Modelling Individual and Place Variations in Residential Moves using Commercial Data and Official Statistics

Michael James Thomas

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

The University of Leeds

School of Geography

July 2014

The candidate confirms that the work submitted is his/her 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.

The work in Chapter 6 of the thesis has appeared in publication as follows:

> Thomas, M., Stillwell, J. & Gould, M. (2014). Exploring and validating a commercial lifestyle survey for its use in the analysis of population migration, *Applied Spatial Analysis and Policy*, 7: 71-95.

I declare that the research for this publication was solely my own work and that I am the lead author. The contribution of the other named authors, John Stillwell and Myles Gould, was purely editorial and advisory.

Acknowledgements

First and foremost I would like to thank my two supervisors John Stillwell and Myles Gould. Their enthusiasm, dedication and generosity in providing their time and sage advice is most appreciated. From my first visit to the university, where Myles spent an afternoon showing me around campus and introducing me to the long list of great bands who've played there, right up to the writing up of this thesis, where John's superhuman ability to make time for me despite his great many other callings continued unabated, I could not have asked for two better supervisors – thank you!

Likewise, Paul Norman and Mark Birkin have also been very kind in providing their time, help and expertise as members of my research support group. As the director of TALISMAN (funded by the ESRC), Mark has also been hugely important in providing me with the necessary resources and encouragement to seek out the best training opportunities across the UK as well as present my work to various audiences both here and abroad. Additional thanks go to Acxiom Ltd., and to Clare Woodvine in particular, for providing the data sets upon which the research in this thesis based and to the ESRC for the TALISMAN +3 postgraduate studentship.

I would also like to thank the Oz gang – Michelle Almond, Luke Burns, Graham Clarke, Nik Lomax, Michelle Morris, Andy Newing and Chris Thompson (in alphabetical order!) – for the experience of a lifetime, as well as their advice and friendship through the last three years. Similarly, thanks go to my fellow (adopted) Sheffielders' Nick Hood (Geordie) and Fran Darlington, and to Nick Addis, Thomas Murphy.

Given that the completion of a thesis is the final stage of a seven year marathon, I would like to thank Dimitris Ballas, Jessica Dubow, Charles Pattie, Deborah Sporton and Dan Vickers for their invaluable contributions during my undergraduate and postgraduate education at the Geography Department, University of Sheffield. I would also like to thank Josie Wilson and Paul Jones at Sheffield Hallam University, as well as Karen Escott for her support and advice.

Huge thanks go to my family, my mum, dad and Rachael – "*who'd have thunk it*?" – also to my friends in Sheffield (Baz, Cal, Danny, Fisher, Jenks, Jose, Josh, McGee, Nash, Oli, Ralph and Tim) for providing ample distractions from the work!

Finally, Alice, I can't thank you enough … this is dedicated to you.

Abstract

Through its ability to transform local area population size, composition and character, residential mobility is a subject of particular relevance for policy makers, service providers, academics and, to some extent, the population at large. Whilst the phenomenon can be understood in very basic terms as the relocation of an individual and/or household from one geographic location to another, the place-based and subject-specific determinants that are said to inform population movement, and the associated propensities and trends, are inherently complex and multifaceted.

There is a long tradition in the quantitative study of population movement in Great Britain, with a great many models calibrated using different data sources of varying detail, size and coverage and designed with the purpose of providing improved interpretation and understanding of either micro (individual/household) or macro (area) processes. In this thesis a new source of commercial data is employed which has the potential to allow for a novel break from the traditional dichotomy of the micro/macro approach. Indeed, through the combined use of detailed geo-referenced and geographically extensive microdata, appropriate statistical methods, and well-informed micro and macro theory, this work is able to simultaneously measure, analyse and interpret a variety of individual and place variations in residential mobility in Britain.

The thesis integrates a previously unused source of commercial data with official statistics and provides unique insights into various multilevel patterns, propensities and characteristics of residential mobility that have, whilst long theorised, often been difficult to demonstrate empirically due to a longstanding dearth in access to suitably detailed data and methods. In particular, new insights are gained through the examination of a number of understudied subjective and behavioural characteristics of movers vis-à-vis stayers across different life-course stages, the detailed interrogation of duration-of-residence effects and associated residential exposure times on future movement propensities and the simultaneous analysis of micro and macrogeographical (origin and destination) variations in the postcode-to-postcode distance travelled by recent movers.

Table of Contents

List of Tables

List of Figures

Chapter 1
Introduction

1.1 Research context

Residential mobility is a key component in the evolution of local population size and structure and is thus a phenomenon of huge social and economic relevance and, unsurprisingly, of academic and policy interest. Broadly speaking, residential mobility involves the relocation of individuals and/or households between geographical locations. The phenomenon itself is inherently complex, not only because of the variety of possible mobility patterns and outcomes, but also because of the myriad of (tied) place-based and subject-specific motivations, events and characteristics that are expected to inform the propensity to change residence. Much has been written about the variations in movement patterns, propensities and trends by age, sex and, more recently, by ethnicity, for example; and many independent variables have been incorporated within different types of explanatory model. For instance, at the aggregate level, the calibration of different macro migration models has provided a fairly detailed understanding of the broad spatial and compositional characteristics of migration flows across Great Britain (GB) (see Rees *et al.*, 2004). However, a greater understanding of the more personal lifestyle and behavioural characteristics, for instance the duration of residence, the subjective evaluation of residential environments and the access to financial resources is essential, especially since these are the issues that are often theorised to be central to the decision-making processes behind individual/household mobility propensities and thus wider movement patterns and outcomes (Rossi, 1955).

However, whilst there exists a large literature focussed on the application of micro approaches, that is, broadly focussing on the behavioural and decision-making processes of the micro unit (individual and/or household), there has been a clear dearth in research focussed on the simultaneous influence of factors operating across levels, from the micro-individual through to the macro-geographic. Whilst macro approaches to migration modelling have emphasised the importance of macro explanatory variables including population size, employment rate or environmental factors at either or both residential origin and destination, micro approaches have suffered from a distinct absence of such factors within their behavioural models. Theoretically, the failure to include relevant contextual factors may be problematic for micro models designed to explore the behavioural and decision-making processes behind residential mobility, particularly if we consider our evaluations of residential environments to extend beyond matters of the individual and household.

Thus, there remains a continuing need to understand the micro and macro variations, and the cross-level interactions, in residential moves. Of course, it should be noted that this continuing need is not driven by a widespread ignorance of context by population researchers; rather it is more likely the result of a particular scarcity in geographically detailed microdata of sufficient sample size and geographic coverage. This thesis draws on a unique opportunity to utilise a previously unused source of commercial microdata, namely the Acxiom Ltd. Research Opinion Poll (ROP), which contains many of the socio-demographic and behavioural characteristics required for a detailed analysis of individual and place variations in residential moves in GB.

1.2 Research questions, overall project aims and specific research objectives

Consequently, following the brief introduction to the research context, a number of detailed research questions can be identified as follows:

1 Can reliable data on residential mobility, socio-demographic and lifestyle characteristics, be extracted from the Acxiom Ltd. Research Opinion Poll that allow for the following research questions to be addressed?

2 What variations occur between different types of people (e.g. demographic, socio-economic and lifestyle groups) in the propensity to move residence and are these variations consistent across the broad stages of the life course?

3 Taking into account individual and area-level characteristics, does an individual's duration of residence at his/her current address (in years) influence his/her likelihood to be planning a future move? [3a] Is there additional evidence of a substantively important cross-level interaction, e.g.

between an individual's duration of residence and the stability of his/her neighbourhood population? [3b] Is there evidence that exposure times (to residential environments) are important in enacting variations in the duration effects?

4 [4a] Are there discernible differences, in terms of individual socio-demographic and lifestyle characteristics, between those who move shorter distances and those who move further? [4b] Do some origin/destination types lose/attract ('send'/'receive') longer/shorter distance movers than others?

These research questions align with the overall project aims:

a.) to investigate individual and place variations in residential mobility and immobility in Great Britain using commercial data and official statistics;

b.) to explore the effects of duration of residence, and additional cross-level interactions, on the propensity for future residential moves; and

c.) to examine the potential variations in migrant origin to destination distance according to individual and place-based characteristics.

In an attempt to fulfil the overall project aims, the research project has a number of more specific research objectives, of which all are sought to be met in the subsequent chapters of the thesis. Table 1.1 summarises the research objectives and the corresponding chapter(s) through which they are addressed.

Table 1.1. Specific research objectives and corresponding chapters

Research Objective	Corresponding chapter(s)
I. To explore and review the existing literature associated with individual and area demographic, socio-economic and lifestyle dimensions of population movement in GB and provide the theoretical and empirical context for the analyses herein	**Chapter 2** – Population movement in GB: Patterns, propensities and trends **Chapter 7** – Modelling mover/stayer characteristics across the life course **Chapter 8** – Modelling the duration of residence and plans for future residential mobility: A multilevel analysis **Chapter 9** – Modelling micro, meso and macro variations in origin to destination distance moved
II. To critically evaluate the existing sources of secondary data (aggregate and micro) for the analysis of population movement in Great Britain	**Chapter 3** – Population movement in GB: Sources of data
III. To review the current methodological approaches to the quantitative analysis of population movement at the macro, micro and multilevel scales in GB	**Chapter 4** – Population movement in GB: Methods for analysis
IV. To benchmark and validate the Acxiom Ltd. Research Opinion Poll, as a source of population migration microdata in GB, using official statistics (census, administrative and survey)	**Chapter 5** – Data validation: Descriptive-based benchmarking **Chapter 6** – Data validation: Model-based benchmarking
V. To determine and quantify any individual and/or contextual variations in residential mobility with an initial detailed focus on micro-level demographic, socio-economic and lifestyle influences before allowing for, and modelling, variance heterogeneity where possible in a multilevel framework	**Chapter 7**– Modelling mover/stayer characteristics across the life course **Chapter 8** – Modelling the duration of residence and plans for future residential mobility: A multilevel analysis **Chapter 9** – Modelling micro, meso and macro variations in origin to destination distance moved
VI. To summarise the aforementioned objectives with a focus on answering the overall research aims	**Chapter 10** – Conclusions

1.3 Thesis structure

As is shown in Table 1.1, aside from the review of the data (Chapter 3) and methodological approaches (Chapter 4), which seek to address objectives II and III respectively, the research objectives are met through a combination of reviews, discussions and analyses presented across multiple chapters in the thesis. The most obvious case is Objective I, where Chapters 2, 7, 8 and 9 are all relevant. Chapter 2 seeks to define the subject matter and introduce the basic theoretical and empirical context of residential movement in GB, presenting the key patterns, propensities and trends observed in the most recent official migration statistics, before introducing the major micro and macro contextual theories that are considered central to explaining them. However, Chapter 2 is designed for the purpose of providing a relatively brief theoretical introduction as well as some justification for the analysis that follows; indeed the latter analysis chapters (Chapters 7, 8 and 9) provide further, and more detailed, reviews and discussions of the major theories that underpin the relevant areas of particular interest. Thus, where possible, Chapter 2 provides clear signposting to the theoretical discussions of relevance found in the latter substantive chapters.

Chapter 3 is concerned with addressing Objective II and thus providing a detailed review of the current data landscape in GB. It focusses on the three main sources, census, administrative and social survey, and offers a discussion of the relative strengths and weaknesses of the micro and aggregate data reviewed. In addition, it introduces the ROP and provides a detailed discussion of its relative strengths and weaknesses, when compared to other current survey data sources, for the analysis of population mobility in GB. Finally, there is a brief discussion of the potential relevance of the ROP within the context of the Office for National Statistics (ONS) 'Beyond 2011' programme, where the ONS is actively engaged in opening up and linking together existing alternative sources of detailed geo-demographic data to complement the decennial population census.

Chapter 4 contributes to the meeting of Objective III which is to review the current methodological approaches to the quantitative analysis of population movement at the macro, micro and multilevel scales in GB. Following a review of the methods, the chapter presents a detailed step-by-step development of a multilevel model with reference to its practical application for the analysis of population movement. Thus, the chapter details the evolution of the regression model, from a simple single level model to a multilevel model with random intercepts, random slopes and cross-level interactions. Finally, a more complex non-hierarchical structure, the cross-classified multilevel model, is presented, and a practical example of its use for the analysis of population movement is given. A major aim of the chapter is to develop the argument that a multilevel modelling approach provides the best opportunity for maximising the utility of the detailed geo-referenced ROP data for addressing the overall project aims. Moreover, accepting that the multilevel modelling approach is the most suitable approach, this chapter also offers regular signposting to the relevant substantive chapters that employ the methods described within.

Using different descriptive and model-based approaches, and drawing on a wider range of population data and official statistics, Chapters 5 and 6 report the extensive cleaning, benchmarking and validation exercises that are necessary for evaluating the value of the ROP as a source of population migration microdata (Objective IV). Chapter 5 discusses the data management and cleaning approaches used before embarking on descriptive-based benchmarking of different aggregate, micro and spatial characteristics and patterns observed in the raw ROP samples. However, since the thesis is concerned with modelling variations in residential moves, Chapter 6 seeks to build on Chapter 5 by employing a practical approach to the validation of data from this commercial source for use in the model-based analysis of population mobility, particularly when little to no information on sample design is available. In particular it presents a method of sample reweighting, based on the use of auxiliary population data, designed with the purpose of adjusting the sampling distributions of key variables in the ROP and checking the effects of the sample adjustments on the estimated model coefficients. Where differences between weighted and unweighted models are small, we can be more confident that the model results drawn from the ROP data are reasonably robust to issues of nonresponse and sample bias. Moreover, a brief model-based benchmarking exercise against data drawn from the 2001 Census is also presented in an attempt to further assess the relative value of the ROP for use in the model-based analysis of population movement in GB.

Building on the positive findings of the previous chapter, the substantive analytical chapters follow. Chapter 7 develops from the validation models used in Chapter 6 in a way that allows for the micro-level analysis of variations in the associational behaviours and characteristics of movers and stayers across broad life-course stages, a required component of Objective V. Chapters 8 and 9 are also critical for addressing Objective V and both focus on the application of multilevel models described in Chapter 4. Chapter 8 aims to utilise a random intercepts and random slopes model in the analysis of an area of longstanding contention within the population migration literature, that of the functional form of the relationship between duration-of-residence and movement propensities. Chapter 9, on the other hand, shifts the analytical focus away from that of the previous chapters, where the emphasis was on the basic decision/ability to move or stay, and towards the variations in the distance of move, once the decision to move has been made. Consequently, Chapter 9 draws on multilevel theories detailing the importance of factors at both the area of origin and destination. Thus, it employs a cross-classified multilevel model, as discussed in Chapter 4, designed for the purpose of exploring simultaneous individual and place-based variations, operating at both the origin and the destination, in the postcode-to-postcode distance moved by migrants.

Finally, Chapter 10 seeks to synthesise the findings of the whole project and provide some overall conclusions (Objective VI). As part of this, the aforementioned research objectives are reviewed, with the major focus being the extent to which each has been achieved. Whilst the findings of the thesis provide some valuable contributions to the existing literature, there is undoubtedly scope for future improvements and research, and these provide a further focus in the final chapter.

Chapter 2
Population movement in GB: Patterns, propensities and trends

2.1 Introduction

This chapter is concerned with reviewing what are some of the key patterns, propensities and trends associated with residential movement in Great Britain (GB). To do this, it provides a broad empirical review of the most recent interregional characteristics and variations in the national migration system, employing a number of common measures of migration. Thereafter, it provides an introduction to the major micro and macro-contextual theories that are considered central for explaining these observed characteristics and variations. Indeed, residential mobility is seen to be a complex and multifaceted phenomenon with variations in the propensity to move, and distances moved, said to be driven by differential micro behaviours, characteristics and influences, as well as important macro-contextual influences thought to operate at, and across, different scales of geographic aggregation. A large part of the discussion here is taken up in more detail in the major substantive analytical chapters (Chapters 7-9) and therefore this chapter, where possible, seeks to provide regular signposting to the latter analysis chapter(s) that include the more detailed substantive reviews, discussions and insights.

2.2 Concepts, definitions and magnitude

Residential movement is something that will affect almost all of us at some point in our lifetime. Of the three demographic processes (i.e. fertility, mortality and migration) internal population movement usually has the largest impact on local area population size and composition (Bogue, 1969; Rees *et al.*, 2009; Poston and Bouvier, 2010). Indeed, beyond the simple change in numbers, the movement of individuals and/or families to new residential locations, whether within the same neighbourhood or to a different city or region, has the ability to transform the character and structure of populations, in some cases affecting real change to the social, cultural, physical and economic characteristics of an area. With this in mind, it is clear that the measurement, analysis and understanding of what drives the flows of different people between different places, is of huge importance. After all, as Rees *et al.* (2009: 1) make clear, such details are *"at the heart of decisions around policy development, resource allocation and service delivery, both nationally and locally"*.

Whilst population mobility can be understood as a particularly important social and economic phenomenon, it is perhaps surprising that one of the first issues to confront a new researcher to population mobility is how to conceptualise and define the topic of interest. Indeed, whilst we may all have a rough idea of what constitutes population mobility, there is no unique and universally agreed upon definition. However, at the most basic level, the distinction is often made between what are termed movers and migrants. Poston and Bouvier (2010: 168) detail the distinction as follows: *"migration differs from local movement in that a migrant leaves his/her community and moves to a new community. Such a move usually involves other changes: in one's school, job, church, doctor, dentist, library, pub, shopping centre, nightclub, automobile mechanic, and other institutional aspects of daily life"*. In contrast, a local movement is not expected to involve changes to the key institutions of the mover's daily life. Of course, whilst this provides us with a rather broad theoretical understanding of the distinction, its operationalization, whilst varying from study to study, is largely based on the use of predefined political/administrative geographical boundaries, wherein a move becomes a migration if the individual/household crosses said boundary lines (Frey, 2003). However, as Fielding (2012: 4) admits, the limits to his conceptualisation of what constitutes a migrant, as opposed to a mover, are not "*nice and sharp"* but instead rather "*fuzzy"*. Indeed on a similar note, Rees *et al.* (2009: 64) argue that *"it is not useful to define a threshold distance below which migration is labelled residential mobility and above which it is labelled 'proper' migration, because such a threshold is arbitrary"*. In reality, when it comes to an operational definition of what constitutes a migrant over a mover, the decision is largely influenced by the migration statistics at hand. For instance, as is discussed in Chapter 3, the individual records in the Patient Register Data System (PRDS) collect information on the National Health Service (NHS) patient as well as their home address at the postcode level. These details are updated annually with a migrant first being identified as a person whose postcode changes between consecutive patient register downloads. However, in terms of developing a practical definition, taking into account the geographies produced, a migrant is finally identified only as a person whose change in postcode takes them across either a former Health Authority (HA) or Local Authority (LA) boundary (Jefferies *et al.*, 2003), regardless of whether the move was in reality very short. With the arguments of Rees *et al*. in mind, for this study, where the data used are measured with detailed postcode identifiers at the individual level (see Chapter 3), the terms movement, mobility and migration are used interchangeably and very generally; wherein they describe the full continuum of distance moved and therefore cover both residential moves and migration.

In the year before the 2001 Census, that is the most recently published population census for which detailed migration data are available, approximately 10 per cent of the usually resident population moved address (Stillwell *et al.*, 2011). Interestingly, while there have been fluctuations in the propensity to migrate over the last 40-50 years, for instance, transition data show that roughly 8.5 per cent of the population were migrants in the year preceding the 1981 Census, this was lower than the number recorded by the 1971 Census where 10.5% of the population changed address (Rees and Stillwell, 1992: 29), the general rates of migration have remained broadly stable. Indeed, this is reflected more recently by time-series data from the National Health Service Central Register (NHSCR) covering the years 1998-2006, where the between LA district migration rates again show great stability with 2.43 million inter-district moves in 1998-99 as compared to 2.44 million moves in 2005-06, and fluctuations around the baseline of approximately 3 per cent across the entire eight year period (Duke-Williams and Stillwell, 2010). However, movements between LA districts account for only one third of all moves, and most of these are within the region of origin (Bailey and Livingston, 2005). Indeed, approximately 60 per cent of the 6 million residential moves recorded in the 2001 Census were moves which took place within the boundaries of a local authority (Stillwell *et al.*, 2011). As is well known, through empirical analysis and even the earliest of theoretical works in population mobility, most notably those of E.G. Ravenstein (1885), most moves tend to take place over particularly short distances (see Table 2.1). Indeed, based on the most up-to-date ONS estimates for interregional migration, using NHSCR/PRDS data for the year ending December 2010, less than 1.2 million people were estimated to have moved far enough to cross Government Office Region (GOR) borders (see Table 2.2).

Table 2.1. Distance moved by migrants within UK

Distance moved (km)	N	Share of within-UK moves
0-2	78,122	44.4
3-4	19,069	10.8
5-6	11,173	6.3
7-9	10,625	6.0
10-14	9,948	5.6
15-19	5,599	3.2
20-29	6,031	3.4
30-49	5,929	3.4
50-99	8,465	4.8
100-149	5,820	3.3
150 – 199	4,542	2.6
200 +	10,775	6.1
Total	176,098	100.0

Source: Bailey and Livingston (2005: 5). (N.B. Data from Census 2001 SARs: Population resident in private households – England, Wales and Scotland).

To provide some context of the macro patterns to migration in Great Britain and Northern Ireland, Tables 2.2 and 2.3 show the interregional migration statistics for the year ending December 2010, the year for which the most recent estimates are available. At first glance, it is clear that some regions like London, the South East and South West have in- and out-migrant counts that are far higher than those in Scotland and Northern Ireland. Moreover, according to the net migration rates, some areas made relatively large net migration gains (South West gained almost four persons through net migration for every 1,000 members of its population) while others witness relatively large net migration losses (London lost almost six persons through net migration for every 1,000 members of its population). Regarding the migration efficiency ratios, Northern Ireland had the highest negative efficiency ratio of -15.1 while the South West had the highest positive efficiency ratio of 8.7. The efficiency ratio for Wales was the lowest in the UK, with a ratio of 2.5. Strictly speaking, areas with high positive efficiency ratios are areas where most migrants have moved in and few have moved out, in contrast areas with high negative efficiency ratios are areas where the majority of migrants have moved out and very few have moved in. Areas with low efficiency ratios are considered ‘inefficient’, that is, there are similar numbers of migrants moving in and out. Thus for Wales, migration is inefficient, there were relatively large numbers of people coming and going in 2010 but due to their similar numbers there was very little in terms of net migration gain/loss.

Table 2.2. Interregional migration in the UK

	Population ('ooo)	In-migrants ('ooo)	Out-migrants ('ooo)	Migrant turnover ('ooo)	Net migrants ('ooo)
North East	2,606.6	37.0	39.2	76.2	-2.2
North West	6,935.7	96.0	101.6	197.6	-5.6
Yorkshire and The Humber	5,301.3	88.8	94.4	183.2	-5.7
East Midlands	4,481.4	99.5	94.5	194.0	5.0
West Midlands	5,455.2	89.1	97.3	186.4	-8.2
East	5,831.8	132.4	118.0	250.3	14.4
London	7,825.2	176.1	220.8	396.9	-44.6
South East	8,523.1	209.4	187.2	396.6	22.2
South West	5,273.7	124.4	104.6	228.9	19.8
Wales	3,006.4	50.8	48.3	99.1	2.5
Scotland	5,222.1	44.3	39.0	83.3	5.3
Northern Ireland	1,799.4	8.2	11.0	19.2	-2.9
UK	62,261.9	1,155.9	1,155.9	2311.8	0.0

Source: Rates based on NHSCR (interregional migration data for year ending December 2010) and ONS (mid-year population estimates 2010).

However, whilst these broad empirical descriptions of aggregate data are useful for providing an account of the migration system in which we live, detailing areas of population increase, change and decline, they leave a great many questions to the imagination. For instance, what types of people are doing the moving in the first place? And, for those people who are moving, what sorts of places are they leaving, what distances are they travelling, and where are they going? Moreover, do the answers to these questions differ depending on the differences between people and the differences between the contexts in which they live? These sorts of questions remain largely unanswered, and it is only through the use of suitable data, theory and methods that such questions can begin to be addressed. With this in mind, the following sections of this chapter are focussed on providing a broad theoretical background and context that will be useful for informing the more detailed discussions and analyses aimed at addressing the questions above, and in particular, the overall project aims and research questions set out in Chapter 1.

Table 2.3. Interregional migration in the UK: Migration measures

	In-migrant rate (per 1,000)	Out-migrant rate (per 1,000)	Migration turnover rate (per 1,000)	Net migrant rate (per 1,000)	Migration efficiency ratio
North East	14.2	15.0	29.2	-0.8	-2.9
North West	13.8	14.6	28.5	-0.8	-2.8
Yorkshire and The Humber	16.7	17.8	34.6	-1.1	-3.1
East Midlands	22.2	21.1	43.3	1.1	2.6
West Midlands	16.3	17.8	34.2	-1.5	-4.4
East	22.7	20.2	42.9	2.5	5.8
London	22.5	28.2	50.7	-5.7	-11.2
South East	24.6	22.0	46.5	2.6	5.6
South West	23.6	19.8	43.4	3.8	8.7
Wales	16.9	16.1	33.0	0.8	2.5
Scotland	8.5	7.5	16.0	1.0	6.3
Northern Ireland	4.5	6.1	10.7	-1.6	-15.1
UK	18.6	18.6	37.1	0.0	0.0

Source: Rates based on NHSCR (interregional migration data for year ending December 2010) and ONS (mid-year population estimates 2010).

2.3 Micro and place based theories of residential movement in GB

The earliest contributions to the analysis of the decision-making processes and patterns of population movement can be dated right back to seminal works by Thomas (1938) and Rossi (1955). The decision to change residence is widely considered to be a utility-maximising behaviour, performed by individuals, either independently or collectively within households, reacting to disequilibrium between the current residential environment and a perceived environment elsewhere (Bartel, 1979; Clark and Dieleman, 1996; Clark, 2013). Thus the decision to move is largely driven by the extent to which the welfare of the individual/household can be maximised, which itself requires the relevant actors to weigh up the expected costs and benefits of moving to an alternative location as oppose to staying put at their current location. However, the factors behind the motivation to move are known to vary greatly depending on personal situation and stage in the life course. Since the pioneering work of Rossi (1955), residential mobility has been theorised to be strongly associated with the transitions between the different stages of the family life course, transitions that, whilst increasingly diverse in their timing and sequence, remain largely observable in the age-mobility trends of the most recent population census (Figure 2.1). Indeed, whilst there is no biological mechanism for the influence of age over the propensity to move, it does act as a rather consistent proxy for timing of certain life course transitions and events, which are themselves associated with shifts in household structure (Feijten and Mulder, 2002; Boyle *et al*., 2008; Mulder and Wagner, 2010); housing tenure (Boyle, 1993), and income, occupational and educational attainment (Fielding, 1992; 2007).

For instance, as is shown in Figure 2.1, migration rates and shares are highest for those who are in the 18-25 age brackets, where moves in these groups are commonly associated with transitions into adulthood; where the high propensity of movement is motivated by the pursuit of early career educational and occupational opportunities, with the majority generally transitioning from school to first employment or university, and/or first employment following university (Champion, 2005a; Smith, 2009). Following this, the subsequent age groups reflect a sharp decline in mobility rates, and are associated with transitions into relative career stability, family formation, child rearing and increased levels of homeownership, all of which can be expected to encourage residential stability and lower mobility propensities (Fielding, 2012). The decline is reduced somewhat for those aged 45-64, where the transition from parenthood to ‘empty nesting’ can be thought to prompt a re-evaluation of the residential environment and, for some, a change of residence (Wulff *et al*. 2010). Moreover, transitions from work into retirement and exit from the labour market emerge, which again often lead to changes in residential preferences, needs and desires (Fielding, 2012). Finally, the mobility rate recovers somewhat with raised propensities for those in the eldest age groups, commonly linked to the desire/need for closer proximity to family members and social/health services (Evandrou *et al*., 2010).

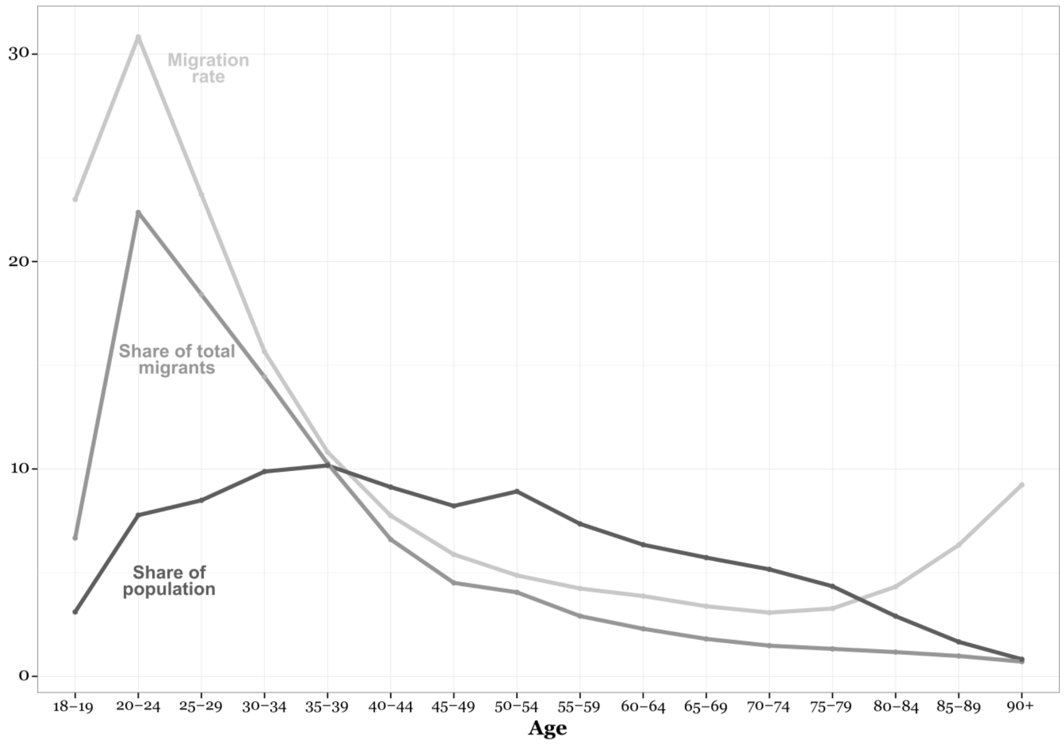


Figure. 2.1. Migration rates, migration shares and population distribution by age. Source: Census 2001 SMS.

However, whilst these very general and normative understandings of the life-course, that is, composed of certain follow-on sequential stages, can be useful for providing some understanding of the commonalities to the patterning of propensities of residential mobility in GB, it provides only a rather blunt interpretation of what is in reality an increasingly complex and dynamic phenomenon. Indeed, it is certainly the case that the timings and order of transitions and events are becoming increasingly diverse (for example, the delaying and/or avoidance of childbearing for occupational reasons), and in some cases recurring (for instance, returning to education in later life) (Clark, 2013). Thus, there is a growing acceptance that whilst there are general patterns to the life course, analysts should be careful in acknowledging the fact that each person has experienced their own unique sequence of events and their own complex and interrelated household, labour force, education and housing careers, all of which embody the key mitigating factors known to inform our mobility behaviours and outcomes (Mulder, 1993; Elder, 1994; Clark and Dieleman, 1996; Dykstra and van Wissen, 1999; Bailey, 2009). Indeed, whether expected or otherwise, life course events and disruptions emerge which can greatly alter our residential preferences, in some cases exacerbating the residential mismatch, and thus potentially lead to a significant increase/decrease in the probability of changing residence. Examples of commonly cited events said to greatly influence residential mobility include: unemployment (Clark and Davies Withers, 1999; Böheim and Taylor, 2002; Fielding, 2012), pregnancy and the birth of children (Kulu and Steel, 2013), union dissolution (Mulder and Wagner, 2010), marriage (Mulder and Wagner, 1993) and widowhood (Chevan, 2005; Evandrou *et al.*, 2010). Indeed, these are all events which can lead to residential adjustments allowing individuals/households to bring their location into equilibrium with their housing, family and occupational needs. Chapter 7 seeks to uncover the variations that occur between different types of people in the propensity to move residence, paying particular attention to how the relationships between different explanatory factors vary according to the broad stage in the life course.

