Internal and cross-border migration in the United Kingdom: harmonising, estimating and analysing a decade of flow data

Nikolas Michan Lomax

Submitted in accordance with the requirements for the degree of Doctor of Philosophy

The University of Leeds
School of Geography

October 2013

Peer Reviewed Paper

Part of the work contained in Chapter 3, Chapter 4 and Chapter 5 has been published in:

Lomax, N.**,** Norman, P., Rees, P. and Stillwell, J. (2013) Subnational migration in the United Kingdom: producing a consistent time series using a combination of available data and estimates. *Journal of Population Research,* 30(3): 265-288

Working Paper

Part of the work contained in Chapter 3 and Chapter 4 has been reproduced in:

Lomax, N., Stillwell, J., Norman, P. and Rees, P. (2011) Data sources and estimation methods for sub-national migration in the United Kingdom. *Working paper 11/4,* School of Geography, University of Leeds.

I declare that the research for these publications was solely my own work, and the contribution of my co-authors was editorial and advisory.

Acknowledgements

This research would not have been possible without the help and support of my supervisors John Stillwell and Paul Norman and without the tireless effort of Phil Rees. All three have been hugely generous with their time, innovative in their recommendations and comprehensive in their feedback on the substantial volulme of material which I produced during the process of writing this thesis, for which I will be eternally grateful.

Thanks also to my research support group members Pete Boden and Paul Vickers, and to colleagues from Southampton: James Raymer, Jacub Bijack and George Disney, whose feedback on my work has been invaluable. To Pete Large, Michelle Bowen and Alistair Davies at my CASE partner the Office for National Statistics, Esta Clarke, Ganka Mueller and Bart Jaron at National Records of Scotland and Jos Ijplaar and David Marshall at the Northern Ireland Statistics and Research Agency for providing information and data.

Thank you to my colleagues and friends, both inside and outside the School of Geography of whom there are too many to name, for their support but especially in providing much needed distraction and beer along the way. Thanks also to the research community which is CSAP, for providing an excellent environment in which to work and helping me to share my research at a number of conferences around the world.

A final big thank you to my family for encouraging me at every step of the way and to Claire, who has listened to every intricate detail of this research process without complaint, has spurred me on to finish this thesis and to whom I owe my sanity.

This research is supported by an Economic and Social Research Council CASE studentship (ref. ES/H03997X/1ES/H03997X/1) awarded in partnership with the Office for National Statistics.

Abstract

Migration is a process which is difficult to measure accurately due to an absence of any mandatory system for registering a move to, from or within the United Kingdom (UK). This problem is exacerbated by inconsistency in statistical reporting, as three national statistics agencies produce migration statistics for the four countries of the UK: the Office for National Statistics in England and Wales, the National Records of Scotland and the Northern Ireland Statistics and Research Agency. They draw upon different data sources, use different estimation methods and produce different outputs. What results is a data landscape which is not consistent and is missing some key information, notably migration where a person moves between local authorities which are located in different UK countries.

This thesis makes the case for a consistent methodology to be employed in estimating migration in the UK. A key contribution is made through the harmonisation of available data and the use of an iterative proportional fitting method to estimate the missing flow data. The resulting output is a consistent UK wide dataset of migration at the local authority level for the first decade of the 21st century, disaggregated by age and sex.

Analysis of the dataset reveals a decline in the longstanding pattern of counterurbanisation which has characterised UK migration for the past 50 years, driven to a large extent by the fall in the intensity of migration from urban to rural areas. Net migration gain in the north from the south is reversed mid-decade, owing largely to an increase in moves from urban north to urban south. Internal migration rates are highest in 2006/07 at the peak of the economic boom, then decline as the financial crisis takes hold. The distance that people migrate falls between 2001/02 and 2010/11.

Table of Contents

List of Tables

List of Figures

Chapter 1
Introduction

The need for accurate and timely population estimates for local authority areas in the United Kingdom (UK) is emphasised by the UK Statistics Authority (UKSA 2009, p. 1) statement that *"these estimates are at the heart of decisions around policy development, resource allocation and service delivery, both nationally and locally".* The substantial impact that migration has on the composition and size of these populations is well summarised by Raymer and Smith (2010, p.705) who stress that there is a need for improving the evidence base for both public policy making and academic research that endeavours to provide better understanding of current migration intensities and patterns since "*migration is currently, and increasingly, the major factor contributing to population change in developed countries".*

Whilst it is well documented that both internal and international migration contribute to local, regional and national population dynamics in different magnitudes, they are both processes that are extremely difficult to measure accurately (ONS 2011a; GROS 2010b; NISRA 2007; UKSA 2009) and there is general consensus that the migration statistics that are collected by the national statistical agencies in the UK need to be improved (National Statistics 2006). A damning report published by the House of Commons Treasury Select Committee (2008, p.47), which heard evidence from a wide selection of experts and public sector bodies, concludes that UK international migration statistics are based on a survey *"designed to provide data primarily for tourism and business travel purposes"* which is not suitable for measuring migration, while they deem that the *"current methods of estimating internal migration are unsatisfactory and lead to decisions on the allocation of funding to Local Authorities being based on inadequate information".* These findings of the Treasury Select Committee led to the formation of the Migration Statistics Improvement Program (MSIP), a cross-agency collaboration, headed by ONS, tasked with improving migration statistics. This collaboration brought about a number of improvements to migration statistics in the UK, especially to international immigration which is the focus of much critical attention from the current administration, press and public. In comparison, estimates of internal migration have received far less attention in the popular press and from policy makers. However, internal migration within the UK has long been the focus of academic researcher's attention and recent examples include Fielding (2012), Champion (2005) and Dennett and Stillwell (2010). A notable absence from much academic literature is analysis of cross-border migration (between the four countries of the UK), which is often excluded due to inadequate data being available. The estimation and analysis of these cross border migrations is one of the key contributions of this thesis.

It is within the context of this drive to improve the evidence base, estimation methods and data output for migration statistics that this thesis is located. Various data sources are available which report migration but they are used differently across the UK, with no single consistent methodology being employed for the estimation of migration at the subnational level. Furthermore, there is currently uncertainty over the future of the census which has given rise to an ongoing debate around suitable alternative data sources, driven by a consensus for the need to produce comprehensive and accurate population statistics (House of Commons Science and Technology Committee 2013). The 'Beyond 2011' program, led by ONS, is tasked with investigating the potential options moving forward, with one of the key underlying principles being *"the UK harmonisation of statistical output as far as possible where there is clear and substantial user need"* (ONS 2012a, p.1). The research presented in this thesis is well timed to contribute to this debate.

With these shortcomings in data availability and consistency in mind, the question of how best to develop a set of migration flows between local authority areas is at the heart of this thesis. This requires an understanding of what datasets are available and what estimation methods can be used. These flows between sub-national local authority areas are termed internal migration, where a person crosses an administrative boundary, either within the same country or between one of the four countries of the UK. The geographical scale of this analysis is consistent with much of the literature on internal migration where data availability dictates the level of spatial disaggregation (Stillwell and Hussain 2010; Dennett and Stillwell 2008; Kalogirou 2005). Additionally, analysis of international migration in to and out of these local authority areas is undertaken in this thesis.

The estimation and analysis of migration patterns presented in this thesis focuses on 2001/02 to 2010/11, a time period during which a number of substantial socio-economic changes occurred in the UK: expansion of the European Union saw a large increase in the number of international immigrants entering the UK from 2003/04 onwards; a period of economic growth gave way to the deepest economic recession since the 1920s in 2006/07; and a coalition government took power in May 2010, imposing an austerity programme on public spending. In this context, the thesis examines the trends in migration that are apparent during the first decade of the 21st Century, both in aggregate terms and disaggregated by age and sex. Given that little attention has previously been given to cross-border migration, the magnitude of these flows can be assessed. The effect of the economic downturn on migration patterns needs to be considered, which gives rise to the question of policy implications for the observed migration dynamics in the 2000s, especially given that respective censuses report the UK population has increased steadily from 59 million in 2001 to 63 million in 2011 and public spending is being reduced as part of the austerity measures. The next section outlines the aim and objectives which will guide the research presented in this thesis.

1.1 Aims and objectives

The aim of this thesis is to produce a comprehensive and consistent database of migration for the entire UK at the subnational level, disaggregated by origin, destination, age and sex which can be used subsequently to analyse migration intensities and patterns and monitor migration change. To achieve this aim, six objectives are proposed:

1. to highlight the need for consistent UK wide subnational migration statistics and review the substantial literature that deals with determinants of migration, data estimation and visualisation;
2. to comprehensively review the data and methods used in the estimation of subnational migration in the UK for each of the four home nations, highlighting where inconsistencies exist and data are missing;
3. to combine and harmonise the available subnational migration data and estimate the missing information;
4. to build on existing techniques to develop a framework and set of measures for effectively presenting the results from the estimated migration database;
5. to analyse the trends and patterns that exist in the UK wide subnational migration system between 2001/02 and 2010/11, using additional data where appropriate; and

6. to provide a discussion of the work in the context of on-going methodological improvements and a changing data landscape within the UK.

1.2 Thesis structure

This thesis builds up to the analysis of a full migration dataset by origin, destination, age and sex between 2001/02 and 2010/11, with a logical progression through methodology, to results and implications. In this section, the way in which the six objectives are addressed in successive chapters of the thesis is briefly described.

Objective 1: to highlight the need for consistent UK wide subnational migration statistics and review the substantial literature that deals with determinants of migration, data estimation and visualisation;

This objective is addressed in Chapter 2, where an overview of the need for comprehensive migration statistics in the context of resource allocation at the local authority level is identified. The problem of inconsistent data reporting due to the involvement of three national statistical agencies (NSAs) who oversee the four countries of the UK is highlighted. A review summarises the substantial body of literature which is concerned with the determinants of migration propensity (which helps to explain patterns seen in the data later in the thesis) and the visualisation and analysis of large migration datasets, which is an essential step in interrogating the migration dataset in subsequent chapters. Finally, literature which deals with the estimation of missing migration data and the application of models is assessed, which aids in the development of the estimation strategy used in this thesis.

Objective 2: to comprehensively review the data and methods used in the estimation of subnational migration in the UK for each of the four home nations, highlighting where inconsistencies exist and data are missing

Having established the need for consistent migration statistics, Chapter 3 provides a review of migration statistics methodology and data availability in the UK, which is gleaned from methodology documents and evidence gathered during meetings with statisticians at the three NSAs. This serves to highlight what is required in the estimation of missing data for the inter-censal time series throughout the 2000s. The structure and coverage of the available data is assessed while at the same time a number of demographic concepts are defined. This provides a theoretical base to the estimation carried out in this thesis.

Objective 3: to combine and harmonise the available subnational migration data and estimate the missing information

Chapter 4 pulls together the available data by origin and destination and, using the framework of a comprehensive matrix of interaction flows, sets out this information which enables the gaps where data are missing to be identified. The available data are harmonised so that information on migration in England, Wales, Scotland and Northern Ireland can be compared. Use of the iterative proportional fitting routine (IPF) is justified and the method is used to fill the gaps in the interaction matrix by combining data that are available for each mid-year between 2001/02 and 2010/11 and utilising the directional structure of the flow matrix collected by the 2001 Census. The estimated origin- destination interaction flow data are then disaggregated by age and sex in Chapter 7 using a similar IPF approach. In Chapter 7, the estimation methodology is extended by clustering local authorities based on their age profiles and smoothing migration schedules, thus contributing to the understanding of how the data are structured.

Objective 4: to build on existing techniques to develop a framework and set of measures for effectively presenting the results from the estimated migration database

In Chapter 5, a number of measures of migration are identified from the literature along with a set of frameworks which have previously been used to interpret and visualise large and complex interaction datasets. Measures of migration distance, migration efficiency and spatial inequality are introduced, and the local authority district time series data are summarised using a broad classification into north-south and urban-rural geographies. In Chapter 6, a city region framework is used to reduce the burden of information being presented when connections between origins and destinations are considered. To ensure that the city region framework is appropriate, a spatial interaction model is used to compare the local authority and city region aggregated data. These measures and frameworks are used again in Chapter 8 where the inter-district flow data are disaggregated by age and sex.

Objective 5: to analyse the trends and patterns that exist in the UK wide subnational migration system between 2001/02 and 2010/11, using additional data where appropriate

Using the measures and frameworks identified in Objective 4, the UK wide migration database is analysed for the period 2001/02 to 2010/11. In Chapter 5, an overview assessment of general UK wide trends and net flows/rates at local authority level is presented. Chapter 6 takes this analysis a stage further by analysing the connection between origins and destinations within a city region framework. In Chapter 8, the full origin, destination, age and sex dataset is analysed using a life course approach, where the five-year of age estimates are aggregated into various stages of the life cycle. Here, data on student migration and armed forces migration are introduced to explain certain patterns in the migration dataset.

Objective 6: to provide a discussion of the work in the context of on-going methodological improvements and a changing data landscape within the UK

As the whole thesis is geared towards an improvement of migration estimation methodology and the harmonisation of data in the UK, this final objective is considered to a varying degree in a number of chapters. This final objective is first covered in Chapter 3 where methodological improvements which are underway or being considered by the NSAs are discussed. Thereafter, in Chapter 4 the best available data are used in the estimation of the aggregate matrix which includes changes to data for Scotland in 2006/07, while data availability by age and sex is considered in Chapter 7. In Chapter 9, a discussion of the thesis findings is undertaken, and consideration of changes in data and methods moving beyond the 2011 Census form an integral part of this discussion. It is this final chapter which offers some overall conclusions from the project and reflects on the extent to which the six objectives presented in this chapter have successfully been achieved.

1.3 Summary

Considerable data gaps and inconsistencies are resolved using the interaction matrix and IPF routine employed in this thesis, and analysis of the complete dataset, estimated for 2001/02 to 2010/11, reveals that a number of changes in the magnitude and structure of migration have taken place over the 2000s. The approach employed in this thesis, which deals with the whole of the UK in a consistent manner, demonstrates these changes in a way that is often overlooked in migration research, i.e. where the UK is treated as a disjointed or unconnected spatial system. The next chapter provides the context for the research carried out in this thesis by outlining in more detail the problem that exists with migration data in the UK, the proposed solution and a review of previous studies which tackle the analysis, estimation and presentation of migration data.

Chapter 2
Understanding and estimating migration: an outline and review

2.1 Introduction

The aim of this thesis is to produce a comprehensive and consistent UK wide database of migration by combining available data and estimating the gaps that exist, and an analysis of migration trends over the past decade is undertaken using this new, complete dataset. This chapter puts the research into context: first, by setting out the problem that exists with migration statistics in the UK; second, by outlining the framework that will be used to provide a solution to this problem; and third, by examining a selection of studies from the substantial body of literature that deals with migration patterns, propensities and estimation. The last of these themes provides a solid theoretical background for the estimation of missing data and analysis of patterns that is undertaken in this thesis, given that both are fields which have received extensive attention in the literature.

A desire and need to understand patterns of migration spans the academic, public and private sector. From an academic point of view, the process of migration underpins social phenomena studied by geographers (Dorling and Rees 2003; Champion *et al.* 2013), sociologists (Berry 2000), epidemiologists (Carballo *et al.* 1998; Evans 1987), environmental scientists (Reuveny 2007) and researchers in any other discipline where the distribution of people or the composition of the population are involved. In the public sector, the formulation of most policy decisions, ranging from resource allocation, such as public health spending, to social cohesion, which encompasses education, housing and a host of other factors, is dependent on a solid evidence base which reports the size and composition of local populations, which are underpinned by the movement of people. In the private sector, business decisions are based on the location of people (and by extension their migration decisions), whether it be choosing the site of a new supermarket or distribution centre to maximise revenue or positioning an office or factory in an accessible area for the workforce. These patterns are, however, not easy to measure and the following section provides an overview of the problem which exists in the UK in relation to the consistency and availability of migration statistics.

2.2 Measuring migration in the UK: difficulties and inconsistencies

Migration is an integral component of population change alongside the natural change components of births and deaths but is the most difficult demographic component to measure or estimate (ONS 2011a). Whilst death is an event that occurs to a person only once and a birth is experienced by mothers only one to three times on average, a person can experience any number of migrations during a lifetime. The problem the UK faces with regard to producing migration statistics is summarised by the UKSA (2009, p.9):

> *"whereas data on the number of births and deaths are well documented and reasonably predictable, the movements of people into and out of the country and between areas are less so. So as well as being the larger component of population change, internal and international estimates of migration are more difficult to estimate with confidence".*

There is currently no compulsory system for registering migration that occurs both internally (a move within the UK) or internationally (a move between the UK and overseas). The data used and estimates produced for migration in the UK are covered in detail in the next chapter but the fundamental problem is that migration estimates are derived from a small sample (in the case of international migration) or depend on individuals voluntarily reporting a change of residential address to their doctor (for all subnational internal migration).

This problem with the recording of migration events is exacerbated by the fact that three different national statistical agencies (NSAs) estimate and administer migration data for the four constituent countries of the UK: the Office for National Statistics (ONS) in England and Wales; the National Records of Scotland (NRS); and the Northern Ireland Statistics and Research Agency (NISRA). Differences between the methods used and data produced by the three NSAs means that there is no single consistent methodology for the production of migration statistics in the UK, especially when moves at the subnational level are considered. These inconsistencies in data and methods are covered in detail in Chapter 3, while the geography of the UK is summarised in Figure 2.1, with the coloured outlines representing the four constituent countries (yellow for England, red for Wales, green for Scotland and blue for Northern Ireland).

Figure 2.1: The four countries of the UK and 406 local authority districts

2.2.1 The UK subnational specification

The boundaries of the subnational administrative areas of the UK which will be used in thesis can be seen in Figure 2.1 and are represented by grey lines. In England, this administrative subnational geography comprises 326 local government areas which include the City of London and 32 London Boroughs, 36 Metropolitan Districts, 56

Unitary Authorities (UAs) and 201 Non-Metropolitan Districts (which may variously be referred to as Shire Districts, Borough Councils or District Councils). Wales comprises 22 UAs, Scotland contains 32 Council Areas (CAs) and Northern Ireland is made up of 26 Local Government Districts (LGDs). For simplicity, in this thesis these administrative geographies will be referred to as local authority districts (LADs). These administrative geographical boundaries are the subject of periodical change, so the 406 LADs used consistently throughout this thesis are the most up to date boundaries (the last major change occurred in 2009). Creating consistency between the beginning (2001/02) and end of the time series (2010/11) requires some adjustment of the data, a process which is dealt with in Chapter 4.

UK-wide subnational migration incorporates four types of migration flows (from a person's origin residential location, to their destination) between/within the LADs identified in Figure 2.1:

(i) intra-LAD flows that occur within each of the LADs;
(ii) inter-LAD flows within each constituent country which can be referred to as 'internal' migration;
(iii) inter-LAD flows between each constituent country which can be referred to as 'cross-border' flows; and
(iv) flows into each LAD in the UK from the 'rest of the world' and out of each LAD to the 'rest of the world' which can be referred to as 'international immigration' and 'international emigration' flows

It is only flows ii, iii and iv that are considered in this thesis, as it is migration across a LAD boundary that impacts on resource allocation and policy decisions at national, regional and local level. It is the responsibility of the NSAs in each country to provide mid-year population estimates (MYEs) at the LAD scale and therefore it is the inter-LAD flows that are particularly relevant, rather than the intra-LAD flows (where, even though a migration has occurred, the financial or other service allocation is not impacted). The MYEs are very important because they inform resource allocation and policy decisions and therefore considerable importance is attached to the natural change and migration components that are fed into the cohort component model used to produce the MYEs. The role of the MYEs and the migration component that informs them is considered in more detail in the next chapter.

Assessment of and improvement to the estimation of the fourth of these flows, moves between each LAD and the ‘rest of the world’ is currently being undertaken by George Disney at the University of Southampton as part of an ONS sponsored doctoral thesis. With a whole thesis being dedicated to improving the overseas component, the focus is on subnational internal and cross-border migration in this piece of work. International migration data do form an integral part of the analysis contained within Chapter 5, but most of the data are included ‘as is’ with no adjustment being made. The exception is international migration to and from Scotland which was reported as a combined flow alongside cross-border migration up to 2006/07; the method used to split the international and cross-border migration flow is covered in Chapter 4. Estimation of international and internal migration in the UK has received a good deal of attention in the past, but internal cross-border flows have received much less consideration. Therefore, a key contribution made by this thesis is in estimating these subnational cross-border flows.

A comprehensive overview of LAD level migration can be obtained from the 2001 Census Special Migration Statistics (SMS), which provide an indication of the magnitude of each of the different types of flow (internal, cross-border and international immigration) occurring in the UK system in the year before the 2001 Census. Flows at the national level (an aggregation of all moves occurring at LAD level) from the Special Migration Statistics (SMS) are shown in Table 2.1. The flows reported in Table 2.1 provide an important benchmark for the magnitude of each type of flow examined in this thesis, as the census offers a once in a decade opportunity to analyse migration data which is consistent for the whole of the UK (in terms of the methodology used and outputs produced). Estimating and analysing UK migration patterns in a consistent way is a key theme running through this thesis, and the 2001 Census is used to benchmark much of the estimation methodology outlined in Chapter 4 and Chapter 7.

The diagonal elements represent the flows between subnational LADs within each of the home nations (internal migration, excluding within LAD moves). Table 2.1 shows that internal migration comprises 80 per cent of the 3.1 million migrants that cross a LAD boundary, whose origins and destinations were both stated in the 2001 Census returns. The off-diagonal elements are the flows between the home nations (cross-border flows) and the immigration flows from the rest of the world which account for a further 6.5 per cent and 13 per cent respectively of the 3.1 million migrants. In addition, the 2001 Census recorded a further 467,000 migrants at destinations within the UK whose origins were not stated on the census forms in 2001 so it is not clear whether these are intra-LAD moves, internal migration, cross-border migration or international immigration flows. Moves between LADs in England account for by far the largest proportion of moves in 2000/01, where 2.3 million migrants (74.3 per cent of all migration) crossed a LAD boundary. To put the size of these flows into perspective, the intra-LAD flows not reported in Table 2.1 (and not reported in this thesis) comprise a further 3.5 million migrants – 62 per cent of all migration that occurred in 2000/01 (although this proportion varies: 60 per cent in England, 83 per cent in Wales, 76 per cent in Scotland and 70 per cent in Northern Ireland). International emigration flows are not reported in the 2001 Census SMS as it is not possible for the census survey to capture people who are not in the country on census day.

Table 2.1: UK migration flows, 2000/01

	Destinations					
Origins	England	Wales	Scotland	Northern Ireland	Rest of the world	**Total outflows**
England	2,315,824	48,248	43,675	7,899	-	**2,415,646**
Wales	42,614	40,835	1,546	325	-	**85,320**
Scotland	42,831	1,396	113,824	2,633	-	**160,684**
Northern Ireland	8,812	360	2,602	37,437	-	**49,211**
Rest of the world	360,531	9,916	28,868	7,461	-	**406,776**
Total inflows	**2,770,612**	**100,755**	**190,515**	**55,755**	**-**	**3,117,637**
No usual address	400,368	19,721	36,562	10,401	-	**467,052**

Source: 2001 Census (Special Migration Statistics Table MG101)

Having outlined problematical elements of migration statistics in the UK and highlighted the three types of migration that will be dealt with at the subnational level in the remainder of this thesis, the following section outlines the proposed solution for collating and harmonising the data provided by the three NSAs.

2.3 The solution: producing a consistent UK dataset

Rees and Willekens (1981; 1986) define the migration estimation problem as the ‘three face (3F)’ problem, where the three visible faces of a cube constitute the available data while the unknown data (which needs to be estimated) sits within the cube (behind the faces). Figure 2.2 shows a cube which is based on this concept of visualising data, and is adapted and developed from a diagram presented in Rees and Willekens (1986, p.53). Figure 2.2, which will be termed the ‘Migration Cube’, serves to illustrate the processes that are undertaken in this thesis, which culminates in an estimated UK migration matrix by origin (O), destination (D), age (A) and sex (S) to produce the full ODAS array of data (shaded grey in Figure 2.2, behind the three faces of the Migration Cube). This dataset will be estimated for the ten year period mid-2001/02 to mid-2010/11, which covers and extends slightly beyond the inter-censal years (between census day 2001 and census day 2011).

The face of the Migration Cube labelled OD is an interaction matrix of aggregate flows between the 406 LADs in the UK, plus a flow to and from overseas (labelled international). Just filling in this face of the Migration Cube is a challenge in itself (with internal Northern Ireland and all cross-border cells missing): the methodology for collating and estimating the missing data for the OD face is presented in Chapter 4 (alongside a detailed account of the OD face in Figure 4.1) while analyses of these aggregate results is presented in Chapter 5 and Chapter 6. The faces of the Migration Cube labelled OAS and DAS represent total out and total in migration by age and sex respectively. These faces have been adapted from the specification set out by Rees and Willekens (1986) with the addition of the sex variable in combination with age (exclusion of sex would require two cubes for representation, one for males and one for females). Estimation of the faces of the Migration Cube labelled OAS and DAS, along with estimation of the full ODAS array is covered in Chapter 7, while analysis of this data is carried out in Chapter 8. In reality, data availability in this study dictates that for a part of the OD face (the portion that covers within Northern Ireland and UK cross-border migration) and for the OAS and DAS faces of the Migration Cube, the problem is one where there is only information for the edges (total out-migration and total in-migration for each LAD, with no origin-destination linkage), defined as the ‘three edge or 3E problem’ by Rees and Willekens (1981). This problem, and the iterative proportional fitting routine used to estimate the missing values is explained in detail within the two methodology chapters (Chapter 4 and Chapter 7).

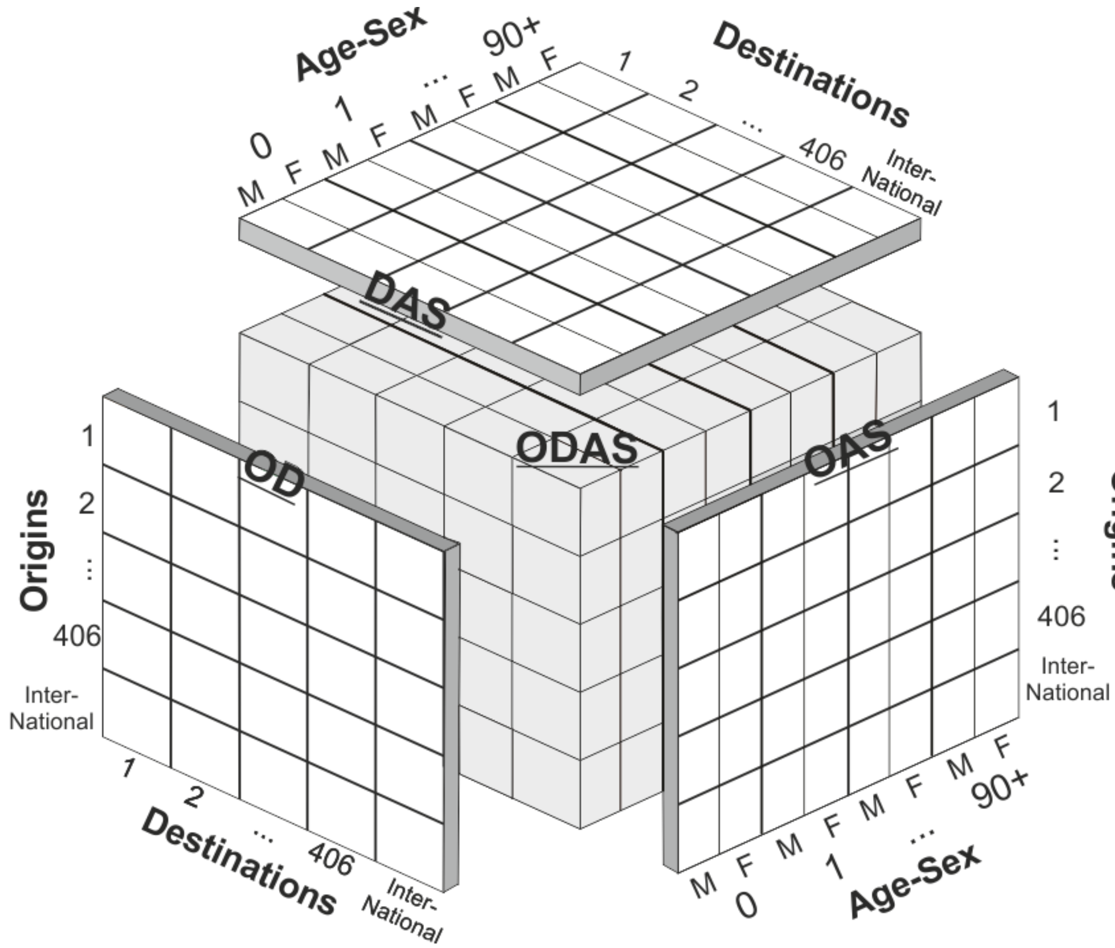


Figure 2.2: Migration Cube representing the combinations of origin (O), destination (D), age (A) and sex (S) data that will be explored in the thesis

Source: adapted from Rees and Willekens (1986)

The decisions made in the formulation of the methodology to estimate the OD and ODAS tables in Chapter 4 and Chapter 7 of this thesis are informed by the large body of literature pertaining to the estimation of inadequate data whilst the way the estimated matrix is analysed (in Chapter 5, Chapter 6 and Chapter 8) is informed by literature pertaining to drivers of migration propensity and measures/frameworks used to distinguish patterns. This literature is the focus of the remainder of the chapter, where first, factors which influence migration propensities are addressed, followed by a review of the ways in which migration patterns are measured and analysed. Finally, various ways of estimating missing or inadequate data are considered.

2.4 Propensities and patterns: why do people migrate and where do they go?

Understanding the factors which drive the movement of people, the spatial patterns that these migrants exhibit and the impact that migrants have on the size and composition of local populations is an area of demography that has received a huge amount of attention in the literature. In two seminal papers, Ravenstein (1885; 1889) concludes that there exist a number of 'laws of migration,' which have been carried forward to form the basis of much contemporary migration research (Grigg 1977). The laws identified by Ravenstein are familiar to most geographers and demographers: most migration occurs over short distances, the majority of migration is driven by economic considerations and there is a preference for movement from agricultural areas of low employment to urban centres where there is high employment. Attempting to review migration literature from Ravenstein to the present day would be a very interesting piece of work but would be a huge undertaking, well outside the scope of this chapter. The focus of this review therefore switches to more contemporary literature, but it is worth considering the assessment of Maier and Weiss (1991, p.17) who stress that *"despite the many different approaches to the analysis of migration there is consensus about one point: all researchers seem to agree that people or households migrate to improve their situation, or more precisely, to be better off in the new location in the future than they would be in the old".* This is a principle that informs variable selection in migration modelling studies, of which a selection is considered in Section 2.5.

In an influential piece of work, Lee (1966), outlines a framework for migration which involves four factors: (1) factors associated with origins; (2) factors associated with destinations; (3) intervening obstacles (the link between origin and destination); and (4) personal factors. Lee (1966, p.51) argues that the positive and negative factors at origin and destination are assessed by a migrant (the value attributed to these factors are different for each individual and varies at different stages of the life course however) and *"while migration may result from a comparison of factors at origin and destination, a simple comparison of +'s and –'s do not decide the act of migration. The balance in favour of the move must be enough to overcome the initial inertia which always exists".* The intervening obstacles can be a variety of factors including distance, physical boundaries or immigration laws. The factors associated with origins and destinations, as well as intervening opportunities are discussed further in the next section, while 'personal factors' are discussed in Section 2.4.2. A substantial review of the determinants of migration is presented by Champion *et al.* (1998) who focus on these various factors under the headings of demographic, social and cultural, economic, housing and environmental variables. Champion *et al.* (1998, p.63) stress that *"in most situations these factors play simultaneous roles of differing significances, depending on the context in space and time".* With this in mind, the following sections summarise some of the many factors that have an influence on migration propensity.

2.4.1 Origins, destinations and intervening factors

Economic variables are intrinsically linked with other factors (the availability of jobs and housing for example) but general measures of prosperity can be associated with the level of migration. GDP per capita and the number of new business registrations are two such measures identified by Stillwell (2005) and the first of these is used in Chapter 5 of this thesis to assess migration levels at the national level. Van Der Gaag and van Wissen (2008) address the relationship between internal migration and a suite of economic determinants including general business cycle indicators (GDP per capita, unemployment), financial variables (inflation, interest rates) and structural labour market developments (female labour force participation, ageing of the labour force) in five EU countries. They find that GDP per capita has the most substantial effect on migration rates, interest rates and to a lesser extent unemployment have an impact, *"while structural characteristics of the labour market… have a strong but complicated effect on internal migration levels"* where increased ageing and increased female participation in the workforce have a strong negative impact on internal migration rates (p. 220). In times of recession, migration activity can decrease: in the case of the 1979-83 recession in the UK this is reported by Stillwell *et al.* (1992) and more recently the negative impact of the 2008/09 recession on UK migration rate is reported by Campos *et al.* (2011). In terms of the national business cycle, Milne (1993), in the case of Canada, identifies the general pattern as one where potential economic growth declines, the overall migration rate falls. At the regional level (provinces in Canada) Milne found that the relationship between economic events and migration was even stronger than at the national level. This impact of the recession and the general health of the economy is particularly relevant for the estimated migration dataset presented in this thesis, as the early years (2001/02 to 2006/07) represent a time of growing prosperity, while 2007/08 represents the beginning of the recession from which the UK is still recovering (at

October 2013). Two key economic indicators, the availability of jobs and house prices, are considered next.

The availability of jobs and wage differences are two factors which have been found to influence migration at the regional level. Rees *et al.* (1996) highlight that moving for employment reasons is a key driver for cross-region migration. McCormick (1997), in a regional study between 1970 and 1990, find that employment growth in the South West and East Anglia were consistently 10-15 per cent above the national average which prompted persistent in-migration from other regions for non-manual workers, especially from Scotland, North, North West, Yorkshire and the West Midlands which experienced negative employment growth. They find that the manual labour market, by contrast, is spatially inflexible. Thomas (1993) finds evidence of preference for migration to areas that offer higher wages, while the inflexibility of manual workers is attributed to *"little geographic flexibility in manual workers' wages"* by Evans and McCormick (1994, p.298), who suggest that, of the manual workers who migrate between regions, only a small proportion do so for job-related reasons. More recently however, Andrews *et al.* (2011) find that there is no significant pattern between wage differentials and migration at the regional level. Using British Household Panel Survey data between 1990 and 2007 they divide migrants into two types: those who move to a region with higher real wage than that of the origin and those who migrate to a region with a lower real wage. Dobson *et al.* (2009) find that international immigration falls while unemployment is rising, but this is only for a limited period. They look at the pattern in the context of the current recession using Worker Registration Scheme data (which records the number of migrants from the A8 countries), finding that the typical early recession trend is emerging: a reduction in immigration of labour and some evidence of a rise in emigration.

House price differentials are investigated by Rabe and Taylor (2009, p.30), who report that *"house price differentials strongly influence migration propensities – relatively high house prices in potential destinations deter migration which is likely to reflect credit constraints. Mortgage holders and social tenants are particularly sensitive to these differentials".* Home owners are directly affected by house price differentials while renters are affected as house price increases drive up rents or exclude them as first time buyers. Thomas (1993) argues that regional house price differences have a substantial influence on the destination choice of migrant retirees, and influence the destination choice of workers who move for job reasons, whereas they have no effect on the destination choice for homemakers, the long term sick and workers who move for other, non-job, related reasons. Cameron and Muellbauer (1998), in a study focusing on the period 1983 to 1995, report that high relative house prices discourage net migration to a region, as does recent experience of negative returns in the local housing market. This is expanded on by Cameron *et al.* (2005), who find that housing market comparisons with contiguous regions are more important than with the average of all regions, given that commuting may be possible from a region with lower house prices to one with higher house prices. They also report that as age increases, house prices matter somewhat more relative to earnings.

The interplay between housing market forces and tenure type, which is identified as a personal characteristic below, is a complex one. If housing markets dictate high prices and economic conditions dictate that credit is hard to acquire then moves between rented and home-ownership tenure types are much more difficult, which may impact on migration. There is also a complex link between the housing market and labour market: Owen and Green (1989, p.125) in a study of labour migration in the 1980s, find that long distance moves are predominantly motivated by the job market, but that this movement is being limited by the housing market. They attribute the fact that house prices in London and the South East are increasing faster than the rest of the UK (as is happening now in October 2013) and the contraction of the rented housing sector to effectively curtail speculative migration to search for work. They suggest that policies promoting owner-occupation ignores *"new and almost insurmountable barriers to migration being erected by the current dynamics of the owner occupied housing market".* In the case of the 1980s, Millington (2000) argues that the problem was further exacerbated by the trend for the retired to stay put in areas that are experiencing rapid house price inflation, postponing their out-migration which had the effect of further restricting the supply of housing.

Environmental factors are identified by Champion *et al.* (2008) as having an impact on migration decisions for people at all ages, but as especially influential at young adult and older ages. They specify that the physical environment (landscape, climate), built environment (type of housing), social environment (access to friends, family, entertainment) and the services environment (access to retailers, entertainment, nursing homes) all have an impact on migration decisions. Much of the literature which deals with migration motivated by environmental factors discusses moves down the urban hierarchy – from more to less urban areas – and this pattern of counterubanisation is discussed in detail alongside the dataset output from this thesis in Chapter 5. The reasons for this preference for a rural environment are covered in detail by Champion *et al.* (1998, p.92-98), and they suggest that all evidence points towards *"a force deep in the English psyche which is driving people to aspire to a rural lifestyle"* (p.96). this is termed the pursuit of a 'rural idyll' which motivates people to migrate out of urban centres by Mingay (1989). Migration for amenity reasons are often reported for elderly and retired people, in the case of England and Wales by Raymer *et al.* (2007) and for moves to rural Scotland by Fleming (2005). In contrast, Rees *et al.* (1996, p.53) report that young people *"leave the smallest places ('boring – nothing to do') on balance and they have the smallest net loss of all ages in the largest cities"* – amenity and environmental factors seem to have a different impact on different age groups.

Finally, policy is reported as having a substantial influence on migration propensity. Internal migration is constrained by job and housing markets, so any policy which promotes (or inhibits) these markets will have an impact on migration propensity. For international migrants, immigration policy is restrictive or permissive, Coleman (2008, p.466) argues, and can "*switch radically as the political pendulum removes one party from government office and installs another".* Coleman (2008) reports that migration policy is often overlooked in analysis of migration that is dominated by economic modelling, a view shared by Hatton (2005) and Mitchell *et al.* (2011) where, in models which include economic variables, both studies find policy to be the dominant influencing factor. A case in point is the accession of Eastern European countries (A8 countries) to the EU in 2004. Labour market access was temporarily restricted to migrants from the A8 countries by all other European countries except for the UK, Ireland and Sweden, and Vagras-Silva (2013b) reports that the 653,000 A8 migrants who entered the UK between 2004 and 2011 represented an unexpectedly high number (based on predictions made in 2003).

This section has revealed that factors at origins, destinations or somewhere in between the two have a complex and substantial influence on migration propensities. It has also touched on the fact that these factors exert different influences on particular groups, based on their personal and demographic characteristics. These characteristics and their impact on an individual's propensity to migrate are considered further in the next section.

2.4.2 Personal characteristics

A number of personal characteristics have been reported to impact on a person's propensity to migrate. Age and sex are the most immediately obvious, and these will be available in the dataset produced in this thesis. Other factors such as ethnicity, health, social mobility, marital status and tenure type are not possible to measure in the data but nonetheless have been found to influence migration decisions. This section looks at these personal characteristics.

As Plane (1993, p.376) reports, age *"has been found to be one of the strongest empirical predictors of geographical mobility behaviour"* and migration propensity is widely reported to vary by age. This variation by age is, however, very stable across space and time, and these regularities are demonstrated in a large number of studies. Rogers *et al.* (1978) introduced a mathematical representation of the age migration schedule which is built on by Rogers and Castro (1981) who note that the regularity of migration by age is no different to the regularity seen in fertility or mortality schedules. These regularities across space and time are reported by Bates and Bracken (1982; 1987), Bracken and Bates (1983) and more recently by Raymer and Rogers (2008). All these studies report that migration intensities peak in the young adult ages; thereafter they decline as age increases until retirement and old age. This age specific pattern of migration is widely reported and analysed using a life course perspective, where certain events happen at different ages which have an impact on migration propensity. Plane and Jurjevich (2009, p.5) summarise that *"the likelihood that an individual will change residences varies dramatically and in broadly predictable ways across the major stages of life"* and these life course stages are used to report and interpret results from the ODAS dataset in Chapter 8.

Stillwell *et al.* (1996), in a study of migration between NUTS2 regions in the UK (Counties in England, groups of LADs in Scotland and Wales and aggregate Northern Ireland) report differences in migration rate at five key stages in the life course: 'family ages' (0-15 and 30-54), the 'ages of leaving home' (16-19), the 'ages where work and careers start' (20-24), the 'age of retirement' (60-69) and the 'elderly ages' (75-85 plus), although various different aggregations of these ages are used in life course analysis. Warnes (1992, p.184) for example, breaks down the life course into 11 stages, which is as detailed as specifying divorce, cohabitation and second marriage at some point between age 27 and 50. The 'family ages' are widely considered as groups who migrate together as a family unit (Dobson and Stillwell 2000; Bushin 2000; Kofman 2004) and where migration propensity is low, owing largely to the desire not to interrupt a child's schooling(Champion 2005). The 'ages of leaving home' show the highest migration rates due largely to moves for higher education (Duke-Williams 2009; Mosca and Wright 2010). Wilson (2010, p.194) argues that student migration is not adequately accounted for in the modelled schedules proposed by Rogers *et al.* (1978) due, in part, to the formulation of the mathematical model being undertaken in the 1980s where higher education student numbers were lower. Wilson (2010) develops a formulation of the Rogers Castro model that includes this student peak. Migration rates rise again at the 'age of retirement' where people move from urban to 'amenity areas' where quality of life is a priority (Raymer *et al.* 2007; Fleming 2005), while migration in the 'elderly ages' is generally associated with poor health and support-related moves, where people move to be close to family or to a nursing home (Glaser and Grundy 1998; Burholt 1999).

The connection between migration at various stages of the life course and the structural factors identified above is made by Champion *et al.* (1998), who specify that higher education and the labour market are driving forces for young adult migrants, family migration is dependent on labour market and environmental factors, while retirement migration is influenced by the housing market and environmental factors. Similarly, Millington (2000) in a modelling exercise that uses the age distribution in inter-county migration flows derived from the National Health Service Central Register (NHSCR, a dataset discussed in the next chapter), finds that *a priori* expectations were confirmed: young migrants fitted with labour market expectations, while elderly migrants responded most to local house price and amenity variables.

The second personal characteristic available in the dataset presented in this thesis is sex. Champion (2005, p.93), in an assessment of 2001 Census data, reports that sex *"is not a major discriminator of migration behaviour except in certain contexts such as the movement of armed forces personnel"*. In order to pick up variation in migration propensity by sex, it needs to be combined with age as, overall, sex ratios do not vary much when aggregate data are reported (Rogers and Castro 1981). Such analysis of age and sex is carried out by Dennett and Stillwell (2010) using 2001 Census data, who find that female migration rates are higher than male rates at ages 18 to 19 and 20-24. They offer explanation for the difference seen in the former as possibly down to the female in a migrating couple being younger than the male while in the latter age group, they cite Faggian *et al.* (2007) who suggest that female graduates are more mobile than males to compensate for gender discrimination in the labour market. A reason for the difference at earlier ages is given by Champion *et al.* (1998) who suggest that females leave the family home one or two years earlier than men and/or marry/cohabit with men who are about two years older. Champion *et al.* (1998, p.69) also report that at the elderly ages, differences emerge as a consequence of men dying earlier than women, so are effectively *"escaping some of the migrations consequent on spousal death"* but that overall (aggregate) differences are small as males and females migrate together over the majority of the life course.

Patterns of migration in the UK have been reported to vary by ethnic group. Finney and Simpson (2008, p.80), using 1991 and 2001 Census data, find that Chinese and Other groups have the highest crude internal migration rates, followed by Black, White and South Asian groups, but that *"the demographic and socio-economic characteristics of those who migrate internally are similar for each ethnic group".* This is elaborated on by Simpson and Finney (2009, p.53) who find that, in terms of percentage impact on the ethnic group's population, the highest proportion of movement into areas with high concentration of White population is by other ethnic groups, suggesting that the predominant driving force for migration decisions are *"common aspirations to improve housing and environmental living conditions away from dense urban areas"* which, they suggest, challenges theories of ethnic minority migration as a mechanism for self-imposed social segregation. Despite the general aspirational driver of ethnic group migration, Finney (2011, p.466) finds that the timings of migration in the life course do vary for young adult ethnic minorities, as *"ethnic groups experience different pathways out of the family home,"* with South Asian groups tending to remain in the family home until married, in comparison to White Britons who tend to live independently from their late teen years. Further differences are identified by Raymer and Giulietti (2009) who, using the Small Area Microdata sample of the 2001 Census, find that while education and employment are important determinants of destination choice, they exert different influences on the

White and ethnic minority population. Education level is an important factor for the White population, while for other ethnic groups, employment status is much more important. The majority of studies of ethnic group migration are carried out based on Census data and most of the studies above report a scarcity of other data which reports migration by ethnic group, a problem echoed by Stillwell (2013), who suggests that as a result there have been few attempts to model ethnic minority migration using explanatory variables.

Health is a factor that influences people's migration decisions. Bentham (1988) reports evidence that migration amongst younger migrants is selective of people with good health, especially over longer distances, while ill health is a motive for migration over shorter distances and among older migrants due to the need to be nearer family or social care. Boyle *et al.* (2002), using 1991 Census data for Scotland, echo these findings and report that overall, young migrants tend to be healthier than young non-migrants, while Boyle *et al.* (2004) confirm that elderly people are more likely to move if they are ill. Norman *et al.* (2005) expand on these findings. Using the census Longitudinal Study between 1971 and 1991 they report patterns of health selective migration, finding that the dominant migration flow is for relatively healthy younger migrants (particularly those aged 20 to 59) moving from more deprived towards less deprived (more economically favourable) areas. The effect of this dominant migration pattern is an increase in health inequality, where there is an increase in ill health and mortality rates in the origin and reduction in rates at the destination. In contrast, people in poor health tend to move from less deprived to more deprived areas which has the effect of exaggerating this inequality. Norman *et al.* (2005, p.2768) also find that moves between deprived areas are made by people in poor health and that a small but important group move from more to less deprived areas, concluding that these migrants *"move to improve their circumstances, perhaps to be cared for by family"*.

Marital status and family type (independent of age) are identified by Champion (2005) as factors that influence migration propensity. Champion, analysing 2001 Census results, finds that single, never married people have the highest migration propensity while the widowed have the lowest, and widowers move the shortest distances, followed by the divorced and separated. In terms of family type, Champion identifies that cohabiting couples with no children have the highest propensity to migrate, while married couples with no children in the household have the lowest (this is attributed to these couples being older families whose children have left home). Lone mothers were found by Champion (2005) to move the shortest distance, while married couples with no children moved the longest distances. In a study comparing nine countries using 1980 or 1981 Census data, Long (1992) finds that the relationship between marital status and residential mobility is similar in Great Britain, the United States and Japan, where the highest rates of migration are made by divorced people, and never-married people make more moves than their married counterparts, a pattern which is most pronounced in mid-life.

Housing tenure has been found to impact on migration propensity. Boyle (1993a), using the 1981 Census, employs a Poisson regression to compare the effect of distance on migration propensity for moves to owner occupied, council housing and private rented housing, reporting that the propensity to migrate varies with tenure type. He finds that amongst owner occupiers, there is a tendency for moves away from urban centres. Many private renters moved into the South East and especially London, but in the North the pattern for private renters was for a move out of metropolitan centres. Council housing tenants are restricted when it comes to moves between LADs (arguing that such moves would usually be special cases) so these moves are generally shorter distance. This trend for shorter distance moves by council tenants is also identified by Gordon and McCormick (1981) and Hughes and McCormick (1991). The role of tenure type and migration in London and the South East is investigated by Hamnett (1991), who finds that private rented tenants were the most mobile, while the majority of migration by council tenants was a move within London. Overall, migration propensity was lowest for owner-occupiers but out-migration from London was higher than for council tenants. As mentioned earlier, for owner-occupiers or those moving from rented to owner-occupied accommodation, house prices and the availability of credit for mortgages have an impact on migration propensity.

Finally, socio-economic status (or the social mobility) of a migrant is found to have an impact on migration propensity. This can be assessed using a number of variables including qualifications held, occupation and industry of employment. As mentioned earlier, McCormick (1997) reports that manual workers are spatially inflexible in relation to non-manual workers. Job-related mobility is reported to be lower for part-time workers and married women by Gordon (1995), the limited mobility of the latter is reportedly due to familial constraints. When it comes to unemployment, Champion *et al.* (1998) find from an analysis of the Longitudinal Study between 1971 and 1981, that the unemployed were near the average in terms of inter-regional migration rates, so lower than most non-manual workers but much higher than manual workers. Dixon (2003, p.199), in a study using Labour Force Survey (LFS) and British Household Panel Survey data between 1992 and 2002, reports that *"people with higher levels of education, and those working in managerial, professional and semi-professional occupations, are much more likely to migrate between regions"* which is attributed to a need to migrate for job related reasons by the highly skilled. Within this group of skilled people, Dixon (2003) found that higher household incomes promote greater mobility.

This section has identified a number of personal factors which exert an influence over migration propensity, the most dominant being age. Identifying these variables and identifying patterns requires robust frameworks for analysis, so the next section briefly addresses some of these measures and frameworks used to analyse and interpret migration patterns that are extracted from large and complex interaction datasets.

2.4.3 Analysing and interpreting patterns

With so many variables being found to have an impact on migration propensity, analysing and interpreting patterns has become a key area in migration research. Tools for measuring migration, frameworks for classifying areas by the type of migration they exhibit and strategies for visualising data are an integral part in the process of understanding migration patterns. In addition, where patterns and trends are identified in a dataset, they become theories and frameworks within which other migration data are examined.

A joint project between the University of Leeds and University of Queensland entitled 'Internal Migration Around the GlobE' (IMAGE) is creating a repository of internal migration datasets from a large number of countries (Stillwell *et al.* 2013), and a 'virtual studio' (the IMAGE Studio) has been created by Daras *et al.* (2013), which provides the tools with which to analyse these datasets using a range of migration indicators. These indicators report a number of quantifiable phenomena, including how efficient migration is as a process for redistributing the population and how equal the spatial distribution of migration is within a system. The IMAGE Studio is also capable of aggregating the data from various countries into different spatial units which allows for analysis of the effect that distance has on migration (Daras *et al.* 2013). The IMAGE Studio is used to calculate a host of indicators for the data presented in Chapter 5 and Chapter 8 of this thesis, and a more detailed assessment of the various indicators is presented alongside this analysis. These indicators are employed in a cross-national comparison of internal migration patterns in the UK and Australia by Stillwell *et al.* (2000; 2001) and by Bell *et al* (2002). The efficiency of migration is a commonly reported measure, being used in an assessment of migration between Standard Metropolitan Labour Areas in Britain by Flowerdew and Salt (1979), in a comparison of different ethnic groups in the UK by Stillwell and Hussain (2010), and assessment of changing migration patterns in the USA by Galle and Williams (1972) and by McHugh and Gober (1992). Inequality within a migration system is measured by Plane and Mulligan (1997) and by Rogers and Raymer (1998), both in the case of the USA.

Clustering of areas based on their geodemographic characteristics has been undertaken at various points by ONS, usually after a census, as is reported by Wallace *et al.* (1995) in the case of 1971, 1981 and 1991. The 1991 results are used by Rees *et al.* (1996) to analyse rates of migration across LADs in the UK. More recently, clustering has been undertaken by Vickers (2006) and Vickers and Rees (2006; 2007) and by Dennett (2010) and Dennett and Stillwell (2010; 2011). Both the Vickers and Dennett classification systems use a collection of demographic, economic, housing and personal variables which group areas which exhibit similar characteristics together; the former is a classification adopted by ONS following the 2001 Census while the latter is specifically created to compare the migration profile of LADs. In addition, various frameworks are employed in migration studies to analyse and theorise patterns of migration: counterurbanisation, reurbanisation, gentrification, studentification, a north-south divide and urban-rural migration are all themes that have been used and re-used in the literature. These are covered in detail in Chapter 5 alongside the analysis of OD results, so a full consideration of these frameworks is reserved until then.

Identifying trends and presenting migration patterns to an audience requires effective strategies for visualising the large datasets that are produced when an origin-destination interaction matrix is used. The visualisation of a large volume of migration data is considered by Rae (2009; 2011), who uses a number of geovisualisation tools to represent migration data from the 2001 Census. Rae (2009, p.177) argues that geovisualisation of migration data *"has been something of a slow starter"* and employs a range of techniques available in standard GIS packages, such as line maps and density raster maps, to represent the intensity of migration between origin and destination. Similarly, Diansheng (2009) develops an interactive system for visualising spatial interaction data in the USA, while Stillwell and Harland (2010) use vector analysis and radar diagrams to visualise pupil and school interactions in Leeds. New and innovative visualisation techniques are constantly developed as access to data (or estimation of data) improves: on-going (as yet unpublished) work by Nikola Sander and Guy Abel at the Vienna Institute for Demography uses Circos, a tool originally designed in the biological sciences, to visualise changes in a genome over time to create a graphic representation of migration flows between countries (Sander 2013, pers. comm.).

This section has briefly summarised strategies for analysing and communicating the results from often large and complex migration datasets once they have been compiled. The next section considers the ways in which these datasets are compiled where there is missing or inadequate data.

2.5 Problems: estimating missing data

The literature presented in previous sections of this chapter has focused on identifying variables that influence migration propensity and on the analysis, reporting and classification of migration patterns and trends. None of these things are possible without a migration dataset and where those data are not available they need to be estimated. It is this estimation which is the focus of this section. Often the process of estimation requires a specific model, so estimation and modelling are terms often used interchangeably in the literature. The premise for modelling migration patterns is summarised by Raymer (2010, p. 73) who stresses that *"the comparative study of migration is hindered by data availability, quality and consistency"* and advocates the use of models *"(i) to correct for the inadequacies and inconsistencies in the available data and (ii) to estimate the missing patterns"*. Rogers (2008) summarises that the migration modelling field has grown enormously over the past 40 years and has split off into too many branches for a comprehensive review to be feasible or appropriate. The article outlines the 'roots' of migration modelling, beginning in 1965 where the 'state-of-the-art' was split into four strands: linear regression models, gravity models, Markov chain models and matrix population models. Rogers (2008) argues that Markov chain models have largely disappeared as a tool for modelling migration, gravity models evolved into spatial interaction models while matrix models evolved from uniregional into multiregional formulations.

Fundamentally, *"a migration flow table can be considered a two-way (that is, origin by destination) contingency table, where the cells represent the counts of migrants"* (Raymer 2007, p.986) and there are a number of ways in which the missing cells of such a contingency table can be estimated. This definition of a contingency table corresponds with the OD face of the Migration Cube presented in Figure 2.2, and to this, age and sex information can be added in various ways. The choice of model form used in this estimation is "*defined as much by the nature of available data as it is by the purpose (e.g. explanation or forecasting) for which the model is intended"* (Stillwell and Congdon 1991, p.1) and choosing the right model with which to estimate missing values depends very much on the data structure and the tools at the researcher's disposal. The following review is split into two sections: first, spatial interaction modelling is discussed alongside the idea of entropy maximisation, which is a concept which frames the method of choice in this thesis, IPF. The second section looks at the family of models termed general linear models, which are widely used in contemporary demographic research.

2.5.1 Spatial interaction models, entropy maximisation and iterative proportional fitting

One way of filling a contingency table (where data may be missing or inappropriate for a variety of reasons) is by using a gravity or spatial interaction model (SIM). The 'gravity variables' of the gravity and SIM are specified by Stillwell (2005, p.7) where *"the characteristics of the origin may act as 'push' factors for potential out-migrants whilst the attributes of the destination reflect 'pull' factors that entice migrants to a particular destination"*. Also important is the frictional effect of distance. Gravity modelling provides a tool to estimate migration and at its most basic, uses the flow of population from origin to destination taking into account the size or mass of the origin/destination and distance between the two (the friction effect). An early gravity model was proposed by Zipf (1946) in a study of migration between US cities and since then the model has been developed extensively. To this basic model, further variables can be added, for example to represent the attractiveness of a certain destination over another (Ewing 1974). Roy and Thill (2004) provide an overview of the development of SIM, from its early form in the context of regional science and its use in measuring consumer behaviour at regional shopping centres through the Huff probabilistic model (1963). Raymer (2007, p.986) argues that *"the spatial interaction model is essentially a statistical form of the gravity model, which includes the factors of population size of the origin and destination regions, the distance between them, and some measure of competition or attractiveness".* The SIM is used extensively in the migration literature. Dennett and Wilson (2013) develop a multilevel SIM where the spatial boundaries are the 287 NUTS2 regions of Europe, which are constrained to inter-country flows and Congdon (2010, p.775) uses a Bayesian methodology to estimate migration between English LADs based on an "*extended random effects gravity model"* which links pull and push scores across all areas. As a predictor of migration flows, a large and complex spatial interaction model using a wide range of variables for 98 zones in England and Wales termed the MIGMOD project has been constructed to help inform policy decisions (Rees *et al.* 2004a).

One of the important considerations of spatial interaction models is the method of calibration. Linear regression was used to derive the early gravity model parameters whereas the family of spatial interaction models derived by Wilson (1967) were based on entropy maximisation principles (Wilson 1970; Wilson 1971) and optimum parameters were calibrated using mathematical methods (Stillwell 1991). Entropy maximisation, *"the most likely configuration of elements within a constrained situation"* (Johnston and Pattie 2009) takes advantage of all information within a spatial system and constrains estimates to known totals: the total number of migrants moving from origin to destination cannot exceed the total number of observed migrants in the system. This, Wilson (1970) argues, is preferable to a gravity model with no bounds which can generate 'nonsense' estimates, with more people travelling between origin and destination than there are in the system.

This principle of entropy maximisation is of itself used in the modelling of incomplete migration tables and as summarised by Raymer (2007, p.986), *"entropy maximisation models borrow most of their strength from historical tables of migration"*. One of the first applications of entropy maximisation in the migration literature is presented by Chilton and Poet (1973), who devise an entropy maximisation model to recover the small flow data (below 10) reported in the 1966 Census. Their contingency table contains known information (flows over 10 and the marginal totals) where small flows are the unknown element to be estimated. Iterative proportional fitting (IPF), the method chosen for the estimation presented in this thesis, can be considered a model where the goal is the maximisation of entropy. Johnston and Pattie (1993) argue that IPF is a means to achieve maximum likelihood estimation, and that the procedure has been employed extensively in the geography literature under the guise of entropy maximisation. Examples of the use of IPF in migration research include Nair (1985), Scoen and Jonsson (2003), Willekens *et al.* (1981) and Willekens (1982). A full consideration of IPF (and the studies that employ the technique), alongside a justification for choosing it over other strategies is reserved until Chapter 4.

2.5.2 Linear regression models

A second family of models used in the estimation of inadequate migration data are termed Generalised Linear Models (GLMs). This encompasses a sub-set of modelling strategies which includes Poisson regression, log-linear modelling, Ordinary Least Squares regression and logit regression (which is often represented using a logistic regression model), all of which are used in migration research and are specified by Flowerdew (1991, p.96) as models *"which involve relating a response variable to a linear predictor"* where the predictor is one or more explanatory variables. GLMs have been implemented in migration research due to their ability to incorporate variables derived from various sources that are known to influence migration decisions (social, economic etc. as specified in previous sections of this review).

An account of the similarities between logit, Poisson and log-linear models is provided by Rogers *et al.* (2001), who find that when all variables in the model are discrete, the results produced are very similar. Poisson regression models (and by extension log-linear models) are generally favoured over standard (OLS) regression models in migration research, as summarised by Lovett and Flowerdew (1989), who argue that standard regression is often inappropriate for count data, and that the Poisson distribution is particularly useful where some observations have very low values, a condition pertinent to migration data, especially at a disaggregated level. Boyle (1993) uses Poisson regression to model a sparse matrix (containing a large number of zeros and small flows) of ward level migration within Hereford and Worcester. He argues that Poisson regression is preferable to OLS regression models which often *"are oversensitive to flows involving very small numbers of people"* (Boyle 1993b, p.1201), an assertion backed up by Flowerdew (2010) in a study of 2001 Census SMS data, who reports that OLS performs as well as Poisson regression where counts are large, but not so well when a large number of the observations are small. The problem that arises when dealing with a large number of zeros in a migration matrix is explored in detail by Bohara and Krieg (1996) who develop a 'zero-inflated Poisson model' for a ten year study of migration in the USA, which they find reduces the under-prediction of migrants found in traditional modelling approaches. In a study of 126 labour market areas in Great Britain, Flowerdew and Aitkin (1982, p. 202) find that a Poisson regression model out-performed a log-normal model as it does not assume that error terms are normally distributed, can deal with unequal variance in error terms and *"avoids the bias of log-normal flow estimates, and can successfully handle zero flows".*

The use of a log-linear model is advocated by Rogers *et al.* (2010, p.30) for the estimation of migration as *"the parameters of that model capture different features of the spatial structure of migration"* allowing for consideration of the characteristics of the origin population, the destination population and the strength of the linkages between the two. Rogers *et al.* (2003, p.67) use a log-linear regression model to *"predict migration from partial data contributed by different data sources"* by origin, destination and age, while Smith *et al.* (2010) combine Patient Register data (covered in the next chapter), the 2001 Census and the LFS in a log-linear model of migration by age, sex and economic activity between counties in England. Raymer *et al.* (2011) take this a stage further, producing an inter-regional age specific and sex specific model of ethnic group migration in England from 1991 to 2007 using a log-linear model to combine migration data from the 1991 and 2001 Censuses and published NHSCR tables from 1991 to 2007. Raymer *et al*. (2011, p. 75-76) argue that the methodology used could be applied to higher levels of disaggregation than GOR, however this would *"require additional efforts to harmonise the Census and NHSCR data over time before combining them"* Past age and spatial structures are used by Raymer and Rogers (2007) to inform log-linear estimates of age-specific migration in the USA and Mexico.

Given that the interaction between origins and destinations is a complex one, it is not unusual for the results from a gravity or spatial interaction model to be incorporated as variables in a regression model. Flowerdew (2010) uses a spatial interaction model for inter-district flows from the 2001 Census to inform a Poisson regression model, while Sarra and Del Signore (2010, p.31) use a 'dynamic gravity model' calibrated using the Poisson regression model and incorporating *"environmental* *variables (housing, transport infrastructures, crime)”* in the modelling of internal migration flows in Poland at the NUTS-2 level.

This short review has shown that different approaches to estimation are undertaken in various studies, and that each has chosen an approach that is suited to the data structure and availability. There is no ‘right’ approach (although as reported, most authors have a preference) and this is an idea explored further in the detailed methodology presented in Chapter 4.

2.6 Conclusion

This chapter has set the context for the remainder of this thesis by: (1) specifying why it is important to understand the magnitude and patterns of migration in the UK; (2) highlighting the problem that exists with the data availability / structure; (3) outlining a solution to this problem; and (4) providing a review of the extensive literature that will be drawn upon in the estimation and analysis of migration in the UK. The next chapter provides a review of the available data and the methods used by the three NSAs in producing these data, thus expanding on the data issues identified in this chapter. The solution to this data problem, discussed briefly in the second section of this chapter, is the focus of Chapter 4 while the literature review, which was the main focus of this chapter, is drawn on in the remainder of the thesis.

Chapter 3

Estimating migration across four home nations: a data review

It was reported in the previous chapter that the data estimation for subnational migration in the UK is carried out by three NSAs (ONS, NRS and NISRA) and that these NSAs use different methods and sources when producing the estimates. This means that the outputs they produce are often not comparable in terms of temporal or spatial focus. This chapter pulls together numerous methodology documents and evidence gathered through meetings and correspondence with statisticians at the NSAs to produce a comprehensive review of the data and methods used in the production of internal, cross-border and international migration estimates. This chapter provides further justification for the core aim of this thesis – to produce a comprehensive and consistent UK database of migration – and provides the context for the next chapter, where the various data sources are combined and the missing data are estimated to produce a database of migration interaction between all LADs in the UK for 2001/02 to 2010/11. Given that much of the data reviewed in this chapter are used in the estimated matrix of the next chapter, assessment of the data accuracy and coverage is undertaken where possible and appropriate. No comprehensive overview of UK wide migration methodology exists, with documentation being produced separately by the three NSAs although a comprehensive audit of available interaction data and their associated methods is presented in Dennett *et al.* (2007), whilst research by Rees *et al.* (2009) provides a summary of migration datasets and a review of migration estimation methods in the UK. More recently, Raymer *et al.* (2012) provide an overview of methods, although their main focus is on England and Wales. The review presented in this chapter adds to this understanding by presenting the most up to date methodology and potential improvements that are being considered by the NSAs.

The remainder of this chapter consists of five sections. First, the role that migration statistics play in the mid-year subnational population estimates is summarised; second, the subnational internal migration methodology and data are considered; third, the cross-border migration methodology and data are addressed; fourth, the methods used in the production of the international migration component are considered; and fifth, a summary of the temporal periods and sub-populations covered by the data is presented. Gathering and collating the evidence was a substantive piece of research in itself, a large number of the methodology documents have no named author or, in some cases, a date of publication. Where this is the case, clarification has been sought from the NSAs and as much detail has been provided in the references as possible.

3.1 Migration in the cohort component model

In order to provide some context for migration statistics in the UK, it is important to understand that their primary use at the subnational level is as a component part of the mid-year estimates (MYEs). The MYEs, produced for the period 1 July - 30 June each year, inform resource allocation and policy decisions at national, regional and local levels. An estimate is produced at a subnational level across the UK – for Local Authorities (LAs) and Unitary Authorities (UAs) in England and Wales, Council Areas (CAs) in Scotland and Local Government Districts (LGDs) in Northern Ireland (collectively termed LADs in this thesis) – broken down by age and sex. The estimates are generated using a cohort component model (Figure 3.1) in which the migration component comprises both international and internal flows.

Figure 3.1: The cohort component method for population estimation

The MYE data for England and Wales are produced by ONS, the Scottish estimates are produced by NRS and the Northern Ireland estimates by NISRA. ONS then compiles a UK-wide MYE assuming that *"the definition, data sources and methods used by NRS and NISRA are broadly consistent, providing comparable population estimates across the UK constituent countries and a coherent UK national compilation"* (ONS 2011b, p.2). This is certainly the case for ageing on the resident population, births and deaths. The usually resident population is aged on one year from the previous MYE. Live births between 1 July of the previous year and 30 June of the reference year are added to population estimates at age zero and are allocated to the LAD where the mother is usually resident. Deaths during the same period are subtracted from the population of the LAD of residence by age at the mid-year reference point. Information on both births and deaths are disaggregated by sex (ONS 2011b; NISRA 2006; GROS 2010b). The 'other changes' specified in Figure 3.1 comprise estimates for special populations such as armed forces and prisoners, plus any boundary changes that happen during the year. Student populations are included as part of the migration component in England, Wales and Northern Ireland, but not in Scotland.

However, the NSAs draw on different data sources and methods for the internal and international migration components of their respective MYEs and these are outlined in more detail in the following sections. It is important to emphasise that available data, assumptions and geographical boundaries used in the estimation of migration are constantly evolving; in this dynamic context, a definitive overview of methodology has limitations, given that historical estimates are often revised based on new information.

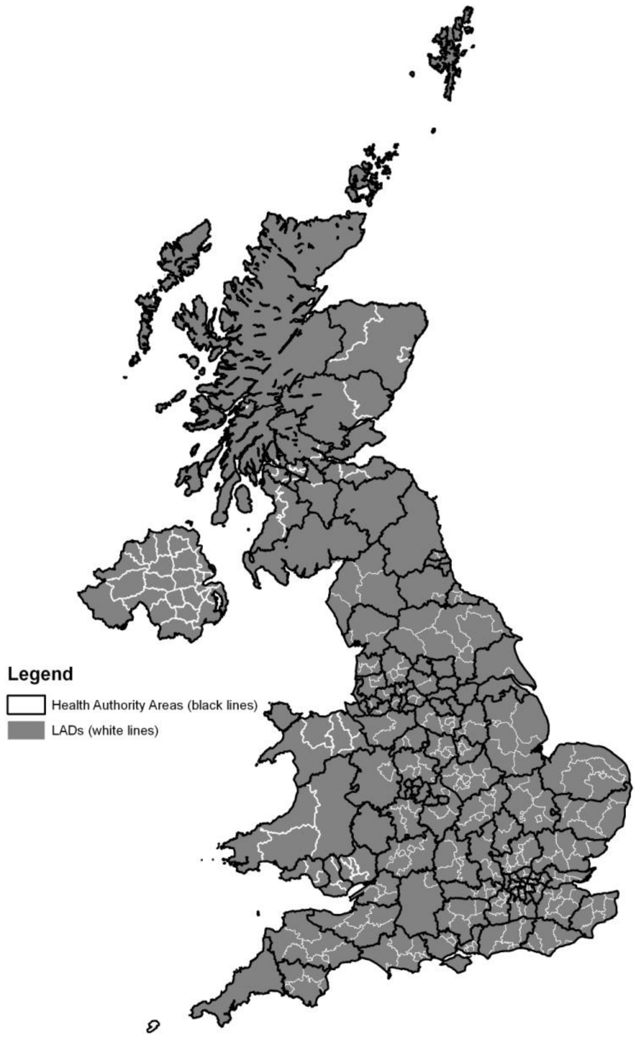


Figure 3.2: Health geography (former Health Authority areas in England and Wales, Health Board areas in Scotland and aggregate Northern Ireland) reported in the NHSCR, overlaid on LAD boundaries for the UK

Both internal and cross-border migration uses a data source called the National Health Service Central Register (NHSCR) in which statistics are reported for health geographies (former Health Service Authority areas in England and Wales and Health Boards in Scotland) and as an aggregate total for Northern Ireland. Figure 3.2 shows the geographical health boundaries reported in the NHSCR, where the black lines represent health areas, overlaid on top of LAD boundaries, shown by white lines. The role of the NHSCR in the context of each type of migration is explained in the following sections.

3.2 Estimation of internal migration in the UK

As there is no compulsory system to record migration in the UK, internal migration (moves between LADs within each country) statistics are derived primarily from National Health Service (NHS) data sources which rely on the re-registration of patients with a doctor when they migrate. They are produced independently by ONS, NRS and NISRA and supplied to ONS for collation at the UK level. ONS and NRS produce LAD to LAD tables of moves, both of which are available in the public domain. NISRA also produces internal migration origin-destination statistics for their MYEs but these are currently not published. The methodologies used in each case and statistics produced are outlined in more detail in the following sub-sections. Here, moves between England and Wales are discussed as internal migration, as the same methodology is applied to moves both within and between each country. In terms of analysis in subsequent chapters however, England and Wales are recognised as separate countries.

3.2.1 ONS estimation method

ONS produce a full matrix of origin-destination flows between LADs in England and Wales which are estimated by combining data from two NHS sources: the National Health Service Central Register (NHSCR) and Patient Register Data System (PRDS) along with data from the Higher Education Statistics Authority (HESA). Estimates are produced by age and sex for ONS use, but age and sex detail is less readily available for academic research purposes. The NHSCR records movements between the former Health Authority (HA) areas in England and Wales, of which there are 104 and can be seen in Figure 3.2; a download is supplied by all Primary Care Trusts (PCTs, the bodies that administer local health service budgets) on a weekly basis, which is then aggregated and reported quarterly by ONS. The combined PCT downloads form the complete NHSCR database for England and Wales.

In 2006, HAs became a redundant health geography but NHSCR estimates continue to be published based on their boundaries (ONS 2010a). Furthermore, following the Health and Social Care Act 2012, the role of PCTs in England is being taken over by Clinical Commissioning Groups, a process that started in March 2013 and which is currently on-going (although this has no impact on the reporting of NHSCR statistics at former HA level). A more detailed discussion of these health geographies can be found in Chapter 4 in relation to geographical consistency throughout the time series.

The extract of the NHSCR which is supplied to ONS does not contain comprehensive enough geographical detail for estimation of migration at a lower level than HA, as it contains no postcode or address information for patients. For this reason the NHSCR is combined with the PRDS which does include the postcode of patients (ONS 2010e). A yearly PRDS download, supplied by the PCTs at the end of July (a date chosen as it fulfils the assumption that there is a delay of one month between a person migrating and registering with a new GP), records all people registered with a GP in England and Wales. The register download in the current year and previous year are compared, with patients being linked between one year and the next by a unique NHS identification number. A migration is recorded when a change in postcode is picked up from one yearly download to the next. Moves within a LAD are discarded, as are any changes that come about through boundary changes. The age and sex of a patient migrant are reported in the PRDS.

The PRDS estimates are then constrained (scaled to agree with) the HA level moves reported in the NHSCR. This scaling procedure is carried out because estimates derived solely from the PRDS miss some migrants due to the download only being supplied by the PCTs on a yearly basis. ONS (2011a, p.5) report that the PRDS misses *"the movement of those migrants who for one reason or another were not registered with a doctor in one of the two years, but who moved during the year".* The largest group of unrecorded migrants is babies who were born part way through the year (so do not appear on the previous year PRDS register) but also people entering or leaving the armed forces (as armed forces personnel are not captured by the PRDS), international immigrants and emigrants and people who die before the end of the year. The NHSCR, as a weekly download, provides better temporal coverage than the PRDS, so effectively *"the more complete information from the NHSCR is combined with the more geographically detailed data from the patient registers"* (ONS 2011a, p.5-6). Finally, an adjustment is made to the constrained estimate using data from the UK Higher Education Statistics Authority (HESA) to take into account student migration, as explained in more detail in the next sub-section.

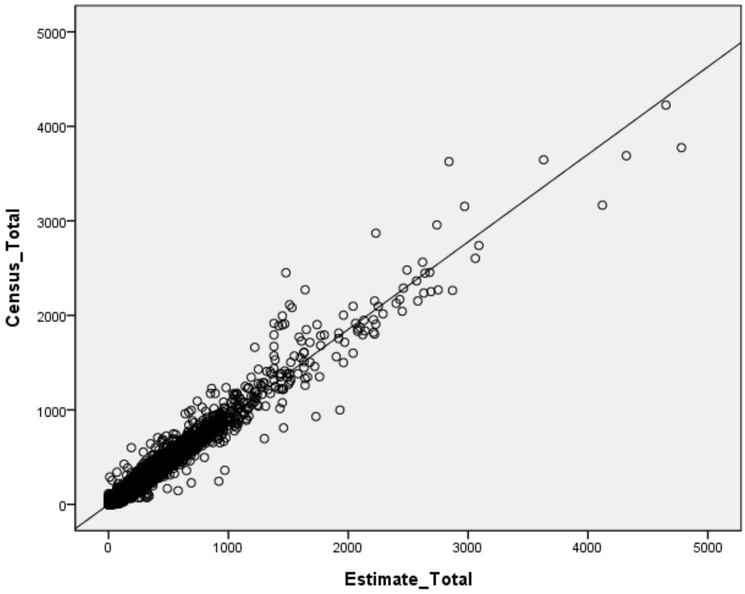


Figure 3.3: A comparison of pairwise LAD-LAD flows within England and Wales, reported in the 2001 Census and 2000/01 PRDS

As the intention is to use the PRDS/NHSCR/HESA estimates as they are supplied by ONS in the complete time-series dataset, as outlined in the following chapter, Figure 3.3 provides a quick assessment of their coverage and accuracy. The graph compares the origin-destination flow for all pairs of LADs (120,756 pairs in total) reported in the 2001 Census to the ONS estimated 2000/01 flows. Although there is a two month time gap between the 2001 Census (29 April) and PRDS/ NHSCR derived (30 June) flows, the correlation is strong and positive ($r=0.97$, $p<0.01$), although some outliers do exist.

3.2.2 ONS student adjustment

The rationale behind applying a student adjustment to internal migration data is set out by ONS (2010c). First, young people, particularly young men, can be slow to change their registration with a GP when they move. Second, movements of students attending higher education can be complex, including transfers to the place of study, moves during the study period and moves after completing their study programmes. Students may have two addresses, a term-time address and a home (domicile) or parental address, both of which they spend time at. For these reasons, ONS introduced the student adjustment using HESA data in 2010. The focus of the adjustment is internal migration moves made by first year undergraduates and students at the end of their studies *"who did not change their GP registration when they moved"* (ONS 2010c, p.2). The adjustment consists of three calculations:

- A start of study adjustment: this is applied only to first-year undergraduate students by comparing the term-time LAD to the domicile LAD by single year of age and sex, and is based on the assumption that most students begin university at age 18 or 19. Where the HESA flows between domicile and term-time LAD are larger than the PRDS flows (for this age group), HESA data are used. A 'flag' is used to identify a student who lives at his/her parents address during term time, and each flagged record was removed if it was a feasible distance from the campus of study (ONS 2010d).
- An end of study adjustment: as there is no source which identifies where students move to at the end of their studies, a set of estimations are undertaken by ONS:
    - the number of people who end their studies each year is collected by HESA, and includes term-time address from 2007/08. For adjustment between 2002 and 2008, the 2007/08 term-time address distribution has been used;
    - the number of former students moving to a different local area after their studies is taken from 2001 Census data, using the question asking for address twelve months ago. A Census record is only used if an individual held an undergraduate degree at age 22 or a postgraduate degree at age 23. These records are used to calculate a rate for graduates leaving a LAD (graduates in the Census who left the LAD divided by Census graduates in the LAD 12 months before the 2001 Census);
    - the number of students who move but do not re-register with a GP – first the rate of students who do re-register is calculated based on moves from the PRDS for mid-2000 to mid-2001, compared to moves from the 2001 Census by sex and age of 17-28 year olds. The rate of moves not

identified on the patient register is then calculated as 1 minus the above; and

  - the destination of former students not re-registering is calculated using 2001 Census data to create a matrix of LAD to LAD moves, disaggregated by sex for an individual who held an undergraduate degree at age 22, or postgraduate degree at age 23.

- A double counting adjustment – as students are likely to re-register with a GP eventually, an investigation into the amount of time this takes was conducted at halls of residence at Bournemouth, Aberystwyth, Newcastle and Northumberland universities. These students were tracked over time to see how long over three years it took to re-register. This includes both a 'start of studies' and an 'end of studies' adjustment.

The adjustment method attempts to deal with problems encountered when producing mid-year population estimates as students move to university after the mid-year reference point (30 June). Assuming students re-register with a GP when they move to university, they will be counted at their home (parents) address in the first year of their study, but their term-time address in the second. At the end of their study, the academic year (particularly for undergraduate students) often ends before the mid-year reference point, *"hence former students may be registered at a new address they have only lived at for a fraction of the mid-year to mid-year period"* (ONS 2010d, p.2).

3.2.3 NRS estimation method

In a similar way to the English and Welsh estimates produced by ONS, an inter-LAD matrix of flows is produced by NRS using two data sources: the Scottish NHSCR and the Community Health Index (SCHI) (GROS 2010a). The SNHSCR, available as a weekly download to NRS, records movements of migrants between 14 Health Board (HB) areas in Scotland (see Figure 3.2) and contains age and sex information. The SNHSCR suffers from the same limitation as the NHSCR used in England and Wales in that it does not contain the postcode information of patients. The SCHI is largely comparable with the English and Welsh PRDS dataset: it is produced as a yearly download for NRS and it records the postcode of patients registered with a GP in Scotland, along with age and sex variables. Comparison of the SCHI register between one year and the next, with patients being linked by a unique identification number, reveals a migration where a patient changes postcode. As a yearly download, the SCHI has the same inherent problem of under reporting certain types of migrant as the PRDS (babies, armed forces, international migrants and people who die at some point between the SCHI downloads).

As is the case for the PRDS/NHSCR derived estimate in England and Wales, the annually downloaded SCHI estimates are 'controlled' (adjusted to agree with) the SNHSCR totals by origin, destination, age and sex (GROS 2010b). No student adjustment is made for inter-LAD flows in Scotland, a deficiency discussed in Chapter 8 where estimates of the student age population are analysed.

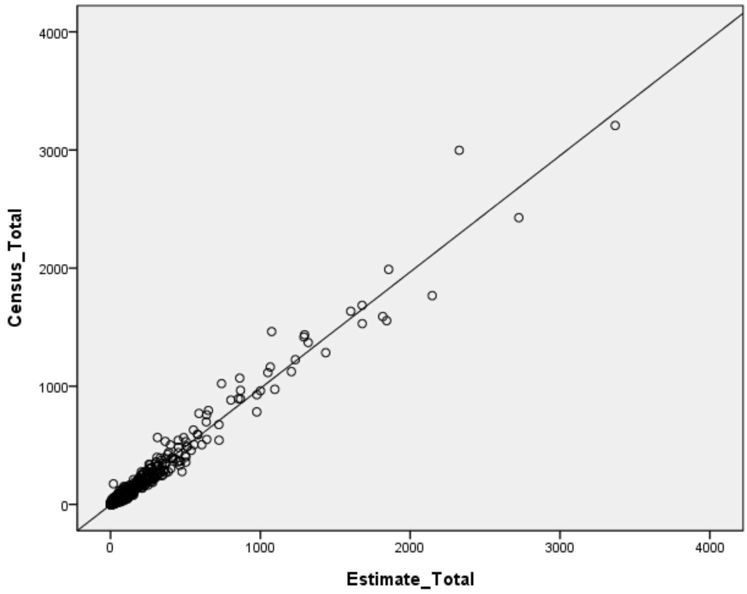


Figure 3.4: A comparison of pairwise LAD-LAD flows within Scotland, reported in the 2001 Census and 2001/02 CHI

Again, as the intention is to use these internal migration estimates as they are, with no adjustment, it is prudent to compare them with 2001 Census data to assess their similarity. Figure 3.4 compares the LAD to LAD flows reported in the 2001 Census and 2001/02 CHI (992 origin/destination combinations in total). The NRS methodology using the SCHI and SNHSCR outlined above only came into effect from 2001/02 onwards with no origin-destination statistics available before this (Dennett *et al.* 2007), even for official purposes within NRS (NRS 2010, p.3). With no CHI data available for 2000/01 this comparison is between data reported one year and two months apart (census day 2001 and 30 June 2001). Despite this inconsistency, the correlation between the reported flows is strong ($r=0.97$, $p<0.01$) which suggests that there is substantial consistency in the structure of origin-destination flows in Scotland.

3.2.4 NISRA estimation method

Unlike in England, Wales and Scotland, only one NHS data source is used in Northern Ireland. NISRA estimate flows at LAD level using the Northern Irish Central Health Index (NICHI) which records changes in address when a patient re-registers with a GP after a migration event and contains detail on the age and sex of the patient. Registration on the NICHI requires a person to obtain a Health Card in order to access medical services, so the data are variously reported as NICHI and 'Health Card data' by different sources. The larger health area equivalent to the HA/HB in Northern Ireland is the Health and Social Service Board (HSSB), but this geography is not used in the production of migration statistics. To account for under registration of adult males, the age distribution reported in the NICHI is adjusted to match that of the young female age distribution (ONS 2011c). In addition to the NICHI derived internal migration estimate, a student adjustment is made, informed by HESA data, by removing a number of people of student age from most LADs and *"adding these to a small number of LGDs with centres of third level education"* (NISRA 2007, p.3). These centres of third level education are identified as Belfast, Newtownabbey and Colerain (NISRA 2006).

Documentation on Northern Irish internal migration methodologies is fairly sparse, but the accuracy of using the NICHI to produce migration statistics was investigated by NISRA (2007) by comparing results from the 2000/01 register with results from the 2001 Census. NISRA found that the NICHI reported 35,500 inter-LAD moves while the Census recorded 37,100 moves. The age and sex breakdown were also reported to show similar patterns and as such it was concluded that the NICHI was a suitable data source for estimating internal migration in Northern Ireland (NISRA 2007). As no origin-destination migration statistics between LADs are available for Northern Ireland, it is not possible to provide a comparison between the NICH reported statistics and the 2001 Census as for England, Wales and Scotland in previous sections. This is a set of flows which are estimated in the next chapter.

3.2.5 Potential improvements to internal migration methodologies by ONS and NRS

The last three sections have explained the current methods that the NSAs use to extract and integrate data from various sources so as to generate the best estimates of internal migration for use in their cohort component models. This section briefly discusses ways that ONS and NRS are considering to improve their internal migration methodologies using new or improved data. No information is available on potential improvements to the NISRA methodology, although Barr and Shuttleworth (2012, p.603), in an assessment of migrant coverage in the Health Card data, recommend that *"better estimates of the whole population are possible if attempts are made to locate some of the groups identified as harder to capture"*. In addition to men and young people, these groups are identified as those who are not married, in higher status occupations and those who are healthy. Barr and Shuttleworth (2012) conclude that further work is needed to assess the accuracy of Health Card data used to measure migration in Northern Ireland.

3.2.6 ONS improvements

ONS are currently assessing the possibility of using data held on patients in the Personal Demographics Service (PDS), a constituent part of the NHS system that makes up the 'spine' of the NHS care records service (NHS 2011). By using the PDS, a wider proportion of the population of England and Wales will be covered as patient address details will be added at more points of contact with the NHS and will not rely solely on registration with a GP in England or Wales. As summarised by the Select Committee on Public Accounts (2007, p. 1), this *"provides more convenience for patients as they need only notify one authorised healthcare organisation of a change of address and this change will be available to all healthcare organisations as and when the patient records are accessed"*. The use of the PDS as a data source would have a substantial positive impact on migrant estimation for hard to measure groups. For example, young males (a group that are consistently undercounted due to poor registration rates, see Smallwood and De Broe 2009; Fotheringham *et al.* 2004) who attend an A&E department would have their address details stored, even if they had not registered (or re-registered following a migration event) with a GP.

