Understanding Internal Migration in Britain at the Start of the 21st Century

by

Adam Raymond Dennett

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

The University of Leeds

School of Geography

September 2010

The candidate confirms that the work submitted is his own, except where work which has formed part of jointly-authored publications has been included. The contribution of the candidate and the other authors to this work has been explicitly indicated below. The candidate confirms that appropriate credit has been given within the thesis where reference has been made to the work of others.

Part of the work in Section 2.4 of Chapter 2 is based on the joint publication:

> Dennett, A. and Rees, P. (2010) 'Estimates of internal migration flows for the UK, 2000-2007', *Population Trends*, **140**, 82-105

I declare that the research for this publication was solely my own work and that I am the lead author. The contribution of the second named author, Phil Rees, was purely editorial and advisory.

Chapter 4 is based on the work found in two joint publications:

> Dennett, A. and Stillwell, J. (2008) 'Population turnover and churn - enhancing understanding of internal migration in Britain through measures of stability', *Population Trends*, **134**, 24-41

> Dennett, A. and Stillwell, J. (2009) 'Internal migration in Britain, 2000-01, examined through an area classification framework', *Population Space and Place*, DOI: 10.1002/psp.554

I declare that the research for these publications was solely my own work and that I am the lead author. The contribution of the second named author, John Stillwell, was purely editorial and advisory.

Acknowledgements

This thesis would not have been possible without the help, support and patience of my two supervisors and CIDER colleagues, John Stillwell and Oli Duke-Williams; who not only have helped guide me through this particular research journey, but who were brave enough to appoint me as Research Assistant on the CIDER project. John's door has always been open and I will be eternally grateful for his words of sage advice and considered use of the infamous red pen! My role within CIDER has made it possible to complete this piece of research and so the funding of CIDER from the Economic and Social Research Council's Census Programme, through award RES-348-25-0005, should also be acknowledged.

I am indebted to the constructive criticism, ideas and encouragement I have received from the members of my research support group; from Linda See and especially from Phil Rees who has provided me with many hours of his time, pages of his notebook and tranches of his wisdom far above and beyond the call of duty.

The School of Geography and especially members of the CSAP research cluster have been immensely supportive over the last few years and have made working here a great experience. During our time working together on the DEMIFER project, Pete Boden, particularly, provided cheery help and advice on areas of demography which have since fed directly into this thesis. Thanks have to go to the original office dream team of Dianna Smith and 'Super' Kirk Harland who provided great friendship, assistance with SQL and countless mid-morning coffee, tea and gossip breaks over my first couple of years in the department; and to Chris Thompson, Luke Burns and Lorna Fraser who have made work a pleasure since then. Thanks must also go to Paul Norman for being a great source of entertaining conversation and more importantly for putting the coffee pot on in the morning for the last few years.

Friends within the department who have made my time here nothing other than enjoyable also deserve thanks. John McCarthy for the laugh-out-loud email distractions, Nick Malleson for the Thursday pints and Alison Heppenstall for a shared outlook on so many things. Rebecca Hughes and Katherine Dunlavy have been two fantastic friends and great housemates providing entertainment and tea in equal measure. Thanks also to the 'come dine with me' crew - Katie, Sophie, Michelle, Kate, Tim and Lindsay - for some fine feasts and a welcome excuse to leave work on time on a Wednesday!

Finally I would like to say a big thank you to my Family: to Mum, Dad, Kim and Rooney who have been very supportive of my various academic endeavours over the years. An especially massive thank you is reserved for Emma Welsh who has put up with me living 200 miles away and the nightly updates about my work when I should really have far more entertaining things to talk about, but who has been my main source of inspiration and determination to finish this piece of research.

Abstract

Along with changes in fertility, mortality and international migration, internal migration acts to affect population change in almost all areas of Britain. As well as changing the numbers of people in localities it will alter the structure and composition of populations, impacting upon the planning and allocation of resources to local communities. In order to plan effectively, knowledge of the flows of people within the country is essential, but with almost ten percent of the population of Britain changing their permanent place of residence every year, a complex system of flows between a multitude of origins and destinations is presented.

There is a long history of studying internal migration in Britain; a history which owes much to the system of flows continually evolving. Monitoring this system can be problematic as unlike births and deaths, there is no compulsory mechanism to record the movements of people within the country. Data are accessed from different sources, each with their own idiosyncrasies which pose challenges for those wishing to build a complete understanding of the flows taking place.

This thesis tackles the problem of building an understanding of internal migration in Britain where data are limited and patterns and processes complex. New methods for estimating incomplete data are presented, along with new techniques for analysing available datasets. Central to the understanding of internal migration patterns is the association of types of migrant with origin and destination areas; therefore one of the central contributions of this piece of work is the development of a new internal migration-based geodemographic area classification framework, designed to both assist in the analysis of internal migration data from the census used to build it and to offer a parsimonious system for the analysis of temporally rich but attribute poor non-census datasets.

Contents

List of Figures

List of Tables

Glossary of terms

ACORN A Classification of Residential Neighbourhoods

CIDER Centre for Interaction Data Estimation and Research

CAS Census Area Statistics

CCS Census Coverage Survey

CHI Community Health Index

DEMIFER DEmographic and MIgratory Flows affecting European Regions and Cities

DEFRA Department for Food and Rural Affairs

ESRC Economic and Social Research Council

EU European Union

FHSA Family Health Service Area

GOF Goodness of Fit

GOR Government Office Region

GROS General Register Office for Scotland

HA Health Authority

HESA Higher Education Statistics Authority

IPF Iterative Proportional Fitting

JSA Job Seekers' Allowance

KS Census Key Statistics

LA Local Authority

LAD Local Authority District

LCCS Leeds Classification for Community Safety

MAUP Modifiable Areal Unit Problem

MIMOSA MIgration MOdelling for Statistical Analysis

MATLAB MAtrix LABoratory

NHS National Health Service

NHSCR National Health Service Central Register

NI-CHI Central Health Index

NISRA Northern Ireland Statistics and Research Agency

NUTS Nomenclature of Territorial Units for Statistics

OA Output Area

OAC Output Area Classification

ONS Office for National Statistics

OPCS Office of Population Censuses and Surveys

PRDS Patient Register Data System

PAR Population at Risk

PCA Principal Components Analysis

PLASC Pupil Level Annual School Census

SCAM Small Cell Adjustment Method

SIM Spatial Interaction Model

SMS Special Migration Statistics

SMDR Standardised Migration Distance Ratio

SMIR Standardised Migration Ratio

SIR Standardised Illness Ratio

SMR Standardised Mortality Ratio

SRMSE Square Root of the Mean Squared Error

ST Census Standard Table

STS Census Special Travel Statistics

SWS Census Special Workplace Statistics

TTWA Travel To Work Area

UK United Kingdom

UN United Nations

U.S. United States

WICID Web-based Interface to Census Interaction Data

Chapter 1

Introduction

In the Regions 2020 report (CEC, 2008) commissioned by the European Union (EU) to assess the future challenges faced by EU regions as we head further into the third millennium, one of key issues identified along with globalisation, climate change and energy is population change. All countries and regions within Europe will be affected by changes in their populations but many in quite different ways. The population of the United Kingdom (UK) is projected to keep rising over the next half century (Rees et al., 2010) against a backdrop of low fertility and continued aging, but these overall patterns will mask much more nuanced sub-national changes in population magnitude and structure. Local variations in fertility and mortality will contribute significantly to the evolving demographic profiles of areas, but migration will also contribute to the altering structure and composition of the local populations across the country. In some cases this migration will involve the movement of people internationally, but far more common-place will be the movement of people to and from other areas within the country through internal migration.

Internal migration is responsible for important changes in the populations of some localities and can have implications in a number of areas. Travers et al. (2007) outline issues for service provision and social cohesion; social cohesion and segregation are discussed by Bailey and Livingston (2007, 2008), with the impact of internal migration on crime analysed by Rotolo and Tittle (2006). Impacts on the physical environment and local economic development are highlighted by Smith (2002) and Smith and Denholm (2006), with the health of local labour markets often dependent on the flow of internal migrants (Dixon, 2003). Therefore knowledge of internal migration continues to be of high importance where effective planning and resource allocation are required.

Understanding internal migration at a relatively small geographical scale is perhaps more important now than it has been at any time. The global financial crisis has led to the deepest recession seen in the UK since the 1920s, posing real problems for those charged with governing the country. A new coalition government, sworn in in mid-2010, is determined to move power and funding away from central and regional government to more local government, scrapping Regional Development Agencies (http://www.bbc.co.uk/news/10391326) and proposing to install Local Enterprise Partnerships (Larkin, 2010), comprised of local authorities, to deal resource allocation and planning.

So there is a continuing and perhaps increasing need to understand, at a local level, the patterns of internal migration that contribute to population change in this country, but this presents a challenge since patterns of internal migration in Britain are complex and varied and thus can evade easy comprehension. Understanding what is happening now - who is moving from where, to where, why the patterns are occurring and how they are persisting or changing - is extremely important, and it is this context which frames the overall research aim of this thesis.

1.1 Aims and objectives

The aim of this thesis is to advance the current understanding of internal migration in Britain. In order to achieve this aim, a number of specific objectives are proposed:

1. To examine and review the current internal migration data landscape of Britain, identifying features in the provision of data which could affect understanding and exploring techniques for improving data where there are deficiencies, resulting in the development of a new partially-estimated national dataset.

2. To review the current methodological techniques and substantive literature surrounding internal migration to form solid foundations upon which to build a more current understanding.

3. To explore the patterns of internal migration in Britain at the start of the 21^{st} century using data from the 2001 Census.

4. To develop a new area classification based on internal migration data to both support analysis of census-based internal migration data and use as a framework for analysis of non-census-based data.

5. To build on existing methods and develop new techniques for understanding internal migration data.

6. To examine recent trends in internal migration in Britain over time using new partially-estimated data.

7. To offer explanations for current internal migration patterns in Britain through the use of mathematical spatial interaction models and life-course theory.

Each of these objectives will be met by the various chapters in the thesis. Table 1.1 below summarises where each objective will be tackled in which chapter(s).

Table 1.1: Thesis objectives and corresponding chapters

Objective	**Corresponding chapter(s)**
1. To examine and review the current internal migration data landscape of Britain, identifying features in the provision of data which could affect understanding and exploring techniques for improving data where there are deficiencies, resulting in the development of a new partially-estimated national dataset	Chapter 2 - Analysing migration: definitions data and methodological approaches
2. To review the current methodological techniques and substantive literature surrounding internal migration to form solid foundations upon which to build a more current understanding	Chapter 2 - Analysing migration: definitions data and methodological approaches Chapter 3 - Analysing internal population migration: substantive analyses - a historical perspective Chapter 8 - Understanding a decade of internal migration in Britain - from spatial interaction to life course explanations
3. To explore the patterns of internal migration in Britain at the start of the 21^{st} century using data from the 2001 Census	Chapter 4 - Internal migration in Britain - a district level analysis Chapter 6 - Developing a migration classification
4. To develop a new area classification based on internal migration data to both support analysis of census-based internal migration data and use as a framework for analysis of non-census-based data	Chapter 5 - The case for a migration classification Chapter 6 - Developing a migration classification
5. To build on existing methods and develop new techniques for understanding internal migration data	Chapter 4 - Internal migration in Britain - a district level analysis Chapter 6 - Developing a migration classification Chapter 7 - Monitoring migration between censuses Chapter 8 - Understanding a decade of internal migration in Britain - from spatial interaction to life course explanations
6. To examine recent trends in internal migration in Britain over time using new partially-estimated data	Chapter 7 - Monitoring migration between censuses Chapter 8 - Understanding a decade of internal migration in Britain - from spatial interaction to life course explanations
7. To offer explanations for current internal migration patterns in Britain through the use of mathematical spatial interaction models and life-course theory	Chapter 8 - Understanding a decade of internal migration in Britain - from spatial interaction to life course explanations

1.2 Thesis structure

As Table 1.1 reveals, most of the research objectives will be met by individual chapters, although some of the broader methodological objectives will be addressed across multiple chapters. Whilst Chapters 2, 3 and 5 in the thesis are the most explicitly review orientated, the introduction of different ideas and concepts with each chapter will necessitate thorough examinations of the relevant literature throughout this piece of work.

Chapter 2 introduces the thesis and will immediately begin to address the first objective through a thorough review of internal migration data sources in Britain. This review will require a detailed discussion of the important relationship between the data and the phenomenon being measured and the effect these have on the theoretical conceptualisation of an internal migrant and the process of internal migration. After setting the data scene, the attention of Chapter 2 will turn to a definition of a set of generic techniques which can be used to observe and analyse migration data, partially addressing objective 2. Knowledge of both data and techniques is needed for the final part of Chapter 2 which addresses the issue of data deficiencies. A new method will be described which enhances currently available patient register-based internal migration datasets though the estimation of missing intra-national flows at the Local Authority District (LAD) level - a dataset which will be of crucial importance to the work carried out in Chapters 7 and 8.

Chapter 3 will complete the second objective of the thesis through examining relevant substantive analyses of internal migration, both in the UK and further afield. The review will conclude the introductory portion of this work through highlighting the patterns of internal migration which have been observed in countries with comparable socio-economic and political landscapes and through outlining the recent history of research into internal migration in the UK, drawing out themes and patterns important for contextualising the work on more recent patterns contained within this thesis.

Chapter 4 is a major contribution towards addressing the third objective of this thesis which is to provide an empirical snapshot of the internal migration scene at the start of the 21^{st} century, through the analysis of inter-district flow data from the 2001 Census. This chapter will also introduce a number of important methodological ideas for the thesis, helping to address objective 5. Analysis of flow data will take place in the context of an existing area classification - a technique central to the analysis in the work that follows and important for enabling a reduction in the complexity of inter-district flows. In addition, alternative measures of migration intensity will be defined and used to enhance the picture of internal migration from the census.

Chapters 5 and 6 provide the core methodological contribution of this piece of work; a contribution bound within the fourth objective and adding to objective 5. Chapter 5 will set out the theoretical rationale for the development of a new migration-based area classification for internal migration analysis, discussing why the classification of areas according to their migration characteristics is both useful for helping understand the constituent internal migration data in even more detail, and important for a more effective reduction in the complexity of migration flows than could be achieved through using a more general purpose classification framework. After deciding upon an appropriate classification methodology and scale of analysis and presenting an argument for the use of 2001 Census data in Chapter 5, Chapter 6 will detail the step-by-step process of choosing variables and selecting an appropriate clustering technique in order that a robust migration-based classification is produced. In the latter half of the chapter the final classification will be presented before an evaluation of the solution is carried out, justifying the position of the ‘Migration Classification’ in migration analysis in place of more general purpose typologies.

In Chapter 7 the new classification will be used as a framework for the analysis of a ten-year time series of the new partially-estimated patient register-based data developed in Chapter 2. This chapter will address the sixth objective of this piece of work, examining the most recent national trends in internal migration, but will also make a significant contribution to objective

5 through the exemplification of the benefits of the new Migration Classification and through developing a number of new analysis metrics to help disentangle further some of the complex internal migration patterns occurring in Britain.

Chapter 8 will build on the findings of Chapter 7, but will look to offer some explanations for the patterns that have been observed. Addressing objective 7 and contributing again to objective 5, part of the chapter will explore the contribution that spatial interaction modelling theory can give to the explanation of flow patterns within the Migration Classification system. Alternative doubly constrained models will be fitted to flow data to explore both the influence of distance decay on migration patterns and to identify where model residuals might point to other influences affecting the patterns that are presented. In explaining some of these other influences, the theory of the interaction between migration and life course events will be explored.

Finally Chapter 9 will look to synthesise the findings of the whole project and draw some overall conclusions. Assessing the overall contribution of this work, the aims and objectives outlined in Chapter 1 will be returned to, and the extent to which each has been met will be examined. There will be areas of success, but undoubtedly questions still to be answered with avenues of future research still to be explored, so the final section of this chapter will offer suggestions for where future research may be usefully diverted.

Chapter 2

Analysing migration: definitions, data and methodological approaches

2.1 Introduction

Accounting for populations within territories has long been a human preoccupation. From the early Roman censuses to the Domesday Book and all the way to modern population registers, we continue to seek knowledge about people in places for a variety of reasons. The study of human populations has led to the development of a set of concepts and techniques collectively referred to as 'demography' (Rees, 2009). Along with fertility and mortality, migration is one of the defining pillars of demographic research, but whilst births and deaths benefit from being events which are relatively easy to define and measure, migration can often defy a definitive definition. Firstly, the migrant individual should be distinguished from the migration event. Courgeau (1973) states that the migrant is an individual who has experienced or is experiencing one (or more) migration event, whereas migration is defined as a change in usual residence, either for this one migrant, or a group of migrants. This thesis is concerned with understanding 'internal migration' and 'internal migrants' as distinct from 'international migration' and 'international migrants'. But how can we define either? Lee (1966) defines migration as 'a permanent or semi-permanent change of residence' of any distance, even if that move is only a few metres. Rees (1977) chooses to define migration as a permanent change of usual residence. A detailed definition is proposed by Rees et al. (2009), which states that:

> *"migration is the event of transfer from one residential location to another by a person who is termed a migrant. In this context, an event is an activity that takes place over a short period of time and a transfer involves travel over some distance from one location to another."* (Rees et al., 2009, p.64)

In this context then, an international migrant will be an individual who changes their permanent residential location and does so crossing a national boundary. This might seem like a logical definition, but consulting the United Nations (UNESCO, 2010), one would discover that an international migrant is defined alternatively as:

> *"any person who lives temporarily or permanently in a country where he or she was not born, and has acquired some significant social ties to this country."*

In this definition the length of time over which the event has happened becomes irrelevant - that a movement event has happened resulting in some kind of social attachment for an individual is what defines them as a migrant. Another definition in the UN glossary states that migrants are:

> *"all cases where the decision to migrate is taken freely by the individual concerned, for reasons of 'personal convenience' and without intervention of an external compelling factor."*

Free-will now becomes the defining attribute. Another UN definition from the same glossary states that the migration event is:

> *"the crossing of the boundary of a political or administrative unit for a certain minimum period of time."*

So defining an international migrant is a difficult task. It follows, therefore, that defining an internal migrant and the process of internal migration might be just as problematic. The issues with defining migration definitively are relayed by Boyle (2009), who observes that in migration research the migration event is often defined by whether it occurs over long or short distances or results in a temporary or permanent change of residence, but that the exact definition of each of these can vary from study to study. He goes on to explain that definitions will frequently be

> *"influenced as much by the data resource at hand as by theoretically guided principals."*

And this indeed will prove to be the case in this piece of work. In principal, the definition of an internal migrant and internal migration might appear obvious - if international migration involves the crossing of a national boundary, then internal migration will be contained within national boundaries and an internal migrant an individual who does not cross national borders. Broadly speaking, of course, this will be the case - internal migration in the UK can be loosely defined as all permanent residential moves occurring within and between the constituent countries of the UK. Some, however, draw the distinction between residential mobility and internal migration. Cadwallader (1992) differentiates between inter-regional internal migration and intra-city internal migration which he terms 'residential mobility' - making a definite distinction between the two. The questions that follow, therefore, are should such a distinction be made and if it should, where do you draw the line between the two? Taking the first of these it could be argued that there should be a distinction as migration (which will be over a longer distance than residential mobility) implies a move that not only involves a change in location, but very probably a change in employment and a departure from previous social groups. Certainly this argument may have held in the past, but today this may not necessarily be the case in a country where the predominance of service sector jobs and advances in communication technologies means it is common for people to move over long distances and retain jobs. Similarly, a long distance migration move does not necessarily mean social circles and friendship groups have to change with online social-networking meaning contact can be maintained far more easily than it once was. Furthermore, it might be that a move to a different location in the same city might involve the same change in status, employment or social ties as a longer distance move. Looking at these moves from a different perspective - from the perspective of local impact - a shorter distance move is probably less likely to involve a change of doctor, school, or council but at if moves are within a local authority district (for example from urban Leeds to rural Otley), very often these move will serve to reinforce social stratification (as the wealthier move to better locations) and will often impact on local school and health services. So the answer to the first question might be ambiguous if one chooses to try and draw the distinction along status changing lines. The task of drawing a distinction becomes even more difficult if one accepts that a distinction should be made and then tries to define where the line should be drawn. If migration is over a long distance and residential mobility is over a shorter distance, when does one turn into the other - further than 100m, 1km, 10km? Any arbitrary cut-off could surely be criticised.

Attempting to draw a distinction between migration and residential mobility is not just difficult, but potentially misleading where acknowledging a distinction indicates an acceptance that an easy partition can be made where in reality no such clear dividing line can be drawn. In this research project different data sets will be used to examine internal migration and each will measure the phenomenon in slightly different ways, so the precise definition of an internal migrant will depend, to an extent, on the dataset being used. In addition, these datasets in many cases will not present information which could be used to determine whether the move of the individual is part of some kind of change in status.

Another practical reason for not drawing such a theoretical distinction exists. As will be seen, the principal primary unit of analysis in this thesis will be the LAD. Were a distinction drawn between residential mobility and internal migration, then the cut-off between the two would have to be flows within Local Authority Districts (LADs) (as no other cut-off would be possible). In the 2001 Census, around 60% of all migration moves were within LADs - ignoring these moves would mean ignoring 60% of the migration story of Britain. For these reasons, in this thesis a theoretical distinction between residential mobility and internal migration will not be drawn. An individual making a permanent residential move within Britain, however measured by the dataset, will be termed a ‘migrant’; and the move they make a ‘migration.’

It is this important idea that shapes the rest of this chapter, the principal aim of which is to set the theoretical, methodological and practical terms of reference for the rest of the thesis. Exploration of the idea that the definition of internal migration is inextricably linked to the data being examined is key to the understanding of what follows. Therefore this will be the preoccupation of Section 2.2, which will provide an overview of the internal migration data landscape in the UK at this time, outlining the datasets in existence and how each will define an internal migrant and internal migration in an individual way. In understanding internal migration data, it is important to also understand the spatial system framework within which the data are bound; therefore Section 2.3.1 will define a generic spatial framework and will describe an accompanying notation to depict flows within this system. This thesis is concerned with analysing internal migration data within such a spatial system, so it follows that a brief overview of the techniques which can be employed to examine the data should be given in Sections 2.3.2 and 2.3.3. In setting these theoretical, methodological and practical terms of reference, some of the shortcomings of the available data will be identified; a problem which leads on to the second aim of this chapter.

Internal migration data in Britain are imperfect, as we shall see, and this poses a problem for anyone wishing to develop an understanding of the internal migration landscape of this country. But given some imperfect data and a framework for exploring the information contained within, a number of pieces of research have proven that it is possible to enhance imperfect data through synthetic estimates. The second aim of this chapter, therefore, is to investigate whether it will be possible to improve upon the data already in existence in Britain to provide a more complete evidence base for analysis in the latter chapters of this thesis. Section 2.4, will address this problem, considering the feasibility of augmenting existing data before proposing an estimation method to achieve these ends; the result being that a new methodological approach will be proposed which will be used to augment existing data producing a new, national, inter-district ten year time-series dataset, disaggregated by broad age groups.

2.2 Internal migration data in the UK

Table 2.1 provides a summary of all of the internal migration datasets currently available in the UK. The principal source of information available to researchers is the UK Census of Population. As it is a legal requirement for every person resident in the UK on census night to fill in an enumeration form, the census is unrivalled in terms of its sampling - capturing almost all of the population directly with the small proportion not captured imputed (a discussion of which can be found in Section 2.4.1) so that a representative 100% sample is produced Rees et al. (2002c). In terms of the internal migration data derived from the census, the spatial resolution (down to Output Area (OA)), coverage (the whole UK), attribute information (migration can be cross-tabulated against every other variable derived from the census survey) and accessibility (all published migration origin/destination flow data are available online from

Table 2.1: Internal migration datasets in the UK

	Reporting period	Coverage	Lowest level of geography	Migrant sample	Flows	Age	Sex	Ethnicity	Data sourced from
Censuses									
Census of Population									
Main tables	2000-01	UK	OA	Total pop		✓	✓	✓	CASWEB/ONS
Special Migration Statistics (SMS)	2000-01	UK	OA	Total pop	✓	✓	✓	✓	CIDER
SARs (including SAM and CAMS)	2000-01	UK	LAD	1-5% pop	✓	✓	✓	✓	SARS
LS (including Scottish)	2000-01	GB	Ward	1% pop (5.5% Scot)	✓	✓	✓	✓	CeLSIUS
Customised commissioned tables	2000-01	UK	OA	Total pop	✓	✓	✓	✓	ONS
School Census (PLASC)	Annual	England & Wales	LSOA (Unit postcode with special permission)	All school children	✓	✓	✓	✓	PLUG/DCSF
Surveys									
General Household Survey	Annual to 2006	GB	GOR	20,000 people, aged 16+	✓	✓	✓	✓	ESDS
Annual Population Survey	Annual since 2004	UK	LAD (special licence)	350,000 people	✓	✓	✓	✓	ESDS/NOMISWEB
Labour Force Survey	Quarterly	UK	LAD (special licence)	125,000 people	✓	✓	✓	✓	ESDS/NOMISWEB
Acxiom lifestyle survey	Annual	GB	SOA	Approx 750,000, aged 18+	✓	✓	✓	✓	Acxiom
Administrative sources									
NHSCR	Annual	GB (separate Scot/Eng and Wales)	2001 HA	All NHS patients	✓	✓	✓		ONS
Patient Register data system (PRDS)	Annual	England & Wales	LAD	All NHS patients	✓	✓	✓		ONS/CIDER
Community Health Index (CHI)	Annual	Scotland	Council Area	All NHS patients	✓	✓	✓		GRO Scotland/CIDER
Central Health Index (NI-CHI)	Annual	Northern Ireland	Local Govt Dist	All NHS patients	✓	✓	✓		NISRA
Higher Education Statistics Agency	Annual	England & Wales	MSOA (unit pc possible)	Approx 900,000 HE students	✓	✓	✓	✓	HESA
Electoral roll/register	Annual	UK	LAD	Voting age population	✓	✓	✓		Local Authorities (no central source)

Source: Rees et al. (2009)

Centre for Interaction Data Estimation and Research (CIDER)), are all incomparable.

The most recent census data dates from 2001 with both internal and international migrants tabulated as counts across a number of standard Census Key Statistics (KS) and Census Area Statistics (CAS) tables. Origin/destination migration flow data are tabulated in a separate set of tables known as the Special Migration Statistics (SMS) - a summary of the data contained within the 2001 SMS is given in Table 2.2. SMS data are published at three geographical levels: level 1, (which contains flows between ‘districts’ - comprised of London boroughs, unitary authorities, metropolitan districts, and other local authority districts in England, unitary authorities in Wales, council areas in Scotland and parliamentary constituencies in Northern Ireland), level 2 (CAS and Census Standard Table (ST) wards) and level 3 (Output Areas (OAs)). As one moves down the geographical hierarchy, the number of tables and the amount of variables contained within each table are reduced. This trade-off between geographical detail vs. attribute detail is necessary in order to preserve the confidentiality of individual census respondents.

Table 2.2: Data contained within the 2001 Special Migration Statistics

Level	**Geography**	**Table Reference**	**Table Name**	**Cells/variables within table**
Level 1	District	Table 1	Age by sex	75
		Table 2	Family status of migrant	54
		Table 3	Ethnic group by sex (GB destinations)	24
		Table 3n	Ethnic group by sex (Northern Ireland destinations)	9
		Table 4	Whether suffering limiting long term illness by whether in household by sex by age	84
		Table 5	Economic activity by sex	42
		Table 6	Moving groups	16
		Table 7	Moving groups by tenure	32
		Table 8	Moving groups by economic activity by sex	336
		Table 9	Moving groups by NS-SEC of group reference person	288
		Table 10	Migrants in Scotland/Wales/Northern Ireland with some knowledge of Gaelic/Welsh/Irish	36
Level 2	Ward	Table 1	Age by sex	51
		Table 2	Moving groups	4
		Table 3	Ethnic group by sex	9
		Table 4	Moving groups by NS-SEC of group reference person	24
		Table 5	Moving groups by tenure	8
Level 3	Output Area	Table 1	Age by sex	12

Source: http://cider.census.ac.uk

In addition to the suppression of more variable detail at higher spatial resolutions, all 2001 Census data are subjected to the Small Cell Adjustment Method (SCAM) - a data perturbation process which is designed to alter small counts to further ensure individual respondents cannot be identified in the data. For each census table at each geographical level of publication, all counts of 1 or 2 are adjusted randomly to either 0 or 3. Small counts are more prevalent in origin/destination flow data as there are two geographies associated with every count, with the effect of SCAM becoming more severe at smaller scale geographies where a greater number of small flow interactions occur. For a full discussion of the effects of SCAM on the 2001 interaction data, see Duke-Williams and Stillwell (2007) and Stillwell and Duke-Williams (2007).

Migration data from the census are derived from the question: *"What was your usual address one year ago"* and as such are counts of single transitions over a one year period (Rees, 1977). If an individual had moved several times within that year, only the transition from the address one year ago would be recorded in the data. There is no scope (in the current census, although this will change in 2011 - (ONS, 2010a)) for multiple addresses - something which is becoming more and more common where modern employment practices mean long distance long time-frame 'commutes' are not unusual (Green, 2004). That said, while only one transition is counted, it can occur over any distance - a move to the next-door address which in some definitions would be classed as 'residential mobility' rather than migration (Dixon, 2003), would still count as a migration move in the SMS.

Despite the benefits of the census, a well documented drawback is that it only occurs on a decennial basis, with no other mandatory system existing to record internal migration between census years (Chappell et al., 2000). Therefore anyone wishing to study internal population movements on a more frequent basis has to turn to a selection of administrative or survey data sources not originally designed for this purpose and thus featuring a number of problems (Dennett et al., 2007). For example, work carried out by Marquis and Jivraj (2009) and Jivraj and Marquis (2009) on measuring internal migration through the Pupil Level Annual School Census (PLASC), shows that whilst a useful source of data producing comparable flows to other data sources, it is greatly limited through sampling only school children. With the survey and administative sources listed in Table 2.1, small samples and coarse geographies mean that most are not generally favoured for direct analysis. The notable exception being those derived from National Health Service (NHS) data, as NHS records are held for the vast majority of the population. Consequently, for many years the National Health Service Central Register (NHSCR) has been the preferred source of inter-censal internal migration information in the UK.

Whenever a patient re-registers with a general practitioner in a different Health Authority (HA) [1], a record of that move is stored on a central database. These data are then collated quarterly, producing summaries of moves between Health Authorities (HAs) in that period. HA geographies can be aggregated to Government Office Region (GOR), and the Office for National Statistics (ONS) (previously the Office of Population Censuses and Surveys (OPCS)) have been producing quarterly tabulations of moves between the constituent regions of the UK for decades (Dennett et al., 2007). As such the reliability of these data have been examined and tested extensively. For example, Boden et al. (1992) following on from the work of Devis and Mills (1986), found a strong association between flow patterns recorded by the NHSCR and the census and concluded that NHSCR data could be used credibly to analyse patterns of inter-regional migration in the UK.

[1] Health Authority geographies are some of the most volatile in the hierarchy of UK geographies, with a huge number of boundary and definitional changes over the years - see Dennett et al. (2007)for a full description of the problems with HA geographies.

The ONS produces estimates (rounded to the nearest 100 people) of total inter-regional migration flows, derived from the NHSCR, every quarter (March, June, September and December) which can be downloaded from its website (http://www.statistics.gov.uk/StatBase/Product.asp?vlnk=10191). Below the level of GOR, however, flows between areas in all constituent countries of the UK or Britain based on NHSCR data are not available. In England and Wales, for a period of time NHSCR flows were available at the sub-GOR - HA - geography, however, a transition from Family Health Service Areas (FHSAs) to HAs between 1998 and 2001, with associated boundary redefinition, means that a consistent time series at this geographical level does not exist (Dennett et al., 2007).

Conscious of the limitations of NHSCR data for monitoring sub-national internal migration flows, the ONS sought new data which could provide internal migration estimates at a finer geographical scale. As well as the central register, records of NHS patients are also maintained locally by HAs. Crucially, these local patient registers also contain address information in the form of unit postcodes and so offer the possibility of monitoring migration at a much more detailed spatial scale than the NHSCR (Chappell et al., 2000). The ONS collects these patient registers from all HAs in England and Wales and collates the information on an annual basis. When a patient is identified on the records in two consecutive years, but with a different address postcode *and* that postcode is in a different Local Authority (LA), a move is recorded between the two Local Authorities (LAs). Conceptually as well as spatially, then, these data differ from the NHSCR. Patient register moves can be defined as 'transitions' Rees (1977) and are more akin to migrants recorded by the census as only the individual migration is counted in the year (multiple moves are not). NHSCR data, on the other hand, record each move event that takes place - more on these conceptual differences will be given in the Section 2.2.1.

Testing the robustness of patient register data, the ONS concluded that the data were accurate from mid-1998. Therefore annual inter-LAD internal migration flow estimates from this Patient Register Data System (PRDS) have been made available from 1998 onwards (ONS, 2005b, 2007a,d, 2009d). However, during testing, some issues with undercounting were identified (Chappell et al., 2000; Scott and Kilbey, 1999), which affected the accuracy of the data. The solution was to constrain flows between LAs to known flows from the NHSCR at the HA level. Consequently the final published PRDS flow data reflect NHSCR *moves* rather than patient register transitions. (Dennett and Rees, 2010; Rees et al., 2009).

The ONS now uses its PRDS to produce three LAD flow data products for England and Wales:

a) normal rounded published PRDS tables (rounded to 10s);

b) unrounded PRDS (rounded to integers);

c) decimal unrounded PRDS (totally unrounded, containing long decimals); a by-product of constraining PRDS data to NHSCR flows and only used internally by the ONS. (Michelle Bowen, ONS Migration Statistics Unit - private communication).

The 'normal rounded' data have been supplied to CIDER (http://cider.census.ac.uk) and made available to all registered academic users in the UK through the Web-based Interface to Census Interaction Data (WICID) system (Stillwell, 2006). These data contain only total flows. 'Unrounded' PRDS data, disaggregated by eight broad age groups (of differing size - equating to life course stages) have been made available for the research in this thesis. 'Decimal unrounded' data are only accessible by the ONS, and used in mid-year population estimates. One of the issues with only having access to the 'unrounded' PRDS data is that rounding errors mean that age disaggregated flows do not sum to the total flows - an issue which will be discussed in more detail later in Section 2.4.3.

Subsequent to the work carried out by ONS on patient registers, a similar methodology was introduced by the General Register Office for Scotland (GROS) to estimate flows between council areas (GROS, 2003; Rees et al., 2009). The Community Health Index (CHI) patient register flows in Scotland are also constrained to NHSCR totals, although the publicly available time series only stretches back to 2001-02. Similarly, Northern Ireland Statistics and Research Agency (NISRA) make use of their Central Health Index (NI-CHI) to estimate sub-national internal migration flows (NISRA, 2005), although unfortunately these flows are not publicly available. In England and Wales the ONS now adjust PRDS data further with additional data on students from Higher Education Statistics Authority (HESA) (ONS, 2010b) and are in the process of assessing whether School Census data can also be used to improve the quality of these estimates (ONS, 2009e). These new adjustments are not included in the data which will be used in this thesis.

In all cases, PRDS-type data only measure flows between districts and not within districts. This is despite it being theoretically possible to do both with all flows being constructed from unit postcodes. As Chappell et al. (2000) note, the main reason for not producing intra-district flows is that it is not possible to constrain these flows to inter-HA NHSCR flows. The other principal drawback of PRDS-based inter-district flow data is that the national statistical agencies of England and Wales, Scotland and Northern Ireland, have not yet collaborated to produce a joined-up set of inter-district flows for either Britain or the whole of the UK. This presents a problem for any researcher wishing to carry out a national analysis, something which is confirmed by a complete lack of published research on a time series of inter-censal, sub-region UK internal migration flows. More importantly, this is also problematic for this thesis - more of which will be addressed in Section 2.4.

2.2.1 Conceptual issues - movement vs. transition data

In the last section the conceptual differences between census 'transition' data and NHSCR 'movement' data were alluded to, but are these differences important and how exactly will they affect the data produced? One of the issues relates to the calculation of migration rates or probabilities and this will be discussed in detail in Section 2.3.2. The other issue relates to the recording of information and is especially apparent when migration data are disaggregated by

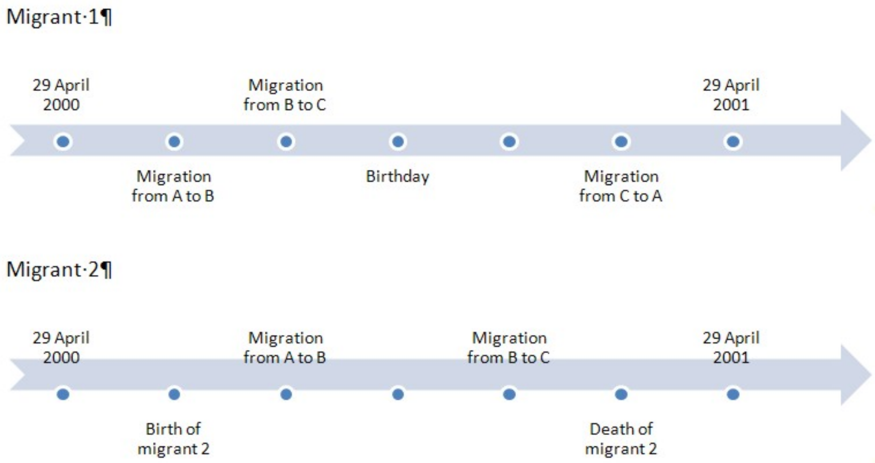


Figure 2.1: Representation of two migrant histories over a 1 year period

age. Movement data counts the number of migration moves that a migrant makes over a given time period - the age of the migrant recorded with each movement. Transition data, on the other hand, records a single migrant transition for a given time period regardless of the number of moves that have been made, and the age of the migrant is recorded at the end of the period. To explain in more detail, consider Figure 2.1.

Both migrant histories in Figure 2.1 are over a one year period. Migrant 1 moves twice (location A to B and B to C), has a birthday and then moves a third time (C back to A). In the NHSCR movement data, not only would all three migration events be recorded, but the age of the migrant would be recorded differently between the first two migration events and the third migration event - each contributing differently to the count collated at the end of the year. The census transition data, on the other hand, would not record the migrant at all as the third migration event sees the migrant move back to original location A and the address at the beginning of the one year period is the same as the address at the end of the period. A migration transition is not recorded even though a number of migration events have occurred. Even if the third migration event either did not happen, or the migrant moved to a fourth location 'D', there would be differences between the NHSCR and census data at the end of the period. If the migration from C to A did not happen, the census would record only one migrant transition and this transition would record the age of the migrant at the end of the period, rather than the different ages when the two migration events actually happened. If a fourth migration event (C to D) happened, then whilst the age recorded would be accurate for that event, previous movements would not be recorded.

Considering the rather unfortunate account of Migrant 2, the individual is not alive at the beginning of the period, but is then born, migrates twice and then dies before the end of the period. The census would not record this migrant. Even if either the death event did not happen, the migrant would still not be recorded as they were not alive at the beginning of the period. The NHSCR, on the other hand, would record both migration events.

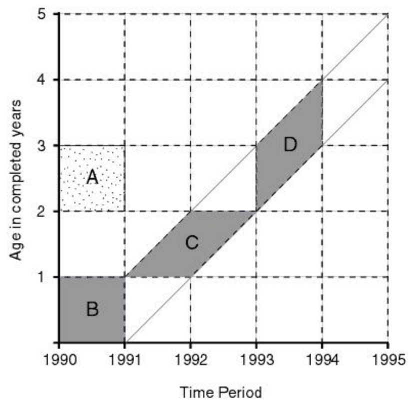


Figure 2.2: Lexis diagram representing period/age/time components of internal migration data
Source: Duke-Williams and Blake (2003)

Whilst movement and transition data then are conceptually quite different, it is theoretically possible to harmonise the two. Duke-Williams and Blake (2003) discuss this in detail, but a summary is presented here. Consider Figure 2.2; this is an example of a Lexis diagram (Carstensen and Keiding, 2005; Vandeschrick, 2001) which can be used to depict the age/time period/cohort elements of demographic data. The different shaded parts of the lexis diagram represent different ways in which migration data can be represented. Shapes A and B show age and period data. As is stated by Duke-Williams and Blake (2003), the dots in A each represent a single movement event of a migrant of a precise age at a precise time. B on the other hand, represents a similar age-period set of data but this time the data represents anyone who moved during the 1990-91 period whose age at the time of migration was between 0 and 1. These representations are comparable to the NHSCR movement data. Shape C represents data classified by age and cohor and represents all migrants (who had moved at some point in the previous year) who turned 2 during the 1992-93 period (Duke-Williams and Blake, 2003) - i.e. whilst they were 2 at the time of asking the question on migration, they would not have been aged 2 at the time of the migration event. Shape D on the other hand, represents data classified by cohort and period and represents all migrants who moved during 1993-94, and were aged at least 3 at the end of the period. This representation is comparable to census transition data available in the UK.

To harmonise movement and transition data, Duke-Williams and Blake propose dividing the age-period square into two component triangles (Figure 2.3). With information about the precise age of the migrant and the exact date of the migration event, it is possible assign migrants recorded in an age-period format to be re-assigned to either an older/earlier cohort element or a younger/later cohort element. These cohort elements can then be re-aggregated into either an age-cohort dataset, or perhaps more commonly into a cohort-period dataset comparable with census transition data. The extent to which these kind of adjustments are made by the ONS when using NHSCR data to constrain PRDS flows is unclear, so the possibility of introducing

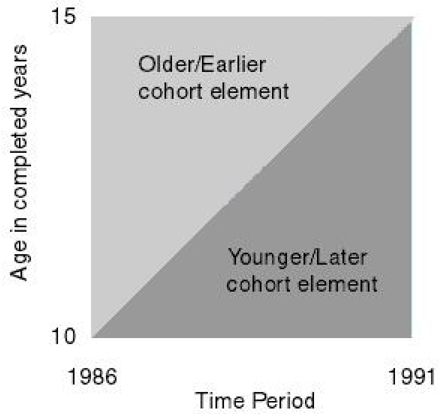


Figure 2.3: Cohort elements of age-period data
Source: Duke-Williams and Blake (2003)

error relating to period, cohort and age should be borne in mind.

So it is clear that whilst internal migration data in the UK broadly measures the movements of individuals within the nation, the two primary data sources (and the ones which will be used in this thesis) - the SMS from the 2001 Census, and the PRDS-based, NHSCR adjusted data - have differences in exactly what they measure, which in turn leads to different specifications of an internal migrant and internal migration. In this thesis, reference may be made variously to the terms ‘migration’, ‘mobility’ and ‘movement’. For each of these, ‘internal migration’ will be being referred to, but the precise definition will be dictated by the particular dataset being studied. Having discussed the definition of internal migration and the internal migrant, the next task in this chapter is to set out a methodological framework which will be adopted for subsequent analysis.

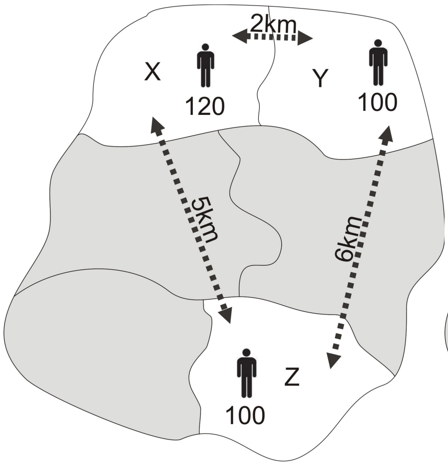
2.3 Methods for analysing migration data

This section sets out a general framework for analysing migration data and covers some of the more commonly used methods and techniques applied to the study of internal migration data. It should be noted that the ‘core’ methodological technique to be used in this thesis - that of cluster analysis and classification building - will not be covered here as these techniques are not commonly used in the analysis of migration. The full rationale for this alternative approach to migration analysis will be provided in Chapter 5.

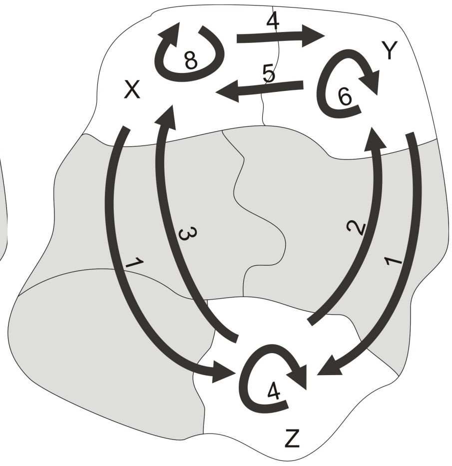
2.3.1 A generic spatial system and notation scheme

Consider Figure 2.4. This represents a hypothetical discrete zone spatial system (Wilson, 2000), containing three zones of interest - X, Y and Z. Each zone has a resident population, which could be measured at any time, but in censuses is generally recorded for the end of a given time period.

Each zone has a centroid point which enables the distance between each zone to be measured. Flows of people between and within each zone can be recorded over a given time period either as migrant transition or migration flow counts. In the UK census time period is over one year, however periods may vary - the Australia census, for example, has previously used a five year migration measure (Bell, 2002), with the French census recently moving from a one year to a five year migration measurement period (Baccaini, 2007).



(a) Population and spatial system



(b) Migration flows

Figure 2.4: A hypothetical spatial system

Figure 2.4b represents the migrant transition counts of people moving within and between each zone which are recorded at the end of a given period (although it could just as easily represent migration movement counts during a period). Each of the elements of these diagrams can be represented numerically in the form of a matrix. Table 2.3a is a matrix representing the population and distance elements of Figure 2.4a, with the interior values showing the distances between each zone, and the marginal values representing the population of each zone. The values in this matrix can be substituted for algebraic notation as shown in Table 2.3b. Here d represents distance between the two subscript zones, and P is the population of each zone.

Table 2.4a is a similar matrix representing the migration elements in Figure 2.4. The total flows into (column) and out of (row) each zone are shown in the table margins. It is important to note that in this system, within zone flows are not included in the in and out flow totals - hence the lower values. In the notation, origin zones are represented with O, destination zones with D and the migration flow between them with M. Subscript indices are used to denote either the origin or destination zone of interest, the convention being that where two subscripts are used, origin precedes destination. Where a column or row margin value is the total in or out flow, the summation over the index it replaces is represented by a $+$ symbol, however, in some instances this notation will be replaced by the origin or destination specific variable, so that, for example the total outflow from zone X could be represented by M_{x+} or O_x.

It is apparent that whether using alphabetical or numerical subscripts, if a zonal system

Table 2.3: Matrix representation of the hypothetical population and spatial system

(a) Populations of zones and the distances between them

	X	Y	Z	
X	0	2	5	120
Y	2	0	6	100
Z	5	6	0	100
	120	100	100	320

(b) An algebraic representation of 2.3a

	X	Y	Z	
X	d_{xx}	d_{xy}	d_{xz}	P_x
Y	d_{yx}	d_{yy}	d_{yz}	P_y
Z	d_{zx}	d_{zy}	d_{zz}	P_z
	P_x	P_y	P_z	P

Table 2.4: Matrix representation of flows within the hypothetical system

(a) Migration transition counts between zones in the system over a defined time period

	X	Y	Z	
X	8	4	1	5
Y	5	6	1	6
Z	3	2	4	5
	8	6	2	16

(b) An algebraic representation of 2.4a

	D_x	D_y	D_z	
O_x	M_{xx}	M_{xy}	M_{xz}	M_{x+}
O_y	M_{yx}	M_{yy}	M_{yz}	M_{y+}
O_z	M_{zx}	M_{zy}	M_{zz}	M_{z+}
	M_{+x}	M_{+y}	M_{+z}	M_{++}

is large then the use of unique subscripts would present problems. Therefore, it is common practice to label typical origin and destination zones with the subscripts i and j respectively, so the total out-migration from an origin zone can be referred to as O_i or M_{i+}; in-migration to any destination, D_j or M_{+j}; and flows between origin and destination zones as M_{ij}. Where flows are within a zone and origin equals destination, the convention M_{ii} will be used. Therefore:

$$\sum_j M_{ij} = M_{i+} = O_i \tag{2.1}$$

$$\sum_i M_{ij} = M_{+j} = D_j \tag{2.2}$$

$$\sum_i \sum_j M_{ij} = M_{++} = M \tag{2.3}$$

where $i \neq j$

For a full description of this nomenclature type, see Wilson (2000) or Champion et al. (1998).

2.3.2 Observing migration

So within any system, gross in, out and within area flows (movements or transitions) can be observed over a period of time. Where the volume of these flows is likely to be a partial function of the size of the population contained in the zone, it is common practice to standardise the flows as an intensity so areas of different size can be compared. If the count of migrants or migration moves is the numerator, then frequently the denominator is the population 'at risk' of migrating during the period of study. The most basic measures of intensity are described as 'crude' in they they do not account for the variation that might occur by age, sex, ethnicity, socio-economic status etc., they merely sum across all such variables to produce a total intensity. Therefore a crude migration intensity *CMI* for the whole system could be calculated as:

$$CMI = 1000 \left(\frac{M}{P} \right) \tag{2.4}$$

Here the intensity is scaled by a constant of 1,000 but any factor of 10 could be adopted. In this thesis by convention an intensity per 1,000 people will be used. Rees et al. (2000) give an excellent overview of the calculation of migration intensities and make an important distiction between a prospective *'crude migration probability'* which can be calculated for transition data using a Population at Risk (PAR) which is the estimated population at the *start* of the time period over which migration is measured, and a retrospective *'crude migration rate'* which can be calculated for movement data using a PAR which is the estimated population at the *end* of the time period. Rees et al. note that whichever measure is used, it is important to include all migrants in the denomenator who appear in the numerator and equally confine the denomenator to those who could validly be included in the numerator. This would mean that emigrants and non-survivors should be excluded from the PAR used in prospective transition probabilities, along with new immigrants - this is especially relevant when comparing intensites between countries which might have different mortality and international migration profiles. When not comparing countries or different times periods this is less important. In practice when census transition data are being analysed, retrospective rates calculated from end of period PAR which automatically exclude emigrants and non-survivors, tend to be used anyway as the beginning of period population is not recorded. Of course this is not ideal as these PAR contain those individuals who migrated into the system or who were born in the system during the period. In the absence of better PAR data, where census data are used in this thesis (Chapters 4 and 6), transition rates will be caluclated using just end of period populations. Where PRDS-based data are used (Chapters 7 and 8), mid-year population estimates will be used as PAR. These data will include some inidvidiuals who, technically, should not appear in either the numerator or denomentaor if wholly accurate intensites are being computed, but again in practice, where more accurate PAR are not available intensites will need to be computed using available data. Whilst Rees et al. (2000) make a semantic distinction between 'probabilities' and 'rates', where such calculations are always a representation of the ratio between two counts the term 'rate' will be used to mean all such standardised calculations.

Table 2.5: Migration rates for the sample system

	X	Y	Z	
X	33.3	18.2	4.6	41.7
Y	22.7	30.0	5.0	60.0
Z	13.6	10.0	20.0	50.0
	66.7	60.0	2.0	

Table 2.5 represents the O_i, D_j and M_{ij} flow rates per 1,000 people, with the PAR in this case being the end of period population P_i (or P_j), from Table 2.3. Rates of migration could be calculated for the whole system, for total in, out or within area flows or for flows between origin and destination pairs. Therefore, in this system, an out-migration rate for a given zone OMR_i could be represented as:

$$OMR_i = 1000\left(\frac{O_i}{P_i}\right) \tag{2.5}$$

The interior cell values of the matrix can be expressed as a rate of either the origin, destination or origin and destination PAR. The exact denominator chosen can vary, but in this example, origin and destination PAR are used, so the MR_{ij} rate per 1,000 people can be expressed as:

$$MR_{ij} = 1000\left(\frac{M_{ij}}{P_i + P_j}\right) \tag{2.6}$$

From these basic observations it is possible to specify a large suite of standardised migration indicators for each zone of interest, including commonly used metrics such as the net-migration rate, NM:

$$NM_i = 1000 \left(\frac{D_j - O_i}{P_i} \right) \tag{2.7}$$

where $i = j$

Sometimes, suitable PAR may not be available (for example where groups of migrants rather than individuals are counted). Where this is the case a measure of migration *effectiveness* (sometimes referred to as migration *efficiency*) can be calculated where the denomenator is the sum of the in-migration and out-migration for a given zone. For example, a rate of net-migration effectiveness NME_i can be calculated as:

$$NME_i = 1000 \left(\frac{D_j - O_i}{D_j + O_i} \right) \tag{2.8}$$

where $i = j$

A large range of other measures are available with an almost exhaustive list documented by Bell et al. (2002). A number of different measures described using this standard notation will be adopted in this thesis, the justification for using and the exact specification of each described at the relevant point. In addition, variations on this standard notation will appear throughout the thesis with precise specification given where necessary.

2.3.3 Explaining migration and predicting migration - models of expected flows

After observation, it generally follows that the next step along the road to understanding is explanation. In the example spatial system shown in Figure 2.4, flows of people are occurring between and within the constituent zones of the system. It might be that a series of observations at different points in time reveal that whilst populations and migration events vary to some extent, broadly speaking similar patterns are presented. For example, it can be observed that more flows are occurring between the two zones - X and Y - which are closer together, or more flows are occurring into the zone with the largest population - X. From these observations, one might be able to develop a hypothesis which attempts to explain the flow patterns present in the system in terms of origin and destination masses and in terms of the distances between zones. This might state that 'more flows occur between zones which are closer together and between zones with larger populations'. A model could then be built to test this hypothesis.

Indeed this kind of explanation for internal migration flows within discrete zone spatial systems has a long history, perhaps dating back to observations made by Ravenstein in his papers on the laws of migration Ravenstein (1889, 1885), but certainly reiterated mathematically and in terms of population, distance and flow by Zipf (1946) and then subsequently by many others (Alonso, 1978; Champion et al., 1998; Flowerdew and Aitkin, 1982; Fotheringham, 1984; Stillwell, 1978; Willekens, 1983) following the development of a new suite of mathematical models based upon Newtonian gravitational principals by Wilson (1970, 1971)[2]. The basic gravity model would state that the flow between any origin and destination would be equal to the product of the mass of the origin and destination and inversely proportional to the square of the distance between them, scaled by a balancing factor k so that the flows do not exceed the capacity of the system. Therefore:

$$M_{ij} = kO_iD_j(d_{ij})^{-2} \tag{2.9}$$

where

$$k = \frac{M}{\sum_i \sum_j O_iD_j(d_{ij})^{-2}} \tag{2.10}$$

Table 2.6: Results of gravity model migration flow estimation for the hypothetical system

	X	Y	Z	
X		5.3	0.3	5.6
Y	8.5		0.2	8.7
Z	1.1	0.6		1.7
	9.6	5.9	0.5	

Using the example system in Table 2.4, the gravity model above produces the model results shown in Table 2.4. Of course it is immediately possible to see that the results are not perfect - such a generalised model will lack enough information to produce precise results, however, it is also possible to see that the results are not very far away from the observed data. Early models of internal migration such as the model proposed by Zipf (1946), were analogous to this gravity principal, but could be criticised, amongst other things, for their relatively poor accuracy. Wilson's suite of 'entropy maximising' models based upon gravity principals were a noticeable improvement in that they were able to take advantage of any known information within a spatial system in order to increase the accuracy of model results . A full discussion of mathematical gravity models and their spatial interaction model derivatives will be given in Chapter 8.

As Flowerdew (2010) points out, one of the useful properties of a mathematical spatial interaction model is that if both sides of the equation are logged, the equation takes the form of a multiple regression equation, with $\ln M_{ij}$ becoming the Y dependent variable, and the other components becoming the independent, X variables, such that the gravity equation shown above becomes:

[2]It should be noted that Alonso's theory of movement technically did not follow the work of Wilson, rather was developed in parallel, although it has since been proven that Alonso's models are essentially an analogue of Wilson's entropy maximising models (Ledent, 1981; Wilson, 1980)

$$\ln M_{ij} = b_0 + b_1 \ln O_i + b_2 \ln D_j + b_3 \ln d_{ij} \tag{2.11}$$

With b_0 - the intercept term - replacing k, the coefficients b_1 and b_2 being around 1 and b_3 around -2 (Flowerdew, 2010).

One of the principal benefits of converting the mathematical spatial interaction model into a statistical multiple regression spatial interaction model is that multiple regression models allow for the inclusion of additional X variables which can improve the results of the model (Field, 2005). In an earlier paper, Flowerdew and Aitkin (1982) argue that mathematical spatial interaction models are relatively inflexible in that the addition of additional explanatory variables would require the augmentation of one of the three components of the model - the origin or destination attractiveness factors or the impedance function. He proposes that using a regression model based on the Poisson distribution retains the benefits of standard entropy-maximising spatial interaction models (i.e. the total flows predicted by the model are equivalent to the total observed flows - something that does not happen with Ordinary Least Squares regression models) but also has the flexibility of being able to incorporate additional explanatory variables (perhaps relating to employment rates, housing conditions or any other origin or destination specific variable) if appropriate, in order to improve the fit of the model. Flowerdew concedes that Poisson regression models are not perfect, with issues of multicollinearity and sparse migration matrices affecting the reliability of model fits, but argues that they present a feasible alternative to traditional spatial interaction models.

Flowerdew and Green (1992) note that Poisson regression models are part of a wider family of generalised linear regression models. Generalised linear models have been used elsewhere to model migration flows (Willekens, 1983). Raymer et al. (2006) use a multiplicative component model (re-expressed as a log-linear model) to examine age-specific interregional flows between regions in Italy over seven five year periods from 1970-1971 to 2000-2001. By using the model to describe the age and spatial structures of migration in Italy over these separate time periods, and then disaggregating the separate flows between origins i, destinations j of ages a into separate components, Raymer et al. (2006) demonstrate that it is then possible to project age-specific interregional migration flows using linear extrapolation techniques from past data to estimate the components of the model in the future. Other examples of log linear modelling to project migration flows can be found in the literature. In another paper, Raymer et al. (2007) use a log-linear model to project elderly, health specific migration flows between ONS district classification areas in Britain. Using census data to impose structure onto an estimated flow matrix, post-census health specific flow estimates are produced from age structured NHSCR data.

The models employed by Raymer et al. (2007), as well as being part of a family of generalised linear models, can be further classified as being part of an approach to migration modelling termed by Van Wissen et al. (2008), as the *'demographic approach'*. Demographic approaches essentially try to find structure in the parameters of the model (for example age or sex), and then use these structures to estimate future migration patterns where the parameters may remain stable. This approach is contrasted with explanatory approaches which introduce variables that could be used to explain the various push and pull influences acting on the movements of migrants. In comparing the relative drawbacks and merits of the two approaches in the context of sub-national, regional migration within three European countries, Van Wissen et al. conclude that for short term predictions, demographic migration models performed better than explanatory models (both for outmigration modelling and destination choice), despite the explanatory approach giving a better model fit. It is surmised that this could be due to the changing influence of explanatory factors in different time periods. Despite these conclusions, Van Wissen et al. do not dismiss the use of explanatory models. Demographic models are less useful where 'what-if' scenarios are defined, where an explanatory model on the other hand could be used to examine the potential impact of any changes in explanatory variables. Where the characteristics of origins and destinations are less stable too (such as in areas of high population turnover) explanatory models, it is suggested, may prove more useful.

The ability of models to both help explain and predict internal migration flows is of key importance to anyone studying the phenomenon, as not only do they allow the exploration of the determinates of migration flows and the projection of migration into the future (Champion et al., 1998; Raymer et al., 2006; Van Wissen et al., 2008), but as is demonstrated by Raymer et al. (2007), Rogers et al. (2003a,b) and Raymer and Rogers (2008) migration models are useful tools for 'filling gaps' in the data where statistics are either poor or incomplete. It was noted in the Section 2.2 of this chapter that a harmonised UK-wide set of internal migration flows between censuses are not available, so it follows that it should be possible to use migration models to fill these gaps in the data. A complete time-series of internal migration flows is important for later analysis in this thesis, so the next section of this chapter will discuss the feasibility and pitfalls of employing models to fill gaps in the data, before proposing a methodology to achieve a new complete time-series dataset.

2.4 Enhancing internal migration data sets in the UK - estimating incomplete datasets

A complete inter-district matrix of internal migration flows within the UK is available from the 2001 Census, but not from any of the other data sources described. Flows at this level within England and Wales are available, as are flows within Scotland and Northern Ireland. Taking England and Wales and the flows between these countries and Scotland, the flows between districts comprise around 8% of the total number of flows recorded in 2001. Where there is this lack of UK-wide inter-LAD migration data, there is a choice to be made; either accept the incompleteness of the data and adapt the research accordingly, or attempt to address the data gaps by estimating data where values are missing. Adapting the research is undesirable, whereas producing estimates of the missing data are both feasible and would be beneficial if the estimates are reliable. But just because something is feasible, does not necessarily make it wise. Is it wise to attempt to estimate missing data and is it sensible to carry out research on data which will only ever equate to a 'best guess'? The answer to these questions will of course depend upon the quality of the results produced, and so by definition, the robustness of any estimation methodology employed. Whilst not providing a definitive answer to these questions, examining some of the recent history of population estimation will assist in evaluating the feasibility of producing estimates of migration for substantive research.

2.4.1 Estimation in studying populations

In the UK, there are a huge number studies conducted on the population, and so a large variety of data exist. A brief tour around the website of the UK Data Archive (UKDA - http://www.data-archive.ac.uk/) reveals hundreds of different datasets sampling a wide range of different population attributes, including crime, work, travel, health, diet, and many other areas of demographic interest. Whilst a number of these surveys feature samples which number in the several thousands (if not millions), drawing these samples from across the geographical space of the UK, none manage (or even attempt) to enumerate the entire population. Only the decennial Census of Population tries to do this, and as Rees et al. (2002b) note, as such represents the 'gold standard' of data collection.

But as mentioned at the beginning of Section 2.2, despite being the gold standard, the census data which arrives on the desks of academics, politicians, business managers, other researchers and decision makers, does not in its entirety represent the aggregation to ticks put into boxes by over 60 million hands. It has long been the case that the censuses have not managed to enumerate the entire population. Diamond et al. (2002) note that in 1991, the undercount was around 2% of the population (the 'missing million'). In 2001, the *'One Number Census'* project estimated that the information for around 3 million people (some 6.1% of the population of England and Wales) was missing (ONS, 2003b). Where these data were missing from the 2001 Census, a range of different techniques were employed to impute information so that a final census, representative of the entire population could be released.

Much of the missing data from the 2001 Census was imputed as part of the One Number Census Project with help from a post-census Census Coverage Survey (CCS) (ONS, 2003b,c) - the survey specifically designed and implemented to ensure that those individuals and households not captured by the census were included in this sample. Whilst the methodologies used to impute different parts of the census were varied and complex, the broad assumption was that there was a standard linear relationship between the census and the CCS which meant that relatively straightforward regression analyses could be employed as the main estimation models (Diamond et al., 2002).

Migration data, of course, were not immune from the undercount; if anything they were affected even more. In 1991 most of the undercount fell into the ages of peak migration (those aged in their 20s) (Rees et al., 2002d). Consequently, when data were imputed as part of the One Number Census project in 2001, a greater proportion of migration data compared to other census count data were estimates. Moreover, the post-tabulation SCAM method discussed in Section 2.2 rendered the origin-destination migration data almost unusable at OA level - certainly outside of London (Stillwell and Duke-Williams, 2007).

So with much of the 2001 Census - the population data source hailed as the most reliable and comprehensive in the UK - actually the result of estimation and imputation, it could be reasonably argued that estimation is a legitimate process to go through where gaps in other data sets exist. The argument may continue that the estimation methods used in the census must be robust and produce accurate results with such a large and diverse user community still making use of 2001 Census data almost a decade after its collection. But can we be absolutely sure that this is indeed the case across the board with all census data products? Just because the decision has been made to make adjustments to the data with estimates, and because, in theory, the methods used were sound and made sense for most of the data (at least that used most frequently), does this legitimise the practice for all data - including those parts of the whole dataset which are used less frequently? And can we therefore extend the justification of estimates to non-census data? Just because some estimates are judged to be more-or-less accurate in one situation, it does not necessarily mean that similar estimates will always be accurate or that different methods will yield equally plausible results.

Certainly, with the 2001 Census, an extensive quality assurance programme was undertaken comparing the estimation results with existing administrative sources (ONS, 2003a), to make sure they were in line with this comparable information. The results of this process suggested that the estimates used in the 2001 Census were reliable for all of the different data products. As part of this process though, a key quality assurance mechanism was to consult users and experts who were able to validate the results using their own experience and knowledge. Interestingly, the census quality assurance programme worked the other way, with the census also testing the accuracy of these other datasets. One of the major surprise findings coming out of this programme was that whilst most census estimates were reliable, other population estimates carried out in the inter-censal period were less so, perhaps casting doubt over the validity of estimates produced without the same stringent quality control measures of the 2001 Census. It was noticed that the mid-year population estimates significantly over-estimated the population leading up to mid-2001, and also that adjustments made to the 1991 Census were too severe, especially in urban areas ONS (2003b). One of the major factors influencing the errors in the mid-year estimates was considered to be unreliable data on international migration (UKSA, 2009), thus confirming that quality estimates rely on quality inputs.

As the mid-year population estimates leading up to the 2001 Census were so poor, then does this mean that where there are not the resources to carry out an extensive quality assurance programme such as that carried out on the 2001 Census, estimates cannot be relied upon? Clements and Whitworth (2008) note the dearth of research on uncertainty and error in non-census population estimates and suggest that this can be attributed, at least in part, to the complexity of the process used to produce such estimates. This may be the case, especially where estimates rely on a number of data sources. However, the errors in the mid-year population estimates before the last census can also be attributed to the nature of the estimates themselves; mid-year estimates might be described as *unconstrained* estimates in that they are not constrained by an overall parameter - i.e. a known total population. Therefore the final figure could, in theory, be anything. Mid-year estimates are an assemblage of demographic data - last year's estimate plus or minus births, deaths and migration data - so generally anything is not a possibility, but where an estimate at year t is in part made of $t-1$, errors in $t-1$, $t-2$, $t-3$ etc. will affect the accuracy of the estimate of t. The further away from the last reliable figure the estimate is, the worse it is likely to be. Before the 2001 Census, some retrospective work carried out by Charlton and Chappell (1999) using the 1981 Census as a base produced uncertainty intervals for the components of population change, each year from the reliable base of the 1981 Census - something which could have been used to assess the quality of mid-year estimates. Unfortunately this work was conducted too late for the results to be used to alter pre-2001 mid-year estimates. In the event, the quality of 1992-2001 mid-year estimates was only checked effectively by comparing the 2001 estimate with the census.

Clements and Whitworth (2008) move on from the work of Charlton and Chappell (1999) and propose a new simulation methodology which could help measure error in the mid-year estimates, improving their accuracy as each estimate is produced, through producing an approximate measure of overall quality. With a quantification of error, it would be possible to produce different results that fall within the error band. This approach may well help improve the quality of any unconstrained, mid-year-type estimate. However, other types of estimate which could be termed *constrained* estimates may not need such a sophisticated quality control methodology.

Constrained estimates differ from their unconstrained counterparts in that they will have to conform to some defined parameter in some way. For example, in the recent development of the inter-LAD migration flow estimates for England and Wales based on patient register records (Chappell et al., 2000; Scott and Kilbey, 1999), the final estimates for LADs were constrained to agree with data from the NHSCR at the HA level. Where NHSCR data are updated annually along with the patient register data, there is not the problem of errors propagating through the estimates over time as they have done with mid-year estimates. Furthermore, the constraint means that the estimates have to agree with the most reliable data source, therefore reducing the likelihood of error. Whilst the chances of reducing error are improved by constraints, problems with the principal data source will always affect the quality of final estimates.

So, it is not necessarily the case that estimates which do not go through the stringent quality control measures of the census are unreliable to use. Yes, there have been issues with some unconstrained mid-year population estimates, but where estimates rely on fewer inputs (population estimates rely on births, deaths, internal and international migration components) and can be constrained to known and reliable sources, it should be possible to limit the problems.

2.4.2 Estimating internal migration flows

Thus far, the discussion of estimation has focused mainly on population estimation in general, rather than the specifics of internal migration estimation. It has been noted that estimates of internal migration comprise part of the annual mid-year population estimate. ONS (2007a) provide a comprehensive guide detailing the methodology employed to arrive at their patient register-based internal migration estimates, a summary of which has already been given at the beginning of this chapter. Whilst the early work of Chappell et al. (2000) indicated that patient register data had the potential to improve upon internal migration estimates as they were, a full analysis of this methodology has only recently been carried out (ONS, 2009b). This analysis evaluates some of the key quality issues associated with using patient register data, and identifies them as:

a) evidence of longer or differential time lags between moving and re-registering - especially resulting in an under count of young male and student migrants (Jefferies et al., 2003);

b) the impact of differences between the 2001 population estimates and the 2001 base population in the NHS data;

c) GP registers not capturing those who move during the year, when they are not registered at one of the two mid-year points e.g. 0 yr olds, and international migrants;

d) uncertainty arising from constraining GP register data to NHSCR figures;

e) errors in Scotland / Northern Ireland migration estimates and allocation to LAs;

f) double counting of moves of school boarders (ONS, 2009b).

All of the issues identified by ONS will have some effect on the quality of the estimates produced. Some of the issues are already being addressed; an adjustment for the undercount of student migrants using data from HESA has been trialled (ONS, 2009a) with indications that the new data set is able to improve one of the least reliable components of the patient register estimate - those migrants aged 18-19. Whilst the other issues identified by ONS are also important, of particular relevance to this piece of work is the issue identified with the allocation of flows between England and Wales and the other UK countries, Scotland and Northern Ireland. The problem, as identified by ONS (2009b) is that the flows between these countries are subject to agreement by the different statistical offices (ONS, GROS and NISRA) with the assumption being that the more accurate estimate will be produced by the country of destination. ONS (2009b) recognise that in order to achieve the best overall estimate, the estimate from the origin country will need to be constrained to the destination country estimate. In practice, doing this requires communication between the different statistical agencies, and as yet this has not been achieved. A full UK matrix at the LAD level is still not published.

Producing estimates of internal migration is not the exclusive preserve of the national statistical agencies. Their focus is principally in producing counts of migrants which can feed into mid-year population estimates, and thus whilst these counts have some age disaggregation, there is little other attribute information attached - these estimates are designed for a very specific purpose. In addition these estimates are grounded in empirical observations which measure the movements of people, but in not all situations are data such as these available. So where data are inadequate (i.e. lacking attribute information) or missing (either partially or completely) other parties interested in studying internal migration have developed their own methods to produce usable data - methods will be based around models of migration systems.

As is noted by both Willekens (1999) and Raymer (2007), it is often necessary to develop models which estimate migration flows where data are either inadequate or missing. Addressing data inadequacy and the lack of attribute information in migration estimates, Raymer and colleagues (Raymer et al., 2007; Raymer and Rogers, 2008) have developed techniques which combine attribute rich data with a temporally rich data to estimate new attribute *and* temporally rich data. Where data are missing altogether, Raymer and others (Raymer (2007); Raymer et al. (2006); Raymer and Giulietti (2010); Raymer and Rogers (2007, 2008) have used similar techniques to produce new data. A key finding in all of the work carried by Raymer and colleagues is that in many cases the structures present in migration data remain remarkably stable over time (Raymer and Giulietti, 2010). By structures, Raymer refers to the interactions between different components of models used to describe migration flows - e.g. origin/destination, origin/time, destination/time, origin/age, destination/time etc. Time and again in his work, Raymer demonstrates that these relationships in the data remain relatively stable - a stability which means that estimates produced using a model incorporating these components should be reliable. If the associations between the components were unstable, then the likelihood of error would be much greater. Certainly the mathematical, multiplicative component models and statistical, log-liner models developed by Raymer and others which rely on these structures present in migration data have shown their practical usefulness in producing migration estimates. The MIgration MOdelling for Statistical Analysis (MIMOSA) project (Raymer and Abel, 2008), funded by the European Union to produce international estimates of migration has produced estimates which have used directly by European decision makers, and indirectly as inputs into other research projects, such as the DEmographic and MIgratory Flows affecting European Regions and Cities (DEMIFER) project (ESPON, 2009).

Returning to the questions posed at the beginning of this section - ‘is it wise to estimate internal migration data and is it wise to carry out research based on this data?’ the answer has to be ‘yes’, but with some qualification. The migration estimates discussed here fall into two families: there are the estimates produced by the national statistical agencies which are estimates in the sense that they cannot be 100% accurate as they use proxy, administrative data to attempt to account for the flows of people moving within the countries of the UK. We might term these *'empirical estimates'* and they have associated with them a number of issues related to the quality of data and its ability to measure flows. Then there are the estimates, as produced by Raymer and others which make use of existing data, but where the data are incomplete, use mathematical and statistical techniques to fill the gaps. We might term these *'theoretical estimates'*. The two types of estimate are very different and have associated with them their own issues. It could be argued that the empirical estimates produced by ONS are more robust in that they are essentially observed data which at most have been adjusted using observations from other data sets, but we know from the issues related to the mid-year estimates leading up to 2001, that this is not necessarily the case. The theoretical estimates employed by Raymer and others, on the other hand, rely on detailed deconstruction of migration flows into their constituent parts (origin, destination, interaction and variable) and use the stable relationships between these elements to estimate flows where they do not exist using other observed data. The issues with these theoretical estimates are that the final products are only as good as the strength of the theoretical relationship between the components. The more components in the model the better the potential fit, although the more that can go wrong and affect the final results.

In producing and using estimates, the final results can only be as good as the data which feeds into the estimate and any theoretical assumptions made, therefore the output must always be treated with a degree of caution. However, despite some well documented problems there is a strong history of using estimation techniques in demographic research, and specifically in migration research. Therefore provided the techniques employed are sound, and the final estimated data are treated with a degree of caution, then it should be feasible to produce estimates which can be used with some confidence.

Of course it may well be the case that estimation is only required for part of a dataset. Certainly in this example around 92% of the inter-district flows for Britain are contained in England and Wales alone. Therefore any theoretical estimates will only complete a small proportion of the total data, thus reducing the effect that any inaccuracies in the estimates will have.

2.4.3 Methodology for estimating LAD level UK migration matrices

So if we accept that it is both feasible and sensible to estimate the 8% of missing inter-district flows, a method for producing these estimates needs to be developed. This section will detail such a methodology - one which is analogous to the multiplicative component framework employed by Raymer et al. (2006) and Raymer and Abel (2008). The the estimation will be carried out using migration data organised in a contingency table (origin/destination matrix), using multiplicative components in the form of interior-cell to table-marginal ratios to produce the estimated data. A key difference between the new estimates created here, and those produced by Raymer, is in the constraint used. NHSCR inter-region flows exist for the whole of the UK in a consistent time series, and will be used to constrain the estimates produced. Using constraints at a different geographical level is a technique which the ONS use to adjust their patient register district level empirical estimates, although, this procedure has not been used in the kind of inter-regional theoretical estimates produced by Raymer et al. (2006) and so can be seen as an extension of both. The technique will produce a new set of data will will be used in the analysis carried out in the latter half of this thesis.

Aggregate matrices

The methodology used here for estimating aggregate, LAD to LAD matrices for inter-censal years is a variation of the methodology developed and employed by Dennett and Rees (2010) to estimate internal migrant flows for the whole UK at the Nomenclature of Territorial Units for Statistics - level 2 (NUTS2) geographical level - estimates which were supplied to Eurostat in the absence of any official national statistics produced by the UK national statistical agencies. Dennett and Rees employed a technique whereby estimates were created using existing, publicly available data sets; here, similar datasets will be used.

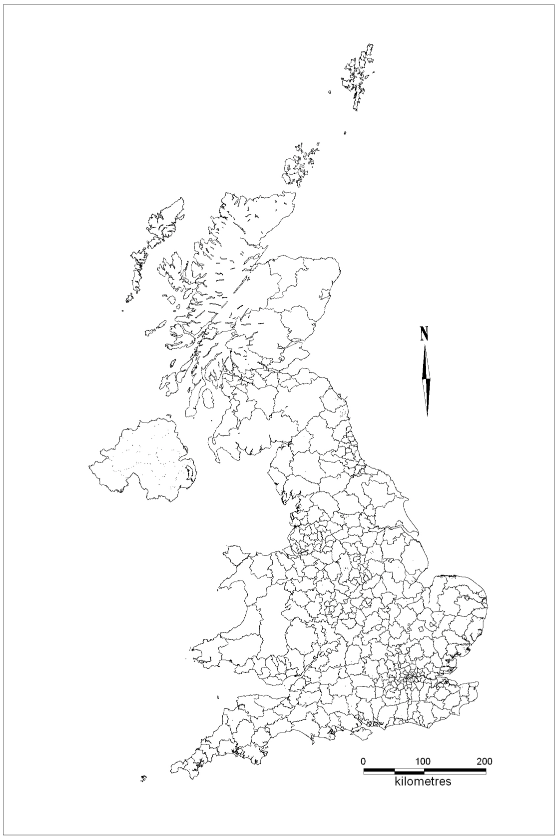


Figure 2.5: The spatial system used for patient register-based internal migration estimates

The spatial system The spatial system used for these estimates is shown in Figure 2.5. It is comprised of the 408 LADs in England, Wales and Scotland with Northern Ireland as a whole zone, making 409 zones in total. The LADs in Britain were chosen as these are the zones which will be used as the principal unit of analysis in this thesis. Northern Ireland does not feature in later analysis for reasons which will be described fully in Chapter 4, so why include this zone, and why not include the districts within Northern Ireland? Firstly, the task of assembling the data was made more straightforward without dropping Northern Ireland. As will be explained in full later, NHSCR matrices for the whole of the UK are used in the estimation process. These are published as whole matrices, so keeping them intact aids data assembly. Furthermore, whilst flows between Britain and Northern Ireland are not needed for the analysis in this thesis, these estimates have been made available more widely through a collaboration between CIDER and the ONS to other users who may find the data of use. The reason why the districts within Northern Ireland were not used is twofold: inter-district flows for Northern Ireland are not readily available from NISRA, despite inter-censal estimates being created in a very similar way to those produced in Scotland, England and Wales from patient registers NISRA (2005). The other reason concerns the very small flows that would need to be estimated between Northern Irish districts and those in the rest of the UK. The 2001 Census indicates that migration flows between Northern Ireland and the rest of the UK are generally only in the order of some some 11,700 outflows and 10,800 inflows in total. When apportioned to individual districts, these very small flow volumes would have wide confidence intervals. Of course, some of the estimated flows between some Scottish districts and those in Wales will also be very small, but census data suggests that the Northern Irish flow estimates would be even smaller thus would likely be even less accurate.

Data Sources The data sources used in the estimation process are set out in Table 2.7, but their relationship to the spatial system and the estimates that are needed to complete the set are shown more clearly in Figure 2.6. The majority of the data matrix is populated with England and Wales patient register data (shown in light blue in Figure 2.6), with these flows being available for a time series of 10 mid-year to mid-year annual periods from mid 1998-1999 to mid 2007-2008. Regional level flows are available from NHSCR data for all mid-year periods (shown in dark blue in Figure 2.6). From the 2001 Census, flows between all origins and destinations depicted in Figure 2.7 are available. The intra-Scotland inter-district level flows shown as data which need to be estimated (coloured grey) in Figure 2.6 are actually available for some years (2001-02 to 2006-07) so will only need to be estimated for the years in the ten year period where no data exist. All data used in the estimation process are publicly available either directly from the websites shown in Table 2.7, or through contacting the organisations concerned.

Estimation Methodology As mentioned at the beginning of this section, the methodology employed is a variation on that used by Dennett and Rees (2010). Whilst implemented in a slightly different way, this method produces identical results; it is adopted here in preference

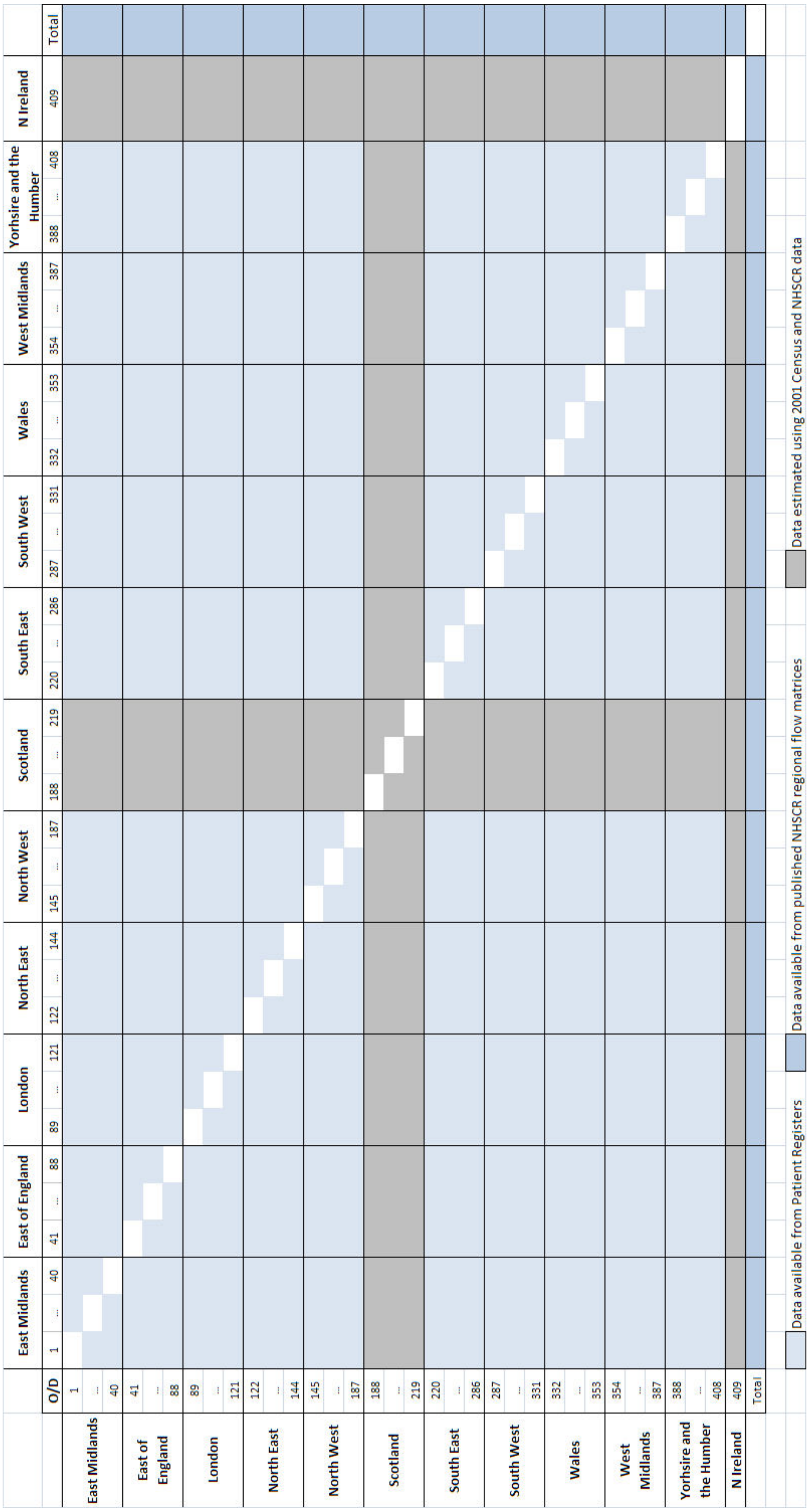


Figure 2.6: Schematic representation of data availability and the gaps which need to be filled through estimation

Table 2.7: Data sources used in the estimation process

Data set	Spatial scale and coverage	Time period(s)	Variables	Source
2001 Census Special Migration Statistics	District UK coverage	Mid 2000-01	Total migrants and migrants by broad age group (8 variables)	http://cider.census.ac.uk
Patient register migration data	District - England and Wales coverage	Annual data from mid 1998-99 – 2007-08. 10 datasets.	Total migrants and by broad age group	http://cider.census.ac.uk
Community Health Index migration data	District - Scotland coverage	Annual data from mid 2001-02 to 2006-07 – 7 datasets	Total migrants and by broad age group	http://www.gro-scotland.gov.uk/
NHSCR migration data (rounded to nearest 100)	Government Office Region. UK coverage	Mid 1998-99 to 2007-08 – 10 datasets	Total migrants	http://statistics.gov.uk/
NHSCR-based in and out flows between Scotland and England and Wales and Northern Ireland (unrounded)	Government Office Region to country and vice versa	Mid 1998-1999 to mid 2007-08	Total migrants	http://www.gro-scotland.gov.uk/

to the original method as it presents an alternative which can be used with data presented in a different format. Consequently an identical algebraic notation to that adopted by Dennett and Rees will be used in this section, with one additional term; this is set out in Table 2.8. The estimation process relies on a key fact and a key assumption. The fact is that within the UK hierarchy of geographies (see Figure 5.1 in Chapter 5) LADs nest neatly within Government Office Regions (GORs). Therefore the population of any given GOR will equal the sum of the population in the constituent districts. Similarly, in or out-migration to or from any GOR will equal the sum of in or out-migration to or from any district within that GOR as long as the flows do not include those between districts within the GOR (see Section 4.2 in Chapter 4 for a full explanation). The assumption is that the relationship between origins and destinations within the migration system remains reasonably consistent over time, i.e. regardless of the total volume of flows, districts in London will always lose more people to districts immediately bordering London than to those in Scotland. The work of Raymer and others described earlier points to this assumption being a safe one to make; examination of the time series of England and Wales PRDS flows adds even more weight, with an average correlation coefficient of 0.99 between the total flow matrices of England and Wales year-on-year from 1999 to 2008.

Armed with this information, it is then quite straightforward to see how the estimation process might be tackled given the data to hand. Consider Figure 2.7. The two matrices C_{ij} and C_{IJ} represent the same census data at two different spatial scales. They are related in that each cell at the finer spatial scale C_{ij} forms part of a cell at the coarser spatial scale C_{IJ}. As such all cells are linked by a number of ratios, including:

C_{ij}/C_{+J} - the ratio of the flow from district origin i to district destination j to the total flows into region J from all regions I. Or:

C_{ij}/C_{IJ} - the ratio of the flow from district origin i to district destination j to the flow from origin region I to destination region J.

Of course a number of other ratios exist, but if we assume that the ratio relationships between cells in the two spatial systems remains consistent over time, the ratios can be used to fill the gaps in the patient register T_{ij} matrix using data from the NHSCR N_{IJ} matrix. In Figure 2.7, these gaps are shaded light grey in the T_{ij} matrix and represent all flows into and out of districts in Scotland and Northern Ireland. The darker grey shading in the centre of T_{ij} represents intra-Scotland flows, which do exist for some years, but not for others. With all cells at the two geographical levels linked by a ratio, it is theoretically possible to estimate all missing data, including these intra-Scotland flows. Other techniques can be adopted for these intra-Scotland flows which may give improved estimates; this will be discussed in full later in this section. Now, a more detailed explanation of the process adopted to arrive at the estimates of flows between England and Wales and Scotland will be given.

Table 2.8: Variables and indices used in the estimation process

Variable, Index	**Description**	**Migration concept used, ranges of indexes**
C	Census migration flows GOR	Transitions
T	Patient register migration flows LAD	Transitions
N	NHS Central Register flows GOR	Moves
M	Target migration flow LAD	Moves
MP	Provisional migration flow before adjustment LAD	Moves
i	Index for district origin	Range of values: $i = 1,409$
j	Index for district destination	Range of values: $j = 1,409$
I	Index for GOR origin	Range of values: $I = 1,12$
J	Index for GOR destination	Range of values: $J = 1,12$
$+$	Indicates summation over index replaced	e.g. $C_{+j} = \sum_i C_{ij}$
$1,2$	Index used to label successive versions of the same M variable	
t	Index for mid-year to mid-year interval in a time series	The time interval extends from mid-year t to mid-year $t+1$ Range of values: t=1998-1999 to 2007-2008
C_{ij}	Census migration flow from district i to district j in year before 2001 Census	Transition between 29 April 2000 and 29 April 2001
T_{ij}^t	Patient register migration flow from district i to district j in year t	Count of reported transitions between 31 July in year t and 31 July in year $t+1$ that cross an area boundary
N_{ij}^t	NHS Central Register migration flow from region I to region J in mid-year t	Count of all moves reported to the NHSCR that cross a region boundary in the time interval indicated
M_{ij}^t	Target migration flow from region i to region j in mid-year t	As target migration estimates are adjusted to NHSCR moves, these estimates can be regarded as move data

The original method used by Dennett and Rees can be explained using the following equations:

$$M_{ij}^t = N_{I+}^t (C_{ij}/C_{I+}) \tag{2.12}$$

$$M_{ij}^t = N_{+J}^t (C_{ij}/C_{+J}) \tag{2.13}$$

$$M_{ij}^t = N_{IJ}^t (C_{ij}/\sum_{i \in I, j \in J} C_{ij}] \tag{2.14}$$

Taking the first two equations, migration between origin i and destination j at the finer geographical scale (NUTS2 in the original paper - LAD in this case) for year t is equal to the ratio of the equivalent flow between i and j measured by the census to the total flows from $I+$

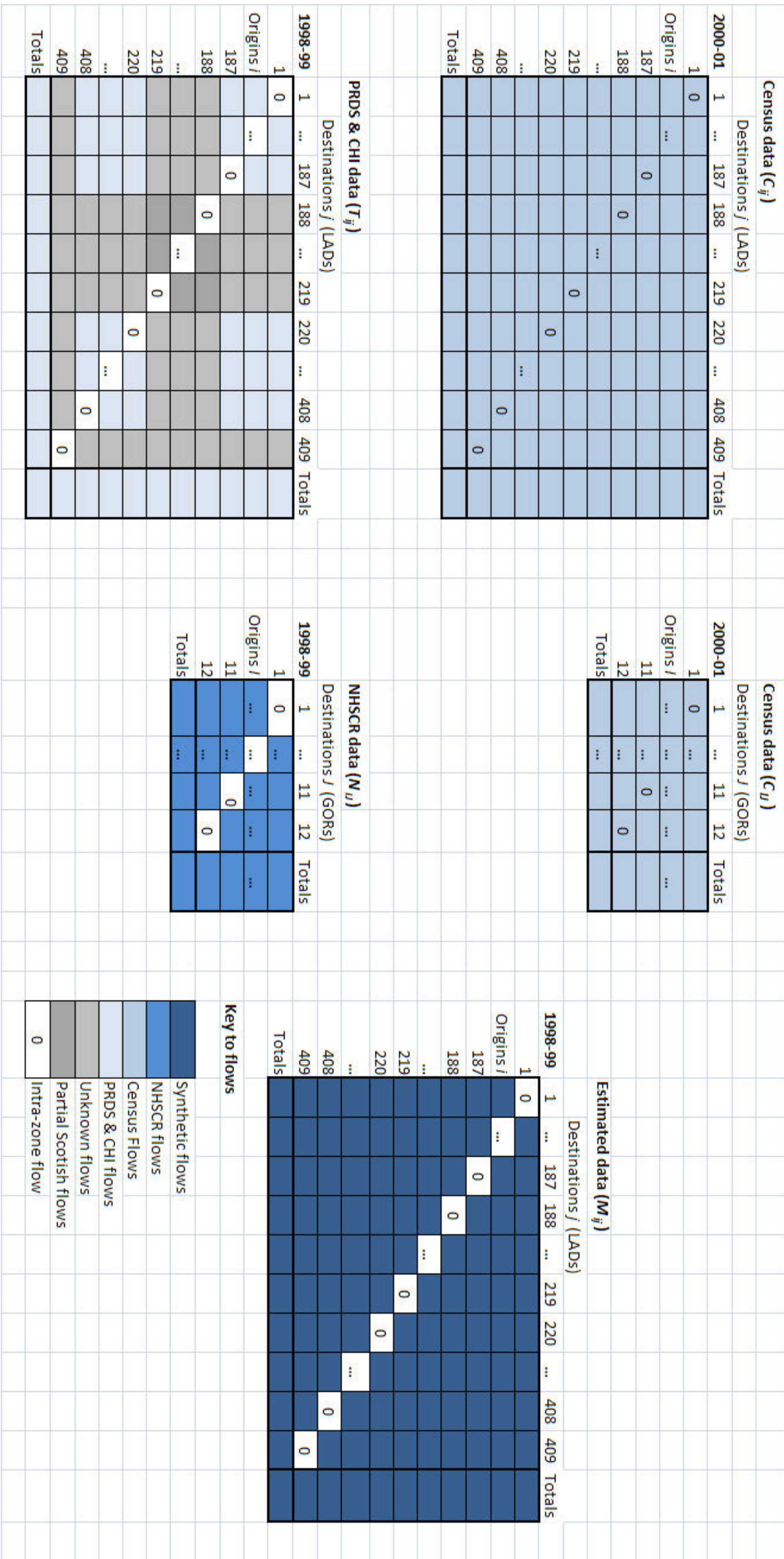


Figure 2.7: The estimation scheme

or to $+J$ the larger GOR (NUTS level 1) zone that i or j is part of, multiplied the NHSCR total flows to or from the equivalent NUTS1/GOR zone. The final equation defines a constraint based on both origins and destinations.

This method was designed to be used on a standard matrix with marginal values being the sum of either the columns or the rows. The method used here produces the same end result, but can be used on a matrix which features two levels of geography. An example of such a matrix is shown in Figure 2.8 where marginal values include flows between districts and regions, as well as row and column sums. Whilst this may appear to be an unnecessary extra layer of complexity, the alternative method is shown here as it could be used with other asymmetric migration data to estimate interior cell values in a symmetrical table[3] - something which was not done in the original article by Dennett and Rees (2010).

So, to estimate the missing inter-district flows, a several stage estimation procedure was designed, such that:

$$M^t_{ij} = M^t_{Ij}(C_{ij}/C_{Ij}) \tag{2.15}$$

or

$$M^t_{ij} = M^t_{iJ}(C_{ij}/C_{iJ}) \tag{2.16}$$

Of course an equivalent doubly constrained model could be specified as in Equation (2.14), but is not shown here for convenience. Either equation could be used to estimate M^t_{ij} flows, with Equation (2.15) using GOR origin to district destination flow ratios to estimate district level flows, and Equation (2.16) using district origin to GOR destination flow ratios to estimate the same flows.

The several stage process is necessary for the formulation of M^t_{Ij} and M^t_{iJ}. These flows are constructed with the following formulae:

$$M^t_{Ij} = MP^t_{Ij}(N^t_{IJ}/MP^t_{IJ}) \tag{2.17}$$
$$M^t_{iJ} = MP^t_{iJ}(N^t_{IJ}/MP^t_{IJ}) \tag{2.18}$$

where

[3]CIDER provides access to tables C0723A and C0723B from the 2001 Census - these tables contain migration flows cross-tabulated by age and ethnicity. This cross-tabulation produces many small flows, so to preserve confidentiality the tables are structured as asymmetric ward-to-region, region-to-ward matrices. The method described here could be applied to tables like these to produce ward-to-ward flow estimates.

$$MP^t_{Ij} = MP^t_{+j}(C_{Ij}/C_{+j}) \quad (2.19)$$
$$MP^t_{iJ} = MP^t_{i+}(C_{iJ}/C_{i+}) \quad (2.20)$$

and where

$$MP^t_{i+} = N^t_{I+}(C_{i+}/C_{I+}) \quad (2.21)$$
$$MP^t_{+j} = N^t_{+J}(C_{+j}/C_{+J}) \quad (2.22)$$

To clarify the process, consider the example estimation scheme set out in Figure 2.8.

Working backwards from Equations (2.21) and (2.22), the first stage in the process was to create a set of ratios from 2001 Census flow information. This is exemplified clearly in Figure 2.8 a to f. These ratios are applied to NHSCR data to create a set of provisional marginal flow totals (Figures 2.8 g and j). These marginal values are then adjusted to the NHSCR distribution (Figures 2.8 k and n), before the marginals are multiplied by the original census distributions to produce a set of M_{ij} flow estimates (Figures 2.8 o and p). In the example shown in Figure 2.8, the in-migration constraint version of the equation (Equation (2.16)) was used. This has the effect that the in-migration (column) marginals sum to the original NHSCR flows, but the out-migration (row) marginals will not. If the example had used the out-migration constrained model, the opposite would be true.

In the estimation process adopted here, both equations were used on different halves of the matrix being estimated, the destination, in-migration constrained model being used on the in-migration estimates, the origin, out-migration model being used on the out-migration estimates. This has the effect of making all marginals in the completed matrix sum to the total flows. Of course an alternative and potentially more reliable way of ensuring origin and destination marginal values sum across both dimensions of the estimated matrix would be to employ an Iterative Proportional Fitting (IPF) method such as the one outlined by Dennett and Rees (2010) and explained in detail by Norman (1999). In their estimation of NUTS2 level migration flows, Dennett and Rees demonstrate only a slight improvement in estimation accuracy on some occasions when IPF is used in preference to either origin or destination constraints, with a drawback of non-convergence for some years. Given this evidence, and given the additional complexity involved in implementing IPF, especially where at the district level flows equalling 0 become far more common and are likely to cause additional problems, the use of origin and destination constraint equations was adopted in this instance. A fruitful avenue of future research would be in the full investigation of IPF procedures in this particular estimation problem.

This identical procedure was carried out for all ten data years, mid 1998-99 to mid 2007-08. Additional data were added to the rounded NHSCR flows in an effort to improve the qual-

a. Example of Census district to district and district to region flow variables

		Scotland				N. Ireland				
	O/D	60QA	60QB	60QC	60QD	N	Scot	NI	C_i	C
East Midlands	00FK	C_{ij}	C_{ij}	C_{ij}	C_{ij}	C_{ij}	C_{iJ}	C_{iJ}	C_{i+}	C_{I+}
	00FN	C_{ij}	C_{ij}	C_{ij}	C_{ij}	C_{ij}	C_{iJ}	C_{iJ}	C_{i+}	
East of England	00JA	C_{ij}	C_{ij}	C_{ij}	C_{ij}	C_{ij}	C_{iJ}	C_{iJ}	C_{i+}	C_{I+}
	00KA	C_{ij}	C_{ij}	C_{ij}	C_{ij}	C_{ij}	C_{iJ}	C_{iJ}	C_{i+}	
	00KF	C_{ij}	C_{ij}	C_{ij}	C_{ij}	C_{ij}	C_{iJ}	C_{iJ}	C_{i+}	
	00KG	C_{ij}	C_{ij}	C_{ij}	C_{ij}	C_{ij}	C_{iJ}	C_{iJ}	C_{i+}	
	09UC	C_{ij}	C_{ij}	C_{ij}	C_{ij}	C_{ij}	C_{iJ}	C_{iJ}	C_{i+}	
	East Mid	C_{Ij}	C_{Ij}	C_{Ij}	C_{Ij}	C_{Ij}	C_{IJ}	C_{IJ}		
	East of Eng	C_{Ij}	C_{Ij}	C_{Ij}	C_{Ij}	C_{Ij}	C_{IJ}	C_{IJ}		
	C_j	C_{+j}	C_{+j}	C_{+j}	C_{+j}	C_{+j}				
	C_J	C_{+J}				C_{+J}				

(a)

b. Example of Census district to district and district to region flow counts

		Scotland				N. Ireland				
	O/D	60QA	60QB	60QC	60QD	N	Scot	NI	Total Out	Total Out
East Midlands	00FK	561	330	114	280	132	1285	132	1417	4079
	00FN	1494	595	180	226	167	2495	167	2662	
East of England	00JA	480	517	61	111	66	1169	66	1235	7168
	00KA	461	272	90	117	78	940	78	1018	
	00KF	803	601	136	217	371	1757	371	2128	
	00KG	634	601	94	195	228	1524	228	1752	
	09UC	310	268	104	133	220	815	220	1035	
	East Mid	2055	925	294	506	299	3780	299		
	East of Eng	2688	2259	485	773	963	6205	963		
	Total In	4743	3184	779	1279	1262				
	Total In	9985				1262				

(b)

c: Example of Census GOR variables

O/D	Scotland	N. Ireland	C_I
East Midlands	C_{IJ}	C_{IJ}	C_{I+}
East of England	C_{IJ}	C_{IJ}	C_{I+}
C_J	C_{+J}	C_{+J}	

(c)

d: Example of Census GOR counts

O/D	Scotland	N. Ireland	Total Out
East Midlands	3780	299	4079
East of England	6205	963	7168
Total In	9985	1262	

(d)

Figure 2.8: Exemplification of the estimation procedure

e. Ratio variables for estimation

		Scotland				N. Ireland				
	O/D	60QA	60QB	60QC	60QD	N	Scot	NI	C_i	C_I
East Midlands	00FK	C_{ij}/C_{Ij}	C_{ij}/C_{Ij}	C_{ij}/C_{Ij}	C_{ij}/C_{Ij}	C_{ij}/C_{iJ}	C_{iJ}/C_i	C_{iJ}/C_i	C_{i+}/C_{I+}	
	00FN	C_{ij}/C_{Ij}	C_{ij}/C_{Ij}	C_{ij}/C_{Ij}	C_{ij}/C_{Ij}	C_{ij}/C_{iJ}	C_{iJ}/C_i	C_{iJ}/C_i	C_{i+}/C_{I+}	
East of England	00JA	C_{ij}/C_{Ij}	C_{ij}/C_{Ij}	C_{ij}/C_{Ij}	C_{ij}/C_{Ij}	C_{ij}/C_{iJ}	C_{iJ}/C_i	C_{iJ}/C_i	C_{i+}/C_{I+}	
	00KA	C_{ij}/C_{Ij}	C_{ij}/C_{Ij}	C_{ij}/C_{Ij}	C_{ij}/C_{Ij}	C_{ij}/C_{iJ}	C_{iJ}/C_i	C_{iJ}/C_i	C_{i+}/C_{I+}	
	00KF	C_{ij}/C_{Ij}	C_{ij}/C_{Ij}	C_{ij}/C_{Ij}	C_{ij}/C_{Ij}	C_{ij}/C_{iJ}	C_{iJ}/C_i	C_{iJ}/C_i	C_{i+}/C_{I+}	
	00KG	C_{ij}/C_{Ij}	C_{ij}/C_{Ij}	C_{ij}/C_{Ij}	C_{ij}/C_{Ij}	C_{ij}/C_{iJ}	C_{iJ}/C_i	C_{iJ}/C_i	C_{i+}/C_{I+}	
	09UC	C_{ij}/C_{Ij}	C_{ij}/C_{Ij}	C_{ij}/C_{Ij}	C_{ij}/C_{Ij}	C_{ij}/C_{iJ}	C_{iJ}/C_i	C_{iJ}/C_i	C_{i+}/C_{I+}	
	East Mid	C_{Ij}/C_+	C_{Ij}/C_+	C_{Ij}/C_+	C_{Ij}/C_+	C_{Ij}/C_{+j}				
	East of Eng	C_{Ij}/C_+	C_{Ij}/C_+	C_{Ij}/C_+	C_{Ij}/C_+	C_{Ij}/C_{+j}				
	C_j	C_{+j}/C_{+J}	C_{+j}/C_{+J}	C_{+j}/C_{+J}	C_{+j}/C_{+J}	C_{+j}/C_{+J}				
	C_J									

(e)

f. Ratios for estimation

	C_{ij}/C_{+J}	Scotland				N. Ireland				
	O/D	60QA	60QB	60QC	60QD	N	Scot	NI	Total Out	Total Out
East Midlands	00FK	0.273	0.357	0.388	0.553	0.441	0.907	0.093	0.347	
	00FN	0.727	0.643	0.612	0.447	0.559	0.937	0.063	0.653	
East of England	00JA	0.179	0.229	0.126	0.144	0.069	0.947	0.053	0.172	
	00KA	0.172	0.120	0.186	0.151	0.081	0.923	0.077	0.142	
	00KF	0.299	0.266	0.280	0.281	0.385	0.826	0.174	0.297	
	00KG	0.236	0.266	0.194	0.252	0.237	0.870	0.130	0.244	
	09UC	0.115	0.119	0.214	0.172	0.228	0.787	0.213	0.144	
	East Mid	0.433	0.291	0.377	0.396	0.237				
	East of Eng	0.567	0.709	0.623	0.604	0.763				
	Total In	0.475	0.319	0.078	0.128	1.000				
	Total In									

(f)

g: Example of NHSCR GOR flow variables for 2006-7

O/D	Scotland	N. Ireland	N_I
East Midlands	N_{IJ}	N_{IJ}	N_{I+}
East of England	N_{IJ}	N_{IJ}	N_{I+}
N_J	N_{+J}	N_{+J}	

(g)

h: Example of NHSCR GOR flow counts for 2006-7

O/D	Scotland	N. Ireland	Total Out
East Midlands	4300	350	4650
East of England	7800	1100	8900
Total In	12100	1450	13550

(h)

Figure 2.8: Exemplification of the estimation procedure

i: Provisional migration flow variables before adjustment to NHSCR data

	O/D	Scotland				N. Ireland				
		60QA	60QB	60QC	60QD	N	Scot	NI	MP_i	MP_I
East Midlands	00FK	MP_{ij}	MP_{ij}	MP_{ij}	MP_{ij}	MP_{ij}	MP_{iJ}	MP_{iJ}	MP_{i+}	MP
	00FN	MP_{ij}	MP_{ij}	MP_{ij}	MP_{ij}	MP_{ij}	MP_{iJ}	MP_{iJ}	MP_{i+}	
East of England	00JA	MP_{ij}	MP_{ij}	MP_{ij}	MP_{ij}	MP_{ij}	MP_{iJ}	MP_{iJ}	MP_{i+}	MP
	00KA	MP_{ij}	MP_{ij}	MP_{ij}	MP_{ij}	MP_{ij}	MP_{iJ}	MP_{iJ}	MP_{i+}	
	00KF	MP_{ij}	MP_{ij}	MP_{ij}	MP_{ij}	MP_{ij}	MP_{iJ}	MP_{iJ}	MP_{i+}	
	00KG	MP_{ij}	MP_{ij}	MP_{ij}	MP_{ij}	MP_{ij}	MP_{iJ}	MP_{iJ}	MP_{i+}	
	09UC	MP_{ij}	MP_{ij}	MP_{ij}	MP_{ij}	MP_{ij}	MP_{iJ}	MP_{iJ}	MP_{i+}	
	East Mid	MP_{Ij}	MP_{Ij}	MP_{Ij}	MP_{Ij}	MP_{Ij}	MP_{IJ}	MP_{IJ}		
	East of Eng	MP_{Ij}	MP_{Ij}	MP_{Ij}	MP_{Ij}	MP_{Ij}	MP_{IJ}	MP_{IJ}		
	MP_j	MP_{+j}	MP_{+j}	MP_{+j}	MP_{+j}	MP_{+j}				
	MP_J	MP_J				MP_J				

(i)

j: Provisional migration flow counts before adjustment to NHSCR data

	O/D	Scotland				N. Ireland				
		60QA	60QB	60QC	60QD	N	Scot	NI	Total Out	Total Out
East Midlands	00FK						1465	150	1615	
	00FN						2844	190	3035	
East of England	00JA						1451	82	1533	
	00KA						1167	97	1264	
	00KF						2182	461	2642	
	00KG						1892	283	2175	
	09UC						1012	273	1285	
	East Mid	2490	1121	356	613	344	4581	344		
	East of Eng	3257	2737	588	937	1106	7519	1106		
	Total In	5748	3858	944	1550	1450				
	Total In									

(j)

k: Provisional migration flow variables aggregated to GOR level

O/D	Scotland	N. Ireland	C_i
East Midlands	MP_{IJ}	MP_{IJ}	MP_{I+}
East of England	MP_{IJ}	MP_{IJ}	MP_{I+}
C_j	MP_{+J}	MP_{+J}	

(k)

l: Provisional migration flow counts aggregated to GOR level

O/D	Scotland	N. Ireland	C_i
East Midlands	4581	344	4924
East of England	7519	1106	8626
C_j	12100	1450	

(l)

Figure 2.8: Exemplification of the estimation procedure

m: Ratio of NHSCR GOR flow variables to provisional GOR flow variables

O/D	Scotland	N. Ireland	
East Midlands	N_{IJ}/MP_{IJ}	N_{IJ}/MP_{IJ}	N_{I+}/MP_{I+}
East of England	N_{IJ}/MP_{IJ}	N_{IJ}/MP_{IJ}	N_{I+}/MP_{I+}
	N_{+J}/MP_{+J}	N_{+J}/MP_{+J}	

(m)

n: Ratio of NHSCR GOR flow counts to provisional GOR flow counts

O/D	Scotland	N. Ireland	C_I
East Midlands	0.939	1.019	0.944
East of England	1.037	0.994	1.032
C_J	1.000	1.000	

(n)

o: Final migration flow variables after adjustment to NHSCR data

		Scotland				N. Ireland				
	O/D	60QA	60QB	60QC	60QD	N	Scot	NI	MP_i	MP_I
East Midlands	00FK	M_{ij}	M_{ij}	M_{ij}	M_{ij}	M_{ij}	M_{iJ}	M_{iJ}	M_{i+}	M_I
	00FN	M_{ij}	M_{ij}	M_{ij}	M_{ij}	M_{ij}	M_{iJ}	M_{iJ}	M_{i+}	
East of England	00JA	M_{ij}	M_{ij}	M_{ij}	M_{ij}	M_{ij}	M_{iJ}	M_{iJ}	M_{i+}	M_I
	00KA	M_{ij}	M_{ij}	M_{ij}	M_{ij}	M_{ij}	M_{iJ}	M_{iJ}	M_{i+}	
	00KF	M_{ij}	M_{ij}	M_{ij}	M_{ij}	M_{ij}	M_{iJ}	M_{iJ}	M_{i+}	
	00KG	M_{ij}	M_{ij}	M_{ij}	M_{ij}	M_{ij}	M_{iJ}	M_{iJ}	M_{i+}	
	09UC	M_{ij}	M_{ij}	M_{ij}	M_{ij}	M_{ij}	M_{iJ}	M_{iJ}	M_{i+}	
	East Mid	M_{ij}	M_{ij}	M_{ij}	M_{ij}	M_{ij}	M_{IJ}	M_{IJ}		
	East of Eng	M_{ij}	M_{ij}	M_{ij}	M_{ij}	M_{ij}	M_{IJ}	M_{IJ}		
	MP_j	M_{+j}	M_{+j}	M_{+j}	M_{+j}	M_{+j}				
	MP_J	M				M				

(o)

p: Final migration flow counts after adjustment to NHSCR data
$M_{ij} = M_{iJ}(C_{ij}/C_{+J})$

		Scotland				N. Ireland				
	O/D	60QA	60QB	60QC	60QD	N	Scot	NI	Total Out	Total Out
East Midlands	00FK	638	375	130	319	155	1375	153	1525	
	00FN	1700	677	205	257	195	2670	194	2866	
East of England	00JA	603	650	77	140	75	1506	81	1582	
	00KA	580	342	113	147	89	1211	96	1304	
	00KF	1009	755	171	273	424	2263	458	2726	
	00KG	797	755	118	245	260	1963	281	2244	
	09UC	390	337	131	167	251	1050	272	1326	
	East Mid	2338	1052	334	576	350	4300	350		
	East of Eng	3379	2840	610	972	1100	7800	1100		
	Total In	5748	3858	944	1550	1450				
	Total In	12100				1450				

(p)

Figure 2.8: Exemplification of the estimation procedure

ity of the estimates. Freely available from GROS (http://www.gro-scotland.gov.uk/statistics/migration/index.html) are unrounded in and out flows between Scotland and England and Wales, and Scotland and Northern Ireland for the four quarters in each year. These are derived from the exact same NHSCR data as the full flow matrices available from ONS. Where these unrounded data were available, NHSCR matrices were updated with the new data. These data were then used as inputs in the estimation model.