However, whilst this introductory discussion of the literature has so far focussed on individual and household influences on residential satisfaction and mobility decision making, the theoretical literature makes clear the relevance of additional contextual place-based influences (Lee, 1966; Massey, 1990; Kearns and Parkinson, 2001; Sampson *et al.*, 2002; Courgeau and Lelievre, 2006; Hedman, 2011). Indeed, thinking in terms of the micro-level behavioural model, one's evaluation of the residential environment undeniably stretches beyond the household, incorporating attributes of the neighbourhood, locality, urban district and region. As of late, the neighbourhood context has received particular attention in studies of residential mobility (Rabe and Taylor, 2010; Hedman, 2011; Hedman *et al.*, 2011; van Ham *et al.*, 2014). Indeed, characteristics such as the relative deprivation and socioeconomic status of the neighbourhood, the demographic and housing profile of the neighbourhood, and the relative stability of the neighbourhood population have all been the subject of empirical and theoretical interrogation with regards their potential influence on individual residential evaluations and mobility outcomes. For instance, greater levels of deprivation have been linked to lower levels of social cohesion and neighbourhood desirability (Taylor *et al.*, 2010; Sturgis *et al.*, 2013) and, interlinked with this, an increase in residential stress which can, in some cases, lead to the outflow of residents to other areas. However, residential mobility is known to be a highly selective phenomenon that often works to filter people into neighbourhoods depending on their personal characteristics. For instance, whilst living in a neighbourhood with high levels of deprivation and crime and low levels of social cohesion and collective efficacy would be unattractive for the vast majority of the population, the opportunity to act on residential dissatisfaction varies greatly depending on the situation of the person involved. For instance, as empirical analysis of the 2001 Census shows, neighbourhood sorting according to income is prevalent in GB, wherein individuals with access to greater financial resources are observed to be significantly more likely to move away from areas of increasing deprivation compared to individuals from lower socio-economic brackets (Bailey and Livingston, 2008). Indeed, where only those with sufficiently high incomes are moving out, the level of neighbourhood deprivation can only be expected to increase. Beyond income and neighbourhood deprivation, the interaction between individual ethnic background and the degree of wider neighbourhood ethnic heterogeneity is a common feature in the literature. Whether for reasons that are positive (importance of access to cultural/social institutions and amenities) or negative (reacting to racism or limited housing/occupational opportunities), the co-location of ethnic minority groups into certain neighbourhoods is a common feature in the urban landscape of GB (Bailey and Livingston, 2005; Simpson and Finney, 2009). Consequently, it is often the case that the more diverse neighbourhoods of the country are the more attractive for members of certain ethnic minority groups. However, in reverse, greater levels of ethnic heterogeneity have been associated with greater levels of residential dissatisfaction and thus adjustment mobility, particularly for the white majority population. However, the vast majority of the literature here, where the relationship is often defined through the ‘white flight’ hypothesis (Ellen, 2000; Crowder, 2000), relates specifically to the unique (space-time) context of the US. Studies of the relationship in GB have questioned the relevance of ethnic heterogeneity, suggesting instead that any significance found for greater diversity is simply the result of the failure to account for important confounding factors, most notably the level of wider neighbourhood deprivation (Harris, 1999; Sturgis *et al.*, 2013). The instability of the neighbourhood population, the intensity of movement into and out of the neighbourhood, is a further characteristic that has received recent attention. Again, as with deprivation, high levels of population turnover and change are linked to a number of negative neighbourhood externalities including greater fears and occurrences of violence and crime (Sampson *et al.*, 1997), and a generally lower residential attractiveness (Andersson and Bråmå, 2004). Moreover, the importance of neighbourhood population (in)stability has also been noted in terms of its relevance for mediating the ability of residents, new and old alike, to form and maintain meaningful place based social ties and networks and the subsequent residential attachments that can supress desires to move (Hedman, 2011; Hedman *et al.*, 2011). It is this latter point that is taken up in Chapter 8, where the relationship between individual duration-of-residence and plans for future mobility is analysed in a way which makes it possible to observe whether higher/lower levels of neighbourhood population instability mediate this relationship.

Yet operating at levels beyond the neighbourhood are a wide variety of additional macro level influences linked, for instance, to the underlying geography of wealth and power, the associated spatial division of labour, the degree of medium- and short-term regional economic robustness, and the differing lifestyle and environmental opportunities afforded for in different macro-geographic areas (Fielding, 1992; Massey, 1995; Champion, 2008; Fielding, 2012). All of these influences can be expected to be of relevance for informing the decision to move, but are of particular importance when the decision to move is motivated by particular factors, most notably factors pertaining to education, the labour market and/or the environment. Moreover, since the original work of Lee (1966), the differential attractiveness of different origin and destination contexts, as measured in terms of push and pull factors, have been fundamental in describing the complex macro migratory system and the patterns, propensities and trends to the migration flows that operate within. The pivotal role of London as an 'escalator region' is a good case in point of how macro-regional differences inform mobility outcomes (Fielding, 1992; Champion, 2008; Duke-Williams and Stillwell, 2010). As can be observed in Tables 2.2 and 2.3, London is unique in the migratory system; the capital tends to attract a large number of young and usually well-educated adults from all parts of the country, largely due to its ability to engender rapid social promotion (Fielding, 1992), yet it loses by far the most people to internal movements out of the city. Despite this, strong natural increase (births) and significant net immigration have maintained a growing total population (Champion, 2008). Indeed, this large net loss has been largely attributed to the desire of people in the latter stages of their career or those at, or close to, retirement seeking pastures new; where they have little more to gain from the ‘escalator effect’, they are instead motivated by opportunities further afield, where living costs are lower and the perceived quality of life higher (for instance returning to their region of origin or to amenity rich rural and coastal environments) (Fielding, 1992; 2007; Champion, 2005b). This mass movement away from London to the more suburban and rural regions has been observed in more recent analysis by Duke-Williams and Stillwell (2010), and is consistent with a far wider phenomenon of urban-rural shift and counterurbanisation in GB (Rees and Stillwell, 1992; Champion, 2005b; Dennett and Stillwell, 2008). Again, these macro-geographic influences are the particular focus of Chapter 9, where the differences between macro regions are used to explain the variations in individual origin-destination distance travelled.

Of course, whilst mobility preferences and behaviours can be expected to vary according to a great range of individual, household and place based processes, characteristics and factors, the final decision/ability to move and, following this, the distance of move, is subject to personal individual/household situations and wider social and economic structural constraints (Hägerstrand, 1975; Boyle, 1995; Massey, 1995; King, 2012; Fielding, 2012). Whilst the example of the selective nature of neighbourhood deprivation on individual residential mobility outcomes has been noted, a further example of selectivity emerges when we stratify mobility rates by occupational class. Indeed, as previous work by Fielding (1995; 2012) has demonstrated, there is a clear hierarchy to mobility rates between the occupational classes, with those at one end of the spectrum, the professional and managerial classes, being roughly three to four times more mobile than those at the other end, the blue collar classes. For the professional and managerial classes, upwards social mobility is often accompanied by a certain degree of spatial mobility. Indeed, individuals with higher educational attainment and associated occupations are known to search over far wider labour markets and have a much greater degree of spatial flexibility associated with the pursuit of career advancement and progression (van Ham *et al.*, 2001; Fielding, 2007; 2012). Conversely, for those in the traditional blue collar working classes, migration rates are observed to be very low. As Fielding (2012) suggests, individuals working in blue collar occupations are often employed in areas where industrial skills and workplace reputations are typically sector and locality-specific and have family and friendship networks that are also particularly spatially restricted. However, the disparity in the mobility rates between these occupational classes uncovers one of the more striking dichotomies of population mobility and its selective nature, namely that between those who *need* to migrate and those who *actually* migrate. One could argue that the professional and managerial classes need to migrate the least, given their relatively well-paid and secure employment. At the same time, it could be argued that those from the blue collar working classes need to migrate the most, due to their relatively low-paid and insecure employment (Fielding, 2007). However, the differing ability of the two groups to act on their mobility needs is the determining factor behind their mobility outcomes and thus their differing mobility rates. Indeed, those with greater educational and occupational attainment are typically those with fewer locality specific ties (occupational and social), those who have more information on opportunities elsewhere, and those with greater access to the critical financial resources that enable a successful migration in the first place.

Further to the basic decision and ability to enact a change of residence, the selective nature of mobility is known to grow in significance as the distance of the move increases. As was detailed in the early work of Lee (1966), the proposed distance between the origin and destination has itself been defined as an important intervening obstacle to residential mobility, with longer distances associated with a variety of increasing restrictions and costs. These include the relinquishing of ties to locality-specific social networks and amenities (Brown, 2002); the increasing likelihood of a change in employment and/or the workplace (Owen, 1992); the financial costs and implications of the actual move itself and the associated costs of the search (Flowerdew, 1976); and, as mentioned above, the requirement of information on opportunities in places far afield (Flowerdew, 1982). Thus, given the many obstacles that longer-distance moves engender, the ability to move long distances is highly dependent on the situation of the respective individual/household. That is, if a long-distance move is indeed the desired outcome, only those individuals/households with sufficient resources and motivation will ultimately achieve their migration to a destination further afield. Chapter 9 provides a more detailed review of the frictional effect of distance on population mobility and explores both micro and contextual variations in migrant origin to destination distances moved.

2.5 Summary and conclusions

This chapter has sought to identify some of the key patterns, propensities and trends associated with residential movement in contemporary Britain. One thing that is immediately apparent from the discussion, is that population mobility is not a simple and consistent phenomenon. Instead, mobility is characterised by marked differences, at all levels, from the various patterns, propensities and trends observed at the interregional level, through to the complex micro processes of residential evaluation and satisfaction, differential selectivity and resultant mobility behaviours and outcomes at the individual and household level. Indeed, as is suggested above, the apparent importance of both the micro and the macro, and the interaction between the two, is widely discussed and supported in the theoretical literature. However, in practice, there is little empirical work recognising the simultaneous effects of different micro processes and contextual effects on movement behaviours and outcomes. Indeed, a large part of this is expected to be the result of what has been a longstanding dearth in suitably detailed microdata with sufficient sample size, geographic coverage and spatial detail. The next chapter provides a thorough review of the current data landscape in GB and, in doing so, introduces a new source of cross-sectional commercial microdata that contains many of the attributes deemed necessary for a detailed analysis of individual *and* place variations in residential mobility in GB.

Chapter 3
Population movement in GB: Sources of data

3.1 Introduction

There is no single comprehensive source of data that can cover all the disparate requirements of those interested in population mobility. Instead, researchers find themselves in a situation where, in order to satisfy their requirements, they must utilise a variety of sources characterised by sharp contrasts in coverage, detail (of both data and geography) and accuracy. This chapter is concerned with providing a review of the key sources of population mobility data, highlighting their general attributes as well as their relative strengths and weaknesses for the specific analysis of population mobility. As a result, the chapter is structured into three main data source subsections: 3.2 on population censuses; 3.3 on administrative sources; and 3.4 on social survey sources. The first two subsections focus on the value and usefulness of data available from the census and a number of selected administrative sources. The final subsection is more comprehensive than the previous two and is focussed on the extensive list of social survey sources that are available in GB today. The rationale to focus more heavily on this latter subsection is twofold; first, this project is heavily based on the application of survey data for mobility analysis, and second, survey data are by far the most varied, ever-changing, and, with respect to population mobility analysis, understudied of all data types. Following this, subsection 3.5 introduces the ROP and provides a detailed discussion of its relative strengths and weaknesses, when compared to other current survey data sources, for the analysis of population movement in GB. Finally, there is a brief discussion on the potential relevance of the ROP within the context of the Office for National Statistics (ONS) ‘Beyond 2011’ programme (Subsection 3.6), after which some conclusions are presented.

3.2 Population censuses

Population censuses in the UK have been taken decennially since 1801, with the exception of 1941, and aim for a complete enumeration of the population. Indeed, given their near comprehensive coverage, their relative reliability and the high detail of their demographic and socioeconomic variables, censuses are currently considered as the optimum points of reference for those interested in population statistics and in small area demographic analysis in particular (Raymer *et al*., 2012).

For those with a particular interest in population movement, censuses provide a variety of data products including: the main census tables, the Special Migration Statistics (SMS), commissioned tables; the Samples of Anonymised Records (SARs) and the Longitudinal Studies (LSs). These products derive their migration data from a single question that has been asked every year since the 1961 Census (with occasional slight variation), namely 'What was your usual address one year ago?'. A detailed discussion on the role of the migration question can be found in Duke-Williams (2011). Through this question, the census identifies a migrant as any UK resident who had a different address in the previous year, regardless of the distance of the move (Champion *et al*., 1998).

3.2.1 Main Census Tables

The main census tables produced by the Office for National Statistics (ONS) from the 2001 Census include the Census Key Statistics (KS), Standard Tables (ST), Theme Tables (TT) and Census Area Statistics (CAS). However, only a handful of tables derived from these sources produce migration statistics, including: KS24 – Migration (All people); ST008 – Resident type by age and sex and migration; ST009 – Age of household reference person and number of dependent children by migration of households; ST010 – Household composition by migration of households; TT033 – Migration (people): All people in the area and those who have moved from the area in the past year, within the UK. The KS tables encompass a limited number of simple univariate tables and are useful in providing a summary and overview of the main topics of the 2001 Census at the smallest level of geography, the output area (OA) level. The ST data sets provide more detailed information and include a large number of cross-tabulations of variables measured in the 2001 Census; however, they are only produced at the ward level in England, Wales and Northern Ireland, and postcode sector level in Scotland. Finally, the CAS data are roughly equivalent to those in the ST data sets but are available at the OA scale. With that said, in order to protect the confidentiality of personal information, the information provided is less detailed than in the ward level ST. Unfortunately for the analysis of population mobility, the main census tables that contain counts of internal movements only allow for detailed breakdowns at either the origin or the destination end of the move, but not both. This makes it impossible to identify origin-destination unit flows (Dennett *et al.*, 2007). However, the census does provide alternative data products that allow for origin-destination flow data to be extracted, one of which is the Special Migration Statistics.

3.2.3 Special Migration Statistics (SMS)

The 2001 Census SMS tables are sorted according to three geographical levels (Table 3.1). Level 1 tables (10 tables totalling 996 cells/variables) contain flows between what are termed 'districts', which comprise a variety of local government authorities including unitary authorities (UAs), local authority districts (LADs), metropolitan districts (MDs) and London boroughs (LBs) in England & Wales, and council areas (CAs) in Scotland, as well as parliamentary constituencies in Northern Ireland (Dennett and Stillwell, 2010). Level 2 tables (5 tables totalling 96 cells/variables) involve flows between 'interaction wards', wards specially designed to minimise the impact of regular electoral ward boundary changes (census area statistics wards in England, Wales and Northern Ireland, and standard table wards in Scotland) (Stillwell and Duke-Williams, 2007). A single level 3 table (containing 12 cells/variables) is available at output area (OA) level which, with roughly 125 households per OA, represents the smallest geographical unit for which 2001 Census data are available (Martin, 2002a; 2002b). As mentioned previously, it is possible, through the ability to cross-tabulate migrants by place of origin (address one year previously) and migrants by place of usual residence (current address), to extract counts for origin to destination migration flow matrices at each SMS level, using the WICID interface via the UK Data Service.

Table 3.1. Tables and cells/ variables for the 2001 SMS

	Level 1 'Districts'	Level 2 'Wards'	Level 3 'OA'
Tables	10	5	1
Cells/ variables	996	96	12

Clearly each level presents different possibilities to the researcher, with level 1 offering more accurate and detailed data but poorer geographical detail than level 3 and *vice versa*. However, one particular problem with the 2001 Census data is that associated with the small cell adjustment method (SCAM) used to adjust small flows with values of 1 or 2 in order to retain individual confidentiality and thus avoid the risk of disclosure. While the exact details have not been made public by the ONS, examination by Stillwell and Duke-Williams (2007) and Duke-Williams and Stillwell (2007) suggests that values of 1 or 2 have been adjusted to values of 0 and 3, thus removing any of these values from the SMS tables. Given the size of flows taking place at different spatial scales, the counts at level 1 will be more robust than those at level 2 and level 3. As Dennett and Stillwell (2010: 519) assert, *"[u]sing a larger primary unit of analysis reduces the chances of small values appearing in the cells of the data tables and thus reduces the effect of SCAM on the data"*. The 'damage' caused by SCAM has meant that the 2001 origin-destination migration matrices are virtually unusable at OA level and, whilst the data are not yet available, it is pleasing to note that ONS have abandoned this form of post-tabular adjustment for the 2011 Census in favour of pre-tabular record swapping, although this may mean more restricted access to multivariate tables particularly at ward and OA levels (Traynor, 2011). However, it should be noted that for the 2001 Census the method was not applied to flows with Scottish destinations.

3.2.4 Commissioned Tables

The ONS produces commissioned outputs on demand for specific cross-tabulations that are not available through the published standard results. These commissioned tables are currently available for the 1981, 1991 and 2001 Censuses; however, each commissioned output incurs a charge for the staff and material costs associated with its production and supply. That said, once a table has been commissioned and paid for, it is listed on the ONS website and available to all uses free of charge. However, to access the tables, a request must be made to the Census Customer Services. Commissioned outputs specifically for Scotland and Northern Ireland must be requested from their respective national statistics authorities (NSAs) although UK-wide requests can be made to any/one of ONS, GROS and NISRA. A list of all commissioned outputs for the 2001 Census is available on the ONS website via the data and product catalogue page. As with the SMS, commissioned data for England & Wales and Northern Ireland are subject to SCAM.

3.2.5 Samples of Anonymised Records (SARs)

A further set of census data products from which mobility data can be obtained is the Sample of Anonymised Records (SARs). Unlike the aggregate data tabulations of the SMS, the SARs allow for the possibility to cross-tabulate individual level migration data with the other demographic and socioeconomic variables included in the census (Norman and Boyle, 2010). However, given the relatively small size of the samples of these micro data, it is only possible to obtain individual origin to destination flow data for a very crude geography, so as to avoid any risk of disclosure. In total, five SARs were produced from the 2001 UK Census including: the Individual SAR (Licensed) (3 per cent sample); the Household SAR (Licensed) (1 per cent sample); the Individual Controlled Access Microdata Sample (Individual CAMS) (1 per cent sample); the Household Controlled Access Microdata Sample (Household CAMS); and the Small Area Microdata (SAM) (5 per cent sample).

As is the case with all of the sources reviewed here, for the analysis of population movement the SARs have both strengths and weaknesses. Indeed, as Bailey and Livingston (2005) have noted, the size and detail of the individual level data allows researchers to explore numerous aspects of migration, but with a specific focus on the variations between relatively small population sub-groups. Bailey and Livingston offer the examples of lone parents, couples with children, minority ethnic groups and regional differences. However, in terms of generating and analysing directional flows, it is only the CAMS (particularly the Individual CAMS) and the SAM that are of major value. For the 2001 Census, the individual CAMS has a LAD based geography for both migrant origins and destinations, thus providing similar spatial detail to the 2001 SMS Level 1. As is evident in Table 3.2, the Household CAMS destination is given at LAD level but the origin is limited to a categorical variable indicating whether the migrant moved from the same district, or not. For the SAM, the destination is also available at LAD level, but unfortunately the origin remains course, at the level of the Government Office Region (GOR). However, the 2001 SAM file does have the advantage of a 5 per cent sample which, taking account of the destination geography, allows for a very detailed look at the characteristics of in-migrants according to a great number of census variables (Dennett *et al.*, 2007). At the time of writing (June 2014), the 2011 Individual SAR (safequarded) remains unavailable whilst the ONS establishes which variables and at what geographic detail the various microdata can be made available. However, based on previous censuses, Norman and Boyle (2010) provide a useful summary of how census microdata (from the SARs and LSs) have been used in research on topics including migration, health and deprivation.

Table 3.2. 2001 SAR comparison table

	Individual SAR	Household SAR	Individual CAMS	Household CAMS	SAM
Sample Coverage	UK	E & W	UK	UK	UK
Sample Size (approx.)	3%	1%	3%	1%	5%
Migrant Origin	GOR (16)	Migration categories	LAD (and SL, NI eq.)	LAD (and SL, NI eq.)	GOR (16)
Migrant Destination	GOR (13)	n/a	LAD (and SL, NI eq.)	LAD (and SL, NI eq.)	LAD (and SL, NI eq.)
Access	End user licence	Special licence	Safe setting	Safe setting	End user licence

N.B. eq.= equivalent.

3.2.6 Longitudinal Studies (LSs)

There are three census based cohort studies in the UK, the ONS Longitudinal Study of England and Wales (ONS-LS), the Scottish Longitudinal Study (SLS), and the Northern Ireland Longitudinal Study (NILS). While they all purport to do the same thing, they do differ in a number of ways. The England and Wales LS consists of an approximate 1 per cent sample drawn from the census of all individuals resident in England and Wales who are born on one of four dates each year (undisclosed for reasons of confidentiality). The information that is used in the LS is based on data collected from the members' census forms as well as linked vital registration systems (e.g. births to female sample members) and cancer registrations. In addition to the sampled individuals, details on the other members of their household are also recorded. However, these household members are not tracked in the following census unless they remain a part of the sample member's household. Additionally, since the first sample was taken in 1971, the LS has included all new births and immigrants who share the relevant four birth dates, and all sample members who die or have emigrated have been removed (ONS, 2003a). For the 2001 Census, there were 539,665 individuals recorded in the LS (ONS, 2003b).

Importantly in the context of research on internal population movement, the large sample allows for full geographical coverage of England and Wales. The potential of the LS for geographical research is discussed by Dale *et al*. (1993) and by Hamnett and Randolph (1987). Given the full geographical coverage and the ability to examine migrations over 10, 20 and 30 years (as opposed to the restriction of a single year for the other census sources) the LS has great potential for the analysis of migration flows over long periods of time (e.g. Ekinsmyth, 1996) and is particularly useful for the analysis of the relationship between migration and longer-term social change (e.g. Fielding, 1989).

The time period covered by the Scottish LS (SLS) is much shorter, having started in 1991, and subsequently incorporates data from only two censuses. The SLS was designed to provide a 5.5 per cent representative sample of the Scottish population based on 20 ‘semi-random’ dates of birth (Hattersley and Boyle, 2007). In order that the SLS has the potential to be compatible with the England and Wales LS, four of the 20 dates of birth match those included in the LS. Again, as with the LS, the SLS draws its data from members’ census forms as well as vital events data (births, deaths, marriages), NHSCR data (migration in or out of Scotland) and NHS data (cancer registrations and hospital discharges – although while these data are available, they are not held as part of the SLS database and are instead linked as required for the specific study) (Hattersley and Boyle, 2007).

The Northern Ireland LS (NILS), run by Northern Ireland Statistics and Research Agency (NISRA), began in 2008 and currently only contains census data from the 2001 Census. However, unlike the ONS-LS and the SLS, the NILS has a representative sample based on 104 selected dates of birth which provides a vastly superior core membership of around 28 per cent of the population. An important distinction to highlight in the NILS *vis-à-vis* the ONS-LS and SLS is that the NILS sample is based on health card registrations that are then linked to the census. The NILS database contains the basic demographic data (age, sex, and home postcode) from the centralised Northern Ireland Health Card registration system and then links this to the census data where the majority of additional cohort attributes are gathered. According to O’Reilly *et al*. (2012: 635), this process provides the NILS with an advantage as *"health card registrations are arguably a more robust source for the basic demographic data as the census data are captured only once and in 2001 depended on accurate interpretation of electronically scanned census forms"*. Beyond this, the NILS is linked to additional administrative data including vital events via the General Register Office for Northern Ireland (births, deaths and marriages), migration events via the Health Card registrations, contextual and area-based data about members' households via the Land and Property Services, as well as various other health and social care data (hospital and laboratory systems, screening services, prescribing data, and uptake of dental services) (O'Reilly *et al*., 2012).

3.3 Administrative sources

For the years between the decennial censuses, there is no single system from which internal migration within the UK is recorded/captured. Instead, the ONS employs a combination of proxy administrative sources from which the estimates are drawn and used subsequently in the estimation of mid-year populations. Two estimates are produced by the ONS based on administrative sources (Rees *et al*., 2009). The first is based on National Health Service Central Register (NHSCR) data and captures events of migration between health authorities (HAs). The second is based on the Patient Register Data System (PRDS) which measures transitions in the NHS patient data. The PRDS data are available from 1999 and are produced at the local authority level (with the potential to aggregate into counties) which means that they are consistent with the geographical units used in the 2001 and 2011 Censuses respectively (Smith *et al*., 2010; Raymer *et al*., 2012). An additional administrative data source, the Higher Education Statistics Agency (HESA), is used to adjust the estimates and control for students in higher education, a subgroup with unique and particularly complex flows (ONS, 2010a). There are comparable registration systems for patients used to estimate internal migration flows in Scotland (Community Health Index data, CHI) and Northern Ireland (Central Health Index, NI-CHI). Beyond this, the cross-border moves to/from England and Wales and Scotland and Northern Ireland are provided to ONS for inclusion in the published estimates by the respective national statistics agencies (ONS, 2011a). Cross-border flows between districts comprising the home countries remain unavailable to the public, though recent work by Lomax *et al.* (2013) has provided the first estimates of inter-censal annual migration flows at the district level across the UK. Generally speaking, administrative data sources are most useful for those interested in measuring more up-to-date migration flows and counts; however, at the same time, they are limited by a distinct lack of demographic and socio-economic detail, and in some cases are only relevant for the analysis of distinct sub-sections of society such as school children (School Census) and HE students (HESA).

3.3.1 National Health Service Central Register (NHSCR)

The NHSCR represents a database wherein each record in the register contains an NHS number, name, age, sex, date of birth and date of registration with the Health Authority (Rees and Boden, 2006: 66). These data are supplied weekly by the NHSCR to the ONS for use in a quarterly rolling year estimate of internal migration. Within the NHSCR data source, a migrant is defined as a person who (re-)registers with a GP in a different former HA/Area Health Board (HB) in Scotland and Northern Ireland (i.e. data only include moves between former HAs). The NHSCR provides counts of moves where a record is made of each movement event and thus all moves within a year are included in the internal migration estimates, i.e. a person who moves from area A to B and then B to C within a year will be counted twice (Jefferies *et al*., 2003; Rees *et al*., 2009).

For estimates of cross-border flows between England and Wales, Scotland, and Northern Ireland, the NHSCR draws on data held by the National Records of Scotland (NRS) and NISRA. For example, migrants moving out of England and Wales to Scotland and Northern Ireland are identified initially where the NHS number of the previous year's patient register is not found on the current year's patient register. The NHS number is then matched to data from NRS and NISRA, and where the number is found in this data, a migrant to Scotland or Northern Ireland is identified and the destination area is recorded (ONS, 2011b).

In terms of limitations, the NHSCR's reliance on the now defunct HAs is perhaps the most problematic. These former HAs no longer exist as administrative entities but continue to be used due to technical constraints of the NHSCR processing system (ONS, 2011a: 2). However, as discussed below, this limitation is in part answered by the PRDS, which allows for the production of estimates at geographical units that are comparable to the census output units at level 1 (i.e. districts).

3.3.2 Patient Register Data System (PRDS)

The individual records in the PRDS collect information on the NHS patient as well as their home address at the postcode level, thus providing far greater geographical detail than the NHSCR. These details are updated annually with a migrant being defined as a *"person whose postcode changes between consecutive patient register downloads"* (Jefferies *et al*., 2003: 5). However, in terms of practical definitions, taking into account of the geographies produced, a migrant is identified only as a person whose change in postcode takes them across either a former HA or LA boundary (Jefferies *et al*., 2003). Clearly, as Rees *et al*. (2009: 113) assert *"[p]atient register data differs from NHSCR data in that it records the transition between the area of residence at the beginning of the annual period and the area of residence at the end, rather than every movement made over a year"*. With the PRDS data being transition data (i.e. change in area of registration), it is closer conceptually to the outputs produced from the census.

An additional issue for migration estimates based on PRDS data is the fact that the data fails to capture a number of migrations by certain groups of people. Indeed, the PRDS data cannot capture the movement of migrants who were not registered with a doctor in one of two consecutive years, but who moved during the year (Jefferies *et al*., 2003). Such migrants may include babies (under 1 year of age), new non-birth registrations (e.g. ex-armed forces personnel and international in-migrants that join the NHS and then move within the same year), and people who move during one year but then leave the NHS register before the end of the second year (e.g. the deceased, new armed forces personnel and international out-migrants) (ONS 2011a). By failing to capture the movements of certain migrants, the PRDS is deemed inadequate as a stand-alone source for internal migration estimates (ONS, 2011b). An additional problem with the PRDS data is its tendency to undercount people of student age, more specifically young adult males, as these people tend to have low rates of registration with a GP (Fotheringham *et al*., 2004); consequently, in 2010, the ONS introduced an additional adjustment for students based on Higher Education Statistics Agency (HESA) (ONS, 2011a). Therefore, with the NHSCR offering existing migration estimates for all within year moves, the more geographically detailed PRDS data are combined at a more aggregate level with the more complete NHSCR data, and adjusted with HESA data, to produce migration estimates for LA and former HA areas (Jefferies *et al*., 2003; ONS 2011b).

The Scottish Community Health Index (CHI) and the Northern Irish Central Health Index (NI-CHI) are similar systems to the PRDS. Yet while the methodologies used by NRS and NISRA to formulate their patient register-based sub-regional migration estimates are very similar to those employed by the ONS, no real effort has been made to harmonise the estimates so as to provide UK-wide internal migration estimates (Rees *et al*., 2009: 114).

More generally, there are a number of limitations that the NHSCR, combined with the PRDS, suffer from. These include: the variation in the delay between a person moving and registering with a new doctor; the fact that some moves may not involve a GP re-registration and therefore will not be recorded; and individuals may move and not register the move with the GP (ONS, 2011a).

3.3.3 School Census (formerly the PLASC)

The School Census (formerly the Pupil Level Annual School Census) is an administrative data source that holds updated records for between seven and eight million state school pupils in England. It is mandatory for all state primary, secondary and special schools to collect data for the Schools Census on pupils aged 5-15 (at the start of the school year); once collected, it is submitted to the Department for Education (DfE) by each respective Local Education Authority (LEA) (Simpson *et al*., 2011). The PLASC data were updated annually between 2002 and 2007 and formed one of the data sets collected within the National Pupil Database (NPD). Other than the School Census, the NPD includes additional information on pupil attainment as well as reference data on schools and LEAs. Within the NPD, each pupil is given a unique pupil number (UPN) which in turn makes the linking of pupil records over time, and across the data sets within the NPD, possible (Dennett *et al*., 2007; Harland and Stillwell, 2007). From 2006, the updates for England have been increased to three times a year (once for each school term), this change also coincided with its renaming from the PLASC to the School Census. Similar systems are set up in Scotland, Wales and Northern Ireland;

however, these are only conducted annually, are less well established, and unfortunately the potential to link the separate pupil databases remains unclear (Rees *et al*., 2009; Simpson *et al*., 2011).

In terms of its potential usefulness for the analysis of internal migration, the locational data attached to the individual pupil record is especially useful. The geographic information includes the pupil's school, and the postcode and associated OA and SOA of their home address (Marquis and Jivraj, 2009). However, it should be noted that while unit postcode level geographic detail is possible, it is controlled by the Department for Education with support from the PLASC/NDP user group (PLUG) at the University of Bristol. Residential movement can be measured using the School Census by identifying a change in postcode between consecutive years so long as the pupil remains in the data set for the two consecutive years (Simpson *et al*., 2011). These locational data are supplemented by a number of other useful individual level pupil attributes including: age, gender, ethnicity, first language and free school meals status.

However, beyond its beneficial features, the School Census has a number of obvious limitations. Firstly, as has been noted repeatedly, it only covers pupils of compulsory school age and only within the state school system. Moreover, consistency issues arise between the constituent countries of the UK, where each country's equivalent School Census is run separately and using slightly different methodologies (Rees *et al*., 2009; Simpson *et al.*, 2011). Finally, care needs to be taken when analysing the School Census data; indeed it cannot be used to generalise about the movement patterns of the wider population because households containing school-aged children are known to be less likely than other households to migrate (Marquis and Jivraj, 2009).

3.3.4 Higher Education Statistics Agency (HESA)

The HESA is the official body for higher education statistics tasked with collecting, analysing and disseminating data related to students, qualifiers and staff as well as information on the destinations of leavers, finance, business and community, and estates management within the higher education sector. As Raymer *et al*. (2012: 75) make clear, *"data available from HESA covers all students attending a public higher education institution in the UK … and provides information on internal* *migration patterns as term time address as well as home (domicile) address is collated"*. In terms of the geographical detail offered by the HESA, while unit postcode detail is recorded, for reasons of confidentiality only data at the middle super output area (MSOA) level are made available (Rees *et al.*, 2009). Yet, as Rees *et al*. (2009: 115) suggest, *"[w]hilst knowledge of precise residential destinations is impossible, estimates of student migration at the local authority level are likely to be relatively accurate"*.

An important future development that is currently under discussion relates to a joint project to undertake a data linkage process between the NPD, UCAS and HESA data. According to HESA (2011) the intention is that such an exercise would produce a valuable data source tracking students right the way through from school to HE and initial destinations having completed higher education qualifications. If the data linkage is successful, this could prove to be a valuable resource for migration analysis, especially given its potential to focus on and track what is a highly mobile and dynamic section of the population. However, it should be noted that this proposal is at a very early stage and no formal details have been made available as yet (HESA, 2011).

3.4 Social survey sources

Surveys are typically rich in variable detail, but lacking in geographical detail and coverage due to their relatively small sample size. As has been noted by Poston and Bouvier (2010: 34)*"[b]y administering surveys to carefully selected random samples of the larger populations, demographers are better able to uncover underlying patterns of demographic behaviour than is possible with materials from censuses and registration systems"*. However, as stated, this level of attribute detail is commonly constrained by the level of spatial detail and coverage included in surveys. As such, all the surveys covered here offer at least some potential for those interested in the analysis of population mobility. However, they are largely restricted to the analysis of micro behavioural aspects of mobility analysis and are often limited in the geographical insights they provide. Nevertheless, over the last few decades, there has been a rapid increase in the availability of increasingly large-scale cross-sectional and longitudinal sample survey data sources which contain attributes relevant for the analysis of population movement (Cushing and Poot, 2004). Indeed, the analyses presented in the latter chapters of this study utilise a previously unused survey source with detailed information on population movement as well as socio-demographic and lifestyle characteristics, namely the Acxiom Ltd. Research Opinion Poll (itself described in detail in subsection 3.5). For a comparative overview of the current social survey data landscape see Tables 3.5 and 3.6 at the end of this section.

3.4.1 Integrated Household Survey (IHS)

The Integrated Household Survey (IHS) is a continuous composite survey with annual updates being released quarterly on a rolling basis (April 2009 to March 2010, July 2009 to June 2010, October 2009 to September 2010, January 2010 to December 2010, etc.). The most recent release covers the period April 2011 to March 2012. According to the ONS (2010c), the aim of the IHS is to produce estimates for particular themes of interest to a higher precision and more detailed geographic level than is currently on offer in alternative ONS social surveys. For the first IHS data release (April 2009 to March 2010), a sample size of approximately 450,000 individuals was achieved. The IHS has been formed from the merging of a number of existing government surveys. All of the component surveys contain a number of similar questions that form what are the ‘core’ questions of the IHS, covering themes including: economic activity, education, health and disability, identity and income (ONS, 2010c; Walthery, 2011). There are roughly 100 ‘core’ questions within the IHS. As outlined in Table 3.3, the component surveys for the April 2009 to March 2010 IHS include: the General Lifestyle Survey (GLF), the Living Cost and Food Survey (LCF), the Opinions Survey (OPN), the English Housing Survey (EHS), the LFS/Annual Population Survey (LFS/APS), and the Life Opportunities Survey (LOS). However, since the first data release, there have been a number of component changes to the IHS. In January 2010, the OPN survey was removed, a decision based on the requirement to reduce the size of the OPN survey by removing the IHS ‘core’ questions (ONS, 2010c). Further to this, both the LOS and EHS were also removed in April 2011. The LOS removal was based on a change in its sampling methodology making it inappropriate for inclusion while the EHS removal was based on funding restrictions in the EHS. Finally, the last data from the GLF were contributed in December 2011 (Jones, 2011). However, while the contributing surveys have been reduced in number, it is expected that the composition of the IHS will be flexible with surveys leaving the IHS in some cases being replaced by others each year (Walthery, 2011).