3.2.7 NRS Improvements

NRS is looking towards the use of a SNHSCR monthly extract which includes the postcode of all people registered with a Scottish GP. As was reported in Section 3.3.3, the SNHSCR extract contains no address details but a revised data specification agreement with the NHS could see the inclusion of postcode information on SNHSCR data (Mueller 2011). As the SNHSCR is deemed to provide better temporal coverage than the SCHI, due to the frequency with which it is reported, this means a potentially more accurate reporting system. The possibility of removing reliance on the SCHI would allow the estimates to be more direct, as SCHI totals would no longer need to be constrained to SNHSCR totals and estimates would no longer need to be constrained to HB areas. Work is currently underway to assess the postcode data provided by the SNHSCR and NRS (2011, p.2) report that migration figures at LAD level can likely be produced *"to an acceptable degree of accuracy despite the existence of a number of postcodes which cannot be validated and allocated to Local Authorities".*

In contrast, NRS (2010) report that there is no short term plan to introduce a student adjustment to their migration estimates. In a report assessing the viability of applying the ONS student adjustment methodology (outlined above), NRS conclude that as the SCHI 2000/01 data are not available, the end of study adjustment (which in England and Wales compares 2001 Census and 2000/01 PRDS data) could not be calculated without a 'viable alternative' dataset (NRS 2010, p.4).

3.3 Cross-border migration in the UK

Cross-border migration statistics in the UK are generally reported by the receiving country as these are seen to be more accurate than those of the sending country (ONS 2011a). The level of detail available for migration across the borders varies between the constituent countries and the following sub-sections outline the methods used and data that are produced. Flows (with an origin and a destination) disaggregated to anything below health geography units are not estimated and so data on cross-border flows between LADs in each of the four constituent countries (internal inter-national flows) do not currently exist. This is a flow estimated in the next chapter and relates to the core aim of this thesis, to produce a comprehensive UK wide dataset.

3.3.1 Cross-border migration reported in the NHSCR

Between England, Wales and Scotland, the NHSCR and SNHSCR are both able to provide counts that distinguish cross-border flows between health geographies (HBs in Scotland and HAs in England and Wales). The SNHSCR, compiled in Dumfries, records moves from HAs in England and Wales to HBs in Scotland, while the NHSCR, compiled in Southport, records the moves in the opposite direction. In addition, both are able to identify a move to and from Northern Ireland, but do not distinguish origins or destinations in Northern Ireland below the national level (Northern Ireland is treated as a single area in the (S)NHSCR data, as shown in Figure 3.2). The data reported in the

NHSCR and SNHSCR are available as counts of moves by age and sex. The register in the receiving country is adjusted first, and notification of the move is then communicated to the sending country and its register is updated. NRS send their matrix of flows from the SNHSCR to ONS who administer the data. In the case of Northern Ireland, because the NHSCR is not used for cross-border migration, (ONS 2011a, p.3) report that *"invariably, the number of migrants moving to Northern Ireland as recorded in the NHSCR is different from the number of moves to Northern Ireland recorded by NISRA… Therefore, ONS apportions the NHSCR data to take account of the differences in the number of moves recorded by the NHSCR and NISRA".* This approach is based on the assumption that the count reported in the country that receives the migrants is better than that in the origin country.

In terms of the accuracy of reporting, the Audit Commission (2006), under the national duplicate registration initiative, assessed the 56 million electronic records of patients registered with a PCT or Local Health Board in England and Wales for 2004/05, identifying 185,000 records (0.3 per cent of the population) which could be deleted based on a number of criteria, including duplicated records and 'gone aways' who no longer lived at the address held and deceased persons. A review of the NHSCR using the Longitudinal Study (LS is a sample of circa 500,000 people at each census date) was carried out by Smallwood and Lynch (2010) who report that 95.7 per cent of ONS LS members enumerated at the 2001 Census resided in the same area as their NHSCR record. They also found that only 1.5 per cent of those enumerated in the 2001 Census did not appear on the NHSCR at all. These results suggest that the overall coverage of the NHSCR is good.

Assessment of the flows between health areas reported in the NHSCR and SNHSCR that are available for use in this study can be undertaken by comparing the data to the 2001 Census, and the result of this assessment is presented in Figure 3.5. Flows reported between LADs in the 2001 Census have been aggregated up to the HA/HB in which they are located, and this aggregated flow is compared to the NHSCR health area to health area data for 2000/01 (with Northern Ireland as one aggregate area, as it is reported in the NHSCR data). The correlation between NHSCR and 2001Census flows is strong and positive ($r = 0.97$, $p<0.01$), as is the correlation between migration rates based on the 2001 population size of the destination HA ($r = 0.98$, $p<0.01$). When the standardised residuals for these rates are analysed (blue points in Figure 3.5), 147 out of 14,055 pair-wise cases fall outside of the 95 per cent confidence interval. Of these 147 cases, only nine constitute cross-border flows (7 from England to Scotland and 2 from Northern Ireland to Scotland). The rest of the outliers are between HAs located within close proximity to one another (all are within the same GOR as each other). In 96 instances, the 2001 Census flow exceeds the NHSCR flow, suggesting that migrants who are picked up in the former did not re-register with a GP when moving, so are not captured in the NHSCR. However, in 51 cases, the NHSCR flow exceeds the 2001 Census flow.

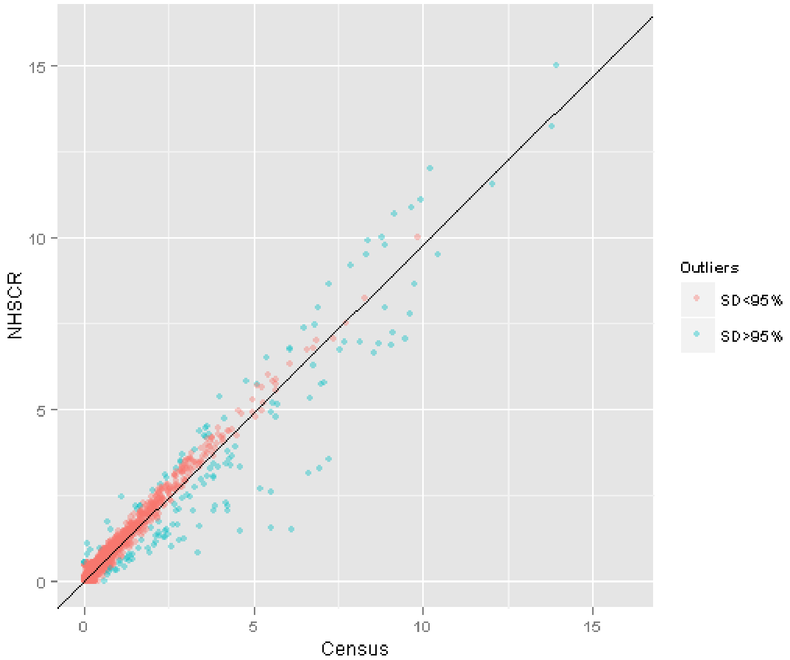


Figure 3.5: The correlation between HA flows derived from the NHSCR and the 2001 Census

Consideration of a methodology that uses this NHSCR data to estimate LAD to LAD cross-border migration flows is undertaken in the next chapter. Aside from the NHSCR, there is no other data source that captures cross-border flow data, so the data sources which capture outflow and inflow at LAD level are assessed in the next section.

3.3.2 LAD level cross-border estimates and data

At LAD level, the most comprehensive information on cross-border migrants is available from ONS for England and Wales. ONS (2011a) report the method used, which relies on the yearly download of the PRDS. For in-migration to a LAD from Scotland or Northern Ireland, a migrant is initially identified where an NHS identification number appears on the PRDS yearly download which was not on the previous year's PRDS and the record indicates that the previous HA was Scotland or Northern Ireland. These data are then constrained to agree with total moves into England and Wales from Scotland or Northern Ireland as reported in the NHSCR. There is however no record of the subnational origin of that migrant.

The process is similar for estimating cross-border emigration from a LAD in England and Wales. Where an NHS identification number that was on the previous year's PRDS is not found on the current year's register, the number is checked against the NHS numbers held by Scotland (on the SCHI) and Northern Ireland (on the NICHI) and where a match is found, the migrant is recorded as moving to either country. These migrants are then constrained to the in-migration totals reported by NRS and NISRA (as the data from receiving country is deemed to be more accurate). Again, no subnational destination is reported for the migrant who has left a LAD in England or Wales. The data reported for LADs in England and Wales is therefore a total inflow and outflow from/to the whole of Scotland or Northern Ireland – age and sex variables are reported in the data.

In Scotland and in Northern Ireland, the level of disaggregation for cross-border migration at LAD scale is even less detailed. Both NSAs report migration between a LAD and the 'rest of the UK' with no reporting of which UK country sent or received a migrant. For Scotland, the SCHI yearly download gives similar detail to the PRDS download used by ONS. By comparing the yearly download, where a migrant's previous address is identified as being outside of Scotland but somewhere else in the UK an in-migration is reported. Similarly for out-migration, a move to the rest of the UK is identified where a patient's NHS registration number appeared on the previous year's SCHI register but not on the current year, and is controlled to the national total from ONS and NISRA. Prior to mid-2007, in Scotland this LAD level 'rest of UK' figure was combined with overseas migration and only an 'outside of Scotland' total was reported for in-migration and out-migration at LAD level. The Population and Migration Statistics Committee (2007) report that although the cross-border data and IPS derived international data were available at HB level, for the distribution to LAD level *"both overseas and UK migrants are combined"*. Following discussion with NRS, it was concluded that it was not possible to disaggregate this LAD level total to UK and non-UK migration (Mueller 2011).

For Northern Ireland, NISRA (2013) report that data for migration to/from Great Britain are based on transfer of Health Cards (the NHS number) and the figures are agreed with ONS and NRS. The migration reported at LAD level is derived from Health Card registration where a person's previous address was in England, Wales or Scotland. Out migration is based on Health Card de-registration and agreed with the in-migration figures collated by the other NSAs. Again, the migration data at LAD level for Northern Ireland is reported as having an origin/destination in the 'rest of the UK,' with no indication of which country a migrant moves to or from.

3.4 Estimation of the international migration component

All three NSAs draw on the United Nations definition of a long-term migrant as *"someone who changes his or her country of usual residence for a period of at least a year"* (ONS 2009b, p.3). ONS compiles a national level estimate of Long-Term International Migration (LTIM) for the whole of the UK as well as estimates of short-term migrants (staying for 3-12 months) and visitors to the UK (staying less than 3 months). It is the long term migrants who are of relevance to this thesis. The national level LTIM estimates produced by ONS are based on the following data sources, as outlined by ONS (2013c):

- The International Passenger Survey (IPS) flows – international migration flows are based on a sample of 0.2 per cent of travellers interviewed on entering or leaving the UK via the major air, sea and Channel Tunnel embarkation points. Of the 230,000 interviews conducted in 2008, 2.2 per cent were migrants, giving a sample size of around 5,000 (ONS 2009b). The IPS is an 'intentions based' survey, asking migrants where they intend to go and how long they intend to stay in the country. Reliance on the IPS has been widely criticised, Kupiszewska and Nowok (2008, p.57) find that analysis of time series IPS data showed *"strong fluctuations, compared with much smoother curves reported by, for example, the Netherlands"* where a population register is used. Problems with the limited sample size of the IPS have been emphasised by Rees (2008, p.354) who states that emigration flows at a regional scale can only be 'guestimated', and by Stillwell *et al.* (2010, p. 2) who suggested that even at regional scale, users are advised to "*smooth out irregularities in the data by calculating three year averages".* A Parliamentary Select Committee (2008, p.26) concluded that the IPS was *"not fit for purpose"* as a source of migration statistics as its primary intended use is for measurement of tourist flows. Nevertheless, the IPS continues to be the primary source of data from which international migration in England, Wales and Scotland is estimated. As part of an ONS sponsored doctoral thesis at the University of Southampton, George Disney is working to improve these IPS derived estimates, using Bayesian modelling to estimate international migration by country of citizenship (Disney 2012, pers. comm.).

- Migrant switcher and visitor switcher flows – because the IPS is an intentions-based survey, an adjustment is made for 'migrant switchers' who state an intention to stay in the UK for over 12 months but who stay, in fact, for less than 12 months and for 'visitor switchers' whose intention is to stay for less than 12 months but who remain for over 12 months (ONS 2007c). these calculations are based on the knowledge that *"the likelihood of a visitor changing their intentions can vary depending on their citizenship and place of last or next residence"* (ONS 2013c, p.9).
- Northern Ireland migration flows – since 2008, Health Card registrations have been used to capture the international migration flows between Northern Ireland and the rest of the world reported in the LTIM. This is because the IPS has not historically sampled air or sea ports in Northern Ireland (although Belfast is a recent addition) and the Health Card registration system is seen as a more accurate measure of flows (ONS 2010e). Two criticisms of the Northern Ireland international migration estimate exist: first, Health Card registration identifies both long-term and short-term migrants (ONS 2010d) resulting in a potential over-count of long-term immigration to Northern Ireland; and second, the emigration estimate derived from Health Card de-registrations underestimates the number of emigrants from Northern Ireland and subsequently needs to be scaled up by 67 per cent (NISRA 2010).
- Home Office asylum seeker data – which are used to adjust IPS data to exclude asylum seekers counted in the IPS and those who returned within one year of their application, the number returned to their country of origin and the number who withdrew their asylum application (ONS 2013c).

Each of the NSAs in the UK has its own method for estimating international migration at the subnational level and these methods have been the subject of substantial revision during the 2000s. Whilst IPS data combined with migrant/visitor switcher and asylum seeker data are used for England, Wales and Scotland at the national level, these national estimates are distributed differently to the subnational level by ONS for England and Wales and NRS for Scotland. NISRA does not use the IPS, relying instead on a distribution method based on Health Card registrations. The following sections outline in more detail the estimation methods that the NSAs use for both subnational immigration and emigration.

3.4.1 ONS estimation methods

In 2013, resulting from work carried out during the MSIP, ONS implemented a new methodology for distributing international immigrants to LADs in England and Wales and back dated the method to all data reported for mid-year 2006 onwards (ONS 2013d, p.9). under this new methodology, the IPS estimate of immigration (adjusted for migrant and visitor switchers) is distributed directly to the LAD level in England and Wales by utilising administrative sources which correspond to the type of migration reported by migrants in the IPS questionnaire (ONS 2011, based on work by Boden and Rees 2010). The main streams identified on the IPS questionnaire are those entering the UK for work, for study, returning migrants and an ‘other’ group who do not state one of the specific reasons for immigration. Where a migrant who states their reason as being for work purposes, the Migrant Worker Scan and the Lifetime Labour Market Database (known as L2) are used to distribute the migrants based on national insurance number (NINo) registrations. For immigrants who state their reason as study, data from HESA and the Department of Business, Innovation and Skills (which records Further Education students) are used. Finally, registrations with a GP (Flag-4 registrations) are used to allocate the ‘other’ migrants. The revised methodology responds to some known estimation issues in certain LADs and the relative complexity of the Poisson regression model for distribution which was in place previously (McGregor 2011).

For all data between mid-year 2001 and mid-year 2005, the previous (pre-2013) methodology still applies, which comprises a three-stage estimation procedure. First, ONS distribute the IPS estimate to the regional level (Wales plus 9 GORs in England) using the Labour Force Survey (LFS) three-year average. The LFS is a sample of 60,000 households per quarter which reports previous country of residence, and is seen to give a more accurate distribution of immigrants than the intention-based inflows captured by the IPS (ONS 2007d). The LFS distribution is used as a control total and the IPS estimate is then allocated to the regional level using the LFS distribution by broad age group and sex. The IPS three-year average estimate is then distributed to an intermediate geography called the 'new migrant geography' for immigration (NMGi). The NMGi is an aggregation of LADs which share a boundary and have a minimum of 20 IPS contacts per year (ONS 2009a). NMGi replaced the increasingly obsolete Health Authority geography (as discussed above) for reporting of international immigration in 2007. In the third stage, immigrants are allocated to the LAD level using a Poisson regression model which incorporates a number of covariates such as Flag-4 General Practitioner (GP) registrations and National Insurance Number (NINo) registrations of overseas immigrants and immigrant counts from the 2001 Census. The covariates vary each year as ONS (2007f) report that fixing the covariates caused volatility in the model over time. The weighted IPS estimate is the response variable and the *"approach reduces the variability in the IPS estimates at local authority level by making use of their relationship with the predictor variables"* (ONS 2011b, p.6). The LAD estimates are constrained to sum to the national and regional IPS estimates. A Poisson model is used as it is able to deal appropriately with count data, where standard regression methods are often not appropriate (see Section 2.5.2 for consideration of the types of regression model used in migration studies).

The method used for estimating immigration to London Boroughs between mid-year 2001 and mid-year 2005 is slightly different; all non-students are allocated to the NMGi level using the LFS three-year average rather than using the IPS three-year average as occurs for the rest of the UK. This is because the sample size of the LFS is seen to be sufficiently large for London boroughs, but not for LADs outside of London (ONS 2007b). Non-UK students are distributed to London boroughs directly without the use of the NMGi based on data supplied by HESA.

The International emigration methodology received no update in 2013, and is implemented as follows. The IPS interview includes a sample of international emigrants at UK air, sea and Channel Tunnel embarkation points. This estimate is used at both the regional level and distributed as a three-year average at the intermediate level (NMGo) where the 'o' stands for out. NMGo areas are based on the NMGi areas with some adjustment made to account for smaller numbers of out-migrants in the IPS (ONS

2007e). The NMGo areas are larger and there are fewer individual areas in order to provide a robust sample size. A Poisson regression model is used at the LAD level, with the IPS direct estimate as the response variable. The specification differs from the immigration model in the selection of variables, since it includes the immigration estimate from the previous year, housing type and housing tenure. Unlike the immigration model, the covariates are fixed for each year (ONS 2010b). When the revised immigration estimate was back-dated to mid-2006 using the new methodology, the immigration input to the emigration Poisson model was revised accordingly, so while the method for estimating emigration did not change, some estimates did (ONS 2013d)

In all estimates of immigration and emigration, detailed data on asylum seekers is provided by the Home Office and are incorporated into the subnational data. While the data are considered high quality, they *"do not correspond directly to the standard ONS definition of a long-term international migrant"* and as such, broad assumptions are made by ONS about the proportion of asylum seeker applicants that actually correspond to the ONS definition (ONS 2011d, p.2). Asylum seekers are distributed subnationally based on the Home Office data.

3.4.2 NRS estimation methods

The approach of the NRS to subnational immigration estimation makes use of SNHSCR and SCHI data. The Scottish share of the IPS (which is adjusted for migrant and visitor switchers by ONS) is initially derived using the LFS (in the same way as the IPS allocation to former GORs is specified in the ONS methodology) which is seen to give a more accurate subnational distribution of international migrants than the intention based counts specified in the IPS (GROS 2010b). The Scottish allocation of IPS migrants is then distributed to Scottish Health Board areas using overseas inflows recorded on the SNHSCR, which includes an age/sex distribution. The distribution of immigrants to LADs is based on the Scottish Community Health Index (SCHI) which records the postcode of patients registering with a GP in Scotland. The SCHI gives the date of registration and a record where an individual previously resided overseas is marked as an international immigration and classed as an international migrant move (GROS 2003). The SCHI is constrained to the SNHSCR in much the same way as it is for internal migration. One problem with the estimate is that the reporting of a previous address that was overseas on the SCHI is not mandatory (GROS 2003). Prior to the mid-2007 immigration estimates, the IPS estimate was used directly without the LFS distribution or adjustment for migrant or visitor switchers (GROS 2010b)

The majority of asylum seekers are assumed to be supported by the National Asylum Support Service (NASS) and as such are removed from the IPS control totals and distributed to Glasgow, which is the only Scottish LAD in contact with the United Kingdom Border Agency (UKBA). A small proportion of non-NASS asylum seekers are distributed around the rest of Scotland and all asylum seeker distributions are based on the five-year age bands provided by ONS (GROS 2010b).

Methodology documentation detailing the subnational emigration estimates produced for the subnational level in Scotland is fairly sparse. GROS (2010b, p.11) report that *"international out-migrants were allocated using a combination of in-migrants to Scotland from overseas and migrants leaving Scotland for the rest of the UK"* while GROS (2010a, p.1) report that the distribution is calculated *"using averaged proportions based on international inflows, outflows to the rest of the UK and the population size of each Health Board"* (GROS 2010a, p.1). The Population and Migration Statistics Committee (2013, p.3) shed some light on this method, suggesting that the total number of people who leave Scotland for overseas are distributed to HB level based on three criteria: (i) the number of out-migrations to the rest of the UK from a HB (which is reported by the receiving country); (ii) the number of in-migrations to that HB from overseas (as reported using the immigration methodology); and the size of the general population of the HB (taken from the MYE). The age/sex distribution of emigrants is based on the distribution of migrants to the rest of the UK, derived from the SNHSCR and SCHI (GROS 2010a, p.1). Prior to mid-2007, de-registration of migrants was used to measure emigration, the coverage of which was poor: the Population and Migration Statistics Committee (2013) report that only around one third of outflow recorded in the IPS was captured by de-registration. The statistics reported prior to mid-2007 also included an 'adjustment for unmeasured migration,' which GROS (2005) report is an adjustment of -2,600 migrants in 2001/02 to 2002/03 and -1,500 migrants in 2004/05 at the national level, which is the result of under estimation of young adult males to overseas destinations identified when comparing the mid-2000 MYE and 2001 Census results.

As was reported in Section 3.4.2 dealing with subnational cross-border migration estimates, prior to mid-2007 moves to and from LADs in Scotland which were not internal migration were reported as having an origin or destination that was 'outside of Scotland.' There is no way of disaggregating this 'outside of Scotland' figure to international and cross-border migrants, so an estimation scheme is incorporated into the methodology presented in the next chapter which takes this into account.

3.4.3 NISRA estimation methods

The methodology in Northern Ireland differs from the rest of the UK as NISRA does not make use of data from the IPS. Instead, Health Card registration data are used in both the immigration and emigration estimates. In the case of immigration, registration with a family doctor requires an international immigrant to apply for a Health Card, at which point he/she must provide information about age, place of residence and time of stay to the Business Services Organisation of Health and Social Care (HSC-BSO) in Northern Ireland (NISRA 2010). Health Card data are seen as the most comprehensive source with which to estimate international migration and give an indication of intention to stay for a period of time, as registration is only possible for a migrant staying for over three months. To account for under-registration by young males, the age distribution is adjusted to be similar to that of young female migrants in the estimates (NISRA 2010). Health Card registrations give detail allowing estimates to be disaggregated by age and sex.

Emigration estimates are also derived from the health Card registration system, which records de-registrations with a family doctor, and which are adjusted for young males in a similar way to the immigration estimates. The reported total is scaled up by 67 per cent, based on the assumption that only three in five people de-register with their GP (NISRA 2013), as de-registration is not mandatory and there is little incentive to do so. The de-registration data are combined with the data from the Central Statistics Office (CSO) Irish Quarterly National Household Survey which provides an estimate of numbers moving from Northern Ireland to the Republic of Ireland. Immigration and emigration by asylum seekers in Northern Ireland is distributed subnationally using the same Home Office data used by ONS for England and Wales (ONS 2011d).

3.4.4 ONS revisions to migrant distribution methodology in 2007 and 2010

As was mentioned at the beginning of this chapter, the methods used and data produced for migration estimates regularly change. Some revisions carried out during the time series by ONS are summarised in this sub-section, while in the next sub-section some potential improvements proposed by NRS and NISRA are summarised.

In 2007, ONS implemented a series of improvements which primarily impacted on the distribution of international migration to the subnational level. It was the 2007 improvements that saw the introduction of the new migrant geography (NMG) to replace the increasingly obsolete Health Authority geography which perpetuated a *"lack of consistency across the country in terms of the numbers of LAs within each intermediate geography"* (ONS 2008, p.2). Improvements to the visitor and migrant switcher methodology (as discussed above) utilised new questions in the IPS, first asked in 2004 to improve on the assumptions made by ONS for visitor/switcher numbers. The questions asked previous migrants *"when you last arrived in* (left) *the UK, how long did you intend to stay* (away) *for?"* (ONS 2007c, p.5).

Two further improvements were implemented, utilising administrative and other data sources. First, the LFS was introduced to improve the regional distribution of migrants. This apportioned the IPS estimate to the GOR for England, and nationally to Scotland and Wales. The new method replaced distributions based on the IPS with the LFS three-year average distribution of migrants, as analysis from the 2001 Census and the LFS showed that *"the distribution of where migrants live by country and region differs from where migrants state they intend to live in the IPS"* (ONS 2007h, p.2). Second, a Poisson regression model was introduced to improve emigration estimates at the LAD level. The model utilised variables available at the LAD level such as population density, prior year immigration figures and variables derived from the 2001 Census (ONS 2007g). The 2007 improvements were applied to produce revised estimates for mid-2002 to mid-2005 and for the mid-2006 estimate going forward. The cumulative effect of the revised estimate between 2002 and 2005 was a net increase of 28,600 migrants in England and Wales (ONS 2007a).

In 2010, a series of improvements were introduced as part of the MSIP programme, most of which are currently used by ONS for the estimation of migration and consequently have been explained already. These constituted an enhanced methodology for the estimation of international immigration, an Irish adjustment and an improved estimate of internal migration by students (covered in detail in Section 3.3.2). The improved methodology for immigration introduced a Poisson regression model for distribution of immigrants to the LAD level which replaced the use of migrant distribution reported in the 2001 Census. The Poisson model for emigration was refined to include a set of fixed covariates, rather than use different variables each year (ONS 2010b). Finally, 2010 saw ONS adopt the NISRA estimates for immigration to and emigration from Northern Ireland, derived from Health Card registrations. By using the Health Card data already utilised by NISRA, ONS sought to *"ensure consistency for users"* (ONS 2010d, p.1). ONS also ceased to use estimates of migration between the UK and the Republic of Ireland derived from the Irish Central Statistical Office, replacing this with an IPS estimate (ONS 2010d).

The revised methodology was used to update population estimates from mid-2002 to mid-2008, and from 2009 onwards. The cumulative impact of the changes to the international estimation methodology between 2002 and 2008 was an increase of 8,300 migrants for England and Wales (ONS 2010f).

3.4.5 NRS and NISRA use of administrative sources

The Interdepartmental Task Force on Migration has also been the catalyst for investigation of ways to improve migration statistics at NRS and NISRA. Since 2006, NISRA has undertaken research to inform the international migration statistics that they produce, drawing on estimates available from a number of different administrative sources. NISRA (2010) identifies the sources that can be used to measure sub-groups of the migrant population, both immigrants and emigrants.

In terms of immigration, workers are identified through the Worker Registration Scheme, the number of NINo registrations and number of applications to work through the Home Office points based system. School children can be identified through the Annual School Census, which identifies a child whose first language is not English and from 2009 asked schools how many children joined the system whose previous address was outside Northern Ireland. NISRA has also assessed HESA data in order to identify higher education students who were domiciled outside of Northern Ireland. Births to mothers and fathers from outside of Northern Ireland can be counted when a new birth is registered, since the mother and father are required to give their country of birth. Northern Ireland Housing Executive data are used to record migrant worker households applying for social housing and finally, the LFS is able to indicate the age structure of the foreign-born population.

A number of sources that measure sub-groups of the *emigrant* population are also identified by NISRA (2010). NINos issued to foreign nationals that fall out of use can be used as a proxy for worker emigration while school children emigrating are measured through the Annual School Census, which asks schools for the *"number of pupils who left Northern Ireland in the previous year"* (NISRA 2010, p.38). Finally, specific out-migration questions have been included in the Continuous Household Survey and Omnibus Survey. The immigration and emigration estimates derived from administrative data only cover sub-groups of the migrant population and are used to quality assure officially published international migration estimates which are based on data from Health Card registrations and de-registrations (discussed in Section 3.2.3).

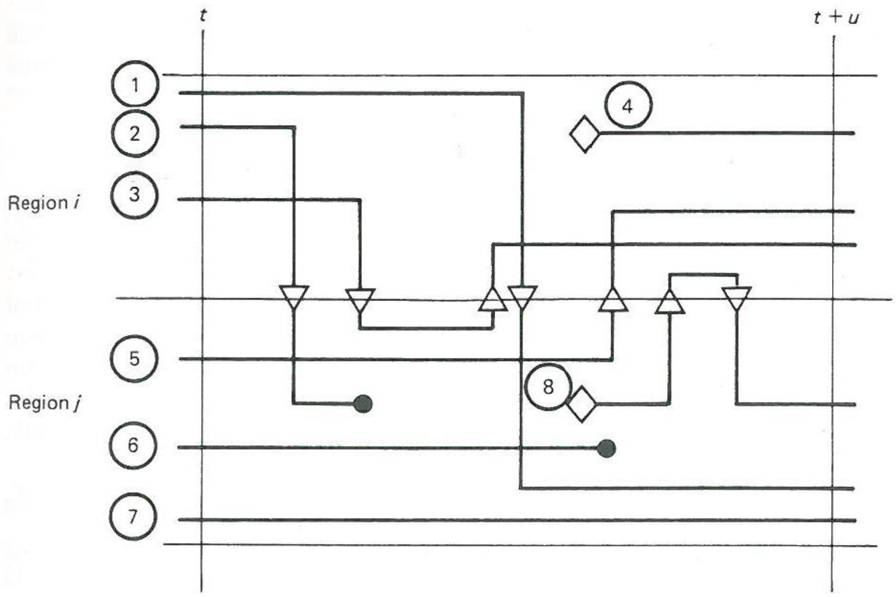
NRS has also undertaken work addressing the potential of using HESA and Annual School Census data to improve the distribution of LTIM (Mueller 2011). However, to date, no revised estimates have been used or published. Rolfe and Metcalf (2009) provide a comprehensive review of data sources available to the Scottish Government. Similar to the work carried out by NISRA, they discuss sources that identify migrant sub-groups highlighting that *"the data is not representative of migrants, but of a self-selecting subset of migrants"* (p.15). Rolfe and Metcalf (2009) suggest the use of the Annual School Census and HESA data to measure immigrants in education and the use of the LFS and WRS for those in employment. They also suggest that the recently launched Integrated Household Survey (IHS) could increase the availability of data on migrants, but will be limited by its sample size.

3.5 Concepts and data coverage

So far this chapter has identified the key datasets used in the estimation of migration at the subnational level in the UK: (S)NHSCR, PRDS, SCHI and NICHI (Health Card data). This section looks at the temporal structure of the data and the populations covered in each of the datasets. This chapter has also discussed the IPS data used in the estimation and distribution of international migrants, but as these data are a sample which is distributed to the subnational level by NHS data sources they are excluded from this assessment. The international estimates produced by the NSAs are used directly in the analysis presented in this thesis, whereas the various NHS sources identified in this chapter are all used in the estimation procedure outlined in Chapter 4 and in Chapter 7, so an understanding of their coverage and structure is important for this estimation process.

Rees and Willekens (1986) emphasise the distinction between movement and transition data, where the former reports a movement when someone changes location, which can occur multiple times within any given time period, whereas a transition is only recorded if a person's location at the beginning of a time period (*t*) and the end of the time period (*t+u*) is different. This distinction is illustrated in Figure 3.6 and shows a variety of lifelines of individuals relating to a time period from *t* to *t+u.* In the context of migration, it is possible to interpret the various lifelines hypothetically by referring to migration data reported in the UK census, which records only those individuals who were in a state of existence at the time of the census (*t+u*) and one year previously (*t*) at a different place of usual residence. Thus, individuals one and five would be recorded as 'transition migrants' as they were in a different region at time *t* and *t+u*. Individual three migrated (moved) twice but the second move was a return to the initial region and therefore neither move would be recorded in the census. Individual 2 would not be counted as a migrant either, because he/she died after moving from one region to another and would not be present on census enumeration day. Thus, whilst the census would have counted two migrant transitions in the period, in fact seven moves took place, one of which (individual 2) was subsequently followed by death and two were associated with someone born during the period who moved to another region and then returned to their region of birth (individual 8).

From the data specification outlined in previous sections, all of these eight movements could have been captured by the NHSCR or SNHSCR as the register is continuous (downloaded weekly by the NSAs), provided the individual reported a change of address to his/her GP every time they moved. All of these moves would not have been captured in the PRDS, SCHI or NICHI however, as the method of comparing a yearly download of the register at time period *t+u* and time period *t* is only capable of identifying transitions: only migrants one and five shown in Figure 3.6 would be picked up. This is the reason that the PRDS in England and Wales and the SCHI in Scotland are constrained to the NHSCR/SNHSCR respectively.



Transitions from the initial state to the final state and movements of individuals represented by lifelines (–). ∇, a move from region i to region j; Δ, a move from region j to region i; ◇, a birth; •, a death.

Figure 3.6: A time-space diagram showing migrant lifelines, from Rees and Willekens (1986)

Coverage in terms of temporal intervals and sub-populations varies between the datasets, and these differences are shown in Table 3.1 and Table 3.2. In both tables, the attributes of the administrative data are presented alongside those of the census. This is because the census provides excellent coverage of the population at a specific time point and will be used in the estimation presented later in the thesis. Table 3.1 summarises the temporal coverage of the datasets: the census provides transition data which is comparable to PRDS, CHI and Health Card data. However, the temporal time frame differs by three months, as the census enumeration year refers to the 12 month period prior to census date in April or March, whereas the mid-year NHS data are reported at the end of June. PRDS, CHI and Health Card data are produced as yearly outputs so changes between one year and the next are counted as migrant transitions. The NHSCR is available weekly, but a rolling mid-year dataset (consistent with the mid-year download of the other NHS data) is used to provide totals with which the PRDS and CHI are adjusted to agree.

Table 3.1: Temporal intervals reported in the available data

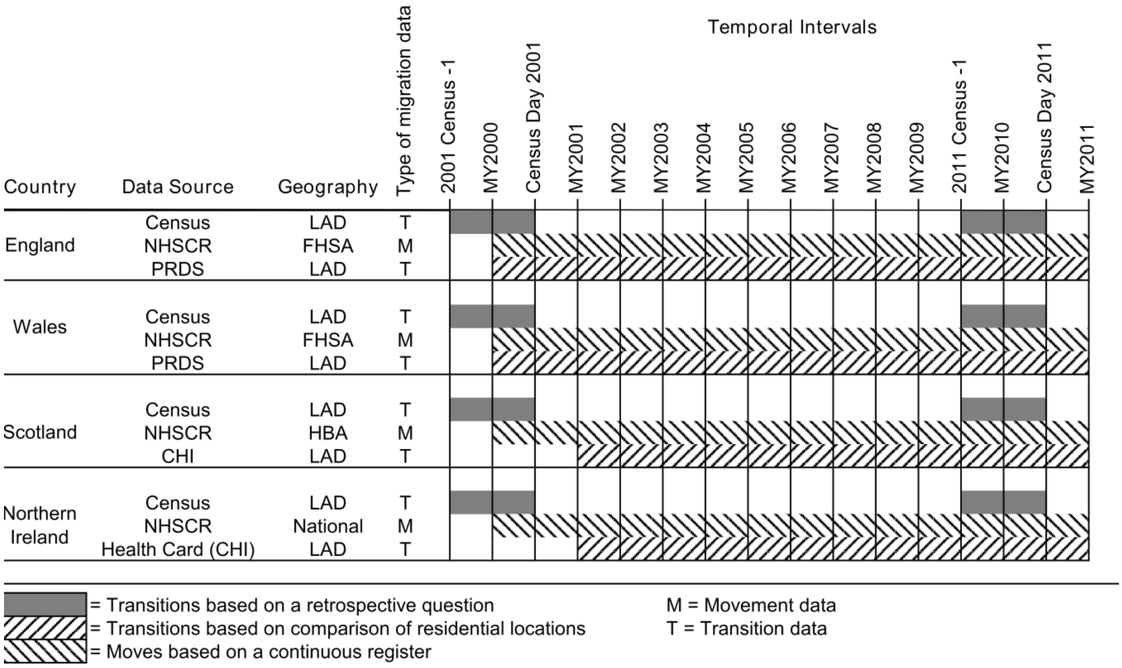


Table 3.2 shows that the UK census of population provides migration information for all sub-populations in the one year prior to the census enumeration date (shown in Table 3.1). These populations are identifiable and sub-settable within the data. All NHS sources undercount young adults, particularly young men, who are often slow to re-register with a GP when they move (ONS, 2010b). For similar reasons, students are undercounted, or counted at their parent's address during term-time. As specified in Section 3, an estimated student adjustment is made by ONS in England and Wales using statistics from HESA, which gives a term time and parental address for all Students in higher education, with a similar adjustment being made by NISRA for Northern Ireland.

Unlike the census, which aims to enumerate all population sub-groups, other migrant populations such as people in prison and in the armed forces are not, as a whole, included in the NHS datasets. These populations are treated separately in the subnational mid-year estimates produced by the NSAs. The exception is that armed forces migrants are included in the to/from the 'rest of the UK' figure reported by NRS for Scotland, which is an inconsistency dealt with in the estimation methodology presented in the next chapter.

Table 3.2: The sub-populations counted in each data source

Migration Data Source	Sub-Populations: NHS Patients: Infant migrants	Young adult	All others	Armed Forces	Prison Population	Students	Others
Census (Transitions)	E	U	Yes	Yes	Yes	Yes	Yes
NHSCR (Events)	Yes	U	Yes			U	
PRDS (Transitions)	Yes	U	Yes	Yes**		U*	
CHI (Transitions)	Yes	U	Yes			U	
Health Card (Transitions)	Yes	U	Yes			U*	

E = Estimate *Adjustment made for sub-population
U = Undercount **Included with all moves to/from the rest of the UK

One final concept needs to be addressed before the data can be used for estimation or analysis: the way that a person's age is reported within the migration dataset. In demography, the Lexis Diagram (Lexis 1875, cited in Keilding 1990) is used widely to understand the age structure of a range of events that occur to people, such as birth, death and migration. In a migration context, DukeWilliams and Blake (2003) devise a version of the Lexis Diagram which is used as the basis of the representation shown in Figure 3.7.

The two grids presented in Figure 3.7 depict age on the vertical axis and time on the horizontal axis. The left hand grid shows data which reports period cohort; while the right hand grid shows how period ages are reported. In the context of data used in this thesis, period cohort is reported in the census, while period age is reported in all NHS data sources (where they are used to produce migration estimates for the MYEs). In the period cohort diagram, the orange block is all migrants aged x+1 sampled at time point t+1 whose location was different at the previous time period (t). The same is true of the blue block, which reports all migrants aged x+3 who moved between time point t+1 and t+2. In the period age grid, the orange block represents anyone who moved during the time period t to t+1 who was aged x+1 to x+2. The blue block represents anyone who moves between time period t+1 and t+2 who was aged between x+2 and x+3. Both data representations are transition data (where a migrant's location is different at the two time points) but the way that age is reported differs.

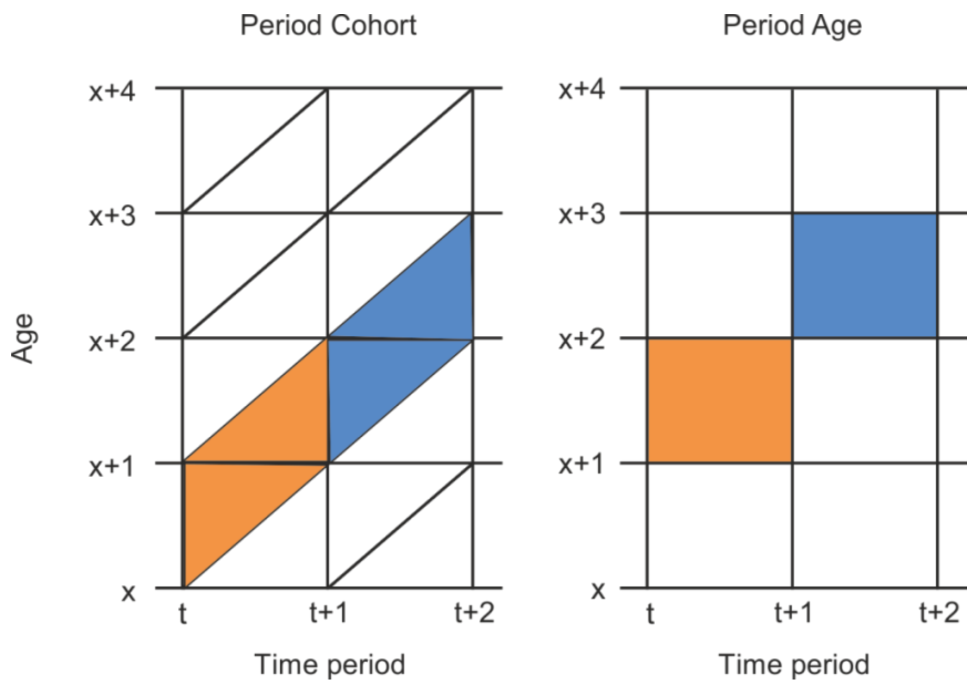


Figure 3.7: An illustration of how migrants are treated in data organised by period cohort and period age

3.6 Conclusion

This chapter has summarised the methodology and data that each of the NSAs use in the production of internal, cross-border and international migration. It has drawn on the large number of methodology documents produced by the NSAs and on evidence collected during meetings and correspondence with statisticians. The final section of this chapter has highlighted the coverage and concepts used in the data sources. Overall, this chapter has served to highlight the complex and disjointed nature of migration statistics in the UK which are created by three different NSAs and draw from a number of different data sources, and has emphasised that the methodology and data availability for migration statistics is constantly evolving. In this context, the next chapter pulls together the available data, and provides details of the methodology used to fill in the data gaps that exist to produce a comprehensive and cohesive UK database of migration.

Chapter 4
Producing a UK wide migration database

After setting out with the explicit aim at the beginning of this thesis to produce a comprehensive migration database for the UK and having established limitations of the existing migration data that are available for the UK in the previous chapter, this chapter details the methodology used to bring together the best available data and the iterative proportional fitting (IPF) routine used to estimate the gaps in an interaction matrix.

The chapter is split into four parts. First, the migration data that are available for the UK will be outlined within the framework of an interaction matrix; second, the IPF routine used to estimate the missing data will be introduced; third, the inconsistencies which are a barrier to the effective function of the routine will be dealt with; and fourth, an assessment of the method will be undertaken. The estimates produced using the methods detailed in this chapter are interrogated in the subsequent two chapters before they are disaggregated by age and sex in Chapter 7.

4.1 The interaction matrix of data availability

Figure 4.1 is a schematic diagram of UK-wide subnational migration, which represents the data that are available for estimating a complete time series database of mid-year to mid-year flows for all 406 LADs between 2001/02 and 2010/11. Rows of the matrix represent origins and columns are destinations, so the leading diagonal cells (represented as AW, BW, CW and DW in Figure 4.1) contain migrations within each LAD for any year in the time series. These within LAD migration flows are excluded from this thesis, given that the focus, as outlined in Chapter 2, is on analysing the redistribution of migrants across the UK and their effect on local authority populations. The data layout presented in Figure 4.1 is known as an interaction matrix because it represents the relationship between origins and destinations (Stillwell and Harland 2010). This schematic is effectively a detailed account of the origin-destination (OD) face of the three face Migration Cube, which was presented in Chapter 2 (Figure 2.2) to conceptualise the data required for compiling a complete UK migration dataset. The notation used within the interaction matrix is described in the following paragraphs and a full listing can be found in Appendix 1.

Origin \ Destination		England LAD 1	England LAD ...	England LAD 326	England Total	Wales LAD 1	Wales LAD ...	Wales LAD 22	Wales Total	Scotland LAD 1	Scotland LAD ...	Scotland LAD 32	Scotland Total	Northern Ireland LAD 1	Northern Ireland LAD ...	Northern Ireland LAD 26	Northern Ireland Total	Rest of the UK	Outside of the UK
England	LAD 1	AW	A	A	AO	E	E	E	EO	G	G	G	GO	H	H	H	HO	U	EE
	LAD ...	A	AW	A	AO	E	E	E	EO	G	G	G	GO	H	H	H	HO	U	EE
	LAD 326	A	A	AW	AO	E	E	E	EO	G	G	G	GO	H	H	H	HO	U	EE
	Total	AD	AD	AD	AT	ED	ED	ED	ET	GD	GD	GD	GT	HD	HD	HD	HT	UT	EET
Wales	LAD 1	F	F	F	FO	BW	B	B	BO	I	I	I	IO	J	J	J	JO	V	FF
	LAD ...	F	F	F	FO	B	BW	B	BO	I	I	I	IO	J	J	J	JO	V	FF
	LAD 22	F	F	F	FO	B	B	BW	BO	I	I	I	IO	J	J	J	JO	V	FF
	Total	FD	FD	FD	FT	BD	BD	BD	BT	ID	ID	ID	IT	JD	JD	JD	JT	VT	FFT
Scotland	LAD 1	K	K	K	KO	L	L	L	LO	CW	C	C	CO	M	M	M	MO	W	GG
	LAD ...	K	K	K	KO	L	L	L	LO	C	CW	C	CO	M	M	M	MO	W	GG
	LAD 32	K	K	K	KO	L	L	L	LO	C	C	CW	CO	M	M	M	MO	W	GG
	Total	KD	KD	KD	KT	LD	LD	LD	LT	CD	CD	CD	CT	MD	MD	MD	MT	WT	GGT
Northern Ireland	LAD 1	N	N	N	NO	O	O	O	OO	P	P	P	PO	DW	D	D	DO	X	HH
	LAD ...	N	N	N	NO	O	O	O	OO	P	P	P	PO	D	DW	D	DO	X	HH
	LAD 26	N	N	N	NO	O	O	O	OO	P	P	P	PO	D	D	DW	DO	X	HH
	Total	ND	ND	ND	NT	OD	OD	OD	OT	PD	PD	PD	PT	DD	DD	DD	DT	XT	HHT
Rest of the UK		Q	Q	Q	QT	R	R	R	RT	S	S	S	ST	T	T	T	TT	YT/ZT	JJT
Outside of the UK		AA	AA	AA	AAT	BB	BB	BB	BBT	CC	CC	CC	CCT	DD	DD	DD	DDT	IIT	KKT

Data available in migration matrix

- T, O, D — Margin total/sub-total available
- T, O, D — Margin total/sub-total not available
- A, ... , P — LAD to LAD flow available
- A, ... , P — LAD to LAD flow not available
- T, O, D — Margin total/sub-total combined 2001/02 to 2005/06

Notation (A-KK) explained in text
Full description of notation available in Appendix A

Figure 4.1: An interaction matrix of the relationship between origins and destinations in the UK, highlighting what data is available

The cells of the matrix labelled A, B, C and D represent inter-LAD flows within each country, shaded according to availability. Cells shaded green denote that data are available, while orange denotes that data are not available. The first data gap can be seen for flows between LADs in Northern Ireland (labelled D). The cells labelled E to P represent the inter-LAD flows which cross the borders of the UK countries, and it is clear that the majority of these flow data are missing. The sub-total margins labelled AO to PO represent total flows out of a LAD to each component part of the matrix, while the sub-total margins labelled AD to PD represent the total flow into a LAD from that component part of the matrix. So, for example, AO represents all migrants moving out of a given LAD in England to the rest of England, while FD represents all migrants moving from a given LAD in England to somewhere in Wales. Grey shading represents data that are available, while blue shading denotes that the data are missing for that sub-section.

The PRDS provides data on the flow of moves from a given LAD in England to somewhere in Scotland (GO) or Northern Ireland (HO) but not which LAD they moved to. Similarly, a move from somewhere in Scotland to a LAD in England (KD) is recorded as is a move from somewhere in Northern Ireland (NO) to a LAD in England, but the specific LAD of origin is not recorded in either case. The same is true for LADs in Wales, but not for Scotland or Northern Ireland, where the sub-total margins are shaded blue to represent the absence of data. Neither the CHI in Scotland nor the Health Card registration system in Northern Ireland report the country of origin for a move from the rest of the UK.

The second to last marginal total of the matrix represents a move to (labelled Q to T) or from (labelled U to X) a LAD from/to somewhere else in the UK, which excludes a move within that particular country. No origin information for an incoming migrant to a given LAD is reported here, nor is the destination of an outgoing migrant, except that they have crossed the border to somewhere else in the UK. So, for example, a move to a given LAD in Northern Ireland from somewhere in the UK, (not including elsewhere in Northern Ireland) is represented by the cells labelled T. Here, grey shading represents data that are available, while red shading denotes that data are missing for one or more years of the time series. Scotland represents the only inconsistency for this sub-total, where between 2001/02 and 2005/06, the rest of UK flow was aggregated with overseas migrants to form an 'outside of Scotland' flow (S and CC for inflow, W and GG for outflow).

The outside marginal total of the interaction matrix represents moves to and from overseas (AA to DD for immigration, EE to HH for emigration) and again is shaded grey or red depending on the data that are available. Apart from the problem of an aggregate figure being reported in Scotland between 2001/02 and 2005/06, all data are available and taken directly from the NSA estimates. The final cells of the matrix represent the sub-total and total for each component part of the matrix (these are labelled with the notation of the sub-section of the matrix followed by a T). So for example, RT represents all moves from somewhere in the UK (but outside of Wales) to Wales. These sub-totals are important in maintaining consistency in the interaction matrix as they provide the control to which all parts of the sub-section must sum.

The data available in the interaction matrix dictates the steps required for estimating a UK-wide matrix, and these steps will be specified in the following sections of the chapter, but stated simply are:

1. Fill in the known sections of the interaction matrix with the best data available;
2. Ensure consistent marginal totals are available for estimation; and
3. Estimate the missing sections (cells) of the interaction matrix using IPF.

4.2 Filling the known parts of the interaction matrix

Where interaction data are available and of sufficient quality, they can be included in the relevant sub-section of the interaction matrix. These data are discussed in detail in the previous chapter, so this section serves to highlight where the data fit into the overall schematic diagram shown in Figure 4.1. First, internal intra-national data are available in both England and Wales (A and B in Figure 4.1 respectively), taken directly from the PRDS. From the same dataset, the flows between LADs that cross the boundary of England and Wales are available, meaning that the first cross-border sub-sections of the matrix can be filled with PRDS data (E and F in Figure 4.1). Moves between LADs within Scotland (sub-section C in Figure 4.1) can be taken directly from the CHI.

Moves to and from overseas are generally available from the NSAs, immigration being represented by cells labelled AA to DD and emigration represented by cells labelled EE to HH. Moves to and from LADs in England and Wales are taken from the best available data produced as part of the subnational immigration and emigration estimates by ONS, while moves to and from Scotland and Northern Ireland are taken from the data produced by NRS and NISRA respectively. The exception is for Scotland between 2001/02 and 2005/06, where the cells labelled CC and GG are estimated in Section 4.6 of this chapter. The control totals for each sub-section (those labelled T) are available from the NHSCR and serve an important role in the checking and balancing of the rest of the matrix.

4.3 Estimating the missing sections of the interaction matrix

Figure 4.1 shows that two main parts of the matrix need to be estimated: first, the internal intra-national flows within Northern Ireland (labelled D); and second, the majority of internal cross-border flows, excluding those between England and Wales (labelled G to P). In both cases, an iterative proportional fitting (IPF) routine can be implemented. IPF is a procedure used to adjust flows in contingency tables so that they are consistent with a set of known marginal constraints. A comprehensive study of the history and application of IPF is provided by Založnik (2011), who emphasises that IPF is a procedure employed across a wide range of disciplines from engineering and transport studies to economics and demography. It is known by different names across the fields e.g. 'Cross-Fratar' and 'Furness' methods in transportation engineering, 'RAS' in economics (Norman, 1999, p.7; Wong, 1992, p.340) and 'raking' in statistics (Cohen 2008). Johnston and Pattie (1993, p.321) conclude that *"other applications have employed different terminology using the IPF procedure as a means to a well known mathematical goal, the maximisation of entropy"*. Entropy maximisation retains the structure of the original contingency table, so the estimated values are the *"maximum likelihood estimates of the unknown values"* (Johnston and Pattie 1993, p.317).

In its classical application (as identified by Bishop *et al.* 1974; Denteneer and Verbeek 1985; Založnik 2011), IPF is used to combine data from two or more sources. The first use of IPF in its classical sense, to fit a contingency table using marginal constraints, is widely accredited to Deming and Stephan (1940) who use the procedure on US census data to extrapolate a 5% sample to the entire population. The initial contingency table is often called the 'seed' as it provides a starting value from which to adjust estimates in subsequent iterations. Consideration of choosing a starting seed value is discussed in Section 4.9. The IPF procedure (after Wong 1992, p.340-341; Norman 1999, p.4) can be expressed as:

$$P_{ij(k+1)} = \left(\frac{P_{ij(k)}}{\sum_j P_{ij(k)}}\right) Q_i \quad (4.1)$$

$$P_{ij(k+2)} = \left(\frac{P_{ij(k+1)}}{\sum_i P_{ij(k+1)}}\right) Q_j \quad (4.2)$$

where $P_{ij(k)}$is the contingency table component in row *i* and column *j* at iteration *k*. Q_i is the row total while Q_j is the column total. Equations (4.1) and (4.2) are employed iteratively and will theoretically stop ('converge') at iteration *m* where:

$$\sum_j P_{ijm} = Q_i \quad (4.3)$$

$$\sum_i P_{ijm} = Q_j \quad (4.4)$$

In practice, the process stops at a pre-defined threshold error (here set at 0.001) or maximum number of iterations (here set at 50), whichever comes first. For the 2001/02 estimate, the 2001 Census provides the initial seed values for $P_{ij(k)}$which are then updated using the marginal in/out totals for the year being estimated. For all years from 2002/03 onwards, the seed value is the prior year's estimated table (so the 2001/02 table is the seed for the 2002/03 estimate). The IPF procedure used to produce the results was operationalised in the statistical software package R, using code developed by Hunsinger (2008) for the Alaska Department of Labour and Workforce Development.

4.4 Using iterative proportional fitting to estimate missing migration data

IPF is a technique that has been widely used in the estimation of missing or incomplete migration data, although it is interesting to note that the approach has not been applied in published research on migration for several years. Previous studies have used the technique to improve existing origin-destination migration flows, to produce estimates for a particular time period where only marginal totals are known and to derive migration estimates for sub-sections of the population. To improve an existing distribution of origin-destination flows, Chilton and Poet (1972) use total in and out marginal totals to estimate the small flows masked by disclosure control for the 33 LADs of London in the 1966 Sample Census (which sampled 10 per cent of the population). Similarly, Rees and Duke-Williams (1997) address suppression of origin-destination flows in the 1991 Census Special Migration Statistics, estimating the missing migration flows using marginal totals and producing a set of revised tables where all subtotals were consistent.

A starting distribution of origin-destination flows can be updated and constrained to marginal totals for a given time period to produce time-series estimates, as summarised by Rogers *et al.* (2003), *"the historical interaction pattern can be imposed onto the current migration patterns using, for example, iterative proportional fitting (IPF)".* Nair (1985), in response to the limitation of many third world countries only reporting lifetime origin-destination migration, uses this distribution in India and Korea to produce one, five and ten year migration matrices based on the marginal totals available. Nair (1985, p.140) concludes that IPF is an approach suited to *"estimating intercensal (usually ten years) migratory flows".* Schoen and Jonsson (2003) use IPF to produce new estimates of interregional migration in the US between 1980 and 1990 as a benchmark against which to test their own estimation methodology. To create origin-destination estimates for sub-sections of the population, Willekens *et al.* (1981) use IPF to derive age-specific flows from an aggregate matrix, as does Willekens (1982). Van Imhoff *et al.* (1997) use IPF to produce a simplified multi-dimensional migration dataset by age and sex.

So why use IPF to estimate the missing flows in the interaction matrix rather than other estimation methods? Chapter 2 provides a literature review of studies that estimate migration data using a wide range of mathematical and statistical methods, but it can be argued that the selection of an appropriate technique for estimating missing data in origin-destination migration tables is largely down to the researcher's preference. Raymer (2007) highlights that log-linear models, gravity models, spatial interaction models, entropy and information maximisation models and IPF are all approaches that have been successfully applied to the estimation of place to place migration flows. He cites Willekens (1983; 1980) as two papers that demonstrate the ‘equivalences’ between all of these techniques The framework is the same, the intention is the same and while the processing is different, the outputs are remarkably similar.

A useful case study in the selection of an appropriate method for estimating migration tables is provided by van Imhoff *et al.* (1997), who favour IPF for modelling a multidimensional age/sex/origin and age/sex/destination dataset for Europe due to the efficiency of the technique when producing a range of model results. They first attempted to use a log-linear approach in the software package GLIM, but found that to run a model *“takes several hours, which is prohibitive for an exploratory analysis”* (p.139). When comparing methods, they conclude that *“the fitted rates of IPF and GLIM are the same. Also, IPF is many times faster”*. For the estimation presented in this thesis, IPF is a suitable approach as consistent marginal totals can be derived for both of the missing sections of the interaction matrix (cross-border and within Northern Ireland migration) and the speed at which the routine can be implemented in *the R Project for Statistical Computing* (widely known as R) allows for efficient estimation for every year across the decade. The necessary tools with which to implement the IPF procedure are available, R is a free package and is flexible and powerful enough for the algorithm to be applied to the specific data requirements of the interaction matrix. This speed and ease of implementation also allows for the estimation of origin/destination/age/sex arrays in Chapter 7, allowing for a consistent methodology to be applied throughout this thesis. Finally, the output from this thesis will be reproducible using the data inputs and R algorithm, and can be replicated in the future when additional data (such as the Special Migration Statistics from the 2011 Census) become available.

4.5 Using iterative proportional fitting to estimate migration in Northern Ireland

Having outlined the IPF procedure, this section shows how it is applied in the case of Northern Ireland to produce a set of estimated migration flows for all years from 2001/02 to 2010/11. Figure 4.2 provides an illustration of the routine on some hypothetical data in any given year (here shown as year x). The first matrix, labelled ‘Start’, shows how the 2001 Census interaction matrix for all 26 LADs in Northern Ireland is inserted as the seed or start value to be adjusted (or the prior year table for estimates from 2002/03 onwards). The diagonal cells which represent moves within the

LAD are set to zero. The orange cells represent the column total for each LAD (the total inflow from the rest of Northern Ireland to that LAD) while the green cells represent the row total for each LAD (total outflow to the rest of Northern Ireland). the column and row totals are available for every year from 2001/02 to 2010/11 from the Health Card data supplied by NISRA, so the 2001 Census distribution (or subsequent table) of LAD to LAD flows is adjusted to agree with the total inflow and outflow for each LAD in that year.

Starting at iteration 1, the seed is first adjusted to agree with the row total, then with the column total. The same happens at iteration 2, and in this example the routine finishes at iteration 3 where the seed values have been adjusted to sum to both the row and column totals. The matrix labelled 'Finish' represents the estimated interaction matrix for a given year. In reality, the full 26 by 26 interaction matrix used in the estimation of migration in Northern Ireland converges at between 12 and 16 iterations.

Start

	LAD 1	LAD 2	LAD ...	LAD 26	
LAD 1	0	6	3	10	20
LAD 2	8	0	10	6	30
LAD ...	9	10	0	9	35
LAD 26	3	14	8	0	15
	35	25	10	30	

Iteration 1

	LAD 1	LAD 2	LAD ...	LAD 26	
LAD 1	0.00	6.32	3.16	10.53	20
LAD 2	10.00	0.00	12.50	7.50	30
LAD ...	11.25	12.50	0.00	11.25	35
LAD 26	1.80	8.40	4.80	0.00	15
	23.05	27.22	20.46	29.28	

	LAD 1	LAD 2	LAD ...	LAD 26	
LAD 1	0.00	5.80	1.54	10.79	18.13
LAD 2	15.18	0.00	6.11	7.69	28.98
LAD ...	17.08	11.48	0.00	11.53	40.09
LAD 26	2.73	7.72	2.35	0.00	12.8
	35	25	10	30	

Legend

(orange)	Column total in year x
(green)	Row total in year x
(white box)	Seed (from 2001 Census)

Iteration 2

	LAD 1	LAD 2	LAD ...	LAD 26	
LAD 1	0.00	6.40	1.70	11.90	20
LAD 2	15.72	0.00	6.33	7.96	30
LAD ...	14.91	10.02	0.00	10.06	35
LAD 26	3.20	9.05	2.75	0.00	15
	33.84	25.47	10.78	29.92	

	LAD 1	LAD 2	LAD ...	LAD 26	
LAD 1	0.00	6.28	1.58	11.93	19.79
LAD 2	16.26	0.00	5.87	7.98	30.11
LAD ...	15.43	9.84	0.00	10.09	35.36
LAD 26	3.31	8.88	2.55	0.00	14.75
	35	25	10	30	

Iteration 3

	LAD 1	LAD 2	LAD ...	LAD 26	
LAD 1	0.00	6.35	1.60	12.06	20
LAD 2	16.20	0.00	5.85	7.95	30
LAD ...	15.27	9.74	0.00	9.99	35
LAD 26	3.37	9.03	2.60	0.00	15
	34.84	25.12	10.04	30	

Finish

	LAD 1	LAD 2	LAD ...	LAD 26	
LAD 1	0.0	6.3	1.6	12.1	20
LAD 2	16.3	0.0	5.8	8.0	30
LAD ...	15.3	9.7	0.0	10.0	35
LAD 26	3.4	9.0	2.6	0.0	15
	35	25	10	30	

Figure 4.2: An illustration of the IPF routine for Northern Ireland (using hypothetical data)

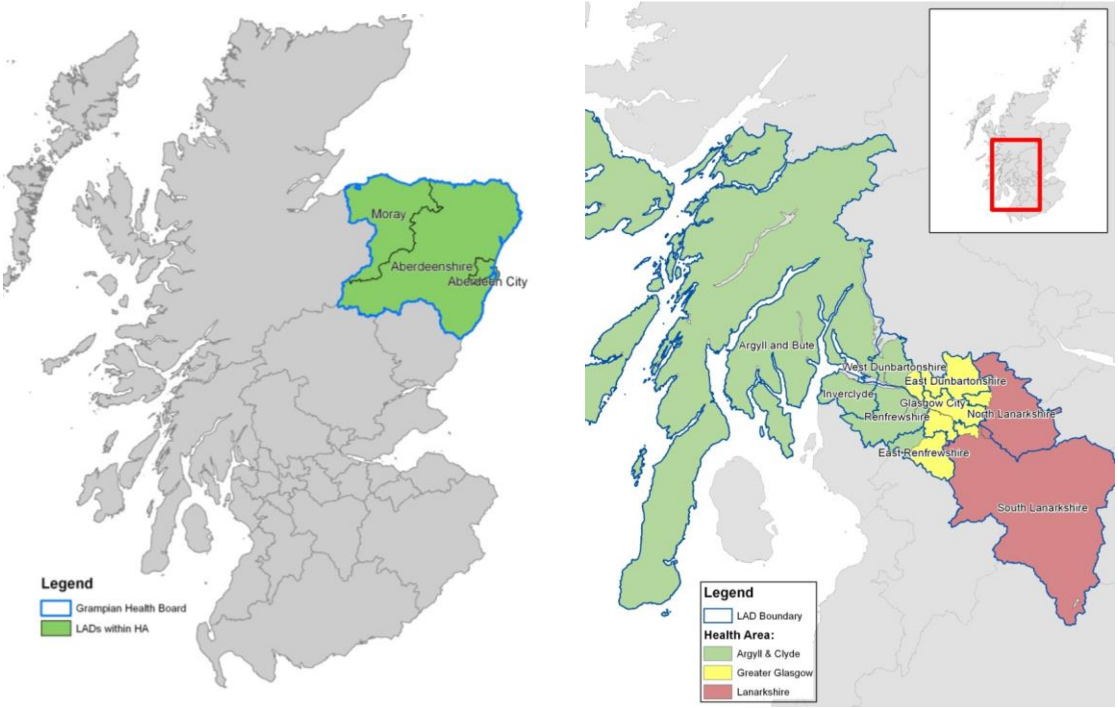
It should be noted that the IPF routine used here produces a final output that contains decimal numbers, rather than 'whole' migrants. This is preferable in terms of completeness and transparency in the data, as the output is an estimate which can easily be rounded to produce a whole number if desired. The analyses presented in this thesis keep all estimated values in their original unrounded format.

4.6 Estimating the missing margins for Scotland between 2001/02 and 2005/06

In order to implement IPF, consistent marginal totals are required for the algorithm to converge. As seen in the previous section, when estimating flows for Northern Ireland this is not a problem, but when estimating rest of UK flows, missing data for Scotland present a challenge for which a solution must be sought. All rest of UK marginal totals (Q to T for the column totals, U to X for the row totals) need to be completed, but prior to 2006/07, the only reported marginal total at the LAD scale for Scotland is an 'outside of Scotland' figure which incorporates both flows between the rest of the UK and overseas (cells labelled S and CC for inflow, W and GG for outflow in Figure 4.1). The methodology was changed by NRS in 2006/07 to enable the flow to be split, but the aggregate flows cannot be redistributed retrospectively (Clarke 2012, pers. comm.). This means that the split between overseas and rest of UK migration included in the 'outside of Scotland' total needs to be estimated.

Two datasets are available for this estimation. First, for every year, a total flow to/from the rest of the UK is available for each HA, which is taken from the SNHSCR and will be used to control the estimates of flows to/from the rest of the UK for each LAD (which nest within HAs). Second, the LAD level proportion of the outside of Scotland flow is derived from post 2006/07 CHI data (for which the UK/non-UK split is reported), and these proportions are used to distribute the SNHSCR data to LADs from 2001/02 to 2005/06. As the CHI migration totals are controlled to the SNHSCR, use of this method ensures that the total proportion of migration allocated to the rest of the UK at HA level is accurate in each year – the challenge lies in ensuring that the distribution to LADs is correct. Figure 4.3a provides an illustration of the method: a total flow to/from the rest of the UK is reported in the SNHSCR for the Grampian HA (outlined in blue) which then needs to be allocated to the LADs within that HA (Moray, Aberdeenshire and Aberdeen) using the average distribution taken from the 2006/07 to 2010/11 data. The same is true for all 13 HAs in Scotland.

Some LADs in the West of Scotland do not nest perfectly within HA boundaries. Figure 4.3b shows how the LADs of West Dunbartonshire, North Lanarkshire and South Lanarkshire are each split between two HA areas, while a small portion of East Renfrewshire crosses the boundary of a third HA. The solution to this problem is to aggregate the total of the three HAs, in effect creating one large HA constraining all nine LADs. This aggregate area still provides a control total for flows to and from the rest of the UK as it is the sum of all three HA level moves.



a: An illustration of the data available at HA and LAD scale in Scotland.

b: Scottish LADs that are split across two or more HAs

Figure 4.3: Data availability and distribution of Scottish LADs across health board areas

The CHI for 2006/07 to 2010/11 (and the census distribution for inflow) for each LAD provide a proportion of the total outside of Scotland migration which can be attributed to UK flows. This proportion is used to split the outside of Scotland figure quoted for 2001/02 to 2005/06 between UK and non-UK migration in each LAD.

This method assumes that an average 2006/07 to 2010/11 UK/non-UK distribution can be applied to 2001/02 to 2005/06 data, which is supported by Figure 4.4 which shows that the UK proportion of the flow is fairly consistent across all LADs between 2006/07 and 2010/11, both for inflow (Figure 4.4a) and for outflow (Figure 4.4b). With five years of data available (six for inflow, including the census), removing the top and bottom value for each LAD reduces the standard deviation of the dataset considerably.

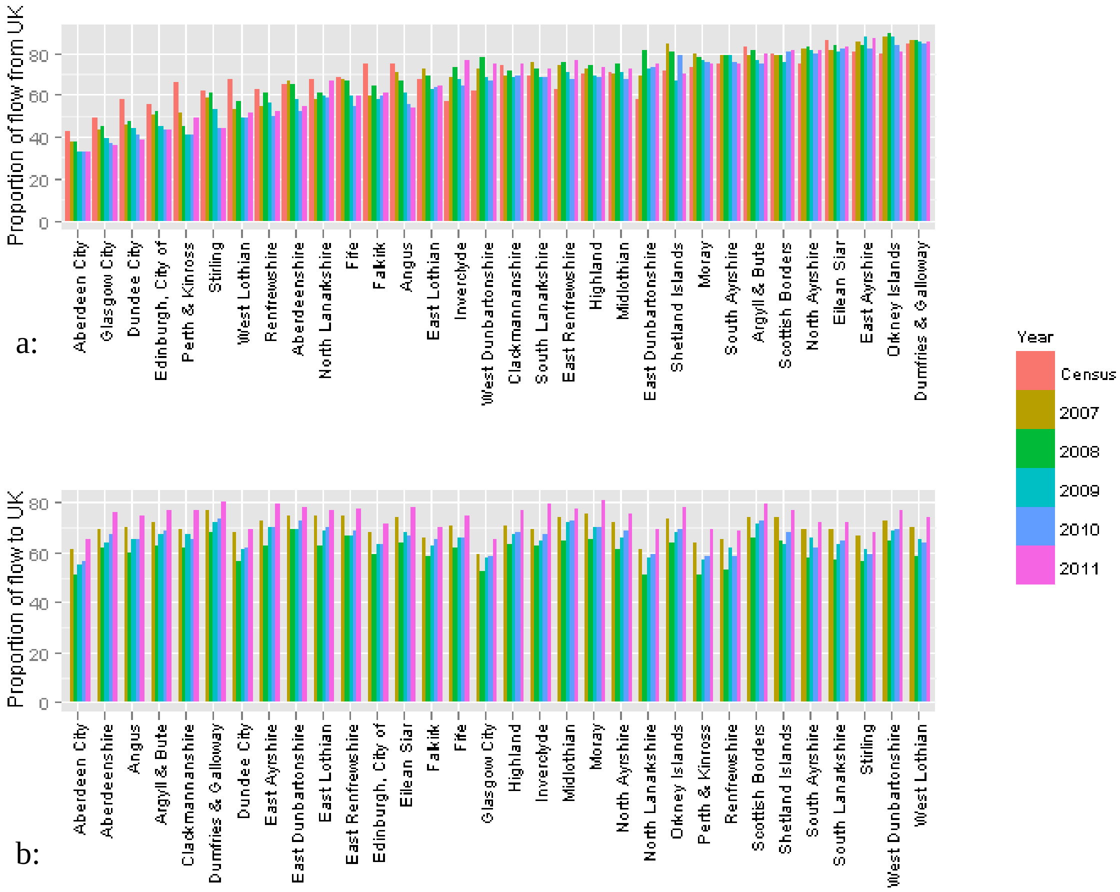
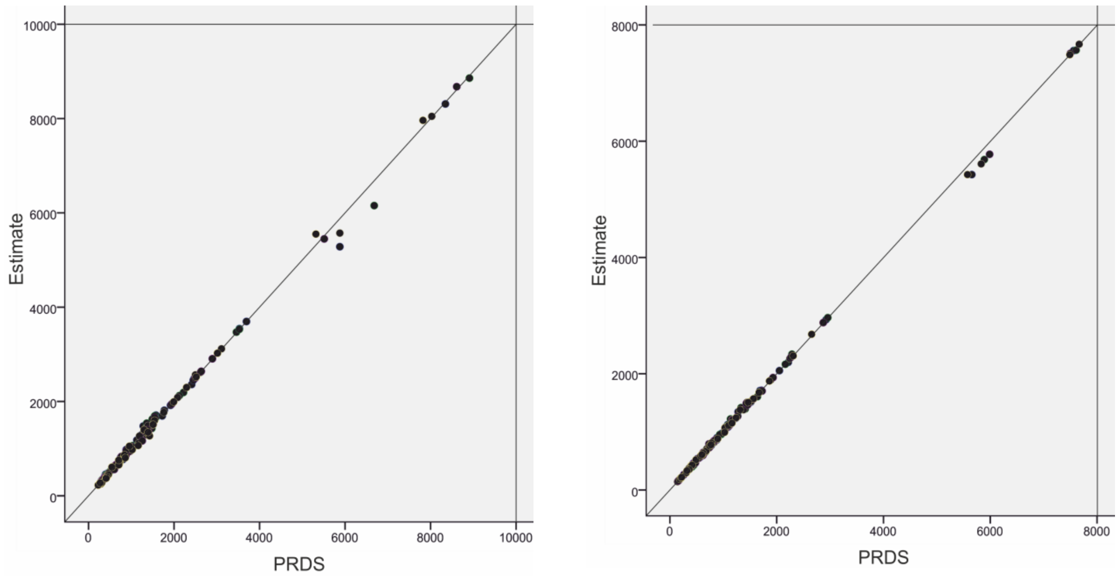


Figure 4.4: The proportion of total outside of Scotland flow that is attributed to UK inflow (4a) and outflow (4b) for 2001 and 2006/07 to 2010/11.

The second step is to take the newly estimated UK flow for 2001/02 to 2005/06 and calculate the proportion of migration that each LAD sends or receives within the HA which it is located in. Finally, the NHSCR total flow to/from the rest of the UK is distributed to the LAD based on its share of flow within the HA. This effectively provides a best estimate of the proportion of migration between the LAD and the rest of the UK while constraining all estimates to the NHSCR data.

The method can be tested by estimating 2006/07 to 2010/11 data and comparing the estimate to the CHI data. The results of this check can be seen in Figure 4.5, where the estimate of inflows (4.5a) and outflows (4.5b) to/from the rest of the UK is compared with the CHI data. The only notable difference is for Glasgow (the distinct cluster at around 6,000 in the PRDS data in Figures 4.5a and 4.5b), where the estimated inflow is 500 people less than the CHI in 2006/07 and in 2007/08 and 300 people less in 2008/09. The Glasgow estimated outflow is between 150 and 200 people less than the

CHI data in each year. The CHI reports an average of 5,800 migrants for inflow and 5,700 migrants for outflow in Glasgow between 2006/07 and 2010/11, meaning that the estimates are up to 8 per cent lower than the reported CHI flow. Given that the estimates are much closer in 2009/10 and 2010/11, that it is not possible to test the accuracy of the 2001/02 to 2005/06 estimates and that the match between estimates and CHI flows for all other LADs is good, no further adjustment is proposed for the estimate in Glasgow.



a: Estimated inflow compared with CHI reported inflow for each LAD between 2006/07 and 2010/11

b: Estimated outflow compared with CHI reported inflow for each LAD between 2006/07 and 2010/11

Figure 4.5: Comparison of estimated vs CHI reported inflow and outflow for LADs in Scotland

Finally, the residual of the ‘outside of Scotland’ figure is taken to be the overseas component (represented as CC for inflow and GG for outflow in Figure 4.1). This residual is agreed to the overseas total reported in the NHSCR for each HA and controlled to the total Scotland overseas migration totals.

4.7 Using iterative proportional fitting to estimate UK-wide cross-border flows

Having estimated the Scotland rest of UK marginal totals for 2001/02 to 2005/06, there now exist a set of consistent totals in each year for which the IPF algorithm can be used to produce estimates of cross-border migration between all LADs in the UK. Ideally one could use the marginal sub-totals for each country to country section of the matrix, but given that these data are not available (represented by all marginal sub-totals shaded blue in Figure 4.1), the solution to use the rest of UK margin for the estimation was suggested by Raymer (2012, pers. comm.) and outlined in this section.

As the UK wide matrix is a closed system where the sum of all moves from one part of the UK to another part should have an overall net effect of zero, the count in the corner cell of the cross-border margin in Figure 4.1, labelled YT/ZT, should equal both total inflows (Q to T in Figure 4.1) and total outflows (U to X in Figure 4.1). This is not the case for two reasons: first, the effect of rounding individual cells to 10 in the ONS supplied PRDS data and second, the inclusion of armed forces moves in the NRS supplied CHI data for Scotland. Moves to and from the armed forces are included in the 'rest of UK' figure for Scottish LADs, but it is not possible to distinguish between an armed forces move within Scotland or armed forces moves to/from another part of the UK. It is the inclusion of armed forces which appears to cause a large proportion of the inconsistency between total inflows and outflows (YT/ZT), as can be seen in Figure 4.6. The comparison for Scotland (light grey bars in Figure 4.6) has been drawn from national level NHSCR data (which do not include armed forces moves) and summing the CHI data (which do include armed forces moves). By taking the difference between NHSCR and CHI, what is left is moves to/from the armed forces for Scotland. These armed forces moves account for the majority of the total difference seen for the UK (black bars in Figure 4.6).

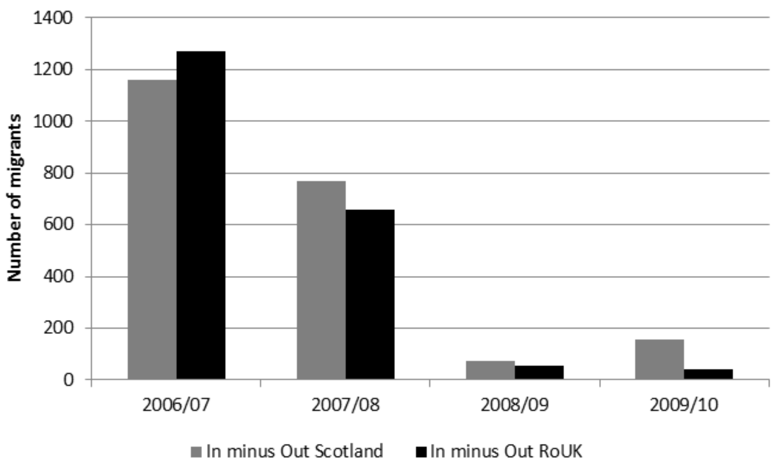


Figure 4.6: A comparison of the difference between origin and destination migration totals for the UK and for Scotland

For the IPF routine to converge, the marginal totals must sum to the same value, so the totals have to be adjusted to ensure consistency. The Scottish data are adjusted to remove the armed forces moves, while the small remaining difference is attributed to the rounding issue in England and Wales. Thus, where $\sum_j D_j$ is total in-migration and $\sum_i O_i$ is total out-migration, if:

$$\sum_j D_j - \sum_i O_i = E \neq 0 \tag{4.5}$$

where E is the difference between total inflow and outflow, then an adjustment needs to be made to ensure the total of all origins and destinations are equal. For all years, total inflow is higher than outflow, so the outflow totals for each LAD in Scotland were adjusted upwards (as were the LADs in England and Wales to account for the small difference in rounding) as follows:

$$\hat{O}_i = O_i + E \times \left(O_i / \sum_i O_i \right) \tag{4.6}$$

where $\hat{O}_i$ is the adjusted outflow for LAD_i. Any error is distributed across origins in proportion to the estimated out-migration total. The error is distributed across origins rather than destinations as the destination totals are more certain in census and survey migration tables because recall bias is avoided. For register based datasets, although this argument does not apply, only the census gives comprehensive coverage of the population groups, so the census logic is followed. Because the 2001/02 to 2005/06 Scottish rest of UK marginal totals are estimated by allocating NHSCR data, no armed forces moves are included so the small difference seen in these years is attributed to rounding differences in England and Wales and adjusted accordingly.

The IPF procedure requires an entire origin-destination matrix, so while there is no need to estimate intra-country flows, or the flows between England and Wales, all cells (A-P in Figure 4.1) need to contain a value. These internal migration and cross-border between England and Wales cell values (A-F in Figure 4.1) are set to 0.001 (the lowest value possible for the IPF routine to work) so that no value is assigned to them in the rest of the UK estimation model. The 2001 Census distribution is then used as the seed value for cross-border migration flows in cells G to P for the 2001/02 estimate, while the prior year's table is used for each subsequent year's estimate. The IPF algorithm is run in exactly the same way as illustrated for Northern Ireland in section 4.5.

4.8 Testing the iterative proportional fitting algorithm on observed data

In order to test the performance of the IPF algorithm, it can be used to estimate a sub-section of the matrix for which there is prior information. In this example, illustrated by Figure 4.7, the cross-border migration between England and Wales can be estimated using IPF in the same way as the whole UK cross-border matrix has been estimated, and compared with the PRDS data which are used to fill sub-sections E and F of the interaction matrix. The PRDS data will be referred to as the 'observed' data, which can be compared with the IPF derived estimate.

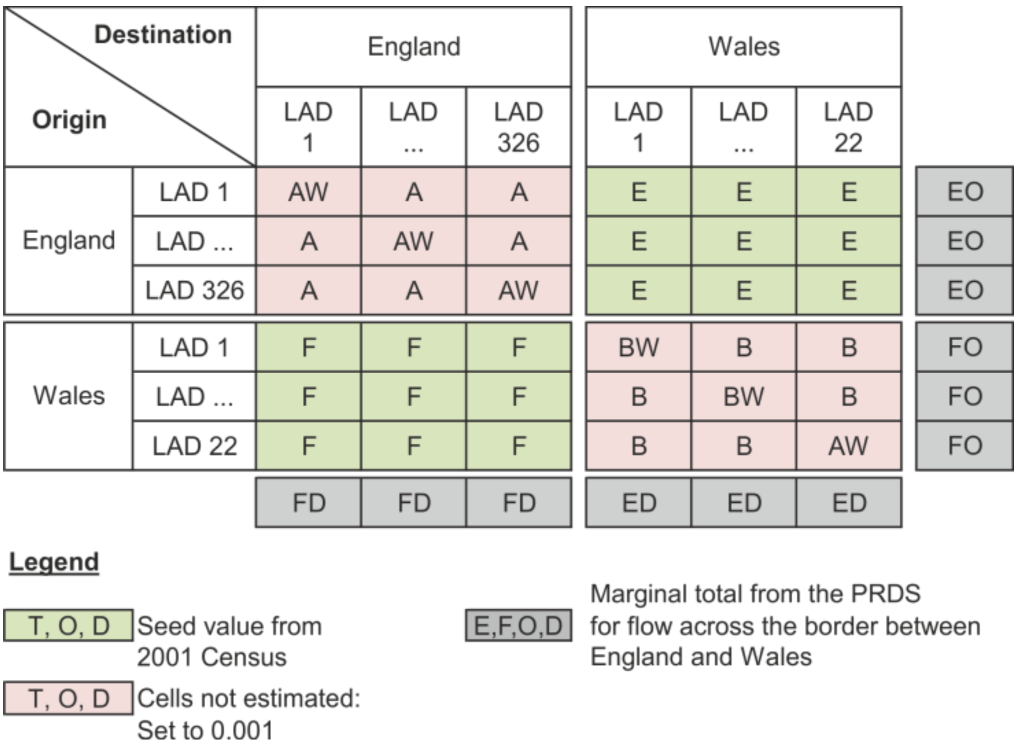


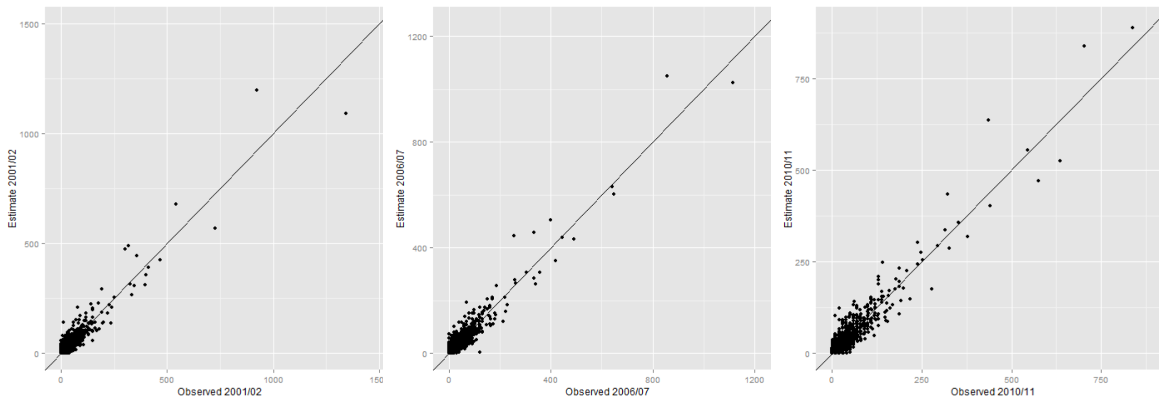
Figure 4.7: An illustration of the cross-border moves between England and Wales to be estimated

Figure 4.7 shows that the cross-border outflow totals (EO and FO) and the cross-border inflow totals (ED and FD) are taken from the PRDS for each LAD in England and Wales, while the cells to be estimated, labelled E and F, are filled with the 2001 Census seed. The intra-country LAD flows which are not required are set to 0.001. When the estimated cell values are compared with the observed cell values, 14,344 pairs of LAD to LAD flows can be compared in each year. A strong positive correlation between observed and estimated migration between England and Wales can be seen in Table 4.1, and the Pearsons coefficient ranges between 0.91 and 0.93 in each year, with all years being statistically significant ($p<0.01$).

Table 4.1: The correlation between IPF estimated and PRDS observed data in each year for migration flows between England and Wales

Year	Correlation
2001/02	0.91
2002/03	0.90
2003/04	0.91
2004/05	0.92
2005/06	0.92
2006/07	0.92
2007/08	0.92
2008/09	0.92
2009/10	0.93
2010/11	0.92

Figure 4.8 shows the correlations in three years: 2001/02, 2006/07 and 2010/11. The majority of flows are relatively small and there are differences where the estimate is above the observed data and *vice versa*.



a: Estimated vs observed data 2001/02

b: Estimated vs observed data 2006/07

c:. Estimated vs observed data 2010/11

Figure 4.8: Estimated vs observed estimates of LAD to LAD migration, 2001/02, 2006/07 and 2010/11

The experiments presented in this section showed that the estimate based on a census seed distribution, PRDS marginals and IPF is very close to the observed data, while not a perfect match.