As mentioned earlier, whilst the method described above could be used to estimate flows between council areas within Scotland (the dark grey square in Figure 2.6), a different technique was chosen for the estimates in this section of the flow matrix. The adoption of an alternative methodology was driven primarily by the availability of district level, intra-Scotland flow data for the years 2001-02 to 2006-07. This meant that only four years of estimated data were required (1998-99 to 2000-01 and 2007-08). As an aside, intra-Scotland flows between HAs are available for a far more extensive time series, but converting these data to the smaller district/council area geography is exceedingly difficult where flows within HAs (which are not recorded) become flows between districts at the lower level. An additional consequence of the different scale of measurement being that the total intra-Scotland flows recorded by HA level flow data are considerably lower than those measured by flows between council areas.

The availability of a short time series of some inter-council area flow data presented the opportunity to compare a trend-based estimate for intra-Scotland flows, with an estimate produced using the main estimation method. The methodology for producing a trend estimate can be described as follows: Firstly, a year-on-year percentage difference between known M_{ij} flows in two years needs to be calculated.

$$df_{ij}^{t} = \frac{M_{ij}^{t} - M_{ij}^{t-1}}{M_{ij}^{t}} \tag{2.23}$$

Where df_{ij}^{t} = the difference between the M_{ij} flow year t, and the M_{ij} flow in year $t-1$ as a proportion of year t. Where several years of data are available, an average of the df_{ij}^{t} proportion - $\overline{df_{ij}^{t}}$ - is taken:

$$\overline{df_{ij}^{t}} = \frac{1}{n}\sum_{ij=1}^{n} df_{ij}^{t} \tag{2.24}$$

Where n = the number of years over which the average is taken, this average proportion (Equation (2.24)) is then used to estimate M_{ij} flows in the years preceding the available data:

$$M_{ij}^{t} = M_{ij}^{t-1}\overline{df_{ij}^{t}} \tag{2.25}$$

The method can be adapted using M_{ij}^{t+1} to estimate flows in years after the available data set. The benefit of having some intra-Scotland flow data available is that it is possible to directly compare the results of the different estimation methods with the observed data. A summary of the comparison between the two estimation methods is shown in Table 2.9

Table 2.9: A comparison between observed 32x32 intra-Scotland flow matrix data and estimates of the same matrix produced by two different estimation methods

		Trend estimate			**Estimate using marginal constraints**		
	Year	Internal cell matrix	in flows	out flows	Internal cell matrix	in flows	out flows
Correlation coefficient	2001-02	0.991	0.996	0.999	0.985	0.989	0.997
	2002-03	0.993	0.992	0.999	0.98	0.988	0.996
	2003-04	0.993	0.995	0.999	0.978	0.978	0.995
	2004-05	0.995	0.997	0.999	0.982	0.987	0.997
	2005-06	0.994	0.996	0.999	0.98	0.985	0.996
	2006-07	0.998	0.999	1.000	0.983	0.989	0.995
Absolute mean difference	2001-02	20	243	220	24	420	358
	2002-03	18	257	140	24	397	323
	2003-04	17	206	195	36	912	894
	2004-05	16	163	115	30	616	556
	2005-06	17	229	185	26	501	372
	2006-07	9	140	115	23	389	302

For each year where a comparison was possible, the trend estimate produced better results than the estimate using marginal constraints, both in terms of the correlation coefficient, and the absolute mean difference in values. For the internal cell values in the matrix (the M_{ij} flows), the trend estimate produced a result with a much stronger correlation with the original data. This was also the case with the matrix marginal values. The much stronger correlation coefficients across the board confirmed the preferred choice of this method of estimation over of the other.

Age Specific Matrices

A set of ten (annual) aggregate flow matrices were constructed using the methods described above. The PRDS data supplied by ONS, however, also contains an age breakdown into 8 broad age groups of varying size - 0-15, 16-19, 20-24, 25-29, 30-44, 45-59, 60-74, 75+. The next challenge in the estimation process was to see whether it was feasible to produce estimates of these age-specific flows for the missing Scotland and Northern Ireland data.

In theory, a similar technique to the one used to produce the aggregate estimates could be used. Where earlier the assumption was that the associations displayed between migrant origins and destinations will remain relatively constant over time, and therefore we can use detailed information from the census at one point in time to help predict flows at another; a similar assumption could be made that the age breakdown of flows between origins and destinations also varies little over time. Of course, this could be a wild assumption, but a number of pieces of research point to the age structure of migrants being relatively constant over time: the work of Raymer and Rogers on North America (2007) Raymer et al. on Italy (2006), Tobler (1995), Rogers et al. (2002, 2003b) and Bates and Bracken (1987) all show that the age structures present in migrant flows (such as the age related propensity to migrate or where migrants of different ages tend to move from and to), remain relatively stable over time. This stability means that given the age patterns of migrants at one point in time, it is feasible to apply these patterns to another point in time to produce estimates.

Migration flows from the census SMS can be broken down into the exact same broad age groups as the England and Wales PRDS data. Therefore it is feasible to use the equation:

$$M_{ij}^{ta} = M_{ij}^{t}(C_{ij}^{a}/C_{ij}) \tag{2.26}$$

To apportion aggregate migration flows to each of the broad age groups in the England and Wales PRDS, where the superscript a refers to an age group in the data.

Whilst this appears to be a straightforward estimation scheme, it is not without issues which could affect the quality of the estimates produced. The first issue relates to the effect of disclosure control on the census tables. There are three main district level individual migrant tables in the SMS (Table 2.2); SMS Table 1 which contains the age information, and then Tables 3 and 4 which disaggregate the data by ethnicity and limiting long term illness by household status respectively. Analysis of these three tables by origin and destination pair reveals that of all the origin/destination pairs in SMS Table 1 with a flow of 0, around 18,000 of the same pairs record a non-zero entry in SMS Tables 3 and 4. This suggests that at very least there are around 18,000 flows which are ignored by Table 1, the actual number will be much higher. This is a problem for the estimates as wherever a zero is recorded, any multiplication or division by zero in the estimation formula means that it propagates through, even where the total flows are non-zero.

At the district level, M_{ij} flows of zero are not uncommon, so will always cause problems in this type of estimate, but where these zeros are the result of small cell adjustment it is possible to adjust at least some of them to a non-zero entry to reduce error as much as possible. The solution chosen was to adjust all C_{ij} flows in SMS Table 1 to the average C_{ij} flow from SMS Tables 1, 3 and 4. In overall terms, where before around 36% of all flows were recorded as zero, this was reduced to 26% after adjustment to the average of the three tables. Where a new non-zero flow was produced, the average C_{ij}^{a}/C_{ij} for all origin/destination pairs was used to allocate the total flow to an age group. This figure was then rounded so given a large enough total, data would also appear in the age disaggregation. Of course, rounding in this way will produce rounding errors which means that the age disaggregated flows will not necessarily always equal the aggregate flow; and substituting the previous SMS Table 1 total for the average of Tables 1, 3 and 4 means that this will be the case even where non-zero values were present in the original data. However, it was decided that this was an acceptable trade off given the benefits of producing these additional flow data. Indeed, the PRDS national statistics released by the ONS suffer an identical error, with age disaggregated flows not equaling the published totals.

So once a set of adjusted census age breakdowns were produced for each origin/destination pair, these were then converted into a ratio of C^a_{ij}/C_{ij}. This ratio was then applied to each M_{ij} flow to produce an estimated age breakdown. These estimates were then rounded into integers, and added to the existing England and Wales PRDS data producing the final set of 90 matrices. Again, with an age breakdown already available in the England and Wales PRDS, only around 8% of the data in these final matrices are the result of this new estimation procedure.

2.5 Assessing the quality of the data - comparison with 2001 Census and 2001 patient register data estimates

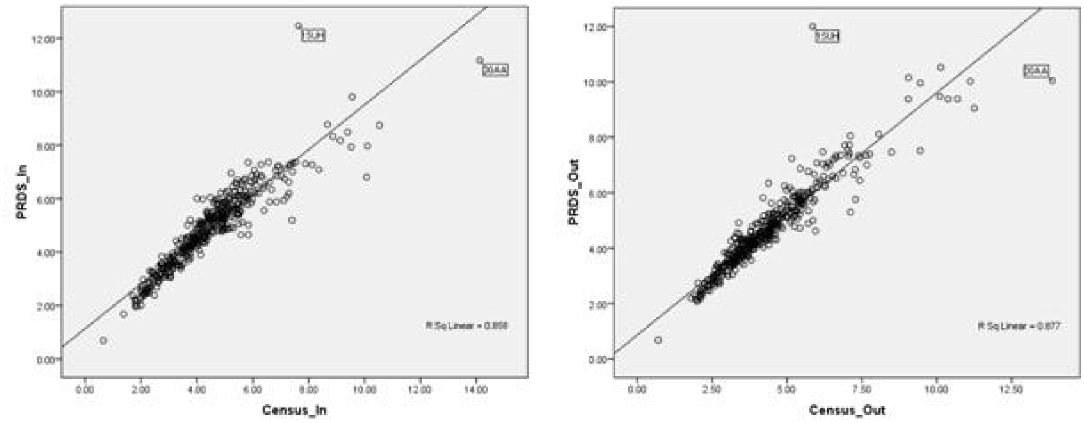
The complete patient register-based data sets now allow for a comprehensive analysis of the migration patterns in Britain over a ten year period, and indeed will be used for this purpose in Chapters 7 and 8. Some potential issues, especially with the age-disaggregated data, were identified during the estimation process, so before embarking upon any analysis later in the thesis, it will be useful to assess the quality of the estimated data. Assessing the quality of estimates is difficult as by their very definition they exist to fill a gap where comparable data do not exist. In this case, however, comparable data from the 2001 Census do exist for the 2001 PRDS estimate year. This of course is not a perfect method of validation given the differing time periods covered by the two data sets (Mid April 2000 to Mid April 2001 for the census and June 2000 to June 2001 for the PRDS data) and inconsistencies in the data such as the undercount of young males in the PRDS. Moreover, the issues identified with the modelling of zero flows from zeros present in the census will not be highlighted by a comparison with the census. But, in the absence of any other data which could be used to validate the estimates, comparison with the census is the only option available.

A summary of the 2001 PRDS-based estimate compared with the 2001 Census data is shown in Table 2.10. The two datasets have similar flow volumes. Overall, the correlation coefficient (calculated by comparing every origin/destination flow) is very high at 0.98. The new patient register based data does estimate some 220,000 more flows than the census in 2001, however, in terms of the total number of flows being measured - some 2.5 million moves overall, this only around 9% more. The reasons for these slight overall differences are difficult to explain definitively, but error in the adjustments made to the census for under-enumeration and the additional flows recorded by the NHSCR (i.e. recording multiple moves in a year rather than single transitions in the same period) are likely to have some effect. The different datasets also cover slightly different populations, for example migrants who die and infant migrants are included in the NHSCR data and excluded from the census, and armed forces migrants, prisoners and psychiatric patients are included in the census but excluded in the NHSCR (Boden et al., 1992).

Whilst the overall new estimate totals are slightly higher than the census, there is variation in the differences between the new estimates and census data by age group. At the three age

Table 2.10: Comparison between UK internal migration flows recorded in the 2001 Census and new PRDS based estimates

	Total	Sum ages	0-15	16-19	20-24	25-29	30-44	45-59	60-74	75+
Census	2469616	2469616	368926	217050	474160	404517	607430	223702	105828	68003
PRDS estimate	2691382	2651640	447471	207808	429558	395944	697186	265395	127769	80509
Difference	221766	182024	78545	-9242	-44602	-8573	89756	41693	21941	12506
% Difference	108.98	107.371	121.29	95.742	90.593	97.881	114.776	118.638	120.733	118.39
Correlation coefficient	0.983		0.972	0.896	0.945	0.967	0.983	0.96	0.914	0.89



(a) PRDS vs census in-flows

(b) PRDS vs census out-flows

Figure 2.9: PRDS versus census inflow and outflow rates across the 409 zone spatial system

groups of peak migration (16-19, 20-24 and 25-29) the new estimates are noticeably lower than the census. A difference exaggerated when one considers that the patient register data estimate more migrants in total than the census. The undercount of young, (especially male) migrants in patient register data has been well documented (Chappell et al., 2000; Scott and Kilbey, 1999), with the received wisdom being that these younger migrants are much less likely to re-register with a new doctor if they move residence than older migrants. Undercounting these migrants has the effect of overcounting migrants who are slightly older, as when the migrants do finally re-register with a doctor, they are older than they were when the migration event occurred. This may account for at least some of the additional migrants in the 30-44 age group. For most age groups, the correlation coefficient between all origin/destination flows is very high. Only the 16-19 and 75+ age groups have coefficients lower than 90%.

It could be argued that this under-counting of young male migrants should be accounted for in these estimates. Raymer and Giulietti (2009) and Smith et al. (2009) propose a methodology which assumes the proportions of male and female migrants are approximately equal (based on the counts of male and female migrants in the 1991 and 2001 censuses) and as such applies a national female/male ratio of migrants from these data to adjust up the male migrant estimates.

For an even more accurate adjustment, these ratios should probably be applied on a district-by-district basis. Here, such an adjustment has not been applied where flows in the age group where the under-count is occurring are already significantly higher than in other age groups and applying such adjustments at this stage would add an extra layer of complexity to the estimation process. That is not to say, however, that this would not be a useful exercise to undertake were subsequent improvements to these estimates embarked upon.

Across the spatial system, there is a high level of association between flows measured by the PRDS and the census. Figure 2.9 plots in- and out-migration rates for each zone in the spatial system, measured by the two data sets. The R^2 correlation coefficient is high for both in (0.86) and out (0.88) migration with only the Isles of Scilly (15UH) and the City of London (00AA) deviating noticeably from the regression line.

Of course a certain amount of error will also have been introduced in the estimation process, but as discussed already only 8% of the data being used is a result of the post-tabulation estimation. Where intra-Scotland flows were available, this figure is even lower. Consequently, the amount of error introduced as a result of the estimation should not affect the data unduly. Both the census and the new patient register based estimates are imperfect measures of the true internal migration flows, but with high correlations between the total and age disaggregated flows for each dataset, even if we cannot quantify the accuracy of either dataset, it is possible to conclude that the imperfections in the datasets are at least similar. Improved methods of estimating especially the age disaggregated flows in the new data could be trialled and tested, but at a cost to subsequent analysis as this would take time. Where the current estimation methodology produces plausible results, then it could be argued that where the aim is to understand general internal migration patterns and processes, more benefit can be gained from analysing the estimates than could be gained from extensive testing and evaluation of alternatives at this stage.

2.6 Summary and Conclusions

At the beginning of this chapter, it was stated that the principal aim was to set the terms of reference for the work which followed, initially though arriving at an understanding of the concept of internal migration through an examination of the datasets available in the UK. Section 2.1 reviewed the datasets available and discovered that there are two main sources of data which are available to researchers - the census and NHS records, either from the NHSCR or the PRDS. When using census data internal migration is recorded as a single transition event from an address recorded at the beginning of a year period to an address at the end of a year period. Internal migration measured by NHSCR data - and by extension PRDS data adjusted to NHSCR totals - is the summation of all moves crossing defined geographical boundaries over a similar annual period. As such internal migration data from these two main systems differs along with the accompanying definition of an internal migrant.

The terms of reference for this thesis were also set though the definition of a generic spatial framework: a discrete zone system. Section 2.3.2 demonstrated how such a system can be represented schematically through the use of origin/destination matrices and how the analysis of data represented in such matrices can be assisted through the calcuation of intensities and through the fitting of explanatory models.

The identification of shortcomings in the data along with the definition of a general schema for analysis and the description of modelling methods led onto the second aim of this chapter which was to investigate whether it was feasible to augment existing inadequate data though modelling an estimation techniques in order that analysis in the latter chapters of this thesis could be carried out more effectively. Section 2.4 of this chapter explored this theme in detail, demonstrating how first it was theoretically feasible before secondly exemplifying a technique to achieve a harmonised UK time-series of internal migration data based upon the PRDS and CHI data available from the ONS and GROS.

This chapter has fully explored the topic of internal migration data, definitions and methodological approaches, but the scene is not yet fully set for the thesis. Before being able to embark upon a study of internal migration in Britain it is imperative to contextualise any analysis though examining substantive work which has gone before. Therefore the next chapter will do this though a selective review of the internal migration literature.

Chapter 3

Analysing internal population migration: substantive analyses - a historical perspective

3.1 Introduction

The aim of this piece of work is to understand internal migration in Britain at the start of the 21^{st} century. The last chapter set some of the theoretical and methodological terms of reference for the thesis through defining internal migration, examining the data available in the UK and outlining some of the techniques which can be used to both explore the data and enhance it where there are inadequacies. As explained in the concluding comments, it will be impossible to situate and make sense of the patterns of internal migration discovered in this thesis without first examining existing substantive examples of internal migration research, both in this country and further afield. 'How much further afield?' is one question that necessarily follows, with the answer being that useful context can only really be given satisfactorily through examining countries with similar demographic, political, social and economic structures to Britain. For example a number of pieces of research have been carried out on internal migration in China at various points in time (Fan, 2005a,b; He and Pooler, 2002; Liang et al., 2002; Liang and White, 1996; Shen, 1996; Wei, 1997). Many of the migration patterns within China have been and still are being influenced by the particular political system in place and the development of the economic system within the country. This means that despite the concessions of Fan (2005a) that theories of migration applicable to capitalist economies are beginning to hold more relevance now that central planning is exerting less influence, the new internal migration flows which are occurring as a result are more akin to the early urbanising moves of western developing capitalist economies of many years ago. China may be an extreme example of a country with a very different profile to Britain, but even in newly emerging post-communist democracies like Russia, different factors act to influence internal migration patterns - as is demonstrated by White (2009).

Indeed these differences might been seen as evidence for the kind of ‘mobility transition’ first postulated by Zelinsky (1971) as a corollary of the demographic and epidemiological transitions occurring through social, economic and political development. Differences which makes comparison of countries at different points on the mobility transition continuum somewhat problematic. Certainly this is something which is identified by Long (1991) who concentrates his review of residential mobility on ‘developed’ countries. Therefore this short review will follow the lead of Long and will concentrate only on internal migration in mature western capitalist democracies which have broadly similar characteristics to Britain. The aim of the first part of the review will be to identify the general features of internal migration which have been observed within a range of different countries. The focus will be on recent patterns from the latter half of the twentieth century and will look to draw together common themes and patterns in relation to both the migration flows and migrant characteristics. In the second half of this chapter the focus will turn to Britain; a historical overview of the internal migration landscape leading up to the turn of the 20th century will be provided. In realising both of these aims, the context for the rest of the thesis will be set, enabling a the work that follows to be effectively positioned in the canon of substantive research which has occurred to date.

3.2 Internal migration across selected industrialised western democracies

Baccaini (2007) gives an overview of inter-regional migration in France between the 1950s and the late 1990s using data from French censuses and observing a number of important patterns. Using net-migration as the measure of choice, she notes the most remarkable story concerns the changing fortunes of the Paris region across the fifty year period of study. Against the background of increased inter-regional mobility from the mid-1950s until the mid-1970s, followed by a decrease in mobility from then until the mid-1980s and then a subsequent recovery (also identified by Donzeau and Shon 2009), Baccaini observes that at the beginning of this period, the Paris region (Ile-de-France) is a net-gainer of migrants. But from the late 1970s until the early 1990s this pattern is reversed, with the city becoming a net-loser of migrants. She notes that this change from net-gain to net-loss is more the consequence of variation in the rate of out-migration than in-migration. The Capital region exhibits relatively stable rates of in-migration across the study period, but large variations in the rate of out-migration. Baccaini notes the variation by age, and explains this pattern by an increase in the out-migration of family and retired migrants to more rural southern and western regions, whilst the draw of the Paris for young migrants in search of higher education and jobs remains constant.

Baccaini’s work is a case study of some variation in volumes of migration over time, but of relatively stable origin-destination patterns. No explanation for the changing attractiveness or repulsiveness of regions resulting in these changing rates is given, but the story of urbanising moves of young migrants searching for employment and education, and counter-urbanising moves of older migrants in search of more desirable living environments remains persistent. The importance of Paris in the inter-regional migration story of France is also made very explicit, with a comment that the most dramatic variation in rates of flow can be observed with this region.

The rates of internal migration reported by Baccaini, even with the recent rise, might be viewed as low to intermediate relative to other European countries according to Pailhe and Solaz (2008). They report that between 1990 and 1999, the average annual inter-department migration rate (not to be confused with the total mobility rate which is somewhat higher) was around 2.8%, with Italy and Greece lower and Scandinavian countries higher. They attribute relatively low inter-regional mobility in France to the high cost of moving and the peculiarities of the French housing market coupled with an inflexible labour market which results in long-term attachments of employees to the companies that employ them.

The case of France can be contrasted with its near neighbour Spain. Whilst the comparison of rates for regions of different sizes should be carried out with extreme care (the larger and fewer in number the regions the fewer inter-regional flows will occur, thus affecting the rates produced), Spain, despite having fewer and in some cases very much larger regions than the French departments, exhibits higher inter-regional flow rates (between 5.63 and 7.91% between 1992 and 2003 - Hierro 2009). Garcia Coll and Stillwell (1999) examine inter-provincial migration in Spain in detail in the 1980s. Following demographic, social and economic restructuring in the 1960s rapid urban industrial growth and related rural to urban migration occurred. Garcia Coll and Stillwell (1999) note a slowing down and in some cases reversal of these urbanising net-migration moves in the 1980s coinciding with an industrial crisis. It is recognised that the migration flows in Spain vary considerably by age at this time, with younger labour market influenced migrants still leaving rural areas, but a considerable number of older retirement and pre-retirement 'return migrants' moving in the opposite direction. Garcia Coll and Stillwell (1999) demonstrate that in many cases, the labour-market peak of migration occurs later than in other countries at around 30-35. Reasons for this may be varied, although in a separate study, Holdsworth (1998) identifies Spain as fitting into the southern European cultural model of individuals leaving home later, normally as the consequence of marriage and family formation, rather than the youth to adulthood transitions more common in northern Europe.

A northern European example which perhaps can be viewed quite differently to other countries is that of Germany. Following reunification in the early 1990s and the destruction of the physical barrier dividing west from east, a significant economically motivated flow from east to west was observed (Kemper, 2004). To a certain extent these east-to-west moves are maintained throughout the 1990s, especially for younger adults, but Kemper (2004) adds to the classic theory of economically influenced motivations by suggesting that regional restructuring and investments coupled with housing and social networks also contributed to continued east-west flows. In the areas which constituted the old West Germany, the process of counterurbanisation seen in France and Spain is also observed, however, even towards the end of the 1990s, this is not apparent in regions located in the former East Germany - findings echoed by Kupiszewski et al. (1998a).

The patterns of general counterurbanisation contrasted with young migrant urbanisation, whilst varying to some extent by region and other by factors such as the threshold ages at which particular volumes and directions of flows occur, can be seen in countries across Western Europe (Champion and Vandermotten, 1997; Fielding, 1982), for example in the Netherlands (Rees et al., 1998c), Norway (Rees et al., 1998b), the Czech Republic (Kupiszewski et al., 1998b) and Portugal (Rees et al., 1998a). Therefore does it follow that similar patterns are also exhibited outside of Europe in countries with similar socio-economic and political profiles?

A number of pieces of work have been carried out on internal migration in Australia. The first point of note is that residential mobility in the antipodes is much higher than it is in Europe. Bell (2002) notes that in Australia and New Zealand in 1981 (and well as the U.S. and Canada but these will be returned to later) the one-year migration transition rate was around 15-18% of the population - this compares to around 9-10% in France and Britain.

In the 1990s, the north-eastern state of Queensland had the fastest growing population in the country, fuelled by net-inward migration from other states (Barker et al., 1998). A continuing flow from south to north along the east coast of Australia is noted elsewhere by Bell (2002) who cites macro and micro economic changes and demographic aging as the driving factors for these moves. Along with this south-to-north move, Bell also notes that rural to major-city-urban migration is a recognised feature of the Australian demographic system with limited employment opportunities in rural areas and small towns, however, recently and in line with patterns already described in Europe, a distinct counterurbanisation stream has been observed from the more heavily populated urban areas to peri-urban areas and, in particular, the coast (Bell, 2002; Drysdale, 1991; Smailes, 1996; Walmsley et al., 1998).

Again, as with Europe age-specific migration propensities are important in the Australian migration story with similar life course influences being cited for the differential propensity to migrate at different ages. As Bell (2002) demonstrates, in Australia the peak ages of migration are in the early 20s, although a very distinct old-age rise in migration can be detected when rates are measured over 5 year transitions in the mid 1990s.

The role of age features heavily in studies of internal migration in North America. Northcott (1985) examines the changing propensity of elderly migrants in Canada between 1961 and 1981, concluding that rates of migration increased over the period, although interestingly many of the moves of these older migrants were into the large cities of Victoria and Vancouver in contrast to the prevalence of counterurbanising moves of elderly migrants in Australia - a difference attributed to the heavy use of services and their urban locations at this time in Canada. Other work on internal migration in Canada, such as that by Coulombe (2006), also highlights some of the age-related influences on migration, however it is the interaction between age and role in the economic system which is drawn out here, with inter-provincial migration for different age groups, such as the 18-24 age group, reacting to economic influences far more than others.

Research on internal migration in Canada comprises only a small proportion of work on internal migration in North America. A considerable amount more work has been carried out on internal migration in the United States (U.S.). Johnson et al. (2005) examine inter-county migration flows in the 1990s and age-specific migration data each decade from the 1950s in an effort to build a profile of the spatial patterns of age specific migration over time. They discovered that whilst considerable variation in age specific net-migration patterns exists between 1950 and 2000, net migration patterns for particular types of county remain remarkably consistent over the years, with the same types of county either gaining or loosing migrants of particular ages in each decade of study. For example, non-metropolitan recreational and agricultural counties were consistent net losers of migrants in their early twenties and conversely net gainers of older migrants (for the most part) over the same period.

Rayer and Brown (2001) examine inter-county migration in the U.S. between 1980 and 1995 and, like Johnson et al. (2005), take the approach of examining county types - although they use a different typology. Ignoring age differentials, nationally they document a fluctuating picture between the 1970s and 1990s - one of counterurbanisation in the 1970s, a certain amount of re-urbanisation in the 1980s before a reversal back to counterurbanisation in the 1990s - a counterurbanising pattern also noticed by Barcus (2004) and Manson and Groop (2000). Rayer and Brown attribute these migration patterns in the 1980s and 1990s principally to a combination of job-related and socio-economic influences acting in parallel with a shift towards de-concentration. Economic restructuring led to a recovery of metropolitan attractiveness followed then by a levelling of the economic and social differences between metropolitan and non-metropolitan areas thus removing the barriers which may have previously stopped the de-concentrating preferences of individuals (Rayer and Brown, 2001).

The movement of population down the U.S. urban hierarchy in the latter half of the twentieth century is something which has been documented in a number of pieces of work, with Henrie and Plane (2008) examining the phenomenon in the context of the west coast and Plane et al. (2005) providing an overview of the whole of the U.S. Like Johnson et al. (2005), Plane et al. (2005) point to the age differentials in migration patterns in the U.S., however a full exposition is given by Plane and Jurjevich (2009) using data from the 2000 U.S. census, with findings which would not be unexpected in the context of other work carried out in the US and in the context of Europe and Australia - namely that younger migrants (those aged between 15 and 29) are more frequently involved in moves which take them up the urban hierarchy towards the largest 'Mega-metro' areas, and older migrants (those aged between 30 and 64) more likely to be moving down the urban hierarchy towards the smallest non-core-based statistical area counties (with populations of less than 10,000).

As a final point of note concerning internal migration in the U.S., as mentioned earlier the rate of migration in North America is high in comparison to Europe. Whilst the rates of migration are high, a decline in the overall rate of inter-state migration is observed between

1950 and 2000. Where in Europe and Australia a reduction in the volume of internal migration might readily be attributed to economic factors, the situation may not be the same in the U.S. As observed by Pingle (2007) much of the decline in volumes of inter-state migration in the U.S. between 1950 and 2000 can actually be attributed to a reduction in military service related migration moves - a situation unlikely to have such an influence in countries with smaller militaries, or indeed countries without enforced national service such as Greece.

So from this brief overview of internal migration in industrialised western democracies, a number of commonalities can be observed. In all of the countries examined, there have been instances of population de-concentration from urban areas - counterurbanisation - at a number of points in the latter half of the twentieth century. Whilst the exact timing of the switch from aggregate net flows into urban areas to net flows out of urban areas, the exact factors influencing internal migration moves and the rates of internal migration vary quite considerably and indeed do not follow constant trajectories, by the turn of the twentieth century in general patterns of counterurbanisation can be observed to a greater or lesser degree across all countries examined. In addition, a number of the pieces of research described have shown how these broad patterns can vary quite considerably by age, with younger migrants tending to migrate up the urban hierarchy towards more urbanised areas, and older migrants tending to move to less densely populated areas. Having provided an overview of internal migration in other western democracies, it remains to examine the history of internal migration in the UK.

3.3 Internal migration in Britain and the UK

Since Ravenstein's seminal papers (Ravenstein, 1885, 1889) where the employment related economic attractions of urban areas were recognised working in partnership with the relative dearth of employment opportunities in rural areas, the former resulting in a migration flow from the latter, research on the factors influencing migration behaviour and flows between particular locations in the UK has been abundant. Any review of the relevant literature in this area could encompass studies examining aspects of internal migration a long way back into the last century. Indeed, Redford (1926) goes as far as to review labour led internal migration in England from as far back as 1750. The process of rapid industrialisation from this point in history saw a population shift from rural to newly urbanising areas, as manufacturing and related industries drew people in with opportunities for employment and higher wages. Redford describes a complex ebb and flow of migrants within England in the 100 years between 1750 and 1850, characterised by a predominance of short distance moves from surrounding rural areas to newly industrialising towns.

The historical national internal migration story is progressed further by Pooley and Turnbull (1996), who, using qualitative data from family historians and focusing on longitudinal data from the 18^{th} to 20^{th} centuries, discovered a high frequency of shorter distance moves within the regional framework they adopted. These shorter distance moves were also characterised, in many cases, by moves between small settlements, and in some cases by counter-urbanising moves.

Whilst historical internal population migrations in the UK are undoubtedly of much interest, and certainly the precursor to all modern population movements with recognisable processes driving these movements, an exhaustive review of all historical UK related internal migration literature would necessarily be a very extensive exercise. As such, and in the context of the first part of this review concentrating on more recent internal migration patterns in other western countries, the focus of this second half of the review will be on studies reporting on internal migration from the latter half of the 20^{th} century onwards.

3.3.1 Aggregate patterns of migration in the UK in the latter part of the 20th century

A number of authors have considered aggregate internal population migration trends in the UK. Stillwell and Boden (1986) used census and NHSCR data to examine aggregate national (British) patterns of migration between the 1961 and 1981 Censuses. They show that, on a national level, the rate and level of migration increased throughout the 1960s before declining during the 1970s from 5.8 million migrants in 1970-71 to 4.7 million in 1980-81. Over the two decades, the majority of migrants were female, although rates were higher for males. Stillwell and Boden acknowledge the limitations of decennial census data, especially where there is a desire to understand inter-censal movements. To address this they use annual movements estimated from the NHSCR between 1975 and 1983, focusing specifically on age-related migration schedules. They conclude that whilst a decline in the level of mobility can be seen in the 1970s when compared to the previous decade, the age and sex characteristics of migrants remained relatively stable.

Rees and Stillwell (1987) extend this national aggregate picture of internal migration through examining internal migration in regional and metropolitan/non-metropolitan contexts using census and NHSCR re-registration data, but from the mid-1960s to the mid-1980s. The regional pattern over this twenty year period is one of the northern periphery (including Northern Ireland and Scotland), the Industrial Heartlands (including the North West, Yorkshire and the Humber and the West Midlands) and Greater London all experiencing net out-migration, with the south (including the East Midlands, East Anglia, the East and the South West) experiencing net in-migration. This north-south shift has also been identified by authors including Champion (1989) and Owen and Green (1992). A regional perspective on migration at this time was also taken by Ogilvy (1982) using the NHSCR data. As might be expected, comparable findings were presented, although Ogilvy placed emphasis on the drop in out-migration from the South East during the 1970s rather than the increase in in-migration as the reason for continued net gains.

Whilst these regional patterns add detail to the national scene depicted by Stillwell and Boden (1986), it is recognised by Rees and Stillwell (1987) that the examination of trends at a regional scale can mask important movements between metropolitan and non-metropolitan areas. During this time period, the dominant trend is one of a decentralisation of population, with movements from the 'core' of metropolitan areas to the 'fringe', and more generally from metropolitan areas to non-metropolitan areas. Indeed, the process of counterurbanisation in Britain at this time has been reported elsewhere in the UK by Kennett (1980), Champion (1989, 1994) and Cross (1990).

The loss of population from metropolitan to non-metropolitan areas recognised throughout the 1970s did not abate in the 1980s (unlike in the U.S.). Stillwell et al. (1992) note the continuing trend of metropolitan out-migration throughout this decade. In volume terms, the last three years of the 1980s saw counterurbanising moves account for 37% of the total internal migration movements (according to NHSCR patient re-registration data) when compared to north-south moves that accounted for only 27% of total internal migration movements. It is further demonstrated that, in the 1980s, the metropolitan areas were only gaining population in the student and immediately post-student quinary age-groups. Whilst an overall trend of counterurbanisation in Britain was identified, when examined closely, the trend was region and age-specific. It is noted that in the north, whilst there was a net out-migration from metropolitan areas, there was not a noticeable corresponding net in-migration to non-metropolitan areas. The trend in the north of Britain was more likely to be the movement from north to south, than the movement from urban to rural. Certainly whilst this north-south movement pattern is the important trend for most of the 1980s, one significant point of interest occurs at the end of the 1980s. Whereas the net flow had been into the south from the north since the mid 1970s, Stillwell et al. (1992) show that an increase in the movement of migrants northwards had in fact tipped the balance slightly in the favour of a net gain to the north in 1989.

Rees et al. (1996) use data from the 1991 Census to show a continuation of the trend of depopulation from urban areas that had been revealed in previous decades. This applied both to the largest metropolitan areas as well as the smaller cities. It was further shown that resource regions associated with much of Britain's dwindling primary industry (such as mining and fishing) were losing population, whilst new 'resource frontier' regions (broadly associated with offshore industry) were gaining migrants. Other trends identified were associated with the predominance of the 1-15 and 30-44 year old age groups in the overall patterns of migrant redistribution, and with lower mobility but clearer patterns of redistribution at retirement and post-retirement ages. There was a noticeable urbanisation of the 16-29 age group, associated with student movements to higher educational institutions which were often found in large urban areas - a trend familiar from previous decades. Rees et al. (1996) highlight that there was a 'downward and outward' redistribution of population from cities, meaning that population was redistributed both down the urban hierarchy from larger to smaller urban centres, and out from urban centres into rural fringes. The latter did not necessarily indicate the return of a desire by people to live rural lifestyles, but rather was the result an expansion of pre-existing urban systems into areas otherwise identified as rural.

Looking into these patterns in more detail it was noted that there appeared to be a propensity within the migrant population to move from higher to lower density areas on the whole; this was coupled with a shift from areas suffering from above average unemployment to areas with below average unemployment. Indeed, all of the findings by Rees et al. echo and bolster the regional level work of Stillwell et al. (1995b) for the same period. Subsequent work by Kalogirou (2005) on England and Wales corroborates these findings.

A further overview of internal (and international) migration in this period is provided by Champion et al. (1998). As would be expected, the aggregate patterns recounted are no different from those already covered in this review. Where this latter overview differs is that it brings into focus the issue of scale. Work already mentioned has examined migration at different spatial scales; however, here the effect of scale on results is discussed explicitly. Due to the preponderance of short distance moves over longer distance moves, inter-area flows become more important when the scale of analysis is smaller. In addition, regional level analysis will, for example, emphasise the importance of international flows as the major contributor to the migration component of population change (certainly when looking at net redistribution if not gross flows), whereas analysis at more disaggregate scales will promote the importance of internal migration.

One key conclusion made by Champion et al. (1998) is that at the regional level, the traditional drivers of flows between areas such as the availability of employment have, in the 1990s, been replaced by the determinants more commonly recognised as influencing shorter distance moves, such as housing or environmental factors. This assertion is backed up by the evidence that the largest inter-regional moves are between adjacent regions, and perhaps more importantly, adjacent counties on either side of regional boundaries. Of course, migration influencing factors such as employment opportunities, housing supply/demand and population density will differentially affect migrants at different stages of the life course, and so the conclusions of Rees et al. (1996) that place emphasis on the role of employment in migration, as well as environmental factors such as population density, should not be discarded in the light of these new findings. Indeed, work elsewhere by Fielding (1992), which places central importance on employment as an explanatory factor for internal migration, and Cameron et al. (2005), which focuses attention on housing, each offer persuading evidence of the influence of employment and housing respectively on regional-level internal migration.

Thus far this review has drawn on data available before the 2001 Census. Despite internal migration data from the 2001 Census being available since 2003, there has been relatively little work carried out on internal migration patterns in the new millennium. Standing out from this relative dearth is a study published as part of the ONS (2005a) ‘Focus On’ publication in which Champion (2005) provides a wide ranging overview of internal migration in the UK, drawing principally, although not exclusively, from the 2001 Census.

Champion indicates that, in 2000-01, there were around 6.7 million internal migrants nationally, but comments that little difference is evident in the migration propensities of males and females; at least at an aggregate level. Where age (the other key demographic indicator often mentioned with sex) is concerned, however, the situation is somewhat different. The trend in 2001 (as in previous years) is that young adults have the greatest propensity to migrate. This coincides with the now familiar movement of many in their late teens to higher and further education institutions, and then away from these locations as students move on to their first jobs after completing higher education courses. The tendency to migrate reduces with age after young adulthood until around the age of 75. As in previous years, age-specific migration follows a familiar pattern with a reduction in the propensity to migrate from the mid-twenties to the mid-thirties, corresponding with family raising and the desire for settled child rearing. This decline in migration propensity continues until around pensionable age. From here, there is a noticeable increase in the rate of migration and this can be attributed to the 'defensive' moves of older individuals as dependency and insecurity increases with age. Migration at this age can readily be attributed to moves associated with a greater need for care or to be within proximity of family.

Champion outlines the broad national migration trends for other demographic variables featured in the 2001 Census. The migratory patterns of individuals classified by marital status, family type, health, housing tenure, economic activity, industry of employment, occupational level, qualifications and ethnicity are all summarised briefly in relation to the whole country, with the overall (highly generalised) picture being that single or childless adults, those who did not own their own homes or individuals who were more qualified or in a higher socio-economic group being generally more likely to migrate. Home owners, parents or lower socio-economic groups were generally less likely to migrate. The white ethnic group also had marginally lower migration tendencies than non-white groups.

Champion examines some of the sub-national migration patterns that are displayed by the results of the 2001 Census. At the district and ward level, the salient point is that districts and wards with the highest proportions of people living at a different address one year ago tend to be those with highest student populations. Unlike the 1991 Census, when students were recorded at their parental domiciles, the 2001 Census recorded students at their term-time addresses and therefore contain counts of student 'migration'. At the other end of the scale, those districts and wards with the lowest proportions of their populations consisting of people who lived at a different address one year previously were frequently located in Northern Ireland. Mapping reveals the relative importance of coastal and rural retirement areas where higher migratory rates are present.

Another key feature of internal migration from the 2001 Census picked out by Champion is the pervasiveness of net urban-to-rural migration across the whole of the UK - not just where London is concerned. Using a classification of districts adapted from work carried out in the early 1980s, he demonstrates that metropolitan areas are continuing to lose migrants to rural areas. The validity of using a classification for areas devised in the 1980s should be questioned to some extent, especially when considering the amount of change that has taken place in the socio-demographic and physical characteristics of many of these areas since then. However, one might suspect that, under scrutiny, these broad patterns are likely to be more-or-less accurate at this aggregate level. Indeed the patterns of metropolitan losses and non-metropolitan gains at the district level in 2000-01 are confirmed by Stillwell and Duke-Williams (2007).

Finally Champion takes a somewhat more detailed look at the geographical variations in the interaction flows of four specific population characteristics: age, student status, ethnic group and higher managerial and professional occupations are examined at a regional scale. He concludes that there is a rural/urban association with age, in that younger age groups may be influenced by the ‘bright lights’ of urban areas with the opposite being true of older age groups. Unsurprisingly, students are identified as being attracted to those districts containing educational establishments, with the inner and outer boroughs of London experiencing some of the most noticeable in-migration and out-migration flows respectively. It is also London which shows the most significant migration patterns in relation to non-white migrants, with the largest absolute increases and decreases of this group occurring here. It is also shown that the south, specifically districts in and around London, recorded significant net gains of people in the highest socio-economic groups, with areas to the south-west and east of the country recording the lowest net gains.

So far this section of the review has been concerned with examining work from the latter half of the twentieth century and early part of this century - work which has looked at aggregate internal migration patterns within the UK. Examining this extensive body of work, a clear story emerges: the story is one of increasing rates of national internal migration until the early 1970s, before a decline until the 1980s, followed by an increase again in migrants in the 1990s leading up to 2001. Over this period, the overall fluctuations in the numbers of migrants within the UK system have only been part of the story, as the forty years or so have seen major movements from the north to the south; from metropolitan areas to non-metropolitan areas; all punctuated by reversals in these general trends at certain times (in the case of the north-south movement) and by specific sections of the population (in the case of younger migrants being positively drawn to metropolitan areas). The latest work by Champion, as well as offering insight into these aggregate patterns, also looks at migration in the context of different defining variables and at different scales. Indeed, Champion is by no means the first to do this, and so moving away from work on aggregate patterns of migration, this review will now focus on research that has been carried out on specific features of internal migration in the UK over this period.

3.3.2 Features of internal migration in the UK

Work by Fielding (1992) seeks to make the link between migration and social mobility in the UK using data from 1971 to 1981 from the (then OPCS, now ONS) Longitudinal Study and the NHSCR. Fielding recounts the phenomenon of young, educated people moving to the South East for employment before migrating back away from the South East as their socio-economic status increases as a result of this initial employment. Fielding concludes that the South East in particular (when compared to other regions) acts as a relatively rapid socio-economic promoter of residents - an escalator region. As these residents move into the region and then up the socio-economic hierarchy there is evidence of an inclination by a significant number to move away again from the South East - a region stereotypically classified as an upwardly mobile service class region, but argued by Fielding (perhaps to the offence of some native South-Easterners) to be *"more banal than the regions of northern and western Britain."* (Fielding, 1992, p15).

The role of the South East is again brought into focus by Cameron et al. (2005) who examine the influence of the housing market on inter-regional migration in England and Wales. Whilst it is recognised by the authors that the drivers behind inter-regional migration can be numerous and complex, an assertion of the greater importance of housing (in terms of the influence of cost and availability) amongst these reasons is made. Furthermore the locational importance of London and the South East in UK internal migration is reiterated. Indeed, whilst it might be expected that higher house prices in economically prosperous areas might mitigate against the further in-migration encouraged by this prosperity, Cameron et al. demonstrate that the expected house price appreciation in the South East helps encourage migrants when they might otherwise be put off by higher average house prices. Both Fielding (1992) and Cameron et al. (2005) agree that the relative economic prosperity of the South East (even when examining data for different time periods as is the case with these two pieces) acts as a significant pull factor to migrants from other parts of the UK. Whilst the drivers of migration are the focus for these papers, the nature of migration is another key area of research in UK internal migration studies.

Work by Norman et al. (2005) examines the relationship between health, deprivation and migration in England and Wales using a closed population sample from the ONS longitudinal study between 1971 and 1991. It is demonstrated that over this twenty year period, Standardised Illness Ratios (SIRs) and Standardised Mortality Ratios (SMRs) increased between the least and most deprived areas, with an accumulation of healthy and surviving people in the least deprived areas and that in the main, this was down to migration rather than changes in the areas themselves. This work showed that young migrants moving from more to less deprived areas are generally healthier than non-migrants, or older migrants, moving from less to more deprived areas. Interestingly, migrants within deprived areas are less healthy than non-migrants and migrants who move from less to more deprived areas. Only a small proportion of migrants in this study suffer from a limiting long term illness and move from more to less deprived areas. Most un-healthy migrants tend to move from less to more deprived areas. These movements have a dual effect on both the numerators and denominators at the origins and destinations, reducing ill health rates at the destination and increasing them at the origin more noticeably. Some limitations are recognised; including the exclusion of non-surviving individuals who would have been more likely to have lived in deprived areas, this leading to a healthier group of remaining people. Despite such limitations the overall message and summary from Norman et al. is that migrants are generally healthier than their non-migrant counterparts. However, as a caveat to this general finding, when the origins and destinations of these migrants are taken into consideration and identified by their level of deprivation, a pattern emerges of a health selective migration with healthy migrants appearing to choose improved destinations, and un-healthy migrants tending to move to more deprived areas, perhaps as a consequence of their ill health.

The nature of ethnic migration is the focus for Finney and Simpson (2007). A comprehensive account is given of the migration characteristics of ethnic groups in the UK, using data from the 2001 census, with the aim of the work to ascertain if the characteristics of migrants and patterns of migration for different ethnic groups are similar to those of the population in general. It is recognised early on that many of the apparent differences in ethnic migration could be ascribed to other characteristics of these groups such as their age structures, socio-economic make-up, tenure characteristics, qualification levels and geographic locations. Indeed regression analysis which controls for these factors confirms this initial theory, leading to the conclusion that, for the most part, differences in the migration patterns of ethnic groups can be attributed to these other explanatory variables. For example the relatively low migration rates of the White population when compared to other ethnic groups such as Bangladeshis and Pakistanis, can be explained by the comparatively old White population and the relatively old age structures of these other ethnic groups. Other differing socio-demographic characteristics between the groups are such that apparently significant differences between the migration propensities of ethnic groups become insignificant when these socio-demographic characteristics are taken account of. It is discovered, however, that even when other possible explanatory factors are taken into consideration, non-White groups tend to migrate less far than White groups. On average, Pakistani, Bangladeshi, African and other Black groups generally moved less than 30km. For other groups the average distance was further. One final conclusion reached from the analysis carried out by Finney and Simpson (2007) is that contrary to conventional theories of ethnic 'self-segregation' and 'white-flight,' all non-white ethnic groups have been found to be migrating away from areas of ethnic concentration to areas of highest white concentration. An interesting pattern, perhaps explained in some way by a desire demonstrated historically by many groups to move away from densely populated urban areas (in which large numbers of non-white groups can be found) to less densely populated, predominantly white areas.

The focus of the study by Bailey and Livingston (2007) moves away from the migrant and onto the nature of the areas between which migrants flow. More specifically they concentrate particularly on flows to and from more deprived neighbourhood areas (defined as Super Output Areas in the UK and Data Zones in Scotland) the effect that movements (or lack of) have on these more deprived origins or destinations, using data from the 2001 Census. One of the first findings made by Bailey and Livingston is that deprived areas are not as disconnected or isolated from, what they term, 'non-deprived' areas as might be expected. Connectivity in this case is defined by the rate of flow between areas identified as either deprived or non-deprived. It is demonstrated that around half of the migrants to more deprived areas come from non-deprived areas, and a similar proportion of migrants move the other way. The connectivity rates are lower for areas with much higher deprivation, although at the same time, for the most deprived decile of neighbourhoods within each city region in the UK, half of all migrants either come from or go to non-deprived neighbourhoods within that same city region.

What is clear, however, is the importance of local context on flows of migrants to and from deprived neighbourhoods. The most deprived areas are seen to have lowest stability (i.e. fewest people remaining there from one year to the next) but also the lowest connection rates (signifying moves from less to more deprived areas and vice-versa). More than the level of deprivation, the demographic mix of a neighbourhood is seen as having the most influence on the stability and turnover of the population in that area. Areas with higher proportions of young adults or young children are likely to experience higher rates of movement. There is an implicit assumption that a stable community is a more sustainable one - something which is likely to have a positive effect on the area in which that community resides. With this being the case, it is suggested that policies designed to improve areas should focus on improving the demographic mix perhaps more than the income or tenure mix. These findings suggest that deprived areas act as transitional areas, especially for young adults. The associated lower cost of housing and living with deprived areas makes them attractive destinations for young migrants either just moving out of the parental home or into their first home after a period in higher education. With a national pattern of increasing house prices, it is projected that this trend may continue. With this though, the loss of qualified people from deprived areas has little affect on the social mix, and it is suggested that local interventions based around improving the socio-demographic mix of deprived areas, could have a noticeable impact on the sustainability of these areas.

3.4 Conclusions

This short review began with the intention of mapping out the recent broad internal migration landscape in the western world and in the UK in order to set the substantive context for the work which will follow in subsequent chapters. Initially, research carried out on internal migration in western Europe was reviewed, revealing remarkable similarities between countries across the continent. In France, Spain and Germany can be found three large, neighbouring countries but with three contrasting social, economic, cultural and historical backgrounds. In spite of these differences and ignoring some of the idiosyncratic migration features such as the east to west flow in Germany and the later peak migration age Spain associated with an older age at which young migrants tend to leave home, very similar internal migration stories can be told in each of these countries. It is a situation of an increasing trend towards counterurbanisation counterbalanced to some extent by urbanising moves of young migrants. Casting the net further afield, very similar stories are told in Australia and North America - the contrasting feature of internal migration in these areas being that rates of movement are much higher than they are in these western European countries.

The second half of this review concentrated on the recent internal migration history of the UK. In many ways the UK exemplifies the experience of all of the other countries included in this review. It could be argued that it has a little more in common with the U.S. and Australia in terms of the overall rate of internal migration when compared with countries in Europe, but although sharing the broad movement towards counterurbanisation the UK has some particular features of its internal migration situation which are worth highlighting. For example, in the UK there has been a historical preponderance of moves from the north of the country to the south-east - particularly London and its immediate hinterland, moves that have been driven by the economic prosperity of this region and by its position as a socio-economic escalator. Whilst economic variation and the availability of particular types of job within the UK is one influencing factor on migration, another is the cost of housing with there being evidence of house prices having an even greater influence on the volume and direction of movements more and more towards the end of the last century. Interacting with these dominant influences, health, ethnicity and socio-economic status have also been shown also to be contributing factors to the internal migration landscape of the UK.

From this short review of the internal migration literature it might be noted that there are some gaps in our understanding of the situation in the UK at the turn of the 21^{st} century. Internal migration is clearly a complex process with flows of people occurring between any number of the zones within the UK spatial system, so consequently much of the analysis has tended towards more simplistic dichotomous spatial divisions - north to south, metropolitan to non-metropolitan for example. In some of the research carried out in the U.S., a typology of urban area size was adopted to enhance the level of spatial detail whilst still retaining an amount of simplicity in the explanation. Work by Rees et al. (1996) on internal migration in Britain from the 1991 census continues down this road but takes the analysis further through the adoption of area classification typologies - both those based on functional regions (the concept of which will be returned to in Chapter 5) and on area types; the advantages of both being they offer a little more explanatory power than more simplistic population size typologies. Despite the clear explanatory advantages of such an approach when exploring the spatial patterns of migration, much recent research on internal migration in the UK has been concerned with explanatory variables or overviews rather than anaylsis of flows and area types - perhaps with the exception of work by Raymer et al. (2007) and Raymer and Giulietti (2009). In addition, whilst a number of migrant characteristics such as ethnicity, socio-economic status and health have been explored in recent analyses, perhaps the characteristic which acts more than any other to influence the volume and direction of migration - age - has yet to be explored in detail using the recent internal migration data. So it is from this point that the next chapter will embark.

Chapter 4

Internal migration in Great Britain - a district level analysis

4.1 Introduction

The two review chapters preceding this have set the scene for the rest of this thesis both theoretically and methodologically through providing an overview of the internal migration data landscape in the UK, defining a spatial system and associated analysis techniques; and substantively though reviewing selected internal migration analyses both within the United Kingdom and further afield. Evident from the last chapter is the relative dearth of research on Internal Migration in Britain following the 2001 Census. Given that the principal aim of this thesis is to understand migration patterns at the beginning of the 21^{st} century, it would seem logical where there is both a lack of existing research and the availability of a rich source in the 2001 Census, to explore the internal migration patterns revealed by it.

As mentioned at the end of the last chapter, to date, work on the influence of age on recent internal migration patterns has not been undertaken, so for this reason and because it has been demonstrated on many occasions that age is one of the largest influencing factors on the propensity to migrate, this chapter will look to address gap in our knowledge. Section 4.2 will outline the data sources to be used in the analysis and will highlight some of the issues which are presented as a result of the data selection. Section 4.3 will provide an aggregate overview of migration in Britain, before Section 4.4 will delve a little deeper through examining national age-specific migration rates for different types of area. The latter half of this chapter from Section 4.5 onwards will look to advance the understanding of internal migration through the introduction of some alternative measures which address the idea of population stability. By the end of this chapter a much clearer picture of internal migration at the beginning of the 21^{st} century will be achieved, but as will be seen, during the course of this analysis some important new issues will be raised.

4.2 Data Sources and Issues

The data used in this analysis have been taken principally from the 2001 Census SMS at level 1 (district or LAD level) - the coarsest level available. This level has been chosen for three main practical reasons: firstly the interaction data are more accurate at this level as SCAM described in Section 2.2 has less of a damaging effect on the data where a larger primary unit of analysis reduces the chances of small values appearing in the cells of the migration matrices. Secondly, as discussed at the end of Chapter 3, a useful way of reducing complexity in exploring migration data is through the use of an area classification - more of these classifications have been produced at the LAD level. Thirdly as shown in Chapter 2, outside of the census, data are not produced at a level below LAD level. As will be seen in subsequent chapters of this thesis, other non-census datasets will be vital in understanding the internal migration story in this country, so it would be illogical to conduct an analysis at a level where future comparisons cannot be made.

The spatial system The analysis of districts will be carried out for Great Britain (England, Wales and Scotland), rather than the whole UK (including Northern Ireland as well). Northern Ireland has been omitted from the analysis for a variety of practical reasons. Firstly, there are no interaction data available for Northern Ireland at level 1 for district areas. Northern Ireland data at level 1 are only available for Parliamentary Constituencies - a geography of comparable size but with incompatible boundaries leading to a set of harmonisation problems.

Why not aggregate ward level data for Northern Ireland up to the level of district? In theory, this aggregation should be possible and doing so would enable an analysis for the whole of the UK to be carried out. In practice, however, this is not a straightforward task. The main problem here is that wards in Northern Ireland do not aggregate perfectly into districts. There are 37 instances of CAS wards not aggregating into districts perfectly in Northern Ireland. Look-up tables provided through the Geo-Convert facility (http://geoconvert.mimas.ac.uk/) give the precise proportions of each ward that feature in each related local authority district. For easy data aggregation, it would be desirable that each ward would fit 100% into a related district. Table 4.1 exemplifies the relative proportions of each ward that feature in each district, where a ward does not fit wholly into one district. Whilst in all cases, one district tends to feature the majority of each ward (well over 90% in all cases), it is impossible to know precisely how to weight the data to assign the correct proportions of the ward data to each district.

There are two issues here. The first issue relates to the geographic location of addresses within the ward. For example, it could be the case that there are no addresses featured in the small proportion of the ward associated with one district. Where this is the case, there would be no need to reallocate a proportion of the data to this district. On the other hand, this very small portion of the ward could contain a considerable proportion of the addresses, requiring the allocation of a considerable proportion of that ward's data to the district. The second issue is that even if the proportion of addresses allocated from a ward to districts is

Table 4.1: Example of the proportions of each ward assigned to each associated district in Northern Ireland

Ward Code	Proportion of Ward in District	District Code
95AA01	0.9487	95AA
	0.0513	95SS
95AA12	0.9904	95AA
	0.0096	95WW
95AA14	0.9894	95AA
	0.0106	95DD
95BB02	0.9906	95BB
	0.0094	95II
95BB10	0.9987	95BB
	0.0013	95XX
95BB16	0.9935	95BB
	0.0065	95II

Source: http://geoconvert.mimas.ac.uk/

known (something that is feasible, if not practically possible for large areas through the address counts available in the all fields postcode look-up directory tables), it is almost impossible to allocate appropriately the correct data to the correct addresses - especially for a large amount of areas where many calculations would be needed. Furthermore, address counts will include communal establishments (such as student halls of residence, hotels, hospitals and prisons) as well as households which more commonly accommodate smaller numbers of residents, making data allocation even more difficult.

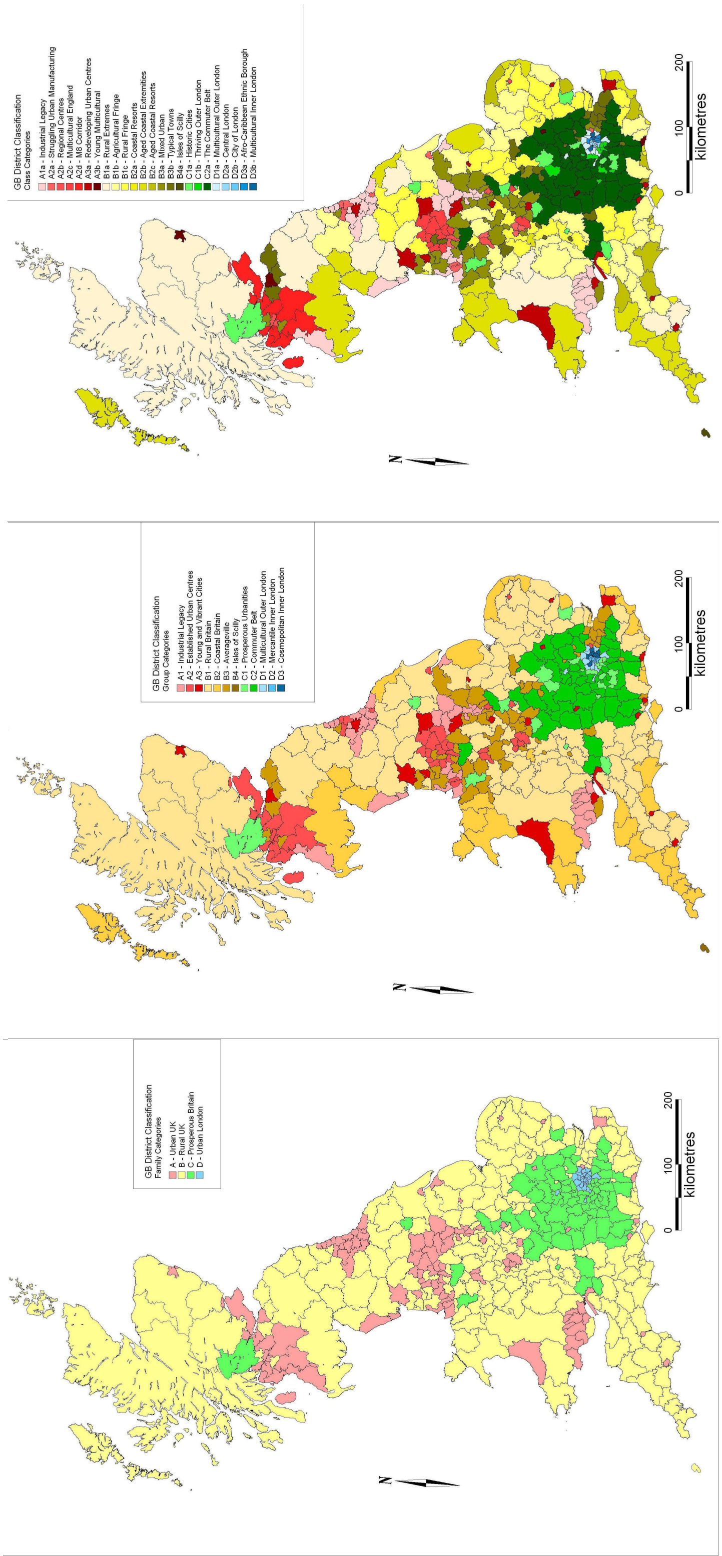
Even if an appropriate way of allocating the correct flow from wards to districts was devised, another hurdle would need to be cleared if Northern Ireland data were to be included. UK analysis would require an understanding of the total flows to and from all of the districts in Northern Ireland. As these districts are not available via WICID, all of the data from Northern Ireland would need to be downloaded as wards for Northern Ireland, for both origins and destinations; this would also need to be accompanied by the information for every district in the rest of the UK. This means that for each of the 990 origins (408 districts in England, Wales and Scotland and 582 wards in Northern Ireland) there would also be 990 destinations. This would be a pairwise list of some 980,100 rows of data if one were downloading a list of every variable by every origin/destination pair. Alternatively it would be a 990 by 990 matrix for each variable selected. Whatever the format used for downloading, the flow data for every Northern Ireland ward would need to be weighted appropriately and then assigned to a new district. This would be a considerable task!

The analysis reported in this chapter makes use of a national classification of districts that has been developed by Vickers et al. (2003) using the 2001 KS which assigns each district in the UK to a different Family, Group or Class based on a range of socio-economic and demographic characteristics (Figure 4.1). There do exist other general purpose district level classifications such as the three tier system developed by the ONS (ONS, 2004) and the urban-rural classification produced by the Department for Food and Rural Affairs (DEFRA) (DEFRA, 2009). The Vickers et al. classification has been selected for this analysis because of its comprehensive and transparent methodology and because it makes a more logical distinction between rural and urban areas than the ONS classification, whilst separating London and prospering commuter areas from other districts in a more effective way than DEFRA do in their purely rural/urban classification.

There are two main reasons why it is beneficial to use an area classification for the study of population flows in Britain. Firstly, within Britain there are 409 LADs. 409 potential origins and destinations means that there are 166,464 possible interaction flows which could take place. By classifying these districts into types a large and potentially difficult to interpret matrix is reduced into something far more manageable. Secondly, the classification is an effective way of characterising areas and, as such, provides a useful backdrop against which to present migration flow data. By looking at population flows in the context of the underlying socio-demographic characteristics of areas between and within which the flows occur, it is possible to understand more about how these underlying areal characteristics may be associated with particular flows of migrants. Since the Vickers et al. classification does not incorporate any migration variables it provides a framework for migration analysis which is independent of the influence of a migration dimension.

Past studies by Champion (1989), Champion et al. (2007) and Fielding (1992) have summarised trends in migration between 'urban' and 'rural' or 'London' and the 'rest of the UK'. Analysing migration using a classification framework allows for the identification of migration trends and patterns in relation to these familiar binaries, but in addition, the sub-classifications (Families, Groups, Classes) allow flows to be further disaggregated, for example, into migration into or out of types of rural and urban area or very specific parts of London. Observations of a general pattern of counterurbanisation which could mask instances of urbanisation in some key urban areas need not do so where the classification will disaggregate flows by more categories. Use has been made of general purpose classifications to analyse migration patterns summaries in the past Champion (2005); Rees et al. (1996). The Vickers et al. classification offers further potential for a more detailed insight into more nuanced flows.

However, using an area classification in the analysis of migration data does present some practical problems: care needs to be taken when using classifications in conjunction with the calculation of rates based on inflows, outflows inter-zone flows, intra-zone flows and net flows. Particular care and attention should be paid to the way that these flows are calculated once analysis moves to a geographical level above that of the basic building block (the district in this analysis). Attention is needed to avoid double counting or undercounting of flows between or within areas identified by the classification. Illustration of this issue using hypothetical flow data and a simplified tiered classification is given in Figure 4.2. A move within an area at a more spatially aggregate level can become a move between one area and another as one moves down the classification hierarchy. So, for example, the 30,916 migrants within Group A1 will



(a) Family level

(b) Group level

(c) Class level

Figure 4.1: Vickers et al. (2003) classification of local authority districts

break down into i) 30,748 moves within Classes A1a and A1b and 168 moves between these Classes or ii) 30,690 moves within districts 1 to 4 and 226 moves between them.

				Family A													
				Group A1				Group A2									
				Class A1a		*Class A1b*		*Class A2a*		*Class A2b*							
				District 1	District 2	District 3	District 4	District 5	District 6	District 7	District 8	District Outflow	Class Outflow	Group Outflow	District Inter	Class Inter	Group Inter
Family A	Group A1	*Class A1a*	District 1	154	0	9	3	9	15	44	6	86	253	3465	193	645	8022
			District 2	0	6844	24	29	52	18	29	15	167			452		
		Class A1b	District 3	19	46	15168	33	1161	88	1174	93	2614	3380		6154	7713	
			District 4	0	38	25	8524	21	657	39	44	824			1675		
	Group A2	*Class A2a*	District 5	11	80	1383	66	13198	75	837	150	2602	4248	4557	5145	8727	8022
			District 6	12	44	62	560	53	13123	84	959	1774			3838		
		Class A2b	District 7	65	45	1909	52	1174	77	10966	90	3412	4797		5750	8271	
			District 8	0	32	128	108	73	1134	131	17798	1606			2963		
District Inflow				107	285	3540	851	2543	2064	2338	1357						
Class Inflow				392		4333		4479		3474							
Group Inflow				4557				3465									
District Intra				154	6844	15168	8524	13198	13123	10966	17798				**98860**		
Class Intra				6998		23750		26449		28985							
Group Intra				30916				59922									

Figure 4.2: District level flow matrix including an example hierarchical geodemographic classification

In this chapter the focus will be on the key demographic characteristic of age; for which the migration data comes by quinary age groups up to 89 and 90+ from SMS level 1, Table MG101. Corresponding PAR have been obtained from Standard Table ST001. As noted in Chapter 2, it is common practice when studying patterns of migration to calculate crude intensities of movement, as these rates give a measure of migration that is independent of the population size in any given area. Here the net migration rate calculation shown in Equation (2.7) will be used.

4.3 Aggregate patterns of internal migration

In this section, the spatial patterns of aggregate internal migration for the year preceding the 2001 Census are examined. Figure 4.3 reveals that patterns of net gain and loss for migrants of all ages at the district scale tend to be associated with areas generally recognisable as rural and urban respectively. Most districts of Greater London and those bordering (including those stretching out along the M4) are experiencing net out-migration. Other metropolitan areas, including Birmingham, Liverpool, Manchester and their surrounding districts, the North East and Glasgow are also experiencing net out-migration. In contrast, more rural areas covering large parts of East Anglia, the South West, Wales, the Midlands, the North and Scotland are all experiencing net in-migration.

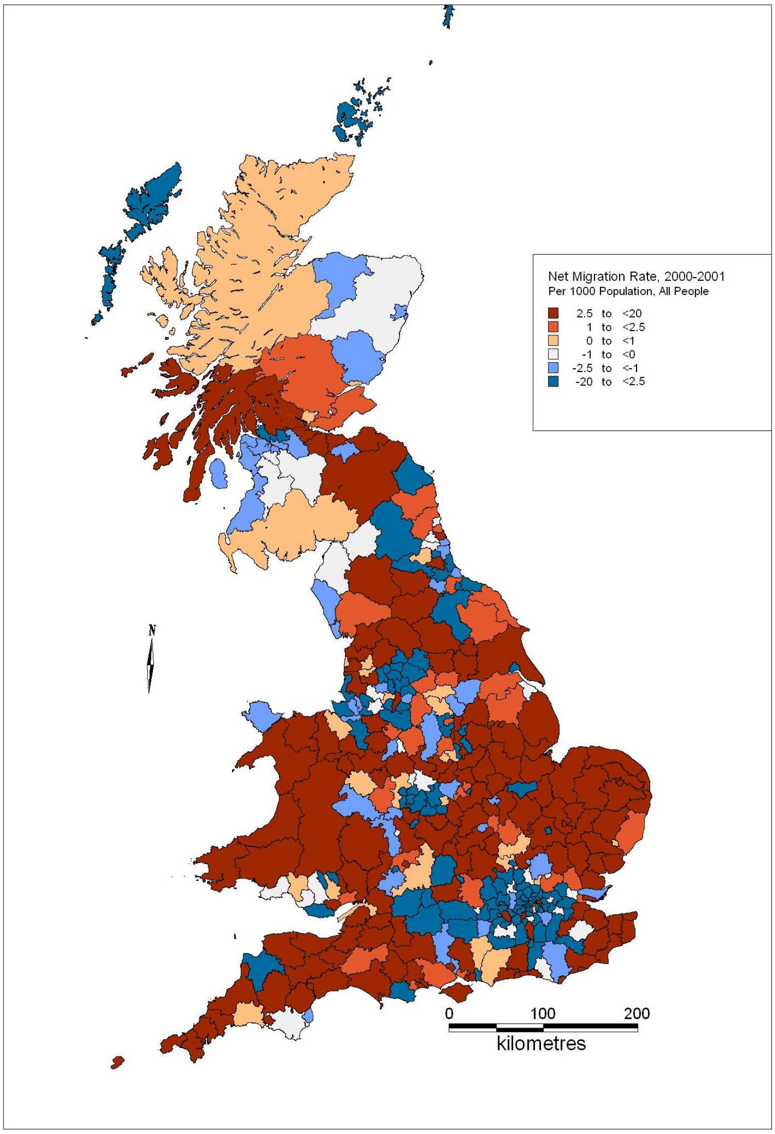


Figure 4.3: LAD net migration rates (per 1,000 population) - all ages, 2000-01

Table 4.2 provides a summary of the net balances displayed on the map using the Families, Groups and Classes from the original Vickers et al. classification (the use of the original classification nomenclature explains the appearance of 'UK' in this analysis, when in fact only Britain is being studied). There is consistency between the summed net balances at each level indicating that the balances at Group and Class level refer to flows between districts in different Families. Consequently, the balances in each column of the net migration section of the table therefore sum to zero at each level of the hierarchy. The patterns of net migration revealed on the map (Figure 4.3) are summarised at the most aggregate Family level, with Urban UK, Urban London and Prosperous Britain exhibiting net out-migration and out-migration rates, and Rural UK exhibiting net in-migration. Of the four Families, Rural UK gains the largest number of net migrants; however, with a larger PAR, its net in-migration rate of 2.7 people per 1,000 population is considerably lower than the net out-migration rate of 8.5 people per 1,000 population from London.

Table 4.2: Net migrants and net migration rates by district classification - all people, 2000-01

District Classification (**Family**, Group, *Class*)	**Net Migrants**	**Net Mig Rate (per 1000 population)**
A: Urban UK	**-4,842**	**-0.23**
A1: Industrial Legacy	-5,263	-0.95
A1a: Industrial Legacy	*-5,263*	*-0.95*
A2: Established Urban Centres	-16,829	-1.69
A2a: Struggling Urban Manufacturing	*-8,292*	*-2.78*
A2b: Regional Centres	*3,195*	*1.96*
A2c: Multicultural England	*-11,264*	*-3.13*
A2d: M8 Corridor	*-468*	*-0.26*
A3: Young and Vibrant Cities	17,250	3.38
A3a: Redeveloping Urban Centres	*15,395*	*3.89*
A3b: Young Multicultural	*1,855*	*1.61*
B: Rural UK	**57,947**	**2.72**
B1: Rural Britain	30,653	3.55
B1a: Rural Extremes	*1,067*	*0.67*
B1b: Agricultural Fringe	*14,762*	*4.32*
B1c: Rural Fringe	*14,824*	*4.09*
B2: Coastal Britain	29,296	6.59
B2a: Coastal Resorts	*7,231*	*7.27*
B2b: Aged Coastal Extremities	*13,987*	*5.2*
B2c: Aged Coastal Resorts	*8,078*	*10.62*
B3: Averageville	-2,040	-0.25
B3a: Mixed Urban	*-2,311*	*-0.45*
B3b: Typical Towns	*271*	*0.09*
B4: Isles of Scilly	38	17.79
B4a: Isles of Scilly	*38*	*17.79*
C: Prosperous Britain	**-4,983**	**-0.52**
C1: Prosperous Urbanites	1,844	0.58
C1a: Historic Cities	*5,606*	*3.48*
C1b: Thriving Outer London	*-3,762*	*-2.41*
C2: Commuter Belt	-6,827	-1.07
C2a: The Commuter Belt	*-6,827*	*-1.07*
D: Urban London	**-48,122**	**-8.53**
D1: Multicultural Outer London	-20,947	-8.01
D1a: Multicultural Outer London	*-20,947*	*-8.01*
D2: Mercantile Inner London	-11,705	-10.21
D2a: Central London	*-11,725*	*-10.29*
D2b: City of London	*20*	*2.79*
D3: Cosmopolitan Inner London	-15,470	-8.22
D3a: Afro-Caribbean Ethnic Borough	*-7,164*	*-6.07*
D3b: Multicultural Inner London	*-8,306*	*-11.81*

Close inspection of the Groups and Classes within each Family reveals that whilst Urban UK as a whole is losing migrants, some areas within Urban UK are gaining significant numbers of migrants and exhibiting positive net migration rates. For example, the Young and Vibrant

Cities Group, (comprising districts in the Redeveloping Urban Centres and Young Multicultural Classes), is experiencing net in-migration of over 17,000 people, and has a net in-migration rate of 3.4 people per 1,000 population. Other examples of net flows masked at the broad Family level include the net out-migration from districts in the Averageville Group within Rural UK which experiences a net gain overall. There are also net gains experienced by districts classified as Historic Cities within the Prosperous Britain Family - a Family which is experiencing overall net out-migration. A similar example can be found within the Urban London Family where the City of London is the only Class within the family to be experiencing net in-migration (albeit very small) - all other areas are experiencing net out-migration.

In much the same way that the areal aggregations discussed above can mask migration patterns, studying all-age migrants can similarly hide important variations. Whilst there are an number of individual attributes that account for migrant behaviour, variation by age captures many of the most important changes of location made during the life course.

4.4 Age-specific patterns of migration

One of the more major attributes affecting an individual's propensity to migrate is their age. A large volume of work, including studies in the 1980s by Bates and Bracken (1982); Raymer et al. (2007, 2006); Rogers and Castro (1981); Rogers et al. (2002), has identified the influence of age on migration behaviour. The seminal work of Rogers and Castro (1981) was important in identifying the similarities in migration rate age 'schedules' across a range of countries and cities. From these common observations, Rogers and Castro were able to construct a model migration schedule consisting of a series of key age-related components. Consider the example shown in Figure 4.4.

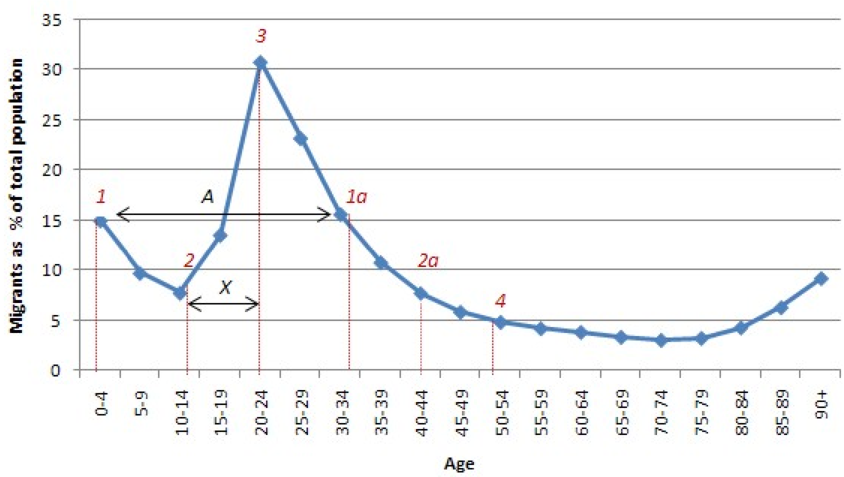
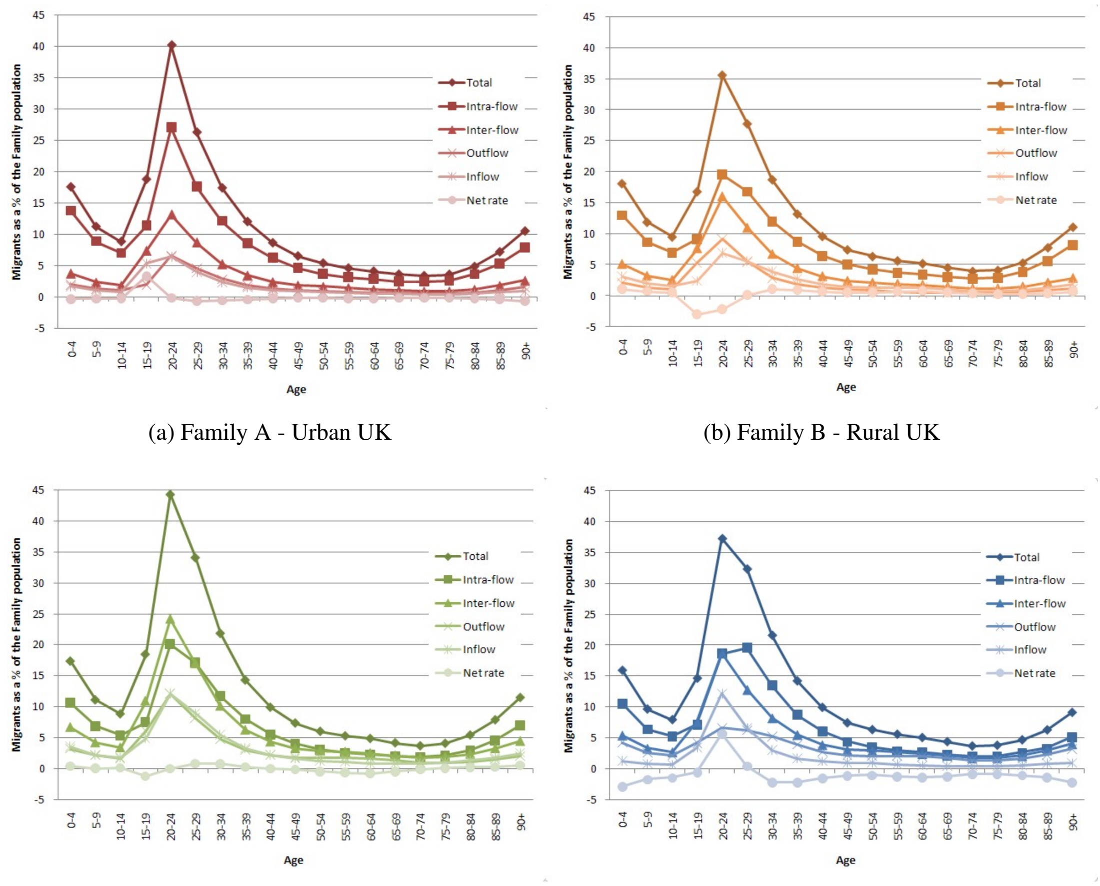


Figure 4.4: Example age-specific migration schedule, Britain, 2000-01
Source: after Rogers and Castro (1981)

Whilst Figure 4.4 is an empirical schedule constructed from 2001 Census data for Britain as a whole based on data for quinary age groups, the model components as identified by Rogers and Castro are included; these identify some distinct phases in the life course where the propensity to migrate fluctuates. Firstly there is a 'pre-labour force' component (between lines 1 and 2), which is characterised by a steady decline in the rate of migration. This decline can be compared directly with the decline shown between lines 1a and 2a. Arrow *A* links these two elements of the schedule and signifies a 'parental shift' where the two comparable rates of migration decline, starting at the average age at which parents have children. After the pre-labour force component. there is a 'labour force' component between lines 2 and 4. Between these lines, line 3 represents the high peak of migration towards the beginning of the labour force component at age 20-24. This peak represents a period of the life course when individuals are more likely to move, and is associated with employment seeking moves associated with labour market influences. Arrow *X* signifies a 'labour force shift' in migration propensities between the pre work-age population and 'first job' age range. The 'post-labour force' component begins after line 4 at age 60-64 when retirement migration is most likely, and shows a steady decline in migration propensity from the initial peak, followed by an increase towards the end of the life course associated with moves to be closer to family or into communal establishments or to be closer to health and other services as the ability to maintain independent living status declines. In the original work by Rogers and Castro (1981), a clear 'retirement peak' was identified. In this example of migration in Britain in 2000-01, the peak is much less evident, suggesting either that retirement is not the catalyst for migration as it has been in the past, or that people are retiring at different ages which has the effect of spreading the effect over several age groups.

4.4.1 Migration age schedules for classification areas

To provide context for the subsequent discussion of broad age-group related patterns of migration between areas in the Vickers et al. classification, it is useful to summarise how migration propensities vary with age by classification area as the effect age has on migration propensity has been well documented (Champion et al., 1998; Stillwell, 2008). Figure 4.5 shows sets of age-specific migration schedules for the four Family groups in the Vickers et al. classification. Pre-age 15, migrants between Families comprise between 2 and 7% of the total population, and in all cases the proportion of migrants decreases as age increases towards 10-14 years. Intra-zonal migrants of this age make up a larger proportion of the total population of all Families (around 5-10% more on average). At age group 15-19, for all Families, inter-zonal migrants begin to comprise a larger percentage of the age group population, attributed differentially to in-migration and out-migration in the case of each Family, as is demonstrated by the net rate schedule. Families A (Urban UK), B (Rural UK) and C (Prosperous Britain) all experience net rates close to 0 in the pre-15 age groups followed by a significant change to a peak of net in-migration (Urban UK) or net out-migration (Rural UK and Prosperous Britain). Whilst Urban UK has more in-migrants than out-migrants at age 15-19, a positive net migration gain is not apparent until age 20-24 for Family D (Urban London). Net losses for Rural UK are more pronounced at age 15-19 due to movements of large numbers of migrants at this age from rural areas to major university towns, most of which are located in Urban UK districts.



(a) Family A - Urban UK

(b) Family B - Rural UK

(c) Family C - Prosperous Britain

(d) Family D - Urban London

Figure 4.5: Age-specific migration rate schedules for total inflow, outflow, intra-Family, inter-Family, total flows and net rates, 2000-01

In the age group of peak migration propensity (20-24) there are differences between Families in terms of inter-zonal and intra-zonal as well as variations in inflow and outflow propensities. For Urban UK and Rural UK, intra-zonal rates are higher, whereas in Prosperous Britain and Urban London, intra-zonal rates are lower (or equal). Relatively high intra-zonal rates in Urban UK can be attributed to young people moving between Britain's larger cities in search of employment opportunities, the same reasons which explain the much higher rates of inflow to Urban London and the peak in outflow from Rural UK districts.

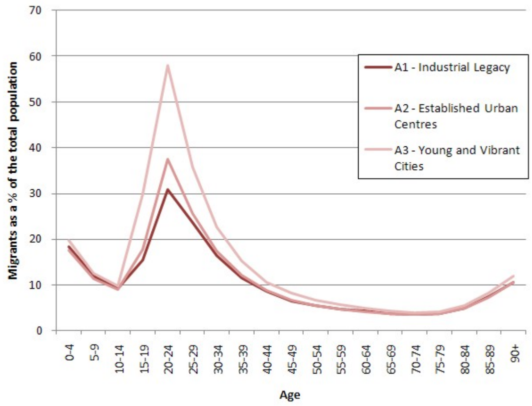
As age increases beyond 20-24 up to the mid 70s, gross migration rates for all flow types decrease across all Families, except for the intra-zonal flow in Urban London where a slight rate increase is evident and likely to be attributed to the inter-zonal migrants at age-group 20-24 increasing in affluence and/or changing personal status and being able to move from their first job residential locations. The decline in in-migration and out-migration rates remain broadly similar in Urban UK whilst the out-migration rates become more significant in Urban London; conversely (but unsurprisingly) the in-migration rates are in Rural UK. Migration rates of all types increase in old age, resulting in higher rates of loss in Urban UK and Urban London and higher rates of net gain in Rural and Prosperous UK.

Whilst Family migration schedules reveal some interesting differences, disaggregating the flows further by district Group or Class offers deeper insights into area-specific internal migration patterns. Figure 4.6 shows Group level schedules for total migration for each Family. Groups A1 (Industrial Legacy), A2 (Established Urban Centres) and A3 (Young and Vibrant Cities) have large differences at the peak migration age group of 20-24. Almost 60% of the population of Group A3 are migrants; this is double the A1 rate and a third greater than the A2 rate. This is a huge variation but it is perhaps not expected that less economically buoyant Industrial Legacy areas are experiencing less young adult migration than Young and Vibrant Cities. Noticeable variations also occur between groups in Prosperous Britain and Urban London, although much smaller variations occur between Groups in Rural UK, except for the Isles of Scilly where the numbers involved are very small. Other points of interest are the continuation of Group D2 (Mercantile Inner London) as an area with a comparatively high proportion of internal migrants from age group 20-24 until age group 35-39. Closer scrutiny of inflows and outflows reveals that this is mainly the result of a much higher outflow rate than any other group. This greater degree of disaggregation at the Group level allows these variations to be observed.

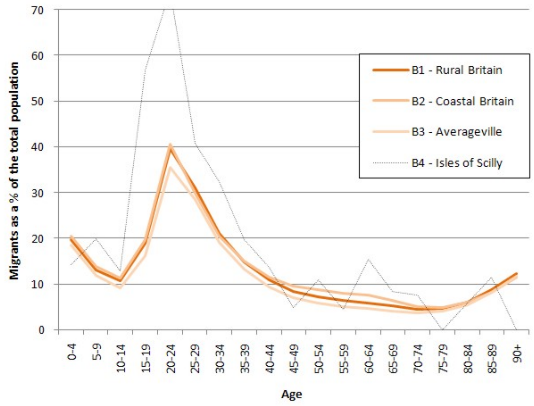
Examination of the age-specific migration schedules for the two most aggregate levels of the Vickers et al. classification has shown that whilst all areas exhibit the same broad patterns, there are distinct variations in rates (particularly at the ages of peak migration) depending on the type of area; Industrial Legacy or Averageville areas experience low migration rates which are especially noticeable at the ages of peak migration propensity and are half the rates of migration in the Young and Vibrant Cities, Prosperous Urbanites and Mercantile Inner London areas. Whilst higher migration at the peak age range might be indicative of active participation in labour markets and education, it could be postulated that lower levels of migration at these key ages in these locations may represent a relative stagnation of the population.

4.4.2 Age-specific migration by area

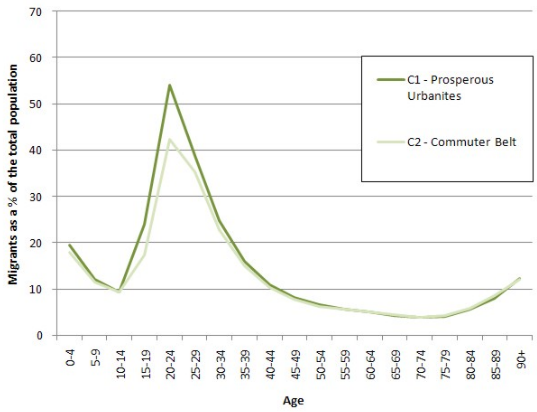
Although useful for assessing the relative magnitude of migration at different ages in different areas, migration schedules tell us little about how age affects the direction and volume of movement in relation to key live stages and give no clues about how areas are linked through migrant flows. In this section this issue is addressed, using migration data for those in five broad age groups: 0-15, 16-29, 30-44, 45-pensionable age (pensionable age in this case defined as 65 for males and 60 for females) and pensionable age and above. These groups were chosen as they represent groupings of around 15 years, making it possible to draw comparisons with the



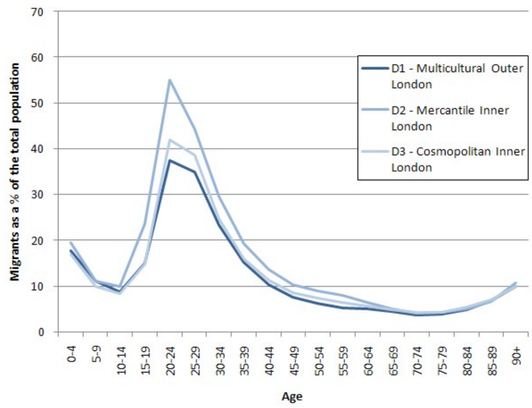
(a) Groups A1 to A3 - Urban UK



(b) Groups B1 to B4 - Rural UK



(c) Groups C1 to C2 - Prosperous Britain



(d) Groups D1 to D3 - Urban London

Figure 4.6: Age-specific migration rate schedules, disaggregated by Group category of district, 2000-01

relative numbers of migrants present in each group, but they also represent recognisable stages in the life course: ages 0-15 are the dependent child years; ages 16-29 contain a number of key life stages in the young adult years: leaving home to study or take a first job; graduating and moving to a first job; moving through the early stages of a career; starting a young family; ages 30-44 are the family rearing years; ages 45-pensionable age are the years after the children have left home; and pensionable age and above are the retirement years. The 16-29 age groups will be disaggregated further where appropriate in this analysis and smaller age groups within this larger group will be looked at separately.

The proportions of each of the chosen age groups that migrate vary appreciably. Perhaps the most striking feature is that almost one quarter of the 16-29 age group are internal migrants. This is by far the largest proportion of the total population in any of the age categories, and is perhaps not surprising in the context of what one might expect given the previous characteristics shown in the age-specific schedule and knowledge about migration as a result of moving to and from higher education institutions and migrations to find a first job. The flows for 0-15 and 30-44 year olds represent 10.5% and 11.5% of their respective populations and are broadly comparable to the 10.7% observed in the total population, whereas the flows for the two oldest age groups are considerably lower than the average at 4.9% and 3.9% respectively.

Beginning with the youngest age group, 0-15 year olds, there is a clear pattern of net out-migration from urban areas - London especially (Table 4.3, columns 2-3). In the year preceding the 2001 Census, Urban London lost almost 23,000 individuals aged 0-15. This was a rate of over 20 people per 1,000. Net out-migration from London also included a loss from Thriving Outer London, part of Prosperous Britain. In all but the Industrial Legacy and M8 Corridor areas of Urban UK, there was also net out-migration. On the other hand, net in-migration of this age group can be found across most of Rural UK and Prosperous Britain, with the highest rates found in the south west of Britain, and outside of the London Commuter Belt area. Paradoxically, the highest rates of gain are to be found in the Aged Coastal Resorts.

In contrast to the 0-15 age group, the 16-29 age group pattern of net migration is virtually the opposite (Table 4.3, columns 4-5). The pattern of individuals in their late teens and twenties moving from rural to more urban areas was identified in the 1991 Census by Rees et al. (1996). The high rate of more than 25 out-migrants per 1,000 of population from rural areas noted in 1991 appears to have continued in 2001. Table 2 shows that in Classes within Rural and Coastal Britain Groups (with the exception of Coastal Resorts), net out-migration rates are around 25 people per 1,000 of population. In fact, it appears that rates of net out-migration increase with increasing rurality, with Rural Extremes experiencing almost double the rate of net out-migration than the Rural Fringe (with the exception of the Scilly Isles which exhibit unusually high in-migration rates due to the very small PAR). Net in-migration rates are high for urban areas as expected in this age category; particularly vibrant urban districts with universities which are likely to attract significant numbers of student migrants and young economic migrants, and London which has always offered education and employment opportunities for young migrants.

The patterns of migration for the 30-44 age group (Table 4.3, columns 6-7) are very similar to those of the 0-15 group, principally because the majority of 0-15 year old migrants will be migrating with parents who are very likely to fall into the 30-44 age category. As with the 0-15 age group, net out-migration is experienced from virtually all of Urban UK (except Industrial Legacy areas), and net in-migration can be observed in all areas defined as Rural UK. Significantly, there is also net in-migration to areas defined as Commuter Belt, as individuals wishing to maintain city jobs move out to areas perhaps perceived more appropriate for raising their families.

The pattern of migration changes for the 45 to pensionable age group (Table 4.3, column 8-9). Whilst there is still a noticeable net out-migration of individuals in this Group from London, the rates of net out-migration are lower (12.1 persons per 1,000 population for this age group compared to 20.2 at 30-44 for Urban London). Similarly, the rate of net in-migration to Rural UK is lower at around 5.9 persons per 1,000 population, although rates are noticeably higher for

Table 4.3: Net migrants and net migration rates by district classification - broad age groups, 2000-01

District Classification (Family, Group, *Class*)	**Net 0-15 Migrants**	**Net 0-15 Mig Rate (per 1,000)**	**Net 16-29 Migrants**	**Net 16-29 Mig Rate (per 1,000)**	**Net 30-44 Migrants**	**Net 30-44 Mig Rate (per 1000)**	**Net 45-PA Migrants**	**Net 45-PA Mig Rate (Per 1,000)**	**Net PA+ Migrants**	**Net PA+ Mig Rate (Per 1,000)**
A: Urban UK	**-10,293**	**-2.45**	**36,864**	**9.44**	**-17,463**	**-3.83**	**-6,762**	**-1.61**	**-7,188**	**-1.92**
A1: Industrial Legacy	3,508	3.1	-14,010	-15.92	3,145	2.58	1,407	1.15	687	0.65
A1a: Industrial Legacy	*3,508*	*3.1*	*-14,010*	*-15.92*	*3,145*	*2.58*	*1,407*	*1.15*	*687*	*0.65*
A2: Established Urban Centres	-8,401	-3.98	13,259	6.98	-11,988	-5.41	-4,974	-2.49	-4,725	-2.68
A2a: Struggling Urban Manufacturing	*-3,083*	*-4.92*	*2,379*	*4.31*	*-4,069*	*-6.18*	*-1,673*	*-2.8*	*-1,846*	*-3.35*
A2b: Regional Centres	*-4,122*	*-13.22*	*15,929*	*42.41*	*-5,481*	*-15.22*	*-1,541*	*-5.25*	*-1,590*	*-5.56*
A2c: Multicultural England	*-2,279*	*-2.78*	*-957*	*-1.42*	*-3,876*	*-4.93*	*-2,374*	*-3.32*	*-1,778*	*-2.93*
A2d: M8 Corridor	*1,083*	*3.08*	*-4,092*	*-13.74*	*1,438*	*3.51*	*614*	*1.57*	*489*	*1.52*
A3: Young and Vibrant Cities	-5,400	-5.61	37,615	33.4	-8,620	-7.69	-3,195	-3.24	-3,150	-3.45
A3a: Redeveloping Urban Centres	*-3,192*	*-4.12*	*27,372*	*32.49*	*-5,119*	*-5.99*	*-1,618*	*-2.1*	*-2,048*	*-2.87*
A3b: Young Multicultural	*-2,208*	*-11.82*	*10,243*	*36.13*	*-3,501*	*-13.14*	*-1,577*	*-7.3*	*-1,102*	*-5.54*
B: Rural UK	**30,076**	**7.19**	**-59,853**	**-18.7**	**39,581**	**8.58**	**28,787**	**5.88**	**19,356**	**4.42**
B1: Rural Britain	17,547	10.54	-35,853	-29.41	23,437	12.63	15,243	7.41	10,279	5.63
B1a: Rural Extremes	*2,272*	*7.46*	*-9,173*	*-41.41*	*3,321*	*9.67*	*3,281*	*8.54*	*1,366*	*4.09*
B1b: Agricultural Fringe	*7,438*	*11.5*	*-14,839*	*-31.88*	*9,755*	*13.82*	*7,609*	*9.32*	*4,799*	*6.16*
B1c: Rural Fringe	*7,837*	*11*	*-11,841*	*-22.25*	*10,361*	*12.86*	*4,353*	*5.08*	*4,114*	*5.76*
B2: Coastal Britain	8,047	9.76	-11,019	-17.44	9,051	10.34	15,270	14.89	8,785	8.07
B2a: Coastal Resorts	*1,094*	*5.85*	*1,095*	*7.02*	*1,451*	*7.09*	*2,175*	*10.32*	*1,416*	*5.99*
B2b: Aged Coastal Extremities	*4,873*	*9.58*	*-9,885*	*-25.82*	*5,646*	*10.57*	*9,210*	*14.44*	*4,981*	*7.96*
B2c: Aged Coastal Resorts	*2,080*	*16.21*	*-2,229*	*-23.94*	*1,954*	*14.29*	*3,885*	*22.01*	*2,388*	*10.56*
B3: Averageville	4,488	2.64	-13,024	-9.65	7,087	3.76	-1,725	-0.95	296	0.2
B3a: Mixed Urban	*4,764*	*4.57*	*-13,254*	*-16.14*	*6,033*	*5.17*	*-1,012*	*-0.86*	*320*	*0.33*
B3b: Typical Towns	*-276*	*-0.42*	*230*	*0.43*	*1,054*	*1.47*	*-713*	*-1.12*	*-24*	*-0.05*
B4: Isles of Scilly	-6	-16.76	43	126.84	6	13.95	-1	-1.78	-4	-8.93
B4a: Isles of Scilly	*-6*	*-16.76*	*43*	*126.84*	*6*	*13.95*	*-1*	*-1.78*	*-4*	*-8.93*
C: Prosperous Britain	**3,200**	**1.66**	**-2,545**	**-1.6**	**7,582**	**3.36**	**-10,235**	**-4.83**	**-2,985**	**-1.79**
C1: Prosperous Urbanites	-2,956	-4.74	13,839	22.58	-3,957	-5.28	-3,687	-5.65	-1,395	-2.63
C1a: Historic Cities	*299*	*0.99*	*6,774*	*22.18*	*-263*	*-0.74*	*-972*	*-2.81*	*-232*	*-0.77*
C1b: Thriving Outer London	*-3,255*	*-10.13*	*7,065*	*22.96*	*-3,694*	*-9.36*	*-2,715*	*-8.87*	*-1,163*	*-5.04*
C2: Commuter Belt	6,156	4.71	-16,384	-16.71	11,539	7.64	-6,548	-4.46	-1,590	-1.4
C2a: The Commuter Belt	*6,156*	*4.71*	*-16,384*	*-16.71*	*11,539*	*7.64*	*-6,548*	*-4.46*	*-1,590*	*-1.4*
D: Urban London	**-22,983**	**-20.08**	**25,534**	**19.71**	**-29,700**	**-20.24**	**-11,790**	**-12.11**	**-9,183**	**-12.05**
D1: Multicultural Outer London	-6,027	-10.89	293	0.55	-7,276	-11.32	-4,685	-9.5	-3,252	-8.29
D1a: Multicultural Outer London	*-6,027*	*-10.89*	*293*	*0.55*	*-7,276*	*-11.32*	*-4,685*	*-9.5*	*-3,252*	*-8.29*
D2: Mercantile Inner London	-6,932	-37.52	11,799	38.93	-11,402	-35.77	-2,819	-14.69	-2,351	-15.84
D2a: Central London	*-6,890*	*-37.43*	*11,765*	*39.03*	*-11,396*	*-35.98*	*-2,855*	*-15.02*	*-2,349*	*-15.95*
D2b: City of London	*-42*	*-62.69*	*34*	*20.62*	*-6*	*-3.05*	*36*	*20.04*	*-2*	*-1.82*
D3: Cosmopolitan Inner London	-10,024	-24.67	13,442	29.18	-11,022	-21.79	-4,286	-14.85	-3,580	-16.17
D3a: Afro-Caribbean Ethnic Borough	*-6,533*	*-26.63*	*11,202*	*39.74*	*-6,982*	*-20.98*	*-2,761*	*-15.23*	*-2,090*	*-15.14*
D3b: Multicultural Inner London	*-3,491*	*-21.7*	*2,240*	*12.53*	*-4,040*	*-23.34*	*-1,525*	*-14.2*	*-1,490*	*-17.87*

Coastal Britain. Within this Group, the Classes of Aged Coastal Extremities and Aged Coastal Resort exhibit in-migration rates of 14.4 and 22.0 persons per 1,000 respectively. The other key change is in the Commuter Belt where the net in-migration rate of around 7.6 persons per 1,000 changes to net out-migration of 4.5 persons per 1,000 for the 45 to pensionable age group. Drilling down through the classification to the districts beneath, it can be seen that particular districts around London, Birmingham and Manchester all show a clear shift to negative net migration for the older working ages.

The final age group includes those of pensionable age and above (Table 4.3, column 10-11). Essentially, the overall migration patterns of this group are very similar to the 45 to pensionable age group, characterised by net out-migration from Urban London and other built-up areas in Urban UK and Prosperous UK, and net in-migration to Rural UK, especially the Coastal Resort areas. The only areas of relatively high in-migration for this age group are districts along the south coast, and in Norfolk and Lincolnshire. Moreover, whilst there is still an overall net out-migration from Commuter Belt areas, this is lower (1.4 people per 1,000 population) than the rate for the preceding age group. Furthermore, for a number of Commuter Belt districts in the Home Counties, the rate of migration has switched from negative in the 44 to pensionable age group to positive in the oldest group - perhaps reflecting the movement of dependent elderly individuals to the homes of younger relatives who are able to provide care.

Whilst Table 4.3 contains much information in relation to the volume and direction of migration of those in different age groups for all areas in the classification, it reveals little about the linkage between areas. For example where Mercantile Inner London experiences a very high net in-migration rate of around 39 migrants per 1,000 population, the origins of the migrants to this area are not clear. As a result attention will now be turned to area linkage. An analysis of the flows between Family areas gives a useful overview of the linkage between these areas overall, and for different age groups. Table 4.4 summarises the flows through net migration rate balances. Rates of migration between areas were calculated using the sum of the origin and destination PAR as the denominator. Consequently these directional net migration rates cannot be compared directly with the single area rates calculated earlier in the chapter:

$$nm_{ij} = 1000\left(\frac{M_{ji} - M_{ij}}{P_i + P_j}\right) \tag{4.1}$$

where:

nm_{ij} = net migration rate between area i and area j;

M_{ji} = in-migration to area i from destination j;

M_{ij} = out-migration from area i to area j;

and P_j = population area j.

Taking the total migrants column first, it is possible to see that the highest positive balance is between Prosperous Britain and Urban London - the former gaining from the latter at a rate

Table 4.4: Inter Family flow rates by broad age group (PAR is summation of origin and destination)

Origin/Destination	Destination/Origin	Total	0-15	16-29	30-44	45-PA	PA +
B Rural UK	A Urban UK	-0.087	-1.524	5.956	-1.965	-0.896	-0.863
C Prosperous Britain	A Urban UK	0.108	-0.147	1.309	-0.368	0.047	-0.154
D Urban London	A Urban UK	-0.17	0.635	-2.436	0.508	0.212	0.145
C Prosperous Britain	B Rural UK	0.997	1.302	-1.692	1.494	1.956	1.143
D Urban London	B Rural UK	0.875	1.751	-2.092	1.858	1.175	1.056
D Urban London	C Prosperous Britain	1.908	3.336	-1.198	4.117	1.225	1.275

of almost 2 people per 1,000 population. Thereafter, rates of movement between Prosperous Britain and Rural UK, occur at a rate of around 1 person per 1,000 population. Examining the age breakdown shows that this prosperous/rural linkage is driven by migration flows in the 30-44 and 45-pensionable age groups. There is a heavy exodus from London to the surrounding Prosperous Britain area at a rate of 4.1 per 1,000 at age group 30-44, mirrored by high rates in age group 0-15. The 45 to pensionable age group exhibits a rate of movement from Prosperous Britain to Rural UK of around 2 per 1,000 population. As might be expected, the 16-29 age group exhibits very different patterns with a very high rate of almost 6 per 1,000 population moving from Rural UK to Urban UK, and relatively high rates to Urban London from Urban and Rural UK. At pensionable age and above, there are high rates moving into Rural UK from all other Families and from London to Prosperous Britain.

Even at the more aggregate Family level, a simplified hypothetical life course model of migration can be postulated as follows: Initially, during age group 0-15, the migrant moves with parents in the 30-44 age group from London to a peri-London/Prosperous Britain or rural area. When he/she has reached the age of leaving home for higher education or first job, a move is made from the rural or Prosperous Britain area associated with childhood into and Urban UK area in order to attend a higher education institution or to seek enhanced employment opportunities. Post-university, the migrant then moves from Urban UK to Urban London in search of a graduate job. After a few years of living in London, the migrant decides to move out to Prosperous Britain or to a Rural area, perhaps to start a family, but very likely keeping a London-based job. After the children have been reared and have left home and the migrant is nearing retirement, there may be another move from the commuter belt to a more rural, perhaps coastal location. If this move does not occur before the age of retirement, then it is very likely to happen post-retirement.