Table 3.3. IHS (April 2009 – March 2010) component surveys and respective size

Component Survey	Observations (Individuals)	Percentage of IHS sample
Annual Population Survey (APS)	334,206	74
English Housing Survey (EHS)	40,753	9
Life Opportunities Survey (LOS)	23,368	5
Opinions Survey (OPN)	20,981	5
General Lifestyle Survey (GLF)	18,033	4
Living Cost and Food Survey (LCF)	11,989	3
Total (IHS)	**449,330**	**100**

Source: ONS (2010a).

Given that the IHS is a composite survey, made up of separate surveys with their own specific designs, a certain level of care is required when attempting to analyse the data it provides. Indeed, Walthery (2011: 7) warns that *"it is not recommended to produce tables or estimates of the data without using the weights, given the heterogeneity of possible source or error within each variable"*. Moreover, the ONS has officially designated the IHS as experimental, wherein the statistics produced by the survey are still undergoing a testing phase for reliability.

In terms of their sample design, the component social surveys can be separated according to the two broad approaches used: stratified random sample (for the APS and LOS) and multi-stage clustered random sample (for the LCF and GLS). The first approach selects random addresses from the Royal Mail's Postcode Address File (PAF), thus the primary sampling units are addresses. The second approach is a combination of two or more stages. The first stage includes the identification of a random sample of 638 postcode sectors which are then stratified by metropolitan/non metropolitan areas and 2001 Census estimates of proportion of head of household in each Socio-Economic Group and car ownership, and a first stage sample is randomly drawn from these (Walthery, 2011). For the second stage, individual addresses are sampled from the 638 postcode sectors. Thus the primary sampling units are postcode sectors while the secondary units are addresses.

In terms of geographical coverage, the IHS includes the whole of the UK including Northern Ireland. However, it should be noted that not all component surveys cover the UK. For instance, of the April 2009 to March 2010 components, only the LFS and associated APS cover the whole of the UK – the GLF, LCF, LOS and OPN cover GB and the EHS simply covers England. The highest geographic detail available to standard users of the IHS is GORs although, if the user is granted Special Licence access, a range of further geographies are available including UA/LA, County, NUTS2 and NUTS3 regions (ONS, 2011c). It should be noted that there is currently no IHS "safe settings" dataset.

With regard population movement, the IHS provides two key questions: 'place of residence three months ago (UA/LA)' and 'place of residence one year ago (UA/LA)'. However, unfortunately, the data derived from these questions are currently only available on the ONS internal research datasets. The best option available publically, via ESDS, is data derived from 'place of residence three months ago (UK/ somewhere else)', which is itself restricted to Special License access. Within the UK Data Service End User dataset, it is possible to gather data on period at current address. Beyond this, it is important to note that the sample population is persons resident in the UK in private households, and young people living away from the parental home in student halls of residence or similar institutions during term time. Indeed, as has been discussed before in this review, the inclusion of the student population within a data source sample is essential for any researcher seeking realistic analysis and conclusions.

For a researcher interested in analysing changes over time, the IHS is currently not appropriate given that it only has a single year of data available (2009-10). However, when more data sets become available, it will be possible to analyse change over time using what will be a series of repeated cross-sectional surveys. The IHS does not contain any panel data, wherein repeated observations for the same individual over multiple time periods could be made.

3.4.2 Labour Force Survey (LFS)/ Annual Population Survey (APS)

The LFS is a continuous quarterly survey with a sample population of approximately 100,000 individuals in 42,000 households (ONS, 2011d: ii). As the ONS (2002: 550) states, the primary purpose of the LFS is the *"the prompt publication of key aggregate, whole economy indicators for the integrated assessment of labour market conditions"*. In its current form, the LFS employs a rotational sampling design, whereby respondents are interviewed five times at 13-week intervals and once they complete wave 5 they drop out and new respondents take their place. However, since its formation in 1973, the LFS has employed a number of other formats. From 1973 to 1983, the survey was carried out every two years in the spring quarter (March to May) on a UK basis. The LFS was carried out annually from 1984 to 1991 and consisted of two elements: a quarterly survey of approximately 15,000 private households, conducted in GB throughout the year, and a 'boost' survey in the spring quarter between March and May, of over 44,000 private households in GB and 5,200 households in Northern Ireland. The quarterly component of the surveys were not published due to concerns about robustness (ONS, 2011e: 3). From 1992 onwards, the LFS has been carried out quarterly and the sample size was extended to over 60,000 households in the UK up to 2006, the quarters used were seasonal (March-May, June-August, September-November, December-February). Additionally, 1992 saw the extension of the sample to include students living in halls of residence and NHS nurses' homes. Finally, in 2006, the LFS was produced according to its current format, as calendar-quarters, following an EU requirement under regulation linked to the EU LFS requirements for cross-country methodological comparability (ONS, 2011e).

The broad socio-economic and demographic categories included in the LFS have remained roughly the same since the major format changes of 1992. Categories in the LFS include, for example, ethnic group, gender, age, religion, education and training, income, health and employment type/location/hours worked. Paying specific attention to its use for migration analysis, the LFS household dataset includes a question on region of residence three months ago and one year ago, making it possible to formulate flow matrices by cross-tabulating previous residence with the region of usual residence. Furthermore, as Champion *et al*. (1998) have noted, the LFS also includes information on the respondents' labour market position one year prior, an interesting characteristic for those interested in analysing the causes and effects of migration.

For those interested in using the LFS for longitudinal analysis, there are currently two datasets made available with individuals linked across two (responded in two consecutive quarters and include their responses at each quarter) or five (responded in waves 1-5) consecutive quarters (ONS, 2011e). These data are available from winter 92/93 as portable SPSS files with a limited set of coded variables via ESDS. Figure 3.1 shows a snapshot of data from the LFS End User access household dataset. Columns A and B represent the region of origin and, when combined with the region of destination, can be used to generate flows. There are 22 codes representing the GORs as well as additional options including ‘baby under 3 months’, ‘outside UK’ and ‘no answer/does not apply’. As an example, the 16 code seen below represents the ‘South West’ region.

	A	B	C	D	E	F	G	H	I	J	K	L
1	Region of residence three months ago	Region of residence one year ago	Region of place of work - 2nd job	Region of place of work	Government office regions	Government Office Region 2 and 3 combined	Government office regions	Person living in halls of residence	No of children in Hhold between 5 and 15	No dependent chldrn in hhld aged 16-18	Number of dependent children in household aged under 19	No of children in Hhold age 4 yr or less
2	16	16	-10	17	10	10	16	2	0	0	0	0
3	16	16	-10	17	10	10	16	2	0	0	0	0
4	16	16	-10	-10	10	10	16	2	0	0	0	0
5	16	16	-10	17	10	10	16	2	1	0	2	1
6	16	16	-10	17	10	10	16	2	1	0	2	1
7	16	16	-10	-10	10	10	16	-10	1	0	2	1
8	16	16	-10	-10	10	10	16	-10	1	0	2	1
9	16	16	-10	17	10	10	16	2	0	2	2	0
10	16	16	-10	17	10	10	16	2	0	2	2	0
11	16	16	-10	-10	10	10	16	2	0	2	2	0
12	16	16	-10	-10	10	10	16	2	0	2	2	0
13	16	16	-10	-10	10	10	16	2	0	0	0	0
14	16	16	-10	17	10	10	16	2	0	0	0	0
15	16	16	-10	17	10	10	16	2	0	0	0	0
16	16	16	-10	17	10	10	16	2	0	0	0	0
17	16	16	-10	17	10	10	16	2	1	0	2	1
18	16	16	-10	17	10	10	16	2	1	0	2	1
19	16	16	-10	-10	10	10	16	-10	1	0	2	1
20	16	16	-10	-10	10	10	16	-10	1	0	2	1
21	16	16	19	17	10	10	16	2	1	0	1	0
22	16	16	-10	-10	10	10	16	2	1	0	1	0
23	16	16	-10	17	10	10	16	2	1	0	1	0

Figure 3.1. Example of LFS household data extracted from ESDS, April - June 2010. *N.B. Titles of columns have been changed from original variable code names to their descriptions.*

As with all sources, the LFS has a number of general strengths and limitations. In terms of strengths, the LFS offers the largest coverage of any stand-alone household survey in the UK, thus providing statistics with a relatively good level of geographic coverage. Moreover, due to the rich socio-economic and demographic characteristics included, the LFS allows for a number of cross-tabulations to be performed (e.g. migration by educational attainment) (ONS, 2011d). Additionally, as a result of the survey wave structure and the size of the sample, the sampling errors are relatively small (ONS, 2011d). From the perspective of population mobility analysis, its key strength lies in the fact that it offers quarterly migration data that can be disaggregated further by a large number of socio-economic and demographic characteristics (individual level and household level). The quarterly continuous production of such data means that the LFS proves a useful source for researchers interested in examining population movement and population change between census dates (Owen and Green, 1992). However, when compared against the IHS and indeed the ROP, a major limitation is its relatively small sample size, making it a less satisfactory source for those interested in estimating mobility counts and flows within the UK. Moreover, the small sample size additionally limits the spatial detail to the region level (GOR), although from March 2005 it has been possible to obtain UA/LAD level data for both previous and usual residence through Special License access. From the fourth edition of the Secure Data Service Access of the LFS (June 2014), non-anonymised postcodes are available, though only in GB (ONS, 2014a). A further limitation relates to the fact that the regional definitions are not constant between the origin (region of residence one year ago) and destination (region of current residence) (Dennett *et al*., 2007). In one such case, the destination 'Rest of Northern Region' had no corresponding origin, with the closest origin being defined as 'Rest of North East'. As Dennett *et al*. (2007: 90) assert, *"[t]his could be dismissed as a labelling error were it not for the unusually high migration to the Northern Region from 'Rest of the North West' … As such it is impossible to tell for certain whether these differences in flows are to be relied upon as accurate differences, or rather the result of boundary change"*.

A survey that is closely related to the LFS is the Annual Population Survey (APS). The APS, published quarterly, is a continuous combined survey of households in GB and has been in existence since 2004. The fundamental aim of the APS is to achieve a sample large enough to gather a minimum number of economically active respondents (510) in each LAD in England (except London boroughs where the target is 450), so as to produce more accurate attribute estimates at the sub-regional level (Werner, 2006; ONS, 2010b). In relation to the APS design and its potential value for migration analysis, Cangiano (2010: 7) notes the following: *"The APS sample is obtained by merging waves one and five of four LFS quarters and data from the Annual Local (Area) Labour Force Survey (LLFS) Boosts for England, Scotland [SLFS] and Wales [WLFS]. There are approximately 350,000 individuals per dataset, which makes estimates based on the APS more robust than those obtained from a single LFS quarter"*. Consequently, the APS shares all the same characteristics that are discussed in relation to the LFS. For instance, the APS suffers from exactly the same issues in terms of its application as a source for internal migration analysis (Rees *et al*., 2009). That is, data are only available at the GOR level, though again LA/UA geographies at origin and destination can be obtained through special license agreement. With that said, given that many of the variables included in the survey are the same as those in the LFS, with special license access the APS offers itself as a more robust data source for population migration analysis, especially at the sub-regional level.

3.4.3 General Lifestyle Survey (GLF, formerly the General Household Survey, GHS)

The GHS was renamed the GLF in 2008, carrying with it the same sample design and a largely similar questionnaire. The main change relates to the fact that the GLF now includes the IHS core questions. The survey started in 1971 as the GHS and was carried out continuously until its closing in January 2012, with breaks to review it in 1997/98 and to redevelop it in 1999/2000 (Dunstan, 2011). The GLF was a multi-purpose continuous household level survey with an annual representative target sample of approximately 13,000 households across GB (ONS, 2011f). However, for the final published data, the sample consisted of just over 8,000 households (19,000 individual interviews). The interview comprised of questions related to the household, completed by the household reference person, and individual questionnaires completed by all resident adults aged 16 and over. Demographic and health information was also collected about children in the household (Dunstan, 2011). The GLF included students living in halls of residence who were identified as part of the household being interviewed. The sample design used by the GLF was similar to that of the LFS in that it follows a rotation, a four-year sample rotation in which people remain in the sample for four years (waves) with one quarter of the sample being replaced each year (N.B. individuals are traced to their new household if they move) (ONS, 2010a: 9). It should be noted here that from 2007 to its completion in January 2012, the GLF data are only available under ESDS Special License Access.

The GLF covers a broad range of topics including: smoking and drinking, pensions, employment, income, social exclusion, material deprivation, poverty, health (including health services) and family information (relationships such as cohabitations and marriages) and fertility (ONS, 2011f). Beyond the continuous survey design, the GLF also includes trailer questions that differ from survey to survey and cover various topics that are dictated by the government department that sponsors them.

Unfortunately from a population migration point of view, the GLF does not allow for the production of an interaction matrix. This is because the survey only asks about the amount of time each respondent has lived at a current address, thus offering no detailed suggestion of an origin other than somewhere else within Britain or outside Britain. Moreover, the best geographical detail possible for destination data is GORs and the small sample size suggests that reliable findings would be restricted to analyses at these more aggregate levels.

Since 2005, the UK has been required to collect some cross-sectional and longitudinal statistical information on income and living conditions. This is required of all EU countries and the resulting data are known as EU Statistics on Income and Living Conditions (EU-SILC) (ONS, 2011g). As such the GLF was chosen as the UK survey vehicle (2005-2011) for the EU-SILC, a decision that acted as the main driver in the GLF's transition to a four-yearly sample rotation design. The result of this decision for the GLF is that it produces both cross-sectional and longitudinal (four-year time periods) micro-data at the household and person level. Therefore the change in the GLF made it possible for the construction of measures of change, for example in household structure, residential mobility, income, employment history and health measures (ONS, 2011f). With that said, the UK EU-SILC datasets do offer some potential for measuring residential mobility, though only at the level of GOR. However, given the small sample size (8,000 households and 19,000 individual interviews), any analysis of the migrant subsample would be very limited.

With the GLF being discontinued in 2012, the Family Resources Survey (FRS) has replaced it as the source for the cross-sectional EU-SILC data. By 2015, it is hoped that all cases for the EU-SILC will originate from the FRS, although between 2012 and 2015, it will contain cases originating from both the GLF and FRS (ONS, 2011g).

3.4.4 Family Resources Survey (FRS)

The FRS is a continuous survey that was formed in October 1992 (gaining UK coverage in 2002/2003) to meet the information requirements of Department for Work and Pensions (DWP) analysts. The survey is sponsored by the DWP and includes sections covering: income and state support receipt; tenure; savings and investments; carers and disability; and occupation and employment (DWP, 2011).

The sample size obtained in the most recent (2011/12) FRS was 20,000 household units (containing 75,000 individuals) with information being held at the household/family level and, under Special License access, at the individual level (ONS, 2011h). Whilst the FRS was designed with the DWP's needs in mind, the survey does contain information that makes it potentially useful for outside researchers and other government departments. However, in terms of its use for measuring population mobility, the FRS is not particularly helpful. As with the GLF, the spatial scale is GOR and the survey only asks respondents about the length of residence at their current address. Again it is not possible to gain any detail about the origin of the respondents who have changed address. However, with its transition into the main source of the longitudinal SILC survey, future FRS data sets, based on their adjusted design (ONS, 2011g), may very well hold potential for the analysis of population movement and residential mobility.

3.4.5 English Housing Survey (EHS)

The EHS is a continuous national survey commissioned by the Department for Communities and Local Government (DCLG). Formed in April 2008 through the merging of the English House Condition Survey (EHCS) and the Survey of English Housing (SEH), the EHS is tasked with collecting information about people's housing circumstances and the condition and energy efficiency of housing in England (ONS, 2011i). From its formation in April 2008 to April 2011, the EHS formed part of the IHS, with the core questions from the IHS being included within the EHS questionnaire. As such, the years of data that include the IHS core questions that can be used for measuring migration, namely how long respondents have lived at their current address and where they lived before, if they have lived in their current accommodation for less than 12 months. However, as noted above, in April 2011 the last contribution was made to the IHS. Moreover, at the same time the questionnaire content was reduced and questions on previous address were removed, though the question on duration at the current address remains.

In its current format, the EHS uses a complex multi-stage methodology consisting of two interlinked data collection methods. The first is derived from an interview survey of approximately 13,300 households a year (17,000 before the cost review in 2001-12) and is produced annually every financial year (DCLG, 2013). The second data collection is based on a rolling two-year sub-sample of respondents to the initial interview survey and involves a physical inspection of around 6,200 homes. This data set is predominantly focussed on collecting data on housing conditions and energy performance (ONS, 2011i). Through the combining of the two datasets, it is possible to produce a comprehensive list of socio-economic and demographic variables including: ethnicity, household income, education, health and various other indicators linked to deprivation and household related questions (ONS, 2011j).

In terms of its application for migration analysis, only analysis at regional and national levels is possible. However, for the years in which the EHS was integrated into the IHS, the potential is there to link the specialist variables in this dataset, to the IHS core module which includes the previous address questions. Thus, for those data sets, it is possible to generate information on the origin and destination at the GOR level. Moreover, data at the unit postcode and LSOA geographies can be generated with access to the highly restricted Secure Data Service access EHS 2008 to 2012 data, though again the relatively small sample size will make reliability an issue here.

3.4.6 Life Opportunities Survey (LOS)

The LOS is a large scale longitudinal survey of disability in Great Britain. Carried out by the ONS on behalf of the Office for Disability Issues (ODI), the survey seeks to explore disability in terms of the social barriers to participation that people experience and can be used to compare the experiences of disabled people with those of non-disabled people (Howe, 2010: 1). The LOS began in June 2009 with a baseline random sample of 23,380 households (37,500 individuals) across GB, interviewing all people aged 16 and over in the household as well as asking parents or guardians to provide some key demographic data about children aged 11 to 15. It should be noted that once these children reach the age of 16 they too will be able to take part in the face to face interview process (ONS, 2010b: 7). The longitudinal design of the LOS enables three distinct groups to be followed over time: disabled group; comparison group of non-disabled people; a larger non-disabled group, monitored for the onset of impairment over time.

The LOS baseline survey started in June 2009 and took two years to complete with the first full wave (20009-2011) now available via the UK data service website. As a contributor to the IHS, the LOS includes the ‘core questions’ relating to basic demographic characteristics and other household information on all members of the household. However, the second part of the LOS questionnaire is administered to each adult in the household and asks a range of detailed questions covering topics including: health, provision of unpaid care, crime, income benefits, and as Raymer *et al*. (2012: 102) assert, *“a unique variable is reported at the household level based on the ability to cope financially”.* Indeed, such characteristics could make for some interesting analysis of mobility patterns for what have been a particularly hard to measure population subgroup.

As noted above, the LOS ceased to be a contributor to the IHS in April 2011. With its removal, it is currently unclear as to whether or not some of the IHS ‘core’ variables will remain within the LOS. From a migration point of view, the End User Licence LOS is unusable due to the fact that it has no geographic identifier, at origin or destination. However, the more restricted Special Licence LOS data does include geographic variables (country, GORs and LAD) and more information on household relationships, country of previous residence, medical conditions and occupations. Of course, given that the LOS follows a longitudinal design, tracking individuals every 12 months (whether they remain resident in the original house or have since moved), it should be possible to gather information on residential movement within Britain as new waves are published. Although again, as a source of internal migration, the LOS sample size restricts it to more aggregate spatial analysis.

3.4.7 Living Cost and Food Survey (LCF)

The Expenditure and Food Survey (EFS) began in 2001-02 through the merging of the Family Expenditure Survey (FES) and the National Food Survey (NFS), both of which had been in existence since the 1950s. However, from January 2008, the EFS changed its name to the LCF upon integration into the IHS. As Rafferty and Acik-Toprak (2011: 3) declare, the LCF, in a similar manner to the EFS, continues to be primarily used to provide *“information for the Retail Prices Index, National Accounts estimates of household expenditure, the analysis of the effect of taxes and benefits, and trends in nutrition”.* With that said, it also contains useful multipurpose data on economic and social topics.

The LCF draws on the Royal Mail's PAF and follows a multi-stage stratified random sample with clustering (ONS, 2010b). The Northern Ireland sample design is slightly different and is drawn from a random sample of addresses from the Valuation and Lands Agency list (Rafferty and Acik-Toprak, 2011). For the most recent data (January 2012 – December 2012) the sample size is 5,425 households in GB, and 171 in Northern Ireland (approximately 11,000 individuals) (ONS, 2013).

The LCF is collected through three main sources: a household questionnaire; an income questionnaire (for each adult household member); and expenditure diaries (for each adult, and for children aged between 7 and 15 years). The household questionnaire includes questions about subjects including family relationships, ethnicity, employment, and expenditure information not recorded as part of the diary (i.e. large infrequently purchased items such as vehicles, package holiday and home improvements). The household questionnaire consists of questions that are asked at the household level with the questions being answered by and large by the household reference person. Demographic information as well as information on accommodation and tenure is collected for every adult in the household. The individual questionnaire follows and asks questions at the person-level covering topics such as income from employment, benefits and assets. Again the questionnaire must be completed by every adult in the household. Finally, the expenditure diaries record daily expenditure for two weeks; however, for reasons of confidentiality, only derived variables from the expenditure diary are available from the UK Data Service, (see Rafferty and Acik-Toprak, 2011 for a more detailed introductory guide). Unfortunately, the variables added as part of the IHS core are not available within the LCF datasets, but can be linked to through the IHS core module. Thus within the LCF dataset the only variable related to migration is the period spent at current address. The level of spatial detail provided is GOR, however, given the extremely small sample size, the LCF cannot be relied upon as a source for those seeking to undertake a comprehensive geographical study of mobility.

3.4.8 Understanding Society – UK Household Longitudinal Study (UKHLS – incorporates the BHPS)

Understanding Society (otherwise known as the UK Household Longitudinal Study – UKHLS) is a longitudinal multi-topic household study conducted by the Institute for Social and Economic Research (ISER) at the University of Essex. The overall purpose of the UKHLS is to provide high quality longitudinal data about topics including health, work, income, education, family and social life and to explore these within the context of long-term social and economic change (McFall, 2011). The study is unprecedented in its size with a UK-wide sample of approximately 40,000 households included in the first wave. Data collection for each wave is conducted over a 24 month period with collection for the first wave having started in January 2009 and ended in January 2011 (McFall, 2011). The overall sample size is made up of a number of smaller components including: the General Population Sample; the Ethnic Minority Boost Sample; the Innovation Panel; and the BHPS Sample (Burton *et al.*, 2011). As is clear from the sample breakdown, the UKHLS incorporates, and indeed builds upon, the British Household Panel Survey (BHPS) that was phased out in 2011. By offering such a large sample, the UKHLS allows for researchers to gain greater insights into particular population sub-groups, such as teenage parents, older workers or the unemployed, all of whom have been hard to measure in previous longitudinal studies with smaller sample sizes (Bryan, 2011). Moreover, its UK-wide household sample allows for more detailed geographical analysis across a number of spatial scales, depending on the dataset used (see below). It should also be noted that preparations for administrative data linkages are underway. During the first wave of interviews, each adult participant was asked to provide their consent for the UKHLS to link their survey data to health and education records. Further, the study requested consent of parents to link health data on children aged 0-15 and education data on children aged 4-15. However, beyond this there are plans for further administrative data linkages including records of benefit receipt, participation in government employment schemes, savings and pensions, earnings and National Insurance contributions (Bryan, 2011). Clearly, once completed and made publically available, the data linkage would vastly increase the scope of the study.

The data from the UKHLS are available at varying levels of spatial detail, and can be accessed through the UK Data Service. The basic End User License allows for analysis at GOR level, although, analysis at LAD level is possible via the Special License dataset. The highly restricted Secure Data Service access allows for any level of geography to be derived due to the availability of National Grid references (easting and northing) on each record.

In terms of its value for migration analysis, the UKHLS offers up a number interesting possibilities. Beyond its basic advantage as a longitudinal study, in that it follows its members from residence to residence, it includes questions on length of time at current address but also questions that have potential for the analysis of *future* migration propensities and *lifetime* migration propensities:

- “If you could choose, would you stay here in your present home or would you prefer to move somewhere else?”
- “Do you expect you will move in the coming year?”
- “How many times have you moved to a new address since you were aged 14 (come to the UK to live) either on your own or with family?”

When compared to other mainstream social survey sources, the detail of the migration questions are unique and along with the sample size and potential data linkages the UKHLS should be taken seriously as a source of data for those interested in studying population mobility behaviours and outcomes in the UK, particularly once a number of waves have been published. Table 3.5 provides a summary of the characteristics of the major social surveys reviewed here, with a focus on their application in the analysis of population mobility.

3.5 The Acxiom Research Opinion Poll

Founded in 1969, Acxiom Ltd. is an international commercial company based in Arkansas, USA. With a worldwide annual turnover exceeding $1 billion, the company is a global leader in interactive multichannel marketing services (Acxiom Ltd., 2014a). In GB, Acxiom Ltd. produces two major annual data products, one being the ‘Aggregate Data’ which, after a process of weighting and manipulation, is argued to be fully representative of the GB population at LAD and LSOA levels and covers key demographic, behavioural, lifestyle, financial and household variables (Acxiom Ltd., 2014b). The other major product is PersonicX Geo, a geodemographic classification system at the level of the postcode (Raper *et al.*, 1992) designed for commercial applications linked to consumer segmentation (Thompson *et al.*, 2010; Acxiom Ltd., 2011). Both of these products derive their data from the Acxiom's biannual lifestyle survey, the Research Opinion Poll (ROP). Given the requirements of their products, the primary aim of the ROP is to gather detailed and up-to-date information on consumer spending habits, preferences, socio-demographic, behavioural, lifestyle and household characteristics with extensive geographical coverage and detailed geo-identifiers (Thompson *et al.*, 2010).

Acxiom's ROP is a very large lifestyle survey carried out across GB (i.e. England, Wales and Scotland, but not Northern Ireland[1]). It is a voluntary and principally paper-based survey (although it is increasingly being distributed via the internet) that is distributed using direct mail (Raper *et al.*, 1992) twice a year, in September and January. One of the key benefits of the ROP lies in its large sample size; for example, the raw sample from the January 2005 ROP contained over 400,000 responses. Whilst the exact operational surveying details are not disclosed by Acxiom Ltd., they employ a number of address sources to ensure that their response is geographically even and reasonably representative of the GB (18 and over) demographic profile (Rees *et al*., 2009). Thompson *et al*. (2010: 13) acknowledge Acxiom Ltd.'s operational success and note that for the 2009 ROP: *"*[…] *only 0.4% of all Middle Super Output Areas (MSOAs) across* [Great Britain] *did not return a response".*

However, beyond the large size and detailed geographical coverage of the sample, the ROP also offers a great deal in terms of attribute detail. The variables used here have been selected from the micro-database for the analysis of population migration (Table 3.4); however, the survey at large asks approximately 130 questions, allowing for over 1,000 possible answers, covering 26 broad topics including for example: groceries; shopping; local area; environment; outgoings; occupation; home; leisure; education; and health. The questions can be broken down into two broad categories: core questions and sponsored questions. The former are repeated from survey to survey and cover such characteristics as respondents' current address, age, sex, household income, occupation, and housing tenure. The latter are questions included

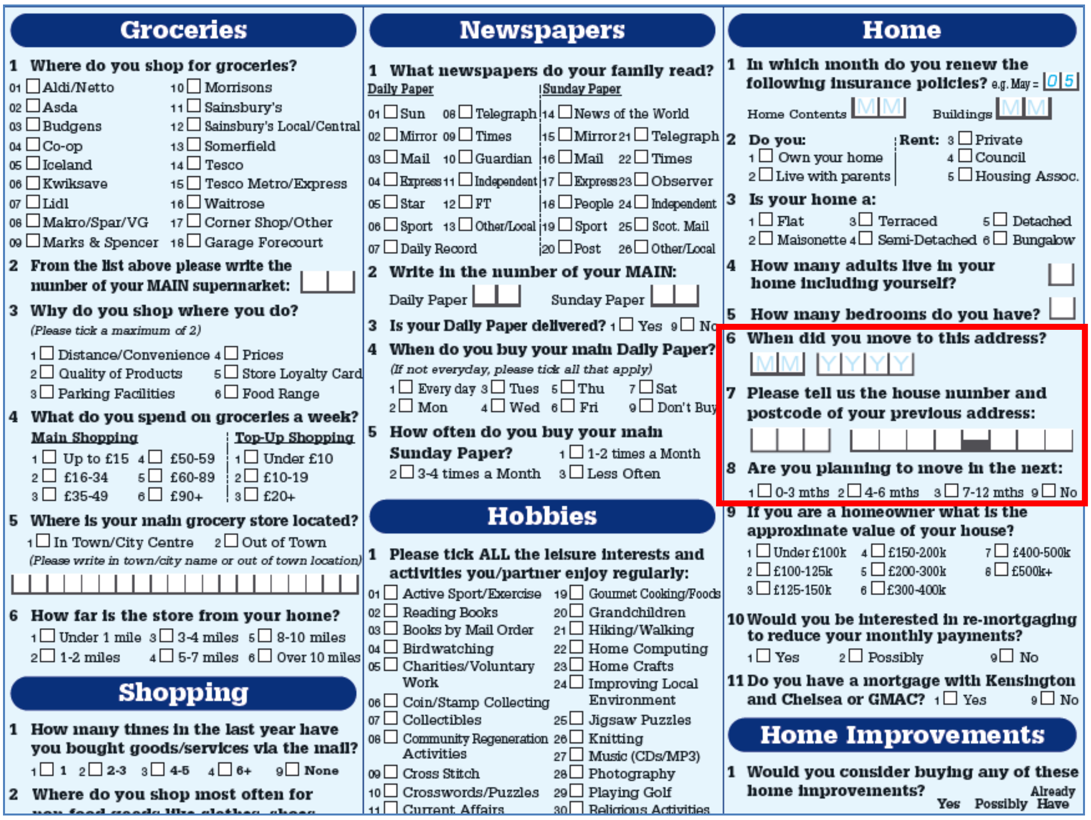
---

[1] Some responses are collected from Northern Ireland in the raw sample; however, the very small sample renders them unreliable for geographical analysis at any scale. Given this, Acxiom Ltd. Products do not include Northern Ireland.

in the ROP that have been paid for by different clients. Yorkshire Forward, the now defunct regional development agency, sponsored a series of questions ranging in topic from specific questions on Yorkshire and The Humber, through to more general questions relevant to the environment and housing tenure. However, importantly for the analysis of population mobility, the following sponsored questions were asked in the ROP in 2005, 2006 and 2007 (Acxiom Ltd., 2007):

- "When did you move to this address? (month and year)"
- "Please tell us the house number and postcode of your previous address"
- "Are you planning to move in the next: 0-3 months; 4-6 months; 7-12 months; No?"

The 'Home' section of the 2007 ROP questionnaire, where the residential mobility questions are presented, is highlighted in Figure 3.2.