4.9 Choosing the correct seed

The IPF methodology outlined in this chapter is reliant on a suitable starting distribution for the seed value and assessment of various data sources was made before the 2001 Census distribution was chosen (Figure 4.9). Ultimately, the 2001 Census provides the best estimates of missing cells and generates a solution which can be updated with results of the 2011 Census in due course. This choice of seed relies on the assumption that it is reasonable to use the distribution of migration which is available in the detailed census data and apply it to the subsequent ten year time period. For estimates from 2002/03 onwards, the prior year estimated table was used as the seed, but the results do not differ from using the 2001 Census distribution as it is.

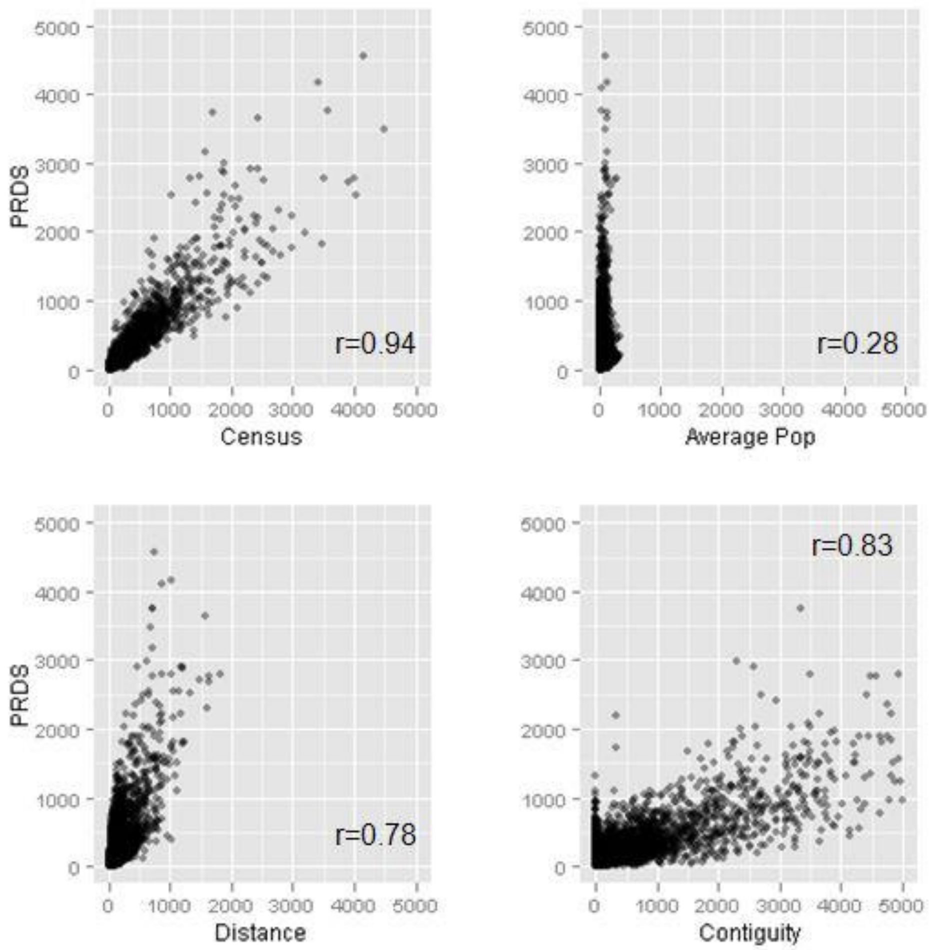
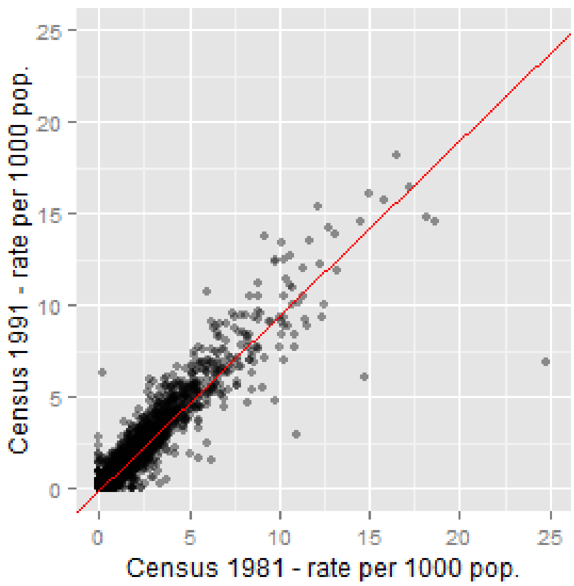


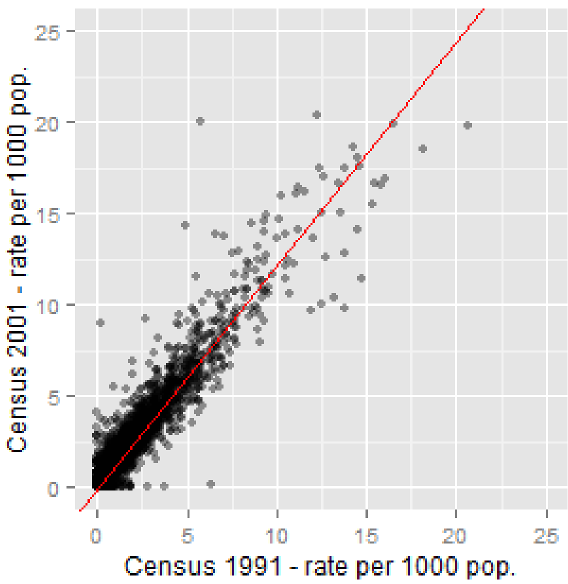
Figure 4.9: The correlation between various seed values and the PRDS data in England and Wales, average of 2001/02 to 2006/07

Figure 4.9 shows that when the estimates created using IPF are compared with PRDS data in England and Wales (where the largest amount of complete data is available for comparison), the 2001 Census distribution is optimal over other starting values. Using the average population size at the origin and destination produces estimates that are far lower than the PRDS data, with a weak correlation (r=0.28). Using the distance between origin and destination LAD tends to produce an under-estimate, while assuming that contiguity is a precursor of migration produces an over-estimate.

The assumption that it is reasonable to take a distribution from the 2001 Census and adjust this for the subsequent 10 year period can be assessed by comparing data from previous censuses; comparisons of all LAD to LAD flows between 1980/81 and 1990/91 flows and between 1990/91 and 2000/01 flows are presented in Figure 4.10. The comparison between 1981 and 1991 (Figure 4.10a) shows a strong positive correlation. When the origin-destination flows for each pair-wise set of LADs in the UK are compared, the r value is 0.97 ($p<0.01$) while the comparison of migration rates (per 1,000 population of destination) shows an r value of 0.95 ($p<0.01$).



a: 1981 compared with 1991



b: 1991 compared with 2001

Figure 4.10: The correlation between LAD-LAD migration rates reported in the 1981, 1991 and 2001 Censuses

Definitional inconsistencies between the 1991 and 2001 Censuses make direct comparison more difficult. These issues have been identified by Stillwell and Duke-

Williams (2007) and Simpson and Sabater (2009) and have been explored in depth by Sabater (2007). The inconsistencies include different treatment of the student population (counted at their home address in 1991 but their term-time address in 2001) and boundary changes between 1991 and 2001. Despite these inconsistencies, the comparison of migration between 1991 and 2001 (Figure 4.10b) shows a similarly strong positive correlation to the 1981-1991 comparison when the total flows are considered ($r = 0.97$, $p<0.01$) and when the rates based on the destination population are used ($r = 0.96$, $p<0.01$).

Overall, this analysis shows that in the absence of any other data, the 2001 Census consistently provides the most robust starting value for estimation using IPF and that it is reasonable to use the distribution for a ten year period, given the strong correlation between three separate census years. The IPF routine constrains the distribution to up to date marginal totals in each year which will control for any outliers seen in the seed.

4.10 Producing consistent geographies

Administrative geographies are subject to frequent change. There are currently 326 administrative areas in England: 56 Unitary Authorities (UAs) which are largely but not exclusively found in medium sized urban areas; 36 Metropolitan Districts (MDs) which represent heavily built up areas outside of Greater London; 201 non-metropolitan Districts and 32 London Boroughs (LBs). Finally, the City of London is a City Corporation, its power is largely consistent with that of LBs, although voting rights differ in that businesses are permitted to vote in local elections, unlike in the rest of the UK. Wales consists of 22 UAs, distinctive from English UAs in that only eight are urban areas, and the remainder are more rural. Scotland consists of 32 Council Areas (CAs) which are also unitary administrations. Northern Ireland is divided into 26 Local Government Districts (LGDs), which ONS refer to as district council areas that are also all unitary administrations, but confusingly have less power than unitary administrations in the rest of the UK. These administrative geographies of the UK are referred to throughout as Local Authority Districts (LADs), and while this definition is not strictly accurate given the complexity of the administrative structure defined above, it is one which is used in this thesis in order to simplify the discussion.

The periodic reorganisation of boundaries and definitions presents considerable challenges when undertaking time series analysis because of the inconsistencies that are created by boundary changes. The latest local government reorganisation took place in 2009 in England with the creation of 10 new UAs. These 'single-tier' entities replaced a number of 'two-tier' systems; a summary of the restructuring process can be found in ONS (2013b). Nine of the new UAs were created by grouping LADs together: Central Bedfordshire, Cheshire East, Cheshire West, Cornwall, County Durham, Northumberland, Shropshire and Wiltshire were created by amalgamating LADs. The Isles of Scilly were separated from Cornwall to form a separate UA for coding purposes. Those LADs involved in the 2009 changes are shown in Figure 4.11.

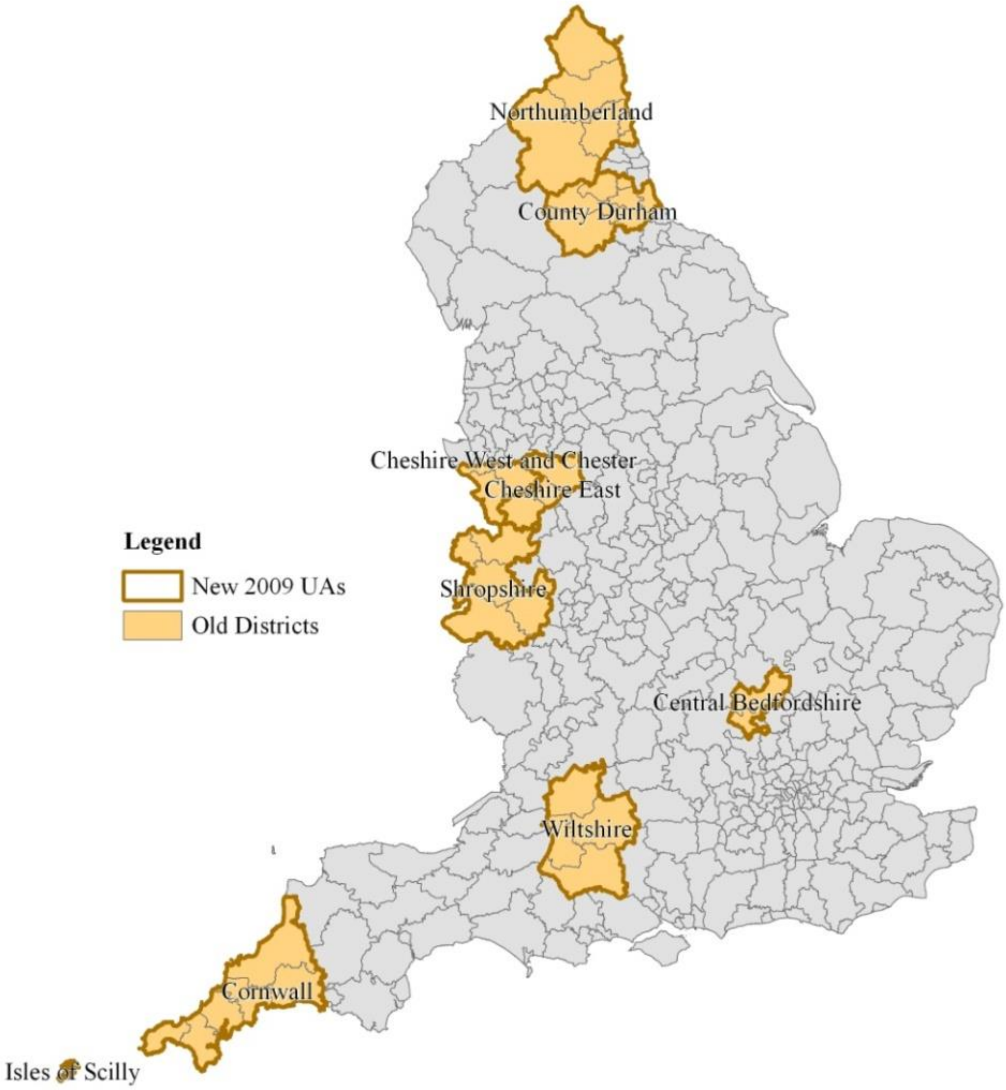


Figure 4.11: Outcome of the reorganisation of local government carried out in 2009

Since the aim of this thesis is to create a consistent time series of migration between mid-2000/01 and mid-2010/11, and disaggregating the data for post-2009 reorganisation would involve large assumptions and be inconsistent with the specification of the MYEs from the NSAs, data for pre-2009 years have been adjusted to agree with the most recent boundaries. This also provides a consistent basis from which the methodology can be taken forward beyond the 2011 Census (at least until further changes occur). To provide an indication of the order of magnitude of the impact of the 2009 boundary adjustments, flow data from the 2001 Census for the areas involved suggests that 23,891 migrants (11,744 males and 12,147 females) that were formerly inter-district migrants in 2000/01 would have become intra-district migrants in that period. These frequent administrative boundary changes in the UK are identified by Norman *et al.* (2003), who advocate the construction of consistent boundaries which enable the analysis of time-series data. Similarly 'geographical harmonization' is a key step undertaken by Rees *et al.* (2004b) in the estimation of small area populations over time. Harmonization of census boundaries is routinely undertaken to enable the comparison of census data over time (Boyle and Feng 2002; Martin *et al.* 2002; Norman 2010).

4.11 An alternative cross-border estimation method

In the previous chapter, the data source identified which provides subnational origin and destination data for migrants that cross the borders of the UK is the NHSCR/SNHSCR. This provides information on flows between Health Authority areas (HA in England) and Health Board areas (HB in Scotland), along with the flow between these health areas and Northern Ireland as a single aggregate area. This section provides a brief summary of a methodology that was developed using these NHSCR flow data, but ultimately was not used in the final estimates presented in this thesis, owing to various problems with its implementation and accuracy. Despite it not making the final cut, a summary of the methodology is presented here for two reasons: first, it builds on and attempts to add a level of disaggregation to a method used successfully by Dennett and Rees (2010) and Dennett (2010); and second, in the development of a final methodology the work carried out on the (S)NHSCR provided an excellent vehicle for understanding the available data (and their limitations) for estimating cross-border migration.

The basic principle of the method is to use the flows between health areas in England, Wales and Scotland and aggregate Northern Ireland reported in the (S)NHSCR in each year to constrain the distribution of LAD to LAD flows reported in the 2001 Census. The estimation equation, adapted from Dennett and Rees (2010) takes the form:

$$M_{ij}^{t} = C_{ij}\left[N_{IJ}^{t} \big/ \textstyle\sum i \in I \mid \sum j \in J \mid \sum Cij\right] \qquad (4.7)$$

where the target LAD to LAD flows (M_{ij}^{t}) at mid-year time point t are estimated by adjusting the 2001 Census LAD-LAD distribution (C_{ij}) by a ratio of the health area origin-destination flow (N_{IJ}^{t}). This ratio is the sum of moves from all LADs in the origin health area ($\sum i \in I$) and all LADs in the destination health area ($\sum j \in J$) where i, j are LADs and I, J are the health areas that constrain them.

Using this method at a different spatial scale, Dennett and Rees (2010) successfully produced estimates of migration between 37 NUTS 2 regions (these are aggregated groups of LADs: 29 in England, 3 in Wales, 4 in Scotland and Northern Ireland as a single area) which were constrained to 13 NUTS 1 regions (the 11 former GORs with Scotland and Northern Ireland as aggregate areas) for calendar years between 2000 and 2007. Dennett (2010) adapted this methodology to produce LAD to LAD flows where the constraint was NHSCR flows reported at former GOR level (again aggregate in Scotland and Northern Ireland). In the work of Dennett (2010), flows at the LAD level in Northern Ireland (both internal and cross-border) were not estimated.

4.11.1 Limitations

Two key geographical limitations exist in the implementation of this methodology. First, there is no data from the NHSCR that is reported at any scale below national level in Northern Ireland. Given that the purpose of implementing this method is to produce a cross-border estimate of migration at the LAD scale, this is a substantial obstacle. A second problem exists in Scotland, where a number of LADs do not nest completely within a single HB area (as described in Section 4.6). While the solution presented above (which involves the creation of a single large area that combines three HBs) is a viable method for estimating marginal totals, it is much less effective when attempting to distribute flows reported in the SNHSCR to LAD level (the result is a large over allocation of migrants destined for Glasgow to more rural LADs).

Another reason for using the main IPF methodology over this alternative is that, given the number of LADs within some health geographies reported in the NHSCR (all 26 LADs in Northern Ireland being the prime example), much of the lower level detail is lost (the ‘rest of UK flow’ reported at LAD scale is not utilised). When the cross-border results from the preferable IPF methodology and the NHSCR adjustment methodology are compared, some large variation exists. Although both methods outlined in this chapter produce estimated results (which cannot be fully verified), the validation of the IPF methodology carried out in Section 4.8 and the major deficiencies to the (S)NHSCR method outlined in this section lend strength to the decision to implement the IPF routine for the dataset. The IPF routine also forms the basis for estimates by age and sex carried out in Chapter 7, a consistency that would not be possible using the available (S)NHSCR data.

4.12 Conclusions

This chapter has detailed the methodology used to estimate the missing sections of the interaction matrix: flows between LADs in Northern Ireland and between LADs which cross the borders of the four UK countries. IPF, implemented in the software package R provided a robust solution for the estimates, using 2001 Census values as a starting seed and adjusting this distribution using up to date information for total in and out migration flows in each year from NHS register data.

A major contribution to the aim of this thesis is the output of these estimations: the dataset comprising internal and cross-border origin-destination flows along with international in and out flows at the LAD scale is therefore available. The following two chapters use this information to analyse the changing pattern of migration between 2001/02 and 2010/11. Chapter 7 revisits the IPF estimation algorithm and explains how it is used to add age and sex information to the migration estimates.

Chapter 5

Understanding migration patterns and processes

The previous chapter has detailed a method for creating a comprehensive UK-wide database of estimates of migration flows between LADs. This chapter presents a review of the extensive literature that covers the patterns of migration and the processes that have underpinned these patterns over the past few decades. Alongside this review, results from the analysis of the migration estimates are presented at the national, regional and LAD level, and established frameworks for migration enquiry found in the literature are used to aid in summarising and interpreting migration trends between 2001/02 and 2010/11. The review and analysis in this chapter is split into six sections: first, the overall numbers of migrants within each of the three migration types (internal, cross-border and international) are noted; second, consideration is given to the temporal consistencies between economic conditions and national level migration rates; third, the focus moves to the subnational level and the flows of migrants taking place between urban and rural areas and the north and south of the UK; fourth, the regional dynamics of migration are investigated; fifth, net patterns of migration at the LAD level are examined; and sixth, a set of migration summary indicators are used to provide further evidence of change throughout the decade. Thus, the chapter establishes the context for migration patterns over the decade before a detailed examination of origin and destination flows is undertaken in the subsequent chapter, and, following an account of their estimation in Chapter 7, the age and sex patterns are analysed thereafter in Chapter 8.

A substantial number of studies look at the patterns of migration in the UK over the past 50 years, and comprehensive reviews of recent migration literature are available in Dennett and Stillwell (2008) and Dennett (2010). This chapter presents a selection of literature that contributes to the discussion of migration in each section in order to complement rather than duplicate the main literature review of this thesis presented in Chapter 2. It also excludes a detailed discussion of the effect of age on migration patterns which is covered alongside the results of age and sex disaggregated estimates in Chapter 8.

5.1 An overview of migration trends

Before embarking on analysis of the migration interaction matrix in the context of established migration theories or examining the subnational patterns that exist, it is necessary to gain an understanding of the general (national level) trend in the number of migrants for each of the three migration streams (internal, cross-border and international). These general trends are presented in Table 5.1 for the beginning (2001/02), middle (2006/07) and end (2010/11) of the estimated time series. Total inflow, outflow and the net result of each type of migration is reported for each of the four countries of the UK.

Table 5.1: Total in, out and net flows for each type of migration by country, 2001/02, 2006/07 and 2010/11

		Internal	**Cross-border**			**International**		
Country	Year	Total	In	Out	Net	In	Out	Net
England	01/02	2,422,040	107,062	-122,423	-15,360	450,747	-302,409	148,338
	06/07	2,566,904	100,540	-118,170	-17,630	530,085	-351,786	178,299
	10/11	2,432,865	98,088	-102,727	-4,638	506,261	-279,049	227,212
Wales	01/02	49,708	64,567	-54,848	9,719	10,533	-8,520	2,013
	06/07	54,010	62,784	-55,756	7,028	18,346	-9,854	8,492
	10/11	53,261	57,034	-54,500	2,534	14,635	-10,278	4,357
Scotland	01/02	118,818	54,408	-49,690	4,717	18,357	-24,400	-6,043
	06/07	117,747	51,542	-42,701	8,840	37,800	-21,000	16,800
	10/11	108,059	43,684	-40,779	2,905	41,000	-16,400	24,600
Northern Ireland	01/02	38,344	12,514	-11,589	924	8,791	-9,613	-822
	06/07	43,251	12,894	-11,131	1,762	19,369	-11,332	8,037
	10/11	36,292	10,322	-11,122	-801	11,414	-13,824	-2,410
UK Total	01/02	2,628,910	238,551	-238,550	-	488,428	-344,942	143,486
	06/07	2,781,912	227,760	-227,758	-	605,600	-393,972	211,628
	10/11	2,630,477	209,128	-209,128	-	573,310	-319,551	253,759

Table 5.1 shows that the majority of the total migration is clearly composed of internal moves for which the net effect is zero and England accounts for a large proportion of the migration in each mid-year to mid-year period. When all migrants are considered, the magnitude of internal migration is over 150,000 higher in 2006/07 than in 2001/02 before it falls back by roughly the same amount between 2006/07 and 2010/11. This pattern is true for England, Wales and Northern Ireland. In Scotland, the first two time periods are relatively consistent but the pattern of decline between 2006/07 and 2010/11 is evident. Wales is the only country where internal migration is lower than cross-border migration (both in and out), which reflects the relationship that LADs in Wales have with English LADs through cross-border migration.

For international migration, total UK immigration and emigration follows the same pattern, with substantial increases in the numbers of both immigrants and emigrants between 2001/02 and 2006/07 (inflow is 117,172 higher in 2006/07 than in 2001/02 while outflow is 49,030 higher). The number of immigrants is 32,290 lower in 2010/11 than in 2006/07 while the number of emigrants falls by 117,172. This pattern of a mid-time period spike is evident for migrant numbers in England, Scotland and Northern Ireland. Overall, cross-border migration falls throughout the decade. These aggregate trends help to inform the patterns discussed in the following sections where aggregate patterns are disaggregated down to LAD level.

5.2 UK internal migration and the economy – a national level analysis

A link between economic conditions and migration propensities is well established in the literature, at least for internal migration, with periods of economic growth coinciding with relatively high migration intensities. Stillwell *et al.* (1992, p.31) highlight the fluctuation in migration propensity between 1971 and 1991, attributing the reduced rate of migration activity in the 1970s to the decline in economic activity in terms of *“changes in the economy on employment, incomes and housing”* where, during the 1979-83 recession, “*migration activity was at its lowest ebb”.* They found that the subsequent increase in the national migration rate from 1981/82 onwards paralleled a decreasing unemployment rate and improving economic conditions. Similar findings are reported by Owen and Green (1992), Oglivy (1982) and by Champion (1987, p. 399), who emphasises that the variability seen in UK internal migration is influenced by *“short term political considerations, as well as business cycles and longer term socio-*

economic developments" which cover a host of variables such as interest rates, mortgage rates and regional employment rates. The impact of recession on migration propensity is revisited in a more recent study by Campos *et al.* (2011) who report that inter-regional and inter-country migration decreased by six per cent in 2008/09 compared with the previous two years, with the largest change taking place in Greater London which experienced a drop of 36,000 people leaving the Greater London GOR in net terms. They attribute this to the unique economic conditions in London, which experienced fewer job losses and lower unemployment rates than many other regions.

Van Der Gaag and van Wissen (2008) address the relationship between internal migration, business cycle indicators, financial variables and labour market developments across Europe at the NUTS2 scale. They found that for all countries, there is a relationship between gross domestic product (GDP) per capita and internal migration which is stable over time and across all countries. Unemployment was found to be significant on its own but not in a pooled model (which incorporated data for all European countries for a year to year time series), which was attributed to multicollinearity with GDP per capita. Stillwell (2005, p.8) concludes that *"the relationship between migration and unemployment remains unclear, depending, in part, on the state of the economy overall"* while Cameron *et al.* (2005) find that for internal migration in England and Wales, the unemployment rate is more relevant than the employment rate, citing the case of the 1990s where regional unemployment rate differentials narrowed more than employment rate differentials as a result of non-participation in poorer regions, including a rise in the number of disability benefits claimants. They suggest that non-economically active people (a category which excludes the unemployed) are less likely to be migrants. Bell *et al.* (2013), in a study of internal migration data from 71 countries around the world, find that GDP per capita has the strongest correlation with migration intensity over both a one and five year time period (as GDP per capita increases, so does migration intensity).

The relationship with economic variables can be tested on the 2001/02 to 2010/11 migration dataset. Figure 5.1 shows the national economic indicators of GDP per capita and unemployment rate for the working age population (here specified as those aged 16-64) alongside the UK internal migration, immigration and emigration rates. For all variables, time series indices are presented with 2001/02 representing the base year (the rate in each year is divided by the rate in the base year and multiplied by

100) so all variables are comparable across the time series. Stillwell (2005, p.7-8) report that *"longer distance migrants tend to have a higher probability of changing their place of work as well as their place of usual residence when they migrate"* and the decision to migrate is influenced by regional economic prosperity meaning that measures such as GDP per capita are important. As all moves in the dataset generated for this thesis are inter-LAD and likely to be longer distance moves, a change of job is also likely to involve a change of house. While housing market variables have been found to be important in influencing migration propensity, their role is clearer at a regional level, so they are excluded from this national level analysis in favour of economic variables.

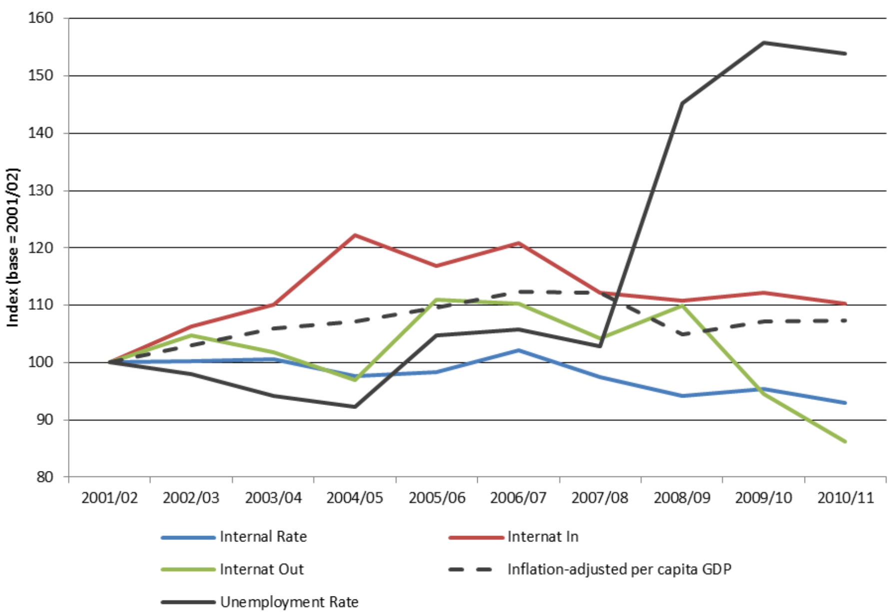


Figure 5.1: Economic indicators for the UK and internal and international migration rates, 2001/02 to 2010/11 indexed to 2001/02 rate

The role of employment as a catalyst for internal migration is apparent with a strong negative correlation between the internal migration rate and the unemployment rate in the same year (r = -0.826, p<0.01, where 10 years of data are compared), this suggests that while national unemployment is low, internal migration is high and *vice versa.* There is, however, no significant correlation between the internal migration rate and GDP per capita, either in the same year or with a one year lag. When trying to interpret these correlations, it is important to remember that unemployment is a variable that affects people directly and immediately, whereas GDP is a combination of economic variables which do not necessarily measure the behaviour or wellbeing of individuals.

However, GDP per capita exhibits a strong correlation with international immigration when figures for the same year are compared ($r = 0.758$, $p<0.05$) but the correlation with emigration is not significant. Looking at the temporal trends shown by the indicators in Figure 5.1, a relationship between unemployment and immigration is apparent (at least up until 2007/08) where falling unemployment seems to coincide with a rise in immigration. The correlation between these variables is, however, not statistically significant. When the migration indices are compared to the prior year economic indices, the only significant relationship is a strong negative correlation between emigration in a given year and unemployment in the previous year ($r = -0.783$, $p<0.05$). This relationship appears to suggest that the emigration rate declines in the year after a rise in unemployment and *vice versa.* At first sight, this relationship seems counter-intuitive; however, the UK is not an isolated system and rising unemployment may well be echoed elsewhere in Europe (and outside of Europe) meaning that migration is less likely if the employment prospects outside of the UK are similarly bleak. This is a view shared by (Dobson *et al.* 2009, p.19) who suggest that as the UK is part of an international network of mobility and the economic downturn is being felt on a global scale *"there will be fewer honey pots to attract mobile people from one country to another".* It could also be argued that migration is a difficult undertaking if a prospective migrant does not have a stable economic base (for example, that prospective migrant became unemployed in the previous year).

A reason for a more tenuous relationship between economic conditions and international migration is presented by Mitchell *et al.* (2011), who develop a Bayesian model of international migration over the past decade and find that the economic cycle of the UK (proxied by the unemployment rate) and other economic determinants play a role but are less important than factors such as immigration policy. Hatton (2005) presents similar findings, suggesting that in an economic model of migration, while improving economic conditions do lead to increases in immigration and better earnings and unemployment conditions do contribute to emigration, it is immigration policy and levels of inequality across the UK which contribute more to the explanation of change. International migration is complicated by the distinction between EU citizens who have freedom of movement between European countries, and non-EU migrants who do not.

The complexities of international migration are discussed by Robinson (2013), who finds that an increase in immigration between 2000 and 2010 has been driven, firstly by an increase in the number of people being granted asylum and, secondly, by the inclusion of EU accession states (A8) from 2003/04 onwards.

In summary, the strong negative correlation between the internal migration rate and unemployment between 2001/02 and 2010/11 is consistent with findings reported in the literature (Owen and Green 1992, Oglivy 1982, Stillwell 1992) whilst a similarly strong negative correlation between emigration and the prior year unemployment rate suggests that unemployment has the effect of dampening migration propensities for people leaving the UK. A strong positive correlation between GDP per capita and immigration is consistent with the migration modelling literature (Mitchell *et al.* 2011, Hatton 2005) and the jump in international immigration seen in 2004/05 is consistent with the accession of new EU states and the policy dimension of immigration identified by Robinson (2013).

5.3 Urban-rural and north-south migration

In this section, the characteristics of migration between two conceptual divides that have been discussed at length in the literature are examined. These are migration between urban and rural areas and migration between the north and the south of the UK. With no definitive definition of either 'divide', the literature is reviewed before analysis is undertaken on the migration dataset to identify patterns of migration between urban and rural locations in the north and south of the UK (which can be considered as four components – urban-rural-north-south) between 2001/02 and 2010/11.

5.3.1 Urban-rural migration

The subnational migration from large metropolitan areas to smaller towns and rural locations, frequently referred to as counterurbanisation, is a predominant theme in the migration literature. Champion (1989b, p.121) charts the trend through the 1960s to the 1980s, describing the *"exodus from cities"* as the *"single most impressive finding of the 1981 Census"*. Champion and Townsend (1994, p.59) describe the 1970s as a decade characterised by counterurbanisation, attributing much of the shift to *"suburban movements that have been forced to become 'extraurban' and inter-*urban because of *pressure on space"*, with Owen and Green (1992) reporting similar findings. Champion (1989c) states that the period 1971-78 saw the most rapid deconcentration of population, in which migration was the most dominant process, with a slowdown in the late 1970s and early 1980s.

Using 1991 Census data, Rees *et al.* (1996, p.78) provide a detailed account of population dynamics in the UK, concluding that the dominant pattern was one of *"deconcentration from the cores of city regions to hinterlands*" and that the strong preference of migrants was for wards with low population densities. The trend of counterubanisation throughout the 1970s and 80s is given detailed attention by Cross (1990), Kennett (1980) and Champion (1989a), whilst the phenomenon in the 1990s is explored by Kalogirou (2005). Similar counterubanisation trends are detected from the results of the 2001 Census by Champion (2005), Stillwell and Duke-Williams (2007) and Stillwell (2013).

The urban-rural migration relationship between 2001/02 to 2010/11 identified in the current work can be assessed by applying a definition of rurality to LADs for the time series migration estimates. Various area classification systems have been used (ONS, NRS and NISRA have their own urban and rural classifications) but a consistent definition across the whole of the UK is not available. To provide a classification that is both consistent across the UK and transparent in methodology, a simple measure of population density has been derived from the 2001 Census to create rural and urban categories, which can be seen in Figure 5.2. The 'rural' areas (green in Figure 5.2) have a population density of between eight and 474 people per square kilometre whereas the 'urban' areas (blue in Figure 5.3) are those with a population density of between 475 and 13,102 people per square kilometre. Each classification contains half (203) of the LADs in the UK.

Population density is used widely as a proxy for the urban-rural dimension (see Stillwell *et al.* 1992, Rees and Kupiszewski 1999) and is an unambiguous and transparent measure which can be applied to any spatial system. The results of this classification will be tested and expanded on when a more sophisticated area classification is used and is picked up again in the following chapter, where the UK is split into 13 city regions, so for now a simple definition which splits the UK in half will provide an unambiguous overview on which to build.

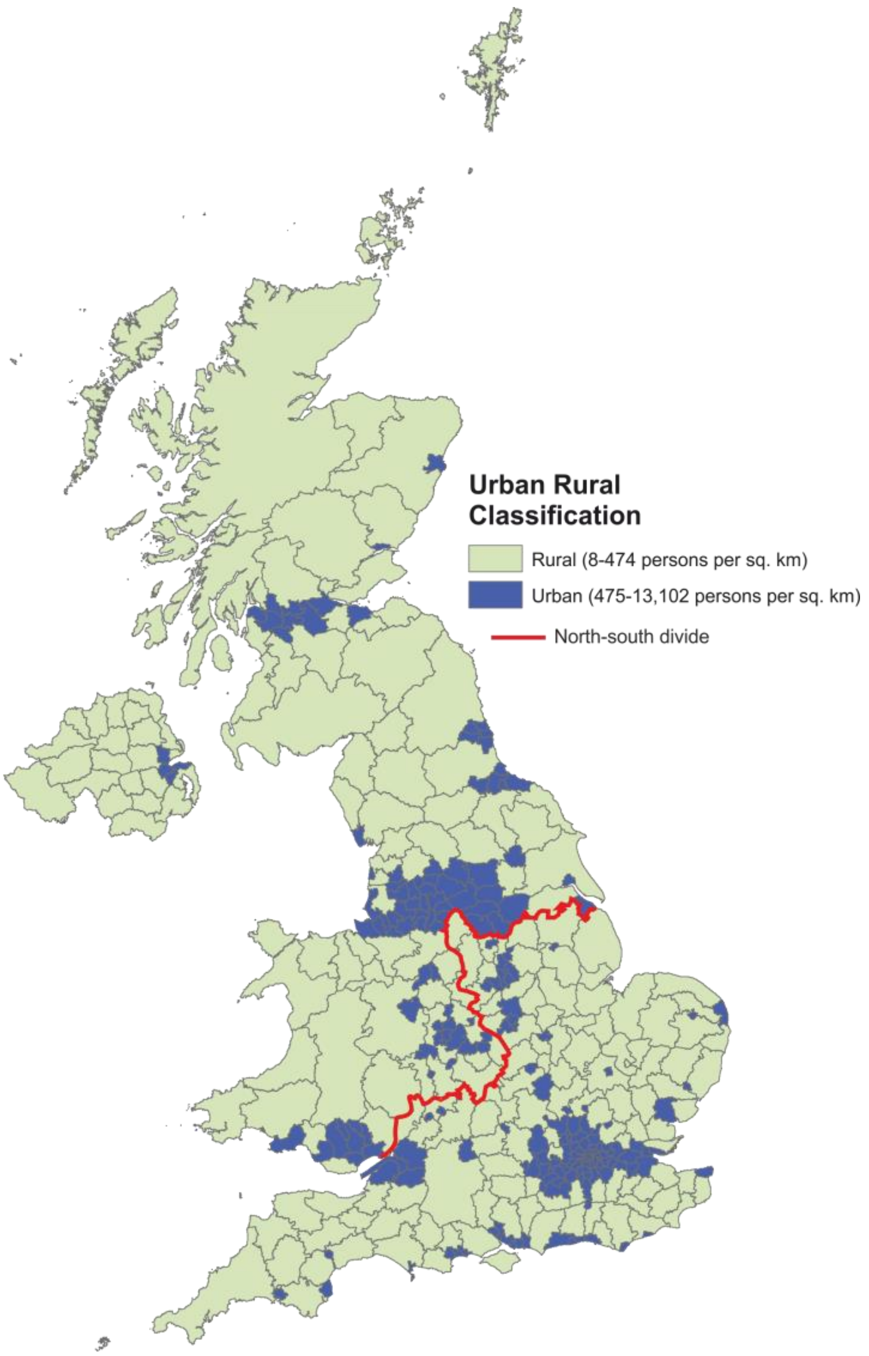


Figure 5.2: Population density by LAD based on the 2001 Census population

Note: See page 99 for a working definition of the north-south divide

Figure 5.3a shows that the general pattern between 2001/02 and 2010/11 is one of net loss from urban areas and net gain in rural areas (as the urban to rural flow is larger than the rural to urban flow) but that this urban-rural flow is in decline. London accounts for a large proportion of UK migration, especially flows between LADs within the Greater London area (which is demonstrated in the next chapter) and to take this into account,

Figure 5.3b shows the flows in both directions expressed as an index of the 2001/02 flows both including and excluding London. The decline in urban to rural migration still holds when London is excluded from the data (the dotted lines in Figure 5.3b), except that the fall from urban to rural is not so accentuated in 2008/09 and the urban to rural migration is a little more stable. This suggests that the changes taking place (the decline in urban to rural migration) is not apparent in London, rather it is being driven by migration from more densely populated LADs outside of London. Between 2001/02 and 2008/09, the net gain in rural areas is falling, driven predominantly by a fall in the urban to rural flow (while the rural to urban flow remains more consistent). There is a brief (and small) resurgence of moves from urban to rural areas in 2009/10, but the trend seen through the rest of the decade resumes in 2010/11. These findings from the migration database are consistent with Rae (2013, p.97) who, in a study comparing small area populations (LSOA) in the 2001 and 2011 Censuses, concludes for England that the inter-censal period *"represents a turnaround from decades of previous population decline"* from metropolitan areas, which is being driven by repopulation of the inner city in particular.

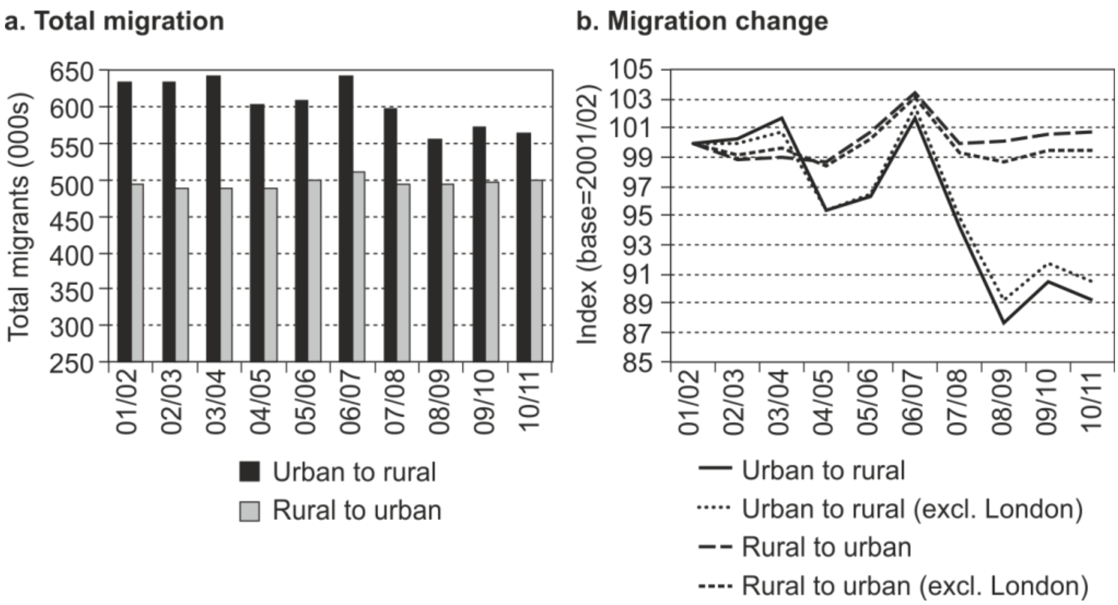


Figure 5.3: (a) Total migration between urban and rural areas and (b) the relative difference with 2001/02

Thus, the pattern of counterubanisation that has been so characteristic of UK internal migration over the last half century appears to have been waning over the last decade, driven primarily by a fall in the number of migrants moving from high density to low density LADs. The timing of this shift (which is most apparent between 2006/07 and 2008/09) coincides with the global financial crisis which resulted in lower GDP, higher unemployment and a slowdown in the UK housing market. This is an important relationship, identified by Rees *et al.* (1996, p.5) who find that in contrast to economic boom seen in the mid-1970s and late 1980s when rising house prices and employment give households the confidence to move out of the main cities, "*recession periods cause house prices to stagnate or fall, remove job opportunities, reduce the gains to be made from migration and increases risk".* This pattern will be explored further in the following chapter and disaggregated by age and sex in Chapter 8 as previous work, using population density as a classification indicator, has revealed different patterns of urban-rural migration by age (Rees *et al.* 1996).

5.3.2 The north-south divide

A divide between the north and the south of the UK (but primarily addressed as a phenomenon in England and Wales) is a theme running through the literature when migration patterns in the 1960s, 1970s and 1980s are being assessed, with London and the South East providing the driving force for migration patterns. The division between north and south is interpreted differently by geographers and the concept is well summarised by Dorling (2007, p.1) stating *"that such an exact line can be drawn is, of course, a fiction but it is also fair to say that moving from North to South is not that gradual an experience".* Green (1988, p.181) defines the divide as *"running roughly between the Severn estuary and Lincolnshire",* much of the literature excludes Scotland and Northern Ireland, whilst the East Midlands is classified as part of the south by Champion (1989b) but is split between north and south by Dorling (2007), where LADs in the counties of Derbyshire and Nottinghamshire are in the north. In the data presented in this chapter, the East Midlands is classified as part of the south.

Champion (1989b) describes the north-south divide as an economic issue, highlighting that the recession of the late 1970s and early 1980s had a more severe impact on the north due to the types of economic sector that predominate, i.e. principally manufacturing industries. Between 1971 and 1986, the overall growth rate of the regions in the south was in excess of the rest of Britain. Champion and Townsend (1994, p.50) identify that the trend for migration from north to south slowed during the 1960s and 1970s, but re-emerged as one of the key features of population change in the 1980s, driven largely by the 'major revival' of the South East since the mid-1970s and the other southern regions in the mid-1980s. Champion and Townsend also comment that, since the 1920s, it appears to be the younger, better qualified people who make the transition from north to south. The concept of London as an ‘escalator region’ proposed by Fielding (1992) fits this model and is picked up in Section 5.5.1. Martin (1988, p.413) argues that the pre-existing economic divide between the north and the south widened in the 1980s due to the policy strategy of the Conservative Thatcher government, where *“wealth creation both requires and generates socio-economic inequalities and differences”*. These inequalities and differences were, Martin (1988, p. 413) argues, preserved by the state as the *“natural order based upon the realities of capitalist production”* which strived towards the creation of private wealth. Whilst Owen and Green (1992) emphasise that a broad trend in migration in the 1980s was movement from the north to the south, Stillwell *et al.* (1992, p.35) report a slowing of the pattern of net gain in the south between 1975/76 and 1986/87, with moves in the opposite direction quickening from 1986/87 onwards, creating net gain in the north. This reversal is attributed to *“shortages of housing, house-price levels, pressures of congestion and increased commuting distances”* and the *“effects of the downturn in the economy being felt earlier in the south than the north”.*

Figure 5.4 shows the north-south divide present in the migration database between 2001/02 and 2010/11. Much of the literature on the north-south pattern excludes Scotland and Northern Ireland, and these moves are represented by the dotted lines on the graph. A more comprehensive, UK wide comparison includes Scotland and Northern Ireland, and these moves are represented as solid lines on the graph. Either way, the clear pattern is for a switch from net gain in the north at the beginning of the time series to net gain in the south at the end of the time series. In the data excluding Scotland and Northern Ireland this crossover occurs in 2007/08, while in the more complete data it occurs in 2006/07, suggesting that their exclusion lowers the magnitude of flows but has little effect on the trends.

The pattern seen here is largely consistent with the pattern of economic indicators seen in Figure 5.1, with the economic shock of 2008/09 having the effect of halting the steady increase of migration from north to south and flattening out the migration in the other direction. This is consistent with literature pertaining to economic conditions and the north-south divide; Champion (1989b) suggests that the pattern of north-south divide was cast in the post-recession recovery period of 1983 to 1986, where the south gained 449,000 extra jobs, while the north gained only 83,000.

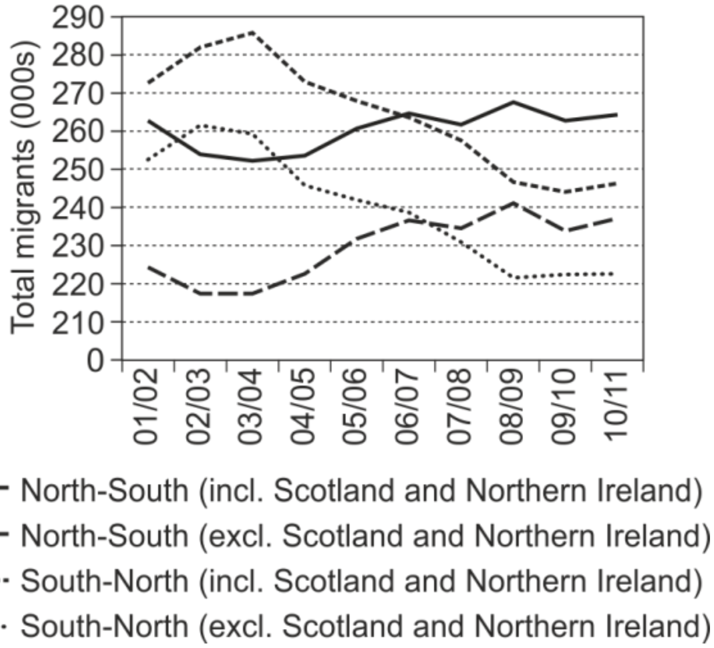


Figure 5.4: Total flows from north to south and south to north, 2001/02 to 2010/11

5.3.3 The effect of an urban-rural and north-south divides on migration rates

With evidence for a reversal in the direction of north-south migration flows mid-decade and the fall in the rate of net urban-rural migration, the two can be combined to give a fuller picture of the changing impact of these established phenomena on migration rates between 2001/02 and 2010/11. Stillwell *et al.* (1992) discuss the counterubanisation pattern of the 1971-81 period using density as a proxy for urbanisation, and finding that low density areas in the south of England experienced an increase in net in-migration between 1980/81 and 1988/89, which mirrored the magnitude of the rate of net out-migration for London. This pattern was found to be less apparent in the north, where low density areas showed relatively small gains. This leads the authors to conclude that *"counterubanisation in the north appears to have been less important than the movement of people from the north to the south"* (p.40). A similar analysis can be carried out on the migration dataset for 2001/02 to 2010/11 by further interrogating the urban-rural and north-south classifications.

The top pair of graphs in Figure 5.5 show the net migration rate, based on the population of the destination LAD, when the LADs are split into urban-north, rural-north, urban-south and rural-south. It is clear from Figures 5.5a and 5.5b that counterubanisation in the south does have a far larger impact on the population in rural areas than counterubanisation in the north, with the rate of migration from urban to rural in the south being twice that of the urban to rural flow in the north. In both the north and the south, the rate of gain for rural areas has declined over the decade; in the south it is 7.6 per 1,000 population in 2001/02 but falls to half this rate in 2010/11. In the north, the rate of gain in rural areas falls from 3.5 to 1.5 over the same period. The pattern seen here is the opposite found by Stillwell *et al.* (1992) for 1980/81 to 1988/89, when counterubanisation was increasing.

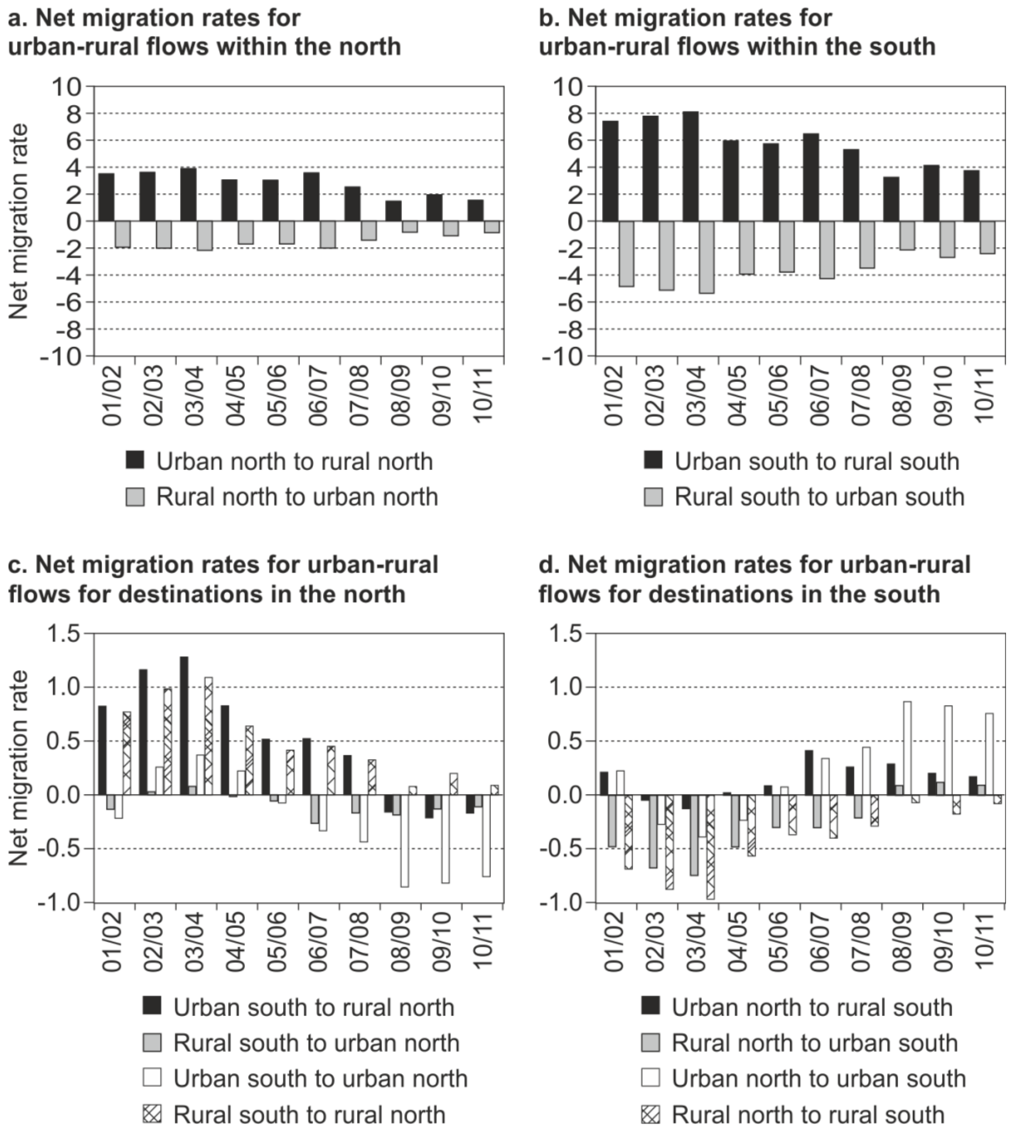


Figure 5.5: Net migration rates (based on receiving population) for urban-north, rural-north, urban-south and rural-south, 2001/02 to 2010/11

The bottom pair of graphs in Figure 5.5 present net migration rates (migrants per 1,000 resident population) between the north and south disaggregated by the density of the origin and destination areas. Figure 5.5c shows net migration rate for moves from the south to the north, and Figure 5.5d shows the moves in the other direction, from north to south. The urban south to rural north net migration rate changes from being positive in the first half of the decade to negative in 2008/09, 2009/10 and 2010/11. Rates of net movement from urban south to urban north show sizeable net losses in the second half of the decade, and particularly from 2008/09 onwards, having been positive in 2002/03 to 2005/06. The net gains in the rural north from the rural south diminished during the decade and rates of net loss from rural south to the rural north became more evident. The trends in rates from a southern perspective are shown in Figure 5.5d. Net migration losses from the urban north to the urban and rural south in most of the early years had been reversed by 2005/06 with gains in urban south from urban north being in excess of 0.5 per 1,000 resident population. Rates of net migration from rural north to urban south also changed from negative to positive during the decade and net losses from rural north to rural south became smaller.

In summary, whilst net urban to rural migration in both the north and the south is still a notable pattern of migration, the rate declined substantially throughout the decade. The reversal of flows from a predominantly south-north to a north-south direction appears to be driven primarily by an increase in migration from the urban north to the urban south. All other flows appear to reduce throughout the decade. The missing flow combination in this analysis is, however, urban to urban migration where the areas are in close proximity (i.e. flows within the urban north and urban south), which is detailed in the context of city regions in the next chapter.

5.4 Regional migration patterns

The national level trends presented in the previous sections mask the differences that occur at the regional and sub-regional scales in the UK. A number of studies suggest that regional migration patterns are largely driven by economic conditions. Inter-regional migration for employment purposes is a process highlighted by Rees *et al.* (1996), Gordon and McCormick (1994) investigate migration between regions in response to regional labour market circumstances, finding that migration is important for the regional adjustment process for non-manual workers, whilst Thomas (1993) finds evidence of preference for migration to areas offering higher wages, both for job and non-job movers.

The patterns of net migration between 2001/02 and 2010/11 presented in Figure 5.6 shows the changing structure of the UK migration system across the decade by England's Government Office Regions (GOR) and the other UK countries. GOR South West is consistently the largest net gainer of migration compared with other regions but shows a trend of declining gains, whilst the South East increases its annual net gains throughout the decade. GOR Yorkshire and the Humber moves from a position of net gain to net loss in 2005/06, as does the North West in 2004/05. Scotland, Wales and the East Midlands consistently gain population, but this net gain declines across the time series. The East has a consistently positive net migration balance, whilst the North East and Northern Ireland have a very small net migration balance which moves from positive at the beginning to negative at the end of the time series. Consistently, the largest net migration balance is net loss from London, although this loss reduces during the decade, particularly between 2006/07 and 2008/09.

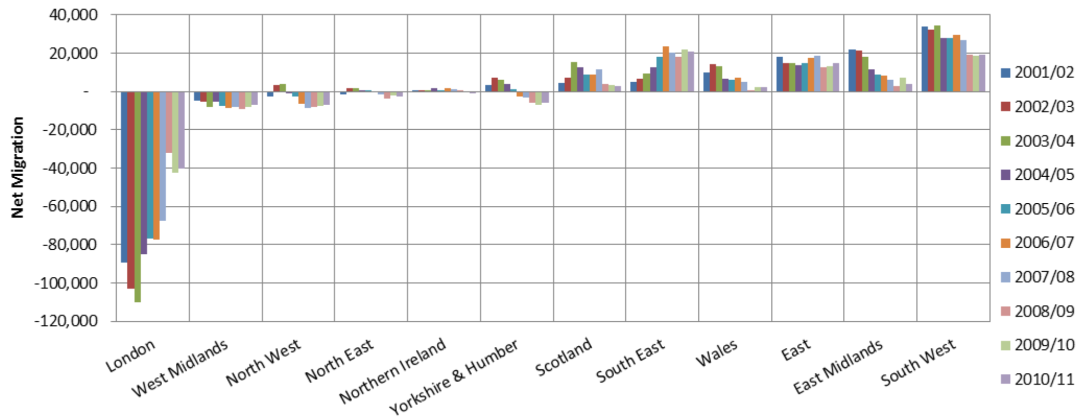


Figure 5.6: Net migration to/from the regions of England & Wales, Scotland and Northern Ireland, 2001/02 to 2010/11

The patterns seen in Figure 5.6 will be picked up later in the thesis, but for the purpose of exploring the literature on regional trends, London and the South East, and Scotland and Northern Ireland will be examined in more detail in the following two sub-sections. London and the South East are generally considered to act as the driving force for UK migration patterns, while Scotland and Northern Ireland receive relatively little coverage but exhibit distinct patterns in the context of UK migration.

5.4.1 London and the South East as an escalator region

The predominance of London within the UK system is emphasised by Champion (2005), who notes that it plays a pivotal role in absorbing international immigrants, and tends to lead economic recovery by being the first region to start redistributing migrants. The primary destinations for out-migrants from London are the South East, South West and East of England whilst attracting in-migrants from across the country. London operates as the powerhouse of the UK migration system, both attracting and generating migrants in large numbers relative to other regions.

The concept of London and the wider South East acting as an 'escalator region' that attracts a large number of young adults from the rest of the country who are (largely) well educated and in the early stages of their career, who then subsequently 'step off' the escalator to move elsewhere having gained the upward mobility offered by the South East, is set out by Fielding (1992). Faggian and McCann (2009a, p.145) emphasise the predominant role of London in this model, suggesting that *"the regions immediately adjacent to London have benefited from human capital spillovers, whereas more peripheral regions are suffering net outflow of human capital"*. The model is critiqued by Champion (2012) who uses time series data between 1966 and 2001 to contest the 'stepping off' phase of the model. Champion suggests that people tend to leave the South East within 15 years, rather than later in their working lives, meaning they are able to actively contribute to the regional economy into which they migrate with the skills they have developed in the South East. Chapter 8 will explore the age dimension of migration in more detail, but it is clear from the analysis presented here that the South East and London in particular play a large role in the redistribution of population in the UK. Indeed, Findlay *et al.* (2009, p.877) suggest that the occupational mobility of the UK workforce makes London *"Scotland's third or Wales's second city"* which has a two-way effect of exchanging highly skilled workers between regional economies while Coombes and Charlton (1992) suggest that London is a 'transit camp', both in terms of a landing point for international immigrants and highly skilled young people from other regions. They emphasise the very high mobility levels of the London population.

Recent work by Champion *et al.* (2013) investigates England's 'second order' cities (those cities in England that are not London) and their emerging role as escalators for migrants, in the same context as London and the South East, where a migrant advances their career faster by moving than staying put. Using Longitudinal Study data for 1991 and 2001, they find that a migration to one of England's second order cities *"raise people's chances of transitioning from WCN to WCC by around ten percentage points on average compared with the longer-term residents of these places"* (p. N/A – early view version); where WCN are White Collar Non-core workers (employers and managers in small firms, ancillary workers) and WCC are White Collar Core workers (a step up in occupational class to employers and managers in large firms and professional workers). They find, however, that the transition from WCC to WCN for all second order cities combined fell short of that seen in London, but that of the second order cities, the rate seen in Manchester was far higher than all others. This finding leads Champion *et al.* (2013) to conclude that Manchester can be seen as a 'mini London' in terms of the employment opportunities it offers. The next chapter investigates the role of England's second order cities (alongside those in Wales, Scotland and Northern Ireland) in more detail.

5.4.2 Scotland and Northern Ireland

Whilst Scotland and Northern Ireland tend to recieve less attention in the migration literature, the unique patterns of migration they experience merit scrutiny here. Jones (1992) highlights two distinctive attributes of Scotland's migration profile: the first is a tradition of overseas emigration and second is low population densities resulting in modest flows between Scotland and adjacent regions of England. He argues that in-migration from the rest of the UK to rural Scottish regions is driven by oil related employment in Highland (especially Aberdeen/Grampian) and residential preference for rural areas. This preference for rural regions has been explored in more detail for the 1970s and 1980s by Forsythe (1980), in the specific case of the Orkney isles by Lumb (1980) and in the case of Mull, Skye and Wester Ross by Jones (1986). Champion (1987) suggests that industry related patterns contributed to decentralisation in the 1970s, citing the North Sea oil boom as a primary example. Rees *et al.* (1996) observe similar patterns in the 1991 Census, referring to the peripheral gains in north east Scotland's 'new resource frontiers' resulting from the development of onshore facilities for offshore gas and oil fields.

Findlay *et al.* (2008) make the connection between Scotland and the South East of England, suggesting that the number of Scottish people in London and the South East has fallen in the 2000s due to the increased level of return migration to Scotland. These migrants are young and educated with Edinburgh their destination of choice, due to the availability of jobs in banking and financial services. This builds on the idea of migrants gaining skills in London before returning to their region of origin. This link is similarly explored by Findlay *et al.* (2002) who address moves from the 'core' of the UK economy to a 'peripheral' region, Scotland, suggesting a strong link between those employed in the service sector in the South East of England and in Edinburgh, whilst recognising the flow of economic migrants in both directions.

Wright (2008) identifies that migrants make up a large proportion of the population in urban areas on the east coast of Scotland (Edinburgh and Aberdeen) as opposed to the Greater Glasgow area which he attributes to the 'economic dynamism' of Edinburgh and Aberdeen in contrast to Glasgow. Wright also highlights the striking difference in the age distribution of migrants to LADs in Scotland, which is picked up in Chapter 8 of this thesis.

Compton (1992) addresses links between Northern Ireland and Great Britain. Looking at a data time series between 1975 and 1990, he finds that the volume of migration fluctuates substantially over time and that inflows and outflows are strongly correlated ($r = 0.81$), with outflow consistently exceeding inflow and the bulk of flows being job-related. He argues that the supply of labour in Northern Ireland has consistently outstripped demand, brought about by rapid labour force growth due to high natural increase. He argues that high unemployment (twice the UK national average) coupled with high net out-migration should be seen as a measure of the *"institutional constraints on labour mobility between Northern Ireland and Britain"* (p.87). For this reason, following an economic recession, *"recovery in Northern Ireland is never sufficiently strong to soak up the available labour supply"* which results in a surge to labour deficient regions in Britain as, he argues, was the case during the economic recovery of the early 1980s.

5.5 Net patterns at the LAD level – making sense of the aggregate trends

In this section, the patterns of net migration at the LAD scale are presented for the three types of migration referred to in the previous chapter: internal, cross-border and international. Looking at the total flows disaggregated in this way, it is possible to decompose some of the national and regional trends identified above. The net migration balances for each of the three years are presented initially and then the net migration rates. The net migration balances illustrate the changing magnitude of migration within the system, but net migration rate is a more useful measure of the effect that migration has on population redistribution at the local level, as it takes into account the size of the population in each LAD. Although the use of net migration means that the changes between component inflows and outflows across the time series are not identified, it does provide a good summary measure of the changing pattern of migration across the decade. In this section data for years at the start (2001/02), middle (2006/07) and end (2010/11) of the time series are used. For the mid-decade analysis, 2006/07 is chosen in preference over 2005/06 as it represents a year in which migration activity was especially high. The gross inflows and outflows that make up the net migration balances and rates shown here are covered in more detail in the next chapter.

5.5.1 Net migration balances

Figure 5.7 shows the pattern of net internal (within each country) migration during each of the three annual periods. The general trend is one of decline in the volume of migrants from the beginning to the end of the decade. Patterns in 2001/02 and 2006/07 are similar, with the same areas losing migrants: most London boroughs, the urban conurbation of the West Midlands, metropolitan LADs in the North West, plus Glasgow, Edinburgh and Belfast. The primary areas of net gain are the LADs in the South West (especially Cornwall), along the south coast and the East of England. Generally the distinction between metropolitan net losses and rural net gains is evident across all three 12 month periods, but is more defined in the two earlier mid-year to mid-year periods. The similarity in pattern seen between 2001/02 and 2006/07 is confirmed by a strong positive correlation between the net flows for all LADs in the two time periods ($r = 0.89$, $p<0.01$), suggesting that the same LADs are losing or gaining a similar number of net migrants.

A shift in the pattern can be seen to have taken place by 2010/11, however, which is confirmed by a weaker correlation between net flows at the beginning and end of the decade ($r = 0.79$, $p<0.01$). The familiar pattern of urban losses and rural gains continues, but with a much smaller net balance for most LADs. This shift is particularly apparent in London (where boroughs in the east are now gaining migrants) and Glasgow, Edinburgh and Belfast which now are losing far fewer migrants to the rest of Scotland and Northern Ireland respectively. In Wales, two predominant LADs for redistribution of migrants in 2001/02 and 2006/07, Cardiff (a net gainer) and Swansea (a net loser), show very little net migration activity in 2010/11. The pattern of net gain in Wales is similar in 2010/11 to previous years but the volume of net incoming migrants has reduced dramatically.

Cross-border migration patterns appear to change substantially between the start and end of the time series (Figure 5.8). The correlation between net flow for all LADs between 2001/02 and 2006/07 is 0.77 ($p<0.01$) and is lower between 2006/07 and 2010/11 ($r = 0.65$, $p<0.01$). The pattern seen at the beginning and end of the decade shows a positive correlation which is significant but weaker still ($r = 0.64$, $p<0.01$). The pattern evident in Figure 5.8 is one of net gain in rural Wales and Scotland, as well as substantial net gain for Belfast in Northern Ireland. Glasgow and its surrounding LADs lose migrants across the border, as do LADs around Belfast. Overwhelmingly the pattern of exchanges between LADs in England and the other UK countries is one of net loss from England. The map for 2010/11 shows a decline in the size of the net loss in English LADs if not a change in the pattern, although the net gain restricted to central London in the earlier time periods spreads to a number of outer London boroughs. The gain seen in the north east of Scotland in 2006/07 has been replaced by a net loss in 2010/11.

Figure 5.9 shows that in contrast to internal and cross-border migration, where the largest change is evident in the last year of the time series, international net migration sees the biggest change between 2001/02 and 2006/07: the correlation between net flows at the LAD level for these two years is 0.73 ($p<0.01$) whereas the correlation between 2006/07 and 2010/11 is 0.86 ($p<0.01$). The most striking change between the beginning and end of the decade is the change for Scotland from a position of large scale net loss to one of net gain for international migrants. Small net gains in Glasgow and Edinburgh in 2001/02 become large net gains in 2010/11 and Aberdeen moves from a position of heavy net loss to net gain. In England, the pattern changes from one where the majority of LADs were losing net migrants in 2001/02 to one where most are gaining in 2010/11, with a clear pattern of net gain that originated in London in the 2001/02 data beginning to spread across the South East. In Northern Ireland, Belfast, after a brief period of net gain in 2006/07, returns to having a negative balance in 2010/11.

The extent to which the pattern of net international migration is opposite to that of net internal migration can be seen by comparing Figures 5.7 and 5.9, and is most

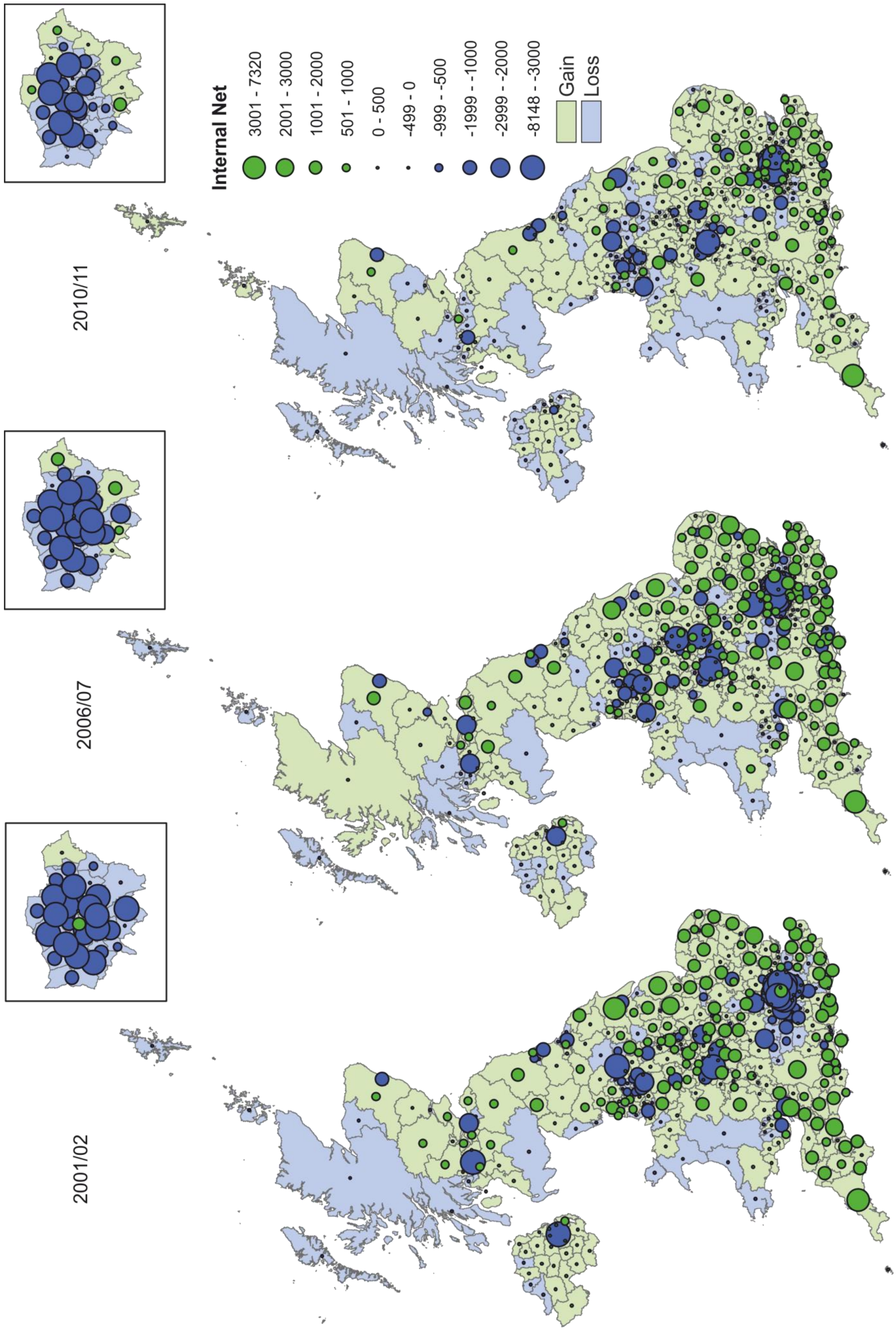


Figure 5.7: Net internal migration balances, LAD level, 2001/02, 2006/07 and 2010/11

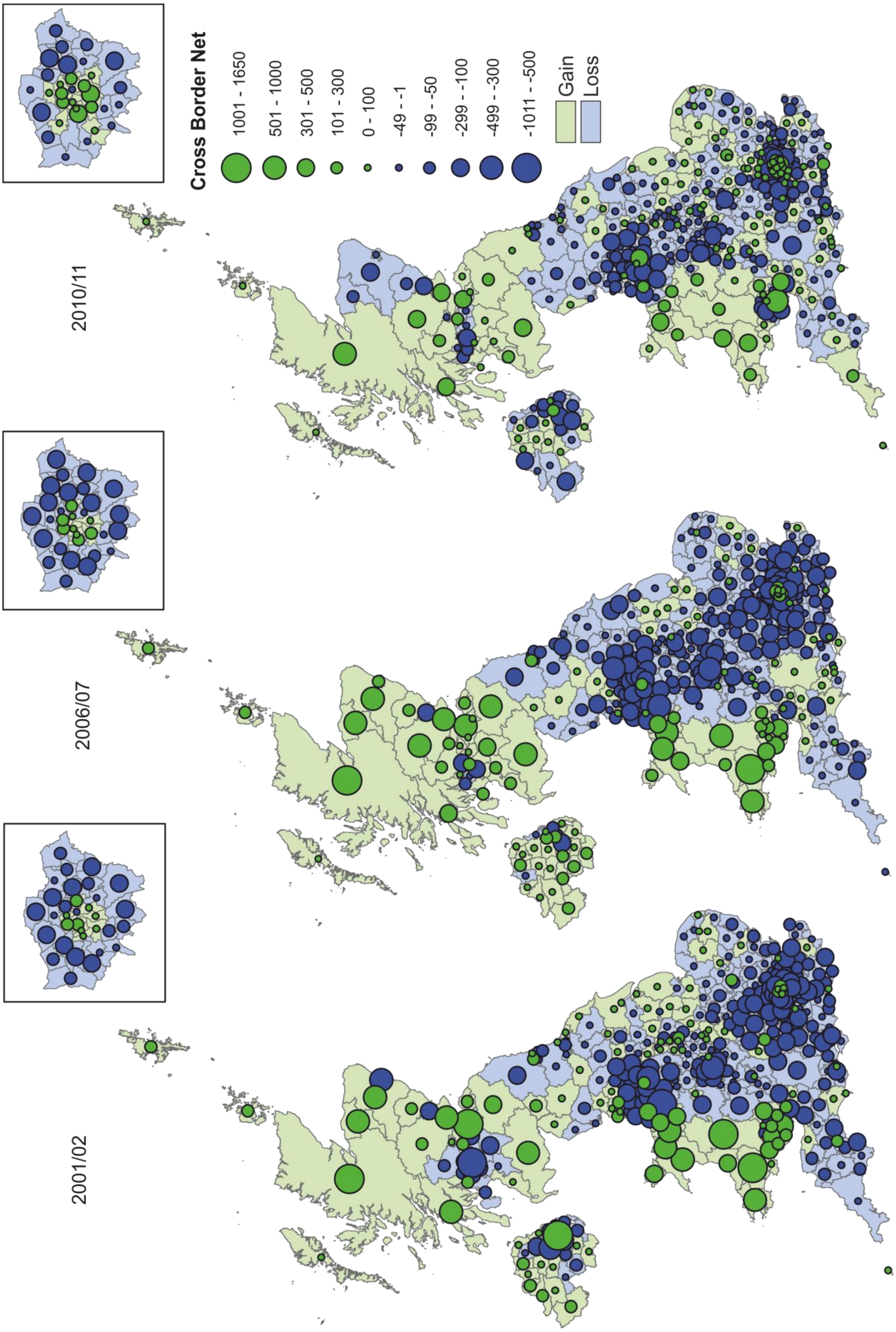


Figure 5.8: Net cross-border migration balances, LAD level, 2001/02, 2006/07 and 2010/11

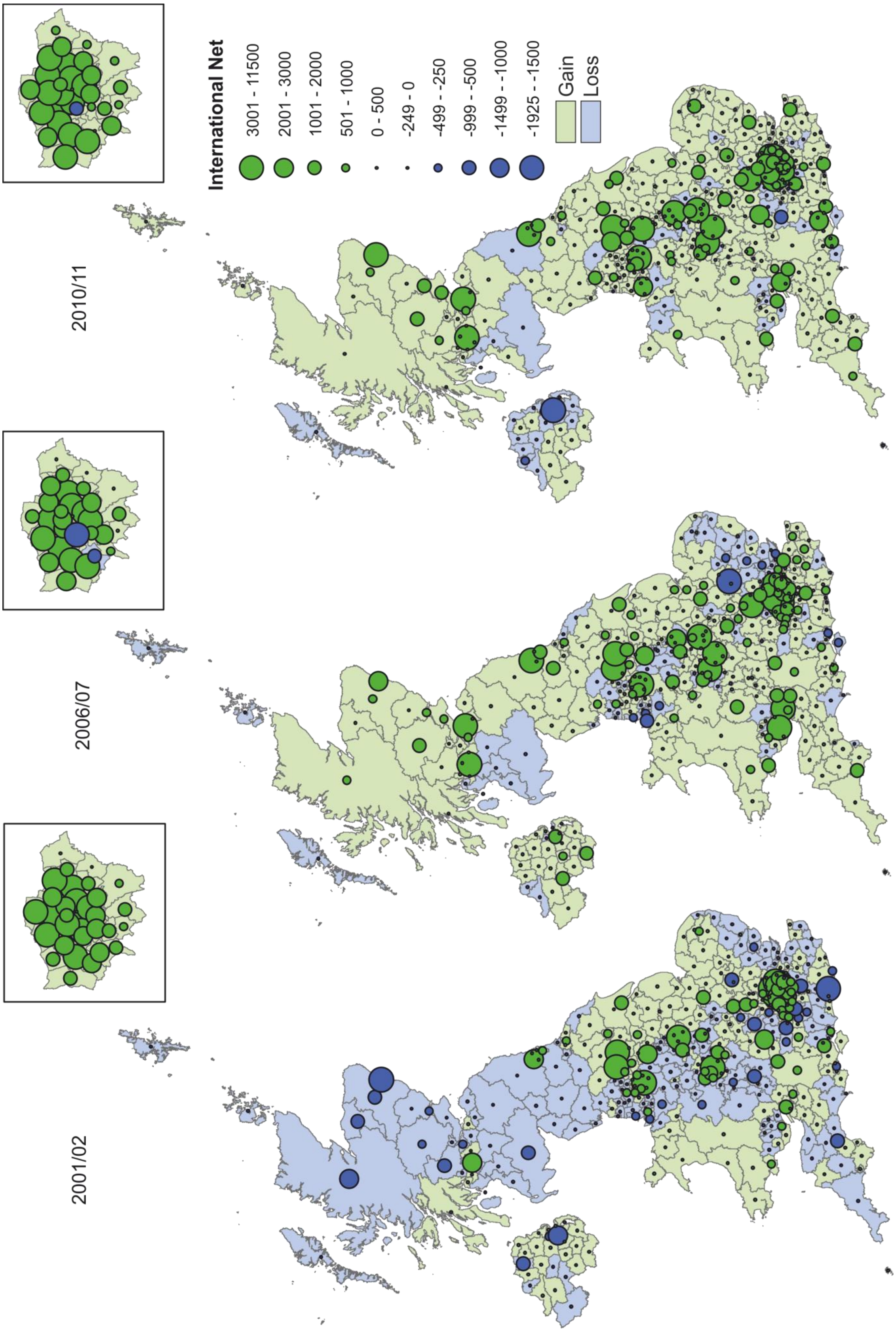


Figure 5.9: Net international migration balances, LAD level, 2001/02, 2006/07 and 2010/11

clear in London and other urban LADs (for example, in the North West) which have net gains of international migrants and net losses of internal migrants. This is consistent with the assertion of London as a transit region (Coombes and Charlton 1992). The negative correlation in each year shows that the relationship holds across the decade ($r = -0.67$ in 2001/02, $r = -0.74$ in 2006/07 and $r = -0.66$ in 2010/11, all $p<0.01$). In contrast, there is no significant relationship between cross-border migration and international migration.

5.5.2 Net migration rates

Looking at net migration rates provides the opportunity to identify patterns in migration across the decade when the balances, seen in the previous section, are standardised for the size of the resident population in each LAD. Figure 5.10 shows the net internal migration rate in 2001/02, 2006/07 and 2009/10. In each year, the rate of loss from the majority of urban areas is more comparable than when the balances are used. The exception is for London boroughs, which exhibit disproportionately high rates of net migration loss in comparison to other LADs. Overall, the fall in net migration rates between 2001/02 and 2010/11 (both loss and gain) is evident in Figure 5.10. This pattern of falling rates is confirmed by the correlation between 2001/02 and 2010/11 ($r = 0.63$, $p<0.01$). The largest fall appears to occur between 2006/07 and 2010/11 where the correlation is 0.76 ($p<0.01$), compared with a stronger correlation between 2001/02 and 2006/07 ($r = 0.81$, $p<0.01$).

Comparison of net migration rates for cross-border migration (Figure 5.11) reinforces the assessment that the most substantial impact can be seen in Scotland, Wales and Northern Ireland, where high rates of net gain can be seen in most LADs. In contrast, net rates in England (both positive and negative) are uniformly low when compared to the other countries. Figure 5.11 also highlights the impact of cross-border migration on Glasgow and Aberdeen, which both exhibit a higher rate of net loss for cross-border migrants than internal migrants in 2001/02 and 2006/07. Similar to internal migration, the fall in rates for cross-border migration is evident in 2010/11, and confirmed by the correlation between 2006/07 and 2010/11 ($r = 0.50$, $p<0.01$) which is lower than the correlation seen between 2001/02 and 2006/07 ($r = 0.74$, $p<0.01$).