Whilst these linkages between Family areas are very revealing, dropping down the Vickers et al. hierarchy to the Group level provides more detail about which areas within the families are gaining and losing population from other areas. Through analysing the linkages in this way, especially when disaggregating the flows by age group, it is possible to start making judgements about the potential impacts of the flows that are taking place.

Figure 4.7 is a representation of the top ten flows in terms of total migrants between Vickers et al. Groups. The largest flow is between the Established Urban Centres in Urban UK and Averageville in Rural UK - a flow of over 60,000 migrants. Rural Britain is an important origin and destination, receiving large flows from the Established Urban Centres and Young and Vibrant Cities, Averageville and Commuter Belt, whilst also losing large number of migrants in the opposite direction to the first three of these Groups. As shown by the size of the circles representing the populations of these Groups, the large flows are taking place between the three Groups with the largest populations. The only flow in the top ten from a Group with a relatively small population is that between the Prosperous Urbanites Group and Commuter Belt. Disaggregating the flows by broad age-group reveals which sections of the population are driving these flows. For example, around 63% of the 51,340 migrants moving between Rural Britain and Young and Vibrant Cities are in the 16-29 age group; perhaps unsurprising as this is the most active age in the migrant population, although this is 17.5% more migrants than would be expected given the average distribution of migrants across all age groups.

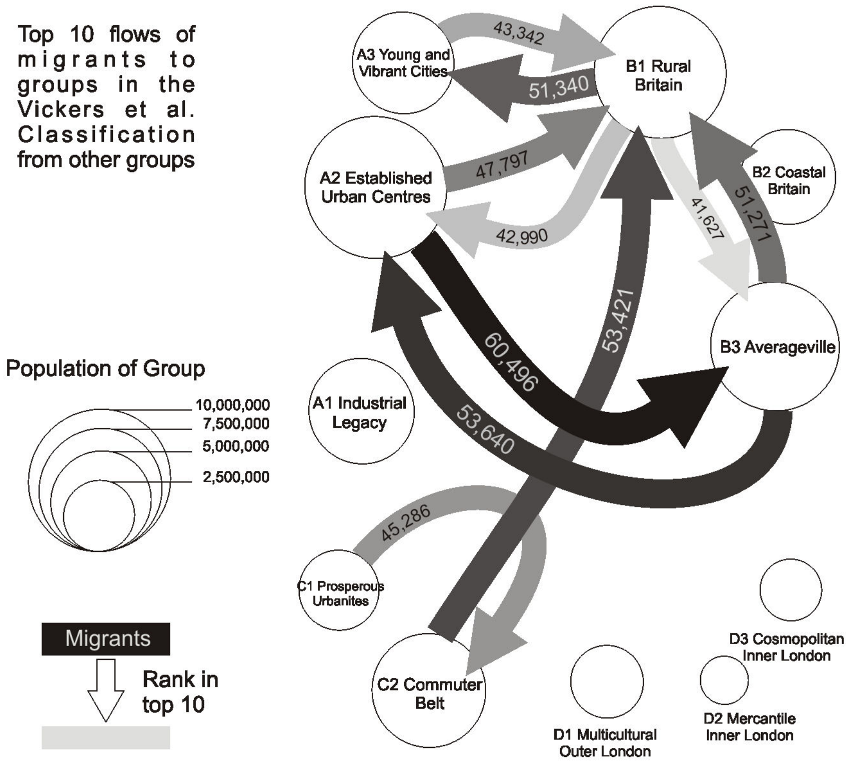


Figure 4.7: The top ten flows of migrants to Groups in the Vickers et al. classification

The interpretation of gross flows should always be carried out with caution, even when we have an idea of the underlying populations of the areas between which flows are taking place as in Figure 4.7. To address this, Figure 4.8 gives an indication of the top and bottom ten net rates of flow between the same Groups. As with the Family level, rates have been calculated using the summation of the origin and destination PAR. In this diagram, this gives the added

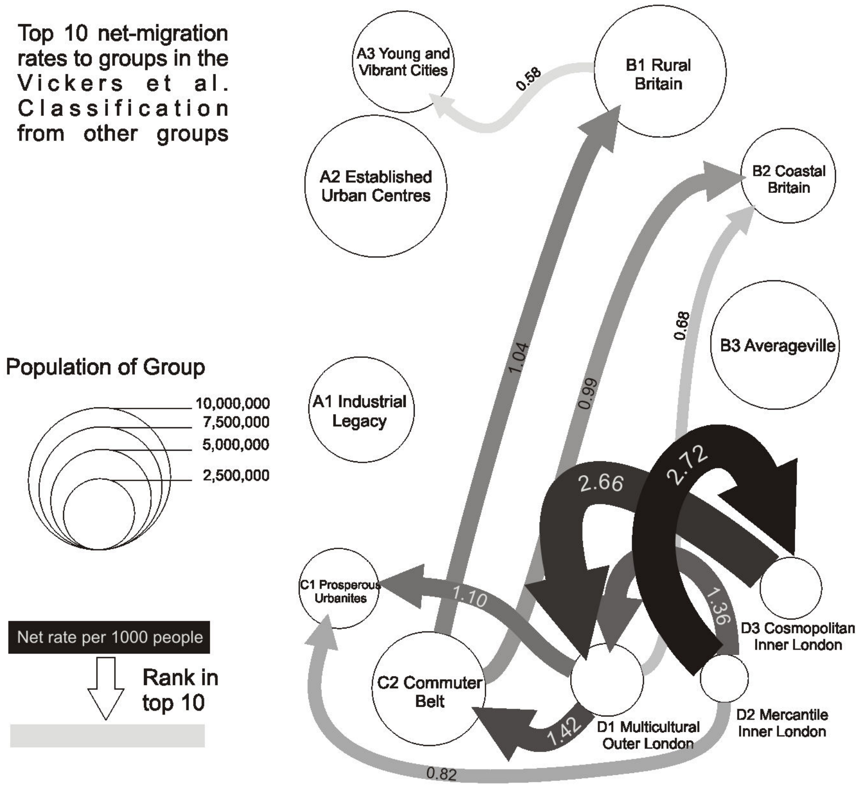


Figure 4.8: The top ten net-migration rates between Groups in the Vickers et al. classification

advantage of revealing the rate of movement in the opposite direction; where the net rate is positive in one direction it is negative by the same amount in the other. Figure 4.8 indicates the shift of focus to Groups in Urban London. The highest rate of net gain is by Cosmopolitan Inner London from Mercantile Inner London (2.72 per 1,000 of their combined populations). In this case, net rates are slightly misleading measures, as whilst there is a very large net gain in one direction, there are still a significant number of migrants moving from Cosmopolitan Inner London to Mercantile Inner London. The net in-migration and out-migration rates shown in Table 4.5 reveal that the highest in-migration rate for all groups is to Mercantile Inner London Cosmopolitan Inner London; it is just that the out-migration in the other direction is even higher. In fact, both Figure 4.8 and Table 4.5 show that there are very high rates of movement between all groups in the Urban London Family - undoubtedly a function of the high populations in all of these areas and their very close proximity to one another.

It is noticeable, with respect to the groups within the Urban London Family, that there is a substantial movement from inner to outer London and from outer London to areas in the London periphery; Commuter Belt and Prosperous Urbanites. Other flows that feature in the top ten rates are from Commuter Belt areas to Rural and Coastal Britain, flows from outer London to coastal Britain, and the key flow picked up in the gross flows in Figure 4.7 between Rural Britain and the Young and Vibrant cities.

Table 4.5: Top and Bottom 10 in- and out-migration rates between pairs of Vickers et al. classification Groups, by broad age group

		Total			**Age 0 to 15**			**Age 16 to 29**			**Age 30 to 44**			**Age 45 to PA**			**Age PA and above**		
	Rank	Orig	Dest	Rate	Orig	Dest	Rate	Orig	Dest	Rate	Orig	Dest	Rate	Orig	Dest	Rate	Orig	Dest	Rate
In migration rate top 10	1	D3	D2	17.92	D3	D1	10.47	D3	D2	40.63	D3	D2	17.31	B1	B2	7.30	B1	B2	5.10
	2	D2	D3	15.29	B1	B2	8.49	D2	D3	35.56	D2	D3	17.08	C2	B2	5.46	C2	C1	3.35
	3	C2	C1	12.94	D3	D2	8.24	C2	C1	34.51	D3	D1	15.95	C2	C1	5.00	C2	B2	3.30
	4	D3	D1	12.58	C2	C1	8.02	B1	A3	28.78	C2	C1	13.10	D3	D2	4.81	D3	D1	3.08
	5	D1	D3	11.12	A2	B3	6.73	D3	D1	25.08	D1	D3	11.15	B3	B2	4.72	C2	B1	3.04
	6	D1	D2	10.10	C2	B1	6.48	D1	D3	24.57	B1	B2	10.47	C2	B1	4.66	B3	B2	2.96
	7	B1	A3	10.05	B3	B2	6.43	A3	D2	24.28	C1	C2	9.39	D3	D1	4.63	B2	B1	2.78
	8	B1	B2	9.25	B3	B1	6.05	D1	D2	23.72	A2	B3	9.04	D2	D3	4.20	B3	B1	2.55
	9	C2	D2	8.58	D1	D3	5.96	C2	D2	23.49	D1	D2	9.04	D1	D3	3.92	A2	B3	2.48
	10	C1	D2	8.08	C1	C2	5.94	C2	A3	21.17	D2	D1	8.08	B1	A3	3.73	B1	C2	2.46
In migration rate bottom 10	1	D2	A1	0.15	D2	A1	0.11	D2	A1	0.36	D2	A1	0.18	D2	A1	0.08	D2	A1	0.05
	2	D3	A1	0.23	D2	A2	0.18	D3	A1	0.50	D3	A1	0.30	D3	A1	0.11	D3	A1	0.07
	3	D2	A2	0.32	D3	A1	0.22	D1	A1	0.77	D2	A2	0.38	D2	A2	0.11	D2	A2	0.07
	4	D1	A1	0.37	A1	D3	0.22	D2	A2	0.82	D1	A1	0.45	A1	D1	0.16	A1	D3	0.09
	5	D2	B3	0.43	A1	D2	0.25	D2	B3	1.06	A1	D1	0.47	D1	A1	0.17	D1	A2	0.12
	6	D2	B1	0.62	D2	B3	0.27	D2	B1	1.07	D2	B3	0.53	D2	B3	0.20	D3	A2	0.13
	7	D2	B2	0.64	A1	D1	0.31	D2	B2	1.21	A1	D3	0.54	D1	A2	0.25	D2	A3	0.13
	8	A1	D1	0.66	B2	D2	0.38	D3	B1	1.32	A1	D2	0.64	D3	A2	0.26	D1	A1	0.14
	9	C1	A1	0.71	D1	A1	0.40	D3	A2	1.77	C1	A1	0.80	A1	D3	0.28	A1	D2	0.14
	10	D3	B1	0.73	D2	A3	0.47	D1	B1	1.80	D2	B2	0.82	D2	A3	0.32	A1	D1	0.17
Out migration rate top 10	1	D2	D3	25.10	D3	D1	14.26	D2	D3	54.05	D2	D3	27.11	D3	D1	7.91	D3	D1	5.46
	2	D3	D1	17.46	C1	C2	12.44	D3	D1	28.96	D3	D1	20.26	C1	C2	6.57	C2	B1	4.89
	3	D2	D1	14.55	D2	D3	11.60	C1	C2	27.52	C1	C2	18.91	C2	B1	6.53	B2	B1	4.66
	4	C1	C2	14.29	D2	D1	11.47	D3	D2	26.73	D2	D1	16.29	D2	D3	6.31	D1	C2	4.31
	5	D2	C1	11.17	D2	C2	10.32	B1	A3	26.58	D1	C2	14.65	D2	D1	6.09	C1	C2	4.14
	6	D3	D2	10.92	D1	C2	9.29	D2	D1	25.17	D2	C1	12.70	B2	B1	5.91	C1	B1	3.92
	7	D1	C2	10.82	B2	B1	8.55	C2	A3	24.31	D2	C2	11.85	D1	C2	5.73	D2	D1	3.91
	8	D2	C2	9.30	D2	C1	8.45	C2	C1	21.57	B2	B1	10.95	C1	B1	5.61	A3	B1	3.83
	9	B2	B1	8.75	C2	B1	8.26	D1	D3	21.27	D3	D2	10.91	A3	B1	5.16	D2	C2	3.80
	10	A3	B1	8.48	A3	B1	7.42	B3	A2	20.42	A3	B1	10.22	D2	C2	4.76	D3	C2	3.61
Out migration rate bottom 10	1	A1	D2	0.24	A1	D2	0.04	D3	A1	0.95	A1	D2	0.17	A1	D2	0.05	A1	D2	0.02
	2	A1	D3	0.28	B3	D2	0.07	D2	A1	1.04	A1	D3	0.22	A1	D1	0.06	A1	D3	0.02
	3	A1	D1	0.31	A2	D2	0.08	A1	D2	1.10	A1	D1	0.25	A1	D3	0.07	A2	D2	0.03
	4	B3	D2	0.49	A1	D3	0.08	A1	D3	1.20	B3	D2	0.33	B3	D2	0.07	B1	D3	0.03
	5	A2	D2	0.52	B2	D2	0.08	A1	D1	1.26	A2	D2	0.35	A2	D2	0.08	B3	D2	0.05
	6	B2	D2	0.53	B1	D2	0.09	D1	A1	1.28	B1	D3	0.39	A2	D3	0.11	B2	D2	0.05
	7	B1	D3	0.57	B1	D3	0.12	A2	D2	2.13	B2	D2	0.40	B1	D3	0.12	B1	D2	0.05
	8	B1	D1	0.57	A1	D1	0.15	C2	A1	2.24	B1	D2	0.45	B3	D3	0.13	A2	D3	0.06
	9	B1	D2	0.61	B3	D3	0.17	A2	B2	2.27	B1	D1	0.50	A2	D1	0.14	A1	D1	0.06
	10	B3	D3	0.65	C2	D2	0.18	B3	D2	2.29	B3	D3	0.52	B2	D3	0.15	B3	D3	0.06

As has already been demonstrated earlier in this chapter, disaggregating the flows of individuals by age offers even greater insight and increases our understanding of the linkage between different types of area. Whilst it is important to examine the areas which exhibit high levels of linkage through migration flows, it is equally as important to look at the areas which show very little linkage. Attention will be turned to this in due course, but with reference to Table 4.5 area linkages displayed at each stage in the life cycle will now be examined. At age groups 0-15 and 30-44, the highest rates of in-migration (destination denominator) are to outer London (D1) from inner London (D2 and D3). These high rates are also exhibited in out-migration (origin denominator) from inner London to outer London. Interestingly there is also a high rate of movement at this age group from Rural Britain (B1) to Coastal Britain (B2), a movement which is in the top ten for the 30-44 age group as well as the 45 to pensionable age group. Indeed, this stream might be seen as an extension of counterurbanisation - an extended ruralisation, or a continued diffusion of population down the urban hierarchy. Counterurbanising moves to rural areas (designated by B prefixes) are common across the 0-15, 30-44, 45-pensionable age and pensionable age and above groups, but it appears that rural coastal areas are gaining population from other rural areas across all these age ranges. This diffusion of the population down the urban hierarchy at the family age ranges (0-15 and 30-44) is further evidenced with moves from Established Urban Centres (A2) to Averageville (B3) in both of these Groups.

As might have been expected, age group 16-29 exhibits very different area linkages. London features importantly as with the other groups, but this importance is increased. As an in-migration destination, areas in Urban London feature in seven of the top ten destinations, six of these being in the two inner London Groups. The only origin not from London is Young and Vibrant Cities (A3). These flows can be linked to the other important flows from Rural Britain (B1) and Commuter Belt (C2) to Young and Vibrant Cities (A3) and together represent an apparently common migration pattern: the migration from rural parental home to university in a Young and Vibrant city, and then from university to the first job in London. Examining the bottom ten inflows and outflows for each age group - equally interesting for different reasons since these flows show where there is very little linkage between pairs of areas - it is immediately clear to see that across all age groups, the Industrial Legacy (A1) Group shows very little linkage with London (D), or indeed many other Groups in Britain, including Prosperous Urbanites (C2). These low volumes of flows could be interpreted in two ways - a lack of employment opportunities or other pull factors enticing migrants from London, but also an isolation of individuals in these areas, especially in the age of peak migration, from the economic, social and cultural escalator that is operating in London. For all age groups apart from the 16-29 group, the other areas which exhibit very low levels of interaction (certainly in one direction) are those in the Rural UK and London. Featuring in the bottom ten out-migration rates in all these groups are moves from Rural UK areas to London. This suggests that whilst the border of London is quite porous in one direction with Rural UK areas gaining from the capital, once people have moved in this direction they will not move back in the opposite direction.

One interesting aside in the Rural/London story is that featuring in the bottom ten across the board (except for in the oldest age group) is the move from Mercantile Inner London (D2) to Averageville (B3), suggesting that once the bright lights of the big city have been sampled, migrants are very reluctant to settle in small town rural Britain.

4.5 Measuring turnover and churn

Migration into and out of areas in Britain will have an effect on the stability of the population in an area, and the net migration variables (balances and rates) that have been used so far provide some measure of the stability of the population, but they do not paint a complete picture. Of course, natural change components such as births and deaths will have an impact as well as international migration, and certainly there will be some areas that feature above or below average birth, death or external migration rates. We know, however, that internal migration has been the most important component driving population change across Britain in many areas in recent decades (Christophersen, 1997; Rees et al., 1996), although the impact of international migration in the south (Champion and Congdon, 1992) has been of considerable importance since the 1980s, with London being the most common destination in Britain for immigrants in the 12 month period before the 2001 Census (Horsfield, 2005). Since 2001, whilst there have been significant inflows from overseas of asylum seekers to London and new labour migrants to other parts of the country, as well as steady flows of emigrants, the average international turnover of population in England and Wales has only been around 800,000 people per year since 2001; on the other hand the volume of inter-district migration in England and Wales, as indicated by patient registration data has remained at over 2.4 million per year (ONS, 2008c).

It is easy to make the assumption that a low level of net in-migration or out-migration means a relatively stable population, but this is not necessarily the case. For example, a hypothetical area with 1 million residents at the end of a year that had seen 100 residents move into the area and 101 residents depart over that period, would have a net migration rate per 1,000 people of -0.001. This rate would be identical if, for the same area, 10,000 residents had moved in and 10,001 residents had moved out. Obviously in this example an identical rate is obscuring a hugely different turnover of population for the area and a massive change in the composition of the resident population. The limitations of net migration as a measure have been recognised (Rogers, 1989) and one alternative to measuring net migration has been to measure migration effectiveness or efficiency (Equation (2.8)). This has been used in previous research both as an alternative to and in conjunction with net migration rates (Stillwell et al., 2000, 2001). Indices of migration effectiveness standardise rates of migration by using gross in-migration plus out-migration flows as the denominator rather than PAR. The direction of flow is standardised by the magnitude of the flows in both directions rather than the population of an area, but in doing so, the direction or symmetry of the flow is still of central importance. Consequently, the nature of the migration effectiveness measure does not make it the most suitable option for assessing the relative stability of underlying area populations.

To address these issues some new metrics will be introduced here: population 'turnover' and 'churn'. Population turnover (TO) by age group a and sex s for area i is defined as:

$$TO_i^{as} = 1000\left(\frac{O_i^{as} + D_j^{as}}{P_i^{as}}\right) \tag{4.2}$$

where:

$i = j$;

D_j^{as} is the total in-migration of those in age group a and sex s to area i;

O_i^{as} is the out-migration of those in age group a and sex s from area i;

and P_i^{as} is the population in age group a and sex s to area i.

Turnover is useful as it takes account of both the inflow and outflow and gives a measure of how the population of an area has changed in a way that standard (inflow minus outflow over population) rate calculations do not. It should be noted that whilst the term 'turnover' is one that can be found throughout the demographic literature, its precise meaning and the way in which it is calculated is not universal. For example, in an article published by the ONS (2007c) turnover (for small areas) is calculated by averaging internal migration flows over a three-year period. This is to avoid the possible distorting effects that might be caused in smaller areas by localised phenomena, such as the building of a new housing estate or the demolition of old housing. In other work by Large and Ghosh (2006) turnover is calculated over one year but using both internal and international migration data. Bailey and Livingston (2007), on the other hand, calculate turnover just from internal migration data, but rather than using only inflows and outflows, also incorporate within area moves. Here, the term population 'churn' (CH) is preferred for a measure that includes within-area migration as well as inflows and outflows:

$$CH_i^{as} = 1000\left(\frac{O_i^{as} + D_j^{as} + M_{ii}^{as}}{P_i^{as}}\right) \tag{4.3}$$

or

$$CH_i^{as} = 1000\left(\frac{O_i^{as} + D_j^{as} + M_{ii}^{as} + M_{ii}^{as}}{P_i^{as}}\right) \tag{4.4}$$

where:

$i = j$;

M_{ii}^{as} = total migrants of age group a and sex s within area i.

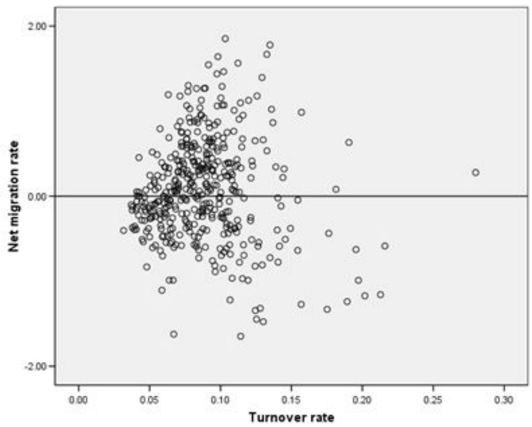
Equation (4.3) derives a measure of churn by counting migration to an area, from an area and within an area as single events. It could be argued, however, that a within-area migration actually encompasses two migration events - leaving the origin and arriving at the destination and therefore, consistent with the inflow and outflow events being counted twice, the within-area migration event should be treated similarly, as indicated in Equation (4.4) (the method used by Bailey and Livingston in calculating their turnover statistics). Providing two methods for measuring churn inevitably invites the question: which is the appropriate measure to use? This would depend upon the reason for measuring churn. It could be that a method for accurate measurement of the total migration events in an area is required; in which case, the latter of the two equations may be more appropriate. If the reason for measuring churn, however, is to get some purchase on the stability of the population in the given area of interest (i.e. how much the area is comprised of the same or different individuals between years) then counting the within-area move as two events could be more problematic, and may be misleading since it is double counting a single event. If an individual is leaving a residence in an area to move to another residence in the same area, then the area actually comprises the same individuals at the beginning of the period as it does at the end; the population size is remaining stable even though people are moving around.

The movement of an individual or individuals is likely to affect the areas they are moving from or to if the origins and destinations differ in demographic composition. If Tobler's 'first law' (Tobler, 1970) is accepted, then closer places are more related than distant places; moves within smaller areas, therefore, are more likely to be between similar places with similar populations than moves within larger areas. Of course, as small and large areas are being used as a proxy for distance, it is difficult to be completely accurate. Accepting this broadly though, it could be said that within-area moves are less likely to be perturbing to the population for small areas than for large areas, therefore counting the within-area move as two events could be more distorting for small areas than for large. With it being difficult to ascertain the exact effect that counting the within-area move twice will have in relation to the size of the area, and because population stability defined by the extent to which populations in an area differ or remain the same over a period of time is to be examined, it is more reliable to count the within-area move once. A within-area move will have some perturbing effect, and counting the migrant once acknowledges this without running into problems of possible over-emphasis. Equation (4.3), therefore, will be adopted in this analysis, although it should be acknowledged that if Equation (4.4) were adopted the interpretation of some results may differ.

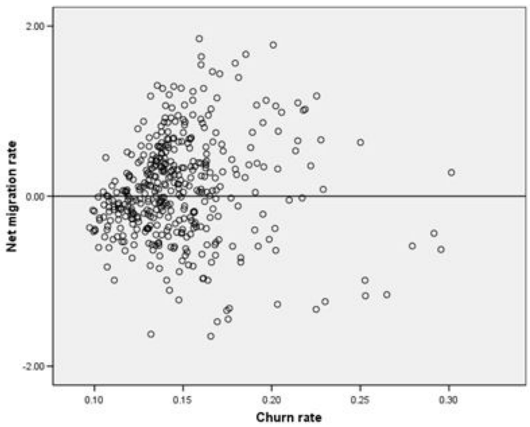
As outlined by Bailey and Livingston (2007), churn is a particularly important aspect of population flow as it is associated more closely with deprivation, especially when small areas are involved. Specific local factors which may agitate local populations at the small scale cease to be important at larger scales. Despite this, measuring population churn even at coarser geographies is important for ascertaining a more accurate measurement of the relative stability of the population in different areas. It could be argued that where two areas with the same levels of population turnover are compared, it would be the area with the higher levels of internal movement relative to the population size that would have the less stable population.

Figure 4.9 plots rates of net migration against turnover and churn for all districts in Britain.

For both turnover and churn there is a relatively even distribution between positive and negative net migration values; however, in both cases there is a positive skew with the frequency of districts with higher rates of turnover and churn tailing off as rates increase. Areas with high turnover and churn but relatively low net migration rates include the university towns of Oxford and Cambridge, as well as inner London boroughs such as Islington, Wandsworth, Hammersmith and Fulham, Camden and Westminster. In these areas, whilst the populations may appear to be relatively stable when examining the low net migration figures, they actually exhibit comparatively unstable populations when the volume of moves is taken into consideration.



(a) Net migration versus turnover rate



(b) Net migration versus churn

Figure 4.9: Net migration rate versus turnover and churn rates for all districts in Britain, 2000-01

In order to gain an appreciation of population stability, is it important to study both turnover and churn together or does the measurement of one reveal something about the other? Figure 4.10 plots rates of turnover against churn and reveals a linear relationship between the two. The outliers represent districts with high scores for both measures of population (in)stability. All of the numbered outliers are located in central London, with the exception of 54 and 251 which are Oxford and Cambridge. Much of this linear relationship, however, is likely to be due to in-migration and out-migration being included in both turnover and churn calculations. It is well-known that area in-migration and out-migration rates are strongly correlated (Cordey-Hayes and Gleave, 1975; Rogers, 1978).

So with Figure 4.10 suggesting a strong relationship between turnover and churn it could be argued that the use of both measures is unnecessary, however, examination of Figure 4.11 provides counter evidence. Here the graph plots intra-district migration rates against turnover and presents a random distribution. This suggests that there is no relationship between the rates of migration within and between districts, thus justifying the use of both turnover and churn in this analysis.

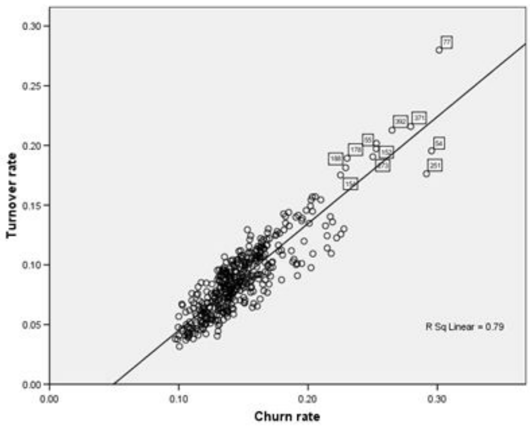


Figure 4.10: Turnover rate versus churn rate, districts in Britain, 2000-01

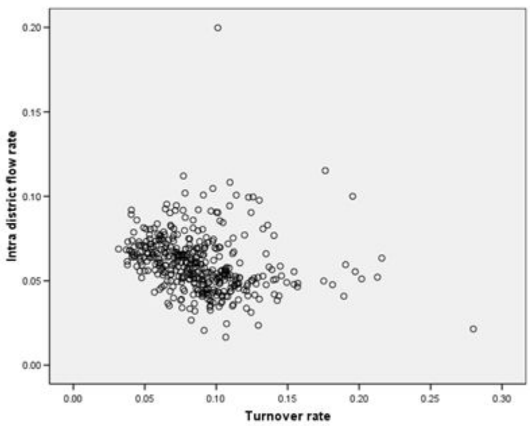


Figure 4.11: Intra-district flow rate versus turnover rate, districts in Britain, 2000-01

4.5.1 Turnover and churn rates by district type

Comparison of the aggregate population turnover and churn statistics with the net migration statistics for districts in Britain shows that these additional measures of population interaction reveal something about the stability of population that cannot be inferred from the standard net migration figures. Table 4.6 gives a comparison of these statistics for total migrants as well as for males and females by the Vickers et al. Family, Group and Class of district.

Table 4.2 has already shown us that the highest net migration rates for district families in Britain are Urban London and Rural UK, with considerably more people moving in than out, and out than in respectively for these two families. As was also noted, examination of the large aggregate numbers of in-migrants and out-migrants would appear to confirm the importance of Urban London and Rural UK in the internal migration story. However, as far as population stability is concerned, the size of the underlying PAR for Rural UK in comparison with the other district Families means that, despite the very large volume of in-migrants and the correspondingly high in-migration rate, the populations of areas within Rural UK are relatively stable compared to other Families. Looking at the aggregate population turnover and churn statistics for the four Families and comparing them to both the averages for all Families, Groups and Classes and each other (Table 4.6), it is clear to see that Rural UK has much greater stability.

From an examination of Table 4.6, it is apparent that Urban London remains a very important location for internal migration in Britain when turnover and churn are taken into consideration along with net migration. With scores of 63 and 151 persons per 1,000 for turnover and churn respectively, Urban London has the second highest turnover score for families and the highest churn score, indicating that not only does it have very significant net out-migration, but with high turnover the change in population is also significant or at least more significant than for all other Families except Prosperous Britain. The high level of churn suggests that movement of population within districts in Urban London is also more significant than it is for other Family categories. Taking all of these measures into consideration it can be concluded that Urban London has the most dynamic population of all Families in the Vickers et al. classification.

In contrast to Urban London, Rural UK shows more stability at this aggregate level. The figures for turnover and churn for this Family are 43 and 120 persons per 1,000 respectively. Not only are these significantly lower than the equivalent figures for Urban London, but they are also lower than the figures for Prosperous Britain (with scores of around 67 and 139 persons per 1,000 for turnover and churn respectively). This means that despite far fewer people moving in and out of districts in Prosperous Britain, the movement is more perturbing than it is for Rural UK. Urban UK, on the other hand, as well as having the lowest rate of net migration, also has the lowest levels of population turnover: a rate of around 38.5 people per 1,000 of population. Levels of population churn are higher than they are for Rural UK but still lower than the mean for all families, groups and classes of district. The districts classified as Urban UK, therefore, can be seen to have far more settled or stable populations than the rest of Britain.

It can be seen in the comparison of nearly all Family, Group and Class categories that males generally have higher rates of population turnover and churn than females, but lower rates of net migration (both in and out). This may seem counter-intuitive and needs explanation. Turnover and churn are measures that take into account total population movements in relation to the underlying population at risk in a way that net migration does not. Net migration will only indicate the balance of movement (either in or out) in relation to the population; this allows one to see if an area is gaining or losing population, and the relative level of this gain or loss. Population turnover and churn will not give an indication of the balance of movement, but will give a standardised measure of the amount of movement in relation to the population at risk. Higher levels of turnover and churn mean that there will be greater numbers of people moving in total in relation to the underlying population, whereas higher levels of net migration just show that there are more people moving in a particular direction. The evidence here suggests that when females move in or out of Family, Group or Class categories, the balance of movement leans more heavily to either in or out. The direction of flow is more asymmetric, but total turnover and churn rates are comparatively low. The net rate of male movement is higher (in the

Table 4.6: A comparison of net migration, turnover and churn statistics for classifications of district in Britain, 2000-01

District Classification (Family, Group, *Class*)	**Total net migration rate**	**Total turnover rate**	**Total churn rate**	**Male net migration rate**	**Male turnover rate**	**Male churn rate**	**Female net migration rate**	**Female turnover rate**	**Female churn rate**
A: Urban UK	-0.23	38.5	128.65	-0.35	40.25	130.4	-0.13	36.86	127
A1: Industrial Legacy	-0.95	37.58	110.5	-1.41	39.36	111.44	-0.52	35.91	109.62
A1a: Industrial Legacy	-0.95	37.58	110.5	-1.41	39.36	111.44	-0.52	35.91	109.62
A2: Established Urban Centres	-1.69	43.92	126.48	-1.85	46.06	128.11	-1.53	41.93	124.96
A2a: Struggling Urban Manufacturing	-2.78	52.03	123.74	-2.88	54.7	125.14	-2.68	49.53	122.42
A2b: Regional Centres	1.96	95.4	184.51	2.58	100.13	189.55	1.4	91	179.83
A2c: Multicultural England	-3.13	47.63	127.37	-3.6	49.09	127.99	-2.69	46.25	126.78
A2d: M8 Corridor	-0.26	35.34	105.99	-0.58	37.51	107.86	0.03	33.36	104.26
A3: Young and Vibrant Cities	3.38	84.92	180.47	3.7	87.99	184.19	3.07	82.01	176.94
A3a: Redeveloping Urban Centres	3.89	81.13	173.63	4.03	84.27	177.44	3.76	78.15	170.01
A3b: Young Multicultural	1.61	108.79	209.38	2.56	112.24	213.17	0.71	105.54	205.81
B: Rural UK	2.72	43.25	120.92	2.59	44.97	122.83	2.85	41.62	119.11
B1: Rural Britain	3.55	65.34	132.74	3.6	67.94	135.46	3.51	62.87	130.14
B1a: Rural Extremes	0.67	65.31	134.88	0.99	67.34	136.68	0.37	63.37	133.16
B1b: Agricultural Fringe	4.32	76.02	138.18	4.32	79.34	141.18	4.33	72.88	135.33
B1c: Rural Fringe	4.09	82.99	140.53	4.07	86.69	144.16	4.11	79.42	137.04
B2: Coastal Britain	6.59	61.82	137.73	6.9	67.32	141.57	7.04	62.2	136.55
B2a: Coastal Resorts	7.27	84.09	160.15	7.84	87.34	163.81	6.75	81.14	156.82
B2b: Aged Coastal Extremities	5.2	59.96	134.37	5.28	66.22	138.1	6.35	61.51	133.97
B2c: Aged Coastal Resorts	10.62	86.84	144.07	11.51	90.82	147.73	9.81	83.24	140.77
B3: Averageville	-0.25	59.77	124.28	-0.76	63.55	126.88	-0.16	59.32	123.59
B3a: Mixed Urban	-0.45	61.85	121.33	-1.17	66.34	124.21	-0.39	62.39	121.3
B3b: Typical Towns	0.09	67.87	135.17	-0.06	70.39	137.22	0.23	65.46	133.2
B4: Isles of Scilly	17.79	134.83	200.84	19.74	138.16	202.07	15.86	131.53	199.63
B4a: Isles of Scilly	17.79	134.83	200.84	19.74	138.16	202.07	15.86	131.53	199.63
C: Prosperous Britain	-0.52	66.56	138.67	-0.22	68.83	141.75	-0.81	64.38	135.7
C1: Prosperous Urbanites	0.58	101.67	169.7	1.1	105.6	174.08	0.09	97.9	165.51
C1a: Historic Cities	3.48	105.79	175.09	3.89	109.75	179.22	3.1	102	171.14
C1b: Thriving Outer London	-2.41	106.56	168.72	-1.77	110.99	173.62	-3.03	102.3	164.01
C2: Commuter Belt	-1.07	76.14	136.78	-0.87	78.94	139.92	-1.26	73.45	133.75
C2a: The Commuter Belt	-1.07	76.14	136.78	-0.87	78.94	139.92	-1.26	73.45	133.75
D: Urban London	-8.53	63.23	151.26	-8.17	65.15	153.93	-8.86	61.42	148.74
D1: Multicultural Outer London	-8.01	92.18	149.87	-7.51	94.5	152.13	-8.49	89.98	147.75
D1a: Multicultural Outer London	-8.01	92.18	149.87	-7.51	94.5	152.13	-8.49	89.98	147.75
D2: Mercantile Inner London	-10.21	147.24	225.15	-11.12	151.25	229.36	-9.36	143.54	221.25
D2a: Central London	-10.29	147.35	225.14	-11.31	151.4	229.42	-9.34	143.62	221.2
D2b: City of London	2.79	279.91	301.35	16.95	276.14	294.65	-13.45	284.22	309.03
D3: Cosmopolitan Inner London	-8.22	111.47	177.96	-7.31	114.94	182.16	-9.08	108.17	173.97
D3a: Afro-Caribbean Ethnic Borough	-6.07	127.47	190.17	-4.74	131.44	194.68	-7.33	123.74	185.93
D3b: Multicultural Inner London	-11.81	118.44	174.39	-11.23	119.49	173.9	-12.04	113.91	169.74

context of nearly every district type), but these movements are more balanced in either direction than those of females. In other words, the female population is being redistributed around the country through net migration at a rate that is faster than for males. One explanation for this may well be associated with the greater number of females than males in the student population.

Focusing on specific Groups and Classes, it becomes clear that (ignoring the Isles of Scilly and the City of London) Classes within Urban London, as well as those urban areas with large student populations and dynamic economies (Regional Centres, Young Multicultural, Redeveloping Urban Centres, Historic Cities) have the highest levels of population turnover and churn and correspondingly least stable populations. Groups and Classes within the Rural UK Family have relatively low levels of turnover and churn, with districts within the Industrial Legacy and M8 Corridor Classes having the lowest levels overall, thus signifying relatively stable populations.

4.5.2 Standardising for age

Age has a significant influence on migration behaviour, as shown in Figures 4.3, 4.5 and 4.6, with propensity to migrate reaching a peak at around the 20-24 age group. Thus far, however, turnover and churn statistics have not taken the effect of age into consideration. It is to be expected that the population compositions of districts and Families, Groups and Classes of district will differ. Urban London, for example, may have a younger age structure than Rural UK and this will necessarily affect the rates of migration associated with these areas.

To deal with these population composition effects it is possible to standardise turnover and churn rates by the age structure present in the total population. A method of 'direct standardisation' is proposed by Rowland (2006), and is used here to produce the standardised rates of turnover and churn shown in Table 4.6. Following the notation proposed by Rees et al. (2000) and taking the standardised turnover rate for area i, STO_i, as the example:

$$STO_i = \frac{\sum_a \left(TO_i^a P_+^a\right)}{\sum_a P_+^a} \tag{4.5}$$

where the turnover rate is calculated as in Equation (4.2) and $P_+^a = \sum_i P_i^a$ = the population in age group a in all zones in the system.

Calculating the age standardised rate of churn follows the same form so that

$$STCH_i = \frac{\sum_a \left(CH_i^a P_+^a\right)}{\sum_a P_+^a} \tag{4.6}$$

where churn is calcuated as in Equation (4.3).

This method of direct standardisation produces the set of age standardised rates shown in Table 4.7. For both turnover and churn, the rates have undergone two methods of standardisation. In the first method, the rates were standardised by the age groups contained in the original data. In the second, the rates were standardised by the 15 year age groups used in the majority of this analysis. These new rates allow for the effect of differing age structures between the Families. A comparison with rates in Table 4.6 indicates some age effects, although these are more pronounced with churn than turnover. Taking turnover first, the relationships between the families do not change. Highest rates of turnover are found in Prosperous Britain and Urban London, with the lowest rates found in Urban UK and Rural UK. The importance of Prosperous Britain and Rural UK increases slightly, whilst for Urban UK and Urban London, the importance decreases - in the case the latter by a relatively large margin. There is little difference between the two methods used to standardise turnover.

Table 4.7: Age standardised turnover and churn for the Vickers et al. classification Families, 2000-01

	Total Turnover rate		**Total Churn rate**	
	Standardised by original age groups	Standardised by 15 year age groups	Standardised by original age groups	Standardised by 15 year age groups
A: Urban UK	36.95	37.29	124.41	125.19
B: Rural UK	46.76	46.11	128.85	127.37
C: Prosperous Britain	67.61	67.38	140.16	139.90
D: Urban London	54.41	55.67	129.55	133.68

In contrast, age standardisation affects the churn rates more noticeably, especially for Urban London. Whereas the non-standardised rates for Urban London show the highest rates of churn, the standardised rate is less than that computed for Prosperous Britain. This suggests that much of the churning of the population in London could be down to a younger age structure. Similarly, whereas Rural UK experiences the lowest rates of churn with the non-standardised calculations, standardised rates are higher than for Urban UK, suggesting conversely that the lower degree of churning could be partially down to an older age structure.

Whilst standardising turnover and churn rates for age results in a small caveat concerning the interpretation of the rates in general, the standardisation does not dramatically alter turnover and churn statistics for district families in most cases. Where differences are more pronounced, as in the case of Urban London, it suggests that much of the population instability demonstrated here is down to the movement of young, labour-force age population. Whilst significant, this does not necessarily mitigate the importance of the characteristics of the Urban London Family in explaining migration patterns, but rather helps stress the interactions between age/life course and the environmental, social and economic characteristics of places. The maps shown in Figures 4.12, 4.13 and 4.14 reveal exactly how these general trends of turnover and churn present themselves spatially for all migrants and for the early working age (16-29) and late working age (45-PA) groups .

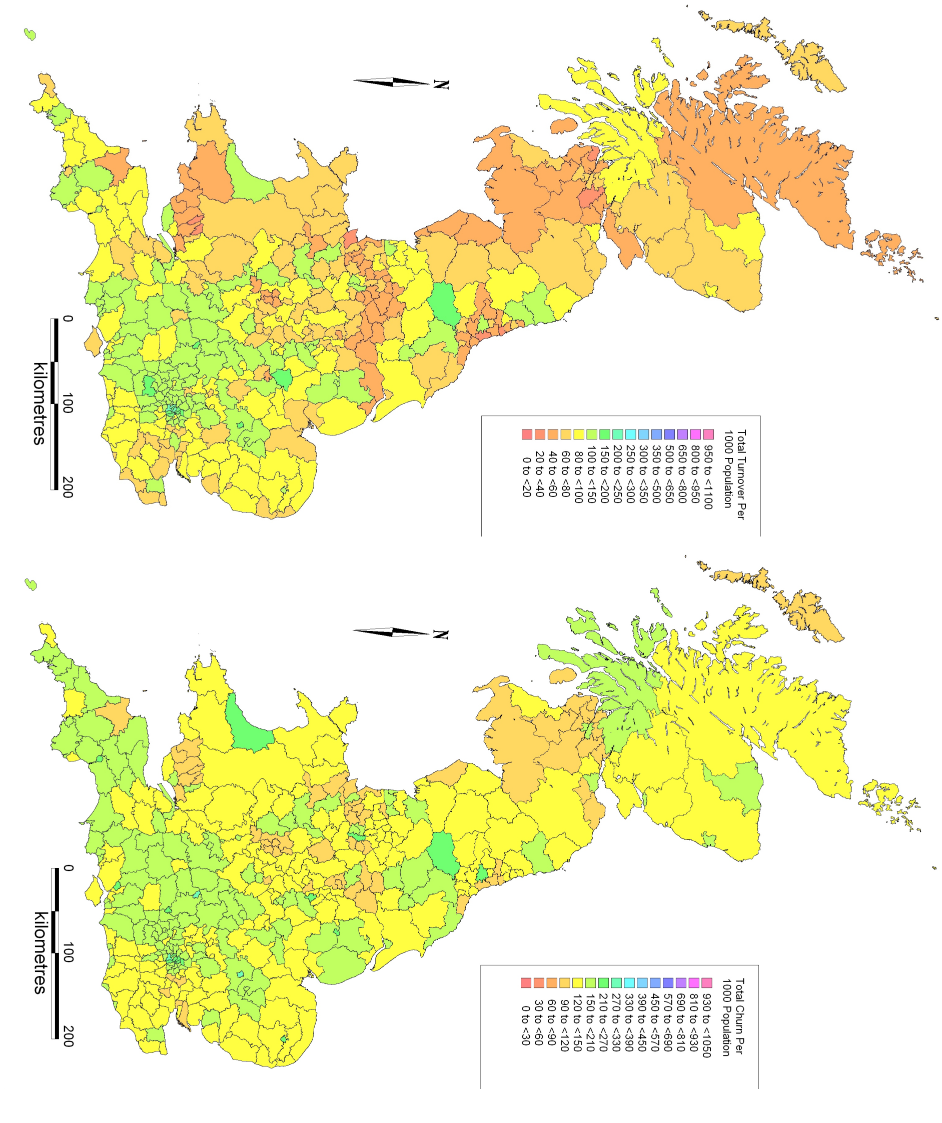
4.5.3 Patterns of turnover and churn by district

In terms of the aggregate patterns of turnover and churn, it is already apparent from studying the figures in relation to the district classification, that highest levels of turnover and churn are found in London and some of the more dynamic urban areas in Britain. It is also noticeable that other areas of relatively high turnover and churn tend to be more concentrated in the South East, around London, and towards the South West. This is evident on both the ‘all ages’ maps shown in Figure 4.12. It is also clear that areas defined as Industrial Legacy and M8 Corridor have particularly low levels of turnover and churn, and these maps show areas close to South Wales, Birmingham, Manchester, Liverpool, West and South Yorkshire, the North East and the M8 corridor between Glasgow and Edinburgh as fitting into the patterns shown in the classification analysis.

It is evident that the gap (in terms of numbers of people moving) between areas of lowest turnover and other higher turnover areas is proportionally greater than when looking at the same gap in churn statistics for the same areas. Put another way, areas of low turnover also have low churn, but the lower turnover is more noticeable than the lower churn when compared with other higher scoring areas. The range between lowest and highest is greater for turnover than churn. As churn takes into account intra-district migrations and turnover does not, it can be concluded that compared with districts in the rest of the country; when migrations occur in these low turnover/churn areas, they are more likely to be shorter-distance, intra-district migrations than longer-distance, inter-district migrations. This is confirmed by looking at the average relationship between turnover and churn for all districts (Table 4.8). On average, there are around 60 more people per 1,000 moving per district when intra-district migrations are taken account of. Districts where the figure is greater than average experience relatively more intra-district migration. The gap between turnover and churn for the majority of the low turnover districts (those in South Wales, the North East, the Industrial North) is greater than 60. In these areas of low turnover and churn, if migrations take place they are more likely to be local, within-district movements - movements less likely to have a perturbing effect on population stability.

The remaining maps in Figures 4.13 and 4.14 show the relative rates of turnover and churn for the early and working age groups, with (as expected) the highest rates in the 16-29 age group and much lower rates older age group (not shown). Two points can be made in summary. Firstly, whilst there are no immediately apparent spatial patterns in the 16-29 age group, closer inspection of the maps reveals the highest levels of turnover and churn are in the spatially diffuse but characteristically homogenous (as confirmed by the Vickers et al. classification) urban areas previously described - the dynamic, growing university towns and urban areas fringing London.

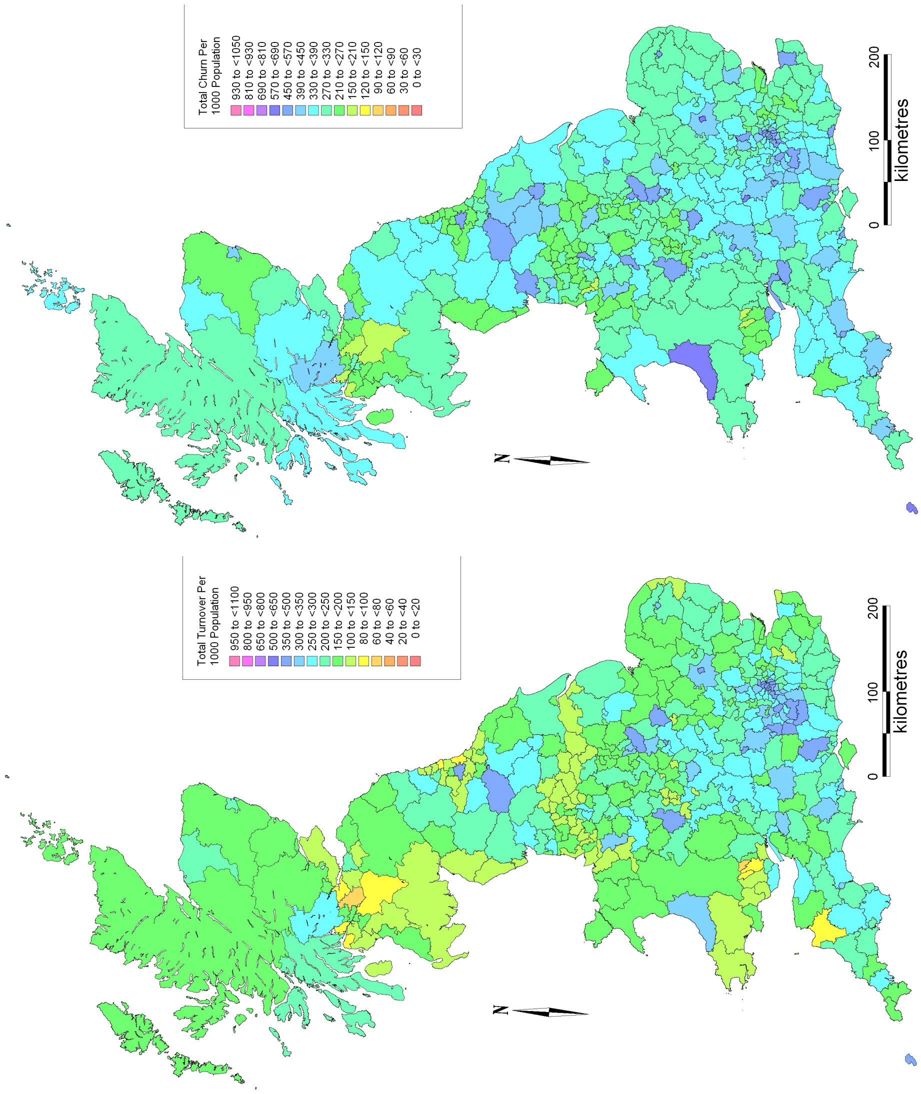
Secondly, there is more clarity in the spatial patterning related to higher levels of mobility in the groups with lowest overall levels of turnover and churn (the two oldest age groups - only 45 to Pensionable age is shown), with coastal areas - the South West in particular, showing higher levels of these measures of stability for these older groups. Low levels of turnover and churn



(a) Turnover rates per 1,000 people

(b) Churn rates per 1,000 people

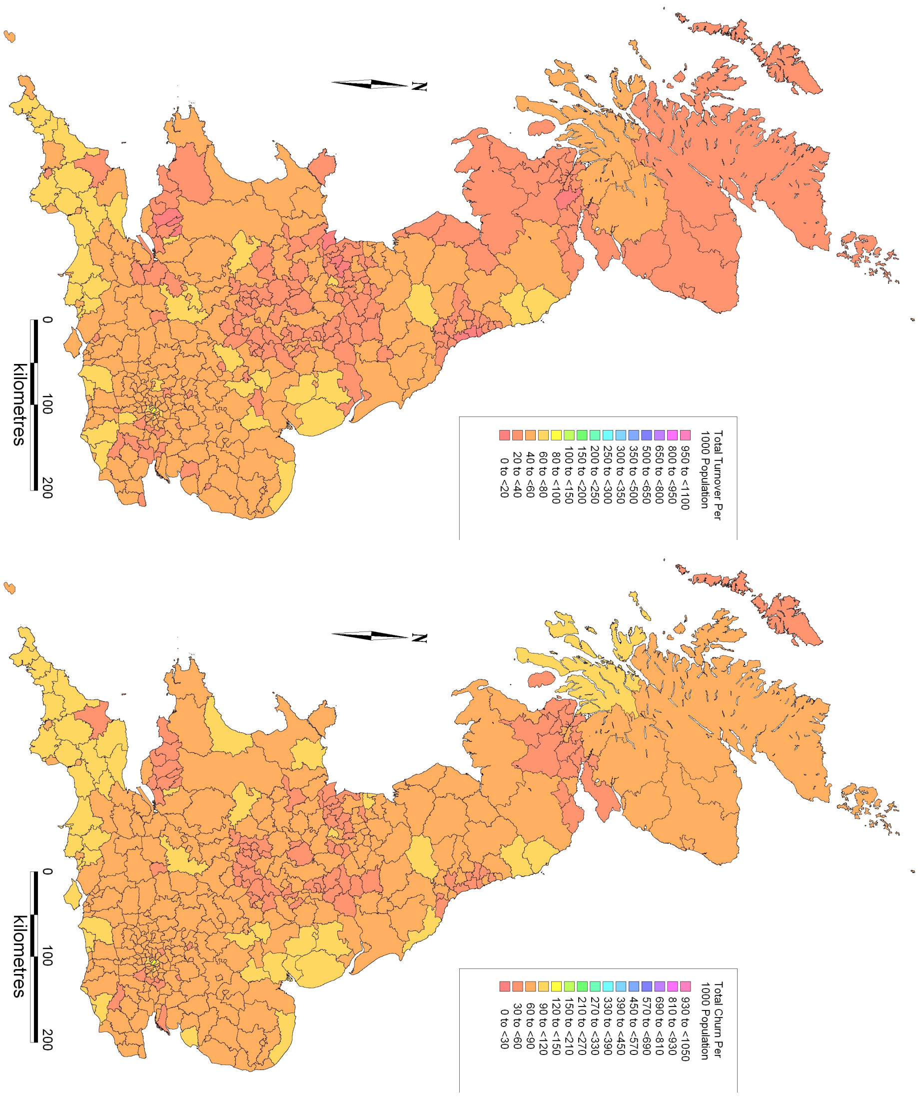
Figure 4.12: Turnover and churn rates by district, all ages, Great Britain, 2000-01



(a) Turnover rates per 1,000 people

(b) Churn rates per 1,000 people

Figure 4.13: Turnover and churn rates by district, 16-29 age group, Great Britain, 2000-01



(a) Turnover rates per 1,000 people

(b) Churn rates per 1,000 people

Figure 4.14: Turnover and churn rates by district, 45-PA age group, Great Britain, 2000-01

Table 4.8: Summary statistics for population turnover and churn for districts in Britain: by broad age group and by breakdown of the 16-29 age group, 2000-01

Summary Turnover statistics for individual GB districts	**Total**	**Male**	**Female**	**0-15**	**16-29**	**30-44**	**45-PA**	**PA+**	**16-17**	**M16-17**	**F16-17**	**18-19**	**M18-19**	**F18-19**	**20-24**	**M20-24**	**F20-24**	**25-29**	**M25-29**	**F25-29**
Max	279.91	276.14	284.22	221.14	613.71	299.64	120.27	67.46	615.38	666.67	545.45	818.18	750.00	1000.00	743.40	748.73	847.46	572.34	548.98	597.78
Min	31.65	32.85	30.51	29.46	75.62	34.53	14.48	10.86	22.92	22.43	18.62	56.42	37.39	77.38	87.91	79.22	90.10	76.59	77.11	76.05
Mean	88.21	91.54	85.02	70.17	216.58	97.03	43.85	31.12	63.74	64.40	62.53	253.66	229.76	277.55	274.64	262.10	287.35	210.68	218.21	203.30
Median	84.71	88.37	82.23	69.78	205.52	94.98	43.61	31.72	56.85	54.11	58.13	237.00	210.50	266.53	264.35	250.89	275.16	196.74	203.68	191.46
Range	248.26	243.29	253.71	191.67	538.09	265.12	105.79	56.59	592.47	644.24	526.83	761.76	712.61	922.62	655.49	669.51	757.36	495.76	471.87	521.72
Without Isles of Scilly/City of London																				
Max	215.88	224.78	213.08	221.14	459.85	211.44	97.25	57.23	291.70	420.51	210.62	641.47	616.65	666.67	638.26	748.73	689.14	432.98	462.14	436.51
Min	31.65	32.85	30.51	29.46	75.62	34.53	14.48	10.86	22.92	22.43	18.62	56.42	37.39	77.38	87.91	79.22	90.10	76.59	77.11	76.05
Mean	87.63	90.97	84.42	69.91	214.96	96.46	43.64	31.02	61.89	62.48	60.82	251.42	228.49	274.37	272.62	260.59	284.74	209.66	217.10	202.35
Median	84.65	88.35	81.97	69.78	205.41	94.94	43.58	31.61	56.81	53.91	58.00	236.49	209.33	266.36	263.23	250.50	275.08	196.49	203.50	190.98
Range	184.23	191.92	182.56	191.67	384.23	176.91	82.77	46.37	268.78	398.08	192.00	585.05	579.26	589.28	550.35	669.51	599.04	356.40	385.03	360.46
Summary churn statistics for individual GB districts																				
Max	301.35	302.33	309.03	298.87	638.57	326.05	141.43	88.05	615.38	666.67	545.45	909.09	750.00	1333.33	818.53	795.85	966.10	596.81	563.27	633.33
Min	97.26	96.85	97.14	100.96	181.75	103.84	42.23	31.96	62.21	57.15	65.37	136.71	96.99	180.48	206.89	173.23	240.32	192.68	183.80	201.67
Mean	148.08	150.90	145.37	141.39	334.40	164.08	72.25	54.93	116.05	109.49	122.55	348.93	300.12	399.52	423.74	391.67	456.56	333.38	341.52	325.42
Median	141.66	144.68	139.32	138.62	319.97	162.56	70.33	54.04	109.71	101.57	117.71	329.20	278.78	379.43	413.54	378.87	447.50	326.17	335.53	316.93
Range	204.09	205.48	211.88	197.91	456.82	222.21	99.19	56.08	553.17	609.52	480.08	772.38	653.01	1152.85	611.64	622.61	725.78	404.13	379.46	431.67
Without Isles of Scilly/City of London																				
Max	295.47	302.33	288.65	298.87	617.45	278.58	128.48	88.05	346.57	449.26	267.40	703.40	667.84	857.14	818.53	795.85	850.95	553.09	548.15	559.02
Min	97.26	96.85	97.14	100.96	181.75	103.84	42.23	31.96	62.21	57.15	65.37	136.71	96.99	180.48	206.89	173.23	240.32	192.68	183.80	201.67
Mean	147.57	150.42	144.83	141.24	333.00	163.54	72.05	54.84	114.46	107.79	121.14	346.87	299.02	396.12	422.10	390.43	454.44	332.54	340.68	324.57
Median	141.62	144.48	139.07	138.58	319.79	162.48	70.29	54.01	109.63	101.39	117.44	328.85	278.49	379.22	412.65	378.38	447.34	325.69	335.29	316.84
Range	198.21	205.48	191.50	197.91	435.70	174.74	86.25	56.08	284.36	392.11	202.03	566.69	570.86	676.66	611.64	622.61	610.63	360.41	364.35	357.36

are apparent for the ex-industrial areas identified in the aggregate analysis, for all age groups. These low levels are, however, particularly pronounced at the 'family-centred' age ranges (0-15 and 30-44 - not shown, although the patterns for these age groups are almost identical to the all ages patterns), less so with the other age ranges.

Table 4.8 provides summary statistics for the main age groups, as well as a more detailed breakdown for the 16-29 age group. Of note here (as with net migration) are the continually higher mean rates of population turnover and churn for females when compared to males for the post-16 age groups. This difference is especially pronounced when looking at the late teen and early twenties age ranges. Aside from this difference, the highest mean rates of turnover and churn for both males and females are found in the 20-24 year old age group, and are around 70 and 121 migrants per 1,000 people higher respectively than the averages for the 16-29 age group.

4.6 Conclusions

The analysis reported in this chapter has sought to provide an introduction to internal migration in Britain at the start of the 21st Century. New insights into population movements have been gleaned from the use of the Vickers et al. classification of districts as a framework for summarising migration flows and rates. Flows have been assessed, not just through familiar binary divisions such as urban and rural, north and south or London and the rest of Britain, but also through more detailed sub-divisions contained within the classification; an exercise which has created an enhanced understanding of migration behaviour.

The chapter has shown that firstly, in relation to net migration, some of what has been discovered in past studies of internal migration in Britain remains constant. London and urban areas in general are net losers of migrants, whereas rural areas are net gainers: at an aggregate level, the process of counterurbanisation appears to be continuing. Moving beyond this, however, through the use of the district classification, it has been possible to deconstruct these aggregate patterns. Young, Vibrant Cities (including major settlements such as Bristol, Canterbury, Cardiff, Derby, Durham, Exeter, Lancaster, Leeds, Lincoln, Plymouth, Portsmouth, Sheffield, Southampton, and Brighton and Hove); Regional centres (such as Manchester, Norwich, Nottingham and Edinburgh) and Historic Cities (Colchester, Warwick and York) are, in fact, all significant net gainers of population through migration in 2001. This certainly runs counter to the trend expected from the aggregate analysis. It has been demonstrated that these gains are principally from Rural Britain areas - areas which, whilst rural, are perhaps not as remotely rural as the Coastal Rural areas in the same Family and that these gains are from the younger sections of the population.

In a similar way that some urban areas are gaining population, it is the case that some rural areas are also losing. 'Averageville' in Rural UK, for example, is losing migrants. It could be argued that many areas in Averageville could be classed as urban rather than rural, as they tend to feature smaller towns surrounded by rural areas (thus supporting the counterurbanisation hypothesis), but this in itself exposes the limitations of broad generalisations such as 'Rural' which inevitably obscure important patterns. Further exemplification of this issue can be found with the definition of a London hinterland, described in the Vickers et al. classification as 'Commuter Belt'. Much of this area, including large swathes of the Home Counties and beyond, would normally be described as rural (see http://www.defra.gov.uk/rural/ruralstats/). However the net out-migration characteristics of this area certainly do not fall into the counterurbanising norm we have come to expect, with many moves between this Group and the Young and Vibrant Cities Group.

This chapter has further reinforced the importance of looking at age when studying migration. Net migration rates and spatial patterns all vary dramatically when age is taken into consideration. Age and stage in the life course affects rates, direction of flow and specific origins and destinations of internal migrants. Studying the linkages between areas through the flows of migrants in relation to broad ages and stages in the life cycle, it has been possible to construct some typical age-specific flows. Between the ages of 0 and 15 migrants tend to move down the urban hierarchy with their parents. If they originate in central London, they are very likely to migrate to outer London; if they originate in outer London, they are likely to move to surrounding Commuter Belt or rural areas, and if they are already in Commuter Belt, then the movement is likely to be into rural areas. Between the ages of 16 and 29, a migration back up the hierarchy from rural or commuter belt areas to university towns and cities is common, as is the move to the centre of London immediately after graduation. Between the ages of 30 and 44, the pattern is largely similar to that of the 0-15 age group, with a continuation of the movement down the urban hierarchy, especially towards coastal areas as migrants approach retirement and post-retirement.

The understanding of internal migration in Britain has been advanced further by the introduction of turnover and churn analysis. These two measures help to quantify the stability of a population in an area; the latter developing the concept of stability slightly further than the former by taking account of intra-zonal as well as inter-zonal flows. The definition of stability used here is based upon the proportion of residents living at an address in an area who remain there from one year to the next. Generally speaking, the areas with the least stable or most transient populations are urban areas. The most stable or least changing populations are found in rural and previously industrial areas. Levels of stability vary greatly between different age groups with the 16-29 age group (and more specifically within that group the 18-19 and 20-24 age ranges) being the least stable, and the older age groups being inherently more stable. Within each age group though there are specific areas within Britain that have more or less stable sections of these populations.

Analysis of turnover and churn statistics for Britain at the Vickers et al. classification Family level has shown that whilst London retains the importance it has when net migration balances and rates are examined, the role of rural Britain becomes less important, even when the distorting effect of differential age profiles across the district classifications is taken into consideration. The vast size of the underlying rural population in Britain means that despite the apparently large net in-migration to rural areas, the disturbing effect that this has on the resident population is relatively low. Rural areas tend to have, on the whole, much more stable populations than areas in Prosperous Britain, though the naming of Prosperous Britain may in this case be misleading. Commuter Belt makes up much of Prosperous Britain, and can in many ways be regarded as rural. There appears to be continuation of the trend recognised by Rees and Boden (2006) of Urban London populations with the relatively footloose Urban London migratory characteristics, occupying space that would otherwise be re recognised as rural. In effect, it could be argued that there is an identifiable two-tier 'rural' in Britain. The 'traditional rural' Britain, with a generally stable population, perhaps experiencing some in-migration from more classically urban areas, and the 'new rural' Britain, which may outwardly exhibit many of the same environmentally rural characteristics as the traditional rural Britain, but that features this 'prospering' population with some of the migratory characteristics associated with the population of Urban London. To complete the narrative at this Family level, turnover and churn statistics have helped reinforce the idea of Urban UK (which should perhaps more accurately be described as 'ex- or declining industrial urban Britain') as being an area where populations are more stable; migrations, if they do happen, tend to be short distance; longer-distance migrations are more rare.

This chapter set out to fill a gap in the knowledge of internal migration in Britain at the start of the 21^{st} century, something which to a large extent has now been achieved. It is the case, though, that a piece of cross-sectional analysis from the census, no matter how detailed, can only present the migration situation over a one year period. For a more complete understanding and to contextualise this cross sectional analysis, a time-series analysis should be carried out. As was outlined in Chapter 2, time-series data does exist from the PRDS - and indeed a new partially estimated national dataset has been created - and so this is where the attention of this thesis will be turned in due course. However, before this is possible a framework for this subsequent analysis needs to be established. It has been demonstrated in this chapter that the three-tier district classification has proved a useful tool in the interpretation of internal migration patterns. But the general purpose classification adopted here has significant limitations in categorising areas with similar migration characteristics as internal migration variables were not included in the suite of variables used to define the clusters in the classification. Indeed it has been noted by Duke-Williams (2009b) that just because migrants can move into and out of areas which have been classified by particular demographic variables, it does not necessarily follow that those migrants will share the characteristics of settled individuals in either the origin or the destination. It follows, then, that a useful avenue of further research would be in the construction of an alternative classification based on a range of migration variables. Such a classification would enable areas with similar migrant characteristics to be identified, and would be extremely useful as a framework for monitoring internal migration flows. In addition, the creation of such a classification would present an alternative method for analysing internal migration in Britain; giving valuable insights into the importance of particular elements of internal migration in different areas in Britain and thus building still further on the work presented in this chapter. The next two chapters, therefore, will be concerned with developing such a classification.

Chapter 5

The case for a migration classification

5.1 Introduction

The preceding chapter has demonstrated the utility of a general purpose area classification in the study of migration. As already discussed, the Vickers et al. (2003) classification adopted does not include any specific migration variables in the clusters of variables used to define the different Families, Groups and Classes present in the hierarchy. Whilst internal migrants will be included amongst the groups of individuals present in each area, their status as such is not explicitly recorded and consequently does not directly influence the clusters represented in the classification. It has been demonstrated that by examining migration flows between areas defined by their socio-demographic characteristics, associations can be made between the flows and the area types. For example, at an aggregate national level, outflows occur frequently from urban area types characterised by poorer health, higher unemployment and lower economic activity. Inflows, on the other hand, tend to be to rural area types characterised (amongst other things according to the classification used) by lower population densities and higher home ownership rates. The argument might continue that by excluding migration variables from the general purpose classification, it is possible to assess the associations between the flows of migrants and the other socio-demographic characteristics of origins and destinations, and that any associations between particular flows and particular area types will be independent of the influence of migration variables already present in the classification. It could be said that observing high in-migration to an area partially defined and characterised by a high proportion of young in-migrants would not tell us a great deal, whereas high in-migration to an area partially defined and characterised by low population densities might tell us something about the aspirations of migrants.

Of course, just because an object - or area - can be classified in one way, it does not mean it resides exclusively within that classification typology. Objects can be classified very differently depending upon the purpose of the classification. For example, a tree might be classified by a biologist as a particular species within a certain genus; by an architect as a source of one type of building material with specific qualities for construction; or by a person sheltering from the rain as a more or less effective shelter than a nearby building. The tree's position in each classification fits a specific purpose. In much the same way, if an area is classified for one particular purpose it could be classified entirely differently for a different purpose - the same area could be classified as a low crime area, or, as an area with a high proportion of elderly residents. In the context of this work, it may well be that in an effort to understand migration between and within defined geographical areas in Britain, the use of a general purpose area classification constructed independently of interaction variables might not be most appropriate. Populations of migrants moving to and from areas may well differ significantly from underlying populations; as such when studying migration flows it may be sensible to classify areas according to the migrants experiencing the flow events and their associated socio-demographic characteristics. This then would counter the argument presented a moment ago that examining flows between areas not defined by migrant characteristics is better for migration analysis. Where migrants may not characterise the underlying population of an area, then studying migration in the context of potentially un-related population characteristics is more likely to in fact confuse the analysis and hinder understanding.

Therefore the aim of this chapter is to make the case for the development of a new, specific purpose, migration classification for Britain. Section 5.2 will discuss the rationale behind classification in general and the history of classifications in geographic research, demonstrating their utility in the analysis of spatial data, with sub-section 5.2.3 examining the important problem of whether to base a classification around the migrant individual or the migration flow; the distinction has already been explained in Chapter 2 and it is in classification where the distinction presents some important choices to be made. After this in Section 5.3 the discussion will move onto why classification is important, both as an end in itself, but also as an integral part of the research process. Finally Section 5.4 will look to explore some of the issues that are presented and considerations which need to be made which are of specific importance for a migration classification, before recommendations are made for proceeding further.

5.2 Why develop classifications?

So if an area classification based specifically around migration variables is an alternative to other more general classifications, one question that might arise is 'should such a classification be created?' Or perhaps taking the question even further - why create a classification at all?

5.2.1 Background to classifications

Taking the latter, it could be argued that human brains constantly classify our lived experiences in order for us to make sense of what is happening around us. Indeed, speaking from a biological perspective, Crowson(2006, p.1) states that *"classifying things is perhaps the most fundamental and characteristic activity of the human mind"*. Through classification we are reducing the amount of data the brain has to deal with, thus aiding us to make sense of situations more readily. Taking an evolutionary perspective, it is understandably an advantage for any animal to mentally classify food and non-food items, or categorise other animals as dangerous or benign. But because classification is useful and indeed necessary for some fairly fundamental life processes, does this necessarily mean it is applicable outside of the sphere of everyday lived experiences?

This question can be answered through examining where else the process of classification has flourished. The introductions to a number of textbooks on the subject (the existence of which already suggest a wider applicability) such as those by Aldenderfer and Blashfield (1984), Kaufman and Rousseeuw (2005), Everitt et al. (2001) and Gordon (1999) all mention the long history of the creation of classifications and taxonomies in fields such as biology, chemistry, physics and astronomy, as well as in the social sciences. When looking at the historical uses of classifications, it becomes clear that whilst they are frequently put to use in a variety of situations, the motivations behind classification creation can differ from those supporting the more common practice of every-day mental classification and the basic desire to sort objects to aid comprehension. Of course, aiding comprehension is one very useful end product, but once comprehension is improved, we are then more able to take what is known and apply it to alternative situations. For example in medicine, Everitt et al. (2001) describe the classification of diseases as both a useful aid to treatment, but also as a basis for research into the causes of disease.

To view the end result of the classification process - the taxonomic groupings - as the only benefit of creating a classification would be to ignore the value in the process which needs to be followed to arrive at this final product. The identification of particular data features which define the groups within the classification may prove even more useful than the classification itself, despite this perhaps not being the reason for embarking upon the classification development process in the first instance. Indeed, it was partially through the process of cataloguing and classification of new species that Charles Darwin began to develop the ideas which led to the publication of arguably one of the most influential works of all time: 'On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life' (Darwin, 1859), containing ideas on the theory of evolution by natural selection - ideas which changed the world as we see it and the received wisdom on the origin and evolution of our entire existence.

5.2.2 Classifications in Geography

Whilst perhaps not as fundamentally significant as the ideas which evolved from Darwin's classification exercise, there has been a long history and development of area-based classifications in geographically related disciplines which have sought to make sense of complex environments and the various population attributes characterising those environments. In all cases, these classifications have stimulated further research. Vickers et al. (2005) cite the work of Charles Booth in the nineteenth century as perhaps the earliest such example. Booth attempted to map and classify areas of London according to the socio-economic characteristics (specifically poverty, employment and religion) of the residents living in those areas - work which influenced both subsequent academic studies (the work of Orford et al. 2002 being one of the more recent examples) and more applied political policy (Bales, 1999). Burgess (1925) in the early twentieth century, whilst focusing on the growth and expansion of Chicago also succeeded in classifying the 'types of areas differentiated in the process of expansion' - area types identified in part by their residents. This seminal work influenced later work on urban structure and classification by authors such as Hoyt (1939) and Harris and Ullman (1945). The work of both Booth and Burgess, whilst having contemporary influence, can also be seen as the forerunner of far more recent work which, whilst perhaps more detailed in its scope, complex in its methodology and arguably more accurate in its definition, actually seeks to do exactly the same thing - to classify areas according to set of particular key characteristics.

A case could be made for one of the main drivers behind much of the recent work on area classifications being commercial interest, which has led to the growth of an industry concerned with developing and applying area classifications for commercial gain. A commonly used term, for both the development and application of these (small) area classifications (as well as the industry stewarding it) is *'geodemographics'*. There is a large literature documenting the development of geodemographics - a development which has largely occurred in parallel with improvements in computational and processing power, software and geographic information. Batey and Brown (1995) and Yano (2001) provide succinct historical overviews. The commercial imperative has helped spawn companies and organisations such as CACI Ltd, CCN marketing (now Experian) and EuroDirect, all producing their own geodemographic area classifications such as A Classification of Residential Neighbourhoods (ACORN) (http://www.caci.co.uk), MOSAIC (http://www.experian.co.uk) and CAMEO (http://www.callcreditmarketing.com) respectively for commercial customers. The continuing growth of the industry (a brief visit to the press release section of any of CACI, EuroDirect or Experian's websites will present a selection of news stories documenting new updates of their classifications and expansions into different countries) might be evidence enough that there is real value in classifying areas according to certain key characteristics, tailored for specific needs and purposes. Other evidence, however, can be found in Harris (2005) where an objective evaluation of whether geodemographic classifications 'work' is carried out though a case study of the application of the ACORN classification. ACORN is used to assess differences in product consumption patterns in a British town with a conclusion that, for this particular application, the classification did indeed work when tested on the ground.

Further evidence as to the utility of geodemographic area classifications outside of the commercial sphere can be found in the renewed academic interest in the creation of area classifications and the ongoing development of area classifications by the Government. Longley (2005) postulates that this revival has been driven through a combination of a desire for evidence-based policy from local government, improvements in data and related infrastructures and a need for setting service delivery targets at a local level. Examples of the former include some early work by Openshaw and Blake (1996) on a 'GB profiler' using 1991 Census data. More recently, a large number of geodemographic profiling projects have been undertaken by the Centre for Advanced Spatial Analysis (CASA) at University College London, including those focusing on health, ethnicity, education, and awareness and access to digital technologies (http://www.spatial-literacy.org/), each with applications in policy and resource targeting. Bespoke classifications have also been created within the School of Geography at the University of Leeds for specific local purposes. Work by Shepherd (2006) profiling neighbourhoods to aid community safety and Debenham (2003) focusing on supply-side variables to extend commercial geodemographic classifications within Yorkshire and the Humber are such examples.

National classifications have been constructed for the major Census geographies by ONS and others: output areas, super output areas/data zones, wards, health authorities and local authorities, (http://www.statistics.gov.uk/about/methodology_by_theme/area_classification/) as well as indices created from multiple variables for other geographical areas such as the Index of Multiple Deprivation . Whilst the Index of Multiple Deprivation does not seek to cluster areas with similar characteristics at the outset as most geodemographic classifications do, by ranking each area by its index score and then dividing the ranked areas into proportions (quintiles or deciles) and allocating areas to these groups, an area classification comparable to those created through clustering methodologies can be created. Rees et al. (2002a) provide a comprehensive historical summary of census-based area classification typologies. In addition, some of the growing range of applications to which the ONS Output Area Classification (OAC) is being put are documented by the OAC User Group (http://www.areaclassification.org.uk/). These include higher education student profiling, analysis of crime and antisocial behaviour, analysis of transport need, various commercial applications and local authority housing market analysis - the breadth of applications no-doubt indicative of the open-source nature of both the classification and its construction methodologies.

Within geographical analysis there is another family of classification techniques concerned less with classifying single areas, but more with identifying the relationships between areas and classifying them by these relationships. These are known as 'functional regionalisation' techniques. Whilst standard geodemographic classifications can classify a number of areas based upon their similar attributes, there is no inference that because areas may fall into the same category they share any kind of connection or interaction. Openshaw (1989) picks up upon this drawback, and, using the example of interaction data in the form of credit card company information on sales (destination) and the residential address (origin) of the customer, he highlights the potential these data have for identifying the catchment areas of shopping centres. By defining catchment areas an indication of the importance of key nodes to surrounding zones is presented.

In identifying the inability of standard geodemographic classifications to deal with interactions such as those which could define catchment areas, Openshaw (1989) highlights functional regionalisation methods as a potential solution. Coombes (2000, p.1502) defines functional regionalisation as a *"form of area classification within which each class is normally a single group of contiguous areas"*. Brown and Holmes(1979, quoted in Feldman et al. 2006) suggest that functional regions are *"areas or locational entities which have more interaction or connection with each other than with outside areas"*. So it is the contiguous nature coupled with the attribute homogeneity of the smaller areas within each class area that sets functional region classifications apart from geodemographic area classifications. Assessing catchment areas for flows is a particularly geographical or spatial problem and functional region classifications deal with grouping common flows rather well.

The functional regionalisation approach is one that has been adopted widely in the analysis of commuting data, principally because it has long been recognised that the poor definition of geographical areas can lead to statistics giving a distorted view of the reality underlying them (Coombes, 2002). The need to create a set of geographical areas relatively consistent with the phenomena being examined in those areas has been recognised more recently by Martin (1998, 2000, 2002) in relation to the creation of relatively socially homogenous census OAs, but was identified by the UK Government in relation to locally specific unemployment rates and the allocation of financial assistance to those areas in greatest need (Coombes et al., 1986) a number of decades ago. The result of this need was that Coombes and others, on behalf of the Department of Employment developed and successively refined a set of Travel To Work Areas (TTWAs). TTWAs were designed such that they reflected labour market areas within which the local supply and demand of labour interact (Coombes, 2002). With commuter flows inextricably linked to labour supply and demand, the origin-destination elements of commuting data can be used to identify labour demand nodes surrounded by labour supply areas such that the boundary of each Travel To Work Area (TTWA) surrounds an area that is relatively self contained in terms of its commuter flows. From 1981 census data and then with each successive wave of the census, Coombes and colleagues (2000; 2002; 1986) have developed new sets of TTWAs using variations on a functional regionalisation algorithm which essentially identifies employment nodes (or foci) through functions of job ratio and residence-based self containment, before amalgamating adjacent foci where they were strongly linked, and then iteratively allocating residual non-foci areas to the foci with which there is the heaviest commuting association. A very similar methodology was used with migration data by Coombes et al. (2004) to create a set of 'Housing Market Areas' for housing policy developments. Here areas of relative in-migration self containment were defined.

5.2.3 A migration flow or migrant based classification?

At this early stage, before a full discussion of the benefits of developing a classification is had, an important decision needs to be made: is this to be a true *interaction data* classification - a classification which concerns itself with both origins and destinations and the interaction flows between them - or is this to be a classification based on *migration variables* - a classification which is concerned more with single areas and the characteristics of migrants associated with those areas? The answer to this question will guide the rest of the discussion and the rest of this chapter. Choosing the former will mean that the work is more likely to continue down the route of functional region creation, whereas the latter will mean that the work will follow the route of geodemographic classification.

Whilst identifying a migration-based functional region makes some sense in that the majority of moves are over shorter distances and therefore identifiable housing market localities could be constructed (Coombes et al., 2004), the zones in the functional region system will represent spatial flow cluster localities rather than clusters based on the attributes of individual migrants. One of the drawbacks, in this context, is that areas with similar origin/destination flows will be in close geographic proximity so the localities tend to be sets of contiguous areas. An area in Scotland would almost certainly not be grouped with an area in London, for example, even if the types of migrant flowing into and out of these areas are similar.

One of the benefits of geodemographic classifications is that although areas in the classification may not be in close geographic proximity, they might share similar characteristics meaning they can be classified similarly. As was demonstrated in Chapter 4, interesting features of the internal migration landscape of Britain could be viewed when similar areas in different locations were grouped together, for example it is possible to extrapolate wider trends such as the flows up and down the urban hierarchy, or the flows from rural areas to Young and Vibrant Cities, at different life stages - such analysis would not be as straightforward with functional regions. It should also be noted that opting for a geodemographic classification in preference to a functional region classification does not preclude the use of flow data. As will be seen in Chapter 6, it is entirely possible to construct variables in relation to the distance of flow, and use these in a geodemographic classification.

So at this point the discussion will continue having chosen the path towards geodemographic classification, but having taken this route much discussion still needs to be had surrounding the rationale behind creating a migration-based geodemographic classification, as well as the pitfalls which might confront such a task.

5.3 Why create a migration data based classification?

The preceding discussion has illustrated why the idea of classification has been appealing and has looked at how the creation of different types of classification (both geodemographic and functional region) for a variety of purposes can be beneficial, both as an initial aid to understanding by reducing the amount of information we need to process and understand in order to appreciate phenomena, but also as a foundation for subsequent analysis or research through the use of the taxonomy directly and also through the by-products of the data clustering process - the key variables which help define the groups and clusters within the classification. A classification, whilst useful in its own right is often only the starting point for additional exploration. Having answered the question as to why create classifications at all, attention must now be turned to the slightly more difficult question of 'why create a classification based upon migration variables?' At least part of this question has been answered in the previous discussion, and in the conclusion to Chapter 4. It is worth spending some time though exploring fully the benefits which could be gained from developing an internal migration classification.

5.3.1 Classification to aid understanding

Dealing first with what has already to an extent been answered; creating a classification (or indeed *classifications*) based upon migration variables will aid the understanding of what are inherently more complex data than the standard counts of people residing in places displaying particular attributes (which comprise the bulk of the data in most social surveys). When examining standard census or other social survey data, the counts will relate to a defined geographical area. Migration data on the other hand, relate to both an origin and a destination or numerous origins and destinations. The two can be seen to be connected through the flow of individuals between these locations. Taking permanent migrants as an example, these individuals residing in an area will of course display many of the same attributes as the non-migrant population: they will be male or female; of a certain age, socio-economic category or ethnicity, etc. In addition, however, they will have a number of attributes associated with them which separate them from the non-migrant population: whether they have moved in, out or within the current area; whether they have moved short or long distances; if they have moved into the area, whether it is from an overseas origin; whether they have moved as part of a household or moving group or as an individual migrant. It is these unique and complex features of migrants that mean areas hosting them can be classified separately from existing classifications. It may be very useful from a policy perspective to know, for example, if an area is particularly prone to receiving relatively high numbers of elderly in-migrants or losing high numbers of skilled workers (human capital). Furthermore, as was mentioned in the conclusion to Chapter 4 it may also be that migrant populations either leaving or moving into areas are not representative of the underlying resident populations, something which could certainly influence policy decisions - a poor inner-city area with a large transient student population may require different resource targeting to an inner city area with a sedentary young population. It is also the case that while some areas are very popular origins or destinations for migrants, there are also areas which are very isolated. There are some areas which will not send or receive very many migrants and consequently will have distinctive characteristics of their own.

5.3.2 Classification as part of the research process

Both the classification itself and the development of a classification can be seen as part of the wider research process. The nature of this research will of course vary, but some key themes can be identified: the concept of change is an important one to consider both as it will provide avenues for this research, as well as obstacles to overcome. It could be argued that due to the fluid nature of populations in most areas (an average of around 10% of the population across the UK lived at a different address in the year preceding the 2001 Census), the moment any data are recorded at a given time-point, the further away from that time point we move, the less likely it is that those data remain relevant. Of course this is precisely why there is a continuing programme of collection with most social surveys and considerable interest in longitudinal analysis across the waves presented by these surveys. This is also why organisations such as CACI and Experian are keen to publicise their ACORN and MOSAIC geodemographic classifications as 'latest' versions and why the ONS have released new sets of area classifications in tandem with census results since 1961 (Rees et al., 2002a).

What a classification with its roots in one specific time period does do, however, is allow for the exploration of change over time. A classification based upon migration variables will be tied inextricably to the time period associated with the collection of those variables. Whilst this means that, potentially, the further away from that time period we move, the less relevant the classification will be; testing a hypothesis along those lines should reveal both information about change over time as well as the extent of the change. Of course it may also be that the underlying structures which define migration in Britain - such as the interregional structures defined by Raymer and colleagues (2007) which demonstrate a certain stability in the origins and destinations of migration flows, might also be applicable at finer geographical scales. In this case, it is the absence of change that is telling. Either way, the classification could be employed very effectively as a framework for monitoring spatial interactions over a period of time.

A classification in one time period also allows for the opportunity for a similar classification to be developed at a later date using a similar methodology and data from later waves of the same sources; the differences and similarities between the two revealing the extent of any change. For example, if a classification is constructed using data from the 2001 census, come the advent of the 2011 Census, a similar classification could be created and compared. Comparison will reveal areas which are more or less susceptible to change, and in doing so will give clues as to the temporal validity of other area classifications. Research by Orford et al. (2002) already cited reveals that in the case of the early classification of London developed by Charles Booth in the nineteenth century, there were significant similarities between areas of poverty and affluence in Victorian London and areas of poverty and affluence today. Indeed, Orford et al. demonstrate that Victorian socio-economic conditions are a strong predictor of present day mortality.

Research stemming from the development of a migration data classification need not be limited to uses directly related to the classification itself. The process of analysing and clustering data to create a classification will produce a set of significant variables (Vickers et al., 2005) which might be put to subsequent, alternative uses. These variables only become apparent through the classification building process: As Everitt and Dunn (2001) relate, a classification will consist of a small number of homogenous groups or clusters. The Vickers et al. (2003) area classification already discussed consists of a number of districts grouped according to the similar characteristics of the individuals residing within, however, whilst the classification groups were created using 2001 Census data variables, all available variables from the census were not used; rather a selection of as limited a number of variables as possible were chosen. These selected variables represent the main dimensions of the parent data source and were chosen to have as much variation as possible across the whole spatial system whilst also showing as little correlation with each other as possible (Vickers et al., 2003). A series of analyses were needed to decide which variables should be chosen to summarise the whole dataset. Principal Components Analysis (PCA) was used to initially to identify which variables drive the dataset. That is, because principal components explain the majority of the variance in any particular data matrix (Kline, 1994), variables which comprise more of any one principal component can be seen to be more important. In addition, correlation matrices were used to exclude highly correlated variables and standard deviation statistics were employed to select variables which varied more across the range of districts in the spatial system. By using these techniques together, an initial large set of variables was reduced to a smaller set of significant variables which could be used to classify any particular area.

Through adopting a similar process of variable selection for a migration data classification, important variables which reveal the most about population flows will be exposed. It is likely that some variables will be distributed over space far more evenly than others, with some showing greater spatial concentration. For example, male and female migrants are likely to be distributed widely and relatively evenly, whereas migrants of certain ages are more likely to be concentrated in particular areas - university towns for those in their late teens and early twenties or coastal resorts for those at retirement age. Where particular variables exhibit much greater spatial variation, they could be used to explore what is determining, maintaining and/or changing population flow patterns.

5.4 Considerations for a migration classification

There are a number of questions that need to be answered before the process of developing a classification can begin. Perhaps the two most important questions, closely related, are: what should be the scale of analysis and which data should be included? The two are linked as data will be available for discrete geographical areas, the choice of scale affecting both the availability of variables as well as the application of the classification and vice versa.

5.4.1 Scale and interaction data

All interaction data and associated migration variables available in the UK are available for discrete geographical units (Dennett et al., 2007). Within the UK there is a complex hierarchy and linkage of geographies, depicted in Figure 5.1, in which all small area geographies eventually aggregate up to the coarsest country level. A problem for some spatial analyses is that not all lower level geographies are compatible, for example, electoral wards aggregate into both districts and parliamentary constituencies, but these two geographies cannot be harmonised. This is an issue for the creation of any national classification where potential data which could be used from disparate sources are produced for different geographies.

As discussed in Chapter 2, different datasets have different levels of geographical detail. Of relevance to the interaction data and associated migration variables for single areas and as has been discussed previously in Section 4.2, are issues of national geographical compatibility, particularly with census data. An unrivalled choice of variables at relatively fine spatial scales means that the 2001 Census will be the most useful source of data for classification purposes, but an issue with some census interaction data is that geographical compatibility problems can occur between the constituent countries of the UK at particular spatial scales. As noted in Section 4.2, in Northern Ireland, data at level 1 are available for parliamentary constituencies rather than district council areas. Therefore any UK classification created from census interaction data could not be compared completely with other district level classifications. It is probable, therefore, that if census data are chosen for a migration classification, as with the analysis in Chapter 4 the classification will be for Britain rather than the whole UK. But aside from spatial data compatibility, looming over the issues of scale and data are the more conceptual issues of the Modifiable Areal Unit Problem (MAUP) and the related problem of the ecological fallacy.

5.4.2 The MAUP and the ecological fallacy

Organising any data into discrete areal units presents a set of problems. The MAUP is one such issue, and one that has been providing problems for spatial analysts since it was first identified by Gehlke and Biehl (1934). As outlined by a number of authors including Openshaw (1984), Wrigley et al. (1996), Green and Flowerdew (1992) and Geddes and Flowerdew (2004), the modifiable areal unit problem can create difficulties when analysing aggregate data for discrete geographical areas. O'Sullivan and Unwin (2002, p.30) describe the problem thus:

> *"aggregation units used are arbitrary with respect to the phenomena under investigation, yet the aggregation units used will affect statistics determined on the basis of data reported in this way."*

The problem is duel faceted: the first relates to scale, the second to zoning. Taking the former, patterns identified in that data at one scale of aggregation may not present themselves at a different level of aggregation. Exemplifying the scale problem in internal migration analysis, Gober-Meyers (1978) examines the influence of different socio-economic, demographic and environmental factors on interregional population movement in the U.S. between 1965 and 1970, at the state and metropolitan/non-metropolitan within-state levels. She demonstrates through analysis of the 1970 U.S. Census, that certain factors (such as fertility and unemployment) can be seen to influence migration at the metropolitan/non-metropolitan level, but not at the state level. Gober-Meyers stops short of offering a solution to this problem, but advises careful consideration of the scale factor when carrying out migration research. For a more detailed

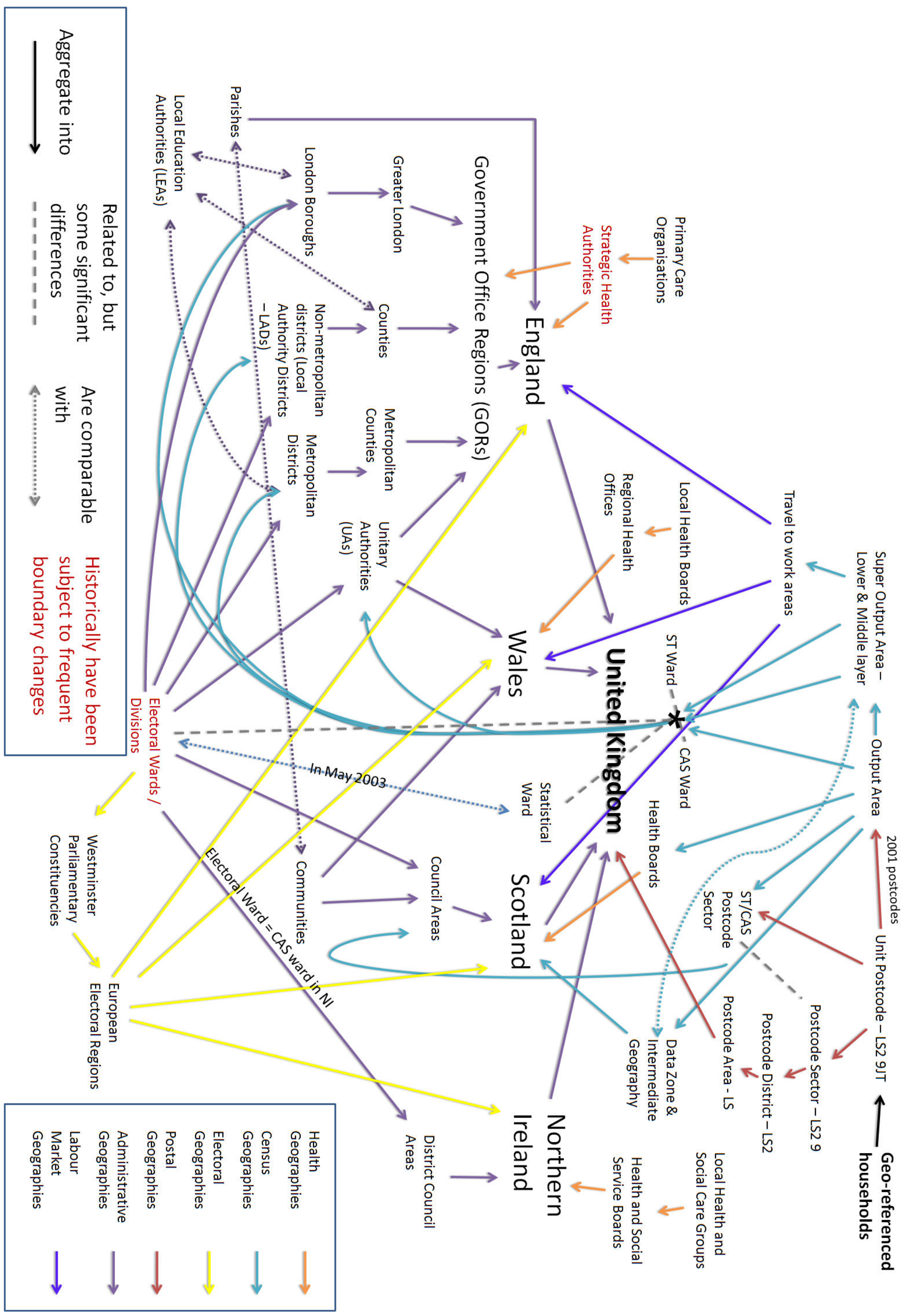


Figure 5.1: A schematic representation of the hierarchy and connectivity of UK geographies, 2008.

explanation of the problem and its effects on spatial data in general, see Openshaw and Taylor (1979).

The ecological fallacy has some commonality with the MAUP, although is a slightly different problem. Where the two are similar, as pointed out by O'Sullivan and Unwin (2002), is that both issues make it apparent that statistical relationships can change at different levels of aggregation. The ecological fallacy emerges from the practice of 'ecological inference,' described by King et al. (2004) in the preface to their book as the *"reconstructing of individual behaviour from group-level data."* That is to say that the ecological fallacy is the problem of inferring something at a lower level of aggregation, from something observed at a higher level. We make ecological inferences commonly in everyday life - perhaps when deciding upon a holiday destination, because a particular country has a reputation for good beaches, and inferring (rightly or wrongly) that because a particular resort lies within that country, it too will have good beaches. The ecological inferences also form the basis of many governmental decisions - the recent ban on smoking in public buildings was in part an effort to reduce the numbers of smokers, and this was largely due to compelling aggregate evidence that cases of heart disease and lung cancer are more prevalent among patients who smoke. Of course at an individual level, there are always exceptions to this general rule and one cannot say that *all* smokers will develop heart or lung problems. It is understandable, however, that ecological inferences are made in this context, as it would be impossible to tailor manageable policy to individual needs.

5.4.3 Additional issues of scale

Bringing the discussion back to the issue of scale and data choice, it is inevitable that any decision has the potential to create problems and these will need to be acknowledged. The question is will any particular scale of analysis create any more or less problems? Harris (2005) asserts that all users of geodemographic typologies will need to contend with issues of representation. Whereas classifications describe areas, some users will try to infer the characteristics of individuals from these areas and it is inevitable that general classifications will not be entirely representative of the whole population. This is perhaps more of an issue where geodemographic classifications are constructed from micro-data and apply to small areas - for example output areas or unit postcodes. The temptation is to assume that as the level of areal aggregation is reduced, the likelihood of generalisations being accurate increases - indeed Farr and Webber (2001) state that analyses have shown data at the level of the person discriminate better than more aggregate data. Of course, even at the household level, generalisations can be inaccurate. The key to the utility of the classification lies in the purpose for which the typology was created and the use to which it is eventually put. If an area classification is created at the district level (for example, the ONS classification of local authorities) and is designed to summarise districts in terms of their key characteristics, then providing the methodology is sound, the classification should be fit for that purpose. Problems will only start to arise if assumptions are made about the population residing within any given classification area below (or indeed above) the level of that described by the classification. Just because the district features high proportions of elderly residents, it does not mean that all areas within the district will also feature high proportions of elderly residents. It sounds a very obvious point to make, but this type of assumption is what lies at the heart of the ecological fallacy. Returning to the original question, however, it is unlikely that any particular level of analysis will create any more or less problems for the user; issues associated with the MAUP and the ecological fallacy will propagate at any level of analysis - the extent to which they will matter will depend upon the final use to which the classification is put.

5.4.4 Decisions on a migration classification

So having discussed the case for a migration classification and explored some of the issues which are presented, it remains that some initial decisions need to be made before it is possible to proceed any further. Firstly, a decision on the spatial system, geographic scale and principal data source needs to be made. Each of these are inter-dependent; a decision on one affecting a decision on another, therefore all three should be addressed together. Following the conclusions of Chapter 4 which pointed to one of the principal uses of the classification being a framework for the analysis of PRDS data, it would be sensible to choose the LAD scale over any other. This decision is further justified when considering the data source, as of the data available at LAD scale, the census offers the largest range of variables - a range which diminishes at lower level geographies. It is desirable to include as many different variables as possible in order that a comprehensive classification is constructed, therefore data from the 2001 Census at LAD scale will be used. These two decisions make the final one on the spatial system an easy one: for reasons already discussed in detail at the beginning of Chapter 4, where 2001 SMS data at level 1 are being used, the inclusion of Northern Ireland is precluded. Therefore, in keeping with the analysis in Chapter 4, this will be a migration classification for Britain rather than the whole UK.

Up until this point reference has only been made to the construction of an internal migration classification - this would seem logical given the aim of the classification is to help enhance the understanding of internal migration in Britain, however, should other interaction data also be considered? Certainly the work of Eliasson et al. (2003) has made links between commuting data and internal migration, and the work of Stillwell and Duke-Williams (2005) has shown links between international migration and internal migration, especially in the context of London. At this stage, the inclusion of commuting data can be discounted. A big argument for excluding commuting data, relates to the attribute data associated with commuting flows. Essentially in the 2001 Census Special Workplace Statistics (SWS) data, the only variables which are different from internal migration data relate to method of travel to work; all other variables such as age and ethnicity will be the same or comparable. Disentangelling some migration events from longer distance, longer time-frame commuting events could be tricky

(Frost and Dennett, 2010), but with the vast majority of commuting moves being short distance and with a similarly high proportion of migrants also likely to be commuters, it could be argued that little is gained from including variables related to commuters in a classificiation where the commuting flows themselves have already been discounted. Where the commuting flows are the key difference between migration and commuting data and it has already been shown that flows are more effectively analysed through functional regionalisation techniques, then at this stage the decision has to be made to focus solely on migration data.

A less clear-cut case can be made for the exclusion of international migration data at this point - international migrants will certainly be differentiated from internal migrants by their origin, with flows of immigrants potentially affecting the flows of some internal migrants; therefore the use of international migration data in a migration classification will not be discounted completely just yet. But, as will be seen in the next chapter, the particular spatial profile of international migration in Britain means that there are problems which could also preclude its use in a final migration classification.

5.5 Concluding Remarks

At the beginning of this chapter, the contention was made that a general purpose classification may not provide the best framework for the analysis of migration flows as general purpose classifications define themselves principally from non-migrant population stocks. Where migrants might have different characteristics to the settled population in an area, then it would make sense to classify these areas separately by their migrant characteristics for the purpose of analysing migration.

The case for the development of a migration classification was made both philosophically, in that it was argued that classification satisfies a natural human instinct towards parsimony and the desire for a more ready appreciation of complex phenomena, and practically in that classifications can be a key part of the research process. Both of these justifications are relevant for this thesis: a more meaningful parsimony can be achieved though reducing complex flows between multiple origins and destiniations to flows between far fewer but relevant, in the context of migration, origin and destination types - an important step in fully understanding migration in Britain. And through creating a classification and during the building process, identifying important characteristics contained within migration data - another important step towards exactly the same end.

The development of a migration classification will provide a new framework which could be used for the analysis of trends over time. These trends could be assessed in an overtly cross-sectional way through comparing a classification created with data available now, with a similar classification created from future census data, but they could also be more time-series through using the classification as a scheme for framing year-on-year data such as that provided by the PRDS. Of course the former will not be possible for this piece of work, but the latter certainly will.

It was discussed early on in the chapter whether a migration classification, in the context of this research, should be concerned with classifying migration flows or migrants and their characteristics; the former likely to result in the classification being a set of functional regions, the latter leading the research down the route of geodemographic classification. It was decided that a geodemographic classification would be the best choice for a number of reasons, the most important being that for a clearer apprieciation of the internal migration landscape of Britain, it would be important to be able to identify non-contiguous zones which exhibit similar migrant characteristics.

In the latter half of the chapter, the various arguments surrounding the scale of analysis and the data to be used in the classification were addressed. Addressing scale first, it was identified that of key importance to the development of a useful typology is the availability of variables. The finer grain the spatial scale, the fewer variables are available for analysis in all candidate interaction datasets. Where one particular scale of analysis will not generate any more or less problems than another, then a slightly coarser scale with more associated migration variables was seen as preferable. With scale and data intrinsically linked, attention was turned to data sources and data types. It was decided that census data would be the best source, with district level data offering the most potential migrant variables for selection. The selection of LADs as the spatial unit of choice for the classification was further influenced by the availability of inter-censal PRDS data at this scale. The final chapters in this thesis will be concerned with analysing these data using the new classification framework, so it would be unwise to pursue a classification at any other geographical level at this stage. The inclusion of commuting data was rejected, despite some association with migration data, on the grounds that very little would be offered in terms of new variables. In addition an early decision not to study flows means that much of what would set commuting data appart from migration data would now not be important to the classification. The decision was taken not to completely discount international migration data at this stage due to the potentially interesting associations between international migration and internal migration - if it is to be discounted, it should only be done so after a far more thorough analysis. This will take place in the next chapter.

So this chapter set out with the aim of making a case for the development of a migration classification. A thorough discussion has taken place, with careful justification of a number of arguments for the development of a new migration data-based geodemographic classification. Local authority districts have been chosen as the level of analysis and so with the census being the source of data, then the classification will be designed to cover Britain rather than the whole of the UK due to the issues of geographical harmonisation with Northern Ireland. The classification, both through the process of building and through the use of the final typology will enhance the understanding of migration in Britain, but for the classification to be successful, care must been taken to ensure the right decisions are made at each stage of the construction process. Chapter 6 will document this process in detail before presenting a series of results in the form of a new migration classification typology.

Chapter 6

Developing a migration classification

> *"If the process of clustering is likened to an animal then it is a very peculiar beast! It has the front legs of automation but the back legs of user intervention; eyes for data-led classification but the ears of a priori expectation; it feeds on a variety of data sources but generally prefers a census; displays a patchwork coat mixing the qualitative and the quantitative, the objective and the subjective; and is born of a cross-breed between art and science!"* Harris (2005)

6.1 Introduction

The previous chapter made the case for the development of a migration classification - this chapter will detail the construction of this new classification, and will proceed with a series of aims. The first is to arrive at a new migration classification typology which can be used as a framework for further analysis of migration flows in Britain - analysis which will take place in Chapters 7 and 8. Another is that through the classification building process, much more is learnt about the migration flows and migrant characteristics which help define British internal migration landscape in 2001. Chapter 4 made some progress into understanding the types of area that are prominent to a greater or lesser degree in the internal migration system, but the definition of new clusters of LADs defined just by their migrant characteristics should, to an even greater extent, differentiate those areas which experience one type of migrantion experience from another. Questions like: 'which are the areas that lose young migrants?' 'Are there areas which attract migrants of differing socio-economic status?' 'Are some areas largly excluded from the internal migration system?' 'Are there any associations between these migrant attributes in particular areas?' And 'how are these areas with similar migrant characteristics distributed across space?' are all ones which could be addressed through the construction of this new classification. Certainly whilst the first three, it could be argued, could be answered without the use of a classification, the associations between particular migrant attributes across space can most effectively be observed through the development of a new area typology. It is hoped that through the definition of a new set of distinct clusters it will be possible to make associations and differentiations between the profiles of LADs which are different from any other made by existing geodemographic classifications.

Building a geodemographic classification requires a number of stages to be completed, with careful decisions made at each point in the process, therefore in Section 6.2 of this chapter, a trial district level classification will be developed using a tried-and-tested methodology; this trial classification will be used as the foundation for a final evolution, arrived at only after a full evaluation of variables, methods and initial results. Section 6.2 will adopt a seven stage procedure to build this experimental classification, with considerable attention devoted to the identification, selection and testing of variables to be included. Section 6.3 will present these initial results before Section 6.4 evaluates the methods and inputs of the trial classification, refining each where necessary in order to produce a robust final solution. In Section 6.5, the *'Migration Classification'*, as it will be known, will be presented along with descriptive portraits of each cluster in the classification in Section 6.6. Finally in Section 6.7 the new Migration Classification will be compared with other geodemographic classifications to assess the extent to which it offers a totally new typology and therefore a new tool for the analysis of internal migration in Britain.

6.2 An initial district level area classification based upon migration variables

As is noted by Založnik (2006, p.10), geodemographic classifications *"invariably produce plausible results"*. That is to say whatever data are input into a clustering procedure, the resulting outputs can often be interpreted in a way that can make intuitive sense. As Založnik points out, this is both a great strength and great weakness of the process. How then can one be sure that the classification output from a clustering procedure is 'optimum' - i.e. most accurately reflects the key patterns in the underlying data? The answer is probably 'never', as with any generalisation, detail will be lost that some may argue is important. In practice though, it is possible to create a more robust classification though careful decision making at each step of the process. This presupposes that the process can be theorised as a series of steps, which indeed it can. A number of authors have considered the process of designing and creating a classification and a general framework for this process which has been suggested several times (Everitt et al., 2001; Shepherd, 2006; Vickers, 2006) is that proposed by Milligan and Cooper (1987) and Milligan (1996). It consists of seven sequential steps which organise the clustering process from start to finish. Creating a classification from the beginning, it would seem appropriate, therefore, to adopt Milligan's approach. The steps outlined below from Section 6.2.1 to Section 6.2.7 are those suggested by Milligan. As with any piece of work, it is unlikely that the first 'draft' will be the same as the final evolution. An initial draft of the classification will be created here and reviewed. Where improvements to the initial methodology and decisions can be made,

Table 6.1: 2001 SMS tables

Table Reference	Table Name	Cells/variables within table
Table 1	Age by sex	75
Table 2	Family status of migrant	54
Table 3	Ethnic group by sex (GB destinations)	24
Table 3n	Ethnic group by sex (Northern Ireland destinations)	9
Table 4	Whether suffering limiting long term illness by whether in household by sex by age	84
Table 5	Economic activity by sex	42
Table 6	Moving groups	16
Table 7	Moving groups by tenure	32
Table 8	Moving groups by economic activity by sex	336
Table 9	Moving groups by NS-SEC of group reference person	288
Table 10	Migrants in Scotland/Wales/Northern Ireland with some knowledge of Gaelic/Welsh/Irish	36

these will be discussed and implemented in Section 6.4.

6.2.1 Objects to cluster

As stipulated at the end of the last chapter, local authorities have been selected as the areas to cluster within the whole spatial system. The whole system is Britain rather than the UK, so the objects to cluster will be the 408 LADs of England, Wales and Scotland.

6.2.2 Variables to be used

In Chapter 2, reference was made to sources of data other than the census being less attribute rich and sampling far fewer individuals, and in the last chapter this was flagged as an issue for classification building. Consequently the classification taxonomy in this initial classification will be developed solely from 2001 Census migration data. A summary of the data tables available from the SMS is given in Table 6.1.