Groceries

1 Where do you shop for groceries?
01 Aldi/Netto 02 Asda 03 Budgens 04 Co-op 05 Iceland 06 Kwiksave 07 Lidl 08 Makro/Spar/VG 09 Marks & Spencer
10 Morrisons 11 Sainsbury's 12 Sainsbury's Local/Central 13 Somerfield 14 Tesco 15 Tesco Metro/Express 16 Waitrose 17 Corner Shop/Other 18 Garage Forecourt
2 From the list above please write the number of your MAIN supermarket:
3 Why do you shop where you do? (Please tick a maximum of 2)
1 Distance/Convenience 2 Quality of Products 3 Parking Facilities 4 Prices 5 Store Loyalty Card 6 Food Range
4 What do you spend on groceries a week?
Main Shopping: 1 Up to £15 2 £16-34 3 £35-49 4 £50-59 5 £60-89 6 £90+
Top-Up Shopping: 1 Under £10 2 £10-19 3 £20+
5 Where is your main grocery store located?
1 In Town/City Centre 2 Out of Town
(Please write in town/city name or out of town location)
6 How far is the store from your home?
1 Under 1 mile 2 1-2 miles 3 3-4 miles 4 5-7 miles 5 8-10 miles 6 Over 10 miles

Shopping

1 How many times in the last year have you bought goods/services via the mail?
1 1 2 2-3 3 4-5 4 6+ 9 None
2 Where do you shop most often for

Newspapers

1 What newspapers do your family read?
Daily Paper: 01 Sun 02 Mirror 03 Mail 04 Express 05 Star 06 Sport 07 Daily Record 08 Telegraph 09 Times 10 Guardian 11 Independent 12 FT 13 Other/Local
Sunday Paper: 14 News of the World 15 Mirror 16 Mail 17 Express 18 People 19 Sport 20 Post 21 Telegraph 22 Times 23 Observer 24 Independent 25 Scot. Mail 26 Other/Local
2 Write in the number of your MAIN:
Daily Paper Sunday Paper
3 Is your Daily Paper delivered? 1 Yes 9 No
4 When do you buy your main Daily Paper? (If not everyday, please tick all that apply)
1 Every day 2 Mon 3 Tues 4 Wed 5 Thu 6 Fri 7 Sat 9 Don't Buy
5 How often do you buy your main Sunday Paper?
1 1-2 times a Month 2 3-4 times a Month 3 Less Often

Hobbies

1 Please tick ALL the leisure interests and activities you/partner enjoy regularly:
01 Active Sport/Exercise 02 Reading Books 03 Books by Mail Order 04 Birdwatching 05 Charities/Voluntary Work 06 Coin/Stamp Collecting 07 Collectibles 08 Community Regeneration Activities 09 Cross Stitch 10 Crosswords/Puzzles 11 Current Affairs
19 Gourmet Cooking/Foods 20 Grandchildren 21 Hiking/Walking 22 Home Computing 23 Home Crafts 24 Improving Local Environment 25 Jigsaw Puzzles 26 Knitting 27 Music (CDs/MP3) 28 Photography 29 Playing Golf 30 Religious Activities

Home

1 In which month do you renew the following insurance policies? e.g. May = 05
Home Contents MM Buildings MM
2 Do you: 1 Own your home 2 Live with parents
Rent: 3 Private 4 Council 5 Housing Assoc.
3 Is your home a: 1 Flat 2 Maisonette 3 Terraced 4 Semi-Detached 5 Detached 6 Bungalow
4 How many adults live in your home including yourself?
5 How many bedrooms do you have?
6 When did you move to this address? MM YYYY
7 Please tell us the house number and postcode of your previous address:
8 Are you planning to move in the next:
1 0-3 mths 2 4-6 mths 3 7-12 mths 9 No
9 If you are a homeowner what is the approximate value of your house?
1 Under £100k 2 £100-125k 3 £125-150k 4 £150-200k 5 £200-300k 6 £300-400k 7 £400-500k 8 £500k+
10 Would you be interested in re-mortgaging to reduce your monthly payments?
1 Yes 2 Possibly 9 No
11 Do you have a mortgage with Kensington and Chelsea or GMAC? 1 Yes 9 No

Home Improvements

1 Would you consider buying any of these home improvements? Yes Possibly Already Have

Figure 3.2. Section of January 2007 ROP with mobility questions highlighted in red (Source: Acxiom Ltd., 2007)

When these questions are combined with each respondent's 'current address' (at postcode level), the potential of this data source for the analysis of population mobility becomes apparent. Not only does the precise geo-referencing of cases allow the researcher to generate an aggregated geography of their choice (at the place of origin and/or destination), it is also possible to join small area functional geographies such as the Census 2001 Output Area Classification (OAC) (Vickers and Rees, 2007), a classification of neighbourhood type which may be helpful in identifying/exploring the influence of the neighbourhood on individual-level mobility behaviours. Moreover, the size of the sample, coupled with its extensive (non-clustered) geographical coverage and detailed geo-identifiers, at both the origin and destination, makes the ROP a source of data with genuine potential for analysing the simultaneous effects of individual (i.e. age, gender, ethnicity, household income) and contextual (i.e. origin and/or destination area effects or origin-destination flows) level phenomena on various characteristics of population movement in GB, be it the propensity to move in the first place or, following this, the postcode-to-postcode distance of the move. It should be noted, however, that the ROP only allows for a single household respondent and therefore multiple members of the household are not measured, although general characteristics about the household, for instance gross annual household income, housing tenure and marital status, are included in the survey. Table 3.4 presents the variables that have been collected for use in this study, covering the period where the necessary residential mobility questions were asked (January 2005 to September 2007).

Table 3.4. Overview of Acxiom ROP variables available for use in this research project

Key ROP variables	Jan. 2005	Sept. 2005	Jan. 2006	Jan. 2007	Sept. 2007
Current address (postcode)	✓	✓	✓	✓	✓
Sex	✓	✓	✓	✓	✓
Age	✓	✓	✓	✓	✓
Ethnic background	✓	×	✓	✓	×
Marital status	✓	✓	✓	✓	✓
Occupation	✓	✓	✓	✓	✓
Highest qualification	✓	✓	✓	✓	✓
House price	×	×	×	✓	✓
Gross household income	✓	✓	✓	✓	✓
Type of home	✓	✓	✓	✓	✓
Housing tenure	✓	✓	✓	✓	✓
Household size	✓	✓	✓	×	✓
Number of cars	✓	✓	✓	✓	✓
Year and month of move	✓	✓	✓	✓	✓
Previous address (postcode)	✓	✓	✓	✓	✓
Neighbourhood satisfaction	✓	✓	✓	✓	✓
Neighbourhood improvement	×	×	✓	✓	✓
Plans for future move	✓	✓	✓	✓	✓

As with all of the sources of migration data discussed here, despite the relative strengths of the ROP, the data do not come free of problems. For instance, is clear from Table 3.4 above, where many of the questions asked are sponsored by outside actors, the ROP struggles to provide consistency across the period of study. For instance, potentially important questions on neighbourhood improvement, household size and house price are not available in all cross-sections. Perhaps more problematic is the lack of the key demographic measure of ethnicity in both of the September ROP cross-sections. Beyond this, unsurprisingly given its form as a voluntary postal/online survey, the raw ROP cross-sections have been found to contain inherent individual- and area-level bias on a number of characteristics including: age, sex, geography, ethnic group and income group (see Thompson *et al.*, 2010). Such bias can be expected to be driven, to a large extent, by survey non-response and errors in the sampling frame. Unfortunately, due to commercial sensitivity, basic survey response rates are not available; nor is it possible to obtain information on the addresses of those who failed to provide a response. Moreover, detailed documentation of the ROP's sampling strategy is not publically available, though from what is known, the ROP does not follow traditional conventional survey approaches with complex multi-stage cluster designs; rather it is an attempt at generating a very large and geographically un-clustered sample with a broadly accurate demographic profile based on the PAF (Thompson *et al.*, 2010). Beyond the raw sample size and characteristics, when delivered, the ROP microdata are in raw format, with only the household representative's current postcode address having undergone prior preparation and cleaning by Acxiom Ltd. As such, all of the above issues are discussed and attended to in detail in the validation Chapters (Chapter 5 and Chapter 6). To make comparisons easier, Table 3.5 (below) provides a broad overview of the characteristics of major social surveys reviewed above, with a focus on their application in for the analysis of population mobility. Similarly, Table 3.6 presents the questions relevant for population movement analysis in the social survey sources reviewed, as well as the access restrictions and corresponding geographical identifiers.

Table 3.5. Overview of social survey data sources

Survey	Reporting period	Coverage	Lowest level of geography	Approximate annual sample	Orig.	Dest.	Broad themes	Download	Components
Integrated Household Survey	2009	UK	GOR (Special License: LAD)	450,000 people	✓	✓	Socio-economic & demographic characteristics	ESDS (OTHER, SPSS, STATA, TAB)	GLF (GHS), LCF (EFS), EHS, LOS, LFS/APS
Labour Force Survey	1973	UK	GOR (Special License: UA/LA)	100,000 people	✓	✓	Labour market characteristics & conditions	ESDS (SPSS, STATA, TAB)	
Annual Population Survey	Annual since 2004	UK	GOR (special license: UA/LA)	325,000 people	✓	✓	Same as LFS	ESDS (SPSS, STATA, TAB)	LFS, LLFS, WLFS, SLFS
General Lifestyle Survey	2008 - 2012	UK	Special License only: GOR	19,000 people		✓	Lifestyle & family information, socio-economic circumstances	ESDS (SPSS, STATA, SAS, TAB)	Formerly GHS
Family Resources Survey	1992	GB 1992 UK 2002	GOR (Special License: same)	75,000 people		✓	Housing, employment, income, general health	ESDS / DWP (SPSS, STATA, TAB)	Full integration of EU-SILC by 2015
English Housing Survey	2008	England	GOR (Special License: N/A)	41,000 people		✓	Housing circumstances & condition	ESDS (SPSS, STATA, SAS, TAB)	Formerly SEH & EHCS
Life Opportunities Survey	2009	GB	Special License only: LAD	23,000 people		✓	Social barriers to participation for disabled people	ESDS (SPSS, STATA, TAB)	
Living Costs & Food Survey	2008	UK	GOR (Special License: N/A)	12,000 people		✓	Household expenditure & socio-economic data	ESDS (SPSS, STATA, SAS, TAB)	Formerly EFS
Acxiom Lifestyle Survey	Current format 2004	GB	Postcode	750,000 people (Jan. & Sept.)	✓	✓	Consumer spending, preferences, socio-demographic data	Acxiom	
Non-Census Longitudinal Studies									
British Household Panel Survey	1991 – 2011	UK (from wave 11 onwards)	GOR (Special Licence: CAS Wards) (Secure Data Service Access: easting & northing)	10,000 households		✓	Social & economic change	ESDS (SPSS, STATA, TAB	
Understanding Society	First wave: 2009	UK	GOR (Special License: LAD & CAS wards) (Secure Data Access: easting & northing)	40,000 Households		✓	Socio-economic circumstances & attitudes	ESDS (SPSS, STATA, SAS, TAB)	BHPS, LSEM

Table 3.6. Overview of social survey migration questions

Survey	Migration questions and access restrictions
Integrated Household Survey	Period at current address; place of residence 3 months ago (UK or somewhere else) (Special License); place of residence 3 months ago (range of codes as UA/LA) (ONS internal only); place of residence one year ago (range of codes as UA/LA) (ONS internal only).
Labour Force Survey	Period at current address; region of residence 3 months ago and 1 year ago (Special License: UA/LA).
Annual Population Survey	Period at current address (Special License: region of residence 3 months and 1 year ago).
General Lifestyle Survey	Period at current address (can be linked through IHS to migration questions in 'core' module).
Family Resources Survey	How long have you lived at the address? (0-12 months / 1 year to more than 20 years).
English Housing Survey	Period at current address (can be linked through IHS to migration questions in 'core' module).
Life Opportunities Survey	Period at current address, and ability to track longitudinally (can be linked through IHS to migration questions in 'core' module).
Living Costs and Food Survey	Period at current address (can be linked through IHS to migration questions in 'core' module).
Acxiom Lifestyle Survey	When did you move to this address (Month and Year); the postcode of previous address; planning to move in the next 12 months. Sponsored questions asked: January 2005, 2006, 2007 and September 2005, 2007.
British Household Panel Survey	Moved in past year; future intention to move; move into residential home, and ability to track moves longitudinally.
Understanding Society	Lived at address whole life; moved to address (month, year); prefers to move house; expects to move house; expects to move in next year, and ability to track moves longitudinally.

3.6 The ONS ‘Beyond 2011’ Programme

In 2010, the UK Statistics Authority asked the National Statistician and the ONS, in collaboration with the statistics offices of Northern Ireland and Scotland, to review the options for the next census and, more generally, the future provision of population and socio-demographic statistics in the UK. Consequently, from its formation in April 2011 to the publication of recommendations in March of this year (2014), the ONS Beyond 2011 Programme has undertaken extensive research, reviewed practices in other countries, engaged in a wide ranging public consultation and commissioned an independent review of methodology (Skinner *et al.*, 2013) into new approaches to counting the population, particularly at small area level. Indeed, the Beyond 2011 Programme was prompted by a number of concerns surrounding the traditional population census methodology and outputs. For example, financial concerns were raised about the ever growing costs of traditional census dissemination and collection, whilst the issues of general applicability and usefulness were also discussed. For instance, with the current census being a once a decade snapshot of the population, the relevance of the statistics produced are known to necessarily deteriorate over time, an issue that is further exacerbated by an increasingly dynamic population (ONS, 2014b). Given these concerns, a major focus of the programme was to explore alternative methods and sources of data collection and provision, investigating the potential for combining existing administrative datasets with survey datasets, both public and commercial (ONS, 2011l). As such, a great many existing sources of population data were explored by ONS, including many that have been discussed above; examples include, NHS Central Register (NHSCR); DWP/HMRC Customer Information System (CIS); electoral roll (18 years and over); School Census (5-16 years); HESA (students); birth and death registrations; and the DVLA. For sources of more detailed small area socio-demographic data, the ONS further explored the use of large scale social surveys; however, other possibilities were also investigated, including: DVLA; utilities; TV licensing; and commercial sources (Calder and Swan, 2011).

Whilst research is ongoing in terms of exploring ways to maximise the use of administrative data and survey sources (ONS, 2014c), the ONS accepted all recommendations of the Skinner *et al.* (2013) report. As a result, on behalf of ONS, the National Statistician made her recommendation to UK Statistics Authority, namely to make use of all sources of current and future data sources, combining data from an online census and administrative data and regular surveys. The specific ONS (2014b: 11) recommendation is as follows:

- "*An online census of all households and communal establishments in England and Wales in 2021, as a modern successor to the traditional, paper-based decennial Census. As in 2011, ONS recognises that special care would need to be taken to support those who are unable to complete the census online.*

- *Increased use of administrative data and surveys in order to enhance statistics from the 2021 Census and improve statistics between censuses".*

Whilst the recommendation awaits parliamentary approval, further ONS research is planned for the coming months and years to determine the optimal blend of methods and data sources (ONS, 2014b). With this in mind, the question emerges as to whether the ROP, as a commercial survey source, can be considered to have any potential in this area. Indeed, whilst it has its limitations, it provides a biannual sample size that far outweighs anything seen in the government survey source datasets. Moreover, along with its postcode geo-identifiers, the lifestyle and socio-demographic information contained within the samples, is equally impressive. Consequently, within the context of ongoing research by ONS, it is hoped that the research undertakings of this study will not only be useful in broadening the evidence base relating to our knowledge of population mobility in GB, but also potentially useful in benchmarking, validating and integrating the Acxiom Ltd. ROP with official statistics.

3.7 Summary and conclusions

As was stated in the introduction, researchers interested in population mobility find themselves in a situation where they must utilise a variety of sources, sources that are characterised by sizable variations in terms of their coverage, detail and accuracy. Consequently, this chapter has sought to provide a detailed review of the various census, administrative, and social survey data sources from which mobility data can be generated. Perhaps most apparent in the review is the fact that all the sources have their own respective strengths and weaknesses, issues that must be considered carefully when deciding upon which data source to use and for what types of analysis. Broadly speaking, it would be fair to argue that census statistics are the most comprehensive and reliable of all; however, in their current guise, they are quickly outdated. Alternative administrative sources, provide up-to-date information, usually combined with good geographical coverage, but can be partial in their population coverage and the variable detail contained within. Surveys, on the other hand, are a timely source of highly detailed socio-demographic and economic data, but are also typically characterised by relatively small sample sizes which often act to restrict their potential for reliable analysis at more detailed geographical levels.

Yet as was discussed in subsection 3.6, the coming years are likely to see existing, though previously unused, alternative sources of geo-demographic data rise to prominence in the social science community. Thus, with the ONS still engaged in ongoing research into the opening up and linking together of existing though underutilised alternative sources of data, the benchmarking and integration of the ROP (with its unique combination of a large micro sample, comprehensive geographical coverage and detailed geo-identifiers and variable attributes) appears to sit quite well within the wider Beyond 2011 context. As a result, the next chapter, Chapter 5, details the extensive data preparation and cleaning exercises employed on the ROP, before revealing the initial validation process, encompassing empirical benchmarking against the 2001 Census, administrative and population survey sources. Following this, Chapter 6 seeks to build on the empirical (descriptive-based) benchmarking of Chapter 5 by assessing the reliability of the ROP data for model-based analyses.

Chapter 4
Population movement in GB: Methods for analysis

4.1 Introduction

As was mentioned in the previous chapter, migration researchers and policy makers frequently find themselves in a situation where, in order to satisfy their analytical requirements, they must utilise data from a variety of sources that are characterised by sharp contrasts in coverage, detail (of both attribute and geography) and accuracy. There is a fundamental dichotomy between micro-level and macro-level approaches to the analysis of population migration (Stillwell and Congdon, 1991). The former is concerned with methods that analyse the behaviour of the individual migrant (or family), the influences on the decision-making process and the consequences of migration as far as the micro unit is concerned, whilst the latter involves approaches that analyse aggregate migrant flows of people and identify the importance of macro explanatory variables including population size, employment rate or environmental factors at either /both places of origin and destination, together with distance moved. Indeed, for analysis at the macro or aggregate scale, spatial analysts of population migration in the UK have primarily sourced data from the aggregate census sources and administrative registers, reviewed in the previous chapter, because national survey data are usually restricted by sample size and geographic detail (Nam *et al.*, 1990; Rees and Kupiszewski, 1999; Stillwell *et al.,* 2011). In the UK context, decennial censuses provide extensive demographic and socio-economic attributes of migrants moving between and within geographical units at different spatial scales (Flowerdew and Green, 1993; Rees *et al.*, 2002; Raymer *et al*., 2012). Given their near comprehensive national coverage, their relative reliability (on enumeration day) and the rich detail of their demographic and socioeconomic variables, censuses are currently considered as the optimum points of reference for those interested in local population statistics (Raymer*et al*., 2012) and in small area demographic analysis, inter-censal population estimation and future population projection (Moon *et al.*, 2000). Alternatively, population registers and administrative sources are extremely useful for those estimating inter-censal aggregated annual migration flows at the district level across the UK (Lomax *et al.*, 2013). However, administrative sources have not been designed explicitly for the purpose of capturing migration and thus suffer from a distinct lack of demographic and socio-economic detail as well as failing to capture short-distance residential moves (e.g. within local authority districts).

In contrast, for the micro-level approach to the study of population mobility, sample surveys represent valuable sources of data with considerable levels of attribute detail, and have traditionally enabled the testing of various hypothesised relationships between individual/household-level characteristics and mobility behaviours and outcomes. However, they are typically characterised by relatively small sample sizes which restrict their usefulness *vis-à-vis* the inclusion of potentially important contextual effects operating at more detailed geographic levels. Similarly, whilst census SARs and longitudinal studies provide rich sources of microdata for undertaking micro analyses, for reasons of respondent anonymity and confidentiality, the samples are again restricted in terms of the geographic detail provided, generally only including national or regional scale geo-identifiers (Gould and Jones, 1996; Dale *et al.*, 2000; Norman and Boyle, 2010). Given its large sample size, extensive geographical coverage and detailed geographic (origin/destination) identifiers, the ROP holds great potential as a source of data that can enable the incorporation of both micro *and* macro-contextual influences on mobility behaviours and outcomes.

As Chapter 2 made clear, much of the literature would suggest that many of the individual/household level factors relevant to residential mobility decision making and outcomes are inextricably tied to complex structural phenomena that interact across various aggregate/spatial scales – for both the origin and destination – from the neighbourhood through to the broader region, nation and possibly beyond. Indeed, to this point, the limitations of the existing migration data landscape have made opportunities for such research extremely limited in the UK. Consequently, with the availability of a sufficiently large-scale geo-referenced microdata source, the ROP, this chapter reviews the (micro/macro) migration modelling approaches traditionally used, before justifying, and explaining in detail, a modelling approach that is deemed most appropriate for the simultaneous estimation and analysis of both micro and macro influences on mobility behaviours and outcomes.

4.2 Macro approaches to modelling migration

In the context of migration analysis, macro theory models can be employed to answer questions relevant to aggregate migration flows to, from or between zones at different geographical scales, for instance relatively short-distance flows between neighbourhoods in a city, or longer-distance flows between districts or regions in a country, or across national borders between countries (Dennett and Wilson, 2013). They allow for a quantification and examination of factors important to migration flow intensities such as the characteristics of origin areas (e.g. labour market, housing market and environment) that generate outflows, the attractiveness of competing destinations and the frictional effect of distance on moves (Stillwell, 2008; Stillwell and Harland, 2010). Champion *et al.* (1998) provide a comprehensive summary of the determinants of migration in GB.

The evolution of macro migration modelling can be traced right back to the early development of Ravenstein's (1885) "laws of migration" wherein the characteristics of different spatial units, and particularly the frictional effect of the intervening distance between them, were seen as fundamental to explaining regional differences in origin-destination migratory flows and wider population redistributions. Indeed, these initial explanations laid the foundation for the early so-called gravity models of the 1940s (Zipf, 1946), which sought to quantitatively measure and test such assumptions though the incorporation of terms relating to differential unit population size, intervening distances and observed flows, most commonly calibrated using log-linear statistical techniques. However, these early gravity models often produced predicted interactions inconsistent with observed flows and thus subsequent mathematical approaches were developed based upon Newtonian gravitational principles (Wilson, 1970; Wilson, 1971), forcing predictions to be consistent with observed flows from each origin and to each destination (Stillwell, 2008). This mathematical tradition of constrained spatial interaction modelling was extended in various ways in migration analysis to allow, for instance, the incorporation of unique origin or destination specific distance decay parameters (Stillwell, 1978) and to account for the potentially destabilising effects of spatial autocorrelation and agglomeration of destinations on the estimation of the distance decay parameter (Fotheringham, 1983; Fotheringham *et al.*, 2001). Parallel to the development of the mathematical formulations of macro theory models of migration streams has been the development of alternatives based on statistical calibration methods. These models bring their own particular advantages. For instance, generalised linear regression approaches, using an appropriate Poisson specification for count data responses, allow for the easy incorporation of additional explanatory variables (e.g. employment rates, housing profiles etc.) which can potentially further improve model fit, whilst at the same time maintaining the benefits seen in similar mathematical approaches, namely constraining the total predicted flows to the total observed flows (Congdon, 1991; Flowerdew, 1991; 2010). Similarly, geographically weighted regression approaches have been used to account for the expected spatial heterogeneity across different zones with respect to the relationship between migration and the predictor variables (Fotheringham *et al.*, 2002). Indeed, whether mathematical or statistical in their tradition, these macro models have been useful for informing our understanding of population dynamics and the evolution of population structures and composition at different spatial levels. Courgeau (1995: 146) has argued that macro models seek to explain: "*migratory streams assuming that the behaviour of migrants is influenced by various characteristics in the departure and arrival areas and by the physical or social distance separating these areas […] it is the characteristics of the areas which alone influence the movements of individuals*". Indeed, whilst macro theory models are designed with the purpose of uncovering large-scale influences on wider systems of movement, for instance whether people move to areas with growing employment prospects or better lifestyle environments, they cannot be used to explore the various individual/household characteristics, behavioural mechanisms and micro motives behind the decision to move itself.

4.3 Micro approaches to modelling migration

Whilst its tradition can be dated back to Rossi's (1955) original study of "*Why families move*", the past 2-3 decades have seen the development of a large number of highly detailed longitudinal and cross-sectional microdata sources, which have in turn encouraged the uptake and application of micro theory modelling approaches to mobility analysis. Indeed, through the use of a family of generalised linear modelling techniques, the most common being those of the binomial and multinomial logistic regression models, it has been possible to explore and test hypotheses pertaining to the central role of different personal characteristics and situations, while holding others constant, for informing the observed variations in different mobility behaviours and outcomes (Cushing and Poot, 2005). In addition to detailed cross-sectional sample surveys, the particular availability of longitudinal panel data has been fruitful in making it possible to link the probability of a migration event occurring to an individual's previous experiences and (often) complex life-course trajectories (e.g. mobility histories, partnerships, employment and housing dynamics) (Courgeau and Lelievre, 2006; Mulder, 2007; Bailey, 2009).

However, whilst micro modelling techniques have been very useful in demonstrating and testing hypotheses at the micro level, there is a danger in considering only characteristics of the individual, and/or household, when analysing mobility behaviours. The danger relates to concerns about omitting from analyses the context in which the behaviours are practiced, a danger which is more formally described as atomistic error or atomistic fallacy (Alker, 1969; Courgeau and Baccaini, 1998; Subramanian *et al*., 2009). Indeed, as was suggested in Chapter 2 and revealed in Chapters 8 and 9, it is fallacious to suppose that mobility behaviours and decisions are developed and informed within a social and economic vacuum devoid of social interactions and routines, local and national institutions, cultural traditions and other place based processes, practices and characteristics. The decision to migrate is likely to depend on a combination of both individual or micro-level characteristics and (perceptions of) macro variables translated into utility functions as documented by Cadwallader (1989).

A somewhat lesser known alternative micro theory modelling approach, which has developed in relative isolation from the statistical techniques above in recent years, is that of the spatial Agent-Based Model (ABM). ABMs are a mathematically derived modelling strategy which, through the incorporation of various micro and macro characteristics, can be argued to hold some potential for micro mobility analysis. Indeed, ABMs aim to explore the complex systems in which autonomous, though interactive, individuals' operate. As computational methods, they seek to provide an explanatory mechanism for the emergence of certain social phenomena, for instance the movement of burglars in a city (Malleson *et al.*, 2008), the discriminatory residential mobility behaviour linked to ethnic diversity in the neighbourhood (Schelling, 1969; 1971) or the housing choice of residential movers in neighbourhoods undergoing regeneration (Jordan *et al.*, 2011). In the latter, and perhaps most relevant example, household agents are assigned attributes including age, housing tenure and social class while macro characteristics such as letting rates, job opportunities, mortgage conditions and lending facilities are also considered, all with the potential to vary over time. The purpose of this simulation is to observe how agents' (households') mobility behaviour and housing choices are informed by interactions with their environment and the mobility and housing decisions of other agents. The repetition of this simple rule-based simulation results in the emergence of trends in the distribution of households, trends which may reveal potentially important and policy relevant patterns, such as segregation (Jordan *et al*. 2011).

ABMs are deemed unsuitable for the research proposed in this study given their focus on rule-based simulation rather than detailed empirical analysis and hypothesis testing. Moreover, if simulation was to be desired, there are significant obstacles that make the application of an ABM, particularly for the analysis of individual and place variations in residential moves in GB, a rather undesirable proposition. Indeed, the usefulness of ABM relies on the ability of the model to incorporate behavioural rules that reflect real world systems and mechanisms, and whilst these models do incorporate a stochastic element, the nonlinear and often quasi-random nature of individuals and their interactions can make their calibration and validation particularly problematic (Crooks *et al.*, 2008). However, if simulation is required, and such obstacles can be overcome, their potential for future micro theory based mobility research should certainly not be ignored.

4.4 Multilevel approaches to modelling migration

Multilevel modelling is an approach that allows for the rigorous quantitative analysis of patterns, propensities, relationships and differences that can operate simultaneously at different levels of aggregation. In its broadest conception, multilevel modelling is a statistical approach that allows for the realistic recognition of social structure, dependency and context, for informing individual behaviour. Whilst its use in geographical analysis can be dated back to the early 1990s (see Bondi and Bradford, 1990; Jones, 1991a; 1991b), the application of multilevel modelling in the sub-disciplines of residential mobility and population migration has, to date, been very rare though exceptions do exist (Boyle and Shen, 1997; Chi and Voss, 2005). This is surprising given its potential, explained below, to integrate certain parts of the traditional dichotomy of micro and macro approaches to mobility modelling. Indeed, the defining factor behind its rarity is most likely the distinct lack in availability of suitably detailed (geographic and attribute) migration data. Given the characteristics of the ROP, the general multilevel modelling framework, described in detail below, can be seen to contain the necessary technical and substantive complexities required to maximise the utility of the data source for exploring individual and place variations in residential moves.

4.4.1 Modelling individual and place variations: Comparing fixed part and random part expansion

To provide a fundamental understanding of the substantive reasoning behind the benefits of applying multilevel modelling, it is useful to focus on a hypothetical example of a simple two-level situation (Figure 4.1), where individuals i (level 1) are nested (or grouped) within neighbourhoods j (level 2).

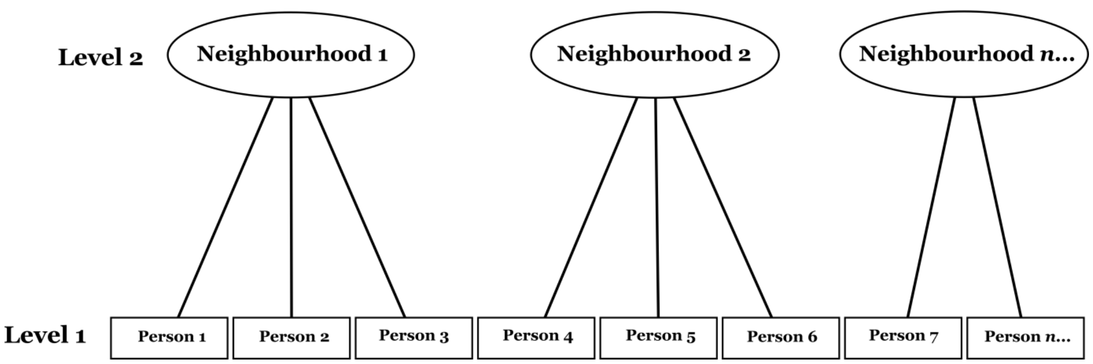


Figure 4.1. A simple two-level hierarchy

Building on this, a bivariate regression model of interest might quantify whether individuals' propensities to move, measured for convenience as a continuous and normally distributed dependent variable, vary according to age (in years). It is possible to run this analysis as a simple single level linear normal theory model (i.e. Ordinary Least Squares (OLS) model) as follows:

$$y_i = \beta_0 + \beta_1 x_{1i} + e_i \tag{4.1}$$

where y_i represents the response variable, a continuous measure of migration propensity for person i, which relates to x_{1i} the value of the explanatory variable which is the age of person i. The estimated coefficient β_1 describes the predicted linear gradient of the relationship between the response and the predictor variable, and in this case represents the average change in y (migration propensity) for a single unit increase in x (age), and the intercept β_0 denotes the point at which this gradient (or line) crosses the y-axis, and gives the average propensity to move where x(age) is equal to 0. The error term (or residual) e_i reflects the extent to which the predicted 'modelled' outcome deviates from the actual 'real-world' outcome for each person i, and is summarised by a single variance term σ_e^2. Fundamentally, the model is made up of two parts, the fixed part ($\beta_0 + \beta_1 x_i$), which reflects the general systemic component of the average relationship between individual movement propensity and age, and the random part (e_i), which reflects the, assumed to be random, remaining differences in individuals' movement propensities having accounted for age.

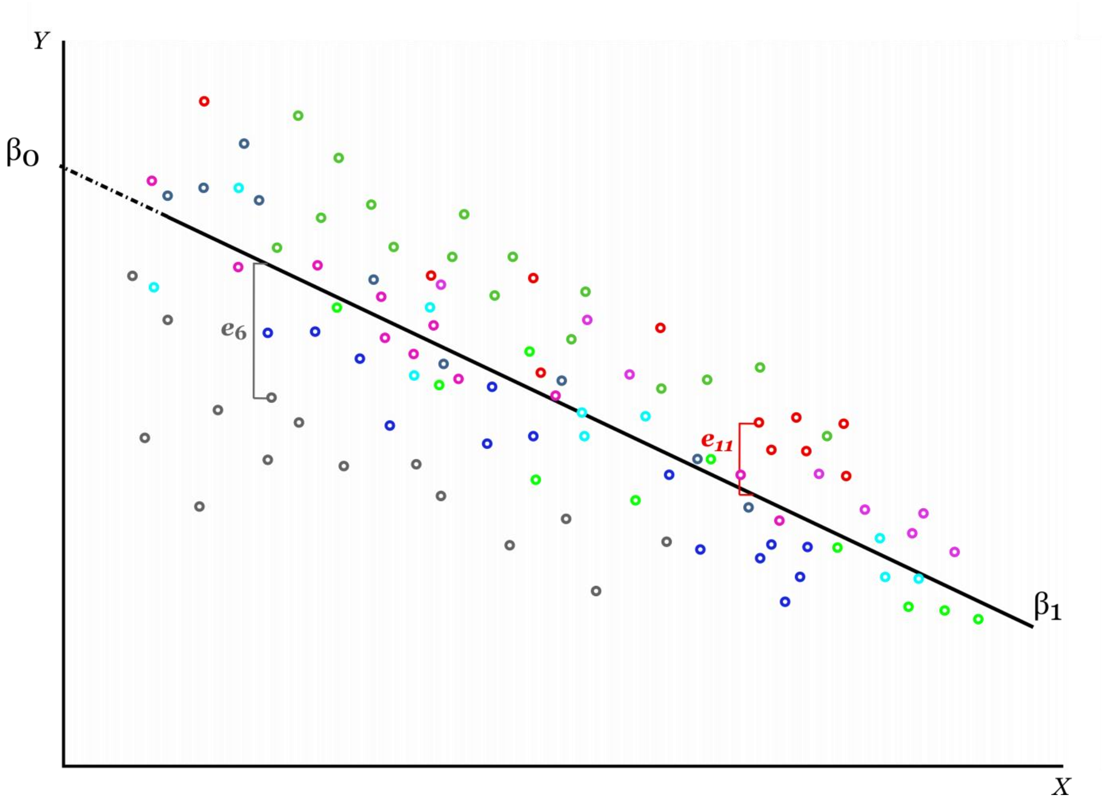


Figure 4.2. Single level regression model[2]

---

2 All figures in this chapter are adaptations of learning materials developed by the Centre for Multilevel Modelling (CMM) at the University of Bristol (http://www.bristol.ac.uk/cmm/).

Figure 4.2 provides a graphical description of this hypothetical single level model where, whilst not recognised in this model, the different colours represent the constituent neighbourhood for which each person is a member, β_0 is the intercept of slope β_1 and for informative purposes e_6 and e_{11} represent the residual model error for persons 6 and 11 respectively. Whilst from a substantive point of view this simple model ignores context and does not allow for an examination of potential neighbourhood differences in migration propensity, it also violates some key assumptions of the simple regression model. Indeed, fundamental to the correct estimation of the regression model is the assumption that all observations (i.e. people) are independently and identically distributed (IID), that is, individuals are expected to come from an unstructured random sample of the population and be completely independent of one another. Where spatial proximity does play a role, and thus individual responses are correlated within contexts, we can expect the estimation of standard errors and significance tests to be overly precise and thus increase the potential of finding statistically significant differences or relationships where none exist (Skinner *et al*., 1989). Moreover, it is assumed that there are to be no trends in the residuals, in this case the residuals should remain constant as age increases and be independent of the response variable (i.e. homoscedastic).

Given that in this example (Figure 4.2) there appears to be a degree of dependency/clustering according to the neighbourhood where each person lives (e.g. all red individuals are located above the overall average slope), a possible solution could be to include a set of dummy indicator variables within the fixed part of the single level model, a technique known as fixed part expansion and an equivalent to a standard ANOVA (analysis of variance) model (Duncan *et al*., 1998). This formulation would lead to the generation of an intercept for each neighbourhood and would allow, for instance, for the calculation of the average additional differential effect of living in a specific neighbourhood (e.g. j_6) as opposed to the reference neighbourhood (e.g. j_1).