Net international migration rates shown in Figure 5.12 highlight the role of London as a destination for international migrants, where net rates of gain for most

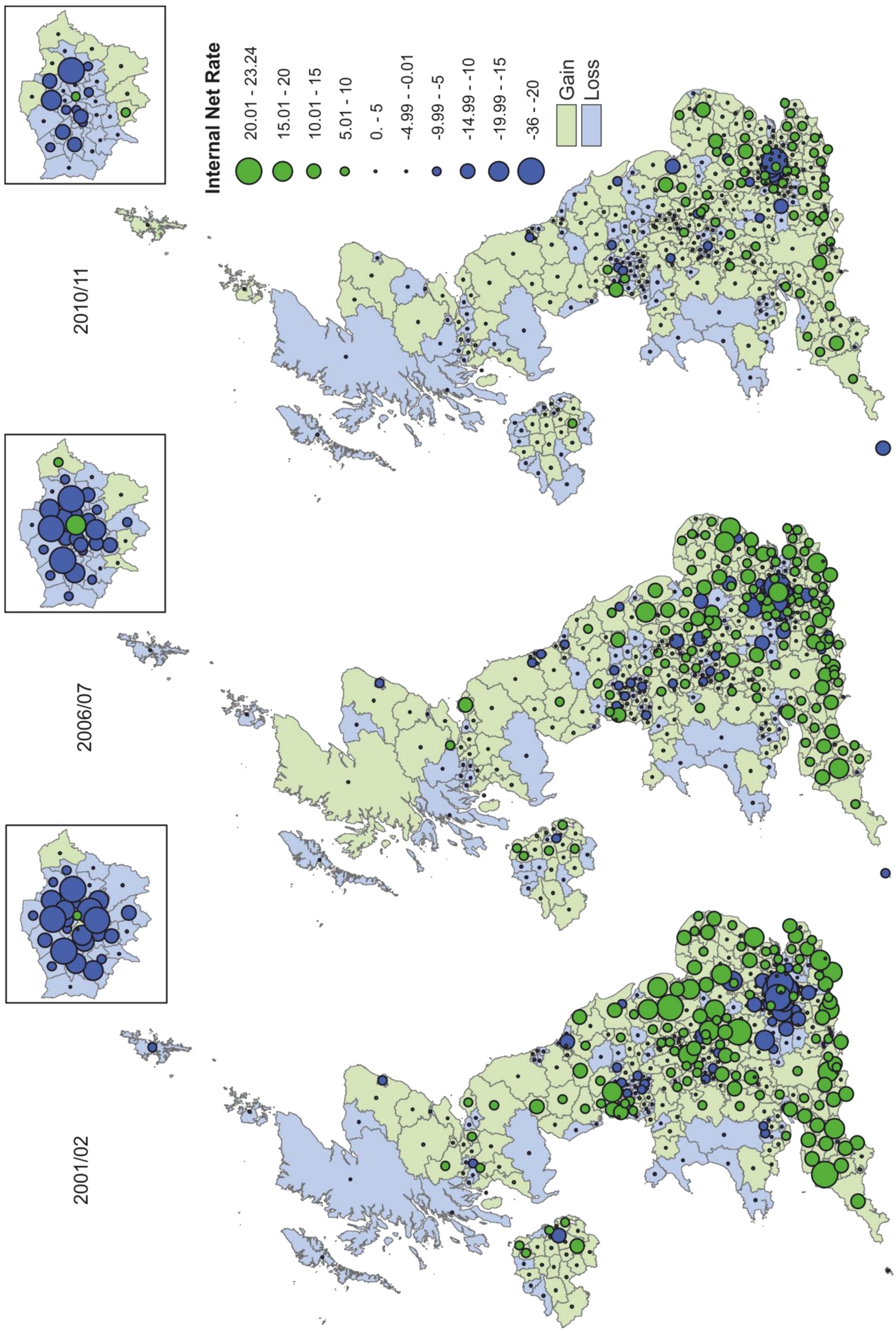


Figure 5.10: Net internal migration rates, LAD level, 2001/02, 2006/07 and 2010/11

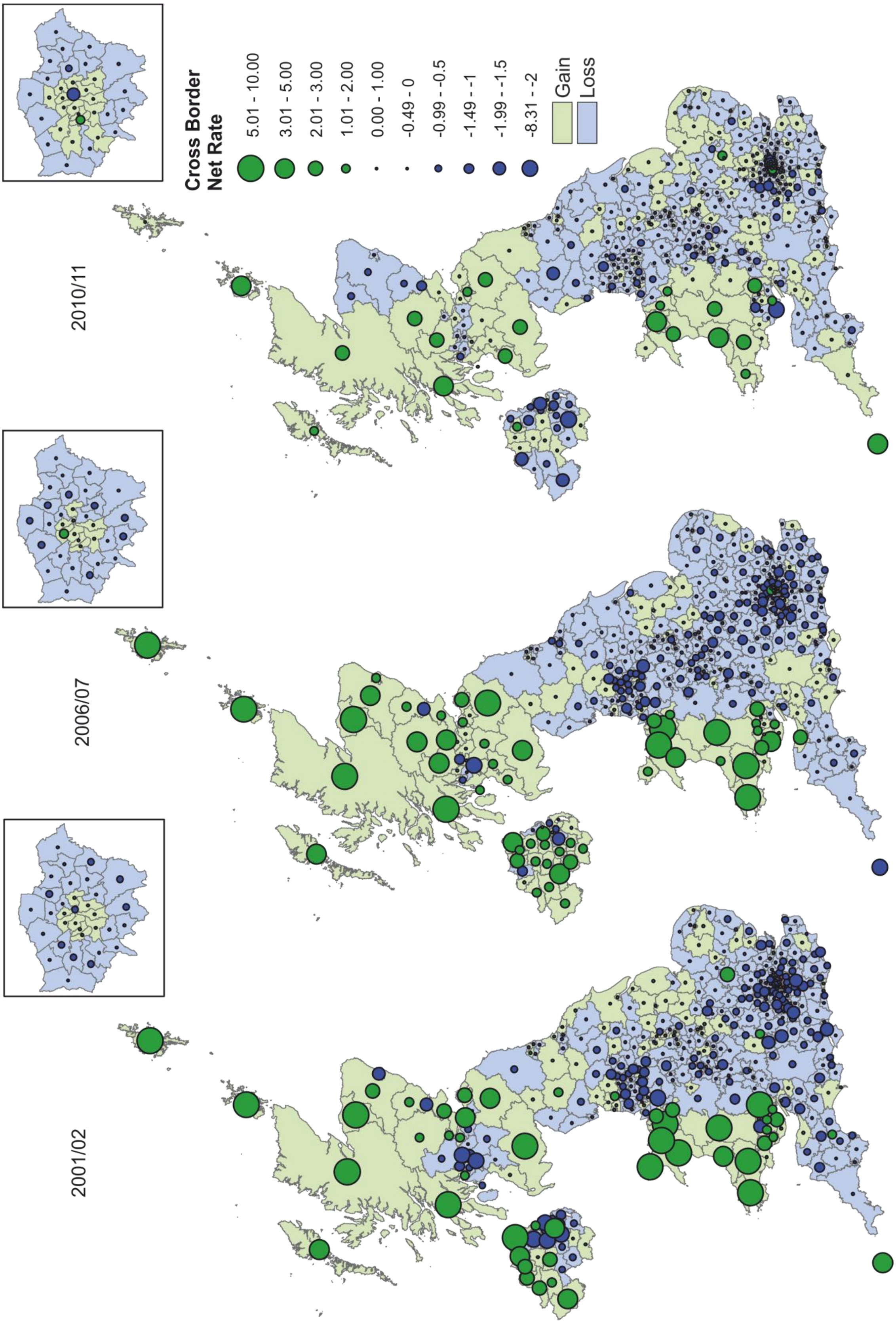


Figure 5.11: Net cross-border migration rates, LAD level, 2001/02, 2006/07 and 2010/11

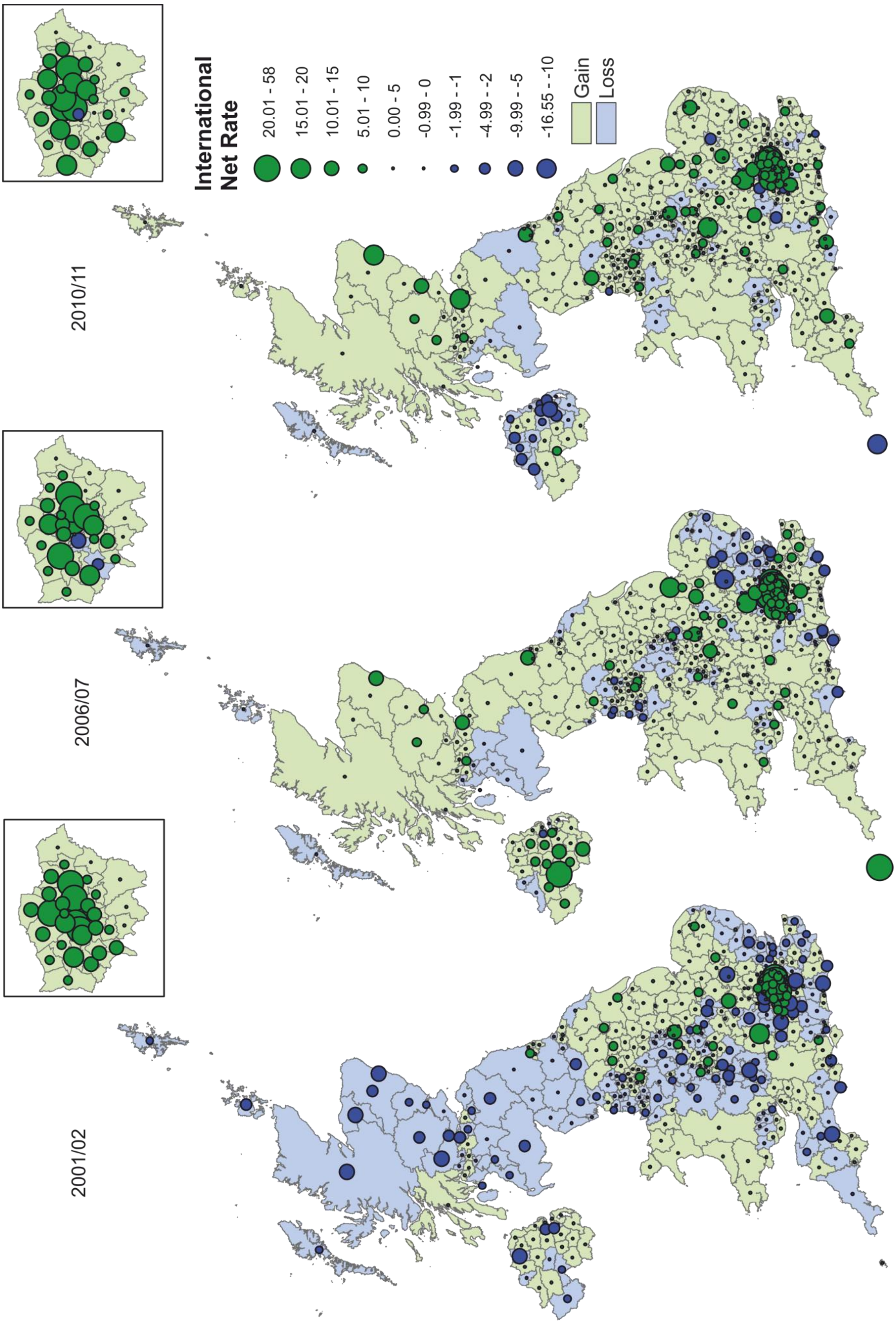


Figure 5.12: Net international migration rates, LAD level, 2001/02, 2006/07 and 2010/11

LADs are high in comparison to other LADs in the UK. The move from net loss to net gain for most LADs in Scotland between 2001/02 and 2006/07 is apparent in the weak correlation between the two years ($r = 0.53$, $p<0.01$) but the correlation between 2006/07 and 2010/11 is still not particularly strong ($r = 0.66$, $p<0.01$). These correlations reflect the volatility of international migration rates, especially when compared to internal migration.

The negative correlation between internal and international migration reported for the net migration balances holds when the net rates are compared but is weaker: in 2001/02 $r=0.45$, in 2006/07 $r=0.41$ and in 2010/11 $r=0.39$ (all $p<0.01$). A weak negative correlation between cross-border and internal migration also emerges when net rates are compared. A time lag exists between the two migration types, so the cross-border rate for 2001/02 is negatively correlated with the internal rate in 2006/07 ($r=-0.14$) and in 2010/11 ($r=-0.13$). The 2010/11 cross-border rate is similarly negatively correlated with internal migration in 2001/02 ($r=-0.13$, $p<0.05$) and in 2006/07 ($r= -0.15$, $p<0.01$). This weak relationship between the two migration streams suggests that areas gaining cross-border migrants lose internal migrants in subsequent years. No significant correlation exists in the same year for internal and cross-border migration, and cross-border migration in 2006/07 has no significant correlation with the other two years.

5.6 Other measures of internal migration

In this final section of the chapter, a set of indicators are used to further understand the patterns of internal migration that are occurring within the UK system between 2001/02 and 2010/11. Moves between all 406 LADs are considered using these measures, so no distinction is drawn between within country and cross-border migration. Bell *et al.* (2002) spell out in detail a set of summary measures that can be used to assess four key dimensions of migration: intensity, distance, connectivity and the impact of migration on the redistribution of the population. The analysis in the following sub-sections draws heavily on Bell *et al.* (2002) and presents a selection of measures to illustrate the changing pattern of migration in the UK between 2001/02 and 2010/11. Some of these measures have been derived using the IMAGE studio, a tool developed for aggregating and comparing migration data across the world (Stillwell *et al.* 2013).

5.6.1 Intensity and distance of migration

In order to assess the intensity and distance dimensions of migration in the 2000s, three measures are presented, the crude migration intensity, mean migration distance and median migration distance. All three measures were computed using the IMAGE studio.

Crude migration intensity (CMI or crude migration probability, CMP) is a simple measure of migration, defined by Bell *et al.* (2002) as:

$$CMI = \left(\frac{M}{P}\right)100 \tag{5.1}$$

where M is the total number of internal migrants between LADs in a given time period and P is the population at risk. The calculation returns a percentage of the population at risk in a given year that migrate. Figure 5.13 shows that over the decade, the crude migration intensity falls. In 2001/02, 4.8 per cent of the population at risk migrated between LADs, compared with 4.5 per cent in 2010/11. CMI rose between 2004/05 and 2006/07, where it peaked at 4.9 per cent.

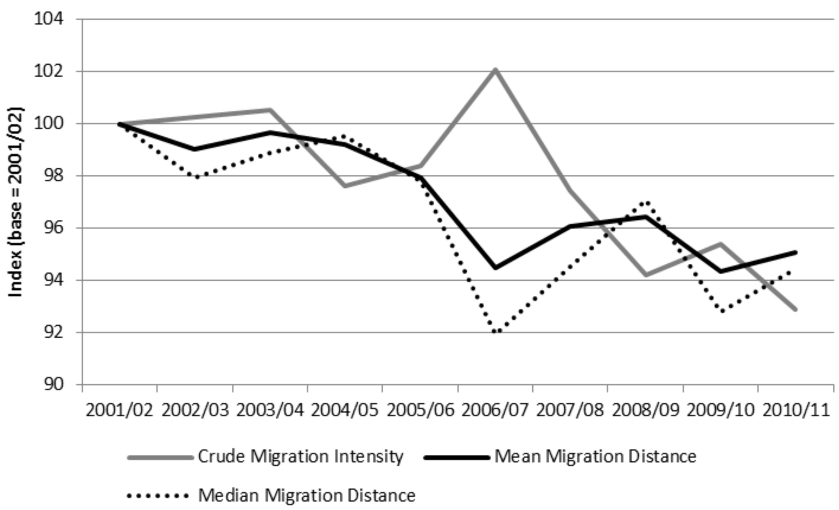


Figure 5.13: Crude migration intensity, mean and median migration distances 2001/02 to 2010/11 (indexed to 2001/02)

The mean and median distance moved by migrants in the system give a clear indication of the spatial dimension of internal migration. The median migration distance is around half the mean migration distance (for example in 2001/02, the distances were 50 km and 107 km respectively). Bell *et al.* (2002, p. 449) point out that *"the median is clearly* *preferable to mean as the distribution of distances is negatively skewed, reflecting the strong distance decay effect which consistently occurs".* On this basis, it can be seen that median migration distance declines between 2001/02 and 2010/11, from around 50 kilometres to 47 kilometres. Figure 5.13 shows that the trend appears to mirror that of the CMI, rising in years where CMI falls and *vice versa.*

These three measures, CMI, mean distance and median distance, show that within the UK system, the proportion of the population who migrate between LADs is falling between 2001/02 and 2010/11 and those that do migrate are not, on average, travelling as far. This fall in migration intensity is consistent with the pattern seen in a number of other countries, as reported by Bell and Muhidin (2009) who present a cross-national comparison of internal migration.

5.6.2 Measures of migration connectivity

To assess the extent to which and the way in which LADs in the UK system are connected, four measures are computed: the index of migration connectivity, the migration inequality index, the coefficient of variation and the Theil index. Aside from the index of connectivity, all measures have been calculated using the IMAGE studio (Stillwell *et al.* 2013).

The index of connectivity (I_c) is defined by Bell *et al.* (2002) as the simplest measure of connectivity within a system by counting all non-zero flows between an origin area and all destination areas or *vice versa*. Stillwell and Hussain (2010) use spatial connectivity in the context of measuring ethnic migration in the 2001 Census, finding that 65 per cent of LADs were connected to one another, with the large metropolitan areas predominant in the linkages. For this analysis, a link of 10 migrants or more has been used to provide the connectivity score to allow for rounding and adjustment (as discussed in the previous chapter) throughout the data time series. Figure 5.14 shows the percentage of LADs that are connected to each other by 10 or more migrants using the following calculation:

$$I_c = 100 \frac{Number\ of\ flows > 10\ migrants\ between\ LADs}{Total\ number\ of\ pairs\ of\ LADs} \quad (5.2)$$

As a system-wide measure for all UK LADs, Figure 5.14 shows that the level of connectivity falls slightly between 2001/02 and 2010/11, from over 31 per cent of LADs being connected to each other at the beginning of the decade to just under 30 per cent in 2010/11.

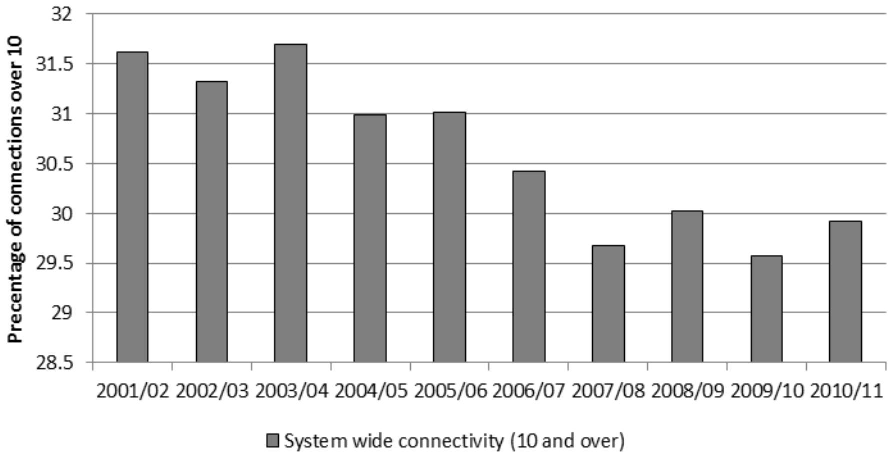


Figure 5.14: The percentage of districts in the UK connected by over 10 migrants in each year 2001/02 to 2010/11

Connectivity is the first of two indicators that can be effectively examined in more detail at LAD level (the second being migration efficiency). Connectivity for inflows to LADs is shown in Figure 5.15, whilst outflows are shown in Figure 5.16.

Figures 5.15 and 5.16 show that LADs in Scotland and Northern Ireland are relatively poorly connected with the rest of the UK, both in terms of inflow and outflow of over 10 migrants, with only Edinburgh consistently showing connections with over 50 per cent of other LADs in the UK. Most London boroughs are consistently well connected, both in terms of inflow and outflow, as are larger LADs in England, including more rural locations such as Cornwall and Wiltshire in the South West.

The number of LADs connected to over 50 per cent of other LADs by inflow is fairly consistent, remaining at 65 in 2001/02 and 2006/06, dropping to 58 in 2010/11. The number of LADs connected by outflow sees a larger decline, from 71 in 2001/02 to 68 and 59 in the two subsequent years. The number of LADs at between 25 and 49 per cent remains fairly constant. The distribution of these areas does appear to change more for outflows than inflows through the decade, with LADs in the East of England, along

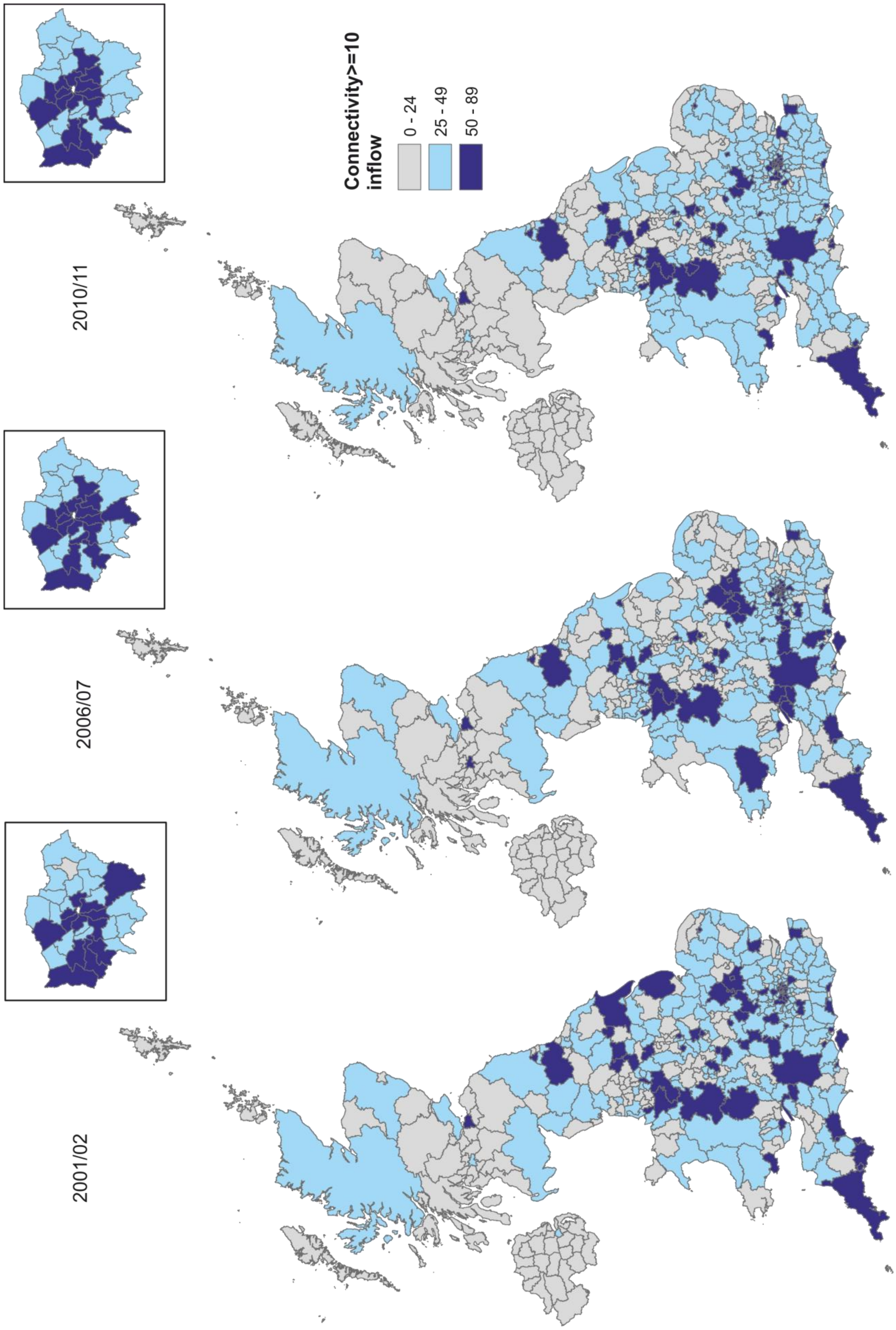


Figure 5.15: Index of inflow connectivity, LAD level, 2001/02, 2006/07 and 2010/11

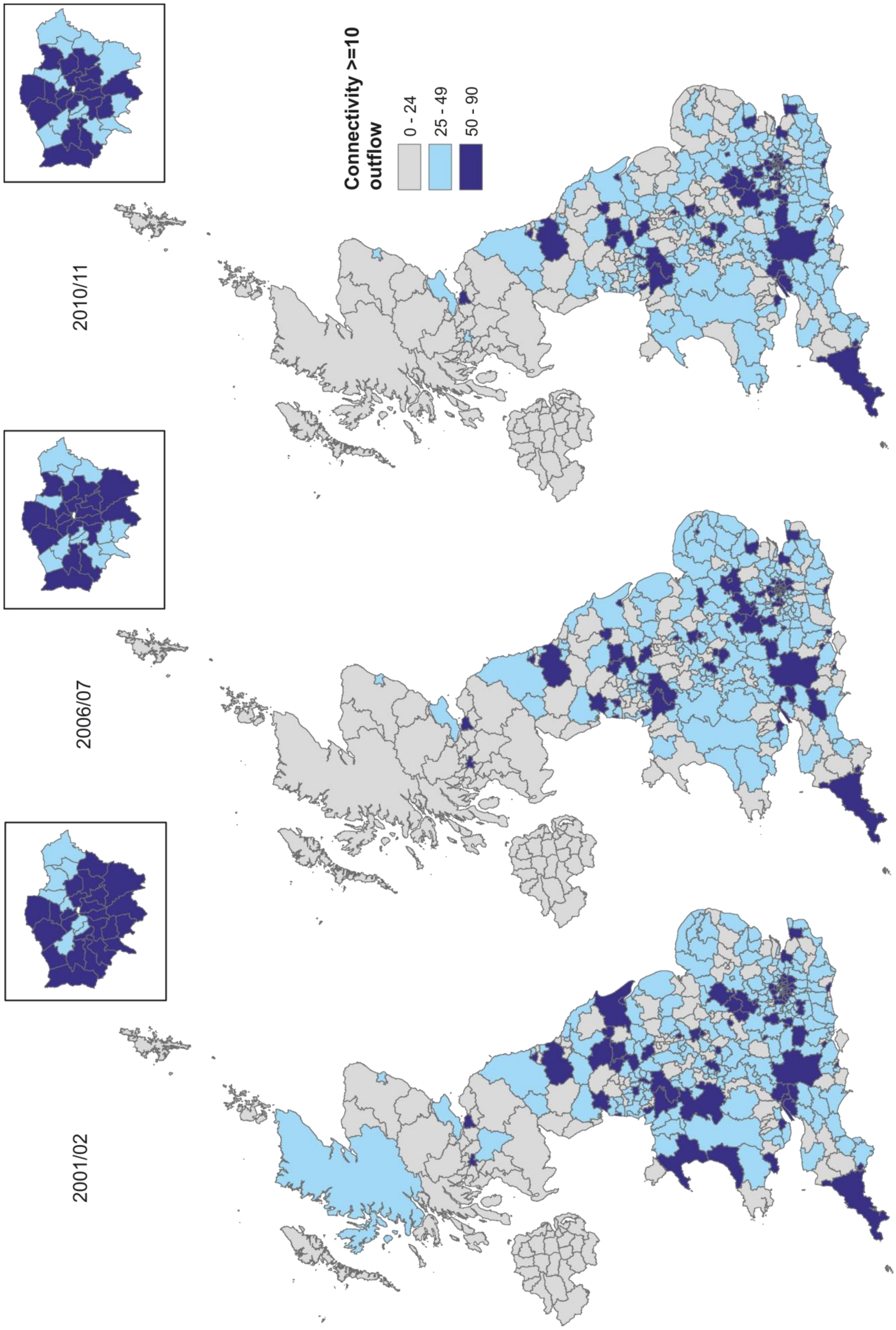


Figure 5.16: Index of outflow connectivity, LAD level, 2001/02, 2006/07 and 2010/11

with a number of LADs in the North East and North West of England and in Wales becoming less connected with the rest of the UK.

The index of migration inequality (I_{MI}) is defined by Bell *et al.* (2002) as the difference between the observed distribution of interregional flows (M) and the expected distribution (M'), where:

$$I_{MI} = 0.5 \sum_{i} \sum_{j \neq i} \left| \frac{M_{ij}}{M_{++}} - \frac{M'_{ij}}{M'_{++}} \right| \tag{5.3}$$

and where $M'_{ij} = \sum_i \sum_{j \neq i} M_{ij} / [n(n-1)]$, n is the number of districts, $M_{++} = \sum_i \sum_{j \neq i} M_{ij}$ and $M'_{++} = \sum_i \sum_{j \neq i} M'_{ij}$. Here, the expected distribution assumes that all LAD to LAD flows are of the same magnitude, irrespective of how far apart are the origins and destinations. The value lies between zero and one, with zero signifying that the observed and expected distributions are identical and values closer to one denote greater inequality. A system with greater inequality is one where a small number of LADs account for a large proportion of total migration. In 2001/02, the I_{MI} score was 0.63, and Figure 5.17 shows that there is a consistent trend of inequality in the system increasing between 2001/02 and 2010/11 (where it reaches 0.66).

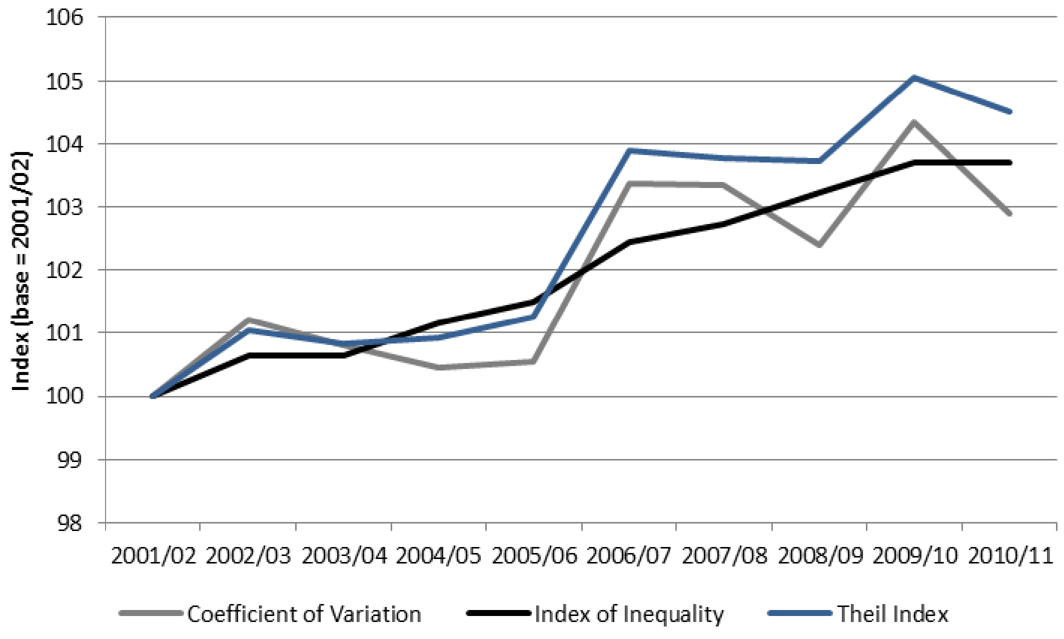


Figure 5.17: The index of inequality, coefficient of variation and Theil index for the UK system, 2001/02 to 2010/11

The coefficient of variation (CV) and the Theil index are measures of spatial concentration, where each interzonal flow is compared with every other interzonal flow to give a dispersion of data points around the mean. Higher values denote greater spatial concentration in the system (the dominance of a certain set of LADs in the context of this study) while a perfectly equal distribution of migration in the system would give a value of zero. Rogers and Raymer (1998) compare CV and the Theil index (alongside another measure, the Gini Index) in a study measuring US interstate migration flows. They argue that all three are measuring the inequality within a system and in the absence of a universal 'best measure,' they favour CV over the Theil and Gini indexes as it is "*transparent, intuitively plausible and well known*" (p.77). Bell *et al.* (2002, p. 457) describe the CV as the square root of "*the standard deviation divided by the mean of a given set of interzonal migration flows*" and specify it as:

$$CV = \sqrt{\left\{\sum_i \sum_{j \neq i} (M_{ij} - \bar{M})^2 / n(n-1)\right\}} / \bar{M} \qquad (5.4)$$

where $\bar{M}$ is the mean migration flow and n is the number of LADs.

The Theil index, developed as an economic measure of inequality (Theil 1967) and often used to measure the distribution of income (see, for example, Hale 2004), is described by Rogers and Raymer (1998, p.65) as *" the mean of* (m_j/m) *[ln* (m_j/m)*] where* m_j *denotes the migration from a particular state to state j, and m is the mean of all* m_j *values".* Figure 5.17 shows that the values of CV and the Theil index are similar across the time series, with Theil consistently higher. Whichever measure is chosen, the trend is clear – all three show that throughout the decade, spatial 'focusing' and inequality in the system are increasing, while connectivity is in decline.

5.6.3 Measures of the impact of migration

Measures of impact include the migration efficiency index (MEI) and the aggregate net migration rate (ANMR). Bell *et al.* (2002, p.459) define the two measures as: *"The MEI essentially indicates the degree of (a)symmetry or (dis)equilibrium in the network of interregional migration flows whereas the ANMR summarizes the extent of population redistribution arising from the net migration balance".* Stillwell *et al.* (2000) employ the MEI to gain an insight into the changing pattern of migration for Australia and the UK between 1976 and 1996. The area-specific migration efficiency ratio (MER_i) is the ratio of net migration in area *i* to the sum of its inflows from all other areas (D_i) and outflows to all other areas (O_i) and represented as a percentage:

$$MER_i = 100(D_i - O_i)/(D_i + O_i) \tag{5.5}$$

MER_i is constrained to +/-100 where a high positive value denotes that in-migration is an efficient force for redistributing population in an area, while a high negative value means out-migration is an efficient force. A score of zero means that in and out migration are equal, so migration is inefficient.

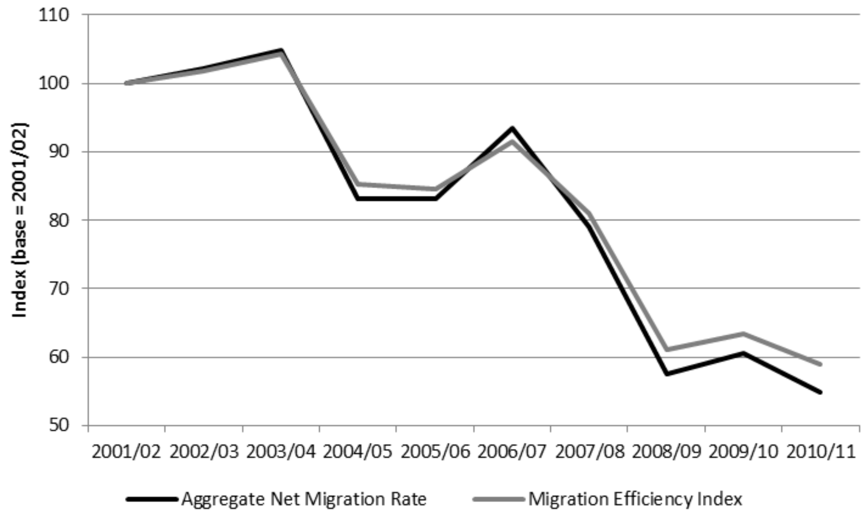


Figure 5.18: The aggregate net migration rate and migration efficiency, indexed to 2001/02

A system-wide migration efficiency index can be defined as:

$$MEI = 100\left(\sum_i |D_i - O_i| / \sum_i (D_i + O_i)\right) \tag{5.6}$$

where D_i is the total in-migration to region *i* from elsewhere in the UK $(= \sum_{j\neq i} M_{ji})$ and O_i is the total out-migration from region *i* to all destinations in the rest of the UK $(= \sum_{i\neq j} M_{ij})$. Note that migrations within regions are excluded from these definitions. The equation for ANMR is similar to the MEI, but the denominator is the population at risk. Thus, ANMR can be defined as:

$$ANMR = 100 \times \frac{1}{2}\sum_i |D_i - O_i| / \sum_i P_i \tag{5.7}$$

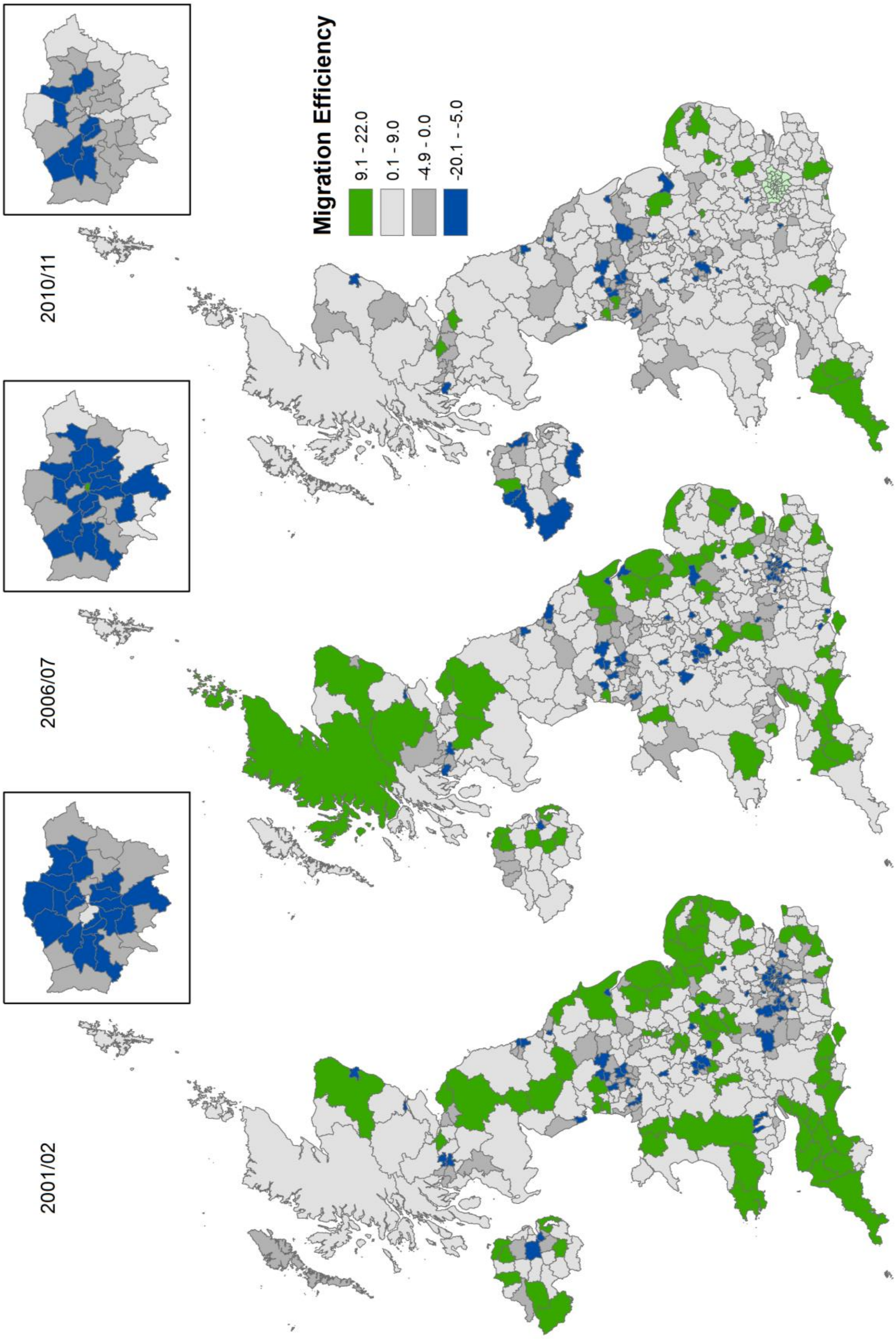


Figure 5.19: LAD-specific migration efficiencies, 2001/02, 2006/07 and 2010/11

where P_i is the population at risk in region i. The ANMR is determined as the product of CMI and the MEI for a spatial system. Figure 5.18 shows that the general pattern for both MEI and ANMR is for a decline in efficiency across the decade, with a slight recovery in 2006/07 before the trend continues to 2010/11. This trend of a decline in efficiency continues the UK pattern seen by Stillwell *et al.* (2000) between 1981 and 2006, in which they conclude that demographic factors such as declining rate of population growth, economic forces such as the demise of older industrial towns and technological factors including improved transport and communications infrastructure have all contributed.

The specification of MER_i in Stillwell *et al.* (2000) can be used to examine migration efficiency at the LAD level. The decline in global migration efficiency that is evident through the decade is shown at the LAD level in Figure 5.19. In 2001/02, rural and peripheral areas appeared to have the highest positive efficiency (over 10) meaning that in these areas, in-migration was an effective process in renewing the population. This can clearly be seen in the South West, along the east coast of England, in Wales and in Scotland. Those LADs showing highest out-migrant efficiency (with scores of over -5) are London boroughs, urban LADs in the South East and the North West of England. The findings for 2001/02 are consistent with those of Stillwell and Hussain (2010) who find in the 2001 Census that for migrants in the White ethnic group, across all LADs the MERi score lay between minus 20 and plus 20.

It is apparent from the pattern in 2010/11 that the number of LADs with a high positive score has decreased substantially, with only some of the most peripheral areas retaining their 10 plus score. Migration in the majority of Scotland and all of Wales has become less efficient at redistributing population, and the number of LADs in London showing high negative scores has fallen. An increase in migration efficiency can be seen in a number of coastal LADs in Northern Ireland where high negative scores have emerged. Glasgow retains a high negative score throughout although Belfast does not.

5.6.4 Making sense of the trends

The seven measures used to assess internal migration between 2001/02 and 2010/11 show a consistent pattern when taken together:

- migration distance and intensity are falling;
- connectivity is falling while spatial focus and inequality are rising; and

- the efficiency of migration as a force for redistributing the population around the system is falling.

When compared alongside the falling net migration rates and balances seen in section 5.4, this points towards a slowdown and spatial focusing in the migration system across the 2000s.

5.7 Conclusion

This chapter has presented a review of the literature pertaining to patterns of migration over the past 50 years which has provided a context in which to analyse the results of the 2001/02 to 2010/11 migration time series estimates as derived in the previous chapter. At the aggregate (national level), internal migration was found to relate closely with unemployment, while international immigration showed a significant relationship with GDP per capita.

A pattern of net migration from urban to rural regions was found across the time series consistent with the body of literature on counterurbanisation which characterised migration patterns in the 1970s, 80s and 90s, but the pattern was found to be in decline across the 2000s. The predominant factor in this decline is the fall in the intensity of migration from urban to rural LADs, while moves from rural to urban LADs remained fairly consistent.

Assessment of the existence of a north-south divide in England and Wales shows an interesting reversal in the second half of the decade, with the north experiencing declining net migration gain from the south until 2007/08, after which the south becomes a net gainer from the north. Combining the two measures of urban-rural and north-south migration reveals that an increase in moves from urban north to urban south has driven the reversal in the north-south migration pattern.

At the regional level, the time series data confirmed the importance of London as a key driver within the UK-wide migration system and London LADs were shown to attract a large number of international migrants and generate a large number of internal migrants to the rest of the UK. Over the 2000s, the general trend at the LAD level has been a general reduction in the level of internal, cross-border and international migration, and the changing magnitude of international and cross-border migration flows have resulted in new patterns of net migration.

Finally, when seven measures of internal migration are taken together, it is clear that migration intensity is falling between 2001/02 and 2010/11, and that migration has become an increasingly less effective mechanism for redistributing the population. Coupled with this, inequality and spatial focus within the system have increased.

The broad overview of patterns and trends presented in this chapter provides the context for more detailed analysis of the connection between areas which is undertaken in the following chapter, while trends disaggregated by age and sex are presented in Chapter 8.

Chapter 6

Using city region functional geographies to link origins and destinations and examine distance

In the previous chapter, the main focus of analysis was on aggregate inflows, outflows and net flows derived from the origin-destination (OD) matrix. The results were interpreted using a number of well-established frameworks taken from past studies of migration. This chapter goes a step further in analysing the OD matrix, examining the linkages between areas which have been estimated for the ten year time period 2001/02 to 2010/11. The volume of information created by estimating the OD matrix (as outlined in Chapter 4) requires a clear strategy for presenting the results in a way that is manageable, and a ‘city region’ geography is used to group LADs into functional clusters. These groupings enable the identification of flows between different ‘types’ of area whilst still allowing for the analysis of individual flows where beneficial. It is important to stress that the aggregate OD flows examined in this chapter are masking a further level of disaggregation – age and sex – which have a material impact on the size and intensity of migration flows between different types of area. These disaggregated flows are covered in detail in Chapter 7 and Chapter 8. Since no information is available on the origin of international immigrants or the destination of emigrants, the focus of this chapter is on internal migration in the main, where both origin and destination are known.

6.1 LAD connectivity between 2001/02 and 2010/11

With a migration dataset of 406x406 origin and destination LAD combinations across a ten year time period, making sense of the connections between these areas and identifying those that have most influence in driving migration around the system is essential. Figure 6.1 shows the connection between LADs in 2001/02, 2006/07 and 2010/11. The lines represent the total number of migrants moving between LADs (so all migrants in either direction between each pair), with only flows of over 500 persons being shown. In 2001/02, flows between 869 pairs of LADs were identified, representing 1.06 per cent of the total number of potential pairs, while in 2006/07, there were 948 OD pairs involved (1.15 per cent) and in 2010/11 the number was 892 pairs (1.09 per cent)

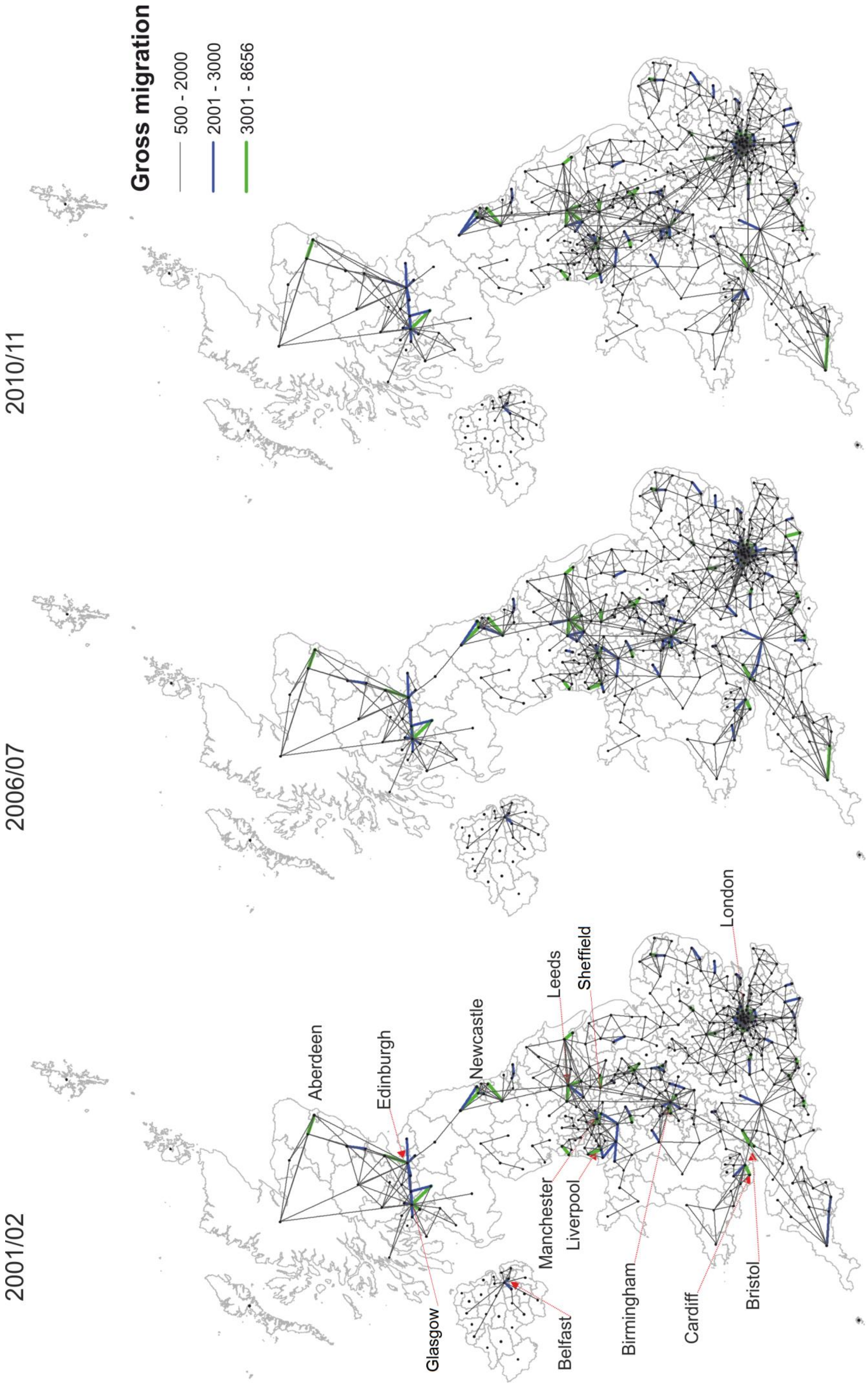


Figure 6.1: The size of the gross migration flow between LADs in 2001/02, 2006/07 and 2010/11 (origin plus destination total)

Two key conclusions can be drawn from these maps: first, there is a consistency in the pattern of connections where the gross flow is large (over 500) in each year; and second, there is a clear dominance of a number of LADs which are at the centres of fairly localised migration networks. These are labelled on the first map and are, of course, the larger urban centres of the UK: Aberdeen, Belfast, Birmingham, Bristol, Cardiff, Edinburgh, Glasgow, Leeds, Liverpool, London, Manchester, Newcastle and Sheffield.

If it can be assumed that the centres of these migration networks are stable over time (as Figure 6.1 suggests), they provide a fixed reference point from which to analyse migration patterns over the decade in a consistent way. The following section introduces a classification of LADs into city regions, for which the main urban areas identified in Figure 6.1 form the core centres.

6.1.2 A case for using city regions

The pattern of 'local' migration seen in Figure 6.1, where the larger flows are predominantly driven by the relationship between urban LADs (the major UK cities) and other LADs that are spatially close or contiguous gives justification for a clustering of LADs which is based on these relationships. That is not to say that other important flows do not exist, but such flows can be more easily identified as unusual when presented within a framework that groups LADs together in a way that exhibit more 'normal' patterns.

The classification of LADs into particular regions or sub-groups can be the focus of an entire PhD thesis. See, for example, Dennett (2010) and Vickers (2006), who have both produced detailed classification systems based on the 2001 Census in research undertaken at the University of Leeds. The Dennett classification clusters LADs based on the characteristics of migrants derived from the 2001 Census. The Vickers classification uses census variables to cluster Output Areas based on their economic and demographic characteristics, so is not explicitly a system used for analysing migration patterns. The Vickers classification is being updated and refined further for 2011 Census geographies in doctoral research currently being completed by Christopher Gale at UCL (Gale 2013, pers. comm.). As the aim of this chapter is to understand the linkage between origins and destinations, it makes sense to use a pre-defined classification system, rather than to invent a new one. This classification system needs to be flexible enough to allow for analysis of individual areas, but also offer the opportunity for aggregation of LADs in order to reduce the burden of information being presented. Both the Dennett and Vickers classifications offer robust area frameworks for analysis, but a major disadvantage for the visualisation of flows is that the areas are not contiguous, so any attempt to analyse a large number of connections at once quickly becomes difficult.

Any solution used to group LADs (or any other spatial unit) needs to consider the effect that aggregation and scale can have on the results produced, known as the Modifiable Areal Unit Problem (MAUP). MAUP, defined by Openshaw (1983) actually comprises two separate but related problems. The scale problem encapsulates variances in results which occur when areal units are aggregated in to "*fewer and larger units for analysis"* while the aggregation problem occurs where *alternative combinations of areal units at equal or similar scales"* are employed (Openshaw, 1983, p.8). A third, closely related, problem is that of ecological fallacy where *"it is inferred that results based on aggregate zonal (or grouped) data can be applied to the individuals who form the groups or zones being studied"* (Openshaw 1983, p.8). All three issues are kept in mind when choosing a solution for analysing the migration data in this chapter.

The solution for analysis and visualisation chosen for this chapter is to use a city region framework. Work commissioned by the Department for Communities and Local Government (Marvin *et al.* 2006, p.5) aimed to define a set of city regions that are "*functional entities"* which have *"greater cultural resonance than current administrative regions and local authority districts"*. They conclude that former Government Office Regions (GORs) are too large, while LADs are too small for strategic decision-making. The city regions they produce are based on travel to work areas derived from 2001 Census data, with a set of core LADs as the focal point for each region. The detailed methodology for creating these city regions can be found in Robson *et al.* (2006). A barrier to using these city regions exists, however, in that that they focus solely on England, excluding the other three countries of the UK. This renders the Marvin *et al.* (2006) classification unsuitable for analysis in this chapter without development of a set of city regions for Wales, Scotland and Northern Ireland. A more comprehensive city region classification has been used by Stillwell *et al.* (2000; 2001) , and Bell *et al.* (2002) which includes all four countries of the UK. These city region areas are shown in Figure 6.2, and are groups of LADs which cluster around 13 'Core' metropolitan areas. The composition of these city regions is as follows:

- City Core areas – these are the 13 major metropolitan centres comprised of 48 LADs (12 per cent of all LADs in the UK), of which 33 are London boroughs which comprise the London Core, three make up the Belfast Core and two make up the Cardiff Core;
- City Rest areas – these are urban areas surrounding or adjacent to the City Cores and are comprised of 98 LADs (24 per cent). These areas have close functional linkages with their City Cores;
- City Near areas – these are comprised of 187 LADs (46 per cent) that are further away from the City Cores than the City Rest areas; and
- Coast and Country areas – these are comprised of the least metropolitan 73 LADs (18 per cent) and are the furthest away from the City Cores.

The LADs in each of the City Rest, City Near and Coast and Country areas are associated with one of the 13 City Cores and labelled accordingly. However, not all city regions have LADs in all four area types; some city regions do not have City Rest areas, for example. Over the decade, between 62.5 and 64.3 per cent of LAD to LAD migration occurred within these 13 city regions (shown in Table 6.1). This proportion is slightly higher in 2010/11 than 2001/02, and peaked in 2006/07.

Table 6.1: The proportion of migrations that cross LAD boundaries which occurred within and between the city regions, between 2001/02 and 2010/11

Year	**01/02**	**02/03**	**03/04**	**04/05**	**05/06**	**06/07**	**07/08**	**08/09**	**09/10**	**10/11**
Within city regions	62.5%	62.8%	62.8%	62.7%	63.0%	64.3%	63.5%	63.1%	64.0%	63.6%
Between city regions	37.5%	37.2%	37.2%	37.3%	37.0%	35.7%	36.5%	36.9%	36.0%	36.4%

This geographical framework of city regions will be used to examine the trends in migration throughout the decade, both between the LADs that fall within different parts of the city region typology and between each component part of these city regions. The analysis that follows is split into six sections: first, the city region typology is used to understand the general trends that are occurring in the UK as a whole (allowing for migration between all LADs, both within and between each part of the city region); second, the patterns occurring within each city region are explored; third, migration between each functional part of the city regions is addressed; fourth, the role that the City Core areas play is examined; fifth, a spatial interaction model is employed to examine the effect of distance at both LAD and city region scale. Interpretation of these patterns is reserved until the final (sixth) section of this chapter, which is presented alongside some policy implications which relate to the trends seen between 2001/02 and 2010/11.

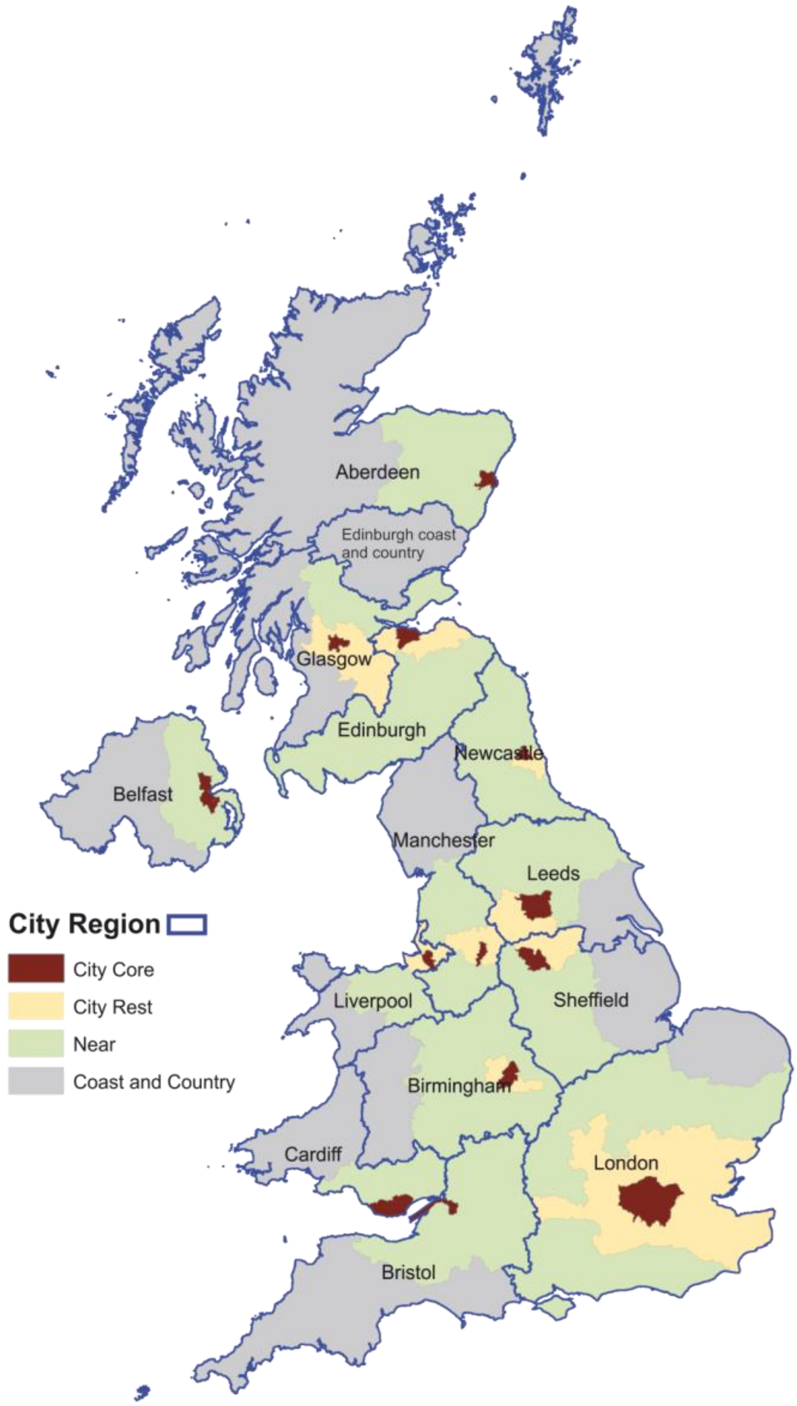


Figure 6.2: UK city region areas

Source: boundaries created by the author using information from the Centre for Interaction Data Estimation and Research

6.2 Migration trends in the UK

It was a pre-requisite of choosing a classification that it allowed for the assessment of migration in the whole system, which is the focus of this section. Figure 6.3 shows the proportion of all migrants in the system who move between the four different parts of the city regions (City Core, City Rest, City Near, Coast and Country) whilst allowing for moves within each of the component parts where applicable (so between all 33 LADs that make up the London Core for example). All four graphs in Figure 6.3 are indexed to 2001/02 to show change over time. Given the dominance of London in the UK migration system (which is examined later in the chapter), all patterns were tested with the 33 LADs that comprise the London Core being excluded and found to be consistent.

Figure 6.3a shows moves out of the City Core areas to all other areas, as well as moves between the City Cores. The proportion of total migrations that occurred between City Core areas increased between 2001/02 and 2008/09, but slowed slightly in 2009/10 and 2010/11. In 2001/02, 13.18 per cent of all migrations occurred between LADs in City Core areas which rose to 15.07 per cent in 2010/11. When moves between LADs in the London Core are discounted (blue line in Figure 6.3a), the trend is similar, except the moves between LADs that form the City Cores which slowed in 2009/10 increases again in 2010/11. The proportion of moves from City Core areas to Coast and Country areas generally declines across the decade. This Core to Coast and Country flow accounted for 1.88 per cent of all migrations in 2001/02 which fell to 1.68 per cent in 2010/11. Flows out of the City Core areas to both City Rest and City Near areas fell slightly between 2001/02 and 2010/11: flows from City Core to City Near areas accounted for 6.72 per cent of migrations in 2001/02 and 6.65 per cent in 2010/11 while flows from City Core to City Rest accounted for 7.52 and 7.37 per cent at the beginning and end of the decade respectively. This decline is not consistent throughout the time series however, the proportion of moves from City Core to City Rest increases between 2001/02 and 2007/08 before it starts to decline, while the decline in proportion of moves from City Core to City Near only begins in 2005/06. This pattern will be examined in relation to international migration patterns in the discussion section later in the chapter.

Figure 6.3b shows the proportion of total flows that involve a move into the City Core areas. The general pattern from 2003/04 onwards is an increase in the proportion of the total flow from City Near and City Rest areas to City Core areas. Flows from City Near to City Core accounted for 5.65 per cent in 2001/02 and 6.46 per cent in 2010/11, while flows from City Rest to City Core rose from 4.90 per cent to 5.58 per cent over the same period. The pattern of flows from Coast and Country to City Core areas fluctuates more throughout the decade but is higher in 2010/11 (1.74 per cent of total migration) than in 2001/02 (1.62 per cent).

Figure 6.3c shows three flows that have declined over the decade. The most substantial relative fall is the flow from City Rest to Coast and Country which declines from 2.18 per cent to 1.64 per cent of total migration between 2001/02 and 2010/11. The City Rest to City Near flow falls from 7.54 per cent at the beginning of the time series to 6.55 per cent at the end, while the City Near to Coast and Country flow falls from 4.25 per cent to 3.74 per cent in the same period. All of these declines are moves from more urban to less urban areas.

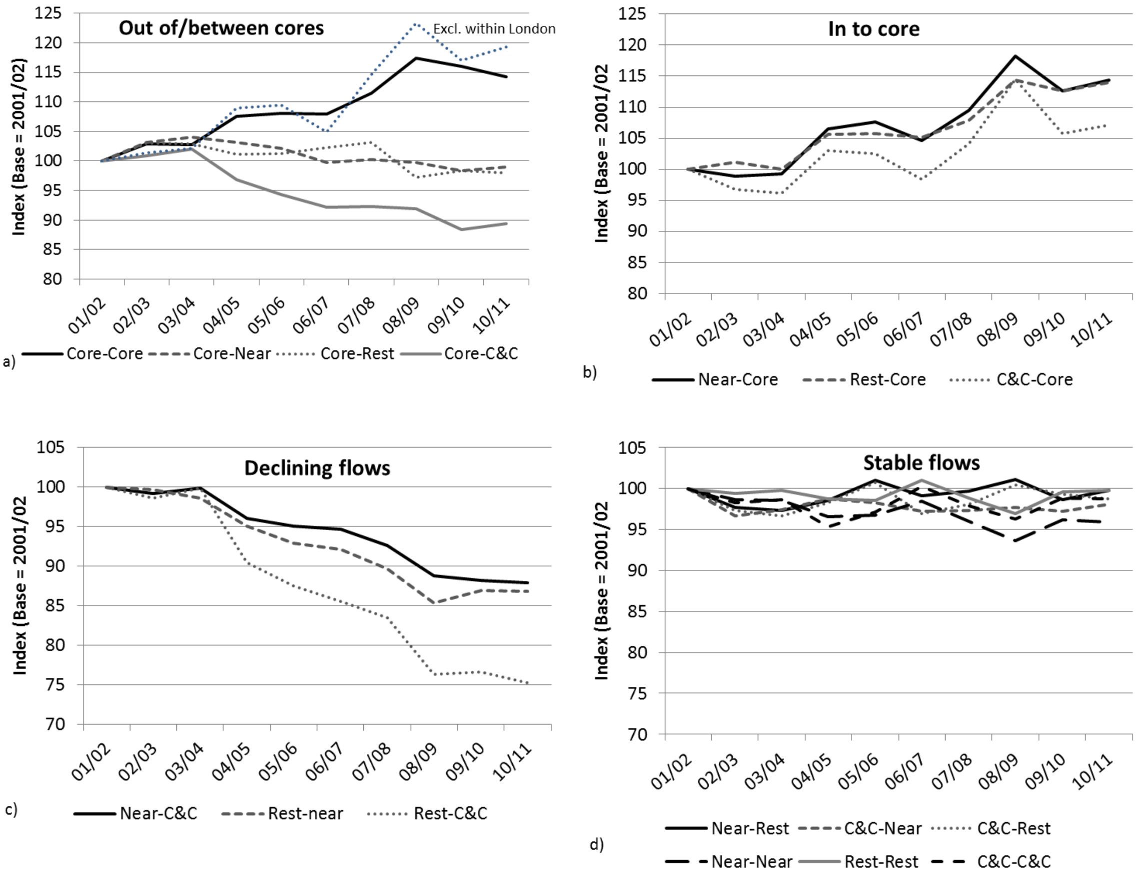


Figure 6.3: Proportions of total migration represented by each type of flow, 2001/02 to 2010/11, indexed to 2001/02

The final graph, Figure 6.4d, shows the remaining migration flows that, as a proportion of total migration in the system, are most stable across the time series (in relation to 2001/02). The flows occurring between (and within) parts of the city regions all fall slightly between 2001/02 and 2010/11: City Near to City Near moves decline from 20.88 per cent to 20.03 per cent of total, City Rest to City Rest moves fall from 8.89 per cent to 8.87 per cent while moves between Coast and Country areas drop from 4.54 per cent to 4.48 per cent of total migration. The other migration flows shown in Figure 6.4d also exhibit a small decline in 2010/11 relative to 2001/02: City Near to City Rest falls from 5.41 per cent to 5.39 per cent, Coast and Country to City Rest falls from 1.30 per cent to 1.28 per cent and Coast and Country to City Near drops from 3.53 to 3.46 per cent of total migration.

The patterns seen in Figure 6.3 can be simplified and summarised by defining areas as either metropolitan or non-metropolitan. The metropolitan category comprises all LADs that are in the City Core and City Rest groups, while non-metropolitan is made up of LADs in the City Near and Coast and Country groups. Figure 6.4 shows the proportion of total migration that can be attributed to each of these moves, and the two most clear trends from 2003/04 onwards are an increase in the proportion of migrants moving between metropolitan LADs (36.89 per cent in 2010/11 compared with 34.5 per cent of all migrations in 2001/02) and a decrease in the proportion of migrants moving from metropolitan to non-metropolitan LADs throughout the decade (18.32 per cent in 2001/02 and 16.52 per cent in 2010/11). The trend for moves from non-metropolitan to metropolitan LADs is a general increase from 2003/04 (a slowing of the trend occurred in 2006/07 and 2009/10). This non-metropolitan to metropolitan migration flow accounted for 13.98 per cent of total migration in 2001/02 and 14.87 per cent in 2010/11. The trend in the proportion of moves occurring between non-metropolitan LADs is one of decline: in 2001/02 these moves accounted for 33.21 per cent of total migration while in 2010/11 fell to 31.72 per cent.

In the interest of summarising the general trends seen in Figure 6.4, it can be said that a greater proportion of people are moving between metropolitan areas as the decade progresses, a smaller proportion are moving out of metropolitan areas to non-metropolitan areas, the proportion of moves from non-metropolitan to metropolitan LADs increases while moves between non-metropolitan LADs decreases. These findings are consistent with the urban-rural migration patterns seen in the previous chapter where, over the decade, moves from high to low density areas are declining, while moves from low to high density areas are on the increase.

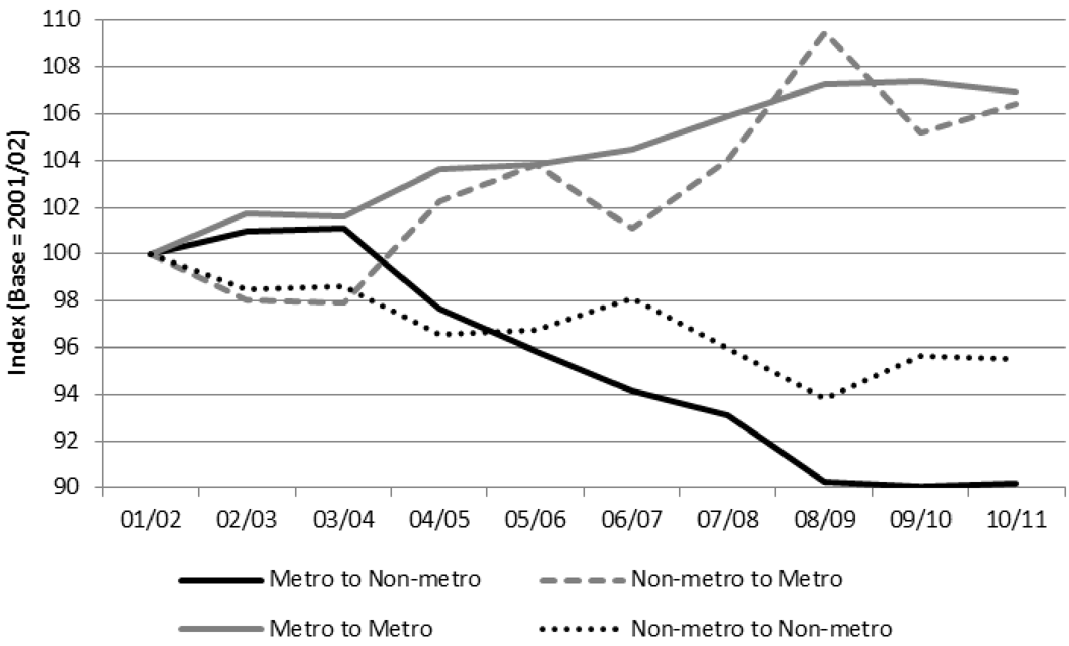


Figure 6.4: The proportion of total flow between metropolitan and non-metropolitan LADs (with the base year as 2001/02)

Table 6.2 provides a comparison of the more sophisticated city region typology used in this chapter and the simpler urban-rural classification used in the previous chapter (where half of LADs were assigned as urban and half as rural, based on their population density). All of the City Core and 69 per cent of City Rest LADs were classed as urban based on their population density, while 60 per cent of City Near and 82 per cent of Coast and Country LADs were classified as rural. As a result, 57 per cent of LADs defined as metro areas in Figure 6.4 were considered as urban in the previous chapter, while 85 per cent of those defined as non-metro were classified as rural. The results of the cross-tabulation show that even though differences between the two classifications means that some LADs were assigned differently, the overall patterns seen are consistent.

Table 6.2: Cross-tabulation of LADs assigned to city region and urban-rural classification

		Urban-rural classification		Total
		Rural	Urban	
City region classification	Coast & Country	60	13	73
	Core	0	48	48
	Near	113	74	187
	Rest	30	68	98
	Total	203	203	406

Returning to the city region classification, the period 2001/02 to 2003/04 represents one of relative stability, with substantial changes from 2004/05 onwards. Some potential reasons for this are explored in the final section of this chapter and the next section looks at these general migration trends in more detail, first addressing the patterns of migration occurring within each city region and then the patterns occurring between each of the city region component parts.

6.3 Moves within and between city regions

To provide an overview of the pattern of migration for the past decade, the previous section used the city region framework to classify LADs into four component parts, City Core, City Rest, City Near and Coast and Country, and allowed for moves both between and within these component parts. This section applies the city region framework more rigidly and excludes moves within each component part of the city region (so, for example, the 10 LADs which comprise the Manchester City Rest area are here treated as a single area). The result allows for analysis of moves between 47 origin and destination areas and will be examined in Section 6.4 for each city region separately and in Section 6.5 for moves within and between city regions. This analysis allows assessment of to what extent the general trends presented in the previous section are occurring for each of the 13 city regions. The migrations occurring between each type of area are summarised in Table 6.3 for 2001/02, 2006/07 and 2010/11. The number of LAD to LAD migrants that are excluded from the analysis because they move within a city region part are shown in the table, and amount to 1,022,658 migrants in 2001/02, 1,133,433 migrants in 2006/07 and 1,050,822 migrants in 2010/11.

The general trends seen in the previous section are confirmed in Table 6.3. The number of migrants moving from City Rest, City Near and Coast and Country areas to City Core areas increases consistently in 2006/07 and 2010/11, as does the number of moves between City Cores. This City Core to City Core migration stream represents the largest percentage increase between 2001/02 and 2010/11, where an additional 11,550 migrants represent an increase of 16.2 per cent. The largest rise in total migrants to the City Core is from City Near areas which is 21,442 migrants higher in 2010/11 than 2001/02 (representing an increase of 13.2 per cent), closely followed by an 18,091 (12.9 per cent) increase in the number of migrants from City Rest areas over the same time period. The number of migrants moving from all city region parts to City Core areas increases between 2006/07 and 2010/11 despite a fall of over 100,000 total migrants in the system in the same time period who move between city region component parts. The largest percentage increase between 2006/07 and 2020/11 is the City Core to City Core migration stream, where 5,150 additional migrants represent an increase of 6.6 per cent.

The number of migrations from City Core, City Rest and City Near areas to Coast and Country areas falls in both 2006/07 and 2010/11, as does the number of moves between Coast and Country areas. The largest of these declines that occurs between 2001/02 and 2010/11 is the number of migrations from City Rest areas to Coast and Country areas (15,930 fewer, representing a fall of 25.5 per cent) and from City Near areas to Coast and Country areas (15,726 fewer, a drop of 12.9 per cent). A fall in number of migrants moving to City Rest and Near areas can be observed between 2001/02 and 2010/11, where the largest falls in absolute number of migrants is for City Rest to City Near (30,250 fewer migrants in 2010/11 compared with 2001/02, representing 14.0 per cent) and between City Near areas (with a fall of 20,138 migrants, 9.9 per cent).

2006/07 represents the year in the time series where there are the highest number of migrants in the UK system (1.8 million cross a city region boundary) and this midpoint has very different pattern when compared with the beginning (2001/02) and end (2010/11) of the time series. Most notable is the change in number of migrants moving from City Core areas to City Rest and City Near areas, which increase between 2001/02 and 2006/07 by 15,766 (7.3 per cent) and 9,137 (4.7 per cent) respectively. In contrast, the number of migrants moving from City Core to City Rest and City Near areas falls substantially between 2006/07 and 2010/11, where it is 22,254 (9.62 per

Table 6.3: Numbers of people moving between the city region component parts, 2001/02, 2006/07 and 2010/11

	Destination	Core	Rest	Near	Coast & Country	Total
2001/02						
Origin	Core	71,260	215,683	192,743	53,890	533,576
	Rest	140,474	40,029	216,228	62,450	459,181
	Near	162,031	155,044	203,865	121,898	642,838
	Coast & Country	46,537	37,158	101,228	24,123	209,046
	Total	420,302	447,914	714,064	262,361	**1,844,641**
Excluded moves within region		*306,780*	*214,904*	*394,913*	*106,061*	*1,022,658*
2006/07						
Origin	Core	77,663	231,449	201,880	52,142	563,134
	Rest	154,998	38,191	209,032	56,093	458,314
	Near	178,045	161,245	201,325	121,080	661,695
	Coast & Country	48,044	37,810	103,327	23,755	212,936
	Total	458,750	468,695	715,564	253,070	**1,896,079**
Excluded moves within region		*350,513*	*232,140*	*417,531*	*113,249*	*1,113,433*
		Percentage change over 2001/02 figure				
	Core	8.99%	7.31%	4.74%	3.24%	
	Rest	10.34%	4.59%	3.33%	10.18%	
	Near	9.88%	4.00%	1.25%	0.67%	
	Coast & Country	3.24%	1.75%	2.07%	1.53%	
2010/11						
Origin	Core	82,813	209,195	188,904	47,696	528,608
	Rest	158,565	37,076	185,978	46,520	428,139
	Near	183,473	153,158	183,727	106,172	626,530
	Coast & Country	49,380	36,332	98,281	21,403	205,396
	Total	474,231	435,761	656,890	221,791	**1,788,673**
Excluded moves within region		*345,105*	*214,707*	*385,106*	*105,904*	*1,050,822*
		Percentage change over 2001/02 figure				
	Core	16.21%	3.01%	1.99%	11.49%	
	Rest	12.88%	7.38%	13.99%	25.51%	
	Near	13.23%	1.22%	9.88%	12.90%	
	Coast & Country	6.11%	2.22%	2.91%	11.28%	
		Percentage change over 2006/07 figure				
	Core	6.63%	9.62%	6.43%	8.53%	
	Rest	2.30%	2.92%	11.03%	17.07%	
	Near	3.05%	5.02%	8.74%	12.31%	
	Coast & Country	2.78%	3.91%	4.88%	9.90%	

cent) and 12,976 (6.4 per cent) lower respectively. As mentioned earlier in relation to general trends, this change may be explained by the pattern of international migration and will be picked up later in this chapter. The total number of migrants moving to Coast and Country areas from all other parts of the city region falls in 2006/07 compared with 2001/02, but the fall in the number of migrants to Coast and Country in 2010/11 over 2006/07 is more pronounced, suggesting that the trend intensifies in the latter half of the decade.

The flows seen in Table 6.3 confirm that when the city region typology is adopted more rigidly, the patterns seen across the whole UK system (migration between all 406 LADs shown in the previous section) hold. The totals reported in Table 6.3 inform the analysis carried out in the following two sections. The next section addresses the moves within each city region where, as seen in Table 6.1, the majority of migration takes place, while Section 6.5 extends the analysis to include moves across the 13 city region boundaries.

6.4 Flows within city regions

The net migration rates for each part of the city region (City Core, City Rest, Near and Coast and Country) in relation to the other parts of that city region are shown in Figure 6.5. Here, the LADs that form each part of the city region have been amalgamated, so moves between LADs in each part of the region are excluded (City Rest to City Rest, City Near to City Near, etc.). The net rates reported in Figure 6.5 effectively show the redistribution of population within each city region and the use of rates rather than balances makes the patterns comparable.

The most notable trend, which is clear for most city regions, is the role of the City Core as a net loser of migrants to other parts of the region. This rate of net loss declines substantially in most regions during the decade between 2001/02 and 2010/11. In some regions, the decline is steady (London, Leeds, Glasgow, Birmingham, Edinburgh, Bristol and Sheffield) while in others, 2007/08 appears to be the year when a marked decline took place in the rate of net loss compared with the previous year (Belfast and Cardiff). The Cardiff City Core moves from a position of net loss in 2006/07 to net gain in 2007/08. The other four city regions (Newcastle, Manchester, Liverpool and Aberdeen) have a less discernible trend across the ten year period, but Aberdeen is the only city region where the net rate of loss from the City Core area is higher in 2010/11 than in 2001/02. Apart from Aberdeen, all City Cores conform to the general trend of a fall in net loss to other parts of the city region.

A second clear trend is a fall in the net rate of gain for the Coast and Country part of each city region. Of the 12 city regions that have a Coast and Country component, only one – Cardiff – exhibits a move from net loss to net gain through the decade. Belfast, Manchester, Bristol and Birmingham Coast and Country all move from a position of net gain in 2001/02 to net loss in 2010/11. The net rate for Glasgow, Leeds and London Coast and Country remains positive but falls substantially through the decade. Liverpool Coast and Country, after some fluctuation through the time series, has a larger rate of net loss in 2010/11 than in 2001/02, while the pattern of net loss in Aberdeen and Edinburgh is fairly consistent throughout the time series. Sheffield Coast and Country shows net gain through the time series which is slightly lower in 2010/11 than 2001/02.

The net rate of gain for the majority of City Near areas has declined over the decade, with the exceptions being Aberdeen Near and Liverpool Near, which fluctuate through the time series, and Sheffield Near, which has a very low rate of net migration throughout the time series. The trend change for the City Rest areas (which are present in nine city regions) is less discernible. Newcastle Rest and Manchester Rest move from a small rate of net loss to net gain, while five City Rest regions show a decline in the net rate of gain (Glasgow, Leeds, Edinburgh, London and Sheffield). The rate of loss in Birmingham Rest is falling while Liverpool Rest’s net gain fluctuates in its magnitude.

In summary, when moves between various parts of each city region are assessed, the net loss from City Cores falls across the decade (except in Aberdeen), the net gain in Coast and Country areas falls (except in Cardiff) and the net gain seen in City Near areas falls (except in Liverpool and Aberdeen). The trend for net rates in City Rest areas is less clear. The next section expands this analysis to selected migration flows between and within the city region areas to provide a picture of the changing trends across the whole UK migration system.

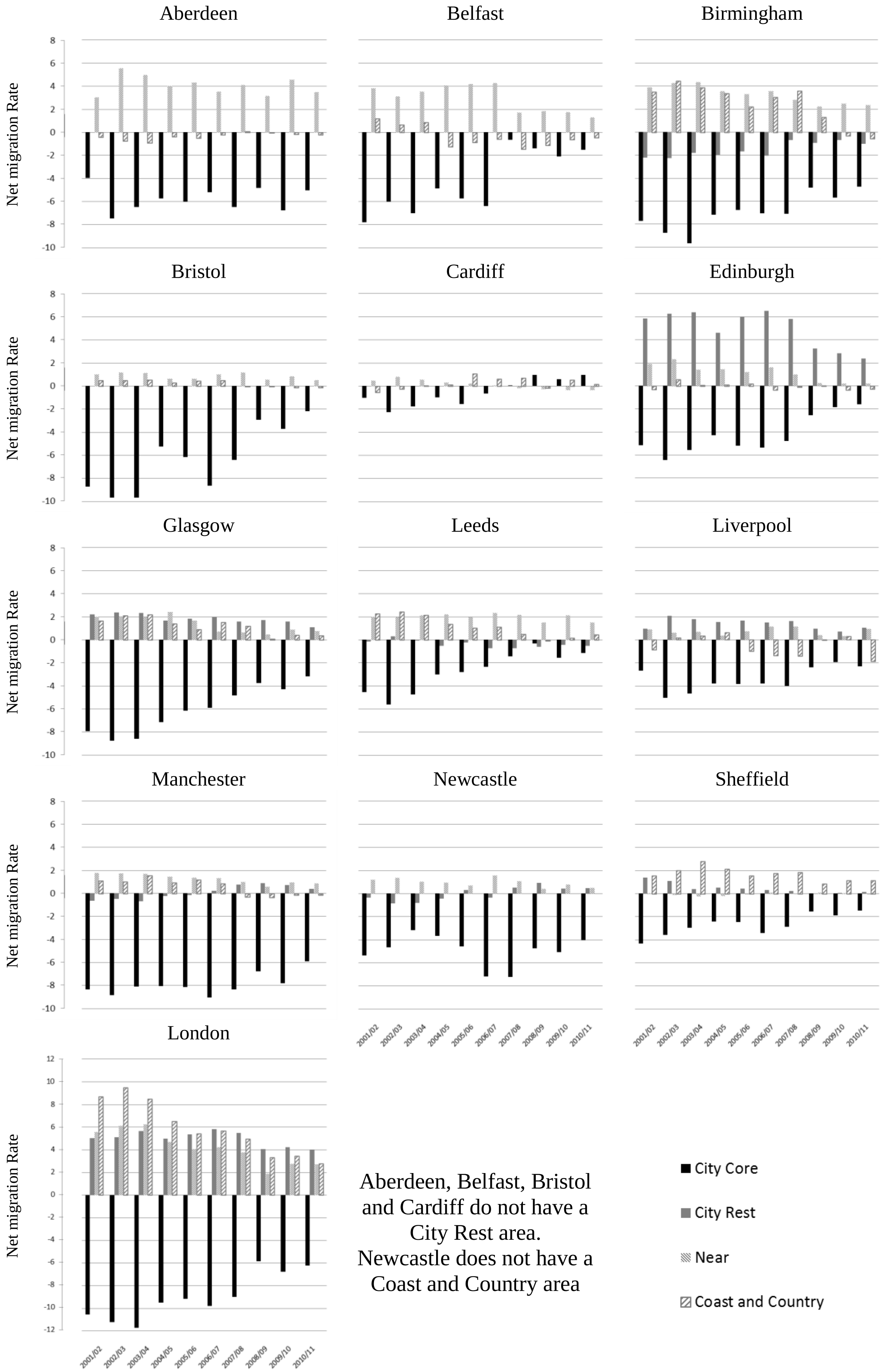


Figure 6.5: Net migration rates within each city region, 2001/02 to 20010/11

6.5 Flows between city regions

This section expands the analysis of city region migration to include migrations that occur across the boundaries of the 13 city regions. Table 6.3 shows that there are effectively 16 sets of origin and destination flow types and only a selection are presented here. As shown in Table 6.3, the largest flow in 2001/02 involved migrations from City Rest to City Near areas, while in 2006/07 and 2010/11, the largest number of migrations occurred between the City Core and City Rest areas. The second of these flows consists largely of within city region moves which have been covered in the previous section and will be examined in relation to City Core migration in Section 6.6. The largest 2001/02 flow (City Rest to City Near) reveals some interesting patterns of cross city region migration and is explored further in this section. By association, migration in the other direction (City Near to City Rest) is also examined here.

The majority of City Core to City Rest, City Near and Coast and Country migration (and the associated moves in the other direction) occurs within each city region and as such was covered in the previous section. These flows are also examined in the context of the role that the City Cores play in migration in Section 6.6. Migration flows between City Rest areas and between City Rest and Coast and Country areas are relatively small so are also excluded from analysis presented in this section.

The migration flows from City Near to Coast and Country (and moves in the other direction) are relatively large and are presented in this section. Similarly, the large number of moves between City Near areas are also examined. Finally, the pattern of migration between City Cores is examined, which leads on to a more detailed analysis of the role of the Core in the migration system in the next section. So to summarise, the following relationships are examined in this section: first between City Rest and City Near areas; second between Coast and Country and City Near areas; third between the different City Near areas; and fourth between the different City Core areas.

6.5.1 Migration between City Rest and City Near areas

Table 6.3 shows that migration from City Rest to City Near areas accounted for the largest number of migrants in 2001/02 (216,228) and while this number fell in both 2006/07 (209,032) and 2010/11 (185,978), the aggregate flow still accounted for a substantial number of migrants. The flow of migrants in the other direction, from City Near to City Rest, also accounts for a large number of migrants in each year (between

153,158 and 161,245). When the relationship between City Rest and City Near areas is assessed in terms of net migration rate (Figure 6.6), it is apparent that the predominant pattern for all city regions is one of net gain in City Near areas from City Rest areas, but this net gain is declining over time. Not surprisingly, the largest net gains for City Near areas come from the City Rest within the same city region, the largest rates being seen in Birmingham, London, Manchester, Leeds and Newcastle. The pattern of a decline in net gain is evident in these City Near areas: London falls from a net gain of 2.9 people per 1,000 population in 2001/02 to under 1.8 in 2010/11 while the Birmingham Near net gain falls by a similar amount from just under 2.8 to under 1.8 persons per 1,000 population.

A number of City Near areas also exhibit fairly large rates of net gain from City Rest areas outside of their own city region. This trend is pronounced in Edinburgh Near which gains from most City Rest areas, especially Glasgow, London and Manchester, where the rate is over 0.6 persons per 1,000 resident population (in 2001/02). Liverpool Near gains over 1 migrant per 1,000 from its nearest neighbour, Manchester Rest, in 2001/02 while Manchester Near gains from Liverpool Rest at a slightly lower rate in 2001/02. Most City Near areas have a positive net migration rate from London Rest; Bristol Near has a large net inflow from London, for example, and the pattern is discernible in Leeds and Edinburgh.

The overall pattern of migration from City Rest to City Near areas in a different city region is one of large scale decline over the decade. One pronounced shift occurs in Aberdeen Near which moves from gaining around 0.6 migrants per 1,000 from London in 2001/02 to a position of small net loss in 2010/11. In summary, while most City Near areas gain from their own City Rest areas and some also gain from other City Rest areas, the pattern is one of decline in this net gain over the decade.