Throughout the literature warnings abound that choosing appropriate variables is very important, if not key, to the success of the final classification produced. Whilst the use of statistical techniques can certainly help whittle down the choice of variables systematically, (as will be shown later on), Openshaw and Wymer (1995, p.244) suggest that *"[t]here is no statistical technique that is a good substitute for thinking about choice of variable, yet!"* Certainly in the case of a migration-based classification, careful thought should be given to whether particular variables are likely to influence migration events or patterns. With this in mind it is useful to assess groups of variables as to their suitability for inclusion.

Age and Sex As has been shown in Chapter 4 and in a number of other pieces of work (Bates and Bracken, 1982; Dennett and Stillwell, 2009; Raymer et al., 2007, 2006; Rogers and Castro, 1981; Rogers et al., 2002), age has a significant influence on the propensity to migrate, as well as the direction and volume of migration, with very large numbers of migrants in their late teens and early twenties gravitating towards larger conurbations; migrants in the family rearing ages moving out of cities into rural areas; and post-retirement migrants moving to coastal areas (Uren and Goldring, 2008). Therefore the inclusion of age variables is of great importance to any migration-based classification.

The case for the inclusion of sex variables is less clear-cut. Evidence from past analysis (Champion, 2005) has tended to indicate that there is little difference between the migration patterns of males and females. However, as has been demonstrated elsewhere, (Dennett and Stillwell, 2010) there are some variations by sex, especially at different ages. The propensity for females to migrate at the age of peak migration (late teens and early twenties) may warrant the inclusion of sex variables in a migration classification.

Family Status Cooke (2008) provides a comprehensive review of research which has been carried out on the many complex family-based influences which can affect migration flows, from marriage to family formation to divorce. The influence of family status can influence both the motivations for moving and the moves themselves (Geist and McManus, 2008), and can interact with other influencing factors. For example, work by Boyle et al. (1999) and Cooke and Bailey (1999) has made the link between the differing employment status of female migrants who move either alone or as a part of a family. Certainly, therefore, a case can be made for the importance of including family status variables in a migration-based classification as origins and destination particularly favoured by migrants moving in families may have labour market implications. Furthermore, as Castro and Rogers (1981 p.vii) note *"many internal migrations are undertaken by individuals whose moves are dependent on those of others"*. It may well be that the origins and destinations of group or family movers are markedly different from those who move independently of others.

Ethnic Group The particular patterns of migrants of different ethnicities within the UK have been the focus of a number of recent pieces of work (Faggian et al., 2006; Finney and Simpson, 2008, 2009; Owen, 1997; Raymer and Giulietti, 2009; Raymer et al., 2008; Simpson and Finney, 2009; Stillwell et al., 2008). All of this work suggests there are differences in the migration propensities between ethnic groups. It may be that in some cases the patterns are confounded by other variables such as age and socio-economic status, although despite this, with concentration of non-white groups predominantly in urban areas, particularly cities, the identification of areas where ethnic minority migrants are more commonly moving in or out will be useful in developing a clearer migration picture for Britain.

Limiting Long-term Illness Research carried out by Norman et al. (2005) has focused on the health of migrants and the implications for the origins and destinations associated with healthy or less healthy migrants. Norman et al. (2005) discovered that whilst (amongst the young) migrants are generally healthier than non-migrants, in less deprived areas migrants are healthier than non migrants, but in more deprived areas migrants are less healthy than non-migrants. They also found that healthier migrants move away from deprived areas, increasing the rates of ill health and mortality in these areas, and interestingly a significant number of unhealthy migrants move into more deprived areas, exacerbating this increase in ill health and mortality rates still further. With this in mind, the inclusion of variables related to limiting long-term illness may certainly highlight areas, which, if characterised by flows of unhealthy migrants could flag important changes in the concentrations of ill health.

Economic Activity Much has been written on the influence of economic activity on direction and volume of migrant flows. From the work of Ravenstein (1889, 1885) well over one hundred years ago which observed the pull of urban areas for rural workers, to more recent work by Fielding (1992) and Findlay et al. (2009) which characterises the south east of Britain as an 'escalator region' for economic migrants, the influence of employment availability on migration flows has been well documented. Whilst the economic condition of origins or destinations may influence the flows of migrants, the economic condition of migrants themselves may also be influential. Work by Bohara and Krieg (1998) in the United States provides evidence of a linkage between levels of income and the propensity to migrate. Dixon (2003) recounts a similar story in the UK, showing through time-series analysis that those in the highest socio-economic groups are far more likely to migrate between regions than those who are less educated and employed in less skilled jobs, with Bheim and Taylor (2002) making an alternative observation of a strong link between unemployment and migration propensity. If economic reasons are the influencing factor for a very large number of migration events, then while examining whether migrants are employed may not tell us a great deal as there are large differences between the earnings of those employed at the bottom of the socio-economic scale and those at the top, examining those who are not employed may tell us more. Furthermore, additional categories of economic activity such as 'retired' or 'student' are likely to present distinct flows for certain areas.

Housing Tenure Links between the housing market/housing tenure and migration events have been observed before. Boyle (1998) notes that whilst those living in council housing are more likely to move than owner occupiers, these moves are likely to be over shorter distances - longer distance moves being constrained by administrative barriers. Other work has shown that housing availability influences flows of owner-occupant migrants (Cameron et al., 2005; Murphy et al., 2006), with new private housing influencing in-flows (Boyle, 1998), and Clark and Huang (2004) linking the distance of migration moves with tenure in the UK. With housing tenure also being a proxy for affluence as well as an indication of the potential ease at which individuals can move, the inclusion of tenure related variables in a migration-based classification can be justified.

Socio-economic Status As outlined by Champion et al. (2007), migration, historically, has been a selective process with (in the case of counterurbanisation) predominantly wealthy people moving out from the cities to the suburbs, or from cities to rural areas. Whilst international migration has frequently involved some more disadvantaged individuals, recent major internal migration flows in Britain have tended to involve those slightly older migrants who have been able to afford to move (the counterurbanisers) and those younger skilled migrants who have been attracted to urban agglomerations perhaps by higher education opportunities and who have then remained, or who have been tempted to move between larger urban areas (particularly in London) in search of tertiary sector jobs with higher salaries.

Champion et al. (2007) suggest that higher skilled migrants are tending to migrate over much longer distances than their lower skilled counterparts, redressing labour supply and demand imbalances in different locations - findings which echo the early work of Sjaastad (1962). With socio-economic status influencing the direction, volume and distance of migration it is evident that the inclusion of such variables in a migration classification will be important.

So it is clear from previous research that a case can be made for the inclusion of at least some variables from all of the main tables. Therefore, selected for initial inclusion in the classification were data from SMS Tables 1, 2, 3, 4, 5, 7 and 9. Table 3n was not included as it only applies to Northern Ireland and is thus irrelevant for this classification. Table 6 was not included as the information in this table was also contained in other tables. Table 10 was not included as it does not apply to the whole of Britain. The eight tables selected, therefore, cover as far as possible the dimensions of the whole dataset; an approach advocated in the methodology adopted for other area classifications (Bailey et al., 2000; Vickers et al., 2003). From these selected tables, a suite of variables to be considered for inclusion in the classification was created. The tables contain a total of 599 count variables, however unlike a standard area classification where the variable relates to a static individual residing in that place, for every district in a migration-based classification, each variable needs to be further defined by its movement component.

For example for any one area in a standard classification there could be a count of individuals of perhaps of a particular age and sex. These individuals could be divided by a total population in that area and be represented as a proportion. When we add an interaction component, the area could be either a destination for in-migrants, or an origin for out-migrants, or indeed both for within-area migrants. Straight away, just by identifying migrants as in, out or within district migrants we have created three times as many variables from our original count. Where these counts can be divided by populations to create in, out and within-area migration rates (Section 2.3.2) this number is doubled again.

In addition to standard rates and counts associated with internal migration, there are a number of additional counts and rates that can be attached to most variables for all areas. These may include immigration from abroad of in-migration from the 'no usual address' category in the census (which comprise a considerable number of migrants). Distance measures such as the mean, maximum or minimum in- and out-migration distance travelled could also be included. If the 14 indices outlined by Bell et al. (2002) alone were calculated for each variable in Table 6.1 then it is possible to see that very quickly the number of variables which could potentially be included in a classification becomes huge, numbering in the many thousands.

This poses a number of practical problems. It is highly unlikely that all of these variables would be relevant for the classification. Examining the numbers of variables included in district level classifications created by Vickers et al. (2003) and ONS (2004), it becomes apparent that considerably less than many thousands of variables are needed to develop a useful area typology. The Vickers et al. district level classification uses 56 variables, whereas the equivalent ONS classification uses only 42, therefore it is imperative that a systematic process of reducing the huge amount of potential variables is employed. Indeed, Aldenderfer and Blashfield (1984) warn that there is a temptation to include as many variables as possible in an analysis and hope that cluster analysis techniques will produce meaningful output, but doing so could cause problems. They state that *"the importance of using theory to guide the choices of variables should not be underestimated"* (Aldenderfer and Blashfield, 1984, p.20) as cluster analysis is beset with unsolved problems and is, ultimately, still a heuristic technique. Whilst there is debate within the literature about whether in the classification process the use of more variables is better with Harris (2005) advocating a general approach of including as many variables as possible, there does appear to be consensus from a number of writers Everitt et al. (2001); Shepherd (2006); Vickers (2006) based on the ideas of Milligan (1996) that variables should only be included if there is a good reason to think they will define the clusters and that *"irrelevant or masking variables should be excluded if possible"* (Everitt et al., 2001, p.179).

Candidates for irrelevant or masking variables would be those which are highly correlated. For example, two highly correlated variables could be 'age group 30-44' and 'economically active' as the majority of 30-44 year olds will also be economically active. The inclusion of highly correlated variables serves to effectively positively weight the underlying common factor. Aldenderfer and Blashfield (1984, p.21) point out that *"if three highly correlated variables are used, the effect is the same as using only one variable that has a weight three times greater than any other variable"*. Such weighting would skew the results of any analysis, therefore it is desirable, where possible, to only include variables that are not highly correlated (Shepherd, 2006; Vickers, 2006).

Whilst it is desirable to drop one variable from a pair of highly correlated variables to reduce the size of the dataset, doing so is not necessarily straightforward, especially when presented with a symmetrical correlation matrix with many thousands of variables on each axis. One approach to data reduction could be 'top-down'. That is, to start with all variables and successively reduce the numbers through a systematic and logical process, perhaps starting where easily identifiable divisions in the data occur. For example, with this approach, a logical first step could be to discard all non-rate data (as differing area size is likely to bias the classification towards larger areas with more flows) and then to discard either the male, female or total individual elements of a variable depending on the correlation coefficient. In many cases, males and females are likely to be highly correlated, so where they are, only the combined male and female counts should be used. Once the data has been cropped substantially in this way, further data reduction through the elimination of other highly correlated variables can take place.

Although the top-down approach to data reduction is logical, it requires the analysis of very large correlation matrices and can be extremely time consuming. An alternative approach is to start from the 'bottom-up'; starting with no variables, successively adding those viewed as most likely to produce a useful classification before changing and adapting the variables selected as necessary. This approach fits with the recommendations of Aldenderfer and Blashfield (1984) to use theory to guide the choice of variables. It is also a much faster process than the top-down approach, dealing with only a few variables from the beginning. Table 6.2 shows the initial collection of 88 variables chosen for inclusion in the classification:

Table 6.2: Initial variables chosen for inclusion in the trial classification

	Variable
1	Internal in-migration rate of persons aged 0 to 15
2	Internal in-migration rate of persons aged 16 to 29
3	Internal in-migration rate of persons aged 30 to 44
4	Internal in-migration rate of persons aged 45 to 59
5	Internal in-migration rate of persons aged over 60
6	Internal out-migration rate of persons aged 0 to 15
7	Internal out-migration rate of persons aged 16 to 29
8	Internal out-migration rate of persons aged 30 to 44
9	Internal out-migration rate of persons aged 45 to 59
10	Internal out-migration rate of persons aged over 60
11	Internal within-area migration rate of persons aged 0 to 15
12	Internal within-area migration rate of persons aged 16 to 29
13	Internal within-area migration rate of persons aged 30 to 44
14	Internal within-area migration rate of persons aged 45 to 59
15	Internal within-area migration rate of persons aged over 60
16	International immigration rate of persons aged 0 to 15
17	International immigration rate of persons aged 16 to 29
18	International immigration rate of persons aged 30 to 44
19	International immigration rate of persons aged 45 to 59
20	International immigration rate of persons aged over 60
21	In-migration rate from no previous address of persons aged 0 to 15
22	In-migration rate from no previous address of persons aged 16 to 29
23	In-migration rate from no previous address of persons aged 30 to 44
24	In-migration rate from no previous address of persons aged 45 to 49
25	In-migration rate from no previous address of persons aged over 60
26	Internal in-migration rate of whites
27	Internal in-migration rate of non-whites
28	Internal out-migration rate of non-whites
29	Internal out-migration rate of whites
30	Internal within-area migration rate of non-whites
31	Internal within-area migration rate of whites
32	International immigration rate of non-whites
33	International immigration rate of whites
34	In-migration rate from no previous address of non-whites

35	In-migration rate from no previous address of whites
36	In-migration rate of economically active individuals
37	In-migration rate of economically inactive individuals
38	Out-migration rate of economically active individuals
39	Out-migration rate of economically inactive individuals
40	Within-area migration rate of economically active individuals
41	Within-area migration rate of economically inactive individuals
42	International immigration rate of economically active individuals
43	International immigration rate of economically inactive individuals
44	In-migration rate from no previous address of economically active individuals
45	In-migration rate from no previous address of economically inactive individuals
46	In-migration rate of individuals with a limiting long term illness
47	In-migration rate of individuals with no limiting long term illness
48	Out-migration rate of individuals with a limiting long term illness
49	Out-migration rate of individuals with no limiting long term illness
50	Within-area migration rate of individuals with a limiting long term illness
51	Within-area migration rate of individuals with no limiting long term illness
52	International immigration rate of individuals with a limiting long term illness
53	International immigration rate of individuals with no limiting long term illness
54	In-migration rate from no previous address of individuals with a limiting long term illness
55	In-migration rate from no previous address of individuals with no limiting long term illness
56	Migration efficiency of wholly moving households whose household reference person is in NS-SEC category 1.1
57	Migration efficiency of other moving groups whose household reference person is in NS-SEC category 1.1
58	Migration efficiency of wholly moving households whose household reference person is in NS-SEC category 1.2
59	Migration efficiency of other moving groups whose household reference person is in NS-SEC category 1.2
60	Migration efficiency of wholly moving households whose household reference person is in NS-SEC category 2
61	Migration efficiency of other moving groups whose household reference person is in NS-SEC category 2
62	Migration efficiency of wholly moving households whose household reference person is in NS-SEC category 3
63	Migration efficiency of other moving groups whose household reference person is in NS-SEC category 3
64	Migration efficiency of wholly moving households whose household reference person is in NS-SEC category 4
65	Migration efficiency of other moving groups whose household reference person is in NS-SEC category 4
66	Migration efficiency of wholly moving households whose household reference person is in NS-SEC category 5
67	Migration efficiency of other moving groups whose household reference person is in NS-SEC category 5
68	Migration efficiency of wholly moving households whose household reference person is in NS-SEC category 6
69	Migration efficiency of other moving groups whose household reference person is in NS-SEC category 6
70	Migration efficiency of wholly moving households whose household reference person is in NS-SEC category 7
71	Migration efficiency of other moving groups whose household reference person is in NS-SEC category 7
72	Migration efficiency of wholly moving households whose household reference person is in NS-SEC category 8
73	Migration efficiency of other moving groups whose household reference person is in NS-SEC category 8
74	Migration efficiency of wholly moving households whose household reference person is in NS-SEC category Full Time Student
75	Migration efficiency of other moving groups whose household reference person is in NS-SEC category Full Time Student
76	Migration efficiency of wholly moving households whose household reference person is in NS-SEC category Not Classified
77	Migration efficiency of other moving groups whose household reference person is in NS-SEC category Not Classified
78	Migration efficiency of wholly moving households moving into owner occupied accommodation
79	Migration efficiency of other moving groups moving into owner occupied accommodation
80	Migration efficiency of wholly moving households moving into socially rented accommodation
81	Migration efficiency of other moving groups moving into socially rented accommodation
82	Migration efficiency of wholly moving households moving into privately rented accommodation
83	Migration efficiency of other moving groups moving into privately rented accommodation
84	Migration efficiency of individuals living alone
85	Migration efficiency of individuals not living in a family but with others in a household
86	Migration efficiency of individuals who are part of a couple family
87	Migration efficiency of individuals who are part of a lone parent family

88	Migration efficiency of individuals living in a communal establishment

Vickers et al. (2003) recommend that for an area classification to be comprehensive, all domains within the dataset need to be included. Here the initial set of 88 variables cover all domains available in the SMS, with age (variables 1-25), ethnicity (26-35), economic activity (36-45), health (46-55), socio-economic status (56-77), housing tenure (78-83) and family status (84-88) all being accounted for. For each variable, rates per 1,000 people rather than absolute numbers have been chosen to avoid area size creating bias. Where it has been impossible to calculate rates using related PAR data (variables 56-88), rates of migration efficiency (Equation (2.8)) have been used.

A note on distance As was explained in Chapter 2, and as will be further exemplified in Chapter 8, distance plays an important role in migration systems. Distance variables, however, were not included in the trial classification as preliminary experiments with clustering the suite of variables in Table 6.2 along with additional variables relating to the average distance moved, tended to create clusters forming concentric rings radiating out from central London. These distance effects are reminiscent of the functional regions based on local flow data described in the last chapter. An interesting avenue of future research would be to compare the results of functional regionalisation clusters (the size of which can be altered according to different distance and flow percentage thresholds) with standard geodemographic clusters including distance variables, unfortunately this is beyond the scope of this thesis, so distance will not feature in the variables included in the Migration Classification.

Before a first 'cluster run' can be carried out on the data, the list of variables needs to be reduced further, specifically to reduce instances of correlation. One solution, which has been used to reduce the number of cross-correlated variables in classifications based upon standard area counts, is to remove one variable from a related family of variables. Vickers (2006) suggests that where there are n groups within a variable, the optimum number of groups from that variable to include in a classification is $n-1$. Where a classification is being constructed from count data this makes sense. Figure 6.1 represents an area x with a count of 18 individuals residing within. These 18 individuals can be grouped according to their age. There are five different age groups from 0-15 to 60+. If there is information about the proportion of total individuals that each age group contains, then in order to obtain information about the number of individuals in all age groups n, only information about $n-1$ is required. The sum of the proportions in the youngest four age groups means that the proportion in the eldest (60+) age group has to be 11%. By only including $n-1$ variables all the information is still included. In this way the variable can be seen as a 'closed' variable.

Where flow data for an area are being used rather than count data, knowing $n-1$ does not mean it is possible to deduce n. As is shown in the example in Figure 6.1, knowing the inflow or outflow rates for the four youngest age groups reveals nothing about the inflow or outflow rates for the oldest age group. Therefore, in this instance, $n-1$ is not automatically the optimum

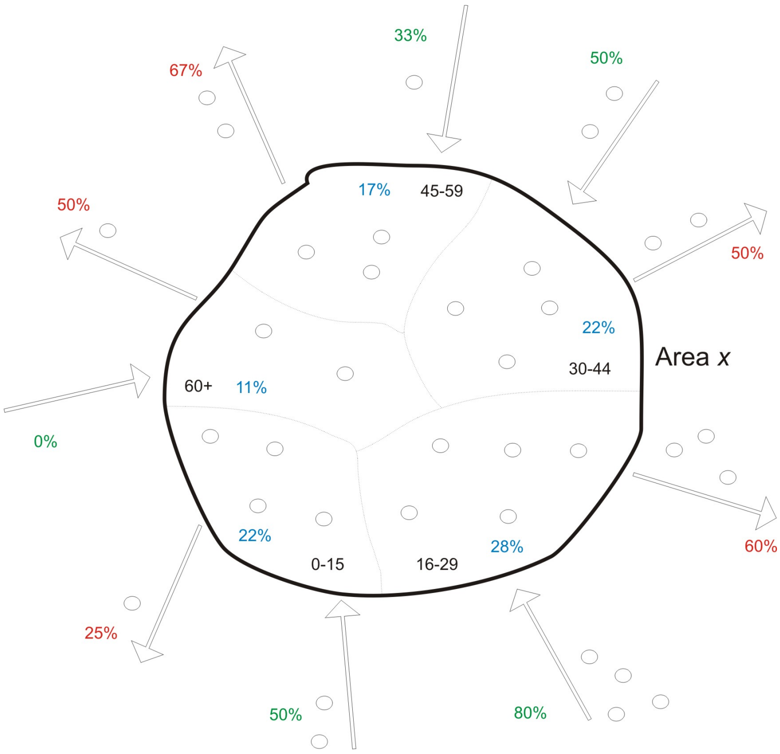


Figure 6.1: Why $n-1$ groups within a variable is not optimal for flow related data

number of groups to use from a parent variable. It may well be that if correlations with other variables are low, then the inclusion of n variables in a family of variables is permissible.

A matrix of Pearson's correlation coefficients has been calculated for every variable with every other variable in the list of 88. From this matrix, pairs of highly correlated variables can be identified in order that one from the pair might be dropped. The question here, however, is what constitutes a 'high' correlation coefficient? A coefficient of +1 or -1 signifies a perfect correlation, whereas 0 signifies a complete lack of correlation. What though is a suitable cut-off? Is anything over 0.5 or under -0.5 a high correlation, or should this figure be higher? The decision that is made will obviously affect subsequent analysis, but is also highly subjective. As a guide, a correlation coefficient of 0.7071 is equivalent to around 50% of one variable being associated with the other (Vickers et al., 2003). A higher coefficient means that even more of a variable's information can be gained from looking at the other variable in the pair. It seems appropriate that where more than 50% of a variable's information can be gained from elsewhere, then this variable would be a candidate for omission from the classification. For each variable in the initial list of 88, a count of the correlation coefficients over 0.7071 was created to flag those variables that it might be useful to omit from the classification. Particularly numerous instances of high correlation were found with variables relating to White ethnicity and an economically active status. This is unsurprising as the majority of individuals in Britain are both White and economically active. As a result, these variables were dropped from the list. Similarly, some variables related to no limiting long-term illness showed higher instances of high correlation, so were also dropped from the list, leaving 73 remaining variables.

Examining correlation, however, should not be the only technique used to choose variables for inclusion in a classification. Shepherd (2006, p.112) notes that in some instances a high correlation *"may not be a good justification for removal"*. For example, the age variables 0-15 and 30-44 (for all interaction types) have a consistently relatively high correlation with each other, as well as with other variables, although the age 30-44 variable shows a slightly higher correlation with other variables. The inclination, based purely on correlations, would be to drop the latter variable; however, with the majority of migrations of young people happening only as a result of parental migration, it is likely that age 30-44 is empirically a more important variable to keep. Another technique, therefore, is required to help make a more effective decision on the inclusion/exclusion of some variables. PCA is a technique advocated by a number of authors (Everitt et al., 2001; Harris, 2005; Shepherd, 2006; Vickers, 2006; Vickers et al., 2003) in the variable selection stage of classification creation, and can be used in conjunction with correlation analysis to choose variables where correlation does not help, as in this example.

It should be noted that both PCA and cluster analysis are data-reduction techniques. The data in this analysis can be conceptualised across two different dimensions - the variable dimension and the object dimension. The variables are the characteristics of the migrant data - age, ethnicity, socio-economic status, etc... the objects are the building blocks of the spatial system - the districts. PCA is commonly used to reduce data across the variable dimension, whereas cluster analysis is used to reduce data across the object dimension, but as both essentially reduce the dimensionality of the data researchers have found relationships between the two (Ding and Xiaofeng, 2004). Normally in a principal components analysis, n correlated variables are reduced to k principal components which represent the uncorrelated essence of the original variables (e.g. inflows of 30-44 year old and inflows of economically active migrants could be reduced to a new variable representing the essence of both perhaps called 'inflows of employed, family-aged adults'). Where multiple correlated variables are included in a PCA, each variable will have a greater or lesser degree of association with this new variable which captures the essence of the others. Whilst this new set of components can be used as surrogates in the analysis, where the interpretation of the clusters created is desirable this is not necessarily a good option. Harris (2005) warns precisely of this problem. He also notes that commercial geodemographic companies such as Experian have avoided PCA claiming that the distinctions between cluster types become blurred when these surrogate variables are used.

Here, however, PCA is used in the variable selection process, as an exploratory analysis tool whereby variables which have higher amounts of their variance accounted for within a particular component can been seen as more important within the dataset. Everitt et al. (2001) refer to this as a measure of *"interestingness"* - more interesting variables are more desirable to include in a cluster analysis. A number of outputs from a PCA can be used to assess the 'interestingness' of the component variables. The amount of variance explained by each factor is explained by its eigenvalue. The larger the eigenvalue the more variance is explained (Kline,

Table 6.3: Variables exhibiting low component loadings in the first 6 rotated components produced by PCA

Variable
Migration efficiency of other moving groups whose household reference person is in NS-SEC category 4
Migration efficiency of other moving groups whose household reference person is in NS-SEC category 5
Migration efficiency of other moving groups whose household reference person is in NS-SEC category 7
Migration efficiency of other moving groups whose household reference person is in NS-SEC category 8
Migration efficiency of wholly moving households whose household reference person is in NS-SEC category 8
Migration efficiency rate of other moving groups moving into socially rented accommodation
Migration efficiency rate of wholly moving households moving into socially rented accommodation
In-migration rate of individuals with a limiting long-term illness
Within-area migration rate of individuals with a limiting long-term illness
In-migration rate from no previous address of individuals with a limiting long-term illness
Internal out-migration rate of non-whites
International immigration rate of non-whites

1994), with initial components having larger eigenvalues than latter components. As well as the list of eigenvalues and the related variance explained by each component, PCA also produces a component loadings matrix whereby the proportion of a variable associated with a component is displayed. Variables with large proportions in the early components are important in the context of the entire dataset. Care should be taken, however, to 'rotate' a component matrix before it is interpreted (Kline, 1994). The purpose of rotation is to pick the most simple principal component solution. Whilst there are a number of ways to rotate a component matrix, the 'Varimax' solution produces for each component variable loadings which are either high or near zero - a feature of a simple solution (Kline, 1994).

PCA was run on the 73 remaining variables, producing 12 components with eigenvalues greater than 1, accounting for around 78% of the total variance in the dataset. From this, a list of variables with low component loadings for the first 6 rotated components (accounting for around 70% of the data) was created (Table 6.3). Featuring consistently low component loadings, these variables could now be considered for omission from the group used in the initial classification.

Returning to the initial problem relating to the choice of age groups, PCA reveals that in the first component, the component loading scores for age group 30-44 are generally higher than they are for age group 0-15, suggesting that it may be more useful to include 30-44 age group variables in the classification rather than 0-15 age group variables.

Before a final decision is made, however, consideration should also be given to the variation of the variables across the areas comprising the spatial system for the classification. Vickers et al. (2003) suggest that by examining the standard deviation of each variable, an appreciation of the extent to which they vary across space can be gained. Shepherd (2006) warns that variables with particularly low standard deviations will probably add little to cluster definitions, whereas those with high standard deviations may feature undesirable outliers. Examining the

Table 6.4: Standard deviation of problematic age variables

Variable	Standard deviation
out_mig_rate_Age_30_44	0.0185
in_mig_rate_Age_30_44	0.0169
within_mig_rate_Age_0_15	0.0161
within_mig_rate_Age_30_44	0.0132
in_mig_rate_Age_0_15	0.0126
out_mig_rate_Age_0_15	0.0124
international_in_mig_rate_Age_30_44	0.0081
international_in_mig_rate_Age_0_15	0.006
no_addr_in_mig_rate_Age_30_44	0.0029
no_addr_in_mig_rate_Age_0_15	0.0023
All interaction categories average 30-44 age group	0.0119
All interaction categories average 0-15 age group	0.0099

standard deviation statistics for age groups 0-15 and 30-44 (Table 6.4), it is evident that whilst all standard deviations are low, age group 30-44, with a higher average standard deviation, is likely to prove a more discriminatory variable across the spatial system than age group 0-15.

So, with the additional information gained from PCA and the examination of standard deviation statistics, it was decided that, for this initial trial classification, the collection of variables that would be used would not include age group 0-15 and would not include those variables with consistently low component loadings from the PCA. These exclusions are in addition to the variables already excluded for having high correlations with greater numbers of other variables. Consequently, the list of variables to be included in the initial classification was reduced to 56. These are shown in Table 6.5:

Table 6.5: Variables used in the initial Migration Classification

	Variable
1	Internal in-migration rate of persons aged 16 to 29
2	Internal in-migration rate of persons aged 30 to 44
3	Internal in-migration rate of persons aged 45 to 59
4	Internal in-migration rate of persons aged over 60
5	Internal out-migration rate of persons aged 16 to 29
6	Internal out-migration rate of persons aged 30 to 44
7	Internal out-migration rate of persons aged 45 to 59
8	Internal out-migration rate of persons aged over 60
9	Internal within-area migration rate of persons aged 16 to 29
10	Internal within-area migration rate of persons aged 30 to 44
11	Internal within-area migration rate of persons aged 45 to 59
12	Internal within-area migration rate of persons aged over 60
13	International immigration rate of persons aged 16 to 29
14	International immigration rate of persons aged 30 to 44
15	International immigration rate of persons aged 45 to 59
16	International immigration rate of persons aged over 60
17	In-migration rate from no previous address of persons aged 16 to 29
18	In-migration rate from no previous address of persons aged 30 to 44
19	In-migration rate from no previous address of persons aged 45 to 49
20	In-migration rate from no previous address of persons aged over 60
21	Internal in-migration rate of non-whites
22	Internal within-area migration rate of non-whites
23	In-migration rate from no previous address of non-whites
24	In-migration rate of economically inactive individuals
25	Out-migration rate of economically inactive individuals
26	Within-area migration rate of economically inactive individuals
27	International immigration rate of economically inactive individuals
28	In-migration rate from no previous address of economically inactive individuals
29	Out-migration rate of individuals with a limiting long term illness
30	International immigration rate of individuals with a limiting long term illness
31	Migration efficiency of wholly moving households whose household reference person is in NS-SEC category 1.1
32	Migration efficiency of other moving groups whose household reference person is in NS-SEC category 1.1
33	Migration efficiency of wholly moving households whose household reference person is in NS-SEC category 1.2
34	Migration efficiency of other moving groups whose household reference person is in NS-SEC category 1.2
35	Migration efficiency of wholly moving households whose household reference person is in NS-SEC category 2
36	Migration efficiency of other moving groups whose household reference person is in NS-SEC category 2
37	Migration efficiency of wholly moving households whose household reference person is in NS-SEC category 3
38	Migration efficiency of other moving groups whose household reference person is in NS-SEC category 3
39	Migration efficiency of wholly moving households whose household reference person is in NS-SEC category 4
40	Migration efficiency of wholly moving households whose household reference person is in NS-SEC category 5
41	Migration efficiency of wholly moving households whose household reference person is in NS-SEC category 6
42	Migration efficiency of other moving groups whose household reference person is in NS-SEC category 6
43	Migration efficiency of wholly moving households whose household reference person is in NS-SEC category 7
44	Migration efficiency of wholly moving households whose household reference person is in NS-SEC category Full Time Student
45	Migration efficiency of other moving groups whose household reference person is in NS-SEC category Full Time Student
46	Migration efficiency of wholly moving households whose household reference person is in NS-SEC category Not Classified
47	Migration efficiency of other moving groups whose household reference person is in NS-SEC category Not Classified
48	Migration efficiency of wholly moving households moving into or from owner occupied accommodation
49	Migration efficiency of other moving groups moving into or from owner occupied accommodation
50	Migration efficiency of wholly moving households moving into or from privately rented accommodation

51	Migration efficiency of other moving groups moving into or from privately rented accommodation
52	Migration efficiency of individuals living alone
53	Migration efficiency of individuals not living in a family but with others in a household
54	Migration efficiency of individuals who are part of a couple family
55	Migration efficiency of individuals who are part of a lone parent family
56	Migration efficiency of individuals living in a communal establishment

Finally, in some instances it may be desirable to weight particular variables depending on their perceived importance. To a certain extent, a weighting exercise has already been undertaken through choosing the variables. All excluded variables have effectively been given a weight of 0, included variables 1. Shepherd (2006) outlines a range of mathematical techniques for weighting variables; other approaches such as those used by Experian in their Mosaic classification, described by Harris (2005), are more down to the judgement of the researcher. A common approach described by Everitt et al. (2001) is to weight variables according to their variability; a technique more often referred to as 'standardisation'. No weighting has been applied to the 56 variables in this initial classification.

6.2.3 Variable standardisation

After the variables to be used have been selected and left un-weighted, it is necessary to standardise them over the same range. This is particularly important when the units used to measure the variables differ. For example, in this initial classification most variables are gross rates measuring the overall magnitude of flow. Efficiencies, however, measure the magnitude of flow in a particular direction rather than the volume of measured by the other rates and as a result will feature some negative flows where out-migration is greater than in-migration. Furthermore, whilst in this initial classification only gross rates are used, it is entirely feasible that, if appropriate, alternative variables measured over different scales could be included in the future. Whenever data measured across different ranges are used, it is appropriate to standardise across the range so that any individual variable will not bias the classification. Whilst Aldenderfer and Blashfield (1984, p.21) debate the necessity of variable standardisation in all situations, they concede that, where units of measurement vary between variables, researchers classifying these data will *"undoubtedly want to standardise them."*

Once the decision has been made to standardise the data, the method of standardisation needs to be chosen. As with all other elements of the clustering process, the literature does not provide consensus on the most appropriate methodology to use. A number of researchers (Everitt et al., 2001; Milligan and Cooper, 1987; Shepherd, 2006) cite work by Milligan and Cooper (1988) which suggests that the most effective way of standardising data is to standardise over the range of data for that variable. That is:

$$Z_i = \frac{[X_i - \min(X)]}{\max(X) - \min(X)} \tag{6.1}$$

where:
Z_i = the standardised variable value for area i,
X_i = the value of variable X for area i,
$\min(X)$ = the minimum value of variable X for all areas and
$\max(X)$ = the maximum value of variable X for all areas.

However, work by Schaffer and Green (1996) contradicts the findings of Milligan and Cooper (1988). As the result of research carried out on empirical datasets (Milligan and Cooper did not use real data), they find that when six different types of standardisation are compared, all but one perform well, leading them to conclude that *"column variable standardization does not seem to affect clustering results nearly as much as other aspects [such as the] choice of clustering algorithm and the presence of noise variables"* (Schaffer and Green, 1996, p.162).

One of the more common ways of standardising data is through the calculation of z-scores. Z-scores standardise variable data for each unit (in this case the local authority district) by its standard deviation from the mean for the entire variable. That is:

$$Z_i = \frac{X_i - \bar{X}}{\sigma_X} \tag{6.2}$$

where:
$\bar{X}$ = the global mean for variable X
σ_X = the standard deviation for variable X
with:

$$\sigma_X = \sqrt{\frac{(X_i - \bar{X})^2}{N}} \tag{6.3}$$

Whilst there are also other methods that can be used to standardise data, in the light of the research by Schaffer and Green (1996) showing little difference in the clustering outcomes when different methods of standardisation were used, z-scores were chosen as the method of standardisation for this initial classification. This is in line with the method chosen to create the Vickers et al. (2003) district classification (although not Vickers' OA classification).

6.2.4 Proximity measure

The decision over which proximity measure to choose in order to judge the distance between the cluster centroids is another important decision which will affect the outcome of any clustering process. Generally, different measures of proximity are suited to different types of data (nominal/binary, categorical or continuous). Where data are continuous (as they are here), Everitt et al. (2001) list six commonly used measures of proximity, the most common of all being Euclidean distance. Whilst Euclidean distance may or may not be the most suitable measure to use, since the objective of this current exercise is to create a trial classification which will be refined at a later stage, for the moment Euclidean distance will be used as the proximity measure. A full discussion and an examination of alternative measures of distance will be discussed in Section 6.4.3.

6.2.5 Clustering method

Any researcher browsing through the literature on clustering will be presented with a plethora of different clustering techniques which can be applied to find groups within data. Choosing an appropriate clustering technique, therefore, can be a challenge, especially when any one particular clustering algorithm will almost certainly produce a different output from another. Aldenderfer and Blashfield (1984) note this problem and suggest that, in such a situation, the wise solution would be to run more than one clustering algorithm and compare the different results from each.

Even if a researcher chooses to use more than one clustering method to analyse a data set it still may be the case that one method may be more logical than another to start with. A review of the literature reveals that there are two main families of clustering method: hierarchical and partitioning. Summarised by Aldenderfer and Blashfield (1984), partitioning methods take n observations and classify them into k clusters or groups which satisfy the requirements of a partition (i.e. that each group must contain at least one data object and each object must belong to one group). When a partitioning method is used, the researcher must decide the value of k before the process begins. This, in itself, can be problematic when the optimum number of groups is unknown, therefore one approach to tackle this could be to go through the process several times with different values of k where the value of k can theoretically be as large as the number of observations. However, running the clustering algorithm for a number of values k could be a lengthy process. As an alternative to partitioning methods, hierarchical methods deal with all values of k at the same time and produce output from $k = 1$ to $k = n$. From this output with all possible solutions for k, the researcher is then able to select the solution with the most appropriate number of clusters k.

The logical way forward for an initial clustering run, therefore, would be to use a hierarchical clustering algorithm in order to ascertain what might be the most appropriate number of clusters, before optimising the solution at a later stage through the use of another clustering algorithm. Indeed, Everitt et al. (2001) suggest an initial partition may be created through a hierarchical technique before an optimisation algorithm such as k-means is used to re-arrange the original solution of k groups into a new solution of k groups - keeping the new partition only if an improvement is made; a methodology adopted in the past by the ONS in their local authority classifications (Bailey et al., 2000).

Within hierarchical methods of classification there are a number of different algorithms to choose from, some of which are agglomerative (i.e. which start with n single member clusters before amalgamating these clusters into successively larger and fewer clusters until a final solution is reached with one single cluster featuring all the data), and some which are divisive (starting with one large cluster before successively splitting the cluster until there are *k* clusters each containing a single data item). One of the more common hierarchical methods employed in the creation of area classifications - used by both ONS (Bailey et al., 2000) and Vickers et al. (2003) - is Ward's method (Ward, 1963). Simple methods of agglomerative clustering join cases to clusters if the case is similar to at least one case already in the cluster (single linkage/nearest neighbour); to all members of the cluster (complete linkage/furthest neighbour); or to the average for all members of the cluster (average linkage). Ward's method, on the other hand, optimises the minimum variance between cases within clusters. Cases are joined to clusters where the addition results in the minimum increase in the error sum of squares (Aldenderfer and Blashfield, 1984). Whilst all methods have their benefits and limitations and a number of studies have found conflicting performance (Everitt and Dunn, 2001), in this initial clustering run Ward's method will be used as it minimises the loss of information associated with each cluster as it is created (Vickers et al., 2003).

6.2.6 Number of clusters

Deciding upon the number of clusters is also a difficult challenge, although as previously pointed out, through using a hierarchical clustering method a solution is produced with all clusters enabling the choice of the appropriate number of clusters to be made subsequent to the clustering process. Dendrogram output (Figure 6.2) can give clues as to the best clustering solutions. As Everitt et al. (2001) describes, the best solutions are likely to be where clusters below a selected distance (from the cluster centre) on the dendrogram are distant from each other by the least amount. In other words, large changes in the distances indicate solutions where the optimum number of clusters are presented. The example in Figure 6.2, suggests that, in this case, it might be sensible to include all cases in one group as the largest distance between clusters includes all cases. When dendrograms are produced for a large number of cases, however, interrogation of the tree to find the optimum number of clusters becomes more difficult. An alternative method therefore is to examine the numerical distance coefficients. Where large jumps occur in the coefficients between clusters, the points where the jumps occur signify the optimum cluster solutions. Figure 6.3 is a graphical representation of the gaps in the coefficients. In this case, the number of clusters just after a steep decline in the graph represent the optimum cluster solutions.

It is clear to see in Figure 6.3 that the steepest declines in the graph occur just before 5 and 8 clusters, signifying that solutions containing either 5 or 8 clusters would be the best for this set of data. Reading the graph from right to left, what this shows is that if there is a steep rise as the number of clusters declines - for example between 8 and 7 clusters - the data within the 7 clusters is more dissimilar than the data within the 8 clusters. Since the objective of the clustering exercise is to find clusters with similar characteristics, it is sensible to select 8 rather than 7 clusters.

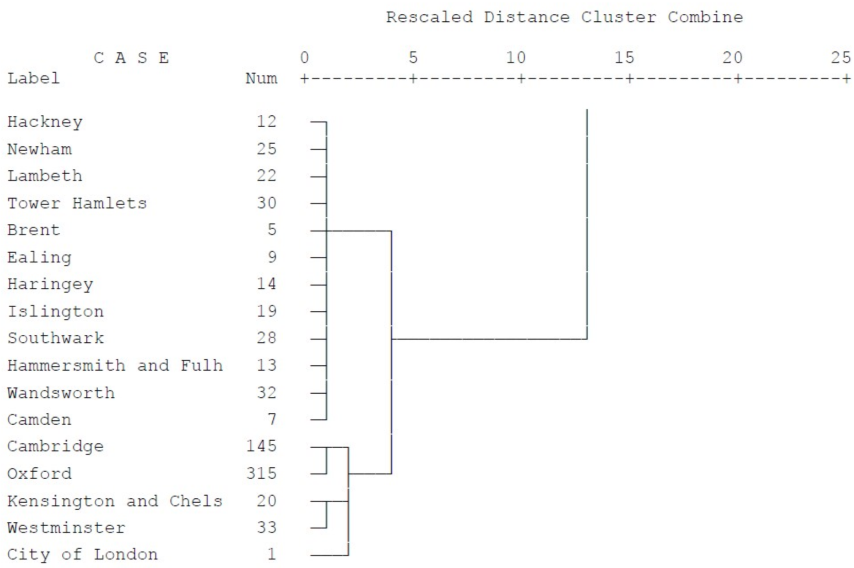


Figure 6.2: Sample dendrogram output

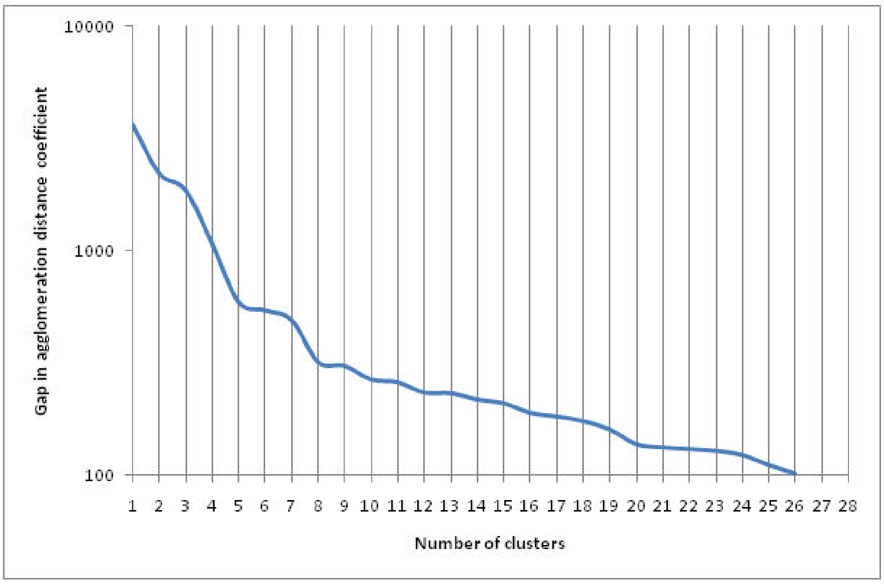


Figure 6.3: Agglomeration schedule representing the distance between the most dissimilar areas within cluster groups

6.2.7 Replication testing and interpretation

The final stage in Milligan and Cooper's framework for carrying out a cluster analysis involves the cross-validation and test of any cluster output that is produced. Validation is a lengthy process and necessarily happens after an initial partition has been made. Consequently discussion related to this stage of the classification process will be carried out later.

6.3 Initial classification results

Following the information given in the agglomeration schedule in Section 6.2.6 an initial, draft classification partition was created for 8 clusters. The spatial distribution of these clusters is shown in Figure 6.4.

Cluster 1 contains the fewest districts - three of which are located in the most central London boroughs; City of London, Westminster, and Kensington and Chelsea; the other two being Oxford and Cambridge. Cluster 2 forms a Greater London hinterland encompassing almost all boroughs bordering the London region as well as a swathe of districts in the Home Counties and a few sprinkled beyond. Cluster 3 is more spatially diverse but perhaps most concentrated in the Midlands, Northern England, South Wales and Scotland. Cluster 4 is a selection of districts found solely in inner London buffering cluster 1 from cluster 2. Cluster 5 is most concentrated around the Northern ex-industrial areas, South Wales and the North East. Cluster 6 is spatially diverse, but features districts mainly associated with thriving cities characterised in many cases by the presence of higher education institutions. Cluster 7 features districts mainly in the rural periphery of Britain, including the South West, Eastern England, central Wales and remoter parts of Scotland. The final cluster, 8, can be found principally in areas to the south, west and north-west of the main body of cluster 2. Other areas are found in the North, south-west and east. Two districts were omitted from the final classification by the software used to run Ward's algorithm; these were Merthyr Tydfil and the Isles of Scilly.

6.4 Refining the initial classification

So it is clear from Section 6.2 that through following a series of methodological steps, it is possible to produce a plausible classification of districts based on migration variables. As was noted earlier, however, geodemographic classification results are often conceivable. Clustering algorithms will always produce results regardless of the data input. Like any piece of work, however, the first draft is very rarely the final product. Once a draft has been produced, then it is reviewed and analysed, with improvements made where necessary to ensure the final result is the one which is based on the best decisions, and therefore produces the most robust end product. This next section, therefore, will review some of the key choices and additional considerations, and where necessary implement changes so that a final, definitive Migration Classification can be achieved.

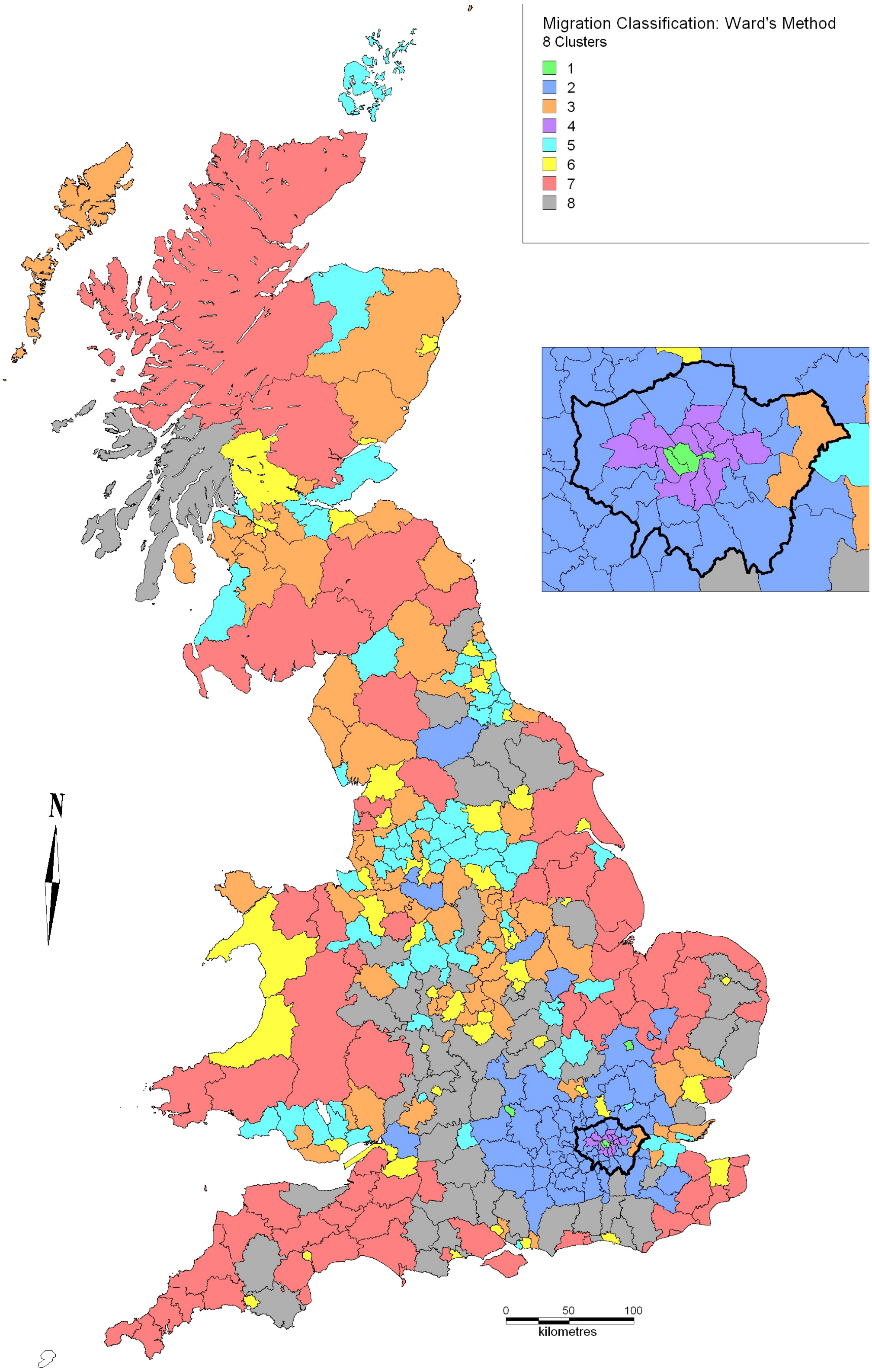


Figure 6.4: Spatial distribution of 8 migration data clusters created using Wards method

6.4.1 Variable transformation

One issue that was not confronted during the initial classification design was the that of variable transformation. Milligan (1996) chose not to address this issue in his seven steps, yet throughout the classification literature there is much discussion about the need (or not) to transform variables which do not meet normal distribution assumptions.

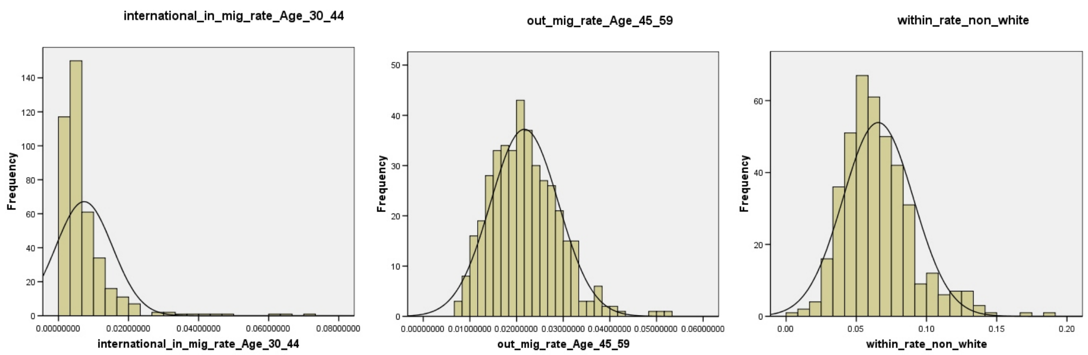
Within the more general statistical literature (Field, 2005, provides a particularly accessible overview) the importance of a normal, Gaussian distribution of frequency observations in data is frequently expressed, especially where parametric tests (which for their accuracy rely on such distributions) are employed. Often, frequency distributions do not follow a normal, symmetrical curve - observations may cluster at one end of the scale exhibiting either a positive skew (more frequent observations are at the lower end of the scale with fewer at the higher end) or a negative skew (more frequent observations at the higher end of the scale with fewer at the lower end). Where data are not distributed normally, typically, statisticians have 'corrected' the data to a more normal distribution in order that further analysis techniques can be used without the results of these analyses being unreliable. Field (2005) outlines the main ways in which skewed data can be corrected - the most common solutions being either to remove outliers (which may be affecting the frequency distribution) or to transform the data using either logarithmic, square-root or reciprocal transformations (Field, 2005, p.80).

Whilst there is a need to transform skewed data for some parametric tests such as ordinary least-squares regression to be reliable, the necessity for such transformations in cluster analysis is less clear-cut. It has been argued that skewed variables will bias cluster membership. In a migration classification, this might be apparent where, for example, inner London boroughs exhibit very high counts of international immigrants compared to all other districts in Britain. These cases may be clustered because of these very skewed variables, meaning that other interesting characteristics that these boroughs may exhibit for other more normally distributed variables will be ignored. The high values for these immigration variables will mean that it is by these variables that they are defined.

Both Vickers (2006) and Založnik (2006) advocate the transformation of skewed variables for the reasons mentioned above. There are others, however, who are less convinced of the case for transformation. Openshaw and Wymer (1995, p.245) remark that:

> *"Some thought may sometimes be given to the possibilities of applying a data transformation. After all, this sounds like the correct statistical thing to do. Well, think carefully about it and, then, perhaps don't do it! It can be argued that there is little to be gained from data transformations, bearing in mind the exploratory nature of classification and the difficulties it might cause during interpretation."*

Grayson (2004) also warns of possible implications of transforming data. By transforming data, whilst relative differences remain the same - i.e. London is still a more popular location for immigrants than Leeds, and Leeds is more popular than Cornwall; exactly how much more



(a) Skew = 4.14 (b) Skew = 0.73 (c) Skew = 1.01

Figure 6.5: Variable distributions and skewness statistics for three example variables

popular is lost in the transformation. Whilst both London and Leeds are more popular than Cornwall, it may appear that Cornwall is less popular than Leeds by the same amount than Leeds is less popular than London, when in reality Cornwall may be vastly less popular than both.

So whilst there are both reasons and advocates for and against the transformation of 'poorly behaved' variables, any decision about whether or not a transformation is necessary for this classification will need to be based purely on the variables being used here, rather than what has or has not been common practice for other classifications. A useful start point may be to assess the impact on the classification of transforming the variables, if indeed poor distributions suggest they need transforming.

Examining variable skewness

One of the main difficulties when deciding whether or not to transform data is assessing 'how skewed is too skewed?' Consider Figure 6.5 below.

Figure 6.5 shows frequency histograms for three variables included in the initial Migration Classification. One common (but perhaps not strictly scientific) way of assessing skewness is to examine the frequency histogram for a variable. To anyone with even a rudimentary training in examining histograms for skew, it would be obvious that the histogram on the left representing the counts of international immigrants aged 30-44, is displaying a significant positive skew, with the majority of observations found to the left of the x axis. With the other two graphs, however, the presence of skew is much less obvious. Indeed the distributions of both graphs look relatively normally distributed. Should these variables be classed as such - is it acceptable to include them in the classification?

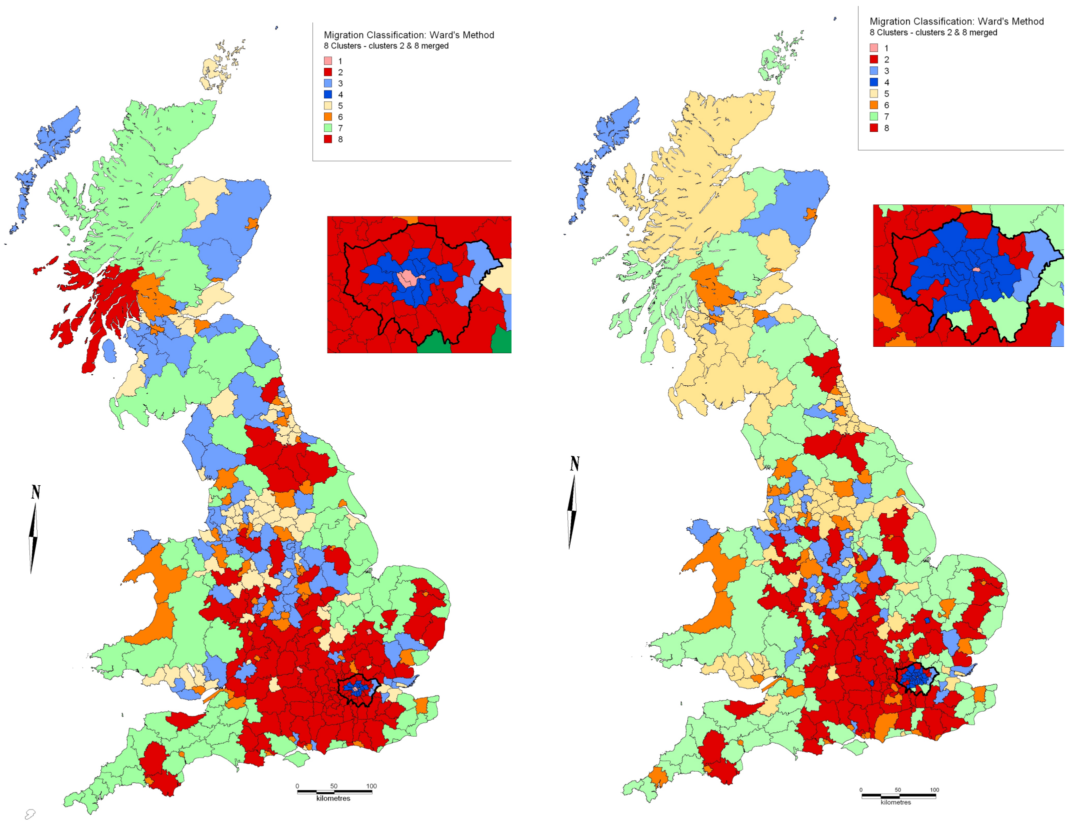
Rather than 'eyeballing' the data in a histogram, perhaps a less qualitative assessment would be to look at the skewness statistic for each variable. Most standard pieces of statistical software (such as SPSS) will provide these statistics as standard descriptive output. SPSS (2006) states that where a distribution is normal, the skewness statistic will equal 0, and as a general rule when this statistic is more than twice its standard error, then the distribution is skewed. As can be seen in Figure 6.5, in all cases the skewness statistic is more than twice the skewness standard error, suggesting all of the variables are skewed. Indeed, analysis of all 56 variables included in the original classification reveals that only 7 have skewness which is less than twice their skewness standard error, despite many variables appearing to display relatively normal distributions in their histograms.

Transforming skewed variables

If we are to accept that the skewness statistics indicate all but 7 variables need transforming, what are results of such a transformation, both on the variables themselves, and on the classification? As previously mentioned, two of the most common transformations used to correct data are logarithmic transformations and square root transformations. Both will be applied to the data and assessed. Before any transformation is applied, however, constant values need to be added to the data. It is pointed out by Vickers (2006), that as the logarithm of zero returns no result, a constant should be added to the data before transformation, if indeed zeros exist in the data. Within the 56 variable dataset being used here, a number of zeros occur, so a constant of 1 was added. All variables were transformed using both methods and the results are displayed in Table 6.6.

The results of the transformations are mixed. In some cases the transformations have reduced the skewness statistics, whereas in others, the original data remains more normally distributed. In fact, for the majority of variables, the original data shows the most normal distribution. As Vickers (2006) points out, it is not an option to apply different transformations to different variables - if any variables are to be transformed, all others included in the suite for clustering must have the same transformation applied to them. With the original data being more normally distributed than the transformed data, then this suggests that the data should be left un-transformed. Before moving on, it will be interesting to explore the effect any transformation has on a cluster solution. If the effect is small, then regardless of whether a transformation improves variable distribution or not, it may not be worthwhile pursuing it; therefore will poorly behaved variables present in the dataset have an undue effect on the clusters produced? One way to explore this would be to compare the cluster solutions produced with the original data and with transformed data.

Such a comparison can be made by comparing the maps in Figure 6.6. In this example, clusters 2 and 8 have been combined to aid exemplification. The two clusters are principally defined by similar variables so combining them here is not an issue. The first point of note is that transforming the data does not drastically alter the overall spatial patterns of the clusters. Both maps show comparable patterns, with the majority of districts remaining in the same cluster after transformation. Table 6.7 quantifies this by showing the number of districts comprising each cluster. At most the clusters change size by 30 districts, and across the whole classification, core



(a) Original data (b) Log transformed data

Figure 6.6: Comparison of areas created by Ward's clustering algorithm

Table 6.6: Results of log and square root transformations on skewness statistics

Variable	Skewness	Absolute difference between skewness and skewness standard error	skewness after log transformation	skewness after square root transformation	data with best skewness statistic (original data = orig; log transformation = lg; square root transformation = sqrt)
in_mig_rate_Age_16_29	1.21	1.09	0.01	0.59	lg
in_mig_rate_Age_30_44	0.60	0.48	-0.49	-0.02	sqrt
in_mig_rate_Age_45_59	0.94	0.81	-0.12	0.38	lg
in_mig_rate_Age_60_plus	0.48	0.36	-0.40	0.04	sqrt
out_mig_rate_Age_16_29	0.71	0.59	-0.19	0.23	lg
out_mig_rate_Age_30_44	1.59	1.47	0.20	0.83	lg
out_mig_rate_Age_45_59	0.73	0.61	-0.30	0.19	sqrt
out_mig_rate_Age_60_plus	0.55	0.43	-0.34	0.09	sqrt
within_mig_rate_Age_16_29	0.98	0.86	-0.25	0.40	lg
within_mig_rate_Age_30_44	0.03	0.09	*	*	orig
within_mig_rate_Age_45_59	0.28	0.16	-0.28	0.00	sqrt
within_mig_rate_Age_60_plus	0.19	0.07	*	*	sqrt
international_in_mig_rate_Age_16_29	2.75	2.63	0.24	1.37	lg
international_in_mig_rate_Age_30_44	4.14	4.02	0.39	1.97	orig
international_in_mig_rate_Age_45_59	5.85	5.72	0.53	2.34	orig
international_in_mig_rate_Age_60_plus	5.15	5.03	-0.12	1.49	lg
no_addr_in_mig_rate_Age_16_29	2.16	2.04	0.87	1.47	orig
no_addr_in_mig_rate_Age_30_44	1.63	1.51	0.67	1.14	orig
no_addr_in_mig_rate_Age_45_59	1.55	1.43	0.44	0.50	orig
no_addr_in_mig_rate_Age_60_plus	1.94	1.82	0.34	0.44	orig
in_mig_rate_non_white	1.51	1.39	-0.72	0.27	orig
within_rate_non_white	1.01	0.89	-0.48	0.02	sqrt
no_addr_in_rate_non_white	0.82	0.70	-0.71	-0.51	orig
in_mig_rate_Economically_Inactive__Total	1.71	1.59	0.52	0.81	orig
out_mig_rate_Economically_Inactive__Total	1.04	0.92	-0.20	0.38	lg
within_rate_Economically_Inactive__Total	2.08	1.96	0.51	1.28	orig
international_in_rate_Economically_Inactive__Total	3.16	3.04	0.27	1.49	orig
no_addr_in_rate_Economically_Inactive__Total	1.78	1.66	0.51	1.23	orig
out_mig_rate_LLTI_in_HH_and_CE_Total	8.24	8.12	0.71	2.21	orig
international_in_rate_LLTI_in_HH_and_CE_Total	3.23	3.11	0.16		lg
efficiency_NS_SEC_11_Wh_move_hh_All_groups	0.30	0.18	0.18	0.24	lg
efficiency_NS_SEC_11_Oth_mvg_grp_All_groups	-0.40	0.52	-0.55	-0.47	orig
efficiency_NS_SEC_12_Wh_move_hh_All_groups	-0.26	0.38	-0.32	-0.29	orig
efficiency_NS_SEC_12_Oth_mvg_grp_All_groups	-0.62	0.74	-0.71	-0.67	orig
efficiency_NS_SEC_2_Wh_move_hh_All_groups	-0.23	0.36	*	*	orig
efficiency_NS_SEC_2_Oth_mvg_grp_All_groups	-0.71	0.83	-0.76	-0.73	orig
efficiency_NS_SEC_3_Wh_move_hh_All_groups	0.11	0.01	*	*	orig
efficiency_NS_SEC_3_Oth_mvg_grp_All_groups	-1.35	1.47	-1.47	-1.41	orig
efficiency_NS_SEC_4_Wh_move_hh_All_groups	-0.13	0.25	*	*	orig
efficiency_NS_SEC_5_Wh_move_hh_All_groups	-0.30	0.42	-0.36	-0.33	orig
efficiency_NS_SEC_6_Wh_move_hh_All_groups	0.30	0.17	0.20	0.25	lg
efficiency_NS_SEC_6_Oth_mvg_grp_All_groups	-0.06	0.18	*	*	orig
efficiency_NS_SEC_7_Wh_move_hh_All_groups	-0.48	0.60	-0.57	-0.52	orig
efficiency_NS_SEC_FT_student_Wh_move_hh_All_groups	0.48	0.36	0.39	0.43	orig
efficiency_NS_SEC_FT_student_Oth_mvg_grp_All_groups	0.90	0.78	0.83	0.87	orig
efficiency_NS_SEC_Not_class_oth_reason_Wh_move_hh_All_groups	-0.40	0.52	*	*	orig
efficiency_NS_SEC_Not_class_oth_reason_Oth_mvg_grp_All_groups	0.74	0.62	*	*	orig
efficiency_Owner_occupied_Wh_mvg_hh_All_groups	-0.47	0.59	-0.53		orig
efficiency_Owner_occupied_Oth_mvg_grp__All_groups	-1.38	1.50	-1.42		orig
efficiency_Private_rented_Wh_mvg_hh_All_groups	-0.60	0.72	-0.69		orig
efficiency_Private_rented_Oth_mvg_grp__All_groups	0.39	0.27	0.34		orig
efficiency_Alone_total	0.02	0.10	*	*	orig
efficiency_Non_Family_Household_Total	0.66	0.54	0.61	0.63	orig
efficiency_In_couple_family_total	-0.60	0.72	-0.63	-0.61	orig
efficiency_In_lone_parent_family_total	-1.64	1.76	-1.85	-1.74	orig
efficiency_Living_in_a_communal_establishment_total	0.50	0.38	0.43	0.47	orig

* = no requirement to transform original data due to normal distribution.

areas appear to remain stable - university towns for cluster 6; ex-industrial areas for cluster 5; coastal and rural areas for cluster 7 etc. These results would suggest that transforming variables (especially where the transformation does not significantly improve skewness) does not have a huge impact upon a final cluster solution. Taking all of this evidence into consideration, the conclusion would have to be to leave the data in its original state, and to not apply a transformation.

Table 6.7: Effect of a log transformation on cluster membership

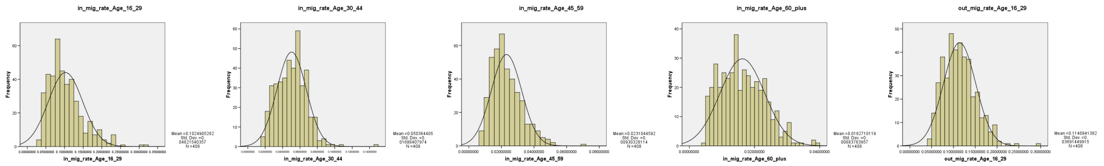
	Count of districts		
Cluster	**original data**	**logged data**	**difference**
1	5	1	-4
2&8	123	93	-30
3	79	49	-30
4	12	24	12
5	67	84	17
6	47	52	5
7	69	99	30

6.4.2 Dropping the most skewed variables?

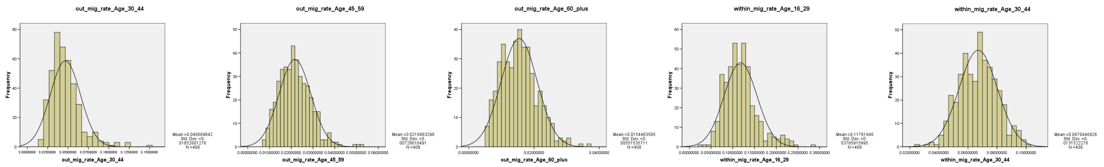
The previous section has shown that transformations make little difference to the skewness of the 56 variables used in the initial classification. But while the decision has been made not to transform variables, there still remain a number of more highly skewed which may affect the final cluster solution. Skewness statistics showed that most variables were skewed, however some variables were significantly more skewed than others. These are shown in Table 6.6 but also can be identified easily by studying the associated histograms (Figure 6.7).

A number of variables are very obviously more skewed than others. The order of the histograms in Figure 6.7 reading left to right and row by row is the same as Table 6.7. It has been suggested (Harris, 2005; Vickers, 2006; Založnik, 2006) that these very skewed variables (which also contain a number of outliers) will bias some of the clusters within the classification. A sensible course of action, therefore, would be to examine whether indeed this is the case. Will the most skewed variables have a detrimental effect on a classification produced? Selectively dropping the most skewed variables and re-running the cluster analysis to examine the effect would allow for an assessment of this type to be made. However, even where the inclusion of very skewed variables creates biased clusters in a classification, there are solutions to the problem. Harris (2005) suggest that if such clusters are created, one approach would be to allow them to form, but then to run a separate cluster analysis on the clusters created by the skewed variables, linking them back to the rest of the classification. The results below describe experiments carried where some of the most skewed variables were dropped from a cluster analysis.

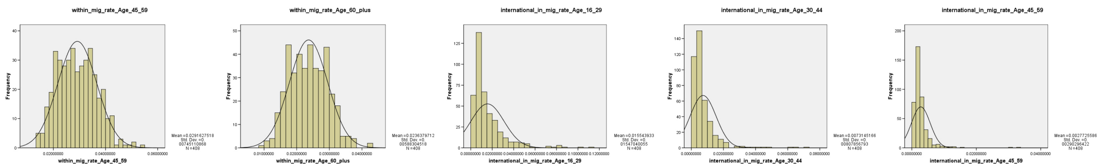
To establish a baseline for the experiment, the 56 variables from the initial classification were clustered using a *k*-means algorithm searching for 8 clusters. *k*-means was used instead of Ward's algorithm in this instance for reasons which will be explained fully in the next section. The algorithm ran through 1,000 iterations with different randomly selected initial cluster centroids in order to find the optimum cluster solution. The 8 clusters produced by this procedure are displayed below in Figure 6.8. In the second part of the experiment, exactly the same *k*-means procedure was applied to the data, although without most variables relating to no previous address, limiting long-term illness and the two most skewed economically inactive



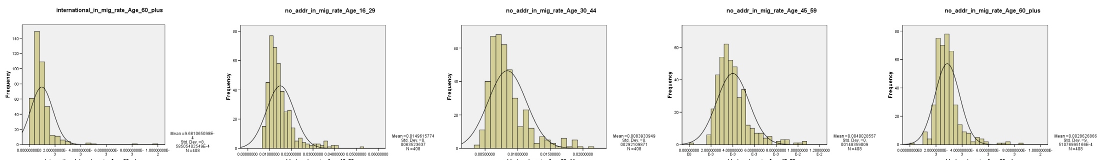
() In Mig Rate 16-29 () In Mig Rate 30-44 () In Mig Rate 45-59 () In Mig Rate 60+ () Out Mig Rate 16-29



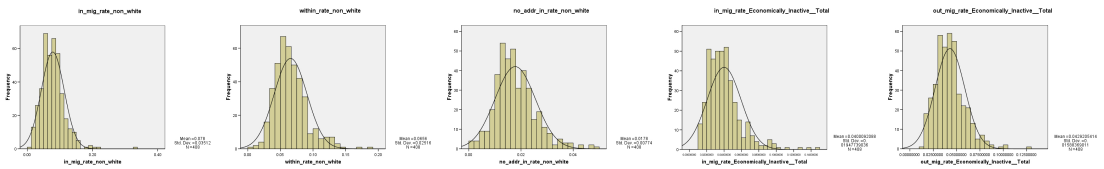
() Out Mig Rate 30-44 () Out Mig Rate 45-59 () Out Mig Rate 60+ () Within Mig Rate 16-29 () Within Mig Rate 30-44



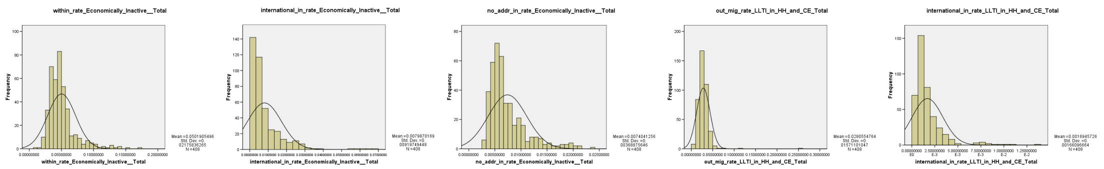
() Within Mig Rate 45-59 () Within Mig Rate 60+ () Int Mig Rate 16-29 () Int Mig Rate 30-44 () Int Mig Rate 45-59



() Int Mig Rate 60+ () No Addr Mig Rate 16-29 () No Addr Mig Rate 30-44 () No Addr Mig Rate 45-59 () No Addr Mig Rate 60+

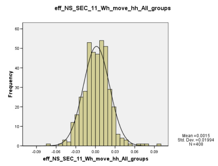


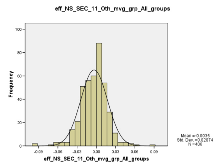
() In Mig Non White () Within Non White () NoAddr Non White () In Econ Inact () Out Econ Inact

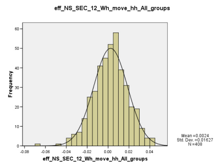


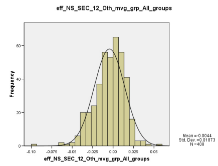
() Within Econ Inact () Int Econ Inact () No Addr Econ Inact () Out LLTI HH and CE () Int LLTI HH and CE

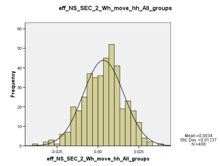
Figure 6.7: Frequency histograms for the 56 variables used in the initial classification

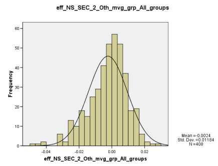

() Eff NSSEC 1.1 WMH

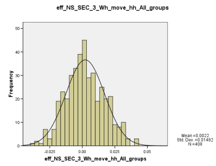

() Eff NSSEC 1.1 OMG

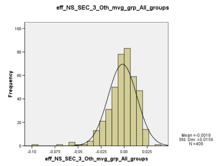

() Eff NSSEC 1.2 WMH

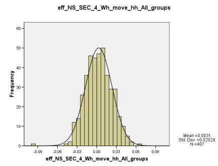

() Eff NSSEC 1.2 OMG

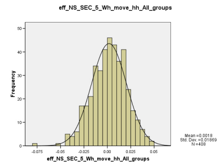

() Eff NSSEC 2 WMH

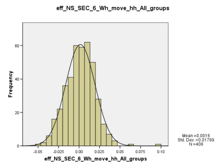

() Eff NSSEC 2 OMG

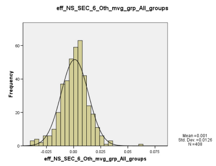

() Eff NSSEC 3 WMH

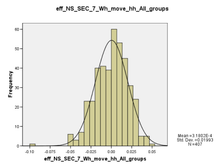

() Eff NSSEC 3 OMG

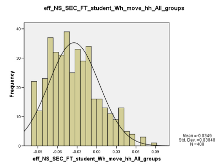

() IEff NSSEC 4 WMH

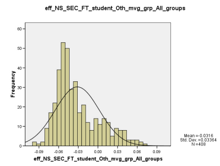

() Eff NSSEC 5 WMH

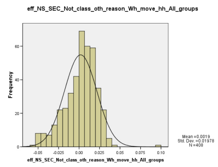

() Eff NSSEC 6 WMH

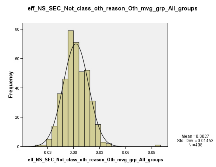

() Eff NSSEC 6 OMG

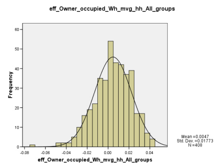

() Eff NSSEC 7 WMH

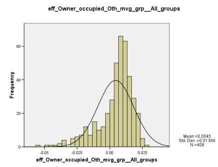

() Eff NSSEC FT Stu WMH

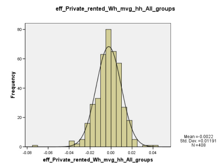

() Eff NSSEC FT Stu OMG

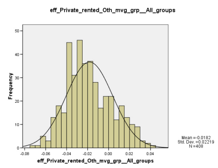

() Eff NSSEC Not Class WHM

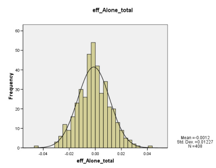

() Eff NSSEC Not Class OMG

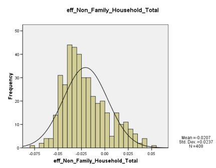

() Eff Owner Occ WMH

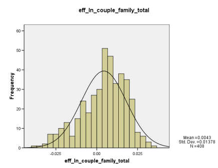

() Eff Owner Occ OMG

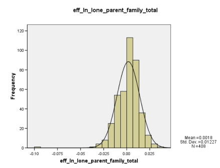

() Eff Priv Rent WMH


() Eff Priv Rent OMG


() Eff Alone


() Eff Non Fam HH


() Eff Couple Family


() Eff Lone Parent

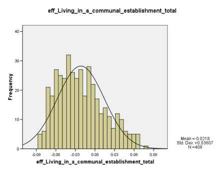

() Eff Communal Estab

Figure 6.7: Frequency histograms for the 56 variables used in the initial classification

variables. This means a total of 8 variables were dropped leaving 48 variables left to be clustered. The skewed international immigration age variables were left in at this stage in order to ascertain their influence on the cluster solutions; in the third part of the experiment they were removed. So in the third and final part of the experiment the data were put through the *k*-means algorithm one final time, this time with the 4 international immigration age variables removed as well - leaving 44 variables. Figure 6.9 below reveals the districts in Britain which moved to a different cluster after the first group of variables were dropped and then after the final international immigration variables were dropped.

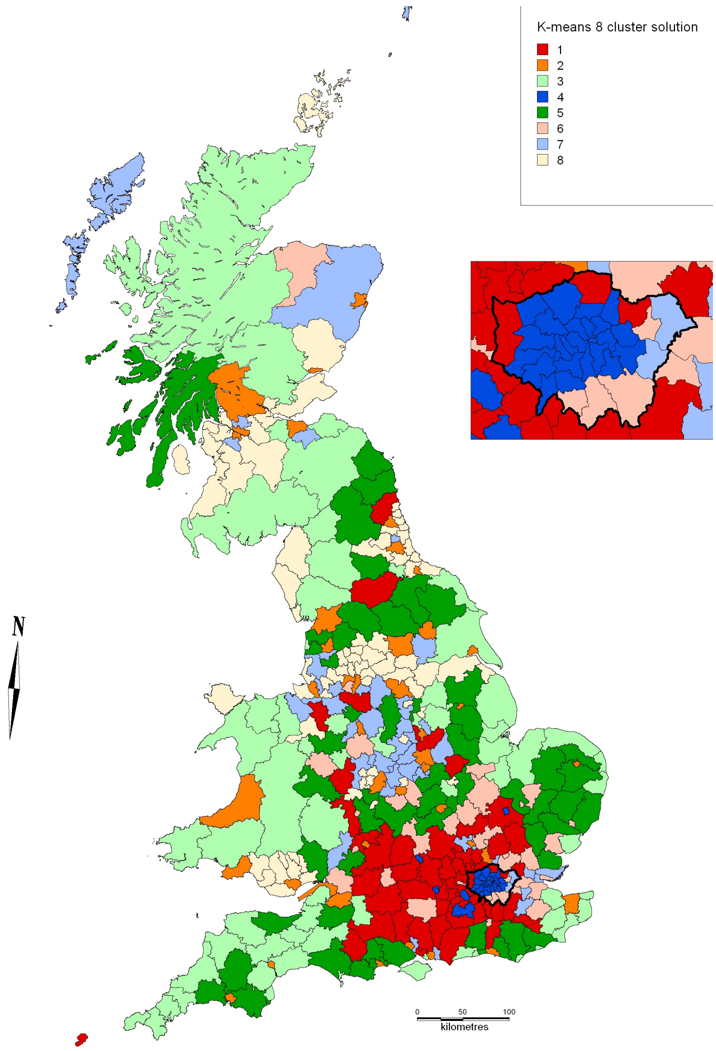
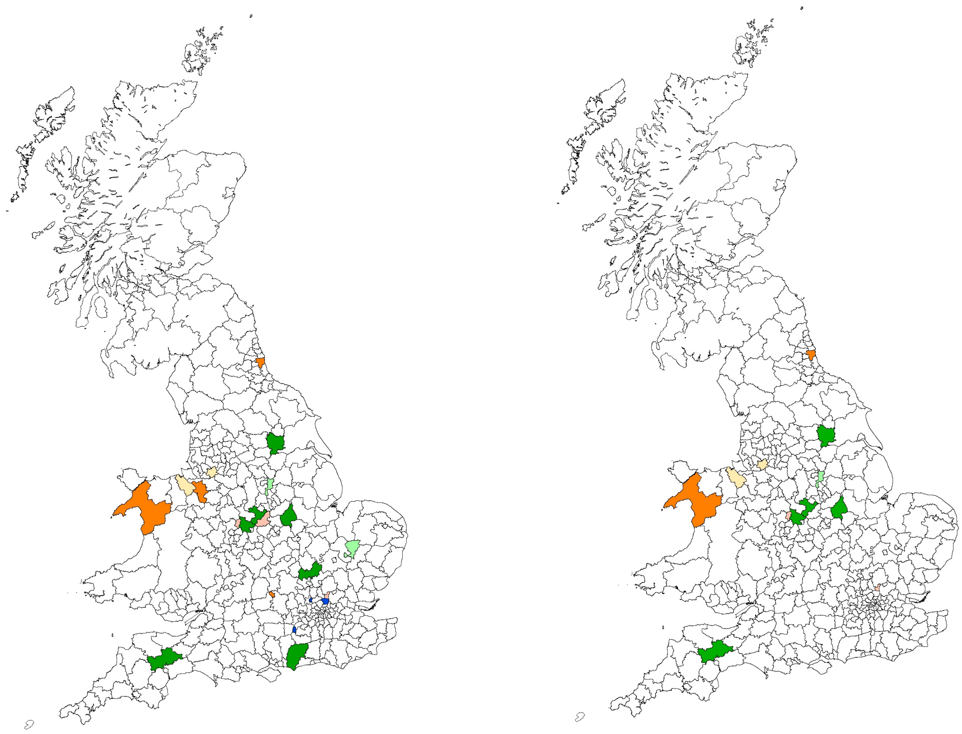


Figure 6.8: Clusters produced from a *k*-means clustering run searching for 8 cluster solutions

As is shown in Figure 6.9, dropping the first set of highly skewed variables has a negligible effect on the final cluster solution. Only 12 districts in Britain change cluster as a consequence. This is an unexpected result given the stated effect (Harris, 2005; Vickers, 2006; Založnik,

2006) of skewed variables on cluster solutions. It might be suspected, for example, that the cluster defined in a large part by no usual address variables (as these were the most numerous dropped) would be affected the most. Cluster 4 was the original cluster defined mostly by these variables, but there were not any districts in cluster 4 which moved group. Most changes, in fact, occurred in cluster 7. Even when the skewed international immigration variables were dropped (Figure 6.9), very few additional districts changed their cluster membership. An additional 9 clusters changed, principally from cluster 1 which forms the London periphery - a cluster not defined heavily by these international variables.



(a) 48 Variables (b) 44 Variables

Figure 6.9: Districts changing cluster group after total number of variables reduced to 48 and 44 variables

Note - colours not representative of previous Figures

What this experiment has shown is that the most skewed variables are not playing a very large role in the formation of clusters in this classification, despite the general consensus in the literature that skewed variables or variables with outliers will tend to create their own biased clusters. Whilst international migration, limiting long-term illness and no usual address variables can characterise some clusters when included, when they are removed, rather than districts being re-appropriated by other clusters, it appears that other variables are maintaining the cluster structures. For example, despite international migration being heavily concentrated in London (hence the heavy skew), other defining characteristics such as high volumes of young migrant internal in-migration are even more important in defining the cluster.

This is a significant discovery. Whilst it is undesirable to have skewed variables dominating some clusters to the detriment of the overall classification, where variables are having more-or-less no effect on the classification, their continued inclusion adds no value to the final solution. International migration can be dropped altogether as a variable within the classification, making the classification an internal migration classification for monitoring internal migration, rather than a classification partially defined by immigration, but designed for studying internal migration. It could be argued that this decision to drop international migration variables could have been made much earlier on in the thesis, but given the potential associations with international migration and internal migration, it was worth exploring the influence of international migration on a trial classification before discounting it altogether.

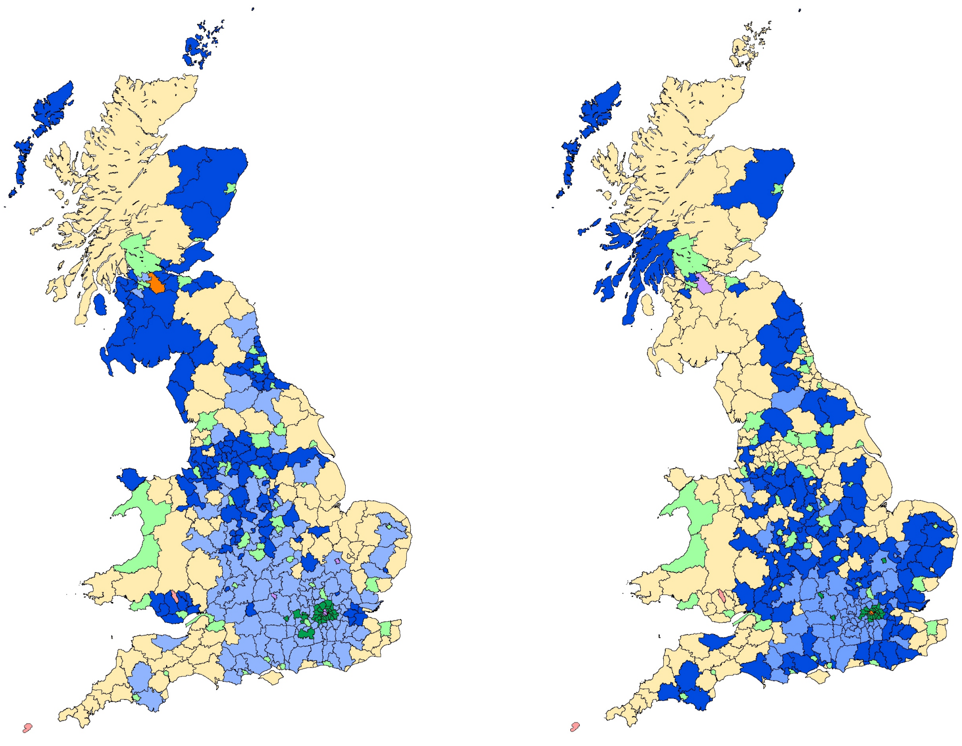
It may be of use in the future to construct an international migration (immigration and emigration) classification, however with data and aggregate estimates of international immigration being unreliable (Boden and Rees, 2009; Rees et al., 2009), details of individual immigrants more so, and estimates of emigration even poorer such a classification may not be feasible. The volatile nature of international immigration and uncertainties about the length of migrant stay - something which caused by changing economic circumstances at both origins and destinations, and is unlikely to become more stable as the current global economic crisis continues (Boden and Stillwell, 2006) - will affect the reliability and usefulness of any classification produced.

6.4.3 Cluster Optimisation: A Different Clustering Algorithm

The previous section of this paper has made it clear that it will be best to use non-transformed, internal migration variables in the final classification. Having dealt with that issue, the next stage in refining and arriving at a final classification concerns the clustering methodology itself. Ward's algorithm was used in the initial classification, principally because the method of creating a hierarchy of clusters allowed for a partition to be selected where the most suitable number of clusters was unknown. The main aim of partitioning data in a classification, as outlined by Gordon (1999), is to group objects that are similar to each other in one class, and dissimilar objects in another class. As mentioned during the development of the initial classification, one of the issues with using a hierarchical algorithm is that it may not necessarily find the optimum solution where all cases/objects are allocated to the class to which they are most similar - i.e. closer to the cluster centroid of one cluster than to the centroid of any other. The nature of the agglomerative hierarchical algorithm means that once a case has been allocated to a particular cluster, it cannot then be removed and re-allocated to a more appropriate cluster. It is for this reason that Everitt et al. (2001) advocate the use of an optimisation algorithm, which, given a specific number of clusters to create, will iteratively allocate and reallocate points to different clusters until no improvement is made to a final solution.

One such algorithm is the *k*-means algorithm (Everitt et al., 2001). For a given initial partition of *k* clusters with randomly selected cluster centroids, the algorithm takes each case in a dataset of n cases, allocates a case to a cluster, recalculates the centroid of that cluster and repeats the process until each case has been allocated to a cluster and the reallocation of any case does not improve the average distance of cases to the cluster centroids (for a more detailed description of the k-means algorithm, see Založnik 2006). Of course the distance to the cluster centroid can be measured in a number of ways, and this can affect the final cluster solution - this will be discussed in full later.

One of the main issues with using a k-means procedure is that the clusters created after the algorithm has been run once may represent a local optimum (i.e. cases are allocated to their closest cluster centroids, but these centroids may not be optimum, merely the artefact of their initial seed or partition), but may not represent a global optimum (where the centroids also represent the best possible solution) (Gordon, 1999). Frequently in k-means clustering, depending on the cases chosen as the initial cluster centroids, it can be that, even when all other elements of the clustering process remain constant, different final cluster solutions will be reached. This is exemplified clearly in Figure 6.10.



(a) Cases sorted by district name

(b) Cases Sorted by district code

Figure 6.10: Alternative outcomes of the k-means clustering of 408 cases, 56 variables due to sorting in SPSS

Figure 6.10 shows the cluster solutions produced when the 56 variables used in the initial classification were clustered into 8 clusters using the k-means algorithm available in the SPSS software package. In Figure 6.10a, the 408 cases are sorted alphabetically by district name, in

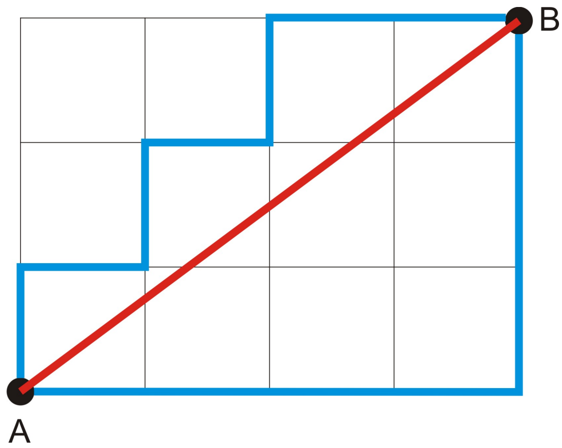


Figure 6.11: A representation of the difference between Euclidean and Manhattan (City Block) distances between two points

Figure 6.10b they are sorted by district code. Clearly the two maps are quite different. Sorting the districts in different ways has drastically altered the solutions produced by exactly the same algorithm using exactly the same data.

The issue with using the *k*-means algorithm as it is implemented in SPSS is that the user does not have control over the selection of initial cluster centroids. The software will allocate these centroids randomly (although the exact method for this is not made clear in the documentation) before iterating through until the end. SPSS (2005) acknowledges that the order of the cases will, in all likelihood, affect the final cluster solution, but only offers the (rather unsatisfactory) solution to sort the cases *"in different random orders"* (SPSS, 2005, p.489) to verify a cluster solution. This is impractical at best, especially where there are a large number of cases. At worst, it is likely that an optimum solution could never be reached intelligently. Work by (Falkenauer and Marchand, 2001, cited in Založnik, 2006) found that in experiments with a dataset using 10,000 different initial partitions, 9,874 different cluster solutions were found.

An additional issue with using the *k*-means algorithm in SPSS is that the distance measure used to measure the distance of cases to cluster centroids is Euclidean distance and cannot be altered (SPSS, 2005). The solution SPSS offers to this problem is to use a hierarchical cluster analysis procedure. However, doing this would obviously be unsatisfactory as *k*-means is being used here to optimise the solution already produced using a hierarchical algorithm! Whilst Euclidean distance was used as the distance metric in the initial classification, it may well be that other distance measures produce better cluster solutions. Indeed research by Aggarwal et al. (2001) points strongly to Manhattan (City) distance between points providing a better cluster solution than standard Euclidean distance where the data has many dimensions - many in the Aggarwal et al. example being more than 20 dimensions.

To explain the difference between the two measures, consider Figure 6.11. Manhattan distance differs from Euclidean distance in that it is the sum of the absolute difference between two coordinates - put another way the distance between two points on a grid system where only the grid can be travelled along (analogous to the distance travelled by a taxi driver along the road grid network in New York City - hence Manhattan), whereas Euclidean distance is the straight line distance between two points in space. The side of each square in the grid represents one unit of space, and points A and B can be located on this two dimensional grid space. The red line represents the straight line Euclidean distance between them; the blue lines demonstrate two alternative ways the same Manhattan distance could be calculated. In this case, the Euclidean distance is 5 units, whereas the Manhattan distance is 7 units. With the data being used in this research containing at least 44 variables (or dimensions) the research of Aggarwal et al. indicates clearly that using Manhattan distance to measure the distance to cluster centroids would result in a better definition of *k* clusters.

Using *k*-means in MATLAB

So with a number of problems inherent in the way that the SPSS software implements the *k*-means clustering algorithm an alternative solution needs to be sought. One possibility would be to write a bespoke piece of software which implements the *k*-means algorithm with the option to run the algorithm through a user defined number of iterations, each starting with different initial cluster centroids, the final solution being the one fitting some defined 'best' criteria, and with the option to choose different measures of distance to the cluster centroids. In an ideal world, every researcher would have the computer programming skills to be able to do this. In reality few do - and even if they could, could rarely afford the time to write such a piece of software - so rely on the skills and expertise of others to create tools for them to use which meet their requirements as far as possible.

Aside from SPSS, a number of other statistical analysis packages are available to researchers which have a version of the *k*-means algorithm pre-programmed. Packages such as Minitab, R and Stata will all run a *k*-means cluster analysis on a given dataset. The program, however, which met the needs of this analysis very well was the MAtrix LABoratory (MATLAB) Statistics Toolbox (MathWorks, 2009). MATLAB incorporates a number of features in its *k*-means algorithm which makes it preferable to SPSS.

Firstly, MATLAB allows for a choice of five different distance measures. Both Euclidean and Manhattan distances are available as well as cosine, correlation and hamming distances. In addition to this, MATLAB also offers a solution to the problem of a local rather than global minima being reached at the end of the clustering iterations. An optional *'replicates'* parameter can be included in the algorithm. This parameter will run the algorithm for the specified number of replicates (1...*n*) with each replicate starting the whole cluster run again with a new set of randomly selected initial cluster centroids. Once the specified number of replicates have been completed, the solution offered up by the program is the one with the lowest total sum of distances to which cluster centroids happen to have been chosen (MathWorks, 2009). This solution is likely to be the global minimum, although obviously the more replicates used, the more confident one can be that this is indeed the case.

6.4.4 Choosing k

One of the difficulties with using k-means over a hierarchial algorithm is that the number of clusters k needs to be defined at the outset. As with many elements of classification building the literature offers no definitive answer for deciding the most appropriate value of k. As stated by Everitt et al. (2001) the initial partition with an associated number of clusters might be chosen through prior knowledge or from result of a previous clustering method. Everitt et al. (2001) and Gordon (1999) review a number of other methods which could be used, ranging from the slightly more subjective, such as the assessment of 'large' differences in the distances between the most dissimilar areas in cluster groups in graphical representations (as used in the initial classification), to the more formal, such as those assessed by Milligan and Cooper (1985, cited in Everitt et al., 2001 and Gordon, 1999), which in general use mathematical procedures to assess the within and between cluster differences - the best results generally being where within cluster distances are minimised and between cluster differences are maximised. Whilst both Everitt and Dunn (2001, p.105) and Gordon (1999, p.63) state (almost too similarly) that no one method assessed by Milligan and Cooper should be used in preference to another, and that researchers should *"synthesise the results of several"* techniques, it is impractical to exhaustively work through a large range of methods. A sensible option would be to select some of the more popular methods and use a combination of those.

One particular method invented by Rousseeuw (1987), recommended by Kaufman and Rousseeuw (2005), espoused by MathWorks (2009) and implemented as the principal method to decide the number of clusters in the classification developed by Shepherd (2006), is the interrogation of 'silhouette' plots and values. A silhouette plot is a graphical representation of the average dissimilarity between any object/case within a cluster and other objects within both its own cluster and those in other clusters. The plots are represented on an index of -1 to +1. A value close to +1 signifies that that object is nearer to its own cluster than any other. A value close to -1 suggests that the object might well be better placed in another cluster. Zero signifies that it is unclear whether that object is better placed in its current cluster or another.

A silhouette S value for any object i in a cluster can be defined thus:

$$S_i = \frac{b_i - a_i}{\max\left(a_i, b_i\right)} \tag{6.4}$$

where:

a_i = the average dissimilarity of i to all other objects in the same cluster

b_i = lowest average dissimilarity of i to the objects in all other clusters in the whole solution

Silhouette plots are ranked (highest to lowest) silhouette values for the objects in each cluster. Better defined clusters will have fewer values close to or below zero compared to others.

In the construction of the ONS OAC, Vickers (2006) chose not to use silhouette data to select the most appropriate value of k, but instead employed a selection of other methods - some more logically than others. For example, whilst Vickers used the average distance of points in a cluster from the cluster centroid as one of his methods, the utility of this method is unclear since the average distance to cluster centroids will always reduce as the number of clusters increases. Of more use is the assessment of cluster size. One key observation made by Vickers (2006, p149) is that it is desirable to have clusters which are similarly matched in size - equal clusters being the optimum solution with very large and very small clusters being much less desirable as small clusters are more likely to contain outliers; therefore groups of more even sizes reflect groupings where outliers do not predominate. This is logical, and Vickers uses the average distance from the mean number of cases in each cluster for a range of cluster solutions to help decide the most appropriate number of clusters for the OAC.

In this classification, silhouette data as well as statistics for the size of clusters were produced for different values of k in order that both methods could be used in parallel to select the most suitable number of clusters for the final solution. These methods used in conjunction with different measures of cluster distance (Euclidean and Manhattan) along the replicates parameter in MATLAB will be discussed in the next section where the final district level Migration Classification is outlined.

6.5 Arriving at a final classification

Now an initial trial classification has been produced, and both the variables selection and methodology reviewed with decisions made about how to improve both, the task still remains to build the final *Migration Classification*. For reasons discussed, variables relating to international immigrants, ill health and most relating to migration from no previous address, were dropped from the initial set of 56 variables, reducing the final set of variables to 44. These are listed below in Table 6.8.

6.5.1 A decision on k

This set of final variables then needed to be clustered using the k-means algorithm in MATLAB. Manhattan (City) distance was selected as the most appropriate distance measure (however Euclidean distance was also tested to compare the solutions produced). Where the most appropriate number of clusters was not known, k-means was run for a range of clusters from 2 to 14 each using 200 replicates to attain a global minimum. This range of clusters was chosen as it was felt likely that the optimum solution would fall somewhere within this range. The initial classification had suggested 8 clusters were most appropriate, Vickers (2006) suggests that around 6 may be a useful place to start, whereas Shepherd (2006) tests a range between 5 and 25. A range somewhere around these numbers would in all probability produce the optimum solution. Silhouette and cluster size metrics were produced for each of the cluster solutions.

Table 6.8: Final selection of internal migration variables used in the classification

	Variable
1	Internal in-migration rate of persons aged 16 to 29
2	Internal in-migration rate of persons aged 30 to 44
3	Internal in-migration rate of persons aged 45 to 59
4	Internal in-migration rate of persons aged over 60
5	Internal out-migration rate of persons aged 16 to 29
6	Internal out-migration rate of persons aged 30 to 44
7	Internal out-migration rate of persons aged 45 to 59
8	Internal out-migration rate of persons aged over 60
9	Internal within-area migration rate of persons aged 16 to 29
10	Internal within-area migration rate of persons aged 30 to 44
11	Internal within-area migration rate of persons aged 45 to 59
12	Internal within-area migration rate of persons aged over 60
13	In-migration rate from no previous address of persons aged over 60
14	Internal in-migration rate of non-whites
15	Internal within-area migration rate of non-whites
16	In-migration rate from no previous address of non-whites
17	In-migration rate of economically inactive individuals
18	Out-migration rate of economically inactive individuals
19	Migration efficiency of wholly moving households whose household reference person is in NS-SEC category 1.1
20	Migration efficiency of other moving groups whose household reference person is in NS-SEC category 1.1
21	Migration efficiency of wholly moving households whose household reference person is in NS-SEC category 1.2
22	Migration efficiency of other moving groups whose household reference person is in NS-SEC category 1.2
23	Migration efficiency of wholly moving households whose household reference person is in NS-SEC category 2
24	Migration efficiency of other moving groups whose household reference person is in NS-SEC category 2
25	Migration efficiency of wholly moving households whose household reference person is in NS-SEC category 3
26	Migration efficiency of other moving groups whose household reference person is in NS-SEC category 3
27	Migration efficiency of wholly moving households whose household reference person is in NS-SEC category 4
28	Migration efficiency of wholly moving households whose household reference person is in NS-SEC category 5
29	Migration efficiency of wholly moving households whose household reference person is in NS-SEC category 6
30	Migration efficiency of other moving groups whose household reference person is in NS-SEC category 6
31	Migration efficiency of wholly moving households whose household reference person is in NS-SEC category 7
32	Migration efficiency of wholly moving households whose household reference person is in NS-SEC category Full Time Student
33	Migration efficiency of other moving groups whose household reference person is in NS-SEC category Full Time Student
34	Migration efficiency of wholly moving households whose household reference person is in NS-SEC category Not Classified
35	Migration efficiency of other moving groups whose household reference person is in NS-SEC category Not Classified
36	Migration efficiency of wholly moving households moving into or from owner occupied accommodation
37	Migration efficiency of other moving groups moving into or from owner occupied accommodation
38	Migration efficiency of wholly moving households moving into or from privately rented accommodation
39	Migration efficiency of other moving groups moving into or from privately rented accommodation
40	Migration efficiency of individuals living alone
41	Migration efficiency of individuals not living in a family but with others in a household
42	Migration efficiency of individuals who are part of a couple family
43	Migration efficiency of individuals who are part of a lone parent family
44	Migration efficiency of individuals living in a communal establishment

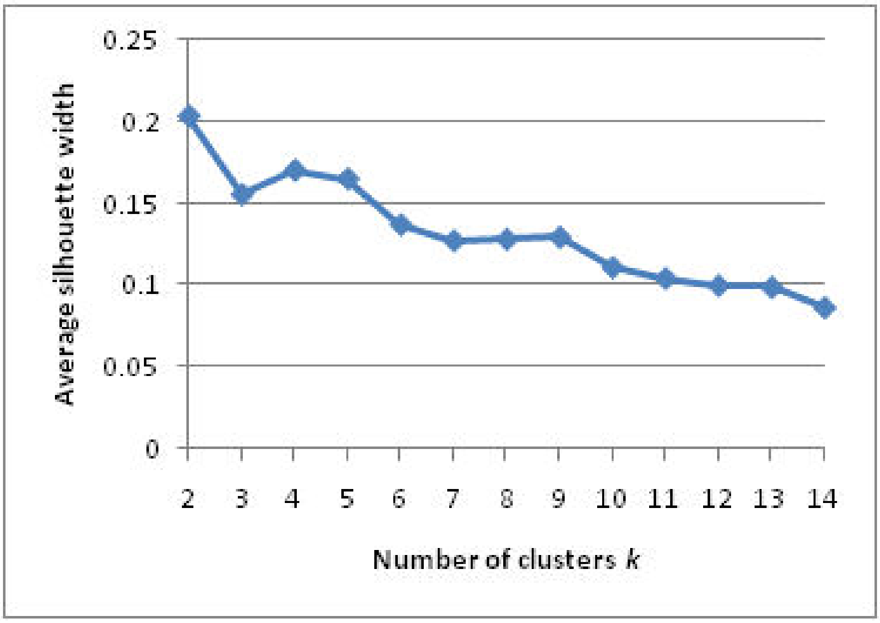


Figure 6.12: Average silhouette width values for solutions between 2 and 14 clusters

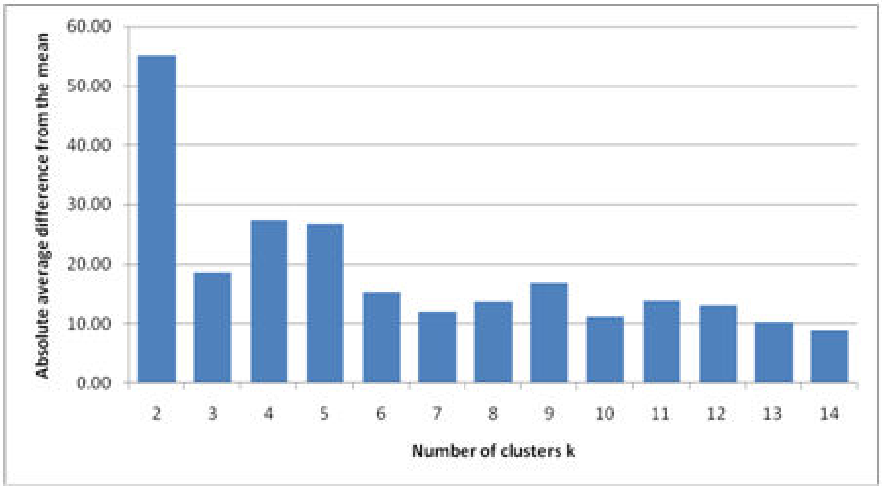
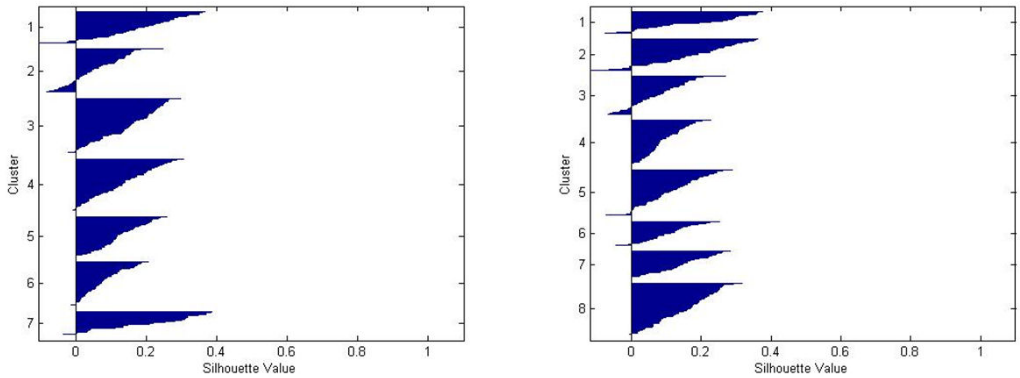


Figure 6.13: Absolute average difference from mean cluster size



(a) 7 Cluster Solution

(b) 8 Cluster Solution

Figure 6.14: Silhouette widths for 7 and 8 cluster solutions - *k*-means, Manhattan distance, 200 replicates

Taking Figure 6.12 first, a better cluster solution will have an average value closer to 1 than 0. Whilst Kaufman and Rousseeuw (2005) state that an average silhouette width <0.25 represents a poor cluster, Shepherd (2006) successfully employs the technique to assess cluster solutions with average values of around 0.1. It is evident that generally, as the number of clusters increases, the average silhouette value decreases, indicating at least for this metric, fewer clusters represent a more desirable solution. Taking Figure 6.13 also into consideration, however, a different conclusion might be reached. If we accept that more evenly sized groups are the most desirable outcome, then lower values in Figure 6.13 represent the better solution. Here clusters of 7, 8, 10, 13 and 14 groups could be candidates for selection.