There are, however, serious limitations to this fixed effects approach to modelling contextual variation (Jones and Bullen, 1994). Firstly, from a simple point of view, where the sample includes a large number of neighbourhoods, the estimation of neighbourhood variation will quickly become unwieldy, wherein a separate parameter is required for each neighbourhood, excluding one for use as the reference; thus, if a study included 600 neighbourhoods, 599 separate dummy terms would need to be estimated. Not only is this an inefficient strategy, it is equivalent to fitting a separate regression model of migration propensity and age for each neighbourhood. This approach greatly limits the scope for detailed geographical analysis, wherein it should perhaps more accurately be interpreted as strategy used for controlling-out the *nuisance* of contextual difference rather than treating it as a subject of genuine substantive interest (Jones and Bullen, 1994). Indeed, whilst the fixed effects approach is one way of incorporating contextual differences, it limits the further inclusion of contextual characteristics (level-2 predictor variables), characteristics that may not only be of substantive analytical relevance but also of potential importance for explaining some of the between-neighbourhood variation in the response. This particular limitation is a result of the fact that the fixed effects of each neighbourhood (i.e. the dummy indicator) will be perfectly confounded with any characteristic measured at the level of the neighbourhood, a situation that makes the identification of either variable a mathematical impossibility (Fielding, 2004).

Multilevel modelling is a form of random part expansion, wherein the neighbourhoods in the sample data are assumed to come from a random sample of a far larger normally distributed population of neighbourhoods about which inferences can be made (Jones and Bullen, 1994). Through stochastic expansion, neighbourhoods are treated as a separate level wherein it is assumed that the residual between-neighbourhood differentials, defined by their intercepts, vary randomly around an overall grand mean (β_0), and can be summarised by a single variance term (σ_{u0}^2). Thus in a multilevel model, the hierarchical structure of the data, in this case individuals (level 1) nested within neighbourhoods (level 2), is explicitly incorporated within the modelling framework by simultaneously specifying regression equations at each level of analysis, commonly defined as the micro (individual) and the macro (area) parts. In algebraic terms, the micro part of the model can be defined as:

$$y_{ij} = \beta_{0j} + \beta_1 x_{1ij} + e_{0ij} \qquad (4.2)$$

where y_{ij} is the outcome and represents the measure of migration propensity for individual i in neighbourhood j, β_{0j} is the mean migration propensity score for the jth neighbourhood and β_1 is the average change in mobility propensity for a single unit increase in age (x_{1ij}), and e_{0ij} is the person or level-1 residual term. Thus within this random intercepts model, the estimated intercept for each neighbourhood is calculated as:

$$\beta_{0j} = \beta_0 + u_{0j} \tag{4.3}$$

where u_{0j} estimates the positive or negative additional differential contribution that neighbourhood j has over the modelled intercept for the grand mean propensity to move (β_0), independent of age. Therefore, with the inclusion of just one extra random parameter (u_{0j}), it is possible to generate the differential neighbourhood effects with the additional benefit of being able to generalise and make inferences to a relevant population of neighbourhoods (Kawachi and Subramanian, 2006). The micro part (Equation 4.2) and macro part (Equation 4.3) of the model can be combined by substituting the latter into the former, and grouping them into the fixed and random parts, resulting in a multilevel random intercepts model:

$$y_{ij} = \beta_0 + \beta_1 x_{1ij} + (u_{0j} + e_{0ij}) \tag{4.4}$$

where the response y_{ij} is the sum of both the fixed part ($\beta_0 + \beta_1 x_{1ij}$) and the random part ($u_{0j} + e_{0ij}$). As with the single level model, the residual terms in a multilevel random intercepts model are assumed to be independent of the covariates, and independent of one another, and follow a normal distribution with a mean of zero. Following these assumptions the allowed to vary residual terms can be summarised through the estimation of their variances σ_{e0}^2 and σ_{u0}^2. Indeed, the estimation of the level 1 and level 2 variance is based on the raw residuals, where the raw residual for a neighbourhood r_j is the mean distance of persons in neighbourhood j from the overall regression line, and the raw residual for level 1 units (r_{ij}) is measured as the distance of the individual units from their respective group mean differentials (Jones and Subramanian, 2013). The raw residuals for the neighbourhood level are the same as those that would be calculated in fixed effects model using dummy indicators for each neighbourhood (Snijders and Bosker, 2012). Figure 4.3 provides a graphical representation of the neighbourhood residual for neighbourhood 4 (u_{04}) and neighbourhood 6 (u_{06}) as well as the person specific residual for individual 4 in neighbourhood 4 ($e_{4,4}$), individual 8 in neighbourhood 4 ($e_{8,4}$), individual 9 in neighbourhood 6 ($e_{9,6}$) and individual 11 in neighbourhood 6 ($e_{11,6}$).

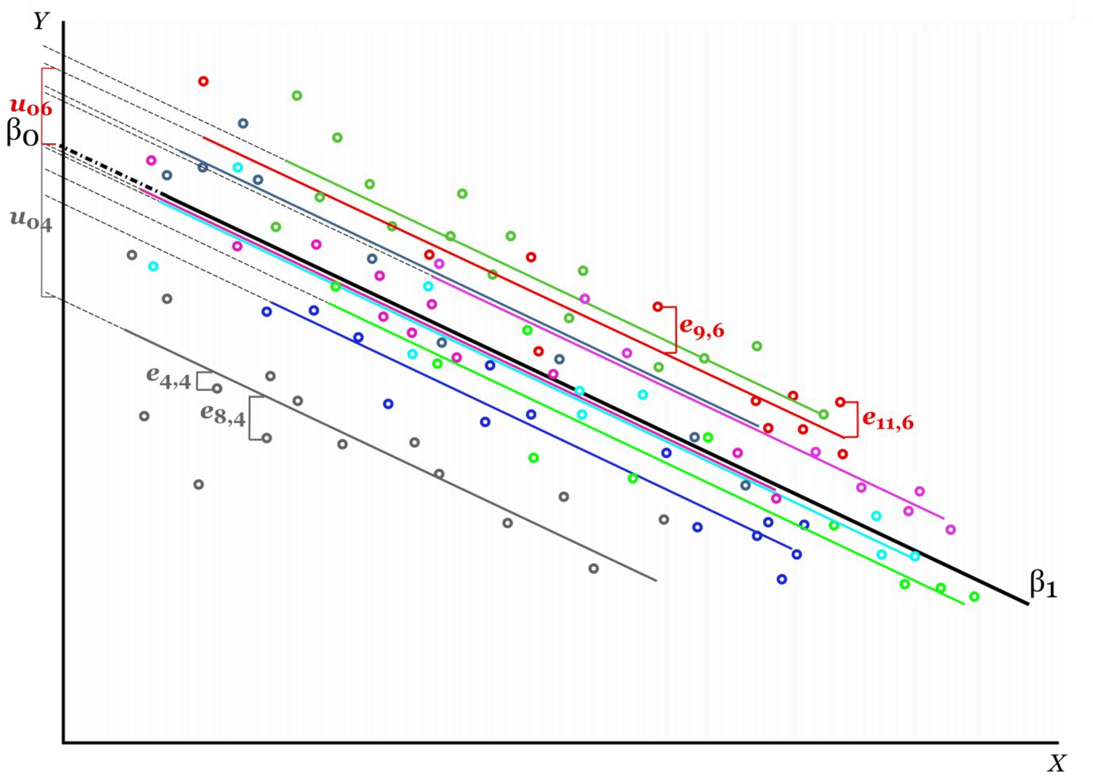


Figure 4.3. Random intercepts regression model

Whilst the estimation of variances is based on the raw residuals and is thus designed to reflect the between-group variance in the population, the calculation of the multilevel random intercepts, that is the estimation of the difference (+ve/-ve) a neighbourhood makes, is more complicated and follows a series of steps designed for the purpose of ensuring reliability through the use of information and distributional assumptions pertaining to the multilevel sample.

Indeed, in a multilevel framework where random-part differentials are included, the additional idea of *shrinkage* is used so as to account for differential group sizes, and the potential impact that this may have on the reliability of the estimated (average) group level residuals, and also allow for the pooling of information derived from the estimation of variability at level 1 and 2 in the model, given that inferences can be made based on their assumed sampling distributions (Jones and Bullen, 1994; Diez Roux, 2002; Jones and Subramanian, 2013).Thus, neighbourhoods with a small sample of individuals from which the average differential can be calculated will be shrunken into the grand mean regression line, neighbourhoods with the smallest sample sizes (i.e. those with the least information from which to base their neighbourhood differential on) will see the greatest shrinkage towards zero, the grand mean across all neighbourhoods. Moreover, the information pertaining to the variance terms at levels 1 and 2 are also important for determining the degree of reliability in the neighbourhood residuals. For instance, where the overall variance at level 1 (σ_{e0}^2) is found to be large, the shrinkage of the level 2 residuals to the grand mean will be greater due to the fact that individual observations will be distributed widely around their neighbourhood line and therefore suggest a degree of inaccuracy in the estimated neighbourhood mean differential. Similarly where the level two variance (σ_{u0}^2) is found to be small, shrinkage will again be greater because the neighbourhood lines are close together and clustered around the grand mean.

Therefore, given the tight distribution of neighbourhood differentials around the overall average, it can be expected that the differential for neighbourhood j should also be close to the grand mean. As a result, the multilevel specification makes use of the information available from the global model to estimate the degree of local reliability (Jones and Subramanian, 2013). The shrinkage most commonly used in multilevel modelling is based on empirical Bayes estimation wherein the group residuals are precision-weighted by multiplying the raw residual of group j (r_j) by its measured reliability, as defined as a group specific shrinkage factor λ_j, which following Snijders and Bosker (2012: 62), can be calculated as:

$$\lambda_j = \frac{\sigma_{u0}^2}{\sigma_{u0}^2 + (\sigma_{e0}^2/n_j)} \tag{4.5}$$

where n_j is the number of people in the given neighbourhood j, σ_{u0}^2 is the between-neighbourhood variance, and σ_{e0}^2 is the within-neighbourhood between individual variance. Consequently, to obtain the estimated shrunken residual for neighbourhood j, the measure of reliability (λ_j) is multiplied with the raw residual (r_j), thus:

$$\hat{u}_{0j} = \lambda_j \cdot r_j. \tag{4.6}$$

The level 1 residual e_{0ij} is calculated as:

$$e_{0ij} = y_{ij} - (\beta_0 + \beta_1 x_{1ij}) - \hat{u}_{0j} \tag{4.7}$$

where the level 1 residual is the actual propensity to migrate (y_{ij}) minus the modelled propensity to migrate ($\beta_0 + \beta_1 x_{1ij}$) minus the estimated level-2 (neighbourhood) residual ($\hat{u}_{0j}$).

A key advantage of random part expansion is its inherent ability to allow for the quantification and partitioning of variance across levels. Indeed, by simultaneously specifying regression equations at each level of analysis, it is possible to generate estimates of dependency between lower level units belonging to the same higher level unit and at the same time explore the extent to which context may influence individual level outcomes.

The variance partitioning coefficient (VPC) (Goldstein *et al.*, 2002; Snijders and Bosker, 2012), also known as the intraclass correlation (ICC) statistic, makes use of both the level 2 variance (e.g. the between-neighbourhood variation) and the level 1 variance (e.g. the within-neighbourhood between-individual variation) in providing a measure of the relative contribution of each level to the total residual variation; in the ongoing example, this is the remaining variation, the left to be explained variation, having accounted for age as a covariate[3]. For the two-level random intercept model, the VPC (ρ) is expressed as the proportion of variation located at the level of the neighbourhood out of the total variation, ($\rho = \sigma_{u0}^2 / \sigma_{u0}^2 + \sigma_{e0}^2$). Thus, where $\rho = 0.05$, 5 per cent of the residual variation is estimated to lie at the between-neighbourhood level, with 95 per cent at the within-neighbourhood, between-person level. The statistic can also be interpreted as representing the degree of similarity in the mobility propensity between two randomly selected people within a neighbourhood. In this case, where ρ is close to 0 the similarity between

[3] A variance components model, or null model, is a special case of a random intercepts model, containing only a constant and no covariates. In this case, the VPC can be used to determine the amount of variance in the response (y) located at each level in the model (Goldstein, 2011).

individuals within a neighbourhood is small, suggesting little contextual dependency and that most variation is at the micro level. Conversely, where ρ approaches 1, the clustering of individuals is implied to be very strong, i.e. individuals within a neighbourhood will be very similar in their mobility propensity, and therefore most variation will be associated with the macro level.

It should be noted that a further method for evaluating the substantive importance of variance attributed at a higher level is the use of coverage intervals. Indeed, based on the assumption that u_j follows a normal distribution, the calculation of a 95 percent coverage interval for the higher level variance ($-1.96 * \sigma_u$ and $+1.96 * \sigma_u$) allows the researcher to get a handle on the additional influence of context by, for instance, comparing the difference in the propensity to move for a typical person in a neighbourhood at the 2.5th percentile of the distribution and a neighbourhood at the 97.5th percentile of the distribution.

Finally, as a way of presenting the hierarchical structure, dependency and clustering assumptions of a multilevel model in summary form, as compared to a single level model, examples of their respective correlation structures can be given. Table 4.1 presents the correlation structure associated with the simple single level linear normal theory model (Equation 4.1) where there are three neighbourhoods containing 10 individuals. As was mentioned above, this model assumes that, having controlled for age, all observations (i.e. people) are IID, that is, individuals are expected to come from an unstructured random sample of the population and be completely independent of one another. Therefore, as shown in Table 4.1, the leading diagonal of the correlation structure represents the correlation of an individual with themselves, and is thus equal to 1, whilst for any pair of different individuals, the correlation is assumed to be zero. In contrast, the correlation structure of the two level random intercepts model (Equation 4.4) relaxes these assumptions and allows for the correlation of lower level units (i.e. people) within a higher level unit (i.e. the neighbourhood), defined by ρ, whilst individuals from different neighbourhoods are assumed to be uncorrelated (i.e. have 0 correlation), having controlled for age (Table 4.2).

Table 4.1. Correlation structure of a single level model

Neighbourhood		1	1	1	2	2	2	2	3	3	3
	Person	1	2	3	1	2	3	4	1	2	3
1	1	1	0	0	0	0	0	0	0	0	0
1	2	0	1	0	0	0	0	0	0	0	0
1	3	0	0	1	0	0	0	0	0	0	0
2	1	0	0	0	1	0	0	0	0	0	0
2	2	0	0	0	0	1	0	0	0	0	0
2	3	0	0	0	0	0	1	0	0	0	0
2	4	0	0	0	0	0	0	1	0	0	0
3	1	0	0	0	0	0	0	0	1	0	0
3	2	0	0	0	0	0	0	0	0	1	0
3	3	0	0	0	0	0	0	0	0	0	1

Table 4.2. Correlation structure of a two-level model

Neighbourhood		1	1	1	2	2	2	2	3	3	3
	Person	1	2	3	1	2	3	4	1	2	3
1	1	1	ρ	ρ	0	0	0	0	0	0	0
1	2	ρ	1	ρ	0	0	0	0	0	0	0
1	3	ρ	ρ	1	0	0	0	0	0	0	0
2	1	0	0	0	1	ρ	ρ	ρ	0	0	0
2	2	0	0	0	ρ	1	ρ	ρ	0	0	0
2	3	0	0	0	ρ	ρ	1	ρ	0	0	0
2	4	0	0	0	ρ	ρ	ρ	1	0	0	0
3	1	0	0	0	0	0	0	0	1	ρ	ρ
3	2	0	0	0	0	0	0	0	ρ	1	0
3	3	0	0	0	0	0	0	0	ρ	ρ	1

4.4.2 Adding more complexity

4.4.2.1 Random slopes

The random intercepts model can be extended so as to incorporate random slopes (or random coefficients), a specification that, in this case, would allow for the person level relationship between age and mobility propensity to vary randomly across the higher level units (e.g. neighbourhoods) around an overall mean effect. Shrinkage is again used for the estimation of the neighbourhood random slopes, where, in addition to evaluations of the neighbourhood sample size and level-1 and level-2 intercepts and slope covariance, the degree of information pertaining to the relevant allowed-to-vary predictor variable is also included. Therefore, in this example, for a neighbourhood with a homogenous age sample, which includes only a small range of ages, and thus has a large sampling variance from which to estimate the differential slope, the shrinkage will be large. Conversely, if a neighbourhood has a heterogeneous age sample, containing a variety of ages, the reliability of the estimate will be greater and the shrinkage to the overall grand mean relationship will be reduced. A discussion and exposition of the matrix algebra necessary for the calculation of the multidimensional shrinkage is provided by Jones and Bullen (1994).

Continuing with the example of age and mobility propensity, age is now represented by a random coefficient (β_{1j}), thus allowing its relationship with mobility propensity to vary across each neighbourhood j. The random intercepts and random slopes model can again be understood to contain both micro and macro parts, with the micro component defined as:

$$y_{ij} = \beta_{0j} + \beta_{1j}x_{1ij} + e_{0ij} \tag{4.8}$$

where the new term β_{1j} is the estimated neighbourhood specific slope term associated with the level 1 predictor age (x_{1ij}), and the subscript j indicates that this term is allowed to vary at level 2. The macro part is defined as:

$$\begin{aligned} \beta_{0j} &= \beta_0 + u_{0j} \\ \beta_{1j} &= \beta_1 + u_{1j} \end{aligned} \tag{4.9}$$

where u_{1j} estimates the positive or negative additional differential contribution that neighbourhood j has on the modelled average slope term (β_1). As with the random intercepts model, the micro and macro parts of the random slopes model can be combined by substituting the latter into the former, and again grouping them into the fixed and random parts:

$$y_{ij} = \beta_0 + \beta_{1j}x_{1ij} + (u_{0j} + u_{1j}x_{1ij} + e_{0ij}) \tag{4.10}$$

The $u_{1j}x_{1ij}$ terms represent another set of neighbourhood level random terms and can again be summarised by their variance σ_{u1}^2. However, with the addition of this extra parameter, the random intercepts and slopes at the neighbourhood level are now assumed to follow a bivariate normal distribution with a zero mean and a variance-covariance structure:

$$(u_{0j}, u_{1j}) \sim N(0, \Omega_u), \text{where} \Omega_u = \begin{bmatrix} \sigma_{u0}^2 & \\ \sigma_{u01}^2 & \sigma_{u1}^2 \end{bmatrix} \tag{4.11}$$

where there are now two variance terms (one for the intercepts σ_{u0}^2 and one for the slopes σ_{u1}^2) and a covariance term (σ_{u01}^2) indicating that the random intercepts and slopes are allowed to covary according to a neighbourhood level, joint distribution. As with the random intercepts model, the between-neighbourhood variance around the grand mean can be derived. However its calculation is more complicated given that the higher level variance is now a quadratic function of a level 1 predictor variable (Goldstein, 2011), in this case age:

$$\sigma_{u0}^2 + 2\sigma_{u01}x_{1ij} + \sigma_{u1}^2 x_{1ij}^2 \tag{4.12}$$

Broadly speaking, if the covariance term is positive then the differences between neighbourhoods will grow with values of x (age), suggesting that the variation between neighbourhoods in the propensity to move for those in the latter stages of life will be greater than for those in younger age groups. Conversely, if the covariance term is negative the between-neighbourhood variance will reduce with increasing values of age, and the slopes will trend towards convergence with the grand mean relationship ($\beta_0 + \beta_1$). Figure 4.4 provides an example of a random intercepts and random slopes model where there is a negative covariance ($-\sigma_{u01}$) between the random slopes and intercepts.

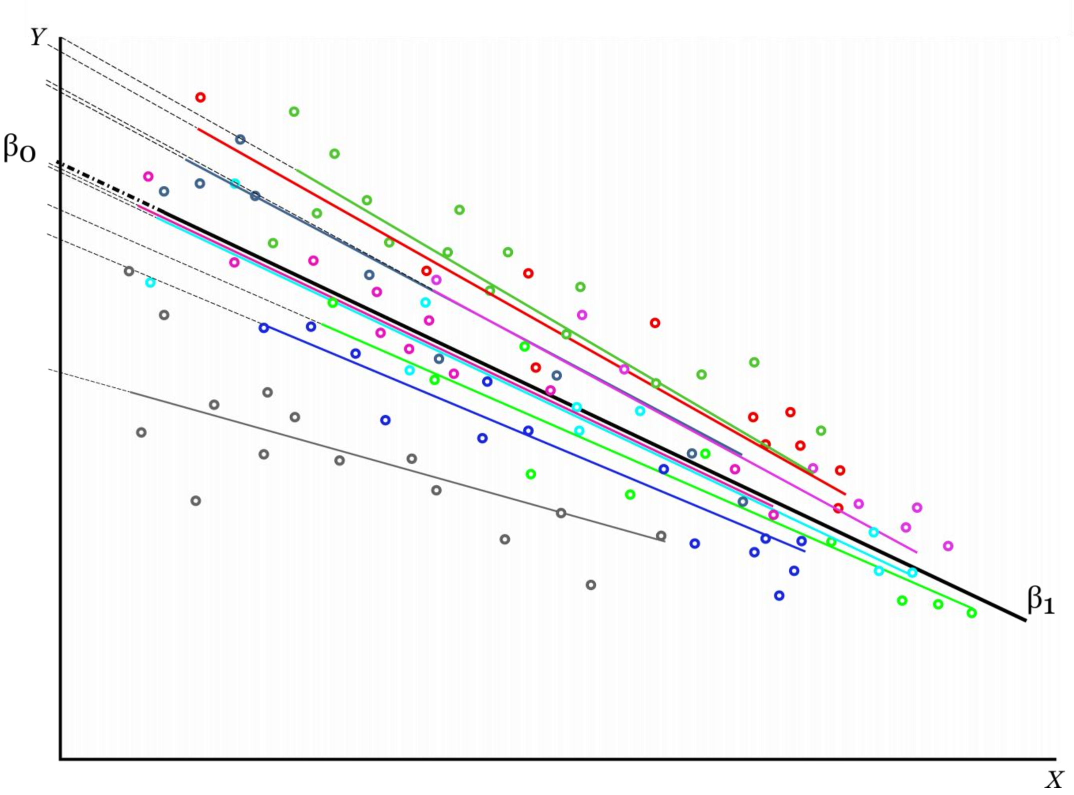


Figure 4.4. Random intercepts and random slopes regression model

4.4.2.2 Higher-level variables and cross-level interactions

The simultaneous analysis across levels means that the effect of place, or context, can be analysed net of the confounding effect of individual and household characteristics – so called compositional variables (Jones and Duncan, 1995). Indeed, the bivariate example above can be extended to include further micro characteristics such as sex, ethnicity, household income and housing tenure, with the remaining variation at the neighbourhood level now conditional on their inclusion. It is also possible to incorporate contextual characteristics, higher level predictor variables, as well as any cross-level interactions that may be of substantive importance for explaining the micro response and thus the residual variance at different levels (Jones and Duncan, 1996; Subramanian, 2004a).

Given the inherent multilevel structure of individuals nested in neighbourhoods, and the fact that neighbourhoods are assumed to come from their own separate random sample of a far larger population of neighbourhoods, for which the correct degrees of freedom can be calculated for use in the estimation of standard errors and so forth

(Jones and Bullen, 1994), multilevel models allow for a more robust specification and estimation of contextual and/or cross-level interaction effects. Contextual variables can be derived from the sample data, by summarising the characteristics of individuals within their higher level units, for instance calculating the mean or proportion across individuals within a neighbourhood. Alternatively, independent macro level variables can be collected, for instance from reliable population census sources, and incorporated into the model at the relevant macro level. Contextual variables may be important for informing the decision to move, for instance high levels of neighbourhood deprivation may encourage individuals to seek alternative residence elsewhere. Moreover, differing neighbourhood demographic and socio-economic profiles may affect the movement propensities of individuals from different age groups in different ways, and therefore the interaction of these micro and macro variables would be important for unravelling such phenomena. The random intercepts and random slopes model of Equation 4.10 can be extended so as to include a contextual neighbourhood level variable and a cross-level interaction between the neighbourhood level variable and the individual level variable:

$$y_{ij} = \beta_0 + \beta_{1j}x_{1ij} + \beta_2 x_{2j} + \beta_3 x_{1ij}x_{2j} + (u_{0j} + u_{1j}x_{1ij} + e_{0ij}) \tag{4.13}$$

where β_2 is the estimated slope term associated with the level 2 predictor variable x_{2j}, and β_3 is the estimated slope term associated with the cross-level interaction between the level 1 predictor variable x_{1ij} and the level 2 predictor variable x_{2j}. To aid interpretation it is recommended that all predictor variables be centred about their mean, though it is particularly beneficial for the interpretation of interaction effects (Snijders and Bosker, 2012).

4.4.2.3 Multilevel models for binary outcomes

For illustrative purposes the dependent variable used in this section so far has been defined as being continuous and normally distributed in nature, whilst in reality, the recorded measurement of migration propensity, whether a person/household has moved or whether they are planning to move, is often based on a binary 0-1 (e.g. 0 = Not moved/ 1 = Moved) outcome. With the dependent variable now defined as a binary outcome, the use of a linear model is no longer feasible for reasons tied to the non-normal distribution of residuals (e_i), the non-linear relationship between the response probability and the predictors and, significantly, the fact that the linear model does not constrain the predicted outcome of the linear equation (Equation 1) to lie between 0 and 1 (Agresti, 2002). Fortunately, as with single level regression, the multilevel regression model for continuous responses can be generalised to handle discrete binary responses through the use of a logit function. The logit function transforms the non-linear relationship between the response probability and the predictors into a linear one, where the conditional probability of y occurring ($P(y_i = 1) = \pi_i$) given a vector of observed predictor variables, X_i, is constrained to lie between 0 and 1.The single level binary logistic regression model with multiple predictor variables ($x_1, x_2, \dots, x_k$) can be written, following Heeringa *et al.* (2010), as:

$$logit[\pi_i(X_i)] = \ln\left(\frac{\pi_i(X_i)}{1 - \pi_i(X_i)}\right) = \beta_0 + \beta_1 x_{1i} + \cdots + \beta_k x_{ki} \tag{4.14}$$

where, in this case, $\pi_i(X_i)$ is the conditional probability of y occurring, that is having moved, given the vector of observed predictor variables, X_i, which, as with normal linear regression, can be measured as dummy categorical variables and/or continuous variables. In the models presented here, β_0 represents the intercept term, which contains all of the reference categories associated with each predictor variable. $\beta_1, \dots, \beta_k$ are the logistic regression coefficients, where, if categorical,$\beta_k$gives the change in the log odds of $y_i = 1$ for a given category k within a predictor variable when compared to the log odds that $y = 1$ for the reference category within that variable. However, when β_k is estimated for a continuous variable, it gives the change in log odds of $y_i = 1$ for a single unit increase in x_{ki}. Once the model is fitted, $\hat{\pi}_i(X_i)$ can be recovered from the log scale through the following function:

$$\hat{\pi}_i(X_i) = \frac{\exp(\hat{\beta}_0 + \hat{\beta}_1 x_{1i} + \cdots + \hat{\beta}_k x_{ki})}{1 + \exp(\hat{\beta}_0 + \hat{\beta}_1 x_{1i} + \cdots + \hat{\beta}_k x_{ki})} \tag{4.15}$$

where $\hat{\pi}_i(X_i)$ now represents the predicted response probability, in this case the probability of having moved, for a person with the specified baseline combination of defined x_i values, values that can be substituted so as to generate response probabilities for different types of people depending on the covariates included. Alternatively, by exponentiating the estimated logits, $\hat{\beta}$, a different interpretation is provided where, for a categorical predictor, $\exp(\hat{\beta})$ (the odds ratio) represents the change in the estimated ratio of the odds of $y_i = 1$ for a given category within a predictor variable, when compared to the odds that $y_i = 1$ for the reference category. Likewise, for a continuous predictor, $\exp(\hat{\beta})$ (the odds ratio) represents the change in the estimated ratio of the odds of $y_i = 1$ for a single unit increase in x. A variant of this single-level model is used in Chapters 6 and 7, though in Chapter 6 an alternative specification, designed for the inclusion of derived sampling weights, is further described.

The extension to a multilevel logistic regression is much the same as for the linear model, where despite there being a non-normal distribution for the random part at level 1, where the level 1 variance is assumed to come from the Bernoulli distribution with mean π_{ij} and a variance $\pi_{ij}(1-\pi_{ij})$, the normality assumptions for the random parts at higher levels remain (Goldstein, 2011). Therefore, the logistic equivalent to the model in Equation 4.13, the full random intercepts and slopes model incorporating a single individual level variable, a single neighbourhood level variable and a cross-level interaction between the two variables, can be written as follows:

$$\ln\left(\frac{\pi_{ij}}{1-\pi_{ij}}\right) = \beta_0 + \beta_{1j}x_{1ij} + \beta_2 x_{2j} + \beta_3 x_{1ij}x_{2j} + u_{0j} + u_{1j}x_{1ij} \qquad (4.16)$$

where $\ln\left(\frac{\pi_{ij}}{1-\pi_{ij}}\right)$ is the log-odds that individual i in neighbourhood j has moved, β_0 is the overall intercept and represents the log-odds that $y = 1$ across all ij units when all predictors are held at their reference (i.e. $x = 0$ and $u = 0$), β_{1j}, β_2 and β_3 have the same meaning as in the linear model (Equation 4.13) except they now reflect changes to the log-odds that $y = 1$. As mentioned, the higher level random part terms maintain their meaning as between-neighbourhood differential random intercepts and random coefficient terms, where the same bivariate normal distribution with a zero mean and a variance-covariance structure shown in Equation

4.11 permits. Again, as with the single level logistic regression, once fitted the predicted response probability $\hat{\pi}_{ij}$ can be recovered:

$$\hat{\pi}_{ij} = \frac{\exp(\beta_0 + \beta_{1j}x_{1ij} + \beta_2 x_{2j} + \beta_3 x_{1ij}x_{2j} + u_{0j} + u_{1j}x_{1ij})}{1 + \exp(\beta_0 + \beta_{1j}x_{1ij} + \beta_2 x_{2j} + \beta_3 x_{1ij}x_{2j} + u_{0j} + u_{1j}x_{1ij})} \tag{4.17}$$

It is important to note that in the multilevel case the contextual effect u is also considered in the prediction of individual responses, and thus when u is held constant, the effect of a change in x is interpreted as the effect for a change in x for individuals within the same neighbourhood (Diez Roux, 2002). Moreover, for a random intercepts logistic regression model, the calculation and interpretation of the VPC is again different from the normal theory equivalent. Indeed, to facilitate an interpretation of the degree of higher level variance, the level 1 variance can be assumed to follow a standard logistic distribution of 3.29 (Snijders and Bosker, 2012), where VPC = $\sigma_u^2/(\sigma_u^2 + 3.29)$[4]. The specification of the model presented in Equations 4.16 and 4.17 forms the basis of the substantive analysis in Chapter 8.

4.4.2.4 Cross-classified structures

Whilst the focus so far has been on the classic hierarchical structure, where lower level units nest perfectly into a higher unit, multilevel models can be specified to incorporate more complex non-hierarchical data structures. The cross-classified version of the multilevel model is one such example, where level 1 units can be simultaneously nested within two higher level units that are themselves exclusive (or non-overlapping) to one another (Jones et al., 1998; Rasbash and Browne, 2001; Fielding and Goldstein, 2006; Goldstein, 2011; Snijders and Bosker, 2012). Again, to provide a hypothetical example, Figure 4.5 depicts a situation where migrants i (level 1) are nested within a cross-classification of their neighbourhood at origin and neighbourhood at destination, where the origin and destination neighbourhoods can be thought of as non-overlapping level 2 units.

[4] Whilst the assumption of a logistic distribution at level one is standard practice, alternative measures of higher level heterogeneity, such the Median Odds Ratio (MOR), are available (see Larsen *et al.*, 2000).

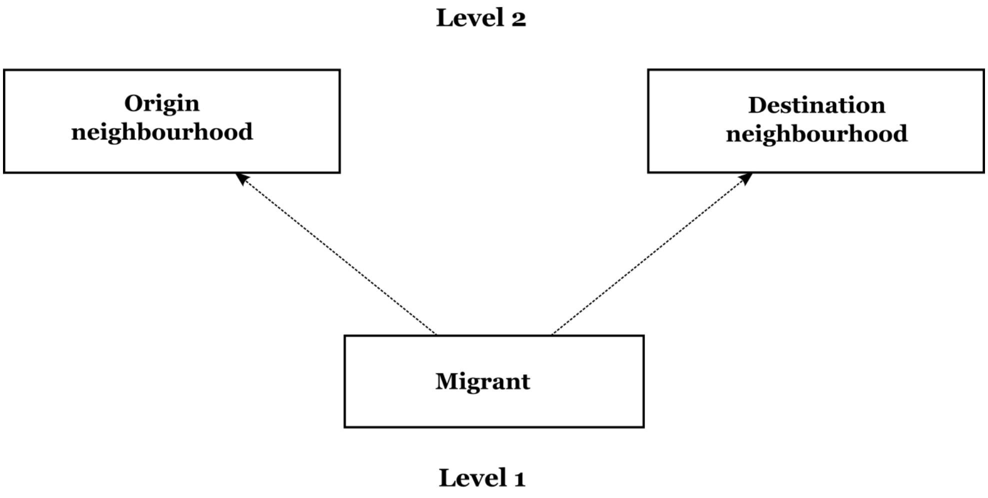


Figure 4.5. Classification diagram of a migrant nested within origin and destination neighbourhoods

In this example, from a statistical modelling perspective, if both origin and destination factors are found to contribute significantly to variations in the outcome, the modelling of only one such context/classification, the origin or the destination, would fail to account for possible confounding effects associated with an underspecified model (Fielding and Goldstein, 2006). For example, if the model only included the multilevel context of the destination, as a simple two-level hierarchy, there is a risk of overstating the importance of the destination as a source of variation at the expense of the origin. More specifically, a simple destination hierarchy would fail to disentangle variation between different destination contexts from that which may be more accurately estimated as variation between different origin contexts. Drawing on the classification notation of Browne *et al.* (2001), the cross-classified model depicted in Figure 4.5 can be presented as follows:

$$y_i = \beta_0 + \beta_1 x_i + u^{(3)}_{orig\,neighbourhood(i)} + u^{(2)}_{dest\ neighbourhood(i)} + e_i$$

$$u^{(3)}_{orig\,neighbourhood(i)} \sim N\left(0, \sigma^2_{u(3)}\right),$$

$$u^{(2)}_{dest\ neighbourhood(i)} \sim N\left(0, \sigma^2_{u(2)}\right), \qquad (4.18)$$

$$e_i \sim N\left(0, \sigma^2_{e_i}\right),$$

where y_i is the observed response for individual i, β_0 is the mean outcome across all origin and destination neighbourhoods, β_1 is the average change in y for a single unit increase in x across all origin and destination neighbourhoods,

$u^{(3)}_{orig\ neighbourhood\ (i)}$ is the additional effect of migrant i's neighbourhood at origin, $u^{(2)}_{dest\ neighbourhood\ (i)}$ is the additional effect of migrant i's neighbourhood at destination, with e_i representing the remaining migrant level residual error. As is the case with the more traditional hierarchical multilevel approaches, all parameters in the random part of the model are assumed to follow a normal distribution with a mean of zero and a constant variance and, for the standard additive cross-classified model, are assumed to be independent across classifications (Goldstein, 2011). Again, to provide some idea of how the dependency and clustering is assumed to operate in a standard cross-classified model, an example correlation structure is given in Table 4.3.

