White	82.5	78.9	20.77	0.000
Black	4.5	5.1	-6.73	0.000
Asian	8.0	10.1	-17.18	0.000
Other	3.4	4.3	-11.03	0.000
Unknown	1.7	1.6	1.60	0.055
N	324,711	60,929	-	-

Table G-2: Analysis of Item Non-Response in Unemployment with Social Background within the DLHE Survey

Background	DLHE– Unemployment - Item Responders (%)	DLHE– Unemployment – Item Non-Responders (%)	Z-Score (diff between responders and non-responders	P-Value
Most Advantaged	15.7	23.1	-44.46	0.000
Advantaged	18.2	20.5	-13.46	0.000
Less Advantaged	23.3	24.0	-3.90	0.000
Least Advantaged	20.8	18.1	15.24	0.000
Unknown	22.0	14.4	42.75	0.000
N	324,711	60,929	-	-

Table G-3: Analysis of Item Non-Response in Unemployment with Gender within the DLHE Survey

Gender	DLHE– Unemployment - Item Responders (%)	DLHE– Unemployment – Item Non-Responders (%)	Z-Score (diff between responders and non-responders	P-Value
Male	41.2	45.8	-21.12	0.000
Female	58.8	54.2	21.12	0.000
N	324,711	60,929	-	-

Table G-4: Analysis of Item Non-Response in Salary with Ethnicity within the DLHE Survey

Ethnicity	DLHE - Salary - Item Responders (%)	DLHE - Salary – Item Non-Responders (%)	Z-Score (diff between responders and non-responders	P-Value
White	84.5	82.7	12.71	0.000
Black	3.9	4.1	-3.47	0.000
Asian	7.0	8.1	-11.03	0.000
Other	3.1	3.3	-3.71	0.000
Unknown	1.6	1.8	-3.66	0.000
N	171,581	112,216	-	-

Table G-5: Analysis of Item Non-Response in Salary with Social Background within the DLHE Survey

Background	DLHE - Salary - Item Responders (%)	DLHE - Salary – Item Non-Responders (%)	Z-Score (diff between responders and non-responders	P-Value
Most Advantaged	14.9	16.3	-10.16	0.000
Advantaged	17.6	18.8	-7.85	0.000
Less Advantaged	23.6	22.6	5.92	0.000
Least Advantaged	21.1	19.9	7.92	0.000
Unknown	22.8	22.4	2.37	0.000
N	171,581	112,216	-	-

Table G-6: Analysis of Item Non-Response in Salary with Gender within the DLHE Survey

Gender	DLHE - Salary - Item Responders (%)	DLHE - Salary – Item Non-Responders (%)	Z-Score (diff between responders and non-responders	P-Value
Male	39.6	41.3	-9.03	0.000
Female	60.4	58.7	9.03	0.000