Using both of these measures, it would appear that 3 clusters could be a good overall solution. A classification with only 3 clusters, however, is undesirable as fewer groups will represent much broader generalisations in the data. If somewhere between the 6 clusters suggested by Vickers (2006) and the 8 clusters suggested by the initial classification is aimed for, then Figure 6.13 suggests that solutions with either 7 or 8 clusters might be suitable as they both perform relatively well in both tests. With comparable scores in both metrics a decision between the two is a difficult one to make. The silhouette plots for both (Figure 6.14), enabling an assessment of the quality of each individual cluster, are also very comparable; where perhaps if one solution featured a cluster with a large negative spike (representing a number of cases which could be very easily associated with another cluster), it would be a clear candidate for being dropped. Here no such spikes are apparent, so additional data are required to assist the decision.

It could be argued that where cluster solutions have similar average silhouette values, as is the case here, the better solution would be the one with fewer values below 0. As Table 6.9 shows, whilst both cluster solutions have similar average silhouette values and similar counts of silhouette values below 0, the sum of the <0 silhouette values for 7 clusters is worse than it is for 8 clusters, indicating that where cases have weak associations with the clusters they have

Table 6.9: Summary of silhouette data for k=7 and k=8 cluster solutions

Cluster	**Cases in cluster**	**Count Silhouette <0**	**Sum Silhouette <0**	**Avg Silhouette**
1	73	1	-0.004	0.133
2	78	3	-0.028	0.146
3	45	4	-0.153	0.188
4	55	0	0	0.123
5	63	2	-0.033	0.08
6	33	2	-0.073	0.204
7	61	14	-0.433	0.065
7 cluster solution	**408**	**26**	**-0.723**	**0.128**
1	65	5	-0.066	0.119
2	39	4	-0.028	0.104
3	45	6	-0.238	0.172
4	37	1	-0.002	0.148
5	75	1	-0.004	0.152
6	53	8	-0.262	0.082
7	31	3	-0.037	0.197
8	63	0	0	0.098
8 cluster solution	**408**	**28**	**-0.637**	**0.13**

been assigned to, these weak associations are worse in a 7 cluster solution. Therefore taking all of this evidence into consideration - as well as the assertion by (Milligan, 1996, quoted in Shepherd, 2006) that where there is doubt, the higher figure should be taken - an 8 cluster solution will be chosen for the final classification.

As a postscript, it should be noted that this process was also carried out using Euclidean distance as the measure of distance between clusters. The main point of note, is that although the average silhouette widths were much higher when using Euclidean distance, the range of cluster sizes was also much higher - in some cases producing single case clusters with silhouette values of 1, or clusters with few cases, some heavily mis-specified (silhouette values very much in the negative). The work of Aggarwal et al. (2001) had already pointed to Manhattan distance providing better cluster solutions - the huge variation in cluster sizes and silhouette values using Euclidean distance confirms this.

6.5.2 The final cluster solution - an internal migration classification for Britain

A final k-means cluster run was carried out in MATLAB, this time using 1,000 replicates to ensure the final solution could be judged with certainty to be the best possible global cluster solution. As Figure 6.15 indicates, the final 1,000 replicates solution varies very little from the earlier 200 replicates solution in terms of the size and shape of the clusters. The clusters are in a different order to Figure 6.14, however, the only small differences occur in the cases with values <0. This is not a surprise as these cases, by their very nature, could very well be assigned to other clusters.

Figure 6.16 maps the final 8 clusters, revealing their spatial distribution across Britain.

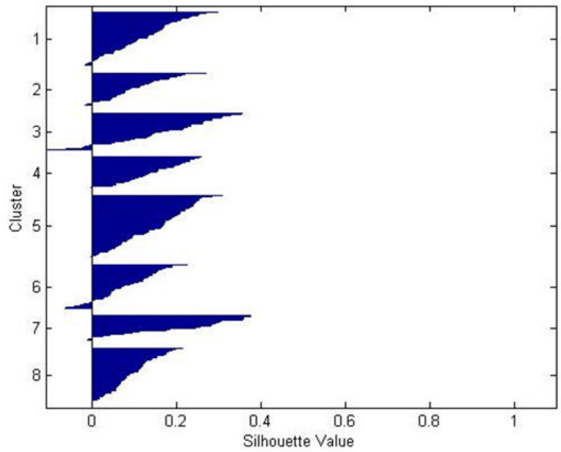


Figure 6.15: Silhouette plot of final 8 cluster solution - *k*-means, 1,000 replicates, Manhattan distance

Although the map gives the impression that each area featured is a firm member of whichever cluster its shading corresponds to, this is a little misleading. Indeed, this is a problem that besets all classifications of this type (whether the end user is aware of the issue or not). The trouble is the degree of membership each district has with the cluster to which it belongs. The silhouette plot in Figure 6.15 shows clearly that each cluster features cases with a greater or lesser degree of membership. This means that where clusters have particular characteristics, the districts within will correspond with these characteristics to a greater or lesser degree depending on the silhouette value.

The potential limitations of the categorical nature of classifications are readily apparent. If one is attempting to classify objects as either red or yellow, an orange object placed in either category will be incorrect to a greater or lesser extent. The solution in this situation would be that the orange object be given a degree of membership to either category - 60% red and 40% yellow, for example. Applying this theory to cluster analysis are a family of techniques known as 'fuzzy cluster analysis'. Within this family various algorithms have been developed designed specifically to create 'fuzzy' partitions in data where 'hard' or 'crisp' partitions might unnecessarily constrain cases to particular clusters. Höppner et al. (1999) give an overview of some of these including the 'fuzzy *c*-means' algorithm, which, given a number of clusters *c* to find in a dataset, will assign cases to clusters in a similar fashion to the *k*-means algorithm but with a membership grade or value determining the degree of membership to that cluster.

Whilst creating a set of fuzzy clusters is attractive as it avoids the rigid allocation that happens with hard clusters, one of the main aims of this exercise was to create a classification which could be used to analyse existing and future migration data. Analysis of flows between clearly defined areas is far more straightforward than the analysis of flows between areas with degrees of membership to a cluster (although the problem is not insurmountable). A parallel

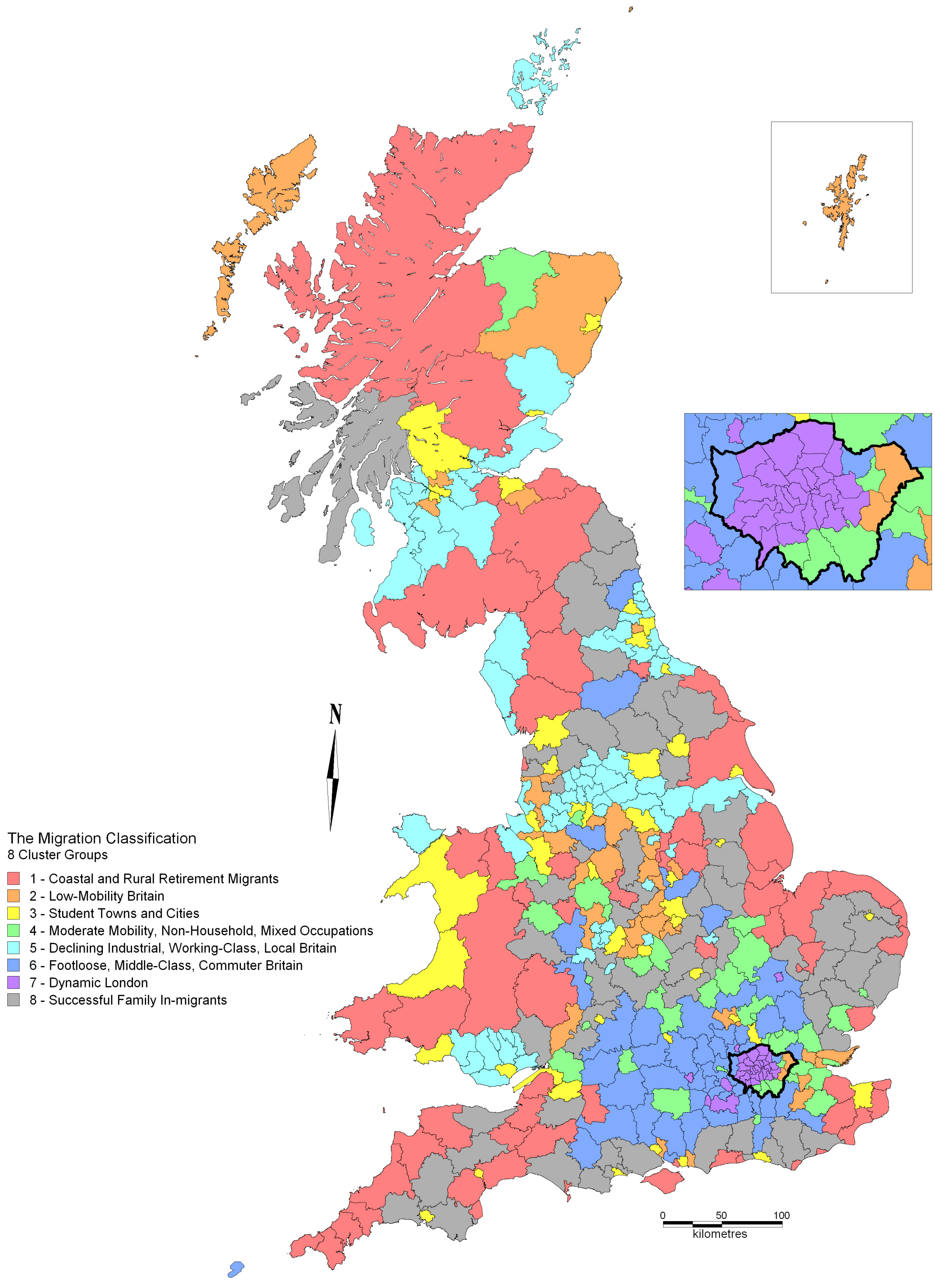


Figure 6.16: Internal Migration District Classification - 8 Cluster Solution

fuzzy classification will not be attempted here, although a degree of fuzziness can be added to the existing classification using the silhouette data already produced. The use of silhouettes in fuzzy clustering is advocated by Everitt et al. (2001) and can be easily applied to the rigid 8 cluster solution here so that cluster membership can be seen as stronger or weaker for each case in the cluster. As discussed earlier, the closer the silhouette value for a particular case is to 1, the more associated it is with that cluster; as values get close to 0 the membership becomes more ambiguous; closer the -1 and the case could more easily be associated with another cluster (although which cluster is not apparent).

Figure 6.17 represents a fuzzy version of the more rigid classification in Figure 6.16. Here the strength of membership is represented by the heaviness of the shading, with more heavily shaded areas having a stronger association with that particular cluster. Details of the silhouette values associated with each case in each cluster, as well as the profiles of each of the clusters in the classification will be presented in the next section.

6.6 Cluster Profiles

The following section outlines the constituent districts of each of the 8 clusters. The key variables defining each cluster are identified by the bars in the associated charts representing the z-scores for each variable. By taking the average z-score value for each variable across the districts comprising each cluster it is possible to ascertain which variables are more and less important within the cluster. The first graph in each cluster portrait contains z-scores for in, out and within area migration rates and should be interpreted with scores >0 showing over-representation of a variable in this cluster, and scores <0, under representation. Larger bars equal greater under or over representation. The second graph in each portrait contains z-scores for migration efficiency rates for moving groups. This graph should be interpreted differently as efficiency rates are directional. So for this graph, a value of 0 means that in/out migration is in balance. A value of >0 represents in-migration for a given variable, and a value of <0 represents out-migration. Larger bars equate to a greater intensity of movement.

It was decided that given the distinct profile of each cluster, a representative name should be chosen. A name which summarises the key features of a cluster will aid in its identification and differentiation from others - something which is important if the typology is to be used in additional analysis. The obvious issue with giving a cluster, characterised by variation across a range of variables, a name, is that any short name is likely to be a subjective generalisation. As Vickers (2006) points out, naming of clusters in geodemographic classifications is often a very contentious issue and can be open to much criticism. In the development of the OAC for the Office of National Statistics - a classification which has the status of a *'National Statistic'*(http://www.statisticsauthority.gov.uk/national-statistician/types-of-official-statistics) and thus is bound by an official code of practice - Vickers (2006) embarked upon an extensive quality assurance exercise through consultation with a range of stakeholders and experts.

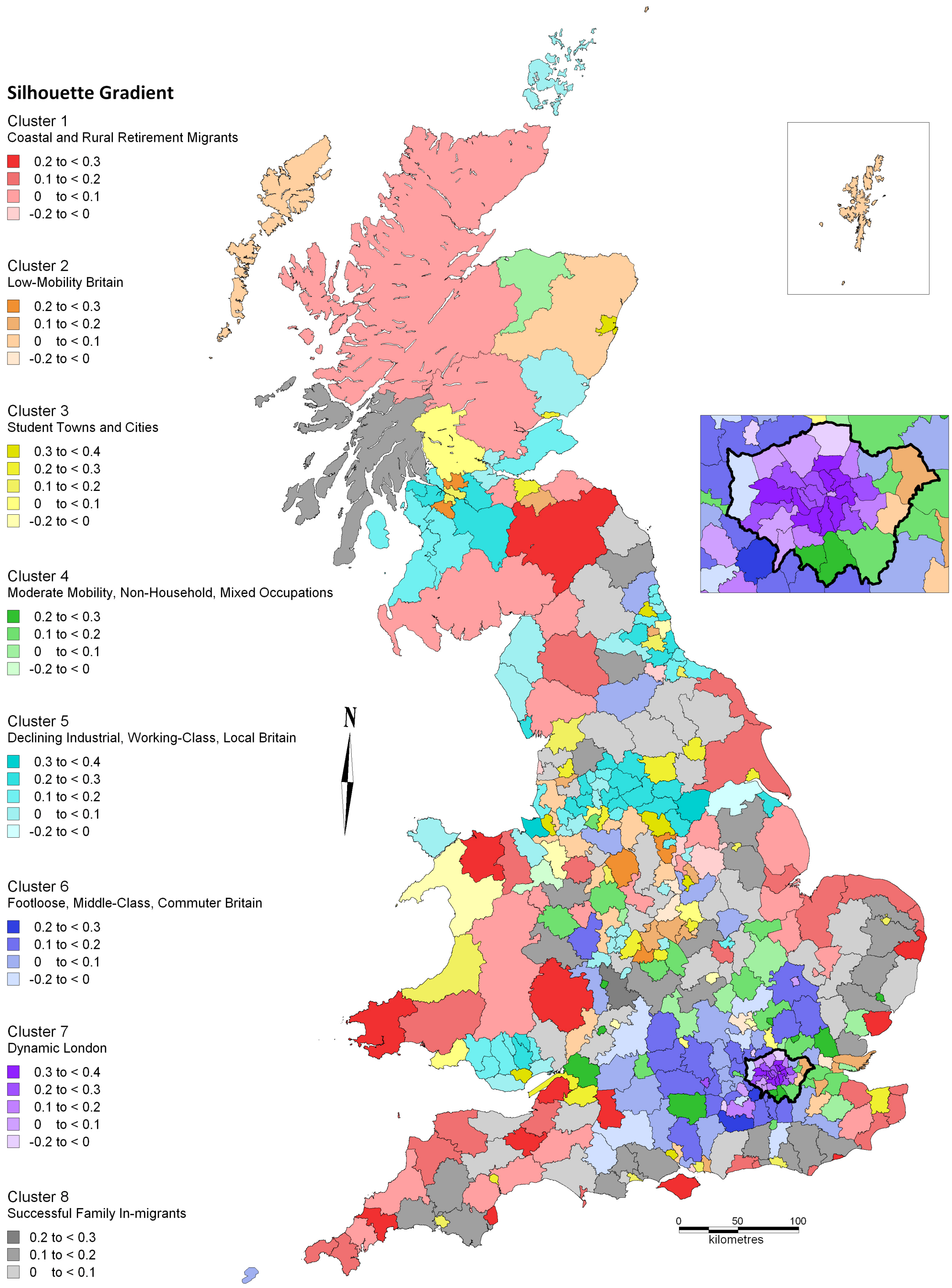


Figure 6.17: A ‘fuzzy’ representation of the internal migration district classification

As the Migration Classification is not governed by such statutes a slightly less exhaustive consultation exercise was undertaken, although that is not to downplay the importance of the exercise which was undertaken. An initial set of names for each cluster were decided upon using the information contained in these cluster profiles. This provisional set of cluster names was presented to a delegation of academics and other experts at an Economic and Social Research Council (ESRC) Census Programme workshop on Social and Spatial Classification (http://www.esds.ac.uk/news/eventdetail.asp?id=2455) where the cluster profiles below were presented along with their provisional names. Delegates were invited to critique existing names and offer alternative suggestions. At the end of the workshop, documents were collected and some minor alterations were made to the cluster names in the light of the suggestions made. The names presented below are the combined efforts of judgements made originally by this author and suggestions made during the quality assurance exercise.

6.6.1 Cluster 1: Coastal and Rural Retirement Migrants

Cluster 1 is dominated by coastal and rural areas, particularly in the South-West, Kent, Norfolk, the South Coast, Wales and Scottish Borders and Highlands. The Isle of Wight is the district most representative of this cluster, with Blackpool the district most unrepresentative. The cluster is characterised by in-migrants and within-area migrants in the older age groups - 45 and above. Younger in-migrants are very much underrepresented. Migrants into these areas are from across the socio-economic spectrum, although the very high socio-economic groups are less common. Migrants preferentially move into owner occupied accommodation, and tend to be either or alone or in couples, far more than parent families.

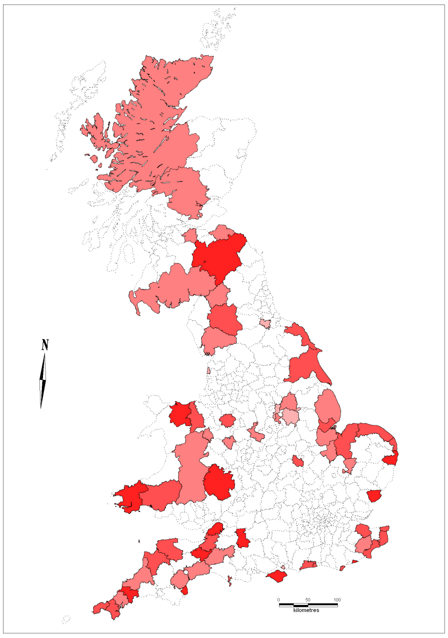
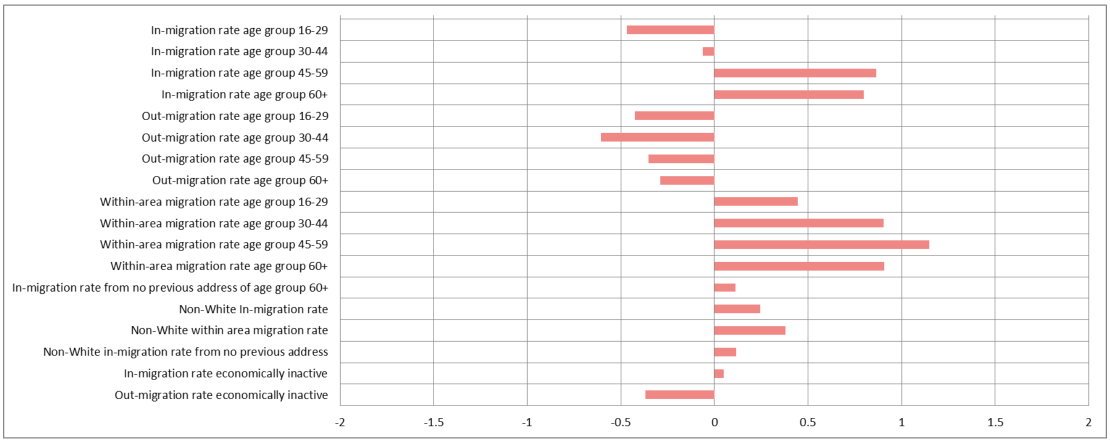
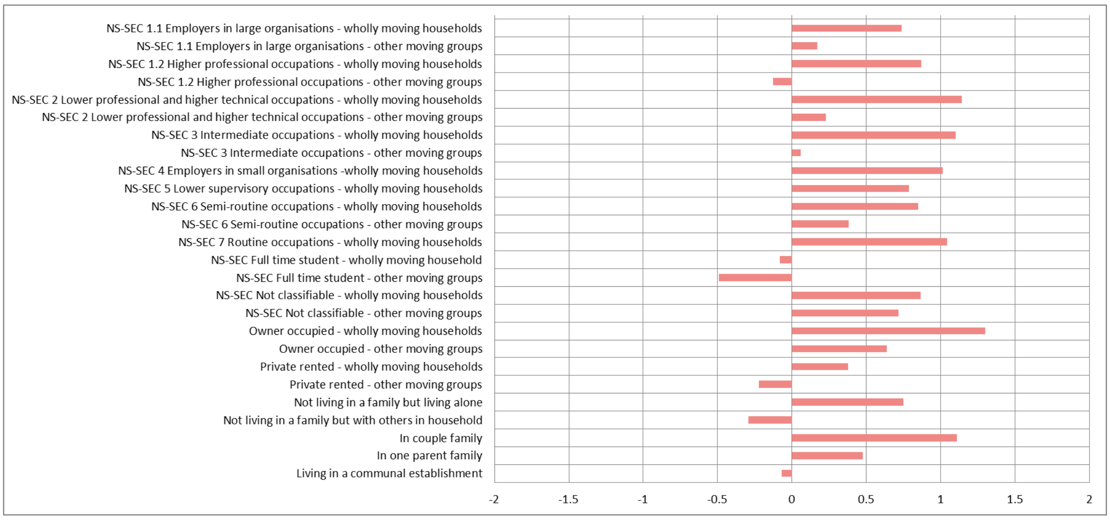


Table 6.10: Cluster 1 Silhouette Values

District	Silhouette Value	District	Silhouette Value	District	Silhouette Value
Isle of Wight	0.297	Worthing	0.162	Telford and Wrekin	0.073
Conwy	0.276	Scarborough	0.16	Ashford	0.067
Torbay	0.27	Arun	0.153	East Devon	0.053
Herefordshire County	0.253	Crewe and Nantwich	0.151	East Staffordshire	0.052
Waveney	0.252	Kings Lynn and West Norfolk	0.148	East Lindsey	0.05
Tendring	0.249	North Devon	0.146	Fenland	0.05
Scottish Borders	0.234	Carmarthenshire	0.135	Highland	0.045
Taunton Deane	0.23	Eden	0.132	West Lothian	0.028
North Somerset	0.228	Kerrier	0.131	Perth & Kinross	0.025
Hastings	0.212	Swale	0.131	Forest Heath	0.025
Pembrokeshire	0.208	Boston	0.114	Dumfries & Galloway	0.017
Restormel	0.202	Sedgemoor	0.109	Teignbridge	0.013
West Wiltshire	0.201	Dover	0.106	South Lakeland	0.012
Eastbourne	0.188	East Riding of Yorkshire	0.103	South Somerset	0.011
Penwith	0.182	North Norfolk	0.102	East Lothian	0.006
Torridge	0.177	South Holland	0.097	Carlisle	0.000
Thanet	0.173	North Cornwall	0.092	Darlington	-0.006
Great Yarmouth	0.171	Oswestry	0.091	Bolsover	-0.007
Kettering	0.171	Powys	0.085	Newark and Sherwood	-0.014
Denbighshire	0.171	Bassetlaw	0.079	Gosport	-0.016
Weymouth and Portland	0.167	Carrick	0.078	Blackpool	-0.023
Shepway	0.164	Ashfield	0.076		



z-scores for in, out and within district migration rates



directional *z*-scores for migration efficiency rates

6.6.2 Cluster 2: Low-Mobility Britain

Cluster 2 is spread around Britain, although small concentrations exist in the Midlands moving into south Merseyside, and to the south and east of London. North East Derbyshire is the most representative district in this cluster, with Erewash and South Bedfordshire most likely to be misclassified. The cluster is characterised by very little internal migration activity, with in-migration and out-migration under-represented across all age groups. Within-area migration is particularly under-represented. Where in-migration does occur it tends to be into owner occupied housing and by migrants in slightly higher socio-economic groups.

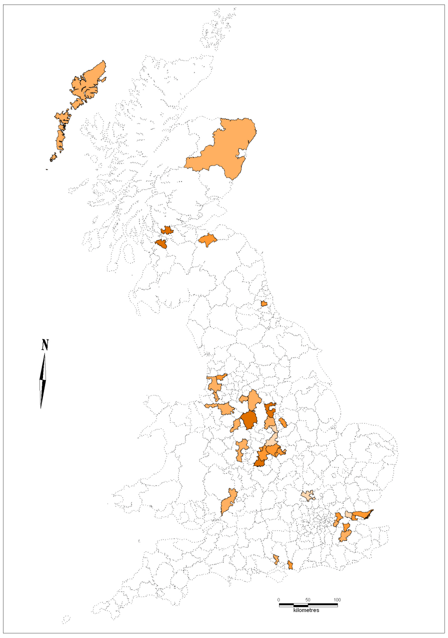
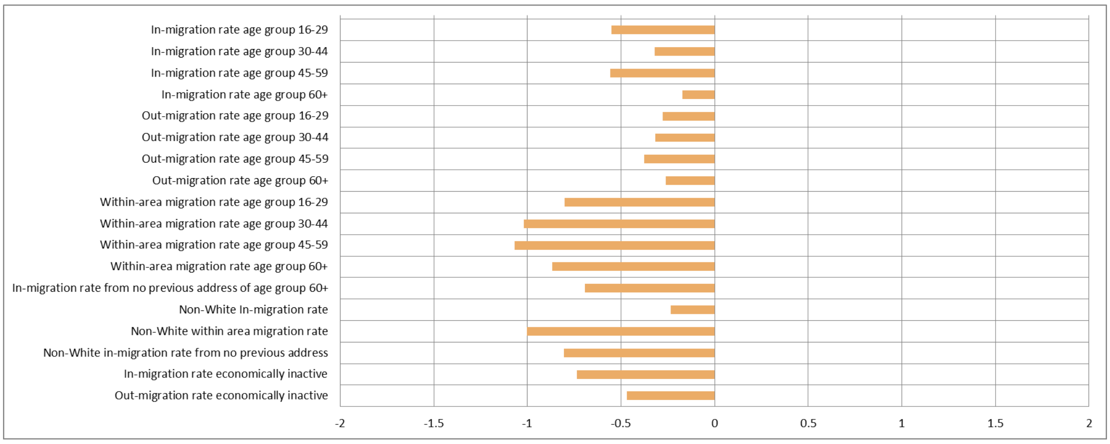
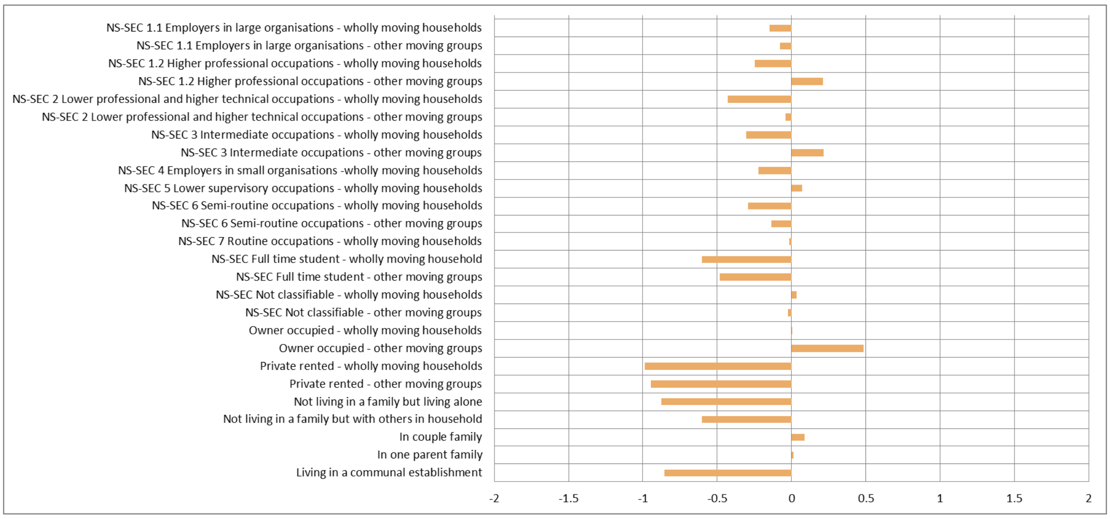


Table 6.11: Cluster 2 Silhouette Values

District	Silhouette Value	District	Silhouette Value	District	Silhouette Value
North East Derbyshire	0.27	Tamworth	0.133	West Lancashire	0.067
East Dunbartonshire	0.225	Rochford	0.122	Tonbridge and Malling	0.06
Solihull	0.219	Blaby	0.121	Ellesmere Port and Nes	0.057
East Renfrewshire	0.212	North Warwickshire	0.115	Forest of Dean	0.054
Staffordshire Moorland	0.201	Havant	0.111	Bexley	0.048
Hinckley and Bosworth	0.186	Newcastle-under-Lyme	0.092	High Peak	0.046
Havering	0.185	Knowsley	0.086	Shetland Islands	0.041
Castle Point	0.18	South Staffordshire	0.086	Eastleigh	0.038
Gedling	0.17	Eilean Siar	0.079	Vale Royal	0.016
Midlothian	0.158	Aberdeenshire	0.078	Basildon	-0.001
South Ribble	0.144	Oadby and Wigston	0.077	North West Leicestershire	-0.002
Chester-le-Street	0.139	Amber Valley	0.076	South Bedfordshire	-0.011
Gravesham	0.134	Stockport	0.068	Erewash	-0.015



z-scores for in, out and within district migration rates



directional z-scores for migration efficiency rates

6.6.3 Cluster 3: Student Towns and Cities

Cluster 3 is comprised principally of larger towns and cities housing universities and higher education institutions. Newcastle Upon Tyne is the most representative district, with Luton being the least representative. Despite a strong average silhouette value, signifying a well defined cluster, 6 districts including Luton have very weak associations. The cluster is characterised by high levels of student in-migration, and young person within-area migration. Non-household moving groups into privately rented accommodation are common in this cluster, as are non-family households and individuals moving into communal establishments - all characteristics of a student population. In addition, non-white within-area migration is important, as is in-migration of economically inactive migrants.

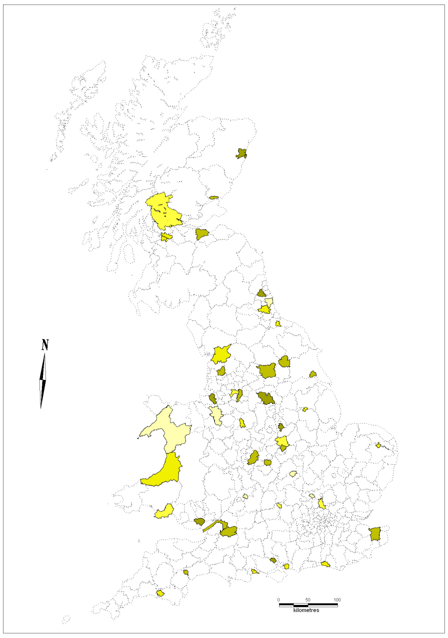
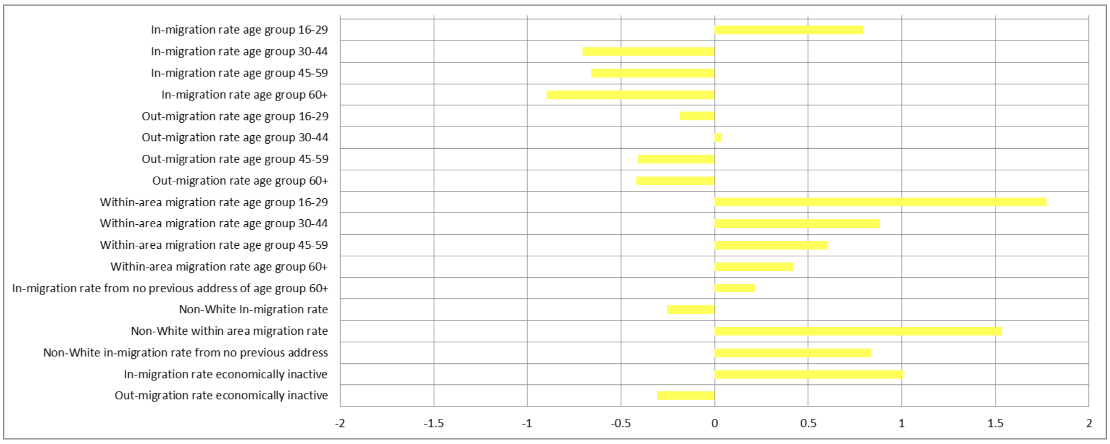
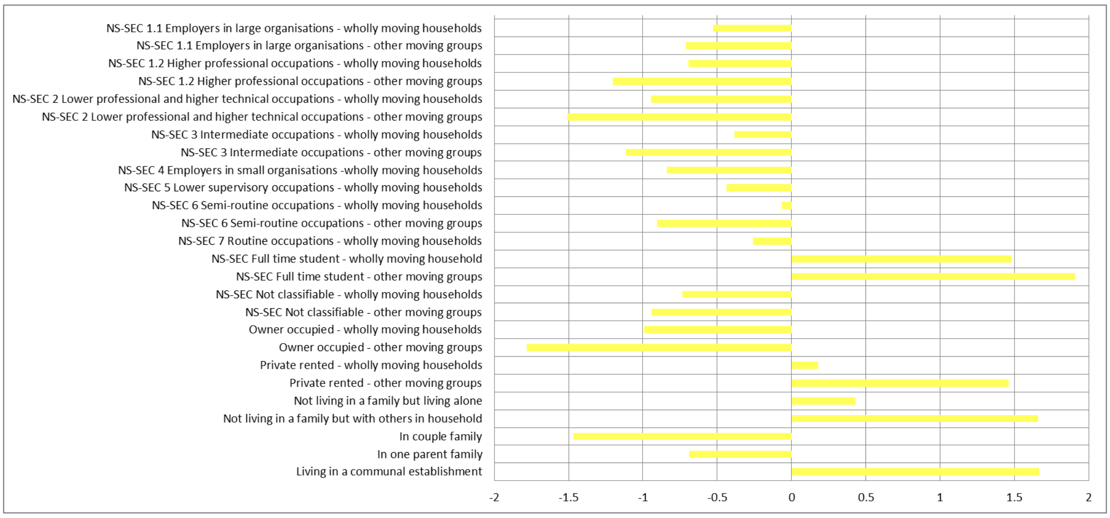


Table 6.12: Cluster 3 Silhouette Values

District	Silhouette Value	District	Silhouette Value	District	Silhouette Value
Newcastle upon Tyne	0.365	Manchester	0.235	Norwich	0.128
Sheffield	0.357	Coventry	0.233	Bournemouth	0.104
Cardiff	0.355	Exeter	0.226	Stoke-on-Trent	0.101
Nottingham	0.349	Edinburgh	0.225	Swansea	0.09
Southampton	0.329	Birmingham	0.218	Charnwood	0.077
Liverpool	0.31	Bath and North East Somerset	0.214	Stirling	0.074
Aberdeen City	0.306	Ceredigion	0.19	Oxford	0.052
Kingston upon Hull	0.294	Glasgow City	0.18	Salford	0.046
Dundee City	0.285	Middlesbrough	0.165	Welwyn Hatfield	0.002
Leicester	0.284	Plymouth	0.151	Gwynedd	-0.005
Leeds	0.271	Brighton and Hove	0.148	Sunderland	-0.016
York	0.264	Lancaster	0.143	Northampton	-0.022
Canterbury	0.261	Durham	0.14	Cheltenham	-0.026
Bristol	0.261	Portsmouth	0.137	Chester	-0.06
Preston	0.254	Lincoln	0.135	Luton	-0.109



z-scores for in, out and within district migration rates



directional z-scores for migration efficiency rates

6.6.4 Cluster 4: Moderate Mobility, Non-Household, Mixed Occupations

Cluster 4 is the second smallest cluster, with districts tending to be found in the south and Midlands. Whilst small, it is reasonably well defined, with only Wrexham having a noticeably ambiguous membership. Croydon is the district with the characteristics most representing this cluster. The cluster features relatively low to moderate levels of migration in general, although net in-migration is more prevalent than net out-migration. Migrants moving into these areas are more likely to be engaged in occupations across the socio-economic spectrum, however, net in-migration from migrants engaged in intermediate occupations is relatively high. Migrants who move alone or in non-family households are also more common in areas in this cluster. Wholly moving households and owner occupiers are less common.

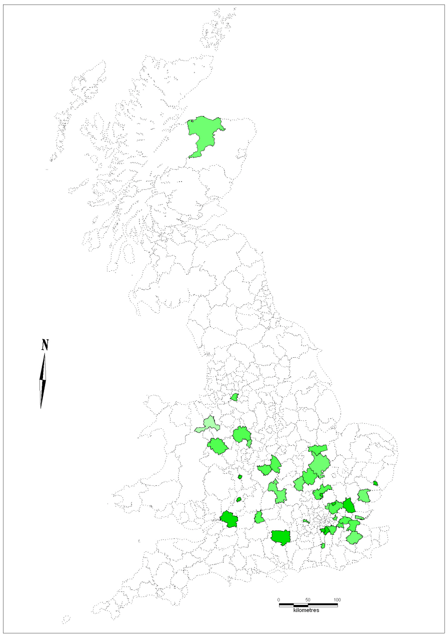
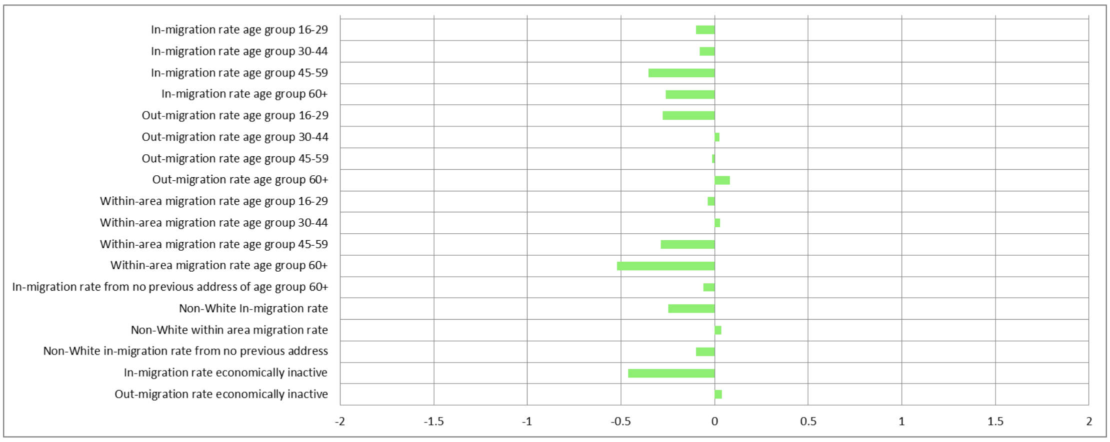
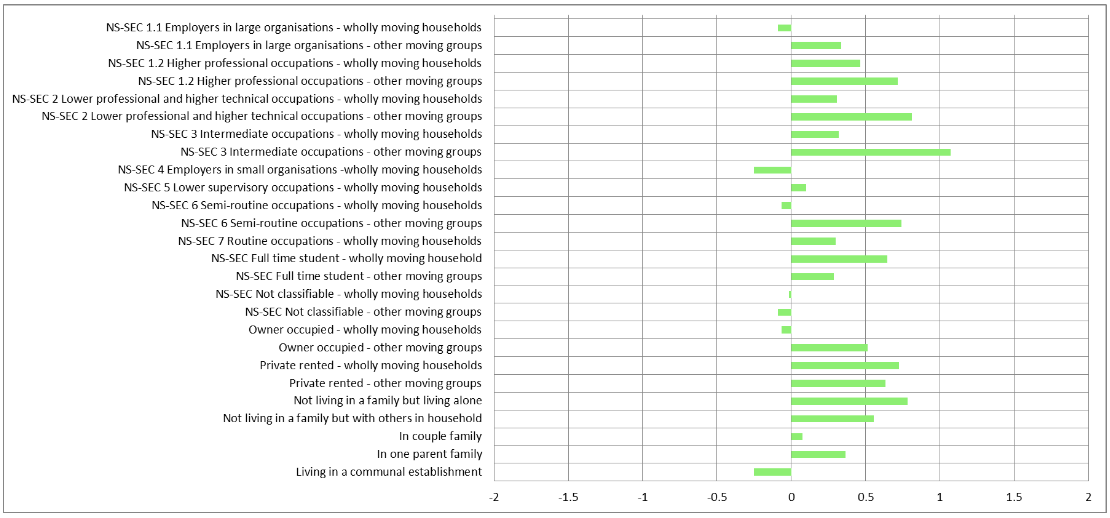


Table 6.13: Cluster 4 Silhouette Values

District	Silhouette Value	District	Silhouette Value	District	Silhouette Value
Croydon	0.26	Bedford	0.187	Rugby	0.101
Gloucester	0.253	Warwick	0.176	Southend-on-Sea	0.1
Sutton	0.242	Medway Towns	0.157	Maidstone	0.092
Stevenage	0.231	Stafford	0.157	Broxbourne	0.087
Harlow	0.231	Barking and Dagenham	0.157	Cherwell	0.076
Worcester	0.23	Shrewsbury and Atcham	0.152	Milton Keynes	0.063
South Gloucestershire	0.227	Dartford	0.138	Cannock Chase	0.057
Ipswich	0.221	Swindon	0.136	Huntingdonshire	0.056
Chelmsford	0.212	Slough	0.13	Colchester	0.053
Basingstoke and Deane	0.205	Trafford	0.126	Moray	0.048
Bromley	0.198	North Hertfordshire	0.12	Wrexham	-0.002
Thurrock	0.195	Epping Forest	0.12		
Peterborough	0.191	Crawley	0.109		

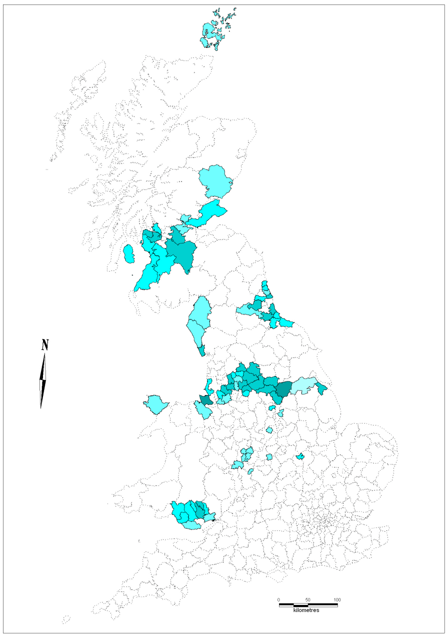


z-scores for in, out and within district migration rates



directional z-scores for migration efficiency rates

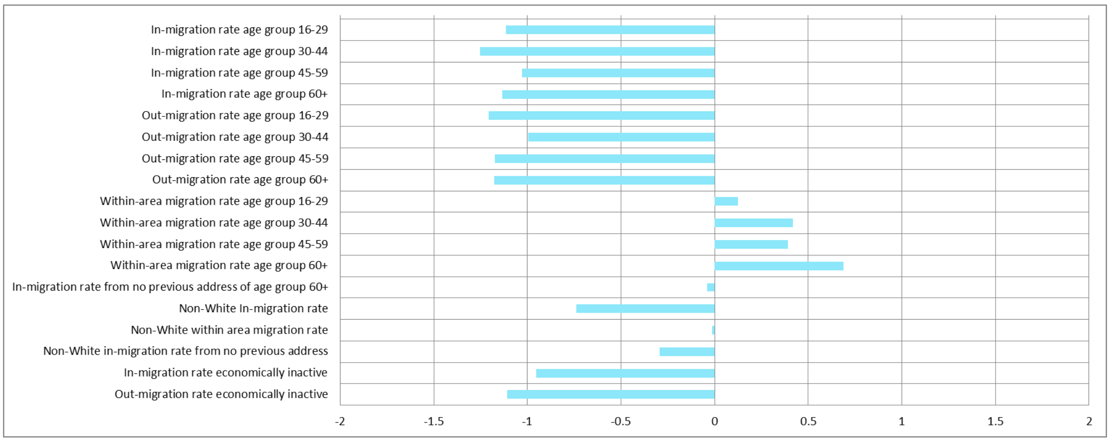
6.6.5 Cluster 5: Declining Industrial, Working-Class, Local Britain



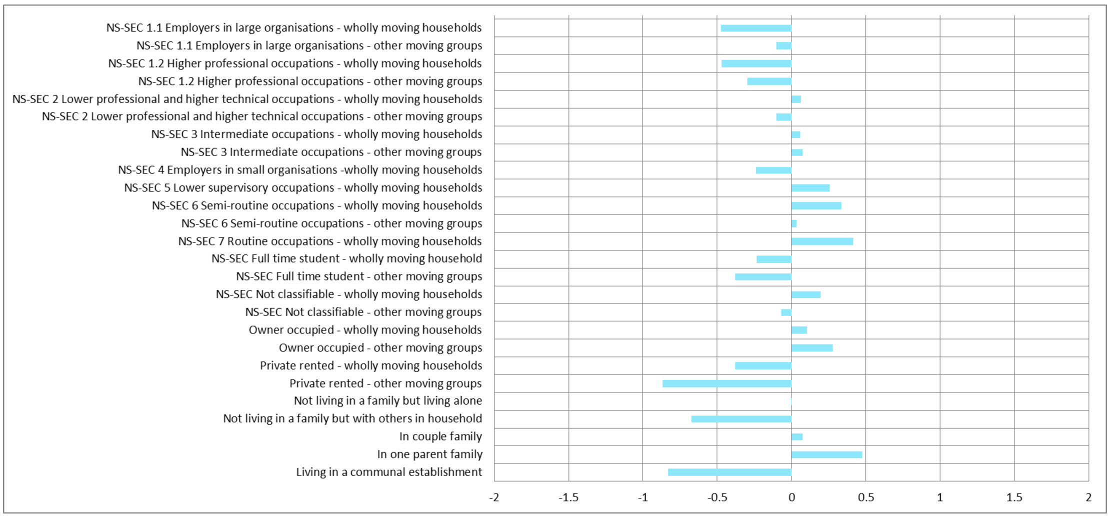
Cluster 5 is the largest cluster and is concentrated mainly around the ex-industrial areas of South Wales, Yorkshire, Greater Manchester and Lancashire, the North-East and Scotland. The cluster is also well defined with only North Lincolnshire having a silhouette value lower than 0. Districts in this cluster have very much below average in-migration and out-migration for all age groups, signifying a degree of isolation from the rest of the clusters in Britain. Shorter distance, within-area migration is slightly above average. Moves into these areas come from individuals in the lower socio-economic groups, with moves of one-parent families being above average. Moves of economically inactive individuals, however, are very much below average.

Table 6.14: Cluster 5 Silhouette Values

District	**Silhouette Value**	**District**	**Silhouette Value**	**District**	**Silhouette Value**
Doncaster	0.324	North East Lincolnshire	0.201	Dudley	0.096
Wirral	0.31	East Ayrshire	0.195	Angus	0.094
Bolton	0.283	Burnley	0.19	Allerdale	0.092
Rochdale	0.266	West Dunbartonshire	0.189	Mansfield	0.089
Bradford	0.26	Hyndburn	0.187	Wolverhampton	0.084
South Tyneside	0.254	Bridgend	0.181	Copeland	0.083
Barnsley	0.253	Rhondda Cynon Taff	0.178	Walsall	0.082
Blaenau Gwent	0.252	Redcar and Cleveland	0.174	Flintshire	0.071
Rotherham	0.249	Neath Port Talbot	0.172	Bury	0.069
Inverclyde	0.246	Wansbeck	0.172	Wyre Forest	0.068
Renfrewshire	0.245	Stockton-on-Tees	0.171	Nuneaton and Bedworth	0.066
Wigan	0.244	Torfaen	0.17	Newport	0.061
Barrow-in-Furness	0.24	Tameside	0.167	Sandwell	0.06
Caerphilly	0.238	St. Helens	0.167	Redditch	0.056
Oldham	0.238	North Ayrshire	0.165	Isle of Anglesey	0.054
Derwentside	0.234	South Ayrshire	0.161	Orkney Islands	0.045
South Lanarkshire	0.229	Sefton	0.15	The Vale of Glamorgan	0.044
Blackburn with Darwen	0.222	Blyth Valley	0.146	Halton	0.04
Wakefield	0.22	Hartlepool	0.134	North Tyneside	0.036
Pendle	0.219	Fife	0.133	Rossendale	0.026
Easington	0.21	Corby	0.129	Warrington	0.016
Sedgefield	0.21	Gateshead	0.129	Clackmannanshire	0.013
Kirklees	0.209	Merthyr Tydfil	0.122	Wear Valley	0.009
North Lanarkshire	0.206	Chesterfield	0.108	Derby	0.005
Calderdale	0.203	Falkirk	0.098	North Lincolnshire	-0.004



z-scores for in, out and within district migration rates



directional *z*-scores for migration efficiency rates

6.6.6 Cluster 6: Footloose, Middle-Class, Commuter Britain

Cluster 6 is the most poorly defined cluster, with the lowest average silhouette value. 8 districts out of the 58 have silhouette values lower than 0. The most representative district of the cluster is Waverley. In general, districts in this are concentrated just outside of London in the Home Counties and heading out west along the M3/M4 corridor. This cluster is characterised by higher rates of in and out-migration, particularly in the below 30 age groups. Within area migration is very much below average. Out-migration rates of economically inactive individuals are much higher than average. Net migration efficiency rates are very negative for lower socio-economic groups, but positive for those in higher groups.

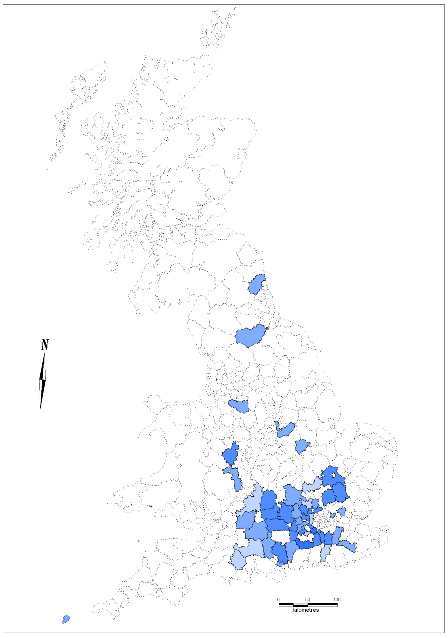
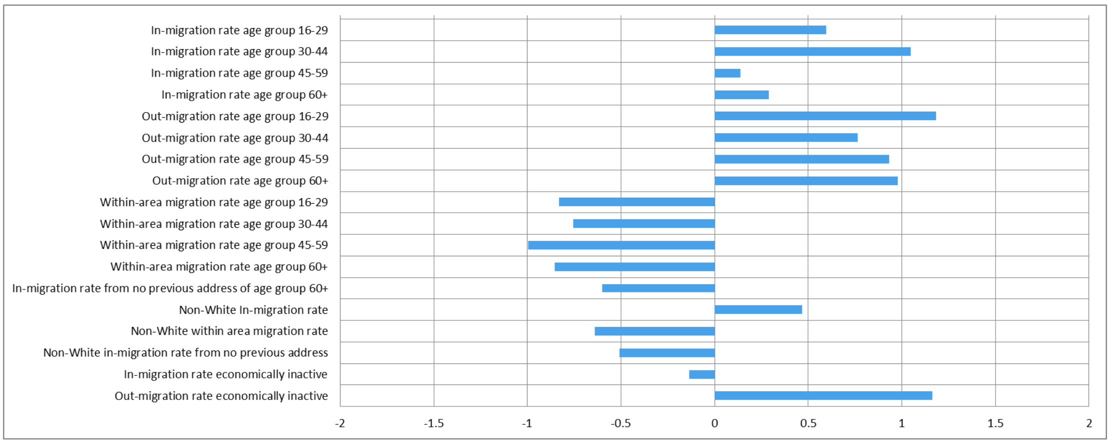
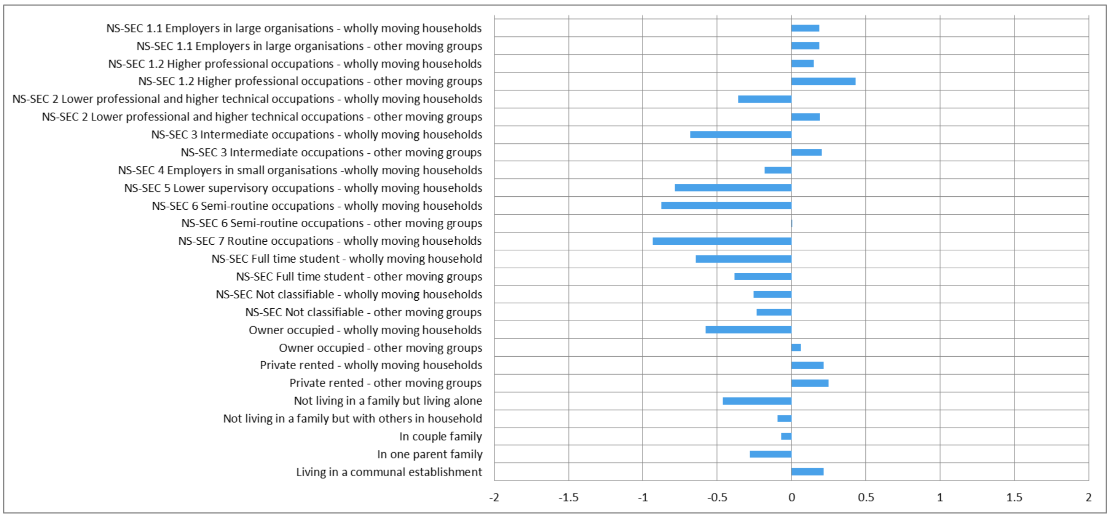


Table 6.15: Cluster 6 Silhouette Values

District	Silhouette Value	District	Silhouette Value	District	Silhouette Value
Waverley	0.249	South Cambridgeshire	0.121	Broxtowe	0.043
Elmbridge	0.225	West Oxfordshire	0.12	Redbridge	0.042
Hart	0.19	Chiltern	0.12	Richmondshire	0.037
Wokingham	0.19	East Hertfordshire	0.112	Aylesbury Vale	0.034
South Bucks	0.182	Three Rivers	0.11	Isles of Scilly	0.027
Epsom and Ewell	0.174	Brentwood	0.098	North Wiltshire	0.025
Winchester	0.164	St. Albans	0.089	Macclesfield	0.012
Mole Valley	0.164	Windsor and Maidenhead	0.076	Rutland	0.006
Uttlesford	0.155	Test Valley	0.074	Fareham	0.001
Surrey Heath	0.153	Malvern Hills	0.07	Woking	-0.005
Spelthorne	0.151	East Hampshire	0.069	Mid Sussex	-0.014
South Oxfordshire	0.148	Kennet	0.066	Cotswold	-0.018
Hertsmere	0.143	Rushcliffe	0.054	Dacorum	-0.028
Tandridge	0.143	Castle Morpeth	0.051	Mid Bedfordshire	-0.031
Vale of White Horse	0.142	Wycombe	0.048	North Dorset	-0.042
Reigate and Banstead	0.133	Tunbridge Wells	0.044	Hillingdon	-0.061
Bridgnorth	0.131	Sevenoaks	0.044	Salisbury	-0.063
West Berkshire	0.123	Bracknell Forest	0.044		

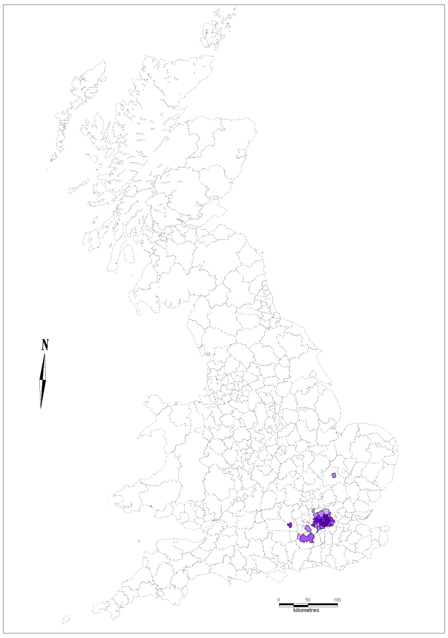


z-scores for in, out and within district migration rates



directional z-scores for migration efficiency rates

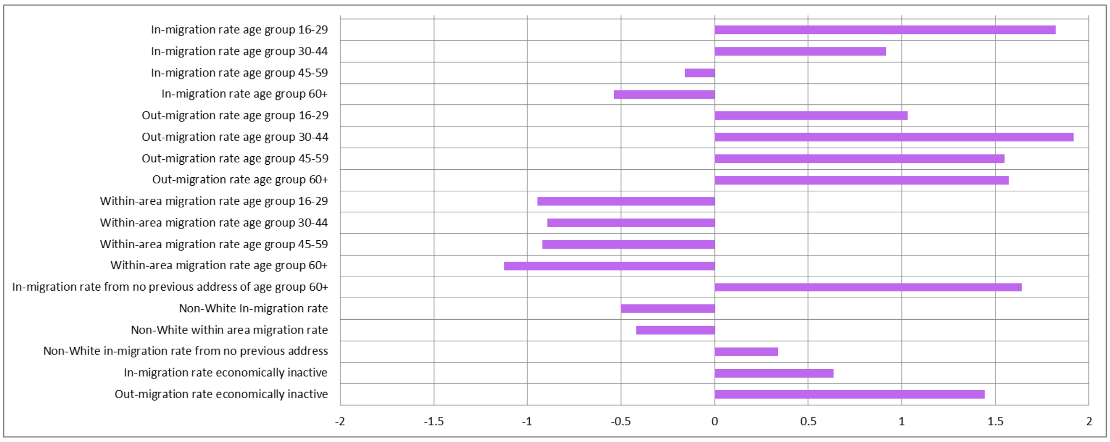
6.6.7 Cluster 7: Dynamic London



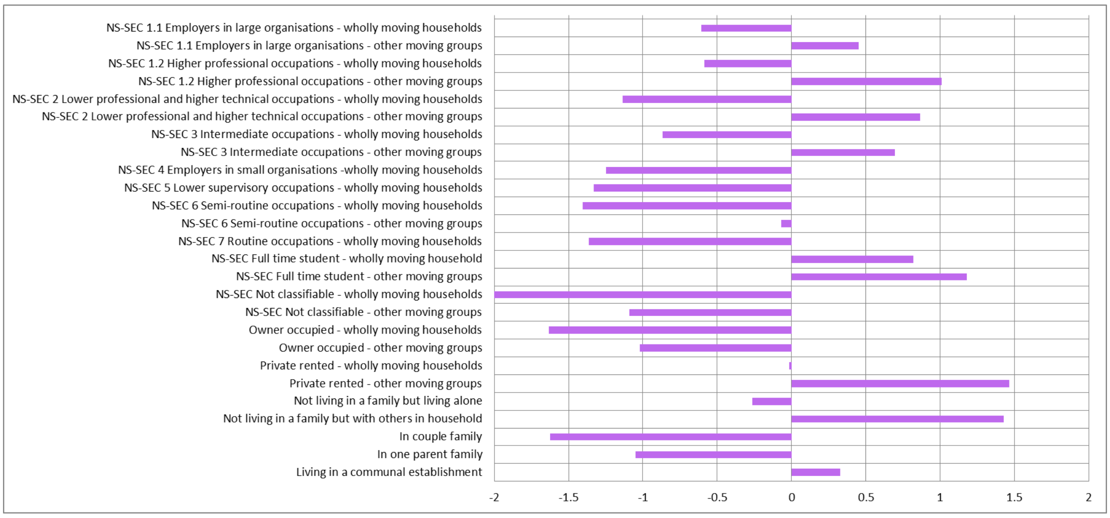
Cluster 7 is almost entirely concentrated within Greater London - only Cambridge, Reading, Guildford and Runnymede fall outside the M25. The cluster is defined by some of the highest and lowest z-score values, indicating it is the most dynamic cluster in the classification. It features very high rates of in-migration for the migrants under 30, but below average in-migration rates for those over 30. Out-migration rates are very high for all groups, but especially for those between 30 and 45. Within area migration rates are much below average, as are those of non-whites (except those with no-previous address). This cluster features the highest rates of movement of the economically inactive. Across the four highest socio-economic groups there are positive efficiency rates for other moving groups, but negative rates for wholly moving households, indicating if whole households move, they leave these areas, whereas non-households individuals tend to move in - especially into privately rented accommodation. Students are also an important group of in-migrants to this cluster. Families (both couples and single parents) are noticeably moving out of this cluster.

Table 6.16: Cluster 7 Silhouette Values

District	Silhouette Value	District	Silhouette Value	District	Silhouette Value
Islington	0.376	Lewisham	0.285	Richmond upon Thames	0.075
Haringey	0.375	Brent	0.281	Barnet	0.05
Hammersmith and Fulham	0.363	Hounslow	0.263	Cambridge	0.041
Southwark	0.358	Kensington and Chelsea	0.256	City of London	0.025
Wandsworth	0.358	Reading	0.245	Harrow	0
Lambeth	0.357	Westminster	0.226	Runnymede	0
Hackney	0.329	Merton	0.183	Rushmoor	-0.008
Tower Hamlets	0.307	Greenwich	0.177	Watford	-0.011
Ealing	0.301	Waltham Forest	0.111	Enfield	-0.018
Camden	0.294	Guildford	0.107		
Newham	0.292	Kingston upon Thames	0.094		



z-scores for in, out and within district migration rates



directional z-scores for migration efficiency rates

6.6.8 Cluster 8: Successful Family In-migrants

Cluster 8 is located mainly in rural areas of England, and is distributed quite evenly across the country. The cluster is relatively well defined as no districts have negative silhouette values. Wychavon is the most representative district in this cluster. In-migration of all age groups above 30 is above average, with importance increasing as age increases. Within-area migration is less significant. There are positive, in-migration balances across all socio-economic groups, although there are noticeable out-migrations of students from this cluster. When migrants move into this cluster, the preference is to move into owner occupied accommodation, with couple families being more important than any other group.

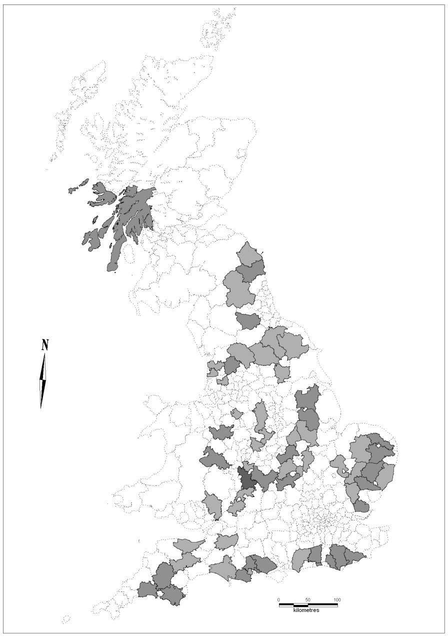
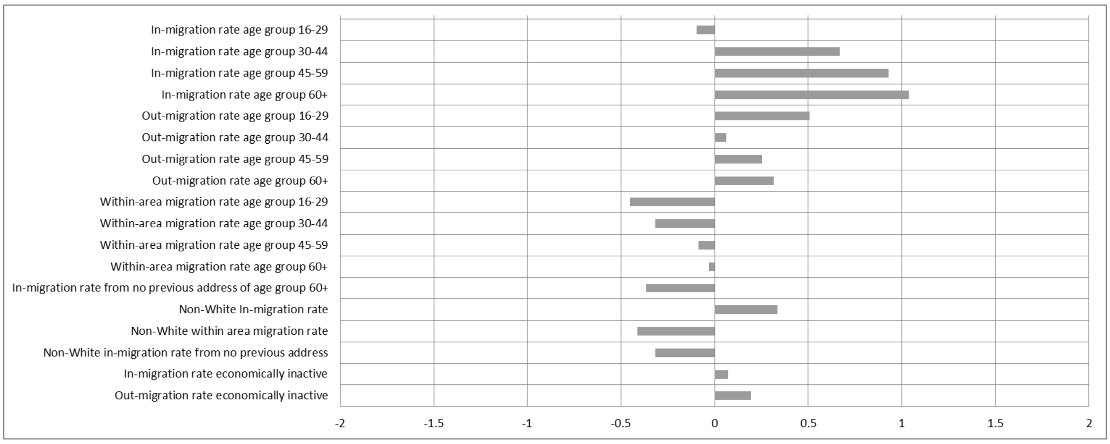
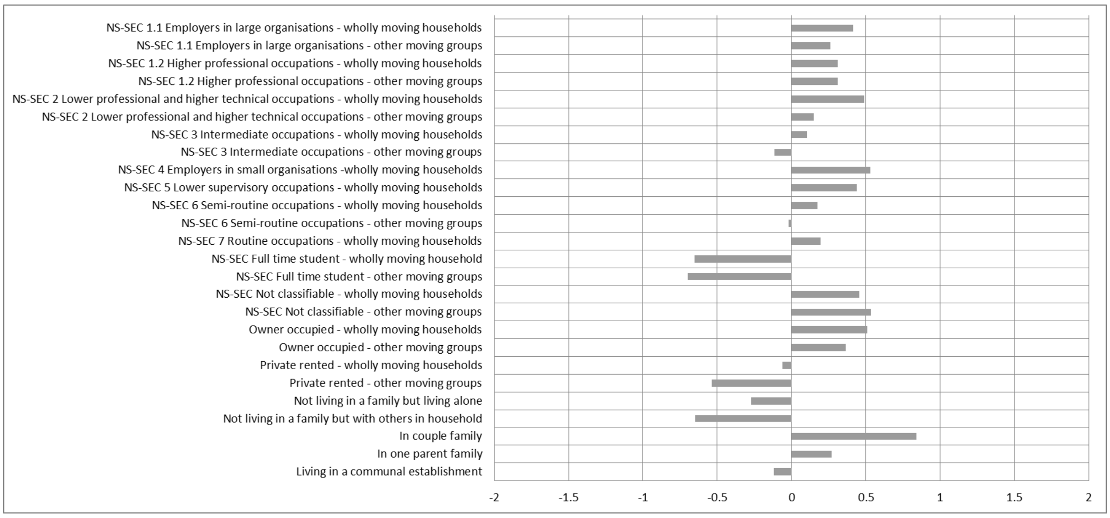


Table 6.17: Cluster 8 Silhouette Values

District	Silhouette Value	District	Silhouette Value	District	Silhouette Value
Wychavon	0.215	Ribble Valley	0.12	Tynedale	0.066
Alnwick	0.198	Teesdale	0.118	Congleton	0.064
West Lindsey	0.197	South Northamptonshire	0.118	Braintree	0.062
South Norfolk	0.189	Stratford-on-Avon	0.114	Poole	0.062
Harborough	0.181	Adur	0.112	South Kesteven	0.06
Christchurch	0.178	North Shropshire	0.11	Tewkesbury	0.059
South Hams	0.176	Horsham	0.11	Berwick-upon-Tweed	0.057
South Shropshire	0.174	Maldon	0.108	Derbyshire Dales	0.056
East Dorset	0.171	Bromsgrove	0.099	Wellingborough	0.056
West Devon	0.159	Craven	0.096	Chorley	0.05
Mid Suffolk	0.154	Monmouthshire	0.09	Mendip	0.045
Babergh	0.153	West Somerset	0.086	Suffolk Coastal	0.044
Argyll & Bute	0.145	East Cambridgeshire	0.082	Lichfield	0.033
Wealden	0.142	Chichester	0.082	Wyre	0.031
Rother	0.139	Hambleton	0.081	Selby	0.028
Lewes	0.13	West Dorset	0.078	Mid Devon	0.025
Caradon	0.128	Daventry	0.078	East Northamptonshire	0.02
Broadland	0.127	St. Edmundsbury	0.075	Stroud	0.016
New Forest	0.124	Breckland	0.07	South Derbyshire	0.016
Purbeck	0.121	Ryedale	0.069	Melton	0.015
North Kesteven	0.12	Fylde	0.068	Harrogate	0.011



z-scores for in, out and within district migration rates



directional z-scores for migration efficiency rates

6.7 Classification evaluation and comparison

Now that a final classification of districts has been achieved, the last stage in the classification building process is to evaluate the final solution. There are a number of ways in which the results of a classification can be examined and tested, all dependent on exactly what the classification is being tested for. For example, it may be desirable to test that the partitions created are robust and represent actual clusters rather than being artefacts of particular algorithms - as in the example earlier where the *k*-means algorithm as implemented in SPSS could produce completely different classifications from the same data, merely through re-ordering the objects being clustered. Alternatively the test may be to see whether a classification is comparable or different to an existing classification - for example, that a migration classification offers something different to a general purpose classification. Or it may be that one wants to test a classification to see whether it is more successful at predicting behaviours than chance - particularly useful in marketing contexts where classifications are used for customer targeting. It could be that the classification needs to be tested to confirm that the variables selected were indeed most appropriate and important for the final solution. Or perhaps it may be that one wishes to assess how well the classification represents the real-world.

All of these reasons for testing classifications are valid in particular contexts, however, it may not be necessary or appropriate to test for all (or indeed any) of them all of the time. So the question that arises is what is the most suitable way of assessing the Migration Classification? The process of variable and algorithm selection was very thorough in this classification, as already described. The *k*-means algorithm as it was implemented in MATLAB means that there can be confidence that the partitions created are robust, given the variables selected for inclusion. Vickers (2006), in creating the OAC from the 2001 Census, employed several different techniques to assess and 'quality assure' his classification. One of these methods to which much time was devoted was sensitivity analysis. In sensitivity analysis, variables are selectively removed from the classification and the algorithm run again. Through examining the change in the average distance to the cluster centre for objects in each cluster, an appreciation of the impact each variable has on the classification can be ascertained, therefore pointing to whether it was wise to include that variable in the first place (in theory). Interestingly, after extensive sensitivity analysis, Vickers (2006, p.173) concludes that *"as long as the reasoning for the original variable selection was sound, removing a variable from the analysis cannot really be justified"*. Vickers argues that even where variables have an apparent small effect on the clusters in a classification, removing them may not be the most sensible solution as whilst they may have a small effect overall, they may be *"vital to the formation of an individual cluster"* (Vickers, 2006, p.173). Thanks to the work of Vickers, it is therefore possible to conclude that sensitivity analysis is not something which will be necessary for the Migration Classification - the original variables were chosen carefully and because they were deemed to add value to the classification. Removing variables will reduce the amount of information present in the classification - something which is important where one benefit of the Migration Classification could be to add value to the attribute poor data available between censuses.

Vickers (2006) also devotes a significant amount of time to assessing the reduction in variability within individual variables afforded by the clusters within the OAC; the idea being that the better the classification, the greater the reduction in variability there will be within each variable. The rationale for this type of validation for the OAC was that work by Voas and Williamson (2001) indicated that with very few variables, an ad hoc general-purpose classification system can offer much of the discriminatory power that a more carefully constructed classification can - i.e. a similar level of reduction in the variability of variables can be achieved. Vickers demonstrated that this was not necessarily the case with the OAC, and where this type of validation was needed to challenge the assertions of Voas and Williamson when constructing a general-purpose classification, a similar exercise is not necessary here. The variables selected for the Migration Classification situate it very much as a 'purpose-built', bespoke framework for analysing migration data, so there is not the need to justify its existence as other, similar classifications do not exist.

The Leeds Classification for Community Safety (LCCS) Shepherd (2006), a bespoke classification more akin to the Migration Classification than the OAC in that its intended use was very specific, employed a very different validation technique to those used by Vickers (2006). Shepherd constructed a series of 'gains charts' to assess relative advantages of the LCCS over general purpose classifications for predicting particular types of crime - a comparison technique also used by See and Openshaw (2001). This type of validation was appropriate, given that the principal purpose of the classification was to improve community safety through predicting patterns of crime. A similar type of analysis is not necessary or appropriate for the Migration Classification. Firstly, it is not the intention that the classification be used for predicting behaviours in the same way that a small area geodemographic classification might - rather the Migration Classification is intended for use primarily as a monitoring and complexity reduction tool. Secondly, where the Migration Classification is based entirely on migration variables, and other general purpose district classifications exclude migration variables, it would be illogical to test to see whether the Migration Classification is better at predicting migration behaviours than classifications which largely exclude migration variables.

Another common method used in the evaluation of small area geodemographic classifications is 'ground-truthing' (Vickers and Rees, 2009). This involves assessing the final solution by examining some of the small areas present in the different cluster groups to see if the real world situation bears any resemblance to that described by the classification group to which the area is assigned. Whilst this approach has some obvious flaws (the extent to which an entire cluster covering many areas can be validated through examining the physical environment of a small number of places is debateable), it is an approach that can only be adopted successfully where the areas being classified are relatively small and self contained (output areas or unit postcodes for example). Ground truthing areas where the smallest spatial unit is a local authority district is not feasible.

So as none of the above approaches to validation appear to be appropriate, an alternative method of validation is necessary. Referring back to the original rationale for developing the Migration Classification, one of the reasons for creating it was that in the context of monitoring migration between censuses, it was argued that migrants do not necessarily exhibit the characteristics of the underlying population and therefore a general purpose classification does not make sense if studying migration flows. It follows, therefore, that an appropriate method of validation would examine this assertion - are we are getting something new from this classification? Is it really an entirely new way of classifying districts, or merely a surrogate for one of the other available district classifications? And whilst sensitivity analysis, whereby individual variables are removed from the classification to assess their individual impact on the final cluster solutions, has been dismissed for reasons already explained, one interesting alternative would be to assess the impact of removing whole groups of variables from the classification and comparing the results with the original to see how different the cluster solutions are (i.e. how different are the clusters created if all socio-economic, or housing or age variables are dropped?). Could fewer variables have produced a similar solution? For example could a classification based entirely on age variables (where age will also be a feature of any migrant exhibiting another feature such as ethnicity) produce a result very similar to the final classification arrived at here? If it does, then this will tell us as much about the relationship between migration variables as it does about the validity of the classification.

6.7.1 Mathematical methods for comparing clusters

A number of techniques have been developed to compare the results of different classification solutions where the same objects have been clustered differently. However, broadly speaking, they all operate similarly in that they assess the extent to which two different classification solutions agree and provide a statistic which quantifies the strength of this agreement. As noted by Everitt et al. (2001), one straight-forward way of measuring the association between two solutions with equal number of clusters is through calculating either a simple percentage agreement or Cohen's Kappa coefficient.

All cluster comparison solutions work on a contingency table of cluster agreement - a cluster 1 x cluster 2 matrix, termed $N = n_{ij}$. Consider Tables 6.11a and 6.11b below which represent a hypothetical dataset and related contingency table (notation adapted from Hubert and Arabie 1985, example adapted from Yeung and Ruzzo 2001)

The similarity between the two classifications u and v can be calculated as the average similarity of u to v and v to u, where:

the similarity of u to v =

$$\sum_{j\max} n_{ij}\left(\frac{100}{N}\right) \tag{6.5}$$

Table 6.18: Dataset and related contingency table for comparing cluster solutions

(a) Dataset containing 10 objects to cluster and two different classification solutions - class u and cluster v

object	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**J**
class u	1	1	2	2	2	2	3	3	3	3
cluster v	1	2	1	2	2	3	3	3	3	3

(b) Contingency table n_{ij} representing the agreement between the two classifications u and v

class/cluster	v**1**	v**2**	v**3**	**sum**	**max**	**%**
u**1**	1	1	0	2	1	0.2
u**2**	1	2	1	4	2	0.4
u**3**	0	0	4	4	4	0.4
sum	2	3	5	10	7	1
max	1	2	4	7		
%	0.2	0.3	0.5	1		

and the similarity of v to u =

$$\sum_{i\max} n_{ij}\left(\frac{100}{N}\right) \tag{6.6}$$

which in the case of the contingency table shown in 6.18b would be:

$$7\left(\frac{100}{10}\right) = 70\% \tag{6.7}$$

for both u to v and v to u

Cohen's kappa coefficient k (Cohen, 1960) uses the contingency table in a slightly different way, this time considering the probability of random agreement between the two classifications. In doing this it can be seen as a more robust measure than the percentage agreement between the two solutions. Cohen's kappa can be calculated thus:

$$k = \frac{Pr(a) - Pr(e)}{1 - Pr(e)} \tag{6.8}$$

Where $Pr(a)$ is the observed percentage agreement between class u and cluster v (or v and u), and $Pr(e)$ is the probability of random agreement. This is the max column and row in 6.18b, but perhaps easier to see in Table 6.18a, where the two partitions u and v agree once for partition 1, twice for partition 2 and four times for partition 3, so:

$$Pr(a) = \frac{1+2+4}{10} = 0.7 \tag{6.9}$$

The probability of random agreement can be calculated as the sum of the probability of being in one partition in one cluster solution multipled by the probability of being in the same partition in another cluster solution. E.g. the product of the probability of being in $u1$ and $v1$. Or, using 6.18b:

$$Pr(e) = (u1 \times v1) + (u2 \times v2) + (u3 \times v3) \tag{6.10}$$

$$Pr(e) = (0.2 \times 0.2) + (0.4 \times 0.3) + (0.4 \times 0.5) \tag{6.11}$$

therefore k = 0.53

The range of k is between 0 and 1; 1 representing complete agreement between the classifications, 0 representing no agreement. In the case of this example, we might interpret that there is moderate agreement between the classifications.

Using metrics such a percentage agreement and Cohen's Kappa are fine when the two solutions being compared have the same number of clusters, these techniques fall down where the numbers of cluster differ. This is important in the context of evaluating the Migration Classification, as other district level general purpose classifications such as the ONS classification of local authorities and the Vickers et al. classification have different numbers of clusters. Where the number of clusters differ, the Rand index (Rand, 1971), can be used as *"it is based on the agreement or otherwise of every pair of n objects"* (Everitt et al., 2001, p.182) rather than a cross-tabulation of the frequencies of agreement.

Given two different partitions of the same objects (class u and cluster v), the Rand index R can be described as:

$$R = \frac{(a+b)}{(a+b+c+d)} \tag{6.12}$$

where:

a = the number of pairs of objects in the same class u and the same cluster v

b = the number of pairs of objects not in the same class u or the same cluster v

c = the number of pairs of objects that are in the same class u and a different cluster v

a = the number of pairs of objects that are in a different class u and the same cluster v

Some have criticised the Rand index as it varies, increasing as the number of clusters being compared increases. It is also does not take into consideration the chance agreement between the cluster allocation. Consequently, Hubert and Arabie (1985) proposed an adjusted Rand index which deals with these problems, with the form as follows:

$$AR = \frac{2\,(ab - cd)}{((a+d)\,(b+d) + (a+c)\,(b+c))} \tag{6.13}$$

Where the Rand index alone may cause problems in comparing classifications with differing numbers of clusters, here both the Rand index and the adjusted Rand index (Hubert and Arabie, 1985) will be computed to compare the solutions. To exemplify the calculation of the Rand and adjusted Rand indices for the example data in Table 6.18a, consider Table 6.19:

so for 6.19:

$$R = \frac{(7+25)}{(7+25+6+7)} = 0.71 \tag{6.14}$$

and

$$AR = \frac{2 \times ((7 \times 25) - (6 \times 7))}{((7+7) \times (25+7) + (7+6) \times (25+6))} = 0.31 \tag{6.15}$$

Similarly to Cohen's Kappa coefficient, both Rand indices range between 0 and 1, 0 indicating no agreement between the two matrices and 1 indicating total agreement. However the adjusted Rand index also gives an indication of how likely the agreement is by chance. For the example used here, the Rand index indicates relatively high agreement between the two indices, with the adjusted Rand index indicating a similarly high probability that the agreement between the two matrices is by chance (0 indicating greater likelihood of chance). Given the ability of Rand and adjusted Rand indices to compare classifications with different numbers of clusters, it is these measures that will be adopted to compare the Migration Classification with other district level classifications.

6.7.2 Comparison with other district level classifications

As was described in Chapter 3, there are two main district level classifications available for the UK: the ONS classification of local authorities (ONS, 2004), and the Vickers et al. (2003) local authority district classification. Other district level classifications do exist, such as the DEFRA rural/urban classification (DEFRA, 2009). But this particular classification only covers districts in England, so will not be used in the comparison with the Migration Classification. Both of these general purpose classifications are hierarchical, with three tiers of clusters, so the Migration Classification will be compared with each tier.

Table 6.20 reveals the results of the comparisons between the classifications using both the Rand and adjusted Rand indices. The first three examples compare the Migration Classification with the three tiers of the Vickers et al. classification. As might be expected with the Rand index, as the number of clusters increases through the Vickers et al. hierarchy, so too does the

Table 6.19: Pairs of objects and associated cluster linkages for calculating Rand and adjusted Rand indices

		a	**b**	**c**	**d**
Object pair		**same** *u* **same** *v*	**not** *u* **not** *v*	**same** *u* **not** *v*	**same** *v* **not** *u*
A	B	-	-	1	-
A	C	-	-	-	1
A	D	-	1	-	-
A	E	-	1	-	-
A	F	-	1	-	-
A	G	-	1	-	-
A	H	-	1	-	-
A	I	-	1	-	-
A	J	-	1	-	-
B	C	-	1	-	-
B	D	-	-	-	1
B	E	-	-	-	1
B	F	-	1	-	-
B	G	-	1	-	-
B	H	-	1	-	-
B	I	-	1	-	-
B	J	-	1	-	-
C	D	-	-	1	-
C	E	-	-	1	-
C	F	-	-	1	-
C	G	-	1	-	-
C	H	-	1	-	-
C	I	-	1	-	-
C	J	-	1	-	-
D	E	1	-	-	-
D	F	-	-	1	-
D	G	-	1	-	-
D	H	-	1	-	-
D	I	-	1	-	-
D	J	-	1	-	-
E	F	-	-	1	-
E	G	-	1	-	-
E	H	-	1	-	-
E	I	-	1	-	-
E	J	-	1	-	-
F	G	-	-	-	1
F	H	-	-	-	1
F	I	-	-	-	1
F	J	-	-	-	1
G	H	1	-	-	-
G	I	1	-	-	-
G	J	1	-	-	-
H	I	1	-	-	-
H	J	1	-	-	-
I	J	1	-	-	-
45		7	25	6	7

Table 6.20: Comparison of district level classifications

Classification pair					
Classification	**Number of clusters**	**Classification**	**Number of clusters**	**Rand Index**	**Adjusted Rand Index**
Migration classification	8	Vickers Family	4	0.64	0.06
Migration classification	8	Vickers Group	12	0.78	0.05
Migration classification	8	Vickers Class	24	0.82	0.03
Migration classification	8	ONS Super-group	7	0.68	0.05
Migration classification	8	ONS Group	12	0.77	0.06
Migration classification	8	ONS Sub-group	23	0.82	0.05
Vickers Family	4	ONS Super-group	7	0.69	0.28
Vickers Group	12	ONS Group	12	0.85	0.36
Vickers Class	24	ONS Sub-group	23	0.92	0.40
Migration classification	8	Age only classification	8	0.86	0.39
Migration classification	8	NS-SEC only classification	8	0.84	0.30
Migration classification	8	Family status only classification	8	0.81	0.20
Migration classification	8	Housing tenure only classification	8	0.83	0.25

index. At the Family level there is only moderate agreement between the two classifications, whereas at the class level agreement is relatively high. However, the usefulness of also including the adjusted Rand index is apparent as for all tiers the figure is very low, and reduces as the number of clusters increases. This suggests that the majority of agreement between the clusters shown by Rand index can be explained by chance. It is a very similar story when comparing the Migration Classification with the ONS classification; a moderate, increasing to relatively high, association between the two classifications according to the Rand index, mitigated by a very high likelihood of chance agreement demonstrated by the adjusted index. In comparison, when the Vickers et al. classification is compared with the ONS classification, as also shown in the table, the Rand index indicates a moderate to very high agreement between the two classifications which is maintained far more convincingly when the adjusted Rand index is also taken into consideration.

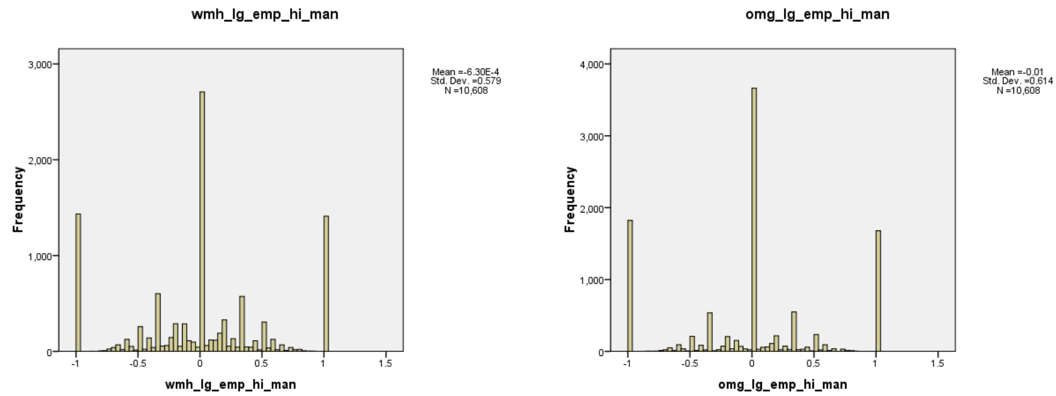
The results of the comparison between the Migration Classification and the two general purpose classifications are encouraging. They show that there is significant difference in the way that the alternative schemas group the districts in Britain. This helps to confirm the original hypothesis that a migration-specific classification will offer something different to a general purpose classification typology. And in answer to the question posed earlier, we can conclude that the Migration Classification is a new way of classifying districts, and not merely a surrogate for one of the other general purpose classifications already available. Certainly it offers a more different way of classifying districts than the two general purpose classifications do when compared with each other.

It remains then that the last task in the validation of the classification is to ensure that the final selection of variables indeed produced a solution distinct from any sub-sets of the same variables. Four alternative classifications were produced using the exact same methodology used in the original Migration Classification, however, this time sub-sets of the original suite of variables were chosen to be clustered. Classifications using only the variables associated with age, socio-economic status, housing tenure and family status were chosen. These classifications were then compared with the results of the original using the Rand and adjusted Rand indices, with the results also displayed in Table 6.20. As would be expected using sub-sets of the original variables, the agreement between the Migration Classification and the sub-classifications is relatively high. The age-only classification achieves the highest Rand index and adjusted figures, followed by socio-economic status, housing tenure and then family status. Whilst in all cases agreement is quite high, it is not so high that one may be attempted to adopt one of the more parsimonious classifications in place of the original. For example, if the age only classification when compared with the Migration Classification scored a Rand index higher than 0.9 and an adjusted Rand index well over 0.5, then the value of including other variables in the classification might be brought into question; it could be argued that age variables explain enough of the final classification for other variables to be dropped. As it is, however, the adjusted Rand index is still relatively low for all of the alternative classifications and with Rand index agreement not above 0.87 for any alternative, then it can be concluded that including all variables and keeping the original classification is offers the best solution.

6.7.3 Comparison with a ward level classification

Part of the discussion in the last chapter concerned the appropriate spatial scale for the classification. It was argued that the greater choice of variables available at district level meant that it was desirable to base the Migration Classification on the district geography. Furthermore, with one of the purposes of the classification being to provide a framework with which to examine migration between censuses, and inter-censal migration data only being available down to the district level, it made even more sense that local authority districts be the scale of analysis. However, during the classification building process, it became evident that some districts were easier to classify than others, something potentially caused by within-area variations. One possible solution to this could be to develop a ward-level classification based upon the same variables (as generalisation of district level variables meant that much of the detail was lost at this level making the variables broadly comparable with ward level variables) and where some districts were potentially misallocated (as evidenced by their silhouette score), then some evidence pointing towards to the reason for this misallocation could be gathered through looking at the classification profiles of wards within the district.

A suite of variables as similar as possible to the set of original variables in the district classification were assembled to be put through the same classification building process. It was at this point that a number of problems presented themselves with the data which meant that a ward classification comparable in variable composition to the original set of variables would not be



(a) Wholly moving households - employers in large organisations and higher managerial occupations

(b) Other moving groups - employers in large organisations and higher managerial occupations

Figure 6.18: Issues with ward level variable distributions caused by SCAM

possible. For exemplification of the problem, consider Figure 6.18 which shows the frequency distribution of the efficiency rate of wholly moving households and other moving groups with the head of the group in NS-SEC group 1. As is immediately evident, the distribution is far from Gaussian. The majority of counts are for efficiency rates of 0, 1 and -1. This is down to the very large numbers of 0s that feature in the in and out migration flows for these variables. Also of note are the other peaks between 0, 1 and -1. These correspond to efficiency rates which are multiples of 3. These occur thanks to the small cell adjustment method which has adjusted all flows of 1 and 2 to 0 or 3 and has resulted in a number of flows being multiples of 3.

Wherever small flows occur in the ward level data, variable distributions similar to the ones shown in Figure 6.18 are found. This includes all NS-SEC related variables and all tenure related variables. Consequently, the only variables left available for building a ward level classification from were the age related variables (although these did not completely match the variables at district level as some of the broad-age ranges were different). Family status variables were also not available.

6.8 Conclusions

This chapter has detailed the development of a new Migration Classification for Britain: a long process with a series of decisions required, the route to a final classification typology is not a straightforward one. What this chapter has shown, however, is that with informed decision making at each stage in the process, it is entirely feasible to produce a robust and useful classification containing clusters which exhibit distinctive profiles, and to this end the chapter has made a number of important contributions.

The first is a methodological contribution. Whilst in Section 6.2 a tried-and-tested route was chosen to produce an initial classification, a number of deficiencies in the process were identified, particularly relating to the software used to carry out the variable clustering. It was shown in Section 6.4 that the MATLAB software package has a number of features which make it preferable to more commonly used packages such as SPSS for finding clusters in multi-variate datasets.

The other contributions are more substantive. For example, one important outcome was the discovery that international migration variables are not of significance when compared alongside internal migration variables in defining clusters of districts. A case could have been made for the exclusion of international migration variables at the beginning of the classification building process as the purpose of the classification was to help improve the understanding of internal migration in Britain, but with some evidence of an association with internal migration the decision was taken to explore the influence of international migration variables before discounting them altogether. Despite international migration variables being highly skewed, they were shown to have very little effect on the final cluster solutions with internal migration variables playing a larger and more important role. International migration variables were dropped and as such the final classification is an internal migration classification for internal migration analysis.

Another substantive contribution was in the identification of variables which are important in describing the internal migration landscape of Britain. The 44 variables included in the final classification, taken from across the census data spectrum, are more important than any which were omitted. The omissions included sizable groups such as the migrants from 'no-previous address', variables relating to migrants with limiting long term illness and those relating to migrants moving into or out of publicly rented accommodation. The domains from which the largest number of variables were selected were age and socio-economic status, pointing to these being some of the more important defining features of internal migration in Britain.

In the introduction to this chapter, a stated aim was that the classification itself would add to the understanding of migration in Britain at the start of the 21^{st} century through the typology defining a series of areas with particular migrant characteristics - characteristics which could answer questions about types of migration area. Careful interpretation of average z-score values reveals very distinctive profiles for each of the eight clusters comprising the Migration Classification, thus realising this aim. Each profile has led to each cluster being given a descriptive name; enabling both identification, but more importantly providing a digestible summary for anyone making use of the clusters.

Whilst silhouette values for each cluster reveal that some LADs are more heavily associated with the overall cluster profile than others, broad observations can be made about each cluster. For example, answering the example questions in the introduction: 'which are the areas which lose young migrants?' - areas such as Dynamic London, Successful Family In-Migrants and Footloose, Middle-Class lose well above average numbers of young migrants; 'are there are areas which attract migrants of differing socio-economic status?' - Declining

Industrial, Working-Class, Local Britain is a desination for significantly more migrants in the lower socio-economic groups than the higher socio-economic groups and the Student Towns and Cities cluster has gained that particular moniker for quite an apparent reason; 'are there are some areas largly excluded from the internal migration system?' - districts in Low Mobility Britain have very little interaction with any other areas in the country, and migrants to districts in Declining Industrial, Working-Class, Local Britain will tend to come from districts in the same cluster; 'are there any associations between these migrant attributes in particular areas?' - yes, for example the heavy inflows of young migrants into the Student Towns and Cities cluster are strongly associated with the predominance of net-inflows into privately rented accommodation; 'how are these areas with similar migrant characteristics distributed across space?' - a map of the new Migration Classification typology clearly defines the spatial extents of each cluster, showing revealing spatial assocations with particular migrant types, for example the stark band across northern England in the Declining Industrial, Working-Class, Local Britain cluster.

The final main contribution of this chapter relates to the last aim of the introduction. It was hoped that the classification created would be significantly different from other general purpose classifications which have preceeded it. Mathematical techniques were employed to compare the Migration Classification with two general purpose classifcations, with the results showing that the Migration Classification *is* a significantly different classification from these and therefore would very likely offer anyone using it to study internal migration in Britain a superior analysis tool to a more general purpose classification. Work can still be carried out to compare in more detail Migration Classification districts with districts in the other classifications to assess the extent to which the migrant profiles depart from the profiles of the rest of the settled population, but this work would detract, at this time, from the main focus of this thesis. It does, however, offer opportunities for future research.

Returning to part of the original rationale for developing the classification laid out in Chapter 5, now the Migration Classification has been developed it remains to use it as a tool for examining other migration data which should further the understanding of migration in Britain at the beginning of the 21^{st} century. In Chapter 2 the process of producing a new partially-estimated PRDS-based dataset was described in detail. Equiped with this new dataset of inter-LAD flows for Britain, the next chapter look to use the new Migration Classification to make sense of a decade of internal migration flows, advancing our knowledge still further.

Chapter 7

Monitoring migration between censuses

7.1 Introduction

In Chapter 4 an analysis of internal migration in Britain was carried out using data from the 2001 Census and a general purpose classification of districts developed by Vickers et al. (2003). The analysis revealed some interesting patterns, but it was concluded that the general purpose framework might not provide the optimal solution to both reducing the spatial complexity of migration flows and exemplifying the important patterns in Britain at the beginning of the 21^{st} Century. Consequently in Chapter 5 the case for a new classification framework was made and in Chapter 6 the new *Migration Classification* was developed and presented. In developing the Migration Classification additional knowledge about the characteristics of migrants defining the internal migration landscape in Britain has been gained through an exhaustive process of data reduction; but despite this extensive analysis, the use of the 2001 Census data in this thesis, thus far, has limited the understanding of internal migration to the transitions occurring over a single year between April 2000 and April 2001. It is impossible to contextualise the situation in 2001 without examining internal migration in neighbouring years; it is only by doing this that a complete understanding encompassing the evolution of migration patterns over time can be achieved.

Time-series analysis of some inter-regional patterns and cross-sectional analysis of some inter-district patterns of internal migration in England and Wales have been reported on a regular basis (ONS, 2006, 2007b, 2008b). Raymer and Giulietti (2009) have analysed a relatively recent time-series of internal migration flows by ethnic group in England and Wales up to 2004, however, assuming it is due to the lack of a national, small-area internal migration dataset, no attempt anywhere else has been made to analyse flows between areas across the whole of Britain using non-census data. In a similar vein to the work in Chapter 4, Raymer and Giulietti (2009) use a general purpose area classification to reduce the complexity of the migration system, but as has been argued already, such divisions may not be the most useful for studying internal migration. Furthermore, analysis of the most recent patterns of internal migration leading up to the start of the global financial crisis which began in 2008, have not been attempted yet.

The substantive aim of this chapter, therefore, is to confront this gap in the current knowledge. In Chapter 2, a method was described to create a partially estimated time-series dataset of inter-district migration flows for the whole of Great Britain. These data are for the ten year period from mid 1998-99 to mid 2007-08 and will be used here to provide an analysis of internal migration flows from the latter part of the last century to the early part of this. The data are disaggregated by broad age group and thus will present the opportunity for an age-specific analysis, which, given then the findings of Chapter 4 pointing to a significant variation in age-specific flows, will be important. As described in Chapter 2 there are some issues with the PRDS data and associated estimates which should be borne in mind before any analysis takes place, perhaps most pertinent is the under-estimation of young migrant flows and over-estimation of older migrant flows. This in mind though, the analysis will proceed addressing a series of aims.

Firstly this chapter will present an overview of internal migration patterns over a ten year period, which whilst not completely, to a large extent covers an inter-censal period. The use of the Migration Classification will demonstrate the benefits of using such a framework to both reduce a large amount of complex data in a useful way and at the same time add value to the temporally rich but attribute poor PRDS-based data. After an aggregate time-series overview of internal migration flows in Britain, drawing out some of the main age-related patterns in Section 7.2, the analysis will turn to the Migration Classification in Section 7.3, describing some of the main characteristics of flows within the Migration Classification system and offering an account of the changing migration patterns over time.

In Chapter 4 turnover and churn were introduced as alternative metrics which can be used to deepen our understanding of internal migration flows. Another aim of this chapter is to introduce new analysis measures which will help to disect, still further, internal migration flows, with the specific methodological objective of detailing two new metrics which borrow techniques from demography to account for the variation in migration flows by distance and age, and which will equip those studying internal migration with additional tools for analysis. As well as equiping other researchers, these new tools will mean that the account of internal migration in this chapter will be able to take some alternative perspectives, enhancing the knowledge of internal migration in Britain still further. In doing all of this, the chapter has a final aim of highlighting the benefits of the new Migration Classification framework for analysing internal migration flows in Britain.

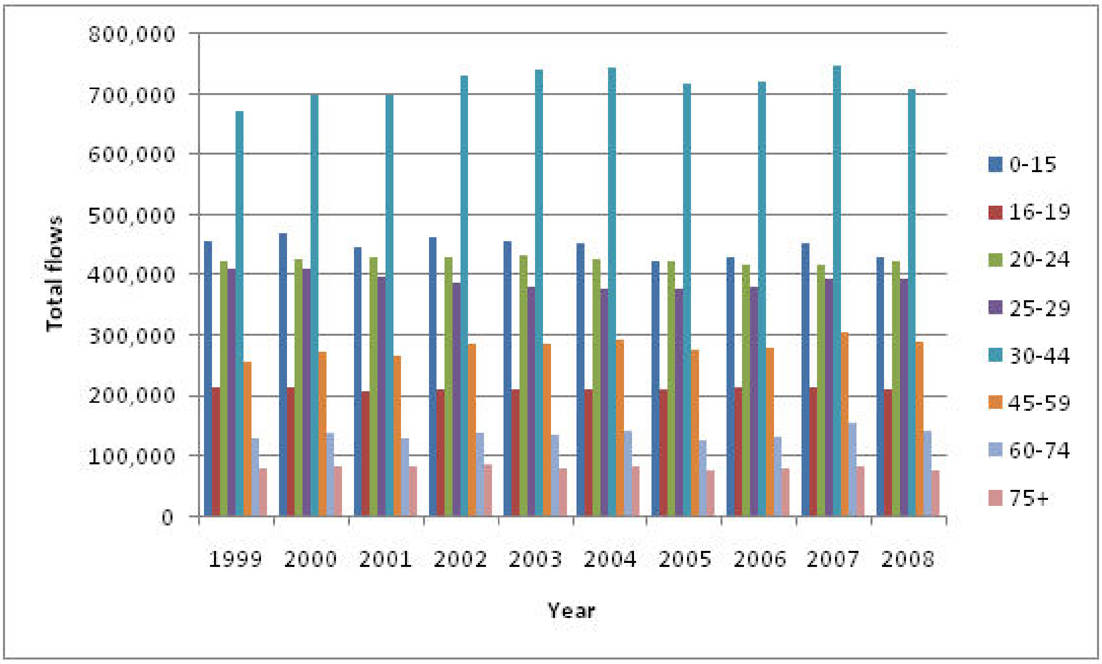


Figure 7.1: Variation in total inter-district migrant flows for defined age groups, 1999-2008

7.2 Variation in aggregate internal migration patterns, 1999-2008

So what patterns do these new data reveal? Assessing the variation in the age-specific flow volumes over the ten year period for the whole of the UK reveals a remarkably stable picture. There is a very little difference in the total volume of migrants over the decade. Figure 7.1 shows the variation in the total flows, with the total number of migrants in each age group scarcely changing year-on-year between 1999 and 2008. The overall proportion of the population who are inter-district migrants also varies very little, however, at the age group of peak migration (20-24) there is a steady and consistent decline in the migration rate from 12% of the population in 1999, to 10% of the population in 2008 (Figure 7.2). The question is, how much of this decline is due to the decline in the numbers of migrants in this age group, and how much is due to a change in the population denominator?

Figure 7.3 provides the answer and reveals that this decline in the overall migration rate of the 20-24 age group is a product of both a declining number of migrants and an increasing population denominator. The trend lines on the graph show clearly the two data trajectories, both acting together to significantly reduce the rate of migration for this age group between 1999 and 2008. Why there is this decline in the propensity to migrate in this age group is interesting. The work of Dennett and Stillwell (2008) suggests that it may be down to a reduction in the attractiveness of London, and an increase in the ability of university towns to retain recent graduates where in the past more would have headed to London in search of employment, however a more detailed examination of this trend with be given in the next chapter, with evidence which could refute this idea.

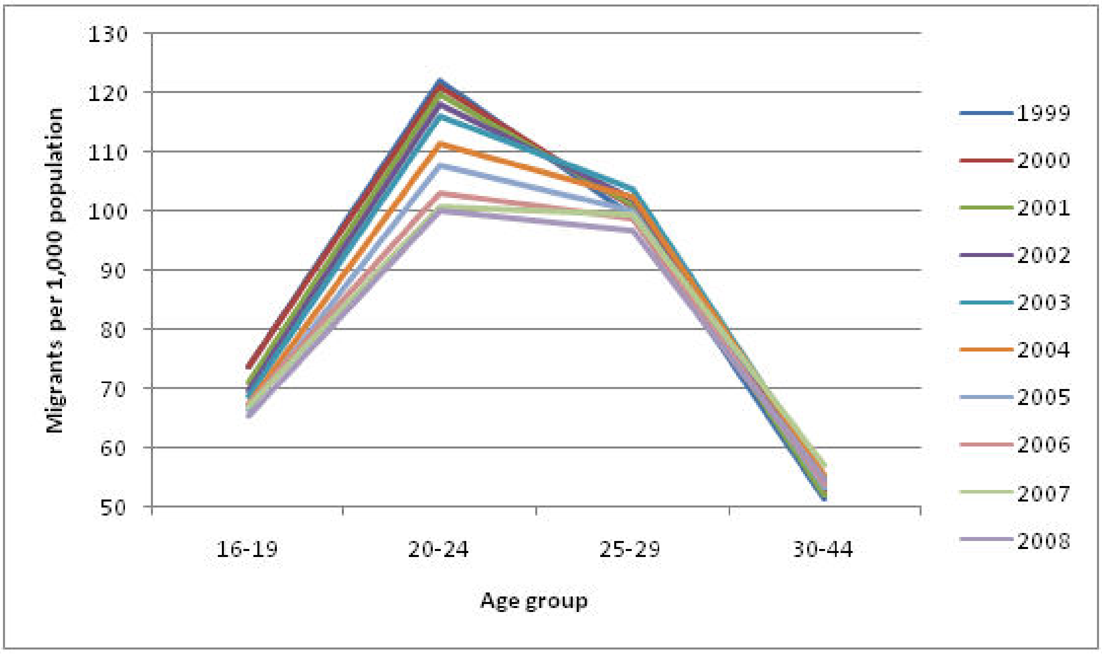


Figure 7.2: Inter-district migrants per 1,000 population, 1999-2008 - peak migration age groups

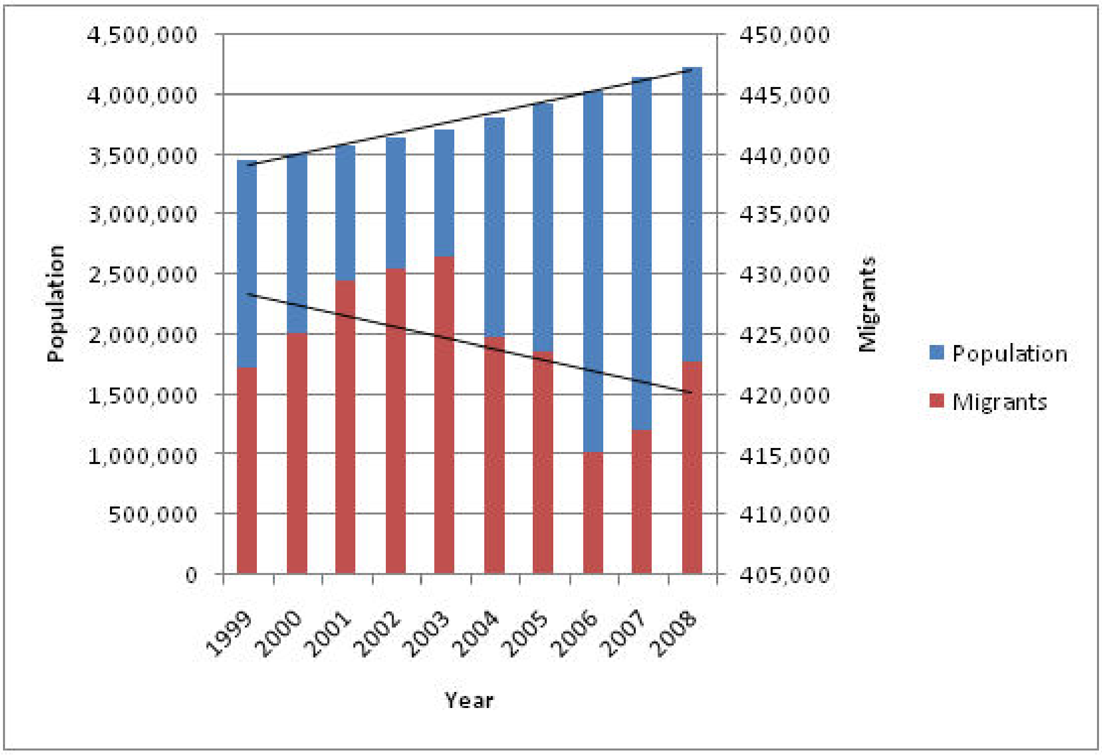


Figure 7.3: Comparison of the population and migration trajectories of the 20-24 age group, 1999-2008

7.3 Time series analysis of internal migration patterns in the context of the migration classification

Time series analysis of migration flows in the UK has a historical precedent with the work of Ogilvy (1982) examining flows in the 1970s, and Stillwell et al. (1995a) in the 1980s, and more recently Dennett and Rees (2010) and Stillwell et al. (2010) in this century. Ogilvy uses the very coarse system of 10 UK regions, Stillwell et al. (1995a) use the less coarse Family Health Service Authorities. Whilst both analyses show the benefits of contextualising any given annual snapshot by the flows in preceding and subsequent years (with relative stability demonstrated in the UK over 20 years) both pieces of work are limited in that the spatial systems employed are somewhat arbitrary in their definition. With the exception of urban areas such as London which have distinctive migrant related characteristics, other areas were defined principally though their geographical location rather than the underlying characteristics of their populations. This analysis will, for the first time, examine migrant flows in the context of areas defined by a set of migrant characteristics.