Table 4.3. Correlation structure of a cross-classified model, migrants within origin and destination neighbourhoods

Origin		1	1	1	2	2	2	2	3	3	3
	Person	1	2	3	1	2	3	4	1	2	3
1	1	1	ρ	ρ	0	0	0	ρ_2	0	0	0
1	2	ρ	1	ρ	0	0	0	0	0	0	0
1	3	ρ	ρ	1	0	0	0	0	0	ρ_2	0
2	1	0	0	0	1	ρ	ρ	ρ	0	0	0
2	2	0	0	0	ρ	1	ρ	ρ	0	0	0
2	3	0	0	0	ρ	ρ	1	ρ	0	0	0
2	4	ρ_2	0	0	ρ	ρ	ρ	1	0	0	0
3	1	0	0	0	0	0	0	0	1	ρ	ρ
3	2	0	0	ρ_2	0	0	0	0	ρ	1	0
3	3	0	0	0	0	0	0	0	ρ	ρ	1

Thus, for the cross-classified model in Table 4.3, ρ represents the correlation between migrants from the same origin neighbourhood, ρ_2 represents the correlation between migrants from the same destination neighbourhood, whilst $\rho + \rho_2$ gives the correlation of migrants with the same origin neighbourhood and the same destination neighbourhood and again 0 is the assumed correlation between migrants with different origin and different destination neighbourhoods. A multiplicative version of this model, the random interaction classification model, can be specified so that the variance at one classification is dependent on the variance at another (Goldstein, 2011). In the multiplicative example, a third ρ would be incorporated into the correlation structure representing the origin$*$destination correlation. A brief discussion on the potential of the random interaction classification specification for migration analysis is given in the concluding section of Chapter 9, which is itself based on the standard additive specification of the cross-classified model.

4.4.3 Estimation procedures, model diagnostics and significance testing

There are two broad approaches to the simultaneous estimation of parameters (fixed effects, random effects, variances of the random effects, and residual variance) in multilevel modelling. Traditionally, estimation has been based on the use of frequentist iterative Maximum Likelihood (ML) procedures where the estimated parameter values are those which maximise the probability of observing the data, which is conceptually defined as being equal (or proportional) to the maximum likelihood of the parameters given the data. Two of the most common methods for general ML estimation are Iterative Generalised Least Squares (IGLS) and Residual or Restricted IGLS (RIGLS), both of which follow iterative schemes of repeated cycles of fixed part and random part estimation. The IGLS algorithmic approach was first outlined in Goldstein (1986), where, following Jones and Bullen (1994: 258), the steps for the first iteration of a normal theory model can be summarised as follows:

Step 1 Estimate initial fixed part parameters using a simple OLS model, ignoring the hierarchical structure and assuming $\sigma_{u0}^2 = 0$;

Step 2 Regress the squared residuals, from the OLS in Step 1, on a set of indicators defining the random part structure to produce initial (random part) parameter estimates of the variance and covariances;

Step 3 Use the random part estimates in a generalised least squares analysis to obtain revised estimates of the fixed part parameters.

Following this initial iteration, further iterations work to repeatedly cycle through Steps 2 and 3, revising the fixed and random parts until convergence (where consecutive estimates are sufficiently close together), in which case, if the random part parameters follow the assumed normal distribution, estimates equivalent to ML are produced (see Goldstein, 1986). As a close alternative to IGLS, RIGLS employs a very similar iterative procedure (see Goldstein, 1989) though this time based on restricted (or residual) maximum likelihood (REML) estimates. Through the use of REML, attempts are made to account for the sampling variation of the fixed parameters (i.e. accounting for the loss of degrees of freedom resulting from the estimation of the parameters), which, particularly for small samples, can be expected to produce biased estimates of the random parameters (Snijders and Bosker, 2012; Goldstein, 2011).

More recently, Bayesian Markov Chain Monte Carlo (MCMC) estimation methods have been developed for use in the application of multilevel analysis. The technical and philosophical details of the Bayesian approach are complex and far beyond the scope of this chapter However, useful discussions of Bayesian methods for multilevel modelling are given in Raudenbush and Byrk (2002), Gelman *et al.* (2004) and Congdon (2010). Broadly speaking, in the Bayesian approach to statistics, modelled parameters are not to be regarded as having fixed values, rather they are expected to be unknown and therefore follow their own probability distributions which are informed by both prior beliefs about the parameter (represented in the model by a prior distribution) and evidence from the data (reflected in a conditional distribution, or likelihood). When fitting a Bayesian model, the prior information is combined with the data driven likelihood which results in the formulation of a distribution known as the posterior, the final distribution detailing the degree of support for different values of the parameter. In reality, however, the aim is to develop a multidimensional joint posterior distribution involving all of the parameters (both βs and σ^2s) for which summaries are required. This is made possible through the use of MCMC methods, which approximate the joint posterior distribution by iteratively sampling from what are called the conditional (or marginal) posterior distributions of each parameter, holding the others constant (Browne, 2012).

Following Browne (2012) and Jones and Subramanian (2013), a simple two-level linear regression example based on the most common MCMC procedure, the Gibbs sampler, using five parameters can be outlined. The five parameters are $\beta_{0(0)}, \beta_{1(0)}, u_{(0)}, \sigma^2_{u(0)}$ and $\sigma^2_{e(0)}$, where the subscript in parentheses signifies iteration zero, and where the ultimate goal is to sample and make inferences from the multidimensional joint posterior distribution, in this case $P(\beta_0, \beta_1, u, \sigma^2_u, \sigma_e^2 | y)$, which is only made possible by iteratively sampling from the respective conditional posterior distributions of each parameter, holding the others constant, to a point of satisfactory convergence. It should be noted that in practice it is often not desirable to use informative prior distributions in the calculation of the posterior; rather the aim may be to only use *evidence* from the empirical data collected, in this case the prior distribution is specified as flat/uniform and therefore uninformative (Browne, 2012). Moreover, before the algorithm can begin, the first step is to provide initial starting values for the parameters which, in a multilevel analysis, can be based on IGLS estimates. With initial parameter estimates, the first iteration of the chain involves a loop through the following steps, though the order is not important:

Step 1 Sample a new value for β_0 from its conditional distribution based on the initial estimates of the other parameters: $P(\beta_0 | y, \beta_{1(0)}, u_{(0)}, \sigma^2_{u(0)}, \sigma^2_{e(0)})$, to generate $\beta_{0(1)}$;

Step 2 Sample a new value for β_1 from its conditional distribution based on the initial (and/or revised) estimates of the other parameters: $P(\beta_1 | y, \beta_{0(1)}, u_{(0)}, \sigma^2_{u(0)}, \sigma^2_{e(0)})$, to generate $\beta_{1(1)}$;

Step 3 Sample a new value for u_0: $P(u_{(0)} | y, \beta_{0(1)}, \beta_{1(1)}, \sigma^2_{u(0)}, \sigma^2_{e(0)})$, to generate $u_{(1)}$;

Step 4 Sample a new value for σ^2_u: $P(\sigma^2_{u(0)} | y, \beta_{0(1)}, \beta_{1(1)}, u_{(1)}, \sigma^2_{e(0)})$, to generate $\sigma^2_{u(1)}$;

Step 5 Sample a new value for σ^2_e: $P(\sigma^2_{e(0)} | y, \beta_{0(1)}, \beta_{1(1)}, u_{(1)}, \sigma^2_{u(1)})$, to generate $\sigma^2_{e(1)}$;

Step 6 Compute e_{ij} by subtraction.

These steps are repeated over and over again with newly generated values replacing the starting values from the previous step. This procedure generates a chain of values for each parameter, hence the Markov chain, which are deemed equivalent to drawing a random sample of values for each parameter from its probability distribution (Browne, 2012). A specified number of initial iterations of the chains are discarded, in a stage of burn-in, to reduce the influence of the initial IGLS estimates and to allow for the chains to settle as they converge towards their posterior distributions (Browne, 2012; Jones and Subramanian, 2013).

Unlike the deterministic convergence of the IGLS algorithm, convergence of MCMC parameters is a subjective matter which the researcher must decide upon; indeed parameter chains can be run for as long (e.g. 1,000,000s of iterations) or short (e.g. 1,000s of iterations) as necessary. Convergence of the remaining iterations, post burn-in, can be assessed with the use of a series of diagnostic tools, checking for serial autocorrelation/dependence and trending in chains that can result in unstable estimates (see Browne, 2012); and following the good practice recommendations of Draper (2006) and Jones and Subramanian (2013)[5]. When convergence is reasonably assumed, summary statistics for the iterations are calculated, where a parameter's point estimate and standard error, both equivalents to frequentist ML estimates (Browne and Draper, 2006),are given by the mean and standard deviation of the parameter's chain. The MCMC procedure for discrete outcomes is based on the more general Metropolis Hastings sampler (see Browne and Draper, 2000; Goldstein, 2011; Browne, 2012), which again repeatedly simulates to create parameter chains that reflect draws from the posterior distribution.

In general, multilevel models for continuous responses are often fitted using the IGLS or RIGLS procedures as they are proven to provide reliable and fast estimation (Goldstein, 1986; Goldstein, 1989). For models with a discrete outcome, simulation studies have shown MCMC estimation techniques to be generally more reliable in terms of parameter estimation (Rodriquez and Goldman, 2001). Moreover, in cases where there are few higher level units or where the models are particularly complex, involving cross-level interactions and/or cross-classified structures, MCMC methods are again recommended, over the more traditional forms of estimation (Goldstein, 2011; Browne, 2012; Stegmueller, 2013). Indeed, for a standard cross-classified model, MCMC methods treat each classification unit (residual) as an additive term

---

[5] Whilst their description is beyond the scope of this review, additional parameter expansion methods, including hierarchical centring and orthogonal parameterisation techniques, can be used to improve the efficiency of parameter estimation and thus increase the speed to convergence, a detailed description of these methods is given in Browne (2012).

in the model therefore making it is no more or less complicated than fitting a nested model using MCMC. Therefore, following Goldstein (2011: 251), the MCMC estimation procedure for the cross-classified model as shown in Table 4.3 and Equation 4.18 is:

Step 1 Sample a new set of fixed effects (β);

Step 2 Sample a new set of origin neighbourhood residuals $(u^{(3)}_{orig\ neighbourhood})$;

Step 3 Sample a new set of destination neighbourhood residuals $(u^{(2)}_{dest\ neighbourhood})$;

Step 4 Sample a new origin neighbourhood classification variance $(\sigma^2_{u(3)})$;

Step 5 Sample a new destination neighbourhood classification variance $(\sigma^2_{u(2)})$;

Step 6 Sample a new level 1 variance (σ^2_e);

Step 7 Compute the level 1 residuals (e_{ij}) by subtraction.

Again, the order to these steps is not important.

Finally, for the assessment of model fit and the significance of fixed and random part parameters, a number of alternatives methods are available. To test the significance of individual parameters or the contribution of sets of parameters when holding everything else constant, individual and grouped Wald tests can be employed. Broadly speaking, non-significance in the individual Wald test suggests that the change associated with the variable of choice is not significantly different from zero, which, in the context of the examples above, can suggest that the variable is not an important predictor of migration propensity, given the other variables included in the model. The grouped parameter Wald test is similar but as the name suggests, involves assessing the contribution of a set of parameters, be they multiple dummy parameters associated with a categorical variable, interactions between variables, or quadratic variance parameters associated with the specification of a random coefficient (Heeringa *et al.*, 2010; Snijders and Bosker, 2012).

It is also possible to check the overall model fit and develop a strategy for measuring improvement in the model fit. For discrete response single level models and normal response multilevel models, deviance statistics (-2∗loglikelihood) can be used to measure how much unexplained information remains after a model is fitted, being roughly approximate to the residual sum of squares in a standard multiple regression (Field *et al.*, 2012). A smaller deviance statistic suggests fewer unexplained observations within the model. The model improvement in the first instance is the difference between the null deviance (constant only model) and the residual deviance (fitted model), both of which follow a Chi-square distribution making it possible to calculate the significance of this value. However, the effect of adding/removing variables on the model fit can also be analysed in this manner by checking the improvement in Model 2 (full suite of variables) when compared to Model 1 (reduced variables).

The Akaike Information Criterion (AIC), allows for checks on the improvement in model fit while effectively penalising the model that contains more explanatory variables, and therefore fewer degrees of freedom (Agresti, 2007; Field *et al.*, 2012). Without penalising, the simple addition of a further variable would increase the model fit while failing to account for the additional complexity the added variable brings. Unfortunately, for discrete response or cross-classified multilevel models the traditional estimation procedures do not allow for the reliable calculation of deviance statistics (Jones and Subramanian, 2013). However, MCMC procedures do provide an equivalent to the AIC based on estimated degrees of freedom. The Deviance Information Criterion (DIC) coefficient again penalises for model complexity, where, when comparing relevant models, a lower value of DIC suggests a better fit (see Spiegelhalter *et al.*, 2002; van Der Linder, 2005; Browne, 2012). Indeed, according to Spiegelhalter *et al.* (2002) a reduction of just 3-7 units should be considered as a potentially important difference in model fit. Given the various benefits mentioned here, the models used in Chapter 8 (a discrete response multilevel model) and Chapter 9 (a cross-classified multilevel model) are fitted using MCMC estimation procedures.

4.5 Summary and conclusions

This chapter has provided a review of the traditional micro-level and macro-level theory modelling approaches used for the analysis of population movement in GB and in doing so has attempted to justify, and explain in detail, a multilevel modelling approach that is deemed most appropriate for maximising the utility of a large-scale geo-referenced microdata source for analysing individual and place variations in movement behaviours and outcomes. Multilevel modelling is a statistical approach that recognises the social dependencies and contextual effects (ecological or area-level) as well as the micro respondents, and allows for the simultaneous analysis of both. Whilst naturally occurring social structures are often a problem for traditional single-level statistical analyses, invalidating the assumptions of independence of observations and randomness, multilevel models exploit the dependencies and clustering of units, identifying the degree of correlation within and between areas and uncovering the extent to which different areas or places vary in their effect on the phenomenon of interest, having taken into account their composition (micro-level characteristics).

As has been shown above, multilevel models provide a flexible framework from which to analyse detailed geo-referenced data; for instance, whilst it is possible to simultaneously analyse individual (e.g. age, sex, ethnicity, income) and place (e.g. neighbourhood deprivation, ethnic mix, population (in)stability) characteristics and cross-level interactions, it is also possible to observe how particular micro-level characteristics of interest (e.g. the length of stay at an address) vary geographically with regard to their effect on the response variable (e.g. the propensity for future residential movement) (see Chapter 8). Moreover, more recent extensions to the traditional hierarchical models have also made it possible to analyse alternative complex structures. For example, cross-classified models make it possible to simultaneously model non-overlapping contexts, for instance migrant origins and destinations (see Chapter 9), and explore the relative importance of origin and destination context whilst controlling for the possible confounding effects that might occur if one or the other were omitted from the analysis. Consequently, multilevel modelling represents the approach of choice used in the major substantive analytical chapters but an alternative model strategy designed for the purpose of data validation is described and applied in Chapter 6.

Chapter 5
Data validation: Descriptive-based benchmarking

5.1 Introduction

This chapter is primarily concerned with detailing the initial cleaning and benchmarking exercises performed on the raw ROP microdata. The benchmarking in this chapter employs relatively simple empirical and descriptive-based methods for comparing the raw ROP sample distributions with those of alternative population data, namely Census 2001, PR-NHSCR, the APS, and the Acxiom Ltd. Aggregate Data product. Whilst the overall project aims are best answered using the model-based strategies discussed in the previous chapter, there is value in exploring the potential of the ROP for the more empirical examination of population mobility patterns in GB. Indeed, whilst descriptive-based benchmarking can be useful in uncovering bias in the different sub-sample distributions of the raw ROP samples, it can also be useful for informing an assessment of how successful the samples are in reflecting real population mobility characteristics and behaviour, as measured by the alternative population data sources.

The chapter presents three separate forms of descriptive-based benchmarking. The first (Subsection 5.3) involves aggregate level benchmarking, exploring the correlation of aggregate migration flows derived from the raw ROP with those in the 2001 Census and PR-NHSCR data. The second (Subsection 5.4) is focussed on micro level benchmarking, selecting certain key micro level characteristics (age, housing tenure and ethnic group), and evaluating their raw sample distributions and mobility patterns as compared to the known population distributions and expected mobility patterns. Finally, given the detailed geo-identifiers and the relatively large raw ROP sample, spatial benchmarking is further employed to explore the value of the raw samples for more general substantive empirical mobility analysis, in this case focussing on the patterns to spatial mobility across deprivation deciles at the district level (Subsection 5.5). However, to begin with, Subsection 5.2 presents the details of the significant data preparation and cleaning exercises that are required in order to make the raw ROP samples useable for the population mobility analyses presented in the later chapters.

5.2 Data management

Excluding the responding household's current postcode address, which is cleaned and imputed using the latest Royal Mail PAF, Acxiom Ltd. delivers the ROP data in raw format. As a result, subsequent concerns surrounding missing values and/or 'impossible' values are left for the end user to decide on. Indeed, given the ROP's formation as a voluntary postal survey, the issues of missing values and/or unusable values are commonplace. Thus, in order to maximise the utility of these data for population mobility analysis, a thorough programme of data preparation and cleaning is required, seeking to retain as much information as possible whilst paying particular attention to the critical address identifiers (origin and destination) that are fundamental in allowing for the benchmarking and validation of this data source as well as the subsequent substantive analyses. Whilst five ROP cross-sections were delivered for use in this project, only three are used in practice due to inconsistencies between the surveys in terms of the questions asked (see Chapter 3, Table 3.4). Therefore, the programme of data preparation and cleaning described herewith, as well as the benchmarking in the latter subsections of this chapter, was applied to all three of the usable cross-sections (January 2005, January 2006, and January 2007). However, due to the limited space available, only the results for the raw January 2005 ROP are reported here.

Whilst each respondent's current postcode address is cleaned by Acxiom Ltd. pre-delivery, any previous address data are left completely unformatted. The ROP provides the previous address data as two separate variables, the postal in-code (postcode sector) and the postal out-code (postcode district), which, once combined, provide the full postcode address of the respondent's previous residence. For the raw January 2005 ROP data, there are approximately 108,000 (26%) records with *usable* out-code responses and 103,000 (25%) records with an in-code response, be it usable or not. This high level of nonresponse may be driven by a great many contributing factors; however, it is likely that the nonresponse will be particularly susceptible to the period of time spent at their previous residence (i.e. people who spent several years at their previous address may find it easier to recall a full postcode address than those who spent just a short time there) and the duration of time since they moved to their current address (i.e. people who moved to their current address more recently may find it easier to recall the details of their previous address). Indeed, these issues are further complicated by the fact that the response to the question on previous address is unconstrained (see Chapter 3, Figure 3.2) and, in reality, open to around 2.5 million postcode combinations (Phelps, 2011). Consequently, a great deal of attention is required on this variable before it can be deemed suitable for use in the following benchmarking exercises and analysis chapters.

The first task is to check which postal out-codes are valid, which appear to be mis-specified and which are broken/incomplete. It is imperative that out-codes are checked first as it is not possible to generate any geography with postal in-codes, indeed, in-codes are only valuable if they are attached to a valid postal out-code. The checking was performed by cross-referencing the response out-codes against a full list of postal areas used by the Royal Mail (Raper *et al.*, 1992) and included in the Office for National Statistics Postcode Directory (ONSPD). A set of codes were produced to indicate those that were usable, those that required cleaning/interpretation and those that were unusable (Table 5.1). Those that are delivered in a usable format were coded ‘1’, those that only included the postal area (e.g. LS for Leeds) were coded ‘2’, those where no information was recorded were coded ‘-99’, and those that are unusable (e.g. including impossible combinations or characters) were coded ‘-9’. In addition to these basic codes, 19 additional codes (3-21) were produced, with reference to Raper *et al.* (1992) and the ONSPD, suggesting a requirement for further bespoke cleaning/interpretation.

Table 5.1. Codes deployed in the cleaning/interpretation of Acxiom ROP previous address 'out-codes'

-99: Non movers	**11**: Interpreted, 1P (1P – IP for Ipswich)
-9: Moved but broken/wrong code	**12**: Interpreted, 1V (1V – IV for Inverness)
1: Usable format	**13**: Interpreted, L5 (L5 – LS for Leeds)
2: Only Postal Area	**14**: Interpreted, LV (LV - LU for Luton)
3: Interpreted, Extra 0 (e.g. B02 – B2)	**15**: Interpreted, P0 (P0 – PO for Portsmouth)
4: Interpreted, S0 (S0 to SO for Southampton)	**16**: Interpreted, T5 (T5 – TS for Cleveland)
5: Interpreted, 0L (0L – OL for Oldham)	**17**: Interpreted, W5 (W5 – WS for Walsall)
6: Interpreted, CU (CU – CV for Coventry)	**18**: Interpreted, Y0 (Y0 – YO for York)
7: Interpreted, 0X (0X – OX for Oxford)	**19**: Interpreted, CRO (CRO – CR0 for Croydon 0)
8: Interpreted, C0 (C0 – CO for Colchester)	**20**: Interpreted, B53, ... (B53 – BS3 for Bristol)
9: Interpreted, HV (HV – HU for Hull)	**21**: Interpreted, WU (WU – WV for Wolverhampton)
10: Interpreted, 1G (1G – IG for Ilford)	

The numbers involved in the cleaning and interpretation exercise for the January 2005 ROP are summarised in Table 5.2, from which it is clear that the cleaning process was successful in boosting the numbers of usable out-codes. Excluding those coded '2' (postal area only), the exercise has increased the number of usable previous out-codes by 4.7% for the January 2005 ROP dataset.

Table 5.2. Counts and percentages for codes used in the cleaning/interpretation of previous address out-code data: Raw UK January 2005 ROP

Code	Count	Percentage	Code	Count	Percentage
-9	3,626	3.24	11	30	0.03
1	107,019	95.61	12	7	0.01
2	337	0.30	13	44	0.04
3	92	0.08	14	20	0.02
4	101	0.09	15	196	0.18
5	12	0.01	16	24	0.02
6	38	0.03	17	27	0.02
7	11	0.01	18	112	0.1
8	102	0.09	19	32	0.03
9	10	0.01	20	55	0.05
10	18	0.02	21	18	0.02
Total usable out-codes		111,931			
Total unusable or not provided		299,394			
Total		411,325			

With the cleaning of the previous address postal out-codes complete, it is possible to join the out-codes and in-codes to create full postcodes, from which exact national grid-reference coordinates can be obtained, distances from origin to destination calculated and aggregate geographies joined. To provide some context in terms of residential mobility, following the cleaning of postal out-codes, the number of recent movers (12 months at current address) with usable full postcode address identifiers at both residential origin and destination is just over 10,000 (or 2.5% of the raw UK sample) for the January 2005 ROP. For the joining of aggregate geographies, the online geography matching and conversion tool GeoConvert (http://geoconvert.mimas.ac.uk/), part of the UK Data Service Census Support programme. For the exercises presented in this chapter, the aggregate geography derived for use is that of the local authority district (LAD) level, however, smaller spatial units and functional geographies were also generated for use in the validation and analysis chapters presented later in the thesis. Primarily, the LAD level of geography was chosen for use in the aggregate and spatial benchmarking exercises in this chapter because it represents the lowest level of geography for which the comparative PR-NHSCR produces data. However, in addition to this, when applying descriptive-based benchmarking exercises, the ROP samples are more robust to small number problems when aggregated to the level of the district, as opposed to the more detailed geographies that could have been applied.

With the addition of a number of aggregate geographies, a further requirement is to define what constitutes a 'mover' as opposed to a 'non-mover/stayer'. As was mentioned in Chapter 3, the defining of such terms vary from study to study, and typically rely on being operationalised via the use of predefined geographical (administrative/political) boundaries and the specification of a time period within which the move can take place. For instance, the PRDS identifies a migrant as a person whose change in postcode takes them across either a former HA or LAD boundary (Jefferies *et al.*, 2003), whereas in the Census 2001, it is possible to define a migrant as anyone who changes address in the 12 months prior to census completion, regardless of having moved across a predefined census area boundary, thus making it possible to explore moves within as well as between different geographical units. With regards the January 2005 ROP, five operational definitions can be derived (Table 5.3). The first definition is based on the response to the question asking for the year that the respondent moved to their current address (see Chapter 3, Figure 3.2). Thus, based on this limited information, those that provided a year of move are defined as a 'mover' while those who did not were classified as a 'non-mover'. For the second definition, those who provided a (post-cleaning) usable previous postal out-code, from which some idea of geographical mobility from one place to another can be derived, are defined as 'movers' and those who did not, as 'non-movers'. The third definition is a more restricted version of the first definition, which only includes those individuals who report having moved in the 12 months prior to the survey completion date (January 2005). The fourth definition is again derived from the first and second definitions, where those who report having moved in the 12 months prior to the survey date and who provide a usable out-code are defined as a 'mover'. Finally, definition 5 is the same as definition 4 but uses those records that provide full and usable postcode address identifiers; with these attributes it is possible to accurately measure the location of current and previous addresses and thus explore and benchmark the sample flows within and between different geographical units over a 12 month period. Thus, definition 5 is used in the aggregate level benchmarking in Subsection 5.3 as well as the spatial benchmarking in Subsection 5.5, although it should be noted that the actual numbers used in these subsections are slightly smaller due to the exclusion of moves to and from Northern Ireland. As was discussed in Chapter 3, the ROP only collects a very small sample for Northern Ireland (e.g. January 2005 ROP n = 1,584) and is thus particularly susceptible to small number issues. Indeed, given this restricted raw sample, the aggregate products produced by Acxiom Ltd. are not inclusive of Northern Ireland (Thompson *et al.*, 2010). For the micro level benchmarking of Subsection 5.4, definition 3 (Table 5.3) is used.

Table 5.3. Five possible mover definitions and their respective counts: Raw UK January 2005 ROP

Definitions:	**1.Ever moved based on YoM**	**2. Ever moved based on Upoc**	**3. Moved in last year based on YoM**	**4. Moved in last year based on YoM & Upoc**	**5. Moved in last year based on YoM & Upc**
Cases matching definition	328,158	107,967	17,435	12,232	10,284
Cases remaining (defined as non-movers)	83,167	303,358	393,890	399,093	401,041
Total	411,325	411,325	411,325	411,325	411,325

N.B. YoM = Year of move; Upoc = Usable previous out-code; Upc = Usable previous postcode.

5.3 Aggregate level benchmarking: District migration counts

As mentioned above, whilst the aggregate level benchmarking has been carried out for all ROP datasets, only those for the January 2005 survey are included for reasons of space. The following aggregate level benchmarking exercise involves comparisons of total inflows over a 12 month period at the LAD level. The sources chosen for the aggregate level benchmarking include the Census 2001 SMS, PR-NHSCR 2005 and the Acxiom Aggregate Data 2005. While the PR-NHSCR data can be considered as less reliable than the comparable Census 2001 data, it is thought important to include comparisons with data that are temporally more consistent. The Acxiom Aggregate data comparisons are also important because they could provide an indication of how much data manipulation has been undertaken by Acxiom in the production of this apparently representative aggregated data. The comparisons against Census 2001 and Acxiom Aggregate data include intra and inter-LAD flows for GB. The Acxiom ROP intra-LAD flows were removed for comparisons against the PR-NHSCR data as this data source does not record intra-LAD flows. Moreover, this comparison is conducted for England and Wales only as the PR-NHSCR data for Scotland remains to be harmonised with the rest of GB (Rees *et al.*, 2009). The decision to focus solely on inflows in this section is based on the fact that the Acxiom Aggregate data only provide district inflow totals, thus making any alternative comparison impossible.

5.3.1 Validation against Census 2001 inflows

In terms of the bivariate regression and scatterplot (Figure 5.1), the results are quite positive with 53.5 per cent of the variation in Census 2001 inflows reflected by the January 2005 Acxiom inflows, and the Pearson correlation coefficient is 0.73. Figure 5.2 represents an assessment of the residuals from a bivariate regression (Figure 5.2) comparing LAD in-migrant counts for the January 2005 ROP and Census 2001. The histogram features a normal curve that represents the probability (i.e., the density) for a given value from a normal distribution of known mean and standard deviation (Field *et al.*, 2012); in this case the mean and standard deviation are calculated using the residuals of the bivariate regression in Figure 5.1. While the distribution is not too far from the normal distribution, when it is compared to the estimated normal curve, a certain degree of positive kurtosis is revealed. The Q-Q plot compares each given value from the sample with the expected value that the score should have if it followed a normal distribution (Field *et al.*, 2012). In a case where the data do in fact follow a normal distribution, the Q-Q plot would show a perfect diagonal line. With this in mind, when focusing on the Q-Q plot, there is clear evidence of non-linearity which is exemplified by the lag towards the lower residual values. Finally, the scatterplot at the bottom of Figure 5.2 shows the raw January 2005 ROP data compared to the residuals of the bivariate model in Figure 5.1. Again, further issues of clear heteroscedasticity (uneven variance) as well as a number of apparent outliers and leverage points become apparent.

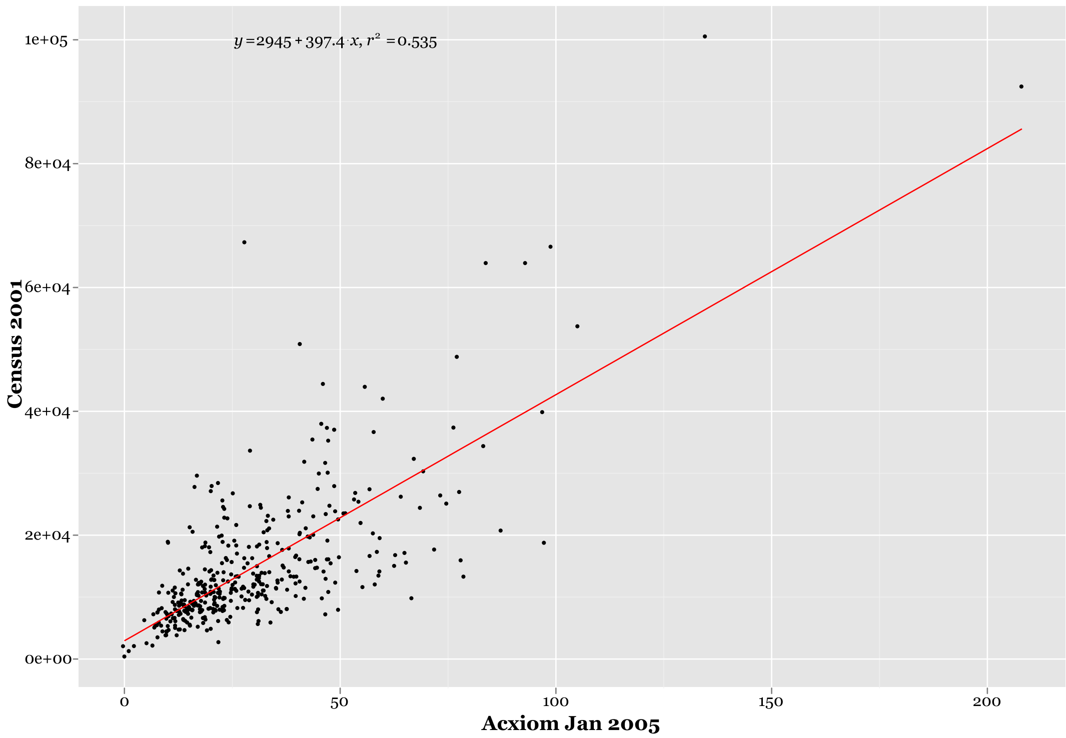


Figure 5.1. Scatterplot showing the relationship between Acxiom January 2005 ROP and Census 2001 in-migrants to LADs

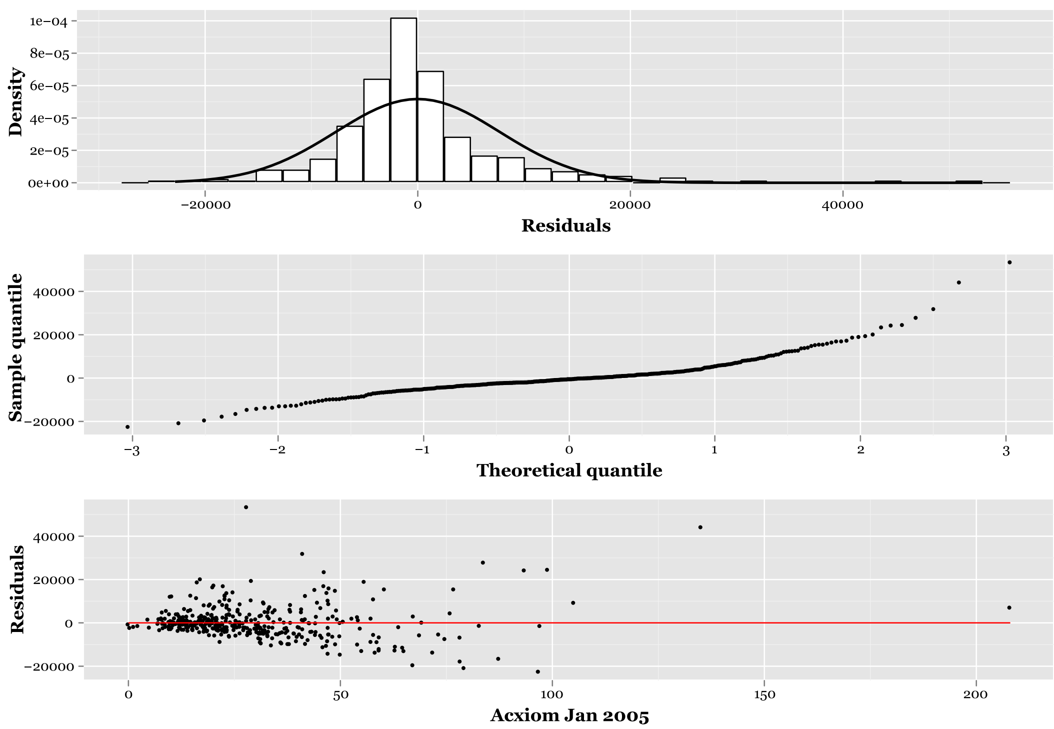


Figure 5.2. In-migrants to LADs data from Acxiom January 2005 ROP and Census 2001: Tests for normality

Given that the bivariate regression in Figure 5.1 does not meet a number the necessary OLS assumptions (see Chapter 4), log10 transformations are used on both variables. It is clear from Figure 5.3 that the log10 transformations have reduced the kurtosis and heteroscedasticity and greatly improved the normality of the residuals. Figure 5.4 shows the scatterplot for the logged variables, and suggests an improved model fit with 55.9 per cent of the variation explained. Interestingly, a number of extreme outliers, identified here as having a standardised residual of ≥ ±3, can be identified despite the transformation. It is apparent that when compared to Census 2001, Acxiom's January 2005 ROP has a significant under-count for Glasgow City and significant over-counts for the Isles of Scilly and Berwick-upon-Tweed.