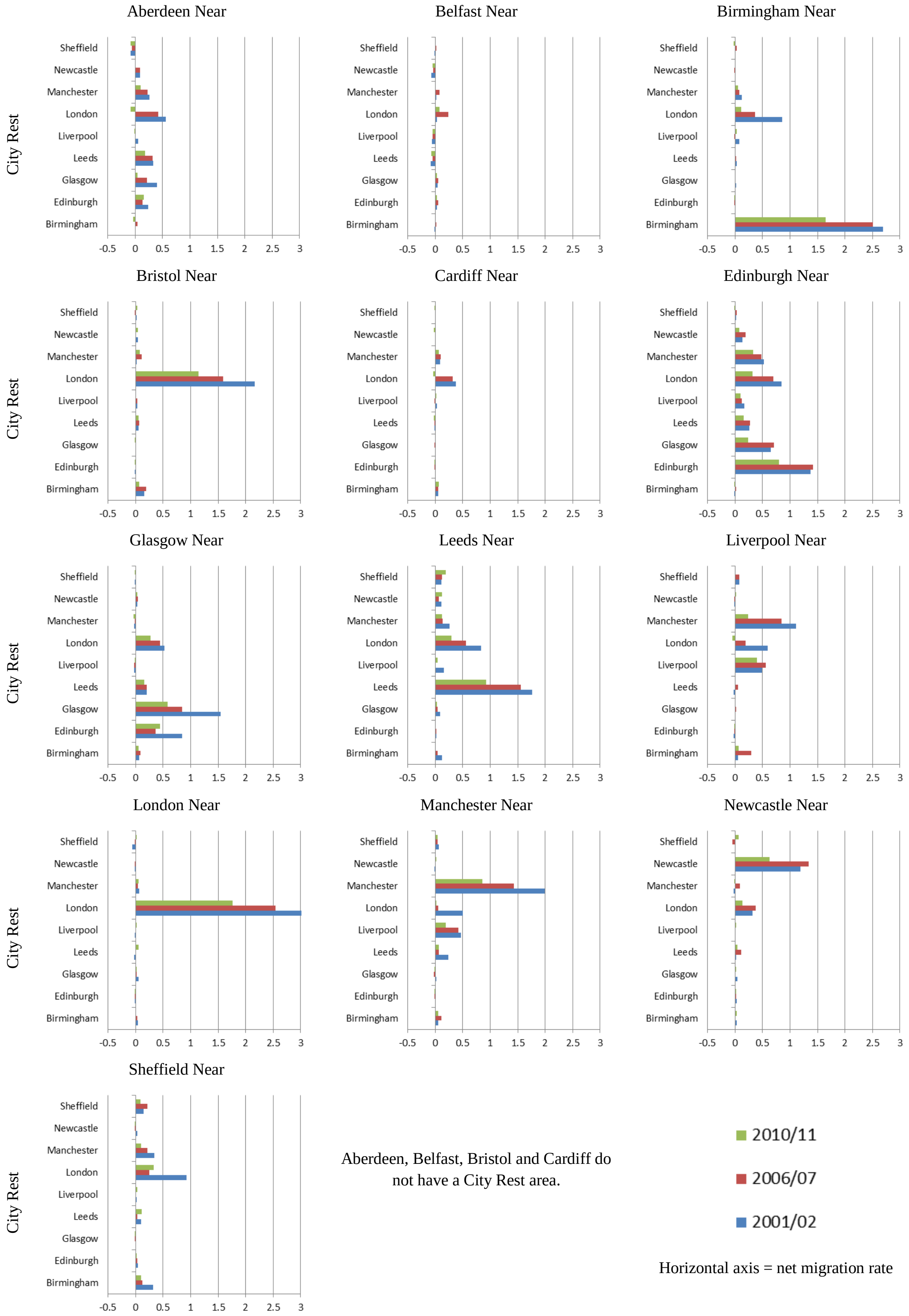


Figure 6.6: Net migration rates for City Near areas in relation to City Rest areas, 2001/02, 2006/07 and 2010/11

6.5.2 Migration between City Near and Coast and Country areas

Total migration between City Near and Coast and Country areas declines over the decade from 121,898 migrants in 2001/02 to 106,172 migrants in 2010/11 (Table 6.3). Over the same time period, the flow in the other direction, Coast and Country to City Near, remains fairly stable at 101,228 in 2001/02 and 98,291 in 2010/11. When these patterns are assessed for each individual area (Figure 6.7), the general trend is one of net gain in Coast and Country areas, with some exceptions: Liverpool Coast and Country loses to Liverpool Near in all three years, increasing to nearly two per 1,000 in 2010/11. Birmingham Coast and Country shifts from a net gain of 2.8 from Birmingham Near in 2001/02 to a net loss of 1.2 per 1,000 in 2010/11. Manchester Coast and Country shifts from net gain to net loss over the decade, but on a smaller scale than Birmingham while Glasgow Coast and Country consistently shows net loss to Glasgow Near.

The relationship of positive net migration for Coast and Country areas in relation to the City Near within the same city region largely declines over the decade, the exceptions being Cardiff and Aberdeen Coast and Country areas, which have a higher rate of net gain in 2010/11 than in 2001/02. This decline is particularly pronounced in London, but for all other areas is a relatively small drop.

London Near plays an important role in the relationship with Coast and Country areas in many other city regions, in 2001/02 providing large net gains for Sheffield Coast and Country (3.8 per 1,000), Birmingham (2.8 people), Bristol (2.4 people) and Cardiff (1.8 people). Aberdeen Coast and Country gains from Sheffield and Birmingham, Liverpool Coast and Country gains from Manchester and Birmingham Near whilst Cardiff Coast and Country gains from Birmingham (the same is true for flows from Cardiff Near to Birmingham Coast and Country). These net gains in Coast and Country areas from City Near areas outside of their city region fell dramatically through the decade however.

A pattern of reversal from small net gain to net loss can be seen in Leeds Coast and Country which begins to lose migrants to London Near by 2010/11. Manchester Coast and Country begins to lose migrants to Bristol Near, as do Sheffield and Leeds Coast and Country areas.

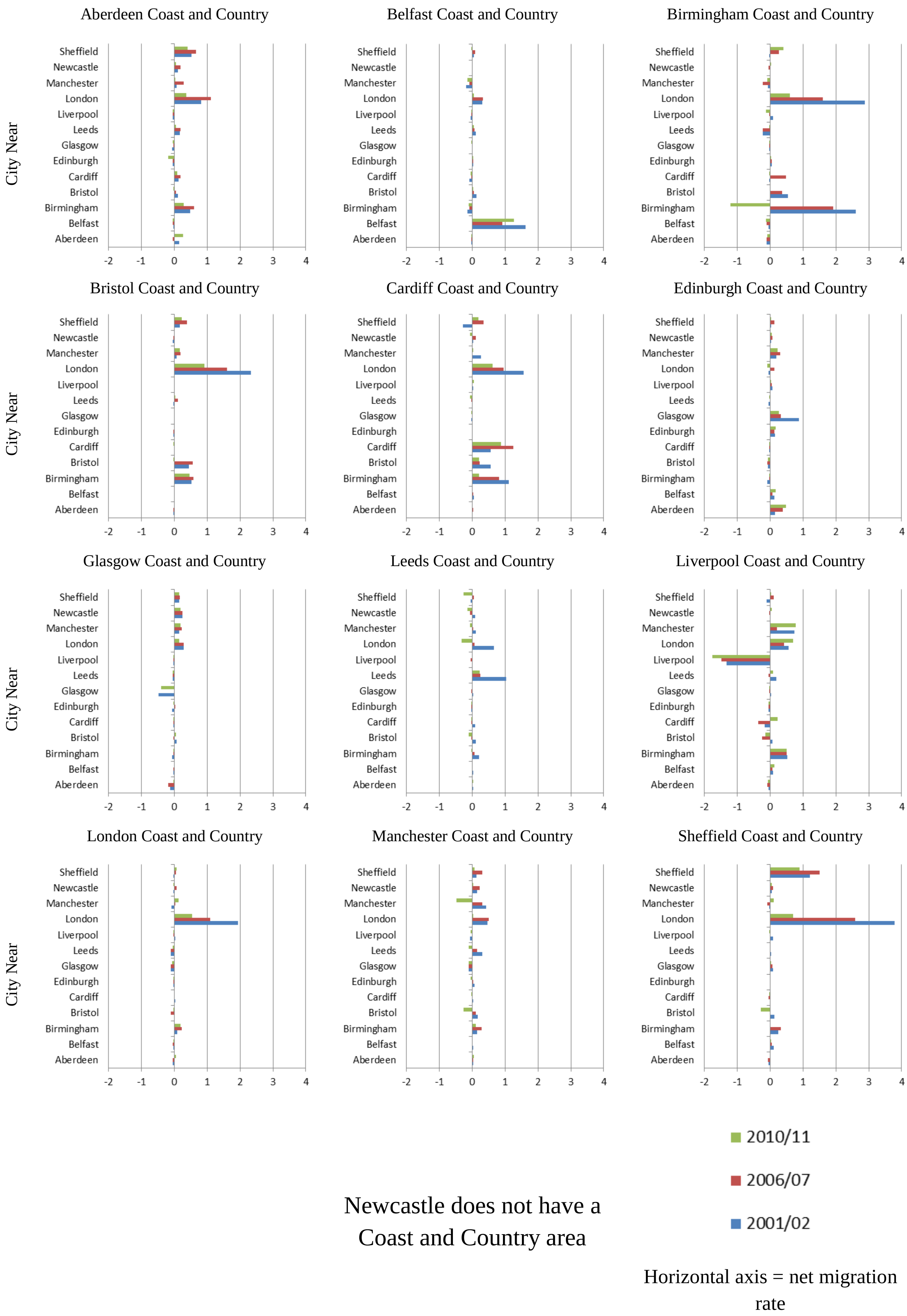


Figure 6.7: Net migration rates for Coast and Country areas in relation to City Near areas, 2001/02, 2006/07 and 2010/11

A mixed picture emerges from these graphs, but the one key finding is that where there is net gain in a Coast and Country area from a City Near area in a different city region, the rate falls substantially over the decade (as exemplified by moves from London Near).

6.5.3 Migration between City Near areas

Migration between City Near areas accounted for 203,865 migrations in 2001/02 which dropped to 183,727 migrations in 2010/11, and Figure 6.8 shows the relationship between all City Near areas. Edinburgh Near stands out as an area that exhibits large scale net gain from other City Near areas, particularly from London and Manchester where the net gain is almost 1 per 1,000 and 0.5 per 1,000 respectively in 2001/02 and 2006/07. This pattern of net gain declines by 2010/11, however, but a fall in net loss to the other two Scottish Near areas negates this pattern somewhat. Aberdeen and Glasgow Near areas exhibit the same net gain as Edinburgh (but at a lower rate), with Glasgow Near appearing to be far more consistent than Aberdeen Near, which moves to a position of net loss along with Newcastle, Manchester, London and Bristol in 2010/11.

Liverpool Near gains substantially from Manchester Near (over 1.5 in 2001/02, falling to 1 per 1,000 in 2010/11). Sheffield, Liverpool, Leeds, Bristol and Aberdeen Near areas gain from London Near in 2001/02 and 2006/07, but all except Bristol show a large decline in this net gain by 2010/11 (when Aberdeen, Leeds and Liverpool begin to lose migrants to London Near, as do Manchester, Belfast, Birmingham, Cardiff and Newcastle Near areas). These reversals from net loss to net gain are less noticeable in London where the resident population is far larger than the other rest areas.

Leeds and Liverpool Near gain a higher net rate of migrants from Sheffield in 2010/11 than in 2001/02, while the net loss to Sheffield Near that London, Birmingham and Bristol Near exhibit falls across the same time period.

The pattern for moves between Near areas can be summarised as one of falling rates (both of loss and gain) in almost all areas, but apart from a small number of areas, the relationships are fairly stable in the direction, if not the intensity, of migration. The role of London Near appears to change across the decade, being far less important as a net gainer of migrants and moving to a position of net loss in some cases.

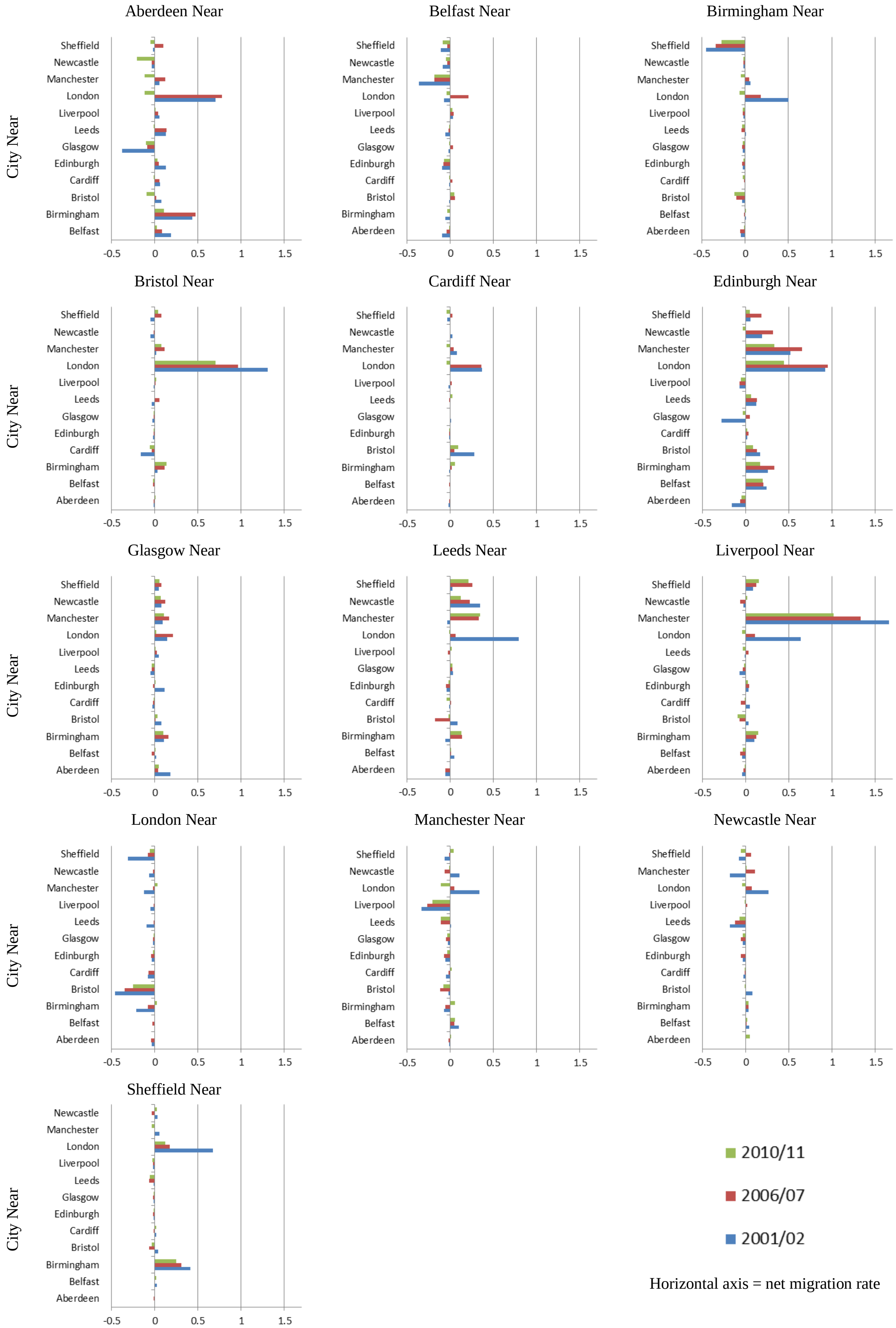


Figure 6.8: Net migration rates for City Near areas in relation to other City Near areas, 2001/02, 2006/07 and 2010/11

6.5.4 Migration between City Cores

Section 6.3 revealed that the proportion of total migration that can be attributed to moves between City Core areas has steadily increased between 2001/02 and 2010/11. Figure 6.9 shows the net migration rate between City Core areas, and the most immediate trend to note is the impact that London exhibits on net migration rates in all other City Core areas. Except for Belfast, all City Cores exhibit a high rate of net loss to London Core, a pattern which generally increases in intensity between 2001/02 and 2010/11. The increase in the size of the net loss is particularly dramatic in Bristol, Cardiff and Leeds where 2010/11 represents a major shift over the rates seen in 2006/07. Manchester and Newcastle are the largest net losers of migrants to London with the rate in all years around two per 1,000 population. Only Aberdeen and Glasgow experience a decline in net loss to London across the decade. Belfast is consistently an area of net gain from London, but this declines across the decade. The effect on the net rate for London is a small one, given the size of the resident population in the London Core.

When the other City Core to City Core moves are taken into account, a number of areas are consistent net losers of migrants: Liverpool and Sheffield show large-scale net loss, although the relationship with Leeds is one of declining loss which becomes gain in Sheffield for 2010/11. Newcastle exhibits mostly loss, apart from with Aberdeen, Liverpool, and Leeds (with which it switches from net loss to net gain through the decade). City Core areas exhibiting largely consistent net gains are Bristol, Belfast and to a lesser extent Manchester (which loses out to the latter two City Cores). Net rates in Birmingham are variously positive and negative but uniformly low (except with London). Cardiff Core loses out to its closest neighbour Bristol, a pattern which increases in its intensity across the decade whilst in Scotland, Glasgow gains from both Edinburgh and Aberdeen, a pattern that is in decline between 2001/02 and 2010/11.

To summarise the patterns seen in Figure 6.9, the City Cores that show net loss to other City Cores do so at a higher rate than those that show net gain, and almost all City Cores lose migrants to the London City Core. This net loss increases over the decade except for Aberdeen and Glasgow where it decreases. The next section examines the role of the City core in the whole system, and the dominant role that London plays is revealed further.

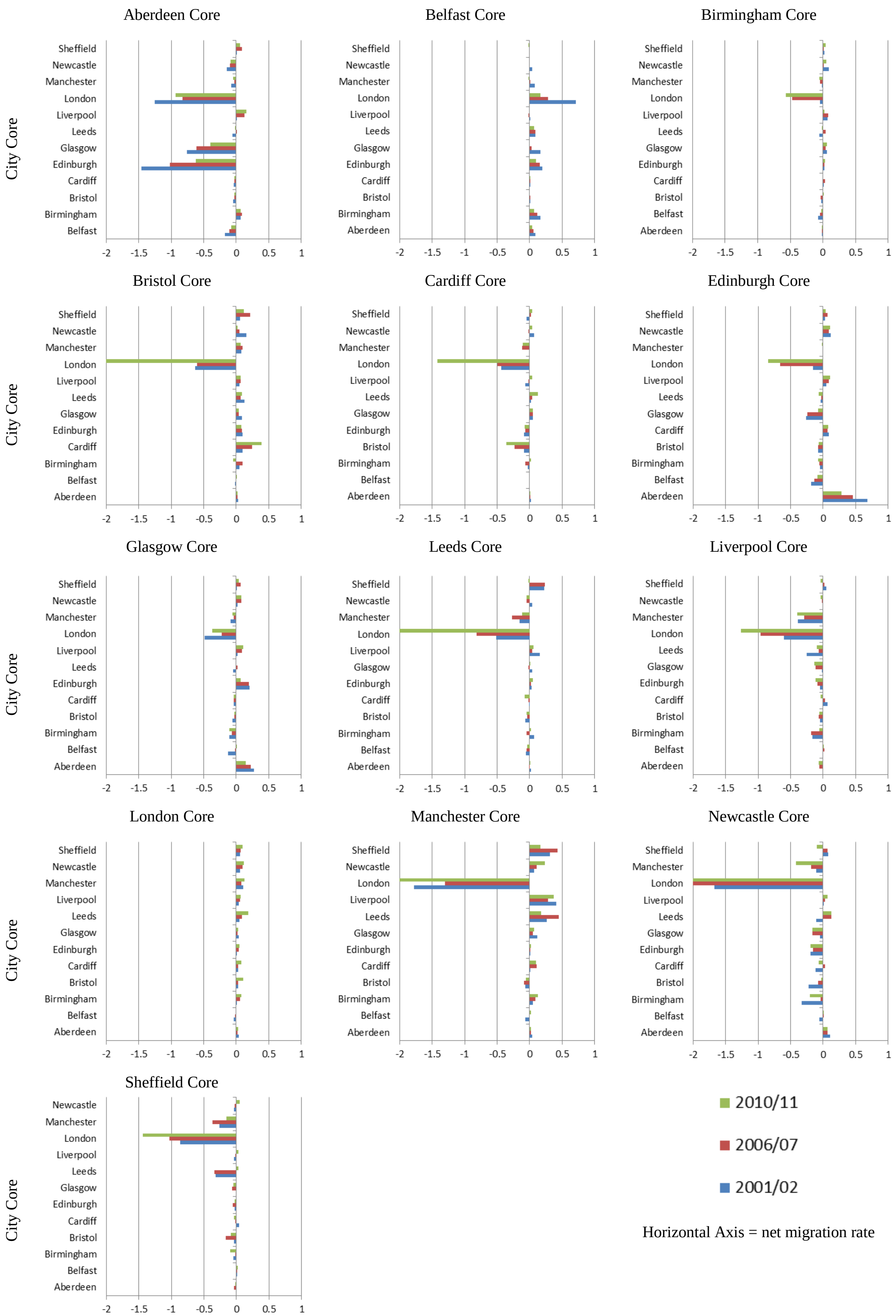


Figure 6.9: Net migration rates for City Core areas in relation to other City Core areas, 2001/02-2010/11

6.6 The role of the City Core

The first section of this chapter showed how the City Cores are the focal point to localised migration networks and Section 6.2 showed how the overall pattern of net loss from the City Cores to other parts of the city regions has declined in its intensity between 2001/02 and 2010/11 while the proportion of moves occurring between City Core areas has increased. Figures 6.10 and 6.11 show the extent of the connections that the Cores of each city region have with all other LADs in the UK. The situations at the beginning (2001/02) and end (2010/11) are compared for flows into the City Cores from other LADs (Figure 6.10) and out of the City Cores to other LADs (Figure 6.11) in the form of the migration rate in the LAD of origin/destination respectively. The general pattern is for LADs closer to the City Core to exhibit higher rates of migration with that Core (both in and out), but the intensity of this trend varies between areas. The geographical influence varies from fairly local, where only a small number of LADs demonstrate high migration rates, to a much broader influence.

Table 6.4 provides a summary of the patterns seen in Figures 6.10 and 6.11, showing the number of LADs with migration rates of between 0.6 and 2 and over 2.1 per 1,000 resident population in 2001/02 and 2010/11 for both outflow and inflow to/from the City Cores. Looking first at inflow to the Cores (that is, the out-migration rate for all other LADs), the biggest influence and largest changes are exhibited by London, where 345 LADs have an out-migration rate of over 0.6 in 2001/02 and 336 have an out-migration rate of over 0.6 in 2010/11. The change is driven by a drop of 23

in the number of LADs in the category of 0.6 to 2 per 1,000, but an increase of 14 in the number of LADs exhibiting a migration rate of 2.1 and over. The only other LAD exhibiting this pattern is Bristol, but on a much smaller scale (a fall of two at the lower rate and increase in one LAD over 2.1 per 1,000). Alongside London and Bristol, Edinburgh is the only other Core to see a fall in the number of LADs with a migration rate of over 0.6 per 1,000 (with a loss of seven in total). The only discernible trend in Figure 6.10 can be seen for Edinburgh, where the fall is driven by LADs in Northern Ireland and in England.

Rates to the City Cores in 2001/02

Rates to the City Cores in 2010/11

Figure 6.10: Out-migration rate for LADs to City Cores in 2001/02 and 2010/11

Rates from the City Cores in 2001/02

Rates from the City Cores in 2010/11

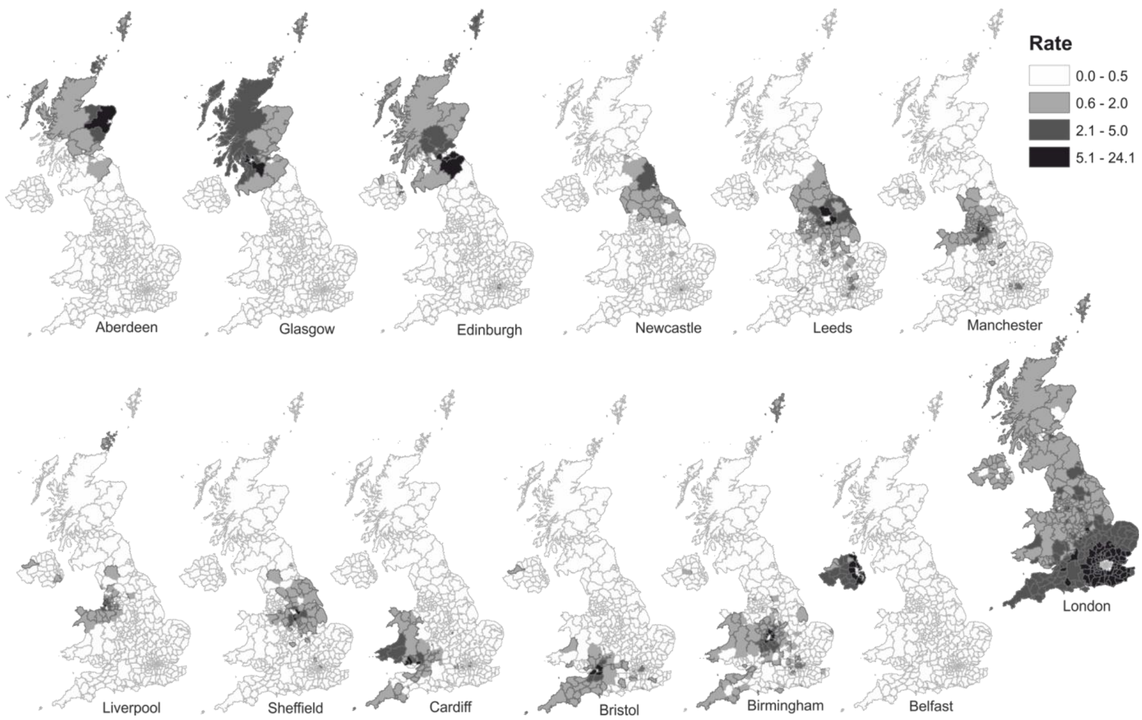


Figure 6.11: In-migration rate for LADs from the City Cores in 2001/02 and 2010/11

All other City Core areas see gains, and this is driven by a rise in the number of LADs exhibiting between 0.6 and 2 out-migrants per 1,000 population. The largest increases are in Leeds (19 LADs), Sheffield (18 LADs) and Liverpool (14 LADs). These increases can be seen in Figure 6.10; Leeds begins to attract migrants from the South East in 2010/11, migration rates to Sheffield appear to increase in a number of LADs in the Midlands while the change for Liverpool can be attributed to LADs in Northern Ireland and the North West. In summary, the draw of most City Core areas increases between 2001/02 and 2010/11 (in terms of migration rates over 0.6 per 1,000), except for London, Edinburgh and Bristol. However, London and Bristol do extend their influence in the number of LADs that exhibit the highest migration rates (over 2.1 per 1,000).

The migration rate for flows from the City Cores into all other LADs reported in Table 6.4 summarise the patterns seen in Figure 6.11. Again, the influence of London is clear: 364 LADs with an in-migration rate of over 0.6 per 1,000 in 2010/11 and 344 LADs in 2010/11. The pattern of change for London is different to that seen for inflow as the decline between the beginning and end of the time series is driven by a drop in the number of LADs exhibiting migration rates of 2.1 and over, of which there are 23 fewer in 2010/11 than 2001/02. The geographical pattern of this decline can be seen in Figure 6.11 and is driven primarily by LADs in the South East. Overall decline in the number of LADs exhibiting a rate of over 0.6 per 1,000 can be seen in Birmingham (five fewer), Edinburgh (three fewer) and Aberdeen (one fewer). All other City Cores see an increase in the number of LADs that have a migration rate of over 0.6, but the change is driven primarily by an increase in migration rates of between 0.6 and 2 per 1,000; only three City Cores show an increase in the number of LADs with rates of over 2.1. The biggest increases here are seen for Leeds (42 more LADs in 2010/11), Manchester (18 more), Sheffield (15 more) and Cardiff (14 more). The patterns in Figure 6.11 show that Leeds appears to extend its influence to LADs in northern England and the Midlands in 2010/11, Manchester's influence extends east to LADs in Yorkshire and south to the Midlands, Sheffield's influence extends to the North West while in Cardiff it is LADs in the South West and Wales that exhibit an increase in migration rates. Overall, the number of LADs exhibiting a migration rate of over 0.6 per 1,000 resident population where the origin is one of the City Cores increases across the decade, except where the London, Birmingham, Edinburgh and Aberdeen Cores are concerned.

Table 6.4: The number of LADs that exhibit medium (0.6 to 2) and high (>2.1) migration rates in relation to flows to and from the City Cores

	To Core 2001/02		To Core 2010/11		To Core change		From Core 2001/02		From Core 2010/11		From Core change	
Core	0.6 to 2	>2.1	0.6 to 2	>2.1	0.6 to 2	>2.1	0.6 to 2	>2.1	0.6 to 2	>2.1	0.6 to 2	>2.1
Aberdeen	10	5	12	4	2	-1	10	4	11	2	1	-2
Belfast	0	23	0	23	0	0	1	22	3	20	2	-2
Birmingham	70	10	75	11	5	1	82	18	80	15	-2	-3
Bristol	51	4	49	5	-2	1	43	5	45	7	2	2
Cardiff	29	13	31	14	2	1	20	8	32	10	12	2
Edinburgh	25	14	21	11	-4	-3	25	12	23	11	-2	-1
Glasgow	13	18	16	16	3	-2	15	15	15	15	0	0
Leeds	64	9	83	9	19	0	38	10	78	12	40	2
Liverpool	16	5	30	6	14	1	12	6	19	6	7	0
London	167	178	144	192	-23	14	148	216	151	193	3	-23
Manchester	48	7	53	9	5	2	40	10	58	10	18	0
Newcastle	21	3	23	3	2	0	14	3	20	3	6	0
Sheffield	31	5	49	6	18	1	21	8	36	7	15	-1
Total	545	294	586	309	41	15	469	337	571	311	102	-26

These patterns reveal that overall the geographical influence of the City Cores is increasing, but that the number of LADs showing the largest in-migration rates of over 2.1 per 1,000 from Core areas has declined by 26 (driven primarily by London). In contrast, the number of LADs where over 2.1 migrants per 1,000 population move to a City Core has increased by 15 (again driven by London). These results appear to corroborate those reported in Section 6.2, that moves to metropolitan areas from non-metropolitan areas are increasing but that much of the change in migration seen for City Core areas is driven by London. The role that London plays is examined further in the next section.

6.6.1 London Core

Table 6.4 and Figures 6.10 and 6.11 show that London is the dominant City Core in the redistribution of the population through migration, with the number of LADs exhibiting high migration rates (both inflow and outflow) being substantially higher than with other City Cores and the extent of the London catchment (in terms of high migration rates) being by far the largest in the UK. While the average rate of migration from all LADs to London Core increases over the decade from 3 to 3.2 people per 1,000, the average rate of migration in the other direction, from London Core to other LADs falls more dramatically, from 5 to 4.2 people per 1,000. It is this out-migration and the role of London as a centre that redistributes the population of the UK which is picked up in this section. This is an important process, identified in the previous chapter where London's role in the South East 'escalator region' was addressed. Stillwell and Hussain (2010) emphasise the importance of London as a centre for redistributing ethnic minority populations, reporting that in 2001, 12.9 per cent of Indian, 12.4 per cent of Chinese and 10.6 per cent of Black migrations were connected with London as either an origin or destination.

Figure 6.12 shows all gross out-migration flows of over 500 from the London Core to other LADs in the UK, overlaid on top of the rate of migration in these LADs. The rates shown are in-migration per 1,000 resident population in each LAD. The beginning (2001/02) and end (2010/11) of the time series can be compared. In 2001/02 the concentration of net outflow destinations is LADs in the south and east of England. The pattern in 2010/11 has changed substantially, however, with the highest net migration rates being more concentrated in the South East of England, with far fewer migrants moving from London to the Midlands and the north of England. The number of LADs that receive over 500 migrants from London falls from 155 in 2001/02 to 134 in 2010/11. This decline is driven primarily by a fall in the number of LADs that receive between 500 and 1,000 migrants (the black lines in Figure 6.12), which falls from 69 in 2001/02 to 51 in 2010/11. The majority of the LADs that fall out of this group are those seen in the South East and East of England but flows to a number of LADs in the South West fall below 500, as do Hull and the East Riding of Yorkshire. The number of LADs with a flow of over 2,001 migrants falls slightly from 31 to 28.

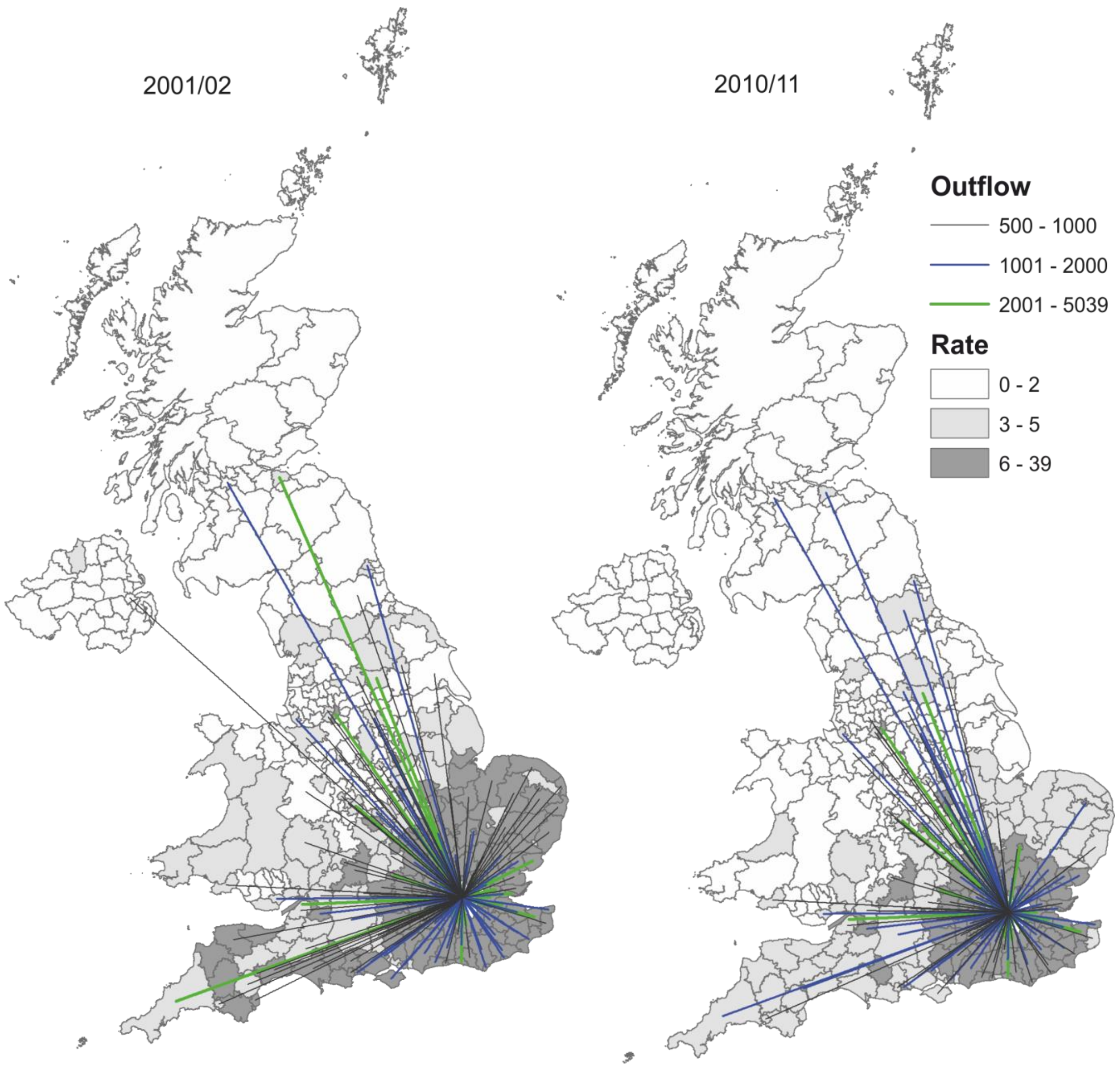


Figure 6.12: Gross flow out of London core to all other LADs and in-migration rate for LADs (from London) in 2001/02 and 2010/11

While London exhibits a substantial influence as a distributor of migrants around the UK migration system, this influence in terms of out-migration rate appears to have declined somewhat between 2001/02 and 2010/11, where the highest net migration rates are much more spatially confined to LADs close to London in the South East of England at the end of the time series.

6.7 Examining distance: implementing a spatial interaction model

In the previous sections of this chapter, results from the migration matrix have been used to analyse the patterns of UK migration between 2001/02 and 2010/11. In order to reduce the burden of information being presented when moves from origins to destinations are analysed, the city region framework has been used to group 406 LADs into 47 functional regions. Given that in each year between 35 and 37 per cent of total LAD-LAD migrants are excluded when only moves across a boundary of one of the city region component parts are considered, this section assesses if the patterns seen at LAD scale hold when the city region framework is used. This analysis is carried out using a doubly constrained spatial interaction model (SIM, which has been implemented using the IMAGE system) and a set of modelled migration flows are presented as a validation tool from which it is possible to analyse how different the frictional effect of distance is when using the LAD and city region geographies.

The basic principle of the SIM in the context of migration is that three variables dictate the magnitude of each modelled flow: the sizes of the origin and destination populations both influence the volume of migration from/to that area and migration is inversely proportional to the distance between origin and destination where *"migrants tend to minimise the uncertainty about a move by favouring closer destinations over more distant ones"* (Fotheringham and O'Kelly 1989, p.9). This preference for shorter distance migration over longer distance migration is termed distance decay (and given the Greek symbol beta, β). The phenomenon is reported by Ravenstein (1885) as one of the laws of migration and confirmed in a large number of contemporary studies (Kalogirou 2005; Dennett and Wilson 2013; Fotheringham and Rogerson 1993; Singleton *et al.* 2012).

Using a doubly constrained SIM (as originally specified by Wilson, 1970; 1971), modelled results at LAD and city region scale have been produced, with the constraints being the marginal origin or destination total for each LAD/city region component part. By producing modelled results at LAD and city region scale, it is possible to assess the effect that choosing a classification which groups LADs has on results as *"the routine is predictive rather than explanatory…and merely seeks to allocate a known number of outflows and inflows to links between these origins and destinations"* (Fotheringham and O'Kelly 1989, p.3). Effectively, using the SIM provides a consistent way of modelling flows within the 406x406 LAD matrix and 47x47 city region matrix, so the results are comparable. Similarity between the results at two spatial scales would suggest that aggregating LADs to city regions still retains the underlying spatial structure of the data. As described by Daras *et al.* (2013), the doubly constrained SIM takes the form:

$$M_{ij} = A_i O_i B_j D_j d_{ij}^{-\beta} \quad (6.1)$$

where M_{ij} is the number of migrations between zone i and zone j; O_i is the total number of out-migrations originating from zone i; D_j is the total number of in-migrations terminating in zone j; A_i and B_j are balancing factors that ensure out-migration and in-migration constraints are satisfied (here the marginal totals of the matrix) while $d_{ij}^{-\beta}$ represents the distance term where β is the distance decay parameter. Such a doubly constrained SIM is implemented in a recent study by Dennett and Wilson (2013) who use it within a multilevel framework to model subnational international flows in Europe between 2002 and 2007. They choose the doubly constrained model as it makes *"use of the maximum amount of available data"* (p.1504) given that the marginal in and out-migration totals are available to them (as they are for the LAD and city region tables used in this chapter).

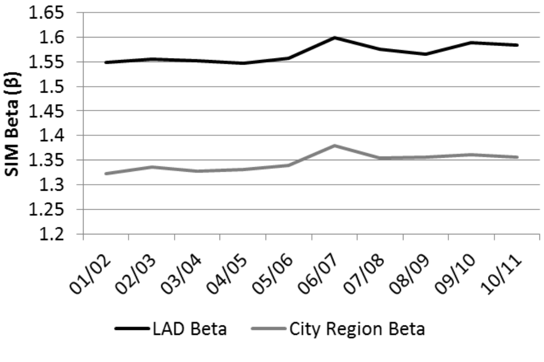
Table 6.5: SIM beta values and mean migration distance for predicted model of LAD and city region flows, 2001/02 to 2010/11

Year	LAD SIM Beta (β)	City region SIM Beta (β)	LAD MMD	City region MMD
2001/02	1.548	1.323	100.96	149.24
2002/03	1.556	1.335	99.89	147.92
2003/04	1.552	1.328	100.6	148.86
2004/05	1.548	1.330	100.02	147.92
2005/06	1.557	1.339	98.75	146.67
2006/07	1.599	1.379	95.26	143.69
2007/08	1.575	1.354	96.81	145.00
2008/09	1.566	1.355	96.89	144.69
2009/10	1.589	1.361	94.74	143.60
2010/11	1.583	1.356	95.42	144.02

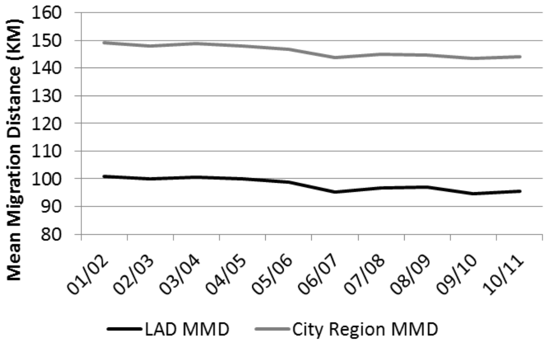
Table 6.5 shows two outputs: the distance decay parameters (beta values) and the mean migration distances for the two spatial interaction models (one at LAD and one at city region scale). Figure 6.13a visualises the beta values and Figure 6.13b visualises the mean migration distance across the time series. The beta value represents the frictional effect of distance, with a higher value meaning that distance has greater effect on the

likelihood of migration. The mean migration distance reports the average distance that all migrants in the system have moved and in both LAD and city region model is the average distance for all moves between area centroids. The mean migration distance is used to calibrate the final SIM output for both LAD and city region, as the mean distance for each SIM is the same as the observed mean distance in the two matrices.

There is considerable consistency in the trend across the decade for the beta value and mean migration distance at both LAD and city region scale. The LAD beta value is higher than the city region beta, which would be expected as there are 406 LADs, compared with 47 city region component areas meaning that distance is more of a hindrance where there are more options for shorter distance migration. For the same reason, mean migration distance is higher for the city region model than the LAD model, showing that the frictional effect of distance is greater on flows between areas that are much larger and involve longer distance movements. Despite these differences, the overall trend through the time series is one of increase for beta values and decrease in mean migration distance at both spatial scales. This suggests that between 2001/02 and 2010/11 the frictional effect of distance increases, which means that people are, overall, moving shorter distances. Both the LAD and city region SIM show a distinct peak of beta values for 2006/07, alongside a dip in mean migration distance.



a: SIM beta vales for LAD and city region areas



b: SIM mean migration distances for LAD and city region areas

Figure 6.13: SIM beta values and mean migration distances for LADs and city regions, 2001/02 to 22010/11

As a validation tool, that the trends seen at LAD and city region are similar is reassuring, as it suggests that using a city region framework to represent the data is a suitable solution that maintains the structure of the migration patterns that are occurring at LAD scale. From the results presented in this section, it is possible to conclude that the similarity in the pattern of beta value and mean migration distance in the two models between 2001/02 and 2010/11 lends strength to the decision to use a city region framework to examine results produced at LAD scale.

6.8 Discussion

The results presented in this chapter show that a number of process are at work in shaping migration patterns between 2001/02 and 2010/11. This section offers some potential explanation of the reasons for these changing patterns using the evidence that is available and looks at the policy implications of these trends continuing.

The broad (start of decade compared with end of decade) trends identified in this chapter are an increase in the number of migrants and proportion of total migration occurring between City Core areas and from City Rest, City Near and Coast and Country to City Core areas; a decline in the number of migrants moving to Coast and Country areas from all other parts of the city region; a fall in the number of migrants moving from City Core to all other parts of the city region; and a decline in moves from City Rest to City Near areas. When looking at other years of the time series, an increase in the first half of the decade (up to 2006/07) for moves from City Core to City Rest and City Near areas is identified, before these moves decline substantially between 2006/07 and 2010/11. The discussion in this section splits these moves into three sections: first, the broad trend of an increase in moves up the urban hierarchy from less urban Coast and Country and City Near areas to more urban City Core areas; second, the trend for a decline in moves down the urban hierarchy to Coast and Country areas and City Near areas from more urban City Core and City Rest areas; and third, the shift in pattern of moves from City Core to City Near and City Rest which changes mid-decade is discussed in relation to the pattern of international migration.

While the patterns of migration discussed in this chapter appear to be changing in the 2000s, counterubanisation is still the prevailing trend which *"has remained a potent force, driving rural social change. It added about 250,000 residents to Britain's rural population between 2001 and 2010"* (Gallent 2011, p.611). The magnitude of this trend does, however, appear to be in decline: in 2001/02, moves from more urban City Core and City Rest to less urban City Near and Coast and Country areas accounted for

525,311 moves, compared with 400,770 in the opposite direction (a difference of 124,541). In 2010/11 moves from more urban to less urban numbered 469,098, whereas moves from less to more urban numbered 422,343 (a difference of 46,755). So while the dominant trend is still one of counterurbanisatrion, the pattern is in decline and evidence suggests a consistent year on year reversal is underway. Couch *et al.* (2009, p.339) outline a lifecycle for British cities: *"urbanisation, suburbanisation, counter-urbanisation and reurbanisation,"* (a set of processes previously covered in detail by Champion, 2001). Given that the previous two decades were characterised by counterubanisation (discussed in the previous chapter) and the results reported in this chapter which show an increase in moves up the urban hierarchy, the 2000s can be described fairly convincingly as a period of reurbanisation.

That a period of reurbanisation is occurring during the 2000s is confirmed by Rae (2013) in a study using results from the 2001 and 2011 Censuses. In an analysis of English cities (the English Core cities identified in this chapter, plus Nottingham), Rae highlights the 'return to the city' which is evidenced by an increase in urban populations and is particularly prevalent in inner city areas. This reurbanisation is similarly identified in the case of Birmingham, Bristol, Cardiff and Swansea by Bromley *et al.* (2007), in the case of Liverpool by Couch *et al.* (2009) and in Birmingham by Barber (2007), who finds that between 2003 and 2007, residential expansion occurred beyond the immediate city centre.

Much of the literature pertaining to the process reurbanisation is framed in the context of gentrification, and there is evidence that the gentrification of city centres in the UK is contributing to the pattern seen in the past decade. This gentrification is widely discussed in terms of segregation – as a displacement of existing inner city populations or a process geared to households with specific demographic characteristics. Buzar *et al.* (2007, p.671) argue that the rise of 'non-traditional' inner city households is driven by the consequences of the 'second demographic transition', namely *"the rise of living alone, delayed child-bearing and the destabilisation and dissolution of traditional family structures".* The problem of heterogeneity is further exacerbated by self-imposed segregation, as *"affluent (childless) families and individuals gravitate towards 'family absent' locations"* (Smith 2011). Specific examples of this process are given by Butler (2007, p.777) in the case of London Docklands where gentrification is a process that has attracted young single people *"who* *do not wish to feel obligations to their neighbours and to socialize”,* and by Couch *et al.* (2009) who, using the example of Liverpool, highlight a problem caused by recent inner city housing development which leads to heterogeneous neighbourhoods, containing adult and student populations with very few young families and children. A similar trend is reported by Seo (2002) for inner city areas of Manchester and Glasgow in the 1980s and 90s where, using a variety of survey sources, Seo finds that new residents in gentrified inner city areas are young single adults or families without children of whom many are also owner-occupiers and white collar workers with high incomes. The gentrified areas are generally separate from existing residential areas, so while not directly displacing existing low-income households, *“systematically discourages low-income households by providing expensive housing”* (p.120) which, along with the provision of better facilities for the gentrified areas, creates spatial segregation.

Alongside this gentrification there is, of course, a more general discourse on the role that City Cores play in terms of employment and economic development. The role of London as an escalator region is discussed at length in the previous chapter, along with Champion (2013)’s recent assessment of other Core cities which act in much the same way – providing improved economic opportunities for people who move to them compared with people who do not. There are also the wider benefits of regeneration projects not directly associated with the process of gentrification: Newton (2009) cites the development of London’s Kings Cross station, Birmingham’s Bull Ring and the Salford Quays as schemes which deliver ‘joined up’ development which includes new homes, improved transport links and the creation of retail, leisure and business facilities. Newton (2009) argues that by providing these combined features, such development contributes to wider regeneration and increases employment opportunities. Gentrification and redevelopment of Core cities may well be contributing to the pattern of reurbanisation seen in the data, but is also driving demographic change in the UK’s cities.

The second trend identified in this chapter, a fall in migration to less urban City Near and Coast and Country areas, from more urban City Core and City Rest areas (alongside an increase in moves from City Near and Coast and Country areas to City Core areas) can be as a consequence of the reurbanisation process. If City Core areas are becoming more attractive destinations, they may draw people who would have previously moved to more rural areas. There is also limited evidence that house prices in rural areas are rising faster than wages, a process which is pricing out families who historically make the move to more rural areas (as investigated in Chapter 8). The National Housing Federation (2013), using average house price sales from the Land Registry, report that average rural house prices have increased by 82 per cent over the past decade, while wages have only risen by 17 per cent in the past five years (as reported in the ONS Annual Survey of Hours and Earnings). In the academic literature, the discussion over gentrification has been expanded to encompass all parts of the urban hierarchy, including rural areas (Smith 2002a; Scott *et al.* 2011) where "*remorseless displacement of lower and middle income groups by those with more resources"* (Shucksmith 2011, p.594) is coupled with planning policy which limits the supply of rural housing (Gallent 2011; Shucksmith 2011) to create an environment that excludes large sections of the rural population and potential in-migrants who would traditionally move from more urban areas.

While the explanation for a decline in moves to Coast and Country and City Near areas is not easily summarised, the implications for rural areas if the trend is to continue are more clear. Bosworth (2006, p.13) highlights the substantial contribution that in-migrant driven business makes to rural economies in the UK, suggesting that it is, *"significantly greater than agriculture, forestry and fishing"* (attributing 1.9 newly created jobs for each new in-migrant microbusiness in the north east of England). These in-migrant business owners are also actively involved in local networks, express a desire for local growth, bring diversity to the local economy and, as Bosworth (2010, p.977) argues, *"are employing local people, trading with local firms, and providing important local services".* Kalantaridis (2010, p.426) echoes these findings, reporting that rural in-migrants contribute to entrepreneurship through their educational attainment, access to networks of contacts established prior to migration and an ability to, *"tap into non-local resource and markets"*, while Kilpatrick *et al.* (2011, p.625) draws similar conclusions. As a result, in-migration can have a substantial positive impact on economic development in rural areas.

Findlay *et al.* (2000, p.346), in a study of labour markets in rural Scotland, concludes that in-migrants make a *"substantial contribution to employment growth and restructuring of rural job markets".* This job creation brought about by rural in-migration is also found by Stockdale *et al.* (2000), and has the effect of raising local wages and investment in the rural housing stock. Stockdale *et al.* (2000, p.254) also find that rural in-migration has the effect of rejuvenating areas as migrant households are *"notably younger and contain a higher number of school age children".* From a policy context, Stockdale (1992), in a study of repopulation of rural areas in Northern Ireland, suggests that state intervention through planning policy and the availability of housing is influential in improving rural mobility.

The third trend identified in the migration dataset is a mid-decade shift in the pattern of moves from City Core to City Rest and City Near areas: between 2001/02 and 2006/07, the number of migrants who made this move increased, while between 2006/07 and 2010/11 the number declined substantially. Here, the pattern of international immigration to the UK and its relationship with internal migration may help to explain this mid-decade change. The number of international immigrants increased between 2001/02 and 2006/07, where it peaked at 605,600. From 2007/08 onwards, the number of international immigrants declined and there were 32,290 fewer immigrants in 2010/11 than there were in 2006/07. In the previous chapter, the negative correlation between the internal migration rate and the international migration rate was identified, so a net loss of internal migrants for a LAD correlates with a net gain in international migrants and *vice versa.* If an international immigrant's first destination is a City Core then a subsequent move has the potential to be to a City Rest or City Near area. That the initial destination of many international immigrants is a City Core area is backed up by results from the 2001 Census, where the 13 City Core areas used in this analysis account for 42 per cent of the 407,548 international immigrants who arrived in the year to 29th April 2001 (29 per cent arrived in London alone). Finney and Simpson (2008, p.80), in a study of ethnic minority migration rates reported in 1991 and 2001 Census data, find that *"a period of high mobility within Britain follows immigration,"* while the predominance of London in the redistribution of international migrants has been identified by Champion (2005, p.102), citing the *"marked acceleration in the UK's net migration gains from overseas since the 1990s and London's predominant role in accommodating this"* as the driving force for the changing migration balances in London to mid-2003.

The largest increase in international immigration to the UK was seen between 2003/04 and 2006/07, a pattern brought about by the enlargement of the European Union (EU): Pollard *et al.* (2008) report that around one million A8 migrants arrived in the UK between 2004 and 2007. The limited research carried out on the internal migration patterns of recent A8 migrants appears to back up the assertion that a high level of internal migration follows a peak in international immigration. Using data from the School Census, Jivraj *et al.* (2012, p.502) find that recent immigrant pupils are very mobile, being three times more likely to move within a one year period of arrival than the wider pupil population and that this move *"is more likely to be over a long distance and/or between local authority district boundaries".* They also find that recent A8 immigrant pupils are moving 'down the urban hierarchy' (from more urban to less urban areas), with the largest internal migration moves being out of inner London. These findings are echoed by Trevena *et al.* (2013) in the case of adult Polish migrants who, *"tend to be most internally mobile in the initial stages of their migration experience, before they develop stronger ties with their place of residence".* Both Jivraj *et al.* (2012) and Trevena *et al.* (2013) report a scarcity of studies examining the internal migration of A8 migrants in the UK, but Jivraj *et al.* (2012, p.492) point towards a consensus in continental European and American literature that "*newly arrived immigrants tend to be much more mobile than established migrants and the non-migrant population".* A large increase in the number of A8 migrants between 2003/04 and 2006/07 (against the backdrop of general increases of other, non-A8 migrants between 2001/02 and 2006/07) could explain the increase in moves from City Core to City Rest and City Near areas, while the fall in international immigration seen from 2007/08 onwards corresponds with a decline in moves from City Core to City Rest and City Near areas.

This section has offered some commentary on three trends that have been identified in the migration dataset: an increase in the number of migrants who move to City Core areas, a decline in the number of migrants moving from more to less urban areas (and the corresponding increase for moves in the other direction) and a mid-decade shift in the pattern of moves from City Core to City Rest and City Near areas. The process of reurbanisation driven by urban gentrification was offered as an explanation for the first trend and as a partial explanation for the second, alongside a link between rising rural house prices squeezing out the families who historically move from more to less urban areas. The final trend for a change in the pattern of migration from City Core to City Rest and City Near areas was linked with the pattern of international migration, a process intensified by the increase in number of A8 migrants between 2003/04 and 2006/07.

6.9 Conclusion

This chapter has focused on the flows and connections between LADs within the migration matrix between 2001/02 and 2010/11. The LADs were grouped into city regions in order to provide a framework in which to analyse the substantial amount of data that is available for the time series (the 406x406 LAD migration matrix was effectively reduced to a 47x47 city region matrix) and a spatial interaction model was used to model these flows at both LAD and city region scale in order to look at the effect of distance on migration.

Within the 13 city regions, the net loss from City Core areas to all other parts of the region fell throughout the decade (except in Aberdeen) while the net rate of gain in Coast and Country areas fell (except in Cardiff). The net rate of gain for City Near areas declined (except in Liverpool and Aberdeen) while the pattern for City Rest areas was less clear. When the analysis was extended to include all UK moves within and between the various parts of the city regions, a number of clear patterns emerged. The proportion of moves from LADs in the City Rest, City Near and Coast and Country to the City Core areas increased over the decade. Coupled with this, the proportion of moves out of City Core areas to Coast and Country areas has declined and the proportion of moves between the City Core areas has increased. This leads to the conclusion that moves between metropolitan areas (the more urban City Core and City Rest LADs) have increased between 2001/02 and 2010/11 while moves between non-metropolitan areas (the less urban City Near and Coast and Country LADs) have declined. Moves from metropolitan to non-metropolitan LADs have declined while moves from non-metropolitan to metropolitan LADs have increased. The trends seen in the time series began in 2004/05, where before this the proportion of migration for each type of move was relatively stable.

The importance of London as a redistributor of migrants in the UK is clear across the decade, but the extent of the catchment (LADs exhibiting high migration rates) for moves out of London declines visibly between 2001/02 and 2010/11. At the same time, the move from the majority of other City Cores to London appears to increase between 2001/02 and 2010/11.

When a spatial interaction model is used to model the migration flows, the pattern of the beta value and mean migration distance at LAD and city region scale are consistent between 2001/02 and 2010/11. Both report beta values rising and migration distance falling. These results add strength to the choice to use a city region framework to represent migration patterns in this chapter.

The underlying age structure of these moves is picked up in Chapter 8, following the detailed methodology used to estimate the age and sex disaggregation of the migration flows which is spelled out in the next chapter.

Chapter 7
Disaggregating the database by age and sex

In the preceding chapters of this thesis, the focus has been on the estimation and analysis of aggregate migration flows in the UK. However, migration propensity varies by age, a phenomenon that has been extensively researched and is well documented in the literature (see Bracken and Bates 1983; Tobler 1995; Rogers *et al.* 2002). Variations in migration propensities are less discernible for males and females but sex may, nevertheless, be an important discriminator in certain contexts, especially when coupled with age (Rogers and Castro 1981). This chapter describes in detail the methods used to add age and sex information to the aggregate migration dataset that was introduced in Chapter 4 and explored in Chapter 5 and Chapter 6. Reference is made to the literature pertaining to estimation and modelling of migration by age and sex in order to inform operational decisions in the development of this methodology.

Estimating migration by age and sex across all 406 LADs creates a huge volume of data and analysing these data presents a challenge in itself. For this reason, the second part of this chapter is devoted to creating an aid to analysis which involves clustering areas based on their age profiles and modelling the age schedules of these clusters to produce a set of simplified results that can be analysed more easily across space and time. This process is not mutually exclusive of the detailed methodology presented in the first part of the chapter as it helped provide insight into the structure of the data and contributed to the formulation of the methods used to estimate age and sex. The analysis of the age/sex disaggregated estimates is carried out in the following chapter with reference being made to cluster associations where it aids interpretation, which is especially useful when assessing the similarity between migration patterns at different ages.

7.1 An overview

Producing an origin-destination-age-sex (ODAS) matrix for each year requires estimation of migration flows using an IPF routine in a similar way to that implemented for the aggregate OD matrix presented in Chapter 4. The difference being that the origin (column) and destination (row) totals are replaced with an age and sex disaggregated total for each LAD (for both OAS and DAS). Full consideration of both the data used as the seed value to be adjusted and the data used for the marginal OAS and DAS totals used to adjust the seed needs to be undertaken before the routine can be implemented effectively for the whole ODAS matrix in each year of the study.

7.1.1 Age bands

It is a good idea to clarify the data specification to which the methodology outlined in subsequent sections will adhere. The matrix needs to be disaggregated by two sexes. However, the choice of age bands requires a little more thought. Work reported by Fotheringham *et al.* (2004), Champion *et al.* (2003) and Rees *et al.* (2004a) explains how the MIGMOD migration modelling system uses seven broad ages, corresponding to life course stages: preschool/school age (0-15), leaving home for university or work (16-19), leaving university for work (20-24), forming couples and stating a family (25-29), raising a family (30-44), older working age (45-59) and approaching and beyond retirement (60+). From experimentation with the MIGMOD system, they conclude that even though it would be possible to produce models for finer age groups (down to single year of age), the sparse nature of the data means that model calibration results would be likely to be unreliable. In the work of van Imhoff *et al.* (1997), five year age groups were found to be sufficient to balance the need for goodness of fit and parsimony. In light of the data available (as spelled out in following sections) and a need to produce estimates that provide a good balance between level of detail and their complexity to implement, five year age bands will be used in the modelling of UK migration in this thesis. These five year age bands can easily be aggregated up to life course stages for analysis (as they are in the following chapter), they provide a fine enough level of detail to dissect and interpret trends and overcome many problems with small numbers and rounding that are present in the data (especially in the full ODAS matrix).

The choice of five year age groups also provides the opportunity to model more detailed single year of age (SYA) data. A modelling strategy is detailed in Section 7.5.9 where migration schedules by SYA are produced for clusters of LADs based on 2001 Census data, and while these SYA results are not used in this thesis, shows the opportunity for further research that exists from the ODAS matrix of migration in the future (a point discussed in Chapter 9).

7.1.2 Specifying the estimation scheme

Figure 7.1 shows the data requirement for the estimation of a full ODAS matrix. Blue cells indicate the distribution to be adjusted by the IPF routine while grey cells denote the marginal in and out values by each five year age group and sex (OAS and DAS). The green cells are the total in and out flows for each LAD which are consistent with the aggregate migration dataset presented in Chapter 4 and include both internal and cross-border flows. The orange cell is the control total, this refers to all migrants within the system and the sum of the final adjusted internal cell values. The sum of all margins must match this total. This schematic represents one year, and needs to be repeated across all years of the time series between 2001/02 and 2010/11. The data requirement for filling this estimation schematic is discussed in the next section.

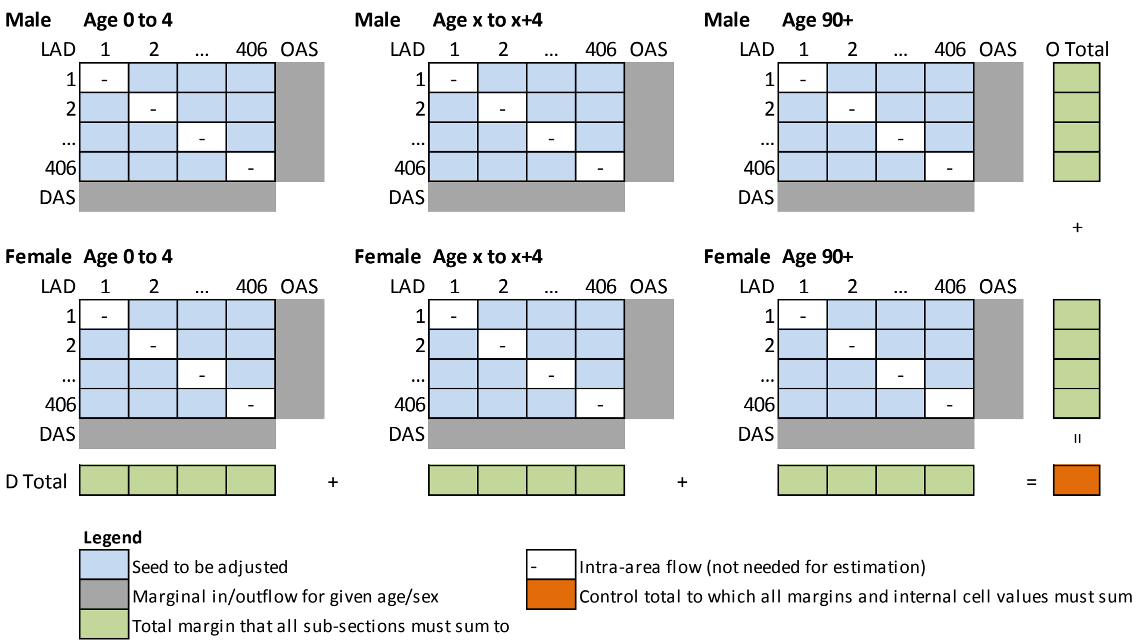


Figure 7.1: The estimation scheme for ODAS migration

7.1.3 The seed value

The seed value (blue cells in Figure 7.1) is the starting distribution which is adjusted to agree with the OAS and DAS margins. Given that *"previous research (Boden 1989; Rees et. al. 1989) has suggested that significant differences exist between the pattern of origins and destinations at different ages"* (Stillwell *et al.* 1996, p.298), it is important to use a starting seed which takes these various patterns into consideration. An ODAS matrix (by five year of age) is available from the 2001 Census in the form of the Special Migration Statistics but is heavily affected by SCAM adjustment (Stillwell and Duke-

Williams 2007) due to the sparseness of the tables when age (and sex) information is disaggregated from the totals. The only part of the SMS unaffected by SCAM is the inflow values for LADs in Scotland. A solution which retains the patterns exhibited at different ages but overcomes the problem of SCAM adjustment is to use seven broad age bands with no sex disaggregation: 0-14, 15-19, 20-24, 25-34, 35-44, 45-64 and 65+. These age bands are a compromise between data availability, SCAM adjustment and accuracy, so the 0-4, 5-9 and 10-14 age group estimates are derived using the 0-14 seed distribution for example. As was the case for the aggregate matrix, the 2001 Census at each age band is used to estimate the 2001/02 interaction matrix, and each subsequent year from 2002/03 onwards uses the previous year matrix as a seed.

7.2 Producing consistent margins for iterative proportional fitting

In Chapter 4, the method used to derive aggregate marginal in and out values for each LAD was outlined. Here, the method used to disaggregate these marginal totals by age and sex is presented. The aim is to retain the aggregate total for each LAD so all estimates will consistently sum to the margins in each given year (the green cells in Figure 7.1). The following sections give an overview of the data available between 2001 and 2011 and provide a solution for this disaggregation by age and sex.

7.2.1 Data

Table 7.1 shows that the most comprehensive age and sex information for the UK is available from the 2001 Census, from which it is possible to derive in and out-migration for every LAD by five years of age and sex, both for OAS and DAS. Internal and cross-border in/out totals are available, as are international inflows. The cross-border data are disproportionately affected by SCAM adjustment as the number of migrations is relatively small, but a robust age distribution can be obtained by combining the rest of UK flows with internal flows.

For the years 2001/02 to 2010/11, the data are far less complete. Table 7.1 shows that for LADs in England and Wales, OA and DA totals are available for flows to and from the rest of England and Wales by five year of age. Flows to and from the rest of the UK are published (as part of a total flow) by ONS, but are rounded to 100 and as such are not particularly useful, especially in areas where flows are small. The exception is for data available in 2010/11, where these cross-border flows are available in an unrounded format (as are the within England and Wales migration flows) and represent a marked improvement in data availability.

Table 7.1: Data available for origin-age-sex and destination-age-sex totals for internal (within country and cross-border) migration

	2000/01			2001/02 to 2010/11		
Country	Age	Sex	Age by Sex	Age	Sex	Age by Sex
England	C	C	C	To/from E&W[1]	Yes	To/from E&W[1]
Wales	C	C	C	To/from E&W[1]	Yes	To/from E&W[1]
Scotland	C	C	C	Average 2009-2011	To/from all migration types[2]	None
Northern Ireland	C	C	C	None	None	None

Data Available:

(green)	All	(light orange)	Most	(orange)	Some	(dark red)	None

C = Census

[1]Except in 2010/11, where an unrounded rest of UK figure is available by age and sex for each LAD

[2]All migration types = internal, cross-border and international

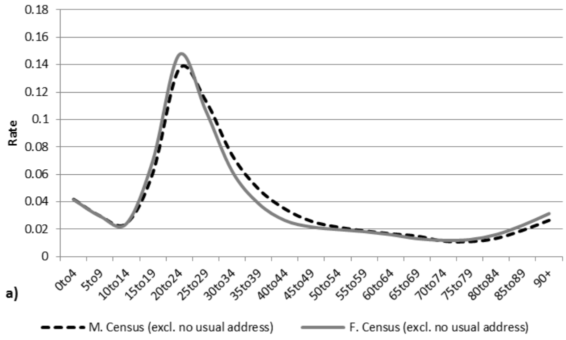
For each LAD in Scotland, inflows and outflows by age are available as an average of 2009 to 2011 flows. In Northern Ireland, no age information is available outside of the census year. Sex information for LADs as origins and destinations is available in all years for England and Wales, whilst in Scotland the distribution is for all migration types (including overseas migrants) and in Northern Ireland no sex information is available. Age by sex information is only available for flows between the 348 LADs of England and Wales. The inconsistency of data available between 2001/02 and 2010/11 for the four UK countries provides a challenge for estimation of the ODAS array over time. Exploration of the more complete migration data from the 2001 Census is undertaken in the following section and an assessment of the register data available after 2000/01 follows on from this.

7.2.2 Census data

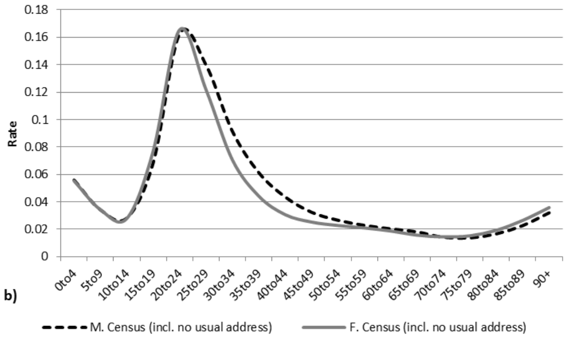
As the 2001 Census provides the most robust data from which to derive marginal OAS and DAS totals for each LAD, some consistency issues need to be explored before the data are used in the estimation process. As part of the 2001 Census questionnaire, migrants were asked where they resided 12 months before the census date. One possible response to this question was a tick box which indicated that the migrant had ‘no usual address’ 12 months before census day. The main aim of including this option was to pick up children aged under one on census day, who were not born 12 months prior to the census and persons without a permanent address (ONS 2004). In practice, over 450,000 people (8% of all migrants) ticked the ‘no usual address’ box and as a result the option was dropped from the 2011 Census questionnaire.

Figure 7.2a shows the migration rate of all UK migrants by five year of age where those who stated ‘no usual address’ are excluded, while Figure 7.2b shows the migration schedule where these migrants are included. It can be seen that the migration rate is disproportionately higher in Figure 7.2b for younger ages (peaking at age 20-24) and that the number of male migrants who stated ‘no usual address’ is also higher than the number of female migrants. This pattern is highlighted in Figure 7.2c which shows the difference between the rates for each sex by age. At the local level, as required for the estimation of an ODAS matrix, census data incorporating no usual address could be used for inflow schedules (DAS), but with no information for the origin of these migrants cannot be used for outflow schedules (OAS). While the more complete data incorporating ‘no usual address’ would be preferable, it is important to ensure consistency between inflow and outflow totals and including these migrants on one margin and not the other would mean that the ODAS table would not balance. The bias towards younger ages and male migrants who reported ‘no usual address’ also suggests that migrants from outside of the UK may be included within this total.

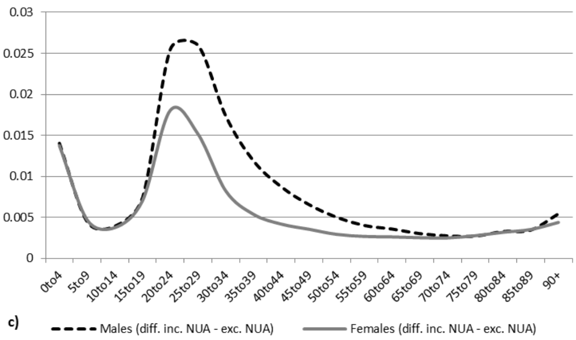
When the number of migrants at each age in the two schedules seen in Figures 7.2a and 7.2b are compared as proportions of the total flow, it can be seen that they are fairly consistent (Figure 7.2d). The need for consistency between in and out flow totals and the possible inclusion of overseas migrants means that the rates which exclude ‘no usual address’ information are preferable in the estimation of OAS and DAS margins.



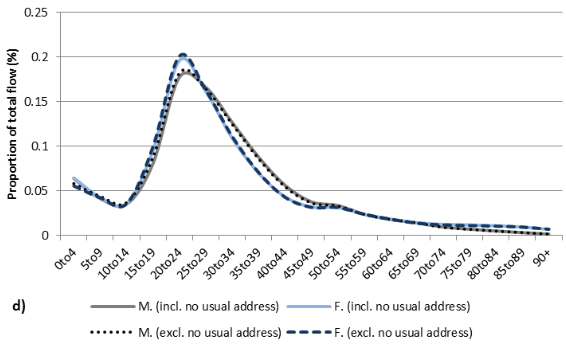
a: Where no usual address information is excluded



b: Where no usual address information is included



c: Difference between rates where no usual address data is included/excluded



d: Difference between the shape of curves when calculated as a proportion of total flow

Figure 7.2: Migration rates by age and sex derived from the 2001 Census

A final check of the census data can be undertaken by assessing the number of small flows in the OAS and DAS totals. The SCAM adjustment applied to the data means that any cell containing a 1 or 2 was adjusted to 3 or zero. Figure 7.3 shows the percentage of cells where the OAS or DAS is three or less.

Until age 55-59, less than 0.5 per cent of all cells in each age band for either sex has a count of 3 or less. This increases gradually to 4.9 per cent for male OAS and 5.7 per cent for male DAS in the 75 to 79 age band. From age 80 to 84 onwards, the proportion of OAS and DAS margins that contain 3 or less increases much faster for males than females, as would be expected given that male life expectancy falls behind female life expectancy. It is interesting to note that over 90 per cent of LADs have three or more female migrants over the age of 90. The pattern in these data suggests that the OAS and DAS flows derived from the 2001 Census would be suitable for modelling

ODAS migration flows for all ages as it points towards a dataset that is largely unaffected by rounding issues.

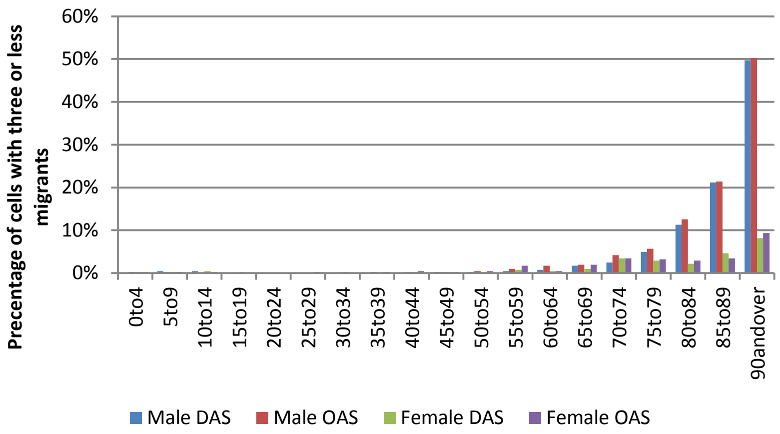
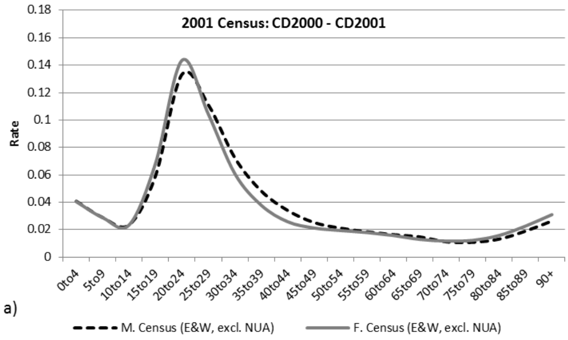
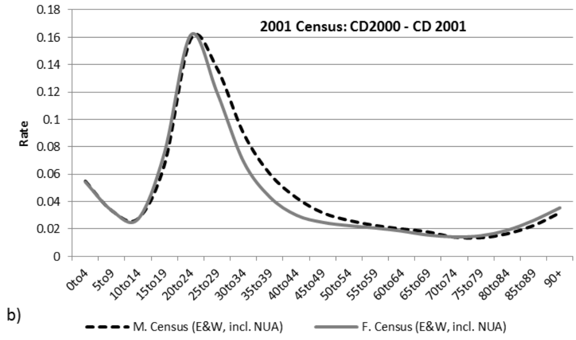


Figure 7.3: The proportion of LADs where the OAS or DAS total is three or less within each age by sex grouping

7.2.3 The problem with using registration data

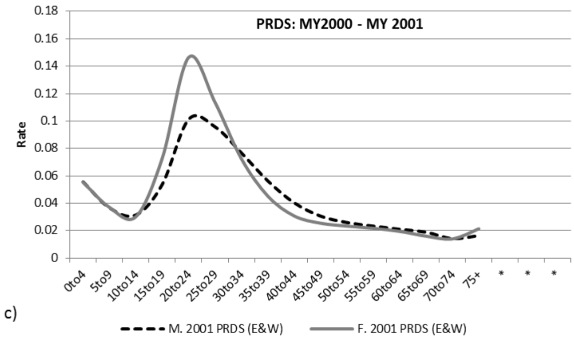
From the picture of data availability presented in Table 7.1, it is clear that inconsistency in data availability across the UK and across the time series presents a real problem when trying to use age and sex information derived from registration data. Problems also present themselves when assessing the usability and reliability of the data. Figure 7.4 shows a comparison of migration rates by age and sex derived from the census (4a shows the data excluding 'no usual address' and for completeness 4b includes this data) and those derived from PRDS patient register data for England and Wales. Data for England and Wales are the only UK information that is consistently available for the time series 2001/02 to 2010/11. The issues with the PRDS are immediately apparent, the schedule for females is fairly consistent with the census data but for males the rate is far lower. This pattern is repeated through the PRDS time series, and Figure 7.4d shows that a similar (albeit slightly less acute) pattern exists in the 2011 PRDS data. This is a well-documented phenomenon (Smallwood and De Broe 2009; Fotheringham *et al.* 2004) attributed to under registration (delays in re-registering with a GP after migration) by young males and is explored in more detail in the next section. Data at older ages are also not available in the 2000/01 to 2009/10 PRDS data, with the oldest category being over 75.

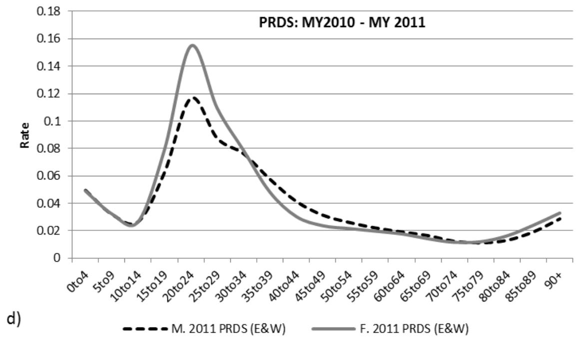




a: Excluding no usual address data

b: includes no usual address data





c: 2001 PRDS data

d: 2011 PRDS data

Figure 7.4: A comparison of migration schedules in England and Wales from the 2001 Census and PRDS data for 2000/01 and 2010/11

7.2.4 Sex Ratios

The sex ratio of migrants does vary by age, Figures 7.2a-c show that for the whole of UK distribution from the 2001 Census, the female peak begins earlier and slightly higher for the 20-24 year old age band. This female peak tails off faster than the male rate, meaning that the proportion of males that migrate between the ages of 25-29 and 50-54 is slightly higher than the female proportion, before females overtake males again at age 80-84. The reasons for these differences are explored more in the next chapter but the point to note here is that the modelling strategy needs to take into account some variation in migrant sex ratio by age. Having said this, Rogers and Castro (1981, p.2) in their seminal modelling paper find that *“sex selectivity is much less pronounced than age selectivity and is less uniform across space and time”.* Stillwell (2005) advocates that sex should be incorporated into any model of migration if it is possible.

A problem with sex ratios derived from data sources other than the census is the effect of under-reporting by males, especially in the young adult age bands. Smallwood and De Broe (2009) look at sex ratio patterns in the 2001 Census, the 2001 MYE and the 2001 PRDS. In the GP register (PRDS), they find that *"the sex ratio pattern is distinct both from a 'natural' population (where only births and deaths determine the pattern) and from the MYEs"* (p.43). They refer to 'missing' young men, with the pattern being replicated in Scotland and Northern Ireland and conclude that trends in the sex ratios of the PRDS are *"driven primarily by the registration and cancellation of female migrants"* who are *"more compliant with administrative requirements to register with a GP on arrival at a new location"* (p.48). This was found to be particularly evident at ages 18-24 where men use doctors less frequently than women and as such do not appear on the PRDS. Smallwood and De Broe (2009) conclude that at subnational level (LAD) errors in sex ratios are much more likely to arise from issues with the GP registration data. By comparison, in Nordic countries (which are generally agreed to have better population recording systems than the UK), sex ratios do not depart substantially until beyond the age of 50.

The under-registration of young adult males in the NHSCR is highlighted by Fotheringham *et al.* (2004) and *"supports the suspicion that young males have a tendency to postpone registering with a GP until later years"* (p. 1639). This undercount is, however, not found to be spatially stable and varies considerably across HAs. The solution presented for the MIGMOD model was to adjust the number of male migrants reported in the NHSCR based on the sex ratio reported in the SMS, therefore adjusting the number of males in each age group based on the assumption that there was *"little or no bias in the recording of female migration in the NHSCR"* (p. 1639). The result of this adjustment in some areas was an increase in the number of male migrants by up to 50 per cent. ONS (2012b) found that, overall, there was little difference between males and females for moves between LADs in England and Wales in 2011, with both peaking at age 19 and a smaller second peak where more males moved at age 23, whereas for females the peak was at age 22. This analysis, carried out at the national level, does mask variations at the subnational level.

Given that the PRDS in England and Wales is the most consistent in terms of availability between 2001/02 and 2010/11 and the problems that are present within the data, the use of 2001 Census age profiles looks to provide a viable solution for estimation in preference to the more timely but unfortunately less reliable register data. The next section discusses the consistency over time for age and sex variables and draws on previous studies which estimate migration with a view to implementing an estimation strategy which uses the 2001 Census migration distribution by age and sex for subsequent years up to 2010/11.

7.2.5 The temporal stability of age and sex variables

Having investigated the data availability for estimating OAS and DAS margins and concluded that the 2001 Census data provide the most viable option for estimation, this section explores the temporal consistency of age and sex variables with a view to extending census age and sex distributions across the decade.

In a large scale investigation, van Imhoff *et al.* (1997) specify a model for multidimensional internal migration matrices in Italy, the Netherlands and the UK. Given the amount of data available for modelling a full ODAS array, they investigate what data needs to be included with a view to producing an efficient modelling strategy (with an acceptable trade-off between goodness of fit and parsimony). They conclude that at least origin-destination, age-origin, age-destination and sex-age information should be included. Their optimum model specification had origin-destination effects independent from age and sex. The temporal stability reported by van Imhoff *et al.* is echoed in projection work by Sweeney and Konty (2002), while in a review of both studies, Wilson and Rees (2005, p.345) conclude that *"one can concentrate on relatively few variables in formulating projection assumptions"* given the temporal stability of the origin-destination effect.

Tobler (1995) recommends that age is a structural process underpinning migration that is worthy of being a 'law' of migration (after Ravenstein 1876) and in a study of inter-state migration in the USA between 1935 and 1990 concludes that the migration system is 'sluggish' with changes in the structure occurring slowly over time. Rogers *et al.* (2002) use data from four consecutive census counts to analyse the age and spatial structure of migration in the USA. They find that *"persistent regularities are exhibited by the age profiles of regional out-migration flows (generation) and by the age-specific destination choices made by these out-migrants (distribution)"* (p.358). These regularities, Rogers *et al.* conclude, can be imposed in empirical studies where adequate migration flow data are not available.

An argument for using a standard age and sex distribution for local authorities is presented by Wright (2008), who outlines the brave NRS assumption that, due to inadequate emigration data, the characteristics of people (in terms of age and sex) who move abroad are the same as those who move to England and Wales – and as such use the England and Wales distribution by age as a proxy. In modelling migration age schedules (which is reported further in Section 7.6), Bates and Bracken (1987) conclude that the reduction in migration flow they observed between 1971 and 1981 is *"independent of flow direction, origin and destination"* which, they argue, justifies the extension of the modelled curve to *"embrace stability over time"* (p.531).

Rogers and Castro (1981) emphasise the *"remarkably persistent regularities"* of age-specific migration schedules, using migration rates in the USA between 1966 and 1971 as an example. They highlight the high rate of migration for infants and young children that mirrors the high rate of their parents, people in their late twenties. Adolescent migration is low but higher than those in their early teens, thereafter rates increased to a peak around age 22. They also find that for migration at different size of aerial unit (within counties, between counties and between states) the age profile appears to be remarkably similar.

The consensus that age and sex variables are relatively stable over time adds further weight to the argument for using 2001 Census data over the less accurate, less complete and more inconsistent (both spatially and temporally) registration data. The following section details the method used to produce consistent OAS and DAS margins from census data and the IPF routine is spelt out.