7.3.1 Patterns of migration for area types over time

Before analysing patterns of migration, it is useful to first contextualise the analysis through examining some basic statistics associated with the geographies and populations of each cluster. In the previous chapter, the precise geographical location of the districts which comprise each cluster is provided, along with the number of districts in each cluster. Information relating to the spatial association that each district in each cluster has with each district in each other cluster is not given. Such information is key to a successful analysis as many of the migration patterns discussed could be confounded by geographical association. For example, a very compact cluster like Dynamic London is likely to experience a higher frequency of intra-cluster migration moves than a diffuse cluster like Coastal and Rural Retirement Migrants. Quantification of the spatial association between districts within and between clusters will allow these associations to be accounted for and dealt with in the analysis.

Table 7.1 details the interaction and distance associations between all of the districts in all of the Migration Classification clusters. It can be seen clearly that Dynamic London is indeed the most compact cluster containing the least number of districts and the lowest average distance between districts contained within the cluster. The average distance between the 31 districts in Dynamic London is around 26km (the maximum potential distance of some 120km due to Cambridge featuring in this cluster). The total number of inter-district/intra-cluster interactions is 930 (31x30). Contrastingly, the Coastal and Rural Retirement Migrants cluster has an average intra-cluster/inter-district distance of 293km, with some 4160 possible interactions between all districts within the cluster. Interestingly the cluster with the most intra-cluster/inter-district interactions (Declining Industrial, Working-Class, Local Britain) does not have the highest average distance between these districts, suggesting that it is (relative to other clusters) more compact. But can this level of compactness be quantified? Taking the information about average distance and total inter-district interactions, it is possible to construct an 'index of association' which defines both the compactness of each cluster but also the potential interaction relationship between districts in that cluster and those districts in others.

If d_{ij} = the distance between LADs in Britain and MC represents the complete matrix for the Migration Classification with:

$$\sum_I \sum_J MC_{IJ} = MC_{++} = MC \tag{7.1}$$

and every district i is a member of Migration Classification cluster I such that $i \in I$

The index can be calculated as:

$$1 - \left(\frac{AD}{TD}\right) \tag{7.2}$$

where AD is the average distance of all inter-district interactions within and between clusters, or:

$$AD = \frac{1}{n} \sum_{ij=1}^{n} d_{ij} \tag{7.3}$$

and TD is the total distance of all inter-district interactions, or:

The closer the index to 1, the stronger the spatial association and the degree of potential migration interaction between the districts contained in each cluster. Examining the index of association in Table 7.1, clearly in terms of intra-cluster interaction, districts in Dynamic London have the strongest association, closely followed by Declining Industrial, Working-Class, Local Britain. The districts with the weakest spatial association are those within Low Mobility Britain. This is principally because districts on the South coast of England and Northern Scotland feature in this cluster.

Of course, as well as spatial association being important in the analysis of migration, so too is underlying population which will provide the rate denominator. Table 7.2 shows the average total populations for each of the clusters, along with minimum and maximum populations for the time period. Declining Industrial, Working-Class, Local Britain is the most populous with over 12 million people resident in the districts which make up the cluster, this is closely followed by Student Towns and Cities, with 11 million residents. All the other clusters have fewer people with between around 4 and 7 million people.

Table 7.1: Flow and distance associations between districts in each cluster

Origin		Destination		Districts in cluster	Relationships in the distances (km) between districts in each cluster							Total inter district flows	Index of association
					Average	Sum	Min	Max	Range	Total Variation	Standard Deviation		
1	Coastal and Rural Retirement Migrants	1	Coastal and Rural Retirement Migrants	65	292.76	1,217,894	12.44	832.61	820.17	24885.34	157.75	4160	0.93
1	Coastal and Rural Retirement Migrants	2	Sedentary Middle-Class Britain		289.01	732,631	10.02	1152.13	1142.11	36858.56	191.99	2535	0.89
1	Coastal and Rural Retirement Migrants	3	Student Towns		277.77	812,466	6.04	809.57	803.52	23649.05	153.78	2925	0.91
1	Coastal and Rural Retirement Migrants	4	Intermediate Single Migrants		249.01	598,860	16.94	844.54	827.59	20968.02	144.80	2405	0.90
1	Coastal and Rural Retirement Migrants	5	Constrained, Working-Class, Local Britain		295.12	1,438,724	7.88	997.39	989.51	25837.23	160.74	4875	0.94
1	Coastal and Rural Retirement Migrants	6	Footloose, Middle-Class, Commuter Britain		244.80	843,325	6.60	863.02	856.42	18704.74	136.77	3445	0.93
1	Coastal and Rural Retirement Migrants	7	Dynamic London		244.05	491,758	28.43	746.73	718.29	18678.80	136.67	2015	0.88
1	Coastal and Rural Retirement Migrants	8	Successful Family In-migrants		261.36	1,070,275	5.49	808.65	803.17	20264.01	142.35	4095	0.94
2	Sedentary Middle-Class Britain	2	Sedentary Middle-Class Britain	39	260.87	386,606	6.91	1041.37	1034.46	49239.93	221.90	1482	0.82
2	Sedentary Middle-Class Britain	3	Student Towns		257.22	451,415	5.00	1108.40	1103.40	34099.79	184.66	1755	0.85
2	Sedentary Middle-Class Britain	4	Intermediate Single Migrants		237.37	342,518	5.71	1015.33	1009.61	42858.24	207.02	1443	0.84
2	Sedentary Middle-Class Britain	5	Constrained, Working-Class, Local Britain		256.05	748,946	0.90	986.94	986.04	30107.34	173.51	2925	0.91
2	Sedentary Middle-Class Britain	6	Footloose, Middle-Class, Commuter Britain		244.62	505,631	4.77	1188.49	1183.72	43063.90	207.52	2067	0.88
2	Sedentary Middle-Class Britain	7	Dynamic London		242.82	293,569	5.51	998.52	993.00	45726.14	213.84	1209	0.80
2	Sedentary Middle-Class Britain	8	Successful Family In-migrants		252.34	619,996	7.73	1108.03	1100.31	35612.20	188.71	2457	0.90
3	Student Towns	3	Student Towns	45	253.93	502,780	9.29	763.05	753.76	21499.99	146.63	1980	0.87
3	Student Towns	4	Intermediate Single Migrants		242.03	402,972	7.97	806.78	798.80	21970.55	148.22	1665	0.85
3	Student Towns	5	Constrained, Working-Class, Local Britain		246.53	832,044	5.53	960.29	954.75	22513.18	150.04	3375	0.93
3	Student Towns	6	Footloose, Middle-Class, Commuter Britain		243.23	580,098	5.41	848.68	843.27	22318.47	149.39	2385	0.90
3	Student Towns	7	Dynamic London		250.15	348,955	17.55	663.77	646.22	21520.53	146.70	1395	0.82
3	Student Towns	8	Successful Family In-migrants		245.72	696,620	4.85	762.50	757.65	20498.91	143.17	2835	0.91
4	Intermediate Single Migrants	4	Intermediate Single Migrants	37	150.48	200,444	6.59	751.18	744.58	19909.01	141.10	1332	0.89
4	Intermediate Single Migrants	5	Constrained, Working-Class, Local Britain		288.76	801,315	12.79	895.54	882.75	22515.02	150.05	2775	0.90
4	Intermediate Single Migrants	6	Footloose, Middle-Class, Commuter Britain		142.80	280,037	4.81	878.90	874.09	15869.15	125.97	1961	0.93
4	Intermediate Single Migrants	7	Dynamic London		109.96	126,127	4.85	729.62	724.77	14175.82	119.06	1147	0.90
4	Intermediate Single Migrants	8	Successful Family In-migrants		203.57	474,522	10.21	808.27	798.06	15934.61	126.23	2331	0.91
5	Constrained, Working-Class, Local Britain	5	Constrained, Working-Class, Local Britain	75	209.18	1,160,938	6.93	842.99	836.06	18404.36	135.66	5550	0.96
5	Constrained, Working-Class, Local Britain	6	Footloose, Middle-Class, Commuter Britain		300.73	1,195,419	8.27	1030.70	1022.43	23207.71	152.34	3975	0.92
5	Constrained, Working-Class, Local Britain	7	Dynamic London		322.89	750,728	65.55	875.49	809.94	19534.84	139.77	2325	0.86
5	Constrained, Working-Class, Local Britain	8	Successful Family In-migrants		267.55	1,264,189	9.80	961.69	951.88	24017.46	154.98	4725	0.94
6	Footloose, Middle-Class, Commuter Britain	6	Footloose, Middle-Class, Commuter Britain	53	125.53	345,966	7.63	658.48	650.85	12077.86	109.90	2756	0.95
6	Footloose, Middle-Class, Commuter Britain	7	Dynamic London		92.38	151,781	3.34	516.70	513.36	8646.54	92.99	1643	0.94
6	Footloose, Middle-Class, Commuter Britain	8	Successful Family In-migrants		200.47	669,370	17.96	703.43	685.47	13721.29	117.14	3339	0.94
7	Dynamic London	7	Dynamic London	31	26.09	24,264	2.06	120.32	118.26	473.84	21.77	930	0.97
7	Dynamic London	8	Successful Family In-migrants		195.01	380,862	19.13	612.78	593.64	12141.01	110.19	1953	0.90
8	Successful Family In-migrants	8	Successful Family In-migrants	63	227.31	887,871	10.15	683.53	673.39	15524.15	124.60	3906	0.94

Table 7.2: Population statistics for clusters, 1999-2008

Cluster	Cluster name	Average population 1999-2008	Minimum population 1999-2008	Maximum Population 1999-2008
1	Coastal and Rural Retirement Migrants	7,418,305	7,180,310	7,649,299
2	Low Mobility Britain	4,410,516	4,359,480	4,480,139
3	Student Towns and Cities	11,291,582	11,134,900	11,609,994
4	Moderate Mobility, Non-Household, Mixed Occupations	5,498,740	5,373,300	5,656,467
5	Declining Industrial, Working-Class, Local Britain	12,314,021	12,232,600	12,439,470
6	Footloose, Middle-Class, Commuter Britain	5,672,490	5,559,300	5,846,800
7	Dynamic London	6,111,870	5,889,100	6,313,598
8	Successful Family In-migrants	5,428,492	5,240,540	5,614,800

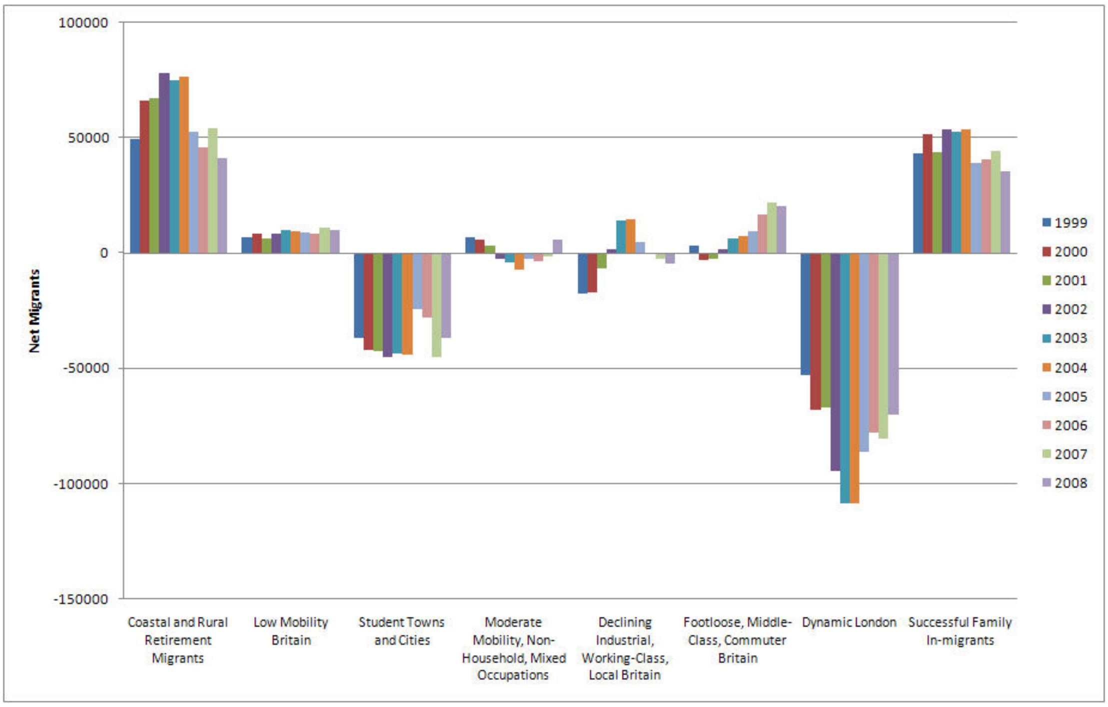


Figure 7.4: Net migrants by cluster, 1999-2008

Table 7.3: Average flows between Migration Classification clusters, 1999-2008, as a percentage of total average flows, 1999-2008

Origin/Destination	**Coastal and Rural Retirement Migrants**	**Declining Industrial, Working-Class, Local Britain**	**Dynamic London**	**Footloose, Middle-Class, Commuter Britain**	**Moderate Mobility, Non-Household, Mixed Occupations**	**Low Mobility Britain**	**Student Towns and Cities**	**Successful Family In-migrants**	**Total**
Coastal and Rural Retirement Migrants	2.41	1.32	0.52	0.81	0.79	0.50	2.44	1.71	**10.49**
Declining Industrial, Working-Class, Local Britain	1.62	4.36	0.49	0.53	0.55	0.97	3.71	1.14	**13.36**
Dynamic London	0.92	0.51	8.50	2.97	2.10	0.66	1.88	0.73	**18.28**
Footloose, Middle-Class, Commuter Britain	1.27	0.53	1.84	2.84	1.45	0.70	1.88	1.34	**11.85**
Moderate Mobility, Non-Household, Mixed Occupations	1.24	0.58	1.14	1.60	1.21	0.82	1.60	1.26	**9.43**
Low Mobility Britain	0.75	0.94	0.37	0.66	0.81	0.63	1.82	0.67	**6.65**
Student Towns and Cities	2.65	4.19	1.97	1.83	1.64	2.21	3.74	2.39	**20.63**
Successful Family In-migrants	1.85	0.87	0.45	0.91	0.88	0.48	2.12	1.75	**9.30**
Total	**12.72**	**13.30**	**15.27**	**12.15**	**9.43**	**6.97**	**19.18**	**10.99**	**100.00**

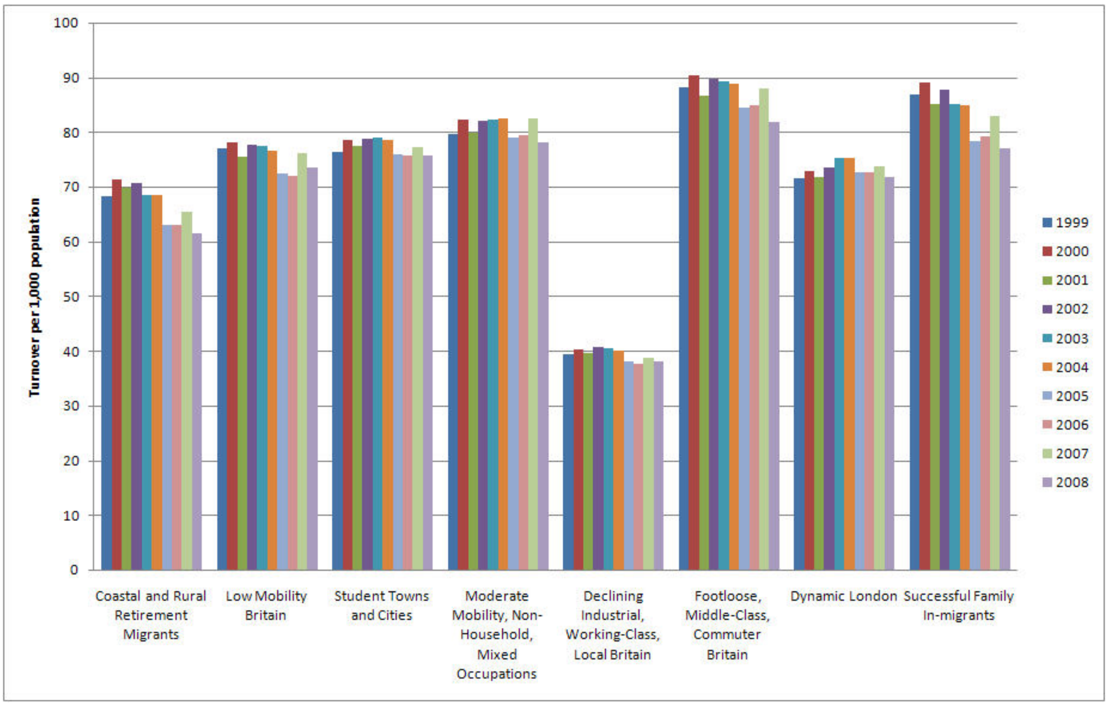


Figure 7.5: Turnover rates by cluster, 1999-2008

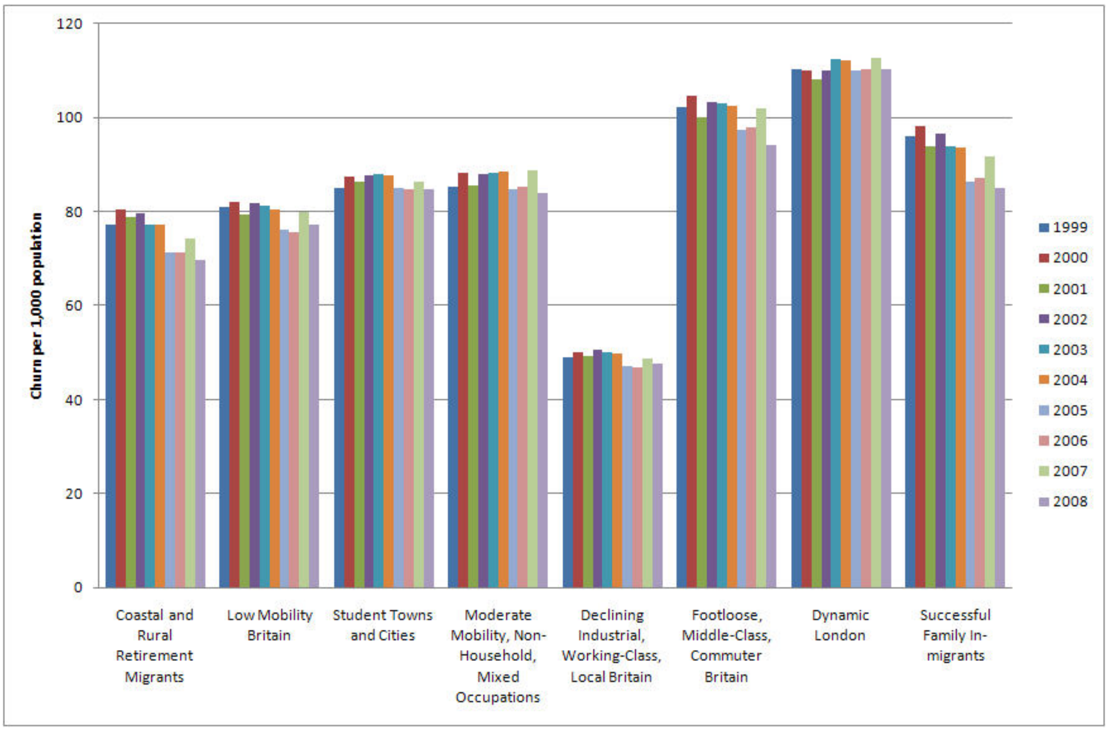


Figure 7.6: Churn rates by cluster, 1999-2008

Aggregate Patterns

Table 7.3 gives an overview of the average total flows between Migration Classification clusters over the decade of study. Each interaction is represented as a percentage of the total flows within the system for ease of comparison. Flows within Dynamic London immediately jump out as being the largest proportion of any inter- or intra-cluster flows at around 8.5% of all flows within the system. In terms of total in- and -out flows, the Student Towns and Cities cluster shows its importance within the system, with around 20% of flows in both directions. Low Mobility Britain lives up to its name with, by some margin, the lowest proportions of both in- and out- migration flows, although as the cluster with the smallest population this is perhaps unsurprising. Figure 7.4 breaks these patterns down by year. Overall, Coastal and Rural Retirement Migrants is a consistent net gainer of individuals, gaining on average around 50,000 migrants in each year of the period, with a peak in 2002 followed by a slight decline. Also gaining consistently over the decade is Successful Family In-migrants, but to a lesser extent. The final consistent net gainer of population over the period is Low Mobility Britain. Net gains in this cluster are modest in comparison with the other two net gainers; however, the gains remain steady at around an average of 500 migrants a year. Consistently a net loser of migrants over the decade is Dynamic London, with a peak loss of around 100,000 migrants in 2004 and 2005. Losses doubled to this point from 1999 and decreased from this point until 2008. However, even at the lowest, the net loss was still over 50,000 people a year. Also losing considerable numbers of migrants between 1999 and 2008 is Student Towns and Cities; consistently making a net loss of around 40,000 migrants a year. The remaining clusters are more varied, with Moderate Mobility, Non-Household, Mixed Occupations and Declining Industrial, Working-Class, Local Britain making both net gains and losses over the decade. Footloose, Middle-Class, Commuter Britain changes from a modest net loss of migrants in 2000 and 2001, to a gradual increase in the net gain of migrants at the end of the decade. Net rates show the exact same patterns as the net flows.

As was demonstrated in Chapter 4, analysis of net migration can obscure total flow volumes, so to address this, overall rates of turnover and churn were also calculated. Turnover is calculated in exactly the same way as in Chapter 4 (i.e. in-migration + out-migration / population at risk), however churn is calculated slightly differently thanks to the structure of the PRDS-based estimates. Whereas in Chapter 4, churn statistics included both the within cluster and within district flows, here as within district flows are not available, only within cluster flows form part of the churn calculation.

Figures 7.5 and 7.6 show the overall turnover and churn statistics for each cluster over the decade; of note first, is that the only relative difference between clusters in the two sets of statistics is that Dynamic London ranks around sixth for its rates of turnover, but first for its rates of churn. This can be attributed to the far higher volume of intra-cluster flows thanks to the increased levels of spatial association between its districts. Apart from Dynamic London, the clusters with the highest rates of turnover and churn over the decade are Footloose, Middle-Class, Commuter Britain and Successful Family In-Migrants. However both of these clusters exhibit reductions in migration rates between 1999 and 2008. Reductions in the rates of migration over this period can also be seen in Coastal and Rural Retirement Migrants and Low Mobility Britain, although the reductions are less pronounced. Lowest rates of migration by some margin can be seen in Declining Industrial, Working-Class, Local Britain, with rates around half of those shown by the clusters with the highest turnover and churn.

Accounting for distance

With the apparent effect of spatial association large in Dynamic London, influencing turnover and churn statistics, the question that follows is 'exactly how large'? Are all turnover and churn statistics partially an artefact of the spatial systems that they operate within? The inabilities of conventional migration measures such as net migration and migration efficiency to account for the spatial systems they inhabit have been noted by Newbold and Peterson (2001) - they propose incorporating distance information into net migration and net efficiency calculations to produce measures of net attraction and attraction efficiency. Here is proposed an alternative which allows for easy comparison between clusters and accounts for the differing spatial associations within the Migration Classification system - a Standardised Migration Distance Ratio (SMDR). This is a technique borrowed from demography where metrics like SMRs are commonly calculated to deal with the different age structures present in different locations. For example, coastal retirement towns will have high mortality rates as they are home to higher proportions of old people. Some may interpret this high mortality rate as reflecting poor general health in the population, however, when the much older age structure of these towns is taken into consideration, it is very often the case that these towns do not contain populations that are any less healthy than other places; more older people simply means more deaths will take place. The basic idea of any standardised ratio is to divide the observed counts of a phenomenon by the expected counts (which are based on some overall global distribution). Using this principal it should be possible to standardise any observed phenomena and thus account for the effect of any distorting local distributions.

To account for the distorting effect of distance on migration within and between clusters in the migration classification, a SMDR is calculated. All inflows and outflows from clusters were separated into 11 equal distance bands (calculated from the population weighted centroid distance between each district - greater or fewer numbers of bands could be used) from 0-100 km to 1000+ km. The total number of flows for each distance band across all clusters is calculated, and then divided by the total population across all clusters. This ratio is then applied to the population of each cluster to produce an expected number of flows. This ratio varies from one that would be calculated for a Standardised Mortality Ratio (SMR), as the PAR for each distance band remains the same. This is in contrast to where, for example, the standardisation is across age groups and the PAR will vary by age group. Here, the population at risk of migrating any distance will always remain the same.

Table 7.4: Observed Migration Classification (MC_{ID}) and expected (E_{ID}) in- and out-migration flow matrices

	Cluster/Distance (km)	**0-100**	**100-200**	**200-300**	**300-400**	**400-500**	**500-600**	**600-700**	**700-800**	**800-900**	**900-1000**	**1000+**	
Observed In-migration MC_{ID}	Coastal and Rural Retirement Migrants	181338	71696	40621	23321	8970	3561	2006	1164	138	21	25	332861
	Declining Industrial, Working-Class, Local Britain	243890	47361	31732	14174	5661	4883	987	91	72	13	0	348864
	Dynamic London	323671	40227	30284	8091	2170	6400	893	274	40	8	0	412058
	Footloose, Middle-Class, Commuter Britain	240810	45277	23404	7507	2472	3699	1290	273	53	70	11	324866
	Moderate Mobility, Non-Household, Mixed Occupations	182483	35719	18072	5898	2342	2330	991	380	39	19	0	248273
	Low Mobility Britain	148229	20071	11355	4126	1774	1576	841	373	150	72	27	188594
	Student Towns and Cities	274820	120869	63260	24337	8872	8545	2378	322	57	5	18	503483
	Successful Family In-migrants	199170	53281	26257	11597	3655	2006	1359	227	36	4	6	297598
		1794411	434501	244985	99051	35916	33000	10745	3104	585	212	87	
Observed Out-migration MC_{ID}	Coastal and Rural Retirement Migrants	156340	60042	33930	19515	7550	3657	1794	942	138	18	20	283946
	Declining Industrial, Working-Class, Local Britain	245260	54115	36842	16875	6972	5316	1264	60	58	20	0	366782
	Dynamic London	383421	41323	25052	7259	1777	5163	588	321	31	11	0	464946
	Footloose, Middle-Class, Commuter Britain	226197	53874	26218	8196	2497	3426	1007	367	42	29	3	321856
	Moderate Mobility, Non-Household, Mixed Occupations	167874	41546	19608	6351	2300	2408	1258	454	66	9	0	241874
	Low Mobility Britain	134227	23982	13444	5014	2000	1916	936	354	150	113	60	182196
	Student Towns and Cities	313021	114035	66257	25087	9549	9255	2682	355	67	4	3	540315
	Successful Family In-migrants	168071	45584	23634	10754	3271	1859	1216	251	33	8	1	254682
		1794411	434501	244985	99051	35916	33000	10745	3104	585	212	87	
Population P_{ID}	Coastal and Rural Retirement Migrants	7180310	7180310	7180310	7180310	7180310	7180310	7180310	7180310	7180310	7180310	7180310	78983410
	Declining Industrial, Working-Class, Local Britain	12242730	12242730	12242730	12242730	12242730	12242730	12242730	12242730	12242730	12242730	12242730	134670030
	Dynamic London	5889100	5889100	5889100	5889100	5889100	5889100	5889100	5889100	5889100	5889100	5889100	64780100
	Footloose, Middle-Class, Commuter Britain	5559300	5559300	5559300	5559300	5559300	5559300	5559300	5559300	5559300	5559300	5559300	61152300
	Moderate Mobility, Non-Household, Mixed Occupations	5373300	5373300	5373300	5373300	5373300	5373300	5373300	5373300	5373300	5373300	5373300	59106300
	Low Mobility Britain	4359480	4359480	4359480	4359480	4359480	4359480	4359480	4359480	4359480	4359480	4359480	47954280
	Student Towns and Cities	11163190	11163190	11163190	11163190	11163190	11163190	11163190	11163190	11163190	11163190	11163190	122795090
	Successful Family In-migrants	5240540	5240540	5240540	5240540	5240540	5240540	5240540	5240540	5240540	5240540	5240540	57645940
		57007950	57007950	57007950	57007950	57007950	57007950	57007950	57007950	57007950	57007950	57007950	
Expected migration E_{ID}	Coastal and Rural Retirement Migrants	226011	54727	30857	12476	4524	4156	1353	391	74	27	11	334606
	Declining Industrial, Working-Class, Local Britain	385358	93311	52612	21272	7713	7087	2308	667	126	46	19	570517
	Dynamic London	185368	44885	25308	10232	3710	3409	1110	321	60	22	9	274435
	Footloose, Middle-Class, Commuter Britain	174987	42372	23890	9659	3502	3218	1048	303	57	21	8	259066
	Moderate Mobility, Non-Household, Mixed Occupations	169133	40954	23091	9336	3385	3110	1013	293	55	20	8	250398
	Low Mobility Britain	137221	33227	18734	7575	2747	2524	822	237	45	16	7	203154
	Student Towns and Cities	351378	85083	47973	19396	7033	6462	2104	608	115	42	17	520210
	Successful Family In-migrants	164954	39942	22521	9105	3302	3034	988	285	54	19	8	244212
		1794411	434501	244985	99051	35916	33000	10745	3104	585	212	87	

Standardised Migration Distance Ratios (SMDRs) can be calculated for both in and out migration flows. As an example, for a matrix of flows between Migration Classification origins MC_I and distance bands D a series of observed migration flows MC_{ID} can be defined (Table 7.4). SMDRs can be calculated as the ratio between these observed flows to expected flows such that:

$$SMDR = 100\left(\frac{MC_{I+}}{E_{I+}}\right) \tag{7.4}$$

where

E_{ID} = the expected migration to or from Migration Classification cluster I in distance band D

and

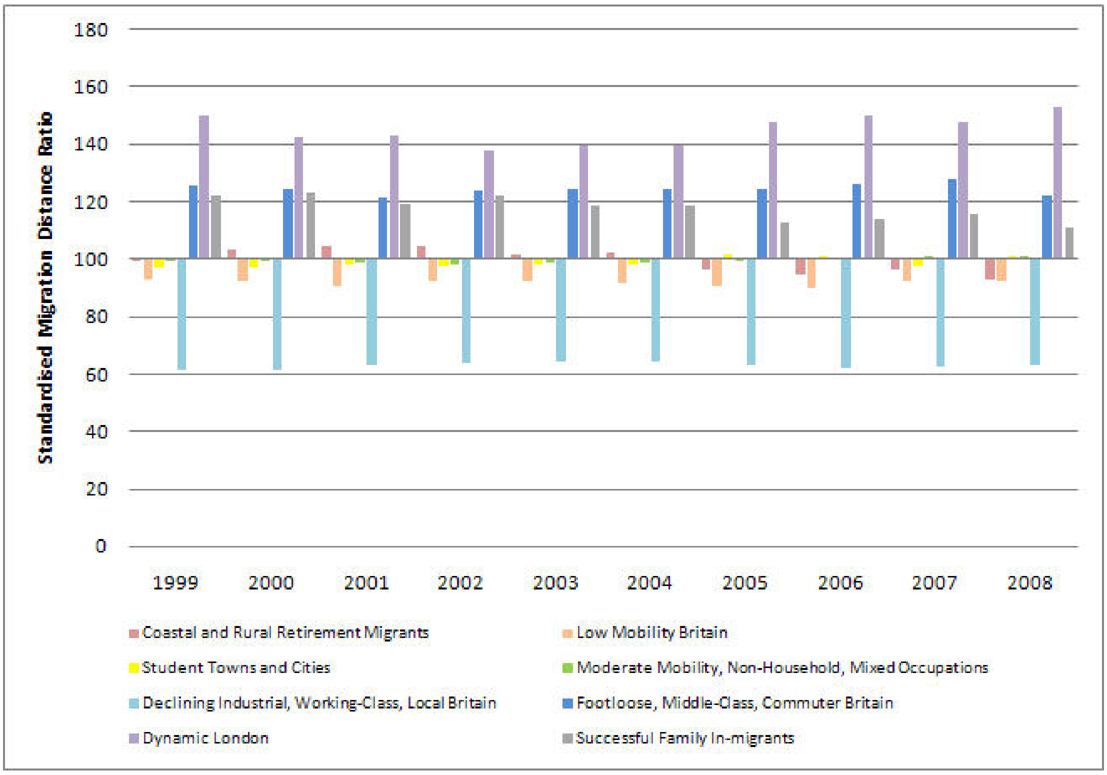
$$E_{ID} = P_{ID}\left(\frac{MC_{+D}}{P_{+D}}\right) \tag{7.5}$$

where

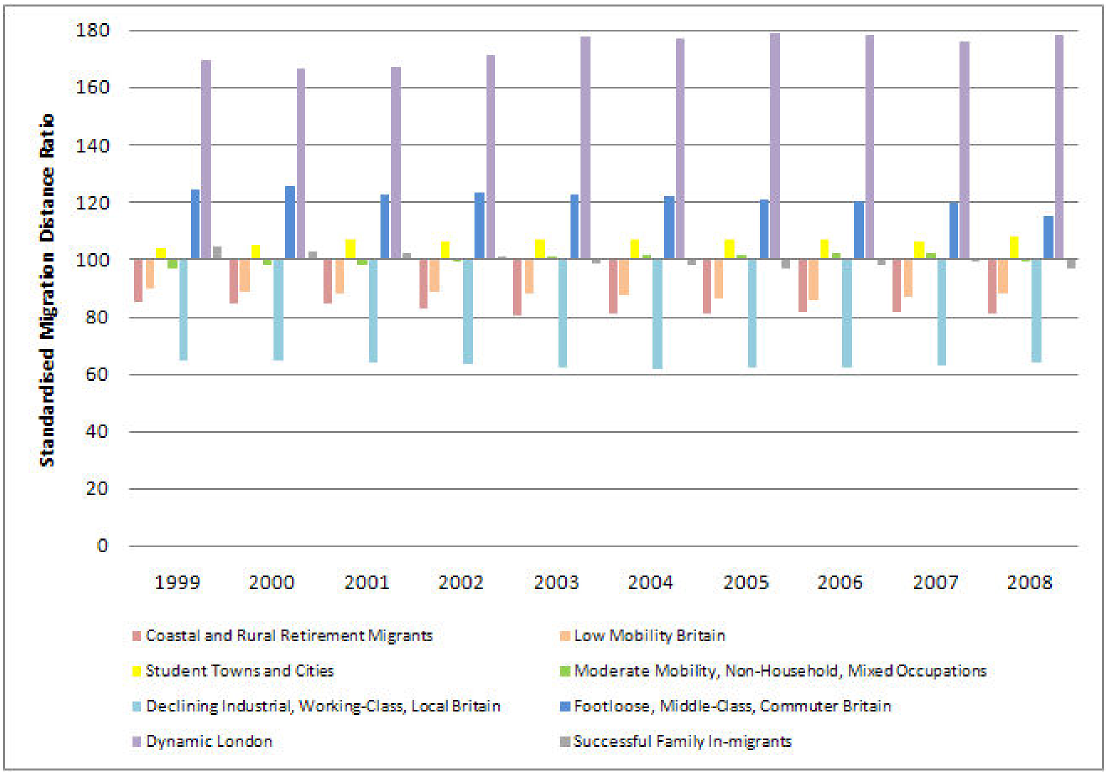
P_{ID} = population associated with Migration Classification cluster I in distance band D.

SMDRs were calculated for both in- and out-migration flows (including both inter and intra-cluster flows in the calculations), with some revealing outcomes. These are displayed in Figure 7.7a and 7.7b. The SMDR ignores the effect that distance might have on encouraging more flows between districts that are closer together, and fewer flows between districts that are further apart. The effect is clear to see. A ratio of 100 represents the expected level of flows given the distribution across the whole system. A ratio of more than 100 shows that there are more flows than expected when distance is ignored, with a ratio of less than 100 showing the opposite. Taking in-migration first, we can see that consistently across the decade, Dynamic London experiences more flows than would be expected, and far more flows than any other cluster. Footloose, Middle Class Commuter Britain and Successful Family In-Migrants also consistently exhibit more in-migration than expected when controlling for distance. On the flip-side, Declining Industrial, Working-Class, Local Britain has a SMDR of well below 100 for in-migration, showing that it experiences considerably less migration than would be expected when controlling for distance. Other clusters are much closer to 100 for in-migration, suggesting that the levels of in-migration that they experience are more or less what would be expected, regardless of the frictional effect of distance. Only Coastal and Rural Retirement Migrants changes from a positive to a negative ratio over the decade. This suggests that there is a slight reduction in the number of migrants that would be expected to move into these areas over the decade.

Taking the out-migration SMDR, clearly Dynamic London is experiencing far more out-migration than would be expected, even when all short-distance out-migration moves are taken



(a) SMDR In-migration



(b) SMDR Out-migration

Figure 7.7: Standardised migration distance ratios by cluster

into consideration. Again this is consistent across the decade, and is in fact a higher ratio than the in-migration SMDR. Footloose, Middle Class Commuter Britain similarly has consistent, increased migration activity. Again, Declining Industrial, Working-Class, Local Britain has very much lower than expected migration. To a lesser extent, but still of note, Coastal and Rural Retirement Migrants and Low Mobility Britain have consistently less out-migration than would be expected.

From this initial analysis of the aggregate patterns, we can start to construct an interesting sub-national migration profile over the decade. Consistently, Dynamic London comes out as the cluster with the highest levels of migration, with the overall balance very heavily towards out-migration rather than in-migration, conforming very much to its classification profile. At the other end of the scale, Declining Industrial, Working-Class, Local Britain also conforms to its cluster definition, exhibiting very low levels of migration overall. Other clusters also appear to have overall internal migration exchanges which also match the profiles presented in the classification with, in many cases, remarkable consistency over the decade. If anything, clusters do seem to be experiencing a drop in migration activity towards the end of the decade. Furthermore, the patterns that come out of this analysis can be attributed to the characteristics of the clusters rather than any spatial association. Examination of SMDRs which control for the effect of distance show that for a cluster like Dynamic London, where high levels of migration could be attributed to the proximity of districts within the cluster, it is in fact the case that these high levels of migration would be experienced anyway.

Cluster associations

Thus far, only overall in-migration, out-migration and net migration patterns for clusters have been analysed; the relationships between the clusters have not been looked at. Before moving on to examine the age disaggregation of flows, it will be interesting to examine the relationships between the clusters. Table 7.5 ranks the net migration relationships by the average across the decade using the same calculation employed by Dennett and Stillwell (2009), which uses the sum of the origin and destination PAR to calculate the net rate, thus meaning a positive net balance in one direction equals the same negative net balance in the other. Table 7.6 ranks the turnover relationships between the clusters. Clearly there is a strong relationship between Dynamic London and Footloose, Middle-Class, Commuter Britain. In net terms, the former is losing to the latter at an average rate of 26 people per 1,000 over the decade (a rate significantly higher than 11 people per 1,000 shown between comparable areas of Outer London and Commuter Belt in the Vickers et al. classification in Table 4.5). This rate peaks around 2003/2004. This association is confirmed with the high levels of turnover between the two clusters. Dynamic London is also losing a considerable number of migrants in net terms to Moderate Mobility, Non-Household, Mixed Occupations. Moving down the hierarchy, these two destinations of out-migration from Dynamic London then become origins for moves to Coastal and Rural Retirement Migrants and Successful Family In-migrants, perhaps suggesting

Table 7.5: Net migration rate balances (per 1,000 population) between origin and destination clusters, 1999-2008

Origin	Destination	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	Average
Dynamic London	Footloose, Middle-Class, Commuter Britain	20.92	23.12	21.26	27.30	30.62	30.42	24.78	26.80	28.76	26.07	26.00
Dynamic London	Moderate Mobility, Non-Household, Mixed Occupations	18.61	21.24	20.38	23.35	25.71	24.86	22.86	20.93	23.21	22.76	22.39
Footloose, Middle-Class, Commuter Britain	Successful Family In-migrants	10.88	13.23	11.96	13.62	12.07	11.52	7.86	7.75	8.86	7.23	10.50
Footloose, Middle-Class, Commuter Britain	Coastal and Rural Retirement Migrants	8.05	11.37	11.35	12.93	12.02	12.00	8.33	7.01	7.48	5.90	9.64
Moderate Mobility, Non-Household, Mixed Occupations	Coastal and Rural Retirement Migrants	7.70	10.19	9.51	11.92	12.12	11.72	8.37	7.69	9.13	6.55	9.49
Moderate Mobility, Non-Household, Mixed Occupations	Successful Family In-migrants	8.33	9.28	9.15	10.51	11.07	11.20	9.21	9.12	8.87	6.92	9.37
Dynamic London	Coastal and Rural Retirement Migrants	6.58	8.32	8.81	10.87	10.82	10.63	7.54	6.04	6.16	5.19	8.10
Dynamic London	Low Mobility Britain	5.51	6.42	6.69	8.52	9.25	9.83	8.20	6.71	8.11	7.05	7.63
Student Towns and Cities	Low Mobility Britain	5.39	7.02	6.54	7.35	8.32	7.97	5.73	5.63	6.98	5.90	6.68
Dynamic London	Successful Family In-migrants	5.46	6.84	6.21	8.69	8.85	8.89	6.02	5.54	5.19	3.97	6.57
Low Mobility Britain	Coastal and Rural Retirement Migrants	5.23	6.40	6.22	6.77	7.01	7.03	5.25	4.55	5.12	4.76	5.83
Student Towns and Cities	Declining Industrial, Working-Class, Local Britain	2.93	3.17	5.15	6.25	8.30	8.73	5.65	5.16	5.36	4.41	5.51
Low Mobility Britain	Successful Family In-migrants	4.88	5.68	5.58	5.67	5.95	6.22	5.17	4.50	5.17	4.46	5.33
Student Towns and Cities	Successful Family In-migrants	3.08	5.05	3.94	6.03	5.54	5.60	2.81	3.29	5.21	2.78	4.33
Declining Industrial, Working-Class, Local Britain	Successful Family In-migrants	4.60	5.13	4.19	4.61	3.73	3.85	3.55	3.74	3.96	3.72	4.11
Declining Industrial, Working-Class, Local Britain	Coastal and Rural Retirement Migrants	4.72	5.20	5.09	4.65	3.48	4.08	3.02	3.44	3.90	3.25	4.08
Moderate Mobility, Non-Household, Mixed Occupations	Footloose, Middle-Class, Commuter Britain	1.09	0.66	2.11	3.33	3.92	5.01	4.43	5.58	6.21	3.51	3.58
Student Towns and Cities	Coastal and Rural Retirement Migrants	1.99	3.19	2.73	4.20	4.70	5.39	2.52	1.90	3.08	1.73	3.14
Successful Family In-migrants	Coastal and Rural Retirement Migrants	1.24	2.98	4.74	5.05	3.71	3.19	2.39	1.38	2.38	1.11	2.82
Student Towns and Cities	Dynamic London	5.98	3.91	3.65	-0.37	-2.71	-3.04	-1.23	0.85	2.93	4.32	1.43
Footloose, Middle-Class, Commuter Britain	Low Mobility Britain	2.08	2.11	1.58	1.51	1.22	1.05	0.87	0.53	0.42	0.37	1.18
Footloose, Middle-Class, Commuter Britain	Student Towns and Cities	-0.16	0.54	0.22	-0.10	1.25	1.56	2.48	1.80	0.00	-0.12	0.75
Student Towns and Cities	Moderate Mobility, Non-Household, Mixed Occupations	1.48	2.19	1.79	1.05	0.15	-0.09	-0.54	-0.39	0.69	0.45	0.68
Declining Industrial, Working-Class, Local Britain	Low Mobility Britain	1.21	0.95	0.30	0.42	-0.07	-0.17	0.11	0.77	0.61	0.57	0.47
Dynamic London	Declining Industrial, Working-Class, Local Britain	-1.19	-0.85	-0.26	0.95	2.10	1.99	1.15	0.46	0.11	-0.33	0.41
Moderate Mobility, Non-Household, Mixed Occupations	Declining Industrial, Working-Class, Local Britain	-0.99	-0.57	-0.33	0.94	1.11	1.45	0.66	0.53	0.22	0.12	0.31
Moderate Mobility, Non-Household, Mixed Occupations	Low Mobility Britain	-0.98	-0.24	-0.26	-0.70	0.17	0.26	1.25	0.03	1.07	1.43	0.20
Footloose, Middle-Class, Commuter Britain	Declining Industrial, Working-Class, Local Britain	-0.95	-0.88	0.03	0.64	0.96	1.12	0.26	-0.02	-0.15	-0.42	0.06

Table 7.6: Turnover exchanges (per 1,000 population) between origin and destination clusters, 1999-2008

Cluster	Cluster	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	Average
Footloose, Middle-Class, Commuter Britain	Dynamic London	10.72	10.91	10.47	10.88	11.18	11.28	10.92	11.13	11.66	10.96	11.01
Student Towns and Cities	Declining Industrial, Working-Class, Local Britain	8.88	9.16	8.99	9.19	9.29	9.20	8.89	8.76	9.07	8.97	9.04
Moderate Mobility, Non-Household, Mixed Occupations	Dynamic London	6.97	7.31	7.18	7.41	7.76	7.69	7.61	7.72	7.92	7.69	7.53
Successful Family In-migrants	Coastal and Rural Retirement Migrants	7.60	7.90	7.67	7.81	7.56	7.66	6.96	7.10	7.62	6.97	7.49
Moderate Mobility, Non-Household, Mixed Occupations	Footloose, Middle-Class, Commuter Britain	7.57	7.65	7.19	7.46	7.45	7.35	7.15	7.21	7.62	6.94	7.36
Student Towns and Cities	Coastal and Rural Retirement Migrants	7.37	7.65	7.57	7.61	7.49	7.46	7.11	7.07	7.25	6.92	7.35
Successful Family In-migrants	Student Towns and Cities	7.44	7.59	7.27	7.48	7.34	7.38	6.95	7.05	7.34	6.98	7.28
Student Towns and Cities	Low Mobility Britain	6.95	7.12	6.92	7.08	7.17	6.93	6.66	6.64	6.91	6.80	6.92
Student Towns and Cities	Dynamic London	5.67	5.81	5.95	5.96	6.02	6.09	6.06	6.05	5.98	6.13	5.97
Student Towns and Cities	Footloose, Middle-Class, Commuter Britain	5.93	5.94	5.87	5.94	5.95	5.98	5.92	5.89	5.86	5.69	5.90
Successful Family In-migrants	Footloose, Middle-Class, Commuter Britain	5.77	6.04	5.72	5.89	5.70	5.57	5.05	5.09	5.34	4.80	5.50
Successful Family In-migrants	Moderate Mobility, Non-Household, Mixed Occupations	5.42	5.50	5.32	5.52	5.33	5.41	4.96	5.07	5.36	4.84	5.27
Student Towns and Cities	Moderate Mobility, Non-Household, Mixed Occupations	5.03	5.23	5.17	5.21	5.28	5.31	5.23	5.22	5.32	5.21	5.22
Low Mobility Britain	Moderate Mobility, Non-Household, Mixed Occupations	4.50	4.56	4.36	4.52	4.44	4.48	4.27	4.28	4.61	4.39	4.44
Footloose, Middle-Class, Commuter Britain	Coastal and Rural Retirement Migrants	4.34	4.70	4.51	4.70	4.53	4.48	3.99	3.95	4.08	3.69	4.30
Moderate Mobility, Non-Household, Mixed Occupations	Coastal and Rural Retirement Migrants	4.14	4.49	4.40	4.47	4.39	4.43	4.00	3.97	4.19	3.88	4.24
Declining Industrial, Working-Class, Local Britain	Coastal and Rural Retirement Migrants	4.22	4.26	4.21	4.24	4.06	4.06	3.76	3.80	3.89	3.75	4.03
Low Mobility Britain	Footloose, Middle-Class, Commuter Britain	3.89	3.93	3.61	3.78	3.76	3.69	3.37	3.42	3.71	3.38	3.65
Successful Family In-migrants	Low Mobility Britain	3.29	3.38	3.19	3.32	3.23	3.23	2.96	2.96	3.15	2.99	3.17
Low Mobility Britain	Declining Industrial, Working-Class, Local Britain	3.24	3.20	3.13	3.16	3.17	3.09	2.93	2.88	3.04	3.00	3.08
Successful Family In-migrants	Declining Industrial, Working-Class, Local Britain	3.13	3.22	3.10	3.24	3.15	3.09	2.88	2.86	2.98	2.79	3.05
Dynamic London	Coastal and Rural Retirement Migrants	2.96	3.10	3.13	3.11	3.05	2.96	2.70	2.64	2.64	2.51	2.88
Low Mobility Britain	Coastal and Rural Retirement Migrants	2.89	3.00	2.96	3.02	2.90	2.96	2.71	2.64	2.77	2.66	2.85
Successful Family In-migrants	Dynamic London	2.92	2.95	2.82	2.94	2.89	2.87	2.64	2.64	2.59	2.47	2.77
Low Mobility Britain	Dynamic London	2.63	2.60	2.58	2.71	2.72	2.80	2.64	2.54	2.66	2.58	2.65
Moderate Mobility, Non-Household, Mixed Occupations	Declining Industrial, Working-Class, Local Britain	1.71	1.74	1.72	1.80	1.80	1.79	1.67	1.65	1.63	1.62	1.71
Footloose, Middle-Class, Commuter Britain	Declining Industrial, Working-Class, Local Britain	1.64	1.65	1.64	1.69	1.66	1.65	1.52	1.46	1.50	1.42	1.58
Dynamic London	Declining Industrial, Working-Class, Local Britain	1.39	1.41	1.48	1.52	1.62	1.57	1.50	1.44	1.36	1.37	1.47

a follow-on move later in the life course for individuals who had moved earlier in their lives. Again, these are relationships that more-or-less hold across the decade, although there is a slight reduction in the net out-flow towards the end of the time period, if not the turnover indicating a small reversal of flows between these areas. Of some surprise is the high turnover ranking between Student Towns and Cities and Declining Industrial, Working-Class, Local Britain, especially as the net flow is from the former to the latter. These two clusters have a relatively high index of spatial association (Table 7.1), which could be one reason for the relatively high overall flows in both directions, however, as is also shown in Table 7.2, these two cluster are, by a considerable way, the most populous clusters. This will be the main driver behind the increased turnover between the two clusters.

In general, Declining Industrial, Working-Class, Local Britain has low turnover and net migration associations with the majority of the other clusters. This is especially the case with Dynamic London - flows of individuals between districts located in these two clusters are consistently rare across the decade. The only pair of clusters where the net relationship varies noticeably across the ten year period is Student Towns and Cities and Dynamic London. In 1999, there was a relatively high net flow from Student Towns and Cities to Dynamic London. This net flow declined for two years before reversing for four, with a peak net outflow from Dynamic London to Student Towns and Cities in 2004. This pattern then reversed and reverted back to a net inflow to Dynamic London by 2008.

Perhaps an easier way to appreciate the key gross flows between clusters is to examine Figure 7.8. The diagram represents the top 5 gross flow rates between clusters at each age group, again using the sum of the origin and destination PAR to calculate the rate. These rates are the average across the ten year time series. The circles which represent each cluster at each age group are proportional to the average populations across the same period and the arrows representing the flows are proportional to the size of the net flow. Taking age group 0-15 first, the largest rate of flow is within the Dynamic London cluster - an undoubted reflection of the flows within this cluster also exhibiting the highest rate at the 30-44 age group. The second and fifth highest rates are also from Dynamic London, but into the Moderate Mobility, Non-Household, Mixed Occupation and Footloose, Middle Class Commuter Britain clusters. Flow rates within the later also very important within the system. Flow rates in the 0-15 age group are also high within the Declining Industrial, Working Class, Local Britain cluster.

At age group 16-19, it is clear that not only do rates of flow increase, as indicated by the size of the arrows, but one destination predominates - the Student Towns and Cities cluster. The highest net flow comes from Footloose, Middle Class Commuter Britain with, perhaps surprisingly given the low associations with other clusters in general, the next highest rate coming from the Declining Industrial, Working Class, Local Britain cluster. High rates into Student Towns and Cities also come from the Coastal and Rural Retirement and Successful Family In-migrants clusters. As with all age groups, flows within the Dynamic London cluster are also important, but age group 16-19 is the only group where the rate is not the highest.

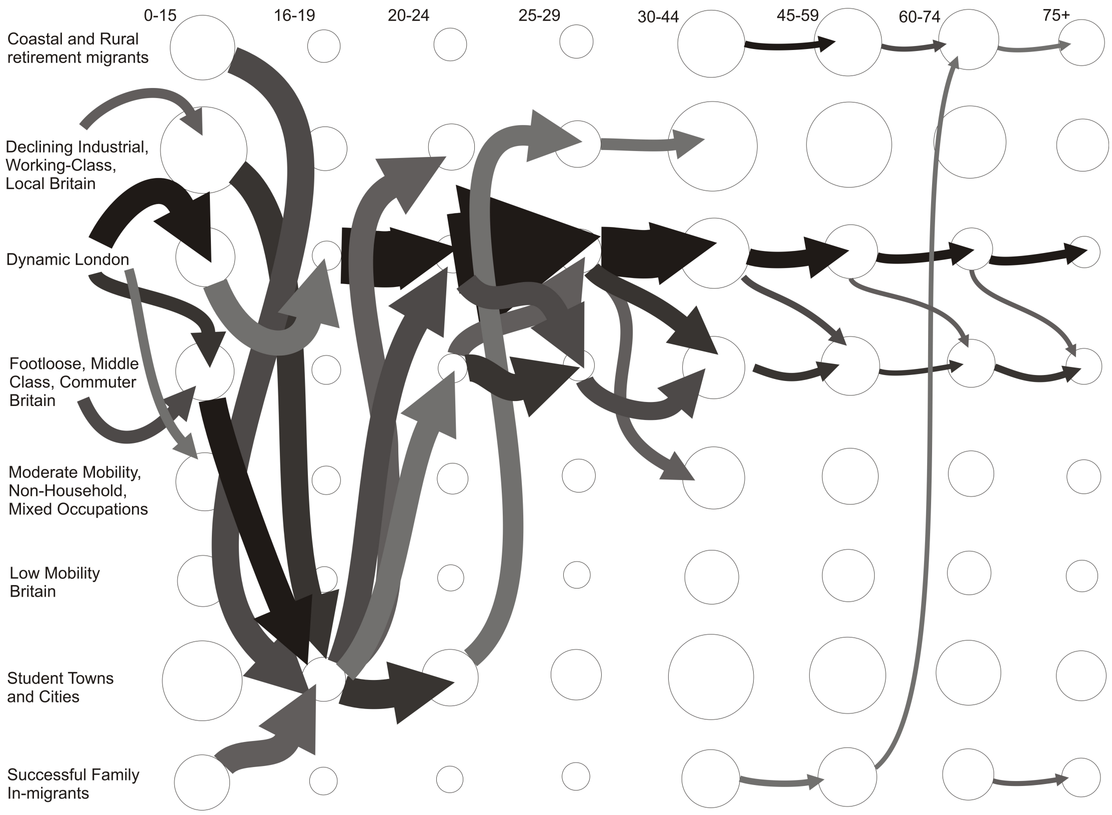
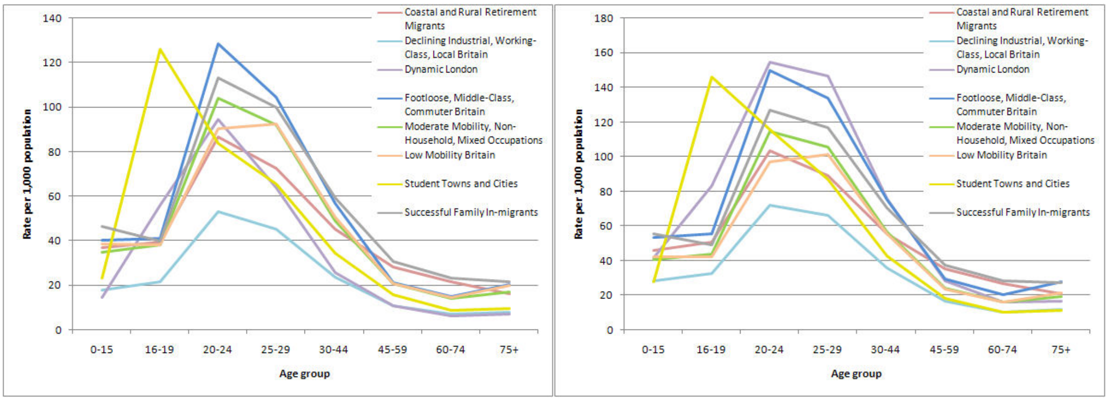


Figure 7.8: Top five gross flow rates between clusters at each age group

At age group 20-24, flow rates increase within London, with the second highest rate coming from the Student Towns Cluster. Aside from flows within the Student Towns and Cities cluster, out-migration rates from this cluster predominate, with flows into Footloose, Middle Class Commuter Britain and Declining Industrial, Working Class, Local Britain. At age group 25-29 the highest flow rates in any of the age groups can be observed, with the highest of all within Dynamic London. High rates are also observed within Footloose, Middle Class Commuter Britain and from Dynamic London to Footloose, Middle Class Commuter Britain. This is a pattern of flow rates within and between these clusters which is maintained until the oldest 75+ age group.

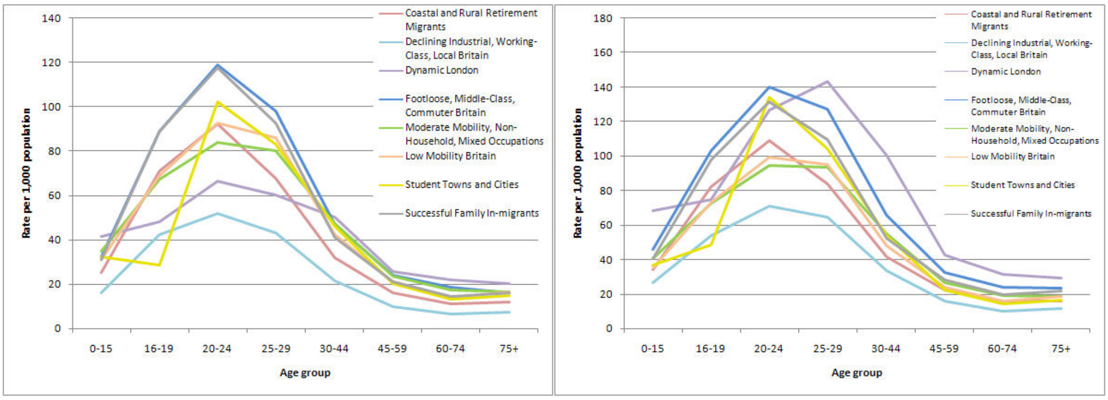
Moving towards the older age groups, the rates of migration drop off considerably, however other clusters begin to increase in importance. From age group 45-59 onwards, rates of flow within the Coastal and Rural Retirement Migrants cluster feature high in the top 5, as do flows within the Successful Family In-migrants cluster, with a net flow from the latter into the former featuring in the top 5 at the 60-74 age group.



(a) Excluding within cluster migrants

(b) Including within cluster migrants

Figure 7.9: Average in-migration rate (per 1,000 people) age schedules, by cluster, 1999-2008



(a) Excluding within cluster migrants

(b) Including within cluster migrants

Figure 7.10: Average out-migration rate (per 1,000 people) age schedules, by cluster, 1999-2008

7.3.2 Patterns by age group over time

Age schedules

Of course, as was shown clearly in the analysis of 2001 Census data in Chapter 4, aggregate migration flows obscure much of the variation in migration flows which happen in different stages of the life course. For a full understanding of the flows which comprise these aggregate patterns, analysis of the age-specific flows should be carried out. It is first useful to look at the average migration age schedules for each of the clusters for both in- and out-migration. Figures 7.9a and 7.9b show the in-migration rate schedules for each cluster. It is immediately obvious in both schedules that the peak for Student Towns and Cities occurs in the 16-19 age group, rather than the 20-24 age group where all other clusters peak. This is entirely expected, as the vast majority of undergraduate students beginning university courses will do so at around 18-19.

In Figure 7.9a, in-migration flows from districts in the same cluster are ignored, hence

Dynamic London having only the fourth highest rate of in-migration at 20-24. The effect of including these flows is clear in Figure 7.10a. As might have been predicted in the light of the earlier analysis, Declining Industrial, Working-Class, Local Britain exhibits the lowest levels of in-migration in almost all age groups, but this is especially noticeable at 20-24 - the age of peak migration propensity. Of particular interest are the clusters which show the highest in-migration rates at each age group, revealing migrant preferences at different stages in the life course. At 16-19, Student Towns and Cities are most important; at 20-24, Dynamic London and clusters associated with successful, middle class and single migrants; at 30-44, Successful Family In-Migrants and Footloose, Middle-Class, Commuter Britain are most important; at 45-49 Coastal and Rural Retirement Migrants begins to increase in importance along with the mainly rural Successful Family In-Migrants, a trend maintained until 75+ when Footloose, Middle-Class, Commuter Britain again increases in importance.

The out-migration schedules shown in Figures 7.10a and 7.10b are equally as revealing. Declining Industrial, Working-Class, Local Britain has very low out-migration rates. At 16-19, Footloose, Middle-Class, Commuter Britain and Successful Family In-Migrants have the highest out-migration rates, indicating that these are the areas of parental domicile for many of the students migrating into Student Towns and Cities. There are then comparatively high rates of out-migration from Student Towns and Cities in the 20-24 age group, suggesting that many graduates do not remain in their place of study after graduating. Interestingly, the peak out-migration rate for Dynamic London when within-cluster migrants are included in the data (Figure 7.10b), is at 25-29 rather than the 20-24 peak shown in all other clusters. As the peak is also different to the Dynamic London peak in Figure 7.9b, it suggests that at 25-29, migrants are tending to move from central to more peripheral areas of the cluster, perhaps in line with an increase in affluence. However, it should also be noted that the undercount of males at the 20-24 age group present in the PRDS data could potentially be having an influence on this particular peak, and so any conclusions should be made cautiously. At 30-44, Dynamic London maintains the highest rate of out-migration, with the rate being considerably higher than for all other clusters when within cluster moves are included. Indeed, for all older age groups and with and without within cluster flows, Dynamic London maintains the highest out-migration rates.

Standardised Migration Ratios

As with the earlier aggregate analysis where the confounding effects of distance were examined, it is useful to attempt to unpick the potentially distorting effects that age may have on the migration profiles of each cluster. The effect of age on migration behaviours is undeniable, but given that each of the clusters in the migration classification will have different age profiles, assessing the extent of the influence of age is important. Of course, unlike distance, age played an important part in the definition of the clusters, so to an extent, the influence of age on each cluster is already known. However, disassociating the effects of age is a useful exercise, given the overwhelming importance of age on migration propensities and patterns.

Where earlier the SMDR was used, here a very similar ratio borrowing its definition even more closely from the SMR will be used to account for the influence of underlying age structures on migration flows. As before the ratio is of the total observed migration flows divided by the total expected migration flows for each cluster. Where before the totals for each cluster were summed across the distance ranges, here they are summed across the age groups. The PAR used to calculated the expected flows this time vary by age group. This ratio will be referred to as the Standardised Migration Ratio (SMIR). As with the SMDR, ratios can be calculated for both in- and out-migration. As an example, for a matrix of flows between Migration Classification origins MC_I and age bands A a series of observed migration flows MC_{IA} can be defined. Standardised Migration Ratios (SMIRs) can be calculated as the ratio between these observed flows to expected flows such that:

$$SMIR = 100\left(\frac{MC_{I+}}{E_{I+}}\right) \tag{7.6}$$

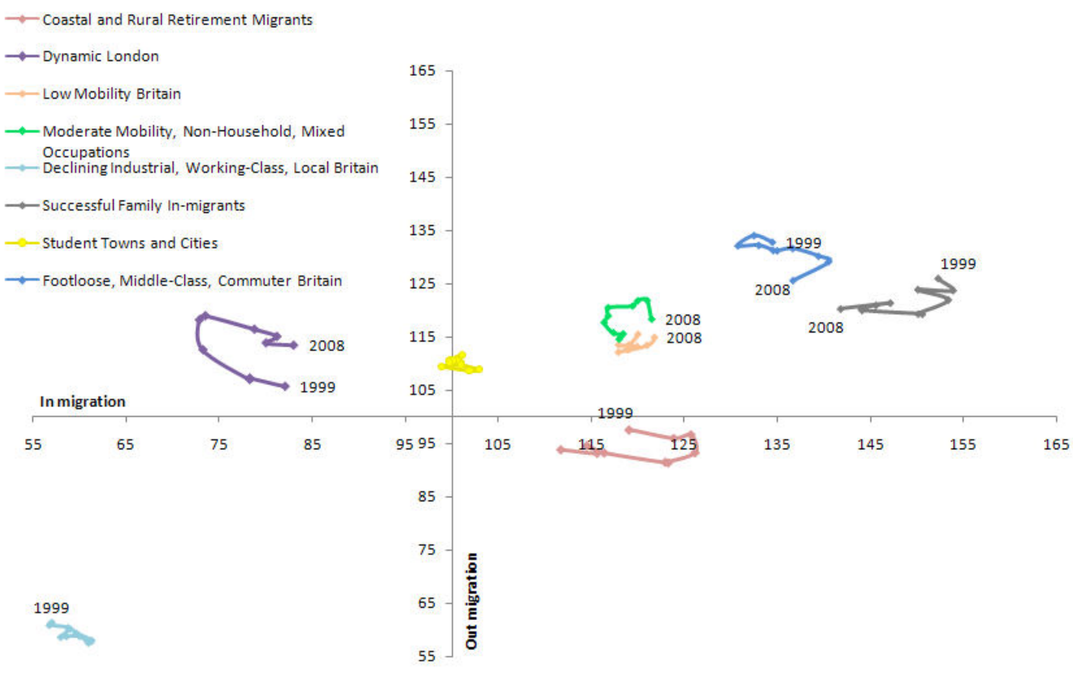
where
E_{IA} = the expected migration to or from Migration Classification cluster I in age group A and

$$E_{IA} = P_{IA}\left(\frac{MC_{+A}}{P_{+A}}\right) \tag{7.7}$$

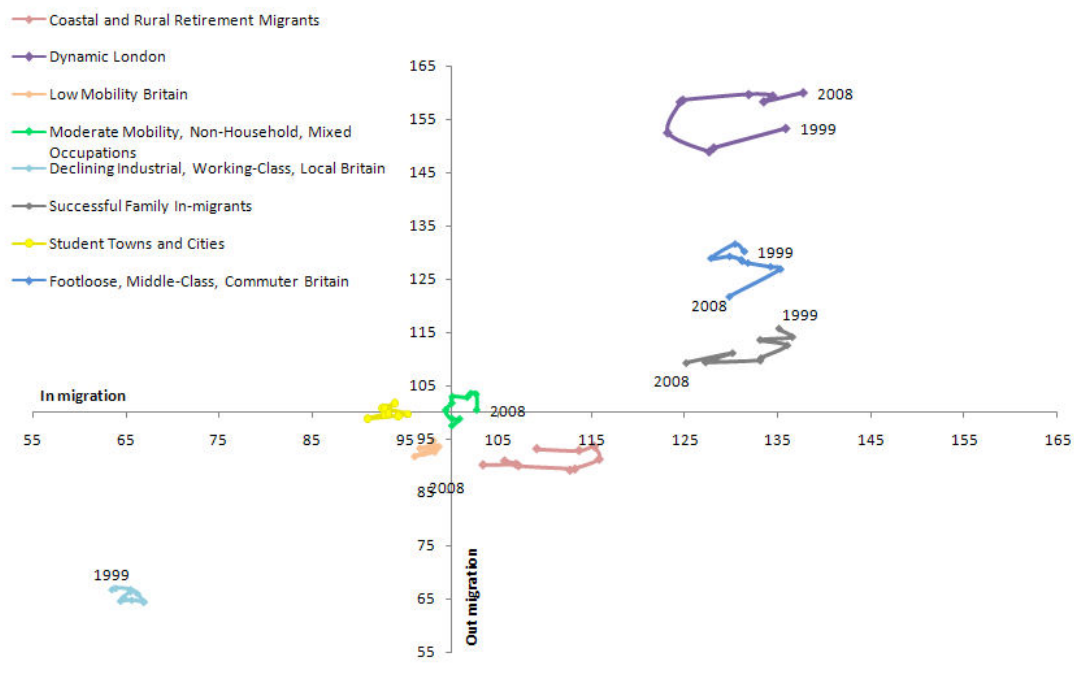
where
P_{IA} = population associated with Migration Classification cluster I in age group A.

Figures 7.11a and 7.11b show the SMIR trajectories for all classification clusters between 1999 and 2008. The graphs are divided into four quadrants, with the point where the x and y axis cross depicting the ratio of 100 for both in- and out-migration. A ratio of 100 represents the system-wide expected migration rate, given the age structure across all clusters. Therefore the top right quadrant represents both in- and out-migration that are higher than would be expected; the bottom left, in- and out-migration lower than expected; the bottom right, in-migration higher than expected, but out-migration lower than expected; and the top right, out-migration higher than expected and in-migration lower than expected. The lines for each cluster represent the full time series of migration from 1999 to 2008 and the trajectory of the migration patterns (e.g. whether in- and out-migration are increasing or decreasing in relation to the year-on-year average over the decade).

As has been seen already, the inclusion or exclusion of within cluster flows has a large effect on Dynamic London. Where within flows are excluded, the cluster has lower than expected levels of in-migration. When they are included, this flips to much higher than expected in-migration. With out-migration the flows become even higher than expected. Where the inclusion of within-cluster flows has an effect on the other clusters, the effect is in the opposite



(a) SMIR Excluding within cluster migrants



(b) SMIR Including within cluster migrants

Trajectory graphs after Baccaini (2007)

Figure 7.11: Standardised migration ratios by cluster type, 1999-2008

direction, with Low Mobility Britain, Moderate Mobility, Non-Household, Mixed Occupations and Student Towns and Cities all moving closer to expected levels of in- and out-migration. This suggests inter-cluster flows are of more importance than intra-cluster flows for these clusters, and especially for in-migration. The trajectory of the SMIRs across the decade for Dynamic London is varied. Between 1999 and 2003 the in-migration SMIR reduces whilst the out-migration SMIR increases, suggesting an acceleration the net loss of all migrants from districts in this cluster over this period. From 2004 onwards the trajectory reverses again with a very similar in-migration SMIR in 2008 to 1999. Out-migration levels in 2008 are still comparatively higher than in 1999, although in all years Dynamic London experiences higher than expected out-migration, with in-migration patterns being much less stable than out-migration.

Following an almost mirrored trajectory to Dynamic London is Coastal and Rural Retirement Migrants. Situated in the opposite quadrant (when within cluster flows are excluded), this cluster experiences lower than expected out-migration, but higher than expected in-migration. Unlike dynamic London the position of this trajectory changes very little with the inclusion and exclusion of intra-cluster flows. From 1999, the SMIR for in-migration steadily increases until 2002. At this time the out-migration SMIR reduces slightly, with the in-migration SMIR then reducing steadily until 2008. Like Dynamic London, it is the in-migration flows that are less stable than the out-migration flows, with far more variation across the *x* axis than up or down the *y* axis of the graph.

Other clusters with varied SMIR trajectories are Footloose, Middle Class, Commuter Britain and Successful Family In-Migrants. Both have in- and out-migration SMIRs above 100 signifying higher than expected migration rates when age structure is accounted for. Taking the latter first, it is possible to see a clear but somewhat erratic decline in both the in- and out-migration SMIRs between 1999 and 2008, indicating a reduction in levels of migration relative to all other clusters. It is a similar story for Footloose, Middle Class, Commuter Britain in that out-migration has declined, although between 2001 and 2007, levels of in-migration relative to other clusters increased.

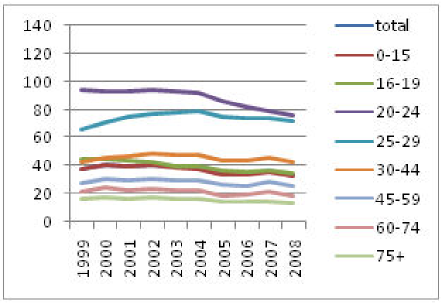
The clusters of Low Mobility Britain, Moderate Mobility, Non-Household, Mixed Occupations, Student Towns and Cities and Constrained, Middle Class, Local Britain all show much less variation in their SMIR trajectories over the decade. Where the inclusion of intra-cluster flows reduces the SMIR closer to 100 for the first three of these clusters, it has very little effect on Declining Industrial, Working-Class, Local Britain. Here the only cluster which consistently displays lower than expected SMIRs for both in- and out-migration also varies very little where intra-cluster flows are included.

What analysis of SMIRs shows, is that some clusters, especially Student Towns and Cities, owe much of their migration profile to the ages of the migrants who move into and out of the cluster. When age is adjusted for, both the in- and out-migration are close to the average national picture. It is a similar case with Low Mobility Britain and Moderate Mobility, Non-Household, Mixed Occupations, but only when intra-cluster flows are considered. When just examining in-flows and out-flows, these clusters exhibit higher levels of migration than would be expected when the age structure of the population is taken into consideration, but with each the level of in- and out-migration relative to the national average changes very little over the decade. Other clusters like Dynamic London and Coastal and Rural retirement migrants vary far more over the decade, suggesting changes in the age structure of (mainly) their in-migrants. Dynamic London experienced an increase and then decrease in the numbers of young, in-migrants (which will push the SMIR down and then up); Coastal and Rural retirement migrants an initial increase followed by a marked decrease in older migrants (which will push the SMIR up and then down). Analysis of SMIRs also confirms the age-unrelated importance of Footloose, Middle Class, Commuter Britain and Successful Family In-Migrants as migrant origins and destinations across the decade, although it is an importance that is in slight decline. It also confirms that, regardless of any age related influence, Declining Industrial, Working-Class, Local Britain remains the cluster with least migrant activity associated with it, and this changes very little across the decade.

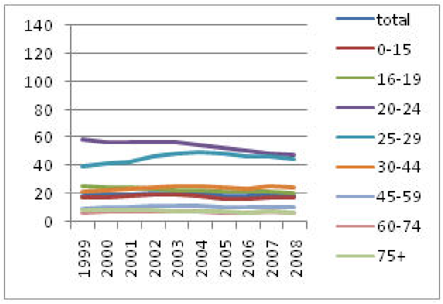
Variation over the decade

It is clear from SMIR analysis that, in the case of some clusters if not all, there has been some variation in migration patterns across the decade. To conclude this commentary of migration between 1999 and 2008, it will be interesting to examine the changes in in- and out-migration propensities by age group for each of the clusters in the Migration Classification. Figures 7.12 and 7.13 detail for each cluster the changes in age-specific migration rates per 1,000 people over the ten year period. Of note first is that the largest variation occurs in the age groups with the highest propensities to migrate. For both in- and out-migration, there is little variation between 1999 and 2008 when the propensity to migrate is low. Where the propensity to migrate is high, the tendency in all clusters is for migration rates to drop off between 1999 and 2008. For all clusters, the rate of in-migration for the 20-24 age group reduces noticeably (except maybe Dynamic London where there is a very slight increase from the middle to the end of the decade). The same is true for out-migration except in the case of Dynamic London where there is little overall decline (and indeed an increase towards the middle of the decade). This is the exact same decline which was shown overall, earlier in Figure 7.2. Reasons for this marked decline are unclear, especially as it is happening across all clusters, and for both in- and out-migration. Another noticeable decline is in the out-migration propensity of the 16-19 age group from clusters where out-migration rates are highest across the decade. The only cluster where this decline in out-migration is not happening is Dynamic London. Here, the trend is very much in the reverse; something which is even more noticeable as it is one of the very few trends where propensity is increasing. Other increases in Migration activity happen to a much lesser extent in the 25-29 and 30-44 age groups for some clusters.

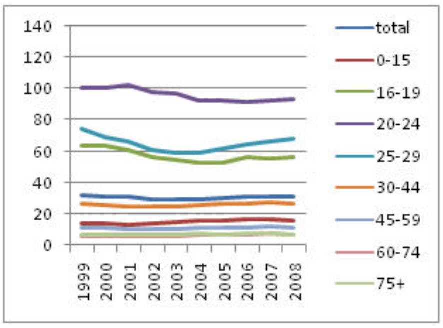
In summary, for the majority of age groups in the majority of clusters, there is little variation in the propensity to migrate across the decade. This lack of variation is useful in the context of



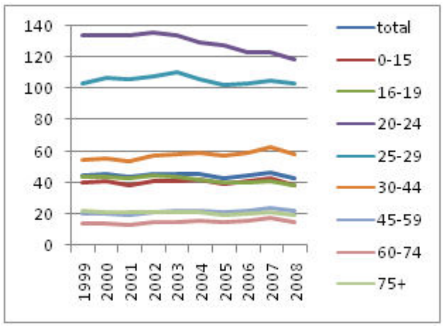
(a) Coastal and Rural Retirement Migrants - In-migration rate per 1,000 population



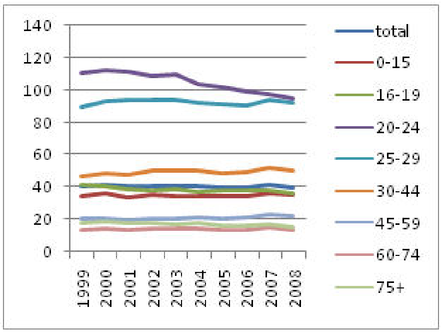
(b) Declining Industrial, Working-Class, Local Britain - In-migration rate per 1,000 population



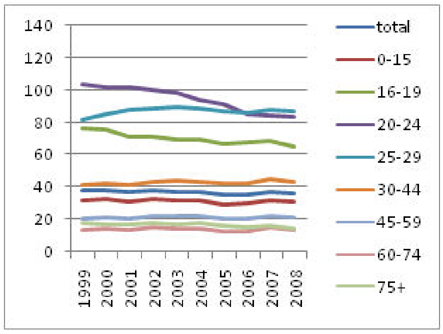
(c) Dynamic London - In-migration rate per 1,000 population



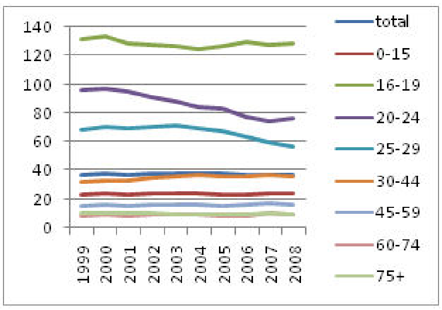
(d) Footloose, Middle Class, Commuter Britain - In-migration rate per 1,000 population



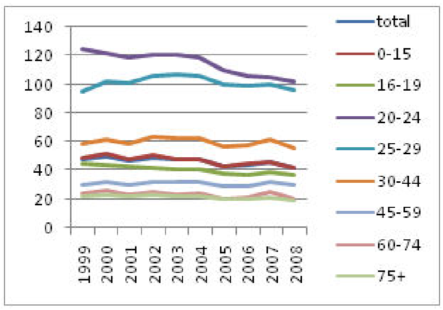
(e) Moderate Mobility, Non-Household, Mixed Occupations - In-migration rate per 1,000 population



(f) Low Mobility Britain - In-migration rate per 1,000 population

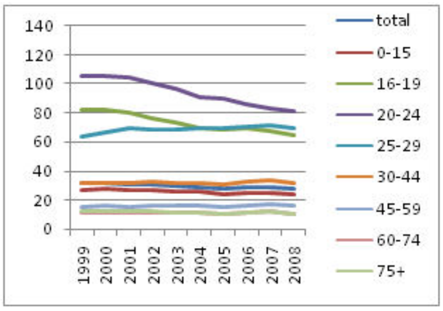


(g) Student Towns and Cities - In-migration rate per 1,000 population

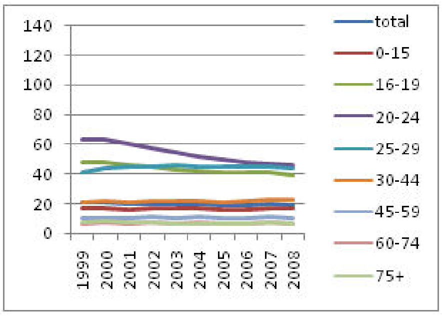


(h) Successful Family In-migrants - In-migration rate per 1,000 population

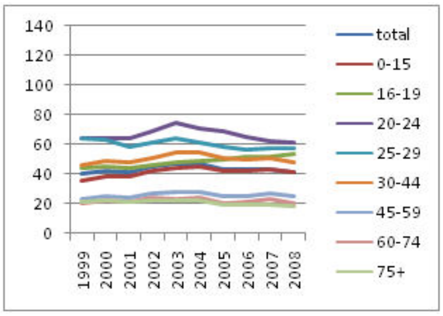
Figure 7.12: In-migration flow rates per 1,000 population, 1999-2008, by age group and cluster



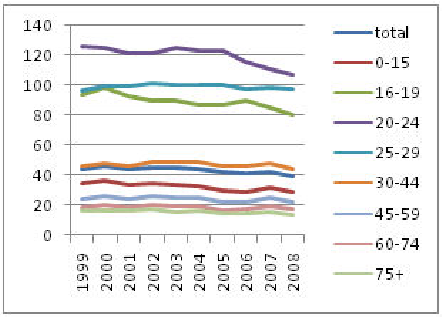
(a) Coastal and Rural Retirement Migrants - Out-migration rate per 1,000 population



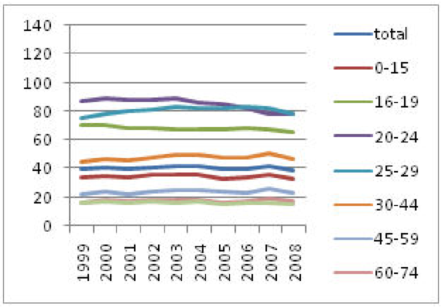
(b) Declining Industrial, Working-Class, Local Britain - Out-migration rate per 1,000 population



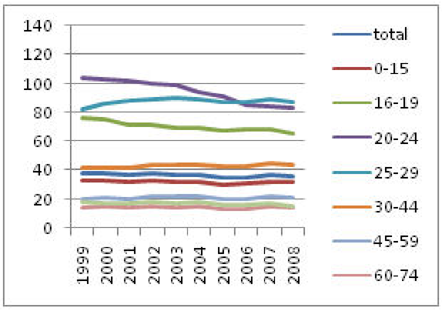
(c) Dynamic London - Out-migration rate per 1,000 population



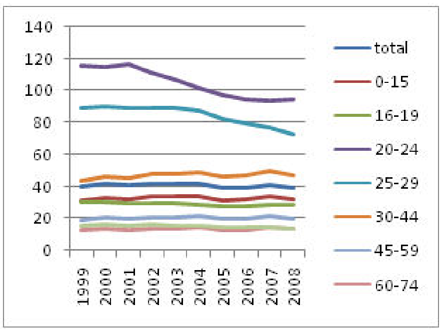
(d) Footloose, Middle Class, Commuter Britain - Out-migration rate per 1,000 population



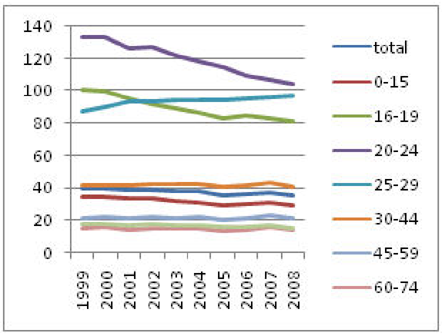
(e) Moderate Mobility, Non-Household, Mixed Occupations - Out-migration rate per 1,000 population



(f) Low Mobility Britain - Out-migration rate per 1,000 population



(g) Student Towns and Cities - Out-migration rate per 1,000 population



(h) Successful Family Out-migrants - Out-migration rate per 1,000 population

Figure 7.13: Out-migration flow rates per 1,000 population, 1999-2008, by age group and cluster

projecting migration into the future. Where there is variation, for example with the 20-24 age group, there appears to be a declining trend.

7.4 Value added by the classification

The preceding discussion has presented an analysis of a ten-year time series of migration data up to 2008. In some respects, the classification based analysis carried out here could be seen as similar to the analysis carried out in Chapter 4 using the Vickers et al. (2003) classification of districts, in that the vast complexity of inter-district flows in the Britain has been reduced to a more manageable set of inter-cluster flows. Where this analysis differs crucially, though, is in the use of the Migration Classification and the definition of the clusters used. As has been discussed in Chapter 5 and also by Duke-Williams (2009b), migrants are not necessarily representative of the general population in an area, certainly in the case of in-migrants, and potentially in the case of some out-migrants (the extent to which a transient student population represents an actual population can be debated). Where other classifications have categorised areas by their resident population, here these sedentary populations are ignored in favour of migrant population characteristics. Part of the real value added by the use of the Migration Classification is that it disentangles some of the complex migration streams occurring in Britain and allows patterns to be separated where before they may have been confounded through the non-migrant properties of a general purpose classification.

For example, consider the two maps shown in Figures 4.1 and 6.16. It is apparent that much of the Urban UK cluster in the Vickers et al. classification features districts that are also in the Declining Industrial, Working-Class, Local Britain Cluster in the Migration Classification. Yet analysis of the two clusters reveals very different patterns. The Urban UK cluster exhibits relatively high in- and out-migration rates, whereas in contrast low are rates exhibited in Declining Industrial, Working-Class, Local Britain. The reason for this difference is clear when the clusters are examined in more detail. A number of high internal migration rate districts feature in Urban UK which are allocated to different clusters in the Migration Classification - combining districts with very different migration profiles confounds analysis by obscuring differences. Of course at the lower Group level in the Vickers et al. classification, some of these districts are separated and patterns more akin to those shown from analysis using the Migration Classification are displayed, but with a general purpose classification, a satisfactory division of areas for a more specific purpose will be hard to achieve. Any useful division is likely to be more through serendipity than design. Even where similar patterns are shown, as was noted in Section 7.3.1 the magnitude of the association may be obscurred by a general purpose classification.

In addition, the migration classification adds value to migration analysis through the detailed cluster profiles which were created during the classification building process. This analysis has demonstrated that internal migration within and between clusters across the decade has remained stable. This stability means that it is possible to utilise the additional details about the types of migrant which helped define the original clusters, to add detail and therefore value to the data presented. As a brief example, migrants moving into districts in the Coastal and Rural Retirement Migrants cluster are less likely to be from the highest socio-economic groups and more likely to move into owner-occupied accommodation and be alone or in couples, rather than in families; or migrants moving into and out of districts in the Moderate Mobility, Non-Household, Mixed Occupations cluster will be more likely to moving into and out of rented accommodation and move alone or in non-family households.

7.5 Conclusions

At the beginning of this chapter a gap in the understanding of recent internal migration patterns in Britain was identified, with the main substantive aim of this chapter being to fill this gap in the current knowledge. The estimation of a full set of 90 inter-district flow matrices for Britain, disaggregated by broad age group for a time-series of ten years in Chapter 2 paved the way for a complete analysis of a decade of migration flows in the context of the new Migration Classification.

This analysis produced a number of key findings. Firstly, overall volumes and rates of migration remain remarkably consistent between and within clusters across the decade. Overall, there is a slight decrease in the volume of migrants between mid 1999 and mid 2008, with some clusters such as Coastal and Rural Retirement Migrants and Successful Family In-Migrants decreasing their migrant exchanges most noticeably. However, each cluster maintains a distinct migration volume and rate profile over the decade.

The second key finding is that each cluster in the Migration Classification continues to exhibit a distinct in- and out-migration profile across the decade and that this profile can be judged to be independent of the spatial associations between districts within the cluster. The introduction of the SMDR allowed for the effects of distance to be disentangled from their influence on the volume of flows into, out of and within clusters. Consistently Dynamic London, even when accounting for its compact nature, exhibits higher than expected levels of in- and out-migration across the decade. At the opposite end of the scale, Declining Industrial, Working-Class, Local Britain has very much lower levels of migration across the decade.

The third key finding is some clusters are more linked through their migration flows than others and that these associations are maintained over the decade. Dynamic London and Footloose, Middle-Class, Commuter Britain have the highest inter-cluster association with the former the source of many migrants to the latter. Dynamic London also has high levels of association with Moderate Mobility, Non-Household, Mixed Occupations, being both a net provider of migrants to this cluster, but also, as turnover statistics show, a net gainer with flows in the other direction. Other high associations were between the clusters which are the destination of choice for those leaving Dynamic London - Footloose, Middle-Class, Commuter Britain; Moderate

Mobility, Non-Household, Mixed Occupations; Successful Family In-Migrants, and Coastal and Rural Retirement Migrants. The cluster least linked with any other cluster is Declining Industrial, Working-Class, Local Britain. This cluster features in the three lowest rates of turnover, but perhaps more surprisingly these low rates of association are with the three most active clusters, suggesting a high degree of detachment from areas where otherwise migration flows are common.

The fourth key finding is that the propensity to migrate with age varies considerably by cluster. In most cases, an age schedule curve reminiscent of those reported by Rogers and Castro (1981) is present for both in- and out-migration, albeit with significant variations in amplitude depending on the cluster. However, for Student Towns and Cities - a cluster heavily defined by its student migrants - the curve is very different. For in-migration the peak occurs in the 16-19 age group rather than the 20-24 group. For out-migration at this age, the rate is significantly lower than it is for all other clusters. Similarly, Dynamic London does not follow the standard schedule for out-migration when within-cluster moves are taken into consideration - the out-migration peak occurring in the 25-29 age group rather than the younger 20-24 group. These variations in both amplitude and the location of the crest of the curve along the age continuum are very important, especially when one considers that similar schedules are by ONS to project migration flows sub-nationally across the UK (ONS, 2008a).

Despite the influence of age on the migration profiles of clusters, the use of SMIRs allowed for this effect to be controlled for, showing that the influence of age was more significant for some clusters than others. For example, the Student Towns and Cities exhibits a profile very close to the system average when the effect of the young student migrants is taken into consideration. When variation in the age profiles of all clusters is accounted for it is possible to compare in- and out-migration profiles over time, with noticably more variation in in-migration than out-migration for Coastal and Rural Retirement Migrants, Successful Family In-migrants and Dynamic London, reflecting perhaps changes in the external factors influencing the decisions migrants make to move into new areas, such as rising house prices.

The final key finding is that examining the migration related age profiles for each cluster over time, there is, on the whole, little variation in the age-specific migration rates across clusters over the decade. The main exception to this general observation is that age group 20-24 in almost all clusters experience a decline in both in- and out-migration rates. As was shown for the whole system in Figure 7.3, this decline is a combination of both a reduction in gross migration flows and an increase in the 20-24 population. Indeed, between 1999 and 2003, there is actually a steady increase in the numbers of migrants, it is just that this increase is offset by an even more severe increase in the population of this age group. Overall, despite there being a downward trend in the volume of migrants at 20-24, an equally important factor is the steady increase of the 20-24 PAR across all clusters.

In all, these findings can be seen as importantly advancing the contemporary understanding of internal-migration patterns in Britain, something which should have wider significance for anyone looking to understand population dynamics in the UK. The patterns and trends shown in this analysis in relation to the age, time and space elements associated with migration enhance our understanding of the present and recent history of the phenomenon within Britain. The classification framework employed works effectively as it is a bespoke framework designed specifically for this job.

As well as the substantive objective of this chapter, a secondary methodological objective was also set out in the introduction, with the specific aims of introducing two new metrics which control for the distorting effects of distance (SMDR) and age (SMIR) on the flows of migration within the Migration Classification system. Both of these measures cast new light onto the effect of age an distance on the migration flows to and from districts within the Migration Classification clusters with, for example, Dynamic London experiencing even more in and outflows and Declining Industrial, Working Class, Local Britain experiencing far fewer in and outflows than would be expected when the distorting effects of distance are accounted for. Similarly for this cluster, the age profiles of migrants inflate the observed flows, with considerably below average flow ratios presented when the effects of age are accounted for.

Whilst a comprehensive time-series account of internal migration within Britain has been attempted in this chapter, much of the analysis has tended towards the descriptive rather than the explanatory. To an extent the use of SMIRs and SMDRs moved the description in the direction of explanation through allowing an appreciation of the influence of both distance and age on internal migration flows to be gained by controlling for these factors, but here explanation was more implicit than explicit. For a complete understanding of the internal migration landscape in Britain attention must be turned more overtly towards explanation, which is where the next chapter will continue.

Chapter 8

Understanding a decade of internal migration in Britain - from spatial interaction to life course explanations

8.1 Introduction

In the chapter preceding this, the utility of the Migration Classification was demonstrated in the descriptive analysis of internal migration flows in Britain. Some illustrative comments were made, but on the whole, the scope of the chapter was more exploratory than explanatory. This penultimate chapter in the thesis will draw on the analysis of earlier chapters, but will look to offer explanations for some of the patterns that are presented. Given the thesis has been concerned with aggregate analysis, the explanations will tend towards those which can be applied to the general rather than the individual. Whilst an individual migrant will choose to move for any number of different personal reasons, accounting for these in a more general theory becomes more difficult, although increasingly micro-level agent-based models of migration which attempt to do exactly this are becoming more common - see Espindola et al. (2006); Makowsky et al. (2006). It has been demonstrated before, recently by Abel (2010), that at more aggregate levels migrant behaviour can be a little more predictable. In order to estimate and predict migration flows, an understanding of the external influences which act upon migrants within the system they move is required. The analysis in the last chapter and earlier on in Chapter 4 has already shown that influences such as the age of the migrant (related to life course stage), the type of area and distance between those areas, can all act to influence aggregate flows of individuals within a spatial system. It is therefore these elements that will be explored further in this chapter.

The principal objective of this chapter is to offer some explanations for the patterns of internal migration observed within the Migration Classification in Chapter 7. These explanations will be tackled from two different but complementary angles. The first will approach the explanation through examining how some of the systemic features of internal migration in Britain act to influence flows between zones within the country, returning to the ideas discussed in Chapter 2 of spatial interaction and the ‘gravitational’ effects of people and places influencing the flows of migrants. In order to achieve this, Section 8.2 will examine the theory underpinning spatial interaction explanations, building on the introduction given in Chapter 2, before proposing and developing a method for modelling spatial interactions in the Migration Classification context in Sections 8.3 and 8.4. In these sections, time will be spent discussing some of the alternative approaches to and the benefits and drawbacks of fitting differently specified mathematical spatial interaction models to observed data; and how the calibration of these models and interpretation of parameters can offer their own particular insights into observed patterns. The results of the models will be discussed and residuals analysed in order to ascertain the level of influence gravitational-type forces are having on the migration landscape of Britain and how these are either maintained or changed over time, this will be followed by a short evaluation of the modelling approach adopted in Section 8.6.

Whilst it will be shown that models are able to offer significant insight, it is likely that a complete explanation cannot be gained from models alone. Therefore the latter half of this chapter will look to offer another explanatory perspective. Much of the analysis in this thesis has highlighted the influence of age on the migration patterns presented, but as is noted by Stillwell (2008), age in itself is really a proxy for the real influences acting on migrants related their stage in the life course. The influence of life course stage on migration has been well documented so it is from this significant pool of research that Section 8.7 will draw, tying the findings from the models in the first half of the chapter to this wider social theory. This final section will examine how the intrinsic life course factors operating in Britain interact with other socio-economic and cultural features of the population to produce distinct migration behaviours and affect changes in these behaviours.

8.2 Models of expected migration

Previously in Chapter 7, an index of spatial association was developed to quantify the size and relative compactness of different area clusters. The index revealed that some clusters were more compact than others and that for some there was an increased potential for interaction to occur. This increasing potential was a function of the number of flows that could take place between districts in the cluster and the distance between those districts. Consequently, standardised migration distance ratios were developed to control for this effect on in- and out-migration flows for different clusters within the classification system. Where age, like distance, had a distorting effect on the flows taking place, similar ratios were developed to control for this. Whilst these metrics are very useful for describing elements migration in Britain and offer some explanation through identifying clusters where factors like age and distance might be having greater effect, for a more complete explanation, different techniques are required.

So given the effect of distance and age and the size of the clusters on the propensity to migrate within the Migration Classification system, how might these effects be explored further? One way this could be achieved is through building each of these elements into an expected model - a simplified version of the real Migration Classification system which will predict flows within the system as a function of just these elements. Given that these elements are known to have some effect, where the model proves to be an accurate representation of reality the level of the effect is likely to be high; on the other hand, where the model falls short of representing reality, the level of the effect is likely to be lower and other explanatory factors are likely to be at work. The question that follows, therefore, is what flows might be expected within and between the clusters - how might such a model be specified? In the SMIR and SMDR calculations, expected migration is calculated using a population at risk - in the SMDR the total system population, and in the SMIR, the age disaggregated population. Where it has already been argued in this thesis that migrants are not necessarily representative of the underlying population, this might be seen as a somewhat crude model of expected migration; crude, but not entirely without precedent: It has been shown that expected migration flows can be estimated as a function of the size of the populations at the origin and destination, and the distance between them. This simple model might be expressed as:

$$M_{12} = \frac{P_1 P_2}{D_{12}} \tag{8.1}$$

where M_{12} is the migration between origin 1 and destination 2, P_1 is the origin population, P_2 is the destination population and D_{12} is the shortest distance between the origin and destination. This model was first proposed by by Zipf (1946), who showed that many of the migration moves between cites in the U.S. - certainly those which occurred by highway - could be modelled in this way. From the early models of Zipf, however, models of expected migration, and more generally spatial interaction, developed and became more and more accurate representations of the systems they were designed to represent. As described in Section 2.3.3, early attempts to model human spatial interactions were based upon models of interaction taken directly from the physical sciences, specifically Newton's law of universal gravitation. Newton's law states that the magnitude of the force between two masses is proportional to the product of those two masses divided by (or inversely proportional to) the square of the distance between them, or:

$$F_{ij} = G\frac{m_i m_j}{d_{ij}^2} \tag{8.2}$$

In Newton's equations, F_{ij} is the interaction force acting between two bodies i and j, G is an empirically derived gravitational constant, m is the mass of the bodies and d_{ij} is the distance between them, which in this case is squared to represent the exponential decay of attraction between i and j as distance increases. At this stage the comparisons with Zipf's model are clear to see, but this model can be re-expressed so that:

$$F_{ij} = Gm_i m_j (d_{ij})^{-2} \tag{8.3}$$

In the study of human spatial interaction, F_{ij} might represent the number of people migrating from one area of residence to another, or travelling to work or to a shopping location from a residential area. The mass of the origin m_i might represent the population, but it could just as easily represent the total number of migrants or commuters leaving area i, and m_j might represent the population or total number of migrants or commuters arriving at j, with d_{ij} representing either some physical distance between i and j or some measure of the cost of travel between the two. In the Newtonian gravity model, the inverse square of the distance is the appropriate distance decay factor. Therefore, taking Equation (8.3) as an example it would mean that given origins and destinations with constant masses - for instance 10 - for every unit of distance increase between the two, the volume of interaction would decrease by a power of -2. E.g.

$$1 \times 10 \times 10 \times (1)^{-2} = 100 \tag{8.4}$$

$$1 \times 10 \times 10 \times (2)^{-2} = 25 \tag{8.5}$$

$$1 \times 10 \times 10 \times (3)^{-2} = 11.1 \tag{8.6}$$

and so on.

Whilst Newton's law of gravitation and its -2 power function associated with the distance term provides a perfectly adequate representation of reality for physical systems (at least at a scale above the sub-atomic), it has been shown that this is not always the case for human systems (Taylor, 1983). Similar models used have incorporated some important alterations, not least to the distance decay term. As described by Roy and Thill (2004), early work on spatial interaction in the context of retail modelling by Huff (1963) led to the d_{ij} term reflecting travel time rather than distance, and the negative power term being calibrated by empirical observations rather than simply adopting the Newtonian -2 (which tended to lead to a power of less than -2).

As Senior (1979) points out, one of the major drawbacks of using gravity-type models to model spatial interactions in geographical/human systems is that the multiplicative nature of the equation means that a doubling of origin and destination masses, rather than lead to a doubling of the interaction, actually leads to a quadrupling of the interaction, e.g.

$$1 \times 10 \times 10 \times (1)^{-2} = 100 \tag{8.7}$$

$$1 \times 20 \times 20 \times (1)^{-2} = 400 \tag{8.8}$$

To deal with this problem, it is possible to constrain the interaction within the system to known information about origins, destinations or both. This technique was first made explicit by Wilson (1971) who proposed a family of spatial interaction models which could take advantage of either complete or partial known information about the system being modelled. Where only information about the total number of interactions in the system, Wilson defines the unconstrained model as:

$$T_{ij} = kW_i^{(1)}W_j^{(2)}f\left(c_{ij}\right) \tag{8.9}$$

The F_{ij} interaction term in the gravity model is replaced by T_{ij} in Wilson's model. m_i and m_j are replaced with $W_i^{(1)}$ and $W_j^{(2)}$ respectively, terms which represent unknown information about the respective origin and destination masses. The negative power function acting on the distance measure in the gravity model is replaced a function f of the cost of travel c_{ij} (which could be distance or any other cost of travel such as time or financial cost). G in the gravity model is replaced by k - a constant which acts as a balancing factor to ensure T_{ij} complies with the known information about the total flows within the system. As noted by Harland (2008), this can be calculated endogenously such that:

$$k = \frac{T}{\sum_i \sum_j W_i^{(1)}W_j^{(2)}f\left(c_{ij}\right)} \tag{8.10}$$

and

$$T = \sum_i \sum_j T_{ij} \tag{8.11}$$

The second spatial interaction model in Wilson's family is the origin or production constrained model. In this model, the total number of flows which leave each origin i is known and so this information is used to constrain output of the model such that where:

$$T = \sum_j T_{ij} = O_i \tag{8.12}$$

the model takes the form

$$T_{ij} = A_i O_i W_j^{(2)}f\left(c_{ij}\right) \tag{8.13}$$

Where O_i is the known information about the origin mass (total outflows) and the balancing factor k is replaced with an origin-specific balancing factor, which can be calculated as:

$$A_i = \frac{1}{\sum_j W_j^{(2)} f(c_{ij})} \tag{8.14}$$

The third model in the family is the destination or attraction constrained model. It is analogous to the origin constrained model, but here the destination mass (total inflows) is the known information constraining the output of the model. Therefore the constraint in this model is:

$$T = \sum_i T_{ij} = D_j \tag{8.15}$$

and the model takes the form:

$$T_{ij} = B_j D_j W_i^{(1)} f(c_{ij}) \tag{8.16}$$

Where D_j is the known information about the destination mass and the origin-specific balancing factor A_i is replaced with a destination-specific balancing factor, which can be calculated as:

$$B_j = \frac{1}{\sum_i W_i^{(1)} f(c_{ij})} \tag{8.17}$$

The final model in the family is the doubly constrained or production/attraction constrained model. In this model, both constraints apply so that the interactions in the model conform to both the origin and destination masses. This model takes the form:

$$T_{ij} = A_i B_j O_i D_j f(c_{ij}) \tag{8.18}$$

and the balancing factors in this model take the form:

$$A_i = \frac{1}{\sum_j B_j D_j f(c_{ij})} \tag{8.19}$$

$$B_j = \frac{1}{\sum_i A_i O_i f(c_{ij})} \tag{8.20}$$

The difficulty with solving this model lies with the balancing factors being mutually dependent. To deal with this, Senior (1979) proposes an iterative algorithm which after setting either A_i or B_j to have an initial value of 1, solves each equation in term successively updating each balancing factor until convergence is reached and a set of balancing factors are produced which ensure both origin and destination constraints can be met.

Wilson's family of spatial interaction models was an important advance on the gravity model and has been adopted and adapted for a range of different human system applications; applications described by Wilson himself (Wilson, 2008) almost forty years after first presenting his ideas, ranging from retail planning to multi-regional demographic and economic modelling. One of these areas has been in the field of migration modelling. Spatial interaction-type models of migration have been used variously to explore the structure of origin/destination migration flows and have ranged from the simple, such as the early model proposed by Zipf, to the very complex, such as the model of migration 'MIGMOD', developed for the then Office of the Deputy Prime Minister (Champion et al., 2003; Fotheringham et al., 2004; Rees et al., 2004). A huge range of different spatial interaction models in between have been specified both as tools for estimating and predicting migration flows, but also as is demonstrated by Stillwell (1978), as tools for examining the effects that the systems themselves have on the flows within them.

The 'entropy maximising' models just presented have been developed and honed in various ways (Flowerdew and Aitkin, 1982; Fotheringham, 1983a; Stillwell, 1978; Willekens, 1983) so that more and more they can account for much of the variation shown in interaction systems. As such this has proved that (in aggregate terms at least), migration and spatial interaction can been seen as an inevitable consequence of the physical systems that individuals inhabit. Individual migrants may indeed exercise individual choice about whether and where to move, but this perceived choice is bound within systems that influence these choices - thus young teenagers who have just finished their A-levels can do anything they like with their life at that point, but the social/educational norm for such individuals in Britain is for them to then attend one of the many higher education institutions located around the country. It is social expectations such as these, or economic imperatives such as the search for employment, cultural influences such as when to start a family or indeed financial constraints which, it could be argued, make individual choice irrelevant when studying whole systems. Moreover, what spatial interaction models have shown is that the physical structure of the systems - the size of the origins and destinations and the physical distances between them - act to influence where these interactions are most likely to take place.

8.3 Expected migration within the Migration Classification

So given these external influences on migrant behaviour, it follows that it should be possible to form a hypothesis of expected migration based upon this knowledge. Specified correctly, such a model hypothesis should account for much of the variation seen within the migration system. Where the model does not perform well, then it would be reasonable to expect that other influences are acting to affect the prediction. For the Migration Classification, ten years of age-disaggregated migration data are available for the full system, so consequently the logical model to fit to the data would be akin to the doubly constrained model proposed by Wilson. This would mean that in the model, flows into and out of zones in the spatial system will always equal the observed totals; variations in the modelled flows could then be compared to observed flows in an explanation of migration flows. The model takes the following general form:

$$\hat{M}_{ij} = A_i B_j O_i D_j f(d_{ij}) \tag{8.21}$$

where $\hat{M}_{ij}$ in this case represents the expected or predicted flow, and the cost of travel c_{ij} measure in Wilson's model is replaced with a more explicit d_{ij} distance measure. The function f in this case can be represented by either a negative power function $d_{ij}^{-\beta}$ or a negative exponential function $\exp(-\beta d_{ij})$ where β is the distance decay parameter which is either estimated externally or calibrated within the model.

8.3.1 Calculating distance between clusters

In traditional empirical studies of spatial interaction, models have generally been used in systems where origins and destinations are discrete entities. These might be administrative areas (Boyle, 1998) or bespoke zones (Baxter and Ewing, 1981), point entities such as cities (Fotheringham, 1984) or combinations of zone and point entities such as residential areas and petrol station forecourts (Heppenstall et al., 2005). Where these discrete entities are locationally tethered, a measure of distance between origin and destination can be easily calculated, either as a straight-line or network distance. Whilst some measures of distance may be more or less appropriate for spatial interaction modelling (see Flowerdew, 2010 for a discussion of the relative merits of different measures), in conventional models whichever method is chosen the distance value is still a real measure of distance between origin and destination. An immediate problem is presented then when specifying a single distance between any two clusters which contain a number of discrete spatial units located in very different places. The argument is usually about whether to use a straight-line Euclidean distance between zonal centroids, or some kind of network distance between the same points. However, where clusters are comprised of a number of different, non-contiguous zones, the definition of a zonal centroid becomes more problematic. Given this situation, the solution could be to take a single centroid value for a cluster based on the x, y coordinates of all districts comprising the cluster. This solution would be empirically difficult to justify as where clusters contain districts which are distributed around the spatial system (for example the Coastal and Rural Retirement Migrants and Student Towns and Cities clusters), it is likely that the centroids would appear in similar locations, not only ignoring the distance between districts at the spatial periphery of the clusters, but ignoring information such as the number of districts in the cluster and therefore the number of potential interactions.