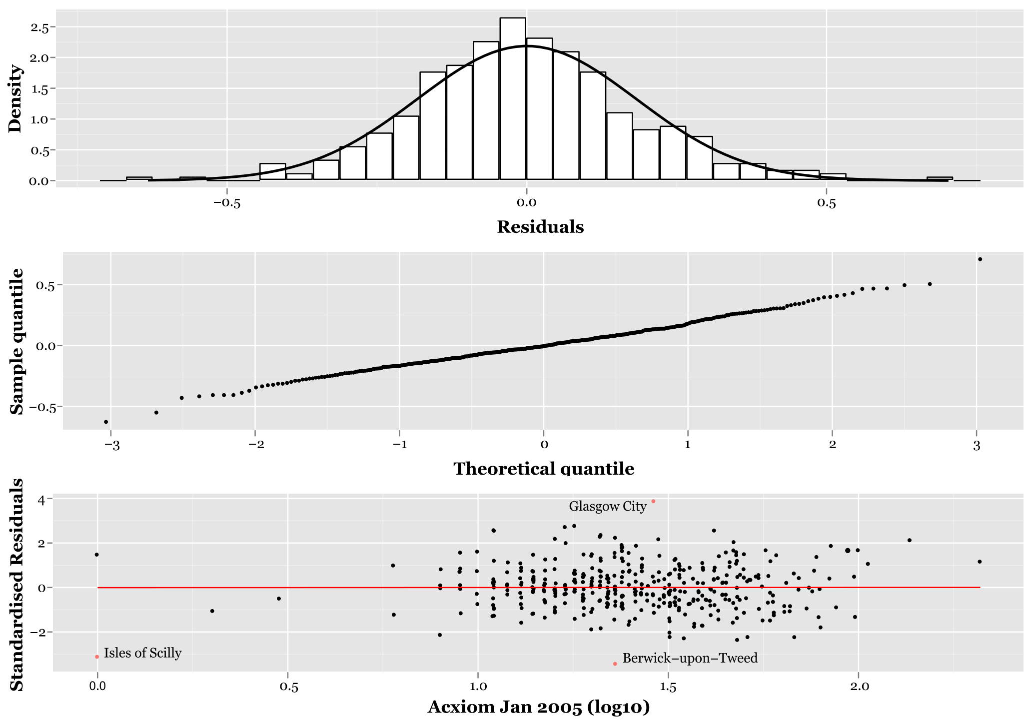


Figure 5.3. In-migrants to LADs from Acxiom January 2005 ROP (log10) and Census 2001 (log10): Tests for normality

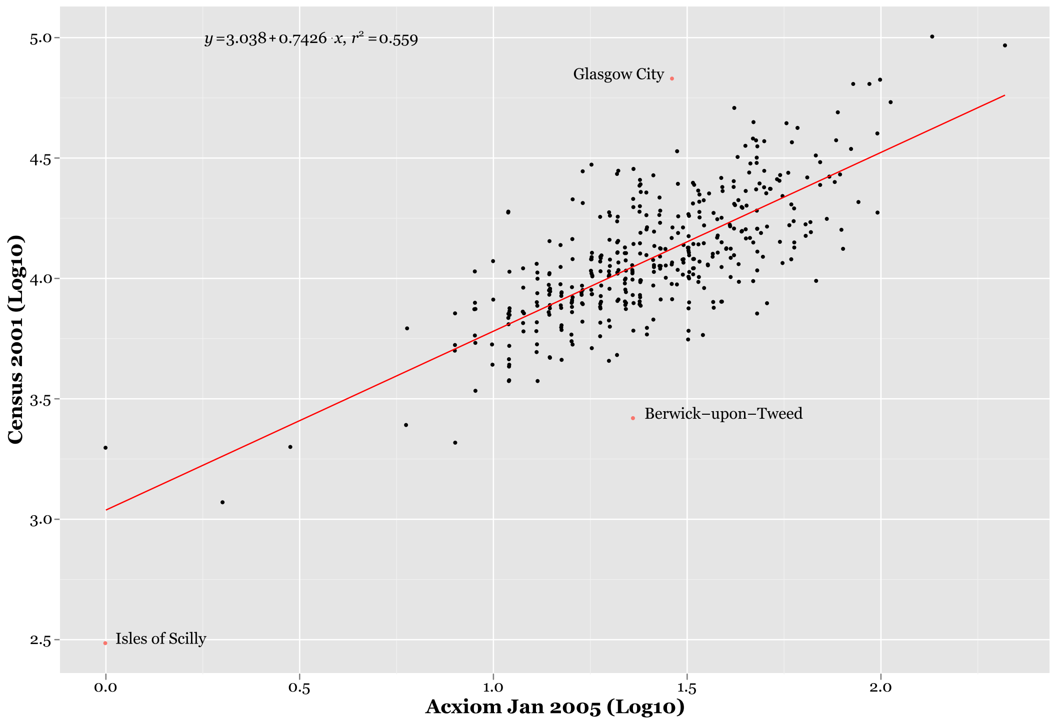


Figure 5.4. Scatterplot showing the relationship between Acxiom January 2005 ROP (log10) and Census 2001 (log10) in-migrants to LADs

Clearly the potential driving forces behind the apparent under/over counts will be complex and without the necessary sampling documentation it is impossible to accurately explore the degree of survey nonresponse in these areas. However, drawing on Figure 5.5, some interesting patterns may suggest potential factors behind the outlier observations.

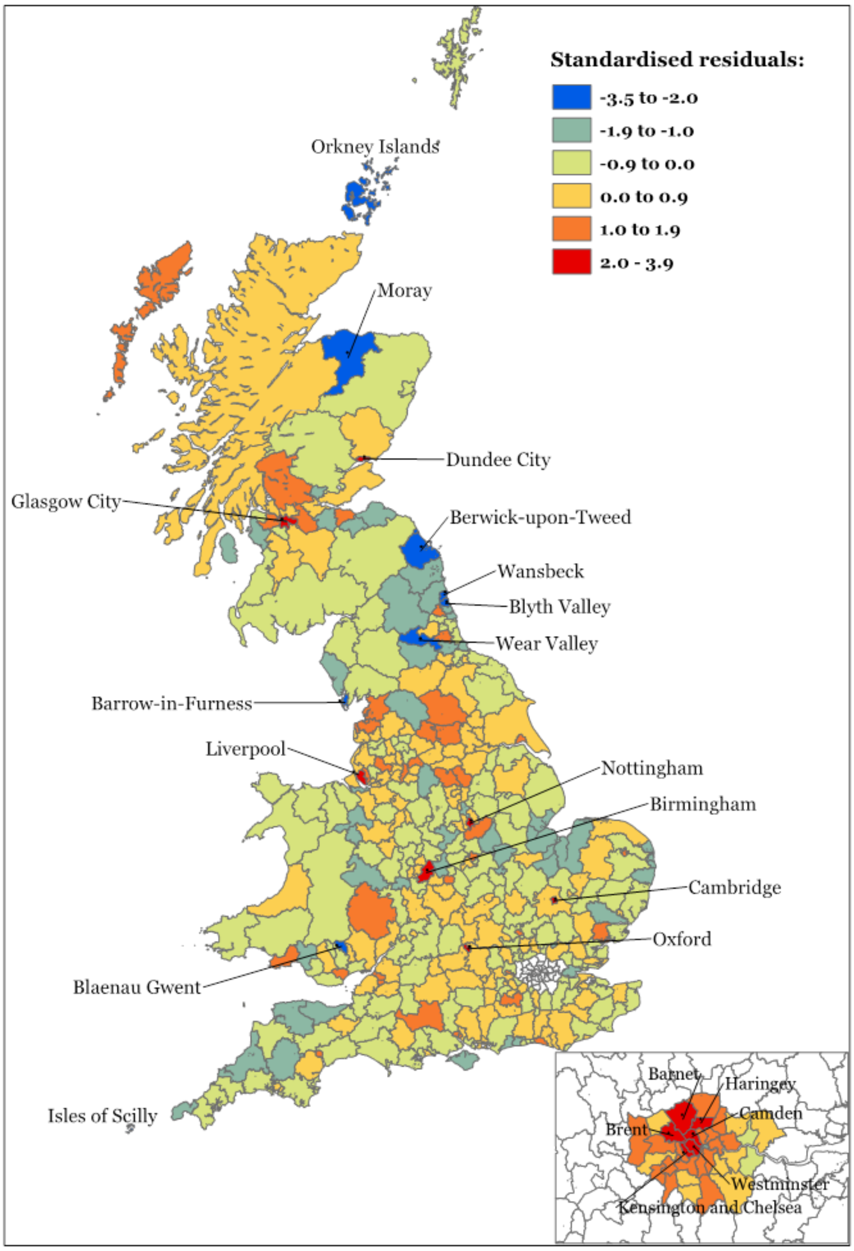


Figure 5.5. Acxiom January 2005 ROP (log10) and Census 2001 (log10): Standardised residuals map

Areas highlighted in blue represent areas with an overestimate in the ROP sample, while areas in red represent an underestimate. Looking at the broad patterns, it is apparent that a number of the major urban districts have significantly lower counts than are observed in the census; moreover, the student towns of Oxford and Cambridge also show significantly lower counts. Conversely, more rural districts, particularly those situated in the North East of England, are characterised by unusually high counts when compared to Census 2001. The island districts of Orkney and Isles of Scilly are also noted as having significantly higher rates than we would expect given the Census 2001 data. These observations most likely are the result of small number distortions linked to the very small samples achieved for these remote districts. The results presented in Figure 5.5 thus appear to point to clear geographical and most likely socio-demographic inconsistencies in the ROP migrant sample. As is made clear later in this chapter, the ROP suffers from an underrepresentation of a number of population sub-groups, most notably young adults and ethnic minority groups, with mobility undercounts in the major urban centres and university towns most likely explained by such factors. For an explanation of the unusual clustering of overcounts in the North East of England, Table 5.4 reveals the relative sample size and percentage of the usual resident population sampled in each GOR. Looking at the figures, it does appear that the ROP has oversampled the North East in comparison with the other regions of GB, something that could explain at least some of the overcounting observed in the in-migrant flows for the districts within this GOR.

Table 5.4. Comparison of the January 2005 ROP raw GB sample against Census 2001 usual resident population: GORs

GB: Government Office Region	Census 2001	Jan 2005 ROP	Sampled (%)
North East	2,515,442	25,585	1.02
North West	6,729,764	46,319	0.69
Yorkshire and The Humber	4,964,833	40,359	0.81
East Midlands	4,172,174	32,030	0.77
West Midlands	5,267,308	36,921	0.70
East of England	5,388,140	41,168	0.76
London	7,172,091	34,155	0.48
South East	8,000,645	51,591	0.64
South West	4,928,434	35,014	0.71
Scotland	5,062,011	39,739	0.79
Wales	2,903,085	26,860	0.93
Total	57,103,927	409,741	0.72

5.3.2 Validation against PR-NHSCR

As mentioned earlier, the comparison between the ROP data and PR-NHSCR data is based simply on in-migrant counts for LADs and therefore excludes those who moved within LADs. Figure 5.6 suggests that 38 per cent of the variation in the PR-NHSCR inflows can be explained by the Acxiom ROP equivalents. The explanatory power of Acxiom *vis-à-vis* PR-NHSCR appears to have suffered somewhat from the reduced numbers, associated with removal of intra-LAD flows. It should also be noted that the PR-NHSCR figures are themselves estimates and therefore susceptible to error too. However, while the r^2 value, together with a Pearson correlation coefficient of 0.62, may be significantly lower than the equivalent statistics for the census comparison, there is still evidence of a reasonable fit between the two data sources. An inspection of the residual plots in Figure 5.7, suggests a number of the OLS assumptions are violated, with positive skew, kurtosis, non-linearity and heteroscedasticity all being apparent.

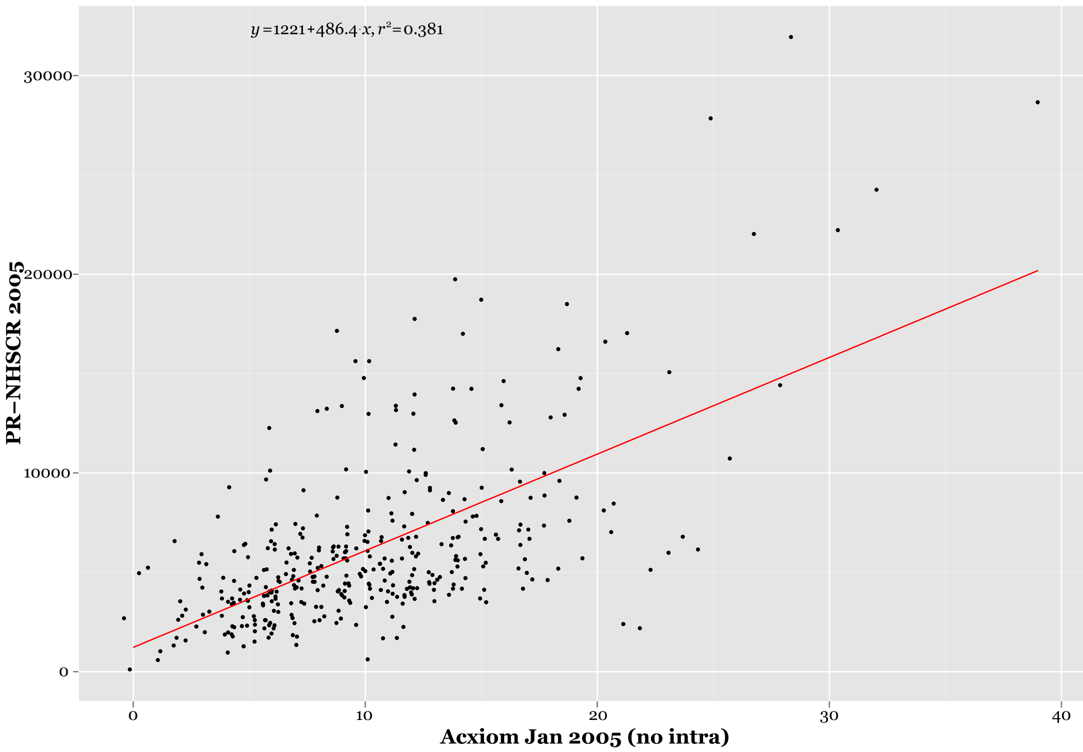


Figure 5.6. Scatterplot showing the relationship between Acxiom January 2005 ROP and PR-NHSCR 2005 in-migrants to LADs

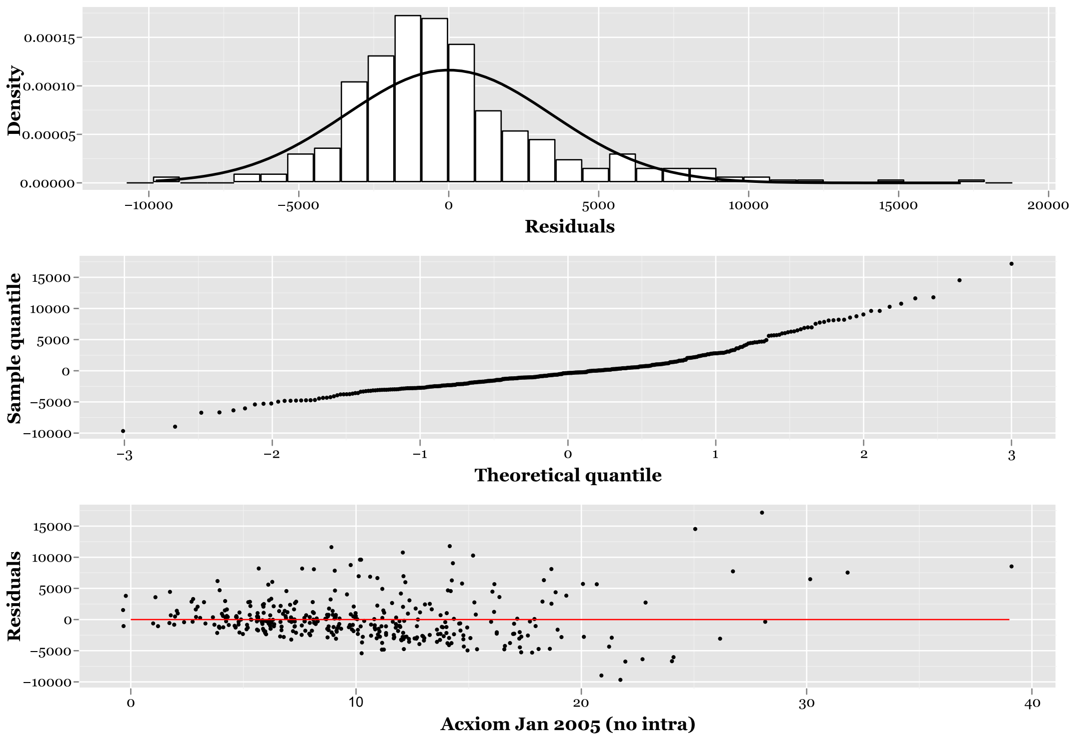


Figure 5.7. In-migrants to LADs from Acxiom January 2005 ROP and PR-NHSCR 2005: Tests for normality

Given that the common OLS regression assumptions are not met, variable transformations are applied. However, this time the lighter square root transformation was employed on the Acxiom data with the PR-NHSCR data being log10 transformed. While the transformations were successful in improving the model specification to meet the regression assumptions (Figure 5.8), the model fit is marginally worse with 35.8 per cent of the variation now explained (Figure 5.9). However, the transformation does allow us to identify the extreme outliers. All three outliers suggest an overcount on the part of the ROP sample and again the two districts of Berwick-upon-Tweed and the Isles of Scilly appear, this time joined by another North East based district in Blyth valley.

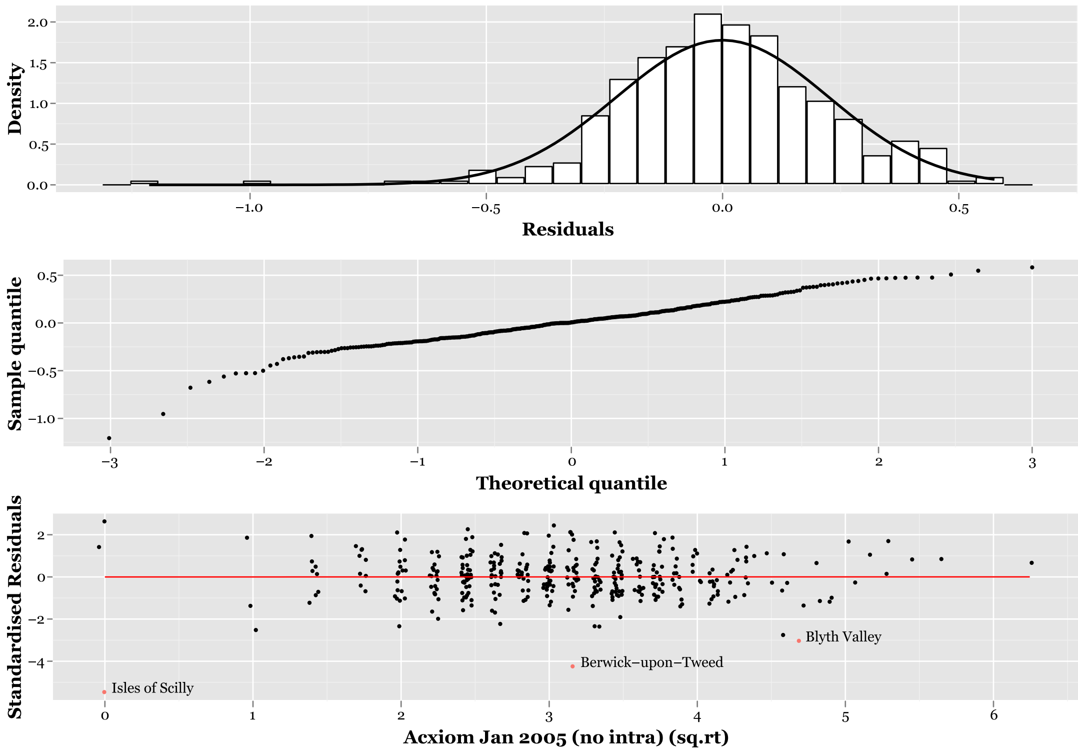


Figure 5.8. In-migrants to LADs from Acxiom January 2005 ROP (sq.rt.) and PR-NHSCR 2005 (log10): Tests for normality

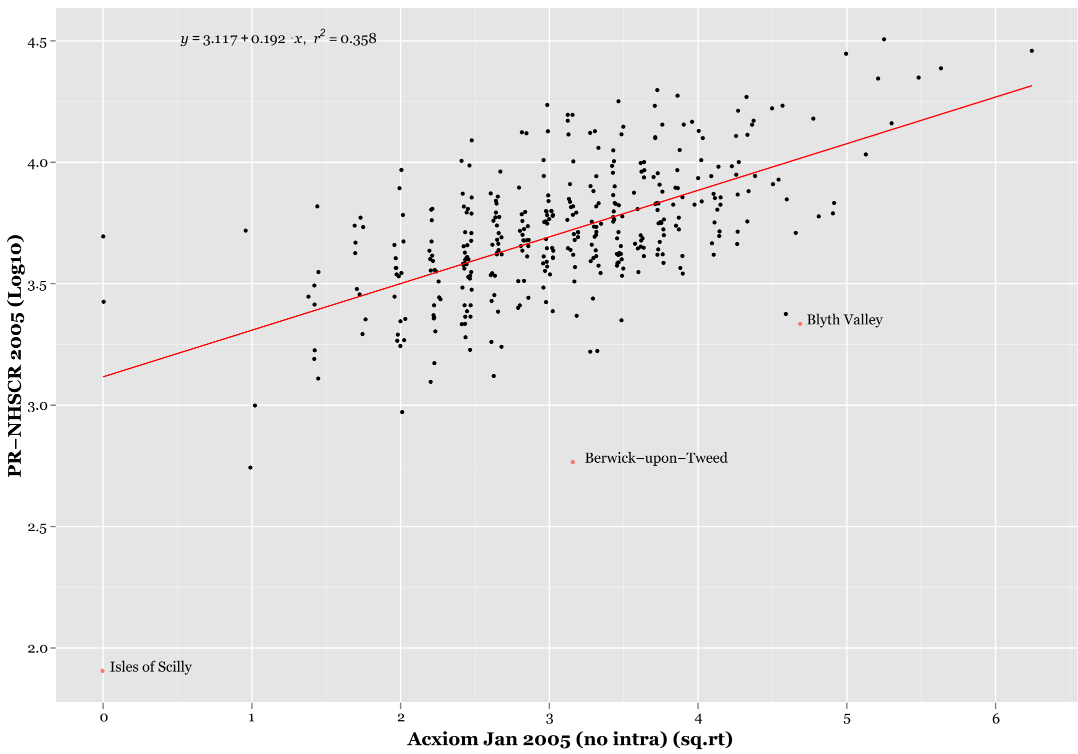


Figure 5.9. Scatterplot showing the relationship between Acxiom January 2005 ROP (sq.rt.) and PR-NHSCR 2005 (log10) in-migrants to LADs

5.3.3 Validation against Acxiom Aggregate Data

As was mentioned in Chapter 3, Acxiom Ltd. produces annual aggregate data that are themselves derived from the ROP microdata. The following checks against the Aggregate Data should give a handle on how much data manipulation has been undertaken by Acxiom in the production of this apparently representative aggregated product; though the precise details of the actual processes of aggregation used by Acxiom Ltd. are confidential and remain unavailable to the public. In terms of the fit, Figure 5.10 suggests that 55.1 per cent of the variation in the Aggregate Data can be explained by the ROP microdata. Thus, there remains 45 per cent of variation in the Aggregate Data left to be explained. Indeed, there are likely two key sources of this unexplained variation. First, as has been noted by Thompson *et al.* (2010), the Aggregate Data are derived using a combination of the prior September ROP and the following January ROP; thus in the case of the 2005 Aggregate Data, the September 2004 and January 2005 ROPs are combined. Moreover, following the combining of the two ROPs, the microdata undergo additional manipulation and reweighting using official data sources in an attempt to remove bias and ensure consistency with Census output (Thompson *et al.*, 2010).

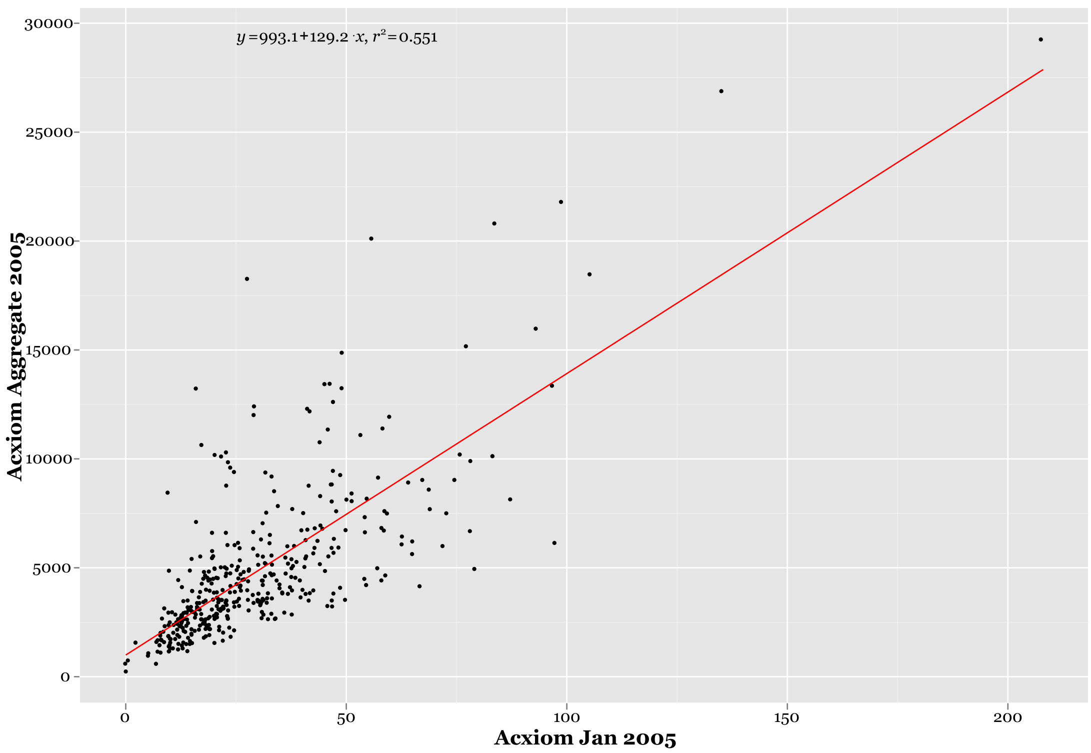


Figure 5.10. Scatterplot showing the relationship between Acxiom January 2005 ROP and Acxiom Aggregate Data 2005 in-migrants to LADs

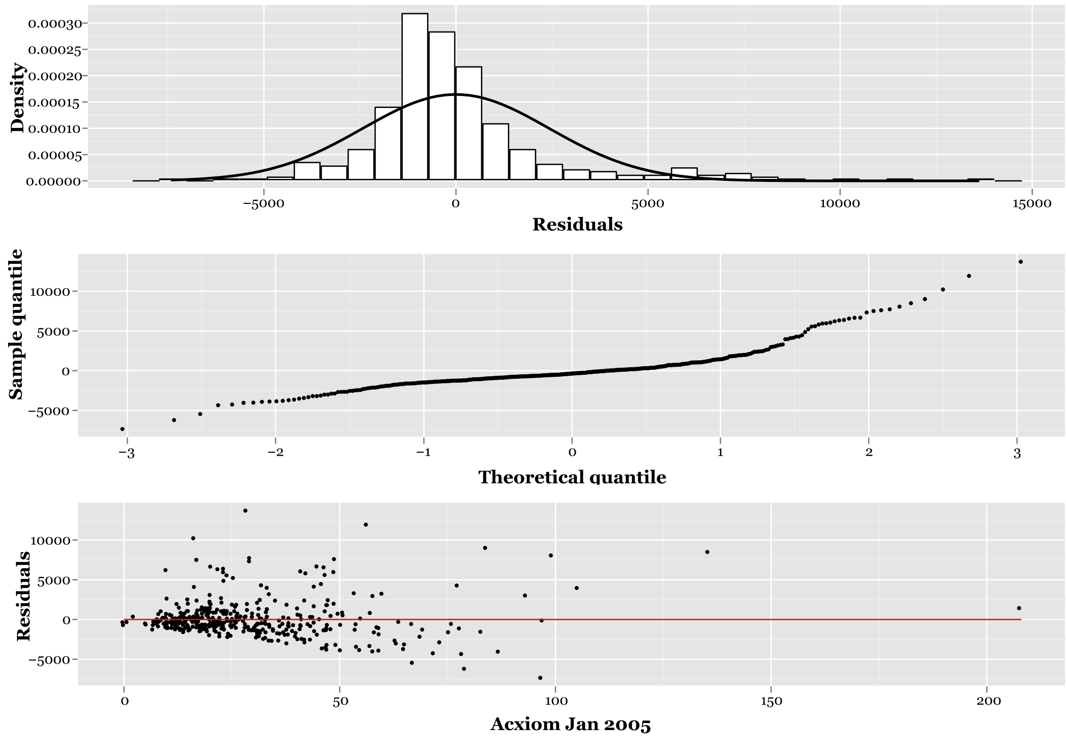


Figure 5.11. In-migrants to LADs from Acxiom January 2005 ROP and Acxiom Aggregate Data 2005: Tests for normality

As Figure 5.11 shows, there are again signs of positive kurtosis, skew, non-linearity and heteroscedasticity, with a number of clear outliers and leverage points also visible. Therefore, as with the other comparisons, the data are again transformed, this time using log10 transformations on both data sources (Figure 5.12). Having transformed the data, Figure 5.13 suggests that the fit has improved with 60.5% of the variation in the aggregate data now explained. Of course, whilst the Aggregate Data are made up of the combination of September 2004 and January 2005 ROP microdata, with an additional manipulation and reweighting process using official data sources, the outliers are potentially important for identifying where the January 2005 ROP is particularly inconsistent with Acxiom Ltd.'s derived aggregate estimates. Unsurprisingly, the outliers shown in Figures 5.12 and 5.13 again suggest an undercount in the raw January 2005 ROP in four urban districts, three central London districts as well as Glasgow City. Amongst other things, the recurrence of Glasgow City along with the undercounts in the central London outliers can again be expected to be the result of inherent selection biases, for instance linked to the underrepresentation of certain subpopulation groups including young adults and ethnic minorities, as shown in the following subsection (5.4) and in previous validation checks by Thompson *et al*. (2010).