7.3 Producing consistent margins: Using a 2001 Census migration distribution

Having decided that the 2001 Census provides the most robust age and sex information for the OAS and DAS margins of the table, the migration rate for each age group, derived from the 2001 Census, can be applied to the population at risk in later years to produce an estimate of migration in that year. A major advantage of this method is that it can be consistently applied across the UK as the data availability is the same for all countries. The first step is to calculate the 2001 Census migration rate (cr) for both OAS and DAS in each LAD across each age group:

$$cr_{s,x\ to\ s,x+4} = \frac{CM_{s,x\ to\ s,x+4}}{P_{s,x\ to\ s,x+4}} \tag{7.1}$$

where $x\ to\ x+4$ represents each five year age group for each sex (s). CM is the census migration in each LAD for that age group and P is the population at risk in each LAD within that age and sex group (in the year being estimated). The OAS and DAS rates are calculated separately so the origins are independent of the destinations. This ratio at each age group is applied to the population at risk in a given year (t) in that particular age group to give the estimated number of migrants (M) in time period t to $t+1$:

$$M_{s,x\ to\ x+4}^{t,t+1} = cr_{s,x\ to\ s,x+4} \times P_{s,x\ to\ s,x+4}^{t} \tag{7.2}$$

the sum of all OAS and DAS migrants are constrained to the total in-migration and out-migration for each area (using the proportion of migrants in each age band). Finally, the outflow (OAS) margin is adjusted to match the inflow (DAS) margin for each age band and both sexes so that migrants in each age and sex group have an origin and a destination. These estimates are constrained to the inflow (D^j) and outflow totals (O^i) from the aggregate migration database presented in Chapter 4:

$$\sum_{s}\sum_{x} \widehat{M}_{s,x\ to\ x+4}^{i+} = O^i \tag{7.3}$$

$$\sum_{s}\sum_{x} \widehat{M}_{s,x\ to\ x+4}^{+j} = D^j \tag{7.4}$$

where $\widehat{M}^{i+}$ is the total estimated outflow and $\widehat{M}^{+j}$ is the total estimated inflow. Within the age band 0-4, migrants aged under 1 have been multiplied by 0.5 to account for over-reporting of babies in the 2001 Census whose migration is based on their mother's location one year before the census. The adjustment of 0.5 assumes that, on average, half of babies were born before their mother migrated and are themselves migrants while half were born after their mother migrated and are therefore non-migrants. This assumption is made due to a lack of empirical evidence on the location of the mother at the time of birth. There is evidence however that infants in general are under reported to some degree in the census (Norman *et al.* 2008).

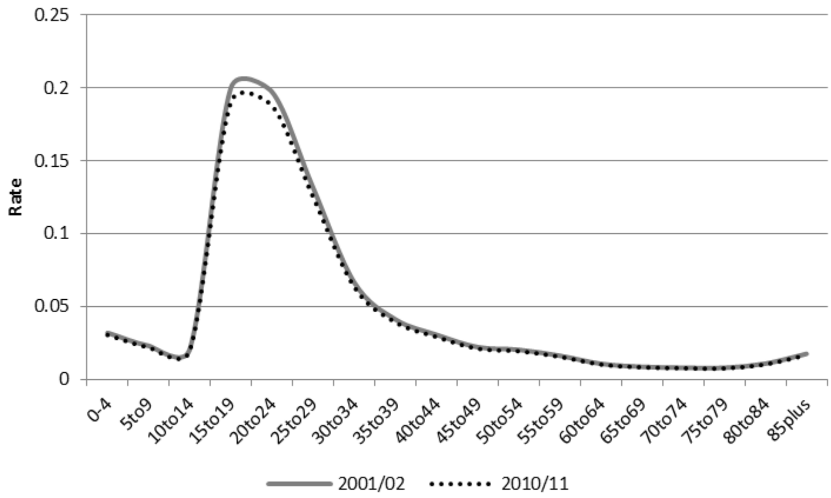


Figure 7.5: Estimated in-migration rate for Manchester in 2001/02 and 2010/11

Figure 7.5 shows (using the example of Manchester) that when these results are converted back to rates, variation is picked up in the estimate, while the overall schedule shape remains the same. This method retains the 2001 Census age profile but allows for the changing demographic structure of an area by applying the age profile to the population at risk each year.

7.3.1 Validation of the estimates

The estimates can be tested at two stages in order to validate the methodology used. The first is to check if the method of applying the total migration rates from the 2001 Census to the total population at risk in each year (in each LAD) is appropriate, and the second is to see how the estimates compare with 2010/11 migration data in England and Wales by each age and sex group. The first check of the data reveals that by simply applying the 2001 Census migration rate to the population in each LAD, the estimated total number of migrants correlates closely with the PRDS, CHI and Health Card data. In each year between 2001/02 and 2010/11, the correlation between the rate derived estimate and migration data from the NSAs does not fall below 0.98 for inflow or 0.98 for outflow ($p<0.01$) and there are no notable outliers (Table 7.2). When the migration rate for these estimates in each year for each LAD is compared the correlation is similarly strong, not falling below 0.90 for inflow or 0.90 for outflow ($p<0.01$). When rates are compared however, some outliers do appear (Figure 7.6).

Table 7.2: The correlation between the estimated and register reported total inflow and outflow for each LAD and the rate in each year compared with the 2001 Census rate between 2001/02 and 2010/11

Year	Inflow Total	Inflow Rate	Outflow Total	Outflow Rate
2001/02	0.981	0.913	0.980	0.922
2002/03	0.982	0.922	0.979	0.921
2003/04	0.981	0.927	0.978	0.931
2004/05	0.982	0.932	0.978	0.916
2005/06	0.982	0.929	0.979	0.926
2006/07	0.980	0.917	0.978	0.917
2007/08	0.980	0.906	0.977	0.917
2008/09	0.978	0.915	0.974	0.904
2009/10	0.980	0.920	0.977	0.922
2010/11	0.980	0.923	0.976	0.924
All $p<0.01$				

Figure 7.6 shows the estimated rate compared with the PRDS/CHI/Health Card derived rate for all LADs in every year (2001/02 to 2010/11). Some outliers do exist, namely Richmondshire, the City of London and the Isles of Scilly. The last two of these exhibit very small numbers in the migration matrix, while Richmondshire is an area with a large armed forces population. These areas are discussed further in Section 7.5.5. It is important to note that the estimates by age created using the methodology in this chapter are constrained to the total migration found in the aggregate matrix, so these outliers will be dealt with in more detail when the age profile is assessed in the following chapter.

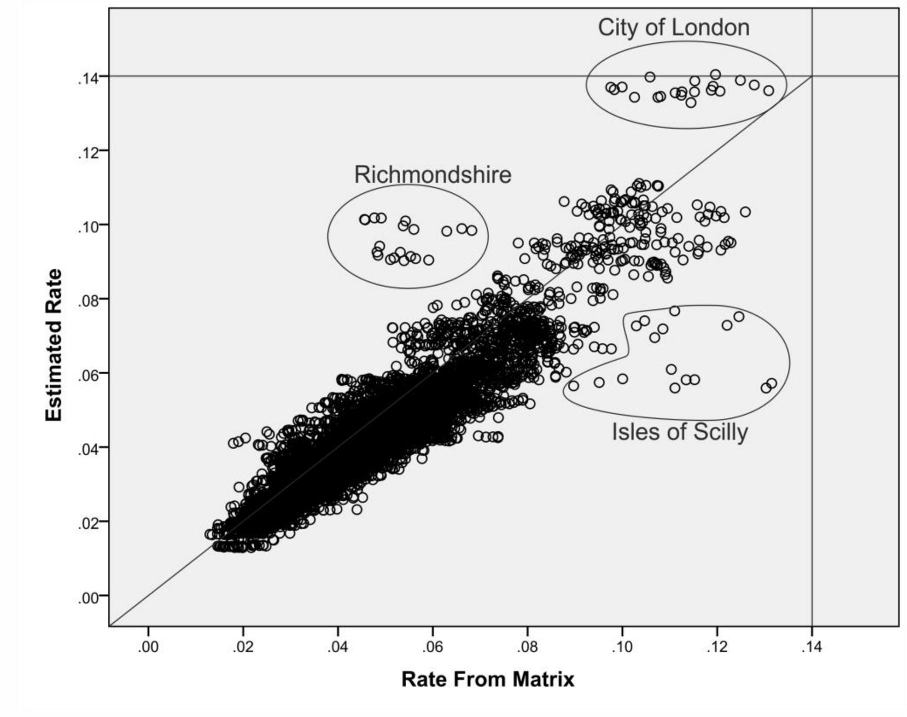


Figure 7.6: A comparison of the estimated migration rate and NHS register reported rate for all years 2001/02 to 2010/11, for all LADs

The second validation is to compare the estimated migration with PRDS data for England and Wales in 2010/11, where age and sex information is available. The estimated in-migration and out-migration for all LADs by each age and sex group is compared with the PRDS data, and Table 7.3 shows the correlation between the two datasets. Generally, the correlation is strong but does fall slightly at older ages, especially for male in-migration estimates (for example, r=0.674 for 80-84 year old males).

By finding a strong positive correlation between the total migration estimated using the 2001 Census migration rate and the figure reported in the registration data, one can be confident that the general principle of the method is robust. The strong positive correlation between the estimates and PRDS data for 2011 in England and Wales provide further confidence that the method is robust when extended to disaggregated sex and age groups, but the correlation falls for older age groups (70+) and more so for males than females.

Table 7.3: The correlation between the estimated migration totals and the PRDS reported totals for England and Wales in 2010/11

Age	Male In	Male Out	Female In	Female Out
0-4	0.995	0.995	0.997	0.997
5-9	0.975	0.989	0.975	0.985
10-14	0.908	0.946	0.918	0.949
15-19	0.969	0.927	0.981	0.939
20-24	0.977	0.973	0.980	0.981
25-29	0.981	0.960	0.991	0.991
30-34	0.987	0.987	0.988	0.989
35-39	0.974	0.984	0.950	0.970
40-44	0.947	0.972	0.936	0.966
45-49	0.925	0.954	0.902	0.931
50-54	0.905	0.939	0.902	0.902
55-59	0.871	0.880	0.880	0.900
60-64	0.885	0.874	0.874	0.885
65-69	0.873	0.849	0.849	0.881
70-74	0.789	0.820	0.820	0.819
75-79	0.762	0.778	0.778	0.797
80-84	0.674	0.788	0.788	0.809
85-89	0.681	0.784	0.787	0.802
90+	0.720	0.804	0.800	0.809

All $p<0.01$

7.4 Balancing parsimony and accuracy – classifying LADs by age profile

Given the aim of this chapter to estimate migration for 19 age groups and two sexes for 406 origins and 406 destinations across ten years, the volume of data which are produced quickly becomes a daunting prospect for analysis. Where just the OAS and DAS margins by five year of age are concerned, 30,856 separate observations will be produced each year. When this is extended to the full ODAS matrix this becomes 6,263,768 individual cell counts. Interpreting these results will require a framework for analysis, and clustering of LADs based on their age and sex profile provides a neat solution through which areas can be more easily compared. Once the LADs are clustered together, a modelled migration schedule will provide an idea of the ‘type’ of migration occurring for that cluster. This section looks at modelling migration schedules from the consistent set of OAS and DAS margins using K-means cluster analysis and modelling the migration schedules for these clusters using variants of the mathematical function proposed by Rogers *et. al.* (1978).

In a similar context to the work presented in this thesis, Bracken and Bates (1983) produce a model of migration flows for 116 local authorities in England and Wales and argue that *“the variation in migration propensity by age is on the whole remarkably constant from one area to another, and a considerable simplification can be made in the model if we can group areas according to the shape of their profiles”.* They find that *“in almost all cases, the female migration profiles are very similar to those for males, except that the female profile is generally in advance of the male profile by about two years”* (p. 252). Where areas were difficult to allocate, the authors relied on subjective judgement based on the consideration of geographical location and the level of urbanisation. When single year of age was modelled, they produced twelve clusters of local authorities. In an extension to their 1983 work, Bates and Bracken (1987) model changes between the 1971 to 1981 Census migration patterns. They note that whilst the change in rates and model profile parameters are small, *“they are of considerable policy significance when translated into quantities of persons”* (p.531). They argue that the modelling of migration curves and clustering of areas is useful in *“reducing the burden of information in migration analysis”* (p.531). Using cluster analysis and modelled age curves could potentially provide an even more simple solution than the use of census rates for each LAD presented above, but certainly allows for useful comparison of areas in a large migration system such as the one presented in this thesis.

7.4.1 K-means clustering

K-means clustering is referred to variously as an iterative partitioning method (Aldenerfer and Blashfield 1984) and an optimization algorithm (Everitt *et al.* 2001) which is used to combine *n* objects into *g* groups. An initial partition is found and then optimized by moving each object from its own to another group. This process is repeated until no move causes an improvement to the cluster criterion.

7.4.2 Implementation of K-means clustering

Before arriving at a final clustering solution, it is necessary to consider a number of variables, namely the software used for implementation of the clustering, the optimum number of clusters and the type of distance measure used to create these clusters. To support these decisions, Figures 7.7 and 7.8 show the exploration of clusters for in-migration and out-migration disaggregated only by age, Figure 7.9 shows the clustering for males by age and Figure 7.7 shows the clustering for females by age. Silhouette values are produced for each cluster solution in order to aid in the assessment of its validity, which assigns a value to each case and is set out by Martinez *et al.* (2011, p.148):

> *"If an observation has a value close to 1, then the data point is closer to its own cluster than a neighbouring one. If it has a silhouette width close to –1, then it is not very well-clustered. A silhouette width close to zero indicates that the observation could just as well belong to its current cluster or one that is near to it."*

Silhouette values are used by a number of authors to assess the clustering of their data (see Dennett 2010; Kaufman and Rousseeuw 2005; Shepherd 2006). From these silhouette values, an average cluster solution can be derived (the closer to 1, the better the overall solution) and the number of cases with a value less than zero, which are likely allocated to the incorrect cluster. Both of these statistics are reported for each silhouette plot.

7.4.3 Software

The way that K-means clustering is implemented differs across a number of software packages, so the selection of a program is an important element of the classification process. The main issue lies in the allocation of the initial seed values, as K-means clustering uses a seed value as a starting point for allocation of cases to clusters. Steinley (2003) investigates the operationalization of K-means clustering in three commercial software packages, SPSS, SYSTAT and SAS. He concludes that the lack of an option to re-order cases and re-run the program multiple times results in a "local optimum" solution that does not necessarily represent the best solution overall. Steinley advocates the use of a purpose built MATLAB program which replicates the clustering numerous times with different case ordering so as to provide a "global optimum" solution.

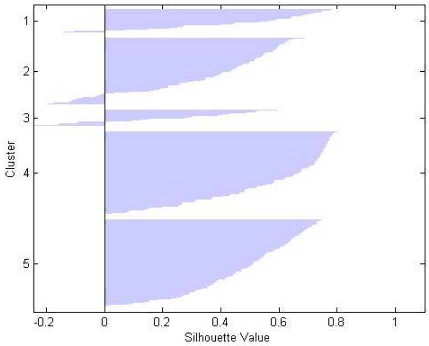
Dennett (2010) carries out a substantial assessment of K-means clustering and draws the same conclusion; he discounts SPSS in favour of MATLAB, which allows for specification in the algorithm for replication of the clustering analysis, allowing one to move from a local optimum to a global optimum solution. Certainly the rather vague guidance supplied in the SPSS user manual that *"the default algorithm for choosing initial cluster centres is not invariant to case ordering"* and the recommendation to *"obtain several different solutions with cases sorted in different random orders to verify the stability of a given solution"* does not suggest that a solution would be suitably robust when using the software. Testing of the data in this thesis using SPSS and manually reordering cases on several runs does indeed produce different clustering results. It is therefore clear that a solution using a purpose written algorithm for MATLAB is preferable to using the off the shelf solutions offered by other software packages.

Figures 7.7 to 7.10 shows the clustering of LADs where the algorithm has been re-run 250 times with random case ordering and seed allocation. The final solutions presented in Figures 7.11, 7.13 and 7.15 re-run the K-means clustering 1,000 times with no change in the allocation.

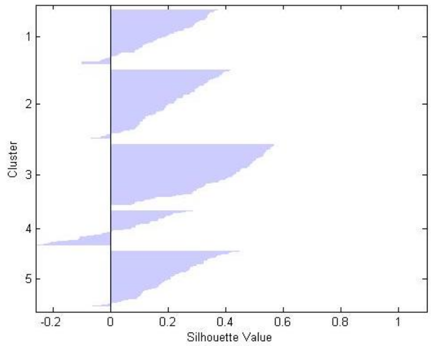
7.4.4 Distance measure

The second consideration is the distance measure used to specify the allocation of cases to cluster seed values. Within the MATLAB program, five distance measures can be specified in the algorithm: Squared Euclidean, City Block (Manhattan), Cosine, Correlation and Hamming measures. Only the first two are relevant to the data presented in this chapter; Squared Euclidean distance allocates each object to the seed with the smallest squared (straight line) difference while Manhattan distance can be visualised as a city block where the calculation is the shortest distance between object and seed on a grid. Aggarwal *et al.* (2001) argue that the choice of distance metric is not obvious and that literature providing guidance on choosing the correct distance measure is sparse. Having said this, both Aggarwal *et al.* (2001) and Dennett (2010) advocate the use of Manhattan distance, finding that it produces better clustering solutions for their data while Loohach and Garg (2012) advocate Manhattan for efficiency reasons. Bracken and Bates (1983) use Squared Euclidean distance to cluster gross migration profiles for England and Wales.

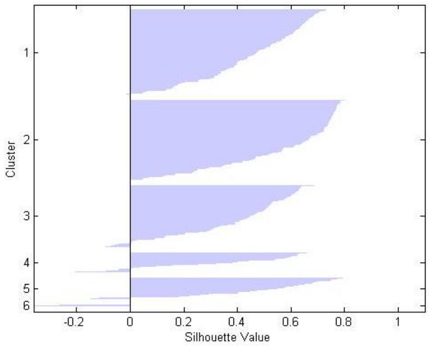
Figures 7.7 and 7.8 show that for the data presented in this thesis, Squared Euclidian distance consistently outperforms Manhattan distance for all cluster solutions. Given that there is no set guidance for the choice of a distance measure and the results seen in Figures 7.7 and 7.8, Squared Euclidian distance have been chosen for the final clustering solutions. The y axis of the silhouette plots shows each cluster, and each bar in the plot represents the LADs associated with that cluster. So the wider the bar at each cluster on the x axis, the more LADs within that cluster. The y axis provides the silhouette value for each LAD within the cluster. The average silhouette value is reported below each plot (higher numbers represent a better fit) along with the number of cases (LADs) with a value of less than one.



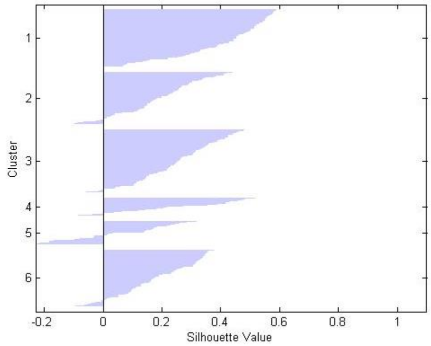
Euclidean 5 Cluster
Average = 0.463, Cases <0 =25



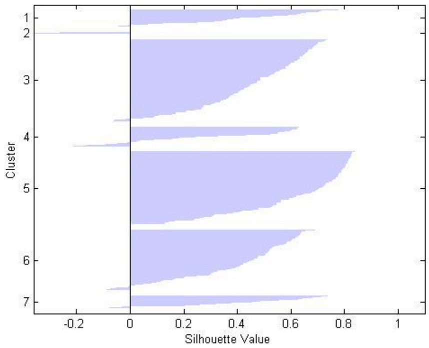
Manhattan 5 Cluster
Average = 0.223, Cases <0 = 40



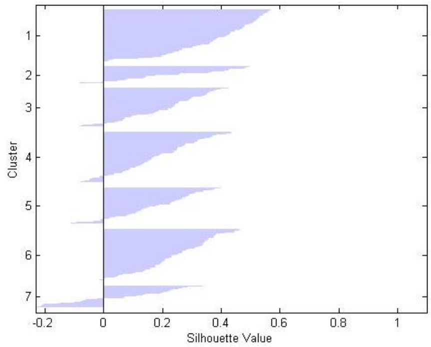
Euclidean 6 Cluster
Average = 0.465, Cases <0 = 18



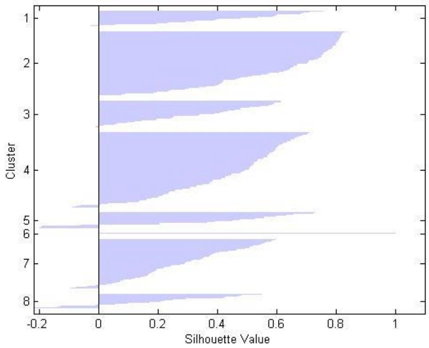
Manhattan 6 Cluster
Average = 0.239, Cases <0 = 38



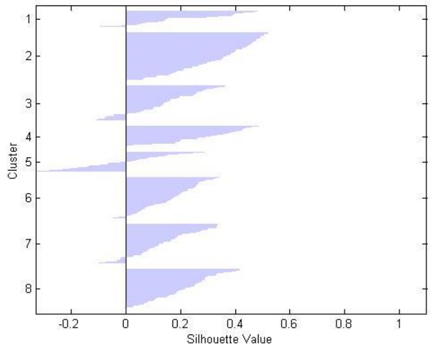
Euclidean 7 Cluster
Average = 0.468, Cases <0 = 25



Manhattan 7 Cluster
Average = 0.229, Cases <0 = 40

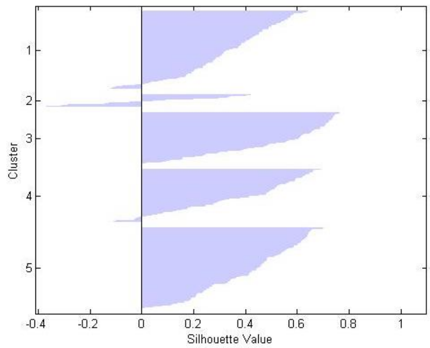


Euclidean 8 Cluster
Average = 0.425, Cases <0 = 22

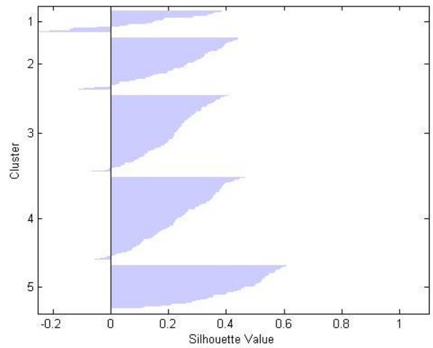


Manhattan 8 Cluster
Average = 0.200, Cases <0 = 37

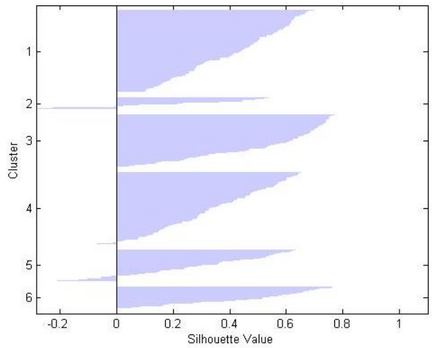
Figure 7.7: Silhouette values for inflow clusters – Euclidean vs Manhattan distance measures



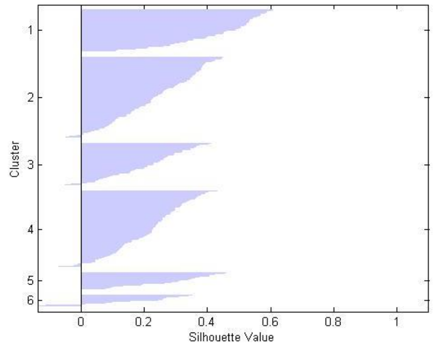
Euclidean 5 Cluster
Average = 0.383, Cases <0 = 23



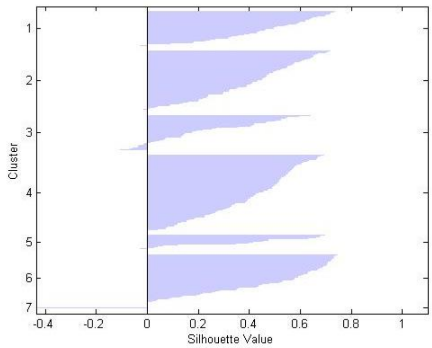
Manhattan 5 Cluster
Average = 0.250, Cases <0 = 24



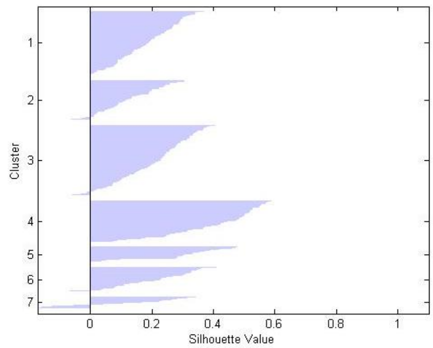
Euclidean 6 Cluster
Average = 0.398, Cases <0 = 15



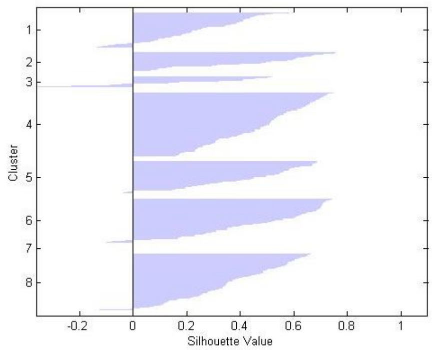
Manhattan 6 Cluster
Average = 0.256, Cases <0 = 16



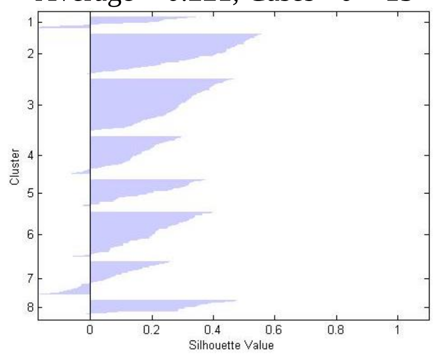
Euclidean 7 Cluster
Average = 0.402, Cases <0 = 17



Manhattan 7 Cluster
Average = 0.221, Cases <0 = 15

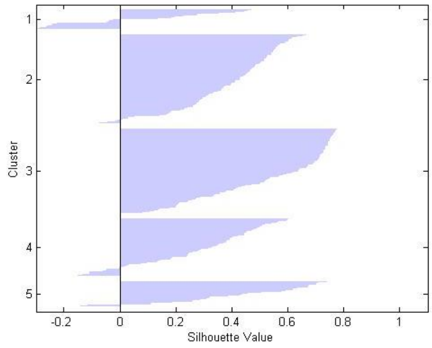


Euclidean 8 Cluster
Average = 0.398, Cases <0 = 18

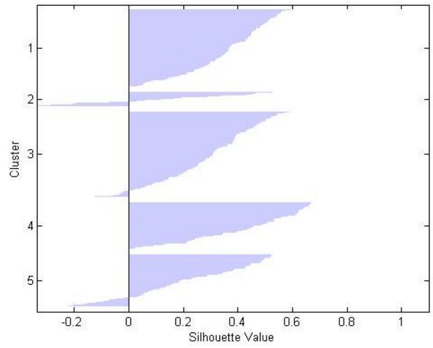


Manhattan 8 Cluster
Average = 0.215, Cases <0 = 37

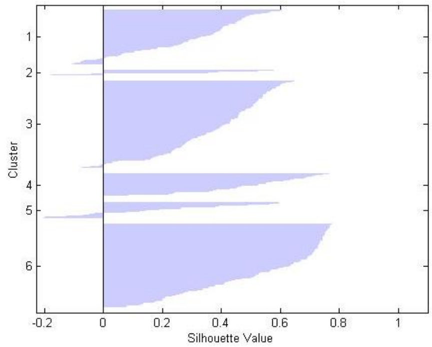
Figure 7.8: Silhouette values for outflow clusters – Euclidean vs Manhattan distance measures



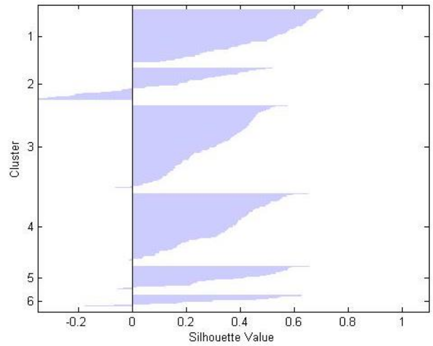
Males 5 Cluster In
Average = 0.394, Cases <0 = 34



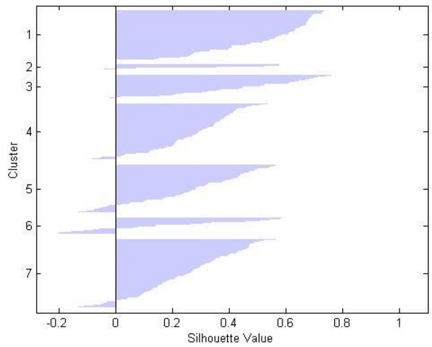
Males 5 Cluster Out
Average = 0.310, Cases <0 = 29



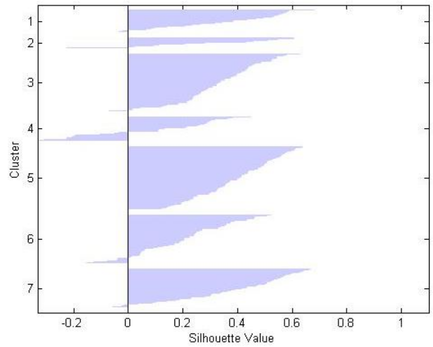
Males 6 Cluster In
Average = 0.406, Cases <0 = 26



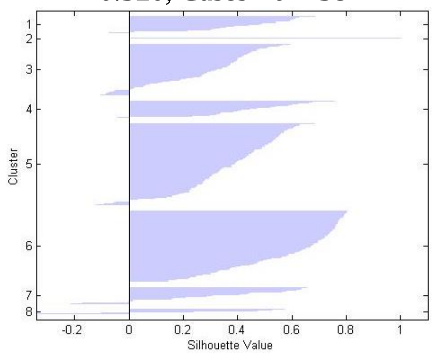
Males 6 Cluster Out
Average = 0.324, Cases <0 = 28



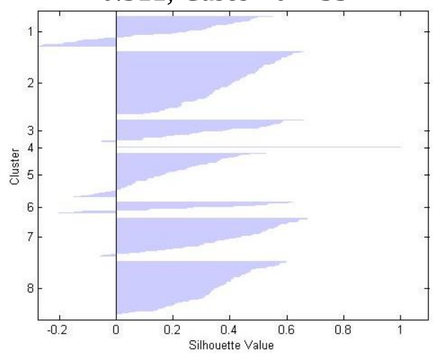
Males 7 Cluster In Average = 0.320, Cases <0 = 38



Males 7 Cluster Out Average = 0.311, Cases <0 = 33

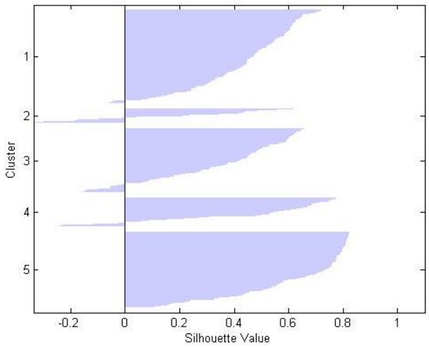


Males 8 Cluster In
Average = 0.414, Cases <0 = 34

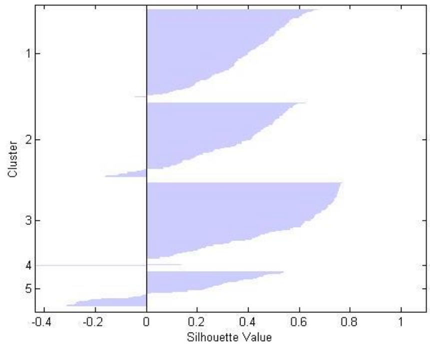


Males 8 Cluster Out
Average = 0.318, Cases <0 = 25

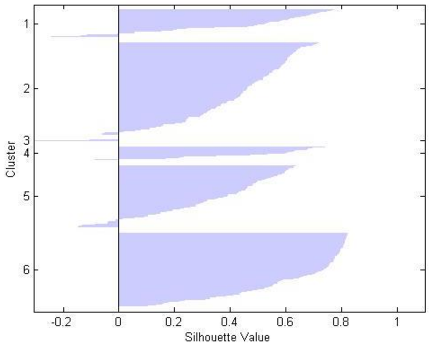
Figure 7.9: Silhouette values for inflow and outflow clusters – Male



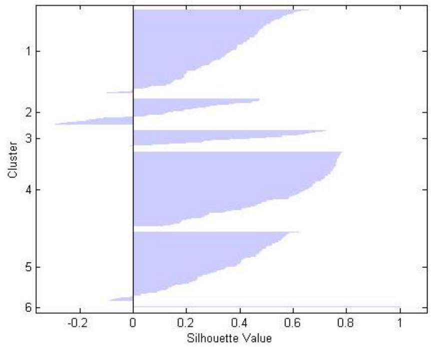
Females 5 Cluster In
Average = 0.455, Cases <0 = 30



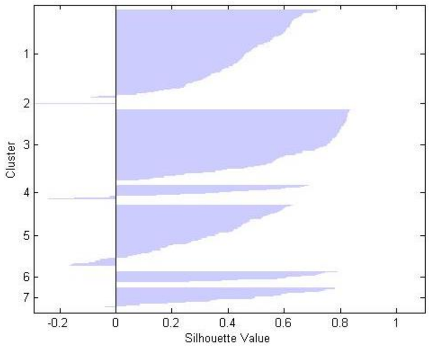
Females 5 Cluster Out
Average = 0.381, Cases <0 = 32



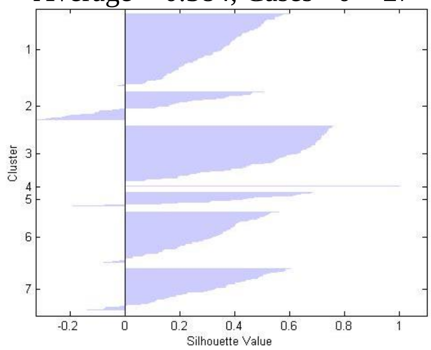
Females 6 Cluster In
Average = 0.461, Cases <0 = 24



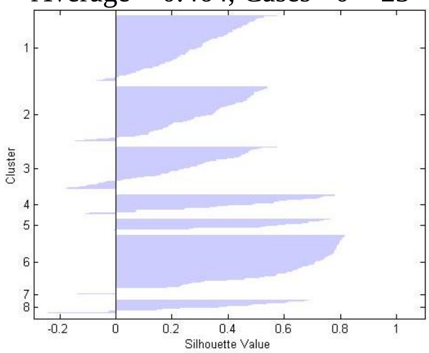
Females 6 Cluster Out
Average = 0.384, Cases <0 = 27



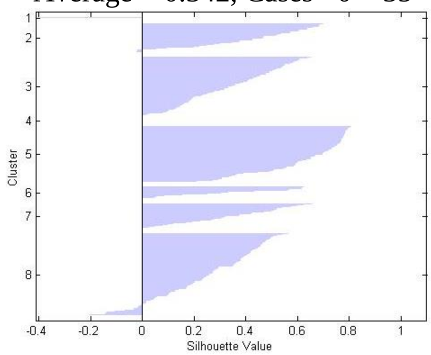
Females 7 Cluster In
Average = 0.464, Cases <0 = 23



Females 7 Cluster Out
Average = 0.342, Cases <0 = 35



Females 8 Cluster In
Average = 0.368, Cases <0 = 34



Females 8 Cluster Out
Average = 0.377, Cases <0 = 30

Figure 7.10: Silhouette values for inflow and outflow clusters – Female

7.4.5 Assessing the number of clusters – age information only

The number of clusters to use for each subset of data can be derived from the silhouette graphs above. For inflow and outflow (with no sex disaggregation) a seven cluster solution provides the best result and can be seen in Figure 7.11).

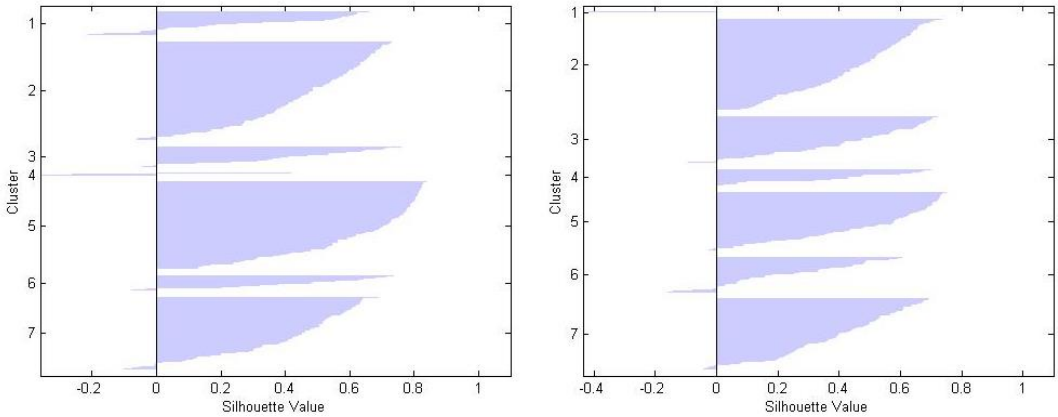


Figure 7.11: Final inflow cluster solution (left graph) and outflow cluster solution (right graph) after 1,000 random re-runs

Both the in and out clustering solutions contain one very small cluster which contains LADs with very specific populations. Inflow cluster 4 contains City of London and Isles of Scilly (both with very sparse data due to small flows) and Richmondshire (which has a disproportionately high young migration peak due to a large number of armed forces personnel and their families). Outflow cluster 1 contains only City of London and Oadby and Wigston (which has a very large student population in halls of residence which are affiliated with the University of Leicester). Apart from these anomalies, the two solutions appear to work well, the average silhouette values are 0.47 and 0.40 respectively, with 17 and 21 cases with a value less than zero.

Figure 7.12 shows the cluster membership of LADs for both inflow and outflow. Overall, many cluster memberships are similar (although it is important to note that the calibration to find a global optimum solution means that the clusters are allocated randomly so inflow cluster 1 does not match outflow cluster 1 and so forth). These memberships are referred to in the following chapter where results of the estimation are presented.

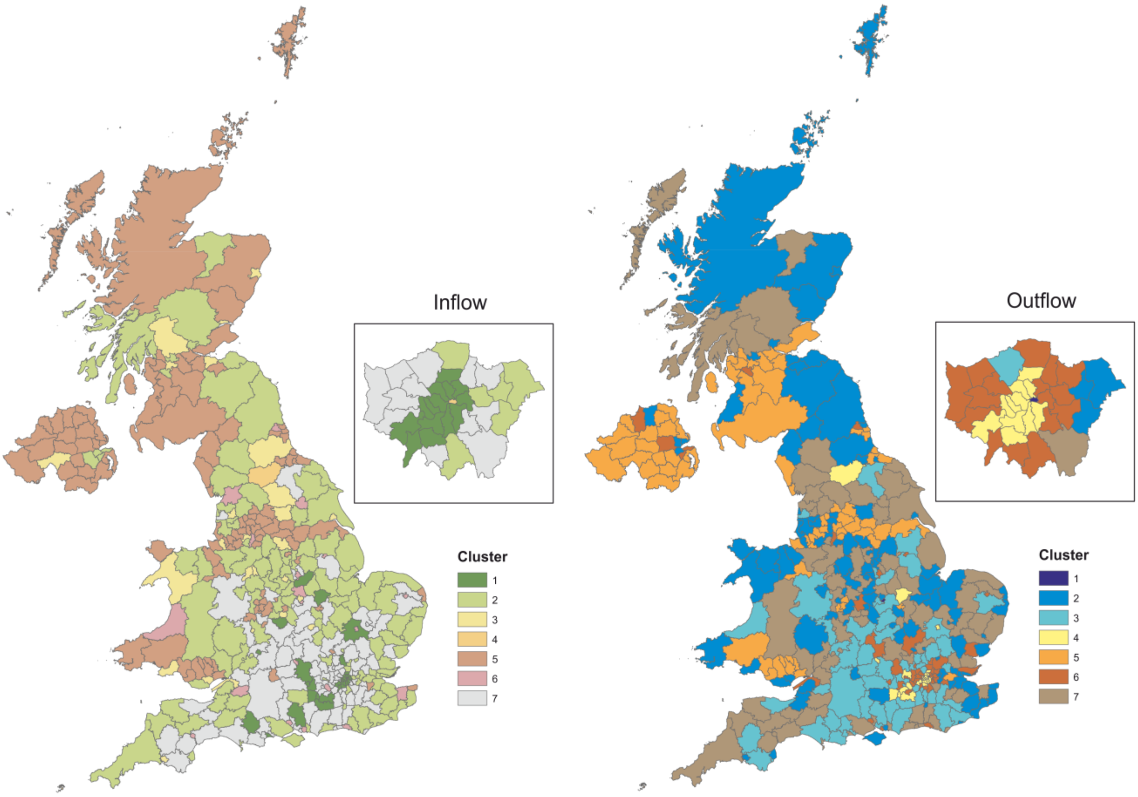


Figure 7.12: Cluster membership for inflow and outflow

7.4.6 Assessing the number of clusters – males

Figure 7.13 shows that a six cluster solution for both inflow and outflow provided the best solution for male migration. the average silhouette values are 0.4 and 0.32 with 26 and 28 cases with a value under zero for inflow and outflow respectively. This suggests that inflow cluster solution is slightly better than the outflow solution.

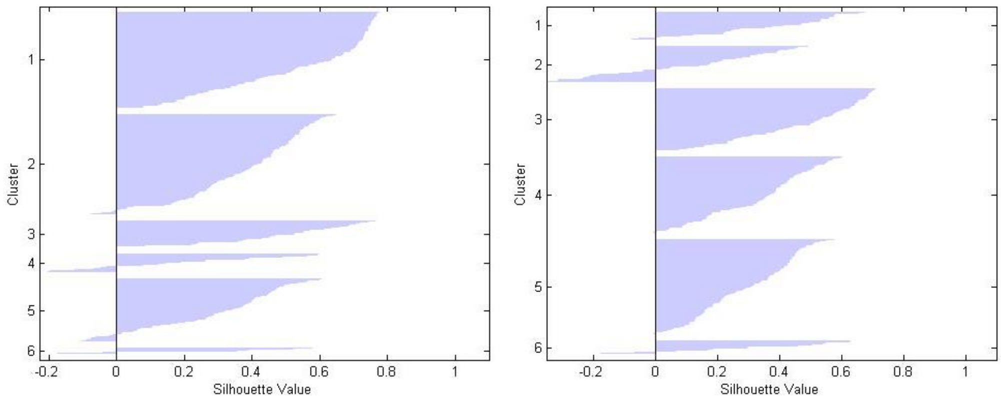


Figure 7.13: Final inflow cluster solution and outflow cluster solution for males (after 1,000 random re-runs)

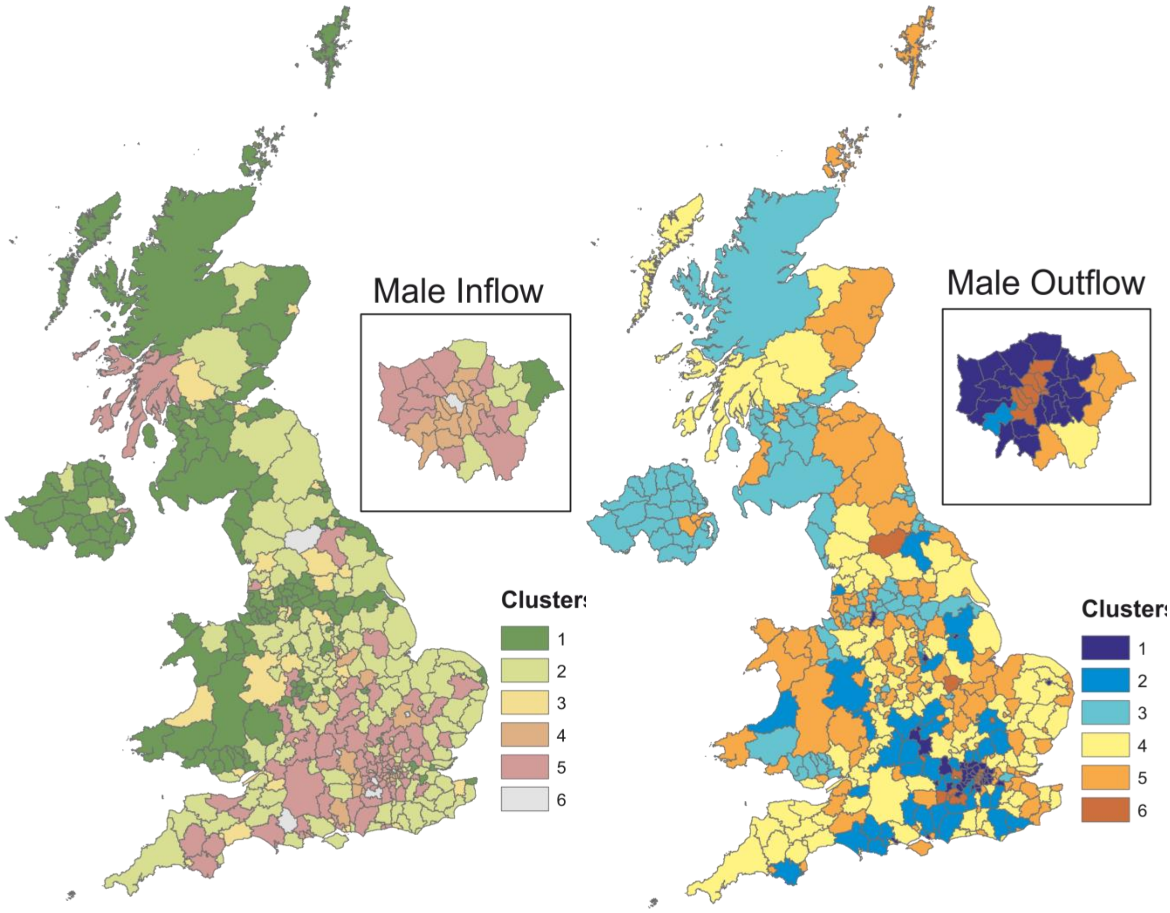


Figure 7.14: Cluster membership for inflow and outflow – males

7.4.7 Assessing the number of clusters – females

Figure 7.15 shows that a seven cluster solution for inflow and six cluster solution for outflow are best for female migration. The average silhouette values are 0.46 and 0.38 with 24 and 27 cases with a value under zero for inflow and outflow respectively. The solution for females is slightly better than that for males.

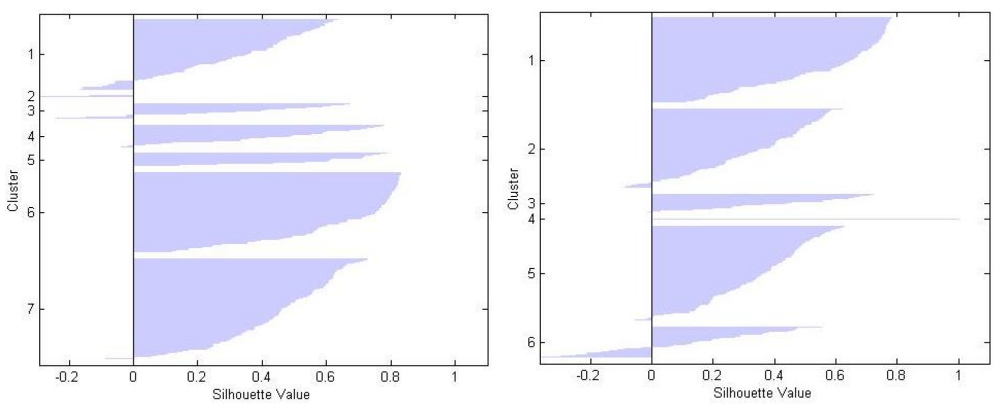


Figure 7.15: Final Inflow cluster solution and outflow cluster solution for females (after 1000 random re-runs)

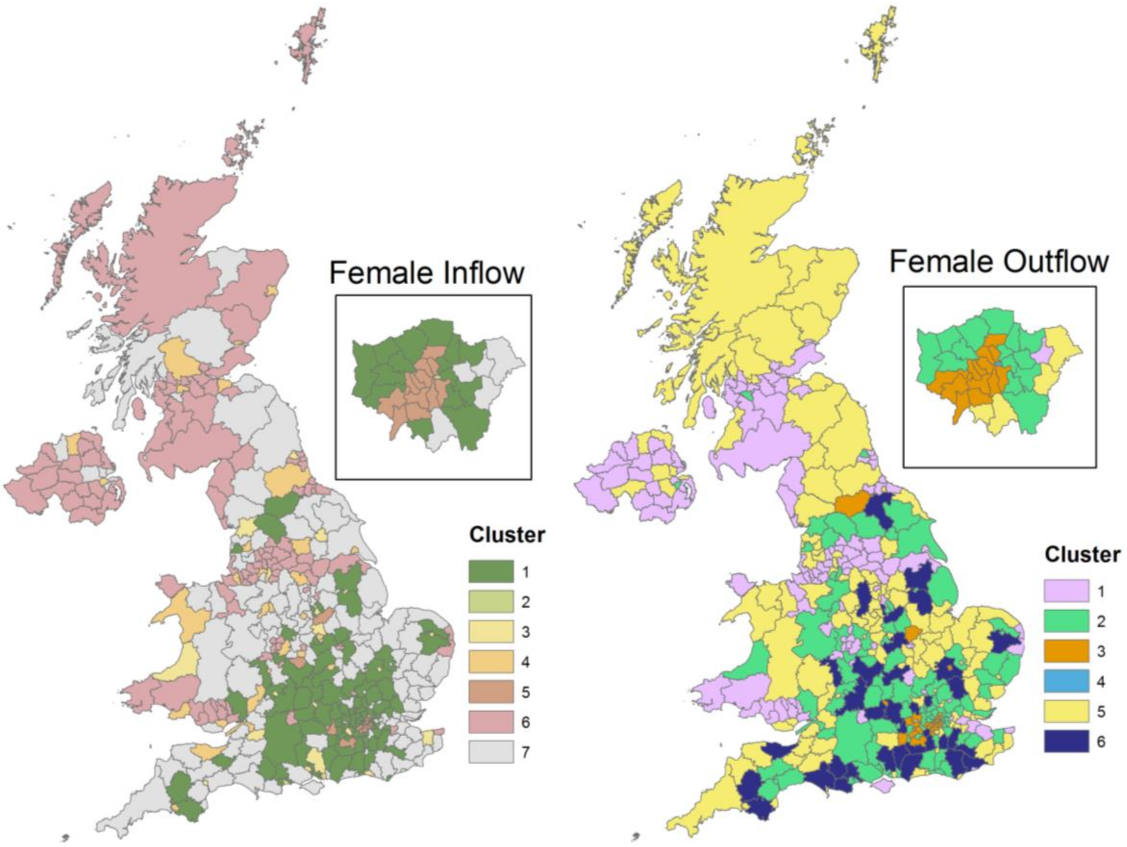


Figure 7.16: Cluster membership for inflow and outflow – females

7.4.8 Comparing cluster memberships

Cross-tabulation of cluster membership can provide a useful tool for understanding the different patterns by sex. By looking for cluster cross-tabulations where over 70 per cent of LADs fall in the same group, the male and female solutions can be compared to the cluster solution where all migrants are included. For inflow, 77 per cent of all LADs are allocated to the same cluster when males and all migrants are compared, and 84 per cent of all LADs have the same allocation for females and all migrants. For male outflow, 77 per cent of LADs are allocated to the same cluster as the aggregate grouping. Female outflow has a far lower association with aggregate outflow, which can be seen in Figure 7.17.

Given the inconsistencies seen here and the requirement set out at the beginning of this section to produce a framework that makes analysis of the large ODAS dataset more simple, the aggregate cluster solutions appear to perform better than the sex disaggregated solutions.

			Female 6 cluster outflow						Total
			1	2	3	4	5	6	
All people 7 cluster outflow	1	Count	0	0	0	1	0	1	2
		% within all people cluster	.0%	.0%	.0%	50.0%	.0%	50.0%	100.0%
		% within female cluster	.0%	.0%	.0%	100.0%	.0%	2.5%	.5%
	2	Count	40	0	0	0	75	0	115
		% within all people cluster	34.8%	.0%	.0%	.0%	65.2%	.0%	100.0%
		% within female cluster	35.7%	.0%	.0%	.0%	60.0%	.0%	28.3%
	3	Count	0	20	2	0	0	37	59
		% within all people cluster	.0%	33.9%	3.4%	.0%	.0%	62.7%	100.0%
		% within female cluster	.0%	19.2%	8.3%	.0%	.0%	92.5%	14.5%
	4	Count	0	1	20	0	0	0	21
		% within all people cluster	.0%	4.8%	95.2%	.0%	.0%	.0%	100.0%
		% within female cluster	.0%	1.0%	83.3%	.0%	.0%	.0%	5.2%
	5	Count	71	0	0	0	3	0	74
		% within all people cluster	95.9%	.0%	.0%	.0%	4.1%	.0%	100.0%
		% within female cluster	63.4%	.0%	.0%	.0%	2.4%	.0%	18.2%
	6	Count	1	35	2	0	7	0	45
		% within all people cluster	2.2%	77.8%	4.4%	.0%	15.6%	.0%	100.0%
		% within female cluster	.9%	33.7%	8.3%	.0%	5.6%	.0%	11.1%
	7	Count	0	48	0	0	40	2	90
		% within all people cluster	.0%	53.3%	.0%	.0%	44.4%	2.2%	100.0%
		% within female cluster	.0%	46.2%	.0%	.0%	32.0%	5.0%	22.2%
Total		Count	112	104	24	1	125	40	406
		% within all people cluster	27.6%	25.6%	5.9%	.2%	30.8%	9.9%	100.0%
		% within female cluster	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

Figure 7.17: Cross-tabulation of cluster membership for aggregate outflow and female outflow

7.5 Modelling migration schedules

Modelling migration schedules gives the opportunity to compare the clusters seen in the previous section based on the type of migration profile that they exhibit. The mathematical function, known as the multiexponential function, used to create migration curves was developed by Rogers (1975) and enhanced by Rogers and Castro (1981) and Rogers *et al.* (1978). These curves are a function of five components (childhood, employment, retirement, old age and a constant) while Wilson (2010) develops the model to include a sixth component, a student migration parameter to account for increases in migration intensities for the late teenage years. Rogers (2007, p. 378) provides an overview of a number of studies that utilise the multiexponential function, given its *"remarkably good fit to a wide variety of empirical interregional* *migration schedules"* and a detailed account of the application of this function can be found in (Norman *et al.* 2012). Taken from Wilson (2010), the model has the form:

$$\hat{m}(x) = a_1 \exp(-a_1 x) + a_2 \exp\{-\alpha_2(x-\mu_2) - \exp[-\lambda_2(x-\mu_2)]\} + a_3 \exp\{-\alpha_3(x-\mu_3) - \exp[-\lambda_3(x-\mu_3)]\} + a_4 \exp(\alpha_4 x) + c \tag{7.5}$$

where:

$\hat{m}$ is the modelled migration intensity, x is age, a_1 to a_4 is the height of each of the component curves (childhood to old age), λ is the rate of ascent, α is the rate of descent for each curve and c is the constant. In Wilson (2010), the student component fits into Equation 4, but is not used in this thesis.

The result of applying this curve fitting equation to each of the seven inflow and outflow clusters can be seen in Figures 7.18-7.20. Figure 7.18 shows the in-migration curves, Figure 7.19 shows the out-migration curves and Figure 7.20 shows the curves for inflow cluster 4 and outflow cluster 1 which contain the small number of LADs that do not fit in elsewhere.

The modelled curves in Figures 7.18 to 7.20 have been estimated for single year of age using the methodology specified in (Norman *et al.* 2012) and represent the best fit based on the number of components used (plus the constant). This single year of age information adds a dimension to the data that was not previously available and would allow for more detailed analysis of the migration patterns by age in the future. Reflecting the different features of the migration curves by area and age for inflow, a four component model has been used for clusters one, two, three and five, while a three component model was used for clusters six and seven. A four component model was used for outflow clusters two, three, five and seven while a three component model was used for clusters four and six.

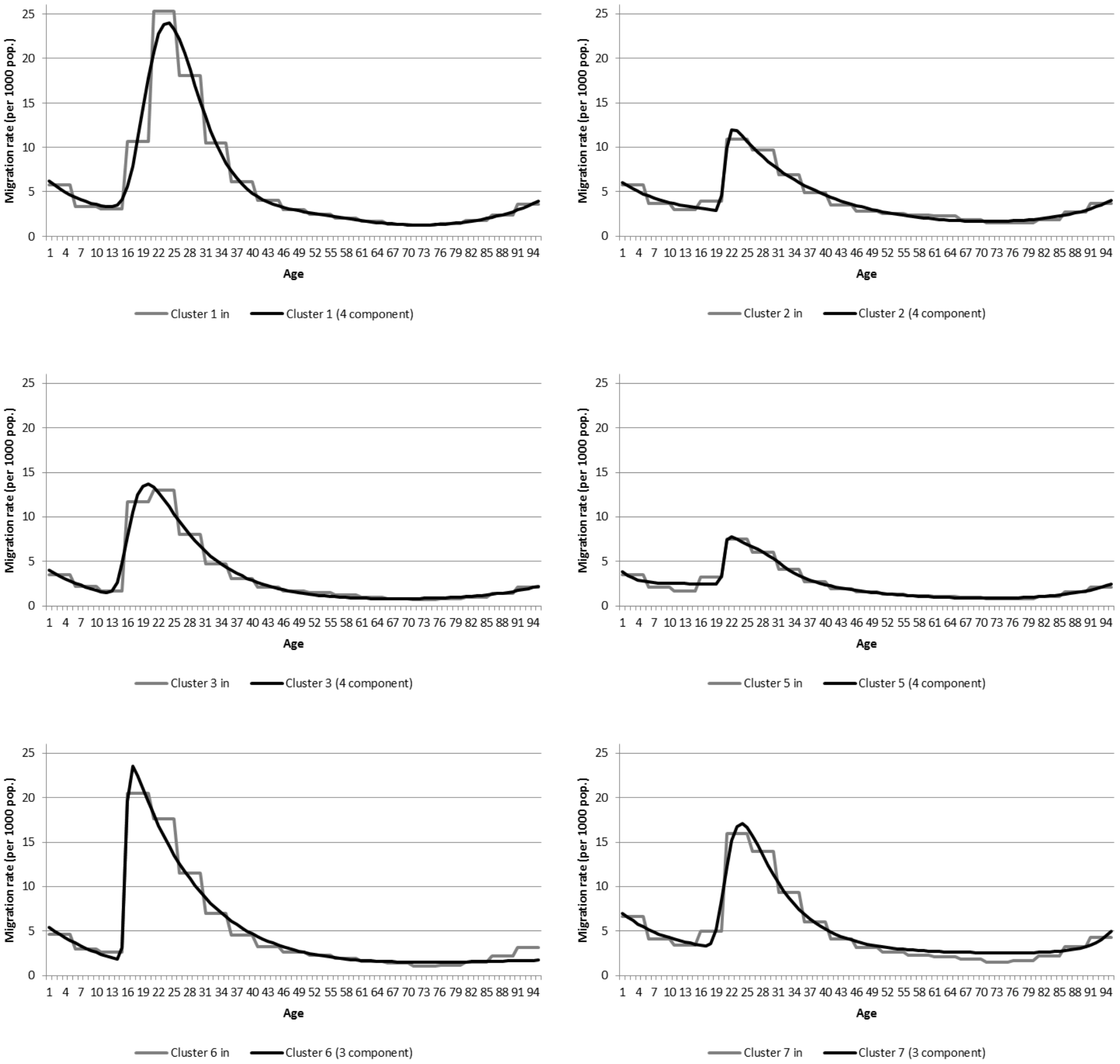


Figure 7.18: Modelled curves for in migration (black line)

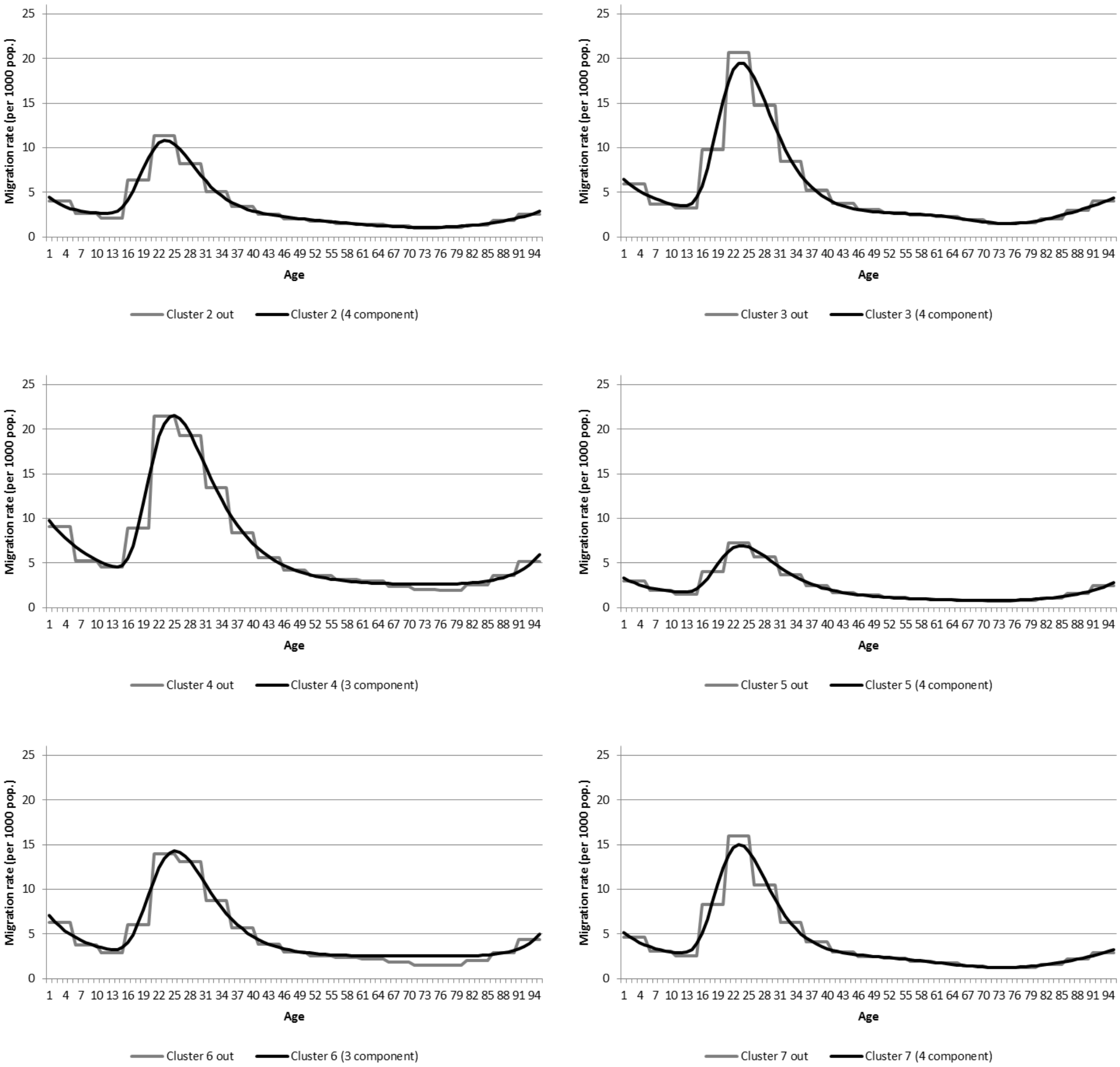


Figure 7.19: Modelled curves for out migration (black line)

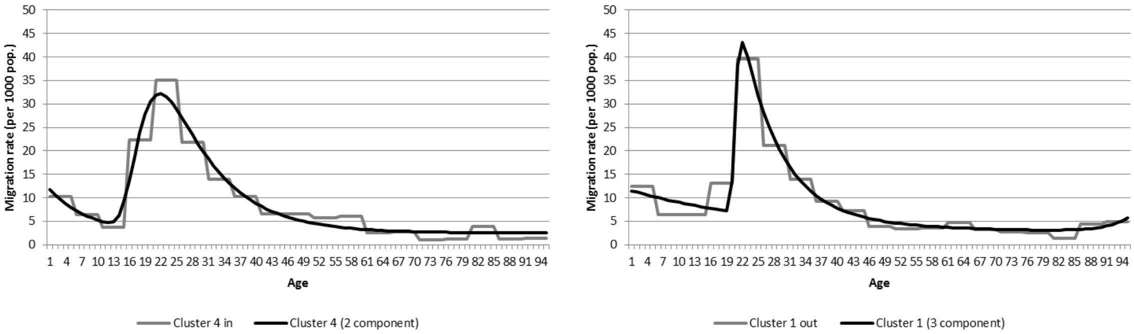


Figure 7.20: Modelled curves for the small inflow and outflow cluster

7.6 Conclusion

This chapter has presented the methodology used to disaggregate the migration database by age and sex. A similar IPF routine to that implemented for the aggregate database is used for migration estimates by ODAS. The challenge has been in finding a solution to produce consistent and robust margins by age and sex (OAS and DAS) from which to adjust the census seed table. The literature presented in this chapter suggests that the underlying structure of migration – in terms of age and sex – is relatively stable over time (or at least slow to change), which has been a factor in the decision to use 2001 Census age rates and the sex distribution which is applied to populations at risk in each year to create the marginal estimates. The other contributing factor in choosing a single methodology that can be applied across all LADs in the UK is the inconsistency in the data sources that are available between 2001/02 to 2010/11 and questions over the quality of disaggregated age/sex information derived from registration data sources.

The second substantial element of this chapter presented a framework for analysing the huge quantity of data that has been produced in the estimation of an ODAS matrix for 406 LADs across ten years. K-means classification has been used to cluster LADs based on their age profile, and in the interest of parsimony it was decided to incorporate both sexes concurrently into these clusters. A multiexponential curve fitting function was used to model the shape of the migration curve for each of the seven inflow and outflow clusters in order to compare these clusters based on the type of migration by age that they exhibit. This curve fitting methodology could also be applied to the estimated ODAS matrix to produce estimates by single year of age, should a finer level of detail be required for future analysis.

The results of the estimated ODAS matrix will be presented in the next chapter, with the seven inflow and outflow clusters of LADs used to aid in the interpretation of the results where appropriate.

Chapter 8

Analysing patterns and trends of migration across the life course

Until now the focus of this thesis has been, in the main, on aggregate migration patterns and trends which mask the variations that occur at different ages (and to a lesser extent by sex). Previous chapters have effectively built up to this point, where the focus shifts to analysing the form and dynamics of the redistribution of population in different ages groups and for males and females, utilising the most detailed matrices of migration flow estimates. The importance of disaggregating the findings of previous chapters is encapsulated in the following paragraph from Plane (1993, p.381): *"When age-aggregate migration flows are disaggregated it is very often the case that the migration system is found to be made up of rather different spatial sets of flows. Whereas there tends to be some correlation across age groups in the migration response to cyclical and longer term economic factors, the movement streams of special populations such as retirees, college students, military personnel and so forth are often quite substantial".*

Whilst estimates of migration by five year age group and sex have been produced using the methodology detailed in the previous chapter, the analysis in this chapter focuses on a broader 'life course stage' classification of ages. The reasons for grouping ages in this way are twofold: first, there is a data presentation issue; there simply is not enough space to discuss patterns of migration by five year age group and sex. The second reason is that migration propensities and patterns are distinctly different at certain key stages in a person's life course which *"is composed of a series of transitions or life events, which are embedded in trajectories or careers (or status passages) that give them a distinct form or meaning"* (Kulu and Milewski 2008, p.568), so to look at each five year bands would be needlessly repetitive in many cases. The analysis carried out in this chapter uses the following eight broad age groups which correspond with stages in the life course:

- 0-14 years – childhood
- 15-19 years – leaving home for education (or first job)
- 20-24 years – leaving education for work
- 20-29 years – early career
- 30-44 years – family formation

- 45-59 years – older working age (or early retirement)
- 60-74 years –retirement (and younger old age)
- 75 and older – older old age

These life course stages loosely follow the bands chosen by Fotheringham *et al.* (2004), Champion *et al.* (2003) and Rees *et al.* (2004a) who defined seven key ages: preschool/school age (0-15), leaving home for university or work (16-19), leaving university for work (20-24), forming couples and stating a family (25-29), raising a family (30-44), older working age (45-59) and approaching and beyond retirement (60+) with the addition of the 60-74 and 75 plus age bands.

The patterns will be analysed and explained for each broad age group using techniques employed in earlier chapters of this thesis: first, a city region framework (as adopted in Chapter 6) will be used to examine flows between areas and address changing trends over the decade; second, key migration indicators (as used in Chapter 5) will be utilised to understand the general patterns of migration seen across the UK at each broad age group; third, net migration rates (as employed in Chapter 5) will be used to identify patterns at the LAD scale and highlight areas that exhibit loss or gain – here the clustering of LADs by age (explained in Chapter 7) will be used to further understand the patterns; and finally, sex ratios will be used to identify any LADs that do not conform to the average UK distribution within each age group. In addition, where other data and information are available – such as data on the movement of higher education students and concentrations of armed forces personnel – they will be used to better understand the patterns observed. Literature pertaining to the determinants of migration across the life course will also be considered in each section.

The arguments for not reporting all findings broken down by sex are similar to those for not reporting all five year age bands; the first is the availability of space in this chapter and the second (repetition of findings) is summarised by Stillwell *et al.* (1996, p.297), who find that although there are small differences between males and females for certain life course events (women's labour force peak is at ages 22 and 23 compared with males 23 and 24 and there is some difference in early adult life and later life, owing to the age of marriage and retirement for example) *"the differences between the gender curves make little difference to overall measures"*. Therefore, identifying anomalies by analysing sex ratios provides a neat solution for analysis which allows a focus on areas that exhibit unexpected patterns.

8.1 UK internal migration trends using the city region typology

Figure 8.1 shows the proportion of total flow which can be attributed to moves between each part of the city region typology in 2001/02 and 2010/11 and allows for moves between all LADs, even where these sit in the same city region part (so for example, all moves between LADs in Manchester Near are included). The first cluster of bars on each graph shows the proportion of all UK migration that occurs between the 33 LADs in the London Core, all subsequent bars show the proportion of total migration *excluding* this within London migration, due to the large flows exhibited within London at some age groups. By allowing for moves between all LADs, these graphs provide a good overview of the dominant migration trends at each age group in the UK system at the beginning and end of the time series. Given the size of City Near areas (in terms of the number of LADs which they incorporate, as reported in Section 6.1.2 of Chapter 6) and an overall preference for short-distance moves above long-distance moves (as discussed in Chapter 6 in the context of a spatial interaction model), it is not surprising to see that migration between LADs situated within City Near areas dominates, accounting for between 20 per cent of moves at age 20-24 and 27 per cent of moves at age 0-14 and age 75 plus. Analysis of the other migration types reveals more varied patterns by age.

A large proportion of total migrations occur between the 33 LADs that make up the London Core (the first cluster of bars on each graph in Figure 8.1). This proportion is particularly high at age 20-29 where it accounts for 18.9 per cent of total UK migration in 2001/02 and 20.1 per cent in 2010/11 and at age 30-44, where it accounts for 13.3 and 16.6 per cent in 2001/02 and 2010/11 respectively. This increase in proportion of total flow attributable to within London migration is seen in each age group and is particularly notable at age 0-14 where it rises from 8.3 to 10.2 per cent of total and at age 45-59 where the it rises from 6.5 per cent to 8.1 per cent at the beginning and end of the time series respectively.

City Core to City Core moves (which exclude moves between Boroughs within London but includes moves to/from London Core) account for 5.2 per cent of total migration for 20-24 year olds in 2001/02, which rises to 5.7 per cent in 2010/11. In the

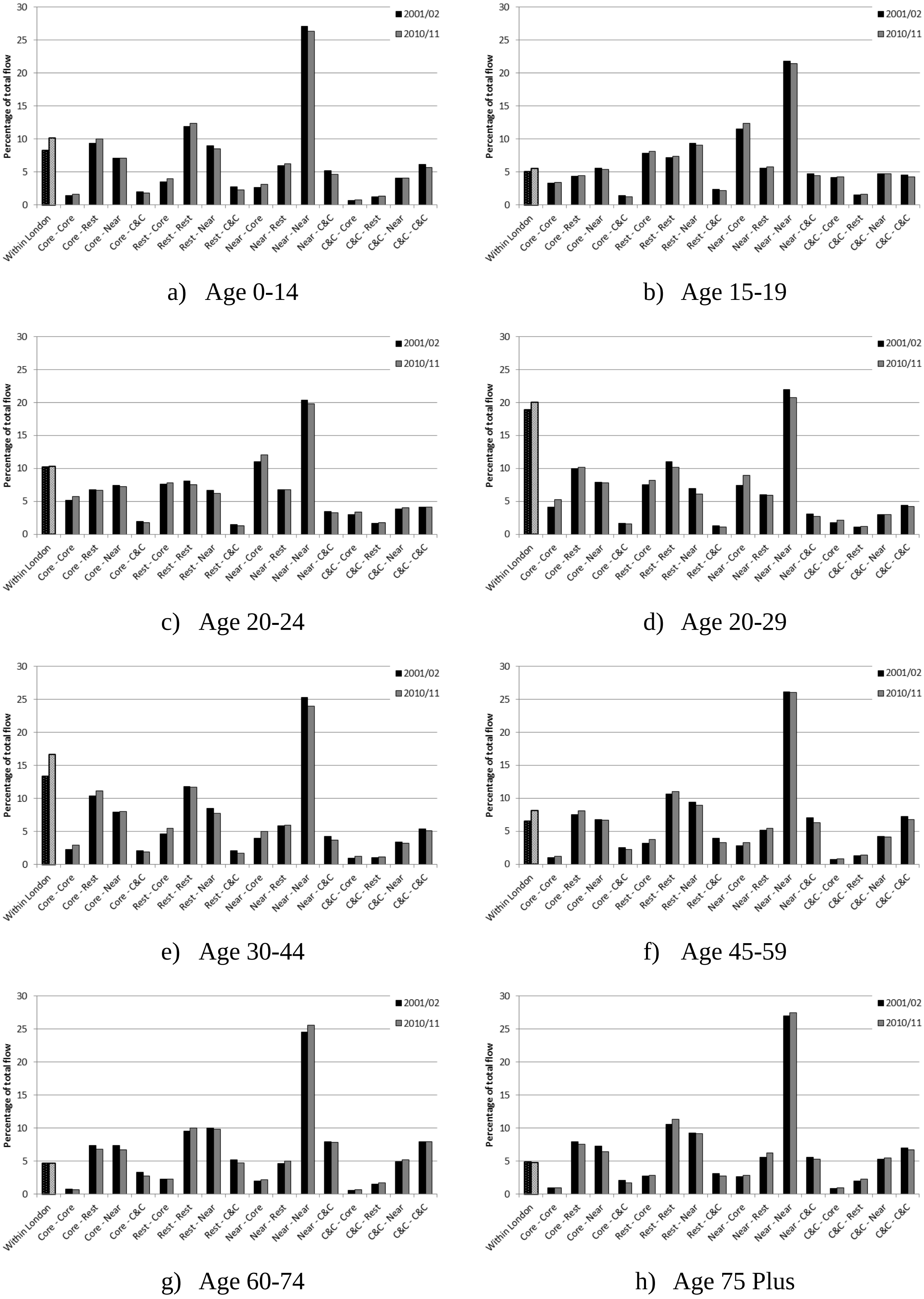


Figure 8.1: The proportion of total migration at each age which can be attributed to flows between city region areas (allowing for all migration including within each city region component part), 2001/02 and 2010/11

20-29 age group, moves between City Core areas account for 4.1 per cent and 5.3 per cent in 2001/02 and 2010/11 respectively, at age 15-19 accounts for 3.3 per cent, rising to 3.4 per cent and at age 30-44 accounts for 2.3 per cent and 3.0 per cent at the beginning and end of the time series respectively. This pattern of general increase in share of total migration between Core areas is true of most age groups between 2001/02 and 2010/11, except at age 60-74 where it falls slightly from 0.7 to 0.6 per cent. The proportion of moves to LADs within City Core areas from LADs in the City Rest, City Near and Coast and Country areas shows a general pattern of increase at all ages.

When all three migration streams are considered cumulatively, the proportion of total moves that involve a migration to a City Core is largest at age 15-19 (23.6 per cent in 2001/02 and 24.7 per cent in 2010/11), at age 20-24 (21.7 per cent and 23.4 per cent) and at age 20-29 (16.8 per cent and 19.4 per cent). The smallest of the three migration streams is the Coast and Country to City Core flow, which accounts for 4.2 per cent of all moves in the 15-19 age group in both 2001/02 and 2010/11, 3.1 per cent and 3.4 per cent of moves for 20-24 year olds in 2001/02 and 2010/11 respectively and for 1.8 per cent and 2.2 per cent of moves in the 20-29 age group. At all other ages, the Coast and Country to City Core stream accounts for less than 1 per cent of total migration. The City Near to City Core move accounts for a large proportion of the flow into the City Cores at age 15-19 (11.6 per cent and 12.3 per cent in 2001/02 and 2010/11), age 20-24 (rising from 11.0 per cent to 12.1 per cent) and at age 20-29 (7.5 per cent in 2001/02 and 9.0 per cent in 2010/11).

Moves from City Core areas to LADs in the City Rest, City Near and Coast and Country can be seen in the third, fourth and fifth group of bars in each graph of Figure 8.1. When all moves out of the City Cores are taken cumulatively, different patterns emerge at each age group. At age 0-14, the proportion of total migration that can be attributed to moves out of the City Core increases slightly from 18.5 per cent in 2001/02 to 18.9 per cent in 2010/11. Similar increases can be seen at age 30-44, where the proportion of total migration rises from 20.4 per cent to 21.1 per cent and at age 45-59 which shows an increase from 16.8 per cent to 17.1 per cent. All other age groups show a decline in the proportion of migration attributable to moves out of the City Core: this is most notable at age 20-24 (where the proportion falls from 16.3 per cent to 15.8 per cent), at age 60-74 (18.1 per cent to 16.2 per cent) and at age 75 plus (17.3 per cent to 15.7 per cent).

8.2 Change in city region flows

The previous section looked at LAD level flows within the city region typology to provide an overview of the general patterns of migration at each age group. This section enforces the city region typology more rigidly by excluding moves within each part of the city region (moves between two LADs that are both situated in Liverpool Near, for example) and focuses on the changing trends across the time series. The goal of this section is to identify which age groups are driving the general patterns of change presented in Chapter 6, which were:

- moves between metropolitan areas (City Core and City Rest) increased steadily between 2001/02 and 2010/11;
- moves from metropolitan City Core and City Rest to non-metropolitan City Near and Coast and Country declined between 2001/02 and 2010/11;
- moves from non-metropolitan to metropolitan areas increased between the beginning and end of the time series; and
- moves between the non-metropolitan City Near and Coast and Country areas declined between 2001/02 and 2010/11.

With the aid of the detailed migration matrices, the changing pattern of city region migration for each broad age group across the decade can be assessed.

Figure 8.2 shows the largest relative changes that occurred between 2001/02 and 2010/11 at each age group, with black and grey lines showing the largest four increases and blue lines showing the largest four decreases. Each migration flow (City Core to City Core, City Rest to City Near etc.) has been calculated as a proportion of total migration within that age group and expressed here as an index based on the 2001/02 proportion. Therefore, Figure 8.2 is showing the relative increase/decrease over the decade for each migration flow within each age band. These changes are summarised in Table 8.1, which presents the difference between the percentage of total flow attributed to a particular migration type in 2010/11 compared with 2001/02. Table 8.2 shows the change in actual number of migrants for each city region move at each age group between 2001/02 and 2010/11: increase is shown in black text while decrease is shown in red text.

The general trend of a rise in City Core to City Core migration (seen in Chapter 6) appears to be driven by all ages from 0 to 59: it is the largest relative increase for 0-

14 year olds, 20-25 year olds and 30-44 year olds across the decade, and is the second, third and fourth largest change for 20-24, 45-59 and 15-19 age bands respectively. This City Core to City Core migration flow is however declining in its relative size for both 60-74 year olds (where it is the second largest decline) and for those aged 75 plus. To put the actual size of these declines into context, City Core to City Core migration for 60-74 year olds accounted for 0.64 per cent of migration in 2001/02 (524 people) and 0.59 per cent in 2010/11 (509 people), while the increase for 25-29 year olds was from 5.38 per cent of total migration in 2001/02 (13,991 people) to 6.85 per cent in 2010/11 (19,075 people). So while the trend differs at the older and younger age groups, the size of the increase at younger ages far exceeds that of the decrease for those aged over 60.

A rise in the proportion of migration to the City Core from all other parts of the city region is seen for most age groups: moves from City Rest, City Near and Coast and Country to the City Core are among the largest relative increases for all age groups up to age 59, with the exception of the Coast and Country to City Rest flow which increases for ages 15-19 and 20-24. It is worth noting that these two age groups exhibit a lower rate of relative change in relation to other ages up to 59, suggesting that the structural higher education (at age 15-19) and work (at age 20-24) influences exhibit a strong influence on the pattern of migration (more on this later in the chapter). The proportion of City Core to City Near moves are also increasing for 60-74 year olds.

When looking at the declining migration flows, it is moves to Coast and Country areas that exhibit the largest relative fall at most age groups. For ages 0-14, 15-19, and 45-59 this decline in the proportion of moves to Coast and Country areas from all others dominates the pattern seen in Figure 8.2. Moves from City Rest to Coast and Country show the largest decline for the age bands 0-14, 20-24, 25-29, 30-44 and 45-59. The percentage change and absolute number of migrants is large at age 0-14 (down from 4.12 per cent in 2001/02 to 3.45 per cent in 2010/11, representing a decline of 2,253 migrants), 30-44 (3.13 to 2.51 per cent, 7,181 people) and 45-59 (5.99 to 5.07 per cent, 2,179 people). The move from City Core to Coast and Country is the largest relative decline for those aged 15-29 and 60-74 (a drop of 0.19 per cent and 0.73 per cent of total migration respectively).

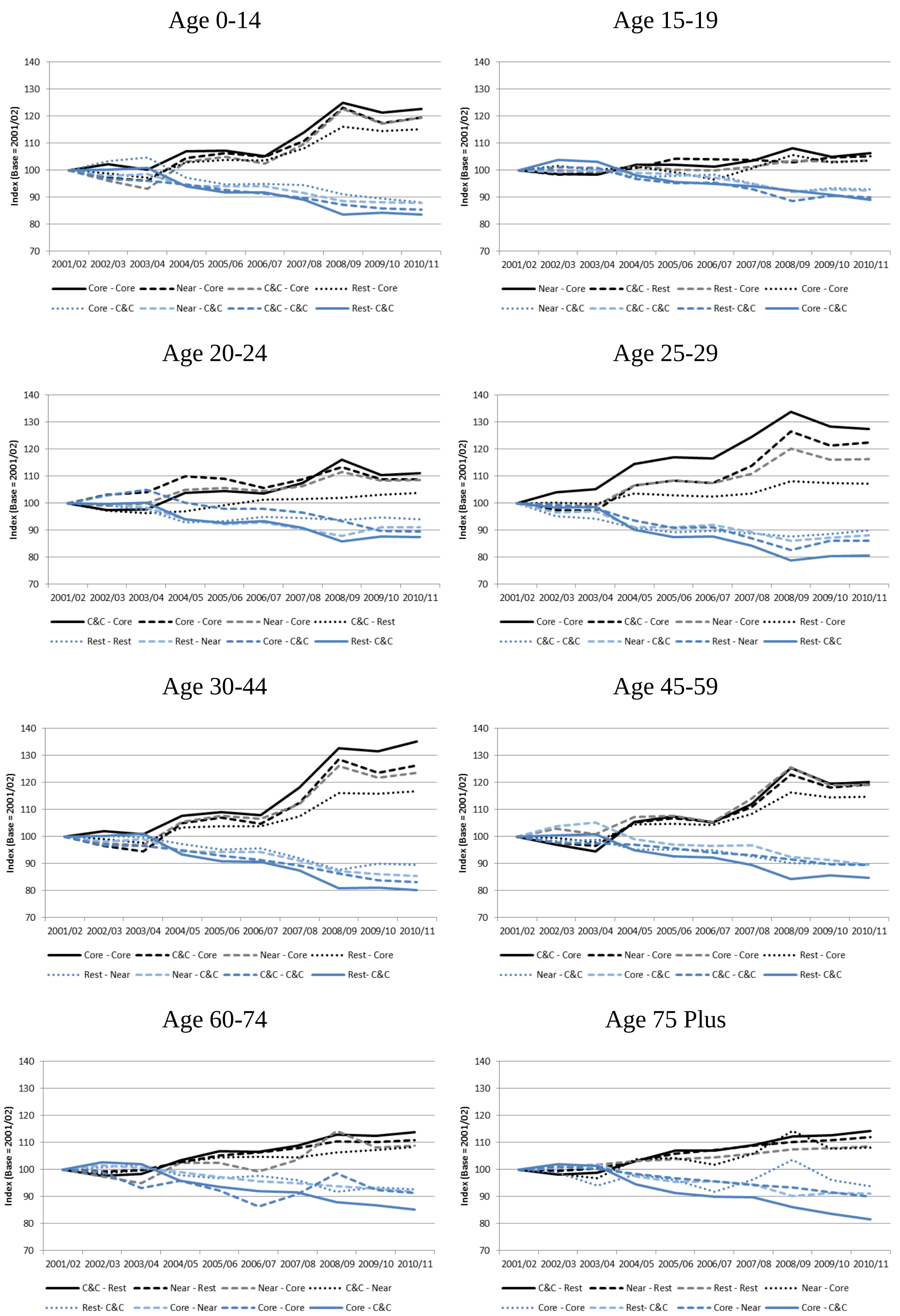


Figure 8.2: The relative change in proportion of migration for each age group between city region component parts, 2001/02 to 2010/11 (largest four increases and decreases)

A good summary of the changing patterns of migration can be obtained from Table 8.1. Moves up the urban hierarchy from less urban to more urban areas are generally increasing: at all ages, the proportion of total moves originating in City Rest, City Near and Coast and Country areas moving to a City Core increases, Coast and Country to City Rest moves increase at all ages increases while City Near to City Rest moves increase at ages 0-14, 15-19, 45-59, 60-74 and 75 plus (but decline at ages 20-24, 25-29 and 30-44). The proportion of total migration that occurs between City Core areas increases at all ages to 59.