An alternative solution would be to take a measure of the distance between every district

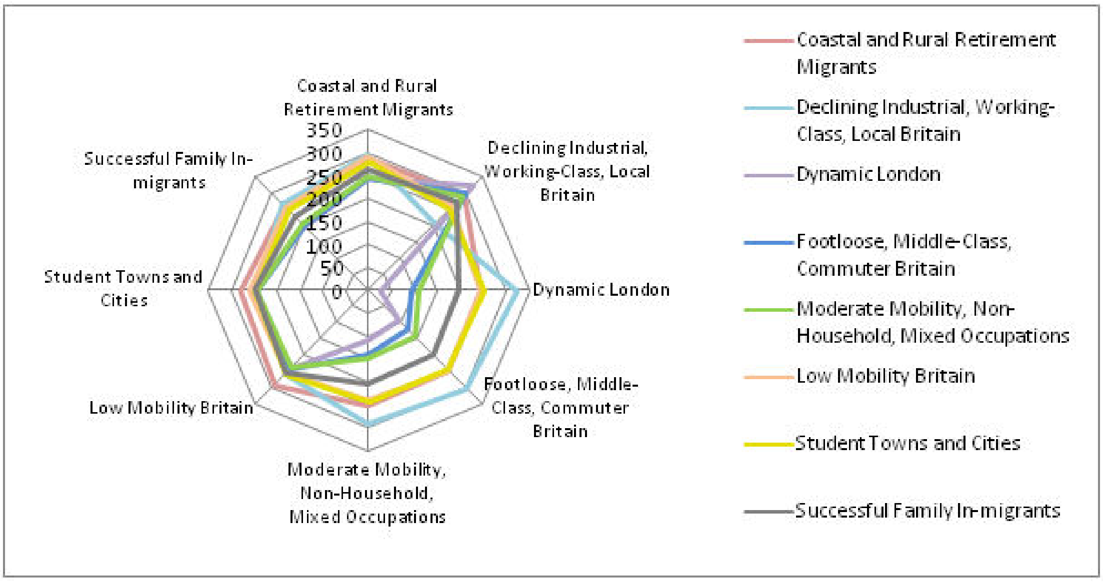


Figure 8.1: Average distance (km) between districts in each Migration Classification cluster

in one cluster with every district in another cluster. This measure could be, for example, the sum, minimum, maximum, mean, modal or median value of all distances between all districts, measured between points such as the district centroids or population weighted centroids, either as straight line Euclidean or a network distances. An interesting avenue of future research would be to examine the most appropriate measure of distance for a spatial system comprised of non-contiguous groups of zones, assessing the impact of different methods on model fits, although having said this, preliminary experiments with the sum of all distances between clusters produced poor model results.

In this research, the distance measure was defined as the mean of the Euclidean distances between the population weighted centroids of the districts in one cluster with the districts in another cluster, or indeed the same cluster. This means that where in a conventional distance matrix the diagonal intra-zonal distance would be zero, in this matrix the intra-cluster flow distance is similar to the inter-cluster distances. A representation of the distance matrix used is shown in Figure 8.1. The largest range of distances can be seen for the Dynamic London cluster, reflecting its spatial concentration - with average distances to districts in the Declining Industrial, Working-Class, Local Britain cluster the greatest at around 325km and distances to other districts within the cluster the least at around 30km. The smallest range of distances is in the Student Towns and Cities cluster (between 250 and 275km) reflecting its spatial distribution and the consequent relatively even distances to districts in all other clusters.

8.4 Implementing the model

The doubly constrained spatial interaction model was constructed in the Microsoft Excel spreadsheet package. Initially, two alternative versions of the model were trialled and the results compared. The first version of the model was a re-interpretation of the doubly constrained model constructed in the Fortran programming language for research carried out by Stillwell (1978). This model calibrates the distance decay β parameter using an iterative Newton-Raphson automatic search routine. There are general formulations of the Newton-Raphson procedure which show that it can be used on any function of a variable such as the R^2 or the SRMSE, but the routine as implemented here works on the mean predicted flow distance. First an initial distance decay parameter value β^1 is set and the model run, calculating a predicted flow distance function $\hat{E}$ for the system using the following formula:

$$\hat{E}^n = \frac{\hat{E}^n_{++}}{\hat{M}^n_{++}} \tag{8.22}$$

where $\hat{E}^n$ is a measure of the predicted flow distance for the system at iteration n, $+$ is the summation over the index it replaces and:

$$\hat{E}^n_{ij} = \hat{M}^n_{ij} D_{ij} \tag{8.23}$$

where $\hat{M}^n_{ij}$ is the predicted flow area (cluster) i and area (cluster) j.

This predicted flow distance variable is compared with an observed flow distance variable E for the system where:

$$E = \frac{E_{++}}{M_{++}} \tag{8.24}$$

and

$$E_{ij} = M_{ij} D_{ij} \tag{8.25}$$

where the difference between $\hat{E}^1$ and E are greater than a predetermined threshold, then on the first pass of the algorithm the Newton-Raphson routine increments β^1 by a predetermined increment *inc* such that:

$$\beta^2 = \beta^1 + inc \tag{8.26}$$

where *inc* is a small value e.g. 0.001

The $\hat{M}_{ij}$ values are calculated again. If the difference between $\hat{E}^2$ and E are too great, then on this second pass through the algorithm β^2 is adjusted more severely using the following formula:

$$\beta^3 = \beta^2 - eps \tag{8.27}$$

where:

$$eps = \frac{del}{r} \tag{8.28}$$

and

$$del = \frac{\hat{E}^2 - \hat{E}^1}{inc} \tag{8.29}$$

and

$$r = \hat{E}^2 - E \tag{8.30}$$

The Newton-Raphson routine iterates though this cycle adjusting β either by $+inc$ at the end of each odd numbered iteration or $-eps$ at the end of each even numbered iteration until the predetermined threshold for the difference between $\hat{E}$ and E is reached - usually less than 0.01 or 0.001.

The Newton-Raphson method is not a perfect calibration solution, however. Under some circumstances - for example where the result at iteration n is a long way from the solution - a subsequent iteration will fail to improve the solution and despite additional iterations it will be impossible to reach the predetermined convergence threshold. This was something discovered when testing the model, but investigation of the literature found other researchers has encountered similar problems. Batty and Mackie (1972), in calibrating spatial interaction models using a Newton-Raphson search routine, note that good approximations of the best parameter value *"are needed to achieve convergence"* (Batty and Mackie, 1972, p.218). Batty and Mackie offer little by way of a solution to this problem, other than testing the size of the increment. This was carried out here along with further empirical investigation with a variety of different observed distance matrices. Experimentation with different values showed that changing the increment had a varying effect on convergence, however, convergence was always hard to achieve where d_{ij} distance values were high, but the division of all distances by 100 resulted in convergence every time with these data. Exploring the reason for this, a definitive answer could not be reached, but the negative exponential or power functions acting on larger values produce very much smaller values (i.e. in many cases very close to zero) where similar functions acting on smaller values produce comparatively higher results. Values closer to zero for distance functions appear to create more problems when attempting to get the model to converge.

Harland (2008) proposes an alternative method of calibration which was also trialled to see if it produced comparable output to the Newton-Raphson routine. Harland's method does not use the balancing factors used in the standard doubly constrained model proposed by Wilson (which are calculated afresh with each round of β parameter estimation), but follows an alternative method proposed by Openshaw (1998). For the doubly constrained model, this method first computes a 'relative flow matrix' M^*_{ij} using either known origin or destination masses:

$$M^*_{ij} = D_j f(d_{ij}) \tag{8.31}$$

or

$$M^*_{ij} = O_i f(d_{ij}) \tag{8.32}$$

The ratio of observed O_i to relative O^*_i or observed D_j to relative D^*_j multiplied by M^*_{ij} is then used to constrain the relative flow matrix to known data and produce an estimated flow matrix $\hat{M}_{ij}$. These new balancing equations are used in place of the originally proposed A_i or B_j balancing factors. The equations for the estimates flow matrix $\hat{M}_{ij}$ then become:

$$\hat{M}_{ij} = \frac{O_i}{O^*_i} M^*_{ij} \tag{8.33}$$

where

$$O^*_i = \sum_j M^*_{ij} \tag{8.34}$$

or

$$\hat{M}_{ij} = \frac{D_j}{D^*_j} M^*_{ij} \tag{8.35}$$

where

$$D^*_j = \sum_i M^*_{ij} \tag{8.36}$$

Either the origin or destination constraint can be used in the first stage of the doubly constrained model, but in order for the M_{ij} flows to satisfy both the origin and the destination constraints, the $\hat{M}_{ij}$ matrix has to be iteratively fitted to the observed O_i and D_j values. Taking Equation (8.35) above as the first iteration of the process, the IPF procedure will proceed as follows until convergence:

$$\hat{M}_{ij}^{2} = \left(\frac{\hat{M}_{ij}^{1}}{\hat{M}_{i+}^{1}} \right) M_{i+} \tag{8.37}$$

$$\hat{M}_{ij}^{3} = \left(\frac{\hat{M}_{ij}^{2}}{\hat{M}_{+j}^{2}} \right) M_{+j} \tag{8.38}$$

$$\hat{M}_{ij}^{4} = \left(\frac{\hat{M}_{ij}^{3}}{\hat{M}_{i+}^{3}} \right) M_{i+} \tag{8.39}$$

Experimentation has shown that whilst starting with either the origin or destination constrained $\hat{M}_{ij}$ matrix can have an effect on the number of iterations the process takes until convergence, the final result will be identical no matter which starting matrix is used.

In Harland's method of calibration, this process is carried out for each change in a parameter value (the parameter in this case being the β distance decay parameter) and the optimum or best-fit parameter is arrived at by comparing the $\hat{M}_{ij}$ estimated matrix with the M_{ij} observed matrix using any one of a range of Goodness of Fit (GOF) statistics. Where the GOF statistic is the mean migration distance measure also used in Stillwell's model, the final β distance decay parameter is always the same albeit to fewer decimal places. The $\hat{M}_{ij}$ matrices are also broadly comparable (Table 8.1), with very minor differences the result of small differences in the β parameter decimal places, and in the balancing factor convergence criterion used in the implementation of Stillwell's model. Harland argues that his method of calibration is preferable where an unknown number of parameters feature in the model as from a very general parameter starting range (for example -2 to 2) the algorithm will search all solutions for all parameters within specified boundaries, however, where in this case only one parameter is being calibrated this advantage is irrelevant.

Harland's method involves the calibration of a parameter for the whole system that applies equally to both origins and destinations. In this case it is a generalised distance decay parameter which represents the average frictional effect of distance between all origins and destinations for a particular migration variable. This is also the case for Stillwell's general model described above. Whilst a generalised distance decay parameter is able to produce a solution for the model, it is not a theoretically or empirically satisfying one since we know that the frictional effect of distance is likely to vary spatially (Stillwell, 1978). Whilst in Chapter 6 it was made clear that distance related variables were not included in the final suite from which the Migration Classification was constructed, the spatial system is partially defined by its flow patterns implicitly though the classification of migrants and the types of moves they make, if not explicitly though the distance of those flows. For example, proxy distance variables in the form of intra-district migration versus inter-district were included. This is important as clusters such

Table 8.1: Comparison of doubly constrained models calibrated using Stillwell and Harland's methods

(a) Doubly constrained model calibrated using Stillwell's method

Generalised β parameter	**-1.31**								
Origin/Destination	**A**	**B**	**C**	**D**	**G**	**I**	**J**	**K**	O_i
A	47645	35812	24839	29726	23872	21665	62714	37674	**283946**
B	44706	117073	9106	14696	14579	34190	96771	35661	**366783**
C	35060	10296	178617	89847	60594	16359	37160	37013	**464945**
D	35643	14116	76325	61710	40516	16404	41756	35386	**321856**
G	27970	13683	50299	39589	32042	14951	35166	28173	**241874**
I	24689	31211	13208	15590	14542	17863	42915	22178	**182196**
J	79196	97893	33246	43976	37903	47556	133571	66975	**540315**
K	37954	28779	26418	29731	24225	19606	53431	34539	**254682**
D_j	**332861**	**348864**	**412058**	**324866**	**248273**	**188594**	**503483**	**297598**	**2656597**

(b) Doubly constrained model calibrated using Harland's method

Generalised β parameter	**-1.32**								
Origin/Destination	**A**	**B**	**C**	**D**	**G**	**I**	**J**	**K**	O_i
A	47779	35743	24672	29678	23853	21681	62795	37744	**283946**
B	44630	117587	8930	14526	14447	34216	96877	35570	**366782**
C	34854	10104	179392	90023	60668	16233	36817	36854	**464946**
D	35602	13955	76441	61913	40609	16353	41602	35381	**321856**
G	27956	13561	50333	39677	32140	14934	35082	28192	**241874**
I	24709	31230	13096	15536	14522	17914	42990	22200	**182196**
J	79304	97984	32912	43798	37801	47638	133837	67042	**540315**
K	38027	28700	26283	29715	24234	19625	53483	34615	**254682**
D_j	**332861**	**348864**	**412058**	**324866**	**248273**	**188594**	**503483**	**297598**	**2656597**

as Constrained, Working Class, Local Britain were defined principally by the preponderance of shorter-distance, local moves, rather than longer distance inter-district moves. It would follow, therefore, that the frictional effect of distance would have a greater effect on flows to and from this cluster than it would on some other clusters - a generalised distance decay parameter would not pick this difference up.

This idea that the effect of distance decay will vary for different origins and destinations is not a new one. Both Stillwell (1978) and Taylor (1983) make reference to the in-migration 'field' - *"the area about some destination from which migrants are drawn"* (Taylor, 1983, p.3) - the idea being that alternative fields will be present for different destinations. Work by Kim et al. (2007) demonstrates this variation for origins, as does Stillwell (1978) for both origins and destinations. Stillwell proposes a 'field' version of the standard doubly constrained model which calibrates different distance decay parameters for origins and destinations once a more conventional model has been calibrated. The model takes the form:

$$M_{Ij} = A_I B_j O_I D_j f(d_{Ij}) \tag{8.40}$$

for an out-migration/origin-specific β model, or

$$M_{iJ} = A_i B_J O_i D_J f(d_{iJ}) \tag{8.41}$$

for an in-migration/destination-specific β model.

The calibration process begins with a model for each origin or destination assuming the best-fit generalised β value from the standard doubly constrained model. Where in the standard model calibration procedure observed and predicted flow distances were calculated for the whole system, here they are calculated for each origin and destination zone separately, so that:

$$E_i = \frac{E_{i+}}{M_{i+}} \tag{8.42}$$

$$E_{i+}^n = \frac{E_{i+}^n}{M_{i+}^n} \tag{8.43}$$

are calculated for the origin-specific model and:

$$E_j = \frac{E_{+j}}{M_{+j}} \tag{8.44}$$

$$E_{+j}^n = \frac{E_{+j}^n}{M_{+j}^n} \tag{8.45}$$

are calculated for the destination-specific model.

Using the same Newton-Raphson calibration routine it is then possible to calibrate either origin or destination specific distance decay β parameters. Preliminary experiments to see whether it was feasible to achieve similar origin/destination specific parameters using Harland's iterative technique were unsuccessful - whilst it was possible to calculate origin and destination specific mean distance, R^2 and SRMSE statistics and update initial β parameters according to these values, a global best solution for origin and destination specific β values could not be reached through passing the new relative flow matrices through the algorithm, even after many iterations.

It would appear sensible to assume that where origin and destination distance decay parameters are calibrated independently and used in the model predictions in place of a generalised parameter, a better overall model fit would be achieved. But is this necessarily the case? Knudsen and Fotheringham (1986) suggest that the best method for assessing the GOF of a spatial interaction model is the SRMSE, although Harland (2008) demonstrates a range of scenarios where different GOF will imply a better or worse fit, suggesting that using a range of these measures is preferable. Table 8.2 below shows the GOF statistics for when doubly constrained Spatial Interaction Models (SIMs) are fitted to three randomly chosen migration matrices.

For each dataset, five different versions of the model are fitted. The first three use Harland's method and calibrate the generalised β parameter using either the R^2, SRMSE or mean distance measure, the final two models in each dataset use origin and destination specific distance decay parameters calibrated using Stillwell's Newton-Raphson method. Note that two different functions were used - the top set of models in Table 8.2 use a negative exponential function, whereas the bottom set of models use an inverse power function. Using the inverse power function, the Newton-Raphson routine employed in Stillwell's origin or destination specific model would not converge and produce results, however, convergence was achieved when using the negative exponential function. Note also that the β parameter values appear high for the exponential decay models - this is a consequence of the original d_{ij} distance values being divided by 100. Distance values were divided by 100 to enable convergence in the origin and destination specific β model - larger distance values lead to very much smaller $f(d_{ij})$ values when the function is an inverse power or negative exponential. These very small values cause problems when attempting to compute balancing factors.

The variation in β between the age groups is clear to see, but this will be returned to in Section 8.5. Here the focus is on the differences between the models given different calibration routines. The first point of note is that for the generalised parameters, the inverse power function appears to offer better model fits than the negative exponential function for the three sample datasets. In all GOF measures but the SRMSE measure in the 60-74 2006/07 dataset this is the case. Unfortunately as the inverse power function was unable to produce results for the origin or destination specific β models, comparisons cannot be made, so attention must be turned to the negative exponential models. Here, for the generalised parameter models, as might be expected the models calibrated using a particular measure perform best on those measures of GOF (e.g. models calibrated using R^2 have the best R^2 GOF), however, both the R^2 and SRMSE tend to perform better than the distance calibrated model across most measures. When these generalised parameter models are compared with the origin and destination specific parameter models, across most GOF measures, the origin and destination specific models tend to perform better. The GOF measures for these models also are better than the generalised models using the negative power function for the most part.

These experiments indicate that it is likely that the best performing model would very likely be an origin or destination specific model calibrated using the SRMSE error between the observed and predicted matrices, using an inverse power function. Where it has not been possible at this stage to produce a model of this exact specification, a compromise must be made. Willekens (1983) observes that a number of authors examining the performance of different distance functions have concluded that, overall, there is relatively little to be gained in using one function over another. With this in mind and with technical limitations precluding the use of inverse power functions of distance and an alternative GOF measure to calibrate origin or destination specific β values, this analysis will proceed using origin and destination specific negative exponential β models, calibrated using the difference between the observed and predicted mean flow distances.

Table 8.2: Goodness of fit statistics for three doubly constrained migration matrices, calibrated using either R^2, SRMSE, average distance or average distance with origin or destination specific β values

	Total 1998/99					16-19 2004/05					60-74 2006/07				
	R^2	SRMSE	Distance	Dist O_i	Dist D_j	R^2	SRMSE	Distance	Dist O_i	Dist D_j	R^2	SRMSE	Distance	Dist O_i	Dist D_j
Exponential β	-1.68	-1.489	-1.317			-0.903	-0.886	-1.007			-1.728	-1.676	-1.505		
Sum of squared deviations	8520684749	8154335867	8501103141	5999444692	7374165924	34959691	34935490	36076868	31870267	29077436	14656596	14598015	15275220	12952942	11377385
Mean absolute % deviation	19.827	19.136	18.496	16.596	17.302	14.593	14.677	14.589	14.086	13.158	15.213	15.079	14.969	13.877	12.993
Index of dissimilarity	9.913	9.568	9.248	8.222	8.651	7.296	7.338	7.294	7.272	6.579	7.607	7.539	7.484	7.376	6.496
Coefficient of correlation R	0.944	0.943	0.94	0.958	0.949	0.981	0.981	0.98	0.975	1.061	0.96	0.96	0.958	0.94	1.372
Coefficient of determination R^2	0.89	0.889	0.885	0.918	0.901	0.962	0.962	0.961	0.95	1.125	0.921	0.921	0.918	0.883	1.882
Square Root Mean Squared Error	56.634	55.403	56.569	52.686	47.522	13.079	13.075	13.286	11.928	12.488	9.76	9.741	9.964	8.599	9.175
Rank exponential β															
Sum of squared deviations	5	3	4	1	2	4	3	5	2	1	4	3	5	2	1
Mean absolute % deviation	5	4	3	1	2	4	5	3	2	1	5	4	3	2	1
Index of dissimilarity	5	4	3	1	2	4	5	3	2	1	5	4	3	2	1
Coefficient of correlation R	3	4	5	1	2	2	3	4	5	1	2	3	4	5	1
Coefficient of determination R^2	3	4	5	1	2	2	3	4	5	1	2	3	4	5	1
Square Root Mean Squared Error	5	3	4	2	1	4	3	5	1	2	4	3	5	1	2
Power β	-1.513	-1.457	-1.748			-1.019	-1.011	-1.358			-2.023	-2.012	-2.22		
Sum of squared deviations	6351689215	6301886167	7496660426	n/a	n/a	33741853	33736288	43061799	n/a	n/a	17683974	17682556	18174415	n/a	n/a
Mean absolute % deviation	17.681	17.812	17.873	n/a	n/a	15.555	15.554	16.736	n/a	n/a	16.941	16.948	16.891	n/a	n/a
Index of dissimilarity	8.84	8.906	8.936	n/a	n/a	7.778	7.777	8.368	n/a	n/a	8.471	8.474	8.445	n/a	n/a
Coefficient of correlation R	0.956	0.956	0.955	n/a	n/a	0.981	0.981	0.977	n/a	n/a	0.951	0.951	0.95	n/a	n/a
Coefficient of determination R^2	0.914	0.914	0.912	n/a	n/a	0.963	0.963	0.955	n/a	n/a	0.904	0.904	0.903	n/a	n/a
Square Root Mean Squared Error	48.897	48.705	53.122	n/a	n/a	12.849	12.848	14.516	n/a	n/a	10.721	10.72	10.868	n/a	n/a
Rank power β															
Sum of squared deviations	2	1	3	n/a	n/a	2	1	3	n/a	n/a	2	1	3	n/a	n/a
Mean absolute % deviation	1	2	3	n/a	n/a	2	1	3	n/a	n/a	2	3	1	n/a	n/a
Index of dissimilarity	1	2	3	n/a	n/a	2	1	3	n/a	n/a	2	3	1	n/a	n/a
Coefficient of correlation R	1	2	3	n/a	n/a	1	2	3	n/a	n/a	1	2	3	n/a	n/a
Coefficient of determination R^2	1	2	3	n/a	n/a	1	2	3	n/a	n/a	1	2	3	n/a	n/a
Square Root Mean Squared Error	2	1	3	n/a	n/a	2	1	3	n/a	n/a	2	1	3	n/a	n/a

8.5 Models of expected migration - results and analysis

8.5.1 The distance decay effect

A series of doubly constrained origin/destination specific models were fitted to observed data matrices to produce both expected data matrices and origin and destination specific distance decay parameters. These models can be specified generally with an age group parameter as follows:

$$\hat{M}^a_{ij} = A^a_i B^a_j O^a_i D^a_j f(d_{ij}) \tag{8.46}$$

where the superscript a represents one of the eight broad age groups in the patient register data. For the purposes of this analysis, age group variables were added to the origin/destination specific models shown in Equations (8.40) and (8.41). Before commenting on the expected distance matrices, it will first be useful to examine the distance decay parameters associated with each of the origin and destination clusters. Table 8.3 shows these parameters for the total migrant matrices over the ten year period of this study. The first point to note is that over the ten year period, there is very little variation in the parameter values for each cluster. The values can be interpreted in that the greater negative values represent a stronger frictional distance effect. So, for example, distance has a much more negative influence over the moves into and out of the Low Mobility Britain cluster than any other cluster in the classification. This cluster has the fewest flows associated with it anyway, but when flows do happen, they are less likely to be happening over longer distances.

The β parameter values, whether high or low for a particular cluster, do not vary very much over time. The largest range can be found in the destination specific parameter for Coastal and Rural Retirement migrants - from a peak in 1999/00 the parameter decreased year-on-year until 2005/06, suggesting that over this period, moves into this cluster were constrained less and less by the distance from the origin. At the other end of the scale but with slightly less variation, the origin specific β grew for Declining Industrial, Working-Class, Local Britain over the same period. This suggests that where moves into this cluster occurred, these became more preferentially over shorter distances over this ten year period. These cases are exceptional cases - over time the frictional effect of distance for both in and out-migration remains stable for most clusters.

As has been mentioned, distance has the greatest frictional effect on moves into and out of the Low Mobility Britain cluster. It has the least effect on the Coastal and Rural Retirement Migrants cluster, although it is even lower for out-migration than in-migration. The effect of distance is surprisingly strong on flows into and out of London given the volume of exchanges taken place, however it is more pronounced for out-migration, and this probably reflects the preponderance of moves both within the cluster and to the Footloose, Middle-Class Commuter

Table 8.3: Origin and destination specific distance decay parameters for each cluster in the migration classification, 1999-2008

β **Origin Specific** (O_i)	**1998/99**	**1999/00**	**2000/01**	**2001/02**	**2002/03**	**2003/04**	**2004/05**	**2005/06**	**2006/07**	**2007/08**
Coastal and Rural Retirement Migrants	-0.717	-0.761	-0.728	-0.708	-0.742	-0.75	-0.693	-0.705	-0.747	-0.693
Declining Industrial, Working-Class, Local Britain	-1.223	-1.307	-1.274	-1.278	-1.277	-1.279	-1.284	-1.304	-1.354	-1.326
Dynamic London	-1.466	-1.448	-1.407	-1.405	-1.397	-1.397	-1.403	-1.429	-1.485	-1.459
Footloose, Middle-Class, Commuter Britain	-1.169	-1.179	-1.145	-1.151	-1.159	-1.162	-1.174	-1.183	-1.233	-1.199
Moderate Mobility, Non-Household, Mixed Occupations	-1.005	-1.008	-0.99	-0.996	-1.007	-1.008	-1.014	-1.036	-1.064	-1.021
Low Mobility Britain	-2.086	-2.053	-1.966	-2.029	-1.988	-1.988	-1.952	-2.066	-2.15	-2.006
Student Towns and Cities	-1.311	-1.307	-1.274	-1.278	-1.277	-1.279	-1.284	-1.304	-1.354	-1.326
Successful Family In-migrants	-0.991	-0.992	-0.949	-0.965	-0.972	-0.976	-0.963	-0.957	-0.997	-0.961
β **Destination Specific** (D_j)										
Coastal and Rural Retirement Migrants	-1.029	-1.307	-1.274	-1.278	-1.277	-0.989	-0.907	-0.898	-0.996	-0.912
Declining Industrial, Working-Class, Local Britain	-1.311	-1.307	-1.274	-1.278	-1.277	-1.279	-1.284	-1.304	-1.354	-1.326
Dynamic London	-1.394	-1.378	-1.345	-1.36	-1.361	-1.363	-1.363	-1.374	-1.422	-1.387
Footloose, Middle-Class, Commuter Britain	-1.28	-1.307	-1.274	-1.252	-1.277	-1.279	-1.284	-1.304	-1.354	-1.326
Moderate Mobility, Non-Household, Mixed Occupations	-1.144	-1.151	-1.113	-1.118	-1.131	-1.12	-1.144	-1.161	-1.214	-1.202
Low Mobility Britain	-2.472	-2.463	-2.346	-2.331	-2.335	-2.356	-2.371	-2.377	-2.582	-2.532
Student Towns and Cities	-1.311	-1.307	-1.274	-1.278	-1.277	-1.279	-1.284	-1.304	-1.354	-1.326
Successful Family In-migrants	-1.168	-1.158	-1.121	-1.092	-1.035	-1.013	-0.994	-1.029	-1.081	-1.04

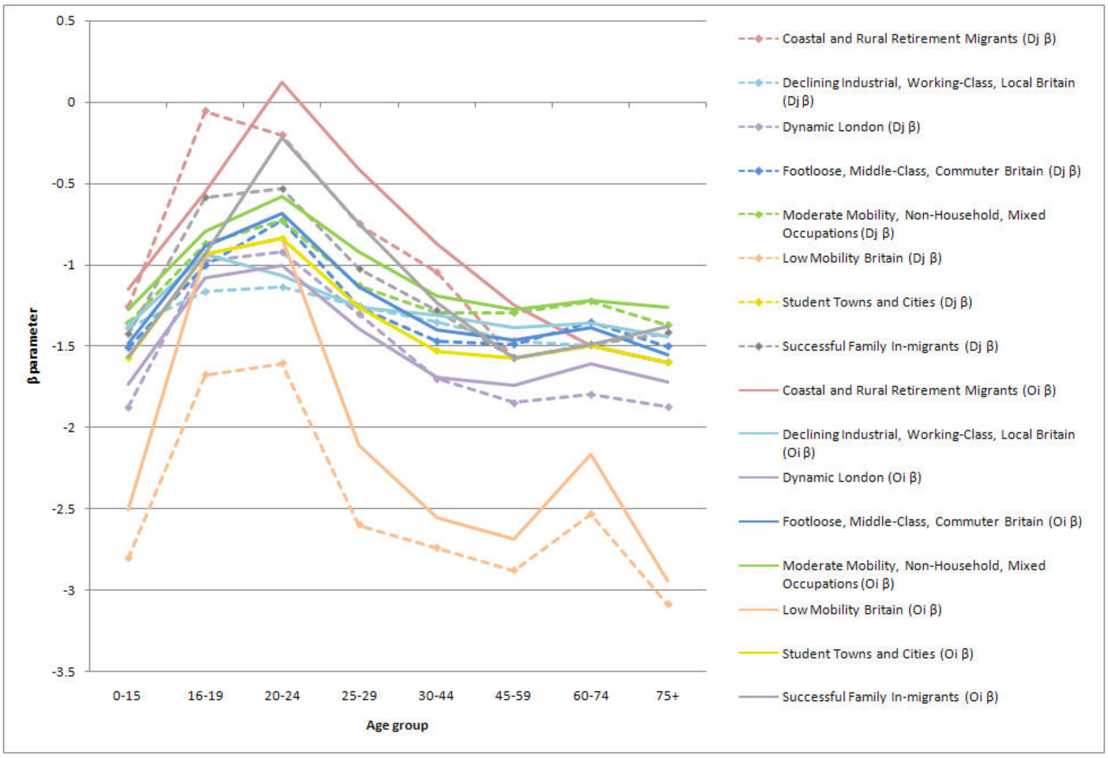


Figure 8.2: Origin and destination specific β values by age group, 1998/99

Britain cluster which, for the most part, closely borders this compact cluster. Interestingly, the Constrained, Working Class, Local Britain and Student Towns and Cities clusters appear to exhibit very similar distance decay parameters also in spite of their very different flow volumes. This will be partially due to the relatively coarse convergence criteria used which sets the origin/destination specific β to the generalised β if origin/destination specific expected average distance measures do not vary significantly from the system average, but by definition then, also partially to their similarity.

Of course as with all analyses of these migration data, there is considerable variation by age and life cycle stage. Figure 8.2 shows the age specific variation in the β parameter values for both origin and destination specific models for the 1998/99 data set. As might be expected, distance has the lowest frictional effect on the age groups where peak migration activity occurs, although the extent of this effect varies by cluster and by whether the β parameter is origin or destination specific. For example, Low Mobility Britain consistently exhibits the highest negative distance decay parameter values across all age groups for in-migration, destination specific flows. This is not the case for the out-migration specific parameter, here, at age 20-24, the frictional effect of distance is higher for the Constrained, Working-Class, Local and Dynamic London clusters.

The patterns shown in Figure 8.2 are somewhat reminiscent of the age-specific migration propensity schedules shown earlier in Figures 7.9 and 7.10. Broadly speaking there is a correlation between the general propensity to migrate and the propensity to migrate over distance, with young migrants in their late teens and early twenties being less affected by the frictional effect of distance than older migrants. This broad pattern is affected by whether migrants are moving out of or into particular clusters, with the Coastal and Rural Retirement cluster being the least constrained by distance - indeed out-migration from this cluster at 20-24 even displays a positive exponential distance decay parameter rather than a negative one, suggesting that if migrants in this age group move away from this cluster, distance will have a positive effect on the move, i.e. the further away the better! For most clusters at age group 20-24, the origin-specific distance decay parameters are noticeably lower than the destination-specific parameters, suggesting that out-migration from these clusters is less constrained by distance than in-migration to them. The only clusters where this is not the case are Dynamic London and Footloose, Middle Class, Commuter Britain where the β parameters suggest that migrants are prepared to move from further away into these clusters than they are from them, which in the context of known life-cycle and employment related migration flow patterns makes empirical sense.

Figure 8.3 reveals the predicted mean migration flow distances used in the calibration of the distance decay parameters for the 1998/99 dataset (the nature of the calibration means that these are very close to the observed mean distances). The lowest mean migration distance is associated with in-migration into the Student Towns and Cities cluster; this is for in-migrants aged 0-15 and 30 and above, at an average distance of around 75km, although the average distance does increase to around 130km at the age group when most moves into this cluster occur (16-19). Mean migration distances are also low for out-migration moves from Dynamic London, although the shape of the age-curve is somewhat different. Here there are two distance peaks at age group 16-19 and 60-74, reflecting the moves out of the cluster to other clusters; the lowest mean distance occurs at the age group when moves within the cluster are most prominent - 25-29.

Both Student Towns and Cities and Dynamic London have very contrasting in- and out-mean migration flow distances. Where Student Towns and Cities has a low in-migration average flow distance, its out-migration flow distance is considerably higher at around 250km across all age groups; the situation is very similar for Dynamic London, but for in-migration. Footloose, Middle-Class Commuter Britain has a very similar age structure and in/out migration average distance pattern to that of London, the difference being the average out-migration distance for this cluster is around 75km further across all age groups. Most other clusters show little difference in their age-related mean migration distances, the main exception being the Moderate Mobility, Non-Household, Mixed Occupations cluster. Whilst generally speaking, the mean migration distance into the cluster is over a shorter distance than that out of it, at the age of peak migration - 20-24 - the mean in and out migration distances are the same. The cluster with the highest mean migration distance associated with both in and out-migration is the Coastal and Rural-Retirement Migrants cluster, with average in and out migration flows of around 275km

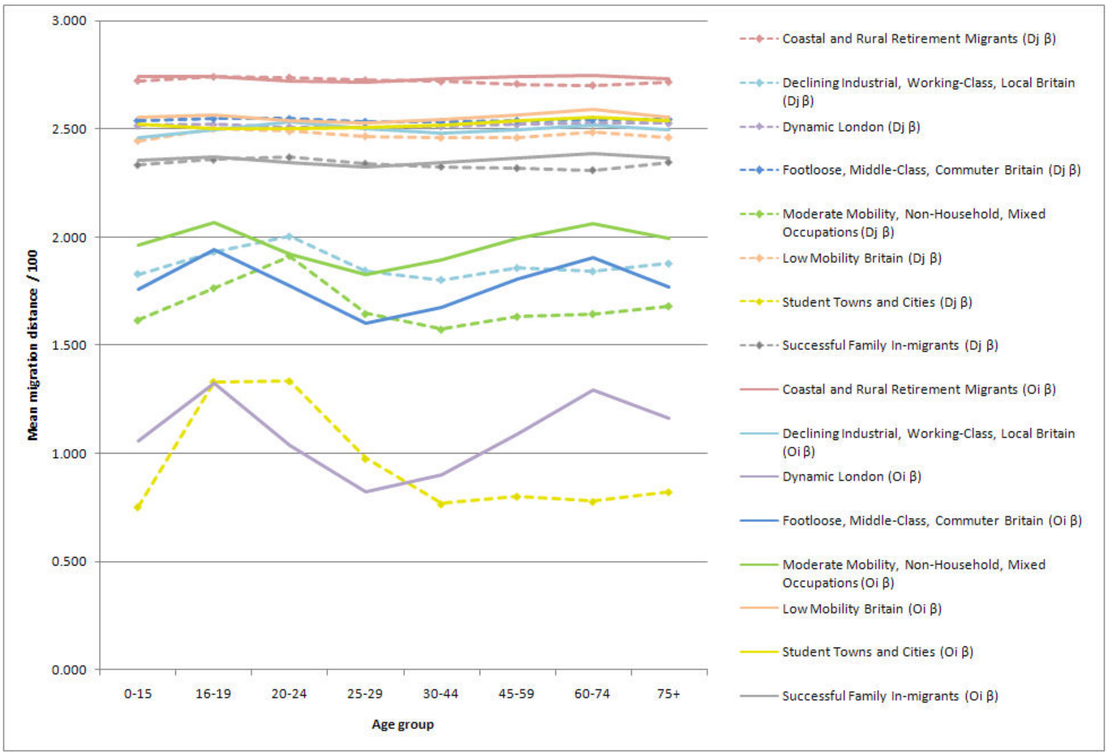


Figure 8.3: Predicted mean migration distances into and out of clusters by age, 1998/99

- a function in part of the peripheral nature of the cluster (the 'map pattern effect' which will be discussed in Section 8.6), but also, possibly, of its functional 'final destination' purpose for many retirement migrants who will be prepared to migrate further if it is just for one last time.

8.5.2 Expected migration flows

One of the main motivating influences for producing models of expected migration was in an effort to better understand the inter-cluster relationships within the Migration Classification system; to understand the influence that the spatial system, in the interplay between origin/destination masses and the spatial association between them, might be having on these flows. Origin and destination specific distance decay parameters and mean migration lengths have opened a window on the role of distance, but where distance and mass relationships alone cannot accurately predict flows, other influencing factors will certainly be at work. The suite of doubly constrained models produced a series of expected flow matrices which model migration as a simple function of the total migrants moving into and out of each cluster and the average Euclidean distance between each district in one cluster with each district in another, other explanatory variables which could have improved the predictive capabilities of the models were not included. Consequently, where observed and expected data do not match it will be important to interrogate both the observed and predicted flows and the variables defining each cluster within the classification to identify reasons why the differences occur.

As advocated by Flowerdew (2010), one way of assessing the relative performance of each model is through examining the residual values which flag the differences between the observed and expected migration flows. Table 8.4 shows a time series of standardised residuals (observed/expected*100) for each of the total migrant origin/destination β models. Some caution should be exercised when interpreting these residual scores as small differences between small flows can register a similar error to larger differences in larger flows. Thus gross residual flows should also be examined (Table 8.5) to assess the numbers involved.

The colour coding of the tables represents the degree to which the results of the model are an under or over-prediction. A value of 100 (white) represents the observed and expected migration flows being equal, a value over 100 (green) shows the observed flows are higher than the model predictions, a value under 100 (red) the observed flows are lower; the darker the shading, the greater the under or over prediction. The first point of note is that the residuals for each of the models are consistent over the ten year period for all origin/destination pairs - the model over and under-predicts in a consistent fashion. But how should this be interpreted? If the model is under-predicting migration flows, then the β parameter is too high and should be reduced for a better reflection of reality. This means that the frictional effect of distance might well be lower than the model dictates and therefore the attractiveness of the destination in question or the repulsiveness of the origin is higher in the observed system. Conversely, if the model is over-predicting migration, then the β parameter is too low and should be increased for a better representation of reality. This could mean that the frictional effect of distance is higher than the model says it is and therefore the attractiveness of the destination or the repulsiveness of the origin is lower in the observed system.

Whilst both of these points could be valid for any given prediction, one must remember that the doubly constrained model being used here is a closed system subject to the marginal constraints imposed upon it. In this context, where the model over or under-predicts for one origin and destination pair, it must have an effect on another pair as the constraints dictate that the total in and out-flows must satisfy the observed data. As such interpretation of the residual error must be carried out with a certain amount of care. Where a plausible explanation for over or under-prediction does not present itself for a given origin/destination pair, one must also look at the other origin or destination interactions between clusters in the same row or column of the inter-cluster matrix in order to theorise a likely reason. Where the patterns of under and over-prediction remain constant over time, reference will also be made during the course of this discussion to Table 8.6a, b, c, d and e which show the gross residuals for the two models and the observed and predicted total migration flows for the first year in the period - 1998/99.

Beginning with the model under-predictions (standardised residuals over 100), there is a fairly consistent under prediction of the flows in both directions between Student Towns and Cities and Dynamic London (Table 8.4a). There are two ways that this can be interpreted. Examining the age specific models which comprise the total flows (not shown), this under-prediction is maintained with each age group. At the late teen and early twenties age groups

Table 8.4: Standardised origin specific model residuals for observed vs. expected inter-cluster total migrant migration rates 1998/99 - 2007/08

		Origin specific distance decay									
Origin O_i	Destination D_j	1998/99	1999/00	2000/01	2001/02	2002/03	2003/04	2004/05	2005/06	2006/07	2007/08
Coastal and Rural	Coastal and Rural	122.20	121.41	121.20	122.21	124.25	125.16	127.77	127.29	128.27	130.46
Coastal and Rural	Working Class Local	91.44	91.10	89.34	89.33	88.78	87.36	87.08	87.84	86.56	87.32
Coastal and Rural	Dynamic London	70.79	71.17	70.07	67.60	67.70	64.87	65.88	67.20	67.02	65.34
Coastal and Rural	Commuter Britain	84.84	84.03	83.08	84.91	84.36	83.15	84.14	83.54	82.03	82.59
Coastal and Rural	Moderate Mobility	97.34	97.11	99.54	97.41	96.43	98.87	99.25	98.18	96.53	98.51
Coastal and Rural	Low Mobility	58.54	58.03	60.00	59.12	56.91	58.83	59.50	59.35	57.76	58.09
Coastal and Rural	Student Towns	102.68	103.54	104.42	103.84	102.66	100.92	99.93	99.38	100.11	98.51
Coastal and Rural	Successful Family	128.97	127.13	127.46	127.36	130.99	133.39	133.87	134.76	134.31	137.00
Working Class Local	Coastal and Rural	98.53	97.58	98.69	97.28	96.58	97.73	97.59	100.35	97.63	98.79
Working Class Local	Working Class Local	104.50	103.37	103.66	102.99	101.40	103.09	102.82	103.49	104.54	104.26
Working Class Local	Dynamic London	137.83	149.35	146.40	147.49	151.62	148.52	146.15	146.52	148.21	144.06
Working Class Local	Commuter Britain	94.79	98.13	95.36	94.57	94.03	93.55	93.58	90.26	91.44	90.95
Working Class Local	Moderate Mobility	99.94	100.16	100.03	98.84	101.36	99.40	98.44	97.92	96.13	95.57
Working Class Local	Low Mobility	82.10	79.43	80.66	80.19	81.27	80.11	80.33	80.85	78.54	79.52
Working Class Local	Student Towns	103.91	104.49	103.34	104.82	105.60	104.32	103.95	102.89	104.47	104.01
Working Class Local	Successful Family	83.95	85.66	85.86	88.09	88.63	87.77	88.51	87.02	85.57	85.24
Dynamic London	Coastal and Rural	87.74	86.58	85.93	82.82	79.72	77.11	75.29	74.59	75.09	74.49
Dynamic London	Working Class Local	166.94	163.79	164.79	166.83	174.26	168.66	170.76	171.27	171.89	165.94
Dynamic London	Dynamic London	113.54	112.88	112.79	113.20	112.99	112.70	112.03	111.20	110.96	110.09
Dynamic London	Commuter Britain	77.92	79.44	78.89	79.66	80.60	81.45	80.48	82.23	82.85	83.30
Dynamic London	Moderate Mobility	81.08	83.82	83.95	84.55	86.43	85.58	86.73	86.93	87.70	87.83
Dynamic London	Low Mobility	130.06	126.67	126.07	126.39	122.53	128.56	127.68	125.34	135.97	131.29
Dynamic London	Student Towns	151.86	153.62	151.66	152.80	150.42	153.15	152.88	155.83	159.44	157.17
Dynamic London	Successful Family	60.46	60.07	58.03	58.55	56.74	56.60	54.67	55.10	52.94	52.48
Commuter Britain	Coastal and Rural	86.02	90.37	90.38	92.31	91.35	90.77	87.81	86.36	85.88	85.59
Commuter Britain	Working Class Local	85.58	84.64	87.98	89.61	88.71	90.30	87.71	86.17	89.07	86.66
Commuter Britain	Dynamic London	72.38	72.42	73.27	73.28	73.58	75.08	76.06	75.54	75.82	75.80
Commuter Britain	Commuter Britain	131.10	130.30	131.53	129.62	130.13	129.83	129.79	130.44	129.31	129.88
Commuter Britain	Moderate Mobility	101.86	99.00	97.03	97.40	97.02	94.89	96.71	96.17	95.74	97.73
Commuter Britain	Low Mobility	116.76	114.72	110.07	109.51	108.56	108.01	106.83	110.05	114.09	111.61
Commuter Britain	Student Towns	110.34	108.38	108.78	107.06	109.11	111.17	115.55	116.12	114.59	115.50
Commuter Britain	Successful Family	100.39	102.88	104.06	103.41	101.43	99.35	96.54	96.79	97.66	98.31
Moderate Mobility	Coastal and Rural	94.53	97.86	97.08	97.75	98.62	97.28	94.45	93.42	94.79	92.93
Moderate Mobility	Working Class Local	79.78	81.19	80.32	84.81	83.47	83.86	82.00	83.51	79.63	81.15
Moderate Mobility	Dynamic London	72.40	74.18	73.85	74.01	74.63	74.46	74.92	76.70	75.15	75.63
Moderate Mobility	Commuter Britain	117.16	114.69	115.21	115.82	114.83	114.38	115.82	114.51	115.36	114.46
Moderate Mobility	Moderate Mobility	107.32	105.85	105.74	104.18	105.43	105.66	105.20	104.51	103.88	106.19
Moderate Mobility	Low Mobility	128.22	128.18	127.78	123.48	121.95	123.70	129.29	127.35	131.89	134.36
Moderate Mobility	Student Towns	100.00	98.89	99.65	98.53	99.34	99.57	100.89	101.67	101.79	102.42
Moderate Mobility	Successful Family	111.82	109.58	112.45	111.17	110.12	110.72	110.49	110.49	108.68	107.65
Low Mobility	Coastal and Rural	104.99	104.85	104.41	106.21	105.88	107.90	105.81	106.26	106.55	105.45
Low Mobility	Working Class Local	85.47	84.54	85.45	82.29	82.79	81.78	83.47	82.13	82.95	84.57
Low Mobility	Dynamic London	71.84	70.82	69.67	69.70	69.34	70.87	70.72	71.30	69.76	69.66
Low Mobility	Commuter Britain	102.08	103.63	101.78	103.65	105.92	105.57	102.07	102.22	105.31	102.85
Low Mobility	Moderate Mobility	127.46	126.17	127.17	130.49	128.51	131.16	127.39	128.78	127.48	125.14
Low Mobility	Low Mobility	97.76	98.89	97.67	99.86	97.75	98.86	99.88	99.93	96.64	96.75
Low Mobility	Student Towns	118.44	117.31	117.61	116.85	116.39	113.64	114.89	116.22	116.77	117.04
Low Mobility	Successful Family	78.03	80.10	81.49	80.35	81.94	83.17	83.55	80.77	80.72	82.06
Student Towns	Coastal and Rural	88.89	86.92	86.22	86.45	87.67	87.58	89.12	88.45	87.30	87.19
Student Towns	Working Class Local	112.32	113.42	112.99	113.11	113.96	113.57	113.92	113.73	113.46	113.37
Student Towns	Dynamic London	158.11	159.46	158.08	156.19	152.93	154.43	154.24	157.34	163.53	162.94
Student Towns	Commuter Britain	107.09	104.38	106.37	107.08	104.93	105.36	106.17	104.51	103.82	105.43
Student Towns	Moderate Mobility	107.70	109.49	109.37	109.53	109.15	109.67	109.18	108.57	109.19	106.71
Student Towns	Low Mobility	124.74	127.08	127.37	127.52	129.63	126.54	124.86	125.16	124.36	123.59
Student Towns	Student Towns	72.45	73.09	73.43	72.95	72.56	73.37	72.84	73.05	72.22	72.90
Student Towns	Successful Family	92.77	93.08	92.11	93.89	93.61	94.06	93.56	93.74	94.34	93.03
Successful Family	Coastal and Rural	118.27	119.15	121.62	121.69	122.20	122.83	123.37	123.18	123.81	123.72
Successful Family	Working Class Local	75.27	75.93	75.04	75.57	75.84	74.28	74.56	73.93	73.84	72.61
Successful Family	Dynamic London	56.04	55.72	54.78	53.74	53.68	53.38	54.43	54.79	53.34	53.97
Successful Family	Commuter Britain	90.29	91.22	92.85	91.24	92.86	91.84	93.26	91.88	89.26	89.98
Successful Family	Moderate Mobility	106.25	104.49	105.72	107.37	104.02	106.28	103.82	105.56	106.81	104.62
Successful Family	Low Mobility	66.56	66.64	65.58	66.53	64.94	65.15	65.18	66.54	64.65	66.40
Successful Family	Student Towns	107.09	105.36	104.38	103.73	103.43	104.27	103.53	103.68	103.35	104.84
Successful Family	Successful Family	143.01	141.43	142.16	140.71	143.89	141.91	146.52	145.08	145.68	148.93

Table 8.4: Standardised destination specific model residuals for observed vs. expected inter-cluster total migrant migration rates 1998/99 - 2007/08

		Origin specific distance decay									
Origin Oi	Destination Dj	1998/99	1999/00	2000/01	2001/02	2002/03	2003/04	2004/05	2005/06	2006/07	2007/08
Coastal and Rural	Coastal and Rural	124.63	128.44	128.31	130.20	131.71	127.21	129.29	128.76	130.34	131.83
Working Class Local	Coastal and Rural	97.06	98.47	99.54	98.52	97.83	93.31	91.32	93.47	91.88	91.72
Dynamic London	Coastal and Rural	73.35	69.58	69.84	67.93	66.57	68.77	68.08	66.91	65.73	65.74
Commuter Britain	Coastal and Rural	97.79	95.99	95.84	97.37	96.17	101.42	99.74	99.52	97.83	99.09
Moderate Mobility	Coastal and Rural	116.48	114.19	111.95	112.15	112.62	116.91	115.67	115.20	117.52	117.47
Low Mobility	Coastal and Rural	76.89	81.55	82.68	82.80	83.65	81.59	79.63	77.13	77.08	78.37
Student Towns	Coastal and Rural	86.38	86.26	85.53	86.02	87.24	85.69	86.54	85.84	84.92	84.37
Successful Family	Coastal and Rural	129.01	127.41	130.21	129.90	130.12	133.55	135.62	136.89	137.38	138.05
Coastal and Rural	Working Class Local	100.34	97.31	95.56	96.06	94.90	93.76	94.80	95.85	94.11	95.50
Working Class Local	Working Class Local	100.27	101.46	101.75	101.22	99.84	102.25	102.01	102.61	103.49	103.20
Dynamic London	Working Class Local	112.67	114.57	117.37	120.13	128.67	122.81	122.89	120.43	118.93	113.86
Commuter Britain	Working Class Local	96.24	93.96	97.60	98.49	96.98	96.70	92.93	92.24	95.38	93.24
Moderate Mobility	Working Class Local	102.66	103.76	100.89	105.65	103.13	102.03	99.58	101.15	98.80	102.03
Low Mobility	Working Class Local	83.65	81.77	82.70	79.75	80.37	79.71	81.10	79.56	80.26	81.73
Student Towns	Working Class Local	109.46	109.46	109.08	109.24	110.22	109.77	109.86	109.50	109.09	108.86
Successful Family	Working Class Local	80.18	80.52	79.78	79.83	79.99	77.52	78.04	77.92	78.07	76.83
Coastal and Rural	Dynamic London	68.14	67.31	66.45	65.05	65.67	64.33	64.04	64.67	63.99	61.49
Working Class Local	Dynamic London	185.90	181.36	178.04	185.24	190.28	190.52	186.33	183.86	185.22	177.88
Dynamic London	Dynamic London	120.25	120.39	119.80	118.96	118.24	117.39	116.78	116.70	116.97	116.57
Commuter Britain	Dynamic London	63.44	64.66	65.38	65.31	65.98	67.20	68.53	67.65	68.16	67.88
Moderate Mobility	Dynamic London	57.28	59.15	59.54	59.67	60.66	60.34	61.00	62.70	60.69	60.32
Low Mobility	Dynamic London	92.49	88.43	86.52	89.23	88.09	91.76	90.55	91.43	89.71	87.51
Student Towns	Dynamic London	184.03	180.91	179.61	181.52	177.30	181.63	180.48	182.22	189.01	186.71
Successful Family	Dynamic London	54.00	53.14	52.05	52.01	51.93	51.93	52.41	51.96	50.40	50.56
Coastal and Rural	Commuter Britain	73.55	74.62	73.87	73.54	74.92	75.22	75.25	74.98	73.37	73.58
Working Class Local	Commuter Britain	103.45	105.42	102.41	98.30	100.69	102.09	102.67	99.52	101.07	100.84
Dynamic London	Commuter Britain	83.91	84.09	83.49	85.87	85.35	85.57	84.29	85.83	86.44	86.89
Commuter Britain	Commuter Britain	126.82	125.94	126.90	125.96	125.42	124.28	124.90	125.18	124.61	125.00
Moderate Mobility	Commuter Britain	106.05	103.68	104.63	105.32	104.12	103.04	104.90	104.26	104.59	103.19
Low Mobility	Commuter Britain	116.30	119.48	116.65	117.46	121.27	122.77	118.37	120.15	124.58	120.33
Student Towns	Commuter Britain	111.24	110.26	112.34	111.03	110.37	112.04	113.22	111.90	111.44	113.40
Successful Family	Commuter Britain	83.83	85.68	86.80	84.80	86.82	86.01	87.00	85.33	82.89	83.48
Coastal and Rural	Moderate Mobility	78.89	79.11	80.96	78.76	78.85	81.42	82.07	81.34	79.80	82.22
Working Class Local	Moderate Mobility	93.73	90.64	90.08	89.16	92.20	90.63	92.11	91.82	90.52	92.02
Dynamic London	Moderate Mobility	93.11	96.42	97.09	97.45	98.80	97.91	97.59	97.42	97.94	97.01
Commuter Britain	Moderate Mobility	106.13	103.85	101.87	101.51	100.58	98.04	99.64	98.98	98.83	100.34
Moderate Mobility	Moderate Mobility	102.91	101.78	102.33	100.04	100.92	100.80	100.27	100.11	99.03	100.28
Low Mobility	Moderate Mobility	142.80	140.17	139.60	144.97	141.87	145.63	142.18	146.39	146.31	142.37
Student Towns	Moderate Mobility	103.15	104.81	104.45	104.61	104.41	104.99	106.06	105.67	106.64	105.62
Successful Family	Moderate Mobility	96.63	95.37	96.04	97.61	94.44	96.18	94.07	95.16	96.31	94.75
Coastal and Rural	Low Mobility	92.66	90.37	90.96	89.67	86.18	90.49	92.85	92.30	94.29	94.55
Working Class Local	Low Mobility	83.95	80.30	81.51	81.36	82.64	82.48	82.77	83.20	80.84	81.84
Dynamic London	Low Mobility	86.90	87.66	89.22	90.50	89.43	92.59	90.93	87.74	92.31	88.80
Commuter Britain	Low Mobility	108.32	106.93	103.43	102.59	101.64	99.42	97.56	101.04	103.14	101.33
Moderate Mobility	Low Mobility	115.76	116.60	116.90	112.82	111.18	111.12	115.86	114.13	116.53	119.84
Low Mobility	Low Mobility	96.91	97.78	96.63	99.33	97.36	99.34	100.00	99.90	96.44	96.27
Student Towns	Low Mobility	125.71	128.04	128.17	128.71	131.09	128.56	126.65	126.65	126.12	125.15
Successful Family	Low Mobility	65.99	66.47	65.64	66.52	64.93	64.74	64.80	66.35	64.31	66.09
Coastal and Rural	Student Towns	103.63	103.37	104.43	104.10	102.77	101.86	101.41	101.05	101.31	99.90
Working Class Local	Student Towns	105.32	105.39	104.28	106.03	106.84	106.67	106.51	105.48	106.94	106.55
Dynamic London	Student Towns	116.91	122.32	122.32	124.22	124.54	125.24	123.92	124.12	125.41	122.90
Commuter Britain	Student Towns	116.59	114.81	115.17	112.58	114.51	114.75	118.63	119.86	118.33	119.58
Moderate Mobility	Student Towns	113.70	112.93	112.70	110.72	111.19	110.02	111.54	112.32	114.01	115.54
Low Mobility	Student Towns	117.13	115.60	116.09	115.53	115.14	113.25	114.39	115.39	115.75	116.15
Student Towns	Student Towns	71.99	72.49	72.88	72.51	72.11	73.11	72.54	72.73	71.81	72.45
Successful Family	Student Towns	108.46	107.17	106.26	105.32	104.88	104.98	104.32	104.75	104.50	106.01
Coastal and Rural	Successful Family	117.19	114.85	115.19	113.80	115.34	117.98	117.35	118.97	118.16	120.07
Working Class Local	Successful Family	85.26	84.72	84.81	86.22	84.83	84.37	84.78	84.01	82.71	82.26
Dynamic London	Successful Family	55.32	56.47	55.10	56.52	56.60	56.31	54.61	54.06	51.51	51.18
Commuter Britain	Successful Family	108.11	111.79	113.10	112.86	112.64	109.62	107.37	107.42	108.58	110.05
Moderate Mobility	Successful Family	121.84	120.34	123.12	122.10	122.76	122.84	123.96	123.62	122.42	122.15
Low Mobility	Successful Family	80.58	82.12	83.36	82.02	82.37	84.00	83.96	81.73	81.87	82.82
Student Towns	Successful Family	93.56	93.52	92.52	94.02	92.79	93.38	92.80	93.32	94.01	92.70
Successful Family	Successful Family	141.40	140.50	141.02	139.95	143.54	141.18	146.04	144.26	144.92	148.41

Table 8.5: Gross origin specific model residuals for observed vs. expected inter-cluster total migrant migration rates 1998/99 - 2007/08

		Origin specific distance decay									
Origin O_i	Destination D_j	1998/99	1999/00	2000/01	2001/02	2002/03	2003/04	2004/05	2005/06	2006/07	2007/08
Coastal and Rural	Coastal and Rural	11429	11713	11410	12153	12662	13363	13569	13512	15023	14952
Coastal and Rural	Working Class Local	-3412	-3557	-4305	-4416	-4619	-5231	-5087	-4761	-5437	-4998
Coastal and Rural	Dynamic London	-6221	-6071	-6403	-6486	-6335	-6947	-6878	-6838	-6905	-7381
Coastal and Rural	Commuter Britain	-4027	-4333	-4418	-3924	-4014	-4362	-3929	-4251	-4889	-4405
Coastal and Rural	Moderate Mobility	-578	-651	-102	-557	-757	-242	-156	-389	-778	-324
Coastal and Rural	Low Mobility	-9683	-9886	-9081	-9491	-9818	-9355	-8862	-8985	-9976	-9566
Coastal and Rural	Student Towns	1717	2303	2841	2459	1689	588	-44	-408	76	-979
Coastal and Rural	Successful Family	10427	10160	9768	9957	10879	11871	11045	11747	12436	12282
Working Class Local	Coastal and Rural	-681	-1156	-613	-1287	-1535	-1025	-993	145	-1041	-500
Working Class Local	Working Class Local	5012	3954	4174	3488	1620	3569	3078	3789	5304	4824
Working Class Local	Dynamic London	3754	4501	4375	4220	4401	4113	4049	4106	4107	4035
Working Class Local	Commuter Britain	-850	-295	-712	-835	-888	-956	-926	-1429	-1287	-1328
Working Class Local	Moderate Mobility	-9	26	5	-178	201	-88	-227	-303	-582	-676
Working Class Local	Low Mobility	-6092	-7082	-6292	-6592	-6080	-6381	-6038	-5879	-7139	-6662
Working Class Local	Student Towns	3785	4443	3204	4611	5269	4067	3744	2756	4366	3952
Working Class Local	Successful Family	-6000	-5472	-5090	-4409	-3999	-4308	-3742	-4316	-5128	-4948
Dynamic London	Coastal and Rural	-3308	-4023	-4377	-5831	-7087	-8087	-7745	-7604	-7489	-7250
Dynamic London	Working Class Local	4623	4704	5243	5948	7152	6614	6199	5737	5369	4998
Dynamic London	Dynamic London	27183	25237	24907	25934	26093	25491	24668	23570	24028	22281
Dynamic London	Commuter Britain	-20780	-19746	-19617	-20308	-20080	-19188	-19290	-17853	-18134	-16535
Dynamic London	Moderate Mobility	-11599	-10350	-10112	-10265	-9366	-9940	-8858	-8698	-8523	-8267
Dynamic London	Low Mobility	3766	3525	3499	3888	3514	4417	3974	3463	4922	4231
Dynamic London	Student Towns	14758	16171	16330	17861	18210	19254	18731	18856	18826	18563
Dynamic London	Successful Family	-12620	-13586	-14140	-15405	-16513	-16644	-15656	-15291	-16342	-15499
Commuter Britain	Coastal and Rural	-5329	-3978	-3863	-3235	-3533	-3791	-4421	-4881	-5319	-4861
Commuter Britain	Working Class Local	-2324	-2528	-2008	-1821	-2001	-1705	-1954	-2120	-1654	-1938
Commuter Britain	Dynamic London	-18849	-18887	-17710	-17367	-17075	-16084	-15771	-16362	-16879	-16222
Commuter Britain	Commuter Britain	18573	18699	18028	17627	17940	17720	16896	17628	18291	16536
Commuter Britain	Moderate Mobility	744	-421	-1176	-1050	-1205	-2060	-1280	-1499	-1774	-880
Commuter Britain	Low Mobility	2920	2640	1720	1709	1537	1418	1118	1618	2377	1834
Commuter Britain	Student Towns	4630	3873	3977	3278	4266	5214	7074	7235	6461	6653
Commuter Britain	Successful Family	144	1119	1473	1311	539	-246	-1170	-1097	-848	-543
Moderate Mobility	Coastal and Rural	-1783	-760	-1020	-834	-506	-1017	-1853	-2186	-1851	-2294
Moderate Mobility	Working Class Local	-3594	-3449	-3665	-3002	-3357	-3310	-3398	-3013	-3801	-3437
Moderate Mobility	Dynamic London	-10957	-10254	-10443	-10270	-10185	-10354	-10426	-10078	-10982	-10445
Moderate Mobility	Commuter Britain	6143	5426	5386	5883	5627	5510	5825	5564	6256	5294
Moderate Mobility	Moderate Mobility	2131	1789	1702	1277	1694	1781	1594	1419	1292	1937
Moderate Mobility	Low Mobility	4716	4878	4626	4160	3959	4281	4959	4600	5737	5872
Moderate Mobility	Student Towns	-1	-467	-147	-631	-292	-190	391	736	791	1054
Moderate Mobility	Successful Family	3507	3003	3711	3570	3221	3460	3079	3151	2791	2218
Low Mobility	Coastal and Rural	936	975	881	1265	1178	1589	1060	1094	1219	984
Low Mobility	Working Class Local	-4410	-4713	-4387	-5594	-5509	-5791	-4848	-5121	-5158	-4545
Low Mobility	Dynamic London	-4174	-4166	-4330	-4219	-4165	-3924	-3992	-4015	-4297	-4417
Low Mobility	Commuter Britain	372	647	302	640	1020	953	337	373	948	478
Low Mobility	Moderate Mobility	4821	4651	4601	5272	4844	5247	4438	4780	4882	4323
Low Mobility	Low Mobility	-396	-197	-391	-25	-396	-198	-19	-12	-590	-552
Low Mobility	Student Towns	7752	7346	7280	7107	6957	5763	6215	6728	7128	7269
Low Mobility	Successful Family	-5110	-4741	-4133	-4634	-4132	-3840	-3398	-4045	-4401	-3792
Student Towns	Coastal and Rural	-8674	-11024	-11539	-11634	-10398	-10613	-8473	-8989	-10509	-10034
Student Towns	Working Class Local	11771	13100	12779	13316	14546	14184	13690	13329	13625	13347
Student Towns	Dynamic London	19630	19769	19917	18365	17185	17695	18285	19728	21638	22703
Student Towns	Commuter Britain	3288	2065	2929	3302	2303	2504	2810	2115	1866	2565
Student Towns	Moderate Mobility	3058	3906	3799	3841	3700	3914	3671	3476	3887	2853
Student Towns	Low Mobility	11533	12928	12629	13102	14254	12687	11350	11487	11904	11342
Student Towns	Student Towns	-36515	-36629	-35860	-36729	-37965	-37028	-38027	-37925	-39537	-39177
Student Towns	Successful Family	-4955	-4949	-5403	-4346	-4466	-4186	-4192	-4176	-4035	-4687
Successful Family	Coastal and Rural	7409	8252	9120	9402	9219	9579	8857	8909	9968	9003
Successful Family	Working Class Local	-7666	-7510	-7832	-7919	-7832	-8331	-7681	-7839	-8247	-8250
Successful Family	Dynamic London	-10366	-10129	-10312	-10177	-9918	-9990	-9934	-10111	-10711	-10555
Successful Family	Commuter Britain	-2720	-2464	-1899	-2385	-1908	-2181	-1723	-2147	-3051	-2605
Successful Family	Moderate Mobility	1432	1050	1281	1660	889	1388	820	1215	1596	1034
Successful Family	Low Mobility	-6764	-6806	-6709	-6752	-6971	-6870	-6482	-6292	-7235	-6500
Successful Family	Student Towns	3874	2961	2374	2044	1867	2332	1916	2022	1889	2664
Successful Family	Successful Family	14606	14466	13814	13955	14472	13891	14034	14028	15527	14969

Table 8.5: Gross destination specific model residuals for observed vs. expected inter-cluster total migrant migration rates 1998/99 - 2007/08

		Origin specific distance decay									
Origin Oi	Destination Dj	1998/99	1999/00	2000/01	2001/02	2002/03	2003/04	2004/05	2005/06	2006/07	2007/08
Coastal and Rural	Coastal and Rural	12435	14707	14390	15513	15624	14217	14141	14079	15870	15461
Working Class Local	Coastal and Rural	-1381	-721	-213	-694	-960	-3170	-3830	-2882	-3785	-3697
Dynamic London	Coastal and Rural	-8599	-11353	-11545	-13274	-13990	-12372	-11061	-11039	-11769	-11030
Commuter Britain	Coastal and Rural	-740	-1561	-1574	-1049	-1486	520	-83	-148	-719	-264
Moderate Mobility	Coastal and Rural	4360	4323	3625	3917	4040	5254	4276	4096	5025	4488
Low Mobility	Coastal and Rural	-5923	-4770	-4371	-4498	-4149	-4898	-4933	-5507	-5894	-5254
Student Towns	Coastal and Rural	-10937	-11670	-12215	-12058	-10821	-12497	-10791	-11359	-12831	-12644
Successful Family	Coastal and Rural	10785	11045	11903	12143	11744	12945	12282	12761	14103	12941
Coastal and Rural	Working Class Local	124	-1007	-1678	-1518	-1962	-2405	-1879	-1488	-2189	-1622
Working Class Local	Working Class Local	313	1738	2037	1449	-187	2616	2210	2854	4118	3657
Dynamic London	Working Class Local	1296	1536	1974	2488	3739	3018	2787	2339	2043	1531
Commuter Britain	Working Class Local	-538	-896	-362	-242	-490	-541	-1061	-1111	-653	-913
Moderate Mobility	Working Class Local	368	539	133	897	515	343	-65	173	-181	295
Low Mobility	Working Class Local	-5069	-5744	-5386	-6601	-6474	-6615	-5705	-6044	-6174	-5567
Student Towns	Working Class Local	9274	9567	9251	9717	11010	10564	10050	9575	9572	9214
Successful Family	Working Class Local	-5767	-5732	-5968	-6191	-6151	-6980	-6336	-6298	-6537	-6595
Coastal and Rural	Dynamic London	-7047	-7277	-7566	-7271	-6942	-7115	-7454	-7653	-7897	-8715
Working Class Local	Dynamic London	6321	6111	6050	6030	6133	5981	5941	5898	5810	5777
Dynamic London	Dynamic London	38371	37458	36293	35450	35021	33512	33023	33492	35279	34553
Commuter Britain	Dynamic London	-28466	-27101	-25706	-25289	-24520	-23662	-23004	-24170	-24723	-24037
Moderate Mobility	Dynamic London	-21444	-20342	-20044	-19767	-19431	-19836	-19911	-19728	-21512	-21326
Low Mobility	Dynamic London	-865	-1323	-1550	-1171	-1273	-858	-1006	-935	-1138	-1447
Student Towns	Dynamic London	24389	23711	24028	22927	21648	22565	23185	24426	26231	27296
Successful Family	Dynamic London	-11259	-11238	-11504	-10910	-10636	-10588	-10773	-11328	-12050	-12101
Coastal and Rural	Commuter Britain	-8107	-7751	-7676	-7940	-7248	-7090	-6859	-7200	-8100	-7506
Working Class Local	Commuter Britain	516	796	344	-251	95	283	351	-63	145	111
Dynamic London	Commuter Britain	-14060	-14429	-14491	-13088	-14316	-14205	-14821	-13635	-13739	-12444
Commuter Britain	Commuter Britain	16558	16567	15940	15900	15704	15069	14678	15193	15941	14373
Moderate Mobility	Commuter Britain	2394	1505	1804	2177	1725	1291	1991	1794	2062	1296
Low Mobility	Commuter Britain	2562	3012	2456	2700	3205	3349	2577	2878	3712	2915
Student Towns	Commuter Britain	5018	4580	5372	4956	4601	5294	5645	5212	5212	5884
Successful Family	Commuter Britain	-4880	-4281	-3751	-4454	-3767	-3993	-3562	-4178	-5235	-4628
Coastal and Rural	Moderate Mobility	-5659	-5789	-5147	-5647	-5485	-4823	-4510	-4811	-5482	-4640
Working Class Local	Moderate Mobility	-1065	-1641	-1709	-1835	-1268	-1512	-1224	-1275	-1514	-1263
Dynamic London	Moderate Mobility	-3679	-1990	-1584	-1469	-724	-1262	-1429	-1533	-1276	-1841
Commuter Britain	Moderate Mobility	2355	1545	707	585	226	-765	-136	-386	-473	128
Moderate Mobility	Moderate Mobility	883	566	714	14	300	264	86	36	-339	93
Low Mobility	Moderate Mobility	6708	6427	6109	6999	6443	6920	6123	6778	7169	6404
Student Towns	Moderate Mobility	1305	2069	1887	1945	1864	2110	2494	2365	2875	2416
Successful Family	Moderate Mobility	-848	-1186	-977	-593	-1354	-933	-1405	-1173	-960	-1296
Coastal and Rural	Low Mobility	-1083	-1457	-1354	-1581	-2079	-1404	-1002	-1095	-826	-765
Working Class Local	Low Mobility	-5341	-6708	-5954	-6114	-5541	-5460	-5132	-5011	-6192	-5739
Dynamic London	Low Mobility	-2456	-2357	-2045	-1955	-2259	-1591	-1829	-2393	-1549	-2240
Commuter Britain	Low Mobility	1563	1334	624	497	314	-112	-437	183	586	231
Moderate Mobility	Low Mobility	2917	3159	3076	2486	2213	2236	2997	2652	3365	3802
Low Mobility	Low Mobility	-552	-399	-572	-118	-465	-114	1	-16	-627	-636
Student Towns	Low Mobility	11891	13288	12917	13540	14791	13439	11994	12025	12586	11939
Successful Family	Low Mobility	-6939	-6860	-6691	-6755	-6974	-6993	-6590	-6344	-7342	-6592
Coastal and Rural	Student Towns	2303	2196	2852	2615	1755	1180	897	679	857	-63
Working Class Local	Student Towns	5071	5286	4071	5702	6359	6134	6025	5097	6622	6301
Dynamic London	Student Towns	6252	8454	8747	10077	10705	11179	10454	10225	10233	9511
Commuter Britain	Student Towns	7033	6463	6490	5556	6476	6669	8255	8637	7860	8117
Moderate Mobility	Student Towns	4859	4744	4661	4108	4416	4054	4609	4906	5528	5999
Low Mobility	Student Towns	7282	6719	6739	6625	6496	5618	6034	6430	6752	6940
Student Towns	Student Towns	-37364	-37757	-36888	-37553	-38825	-37532	-38602	-38558	-40365	-40079
Successful Family	Student Towns	4564	3895	3329	2870	2618	2699	2329	2584	2514	3275
Coastal and Rural	Successful Family	6809	6156	5980	5620	6115	7227	6452	7261	7481	7602
Working Class Local	Successful Family	-5423	-5894	-5534	-5215	-5576	-5724	-5178	-5507	-6355	-6166
Dynamic London	Successful Family	-15585	-15752	-15937	-16744	-16606	-16840	-15694	-15948	-17309	-16326
Commuter Britain	Successful Family	2780	4216	4371	4535	4278	3280	2239	2280	2800	2892
Moderate Mobility	Successful Family	5948	5806	6295	6434	6497	6646	6266	6342	6402	5657
Low Mobility	Successful Family	-4375	-4156	-3632	-4155	-4014	-3614	-3296	-3799	-4079	-3598
Student Towns	Successful Family	-4373	-4610	-5093	-4243	-5085	-4695	-4725	-4479	-4287	-4923
Successful Family	Successful Family	14219	14234	13550	13769	14391	13721	13936	13851	15347	14862

Table 8.6: Gross residuals, observed and predicted flows, 1998/99

a) Gross residuals (origin specific β)	Coastal and Rural Retirement Migrants	Declining Industrial, Working-Class, Local Britain	Dynamic London	Footloose, Middle-Class, Commuter Britain	Moderate Mobility, Non-Household, Mixed Occupations	Low Mobility Britain	Student Towns and Cities	Successful Family In-migrants
Coastal and Rural Retirement Migrants	11429	-3412	-6221	-4027	-578	-9683	1717	10427
Declining Industrial, Working-Class, Local Britain	-681	5012	3754	-850	-9	-6092	3785	-6000
Dynamic London	-3308	4623	27183	-20780	-11599	3766	14758	-12620
Footloose, Middle-Class, Commuter Britain	-5329	-2324	-18849	18573	744	2920	4630	144
Moderate Mobility, Non-Household, Mixed Occupations	-1783	-3594	-10957	6143	2131	4716	-1	3507
Low Mobility Britain	936	-4410	-4174	372	4821	-396	7752	-5110
Student Towns and Cities	-8674	11771	19630	3288	3058	11533	-36515	-4955
Successful Family In-migrants	7409	-7666	-10366	-2720	1432	-6764	3874	14606
b) Gross residuals (destination specific β)								
Coastal and Rural Retirement Migrants	12435	124	-7047	-8107	-5659	-1083	2303	6809
Declining Industrial, Working-Class, Local Britain	-1381	313	6321	516	-1065	-5341	5071	-5423
Dynamic London	-8599	1296	38371	-14060	-3679	-2456	6252	-15585
Footloose, Middle-Class, Commuter Britain	-740	-538	-28466	16558	2355	1563	7033	2780
Moderate Mobility, Non-Household, Mixed Occupations	4360	368	-21444	2394	883	2917	4859	5948
Low Mobility Britain	-5923	-5069	-865	2562	6708	-552	7282	-4375
Student Towns and Cities	-10937	9274	24389	5018	1305	11891	-37364	-4373
Successful Family In-migrants	10785	-5767	-11259	-4880	-848	-6939	4564	14219
c) Observed flows								
Coastal and Rural Retirement Migrants	62916	36426	15074	22540	21149	13673	65743	46425
Declining Industrial, Working-Class, Local Britain	45587	116337	13679	15474	15919	27937	100475	31374
Dynamic London	23672	11528	227888	73342	49713	16294	43216	19293
Footloose, Middle-Class, Commuter Britain	32800	13791	49391	78296	40756	20345	49423	37054
Moderate Mobility, Non-Household, Mixed Occupations	30818	14178	28750	41951	31236	21429	40334	33178
Low Mobility Britain	19710	25936	10650	18279	22381	17309	49782	18149
Student Towns and Cities	69391	107332	53411	49683	42784	58144	96013	63557
Successful Family In-migrants	47967	23336	13215	25301	24335	13463	58497	48568
d) Model predicted flows origin								
Coastal and Rural Retirement Migrants	51487	39838	21295	26567	21727	23356	64026	35998
Declining Industrial, Working-Class, Local Britain	46268	111325	9925	16324	15928	34029	96690	37374
Dynamic London	26980	6905	200705	94122	61312	12528	28458	31913
Footloose, Middle-Class, Commuter Britain	38129	16115	68240	59723	40012	17425	44793	36910
Moderate Mobility, Non-Household, Mixed Occupations	32601	17772	39707	35808	29105	16713	40335	29671
Low Mobility Britain	18774	30346	14824	17907	17560	17705	42030	23259
Student Towns and Cities	78065	95561	33781	46395	39726	46611	132528	68512
Successful Family In-migrants	40558	31002	23581	28021	22903	20227	54623	33962
e) Model predicted flows destination								
Coastal and Rural Retirement Migrants	50481	36302	22121	30647	26808	14756	63440	39616
Declining Industrial, Working-Class, Local Britain	46968	116024	7358	14958	16984	33278	95404	36797
Dynamic London	32271	10232	189517	87402	53392	18750	36964	34878
Footloose, Middle-Class, Commuter Britain	33540	14329	77857	61738	38401	18782	42390	34274
Moderate Mobility, Non-Household, Mixed Occupations	26458	13810	50194	39557	30353	18512	35475	27230
Low Mobility Britain	25633	31005	11515	15717	15673	17861	42500	22524
Student Towns and Cities	80328	98058	29022	44665	41479	46253	133377	67930
Successful Family In-migrants	37182	29103	24474	30181	25183	20402	53933	34349

where the highest volume of migration occurs, this can be interpreted as the excessive influence of life-course related factors on flows between districts in the clusters. The pull to university towns of migrants in their late teens is strong with flows from rural areas already recognised earlier in this analysis. Perhaps less obvious thanks to the significance of many other flows related to Urban London, but no less important, are the flows between London and university towns (demonstrated for selected wards in Britain by Duke-Williams 2009a). The prohibitive cost of living in the capital, along with a desire to move away from the parental domicile, will be push factors for many students who grew up in London and certainly will contribute to a small (an average of around 6,500 migrants annually over the decade of study), but still significant (compared with other flows) number of moves out of the city. Those students attracted to any of London's multitude of universities from elsewhere in Britain may feel more of a desire to leave the capital after finishing their study for similar reasons. The attraction of London to both students and those just leaving university and looking for a first job is strong. The strength of this attraction is likely to be a significant contributing factor to the large under-estimate of flows from Student Towns and Cities to Dynamic London. Less clear is why this under-estimate by the model continues though-out all older age groups. Interrogation of the raw data (not presented here) reveals that beyond age 45, gross flows decline quite considerably, from an average of around 10-15,000 at age group 30-44 to an average of around 2,000 at age 45-59. Where flows are much smaller, smaller errors in the model can account for larger differences in the standardised residuals (as comparisons between Table 8.4a and Table 8.5b show), however the larger observed flows from Dynamic London to University Towns at age group 30-44 are more difficult to account for where all flows in this direction between the clusters are dominated by migrants in younger age groups. Looking across the residual matrices (Table 8.6a and b), part of this model under-prediction could be a consequence of significant over-prediction of flows within the Student Towns and Cities cluster.

The other stark under-prediction of the model concerns the flows between Declining Industrial, Working-Class, Local Britain and Dynamic London and vice versa (Tables 8.4 and 8.5). These under-predictions are across all age groups, but interrogation of the gross flows reveals that many of the largest percentage differences are actually the result of some very low observed and predicted flow volumes, some numbering in the hundreds rather than thousands. A similar situation is presented in the standardised residuals with an under-prediction of flows within the Footloose, Middle-Class, Commuter Britain cluster. Here the reasons for this under-prediction cannot be attributed in the same way as the volume of flows within this cluster is large. The under-prediction is across all age groups, and is likely to be a reflection of the particular attributes of migrants associated with this cluster: as is shown in Chapter 6, the cluster is characterised by net in-migration of migrants in the higher socio-economic groups. The data shows that many of these migrants arrive from London at age group 25-29, which is then succeeded by a large volume of intra-cluster migration at age group 30-44. Whether the migrants moving at 30-44 will have also been those who arrived at 25-29 is irrelevant, as the cluster profile suggests that regardless of this they will be of a similar type - they may still be maintaining higher status and well paid jobs in the capital, or in urban areas situated within or adjacent to this fairly compact cluster. With migrants tethered to the cluster though employment, it is likely that moves within the cluster are to be related to residential improvement rather than any big change in status (such as student to employed or employed to retired), therefore moves are likely to be less affected by physical distance, meaning that the frictional effect that distance has on moves between districts within the cluster should be reduced for a better model fit.

Other noticeable under-predictions in the model concern the Successful Family In-Migrants and Coastal and Rural Retirement Migrants clusters, both concerning flows between the clusters and within. The interaction between the two clusters is perhaps not too surprising when one considers their high index of association values (Table 7.1) - values of 0.94 being as high or higher than their intra-cluster association values, and not very much lower than the highest value of 0.97 for Dynamic London (the lowest being 0.84). These clusters are both located in relatively rural, peripheral areas, with a number of districts in one cluster adjacent to districts in the other. The model predictions are better (i.e. exhibiting much smaller residuals) for the older age groups where one would expect migration between these clusters to have more relative importance in the system - the clusters being characterised by flows of migrants in the older age groups. The model under-predictions are occurring more severely where moves within and between the clusters are happening in the younger age groups, perhaps reflecting the greater difficulty of predicting the much less obvious (but still occurring), migration flows in these age groups; flows which might be regarded as a noisy aside in any parsimonious explanation, but which still contribute (as shown in Table 8.6c) considerable numbers of migrants to the system at the ages of peak migration. Again though, the model under-predictions here may also be influenced by the model over-predictions of flows between Coastal and Rural Retirement migrants and other clusters in the system - particularly Low Mobility Britain; a cluster where the classification shows migration flows are much less likely to happen.

There is considerable over-prediction of flows in both directions between Dynamic London and the Successful Family In-migrants cluster and for the flows between Dynamic London and Footloose, Middle Class, Commuter Britain. The two cases are somewhat different in that the migration exchanges between Dynamic London and Footloose, Middle Class, Commuter Britain are some of the highest in the whole system, whereas the flows between Dynamic London and the Successful Family In-migrants are some of the lowest (Table 8.6c), therefore the context for the explanation differs somewhat. Spatially, the two clusters occupy very different locations in relation to London. Much of Footloose, Middle Class, Commuter Britain cluster is located in the South, bordering Dynamic London - indeed around half of the districts contained in this cluster are located within the London City Region (Champion et al., 2007), with a similar number falling within the London travel-to-work catchment defined at the 25% self containment cut-off (ODPM, 2006); areas defined by their flow linkages. With this strong association it is not a surprise that more flows would take place between these clusters, the over-prediction is a likely artefact of the large under-prediction of flows within the Dynamic London cluster; a cluster not spatially dissimilar to Footloose, Middle Class, Commuter Britain and an area with enough physical, social and economic variation within a small space to cater for the aspirations of migrants who, for employment related and other reasons, might want to remain physically close to the city. The over-prediction of flows into the Successful Family In-migrants cluster is less severe in gross migrant terms, but still large and there is little change in this over-prediction by age group. One interpretation of this could be that the attractiveness of this cluster or repulsiveness of Dynamic London to migrants who do make the move is lower in reality. Given the much older age profile of those migrants who tend to move into this cluster, it might be that if you remain in London to an older age there is a good chance that you will have more of an attachment to the city and be less influenced by the factors which either drive others out to more rural peripheral areas, or indeed those factors which draw them in. The frictional effect of physical (and by extension the socio-economic/cultural) distance might be higher in reality.

The very large over-prediction of flows which stands out, certainly for gross residuals but also for standardised residuals is seen in the intra-Student Towns and Cities flows. Important to an extent across all age groups, the effect is especially pronounced at the age group 16-19 and 20-24 - the ages where significant moves occur into and out of this cluster from other clusters. As a cluster, Student Towns and Cities has the largest number of in and out-migrants of all clusters - an average of over 500,000 in and out migrants over the ten year period - larger even than the Dynamic London cluster. These large constraints lead to the possibility of a larger gross residual error within the cluster. It is also possible that fewer migrants are moving within the cluster as the Student Towns and Cities cluster is synonymous with change of status moves (from parental guidance to independence or from Student to employed worker) and a move within the cluster is less likely to be associated with a significant change in status. More likely, however, is that a large contribution to this effect is because the model fails to fully capture the attractiveness of Dynamic London to migrants leaving Student Towns and Cities and the push influences encouraging students out of London discussed earlier. Scanning across the residuals matrices (Tables 8.6a&b), a large proportion of the over-prediction of the model could be attributed to large under prediction in the same row and column where Dynamic London crosses Student Towns and Cities.

8.6 A discussion of spatial interaction modelling results

Doubly constrained spatial interaction models with origin/destination specific parameters were fitted to known data in the Migration Classification system in an effort to explore the effects of the clusters on migration flows in Britain. As discussed, the Migration Classification system is a very different spatial system to those which are normally used in spatial interaction modelling in that the modelled flows were not between discrete geographic entities with one location for each origin and destination, but rather between aggregations of smaller origin and destination zones aggregated according to their migrant characteristics. The non-standard nature of the system and the particular models used to examine the flows could have contributed to errors in the modelled expected migration flows. For example, the results have shown that whilst the models have displayed reasonable goodness-of-fit to observed data overall, there are in some cases large residual values where the model either over or under predicts flows. Where conventional spatial interaction models have fallen short, one of the more well know criticisms and potential solutions has been tabled by Fotheringham (1983a,b, 1984, 1986a,b); Fotheringham et al. (2001). Fotheringham's specific criticism relates to the assumption in standard spatial interaction models used to model human systems, that the decisions made by individuals do not involve any kind of comparative evaluation - i.e. all possible destinations are treated in the same way. He argues that individuals do in practice make far more evaluative decisions, although due to the multitude of possible destinations they tend to select final destinations from a relatively small pool. This pool is determined through a process of hierarchical information processing, where clusters of destinations are first identified and then individual destinations selected from within that pool. Fotheringham (1983a) demonstrates that under certain conditions spatial clustering of destinations can either encourage (the effect of agglomeration) or discourage (the effect of competition) interaction flows to individual destinations - the latter being more likely than the former. To cope with this issue, Fotheringham proposed a variation on the standard spatial interaction model called the 'competing destinations' model. This model builds in another variable which measures the accessibility of one destination to all other destinations in the system.

Fotheringham has repeatedly shown that for singly constrained models his competing destinations model offers an improvement over the more conventional spatial interaction model (Fotheringham, 1986b; Fotheringham et al., 2001). He also argues that where these 'map pattern effects' are operating to influence the flow of individuals, then the interpretation of distance decay parameters becomes dangerous until the influence (or not) of agglomeration or competition is accounted for. Fotheringham's work, however, concerns itself with the singly constrained versions of spatial interaction models, which given their continued application in predictive trip distribution makes sense as with these models the output is of most importance. The work is also applied to more conventional spatial systems. The extent to which these observations on competing destinations and map pattern effects are maintained for doubly constrained models and for non-standard spatial systems such as the classification system used here are unclear. Exploration of these issues cannot take place now, but they do offer potentially fruitful avenues of future research.

At the beginning of this section, a question was posed asking that given the known effect of distance and age on migration flows, what flows might be expected between clusters in the migration classification system. Bearing in mind the caveat the theory of competing destinations and the map pattern effect places upon the interpretation of the results, fitting doubly constrained spatial interaction models has added a new dimension to the analysis of flows between clusters in the system. It has shown that a large amount of variation in the migration patterns in the system, across all age groups, can be explained by the effects of the origin and destination masses, and the average distance between districts within the clusters. Where the model has fallen short of providing perfect estimates of flows, the residual error data has given clues where the particular features of the clusters in the Migration Classification system have influenced the observed flows. From observations made during the analysis, it was apparent that many of these features were linked to stages in the life course, therefore the last part of this chapter will look to offer a life course perspective on the flows within the Migration Classification system.

8.7 Expected migration - the role of age and life course stage

Much of the analysis in this thesis has made reference to the effect of age on the propensity to migrate. It could be argued that numerical age is not in itself a causal influence on migration, but it is instead the different interlinked social, cultural and economic influences which appear at certain ages in line with stages in the life course which influence migration moves - a point emphasised by Stillwell (2008). For example, retirement moves occur at around age 60-65 in Britain currently, as this is the legal age at which retirement can happen. At the time of writing this, news stories abound in the press that this age will increase soon in line with demographic aging trends. One would assume that when this happens, retirement migration will occur at older ages in Britain. Indeed Courgeau (1985) concludes that when family life-cycle and work are controlled for in a statistical model of migration, the effect of age either disappears or at least is greatly reduced, a finding echoed by Sandefur and Scott (1981). A number of studies make explicit the link and indeed the complex interactions between life course and migration events (De Jong and Graefe, 2008; Geist and McManus, 2008), but here the explicit linkages will be reviewed.

Bailey (2009) reviews a number studies which make explicit links between migration and particular stages in the life course, pointing out the associations with moves in and out of the labour market, partnership, family formation and childbirth, housing careers, separation and retirement. One move not mentioned by Bailey but which in the context of Britain at least is probably the first significant life course migration move, is that of the student away from home to study. Evidence from the UK census and HESA indicates that a very large proportion of the moves at the end of the teenage years are associated these student moves. Little research has been carried out internationally on these migration flows, perhaps reflecting the particularly British cultural phenomenon of both large volumes of students and the propensity to attend an institution which is not necessarily located near to the parental domicile - see Dorling (2010) for an interesting perspective on why this phenomenon exists. Duke-Williams (2009a) presents one of the few quantitative analyses of this phenomenon.

The post-educational phase of the life-course is inextricably linked to the economic imperative of finding employment. Work on employment influenced graduate migration has been carried out in the UK by Faggian and others (Faggian et al., 2006, 2007), and even more research has been carried out in relation to economic influences on migration in general. For example Clark and Hunter (1992) examine the economic influences on migration in the United States. Plane and Jurjevich (2009) also note this link to internal migration in the United States in the young adult ages observing patterns of young adult migrants moving up the urban hierarchy towards the largest 'megametro' areas, with Fielding (1992) and Findlay et al. (2009) noting a similar move of young economic migrants towards the South East metropolitan region in England. Other research has shown that where migrants are more educated, their moves are likely to be over longer distances; Sjaastad (1962), pointing to the increased propensity to migrate over distance of university graduates. Economically motivated migration at the post-educational phase of the life course need not always be associated with moves to areas with increased employment opportunities; migration moves of young adults have also been shown to be influenced by the availability of parental support (Michielin et al., 2008).

Plane and Jurjevich go on to describe the next phase of the life course - the settled career, family rearing phase - in the United States, which sees a move to more suburban areas. In the United Kingdom, Grundy and Fox (1985), making observations from the 1971 Census, note an 'almost universal' migration of women after a marriage event. This pattern is also noted in Germany by Mulder and Wagner (1993) who demonstrate that marriage influences women to migrate over longer distances than men, and that whilst being married negatively influences the propensity to migrate in general, over shorter distances this is reversed. Marriage related migration is also observed in a review of several papers by Kulu and Milewski (2007), but in their review they also note also evidence of increased migration flows after marriage and immediately before first-births, backing up the observations of Plane and Jurjevich (2009) in the United States that a significant number of 'nesting' moves are made to places deemed more appropriate for child rearing, and the family formation moves noted in the UK by Bramley et al. (2006). Of course, the other family-associated life course event which is occurring more and more frequently is that of family breakup. Flowerdew and Al-Hamad (2004) study the effects of family formation and dissolution on migration behaviours in Britain, discovering that although the migration peaks associated with marriage are higher, there are still noticeable peaks also associated with separation and divorce events.

The next influential life course event linked to migration which has been widely studied is that of retirement. Bures (1997) makes the distinction between the 'pre-elderly' (55-64), 'young old' (65-74) and 'old-old' (75+) age groups which exhibit different migration behaviours in the United States. She notes that as the job-related reasons for residential location diminish as individuals head towards retirement and 'empty nests' become the norm with children having reached adulthood and left the family home, there is increased propensity for pre-elderly groups to move to new residential locations as their priorities change. Retirement migration has been one of the more well studied life course related migration events, with a number of pieces of research studying the phenomenon both internally within Countries such as the U.S. (Duncombe et al., 2003; Glasgow and Brown, 2008; Haas and Serow, 1997), Canada (Northcott, 1985), Australia (Drysdale, 1991) and Sweden (Ekstrom and Danermark, 1993) to name but a small selection, and externally with many British retirement migrants moving overseas (Oliver, 2007; Zasada et al., 2010). In the U.S., the motivations for these moves and the means by which they can be achieved are clear. As Plane and Jurjevich (2009, p.6) write succinctly:

> *"Active 'young elderly' can now count on perhaps two decades during which they may zealously pursue their later-in-life passions. They might choose to continue working, on at least a part-time basis. This, like the early adult stage, is a phase in life when intergenerational ties no longer bind so strongly. At this time, when people are freed to move about the country, long-distance exploration and migration occur."*

Whether the migration moves are internal or international there is a broad trend shown by much of this research whereby young elderly, retirement migrants tend to move down the urban hierarchy Plane et al. (2005). Destination areas are more likely to be characterised by *"lower-density living environments, less congestion, higher natural amenities, and cheaper housing"* (Plane et al., 2005, p.15317), certainly in the U.S., although the experience in Sweden - a country with a well developed welfare state - does not echo that of the U.S. (Ekstrom and Danermark, 1993). In Britain there has also been a noticeable move to coastal areas of retirement migrants (Uren and Goldring, 2008).

The final life course migration is that associated with increasing old age and in many cases a growing need for support. Migration flows of the elderly, however, can be nuanced - as shown by research in Sweden by Pettersson and Malmberg (2009). Mobility increases with age and varies by gender, but migration to locations near adult children were more prevalent among the 'young-old' - moves they see associated with close family ties and social interaction rather than moves motivated by a need for care. They note that these types of moves occur less frequently as the 'young-old' move towards being 'old-old'. They interpret the relatively high instances of 'old-old' migration moves as being associated with moves into care institutions - something which in Sweden at least is handled by the State more commonly than the family.

8.7.1 Life course influences on flows within the Migration Classification

So research into life course influences on migration offer a useful micro perspective from which to examine the more macro level migration flows within the migration classification system. The age-specific patterns of migration have already been examined in some detail in Chapter 7, but it remains to link these age-specific flows with life-course events for a more complete understanding of contemporary internal population migration in Britain.

Age 0-15 When examining the migration patterns of the 0-15 age group, it should be borne in mind that in Britain, these migrants will very rarely be migrants who move under their own influence - unlike the situation in developing world countries where independent child migration is far more common place (Whitehead et al., 2007). Most frequently this will be the influence of parents or guardians, either through family household moves where children move with a parent or both parents, or through the dissolution of a family household and the move of children with one or other parent. It is the case, therefore, that in examining this group the parents of the moving child are also being studied. This is important as it is difficult to separate parents from non-parents in the older age groups. It is likely that many of the parents of children in the 0-15 age group will fall in the 30-44 age group, however a significant number will also fall in the 20-24 and 25-29 age groups, as well as some in the 16-19 and 45-59 groups. Moreover, not all of the migrants in the family formation age groups will actually have children and be migrating with them. Consequently the 0-15 age group will be the best to examine for family-related life course influences on migration.

Evidence from Chapter 7 suggests that the highest rates of out-migration of this age group are from Dynamic London - a rate which increases over the decade - with the highest rates of in-migration into Successful Family In-migrants cluster and Coastal and Rural Retirement Migrants - rates which reduce over the decade - and into the Footloose, Middle Class, Commuter Britain cluster. In terms of gross flows, Figure 7.8 shows that most migrants move within Dynamic London and out from Dynamic London to the Footloose, Middle-Class, Commuter Britain and Moderate Mobility, Non-Household, Mixed Occupations clusters. These moves have to be linked to the family formation, career progression-type life course moves discussed earlier - Stillwell and Hussain (2008) have already demonstrated the retreat to the suburbs of many migrants within the London region. This retreat is completely in line with the move down the urban hierarchy observed by Plane et al. (2005) in the U.S. of 30-39 year olds. Analysis of flows out of Dynamic London over the decade has shown that migrants in this age group moving out of districts in the cluster are more affected by the frictional effect of distance than they are for any other cluster other than Low Mobility Britain (Figure 8.2). At the same time, the average flow distance for migrants moving out of this age group is lower than it is for any other age group (Figure 8.3), both measures reflecting the trade-off parents of migrants in the 0-15 age group are making between the economic pull of the capital, with the environmental and quality of life pull of more peripheral areas. Indeed, modelling results shown that observed migration moves within the Dynamic London are considerably higher than would be expected if migration were just a function of mass and distance. With the high rates of migration (but lower gross flows) into the Successful Family In-migrants cluster and Coastal and Rural Retirement Migrants clusters, similar life-course related influences will be occurring within the urban districts elsewhere in Britain, but on a much smaller scale, with slightly higher average distances and lower distance decay effects pointing to weaker attraction forces operating outside of Dynamic London, with modelled flows more often under predicting for these clusters.

Age 16-19 At 16-19, as mentioned several times already, the pull to the Student Towns and Cities cluster is considerable. Not all migrants in this age group migrating to this cluster will be students, but as the work of Duke-Williams (2009a) has shown (albeit at ward level, so one should be cautious of making an ecological inference), many student age migrants moving into areas characterised by high numbers of students are likely to be students. Where migrants are not students, other life course factors will be exerting an influence. Again, referring back to the work of Plane et al. (2005) on life-course influenced moves of the young up the urban hierarchy, it is probable that districts in the Student Towns and Cities cluster, being large urban centres in their own right, will attract young economic migrants. Flows into the Student Towns and Cities cluster are high from all clusters so unpicking particular life-course/cluster associations from conventional metrics is difficult. However, the model residuals presented earlier in this chapter give some clues. The highest net in-migration rate comes from Footloose, Middle Class Commuter Britain, followed by Coastal and Rural Retirement Migrants. Whilst both sets of flows are important, examination of the model residuals points to the observed flow from Footloose, Middle Class Commuter Britain being much higher than the modelled expected. This might point to particular influences affecting the life course of some migrants more than others, such as the increased expectation of higher education participation amongst the families of those young migrants whose parents also attended higher education institutions. Given the composition of the population in Footloose, Middle Class Commuter Britain this kind of differentiated life course expectation would make sense - where attending university is not part of the established family life course, then the life course imperative to move will be lower.

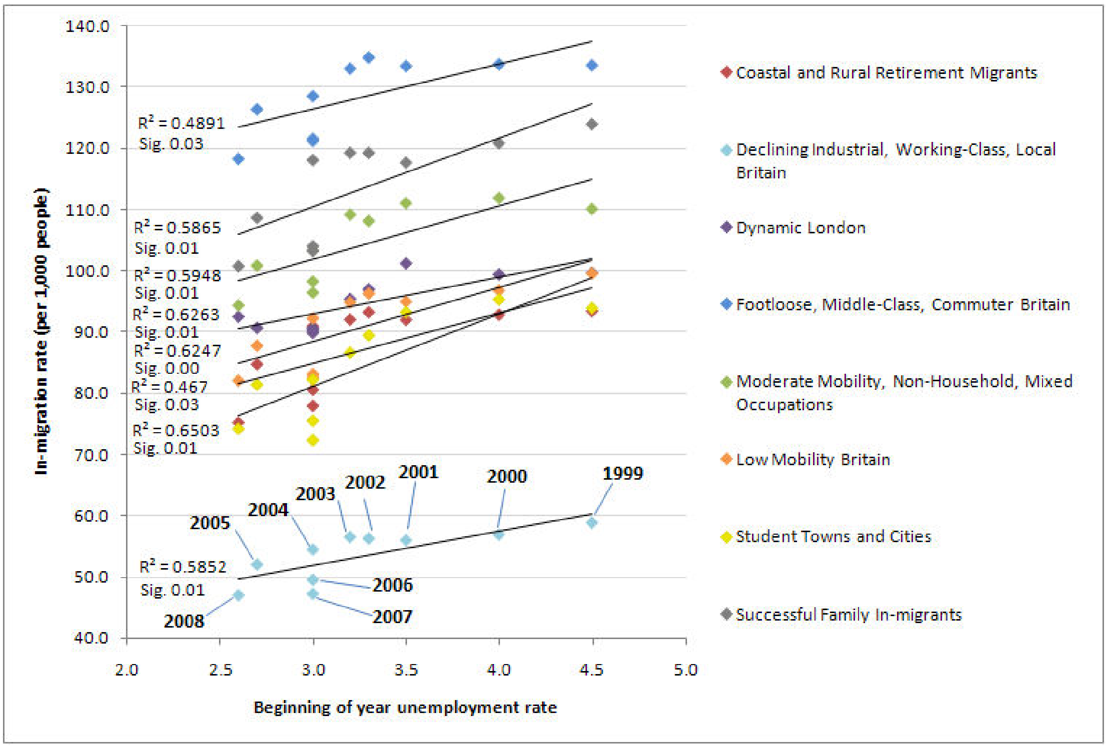
Age 20-24 At 20-24, there is a considerable exodus from the Student Towns and Cities cluster, fuelled by the search for employment. As has been shown in this chapter and the previous chapter a large proportion of this exodus heads for London as the major centre for high-skilled employment in the UK. But the life course perspective in this instance does not offer a complete explanation when the time series of data are analysed. As is shown in Figure 7.10 there is a considerable reduction in the rate of out-migration of 20-24 year olds from Student Towns and Cities between 1998/99 and 2007/08. It is tempting to attribute at least some of this rate reduction to a concurrent year-on-year increase in the number of students remaining in the higher education system for postgraduate study (http://www.hesa.ac.uk/), and where students continue their education at the same institution. This trend could be translated into a life course explanation, with more people attending universities each year and degrees not offering the workplace advantage they once did, combined with grade inflation (or improving standards) meaning more people achieving higher grades, it is becoming more and more the case that postgraduate study is the norm. This would effectively extend the student portion of the life course into the mid-twenties. The non-life course explanation in this context would be that Student Towns and Cities are managing to retain more graduates, potentially through offering increasing opportunities for graduate level employment. However, with reductions in the rate of out-migration across almost every cluster, both of these explanations are unsatisfactory, especially when it is considered that Dynamic London does not suffer a reduction in in-migration at the same rate as the reduction in out-migration from other clusters. Figure 7.3 Shows that the reduction in the rate of migration is a real one representing a reduction in flows and is not an artefact of increasing population in this age group.

So how might life course offer a perspective on this decline in migration rates at 20-24? The key could still lie partially with the life course influence which acts greatest on this age group - that of the search for employment. Aggregate data on Job Seekers' Allowance (JSA) claimant rates for from the ONS, available as part of the NOMIS service (https://www.nomisweb.co.uk/) for the period mid 1998 to mid 2007, may offer some clues. Figure 8.4 shows the beginning-of-year total unemployment rates (the mid-point in the mid-year to mid-year migration data) for each year in the decade of study, plotted against the in and out migration rates for the 20-24 age group in the same period [1]. The first point of note is that for this ten year dataset, there is a steady decline in unemployment between 1999 and 2008, this means that the time-series runs from right to left.

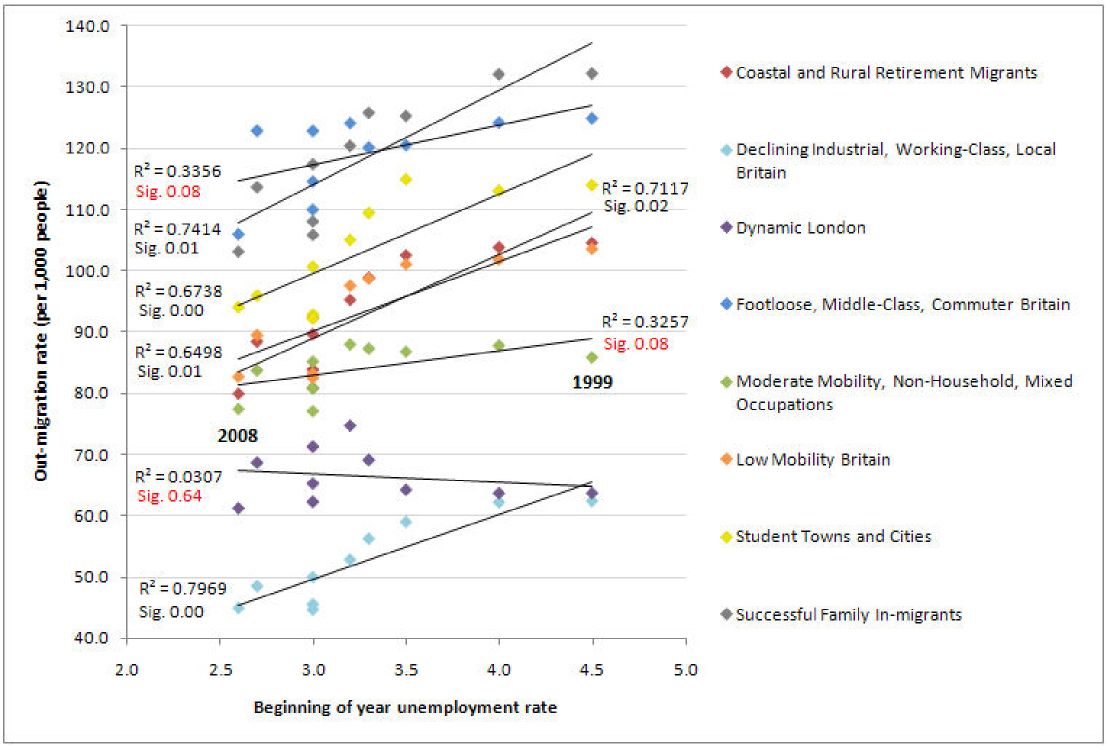
Unemployment rates are at their highest at the beginning of 1999 at 4.5% of the working age population, they then reduce at a steady rate (with some small variation) to 2.6% at the beginning of 2008. Figure 8.4a shows that this decline in unemployment rates is moderately correlated with a decline in in-migration rates across all clusters. The degree of correlation varies from an R^2 of around 47% for Low Mobility Britain, so around 65% for Student Towns and Cities, with all correlations statistically significant at <0.05. It is a similar situation for out-migration rates, although whilst the R^2 value is generally higher across most clusters, for Dynamic London there is almost no correlation. For Footloose, Middle Class, Commuter Britain and Moderate Mobility, Non-Household, Mixed Occupations, correlations are low and not statistically significant.

So what light can this information throw onto the explanation of migration flows? Research elsewhere on unemployment and migration has been inconclusive, some studies showing unemployment encouraging migration (Antolin and Bover, 1997), some showing it discourages flows (Bheim and Taylor, 2002). The positive correlation here could suggest that one of the key factors influencing the flow of migrants in this age group is the availability of employment. When jobs are more plentiful, there is less impetus to leave a location in search of new employment. This would certainly make sense with the 20-24 age group more than any other as this is the age group who are just starting out in employment and who are more likely not to be in a settled job or career. The lack of correlation with Dynamic London and the London hinterland clusters of Footloose, Middle Class, Commuter Britain and Moderate Mobility, Non-Household, Mixed

[1] As is stated on the Nomis website, JSA claimant counts are not an official measure of unemployment, official unemployment rates are taken from the labour force survey, however, JSA claimant counts sample the whole population. Aggregate rates are used here as these data are more easily accessible than age specific rates. Whilst the unemployment rate tends to be higher for younger people, the general trend in rates, which is of interest here, is comparable over time.