Table 8.1: Change in the proportion of total migrants in each age group that can be attributed to flows between component city region parts, 2010/11 compared with 2001/02

Flow/Age	**0-14**	**15-19**	**20-24**	**20-29**	**30-44**	**45-59**	**60-74**	**75 Plus**
Core - Core	0.37	0.14	0.57	1.47	1.02	0.24	-0.06	-0.05
Core - Rest	1.04	0.11	-0.20	-0.18	0.96	0.88	-0.70	-0.48
Core - Near	-0.06	-0.23	-0.36	-0.46	-0.04	-0.03	-0.89	-1.11
Core - C&C	-0.36	-0.19	-0.27	-0.24	-0.30	-0.40	-0.73	-0.60
Rest - Core	0.81	0.36	0.15	0.77	1.17	0.73	0.01	0.16
Rest - Rest	0.03	0.03	-0.12	-0.12	-0.14	0.08	0.07	0.16
Rest - Near	-0.76	-0.48	-0.78	-1.38	-1.34	-0.68	0.04	0.03
Rest- C&C	-0.67	-0.30	-0.24	-0.38	-0.62	-0.92	-0.55	-0.42
Near - Core	0.79	0.92	1.22	1.75	1.41	0.83	0.27	0.34
Near - Rest	0.45	0.25	-0.15	-0.39	-0.06	0.49	0.74	1.05
Near - Near	-0.59	-0.19	-0.25	-0.72	-1.01	-0.23	0.74	0.44
Near - C&C	-0.96	-0.43	-0.21	-0.52	-0.93	-1.11	0.09	-0.38
C&C - Core	0.19	0.03	0.44	0.56	0.37	0.22	0.07	0.11
C&C - Rest	0.10	0.10	0.09	0.04	0.01	0.13	0.31	0.44
C&C - Near	-0.11	-0.02	0.15	-0.10	-0.27	-0.03	0.61	0.39
C&C - C&C	-0.25	-0.09	-0.02	-0.09	-0.22	-0.21	0.00	-0.07

*Red denotes decline

The proportion of total moves that occur down the urban hierarchy (from more to less urban) generally declines between 2001/02 and 2010/11: moves from City Core to City Near and Coast and Country areas decline at all ages while moves from City Core to City Rest areas falls at ages 20-24, 25-29, 60-74 and 75 plus (at all other ages it increases). Moves from City Rest to City Near areas decline in their share of total migration at all ages up to 59 while moves from City Rest to Coast and Country decline at all ages. Moves between the less urban City Near areas falls at all ages up to 59, while the proportion of moves occurring between Coast and Country areas declines at all ages.

Table 8.2: Change in the number of total migrants in each age group that can be attributed to flows between component city region parts, 2010/11 compared with 2001/02

Flow/Age	**0-14**	**15-19**	**20-24**	**20-29**	**30-44**	**45-59**	**60-74**	**75 Plus**
Core - Core	482	145	5,167	5,084	1,840	333	-15	-21
Core - Rest	-514	33	2,805	2,105	-6,043	962	-40	-179
Core - Near	-2,234	-708	2,477	789	-7,343	-438	-194	-505
Core - C&C	-1,370	-448	-94	-210	-2,932	-782	-376	-288
Rest - Core	726	376	4,757	4,088	79	981	189	105
Rest - Rest	-432	-20	320	48	-1,907	56	148	92
Rest - Near	-4,296	-1,372	281	-2,026	-12,512	-1,643	806	94
Rest- C&C	-2,253	-715	-230	-696	-4,110	-1,692	-78	-190
Near - Core	927	1,312	11,193	6,830	1,365	1,154	390	197
Near - Rest	-776	246	3,004	487	-5,550	478	999	581
Near - Near	-3,733	-814	3,137	-394	-10,293	-812	1,199	283
Near - C&C	-3,622	-1,069	947	-653	-7,181	-2,180	687	-148
C&C - Core	225	-131	3,527	2,028	452	315	108	61
C&C - Rest	-161	133	1,278	404	-970	135	384	243
C&C - Near	-1,446	-256	2,746	512	-4,007	-293	907	243
C&C - C&C	-879	-228	307	-92	-1,572	-412	132	-26
Total	-19,357	-3,517	41,621	18,305	-60,842	-3,837	5,246	542

*Red denotes decline

The change in absolute number of migrants between 2001/02 and 2010/11, summarised in Table 8.2, reveals some contrasting trends when the age groups are compared. The number of migrants crossing between the boundaries of the city region component parts increases by 41,621 at age 20-24, 18,305 at age 25-29, by 5,246 at age 60-74 and by 542 at age 75 plus. At all other ages, the total number of migrants declines, most notably at age 30-44 where there are 60,842 fewer migrants in 2010/11 than 2001/02 and at age 0-14, where there are 19,357 fewer migrants.

The largest increases in terms of absolute numbers can be seen for City Near to City Core moves at age 20-24 (up by 11,193) and at age 25-29 (up 6,830) and for these age groups where City Core to City Core moves are considered (both show an increase of over 5,000). The largest decreases are evident in the 30-44 age group: moves from City Rest to City Near areas falls by 12,512 migrants and moves from City Near to City Near areas fall by 10,243.

To summarise the previous two sections, the trend of increased City Core to City Core migration is apparent at all ages except 60 to74, and is most pronounced at ages 25-29 and 30-44. The pattern of increase in proportion of moves to City Core LADs from LADs in the City Rest, City Near and Coast and Country regions is true at all ages, and most pronounced at age 25-29. The general decrease in the proportion of moves from City Core areas to all others is masking a more mixed pattern by age: small increases can be seen at ages 0-14, 30-44 and 45-59, while a decline is seen at all other age groups, and is most notable at age 60-74. The following section uses the migration indicators introduced in Chapter 5 to look at patterns across the decade.

8.3 Migration indicators

This section picks up the migration indicators discussed in Chapter 5, where some key trends were identified across the time series for the aggregate origin-destination (OD) matrix. Here the analysis reverts back to migration in the whole UK system, between all 406 LADs. The trends identified for the aggregate OD matrix were, briefly:

- migration distance and intensity are falling (measured by crude migration intensity, mean migration distance and median migration distance);
- connectivity is falling while spatial focus and inequality are rising (measured by the index of migration connectivity, migration inequality index, coefficient of variation and Theil index); and
- the efficiency of migration as a force for redistributing the population around the system is falling (measured by the migration efficiency index and aggregate net migration rate).

It is possible to interrogate these trends in more detail by looking at the migration indicators by age. Figure 8.3 shows the index of connectivity while Figures 8.4a – h show each of the other migration measures by broad age in 2001/02, 2006/07 and 2010/11, thus comparing beginning, middle and end of the time series. 2006/07 is chosen over 2005/06 here to maintain consistency with the results reported in the rest of the thesis and represents the year when overall migration activity was at its peak. A detailed description of the indicators can be found in Chapter 5, but where necessary a recap is provided in this section.

Crude migration intensity (CMI) is a measure of the proportion of population at risk that migrate in a given year. The pattern by age seen in Figure 8.4a is consistent with the migration literature, a rise from childhood to the most active 20-24 age band, where it then declines to retirement age. There is a small upturn at age 75 plus as a larger proportion of people move in retirement and old age. Mean and median migration distances (Figures 8.4b and 8.4c respectively) show that there is a peak in distance moved at age 15-19 which declines to age 30-44 before increasing steadily to reveal that the longest migration distances are for those aged 75 and over.

The fall in CMI between 2001/02 and 2010/11 seen for the aggregate migration matrix is evident at all ages (Figure 8.4a) but is most pronounced at ages 20-24 and 25-29. There is very little difference between 2001/02 and 2006/07, but in 2010/11 CMI has fallen by 1.0 at age 20-24 and by 0.91 at age 25-29 suggesting that at these ages, a smaller proportion of the population at risk are migrating. Mean and median migration distances fall in both 2006/07 and 2010/11 at all ages compared with 2001/02, and in the latter year the drop is most pronounced at age 45-59 (which falls 2.3 km in mean distance and 4.4 km in median migration distance) and at age 60-74, where it falls 4.0 km and 5.0 km for mean and median migration distances respectively. So in summary, a general decline in migration intensity and in mean and median migration distance is true for all ages, intensity falls most notably at ages 20-29 while for distance it is ages 45-74 which show the largest decline.

Connectivity, spatial focus and inequality can be measured by the index of connectivity, migration inequality index, coefficient of variation (CV) and Theil index. Figure 8.3 shows the index of connectivity at each age group, with the percentage of LAD to LAD flows that are connected by over 10 migrants being reported. It is difficult to compare the connectivity of each age group, given the irregular banding used (age

30-44 will always be more connected than age 25-29, given the 15 ages included in the former and only five in the latter). However, it does give a good overview of change over time: the general trend was one of decline for the aggregate OD matrix between 2001/02 and 2010/11, and Figure 8.3 shows some variation by age. The fall in connectivity can be seen at age 0-14, most notably at age 30-44 where it falls from 7.7 per cent in 2001/02 to 6.3 per cent in 2010/11 and at age 45-59. At all other ages there is an increase, most notable at age 20-24 where it increases from 7.7 per cent in 2001/02 to 8.5 per cent in 2010/11.

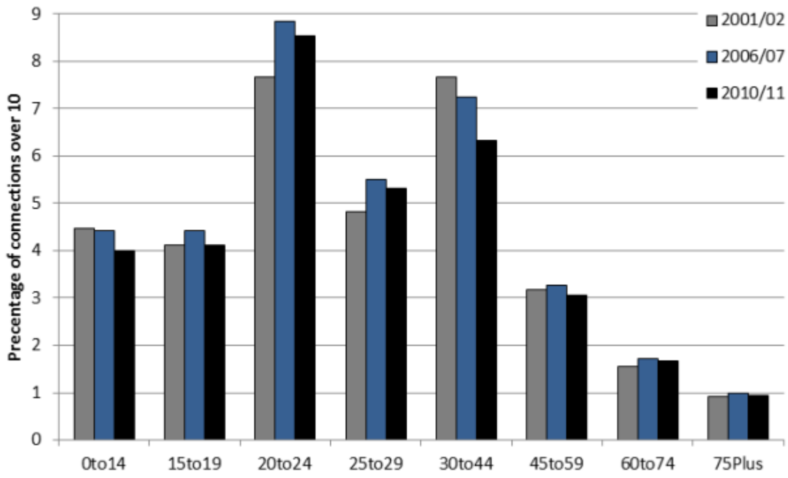


Figure 8.3: The percentage of LADs in the UK connected by over 10 migrants, by broad age in 2001/02, 2006/07 and 2010/11

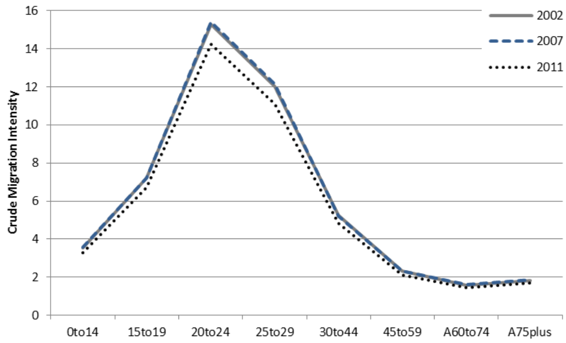
The index of inequality (Figure 8.4d), which compares the observed distribution to an expected distribution (here the expected distribution is one where the magnitude of all LAD to LAD flows is the same) shows that inequality is highest at young ages (0-19) and again at older ages (45-75 plus) but that the variation over the time series is fairly small. The inequality score increases across all age groups in 2010/11 compared with 2001/02 but is only discernible in Figure 8.4d at ages 20-44. The CV and Theil index are measures of spatial concentration, but the scores by age differ between the two measures. The lowest score at age 15-19 seen in the CV is an age group later in the Theil at 20-24. Both increase to age 25-29, where the CV score peaks. The Theil index, however, increases from age 30-44 onwards to peak at age 75 plus.

A more pronounced change across time can be seen in the CV (Figure 8.4e) and Theil index (Figure 8.4f) than in the index of inequality. The CV is actually slightly lower in 2010/11 than 2001/02 for ages 15-19, 60-74 and 75 plus. There is, however, a sharp increase at age 25-29 (a difference of 0.33 between 2001/02 and 2010/11) and at age 30-44 (where the difference is 0.44). The CV is also higher at the end of the time period at ages 0-14, 20-24 and 45-59. This pronounced increase at ages 25-29 and 30-44 is also evident when comparing the Theil score in 2001/02 and 2010/11. The difference in Theil index score between the beginning and end of the time series for all other ages is low. In summary, connectivity is increasing at ages 20-24 and 25-29 and decreasing at ages 0-14, 30-44 and 45-59, while inequality appears to be increasing across all age groups, but most substantially at ages 20-24 and 30-44.

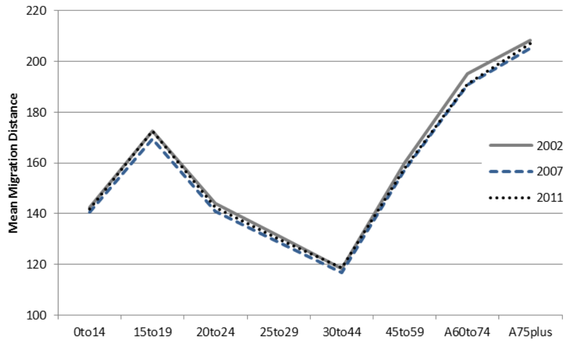
The migration efficiency index measures how efficient migration is as a force for redistributing the population, while the annual net migration rate (ANMR) summarises the extent of the population redistribution. The highest efficiency score (Figure 8.4g) can be seen at age 15-19, which is also true of the ANMR (Figure 8.4h). Both decline to age 25-29, but while the ANMR continues to fall, migration efficiency begins to rise again to age 60-74, before declining again at age 75 plus.

It was reported in Chapter 5 that, overall, the ANMR and the migration efficiency index fell across the time series from 2001/02 to 2010/11. This suggests that the efficiency of migration as a force for redistributing the population has fallen. In terms of the migration efficiency index, at ages 15-19, 20-24 and 25-29, migration appears to be more efficient in redistributing population in 2010/11 than in 2001/02, with the biggest increase being seen in the 20-24 age band. At all other ages, migration efficiency has fallen across the time series, most notably at ages 45-59 and 60-74. The ANMR also falls most notably at ages 30-44 and 45-59, and declines at all ages except 20-24, where the rate is higher in 2010/11 than in 2001/02.

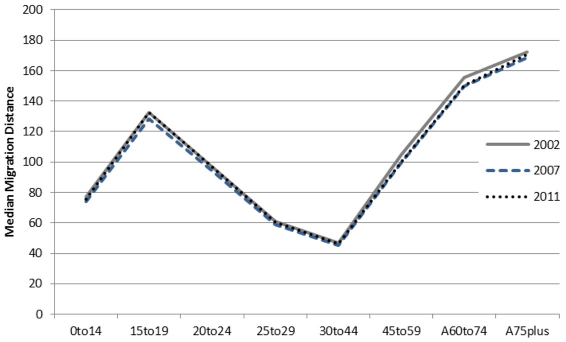
In summary, migration efficiency and net migration rate falls at most ages and is most notable from age 30-44 onwards. Ages 15-29 see an increase in migration efficiency, but the aggregate net migration rate only increases at age 20-24 (which shows the biggest increase in efficiency). The peak in efficiency at age 15-19 becomes easier to interpret when looking at the pattern of net migration rates presented in Figure 8.8, which is one of large-scale loss from many areas and large scale gain in a small number of university towns. This pattern is discussed further below.



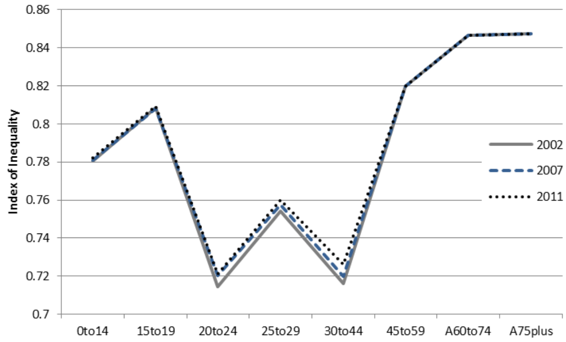
a) Crude migration intensity



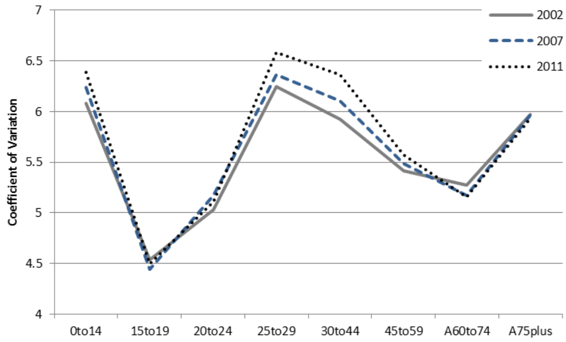
b) Mean migration distance



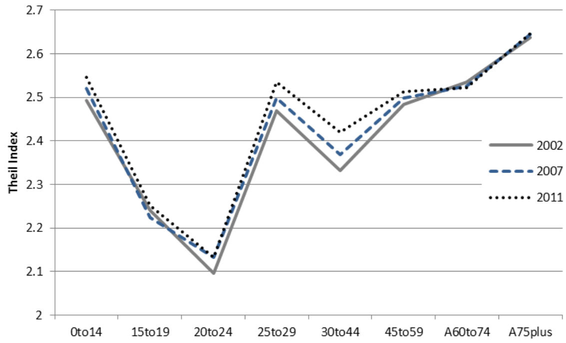
c) Median migration distance



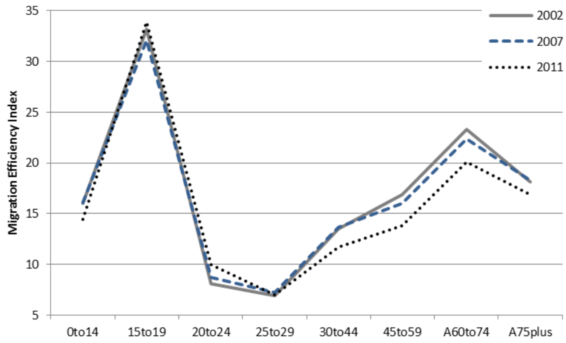
d) Index of inequality



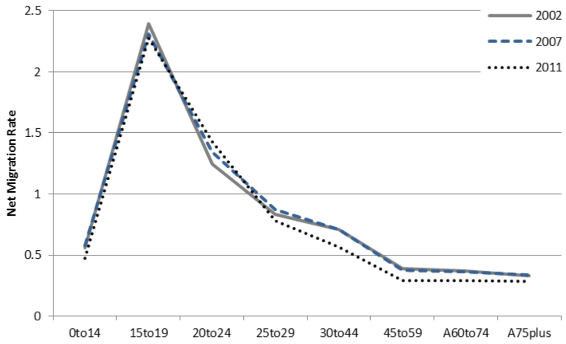
e) Coefficiant of variation



f) Theil index



g) Migration efficiency index



h) Net migration rate

Figure 8.4: Migration indicators at each broad age group for 2001/02, 2006/07 and 2009/10

Having established the general trends at each age group, the following six sections look at the net migration rates at LAD scale for each age and incorporates discussion from the literature on the patterns that exist. The net rate maps presented are for 2001/02 and 2010/11 in order to provide comparison in the changing patterns at the beginning and end of the time series, and the scale for all maps is the same, making comparison of age groups possible.

8.4 Family migration (children aged 0-14 and parents aged 30-44)

The migration pattern of children aged 0-14 (Figure 8.5) is largely influenced by their parents' (or guardians') migration decisions, and immediate comparison can be drawn with the pattern of net migration rates exhibited by the 30-44 age group (Figure 8.6), who make up the majority of the parents/guardians of this 0-14 age group. Dobson and Stillwell (2000) argue that, as a distinct group, migration analysis and theory have largely ignored children, suggesting that in order to find studies that investigate the role of children, one needs to look at literature which considers children in the context of family migration. Even where they are considered in the family context, Smith (2011) argues that the scope of most studies tend to be narrowly focused on economic motivations, advocating instead *"a more flexible approach to studies…to unpick the growing complexities of family migration processes",* for example, where a whole household does not move and a rise in 'live-apart-together' (LAT) couples. Although migration rates for 0-14 and 30-44 year olds are lower than at other ages, Tunstall *et al.* (2010, p.786) argue that family formation is a substantial motivation for residential mobility, while *"pregnancy and childbirth are important triggers for mobility of households, with the birth of a first child most likely to result in relocation".*

Champion (2005, p.94) points towards the tendency of under 16s to move very short distances, citing evidence from the 2001 Census that suggests 74.2 per cent of child migrants move less than 10 km. In the absence of intra-LAD moves in the dataset (which represent the majority of these very short distance migrations), this is a difficult assertion to test. However, Figure 8.4 does show that mean and median migration distances for children in the 0-14 age group is the third lowest, behind those aged 25-29 and those aged 30-44. At the 25-29 age group, this short distance may be attributable to the large proportion of migration that occurs between London boroughs (around 20 per

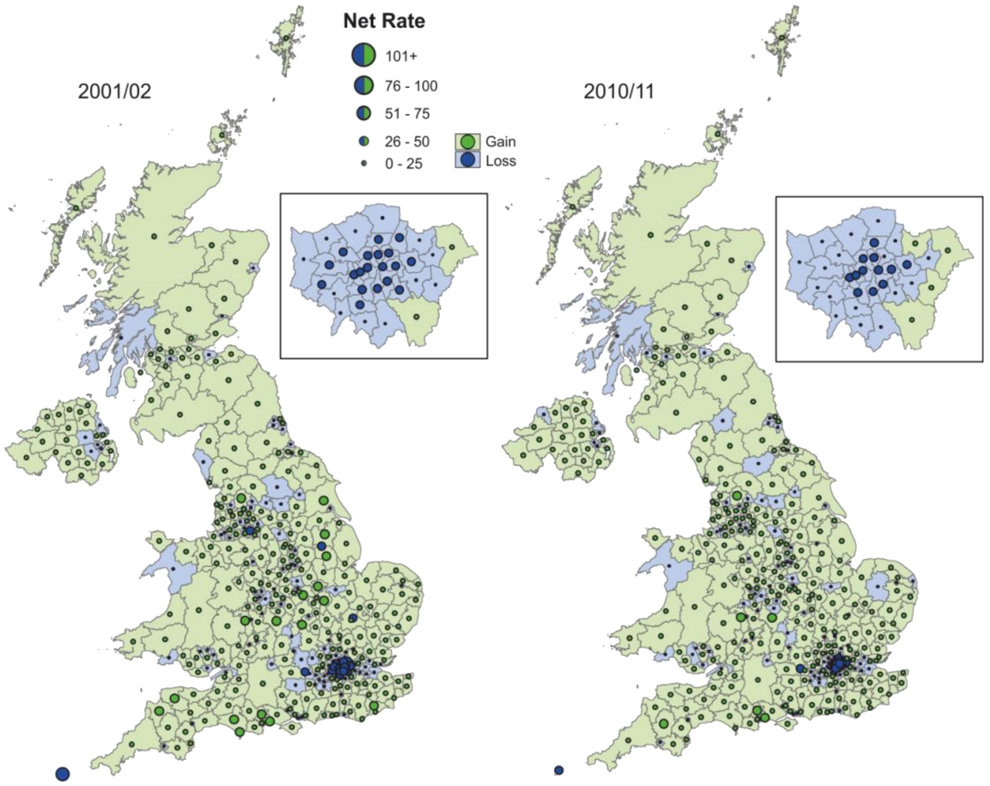


Figure 8.5: The net migration rate for persons aged 0-14 in each LAD, 2001/02 and 2010/11

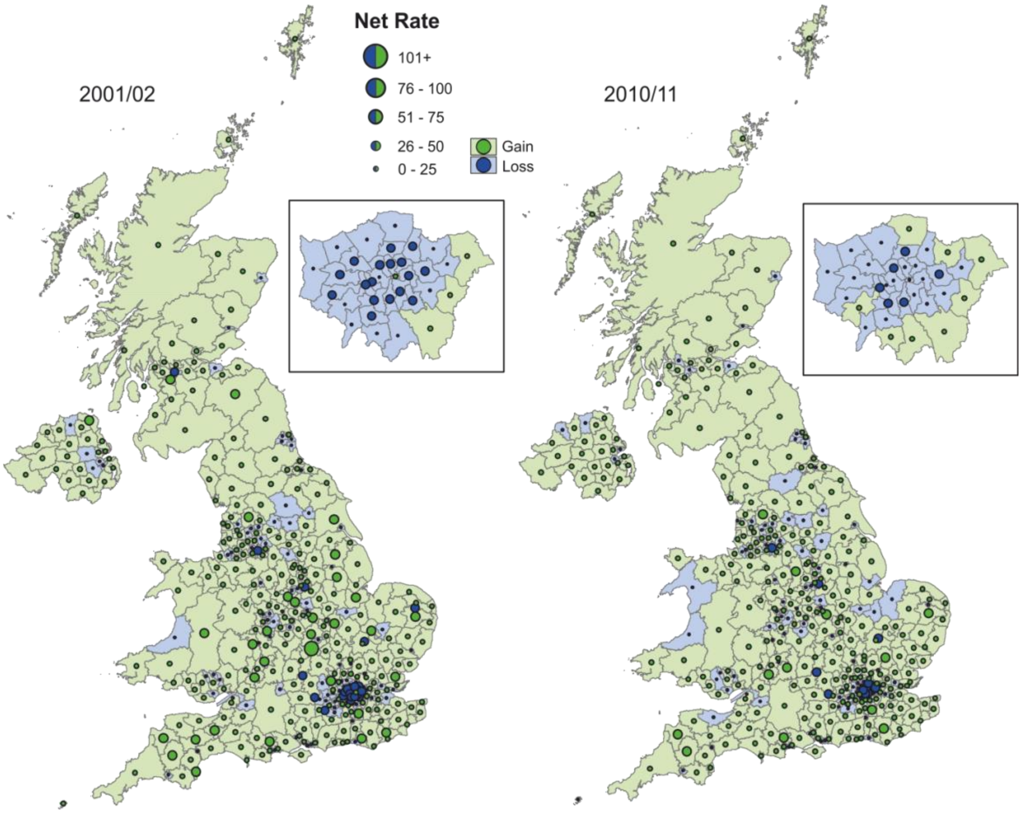


Figure 8.6: The net migration rate for persons aged 30-44 in each LAD, 2001/02 and 2010/11

cent in each year) while migration distances for those aged 30-44 are consistently the shortest in the dataset.

Champion (2005, p.105) attributes the majority of child migration to decisions made by their parents, namely *"minimising the disruption to their children's schooling"* and *"that families with children are more likely to move out of large cities than to move to them".* The role of a child's schooling in the migration decision of a family is picked up by Dobson and Stillwell (2000) and Harland and Stillwell (2007) use data from the Pupil Level School Census for Leeds to show how residential migration is associated with a change of school, particularly at age 11. Smith and Higley (2012, p.54) make the link between access to high quality schools and a move down the urban hierarchy as *"the provision of high-quality school education in gentrified rural hinterlands of metropolitan centres is a factor mediating the in-migration of affluent middle-class families, triggered by constraints in exclusive urban education and housing markets".* Similar findings are reported by Plane *et al.* (2005, p.15317) who, in a study on migration data in the USA, identify the dominant process for those in the age groups 30-34 and 35-39 (defined as mid-career and childrearing years) as one of movement down the urban hierarchy, with housing costs, school quality, suburban road congestion and a preference for lower population densities being cited as the driving factors. In the dataset presented here, this move down the urban hierarchy is evident for both the 0-14 (Figure 8.5) and 30-44 (Figure 8.6) age groups. Net loss can be seen in most London boroughs (especially central London) and also from the urban LADs of the North West and urban conurbation of the West Midlands. Net gain in rural areas is evident across the UK in both 2001/02 and 2010/11 (the largest rates of gain in Figures 8.5 and 8.6 are seen in areas with relatively low population density). Figure 8.1 shows that behind moves between LADs within City Near and City Rest areas, counterurbanisation moves from City Core to City Rest and Near and from City Rest to Near constitute a large proportion of total migration, and exceeds moves in the other direction (up the urban hierarchy).

The clustering of LADs based on their age profile, undertaken in the previous chapter, sheds light on the similarity between the areas that lose and gain 0-14 and 30-44 year olds. In 2001/02, of the 40 largest net gaining LADs in the 0-14 age group, 15 are within Cluster 2 and 14 are within Cluster 7. In the 30-44 age group the cluster, association is very similar: of the 40 LADs exhibiting the largest rate of gain, 16 are within Cluster 2 and 14 are in Cluster 7. The clustering of LADs that lose net migrants at the highest rate in 2001/02 is also similar between the age groups: 12 LADs fall within outflow Cluster 4 for both 0-14 and 30-44 year olds, while 16 LADs for age 0-14 and 15 LADs for age 30-44 fall within outflow Cluster 6. The same is true for 2010/11: seven of the largest net gainers at age 0-14 are in Cluster 2, while ten of the largest net gaining LADs of 30-44 year olds are in the same cluster. For both ages, 18 LADs sit within inflow Cluster 7. For the LADs exhibiting largest net loss in 2010/11, at both age 0-14 and 30-44, 12 are in outflow Cluster 4. Outflow Cluster 6 contains 16 LADs losing 0-14 year olds at a high rate and 13 LADs losing 30-44 year olds. These results lend strength to the decision to consider 0-14 and 30-44 year olds together as a family unit, although there are clearly some differences in net migration rates at LAD level seen in Figure 8.5 and Figure 8.6.

8.5 Leaving home for education (or first job) – age 15-19

Those aged 15-19 are identified in much of the literature as an age group whose migration decisions are driven by access to further or higher education (HE), but a proportion of migrants in this age group move to take up their first job outside of the family home. Kalogirou (2005) finds that the existence of college and HE establishments in an area is strongly related to migration trends for the 16-29 age group (inflow for 16-19 year olds and out migration for 20-29 year olds) but even where the main reason for migration is not education, there is some evidence that the destination choice is similar for non-student migrants (McHugh and Morgan 1984).

Figure 8.7 shows the net rates of gain/loss in each LAD for 15-19 year olds. The connection between HE and migration in this age group becomes clear: the top 40 LADs in terms of net gain (over 40 persons per 1,000 resident population) all contain a higher education institution. The largest gains can be seen in Cambridge (268 per 1,000 in 2001/02 and 271 per 1,000 in 2010/11), Oxford (254 per 1,000 and 270 per 1,000) and Oadby and Wigston (209 per 1,000 and 193 per 1,000). The latter was picked up as an anomaly during the cluster analysis of the previous chapter, due to the large proportion of migratory students who attend halls of residence associated with the University of Leicester.

This age group tends to dominate the migration profile of an area: of the LADs showing the largest net gains, 19 are in inflow Cluster 6 and 13 are in inflow Cluster 3. The majority of the LADs exhibiting large net loss fall in outflow Cluster 3 (17 LADs) and Cluster 7 (15 LADs) and the 40 LADs showing the largest net losses (losing over 62 per 1,000) were all identified in Chapter 5 as having low population densities (8 to 474 persons per sq. km.). The largest net rates of loss are exhibited by Elain Sair (119 per 1,000 in 2001/02 and 109 per 1,000 in 2010/11), Maldon (95 and 84 per 1,000) and Dungannon (88 and 80 per 1,000).

For LADs with the highest inflow rates, the substantial influx of 15-19 year olds to an area can exhibit a large influence on parts of the town or city where students tend to cluster, giving rise to a body of literature that addresses the phenomenon of 'studentification.' The term, established by Smith (2002b, p.15) embodies "*a growing concentration of student residences, associated with a proliferation of houses in multiple occupation (HMOs)"*. This studentification, Smith reports, has an impact on local and wider housing markets, leisure and retail and implications for local communities and public service infrastructures. Duke-Williams (2009) highlights some negative aspects of the process for local non-student residents, such as local properties being snapped up for investment purposes, reducing the local housing stock, while the problem of segregation is raised by Hubbard (2009), where the recent expansion of purpose-built student development attracts students away from more established, mixed communities. This problem is exacerbated when purpose-built student housing is poorly located and has the effect of displacing established local residents. These social, economic and housing disadvantages are particularly felt by *"established populations in deprived urban areas located in close proximity to university campuses"* (Sage *et al.* 2012, p.1076). Some potential strategies to combat the negative effects of studentification within the wider gentrification agenda are proposed by Smith (2008), which include mandatory licencing of student homes, landlord accreditation and most importantly developing policy which restricts the 'ghettoisation' of student communities while pursuing the goal of promoting sustainable communities.

Figure 8.8 shows the number of students aged 15-19 (per 1,000 resident population aged 15-19) who migrated to a different LAD to attend a higher education institution (HEI), reported in data provided by the Higher Education Statistics Agency (HESA). All students whose home domicile and term-time address are the same (those living at home while studying) are excluded from the data, so only moving students are reported. The left hand map shows all LADs as origins prior to the student commencing study (usually the parental address of a student) while the right hand map shows LADs as destinations (LAD of term time residence). While it is usual for students in HEIs to be over 18 (over 17 in Scotland), a very small number of migrants are reported in the HESA data as ‘16 and under’. These students could be attending access to HE courses, or the age could be reported incorrectly in the data. The vast majority appear to be living at their home address while studying, and the universities with the highest number of students under 16 offer foundation level courses.

Figure 8.8 shows that, apart from five LADs in the South East of England (Chiltern, Elmbridge, Mole Valley, South Bucks and St Albans) and one in London (Richmond upon Thames), all out-migration rates are below 100 per 1,000 resident population. The spatial concentration of in-migration is however much more focused on a small number of university towns and cities. On the right hand map, all LADs highlighted in dark blue have between 201 and 526 students aged 15-19 migrating into the area, per 1,000 resident population. It is clear that whilst migration for study in higher education institutions is a substantial factor in both in and out migration at LAD level, outflow is more uniform across the UK while inflow is spatially concentrated to a small number of university towns and cities.

The inflow pattern for students attending HE establishments seen in Figure 8.8 mirrors that seen in the net migration rate maps seen in Figure 8.7. The data used in Figure 8.7 for England, Wales and Northern Ireland contain an adjustment for student migrants, but data for Scotland does not. Gains of between 101 and 200 students per 1,000 resident population in Edinburgh, Aberdeen, Dundee and Stirling (and a rate of 76-100 in Glasgow) reported in the HESA data suggest that the number of student migrants is substantial enough to warrant an adjustment for student populations in Scotland and that such an adjustment, incorporated by NRS would improve migration statistics for Scotland.

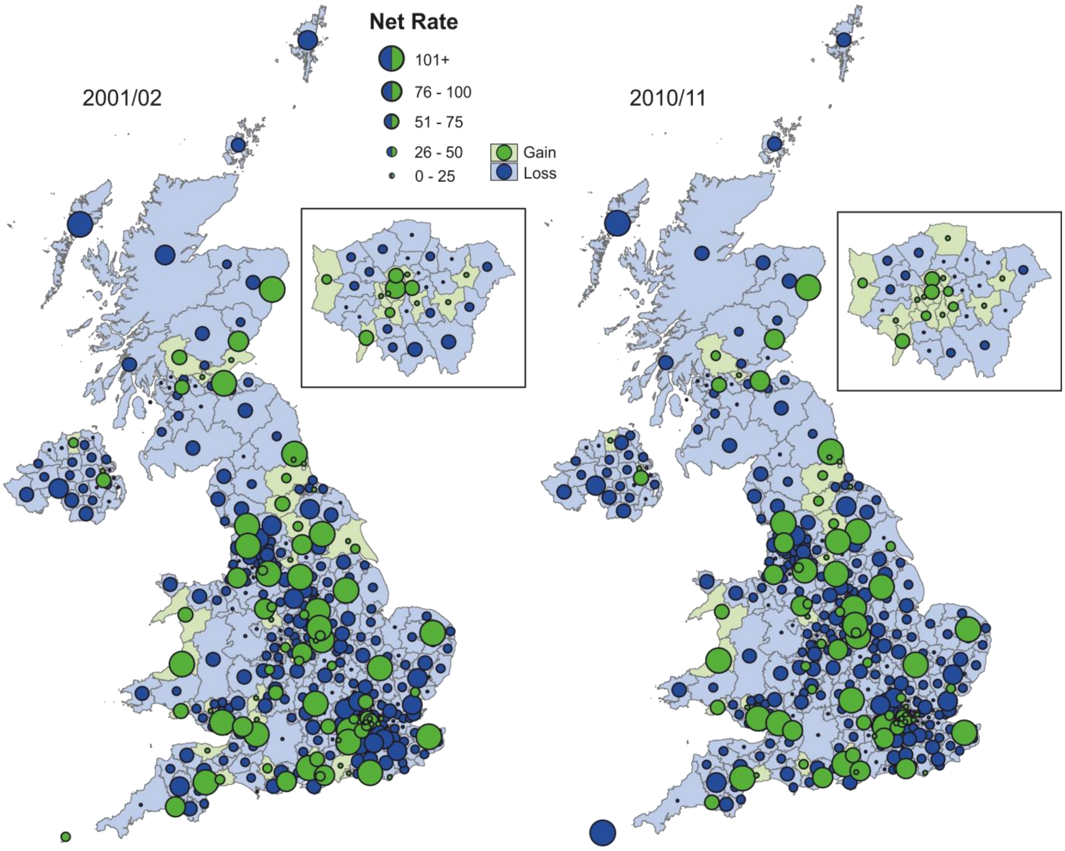


Figure 8.7: The net migration rate for persons aged 15-19 in each LAD, 2001/02 and 2010/11

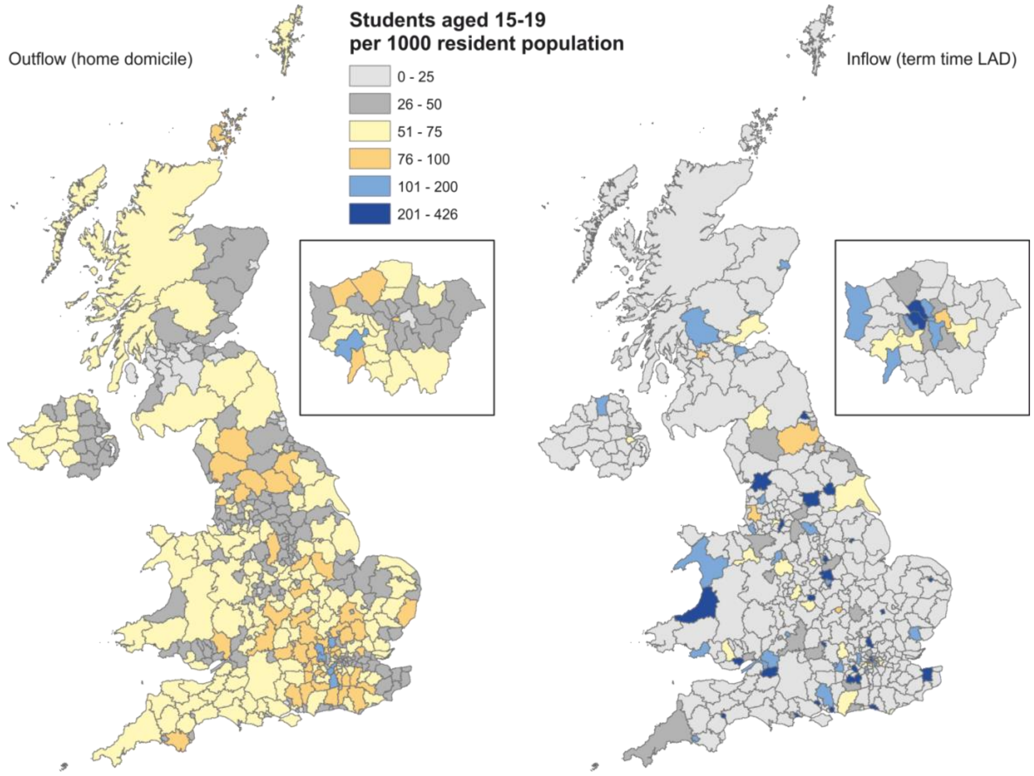


Figure 8.8: Students aged 15-19 making their first move to higher education as a proportion of the resident population aged 15-19 in each LAD (average of 2008/09 and 2009/10 data)

8.6 Leaving education for work – age 20-24

Figure 8.4 shows that at age 20-24, outflow from student towns, as identified by Kalogirou (2005) and Dennett (2010), is demonstrated by widespread net loss from LADs with HE institutions. Of the 40 LADs showing largest net gain for 15-19 year olds in 2001/02, 22 show a net loss of migrants in the 20-24 age range. The same is true in 2010/11, where of the top 40 net gains for 15-19 year olds, 20 show net loss of 20-24 year olds. In both years, Oadby and Wigston (Leicester University), Ceredigion (Aberystwyth University) and Lancaster show the largest rates of net loss for 20-24 year olds.

While the close association between areas gaining 15-19 year olds and losing 20-24 year olds provides strong evidence of graduates migrating out of areas of HE after leaving university, attributing destinations to university leavers is not so easy. Hoare and Corver (2010) shed some light on the post-university trend of graduates. They compile a dataset of recent graduates, comprising region of HE education and first destination (region) of full-time employment post-HE, finding that, overwhelmingly, London recruits graduates from other UK regions, whilst all other regions except Yorkshire and the Humber and Scotland lose recent graduates. Yorkshire Forward (2010), the former regional development agency, report that Yorkshire and Humber had the highest regional gradual retention rates in 2008/09 and emphasise the economic benefits of retaining these skilled people in the labour force. Faggian and McCann (2009b, p.320) report that in London, there are 40 per cent more university graduates employed there than were actually educated there. This pattern of net gain in London plays out in Figure 8.9: it appears that London is the destination of choice for many migrants, with all 33 LADs in London (especially those in central London) gaining 20-24 year olds in 2010/11. This net gain for London is more pronounced in 2010/11 than in 2001/02, suggesting that London has become a more attractive destination for recent graduates as the decade progresses. The position of equilibrium in Yorkshire and the Humber (where inflow of recent graduates is the same as outflow) reported by Hoare and Corver (2010) is consistent with the pattern seen in Figure 8.9 for 2001/02, where Leeds, Wakefield, Doncaster, Richmondshire and Hull all show positive net migration rates for 20 to 29 year olds. In 2010/11, however, only Leeds and Hull are showing a net gain in this age group. In Scotland, however, only Glasgow, Falkirk and Edinburgh consistently show small rates of net in-migration for 20-24 year olds.

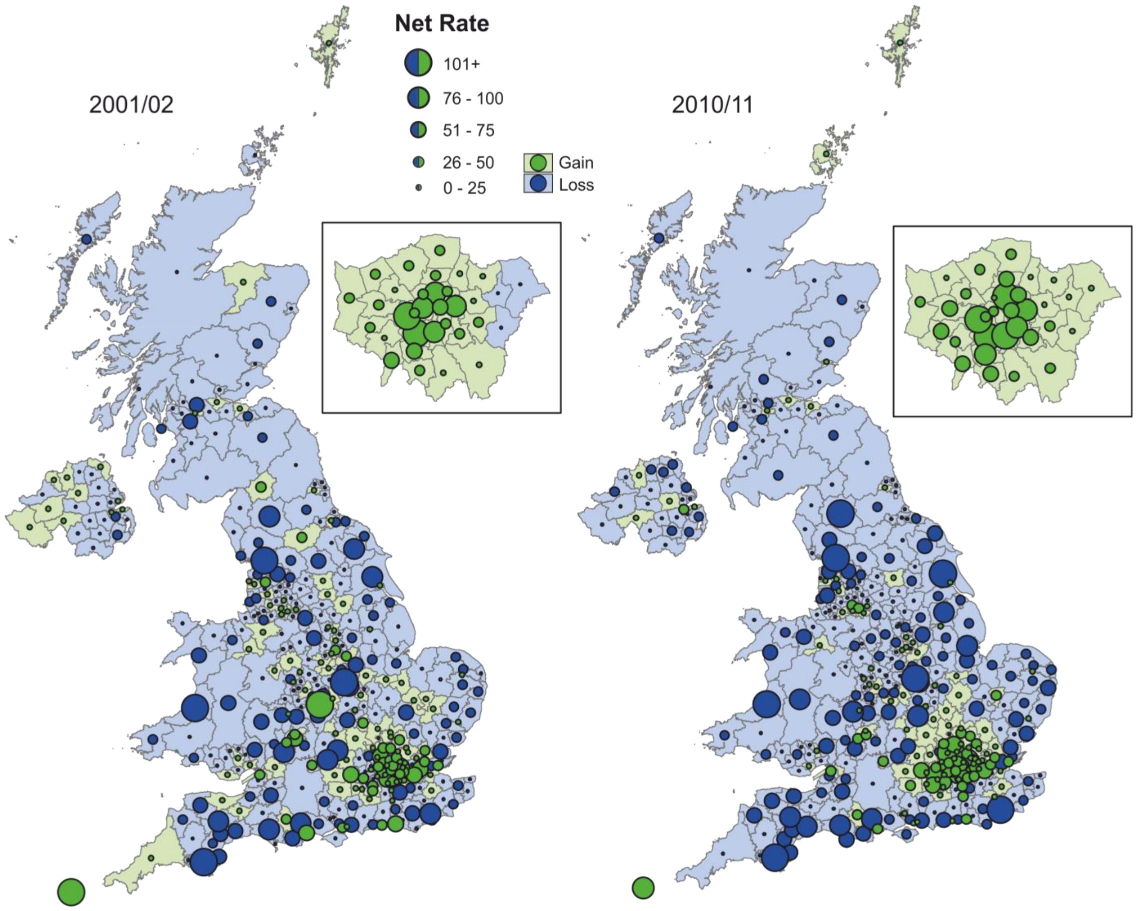


Figure 8.9: The net migration rate for persons aged 20-24 in each LAD, 2001/02 and 2010/11

It is not just university towns which display high out-migration rates for 20-24 year olds. The majority of LADs exhibiting the highest out-migration rates are not associated with high inflow for 15-19 year olds. Of the top 40 LADs in terms of out-migration rate for 20-24 year olds that are not associated with a HEI in 2001/02 (32 in total), 11 are situated within Coast and Country areas, 17 are in City Near areas while 4 are in City Rest areas. The pattern is similar in 2010/11: of the remaining 35 LADs not associated with the highest in migration for 15-19 year olds, 16 LADs are in Coast and Country areas 16 are in City Near areas and 3 are in City Rest areas. In summary, it appears that for the 20-24 age group, out-migration from university towns is a key trend, but in terms of the highest out-migration rates the move from less urban areas to more urban areas is the dominant pattern.

Two LADs exhibit notable shifts between 2001/02 and 2010/11 which exemplify the instability that can occur when institutional populations are concerned: Warwick (in the West Midlands), which hosts a university catering for around 23,000 students, falls from net gain of 105 per 1,000 to 17.3 per 1,000; and Richmondshire in North Yorkshire, home to the Catterick garrison which, at the time of the 2001 Census, housed 2,648 armed forces personnel living in a communal establishment (the largest total in the UK) moves from a net gain of 29.9 per 1,000 in 2001/02 to a net loss of 11.9 in 2010/11. Flowerdew and Salt (1979) report that the movement of large numbers of armed forces personnel between bases needs to be identified as it creates patterns that can run counter to the expected pattern of migration. In Warwick, inflow of 20-24 year olds remains relatively stable (2,959 in 2001/02 and 2,979 in 2010/11) whereas the number of out migrants increases from 1,961 to 2,772 over the same period. At the same time, the usual resident population of 20-24 year olds increases from 9,478 to 11,969, so the combination of a larger population and larger outflow has the net effect of decreasing the rate of migration gain. The number of postgraduate students enrolled at the university increased from 8,280 in 2001/02 to 10,195 (of which 4,425 were from overseas) in 2010/11 (HESA 2012) which may help to explain the increase in the size of the resident population.

8.7 Early career – age 20-29

The dominant pattern for migrants aged 25-29 is for moves between the 33 LADs that make up the London Core (which accounts for 18.5 and 19.4 per cent of all moves in this age group in 2001/02 and 2010/11 respectively). This can be seen as a progression from the dominant trend seen in the 20-24 age group, where people favour a move to London and also reflects the high propensities for ethnic minorities (as a proxy for international migrants) to move between London boroughs. This within London migration by ethnic minority groups is identified by Stillwell and Hussain (2010) who report that in 2001, 57 per cent of all migration by Black, 31 per cent of migration by Pakistani and other South Asian and 30 per cent of migration by Indian ethnic groups occurred between the 33 London boroughs. Vagras-Silva (2013a) reports that over the past decade, around half of international migrants to the UK were aged between 25-44, and in 2011, 32 per cent of all migrants arrived in London.

It can be hypothesised that many of the internal migrants within London aged 25-29 will move to take advantage of the relatively buoyant job market and within this age band are predominantly occupying rented accommodation which is associated with higher mobility than owner-occupied accommodation (Hamnett 1991). The first of these assertions is supported by the finding that a larger proportion of total moves occur within London in 2010/11 than 2001/02, which corresponds with evidence reported by Campos *et al.* (2011) that London exhibited fewer job losses and lower unemployment rates compared with other UK regions during the recession. The progression of this argument is that London retains educated and skilled migrants who might otherwise move elsewhere to work and live.

At age 25-29, the pattern of exodus from university towns seen at age 20-24 appears to be continuing. Of the 40 LADs that gained 15-19 year olds at the highest rate in 2001/02, 33 show a net loss of 25-29 year olds. In 2010/11, this pattern intensifies: of the 40 largest net gainers of 15-19 year olds, 37 become net losers of 25-29 year olds. This pattern is in line with Dennett (2010), who observes a fall in the intensity of out-migration from university towns between 1998/99 and 2007/08, attributing the change to an increased number of 20-24 year olds remaining in HE, suggesting an increased level of postgraduate study. If this is the case, the problems associated with studentification identified in Section 8.5 may be exacerbated by increasing demand for student housing stock from those aged 20-24.

In Section 8.1, it was reported that the proportion of moves occurring between City Core areas for 25-29 year olds are comparable to those seen for 20-24 year olds (3.8 per cent in 2001/02 and 5.0 per cent in 2010/11). However, moves into and out of City Core areas from/to City Rest, City Near and Coast and Country are relatively balanced: moves into City Core areas cumulatively account for 13.6 and 15.5 per cent at beginning and end of the decade, while moves out of City Core areas account for 16 and 15.7 per cent respectively). This can be seen in Figure 8.10 and Figure 8.11, where rates are far lower than they were at age 20-24. Between the beginning and end of the decade there is also a general impression of rates (both positive and negative) falling to produce a picture of lower migration activity for 25-29 year olds.

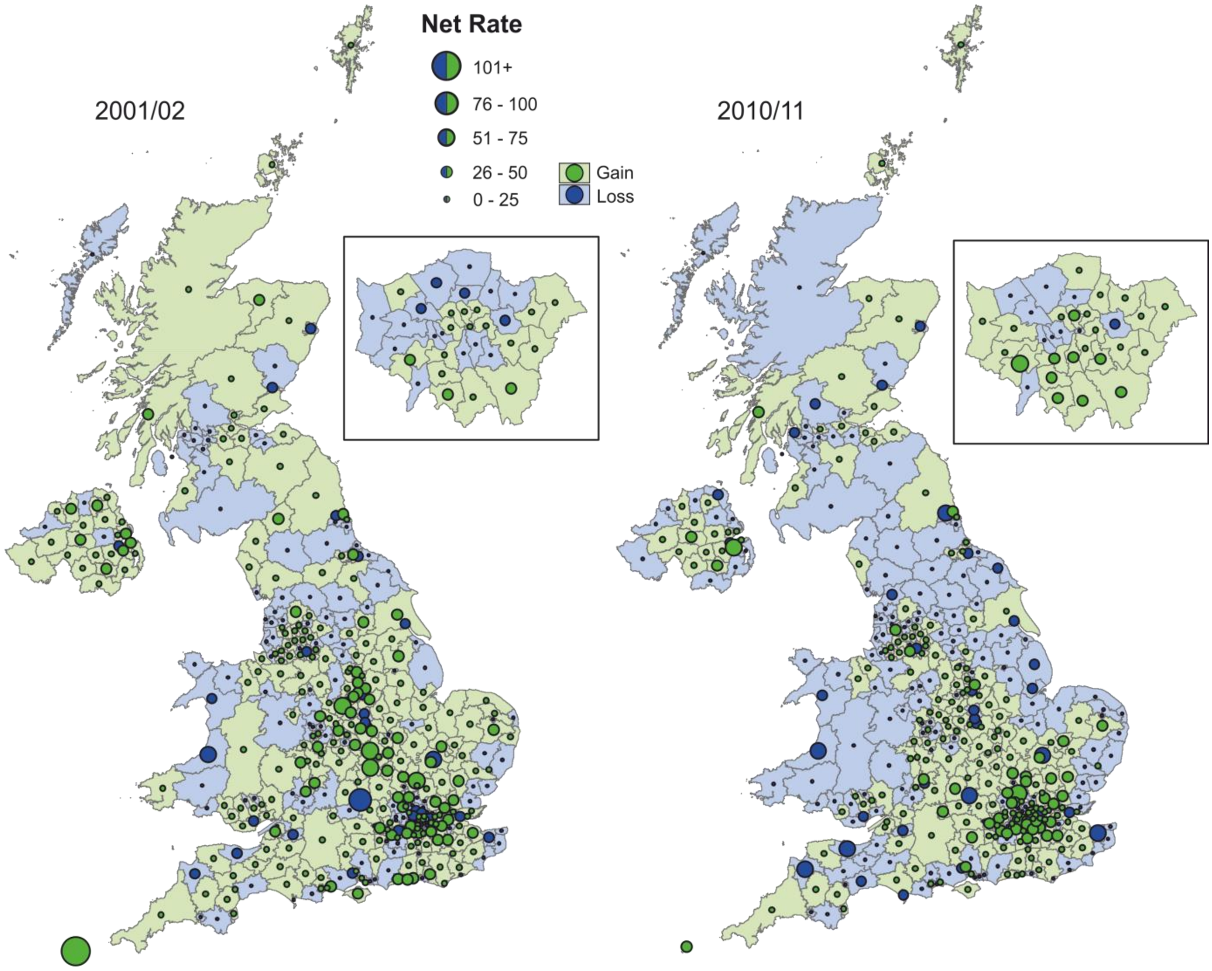


Figure 8.10: The net migration rate for persons aged 20-29 in each LAD, 2001/02 and 2010/11

The pattern in Northern Ireland and Scotland is slightly different to that seen in England and Wales where, in 2001/02, rural LADs appear to be exhibiting net gain of 25-29 year olds. This pattern slows substantially in 2010/11 but is still evident. Stockdale and Catney (2012, p.14) find that in the case of Northern Ireland, migrants in their 20s and 30s are more likely to participate in an urban to rural flow than their older counterparts. They suggest that this is down to unique features on Northern Ireland's geography and planning context in terms of *"settlement hierarchy, rural planning policy, and family farming tradition, alongside a strong tradition of rural 'self-build homes (intended for life) by those at the household formation stage of their lives"*. There is some evidence of this trend occurring at the 20-24 age group (Figure 8.9) but similar to the pattern for 25-29 year olds, diminishes in 2010/11.

In conclusion, at age 25-29, a large proportion of migrants move between the 33 London boroughs while at the same time the out-migration from university towns continues, a pattern that appears to intensify in 2010/11. This increase is perhaps evidence of students staying on to study a postgraduate course at university and making the transition from university to first job at a later age band than people had done previously.

8.8 Older working age – 45-59 and retirement – 60-74

The age bands 45-59 and 60-74 will be considered together in this section, first because of the similar net migration patterns that these age groups exhibit (Figures 8.11 and 8.12) and second because much of the literature relating to these age bands covers various aggregations of ages between 45 and 74. It is however recognised that different processes underpin the migration patterns seen at age 45-59 and age 60-74. The age band 45-65 is often used in migration literature to capture the pre-retirement age bracket, with 65 and over representing pensionable age. In 2011, the default retirement age of 65 was repealed in the UK, allowing people to work for as long as they like (or need to). Although these analyses are concerned with migration between 2001/02 and 2010/11, using the age band 60-74 allows for consideration of those taking early retirement and moving forward this offers a realistic age bracket for a population who are working longer. The state pension age (the earliest age that someone may receive their state entitlement) is being increased incrementally to 66 years by 2020 for both men and women (DWP 2013), but many people will work beyond this age. Sander *et al.* (2010, p.11) argue that the time spent in retirement is expanding, where *"for some, retirement is an extended process involving one or more intervening statuses outside the labour force before they retire fully".* They also point towards evidence that a large proportion of the workforce in the UK currently retire before the age of 65.

A scarcity of literature on the pattern of migration for those aged 45 to 64 in the UK leads to reliance on an Australian case study: Wulff *et al.* (2010, p.308) argue that because of their *"presumed relative stability,"* persons aged between 45 and 64 receive relatively little attention in migration literature, with the focus of studies on older households being directed at the retired or very elderly. Using a case study of Melbourne, they argue that *"the transition from parent to 'empty nester' can prompt the desire for making a residential move"* (p. 319). 'Empty nester' in this context refers to a household where grown-up children have moved away. Those they describe as empty nesters were found to move at higher rates than the general 45-64 year old population (with 45-54 year old empty nesters more likely to move than their 54-65 year old counterparts). This sentiment is echoed by Stockdale and MacLeod (2013), who summarise recent calls for mid-life to be taken as a new life course stage in migration research (they focus on 'pre- retirement' 50-64 year olds). They report that early retirement facilitates a move to rural areas, where *"overall, pre-retirement age migrants were motivated to move for actual or impending retirement and/or quality of life considerations"* (p.90). These migrants are empty nesters and some move to 'bridging jobs' (a final pre-retirement working phase) before retiring into rural communities. These pre-retirement migrants are largely self-employed which leads to business set up and consequently some job creation (as discussed in Chapter 6 in relation to the economic benefits of rural migration). Pre-retirement migrants also participate in the local community through voluntary work and through taking positions in local community groups. Stockdale and MacLeod (2013), argue that a move towards later retirement ages will have consequences for rural in-migration flows, both in terms of the benefits offered and also for the provision of services for a growing population who choose to age in rural areas (given the demographic trend for population ageing in the UK).

Figure 8.11 shows the net migration rates for 45-59 year olds in 2001/02 and 2010/11. Compared with the pattern seen at age 30-44, the most notable shift is to one of net loss from London and LADs in the South East, net loss from other urban areas and net gain in rural LADs. At age 45-59, the proportion of moves that occur between City Core areas drops substantially, and moves down the urban hierarchy begin to emerge (as seen in Figure 8.1). Moves from City Rest to City Near, City Near to Coast and Country and moves within these parts of the city region increase. The same is true of the pattern for 60-74 year olds (Figure 8.12) where the general trend of counterurbanisation continues. At both 45-59 and 60-74, the pattern in 2001/02 and 2010/11 is similar, but the magnitude of net gain in some rural LADs can be seen to decrease. This is especially noticeable at both ages for LADs on the south coast (Eastbourne, Rother, Arun) and the east coast (East Lindsey, North Kesteven). Generally speaking, the pattern of net migration seen in Figures 8.11 and 8.12 appears to be most stable of all age groups across the decade, but this masks some of the large relative changes seen in Figure 8.2.

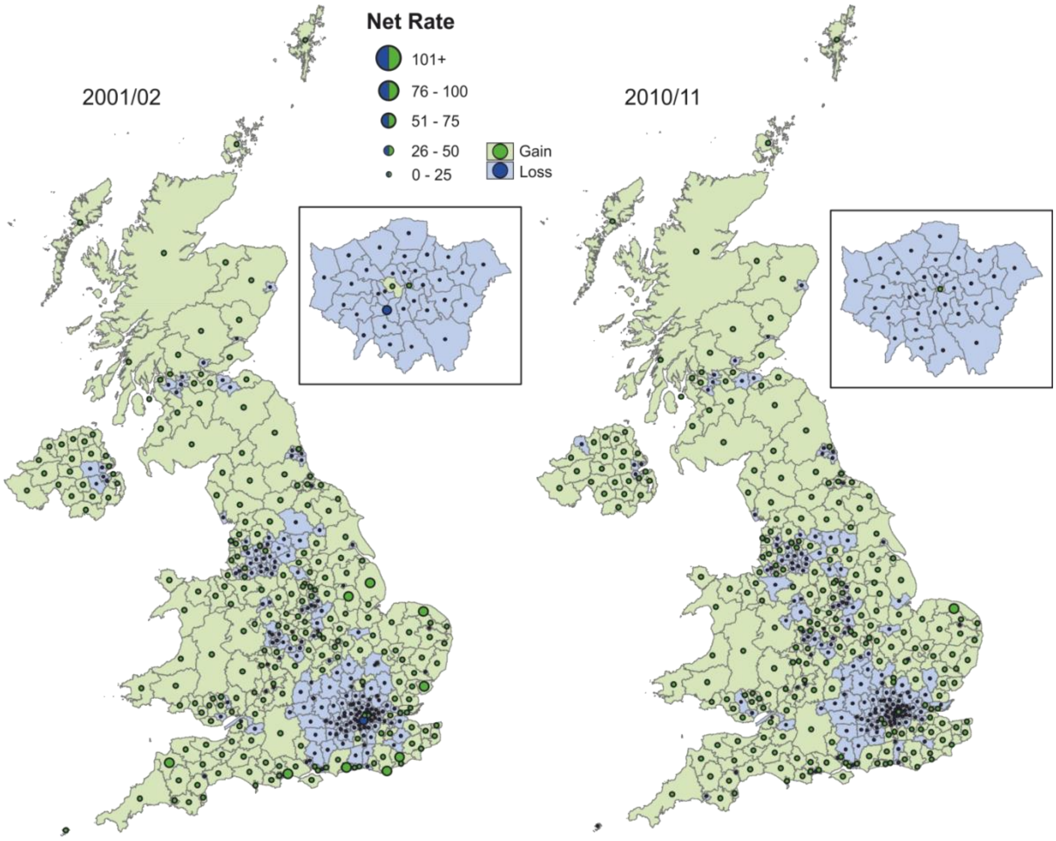


Figure 8.11: Net migration rate for persons aged 45-59 in each LAD, 2001/02 and 2010/11

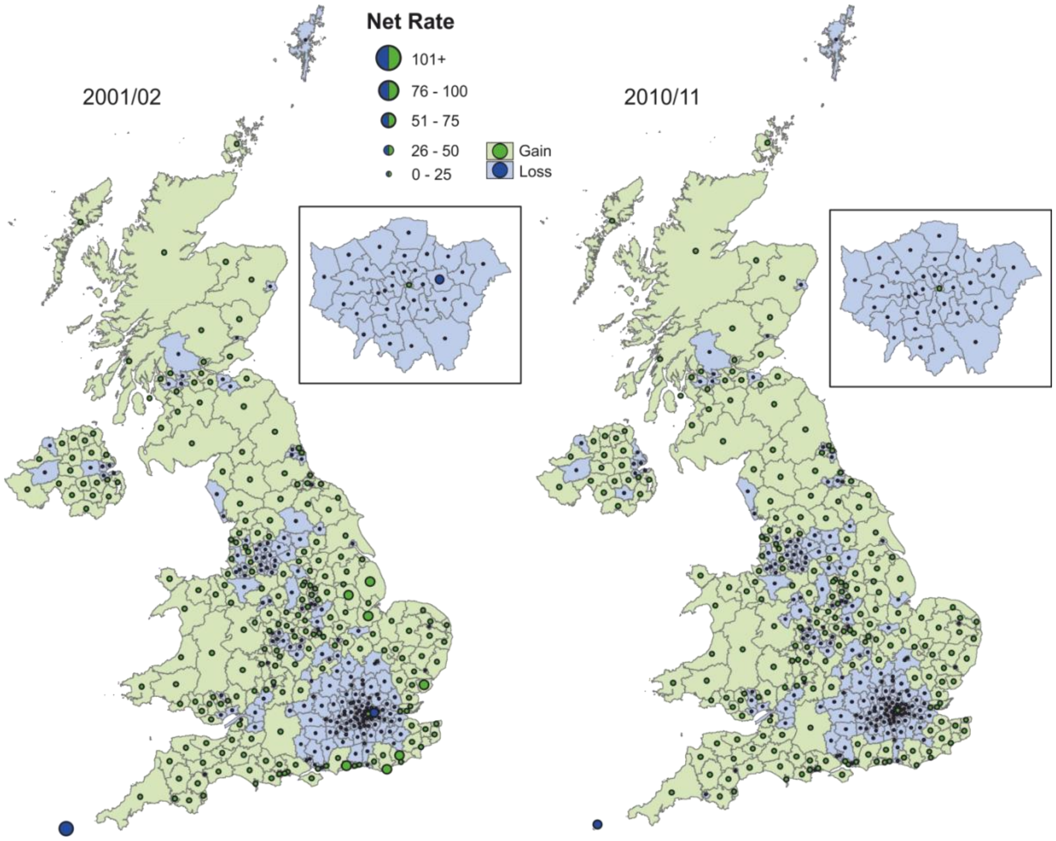


Figure 8.12: Net migration rate for persons aged 60-74 in each LAD, 2001/02 and 2010/11

At age 45-59, 17 of the 40 highest rates of net loss are from London boroughs in both 2001/02 and 2010/11 (the eight largest rates of loss are London boroughs in 2001/02). This pattern is more concentrated at age 60-74, 19 of the top 40 losses occur from London boroughs while of the remaining 21, all except Bristol are in the South East of England (in close proximity to London). In 2010/11 the pattern is similar, except that Manchester has entered the top 40 net rates of loss (at 39th). Within the top 40 LADs for net rate of gain, for both age bands and in both years, there are no UK cities and the majority are rural LADs in terms of population density.

These two age groups are very similar in their cluster associations. For the top 30 inflows, 45-59 year olds have 24 LADs in Cluster 2 in 2001/02 and 21 in Cluster 2 in 2010/11. Of the top 30 LAD inflows at age 60-74, 24 are in Cluster 2 in both years. The other dominant inflow cluster is Cluster 2 (10 and 13 LADs for 45-59 year olds and 10 and 11 LADs for 60-74 year olds in 2001/02 and 2010/11 respectively). Outflow Cluster 6 dominates both age groups (at age 45-59 accounting for 19 and 18 LADs in 2001/02 and 2010/11; at age 60-74 accounting for 20 and 19). Outflow Cluster 4 accounts for 14 LADs in both age bands in both years.

Compared with the other age groups, the UK population aged 60-74 has increased at the fastest rate in relative terms between 2001/02 and 2010/11 (from 7,858,887 in 2001/02 to 9,313,413 in 2010/11). It can be seen in Table 8.2 that there were 5,246 more migrants aged 60-74 at the end of the time series than at the beginning; however, the pattern seen in the rate maps in Figure 8.12 suggest a slow-down in the intensity of migration. In comparison, while the population aged 45-59 has also increased (from 11,305,797 in 2001/02 to 12,377,578 in 2010/11), the number of migrants has fallen by 3,837. It appears that the impact of an ageing population is facilitating the slowdown in the pattern of migration rate seen in these age groups, but this is likely to be masking the migration activity of people as they near retirement. Take Eastbourne on the South Coast, for example, which shows a decline in net rate of gain for both age groups between 2001/02 and 2010/11. In terms of net migration, the difference between 2001/02 and 2010/11 is relatively small (age 45-59 net migration falls from a gain of 466 to 376 persons, at age 60-74 falls from 418 to 353). The size of the resident population at these age groups increases, however, by a larger proportion: the population aged 45-59 increases from 16,138 to 18,309 while the population aged 60-74 increases from 14,227 to 16,808. What can be seen in 2010/11 in Eastbourne, as well as numerous other LADs, is the impact of a large post-war cohort reaching retirement age who are largely healthy and can make plans for long-term retirement (Lundholm 2012). This cohort effect of the post-war 'baby boom' generation has been identified by Smallwood (2012, pers. comm.) as a contributing factor to changing patterns of migration throughout the 2000s.

The impact that the elderly population boom in established retirement areas such as Eastbourne has on resource allocation is highlighted by the Select Committee on Public Service and Demographic Change (2013, p.48) who suggest that *"the middle-class, healthier older people on the south coast are very demanding"* which leads to resources being shifted from poorer areas which has the effect of widening inequality. When assessing the migration patterns for the 60-74 age group (and the 75 plus age group in the next section) Uren and Goldring (2007, p.40) stress the importance of understanding what impact they have on the size of the local elderly population as they are the biggest consumers of health and social services. They find that *"change in marital status, living arrangements and health are important triggers of elderly migration"*. Using the ONS Longitudinal Study (LS), they find that the most popular retirement destinations are seaside towns and rural coastal areas. A move to a communal establishment from a private residence for those aged 50 or over correlated strongly with a chronic illness in the same period.

In summary, migration patterns age 45-59 and age 60-74 are very similar, but in the older age group the number of migrants has increased substantially between 2001/02 and 2010/11. The increasing population at both age groups has the effect of reducing migration rates seen in Figures 8.10 and 8.11 across the time series, but the loss from urban LADs in preference for moves to rural and coastal LADs remains the key trend for both these age groups.

8.9 Old age – 75 and older

Raymer *et al.* (2007, p.892) identify two types of elderly migration flows: young and relatively healthy retirees moving from urban to amenity areas and *"older retirees who have been affected by bereavement or decline in health status returning to places where they previously lived"*. The second of these moves, elderly who migrate for support related reasons, is addressed by Glaser and Grundy (1998, p.337) who report that *"poor health is positively associated with the greater likelihood of changes in both living* *arrangements and address among men and women over the age of 65, suggesting that a substantial proportion of these moves are likely to be support-related".* A preference for long-distance moves for both amenity and support-related reasons are corroborated by the mean and median migration distances reported in Figure 8.3, which, at age 75 plus are the highest for any age group (over 200 km and 170 km respectively). Burholt (1999), in a study of rural Welsh LADs, finds that migration for low-level assistance (moving to live close to relatives) was associated with long distance moves, while migration to facilitate a high level of assistance (to move in with kin, to sheltered accommodation or to an institution) was associated with short-distance moves. In terms of long-distance moves for amenity, Fleming (2005, p.3) in a study of remote areas in Scotland, reports that results from the 2001 Census show *"those moving into remote areas from the rest of the UK were substantially older than those who moved to these areas from other parts of Scotland".*

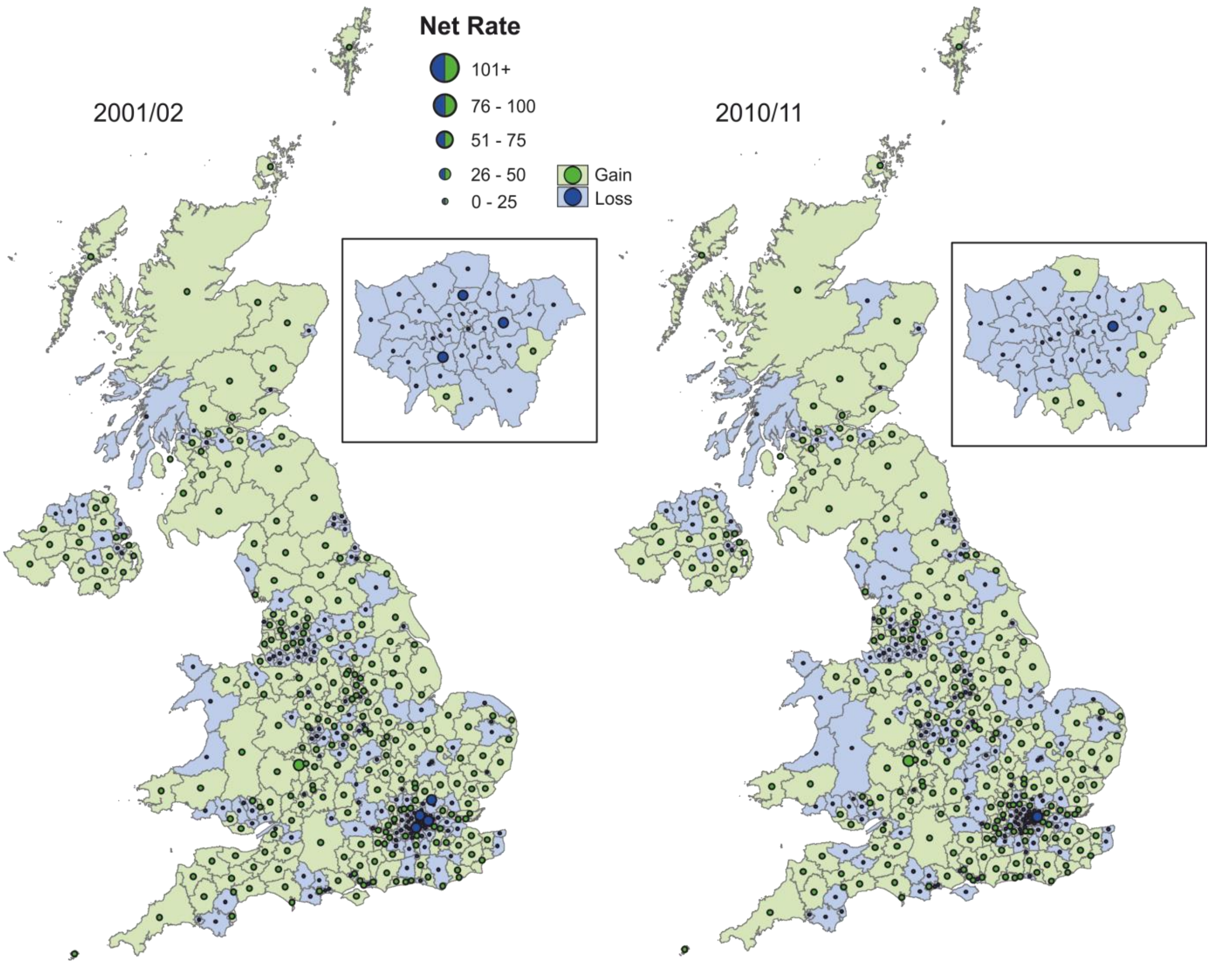


Figure 8.13: The net migration rate for persons aged 75 and over in each LAD, 2001/02 and 2010/11

Certainly net gain in rural Scottish LADs in both 2001/02 and 2010/11 seen in Figure 8.13 supports this.

The patterns of net rates seen in 2001/02 and 2010/11 appear to be very stable at age 75 plus (Figure 8.13). Of the top 40 rates of net loss, in 2001/02, 20 are London boroughs and six are City Core LADs (Bristol, Glasgow, Manchester, Belfast, Newcastle, Liverpool) whilst in 2010/11, 21 are London boroughs and Leeds becomes the seventh City Core LAD. In terms of net gain, in both 2001/02 and 2010/11 no UK cities appear and very few LADs containing larger towns, with the majority gaining in both years being rural LADs.

8.10 Sex ratios

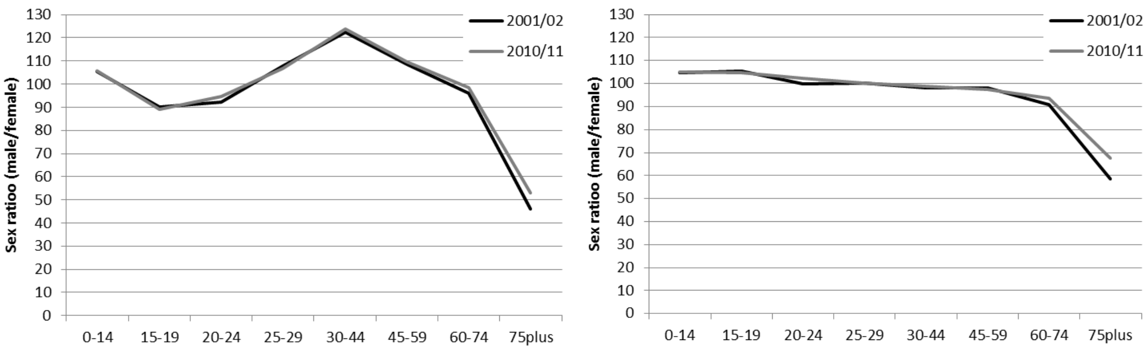
In this section, the sex ratio of in and out-migration for each LAD is assessed in order to identify areas where the ratio of male to female migrants is especially large. The ratio calculation consists of:

$$100 \times \frac{Number\ of\ male\ migrants}{Number\ of\ female\ migrants}$$

where the expression returns the number of male migrants for every 100 female migrants and is calculated for each LAD in the UK. A value of 100 means that the male migration flow is exactly the same as the female migration flow, while a value of over 100 means the male flow is larger than the female flow.

The national sex ration for all UK migration in 2001/02 and 2010/11 is shown in figure 8.14a, alongside the sex ratio for the UK population at each broad age group. When all ages are considered cumulatively, the average sex ratio for migrants is 96.1 in 2001/02 and 97.6 in 2010/11 which reflects a slightly higher propensity for females than males to migrate. In comparison, the sex ratio of the UK population is 94.5 in 2001/02 and 96.2 in 2010/11 which reflects that overall, there are more females than males. Age 30- 44 stands out in Figure 8.14a as the migration sex ratio is just over 120 (i.e. 120 males for every 100 females migrating). At age 25-29 and 45-59, the propensity of males to move is also, on average, higher than that for females. The opposite is true at age 15-19, and age 20-24 where the number of female migrants exceeds the number of male migrants. From age 45-59 onwards, the sex ratio falls considerably, while at age 75 plus the ratio falls to 46.1 in 2001/02 and 53.0 in 2010/11. This pattern is largely due to the life expectancy of females being greater than that of males (as seen in Figure

8.14b, where the ratio is 58.6 and 67.5 in 2001/02 and 2010/11 respectively) and reflects the assertion of Raymer *et al.* (2007) that bereavement is a factor in older age that can trigger a migration event.



a. Migrant sex ratio   b. Population sex ratio

Figure 8.14: Sex ratio for all moves at the UK national level (a) and sex ratio for the UK population (b), 2001/02 and 2010/11

Two values stand out in Table 8.3: the sex ratio for in-migration for 15-19 year olds in 2001/02 has a standard deviation of 87.59 (although the mean is close to 100 at 104.7), while the sex ratio for in-migration in 2010/11 at the 15-19 age band is 86.13 (mean = 103.9). The lowest value at age 15-19 in both years is exhibited by the City of London (10.94 in 2001/02 and 25.88 in 2010/11), which can be accounted for by very small numbers (three males compared with 24 females in 2001/02 and seven compared with 26 in 2010/11). City of London consistently experiences large and fluctuating sex ratios: at age 60-74 for example the ratio is 551 in 2001/02 which results from an inflow of 24 males and just four females and was a LAD that did not fit into the cluster analysis carried out in the previous chapter. The largest sex ratio at age 15-19 in both 2001/02 and 2010/11 is reported in Richmondshire (1,502.61 and 1,505.26) where the presence of a large armed forces base has the effect of skewing sex ratios. This armed forces effect on sex ratios in the 2001 Census results is documented by ONS (2004), who report that a large (predominantly male) armed forces presence at age groups 15-19 to 35-39 has a substantial impact on sex ratios at the local level. At age 15-19 in 2001/02, Richmondshire received 456 male migrants compared with only 30 females while in 2010/11 this was 327 and 21 respectively. Aside from the City of London and Richmondshire anomalies, Blaby, North Dorset (with the seventh largest armed forces population living in a communal establishment in England of 1,237 at the time of the 2001 Census), Hart (armed forces population of 1,106) and Harrogate (third largest armed forces population in England of 1,656) all have an inflow sex ratio of over 400, while Fermanagh and Merthyr Tydfil have low ratios (under 40). The LADs that are not home to armed forces bases are relatively peripheral with small migration flows in the 15-19 age group.

Table 8.3 provides an overview of the sex ratio patterns at each age in 2001/02 and 2010/11. For both in-migration and out-migration, the minimum, maximum and mean sex ratios are reported alongside the standard deviation where all 406 LADs are considered. The minimum and maximum ratios and standard deviation show the variance within each age group when all LADs are considered, so a high standard deviation shows a wide range of values within an age group and merits an investigation of these outliers. A high or low mean (either side of 100) suggest that there is a dominance of males or females in that particular subset of the data.

Richmondshire also accounts for the largest sex ratios in the 20-24 year old group for in and out-migration in both years, where it exhibits sex ratios of between 303 and 454. The remaining high ratios in the age 20-24 age band are rural or peripheral (Moray, Bridgend, West Oxfordshire) with relatively small migration flows. The same is true of the lowest sex ratios, except that the City of London consistently exhibits low sex ratios for out-migration (55.02 in 2001/02 and 57.61 in 2010/11).

The age groups 0-14 and 30-44 have relatively consistent sex ratios. 0-14 year olds have a standard deviation of less than 13 for both in-migration and out-migration in both years, while 30-44 year olds have similar outflow values but higher inflow values (standard deviation of over 16). The City of London accounts for the highest inflow ratio at age 30-44 (197.11 in 2010/11) where there were 170 males compared with 86 females. The LAD of Mendip accounts for the largest inflow sex ratios at age 0-14 (157.24 in 201/02 and 158.98 in 2010/11) which, in 2001/02 had 680 male in migrants compared with 433 female in migrants.

At age 45-59, the standard deviation for the in-migration sex ratio is lower than the out-migration ratio (16 compared with over 20). City of London again accounts for the largest sex ratios in both years (362.83 and 420.57 in 2001/02 and 2010/11) while the lowest ratios are in rural Northern Ireland (Magherafelt with a ratio of 46.70 and

Table 8.3: Sex ratio minimum, maximum, mean and standard deviation for each age band, inflow and outflow in 2001/02 and 2010/11

Inflow/ Outflow	Year	Age Band	Minimum	Mean	Maximum	Std. Deviation
In	2001/02	0-14	70.029	106.349	157.242	12.493
In	2001/02	15-19	10.944	104.657	1502.613	87.585
In	2001/02	20-24	44.165	92.88	303.168	25.792
In	2001/02	25-29	60.017	105.764	277.628	19.84
In	2001/02	30-44	68.616	120.276	175.104	16.087
In	2001/02	45-59	74.835	111.552	182.235	16.512
In	2001/02	60-74	43.8	98.25	551.635	31.702
In	2001/02	75 plus	0	47.906	185.439	17.612
Out	2001/02	0-14	69.132	106.3	153.619	12.854
Out	2001/02	15-19	27.892	89.931	194.51	17.855
Out	2001/02	20-24	55.017	96.545	235.13	18.779
Out	2001/02	25-29	63.458	108.853	222.871	16.762
Out	2001/02	30-44	79.979	120.632	162.103	12.322
Out	2001/02	45-59	46.699	110.914	362.83	22.195
Out	2001/02	60-74	29.663	98.489	210.381	20.705
Out	2001/02	75 plus	11.589	48.227	142.166	14.008
In	2010/11	0-14	65.62	106.284	158.975	12.381
In	2010/11	15-19	25.885	103.88	1505.285	86.125
In	2010/11	20-24	41.326	94.021	454.495	28.559
In	2010/11	25-29	56.74	106.718	267.909	19.473
In	2010/11	30-44	66.936	120.369	197.112	16.741
In	2010/11	45-59	73.107	111.277	193.294	16.932
In	2010/11	60-74	46.055	100.89	469.494	29.154
In	2010/11	75 plus	0	55.538	219.016	20.44
Out	2010/11	0-14	67.932	106.303	156.703	12.94
Out	2010/11	15-19	31.829	89.107	168.858	17.022
Out	2010/11	20-24	54.177	98.019	353.176	21.373
Out	2010/11	25-29	60.171	110.639	236.021	19.938
Out	2010/11	30-44	79.089	121.615	165.153	13.171
Out	2010/11	45-59	48.216	111.298	420.566	24.202
Out	2010/11	60-74	34.027	100.822	223.391	21.03
Out	2010/11	75 plus	14.94	55.439	137.826	16.073

Dungannon with a ratio of 53.37). Apart from these two LADs, the lowest outflow ratios are over 70.

At age 25-29, the standard deviation across inflow and outflow sex ratios in both years is below 20. The lowest values are around 60 but the majority are over 70. The highest in-migration ratios are rural English LADs (West Somerset, Rutland and West Devon, all with a ratio of over 180) while the largest outflow ratios (over 150) are similarly rural and exhibit small flows (Isle of Wight, Fylde, North Norfolk).

At age 60-74, the pattern of rural LADs showing the highest in-migration sex ratios changes somewhat: a number of London boroughs exhibit the highest in-migration sex ratios (City of London, Tower Hamlets, Hounslow) but the rural pattern remains for out-migration.

This section has shown that while the UK average migration sex ratio exhibited in the dataset is close to 100, some substantial differences do exist in some LADs. These LADs are mainly rural and exhibit smaller numbers of migrants which highlights the problem with small numbers inherent within migration datasets. Richmondshire and other LADs containing large populations of armed forces personnel living in communal establishments consistently showed high migration sex ratios for 15-24 year olds.

8.11 Conclusion

This chapter has provided an overview of the pattern of migration at the beginning and end of the time series broken down by age, while sex ratios have been used to identify LADs that exhibit patterns which are not consistent with the average UK pattern within each age band. The various analysis techniques and city region framework used for the aggregate migration dataset have been employed for the disaggregated dataset, which reveals varied patterns by age.

Using a city region typology, the general trend for an increase in the proportion of migration which involves a move up the urban hierarchy is true at all ages, except the move between City Near and City Rest areas which decline for ages 20-44. City Core to City Core moves account for a greater proportion of migration in 2010/11 than 2001/02 at all ages up to 59, where it declines slightly at age 60 plus. The proportion of total migration that involves a move down the urban hierarchy is generally in decline, except at age 60 plus where moves from City Rest to City Near areas show a slight increase in their share of total migration.

Using a set of migration indicators, it was found that migration intensity, mean and median migration distances fell at all age groups between 2001/02 and 2010/11. Connectivity increased for 20-24 and 25-29 year olds but declined at all other ages. Inequality increased at all ages, especially ages 20-24 and 30-44. Migration efficiency fell at younger and older ages but rose at ages 15-29, while the net migration rate fell at all ages except for 20-24 year olds.

When net migration rates were compared in 2001/02 and 2010/11 for all LADs at each age group, a strong relationship between the pattern exhibited by 0-14 and 30-44 age groups was found. Similarly, the patterns of net migration for 45-59 and 60-74 year olds were found to be much the same. This was supported by the similarity in the clustering of LADs for both in-migration and out-migration. A growing population at age 45-74 leads to an apparent slow-down in the net migration rate seen in most LADs, even though the number of migrants aged 60-74 increased by 5,246 (in the city region framework) between 2001/02 and 2010/11. LADs with HEIs dominated the high in-migration rates for 15-19 year olds, a pattern supported by data from HESA. The exodus from university towns was strong at age 20-24 but arguably stronger at age 25-29, suggesting that students are delaying their post- university migration decision. All of the patterns seen in the net migration rate maps are supported by the literature, adding strength to the robustness of the dataset when disaggregated by age.

Finally, sex ratios were used to address the pattern within each age group. While the national level sex ratio was almost 100 in each year, there was some considerable variation at the local level. Many of the largest and smallest sex ratios occurred in LADs with small migration flows and were predominantly rural (or less urban). The largest sex ratios can be attributed to the presence of armed forces bases (particularly in Richmondshire) which impact the ratio reported at age 15-29.

Chapter 9
Conclusions and further research

9.1 Introduction

The process of migration has long been the focus of attention by researchers who attempt to estimate, explain and predict patterns of movement undertaken by the population. Migration is the single largest contributor to the change in size and composition of population at subnational geographies in the UK (where it is acts alongside the natural change components of births and deaths), so a comprehensive and accurate understanding of the process is essential for delivering effective resource allocation and the formulation of policy. Migrants influence, and are influenced by, housing markets and job markets; they use local services, shop in certain areas and contribute to communities in ways that are both quantifiable and unquantifiable. Young people migrate for work and education; couples migrate to raise a family while the elderly migrate for amenity reasons or a requirement for social care. Immigrants or in-migrants, like the existing usually resident population, require schools, hospitals, refuse collection and policing. All of these processes are linked and all have far reaching social and economic implications. Migration is, however, the most difficult component of change to measure and has been the focus of considerable attention in recent times, both in terms of scrutiny of data and methodological improvement by the NSAs. This attention has been formalised through the Migration Statistics Improvement Programme (MSIP) and the Beyond 2011 programme, both led by ONS and involving the other two NSAs: NRS and NISRA. The first of these programmes provides the vehicle for a number of recent methodological improvements (primarily for international immigration) while the latter is an on-going process geared towards the provision of robust statistics in light of uncertainty over the provision of a census after 2011.

The work undertaken in this thesis coincides conveniently with the emerging debate over improving migration statistics in the UK, drawing on the 2001 Census and other more timely data, summarising methodological improvements made over the past few years as part of the MSIP and presenting a consistent UK wide approach which is growing in importance, as the NSAs push to increase consistency in their use of data, reporting of methodology and generation of statistical outputs. This thesis contributes to the understanding of UK migration at the subnational level by presenting and analysing results based on the best available data and an innovative estimation approach. This estimation and the results of the analyses undertaken are summarised in the next section (9.2) alongside the aim and objectives which were set out in Chapter 1. Section 9.3 summarises a set of policy reccommendations which arise from the work undertaken in this thesis, Section 9.4 presents some ideas for future work, while some final concluding remarks are offered in Section 9.5.

9.2 Summary of research findings

This thesis set out with the aim to produce a comprehensive and consistent database of migration flows for the entire UK at the subnational level, disaggregated by origin and destination and by age and sex which can be used subsequently to analyse migration intensities and patterns and monitor migration change. This aim has been successfully achieved through the implementation of six objectives. The remainder of this section discusses each objective in turn, summarising the research findings from each relevant thesis chapter.

Objective 1: to highlight the need for consistent UK wide subnational migration statistics and review the substantial literature that deals with determinants of migration, data estimation and visualisation;

The need for consistent UK wide migration statistics is highlighted in Chapter 2, where the problems of inconsistency, accuracy and availability in migration data are summarised. The key issue identified is that migration statistics are not consistent because they are produced by three different NSAs, who rely on different data, use different methods and produce different outputs. What results is a data landscape which treats the UK's constituent countries as separate entities, with little cohesion, consistency or focus on the whole UK picture. The review provided in Chapter 2 highlights the need for consistency in methodology and output for migration statistics in the UK, and provides a review of the substantial literature which is concerned with migration. This review provides a basis for the work carried out in the rest of the thesis: literature pertaining to the estimation of inadequate data help in the formation of the methodology presented in Chapter 4 and Chapter 7; determinants of migration propensity help with the interpretation of results presented in Chapter 5, Chapter 6 and Chapter 8; and literature which measures, classifies and visualises migration data informs the choices made in the interrogation of the dataset in this thesis.

Objective 2: To comprehensively review the data and methods used in the estimation of subnational migration in the UK for each of the four home nations, highlighting where inconsistencies exist and data are missing

The review of data and methods is undertaken in Chapter 3 and draws on numerous methodology documents produced by the NSAs, alongside evidence gathered during discussions and meeting with statisticians at ONS, NRS and NISRA. Much of the methodology documentation is produced independently by each NSA, alongside their own data, so very few sources provide a comprehensive review of data available and the estimation methods involved from a UK wide perspective. Much of the documentation is not aimed at a user audience, or is missing key information including authorship and date of writing or publication. Thus, by collating a wide range of methodology documentation and clarifying inconsistencies, the data review presented in Chapter 3 adds to existing understanding as it can be seen as a single place from which to find information on data and outputs for all three NSAs. The data review is also an essential step in the creation of a UK wide dataset as it contributes to the understanding of what datasets are available, and where these data come from.

Objective 3: To combine and harmonise the available subnational migration data and estimate the missing information

Having developed a detailed understanding of the data that are available and their limitations, a substantial part of this thesis is devoted to the collation and harmonisation of available data, and the estimation of missing data. In Chapter 2, a Migration Cube is introduced to outline the data availability and estimation requirement. This Migration Cube is reproduced in Figure 9.1 and was adapted from Rees and Willekens (1986). The Migration Cube consists of three faces where known (or partially known) information is available and a full origin-destination-age-sex array (ODAS, shaded grey in Figure 9.1) which requires estimation. The Migration Cube serves to succinctly summarise the steps required in the estimation and harmonisation of migration data in the UK and is adapted from the Rees and Willekens (1986) cube with the addition of a sex variable alongside the age variable (meaning only one cube is required for visualising the data requirement).

The origin-destination face of the Migration Cube (OD) is tackled first in Chapter 4: an interaction matrix is used to visualise the data structure for the whole UK in more detail, and is first populated with available data for total in-migration, total out-migration and origin-destination migration interactions at the LAD scale.

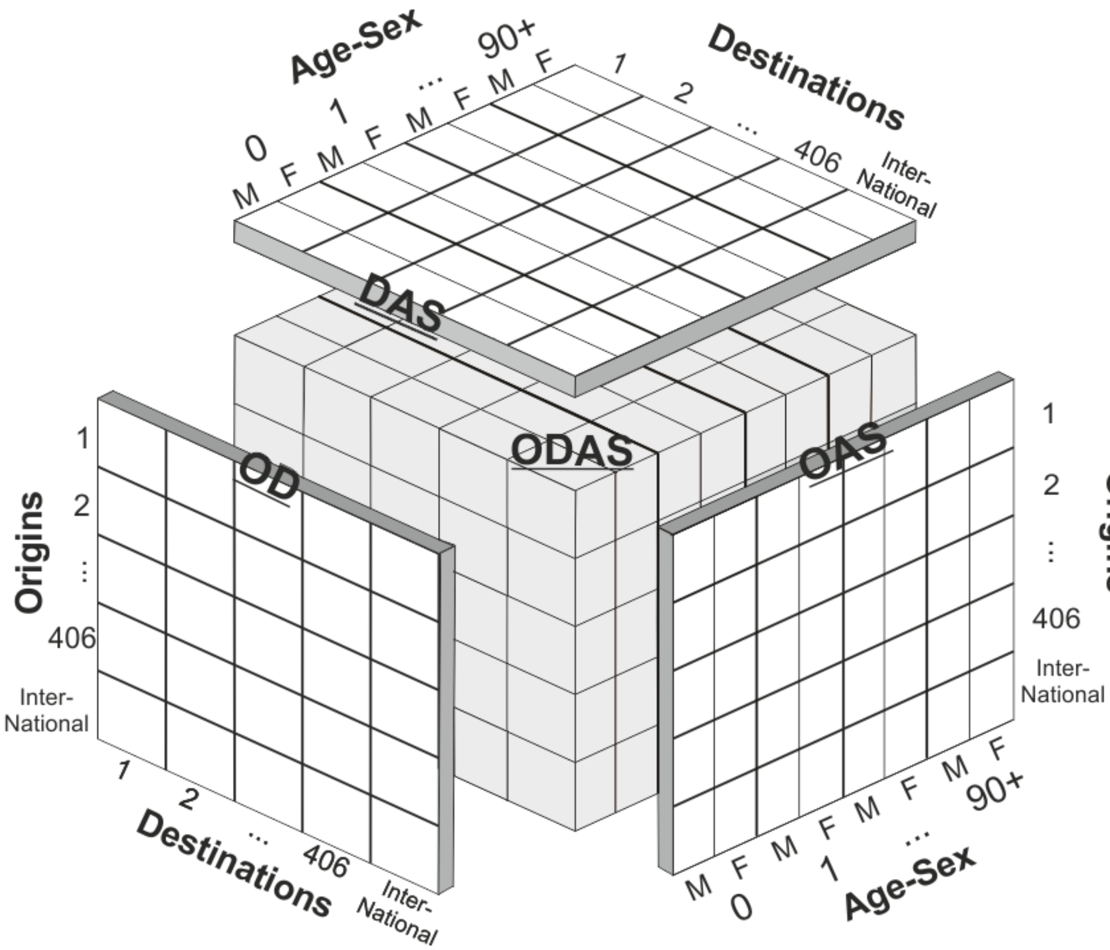


Figure 9.1: Migration Cube representing the combinations of origin (O), destination (D), age (A) and sex (S) data

Source: adapted from Rees and Willekens (1986)

The IPF routine used to estimate the missing data is outlined and justified in Chapter 4. While a number of methods can be used to estimate missing data (as described in Chapter 2), IPF was chosen as it utilises the available data (most marginal in-migration and out-migration totals are available in the interaction matrix) and can be implemented quickly and efficiently across the entire time series (2001/02 to 2010/11). The efficiency of the IPF routine means that it is successfully implemented in estimating the missing sections of the OD face and the full ODAS array of the Migration Cube. Internal migration flows in Northern Ireland and all cross-border migration flows are estimated at LAD scale using this IPF routine in Chapter 4. Using this now complete UK dataset, the IPF routine is extended in Chapter 7 to disaggregate the data by 19 age groups and sex for all LADs in the UK. The use of IPF for both the OD and ODAS estimates provides methodological consistency throughout the thesis. With the available data and the IPF code, the estimated results presented in this thesis are reproducible and easily adaptable (which is discussed further under the heading of Objective 6).

Where possible, validation of the estimated dataset has been undertaken. In Chapter 3, the data inputs (NHSCR, PRDS, SCHI) are validated against the 2001 Census. In Chapter 4, the IPF algorithm is implemented for known cross-border flows, between England and Wales, and the estimated results compared with data reported in the PRDS (as used by ONS) while in Chapter 7, a similar exercise is undertaken for the age disaggregated results. In all cases, the results were found to be robust.

The construction of this complete UK wide dataset of migration is one of the key original contributions of this thesis, and represents a marked improvement in the availability of data for measuring migration in the UK. It is the cross-border estimation in particular that highlights the innovation in using IPF, which draws on all available information, as such data have not been consistently available for the whole of the UK until now.

Objective 4: To build on existing techniques to develop a framework and set of measures for effectively presenting the results from the estimated migration dataset

The UK wide migration dataset contains a large amount of information: flow matrices have been estimated for 406 origins by 406 destinations, for each of 19 age bands and two sexes across a ten year time period. For this reason, a framework for analysis and a number of migration measures are used which aid in the effective interrogation of the time series dataset.

Measures of migration intensity and distance, connectivity, inequality and efficiency are used in Chapter 5 to provide summary measures of the aggregate dataset and again in Chapter 8 for the age disaggregated data. The aggregate data are also classified into urban/rural and north/south geographies in Chapter 5, based on established lines of investigation identified in the literature. Linking of origins and destinations undertaken in Chapter 6 increases the wealth of information that needs to be presented and understood, so a city region framework is used to group LADs into clusters which have a city at their core. This framework is selected as it provides a way of consistently analysing the whole of the UK. The city region framework is again successfully applied to the age disaggregated data presented in Chapter 8.

A spatial interaction model is used to test the validity of the city region framework, which was found to reliably represent the interactions occurring at LAD level. In the formulation of the age and sex disaggregation methodology, LADs are clustered using K-means classification modelled migration schedules (after Rogers and Castro) are applied to the data (Chapter 7). All of these techniques are used to provide a better understanding of the data structure, inform operational decisions in the formulation of methodology and are based on approaches used in the relevant literature.

Objective 5: To analyse the trends and patterns that exist in the UK wide subnational migration system between 2001/02 and 2010/11, using additional data where appropriate

With a substantial volume of data being produced in the estimation of the migration dataset, three chapters of this thesis are devoted to analysis of migration trends. The aggregate (national) patterns and net rates and balances at LAD level are reported in Chapter 5 for internal, cross-border and international migration. International immigration rises steadily between 2001/02 and 2006/07, with the sharpest increase attributable to A8 accession in 2003/04. It is higher at the end of the time series than at the beginning, while international emigration is at its lowest level in 2010/11. The pattern at subnational level reveals a reversal from net loss in 2001/02 to net gain in 2010/11 for a substantial number of LADs, a change that is particularly discernible in Scotland. GDP per capita exhibits a strong positive correlation with immigration, while emigration has a strong negative correlation with the unemployment rate in the previous year. For all years of the time series, there is also a strong negative correlation between internal and international migration at the LAD level, leading to the conclusion that LADs which gain international migrants lose internal migrants and *vice versa.*

Internal migration, the largest of all three migration flows, peaks mid-decade, in 2006/07, before falling back to a similar level as 2001/02 at the end of the decade, while the number of people migrating across the borders of the UK countries falls consistently through the decade. The subnational composition of these migration flows changes substantially over the decade however, as when net rates and balances at the LAD scale are reported, a quietening down in the level of net migration can be seen across the UK. The patterns of internal and cross-border migration are very different: a (declining) pattern of urban net loss and rural net gain summarises internal migration, while for cross-border migration, LADs in England are losing migrants while the main gains can be seen in Wales and Scotland. Internal and cross-border migration are combined for much of the remaining analysis, where the consistent UK wide dataset is interrogated. For all moves within the UK, a strong negative correlation with the unemployment rate exists, suggesting that when unemployment is high, migration is low and *vice versa.*

A set of migration indicators are adopted in Chapter 5 to quantify the UK migration dataset (where both moves within each country and cross-border are combined). These measures, defined in Bell *et al.* (2002) reveal that across the time series, the proportion of the total population who migrate (the aggregate net migration rate) declines year on year, while those who do migrate are travelling shorter distances. Fewer LADs in 2010/11 were connected by 10 or more migrants than they were in 2001/02. Migration across the system is becoming more unequal and increasingly spatially concentrated, while the efficiency of migration as a force for redistributing the population around the UK is in decline.

In Chapter 5, LADs have been classified as either urban or rural based on their population density, and the general pattern across the time series is one of net loss from urban areas and net gain in rural areas. However, between 2001/02 and 2010/11, net gain in rural areas is in decline, accounted for predominantly by a fall in urban to rural migration. The rural to urban flow remains more consistent. At the same time, urban to urban migration increases substantially, which is a process driving a reversal in north-south migration patterns. A pattern of (declining) net gain in the north of the UK from the south is evident up to 2005/06, at which point a reversal occurs and the south begins to gain from the north. This is a pattern which can be accounted for predominantly by an increase in migration from urban north to urban south. This process can be better understood when the dataset is broken down into a city region typology in Chapter 6. Here it is revealed that the proportion of total migration which occurs between City Core areas is increasing, with London being the predominant net gainer from other City Cores. As a net distributor of migrants as a whole however, London’s geographical influence declines across the decade. Moves between non-metropolitan (City Near and Coast and Country) areas reduced over the decade.

Chapter 8 presents analyses using the age disaggregated dataset. Overall, the trend for an increase in moves from less urban to more urban can be seen at all ages, as can the decline in moves from more urban to less urban areas. The aggregate increase in City Core to City Core migration is evident at all ages to 59, while it declines slightly at ages 60 plus. The net migration rate falls for all ages except the 20-44 age group, while the pattern of falling migration intensity and distance is true for all ages. The decline in connectivity seen in the aggregate data is true for all ages except 20-24 and 25-29, where it increases. Migration inequality rises at all ages and is especially prevalent at ages 20-24 and 30-44, while migration efficiency falls at most ages, except 15-29 where it rises. Some considerable variation exists in the sex ratio of migrants at the LAD level, with many of the largest differences being seen in rural areas. The largest sex ratios corresponded with areas that contained armed forces bases and related to migrants aged 15-29.

Based on these results, the first decade of the 2000s can be characterised as one where patterns of migration are changing substantially. A general slow-down in all types of migration is evident, while the process of counterurbanisation which characterised migration patterns over the past 50 years is in decline.

Objective 6: To provide a discussion of the work in the context of on-going methodological improvements and a changing data landscape within the UK

The introduction to this chapter touched on two programmes led by ONS: the Migration Statistics Improvement Program (MSIP) which came about following criticism of UK migration statistics and the Beyond 2011 Programme, a response to the continued need for comprehensive data on the UK population, where the catalyst is the debate around uncertainty of a 2021 Census being undertaken. The first of these is relevant to the work undertaken in this thesis as it represents a process in which administrative and other data are being explored and utilised. The main example of this is improvement to the distribution of international immigrants to LADs. The second highlights the need for access to new and improved data moving forward, along with the need for innovative strategies to collate and harmonise such data. At the heart of the Beyond 2011 programme is a recognition of the need to improve collaboration between the four countries of the UK when exploring alternatives to a traditional census and the need to harmonise UK outputs, not least because the provision of these collated outputs is required for Eurostat, the statistical reporting arm of the European Commission (ONS 2013a). With a push to improve migration statistics, recognition of a need for harmonised UK outputs, a requirement for exploitation of additional data and innovative approaches, the work presented in this thesis is well timed to contribute towards this requirement. The next section outlines ways in which the work presented in this thesis can be taken forward and fulfil future requirements.

9.3 Policy recommendations

This section outlines briefly some policy recommendations which have arisen from the work undertaken in this thesis. First and foremost is the recommendation that the three NSAs who produce migration statistics in the UK work together to develop a consistent methodology for producing internal migration statistics. The variances in methodology and data output which define the current situation mean that the single country of the UK does not produce consistent statistical output. The data review presented in Chapter 3 demonstrates how sub-groups of the population are accounted for differently (students and young males for example), making comparisons difficult. Furthermore, the documentation related to these methods is copious and not always easily accessible to the users of these migration statistics. The same can be said for international migration estimates. The second recommendation is that a consistent UK wide estimate of cross-border migration at the sub-national level is produced to provide completeness to the statistical reporting of migrants within the UK.

The internal and cross-border migration estimates could be made consistent by using the NHS register data differently. It was reported in Chapter 3 that NRS are investigating the use of an NHSCR extract which reports the postcode of patients (opposed to the current system of scaling up the SCHI to NHSCR estimates). Similarly, ONS are investigating the use of data from the NHS ‘Spine’ which would provide more timely data on an individual’s current address whenever they come into contact with an NHS service. Both of these options may provide a viable solution which improves the statistical reporting of migration in the UK, but a key (third) recommendation is that any improvement to methodology is carried out consistently for England, Wales, Scotland and Northern Ireland and which involves all three NSAs. A consistent UK wide database could prove to be more accurate, timely and efficient (cost effective) than the current system that is in place.

9.4 Future work

The main output from this thesis is a comprehensive dataset of migration for the UK which is disaggregated by age and sex between 2001/02 and 2010/11. These data have been analysed in three chapters of the thesis, however there is scope to use them for a variety of purposes and as such, validation, extension and dissemination of the data are a focus for future work.

The estimated time series presented in this thesis draws heavily on the structure of interaction data reported in the 2001 Census, which is updated year on year based on up to date marginal (in-migration and out-migration) totals derived from administrative data sources. Much of the updating of previous origin-destination interactions and age-sex migration schedules relies on advice in the literature and on checks made during this thesis that previous relationships and patterns to a large extent persist. The opportunity to benchmark these estimates against 2011 data is an essential way of validating the estimates, but will not be possible until the special migration statistics (or an equivalent dataset) are compiled and made available by ONS. As an extension of this, the method used in this thesis means that it could be readily applied to 2011 Census data, allowing for a continuation of the estimated UK wide migration dataset into the foreseeable future. In the context of the Beyond 2011 programme, this would present the opportunity to maintain a migration dataset which could be improved as additional administrative and other data sources become available.

The 2001/02 to 2010/11 estimates will be used in the update of ethnic projections in a project which builds on work undertaken as part of the ETHPOP database (Wohland *et al.* 2011). The follow-on project is part of a funding bid made to the ESRC Secondary Data Analysis Initiative (Phase 2) entitled *Evaluation, Revision and Extension of Ethnic Group Projections*. The estimated dataset will be offered to ONS for consideration as official or experimental statistics as the first release of a consistent UK wide dataset. The data will also be supplied to the Census Support section of the UK Data Service for dissemination to other researchers through the Web-Based Interface for Census Interaction Data (WICID).

9.5 Concluding remarks

This thesis has contributed to migration research in the UK by presenting a dataset which was not previously available: migration by origin, destination age and sex for all four countries which is consistent over a ten year period between 2001/02 and 2010/11. In so doing, it has highlighted the inconsistencies that exist in data and estimates produced by three different NSAs, and contributes to two relatively new priorities in UK statistical reporting: for improved collaboration between ONS, NRS and NISRA, and an increase in the consistency of statistical output. Analysis of this dataset reveals distinct changes in the pattern of UK migration which occurred over the first decade of the 21^{st} century. Further work with the dataset, improvements based on better access to administrative data sources and the potential to apply the methodology to 2011 Census data mean that there is considerable scope to develop the work presented in this thesis further.

The work presented in this thesis has contributed to international debates on internal migration in two ways. First, it emphasises the requirement for accurate, timely and consistent migration data and the impact that such data has on resource allocation and policy decisions. This requirement is not unique to the UK, and work undertaken in this thesis could usefully provide a theoretical background for studies of migration in most countries. Second, in implementing an IPF routine to estimate unknown migration flows by origin, destination, age and sex it provides a methodology which could be usefully applied to any spatial system where only marginal in and out flows are available. The method outlined in this thesis also effectively deals with data where some of these marginal totals are incomplete. Estimation of migration within Northern Ireland is analogous to the process required to estimate internal migration in many countries, while the method used to create cross border estimates (harmonising data from different NSAs and filling in the gaps) could be implemented to estimate international migration between countries

.

Bibliography

Aggarwal, C., Hinnenburg, A. and Keim, D. A. (2001) On the surprising behaviour of distance metrics in high dimensional space. *Lecture Notes in Computer Science,* **1973**, pp.420-434.

Aldenerfer, M. and Blashfield, R. (1984) *Cluster analysis.* California: Sage University Papers.

Andrews, M., Clark, K. and Whittaker, W. (2011) The determinants of regional migration in Great Britain: A duration approach. *Journal of the Royal Statistical Society. Series A: Statistics in Society,* **174**(1), pp.127-153.

Audit Commission. (2006) National duplicate registration initiative. [online]. Available from: http://www.audit-commission.gov.uk/SiteCollectionDocuments/AuditCommissionReports/NationalStudies/ndri2004report.pdf.

Barber, A. (2007) Planning for sustainable re-urbanisation: policy challenges and city centre housing in Birmingham. *Town Planning Review,* **78**(2), pp.179-202.

Barr, P. J. and Shuttleworth, I. (2012) Reporting address changes by migrants: The accuracy and timeliness of reports via health card registers. *Health & place,* **18**(3), pp.595-604.

Bates, J. and Bracken, I. (1982) Estimation of migration profiles in England and Wales. *Environment and Planning A,* **14**(7), pp.889-900.

Bates, J. and Bracken, I. (1987) Migration age profiles for local authority areas in England, 1971 - 1981. *Environment and Planning A,* **19**(4), pp.521-535.

Bell, M., Blake, M., Boyle, P., Duke-Williams, O., Rees, P., Stillwell, J. and Hugo, G. (2002) Cross-national comparison of internal migration: issues and measures. *Journal of the Royal Statistical Society Series a-Statistics in Society,* **165**(3), pp.435-464.

Bell, M., Charles-Edwards, E., Ueffing, P., Stillwell, J., Kupiszewski, M. and Kupiszewska, D. (2013) Comparing Internal Migration Intensities around the Globe. In *XVII IUSSP International Population Conference, 26-31 August, Busan*.

Bell, M. and Muhidin, S. (2009) Cross-national comparisons of internal migration. *Human development research paper 2009/30* [online]. Available from: http://hdr.undp.org/en/reports/global/hdr2009/papers/HDRP_2009_30.pdf.

Bentham, G. (1988) Migration and morbidity: implications for geographical studies of disease. *Social Science & Medicine*, **26**(1), pp.49-54.

Berry, H. (2000) Review Essay:Rural Sociology, Migration, and Community Change. *Rural Sociology,* **65**(4), pp.658-667.

Bishop, Y., Fienberg, S. and Holland, P. (1974) *Discrete multivariate:analysis: theory and practice.* Cambridge: MIT Press.

Bohara, A. K. and Krieg, R. G. (1996) A zero-inflated Poisson model of migration frequency. *International Regional Science Review*, **19**(3), pp.211-222.

Bosworth, G. (2006) Counterurbanisation and job creation: entrepreneurial in-migration and rural economic development. *Centre for Rural Economy Discussion Paper Series*, **4**.

Bosworth, G. (2010) Commercial counterurbanisation: an emerging force in rural economic development. *Environment and Planning A*, **42**(4), pp.966-981.

Boyle, P. (1993a) Modelling the relationship between tenure and migration in England and Wales. *Transactions of the Institute of British Geographers*, **18**(3), pp.359-376.

Boyle, P. (1993b) A three-dimensional approach to multistream migration modelling: local-level flows in Hereford and Worcester. *Environment and Planning A*, **25**(9), pp.1279-1293.

Boyle, P. and Feng, Z. (2002) A method for integrating the 1981 and 1991 British census interaction data. *Computers, Environment and Urban Systems*, **26**(2-3), pp.241-256.

Boyle, P., Norman, P. and Rees, P. (2002) Does migration exaggerate the relationship between deprivation and limiting long-term illness? A Scottish analysis. *Social Science & Medicine*, **55**(1), pp.21-31.

Boyle, P., Norman, P. and Rees, P. (2004) Changing places. Do changes in the relative deprivation of areas influence limiting long-term illness and mortality among non-migrant people living in non-deprived households? *Social Science & Medicine*, **58**(12), pp.2459-2471.

Bracken, I. and Bates, J. (1983) Analysis of gross migration profiles in England and Wales: some developments in classification. *Environment and Planning A*, **15**, pp.343-355.

Bromley, R., Tallon, A. and Roberts, A. (2007) New populations in the British city centre: Evidence of social change from the census and household surveys. *Geoforum*, **38**(1), pp.138-154.

Burholt, V. (1999) Testing a behavioural and a developmental model of migration: a reevaluation of migration patterns among the elderly and why older people move. *Environment and Planning A*, **31**(11), pp.2071-2088.

Bushin, N. (2009) Researching family migration decision-making: A children-in-families approach. *Population, Space and Place*, **15**(5), pp.429-443.

Butler, T. (2007) Re-urbanizing London Docklands: gentrification, suburbanization or new urbanism? *International Journal of Urban and Regional Research*, **31**(4), pp.759-781.

Buzar, S., Ogden, P., Hall, R., Haase, A., Kabisch, S. and Steinfiihrer, A. (2007) Splintering urban populations: emergent landscapes of reurbanisation in four European cities. *Urban Studies*, **44**(4), pp.651-677.

Cameron, G. and Muellbauer, J. (1998) The housing market and regional commuting and migration choices. *Scottish Journal of Political Economy*, **45**(4), pp.420-446.

Cameron, G., Muellbauer, J. and Murphy, A. (2005) Migration within England and Wales and the housing market. *Economic Outlook*, **29**(3), pp.9-19.

Campos, C., Dent, A., Fry, R. and Reid, A. (2011) Impact of the recession. *Regional Trends*, **43**, pp.1-69.

Carballo, M., Divino, J. J. and Zeric, D. (1998) Migration and health in the European Union. *Tropical Medicine & International Health*, **3**(12), pp.936-944.

Champion, T. (1987) Recent changes in the pace of population deconcentration in Britain. *Geoforum*, **18**(4), pp.379-401.

Champion, T. (1989a) *Counterurbanisation: the changing pace and nature of population deconcentration.* London: Edward Arnold.

Champion, T. (1989b) Internal migration and the spatial distribution of population. In Joshi, H., ed. *The Changing Population of Britain.* Cambridge: Basil Blackwell, pp.110-132.

Champion, T. (1989c) United Kingdom: population decentralisation as a cyclic phenomenon. In Champion, T., ed. *Counterurbanization: The changing pace and nature of population deconcentration.* London: Edward Arnold, pp.83-102.

Champion, T. (2001) Urbanization, suburbanization, counterurbanization and reurbanization. In Padison, R., ed. *Handbook of Urban Studies.* London: SAGE, pp.143-161.

Champion, T. (2005) Population movement within the UK. In Chappell, R., ed. *Focus on People and Migration.* Basingstoke: Macmillan, pp.92-114.

Champion, T. (2012) Testing the return migration element of the 'escalator region' model: an analysis of migration into and out of south-east England, 1966–2001. *Cambridge Journal of Regions, Economy and Society*, **5**(2), pp.255-270.

Champion, T., Bramley, G., Fotheringham, A. S., Macgill, J. and Rees, P. (2003) A migration modelling system to support government decision-making. In Geertman, S. and Stillwell, J., eds. *Planning support systems in practice.* Berlin: Springer, pp.269-290.

Champion, T., Coombes, M. and Gordon, I. (2013) How far do England's second-order cities emulate London as human-capital 'escalators'? *Population, Space and Place*, **Early view**, pp.n/a-n/a.

Champion, T., Fotheringham, A. S., Rees, P., Boyle, P. and Stillwell, J. (1998) The determinants of migration flows in England: a review of existing data and evidence. [online]. Available from: http://www.geog.leeds.ac.uk/publications/DeterminantsOfMigration/report.pdf.

Champion, T. and Townsend, A. (1994) *Contemporary Britain: A geographical perspective.* New York: Routledge.

Chilton, R. and Poet, R. (1973) An entropy maximising approach to the recovery of detailed migration patterns from aggregate census data. *Environment and Planning*, **5**(1), pp.135-146.

Cohen, M. (2008) Raking. In Lavrakas, P., ed. *Encyclopedia of survey research methods.* Thousand Oaks, CA: Sage Publications, pp.672-674.

Coleman, D. (2008) The demographic effects of international migration in Europe. *Oxford Review of Economic Policy*, **24**(3), pp.452-476.

Compton, P. (1992) Migration trends for Northern Ireland: links with Great Britain. In Stillwell, J., Rees, P. and Boden, P., eds. *Migration processes & patterns V2. Population redistribution in the United Kingdom.* London: Belhaven Press, pp.81-99.

Congdon, P. (2010) Random-effects models for migration attractivity and retentivity: a Bayesian methodology. *Journal of the Royal Statistical Society: Series A (Statistics in Society),* **173**(4), pp.755-774.

Coombes, M. and Charlton, J. (1992) Flows to and from London: a decade of change? In Stillwell, J., Rees, P. and Boden, P., eds. *Migration processes & patterns V2. Population redistribution in the United Kingdom.* London: Belhaven Press, pp.56-77.

Couch, C., Fowles, S. and Karecha, J. (2009) Reurbanization and housing markets in the central and inner areas of Liverpool. *Planning Practice & Research,* **24**(3), pp.321-341.

Cross, D. (1990) *Counterurbanization in England and Wales.* Gower Publishing Ltd: Aldershot.

Daras, K., Stillwell, J. and Bell, M. (2013) The effect of Initial Random Aggregation (IRA) on the analysis and modelling of internal migration. In *GISRUK, 3 - 5 April, Liverpool*.

Deming, W. and Stephan, F. (1940) On a least squares adjustment of a sampled frequency table when the expected marginal totals are known. *The Annals of Mathematical Statistics,* **11**(4), pp.427-444.

Dennett, A. (2010) *Understanding internal migration in Britain at the Start of the 21st century*. Doctoral thesis, University of Leeds.

Dennett, A., Duke-Williams, O. and Stillwell, J. (2007) *Interaction data sets in the UK: an audit*. Working Paper 07/05, University of Leeds.

Dennett, A. and Stillwell, J. (2008) *Internal migration in Great Britain: a district level analysis using 2001 Census data.* Working Paper 01/08: School of Geography, University of Leeds.

Dennett, A. and Stillwell, J. (2010) Internal migration in Britain, 2000-01, examined through an area classification framework. *Population Space and Place,* **16**(6), pp.517-538.

Dennett, A. and Stillwell, J. (2011) A new area classification for understanding internal migration in Britain. *Population Trends,* **145**, pp.1-26.

Dennett, A. and Wilson, A. (2013) A multilevel spatial interaction modelling framework for estimating interregional migration in Europe. *Environment and Planning A,* **45**(6), pp.1491-1507.

Denteneer, D. and Verbeek, A. (1985) A fast algorithm for iterative proportional fitting in log-linear models. *Computational Statistics & Data Analysis,* **3**(1), pp.251-264.

Diansheng, G. (2009) Flow mapping and multivariate visualization of large spatial interaction data. *Visualization and Computer Graphics,* **15**(6), pp.1041-1048.

Dixon, S. (2003) Migration within Britain for job reasons. *Labour Market Trends,* **111**(4), pp.191-202.

Dobson, J., Latham, A. and Salt, J. (2009) On the move? Labour migration in times of recession. *Policy Network Paper* [online]. Available from: http://www3.geog.ucl.ac.uk/research/mobility-identity-and-security/migration-research-unit/pdfs/on_the_move.pdf.

Dobson, J. and Stillwell, J. (2000) Changing home, changing school: towards a research agenda on child migration. *Area*, **32**(4), pp.395-401.

Dorling, D. (2007) The north-south divide - where is the line? [online]. Available from: http://www.sasi.group.shef.ac.uk/maps/nsdivide/index.html.

Dorling, D. and Rees, P. (2003) A nation still dividing: the British census and social polarisation 1971 - 2001. *Environment and Planning A*, **35**(7), pp.1287-1313.

Duke-Williams, O. (2009) The geographies of student migration in the UK. *Environment and Planning A*, **41**(8), pp.1826-1848.

DukeWilliams, O. and Blake, A. (2003) Database fusion for the comparative study of migration data. *Geocomputation 99* [online]. Available from: http://www.geocomputation.org/1999/068/gc_068.htm.

DWP. (2013) The State Pension age timetable. [online]. Available from: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/181343/spa-timetable.pdf.

Evans, J. (1987) Introduction: Migration and health. *International Migration Review*, **21**(3), pp.v-xiv.

Evans, P. and McCormick, B. (1994) The new pattern of regional unemployment: causes and policy significance. *The Economic Journal*, **104**(424), pp.633-647.

Everitt, B. S., Landau, S. and Leese, M. (2001) *Cluster analysis.* 4th ed. London: Arnold.

Ewing, G. O. (1974) Gravity and linear regression models of spatial interaction: A cautionary note. *Economic Geography*, **50**(1), pp.83-88.

Faggian, A. and McCann, P. (2009a) Human capital and regional development. In Capello, R. and Nijkamp, P., eds. *Handbook of regional growth and development theories.* Cheltnham: Edward Elgar.

Faggian, A. and McCann, P. (2009b) Human capital, graduate migration and innovation in British regions. *Cambridge Journal of Economics*, **33**(2), pp.317-333.

Faggian, A., McCann, P. and Sheppard, S. (2007) Some evidence that women are more mobile than men: gender differences in U.K. graduate migration behavior. *Journal of Regional Science*, **47**(3), pp.517-539.

Fielding, A. J. (1992) Migration and social mobility: South East England as an escalator region. *Regional Studies*, **26**(1), pp.1-15.

Fielding, T. (2012) *Migration in Britain: Paradoxes of the Present, Prospects for the Future.* Cheltenham: Edward Elgar Publishing.

Findlay, A., Mason, C., Harrison, R., Houston, D. and McCollum, D. (2008) Getting off the escalator? A study of Scots out-migration from a global city region. *Environment and Planning A*, **40**(9), pp.2169-2185.

Findlay, A., Mason, C., Houston, D., McCollum, D. and Harrison, R. (2009) Escalators, elevators and travelators: the occupational mobility of migrants to South-East England. *Journal of Ethnic and Migration Studies*, **35**(6), pp.861-879.

Findlay, A. M., Short, D. and Stockdale, A. (2000) The labour-market impact of migration to rural areas. *Applied Geography,* **20**(4), pp.333-348.

Findlay, A. M., Stockdale, A. and Stewart, E. (2002) Professional and managerial migration from core to periphery: the case of English migration to Scottish cities. *International Journal of Population Geography,* **8**(3), pp.217-232.

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), pp.455-470.

Finney, N. and Simpson, L. (2008) Internal migration and ethnic groups: evidence for Britain from the 2001 Census. *Population, Space and Place,* **14**(2), pp.63-83.

Fleming, A. (2005) Scotland's Census 2001 statistics on migration. *Occasional Paper No. 11*.

Flowerdew, R. (1991) Poisson regression modelling of migration. In Stillwell, J. and Congdon, P., eds. *Migration models: macro and micro approaches.* London: Bellhaven Press, pp.92-112.

Flowerdew, R. (2010) Modelling migration with poission regression. In Stillwell, J., Duke-Williams, O. and Dennett, A., eds. *Technologies for migration and commuting analysis : spatial interaction data applications.* Hershey: Business Science Reference, pp.261-279.

Flowerdew, R. and Aitkin, M. (1982) A method of fitting the gravity model based on the Poisson distribution. *Journal of Regional Science,* **22**, pp.190-198.

Flowerdew, R. and Salt, J. (1979) Migration between labour market areas in Great Britain, 1970–1971. *Regional Studies*, **13**(2), pp.211-231.

Fotheringham, A. S. and O'Kelly, M. (1989) *Spatial interaction models: formulations and applications.* Dordrecht: Kluwer Academic Publishers.

Fotheringham, A. S., Rees, P., Champion, T., Kalogirou, S. and Tremayne, A. R. (2004) The development of a migration model for England and Wales I: overview and modelling out-migration. *Environment and Planning A,* **36**, pp.1633-1672.

Fotheringham, A. S. and Rogerson, P. (1993) GIS and spatial analytical problems. *International Journal of Geographical Information Systems,* **7**(1), pp.3-19.

Galle, O. R. and Williams, M. W. (1972) Metropolitan migration efficiency. *Demography,* **9**(4), pp.655-664.

Gallent, N. (2011) Gentrification and the discourses of housing affordability, localness and priority in rural England. *Planning Theory & Practice,* **12**(4), pp.611-618.

Glaser, K. and Grundy, E. (1998) Migration and household change in the population aged 65 and over, 1971–1991. *International Journal of Population Geography,* **4**(4), pp.323-339.

Gordon, H. and McCormick, B. (1981) Do council housing policies reduce migration between regions? *The Economic Journal,* **91**(364), pp.919-937.

Gordon, H. and McCormick, B. (1994) Did migration in the 1980s narrow the north-south divide? *Economica*, **61**(244), pp.509-527.

Gordon, I. (1995) Migration in a segmented labour market. *Transactions of the Institute of British Geographers*, **20**(2), pp.139-155.

Green, A. E. (1988) The north-south divide in Great Britain: an examination of the evidence. *Transactions of the Institute of British Geographers*, **13**(2), pp.179-198.

Grigg, D. B. (1977) E. G. Ravenstein and the "laws of migration". *Journal of Historical Geography*, **3**(1), pp.41-54.

GROS. (2003) Suppliment to PAMS (03) 10. [online]. Available from: http://www.gro-scotland.gov.uk/files/pams0310-supplement.pdf.

GROS. (2005) Mid-2005 population estimates Scotland: 6. Notes and definitions - unmeasured migration. [online]. Available from: http://www.gro-scotland.gov.uk/statistics/theme/population/estimates/mid-year/archive/2005/notes-and-definitions/unmeasured-migration.html.

GROS. (2010a) Methodology used to produce GROS official migration estimates [online]. Available from: http://www.gro-scotland.gov.uk/statistics/theme/migration/methodology.html.

GROS. (2010b) Mid-year population estimates for Scotland: methodology guide. [online]. Available from: http://www.gro-scotland.gov.uk/files2/stats/population-estimates/mid-year-pop-est-methodology.pdf.

Hale, T. (2004) An extended example using Theil's T statistic: U.S. income inequality by county. *University of Texas Inequality Project (UTIP)* [online]. Available from: http://utip.gov.utexas.edu/tutorials/extended_example.doc.

Hamnett, C. (1991) The relationship between residential migration and housing tenure in London, 1971 - 81: a longitudinal analysis. *Environment and Planning A*, **23**(8), pp.1147-1162.

Harland, K. and Stillwell, J. (2007) Using PLASC data to identify patterns of commuting to school, residential migration and movement between schools in Leeds. *School of Geography Working Paper*, **07/03**.

Hatton, T. J. (2005) Explaining trends in UK immigration. *Journal of Population Economics*, **18**(4), pp.719-740.

HESA. (2012) Students and qualifiers data tables. [online]. Available from: http://www.hesa.ac.uk/index.php/content/view/1973/239/.

Hoare, A. and Corver, M. (2010) The regional geography of new young graduate labour in the UK. *Regional Studies*, **44**(4), pp.477-494.

House of Commons Science and Technology Committee. (2013) *The Census and social science: third report of session 2012-13.* London: HMSO.

House of Commons Treasury Select Committee. (2008) *Counting the Population: Eleventh Report of Session 2007-08.* London: HMSO.

Hubbard, P. (2009) Geographies of studentification and purpose-built student accommodation: leading separate lives? *Environment and Planning A*, **41**(8), pp.1903-1923.

Huff, D. (1963) A probabilistic analysis of shopping center trade areas. *Land Economics*, **39**, pp.81-90.

Hughes, G. and McCormick, B. (1991) Housing markets, unemployment and labour market flexibility in the UK. In *International Volatility and Economic Growth: The First Ten Years of The International Seminar on Macroeconomics*: Elsevier Science Publishers, pp.83-114.

Hunsinger, E. (2008) Iterative Proportional Fitting - R code. [online]. Available from: http://www.demog.berkeley.edu/~eddieh/datafitting.html.

Jivraj, S., Simpson, L. and Marquis, N. (2012) Local distribution and subsequent mobility of immigrants measured from the School Census in England. *Environment and Planning-Part A*, **44**(2), p491.

Johnston, R. and Pattie, C. (1993) Entropy-maximizing and the iterative proportional fitting procedure. *The Professional Geographer*, **45**(3), pp.317-322.

Johnston, R. and Pattie, C. (2009) Entropy-maximising models. In Kitchen, R. and Thrift, N., eds. *International encyclopedia of human geography*. Amsterdam: Elsevier, pp.499-504.

Jones, H. (1986) Evolution of Scottish migration patterns: A social-relations-of-production approach. *The Scottish Geographical Magazine*, **102**(3), pp.151-164.

Jones, H. (1992) Migration trends for Scotland: central losses and peripheral gains. In Stillwell, J., Rees, P. and Boden, P., eds. *Migration processes & patterns V2. Population redistribution in the United Kingdom*. London: Belhaven Press, pp.100-14.

Kalantaridis, C. (2010) In-migration, entrepreneurship and rural–urban interdependencies: The case of East Cleveland, North East England. *Journal of Rural Studies*, **26**(4), pp.418-427.

Kalogirou, S. (2005) Examining and presenting trends of internal migration flows within England and Wales. *Population, Space and Place*, **11**(4), pp.283-297.

Kaufman, L. and Rousseeuw, P. (2005) *Finding groups in data - an introduction to cluster analysis*. New Jersey: John Wiley and Sons.

Keiding, N. (1990) Statistical inference in the Lexis diagram. *Philosophical Transactions of the Royal Society of London. Series A: Physical and Engineering Sciences*, **332**(1627), pp.487-509.

Kennett, S. (1980) Migration within and between the Metropolitan Economic Labour Areas of Britain, 1966-1971. In Hobcraft, J. and Rees, P., eds. *Regional demographic development*. London: Croom Helm, pp.165-85.

Kilpatrick, S., Johns, S., Vitartas, P. and Homisan, M. (2011) In-migration as Opportunity for Rural Development. *Planning Theory & Practice*, **12**(4), pp.625-629.

Kofman, E. (2004) Family related migration: a critial review of European Studies. *Journal of Ethnic and Migration Studies*, **30**(2), pp.243-262.

Kulu, H. and Milewski, N. (2008) Family change and migration in the life course: an introduction. *Demographic Research*, **17**, pp.567-590.

Kupiszewska, D. and Nowok, B. (2008) Comparability of statistics on international migration flows in the European Union. In Raymer, J. and Willekens, F., eds. *International migration in Europe.* Chichester: Wiley and Sons, pp.41-71.

Lee, E. S. (1966) A theory of migration. *Demography,* **3**(1), pp.47-57.

Lexis, W. (1875) *Einleitung in die theorie der bevölkerungsstatistik (Introduction to the theory of population statistics).* Strassburg: Karl J. Trubner.

Long, L. (1992) Changing residence: comparative perspectives on its relationship to age, sex, and marital status. *Population Studies,* **46**(1), pp.141-158.

Loohach, R. and Garg, K. (2012) Effect of distance functions on K-Means clustering algorithm *International Journal of Computer Applications,* **49**(6), pp.7-9.

Lovett, A. and Flowerdew, R. (1989) Analysis of count data using Poisson regression. *Professional Geographer,* **41**, pp.190-198.

Lundholm, E. (2012) Returning home? Migration to birthplace among migrants after age 55. *Population, Space and Place,* **18**(1), pp.74-84.

Maier, G. and Weiss, P. (1991) The discrete choice approach to migration modelling. In Stillwell, J. and Congdon, P., eds. *Migration models: macro and micro approaches.* London: Belhaven Press, pp.17-33.

Martin, D., Dorling, D. and Mitchell, R. (2002) Linking censuses through time: problems and solutions. *Area*, **34**(1), pp.82-91.

Martin, R. (1988) The political economy of Britain's north-south divide. *Transactions of the Institute of British Geographers,* **13**(4), pp.389-418.

Martinez, W., Martinez, A. and Solka, J. (2011) *Exploratory data analysis with MATLAB, second edition.* London: CRC Press.

Marvin, S., Harding, A. and Robson, B. (2006) A framework for city-regions. [online]. Available from: www.eukn.org/dsresource?objectid=149367.

McCormick, B. (1997) Regional unemployment and labour mobility in the UK. *European Economic Review,* **41**(3–5), pp.581-589.

McGregor, I. (2011) *Visit to ONS Titchfield.* 1 April. Titchfield.

McHugh, K. E. and Gober, P. (1992) Short-Term Dynamics of the US Interstate Migration System, 1980–1988. *Growth and Change,* **23**(4), pp.428-445.

McHugh, R. and Morgan, J. N. (1984) The determinants of interstate student migration: A place-to-place analysis. *Economics of Education Review,* **3**(4), pp.269-278.

Millington, J. (2000) Migration and age: the effect of age on sensitivity to migration stimuli. *Regional Studies,* **34**(6), pp.521-533.

Milne, W. (1993) Macroeconomic influences on migration. *Regional Studies,* **27**(4), pp.365-373.

Mingay, G. E. (1989) *The rural idyll.* London: Routledge.

Mitchell, J., Pain, N. and Riley, R. (2011) The drivers of international migration to the UK: A panel-based Bayesian model averaging approach. *Economic Journal*, **121**(557), pp.1398-1444.

Mosca, I. and Wright, R. E. (2010) National and international graduate migration flows. *Population Trends*, **141**, pp.36-53.

Mueller, G. (2011) *Meeting at Ladywell House.* 11 May. Edinburgh.

National Housing Federation. (2013) Thirty-somethings abandon villages as rural housing prices soar. [online]. Available from: http://www.housing.org.uk/media/press-releases/thirty-somethings-abandon-villages-as-rural-housing-prices-soar#sthash.9zlSqgWH.dpuf.

National Statistics. (2006) Report of the Inter-Departmental Task Force on Migration Statistics. [online]. Available from: http://www.statistics.gov.uk/about/data/methodology/specific/population/future/imps/updates/downloads/TaskForceReport151206.pdf.

Newton, A. (2009) Joining it all up. In Hackett, P., ed. *Regeneration in a downturn: what needs to change?* London: The Smith Institute, pp.27-34.

NHS. (2011) *Spine factsheet* [online]. [Accessed 16 May 2013]. Available from: http://www.connectingforhealth.nhs.uk/resources/systserv/spine-factsheet.

NISRA. (2006) How are population estimates created? The methodology. [online]. Available from: http://www.nisra.gov.uk/archive/demography/population/midyear/mye_methodology.pdf.

NISRA. (2007) Development of methods/sources to estimate population migration in Northern Ireland [online]. Available from: http://www.nisra.gov.uk/archive/demography/population/migration/dev_est_mig.pdf.

NISRA. (2010) Migration statistics for Northern Ireland (2009). [online]. Available from: http://www.nisra.gov.uk/archive/demography/population/migration/Migration%20Report%202009.pdf.

NISRA. (2013) How are population estimates created? The methodology. [online]. Available from: http://www.nisra.gov.uk/archive/demography/population/midyear/Methodology_2013.pdf.

Norman, P. (1999) Putting Iterative Proportional Fitting on the researcher's desk. *School of Geography, University of Leeds Working Paper 99/3*.

Norman, P. (2010) Identifying change over time in small area socio-economic deprivation. *Applied Spatial Analysis and Policy*, **3**(2-3), pp.107-138.

Norman, P., Boyle, P. and Rees, P. (2005) Selective migration, health and deprivation: a longitudinal analysis. *Social Science & Medicine*, **60**(12), pp.2755-2771.

Norman, P., Marshall, A., Thompson, C., Williamson, L. and Rees, P. (2012) Estimating detailed distributions from grouped sociodemographic data: 'get me started in' curve fitting using nonlinear models. *Journal of Population Research*, **29**(2), pp.173-198.

Norman, P., Rees, P. and Boyle, P. (2003) Achieving data compatibility over space and time: creating consistent geographical zones. *International Journal of Population Geography,* **9**(5), pp.365-386.

Norman, P., Simpson, L. and Sabater, A. (2008) 'Estimating with confidence' and hindsight: new UK small-area population estimates for 1991. *Population, Space and Place,* **14**(5), pp.449-472.

NRS. (2010) Migration estimate adjustment using HESA data. *Population and Migration Statistics Committee (Scotland)* [online]. Available from: http://www.gro-scotland.gov.uk/files2/stats/consultation-groups/pams-10-14-mig-est-adjustment-hesa.pdf.

NRS. (2011) Using NHSCR for statistical purposes. [online]. Available from: http://www.gro-scotland.gov.uk/files2/nhscr/nhscr-gb-paper2-2-11-stats-purp.pdf.

Ogilvy, A. (1982) Population migration between the regions of Great Britain, 1971–9. *Regional Studies,* **16**(1), pp.65-73.

ONS. (2004) Report on research into revising internal migration estimates. [online]. Available from: www.ons.gov.uk/ons/rel/migration1/nhscr-/report-on-research-into-revising-internal-migration-estimates/internal-migration--nhscr----report-on-research-into-revising-internal-migration-estimates.pdf.

ONS. (2007a) Actual impacts of improved methods for population statistics in 2007 [online]. Available from: http://www.statistics.gov.uk/downloads/theme_population/FinalImpactTables.xls.

ONS. (2007b) The construction of intermediate geographies in London - amendment. [online]. Available from: http://www.statistics.gov.uk/downloads/theme_population/LondonGeography.pdf.

ONS. (2007c) Improved methods for estimating international migration: estimation of people whose intentions change with respect to their length of stay. [online]. Available from: http://www.statistics.gov.uk/downloads/theme_population/Mig_vis_switching.pdf.

ONS. (2007d) Improved methods for estimating international migration: geographical distribution of estimates of in-migration. [online]. Available from: http://www.statistics.gov.uk/downloads/theme_population/Geog_distn_in-migs.pdf.

ONS. (2007e) Improved methods for estimating international migration: geographical distribution of estimates of out-migration. [online]. Available from: http://www.statistics.gov.uk/downloads/theme_population/Geog_distn_out-migs.pdf.

ONS. (2007f) Improved methods for estimating international migration: improved local age distributions of international in-migrants. [online]. Available from: http://www.statistics.gov.uk/downloads/theme_population/Geog_distn_in-migs.pdf.

ONS. (2007g) Improved methods for estimating international migration: out-migration - propensity to migrate models for 2002 to 2006. [online]. Available from: http://www.ons.gov.uk/ons/guide-method/method-quality/imps/archive-material/archive-background-information/improved-methods-for-population-statistics-revisions--archive-material-/propensity-to-migrate-models-used.pdf.

ONS. (2007h) Improved methods for estimating international migration: the use of the Labour Force Survey to improve estimates of international in-migration - coverage and quality. [online]. Available from: http://www.ons.gov.uk/ons/guide-method/method-quality/imps/archive-material/archive-background-information/improved-methods-for-population-statistics-revisions--archive-material-/the-use-of-the-labour-force-survey.pdf.

ONS. (2008) Improved methods for estimating international migration: geographic areas used in estimation (amendment). [online]. Available from: http://www.ons.gov.uk/ons/guide-method/method-quality/imps/archive-material/archive-background-information/improved-methods-for-population-statistics-revisions--archive-material-/geographic-areas-used-in-estimation.pdf.

ONS. (2009a) Estimating international long-term immigration by local authority (LA). [online]. Available from: http://www.ons.gov.uk/ons/guide-method/method-quality/imps/msi-programme/communication/improvements-mid-2008/methodology-papers/immigration-detailed-methology-update.pdf.

ONS. (2009b) Methodology to estimate long-term international migration: 1991 to 2008. [online]. Available from: http://www.statistics.gov.uk/downloads/theme_population/Methodology-to-estimate-LTIM.pdf.

ONS. (2010a) Estimating internal migration: customer guidance notes. [online]. Available from: www.statistics.gov.uk/statbase/.../EstimatinginternalMigration.doc.

ONS. (2010b) Estimating international long-term emigration by local authority: update. [online]. Available from: http://www.ons.gov.uk/about-statistics/methodology-and-quality/imps/mig-stats-improve-prog/comm-stakeholders/improvements-2008-pop-est/detailed-methodology-papers/emigration-detailed-methodology-update.pdf.

ONS. (2010c) Improved estimation of student migration within England and Wales. [online]. Available from: http://www.ons.gov.uk/about-statistics/methodology-and-quality/imps/mig-stats-improve-prog/comm-stakeholders/improvements-2008-pop-est/detailed-methodology-papers/student-adjustment-detailed-methodology.pdf.

ONS. (2010d) Improving estimates of international migration in Northern Ireland, and between the UK and Republic of Ireland. [online]. Available from: http://www.statistics.gov.uk/downloads/theme_population/Irish_methodological_changes.pdf.

ONS. (2010e) Summary quality report for internal migration. [online]. Available from: www.ons.gov.uk/.../mid-year-population-estimates--mye-.pdf.

ONS. (2010f) Table 2: Cumulative effect mid-2008 by type of adjustment. [online]. Available from: http://www.ons.gov.uk/about-statistics/methodology-and-quality/imps/mig-stats-improve-prog/comm-stakeholders/improvements-2008-pop-est/tables-and-charts/t2-cumulative-effect-2008-by-type-of-adjustment.xls.

ONS. (2011a) Internal migration estimates methodology. [online]. Available from: www.ons.gov.uk/ons/guide-method/method-quality/specific/population-and-migration/estimating-internal-migration-customer-guidance-notes.pdf.

ONS. (2011b) Mid-year estimates short methods guide. [online]. Available from: www.ons.gov.uk/ons/guide-method/method-quality/specific/population-and-migration/pop-ests/mid-year-population-estimates-short-methods-guide.pdf.

ONS. (2011c) Population estimates for local authorities across UK constituent countries: a comparison of data sources and methods. [online]. Available from: www.ons.gov.uk/ons/guide-method/method-quality/specific/population-and-migration/pop-ests/population-estimates-for-las/population-estimates-for-local-authorities-across-uk-constituent-countries.pdf.

ONS. (2011d) Summary quality report for long-term international migration releases. [online]. Available from: http://www.ons.gov.uk/ons/guide-method/method-quality/quality/quality-information/social-statistics/summary-quality-report-for-international-migration.pdf.

ONS. (2012a) Inter administration agreement. [online]. Available from: http://www.ons.gov.uk/ons/about-ons/who-ons-are/programmes-and-projects/beyond-2011/beyond-2011---the-uk-context/index.html.

ONS. (2012b) Internal migration by local authorities in England and Wales, year ending June 2011. [online]. Available from: http://www.ons.gov.uk/ons/dcp171778_280054.pdf.

ONS. (2013a) Beyond 2011 - The UK context. [online]. Available from: http://www.ons.gov.uk/ons/about-ons/who-ons-are/programmes-and-projects/beyond-2011/beyond-2011---the-uk-context/index.html.

ONS. (2013b) Local government restructuring. [online]. Available from: http://www.ons.gov.uk/ons/guide-method/geography/beginner-s-guide/administrative/our-changing-geography/local-government-restructuring/index.html.

ONS. (2013c) Long-term international migration estimates: methodology document 1991 onwards. [online]. Available from: www.ons.gov.uk/ons/guide-method/method-quality/specific/population-and-migration/international-migration-methodology/long-term-international-migration-estimates-methodology.pdf.

ONS. (2013d) Population estimates for England and Wales, mid-2002 to mid-2010 revised (subnational): statistical bulletin. [online]. Available from: http://www.ons.gov.uk/ons/dcp171778_307723.pdf.

Openshaw, S. (1983) *The modifiable areal unit problem.* Geo books Norwich.

Owen, D. and Green, A. (1992) Migration patterns and trends. In Champion, T. and Fielding, T., eds. *Migration processes & patterns: research progress & prospects.* London: Belhaven Press, pp.17-40.

Owen, D. W. and Green, A. E. (1989) Spatial aspects of labour mobility in the 1980s. *Geoforum*, **20**(1), pp.107-126.

Plane, D. A. (1993) Demographic Influences on Migration. *Regional Studies*, **27**(4), pp.375-383.

Plane, D. A., Henrie, C. J. and Perry, M. J. (2005) Migration up and down the urban hierarchy and across the life course. *Proceedings of the National Academy of Sciences of the United States of America*, **102**(43), pp.15313-15318.

Plane, D. A. and Jurjevich, J. R. (2009) Ties that no longer bind? The patterns and repercussions of age-articulated migration. *The Professional Geographer*, **61**(1), pp.4-20.

Plane, D. A. and Mulligan, G. F. (1997) Measuring spatial focusing in a migration system. *Demography*, **34**(2), pp.251-262.

Pollard, N., Latorre, M. and Sriskandarajah, D. (2008) Floodgates or turnstiles? Post-EU enlargement migration flows to (and from) the UK. *Institute for Public Policy Research* [online]. Available from: http://www.bnegroup.org/images/uploads/publications/files/floodgates_or_turnstiles.pdf.

Population and Migration Statistics Committee. (2007) Options for the distribution of migrants. *PAMS Committee minutes*, **(07) 12**.

Population and Migration Statistics Committee. (2013) International migration distribution methodology. *PAMS Committee minutes*, **(13) 05**.

Rabe, B. and Taylor, M. (2009) Residential mobility, neighbourhood quality and life-course events. *ISER Working Paper Series* [online]. Available from: https://www.iser.essex.ac.uk/files/iser_working_papers/2009-28.pdf.

Rae, A. (2009) From spatial interaction data to spatial interaction information? Geovisualisation and spatial structures of migration from the 2001 UK census. *Computers Environment and Urban Systems*, **33**(3), pp.161-178.

Rae, A. (2011) Flow-data analysis with geographical information systems: a visual approach. *Environment and Planning-Part B*, **38**(5), p776.

Rae, A. (2013) English urban policy and the return to the city: A decade of growth 2001-2011. *Cities*, **32**, pp.94-101.

Ravenstein, E. G. (1885) The Laws of Migration. *Journal of the Statistical Society of London*, **48**(2), pp.167-235.

Ravenstein, E. G. (1889) The Laws of Migration. *Journal of the Royal Statistical Society*, **52**(2), pp.241-305.

Raymer, J. (2007) The estimation of international migration flows: a general technique focused on the origin-destination association structure. *Environment and Planning A*, **39**(4), pp.985-995.

Raymer, J. (2010) A general framework for estimating population movements. In Salzmann, T., Edmonston, B. and Raymer, J., eds. *Demographic aspects of migration.* Wiesbaden: VS Research, pp.73 - 96.

Raymer, J., Abel, G. and Smith, P. W. F. (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**, pp.891-908.

Raymer, J. and Giulietti, C. (2009) Ethnic migration between area groups in England and Wales. *Area*, **41**(4), pp.435-451.

Raymer, J., Rees, P. and Blake, A. (2012) A conceptual framework for UK population and migration statistics. [online]. Available from: http://www.ons.gov.uk/ons/guide-method/method-quality/imps/latest-news/conceptual-framework/index.html.

Raymer, J. and Rogers, A. (2007) Using age and spatial flow structures in the indirect estimation of migration streams. *Demography*, **44**(2), pp.199-223.

Raymer, J. and Rogers, A. (2008) Applying model migration schedules to represent age-specific migration flows. *International migration in Europe: Data, models and estimates*, pp.175-192.

Raymer, J. and Smith, P. W. F. (2010) Modelling migration flows. *Journal of the Royal Statistical Society Series a-Statistics in Society*, **173**(4), pp.703-705.

Raymer, J., Smith, P. W. F. and Giulietti, C. (2011) Combining census and registration data to analyse ethnic migration patterns in England from 1991 to 2007. *Population Space and Place*, **17**(1), pp.73-88.

Rees, P. (2008) What happens when international migrants settle? Projections of ethnic groups in United Kingdom regions. In Raymer, J. and Willekens, F., eds. *International Migration in Europe.* Chichester: Wiley and Sons, pp.329-358.

Rees, P., Durham, H. and Kupiszewski, M. (1996) Internal migration and regional population dynamics in Europe: United Kingdom case study. *School of Geography, University of Leeds Working Paper 96/20*.

Rees, P., Fotheringham, A. S. and Champion, T. (2004a) Modelling migration for policy analysis. In Stillwell, J. and Clarke, G., eds. *Applied GIS and spatial analysis.* Chichester: Wiley, pp.259-296.

Rees, P. and Kupiszewski, M. (1999) *Internal migration and regional population dynamics in Europe: a synthesis.* Strasbourg: Council of Europe Publishing.

Rees, P., Norman, P. and Brown, D. (2004b) A framework for progressively improving small area population estimates. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, **167**(1), pp.5-36.

Rees, P., Stillwell, J., Boden, P. and Dennett, A. (2009) A review of migration statistics literature. In UK Statistics Authority, ed. *Migration statistics: the way ahead, report 4 (Final).* London: UKSA.

Rees, P. and Willekens, F. (1981) Data bases and accounting frameworks for IIASA's comparative migration and settlement study. *Collaborative Paper CP-81-39* [online]. Available from: http://webarchive.iiasa.ac.at/Admin/PUB/Documents/CP-81-039.pdf.

Rees, P. and Willekens, F. (1986) Data and accounts. In Rogers, A. and Willekens, F., eds. *Migration and settlement: A multiregional comparative study.* Dordrecht: Reidel, pp.19-58.

Rees, P. H. and Duke-Williams, O. (1997) Methods for estimating missing data on migrants in the 1991 British census. *International Journal of Population Geography*, **3**(4), pp.323-368.

Reuveny, R. (2007) Climate change-induced migration and violent conflict. *Political Geography*, **26**(6), pp.656-673.

Robinson, D. (2013) United Kingdom, migration 1990s to present. *The encyclopedia of global human migration.* Blackwell Publishing Ltd.

Robson, B., Barr, R., Lymperopoulou, K., Rees, J. and Coombes, M. (2006) A framework for city-regions working paper 1: mapping city-regions. [online]. Available from: http://www.ncl.ac.uk/curds/publications/pdf/AFrameworkforCityRegionsWorkingPaper1PDF3314Kb_id11635961.pdf.

Rogers, A. (1975) *Introduction to multi-regional mathematical demography.* New York: John Wiley.

Rogers, A. (2007) Demographic modeling of the geography of migration and population: a multiregional perspective *Working Paper* [online]. Available from: http://www.colorado.edu/ibs/pubs/pop/pop2007-0002.pdf.

Rogers, A. (2008) Demographic Modeling of the Geography of Migration and Population: A Multiregional Perspective. *Geographical Analysis*, **40**(3), pp.276-296.

Rogers, A. and Castro, L. (1981) Model migration schedules. *Research Report 81-30* [online]. Available from: www.iiasa.ac.at/Admin/PUB/Documents/RR-81-030.pdf.

Rogers, A., Little, J. and Raymer, J. (2010) *The indirect estimation of migration.* Dordrecht: Springer.

Rogers, A., Raquillet, R. and Castro, L. J. (1978) Model migration schedules and their applications. *Environment and Planning A*, **10**, pp.475-502.

Rogers, A. and Raymer, J. (1998) The spatial focus of US interstate migration flows. *International Journal of Population Geography*, **4**(1), pp.63-80.

Rogers, A., Raymer, J. and Willekens, F. (2002) Capturing the age and spatial structures of migration. *Environment and Planning A*, **34**(2), pp.341-360.

Rogers, A., Willekens, F. and Raymer, J. (2001) Modeling interregional migration flows: continuity and change. *Mathematical Population Studies: An International Journal of Mathematical Demography*, **9**(3), pp.231 - 263.

Rogers, A., Willekens, F. and Raymer, J. (2003) Imposing age and spatial structures on inadequate migration-flow datasets. *The Professional Geographer*, **55**(1), pp.56-68.

Rolfe, H. and Metcalf, H. (2009) *Recent migration into Scotland: the evidence base.* Edinburgh: Scottish Government Social Research.

Roy, J. R. and Thill, J.-C. (2004) Spatial interaction modelling. *Papers in Regional Science*, **83**(1), pp.339-361.

Sabater, A. (2007) *Estimation of ethnic groups in sub-national areas for analysis of population change, England and Wales: 1991-2001*. Doctoral thesis, University of Manchester.

Sage, J., Smith, D. and Hubbard, P. (2012) The diverse geographies of studentification: living alongside people not like us. *Housing Studies*, **27**(8), pp.1057-1078.

Sander, N., Skirbekk, V. and Erling Lundevaller, S. (2010) Prospects for laterlife migration in urban Europe. *Driving forces and global trends* [online]. Available from: http://www.plurel.net/images/D124.pdf.

Sarra, A. and Del Signore, M. (2010) A dynamic origin-constrained spatial interaction model applied to Poland's inter-provincial migration. *Spatial Economic Analysis,* **5**(1), pp.29 - 41.

Scott, M., Smith, D. P., Shucksmith, M., Gallent, N., Halfacree, K., Kilpatrick, S., Johns, S., Vitartas, P., Homisan, M. and Cherrett, T. (2011) Interface. *Planning Theory & Practice,* **12**(4), pp.593-635.

Select Committee on Public Accounts. (2007) *Progress with demographics in the NHSCRS since Dr Nowlan's departure* [online]. Available from: http://www.parliament.the-stationery-office.co.uk/pa/cm200607/cmselect/cmpubacc/390/6062623.htm.

Select Committee on Public Service and Demographic Change. (2013) Ready for Ageing? [online]. Available from: http://www.publications.parliament.uk/pa/ld201213/ldselect/ldpublic/140/140.pdf.

Seo, J. (2002) Re-urbanisation in regenerated areas of Manchester and Glasgow: new residents and the problems of sustainability. *Cities,* **19**(2), pp.113-121.

Sheperd, P. (2006) *Neighbourhood profiling and classification for community safety*. Doctoral thesis, The University of Leeds.

Shucksmith, M. (2011) Exclusive rurality: planners as agents of gentrification. *Planning Theory & Practice,* **12**(4), pp.605-611.

Simpson, L. and Finney, N. (2009) Spatial patterns of internal migration: evidence for ethnic groups in Britain. *Population, Space and Place,* **15**(1), pp.37-56.

Simpson, L. and Sabater, A. (2009) Enhancing the population census: a time series for sub-national areas with age, sex, and ethnic group dimensions in England and Wales, 1991-2001. *Journal of Ethnic and Migration Studies,* **35**(9), pp.1461-1477.

Singleton, A., Wilson, A. and O'Brien, O. (2012) Geodemographics and spatial interaction: an integrated model for higher education. *Journal of Geographical Systems,* **14**(2), pp.223-241.

Smallwood, S. and De Broe, S. (2009) Sex ratio patterns in population estimates. *Population Trends,* **137**, pp.41-50.

Smallwood, S. and Lynch, K. (2010) An analysis of patient register data in the Longitudinal Study – what does it tell us about the quality of the data? *Population Trends,* (141), pp.152-169.

Smith, D. (2002a) Extending the temporal and spatial limits of gentrification: a research agenda for population geographers. *International Journal of Population Geography,* **8**(6), pp.385-394.

Smith, D. (2002b) Patterns and processes of studentification in Leeds. *Regional Review,* **12**(1), pp.14-16.

Smith, D. (2008) The politics of studentification and `(un)balanced' urban populations: lessons for gentrification and sustainable communities? *Urban Studies,* **45**(12), pp.2541-2564.

Smith, D. (2011) Geographies of long-distance family migration: Moving to a 'spatial turn'. *Progress in Human Geography,* **35**(5), pp.652-668.

Smith, D. and Higley, R. (2012) Circuits of education, rural gentrification, and family migration from the global city. *Journal of Rural Studies,* **28**(1), pp.49-55.

Smith, P., Raymer, J. and Giulietti, C. (2010) Combining available migration data in England to study economic activity flows over time. *Journal of the Royal Statistical Society: Series A (Statistics in Society),* **173**(4), pp.733-753.

Steinley, D. (2003) Local optima in K-Means clustering: what you don't know may hurt you. *Psychological Methods,* **8**(3), pp.294-304.

Stillwell, J. (1991) Spatial interaction models and the propensity to migrate over distance. In Stillwell, J. and Congdon, P., eds. *Migration models, macro and micro approaches.* London: Belhaven, pp.34-56.

Stillwell, J. (2005) Inter-regional migration modelling: A review and assessment. In *45th Congress of the European Regional Science Association, 23-27 August 2005, Vrije Universiteit Amsterdam, The Netherlands*.

Stillwell, J. (2013) Internal migration, ethnic groups in the UK. *The encyclopedia of global human migration.* Blackwell Publishing Ltd.

Stillwell, J., Bell, M., Blake, M., Duke-Williams, O. and Rees, P. (2000) Net migration and migration effectiveness: A comparison between Australia and the United Kingdom, 1976–96 Part 1: Total migration patterns. *Journal of Population Research,* **17**(1), pp.17-38.

Stillwell, J., Bell, M., Blake, M., Duke-Williams, O. and Rees, P. (2001) Net migration and migration effectiveness: A comparison between Australia and the United Kingdom, 1976–96. *Journal of Population Research,* **18**(1), pp.19-39.

Stillwell, J. and Congdon, P. (1991) Migration modelling: concepts and contents. In Stillwell, J. and Congdon, P., eds. *Migration models: macro and micro approaches.* London: Belhaven Press, pp.1-16.

Stillwell, J., Daras, K., Lomax, N., Bell, M. and Wilson, T. (2013) The IMAGE studio: A tool for comparative internal migration analysis and modelling. In *13th International Conference on Computers in Urban Planning and Urban Management, Utrecht, The Netherlands*.

Stillwell, J. and Duke-Williams, O. (2007) Understanding the 2001 Census interaction data: the impact of small cell adjustment and the problem of comparison with 1991. *Journal of the Royal Statistical Society Series A (General),* **170**(2), pp.1-21.

Stillwell, J., Duke-Williams, O. and Dennett, A. (2010) Interaction data: definitions, concepts and sources. In Stillwell, J., Duke-Williams, O. and Dennett, A., eds. *Technologies for migration and commuting analysis: spatial interaction data applications.* New York: IGI Global, pp.1-30.

Stillwell, J. and Harland, K. (2010) Analysing interaction data. In Stillwell, J., Duke-Williams, O. and Dennett, A., eds. *Technologies for migration and commuting analysis.* New York: IGI Global, pp.69-88.

Stillwell, J. and Hussain, S. (2010) Exploring the Ethnic Dimension of Internal Migration in Great Britain using Migration Effectiveness and Spatial Connectivity. *Journal of Ethnic and Migration Studies*, **36**(9), pp.1381-1403.

Stillwell, J., Rees, J. and Duke-Williams, O. (1996) Migration between NUTS level 2 regions in the United Kingdom. In Rees, P., Stillwell, J., Convey, A. and Kupiszewska, D., eds. *Population migration in the European Union.* Chichester: John Wiley and Sons, pp.275-307.

Stillwell, J., Rees, P. and Boden, P. (1992) Internal migration trends: an overview. In Stillwell, J., Rees, P. and Boden, P., eds. *Migration processes & patterns V2. Population redistribution in the United Kingdom.* London: Belhaven Press, pp.28-55.

Stockdale, A. (1992) State intervention and the impact on rural mobility flows in Northern Ireland. *Journal of Rural Studies,* **8**(4), pp.411-421.

Stockdale, A. and Catney, G. (2012) A life course perspective on urban–rural migration: the importance of the local context. *Population, Space and Place,* **Early view**, pp.n/a-n/a.

Stockdale, A., Findlay, A. and Short, D. (2000) The repopulation of rural Scotland: opportunity and threat. *Journal of Rural Studies,* **16**(2), pp.243-257.

Stockdale, A. and MacLeod, M. (2013) Pre-retirement age migration to remote rural areas. *Journal of Rural Studies,* **32**(0), pp.80-92.

Sweeney, S. and Konty, K. (2002) Population forecasting with nonstationary multiregional growth matrices. *Geographical Analysis,* **34**(4), pp.289-312.

Theil, H. (1967) *Economics and information theory.* Chicago: Rand McNally and Company.

Thomas, A. (1993) The influence of wages and house prices on British interregional migration decisions. *Applied Economics,* **25**, pp.1261-68.

Tobler, W. (1995) Migration: Ravenstein, Thornthwate and beyond. *Urban Geography,* **16**(4), pp.327-343.

Trevena, P., McGhee, D. and Heath, S. (2013) Location, location? A critical examination of patterns and determinants of internal mobility among post-accession Polish migrants in the UK. *Population, Space and Place,* **Early view**, pp.n/a-n/a.

Tunstall, H., Pickett, K. and Johnsen, S. (2010) Residential mobility in the UK during pregnancy and infancy: Are pregnant women, new mothers and infants ‘unhealthy migrants’? *Social Science & Medicine,* **71**(4), pp.786-798.

UKSA. (2009) *Migration statistics: the way ahead? Report 4 (final).* London: UKSA.

Uren, Z. and Goldring, S. (2007) Migration trends at older ages in England and Wales. *Population Trends,* **130**.

Vagras-Silva, C. (2013a) Geographical distribution and characteristics of long-term international migration flows to the UK. *The Migration Observatory* [online].

Available from: http://www.migrationobservatory.ox.ac.uk/sites/files/migobs/Briefing%20-%20Geographical%20Distribution%20and%20Characteristics_0.pdf.

Vagras-Silva, C. (2013b) Migration flows of A8 and other EU migrants to and from the UK. *The Migration Observatory* [online]. Available from: http://www.migrationobservatory.ox.ac.uk/sites/files/migobs/Migration%20Flows%20of%20A8%20and%20other%20EU%20Migrants%20to%20and%20from%20the%20UK.pdf.

van Der Gaag, N. and van Wissen, L. E. O. (2008) Economic determinants of internal migration rates: a comparison accross five European countries. *Tijdschrift voor economische en sociale geografie*, **99**(2), pp.209-222.

van Imhoff, E., van der Gaag, N., van Wissen, L. and Rees, P. (1997) The selection of internal migration models for European regions. *International Journal of Population Geography,* **3**(2), pp.137-159.

Vickers, D. (2006) *Multi-level integrated classifications based on the 2001 Census*. Doctoral thesis.

Vickers, D. and Rees, P. (2006) Introducing the area classification of output areas. *Population Trends,* **125**, pp.15-29.

Vickers, D. and Rees, P. (2007) Creating the UK National Statistics 2001 output area classification. *Journal of the Royal Statistical Society Series A: Statistics in Society,* **170**(2), pp.379-403.

Wallace, M., Charlton, J. and Denham, C. (1995) The new OPCS area classifications. *Population Trends,* **79**, pp.15-30.

Warnes, A. (1992) Migration and the life course. In Champion, A. and Fielding, A., eds. *Migration processes and patterns. volume 1: research progress and prospects.* London: Belhaven.

Wilson, A. (1970) *Entropy in urban and regional modelling.* London: Pion.

Wilson, A. G. (1967) A statistical theory of spatial interaction models. *Transportation Research,* **1**, pp.253-269.

Wilson, A. G. (1971) A family of spatial interaction models, and associated developments. *Environment and Planning,* **3**(1), pp.1-32.

Wilson, T. (2010) Model migration schedules incorporating student migration peaks. *Demographic Research,* **23**(8), pp.191-218.

Wilson, T. and Rees, P. (2005) Recent developments in population projection methodology: a review. *Population, Space and Place,* **11**(5), pp.337-360.

Wohland, P., Burkitt, M., Norman, P., Rees, P., Boden, P. and Durham, H. (2011) ETHPOP Database, ESRC Follow on Fund "Ethnic group population trends". [online]. Available from: www.ethpop.org.

Wong, D. (1992) The reliability of using the iterative proportional fitting procedure. *The Professional Geographer,* **44**(3), pp.340-348.

Wright, N. (2008) Levels and distribution of migrants in and out of Scotland: an overview of internal and international migration. *Radical Statistics,* **97**, pp.30-46.

Wulff, M., Champion, A. and Lobo, M. (2010) Household diversity and migration in mid-life: understanding residential mobility among 45–64 year olds in Melbourne, Australia. *Population, Space and Place,* **16**(4), pp.307-321.

Yorkshire Forward. (2010) The role of Yorkshire's universities in driving the region's economy. [online]. Available from: http://3cc8c9788db88691bf02-327e85584ac866108a24ce6f795317b9.r77.cf3.rackcdn.com/theroleofyorkshiresuniversitiesindrivingtheregionseconomy.pdf.

Zaloznik, M. (2011) *Iterative proportional fitting: theoretical synthesis and practical limitations*. Doctoral thesis, University of Liverpool.

Zipf, G. (1946) The P1 P2/D hypothesis: on the intercity movement of persons. *American Sociological Review*, **11**(6), pp.677-686.

Glossary of Terms

APS	Annual Population Survey
ASC	Annual School Census
CA	Council Area (district in Scotland)
CHI	Community Health Index
CIDER	Centre for Interaction Data Estimation and Research
DAS	Destination – Age – Sex
District	Local Authority, Unitary Authority, Council Area or Local Government District
FHA	Former Health Authority
Flag-4	The marker put on a GP registration record when the person's previous address was overseas
GOR	Government Office Region
GP	General Practitioner
HA	Health Authority
HB	Health Board
HE	Higher Education
HESA	Higher Education Statistics Authority
IMAGE	Internal Migration Around the Globe
IPF	Iterative Proportional Fitting
IPS	International Passenger Survey
L2	Lifetime Labour Market Database
LA	Local Authority
LAD	Local Authority District
LFS	Labour Force Survey
LGD	Local Government District in Northern Ireland
LTIM	Long Term International Migration
MAUP	Modifiable Areal Unit Problem
MSIP	Migration Statistics Improvement Program
MWS	Migrant Workers Scan
MYE	Mid-Year Estimate
NHS	National Health Service

NHSCR	National Health Service Central Register
NICHI	Northern Irish Community Health Index
NINo	National Insurance Number
NMGi	New Migrant Geography In
NMGo	New Migrant Geography Out
NRS	National Records of Scotland
NSAs	National Statistical Agencies
NUTS	Nomenclature of Territorial Units for Statistics
OAS	Origin – Age – Sex
OD	Origin – Destination
ONS	Office for National Statistics
PC	Parliamentary Constituency in Northern Ireland
PCT	Primary Care Trust
PRDS	Patient Register Data System
R	The R Project for Statistical Computing
SAS	Census Small Area Statistics
SCAM	Small Cell Adjustment Method
SCHI	Scottish Community Health Index
SMS	Special Migration Statistics
SNHSCR	Scottish National Health Service Register
TIM	Total International Migration
UA	Unitary Authority
UKSA	United Kingdom Statistics Authority
US	United States
WICID	Web-based Interface to Census Interaction Data
WRS	Workers Registration Scheme

Appendix A

List of terms associated with Figure 4.1

Term	Description
A	Migration between LADs in England
B	Migration between LADs in Wales
C	Migration between LADs in Scotland
D	Migration between LADs in Northern Ireland
E	Migration from LADs in England to LADs in Wales
F	Migration from LADs in Wales to LADs in England
G	Migration from LADs in England to LADs in Scotland
H	Migration from LADs in England to LADs in Northern Ireland
I	Migration from LADs in Wales to LADs in Scotland
J	Migration from LADs in Wales to LADs in Northern Ireland
K	Migration from LADs in Scotland to LADs in England
L	Migration from LADs in Scotland to LADs in Wales
M	Migration from LADs in Scotland to LADs in Northern Ireland
N	Migration from LADs in Northern Ireland to LADs in England
O	Migration from LADs in Northern Ireland to LADs in Wales
P	Migration from LADs in Northern Ireland to LADs in Scotland
Q	Migration from somewhere in the UK to LADs in England (excluding migration from elsewhere in England).
R	Migration from somewhere in the UK to LADs in Wales (excluding migration from elsewhere in Wales).
S	Migration from somewhere in the UK to LADs in Scotland (excluding migration from elsewhere in Scotland).
T	Migration from somewhere in the UK to LADs in Northern Ireland (excluding migration from elsewhere in Northern Ireland).
U	Migration from LADs in England to somewhere in the UK (excluding migration to somewhere else in England).
V	Migration from LADs in Wales to somewhere in the UK (excluding migration to somewhere else in Wales).
W	Migration from LADs in Scotland to somewhere in the UK (excluding migration to somewhere else in Scotland).
X	Migration from LADs in Northern Ireland to somewhere in the UK (excluding migration to somewhere else in Northern Ireland).
YT/ZT	All migration occurring within the UK

AW	Migration within a LAD in England
BW	Migration within a LAD in Wales
CW	Migration within a LAD in Scotland
DW	Migration within a LAD in Northern Ireland
A**O**...P**O**	All margins labelled 'O' represent the total outflow from a LAD to that sub-section of the matrix
A**D**…P**D**	All margins labelled 'D' represent the total inflow to a LAD from that sub-section of the matrix
AA	Migration from outside the UK to LADs in England
BB	Migration from outside the UK to LADs in Wales
CC	Migration from outside the UK to LADs in Scotland
DD	Migration from outside the UK to LADs in Northern Ireland
EE	Migration from LADs in England to outside the UK
FF	Migration from LADs in Wales to outside the UK
GG	Migration from LADs in Scotland to outside the UK
HH	Migration from LADs in Northern Ireland to outside the UK
IIT	Total migration from outside of the UK
JJT	Total migration from the UK to outside of the UK
KKT	For completeness, this cell represents all migration which occurs outside of the UK
A**T**… HH**T**	'T' represents the total migration occurring within each sub-section of the interaction matrix