(a) In-migration rates



(b) Out-migration rates

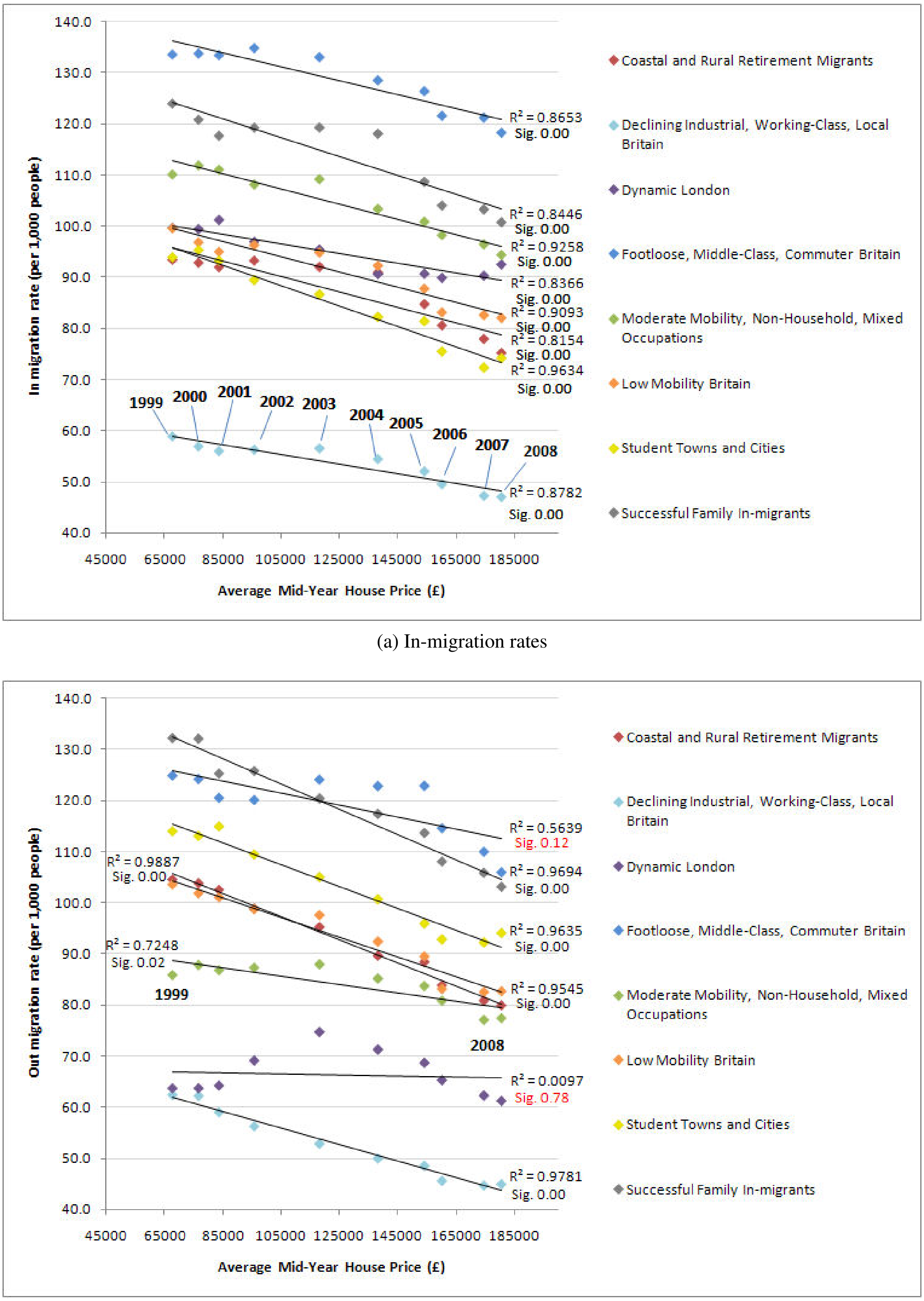
Figure 8.4: Correlation between unemployment rates and 20-24 age group in- and out-migration rates across clusters

Occupations with out-migration is interesting. It suggests that the composition of the employment market in this area differs somewhat from elsewhere in Britain (aggregate unemployment rates are more-or-less the same as the rest of the country). Indeed this assertion is borne out by research by the Greater London Authority (GLA, 2008) which notes that Londoners are more likely to employed in professional and managerial jobs. Whilst 20-24 year olds are less likely to be in at the top of the managerial or professional scale, where higher paid quaternary sector jobs are more prevalent in London and the South East, 20-24 year olds may still be heading towards this socio-economic group if they are not already in it and it could be that employment sector is playing a part here. Certainly it has been noted elsewhere that high relative earnings encourage in-migration to regions and that the effects of higher unemployment affect younger migrants more than old (Murphy et al., 2006), it is likely therefore that similar forces act to discourage out-migration. At this stage though a causal link between employment sector and earnings countering the influence of unemployment related migration of the 20-24 age group in Dynamic London, Footloose, Middle Class, Commuter Britain and Moderate Mobility, Non-Household, Mixed Occupations is just speculative - further research would be required to make a definitive case.

Unemployment offers a plausible, if not wholly convincing, explanation for the reduction in in- and out-migration rates over the decade of study. The economic route to explanation has not yet been fully explored, however. At age group 20-24, not only are young people looking for employment, but for many - especially those who did not do so as a result of entering into the higher education system - they are also looking to move away from the parental home. It follows, therefore, that possible influence on the propensity to migrate at this age would be the housing market. As already noted much earlier in this thesis, the effect of the housing market on migration within the British Isles has been documented before (Cameron et al., 2005; Champion et al., 1998). Cameron et al. note that as age increases, house prices relative to earnings have more of an effect on the propensity to migrate. Despite this, it will be interesting to look at the effect of house prices on the migration propensities of 20-24 year olds.

Despite owner occupation being more prevalent in Britain than it is in many other European countries, it might be expected that owner occupation is proportionally lower in the 20-24 age group and thus some could argue that studying house prices is less relevant for this group. The rental value of property, however, is intrinsically linked to its value and so as the price of property rises, so too does the cost of rental, therefore it is entirely legitimate to look at property prices in general. Quarterly average house price data for the UK was obtained from the Nationwide Building Society (http://www.nationwide.co.uk/hpi/historical.htm) and aggregated for mid-year to mid-year periods for 1998/99 to 2007-08. Comparison of these data with in- and out-migration rates is shown in Figure 8.5 below:

With these graphs, the time-series runs in the opposite direction to those shown for unemployment. This is because from 1999 until 2008 there was a year-on-year increase in the average house price. As with the unemployment rates a correlation exists, however with house prices,



(a) In-migration rates

(b) Out-migration rates

Figure 8.5: Correlation between average house prices and 20-24 age group in- and out-migration rates across clusters

for both in- and out-migration the correlation is much stronger. With in-migration rates (Figure 8.5a) in all cases there is a strong negative correlation ranging from around 82% to around 96%, showing that consistently as house prices increase, the propensity to migrate into a cluster decreases. The negative correlation is also high for most clusters when examining out-migration. As with unemployment, Dynamic London exhibits no relationship between migration and the independent variable and similarly Footloose, Middle-Class, Commuter Britain exhibits a much weaker and statistically insignificant correlation (Figure 8.5b).

The much stronger correlation between increasing house prices and decreasing migration propensity are somewhat easier to understand and offer more simple interpretation, especially in relation to in-migration. As the cost of moving into new accommodation increases this will begin to discourage some from moving - especially those who can least afford to do so. At 20-24, potential migrants will be at the bottom of the career ladder and have lower earnings potentials than those in older age groups. As house prices and rental costs increase relative to earnings, more and more of those at the bottom will be discouraged from moving - maybe opting to live with parents for a few more years until they can. Indeed research by the Skipton Building Society (Skipton, 2003) pointed to this being the case, and research by Berrington et al. (2009) confirms this for older 'young adults'. However, recent data from the ONS (2009c) has shown that in percentage terms the number of 20-24 year olds living with parents has remained quite constant. In their Social Trends report, the ONS detail the increasing number of 20-24 year olds living with parents in 2001 and 2008 - an increase of some 258,000 was recorded, although in percentage terms the proportions in 2001 and 2008 were similar at just over half of men and 35% of women. That said it is also reported that 44% of young adults believe that there is a lack of affordable housing - a perception that could certainly influence potential migrants to remain in one location for longer.

The lack of apparent house price effect for out-migration Dynamic London and Footloose, Middle-Class, Commuter Britain can probably be explained by the differential cost of housing in the South East compared to the rest of the country. As is noted by Murphy et al. (2006) the relative cost of housing in the South East of England is much higher than the rest of the country. With house prices and relative incomes proportionally higher in London and the surrounding commuter belt, out-migration from these areas to other areas in Britain, even when the overall cost of housing is rising, would not be as prohibited by cost.

Where house prices and unemployment are both shown to have an effect on the decline in migration propensities of 20-24 year olds between 1999 and 2008, the logical next step would be to explore whether acting together, both of these economic influences can account for even more of the variation. A simple ordinary least squares regression model was run to explore this with unemployment rate and average house price included as the independent variables and the rates of in- and out- migration (excluding within-cluster flows) from each cluster as the dependent variables. The results of the regression model are shown in Table 8.7 below.

Assessing the R^2 values for each cluster, it is clear that in all but one case (Dynamic

Table 8.7: Results of OLS regression model for house price and unemployment effects on the rate of in- and out- migration decline between 1999 and 2008

		In-migration						Out-Migration					
	Parameter	**Estimate**	**t-value**	**p-value**	R^2	**F**	**p**	**Estimate**	**t-value**	**p-value**	R^2	**F**	**p**
Coastal and Rural Retirement Migrants	Constant	134.15	7.956	0.000	0.868	23.050	0.001	131.74	24.837	0.000	0.993	493.231	0.000
	Slope: Average House Price	-0.22	-4.624	0.002				-0.25	16.744	0.000			
	Slope: Unemployment	-5.86	-1.701	0.133				-2.26	-2.08	0.076			
Declining Industrial, Working-Class, Local Britain	Constant	74.44	8.105	0.000	0.894	29.631	0.000	67.07	10.442	0.000	0.980	173.539	0.000
	Slope: Average House Price	-0.12	-4.523	0.003				-0.15	8.058	0.000			
	Slope: Unemployment	-1.91	-1.013	0.345				1.19	0.908	0.394			
Dynamic London	Constant	108.56	9.317	0.000	0.837	18.034	0.002	108.90	4.565	0.003	0.313	1.593	0.269
	Slope: Average House Price	-0.1	-3.016	0.019				-0.12	1.699	0.133			
	Slope: Unemployment	-0.43	-0.181	0.862				-8.57	-1.754	0.123			
Footloose, Middle-Class, Commuter Britain	Constant	172.42	15.701	0.000	0.928	45.158	0.000	152.92	5.539	0.001	0.595	5.136	0.042
	Slope: Average House Price	-0.21	-6.551	0.000				-0.17	2.111	0.073			
	Slope: Unemployment	-5.59	-2.485	0.042				-4.01	-0.709	0.501			
Low Mobility Britain	Constant	120.76	9.747	0.000	0.918	39.424	0.000	133.35	13.396	0.000	0.967	103.566	0.000
	Slope: Average House Price	-0.18	-5.018	0.002				-0.23	8.242	0.000			
	Slope: Unemployment	-2.35	-0.927	0.385				-3.33	-1.633	0.146			
Moderate Mobility, Non-Household, Mixed Occupations	Constant	140.85	14.53	0.000	0.950	66.837	0.000	121.08	11.944	0.000	0.864	22.251	0.001
	Slope: Average House Price	-0.2	-7.073	0.000				-0.15	5.256	0.001			
	Slope: Unemployment	-3.71	-1.868	0.104				-5.54	-2.667	0.032			
Student Towns and Cities	Constant	126.65	14.779	0.000	0.971	149.207	0.000	144.20	13.89	0.000	0.971	116.439	0.000
	Slope: Average House Price	-0.24	-9.978	0.000				-0.25	8.467	0.000			
	Slope: Unemployment	-3.62	-2.065	0.078				-2.99	-1.405	0.203			
Successful Family In-migrants	Constant	149.88	6.861	0.000	0.851	19.912	0.001	151.04	12.278	0.000	0.969	109.714	0.000
	Slope: Average House Price	-0.22	-3.523	0.010				-0.25	7.187	0.000			
	Slope: Unemployment	-2.72	-0.607	0.563				-0.39	-0.156	0.880			

London for in-migration) the inclusion of unemployment data increased the model correlation; however, in all but two cases the contribution of unemployment is statistically insignificant. In some cases, R^2 values for the combined model are extremely high (99% of the variation in out-migration rates of 20-24 year old from Coastal and Rural Retirement Migrants can be accounted for by variations in unemployment rates and house prices), and for most they are high, but in most cases the high R^2 values are the result of the association with house price rises. The only clusters where unemployment has a significant contribution to the overall model are Footloose, Middle-Class, Commuter Britain for in-migration and Moderate Mobility, Non-Household, Mixed Occupations for Out-migration. These models suggest that for most clusters, much of reduction in migration rates, if not the rates themselves, can be attributed to house price increases. Unemployment decreases have a mostly low, and an often varied and un-reliable effect, but for some clusters, the combined variables act together to influence migration rates.

It should be acknowledged at this point that whilst statistically significant and very interesting, the analysis was only carried out on ten years of data. Examination of residual plots pointed to an absence of hetroscedasticity, but in some cases non-linearity may have been present (see Jones, 1984 or Field, 2005 for an explanation of these issues). The small number of cases, means it is impossible to verify completely satisfactorily that all data meet the distribution assumptions necessary for the results to be treated absolutely reliably. That said, the strength of the relationship at this stage means that it would be very interesting to continue this analysis as more data becomes available in future years - this evidence may pour cold water on the hopeful observations made by Dennett and Stillwell (2008) that provincial regions such as Yorkshire and the Humber are retaining young migrants, with the implication being that it is positive developments in the region which are affecting the change. It may be that these changes are related to nothing other than housing economics.

Linking all of this back to the life course paradigm, it is clear that the employment/leaving home/gaining independence nexus is the key to understanding migration patterns in the 20-24 age group. At this age the frictional effect of distance is felt least by migrants, with the attractiveness and repulsiveness of different clusters varying considerably, with clusters like Coastal and Rural Retirement Migrants exhibiting a positive exponential distance decay parameter for outmigration! The volatility of migrants in this age group and their susceptibility to the external pressures exerted at this stage in the life course mean that conventional spatial interaction-type explanations of migration are less successful at predicting flows - model residuals being larger in this age group than in any other. The particular life course pressures of finding first employment and first home away from parents mean that this age group, more than any other, are susceptible to external economic changes which have a direct impact on the propensity of migrants in this age group to move.

Age 25-29 Age group 25-29 could be viewed as the last of the high-mobility age groups. The almost frenetic movement exhibited at 20-24 drops off considerably in many clusters, but still remains high. Life course influences are less strong in this age group, but still exert some pressure. In terms of gross flow volumes, Figure 7.8 shows that at this age the largest volume of flows at any age group and for any cluster occurs within Dynamic London. This is undoubtedly a function of the particular job, housing market and environmental conditions within Dynamic London. As mentioned earlier, the higher proportion of managerial and professional jobs in London means there are more well-paid and mobile young people within the cluster. High property prices mean that many individuals will be living in rented accommodation, which will also increase their mobility (Courgeau, 1985). The 25-29 age group will be beginning to move up the career and income ladder which will allow some individuals more choice over their residential location - often this will be a centrifugal movement away from the centre of the city (and the geometric centre of the cluster), although as the flow rates in Figure 7.8 show, there are also a large number of centripetal moves towards Dynamic London from the surrounding Footloose, Middle-Class, Commuter Britain cluster.

Unlike at age group 20-24, there does not appear to be much of a change in the rates of in and out migration over the decade across most clusters, perhaps reflecting the more settled employment positions of many in the 25-29 age group. Where variation does occur there are often slight increases in migration rates, the main exception being in out-migration from Student Towns and Cities. This may well reflect the delayed entry into the job market of those who have spent longer in higher education, either as a result of late entry into the system or postgraduate study; indeed in net cluster to cluster terms, six of the top eight flows in this age group are from Student Towns - the most popular still being Dynamic London, all of which may well be reflecting delayed entry into the job market, or of course the lack of life course influences pulling migrants into other areas. From all origins, the highest rates of in-migration across the decade can be seen in the Footloose, Middle-Class, Commuter Britain cluster and the Successful Family In-migrants cluster. These flows could well be linked to the start of the settled career, family rearing phase outlined earlier and described by Plane and Jurjevich (2009), Grundy and Fox (1985), Mulder and Wagner (1993) and Bramley et al. (2006).

Age 30-44 At age 30-44 another system-wide drop in the level of migration can be observed, single year of age migration propensities observed in the 2001 Census suggest a continued and steady decline from age 30 to age 44. At the younger end of this age group where mobility rates are at their highest, the life course influences of partnering, marriage, household formation and the starting of families will be exerting pressure on migration moves. The influence of these family formation life course events have already been discussed in the context of the 0-15 age group, with clusters such as Successful Family In-migrants and Footloose, Middle-Class, Commuter Britain exhibiting consistently higher in-migration rates of this age group than any other cluster, with big net gains from Dynamic London and Student Towns and Cities respectively.

Age 45-59 At age group 45 to 59, the continuation of moves down the urban hierarchy can be observed. Whilst in gross flow terms the flows within and between Dynamic London, Footloose, Middle-Class, Commuter Britain and Moderate Mobility, Non-Household, Mixed Occupations predominate, in net terms the Successful Family In-migrants cluster and particularly the Coastal and Rural Retirement Migrants clusters feature high up in ranked cluster to cluster flows. Indeed net migrant gains to these two clusters account for eight of the top ten net gains in this age group. Flows into these clusters - clusters characterised by mid-life to pre-retirement migrants across the socio-economic spectrum - fits entirely with the retirement from parenthood, withdrawing from the workforce 'retirement transition' documented by Bures (1997). Bures cites the increase in two-career families increasing personal savings (and therefore options) for many pre-retirement families in the U.S.. This coupled with lower fertility rates meaning fewer children to leave home (and support while at home) means that a move to a more desirable location in preparation for retirement becomes more feasible - as Plane and Jurjevich (2009, p.6) put it a *"revivified impetus for movement occurs"*. In the UK, evidence of similar driving forces can be found; the ONS in their Social Trends publication ONS (2009c) document the 1.5% rise in female workforce participation between 1998 and 2008. Other data from the ONS shows a sharp drop in the age specific fertility rates of women from the mid 1970s (given average age at first birth, the age about which now those women would be reaching pre-retirement). This evidence, coupled with the observed flows into the more rural areas typified by the Successful Family In-migrants and Coastal and Rural Retirement Migrants clusters would point to similar life course migration moves occurring in Britain.

Age 60-74 Within the migration classification system, for most clusters across the decade this age group exhibits the lowest in- and out-migration rates of all (Figures 7.12 and 7.13). However, there are some clusters where this is not the case - Coastal and Rural Retirement Migrants and Successful Family In-migrants for in-migration and Footloose, Middle Class, Commuter Britain and Dynamic London for out-migration. Indeed when examining the net-migration exchanges between clusters across the whole decade, the flows from Footloose, Middle Class, Commuter Britain to Coastal and Rural Retirement Migrants and to Successful Family In-migrants are the second and fourth largest respectively. Also taking into consideration the large gross and net flows from Dynamic London down the urban hierarchy into Footloose, Middle Class, Commuter Britain and Moderate Mobility, Non-Household, Mixed Occupations as well, it is clear that many of the influences leading to urban-to-rural and coastal in-migration in the U.S. must also be having an effect in Britain.

Results from the modelled migration flows between clusters confirm that at this stage of the life course there is a noticeable change in the patterns of migration within the Migration Classification system. For Dynamic London there is a reduction in the β distance decay parameter for out-migration (Figure 8.2) after a continuing increase from 20-24. This indicates that at the age of retirement, distance has much less of a frictional effect on migration flows out of the cluster. In parallel with this, the mean migration distance for flows out of the cluster leaps up to a peak only comparable at age group 16-19. Both of these measures indicate that come retirement, many individuals who may have remained in the Dynamic London cluster for many years, are suddenly freed from the constraining influences of employment and then choose leave for destinations much further down the urban hierarchy. Whilst they both exhibit lower β distance decay parameters and longer mean migration distances than London, the two London hinterland clusters of Footloose Middle Class, Commuter Britain and Moderate Mobility, Non-Household, Mixed Occupations display similar patterns in the peaks of mean migration distance and slight reductions in distance decay parameters at this age.

Age 75+ At the final age group, we might expect to find evidence of the documented life course related moves which associated with priorities in the later stages of life - those related to the desire to be nearer to family or the need for dependency related care (Bures, 1997; Pettersson and Malmberg, 2009). What is presented by the data, however, is something which is not as clear cut. Figures 7.9 and 7.12 show that taking the whole system across the decade, the highest rate of in-migration for the 75+ group can be seen in Footloose Middle Class, Commuter Britain - a rate that shows a marked increase from age group 60-74. In the light of the life course research by Bures (1997) and Pettersson and Malmberg (2009) suggests this might be interpreted as a move towards family who may well be the younger migrants so populous in this cluster. On the other hand, when examining the cluster linkages at this age, it can be seen that many of the flows into the cluster are from Dynamic London, as they have been in many other age groups. Could this just be a continuation of the trends from the previous age group - perhaps the result of longer life expectancies and a delay in retirement migration? It is difficult to say. What is clear is that the migrants who will be moving for defensive, care related reasons towards the later portion of the 75+ age group will be very few in number and their destinations varied. As a result, discerning any migration patterns within the system as a consequence of this particular life-course related move will be very difficult if not impossible.

8.8 Concluding remarks

This penultimate chapter has sought to further the analyses of previous chapters and offer some explanations for the patterns that can be observed within the Migration Classification system. The objective of the chapter set out in the introduction was to see whether explanations for the internal migration patterns described in Chapter 7 could be offered through the rather different but complimentary theories of spatial interaction and life course. Firstly it has been shown that a large proportion of the variation within the Migration Classification system across the decade of study *can* be explained by spatial interaction theory. Many of the moves between specific cluster origins and destinations can be seen as the inevitable consequence of consistent numbers of migrants moving in or out of the clusters each year; and of the distances between districts within the clusters; and of the differential decay parameters associated with these distances. That is to say regardless of the individual characteristics of clusters and places within clusters and the migrants making the moves, the effects of social gravity will act to move large numbers of migrants within the system in a relatively predictable way.

The analysis has shown that the basic forces acting within the system do vary by age and by cluster type. For example, the frictional effect that distance has on moves fluctuates considerably by age and by cluster with the force acting upon old migrants moving out of Low Mobility Britain very much greater than that acting upon young migrants moving out of Coastal and Rural Retirement Migrants. These differences help characterise the clusters within the system beyond the characteristics they already have from the variables used in the clustering process - Low Mobility Britain as well as being characterised by below average migration moves can now also be seen as a cluster where distance has a particularly negative effect on in-migration; Coastal and Rural retirement migrants as a cluster where for those in their early twenties migrating out over long distances is the norm.

Models fitted to the data were not only useful where they were able to reproduce reality; in many ways they were even more useful where they did not. Over the decade of analysis and across all age groups the results of the models showed consistent patterns of over and under-prediction. Interpretation of residual values within a closed system must be handled with caution with over predictions in one part of the system causing under predictions in another and vice versa but for some of the larger errors, such as the under-prediction of flows within Dynamic London and between Student Towns and Cities and Dynamic London, the residuals point to external influences over and above the social gravity influences which are causing more migration moves than would be expected to occur - in this case the pervasive interaction of economic and life course influences acting to increase flows within and between these clusters every year in the decade of study.

Where models of spatial interaction reach their explanatory limit, it has been shown that theories of life course can offer a useful perspective on the various age-related influences which act upon migrants within the Migration Classification system. This chapter has shown how different clusters become more strongly associated with life course influences at different points. Importantly the interactions between clusters and life course events has shown how in the case of Dynamic London, its particular position in the life course of many migrants makes it almost immune from the external economic influences of house prices and unemployment. Despite a short time series which casts some doubt on the strenght of the conclusions that can be made, regression analysis has indicated strongly that economic factors such as these can act to suppress the migration activity of an entire age group over time for most areas in Britain.

This chapter has demonstrated, again, the value of the Migration Classification in providing a parsimonious system which allows for the distillation of a huge number of complex flows across the geographic space of Britain over an entire decade. We see a system with a few hubs of migration activity - Dynamic London; Footloose, Middle-Class Commuter Britain; Successful Family In-Migrants; Student Towns and Cities; Coastal and Rural Retirement Migrants - hubs which one migrant may visit each of at a different stage in their life course. And a system which features areas which are almost isolated from this wider system - Low Mobility Britain which interacts very little with anywhere else; Declining Industrial, Working-Class, Local Britain which exhibits a noticeable amount of localised interaction, but very little external interaction; and Moderate Mobility, Non-Household, Mixed Occupations which interacts mainly with Dynamic London and Footloose, Middle-Class Commuter Britain.

Finally it should be reiterated that much of the analysis in this chapter can be see as just beginning the explanation of internal migration patterns in Britain. A number of questions remain about the suitability of unconventional spatial systems for spatial interaction analysis - the appropriate measure of distance, the appropriate model specification, the appropriate calibration statistics, for example. Here, a mathematical model based on Wilson's entropy maximising method was selected as cluster-specific parameters were sought to aid explanation of patterns, but it could be that a model which accounts for the effects of spatial clustering a little more, such as Fotheringham's competing destinations model might offer even better explanations, or a statistical derivation of the spatial interaction model such as those proposed by Willekens or Flowerdew might offer the opportunity of including explanatory variables such as house prices or unemployment, which again could lead to an even more detailed explanation. Furthermore, many of the life course explanations proposed in this explanation could certainly be explored in more detail and over a longer period of time as more annual PRDS migration data are produced, especially where changes to life course influences such as retirement or moving away to study are likely to change as Britain enters a new economic era.

Chapter 9

Thesis discussion and final conclusions

9.1 Introduction

The work within this thesis has successfully addressed the aim set out in the introduction: to advance the current understanding of internal migration in Britain. In completing this research there have been a number of specific achievements: a new national inter-district internal migration dataset has been created; a new internal migration-based geodemographic classification of local authority districts has been produced; alternative migration analysis metrics have been specified; and unconventional applications of spatial interaction models have cast new light on the internal migration landscape of Britain.

This chapter concludes the thesis through summarising the main research findings and achievements, addressing first in Section 9.2 the seven specific objectives laid out in Chapter 1. Section 9.3 will then address some of the limitations of this piece of research through a critique of the methodology employed, before Section 9.4 reflects on these of the successes and limitations of the thesis through suggesting a possible research agenda for the future.

9.2 Summary of research findings

In the introduction to this thesis, a broad aim was set to advance the current understanding of internal migration in Britain. This aim would be met through addressing a series of research objectives; this section will now take each of these objectives in turn and demonstrate how each was met through the research carried out within the thesis.

> **1. To examine and review the current internal migration data landscape of Britain, identifying features in the provision of data which could affect understanding and exploring techniques for improving data where there are deficiencies, resulting in the development of a new partially-estimated national dataset.**

In order to contextualise the data-based analysis of this piece of work, Chapter 2 began with a thorough examination of the current state of internal migration data provision in the UK. The range of datasets available to researchers was presented with it becoming apparent that whilst a number of census, survey and administrative sources containing internal migration data are in existence, geographical resolution, spatial coverage and sample size issues mean that two main data sources are used over and above any other in internal migration research: the decennial Census of Population which contains unrivalled data in many ways, but is limited through being a cross-sectional snapshot presenting a picture of one in every ten years; and data derived from NHS records, which are temporally rich but lack the geographical and attribute detail of the census. Behind these standard characterisations lie more nuanced differences which affect the both the usability of the data and the definition of the phenomenon being studied.

Chapter 2 showed that the way census and NHSCR data record migration events and migrant individuals are quite different, with census data recording a single migrant transition from one address to another, over any distance, over the period of a year, and NHSCR data (including the adjusted patient register data) recording all moves occurring across LAD boundaries: differences which are important in a definitional sense but that have a relatively small effect on the aggregate data collected. Of more importance is the quality of the data. Census data suffer from a trade-off between attribute detail and geographical detail, with a higher resolution in one resulting in a lower resolution in the other. In combination with this, all data are subjected to post-tabulation adjustment by the ONS in order to preserve the confidentiality of individual respondents; the net outcome for researchers being that data at the most detailed geographical resolution are very unreliable for all but the most populous of areas. NHS-based data, on the other hand, are affected by other issues. Coarser geographical resolutions and far fewer attributes mean that data perturbation for confidentiality is not an issue, but NHS-based data have their own drawbacks such as only recording moves between defined geographical areas and not within them, and the persistent and well documented undercount of young males.

Perhaps most importantly though, despite the national statistical agencies of the constituent countries of the UK each producing their own NHS data-based internal migration estimates down to the LAD level; they have not collaborated to produce a UK sub-regional dataset. This is an issue for any researcher wishing to examine national internal migration moves over time at a relatively small scale and led to one of the major substantive contributions of this piece of work. Addressing the issue in the latter half of Chapter 2 a new methodology for estimating a complete LAD level intra-Britain (plus Northern Ireland as a single zone) time-series of internal migration matrices, disaggregated by broad age group was presented. Building on the work of Raymer and others which has shown relative stability in the structures present within internal migration data, and concentrating principally on flows between districts in Scotland, England and Wales, the methodology essentially apportioned inter-regional flows from the NHSCR to LADs using in- and out-flow ratios borrowed from the 2001 Census. Missing flows within Scotland were estimated by taking a data trend from years where there were data available.

These combined techniques resulted in the completion of a ten year time-series running from mid-1998 to mid-2007, disaggregated by 8 broad age groups - data which were important for this research, but which have also been used in other ESRC research projects such as the 'What happens when international migrants settle? Ethnic group population trends and projections for UK local areas' project (RES-165-25-0032) and are now available for general use by any interested parties through the ONS migration statistics unit and CIDER.

2. To review the current methodological techniques and substantive literature surrounding internal migration to form solid foundations upon which to build a more current understanding.

In order to make sense of the internal migration data from sources explored as part of the first objective, an understanding of the quantitative methods which can be employed to make sense of the patterns contained within the data was essential. All internal migration can be conceptualised as occurring within a closed system of origins and destinations, this closed system can be represented mathematically as an n by n matrix of flows. Section 2.3.1 defined an example of such a system before Section 2.3.2 explored some of the ways in which internal migration can be represented; often standardised as rate intensities using a population denominator in order that areas of differing size can be compared - important in a national system like Britain where the historical boundaries which form the basis of many of the current geographical divisions have resulted in numerous areas of varying size. This overview was essential in order that the techniques described and indeed variations on these techniques could be applied successfully and developed in subsequent chapters.

Aside from the various observational techniques which can be applied to internal migration data, a number of explanatory modelling methods can be used to develop understanding still further. Section 2.3.3 offered a short introduction to the theory of spatial interaction; a theory derived from Newtonian gravitational principles which states that the level of interaction between two bodies is a function of their size and the distance between them. Brief exemplification of the idea using the example system from Section 2.3.1 was given, although a far more thorough review was presented at the beginning of Chapter 8. The history of spatial interaction modelling techniques was presented explaining how, in particular, developments following on from the work of Wilson (1970, 1971) have led to the successful specification of models which have been used to explain and predict interaction flows in a variety of contexts. The review of the mathematical modelling techniques in this chapter showed that it was feasible to apply similar models to data within the derived Migration Classification system in order to explore some of the features of migration within the system; for example, the effect of distance on migration flows and the influence of other factors where the models fall short of a good explanation.

In contrast to the explanatory mathematical spatial interaction modelling techniques which followed Wilson, the methodological techniques employed by Raymer and colleagues make use of the demographic and spatial structures inherent in internal migration data. The general approach was reviewed in Section 2.4 and was of particular importance to the estimation of age-specific flows later in Chapter 2 where knowledge that structures such as the relationship between migration and age remain stable over time was of crucial importance in producing a set of new internal migration estimates.

Whilst it was of great importance to review the methodological techniques associated with internal migration analysis, it was equally important to contextualise the work in subsequent chapters of the thesis through a review of the substantive literature. A number of important themes emerged from a short review focusing on patterns of internal migration in selected western democracies and the last fifty years or so in the UK. The principal one was of a continuing recent experience of aggregate counterurbanisation flows, with a common counter flow of urbanising young migrants. Work on migration in the UK emphasised the importance of London and its South East hinterland in the migration profile of the country, with moves from the north to the south of the country also featuring in the recent internal migration history. One of the main findings of the short review was that where internal migration is often a complex process, much of the work tended towards binary simplifications when summarising flows, with some notable exceptions. In these exceptions, work by Rees et al. (1996), Raymer et al. (2007) and Raymer and Giulietti (2009) applies area classification frameworks to aid simplification whilst retaining a little more origin and destination detail. The methodological and substantive reviews provided in Chapters 2, 3 and 8 served to highlight where there were gaps in the current knowledge of internal migration in Britain and provided the methodological background necessary to facilitate tackling problem; both important precursors to the third thesis objective.

3. To explore the patterns of internal migration in Britain at the start of the 21^{st} century using data from the 2001 Census.

Following the reviews in the first two chapters of the thesis, a clear gap in the current understanding of internal migration in Britain emerged, with a relative dearth of work on the patterns emerging from the data collected in the 2001 Census. In addressing this third objective through carrying out a detailed analysis of data from the 2001 Census, a substantial gap in the current knowledge of internal migration at the beginning of the 21^{st} century has been filled and can be seen as the second major achievement of the thesis. A large part of this objective was achieved though the work featured in Chapter 4. In this chapter, data from the 2001 SMS at the district level were examined in detail from two overlapping perspectives: a spatial perspective which analysed the flows through the lens of a general purpose district level area typology; and a life course perspective through analysing the age-specific patterns of internal migration.

A number of important findings arose from this analysis. Firstly, it was shown that, to a certain extent, the historically observed patterns of migration reported in Chapter 3 remained constant; with the central role played by London, the net loss of migrants from urban areas and the net gain to rural areas. These broad generalisations, however, could be deconstructed though the use of a classification typology, with districts classified as ‘Young Vibrant Cities’ actually net gainers of significant numbers of migrants and some ‘rural’ areas on the fringe of London gaining population and actually exhibiting populations with rather more urban characteristics. Other examples of flows counter to the metropolitan to non-metropolitan average were present, with rural ‘Averageville’ areas exhibiting noticeable net internal migration loss.

The importance of accounting for the influence of age on the volume and direction of migration was also reinforced through this analysis of 2001 Census data. Many of the aggregate patterns presented, even where broken down spatially into area type Classes - the smallest unit in the Vickers et al. classification hierarchy - varied hugely when the age of the migrants was taken into consideration. It was shown in Chapter 4 that age and area type interact so that, broadly speaking, migrants tend to move down the urban hierarchy for all age groups except the 16-29 age group: in this age group migrants mainly moving in the opposite direction, up the urban hierarchy into larger urban areas - particularly London.

Whilst Chapter 4 contributed significantly to achieving objective 3, the patterns of internal migration from the 2001 Census were explored in a novel way through the classification-building process detailed in Chapter 6. Starting with a suite of some 88 migration-related variables taken from across the various domains in the 2001 Census, an eventual list of variables half the size of the original was used the build the ‘Migration Classification’. Whittling down the variables though processes such as PCA meant that only those variables important to the overall internal migration landscape at the district scale were preserved. This revealed that, for example, the patterns of migrants with limiting long-term illness or from no usual address were of little importance in the British internal migration story when compared to those of different age, socio-economic status, ethnicity and housing tenure. In the definition of 8 distinct cluster profiles, individual districts become associated with particular patterns of migration; for example, the Coastal and Rural Retirement Migrants areas, tend to contain more in-migrants in the older age groups, from across the socio-economic spectrum; or the Low-Mobility Britain areas show little internal migration interaction with other areas at all. The common patterns of migration and types of migrant experienced by some areas made explicit by the Migration Classification typology serve to offer new and explicitly area-linked perspective on the patterns of migration at the start of the 21^{st} century.

4. To develop a new area classification based on internal migration data to both support analysis of census-based internal migration data and use as a framework for analysis of non-census-based data.

The production of a new area typology based on internal migration data can be seen as the third main achievement of the thesis. Chapter 5 made the philosophical and practical case for building a new migration data-based classification. It was argued that the construction of such a classification satisfies a natural instinct towards parsimony, but far more than this allows for the distillation of a rather complex brew to something more manageable through both the clustering process and the variable selection itself. One of the big decisions which needed to be made was in relation to whether the classification would be concerned with classifying migrant flows or migrant based events. The decision was taken to focus on migrant individuals rather than flows as flow classification will usually lead down the route of functional region creation; functional regions have the property of being discrete zone entities without any association with other zones - the association between non-contiguous zones was one of the central points of interest for this piece of work. Flow information can be included in standard geodemographic classifications in the form of distance variables, but as mentioned in Chapter 6, early experiments with these data produced cluster solutions with a heavy bias towards London and so were not pursued any further. Certainly though, these data could be explored in far more detail in the future, especially in the context of the varied distance profiles produced by the spatial interaction models explored in Chapter 8. Aside from the flow data issue, the other main consideration in the classification-building process concerned the scale of analysis for the typology. For a number of reasons relating to the availability of variables, accuracy of data and usability of the typology LADs were chosen as the spatial unit of analysis.

Chapter 6 details the process of building a migration data-based geodemographic area classification. Another notable contribution of this thesis was that important incremental advances in geodemographic classification building were achieved. For example, the ONS OAC was developed using the SPSS software package, but it was demonstrated in Section 6.4 that there were significant drawbacks in using the clustering algorithm in this particular programme. The chapter goes on to detail an alternative software and methodological approach which helped produce a more reliable final solution.

In Section 6.6, a final geodemographic Migration Classification is presented, assigning each LAD in Britain to one of 8 distinct cluster types, each typified by different migration characteristics - high rates of old in-migration; net out-migration of those in higher socio-economic groups; net in-migration into owner occupied accommodation, for example. As highlighted under the last objective, one of the contributions of this new classification was in offering a new analysis methodology for 2001 Census migration data. The other, and perhaps more important, contribution was in producing a new framework which can be used in the analysis of alternative migration data. This framework makes a more logical partition of areas for migration flows, thus avoiding the problems that might be encountered when, for example, districts which are heavily associated with certain types of migration flows (such as student inflows) are allocated to very different cluster types in general purpose classifications, confounding the interpretation of patterns. In addition to this, the new classification adds value to less attribute rich internal migration data analysed using it. Both of these benefits are demonstrated in Chapter 7, with a brief comparison of the patterns observed for clusters featuring similar districts in two different classifications. It is shown that despite the clusters featuring similar areas, the allocation of just some districts to different clusters leads to very different overall patterns being observed - the particular distribution of districts within clusters in the general purpose classification confounds the analysis of internal migration patterns.

5. To build on existing methods and develop new techniques for understanding internal migration data.

Throughout most chapters in this thesis, existing techniques for understanding internal migration data have been employed and developed. In some cases, these techniques were taken and applied directly, such as the in the calculation of migration intensities, but in many cases existing ideas have been built upon for specific purposes. For example, the estimation techniques employed in Chapter 2, can be viewed as a development of similar techniques which have been used elsewhere to produce new data and enhance existing data. The precise implementation, however, in this instance is novel. Similarly, the use of turnover and churn measures in Chapter 4 whilst offering a new perspective on internal migration, were a straightforward development of existing migration intensity calculations.

As has been discussed already, the use of cluster analysis techniques in the study of internal migration data was an example of a new application of an existing technique. Certainly, no attempt has been made before to employ cluster analysis in the examination of internal migration data in Britain before. This technique was particularly successful in linking areas to the characteristics of migrants.

Chapter 7 took the development of new techniques a step further with the introduction of two novel analysis metrics. Standardised Migration Distance Ratios (SMDRs) and Standardised Migration Ratios (SMIRs) were born out of similar techniques used in demography such as Standardised Mortality Ratios. The new ratios were designed to allow for the distorting effects of distance and age respectively on the migration flows to, from and within classification clusters, and enabled an evaluation of the level of these flows compared to the national average. The ratios showed, for example, that Dynamic London has considerably more migrant in and out flows than would be expected even when distance is accounted for, and conversely, Declining Industrial, Working Class, Local Britain, has far fewer in and out flows than would be expected. Similarly, when age is taken into consideration, the Student Towns and Cities Cluster experiences around the average national in and out migration flows in contrast to the inflated gross flows that are seen when the large numbers of young student migrants are included in the analysis.

The work reported in Chapter 8 takes forward spatial interaction modelling techniques though successfully applying a method normally only used in discrete zone systems, to clusters of zones. It is shown that a doubly constrained model fitted to data aggregated into Migration Classification zones can be used to assess the frictional effect of distance on moves into and out of districts within Britain and to give clues, where the models produce poor estimates, as to the types of cluster where other external influences may be having a greater or lesser impact on the flows which occur. This is certainly a useful development of the application of spatial interaction models.

6. To examine recent trends in internal migration in Britain over time using

new partially-estimated data.

Work carried out in the early part of the thesis was brought together with work featured in Chapters 7 and 8. This sixth objective was tackled principally though the research which appeared in Chapter 7 and constitutes an important empirical contribution of this thesis. Here a detailed descriptive analysis of recent trends in internal migration in Britain was presented using the partially-estimated dataset produced in Chapter 2, resulting in a number of important findings. Firstly, it was found that over the decade of study between mid-1999 and mid-2008, the rates of internal migration between Migration Classification clusters remain remarkably consistent, with perhaps a slight decline in the overall level of migration over the decade. Each cluster in the classification exhibits a very distinct profile over the period with, for example, Dynamic London consistently experiencing far higher in- and out- migration levels than all other clusters. It was also discovered that some clusters are more linked through their migration exchanges than others, but that the associations between clusters are maintained over time. For example, there is a very strong link between Dynamic London and Footloose, Middle-Class, Commuter Britain, with significant flows from the former to the latter. Conversely the links between Declining Industrial, Working Class, Local Britain and the clusters with the highest levels of internal migration activity were almost non-existent, showing a persistent detachment from the national migration system.

Analysis in Chapter 7 showed noticeable variation in the propensity to migrate by age for different Migration Classification clusters, as would be expected given the analysis earlier on in the thesis. For example, the Student Towns and Cities cluster features an in-migration peak in the 16-19 age group, an age group earlier than the peak for all other clusters at 20-24, and Dynamic London features an out-migration peak an age group later at 25-29. Controlling for the effect of age, it was shown that clusters more associated with in-migration moves - Coastal and Rural Retirement Migrants and Successful Family In-migrants - showed a noticeable convergence towards the national average from a position of well above the average over the decade of study; a pattern which reflects changes in the external influences affecting the increased in-migration to these clusters. Other clusters such as Student Towns and Cities, Low Mobility Britain and Declining Industrial, Working-Class, Local Britain show little variation at all in their in- and out-migration ratios compared to the national average across the decade, perhaps reflecting the stability of factors which influence migrants moving in and out of districts in these clusters. Over the decade, rates of in- and out-migration by age show very little variation for most age groups and most clusters, with the noticeable exception of a clear decline in the migration propensities of 20-24 year olds across the decade of study - a particularly intriguing pattern amongst a host of patterns which called for additional analysis and explanation.

7. To offer explanations for current internal migration patterns in Britain through the use of mathematical spatial interaction models and life-course theory.

The final objective of the thesis was to consolidate the analysis of previous chapters through offering some explanations for the patterns presented. Chapter 8 addressed this objective through a combined spatial interaction modelling and life-course approach. Fitting doubly constrained spatial interaction models to internal migration data aggregated to the Migration Classification cluster framework, it was shown that many of the flows between combinations of clusters within the system could be viewed as the inevitable consequence of numbers of migrants moving in and out of the clusters each year, and each cluster exhibiting different levels of 'social gravity'. The variation in gravitational pull and average physical distance between districts in different clusters varies and interacts differently with migrants of different ages, meaning that for most clusters the frictional effect of distance and the distances over which migrants travel to arrive and leave is very different.

It was shown in Chapter 8 that the failure of models to reproduce reality often provided more insights into the reasons for particular migration flows than where they were successful. For example, the models continually under-predicted the flows within Dynamic London and between Dynamic London and Student Towns and Cities, suggesting that where this is happening, influencing factors over and above those of social gravity accounted for by the models are acting to affect migration.

Where models were unable to give a complete explanation for the flows experienced within the Migration Classification system over the decade, theories of life-course influences on the propensity to migrate were turned to. A compelling case for the interaction between life-course and cluster types was presented. In doing so, explanation for the decline in the migration propensities of 20-24 year olds was offered, with the influence of unemployment rates and more importantly house prices, shown to manipulate the internal migration patterns of this age group for all clusters except Dynamic London.

9.3 A critique of the methodology

This thesis has advanced the both understanding of internal migration in Britain and the processes involved in achieving this understanding, but in spite of these successes there are some inevitable issues with the methodological approach which, whilst noted in various places within the thesis, should be acknowledged explicitly here.

The process of building a classification is fraught with problems. No matter how carefully decisions were made at each stage, it is often the case that an argument could be presented for an alternative route to be taken. It is a certainty that whilst the decisions that were made in relation to the number of variables, clustering algorithm, distance measure, etc. were all made for the best reasons given the available information at the time, were someone else to tackle the same problem from the beginning, it is unlikely that the exact same final solution would be achieved. It could be that some decisions were taken too lightly, for example the omission of distance data happened after initial experiments indicated it affected cluster solutions unduly;

it could be argued that this decision was taken too early on in the clustering process and a far more detailed analysis of the effect of distance-based variables on the cluster solutions could have been carried out. If this had happened, it could be that a different conclusion would have been reached and distance data used, leading potentially to a different final cluster solution.

As has been pointed out, classification building is an exercise in parsimony and generalisation. Wherever generalisations occur, error is inevitably introduced with there always being exceptions to the general rule. This is especially true where the spatial unit of analysis is relatively large. The MAUP and the ecological fallacy have already been discussed, but it is worth reiterating that the LAD unit of analysis means that spatial generalisations are being made in conjunction with data attribute generalisations (age group, ethnic group, etc.), thus meaning it would be very unwise to attempt to apply these aggregate observations to alternative spatial and data resolutions.

Another potential problem is with one of the stated benefits of the Migration Classification: adding value to attribute poor data. The classification was constructed using data from the 2001 Census - data which are a snapshot of the population at this time. The further away from this point in time, the less likely the particular situation recorded by the Census remains applicable. It was shown in Chapter 7 that the migration patterns over time in Britain are remarkable stable, however, there is some inevitable variation, and therefore it could be argued that additional migration attributes such as socio-economic status which characterise districts in the migration classification and were used at the end of Chapter 7 to make judgements about the migrants moving between clusters, must be treated with caution, especially the further one gets from 2001.

Another common criticism for any classification surrounds the naming of the clusters. Easily digestible by users, the names of the clusters are more readily recalled by users than any of the data behind their definition and so accurate descriptive names are vital. Care was taken to apply names to the clusters which provided a true representation of both the variables characterising them and of the locations represented. Despite a quality assurance exercise carried out to test the names of the clusters, a relatively small number of interested parties were consulted with many agreeing with the original names given. Even if a much larger quality assurance exercise were carried out, the names given would still largely the subjective interpretation of the author and therefore would always be subject to disagreement. The names of some clusters changed after the consultation process, for example ‘Low Mobility Britain’ was changed from ‘Sedentary Middle-Class Britain’ as it was decided the middle-class element was not a strong enough feature of the cluster profile to warrant inclusion. Had the original name remained, very different emphasis would have been put on the cluster. Further consultation could have led to more name changes, affecting user interpretation still further.

Other criticisms can be made in relation to other methodological elements of the thesis. For example, the data estimation process described in Chapter 2 can almost certainly be improved upon. An IPF methodology was rejected at the time for a number of practical reasons. It is probable, however, that this technique would improve the accuracy of the estimates. In addition, alternative datasets (unrounded NHSCR data or decimal unrounded PRDS data, for example) constituting better model inputs could be utilised to enhance the model outputs.

A number of improvements could be made to the spatial interaction modelling methodology employed in Chapter 8. Problems with model convergence in the Microsoft Excel medium the model was programmed in meant that the models were calibrated using the average distance goodness of fit measure. Whilst this measure provided better fits when used in conjunction with the origin/destination specific β values, experiments with the generalised β parameters suggest that were a method developed to calibrate origin and destination specific β values using either R^2 or SRMSE, better model fits could probably be achieved. Similarly, experimentation indicated that were similar convergence problems overcome, perhaps through the implementation of the models in a different medium, negative power distance decay functions could potentially offer better solutions as well.

Further criticisms could be levelled at the modelling methodology where models which have traditionally been applied to discrete zone systems have been applied here to a system of clustered origins and destinations. No other research has been carried out which documents the fitting of mathematical spatial interaction models to non-standard spatial systems, so consequently it is difficult to assess all of the drawbacks of this approach. It could be that some of the larger residuals (where the model predicted poorly) are the result of the average distance between districts in each cluster being used to assess the cost of travel between clusters, and that other more appropriate measures could be used. Furthermore, the work of Fotheringham (Fotheringham, 1983a,b, 1984, 1986a,b; Fotheringham et al., 2001) on competing destinations, which has shown spatial interaction to be affected by the clustering of origins and destinations, could be used as the basis for the criticism of any models which do not take full account of these patterns.

9.4 Recommendations for future work

The research reported in this thesis has the potential to be taken forward in a number of directions. Firstly, there is much more work which can be carried out in the area of migration-based area classifications. The broad methodology outlined here can certainly be applied to new data which will emerge from the 2011 Census to assess both changes in the internal migration landscape of Britain, and to provide an updated framework for analysis. Given the likely demise of the census post-2011, there is also potential for similar work to be carried out using new and emerging datasets, of which more and more use will need to be made. Lifestyle survey data such as that collected by Acxiom (http://www.acxiom.co.uk/) upon which research is currently underway at the University of Leeds, and which includes origin/destination flow data in an annual sample of around one million respondents and has the potential to extend census-based migration classifications through the use of a number of lifestyle variables such as income which may have a direct influence on an individual's propensity to migrate. A classification built using such data would also have the potential to offer a much higher spatial resolution than the census.

There is a continuing lack of cooperation between the national statistical agencies concerning the creation of a joined-up national internal migration database based upon the patient register data. This is in contrast to the desire of users to access such data, as has been evidenced by the use of the data created in Chapter 2 in other academic projects, and by the interest of the ONS in supplying these data in cooperation with CIDER. The estimates described in Chapter 2 were not without issue, however. Despite only around 8% of the flows in the resulting national datasets the result of the estimation process, meaning that even where errors were introduced in this process, their effect of error on the overall patterns being examined would be small, the technique described in Chapter 2 to estimate these new datasets could almost certainly be improved upon. For example, an IPF method may have produced improved results but was not fully explored in this piece of research following earlier work by Dennett and Rees (2010) showing little improvement in similar estimates at the NUTS2 level and due to difficulties which would have been caused by zero flows in some matrix cells. In addition, the results of the estimation could certainly be further improved with the inclusion of move detailed data sources in the model inputs. For example, the inter-regional NHSCR data which were adjusted to produce the estimates were rounded to the nearest 100 people. Unrounded data do exist, but are not publicly available. A detailed exploration of IPF or alternative maximum likelihood methods and the inclusion of alternative data would certainly be a useful avenue of future research.

Of course the other methodological direction that future research could go in relates to the spatial interaction modelling issues. Certainly, the evidence is that competing destinations models produce better representations of reality where there is only one constraint and where the system is comprised of discrete zones. Exploration of whether this is also the case for cluster-based spatial systems could shed further light onto the factors affecting migration flows between clusters in the classification system. As well as exploring the effects of agglomeration and competition in the system, the less chance there is of model residuals being the result of model error, the more accurately one is able interpret the residuals with other evidence.

Finally, this work made some interesting time-series observations in relation to the factors influencing the propensity to migrate of 20-24 year olds. There was relatively strong evidence of an association with unemployment and especially house prices. The time series over which these observations were made was only ten years, however. With more years of patient-register data becoming available and dramatic recent changes in the rate at which house prices are rising during the global financial crisis, greater insight into the reasons behind these migration patterns could be gained. Certainly, there would also be scope for departing from the mathematical spatial interaction models used in this work and incorporating additional predictor variables such as house prices into statistical regression-based derivations of spatial interaction models to assess the effects of each independent variable on flows within the Migration Classification system. If this type of analysis could be incorporated into future empirical analysis of internal migration flows, our understanding could be advanced still further.

9.5 Concluding remark

This thesis has explored internal migration in Britain at the start of the 21^{st} century and has developed our understanding of the patterns and processes in existence at the local level at a time when there is a continuing need to understand population change in this country. There is of course research still to be continued, methods honed, techniques to be improved and every year new patterns to be observed, but this work has succeeded in unravelling some of the complexities inherent in the internal migration landscape of Britain at this time. It is hoped that through the novel approaches described in this work, those who will inevitably want to make similar sense of future patterns of internal migration, both in Britain and further afield, will now have new tools to assist them in achieving those ends.

Bibliography

Abel, G. J. (2010). ‘Estimation of international migration flow tables in Europe’. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*. 10.1111/j.1467-985X.2009.00636.x.

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

Aldenderfer, M. and Blashfield, R. (1984). *Cluster analysis*. Sage Publications, Inc., London.

Alonso, W. (1978). ‘A theory of movements’. In Hansen, N. (ed.), *Human Settlement Systems: International Perspectives on Structure, Change and Public Policy*, pages 197–212. Ballinger, Cambridge, Mass.

Antolin, P. and Bover, O. (1997). ‘Regional migration in Spain: the effect of personal characteristics and of unemployment, wage and house price differentials using pooled cross-sections.’. *Oxford Bulletin of Economics and Statistics*, 59:215–235.

Baccaini, B. (2007). ‘Inter-regional migration flows in France over the last fifty years’. *Population-E*, 62(1):139–156.

Bailey, A. (2009). ‘Population geography: lifecourse matters’. *Progress in Human Geography*, 33(3):407–418.

Bailey, N. and Livingston, M. (2007). ‘Population turnover and area deprivation’. Report, Joseph Rowntree Foundation.

Bailey, N. and Livingston, M. (2008). ‘Selective Migration and Neighbourhood Deprivation: Evidence from 2001 Census Migration Data for England and Scotland’. *Urban Studies*, 45(4):943–961.

Bailey, S., Charlton, J., Dollamore, G., and Fitzpartick, J. (2000). ‘Families, groups and clusters of local and health authorities: revised for authorities in 1999’. *Population Trends*, 99:37–52.

Bales, K. (1999). ‘Popular reactions to sociological research: the case of Charles Booth’. *Sociology*, 33(1):153–168.

Barcus, H. (2004). 'Urban-Rural Migration in the USA: An Analysis of Residential Satisfaction'. *Regional Studies*, 38(6):643 – 657.

Barker, R., Ward, G., and Moore, I. (1998). '1996 census findings on interstate migration to Queensland focusing on South-East Queensland'. *People and Place*, 6(4):15–24.

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

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

Batey, P. and Brown, P. (1995). 'From human ecology to customer targeting: the evolution of geodemographics'. In Longley, P. and Clarke, G. (eds.), *GIS for business and service planning*. GeoInformation International, Glasgow.

Batty, M. and Mackie, S. (1972). 'The calibration of gravity, entropy, and related models of spatial interaction'. *Environment and Planning*, 4(2):205–233.

Baxter, M. and Ewing, G. (1981). 'Models of recreational trip distribution'. *Regional Studies*, 15(5):327 – 344.

Bell, M. (2002). 'Comparing population mobility in Australia and New Zealand'. *Journal of Population Research*, September(Special Issue):169–193.

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*, 165:435–464.

Berrington, A., Stone, J., and Falkingham, J. (2009). 'The changing living arrangements of young adults in the UK'. *Population Trends*, 138(Winter 2009):27–37.

Bheim, R. and Taylor, M. (2002). 'Tied Down Or Room To Move? Investigating the Relationships between Housing Tenure, Employment Status and Residential Mobility in Britain'. *Scottish Journal of Political Economy*, 49(4):369–392.

Boden, P. and Rees, P. (2009). 'International migration: the estimation of immigration to local areas in England using administrative data sources'. *Paper submitted for consideration by the Journal of the Royal Statistical Society*.

Boden, P. and Stillwell, J. (2006). 'New migrant labour in Yorkshire and the Humber'. *Yorkshire and Humber Regional Review*, 16(3):18–20.

Boden, P., Stillwell, J., and Rees, P. (1992). 'How good are the NHSCR data?'. In Stillwell, J., Rees, P., and Boden, P. (eds.), *Migration processes and patterns. Volume 2. Population redistribution in the United Kingdom.*, pages Chapter 2, 13–27. Belhaven, London.

Bohara, A. and Krieg, R. (1998). ‘A simultaneous multinomial logit model of indirect internal migration and earnings’. *Journal of Regional Analysis and Policy*, 28(1):60–72.

Boyle, P. (1998). ‘Migration and housing tenure in South East England’. *Environment and Planning A*, 30:855–866.

Boyle, P. (2009). ‘Migration’. In Kitchin, R. and Thrift, N. (eds.), *International Encyclopedia of Human Geography*, pages 96–107. Elsevier, Oxford. doi: DOI: 10.1016/B978-008044910-4.00821-X.

Boyle, P., Cooke, T., Halfacree, K., and Smith, D. (1999). ‘Intregrating GB and US census microdata for studying the impact of family migration on partnered women’s labour market status’. *International Journal of Population Geography*, 5:157–178.

Bramley, G., Champion, T., and Fisher, T. (2006). ‘Exploring the household impacts of migration in Britain using panel survey data.’. *Regional Studies*, 40(8):907–926.

Brown, L. and Holmes, J. (1971). ‘The delimitation of functional regions, nodal regions and hierarchies by functional distance approaches’. *Journal of Regional Science*, 11:57–72.

Bures, R. M. (1997). ‘Migration and the life course: is there a retirement transition?’. *Int J Popul Geogr*, 3(2):109–19. 12321163.

Burgess, E. (1925). ‘The growth of the city: an introduction to a research project’. In LeGates, R. and Stout, F. (eds.), *The city reader*, pages 156–163. Routledge, London.

Cadwallader, M. (1992). *Migration and residential mobility: macro and micro approaches*. The University of Wisconsin Press, Madison.

Cameron, G., Muellbauer, J., and Murphy, A. (2005). ‘Migration within England and Wales and the housing market’.

Carstensen, B. and Keiding, N. (2005). ‘Age-period-cohort models: statistical inference in the Lexis diagram’. Report, URL: http://staff.pubhealth.ku.dk/~bxc/APC/notes.pdf
Institute of Public Health.

Castro, L. and Rogers, A. (1981). ‘Status-specific age patterns of migration: family status’. Report, International Institute for Applied Systems Analysis.

CEC (2008). ‘Regions 2020 - An Assessment of Future Challenges for EU Regions.’. Report, URL: http://ec.europa.eu/regional_policy/index_en.htm
Commission of the European Communities.

Champion, A. (1989). *Counterurbanisation: the changing pace and nature of population deconcentration*. Edward Arnold, London.

Champion, A. (1994). 'Population change and migration in Britain since 1981: evidence for continuing deconcentration'. *Environment and Planning A*, 26(10):15011520.

Champion, A. (2005). 'Population movement within the UK'. In Chappell, R. (ed.), *Focus on people and migration*, Focus On, pages 92–114. Palgrave Macmillan, Basingstoke.

Champion, A., Bramley, G., Fotheringham, A., 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*, pages Chapter 15, 269–290. Springer, Berlin.

Champion, A. and Congdon, P. (1992). 'Migration trends for the South: the emergence of a Greater South East?'. In Stillwell, J., Rees, P., and Boden, P. (eds.), *Migration Processes and Patterns, Volume 2 - Population Redistribution in the United Kingdom*, pages 178–201. Belhaven Press, London.

Champion, A., Coombes, M., Raybould, S., and Wymer, C. (2007). 'Migration and socio-economic change: a 2001 census analysis of Britain's larger cities.'. Report, Joseph Rowntree Foundation.

Champion, A., 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.'.

Champion, A. and Vandermotten, C. (1997). 'Migration, counterurbanization and regional restructuring in Europe'. In Blotevogel, H. and Fielding, A. (eds.), *People, jobs and mobility in the new Europe*, pages 69–90. John Wiley and Sons, London.

Chappell, R., Vickers, L., and Evans, H. (2000). 'The use of patient registers to estimate migration'. *Population Trends*, 101:19–24.

Charlton, J. and Chappell, R. (1999). 'Uncertainty estimates for national demographic estimates'. Report, One Number Census Steering Committee Report - Office for National Statistics.

Christophersen, O. (1997). 'Population review of 1996: England and Wales'. *Population Trends*, 90.

Clark, D. E. and Hunter, W. J. (1992). 'The impact of economic opportunity, amenities and fiscal factors on age-specific migration rates'. *Journal of Regional Science*, 32(3):349–365. 10.1111/j.1467-9787.1992.tb00191.x.

Clark, W. and Huang, Y. (2004). 'Linking Migration and Mobility: Individual and Contextual Effects in Housing Markets in the UK'. *Regional Studies*, 38(6):617 – 628.

Clements, J. and Whitworth, A. (2008). 'Understanding and measuring uncertainty associated with the mid-year population estimates'. Report, URL: http://tinyurl.com/clements-and-whitworth-ONS

Office for National Statistics.

Cohen, J. (1960). ‘A coefficient of agreement for nominal scales’. *Educational and psychological measurement*, 20(1):37–46.

Cooke, T. and Bailey, A. (1999). ‘The effect of family migration, migration history, and self-selection on married women’s labour market achievement’. In Boyle, P. and Halfacree, K. (eds.), *Migration and Gender in the Developed World*, pages 102–113. Routledge, London.

Cooke, T. J. (2008). ‘Migration in a family way’. *Population Space and Place*, 14(4):255–265.

Coombes, M. (2000). ‘Defining locality boundaries with synthetic data’. *Environment and Planning A*, 32:1499–1518.

Coombes, M. (2002). ‘Travel to work areas and the 2001 census’. Report, Centre for Urban and Regional Development Studies, University of Newcastle.

Coombes, M., Green, A., and Openshaw, S. (1986). ‘An efficient algorithm to generate official statistical reporting areas: the case of the 1984 travel-to-work areas revision in Britain’. *Journal of the Operational Research Society*, 37(10):943–953.

Coombes, M., Raybould, S., and Wymer, C. (2004). ‘Analysis of census migration data to assist in defining housing market areas for Tyne and Wear’. Report, Centre for Urban and Regional Development Studies, University of Newcastle Upon Tyne.

Cordey-Hayes, M. and Gleave, D. (1975). ‘Dynamic models of the interaction between migration and the differential growth of cities.’. In Baxter, R., Echenique, M., and Oivers, J. (eds.), *Urban development models*. LUBFS Conference Proceedings No.3, The Construction Press Ltd.

Coulombe, S. (2006). ‘Internal migration, asymmetric shocks, and interprovincial economic adjustments in Canada’. *International Regional Science Review*, 29(2):199–223.

Courgeau (1985). ‘Interaction between spatial mobility, family and career life-cycle: A French survey’. *European Sociological Review*, 1(2):139–162.

Courgeau, D. (1973). ‘Migrants et migrations’. *Population*, 28:95–129.

Cross, D. (1990). *Counterurbanization in England and Wales*. Gower Publishing Ltd, Aldershot.

Crowson, R. (2006). *Classification and biology*. Aldine Transaction, New Brunswick.

Darwin, C. (1859). *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life*. John Murray, London.

De Jong, G. F. and Graefe, D. R. (2008). 'Family life course transitions and the economic consequences of internal migration'. *Population Space and Place*, 14(4):267–282.

Debenham, J. (2003). *Extending geodemographics : new small area classifications for Yorkshire and the Humber*. PhD thesis.

DEFRA (2009). 'Defra Classification of Local Authority Districts and Unitary Authorities in England - An Introductory Guide'. Report, Department for Environment Food and Rural Affairs.

Dennett, A., Duke-Williams, O., and Stillwell, J. (2007). 'Interaction data sets in the UK: an audit'. Report, University of Leeds.

Dennett, A. and Rees, P. (2010). 'Estimates of internal migration flows for the UK, 2000-2007'. *Population Trends*, 140(1):82–105.

Dennett, A. and Stillwell, J. (2008). 'Yorkshire and Humber's internal migration exchanges'. *Yorkshire and Humber Regional Review*, 18(3):17–19.

Dennett, A. and Stillwell, J. (2009). 'Internal migration in Britain, 2000-01, examined through an area classification framework'. *Population Space and Place*, (DOI: 10.1002/psp.554).

Dennett, A. and Stillwell, J. (2010). 'Internal Migration Patterns by Age and Sex at the Start of the 21st Century'. In Stillwell, J., Duke-Williams, O., and Dennett, A. (eds.), *Technologies for Migration and Commuting Analysis: Spatial Interaction Data Applications*. IGI Global.

Devis, T. and Mills, I. (1986). 'A comparison of migration data from the National Health Service Central Register and the 1981 census'. Report 35, OPCS.

Diamond, I., Cruddas, M., and Woolford, J. (2002). 'A one number census'. In Rees, P., Martin, D., and Williamson, P. (eds.), *The Census Data System*. John Wiley and Sons Ltd., Chichester.

Ding, C. and Xiaofeng, H. (2004). 'K-means Clustering via Principal Component Analysis'. Report, URL: http://ranger.uta.edu/~chqding/papers/KmeansPCA1.pdf
Lawrence Berkeley National Laboratory.

Dixon, S. (2003). 'Migration within Britain for job reasons'. *Labour Market Trends*, April:191–201.

Donzeau, N. and Shon, J. (2009). 'Residential mobility trends in France, 1973-2006: new estimates'. *Population*, 64(4):779–795.

Dorling, D. (2010). *Injustice: why social inequality persists*. Policy Press, Bristol.

Drysdale, R. (1991). 'Aged Migration to Coastal and Inland Centres in NSW'. *Australian Geographical Studies*, 29(2):268–284. 10.1111/j.1467-8470.1991.tb00720.x.

Duke-Williams, O. (2009a). 'The geographies of student migration in the UK'. *Environment and Planning A*, 41(8):1826–1848.

Duke-Williams, O. (2009b). 'Mapping the geodemographic classifications of migrants' origins and destinations'. *Journal of Maps*, (Paper Submitted).

Duke-Williams, O. and Blake, M. (2003). 'Database fusion for the comparative study of migration data'. Report, URL: http://www.geocomputation.org/1999/068/gc_068.htm
gisca.

Duke-Williams, O. and Stillwell, J. (2007). 'Investigating the potential effects of small cell adjustment on interaction data from the 2001 Census'. *Environment and Planning A*, 39:1079–1100.

Duncombe, W., Robbins, M., and Wolf, D. (2003). 'Place characteristics and residential location choice among the retirement-age population'. *Journal of Gerontology: Social Sciences*, 58B(4):S244–S252.

Ekstrom, M. and Danermark, B. (1993). 'Migration patterns and migration motives among the elderly - Swedish data in a comparative perspective'. *Scandinavian Housing and Planning Research*, 10(2):75 – 89.

Eliasson, K., Lindgren, U., and Westerlund, O. (2003). 'Geographical labour mobility: migration or commuting?'. *Regional Studies*, 37(8):827–837.

Espindola, A., Silveira, J., and Penna, T. (2006). 'A Harris-Todaro Agent-Based Model to Rural-Urban Migration'. *Brazilian Journal of Physics*, 36(3A).

ESPON (2009). 'DEMIFER - DEmographic and MIgratory Flows affecting European Regions and cities'. URL: http://www.espon.eu/mmp/online/website/content/programme/1455/2233/2236/2241/index_EN.html
Report.

Everitt, B. and Dunn, G. (2001). *Applied multivariate data analysis - second edition*. Hodder Arnold, London.

Everitt, B., Landau, S., and Leese, M. (2001). *Cluster Analysis*. Arnold, London, 4th edition.

Faggian, A., McCann, P., and Sheppard, S. (2006). 'An analysis of ethnic differences in UK graduate migration behaviour'. *Annals of Regional Science*, 40:461–471.

Faggian, A., McCann, P., and Sheppard, S. (2007). 'Some evidence that women are more mobile than men: gender differences in UK graduate migration behaviour'. *Journal of Regional Science*, 47(3):517–539.

Falkenauer, E. and Marchand, A. (2001). 'Using K-means? Consider ArrayMiner'. *Proceedings of the 2001 International Conference on Mathematics and Engineering Techniques in Medicine and Biological Sciences*.

Fan, C. C. (2005a). 'Interprovincial migration, population redistribution, and regional development in China: 1990 and 2000 census comparisons'. *Professional Geographer*, 57(2):295–311.

Fan, C. C. (2005b). 'Modeling interprovincial migration in China, 1985-2000'. *Eurasian Geography and Economics*, 46(3):165–184.

Farr, M. and Webber, R. (2001). 'MOSAIC: From and area classification system to individual classification'. *Journal of targeting, measurement and analysis for marketing*, 10(1):55–65.

Feldman, O., Simmonds, D., Troll, N., and Tsang, F. (2006). 'Creation of a system of functional areas for England and Wales and for Scotland'. Report, URL: http://www.mvaconsultancy.com/papers/Creation%20of%20a%20system%20of%20functional%20areas%20for%20England%20and%20W%85.pdf
MVA Consultancy.

Field, A. (2005). *Discovering Statistics Using SPSS*. Sage, London, second edition edition.

Fielding, A. (1982). 'Counterurbanization in Western Europe'. *Progress in Planning*, 17:1–52.

Fielding, A. (1992). 'Migration and social mobility: South East England as an escalator region'. *Regional Studies*, 26(1):1–15.

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):861–879.

Finney, N. and Simpson, L. (2007). 'Internal migration and ethnic groups: evidence for the UK from the 2001 Census'. Report, Cathie Marsh Centre for Census and Survey Research.

Finney, N. and Simpson, L. (2008). 'Internal migration and ethnic groups: evidence for Britain from the 2001 census'. *Population Space and Place*, 14:63–83.

Finney, N. and Simpson, L. (2009). 'Population Dynamics: The Roles of Natural Change and Migration in Producing the Ethnic Mosaic'. *Journal of Ethnic and Migration Studies*, 35(9):1479–1496.

Flowerdew, R. (2010). ‘Modelling migration with poisson regression’. In Stillwell, J., Duke-Williams, O., and Dennett, A. (eds.), *Technologies for Migration and Commuting Analysis: Spatial Interaction Data Applications*. IGI Global.

Flowerdew, R. and Aitkin, M. (1982). ‘A method of fitting the gravity model based on the Poisson distribution’. *Journal of Regional Science*, 22(2):191–202.

Flowerdew, R. and Al-Hamad, A. (2004). ‘The relationship between marriage, divorce and migration in a British data set’. *Journal of Ethnic and Migration Studies*, 30(2):339 – 351.

Flowerdew, R. and Green, M. (1992). ‘Developments in areal interpolation methods and GIS’. *Annals of Regional Science*, 26:67–78.

Fotheringham, A. S. (1983a). ‘A new set of spatial-interaction models: the theory of competing destinations’. *Environment and Planning A*, 15(1):15–36.

Fotheringham, A. S. (1983b). ‘Some theoretical aspects of destination choice and their relevance to production-constrained gravity models’. *Environment and Planning A*, 15(8):1121–1132.

Fotheringham, A. S. (1984). ‘Spatial flows and spatial patterns’. *Environment and Planning A*, 16(4):529–543.

Fotheringham, A. S. (1986a). ‘Further discussion on distance-deterrence parameters and the competing destinations model’. *Environment and Planning A*, 18(4):553–556.

Fotheringham, A. S. (1986b). ‘Modelling hierarchical destination choice’. *Environment and Planning A*, 18(3):401–418.

Fotheringham, A. S., Nakaya, T., Yano, K., Openshaw, S., and Ishikawa, Y. (2001). ‘Hierarchical destination choice and spatial interaction modelling: a simulation experiment’. *Environment and Planning A*, 33(5):901–920.

Fotheringham, A. S., Rees, P., Champion, T., Kalogirou, S., and Tremayne, A. R. (2004). ‘The development of a migration model for England and Wales: overview and modelling out-migration’. *Environment and Planning A*, 36(9):1633–1672.

Frost, M. and Dennett, A. (2010). ‘Issues associated with the analysis of rural commuting’. In Stillwell, J., Duke-Williams, O., and Dennett, A. (eds.), *Technologies for Migration and Commuting Analysis: Spatial Interaction Data Applications*. IGI Global, Hershey.

Garcia Coll, A. and Stillwell, J. (1999). ‘Inter-provincial migration in Spain: temporal trends and age-specific patterns’. *Int J Popul Geogr*, 5(2):97–115.

Geddes, A. and Flowerdew, R. (2004). 'The effect of the modifiable areal unit problem in modelling the distribution of limiting long-term illness in northern England'. In Boyle, P., Curtis, S., Graham, E., and Moore, E. (eds.), *The geography of health inequalities in the developed world: views from Britain and North America*. Ashgate Publishing Ltd, Aldershot.

Gehlke, C. and Biehl, K. (1934). 'Certain effects of grouping upon the size of the correlation coefficient in census tract material'. *Journal of the American Statistical Association Supplement*, 29:169–70.

Geist, C. and McManus, P. A. (2008). 'Geographical mobility over the life course: motivations and implications'. *Population, Space and Place*, 14(4):283–303. 10.1002/psp.508.

GLA (2008). 'Focus on London 2008'. Report, URL: http://www.london.gov.uk/who-runs-london/mayor/publications/society/facts-and-figures/focus-on-london/focus-2008
Greater London Authority.

Glasgow, N. and Brown, D. (2008). 'Grey gold: do older in-migrants benefit rural communities?'. Report, Carsey Institute, University of New Hampshire.

Gober-Meyers, P. (1978). 'Migration analysis: the role of geographical scale'. *The Annals of Regional Science*, 12(3):52–61.

Gordon, A. (1999). *Classification - Second Edition*. Monographs on Statistics and Applied Probability 82. Chapman and Hall, London, second edition edition.

Grayson, D. (2004). 'Some myths and legends in quantitative psychology'. *Understanding Statistics*, 3(1):101–134.

Green, A. (2004). 'Is Relocation Redundant? Observations on the Changing Nature and Impacts of Employment-related Geographical Mobility in the UK'. *Regional Studies*, 38(6):629 – 641.

Green, M. and Flowerdew, R. (1992). 'New evidence on the modifiable areal unit problem'. In Longley, P. and Batty, M. (eds.), *Spatial analysis: modelling in a GIS environment*, pages 41–54. GeoInformation International, Cambridge.

GROS (2003). 'Improving the GROS migration data for Council Areas - proposals for 2002'. Report, URL: http://www.gro-scotland.gov.uk/statistics/publications-and-data/population-estimates/02-chi-migration.html
General Register Office Scotland.

Grundy, E. and Fox, A. (1985). 'Migration during early married life'. *European Journal of Population*, 1:237–263.

Haas, W. H. and Serow, W. J. (1997). 'Retirement migration decision making: life course mobility, sequencing of events, social ties and alternatives'. *Journal of the Community Development Society*, 28(1):116 – 130.

Harland, K. (2008). *Journey to learn: geographical mobility and education provision*. PhD thesis.

Harris, C. and Ullman, E. (1945). 'The nature of cities'. *Annals of the American Academy of Political and Social Science*, 242:7–17.

Harris, R. (2005). 'Considering (mis-)representation in geodemographics and lifestyles'.

He, J. S. and Pooler, J. (2002). 'The regional concentration of China's interprovincial migration flows, 1982-90'. *Population and Environment*, 24(2):149–182.

Henrie, C. J. and Plane, D. A. (2008). 'Exodus from the California core: Using demographic effectiveness and migration impact measures to examine population redistribution within the western United States'. *Population Research and Policy Review*, 27(1):43–64.

Heppenstall, A., Evans, A., and Birkin, M. (2005). 'A Hybrid Multi-Agent/Spatial Interaction Model System for Petrol Price Setting'. *Transactions in GIS*, 9(1):35–51. 10.1111/j.1467-9671.2005.00204.x.

Hierro, M. (2009). 'Modelling the dynamics of internal migration flows in Spain'. *Papers in Regional Science*, 88(3):683–692.

Holdsworth, C. (1998). 'Leaving home in Spain: a regional analysis'. *Int J Popul Geogr*, 4(4):341–60.

Höppner, F., Klawonn, F., Kruse, R., and Runkler, T. (1999). *Fuzzy Cluster Analysis*. Wiley, Chippenham.

Horsfield, G. (2005). 'International migration'. In Chappell, R. (ed.), *Focus on Migration and People*, pages 115–129. Palgrave Macmillan, Basingstoke.

Hoyt, H. (1939). *The structure and growth of residential neighbourhoods in American cities*. Federal Housing Administration, Washington.

Hubert, L. and Arabie, P. (1985). 'Comparing partitions'. *Journal of Classification*, 2(1):193–218. 10.1007/BF01908075.

Huff, D. (1963). 'A probabilistic analysis of shopping centre trade areas'. *Land Economics*, 39:81–90.

Jefferies, J., Horsfield, G., Newman, J., and Stokes, K. (2003). 'Report on research into revising internal migration estimates'. Report, URL: http://www.statistics.gov.uk/downloads/theme_population/Revising_Internal_Migration_Estimates.pdf
ONS.

Jivraj, S. and Marquis, N. (2009). 'A comparison of internal migration data derived from the Pupil Level Annual School Census with the National Health Service Central Register and 2001 Census data'. Report, URL: http://www.ccsr.ac.uk/publications/working/2009-04.pdf
Cathie Marsh Centre for Census and Survery Research, CCSR Working Paper 2009-04.

Johnson, K., Voss, P., Hammer, R., Fuguitt, G., and McNiven, S. (2005). 'Temporal and spatial variation in age-specific net migration in the United States'. *Demography*, 42(4):791–813.

Jones, K. (1984). 'Graphical methods for exploring relationships'. In Bahrenberg, G., Fischer, M., and Nijkamp, P. (eds.), *Recent developments in spatial analysis: methodology, measurement, models*, pages 215–227. Gower, Aldershot.

Kalogirou, S. (2005). 'Examining and presenting trends of internal migration flows within England and Wales'. *Population Space and Place*, 11(4):283–297.

Kaufman, L. and Rousseeuw, P. (2005). *Finding groups in data - an introduction to cluster analysis*. John Wiley and Sons, New Jersey.

Kemper, F.-J. (2004). 'Internal migration in eastern and western Germany: convergence or divergence of spatial trends after unification?'. *Regional Studies*, 38(6):659 – 678.

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*, pages 165–85. Croom Helm, London.

Kim, C., Kim, S., and Lee, S. (2007). 'Regionalized distance decay parameter estimation of spatial interaction models'.

King, G., Tanner, M., and Rosen, O. (2004). *Ecological inference: new methodological strategies*. Cambridge University Press, Cambridge.

Kline, P. (1994). *An easy guide to factor analysis*. Routledge, London.

Knudsen, D. and Fotheringham, A. (1986). 'Matrix Comparison, Goodness-of-Fit, and Spatial Interaction Modeling'. *International Regional Science Review*, 10(2):127–147.

Kulu, H. and Milewski, N. (2007). 'Family change and migration in the life course: An introduction'. *Demographic Research*, 17:567–590.

Kupiszewski, M., Bucher, H., Durham, H., and Rees, P. (1998a). 'Internal migration and regional population dynamics in Europe: Germany case study'. Report, School of Geography, University of Leeds.

Kupiszewski, M., Drbohlav, D., Rees, P., and Durham, H. (1998b). 'Internal migration and regional population dynamics in Europe: Czech case study'. Report, School of Geography, University of Leeds.

Large, P. and Ghosh, K. (2006). 'Estimates of the population by ethnic group for areas within England'. *Population Trends*, 124:8–17.

Larkin, K. (2010). 'Local enterprise partnerships: centre for cities' 6-step plan'. Report, URL: http://www.centreforcities.org/
Centre for Cities.

Ledent, J. (1981). 'On the relationship between Alonso's theory of movement and Wilson's family of spatial-interaction models'. *Environment and Planning A*, 13(2):217–224.

Lee, E. (1966). 'A theory of migration'. *Demography*, 3(1):47–57. 0070-3370 Article type: Full Length Article / Full publication date: 1966 (1966). / Copyright 1966 Population Association of America.

Liang, Z., Chen, Y. P., and Gu, Y. M. (2002). 'Rural industrialisation and internal migration in China'. *Urban Studies*, 39(12):2175–2187. Conference of the Chinese-Economists-Society (CES) JUN, 2001 XIAMEN, PEOPLES R CHINA.

Liang, Z. and White, M. J. (1996). 'Internal migration in China, 1950-1988'. *Demography*, 33(3):375–384.

Long, L. (1991). 'Residential mobility differences among developed countries'. *International Regional Science Review*, 14(2):133–147.

Longley, P. (2005). 'Geographical information systems: a renaissance of geodemographics for public service delivery'. *Progress in Human Geography*, 29:57–63.

Makowsky, M., Tavares, J., Makany, T., and Meiser, P. (2006). 'An agent-based model of crisis-driven migration'. Report, URL: http://www.santafe.edu/events/workshops/images/6/67/Sf_csss06_makowsky_et_al.pdf
Santa Fe Institute.

Manson, G. A. and Groop, R. E. (2000). 'US intercounty migration in the 1990s: People and income move down the urban hierarchy'. *Professional Geographer*, 52(3):493–504.

Marquis, N. and Jivraj, S. (2009). 'Preparation of Pupil Level Annual School Census data for the analysis of internal migration'. Report, URL: http://www.ccsr.ac.uk/publications/working/2009-03.pdf

Cathie Marsh Centre for Census and Survery Research, CCSR Working Paper 2009-03.

Martin, D. (1998). 'Optimising census geography: the separation of collection and output geographies'. *International Journal of Geographical Information Science*, 12(7):673–685.

Martin, D. (2000). 'Towards the geographies of the 2001 UK Census of Population'. *Transactions of the Institute of British Geographers*, 25:321–332.

Martin, D. (2002). 'Output areas for 2001'. In Rees, P., Martin, D., and Williamson, P. (eds.), *The census data system*. John Wiley and Sons, Chichester.

MathWorks (2009). *MATLAB Statistics Toolbox 7 - User's Guide*. The MathWorks Inc., Natick.

Michielin, F., Mulder, C., and Zorlu, A. (2008). 'Distance to parents and geographical mobility'. *Population Space and Place*, 14:327–345.

Milligan, G. (1996). 'Clustering validation: results and implications for applied analyses'. In Arabie, P., Hubert, L., and De Soete, G. (eds.), *Clustering and Classification*, pages 341–375. World Scientific, Singapore.

Milligan, G. and Cooper, M. (1985). 'An examination of procedures for determining the number of clusters in a dataset'. *Psychometrika*, 50:159–179.

Milligan, G. and Cooper, M. (1987). 'Methodology Review: Clustering Methods'. *Applied Psychological Measurement*, 11:329–354.

Milligan, G. and Cooper, M. (1988). 'A study of standardization of variables in cluster analysis'. *Journal of Classification*, 5(181-204).

Mulder, C. and Wagner, M. (1993). 'Migration and marriage in the life course: a method for studying synchronised events'. *European Journal of Population*, 9:55–76.

Murphy, A., Muellbauer, J., and Cameron, G. (2006). 'Housing market dynamics and regional migration in Britain'. Report, Department of Economics, University of Oxford.

Newbold, K. B. and Peterson, D. (2001). 'Distance weighted migration measures'. *Papers in Regional Science*, 80:371–380.

NISRA (2005). 'Development of methods/sources to estimate population migration in Northern Ireland'. Report, URL: http://www.nisra.gov.uk/archive/demography/population/migration/dev_est_mig.pdf
Northern Ireland Statistics and Research Agency.

Norman, P. (1999). 'Putting iterative proportional fitting on the researchers desk'. Report, URL: http://www.geog.leeds.ac.uk/wpapers/99-3.pdf
School of Geography, University of Leeds.

Norman, P., Boyle, P., and Rees, P. (2005). 'Selective migration, health and deprivation: a longitudinal analysis'. *Social Science and Medicine*, 60(12):2755–2771.

Northcott, H. (1985). 'The geographic mobility of Canada's elderly'. *Canadian Studies in Population*, 12(2):183–202.

ODPM (2006). 'A framework for city-regions'. Report, URL: http://www.communities.gov.uk/publications/citiesandregions/framework
Office of the Deputy Prime Minister.

Ogilvy, A. A. (1982). 'Population migration between the regions of Great Britain, 1971-9'. *Regional Studies*, 16(1):65 – 73.

Oliver, C. (2007). *Retirement Migration*. Routledge Research in Population and Migration. Routledge, London.

ONS (2003a). 'Census 2001 - frequently asked questions'. Report, URL: http://www.statistics.gov.uk/census2001/pdfs/onc_qu_ans.pdf
Office for National Statistics.

ONS (2003b). 'Key findings and actions from the one number census quality assurance process'. Report, URL: http://www.statistics.gov.uk/census2001/pdfs/onc_key_findings.pdf
Office for National Statistics.

ONS (2003c). 'One number census methodology and quality assurance process report'. Report, URL: http://www.statistics.gov.uk/census2001/pdfs/onc_qa_process.pdf
Office for National Statistics.

ONS (2004). 'National Statistics 2001 area classification for local authorities'.

ONS (2005a). 'Focus On People and Migration'. Report, URL: http://www.statistics.gov.uk/focuson/Migration/
Office for National Statistics.

ONS (2005b). 'Making a population estimate in England and Wales'. Report, URL: http://www.statistics.gov.uk/StatBase/Product.asp?vlnk=575
Office for National Statistics.

ONS (2006). 'Internal migration estimates for local and unitary authorities in England and Wales, health authorities in England and former health authorities in Wales, year to mid-2005'. *Population Trends*, 125:77–92.

ONS (2007a). 'Estimating internal migration - customer guidance notes'. Report, URL: http://www.statistics.gov.uk/statbase/EXPODATA/commentary/EstimatinginternalMigration.doc
Migration Statistics Unit, Office for National Statistics.

ONS (2007b). 'Internal migration estimates for local and unitary authorities in England and Wales, health authorities in England and former health authorities in Wales, year to mid-2006'. *Population Trends*, 129:66–82.

ONS (2007c). 'Population turnover figures reveal large changes in the population of some seaside towns'.

ONS (2007d). 'Quarterly population estimates - methodology'. Report, URL: http://www.statistics.gov.uk/downloads/theme_population/QPE_Methodology_June07.pdf
Office for National Statistics.

ONS (2008a). '2006-based subnational population projections for England - methodology guide'. Report, URL: http://www.statistics.gov.uk/downloads/theme_population/SNPP-2006/2006_Methodology_Guide.pdf
Office for National Statistics.

ONS (2008b). 'Internal migration estimates for local and Unitary Authorities in England and Wales, year to mid-2007'. *Population Trends*, 133:81–97.

ONS (2008c). 'Migration Indicators by LA in England and Wales, 2001-2007'.

ONS (2009a). 'Improved estimation of student migration within England and Wales'. Report, URL: http://www.ons.gov.uk/about-statistics/methodology-and-quality/imps/mig-statsimprove-prog/comm-stakeholders/improvements-to-the-mid-2008-population-estimates/detailed-methodology-papers/student-methods.pdf
Office for National Statistics.

ONS (2009b). 'Measuring uncertainty in the local authority population estimates - interim report focusing on internal migration'. Report, URL: http://tinyurl.com/ons-internal-interim
Office for National Statistics.

ONS (2009c). 'Social Trends'. Report, URL: http://www.statistics.gov.uk/StatBase/Product.asp?vlnk=5748
Office for National Statistics.

ONS (2009d). 'Summary quality report for internal migration'. Report, URL: http://www.statistics.gov.uk/about/data/methodology/quality/downloads/IM_SQR.pdf
Office for National Statistics.

ONS (2009e). 'Use of School Census Data to Improve Population and Migration Statistics'. Report, Office for National Statistics.

ONS (2010a). 'Final recommended questions for the 2011 Census in England and Wales - Second addresses'. Report, URL: http://tinyurl.com/census2011-second-home

Office for National Statistics.

ONS (2010b). ‘Improving Migration and Population Statistics: Overview of the package of improvements’. Report, Office for National Statistics.

Openshaw, S. (1984). *The modifiable areal unit problem*. Geo Books, Norwich.

Openshaw, S. (1989). ‘Making geodemographics more sophisticated’. *Journal of the market research society*, 31:111–131.

Openshaw, S. (1998). ‘Neural network, genetic, and fuzzy logic models of spatial interaction’. *Environment and Planning A*, 30(10):1857–1872.

Openshaw, S. and Blake, M. (1996). ‘GB Profiler 91’.

Openshaw, S. and Taylor, P. (1979). ‘A million or so correlation coefficients: three experiments on the modifiable areal unit problem’. In Wrigley, N. and Bennet, R. (eds.), *Statistical applications in spatial sciences*, pages 127–144. Pion, London.

Openshaw, S. and Wymer, C. (1995). ‘Classifying and regionalizing census data’. In Openshaw, S. (ed.), *Census Users’ Handbook*. Geoinformation International, Cambridge.

Orford, S., Dorling, D., Mitchell, R., Shaw, M., and Smith, G. (2002). ‘Life and death of the people of London: a historical GIS of Charles Booth’s inquiry’. *Health and Place*, 8(1):25–35.

O’Sullivan, D. and Unwin, D. (2002). *Geographical information analysis*. John Wiley and Sons, London.

Owen, D. (1997). ‘Migration by minority ethnic groups within Great Britain in the early 1990s’.

Owen, D. and Green, A. (1992). ‘Migration patterns and trends’. In Champion, A. and Fielding, A. (eds.), *Migration Processes and Patterns*, volume 1, pages 17–40. Belhaven Press, Kings Lynn.

Pailhe, A. and Solaz, A. (2008). ‘Professional outcomes of internal migration by couples: Evidence from France’. *Population Space and Place*, 14(4):347–363.

Pettersson, A. and Malmberg, G. (2009). ‘Adult children and elderly parents as mobility attractions in Sweden’. *Population, Space and Place*, 15(4):343–357. 10.1002/psp.558.

Pingle, J. (2007). ‘A note on measuring internal migration in the United States’. *Economics Letters*, 94(1):38–42.

Plane, D., Henrie, C., and Perry, M. (2005). ‘Migration up and down the urban hierarchy and across the life course’. *PNAS*, 102(43):15313–15318.

Plane, D. and Jurjevich, J. (2009). 'Ties That No Longer Bind? The Patterns and Repercussions of Age-Articulated Migration'. *The Professional Geographer*, 61(1):4 – 20.

Pooley, C. and Turnbull, J. (1996). 'Migration trends in British rural areas from the 18th to the 20th centuries'. *International Journal of Population Geography*, 2(3):215–237.

Rand, W. (1971). 'Objective Criteria for the Evaluation of Clustering Methods'. *Journal of the American Statistical Association*, 66(336):846–850.

Ravenstein, E. (1885). 'The laws of migration'. *Journal of the Statistical Society of London*, 48(2):167–235.

Ravenstein, E. (1889). 'The laws of migration'. *Journal of the Royal Statistical Society*, 52(2):241–305.

Rayer, S. and Brown, D. L. (2001). 'Geographic diversity of inter-county migration in the United States, 1980-1995'. *Population Research and Policy Review*, 20(3):229–252.

Raymer, J. (2007). 'The estimation of international migration flows: a general technique focused on the origin - destination association structure'. *Environment and Planning A*, 39:985–995.

Raymer, J. and Abel, G. (2008). 'Methods to improve estimates of migration flows - the MIMOSA model for estimating international migration flows in the European Union'. Report, UNECE/Eurostat work session on migration statistics - Working Paper, 8.

Raymer, J., Abel, G., and Smith, P. (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(4):891–908.

Raymer, J., Bonaguidi, A., and Valentini, A. (2006). 'Describing and projecting the age and spatial structures of interregional migration in Italy'. *Population Space and Place*, 12(5):371–388.

Raymer, J. and Giulietti, C. (2009). 'Ethnic migration between area groups in England and Wales'. *Area*, 41(4):435–451.

Raymer, J. and Giulietti, C. (2010). 'Analysing structures of interregional migration in England'. In Stillwell, J., Duke-Williams, O., and Dennett, A. (eds.), *Technologies for Migration and Commuting Analysis: Spatial Interaction Data Applications.* IGI Global, Hershey.

Raymer, J. and Rogers, A. (2007). 'Using age and spatial flow structures in the indirect estimation of migration streams'. *Demography*, 44(2):199–223.

Raymer, J. and Rogers, A. (2008). 'Applying model migration schedules to represent age-specific migration flows'. In Raymer, J. and Willekens, F. (eds.), *International Migration in Europe: Data, Models and Estimates*, pages 175–205. Wiley, Chichester.

Raymer, J., Smith, P., and Giulietti, C. (2008). 'Combining census and registration data to analyse ethnic migration patterns in England from 1991 to 2007'. Report, URL: http://eprints.soton.ac.uk/63739/
University of Southampton, Southampton Statistical Sciences Research Institute.

Redford, A. (1926). *Labour Migration in England, 1800-50.* University of Manchester Press, Manchester.

Rees, P. (1977). 'The measurement of migration from census and other sources'. *Environment and Planning A*, 9:257–280.

Rees, P. (2009). 'Demography'. In Kitchin, R. and Thrift, N. (eds.), *International Encyclopedia of Human Geography*, pages 75–90. Elsevier, Oxford. doi: DOI: 10.1016/B978-008044910-4.00815-4.

Rees, P., Bell, M., Duke-Williams, O., and Blake, M. (2000). 'Problems and solutions in the measurement of migration intensities: Australia and Britain compared'. *Population Studies-a Journal of Demography*, 54(2):207–222.

Rees, P. and Boden, P. (2006). 'Estimating London's new migrant population. Stage 1 - review of methodology'. Report, Greater London Authority.

Rees, P., Boden, P., Dennett, A., Stillwell, J., Jasinska, M., de Jong, A., ter Veer, M., Kupiszewski, M., and Kupiszewska, D. (2010). 'Regional population dynamics: a report assessing the effects of demographic developments on regional competitiveness and cohesion'. Report, ESPON.

Rees, P., Carrilho, M., Peixoto, J., Durham, H., and Kupiszewski, M. (1998a). 'Internal migration and regional population dynamics in Europe: Portugal case study'. Report, School of Geography, University of Leeds.

Rees, P., Denham, C., Charlton, J., Openshaw, S., Blake, M., and See, L. (2002a). 'ONS classifications and GB profiles: Census typologies for researchers'. In Rees, P., Martin, D., and Williamson, P. (eds.), *The Census Data System*. John Wiley and Sons Ltd., Padstow.

Rees, P., Durham, H., and Kupiszewski, M. (1996). 'Internal migration and regional population dynamics in Europe: United Kingdom case study'. Working Paper 96/20, URL: http://www.geog.leeds.ac.uk/wpapers/96-20.pdf
School of Geography, University of Leeds.

Rees, P., Fotheringham, A., and Champion, T. (2004). 'Modelling migration for policy analysis'. In Stillwell, J. and Clarke, G. (eds.), *Applied GIS and Spatial Analysis*, pages 259–296. Wiley, Chichester.

Rees, P., Martin, D., and Williamson, P. (2002b). 'Census data resources in the United Kingdom'. In Rees, P., Martin, D., and Williamson, P. (eds.), *The Census Data System*. John Wiley and Sons Ltd., Chichester.

Rees, P., Martin, D., and Williamson, P. (2002c). *The census data system*. John Wiley and Sons Ltd., Padstow.

Rees, P., stby, L., Durham, H., and Kupiszewski, M. (1998b). 'Internal migration and regional population dynamics in Europe: Norway case study'. Report, School of Geography, University of Leeds.

Rees, P. and Stillwell, J. (1987). 'Internal migration in the United Kingdom'. Report 497, School of Geography, University of Leeds.

Rees, P., Stillwell, J., Boden, P., and Dennett, A. (2009). 'A review of migration statistics literature'. Report, UK Statistics Authority.

Rees, P., Thomas, F., and Duke-Williams, O. (2002d). 'Migration data from the Census'. In Rees, P., Martin, D., and Williamson, P. (eds.), *The Census Data System*. John Wiley and Sons, Ltd., Chichester.

Rees, P., Van Imhoff, E., Durham, H., Kupiszewski, M., and Smith, D. (1998c). 'Internal migration and regional population dynamics in Europe: Netherlands case study'. Report, School of Geography, University of Leeds.

Rogers, A. (1978). 'Demometrics of migration and settlement'. In Batey, P. (ed.), *London Papers in Regional Science 8 - Theory and method in urban and regional analysis*, pages 1–30. Pion, London.

Rogers, A. (1989). 'Requiem for the net migrant'. Report, University of Colorado, Institute of Behavioural Science, Population Programme.

Rogers, A. and Castro, L. (1981). 'Model migration schedules'. Report, International Institute for Applied Systems Analysis.

Rogers, A., Raymer, J., and Newbold, K. B. (2003a). 'Reconciling and translating migration data collected over time intervals of differing widths'. *Annals of Regional Science*, 37(4):581–601.

Rogers, A., Raymer, J., and Willekens, F. (2002). 'Capturing the age and spatial structures of migration'. *Environment and Planning A*, 34(2):341–359.

Rogers, A., Willekens, F., and Raymer, J. (2003b). 'Imposing age and spatial structures on inadequate migration-flow datasets'. *Professional Geographer*, 55(1):56–68.

Rotolo, T. and Tittle, C. (2006). 'Population size, change, and crime in U.S. cities'. *Journal of Quantitative Criminology*, 22:341–367.

Rousseeuw, P. J. (1987). 'Silhouettes: A graphical aid to the interpretation and validation of cluster analysis'. *Journal of Computational and Applied Mathematics*, 20:53–65. 0377-0427 doi: DOI: 10.1016/0377-0427(87)90125-7.

Rowland, D. (2006). *Demographic methods and concepts*. Oxford University Press, Oxford.

Roy, J. and Thill, J. (2004). 'Spatial interaction modelling'. *Papers in Regional Science*, 83:339–361.

Sandefur, G. and Scott, W. (1981). 'A dynamic analysis of migration: an assessment of the effects of age, family and career variables'. *Demography*, 18(3):355–368.

Schaffer, C. and Green, P. (1996). 'An empirical comparison of variable standardisation methods in cluster analysis'. *Multivariate Behavioural Research*, 31(2):149–167.

Scott, A. and Kilbey, T. (1999). 'Can patient registers give an improved measure of internal migration in England and Wales?'. *Population Trends*, 96:44–55.

See, L. and Openshaw, S. (2001). 'Fuzzy geodemographic targeting'. In Clarke, G. and Madden, M. (eds.), *Regional science in business*, Advances in Spatial Science, pages 269–281. Springer, Berlin.

Senior, M. (1979). 'From gravity modelling to entropy maximizing: a pedagogic guide'. *Progress in Human Geography*, 3(2):175–210.

Shen, J. (1996). 'Internal migration and regional population dynamics in China'. *Prog Plann*, 45(3):123–88.

Shepherd, P. (2006). *Neighbourhood profiling and classification for community safety*. Phd thesis.

Simpson, L. and Finney, N. (2009). 'Spatial patterns of internal migration: evidence for ethnic groups in Britain'. *Population, Space and Place*, 15:37–56.

Sjaastad, L. (1962). 'The costs and returns of human migration'. *The journal of Political Economy*, 70(5):80–93.

Skipton (2003). 'Parents shell out to help children reclaim their home'. Report, URL: http://www.skipton.co.uk/press_office/publicity_campaigns/kick_out_your_kids/newsRelease.aspx Skipton Building Society.

Smailes, P. J. (1996). 'Demographic response to rural restructuring and counterurbanisation in South Australia, 1981-1991'. *Int J Popul Geogr*, 2(3):261–87.

Smith, D. (2002). 'Patterns and processes of 'studentification' in Leeds'. *The Regional Review*, 12(1):14–16.

Smith, D. and Denholm, J. (2006). 'Studentification: A guide to challenges, opportunities and practice'. Report, Universities UK.

Smith, P. F., Raymer, J., and Giulietti, C. (2009). 'Combining available migration data in England to study economic activity flows over time'. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*. 10.1111/j.1467-985X.2009.00630.x.

SPSS (2005). *SPSS Base 14.0 User's Guide*. SPSS Inc., Chicago.

SPSS (2006). *SPSS 14.0 for Windows*. SPSS.

Stillwell, J. (1978). 'Interzonal migration: some historical tests of spatial-interaction models'. *Environment and Planning A*, 10:1187–1200.

Stillwell, J. (2006). 'Providing access to census-based interaction data: that's WICID'. *The Journal of Systemics Cybernetics and Infomatics*, 4(1):63–68.

Stillwell, J. (2008). 'Inter-regional migration modelling: a review and assessment'. In Poot, J., Waldorf, B., and Van Wissen, L. (eds.), *Migration and human capital: regional and global perspectives*. Edward Elgar.

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):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. Part 2: age related migration patterns'. *Journal of Population Research*, 18(1).

Stillwell, J. and Boden, P. (1986). 'Internal migration in the United Kingdom: characteristics and trends.'. Report 470, School of Geography, University of Leeds.

Stillwell, J. and Duke-Williams, O. (2005). 'Ethnic population distributions, immigration and internal migration in Britain: What evidence for linkage at district scale'.

Stillwell, J. and Duke-Williams, O. (2007). 'Understanding the 2001 UK census migration and commuting data: the effect of small cell adjustment and problems of comparison with 1991'. *Journal of the Royal Statistical Society Series A (Statistics in Society)*, 170(2):425–445.

Stillwell, J., Duke-Williams, O., and Dennett, A. (2010). *Technologies for Migration and Commuting Analysis: Spatial Interaction Data Applications*. IGI Global, Hershey.

Stillwell, J., Duke-Williams, O., and Rees, P. (1995a). 'Time series migration in Britain: the context for 1991 Census analysis.'. *Papers in Regional Science*, 74(4):341–359. 10.1111/j.1435-5597.1995.tb00645.x.

Stillwell, J. and Hussain, S. (2008). 'Ethnic group migration within Britain during 2000-01: a district level analysis'. Report, URL: http://www.geog.leeds.ac.uk/wpapers/index.html School of Geography, University of Leeds.

Stillwell, J., Hussain, S., and Norman, P. (2008). 'The internal migration propensities and net migration patterns of ethnic groups in Britain'. *Migration Letters*, 5(2):135–150.

Stillwell, J., Rees, P., and Boden, P. (1992). 'Internal migration trends: an overview'. In Stillwell, J., Rees, P., and Boden, P. (eds.), *Migration Processes and Patterns*, volume 2. Belhaven Press, Kings Lynn.

Stillwell, J., Rees, P., and Duke-Williams, O. (1995b). 'Migration between Euro-Regions in the United Kingdom'. Report 95/19, School of Geography, University of Leeds.

Taylor, P. (1983). *Distance decay in spatial interactions*, volume 2 of *CATMOD*. Geo Books, Norwich.

Tobler, W. (1970). 'A computer model simulation of urban growth in the Detroit region'. *Economic Geography*, 46(2):234–240.

Tobler, W. (1995). 'Migration: Ravenstein, Thornthwaite, and beyond'. *Urban Geography*, 16:327–343.

Travers, T., Tunstall, R., Whitehead, C., and Pruvot, S. (2007). 'Population mobility and service provision: a report for London councils'. Report, LSE London.

UKSA (2009). 'Migration statistics: the way ahead'. Report, UK Statistics Authority.

UNESCO (2010). 'UNESCO glossary on migration'.

Uren, Z. and Goldring, S. (2008). 'Migration trends at older ages in England and Wales'. *Population Trends*, 130:31–40.

Van Wissen, L., Van der Gaag, N., Rees, P., and Stillwell, J. (2008). 'In search of a modelling strategy for projecting internal migration in European countries: demographic versus explanatory approaches'. In Poot, J., Waldorf, B., and Van Wissen, L. (eds.), *Migration and human capital: regional and global perspectives*. Edwards Elgar.

Vandeschrick, C. (2001). 'The Lexis diagram, a misnomer'. *Demographic Research*, 4:97–124.

Vickers, D. (2006). *Multi-level integrated classifications based on the 2001 Census*. PhD thesis.

Vickers, D. and Rees, P. (2009). 'Ground-truthing Geodemographics'. *Applied Spatial Analysis and Policy*. 10.1007/s12061-009-9037-5.

Vickers, D., Rees, P., and Birkin, M. (2003). 'A new classification of UK local authorities using 2001 Census key statistics'. Report 22/3/07.

Vickers, D., Rees, P., and Birkin, M. (2005). 'Creating the national classification of census output areas: data, methods and results'. Report, URL: http://www.geog.leeds.ac.uk/wpapers/05-2.pdf
University of Leeds.

Voas, D. and Williamson, P. (2001). 'The diversity of diversity: a critique of geodemographic classification'. *Area*, 33(1):63–76.

Walmsley, D. J., Epps, W. R., and Duncan, C. J. (1998). 'Migration to the New South Wales North Coast 1986-1991: Lifestyle motivated counterurbanisation'. *Geoforum*, 29(1):105–118.

Ward, J. (1963). 'Hierarchical grouping to optimize an objective function'. *Journal of the American Statistical Association*, 58:236–244.

Wei, Y. (1997). 'Interregional migration in socialist countries: the case of China'. *GeoJournal*, 41(3):205–14.

White, A. (2009). 'Internal Migration, Identity and Livelihood Strategies in Contemporary Russia'. *Journal of Ethnic and Migration Studies*, 35(4):555–573.

Whitehead, A., Hashim, I., and Iversen, V. (2007). 'Child Migration, Child Agency and Intergenerational Relations in Africa and South Asia'. Report, URL: http://www.migrationdrc.org
Development Research Centre on Migration, Globalisation and Poverty, University of Sussex.

Willekens, F. (1983). 'Log-linear modelling of spatial interaction'. *Papers in Regional Science*, 52(1):187–205. 10.1007/BF01944102.

Willekens, F. (1999). 'Modeling approaches to the indirect estimation of migration flows: from entropy to EM.'. *Mathematical Population Studies*, 7(3):239–278.

Wilson, A. (1970). *Entropy in urban and regional modelling*. Monographs in spatial and environmental systems analysis. Pion, London.

Wilson, A. (1971). 'A family of spatial interaction models, and associated developments'. *Environment and Planning A*, 3:1–32.

Wilson, A. (2000). *Complex spatial systems: the modelling foundations of urban and regional analysis*. Pearson Education Ltd., Harlow.

Wilson, A. (2008). 'Boltzmann, Lotka and Volterra and spatial structural evolution: an integrated methodology for some dynamical systems'. *Journal of the Royal Society Interface*, 5:865–871.

Wilson, A. G. (1980). 'Comments on Alonso's 'theory of movement''. *Environment and Planning A*, 12(6):727–732.

Wrigley, N., Holt, T., Steel, D., and Tranmer, M. (1996). 'Analysing, modelling, and resolving the ecological fallacy'. In Longley, P. and Batty, M. (eds.), *Spatial analysis: modelling in a GIS environment*, pages 25–40. GeoInformation International, Cambridge.

Yano, K. (2001). 'GIS and quantitative geography'. *GeoJournal*, 52:173–180.

Yeung, K. and Ruzzo, W. (2001). 'Details of the Adjusted Rand index and Clustering algorithms Supplement to the paper An empirical study on Principal Component Analysis for clustering gene expression data'. Report, URL: http://faculty.washington.edu/kayee/pca/supp.pdf University of Washington.

Založnik, M. (2006). *Geographical variation of geodemographic classifiability*. Ma thesis.

Zasada, I., Alves, S., Muller, F. C., Piorr, A., Berges, R., and Bell, S. (2010). 'International retirement migration in the Alicante region, Spain: process, spatial pattern and environmental impacts'. *Journal of Environmental Planning and Management*, 53(1):125 – 141.

Zelinsky, W. (1971). 'The Hypothesis of the Mobility Transition'. *Geographical Review*, 61(2):219–249.

Zipf, G. (1946). 'The P1 P2 / D hypothesis: on the intercity movement of persons'. *American Sociological Review*, 11(6):677–686.