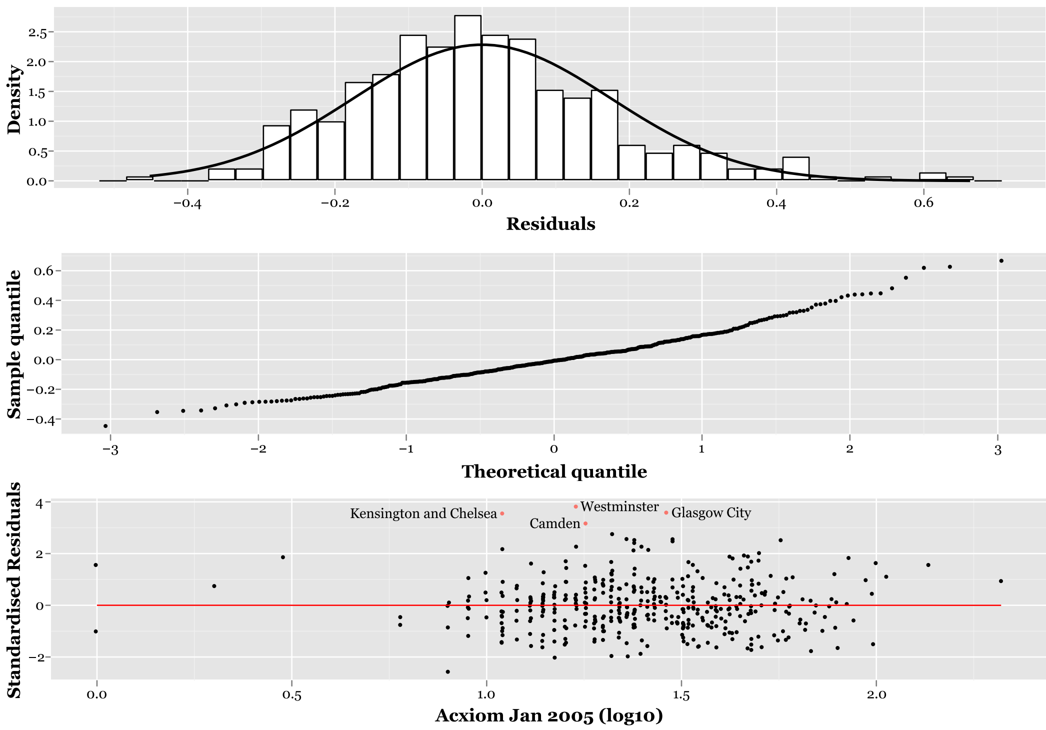


Figure 5.12. In-migrants to LADs from Acxiom January 2005 ROP (log10) and Acxiom Aggregate Data 2005 (log10): Tests for normality

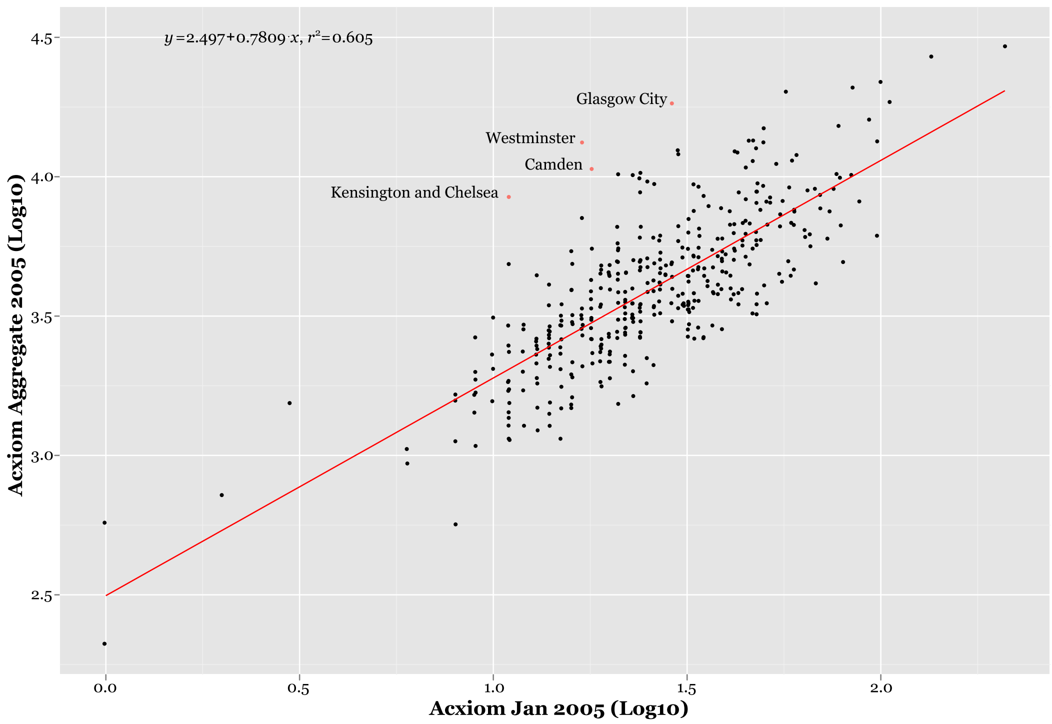


Figure 5.13. Scatterplot showing the relationship between Acxiom January 2005 ROP (log10) and Acxiom Aggregate Data 2005 (log10) in-migrants to LADs

5.4 Micro level benchmarking: Selected variables

As was mentioned in the introduction, this chapter is focussed on the empirical exploration and benchmarking of the raw ROP samples and, within this, it is also concerned with assessing the value of the microdata for descriptive-based analyses of population mobility. Consequently, this subsection focusses on three specially selected micro variables: age, ethnic group and housing tenure. Whilst alternative variables could have been selected, the main rationale behind this selection is based on the fact that they have all been observed to have particularly strong differential associations with mobility propensities (see Chapter 2). Furthermore, as has been mentioned numerous times above, previous validation exercises on the data have shown the ROP to struggle in capturing certain population subgroups including, for instance, younger people and ethnic minority groups (Thompson *et al.*, 2010). The alternative sources used in this subsection are the Census 2001 and the October to September 2005 APS (see Chapter 3). As is the case with the aggregate benchmarking above, the 2001 Census SMS is chosen as it represents the most accurate source of demographic and socio-economic data from which comparisons can be drawn, whereas the APS provides a timelier alternative with the sufficient variable detail required for micro level benchmarking.

5.4.1 Age

Observing the overall share of population by age is perhaps a good place to start in terms of benchmarking the age variable. Figure 5.14 breaks down the population by age for the Census 2001 usual resident population, the APS 2005 weighted population estimates and the January 2005 ROP sample. What becomes immediately apparent is an underrepresentation of young people (18-35 particularly) and an overrepresentation of the older age groups (60-85 particularly). When observing the share of total migrants by age in Figure 5.15, the bias in the ROP total sample is reflected in the migrant subsample. Again a clear underrepresentation of young people emerges, with the share of total migrants peaking at 25-29 instead of the 20-24 age group observed in Census 2001 and APS 2005. Moreover, as with the population share, an overrepresentation of older people (40-80) is apparent in the migrant subsample. Clearly, whilst this thesis is focussed on model-based analysis, for which extra validation is employed in the following chapter (Chapter 6), for any descriptive empirical analysis, such biases will undoubtedly have an influence on the results. With this in mind, for any researcher interested in utilising the ROP for descriptive analyses, some sort of reweighting strategy would be essential. Indeed the use of spatial microsimulaton techniques (Harland *et al.*, 2012) may be one possibility for those interested in descriptive-based empirical analysis.

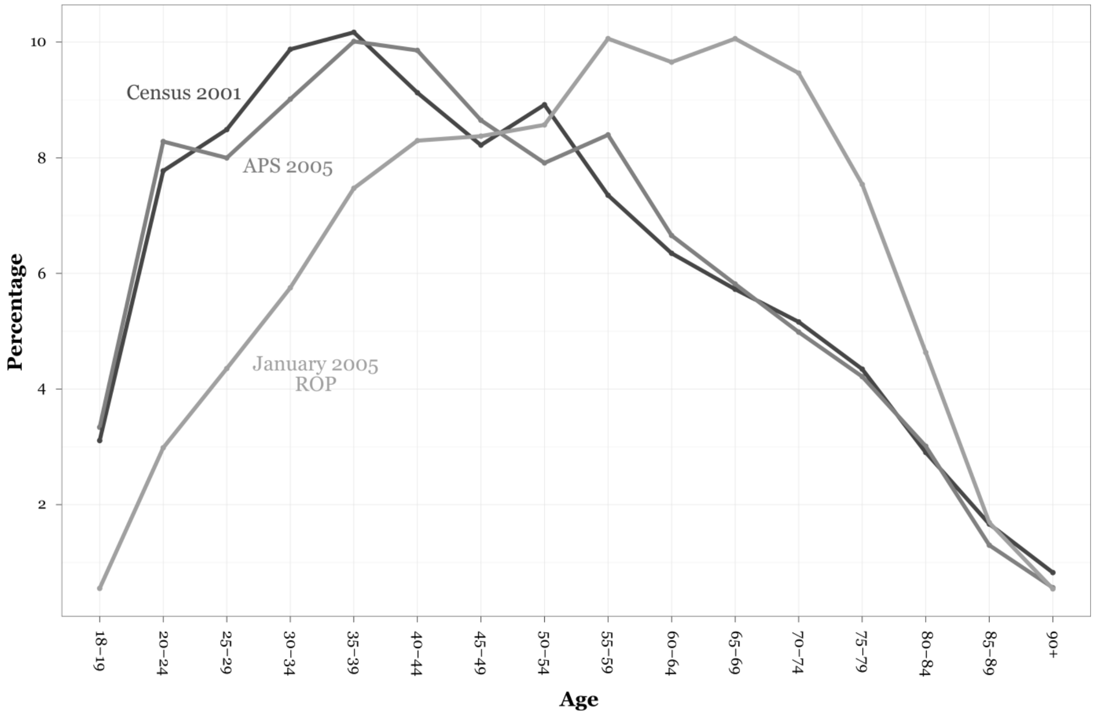


Figure 5.14. Share of population by age for Census 2001, APS 2005 and January 2005 ROP

However, even for model-based analyses that can to some extent control for sample distortions, given the centrality of age as a rather consistent proxy for important certain life course transitions and events (Chapter 2), a reasonable correlation with the known age trends to residential mobility propensities is essential if the ROP is to be taken seriously as a source of population mobility microdata.

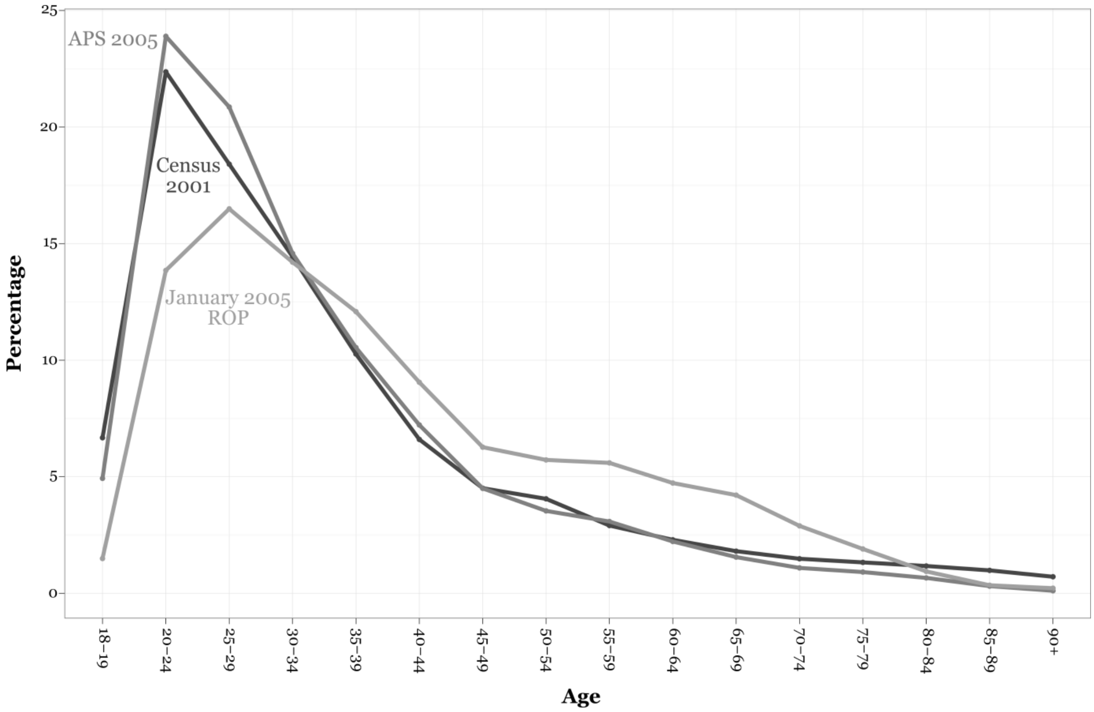


Figure 5.15. Share of total migrants by age for Census 2001, APS 2005 and January 2005 ROP

Figure 5.16 shows that, despite the known underrepresentation of migrants across the age categories, the general life course patterns are reassuringly close to those shown in the alternative sources of microdata. Indeed, the ROP closely matches the other source in picking up the higher propensities for the younger age groups with migration rates peaking during the years 18-25. As was mentioned in Chapter 2, this is an age where it is common for young adults to either move to university, employment or, subsequently, employment following university. The years from the mid-20s to the mid-40s are characterised by a relatively sharp reduction in migration rates and are generally considered the years of family formation and child rearing. The decline then reduces somewhat for the years 45-64, with research associating this easing in the decline with the transition from parenthood to 'empty nester', prompting the desire, at least for some, to make a residential move (Wulff *et al.*, 2010). The decline finally levels out to a slight increase at 75+ in Census 2001, an age commonly associated with a need for closer proximity to family members and services, given the greater requirement of help for the very elderly age groups. Interestingly, while APS 2005 does not pick up on this trend of increased mobility for the older age groups, the ROP does record some increase at least for the oldest age group (90+).

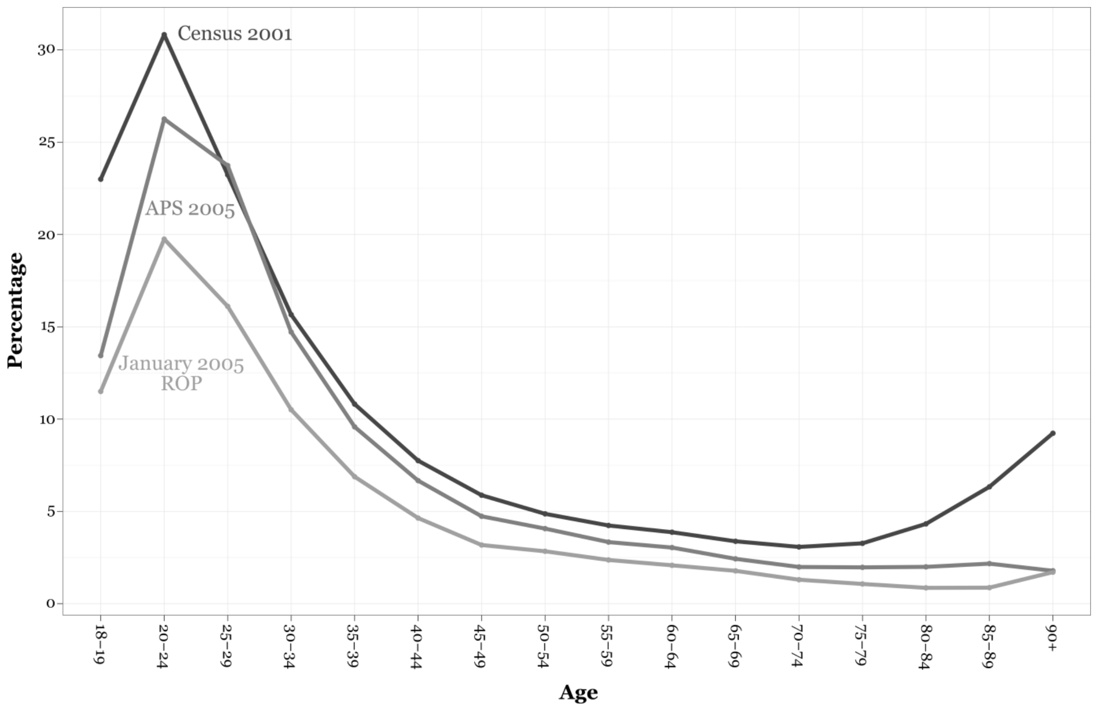


Figure 5.16. Migration rate by age for Census 2001, APS 2005 and January 2005 ROP

5.4.2 Ethnic group

The influence of ethnicity on migration propensities is a further topic of increasing interest (Large and Ghosh, 2006; Simpson and Finney, 2009; Stillwell and Hussain, 2010). To provide some context, research by Stillwell and Hussain (2010) has shown total migration rates are higher for all non-White ethnic minorities, apart from the Indian population, than they are for the White majority. However, some of this can be explained by the demographic structures of the populations. Indeed, all ethnic minority groups have younger populations than the White majority and thus, given the above discussion on age, one would expect higher propensities for these populations. The variation between the ethnic minority groups has been shown to be considerable. For instance, the Chinese population are known to have significantly higher migration rates than the Indian and POSA (Pakistani; Bangladeshi; Other Asian) populations (Stillwell and Hussain, 2010).

These general patterns can be observed in Figure 5.17 where, to make for easier comparison, the ethnic groups included in each dataset are aggregated into broad ethnic group categories. In the case of the January 2005 ROP, the original ethnic groups include: White; African; Pakistani; Chinese; Other Asian; Caribbean; Indian; Bangladeshi; and Other. However, the ethnicity question in the ROP allows the respondent to tick as many boxes as apply to the respondent, thus opening up the potential for more detailed categorisations. Yet this flexibility can also be problematic, for instance it is hard to discern whether or not an individual with reported membership to three or more ethnic groups is genuine, or simply a wrongly specified record. Due to this, and a need to form categories that match as closely as possible to those in Census 2001 and APS 2005, those who reported three or more ethnicities were grouped, along with the 'Chinese' population, as 'Other'. The 'Black' group includes those who described themselves as Caribbean or African while the 'Asian' group includes those recorded as 'Pakistani, 'Other Asian', 'Indian' and 'Bangladeshi'.

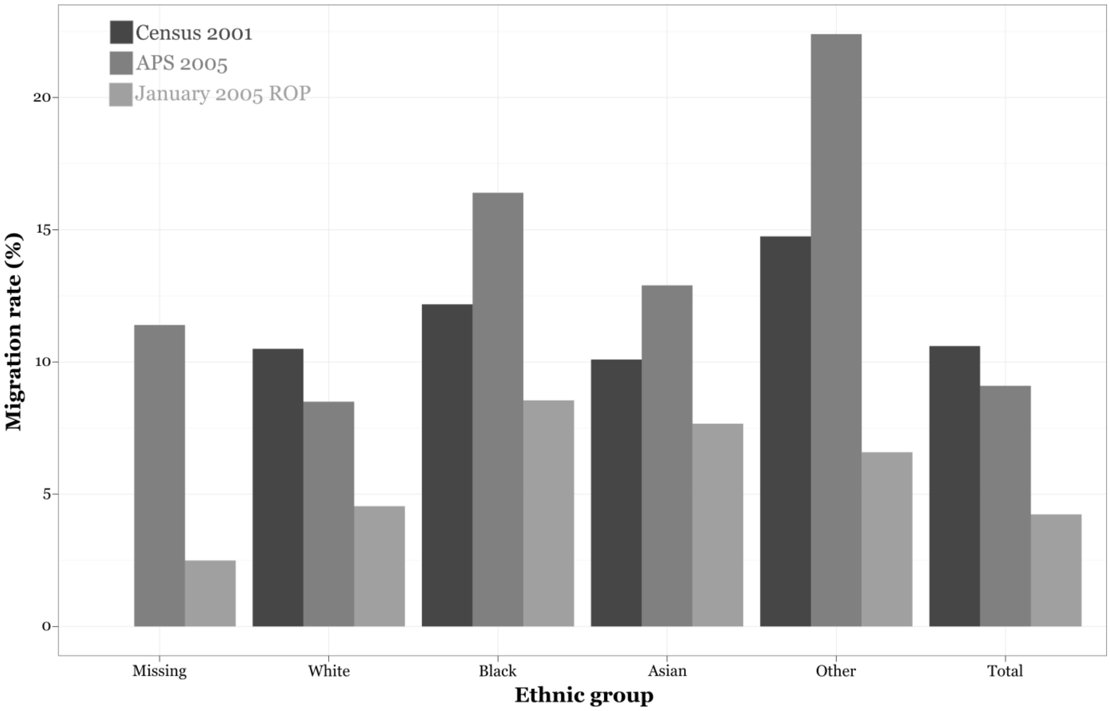


Figure 5.17. Migration rate by age for Census 2001, APS 2005 and January 2005 ROP

Table 5.5 shows some general statistics for the raw January 2005 ROP. When comparing the statistics in Table 5.5 with those for the 2001 Census SMS (Table 5.6), it is clear that, along with a large proportion of missing values, the ROP sample suffers from an underrepresentation of ethnic minorities in its sample. Moreover, whilst one would expect a certain degree of bias in all sample survey data (Crockett *et al.*, 2011), when compared to the population data, it is clear that the raw January 2005 ROP sample contains particularly low migration rates for all ethnic groups, ethnic minorities as well as the White majority. Aside from the 'Other' ethnic group, the patterns across the groups are reasonably similar to those in the APS 2005, with the White majority having the lowest mobility rates, the Black group having the highest and Asian groups falling somewhere between. Moreover, the share of movers for the raw January 2005 ROP (Table 5.5) appears to be more in agreement with the population data than the wider share of population statistics, with, as in the Census 2001 SMS and the APS 2005, the White ethnic groups clearly representing the largest share of movers followed by the Asian, Other, and Black ethnic groups.

The raw January 2005 ROP does reflect an underrepresentation of ethnic minority groups, which is further distorted in the statistics presented in Table 5.5 by the inclusion of the missing record category. Moreover, previous work by Thompson *et al*. (2010) also noted some comparative weakness in the ROP ethnic group sub-sample when checked against the EHS and LFS for the Yorkshire and Humber region. However, beyond the ethnic minority population share bias, the low mobility rates observed across the ethnic groups again reveal more general underrepresentation of migrants/movers in the sample. If the focus of this thesis was on descriptive-based empirical analyses, these issues would necessitate some sort of sample reweighting strategy. However, given that the focus here is on model-based approaches, exploring the directional associations of various variables of interest relevant to population mobility, the use of a regression framework incorporating suitable adjustment confounders may well provide a platform from which reasonably robust results can be drawn from the ROP. Chapter 6 explores this model-based approach in particular detail.

Table 5.5. Migration statistics for ethnic groups: Raw January 2005 ROP

	White	Black	Asian	Other	Missing	Total
Population	316,719	2,912	4,695	6,131	80,868	411,325
Share of population	77.0	0.7	1.1	1.5	19.7	100
Non-movers	302,313	2,663	4,335	5,727	78,852	393,890
Movers	14,406	249	360	404	2,016	17,435
Migration rate	4.6	8.6	7.7	6.6	2.5	4.2
Share of movers	82.6	1.4	2.1	2.3	11.6	100

Table 5.6. Migration statistics for ethnic groups: 2001 Census SMS

	White	Black	Asian	Other	Missing	Total
Population	52,481,200	1,147,589	2,328,757	1,146,387	0	57,103,933
Share of population	91.9	2.0	4.1	2.0	0	100
Non-movers	46,970,538	1,007,778	2,093,682	977,267	0	51,049,265
Movers	5,510,662	139,811	235,075	169,120	0	6,054,668
Migration rate	10.5	12.2	10.1	14.8	0	10.5
Share of movers	91.0	2.3	3.9	2.8	0	100

Table 5.7. Migration statistics for ethnic groups: APS October-September 2005

	White	Black	Asian	Other	Missing	Total
Population	42,403,284	900,038	1,839,915	1,017,523	24674	46,185,434
Share of population	91.8	1.9	4	2.2	0.05	100
Non-movers	38,815,790	752,688	1,602,059	789,912	21869	41,982,318
Movers	3,587,494	147350	237,856	227611	2805	4,203,116
Migration rate	8.5	16.4	12.9	22.4	11.4	9.1
Share of movers	85.4	3.5	5.7	5.4	0.1	100

N.B these are weighted estimates of the population aged 18 and over.

5.4.3 Tenure

Different housing tenure types have long been observed to reflect differing levels of mobility propensities (Hughes and McCormick, 1985; Boyle, 1995; Champion *et al.*, 1998; van Ham and Feijten, 2008; Mulder, 2013). Traditionally, those living in privately rented accommodation tend to have a greater propensity to move than those in publically rented or privately owned accommodation. Reasons for this revolve around the relative flexibility of the private renting, where lower transaction costs and short-term contract durations, and insecurity of tenure for some, lead to raised movement propensities. Conversely, home ownership tends to be a particularly inflexible tenure group, with high transaction costs and a level of long-term tenure security not possible in the other groups (Mulder, 2013). Moreover, in the British context, the restrictive nature of social housing provision, operating according to strict local access rules, has also been a topic of interest, with council renters observed to have reduced mobility propensities, particularly over longer distances and between authority districts (see Chapter 9; also Boyle, 1995; Hughes and McCormick, 2000). In addition to the tenure group itself, there are more general compositional patterns associated with those who tend to rent privately, those who rent publically and those who own their home. For instance, as one would expect, private renters tend to have a younger age profile than owner occupiers while owner occupiers tend to be more (asset) affluent than public renters; all these issues can be expected to inform, indirectly, the differing mobility rates between the housing tenure groups (Bailey and Livingston, 2005).

The longstanding tenure based variations in mobility propensities are apparent in the raw January 2005 ROP microdata. Table 5.8 is a contingency table comparing propensities to move between renters and home owners, the tenure categorisations of ‘own home’ and ‘rent home’ are aggregations of the original ROP categories that are examined in Table 5.10. Looking at the contingency table (Table 5.8), for the 245,915 individuals who own their own home, just 3.3% moved in the last 12 months while 96.7% did not. However, for the 117,978 renters, 7.6% moved with 92.4% remaining in place. In terms of those who did make the move, renters represented 51.4% of the sample with owners representing 47.1%. This is in contrast to those who did not move, who were predominantly home owners (60.4 %) with renters representing just 27.7% of the sample (Table 5.10). On the whole, the expected difference between renters and homeowners with regards the propensity to move is clear.

Table 5.8. Move/not move by own/rent contingency table: Raw January 2005 ROP

	Not Moved	Moved	Row Total
Own Home			
Count	237,698	8,217	245,915
Row Per cent	96.66%	3.34%	59.79%
Column Per cent	60.35%	47.13%	
Rent Home			
Count	109,016	8,962	117,978
Row Per cent	92.40%	7.60%	28.68%
Column Per cent	27.68%	51.40%	
Missing			
Count	47,176	256	47,432
Row Per cent	99.46%	0.54%	11.53%
Column Per cent	11.98%	1.47%	
Column Total			
Count	393,890	17,435	411,325
Per cent	95.76%	4.24%	100.00%

Comparably broad categorisations of own home and rent home are derived from the APS October-September 2005 (Table 5.9). Despite the fact that the weighted APS suffers far less from the problems of item (question) non-response (only 0.1% missing) broadly similar patterns are apparent in both data sources. Renters in the APS sample record a significantly higher propensity to move, at 20.6 per cent, than home owners (5.1%). Moreover, as with the raw January 2005 ROP, the APS 2005 sample suggests that for those who did move, a small majority were renters (58.43%), with home owners representing 46.2 per cent.

Table 5.9. Move/not move by own/rent contingency table: APS October - September 2005

	Not Moved	Moved	Row Total
Own Home			
Count	32,465,304	1,745,428	34,210,732
Row Per cent	94.90%	5.10%	74.07%
Column Per cent	77.33%	41.53%	
Rent Home			
Count	9,476,505	2,455,730	11,932,235
Row Per cent	79.42%	20.58%	25.84%
Column Per cent	22.57%	58.43%	
Missing			
Count	40,509	1,958	42,467
Row Per cent	95.39%	4.61%	0.09%
Column Per cent	0.10%	0.05%	
Column Total			
Count	41,982,318	4,203,116	46,185,434
Per cent	90.90%	9.10%	100.00%

N.B these are weighted estimates of the population aged 18 and over.

One comparative advantage of the ROP is its detailed breakdown of housing tenure, and specifically of the renter bracket. Indeed, the data allows for comparisons of renter mobility propensities based on whether they are private, housing association, or council. In an attempt to explore the directional patterns of the finer grained categories in a little more detail, Chi-squared analysis is performed on the contingency table shown in Table 5.10. The Pearson’s Chi-squared test result is highly significant at the 99% level, reaffirming the expectations that there is a significant association between tenure and residential mobility. It is apparent that the group with the highest propensity to move is private renters (11.6%), where the standardised residuals suggest that significantly more private renters moved than we would expect (z = 75.89, p = 0.01) with significantly fewer remaining in place (z = -15.97, p = 0.01). Likewise, both council and housing association tenants had higher than expected propensities to move, although to a lesser extent than those who rent privately. On the other hand, home owners represented significantly lower numbers of movers (z = -21.61, p = 0.01) and significantly higher numbers of non-movers (z = 4.55, p = 0.01) than would be expected. Reassuringly, these results appear to support the assertions made above, namely that those living in privately rented accommodation tend to have a greater propensity to move than those in publically rented or privately owned accommodation.

Table 5.10. Move/not move by detailed tenure contingency table: Raw January 2005 ROP

	Not Moved	Moved	Row Total
Own Home			
Count	237,698	8,217	245,915
Row Per cent	96.66%	3.34%	59.79%
Column Per cent	60.35%	47.13%	
Std Residual	4.55	-21.61	
Rent (Council)			
Count	48,520	2,451	50,971
Row Per cent	95.19%	4.81%	12.39%
Column Per cent	12.32%	14.06%	
Std Residual	-1.32	6.25	
Rent (Housing Association)			
Count	21,039	1,317	22,356
Row Per cent	94.11%	5.89%	5.44%
Column Per cent	5.34%	7.55%	
Std Residual	-2.53	12.00	
Rent (Private)			
Count	39,457	5,194	44,651
Row Per cent	88.37%	11.63%	10.86%
Column Per cent	10.02%	29.79%	
Std Residual	-15.97	75.89	
Missing			
Count	47,176	256	47,432
Row Per cent	99.46%	0.54%	11.53%
Column Per cent	11.98%	1.47%	
Std Residual	8.23	-39.13	
Column Total			
Count	393,890	17,435	411,325
Per cent	95.76%	4.24%	100.00%

Pearson's Chi-squared test: $X^2(4) = 8291.38$, $p < .001$

5.5 Spatial benchmarking

As was detailed in Chapter 3, when it comes to the spatial detail allowed for, the ROP has a major advantage over many of the traditional sample survey sources. Indeed through the availability of postcode identifiers, it is possible to aggregate to any predefined geography. Moreover, given the relatively large migrant subsample, it is also possible to explore and benchmark flows within and between different geographical units, although in reality the district level is most appropriate. Consequently, in order to benchmark the spatial elements of the raw ROP, the following section reports the relationship between district level deprivation and net migration rates (per 1,000) across England using Census 2000-01 SMS inter-district moves and resident populations, PR-NHSCR 2004-05 inter-district moves and ONS mid-year population estimates for 2004, and the raw January 2005 ROP sample for 12 month movers. The Acxiom ROP sample includes individual who changed residence within England in the 12 months prior to the survey date (January 2005) and for whom usable origin and destination identifiers at the LAD level are available (n = 8,224), with the population denominator being the ROP sample population at risk for each LAD. The deprivation measure deemed most suitable for use here is the Index of Multiple Deprivation 2004 (IMD 2004) with the district scores being the population weighted average of the combined IMD scores for the Super Output Areas (SOAs) contained within each LAD (ODPM, 2004). Employing a similar methodology to that used by Bailey and Livingston (2008), the districts were grouped into four broad regions in an attempt to emphasise the differences in labour and housing market context. These regions are the North (North-East, North-West, and Yorkshire and the Humber), the Midlands (West Midlands and East Midlands), London, and the remainder of the South (East, South-East, and South-West). The districts within each region were ranked into equal deciles (based on the number of LADs) according to their IMD score to avoid a concentration of deprivation in the North. As the decile averages in Table 5.11 show, London's most deprived LADs are on average the most deprived in the country whereas the most deprived LADs in the South have average deprivation scores that would be situated in the middle deciles of the North and London. London also has the largest disparity (between lowest and highest) in deprivation scores with the South having by far the smallest disparity. Beyond this, when we observe the number of LADs in each region, the

South includes roughly five times the number of LADs we see for London, while the North and Midlands are also more than twice the size in terms of the LADs they contain. It is possible that the small number of LADs included in London is having an effect on the more exaggerated patterns we observe in Figure 5.18; after all we could expect just a few LADs with more extreme values to have a significant effect on decile averages composed of relatively few LADs.

Table 5.11. Region average IMD 2004 score for each deprivation decile

	No. LADS	Lowest	2	3	4	5	6	7	8	9	Highest
North	87	11.31	14.78	17.61	21.16	24.70	27.46	29.32	31.66	33.65	43.09
Midlands	74	8.34	11.09	12.46	15.02	16.20	17.72	19.59	23.07	28.26	35.36
London	33	11.94	14.09	15.29	19.05	22.60	25.72	30.40	33.52	37.83	44.53
South (rest of)	160	6.20	8.03	9.19	10.45	12.10	14.14	16.34	19.06	21.65	25.87

In terms of analysing the success of the ROP in matching the patterns observed in the PR-NHSCR 2005 and Census 2001 data, we can be reasonably satisfied. For London (Figure 5.18), we observe a close match between all three data sources, with significant net losses for most areas – a somewhat familiar observation given the large net losses associated with London as a whole (see Chapter 2, Table 2.2 and 2.3). Beyond this, however, there is a clear pattern of greater net losses in the more deprived deciles with the losses reducing somewhat as we move towards the less deprived deciles. Looking at the Midlands (Figure 5.19), a similar pattern emerges again, with all three sources showing the greatest net losses in the most deprived deciles which steadily turn to net gains as we move towards the least deprived deciles. However, the overall picture for the North (Figure 5.20) is somewhat less impressive in terms of the comparability of results. Indeed, whilst the common pattern of net migration shift from the most deprived to least deprived deciles is observed in the Census and PR-NHSCR data, a less clear-cut relationship is seen for the ROP. For the South, on the other hand (Figure 5.21), the Census 2001 and PR-NHSCR data appear to contradict one another with net gains observed across the deciles in the PR-NHSCR and Census 2001 suggesting a reversal of the usual shift from most deprived to least deprived. While this could be a result of the small temporal differences in the data, although research has shown migration trends to be surprisingly stable over the 2000s (Duke-Williams and Stillwell, 2010), it is perhaps a reminder that no single source can be relied upon to provide completely error free estimates. The ROP in relation to the two contradictory sources appears to be successful in following the ground between both trends. For instance, small net gains from deciles 5-10 are observed with a peak at decile 5, also observed in the PR-NHSCR, but in addition the net losses in the two least deprived deciles are also clear, as observed also in the Census 2001 data. As such, the raw ROP appears to be successful in picking up the general patterns that Bailey and Livingston (2008) observed in their work, with net gains in the most deprived and net losses in the least deprived.

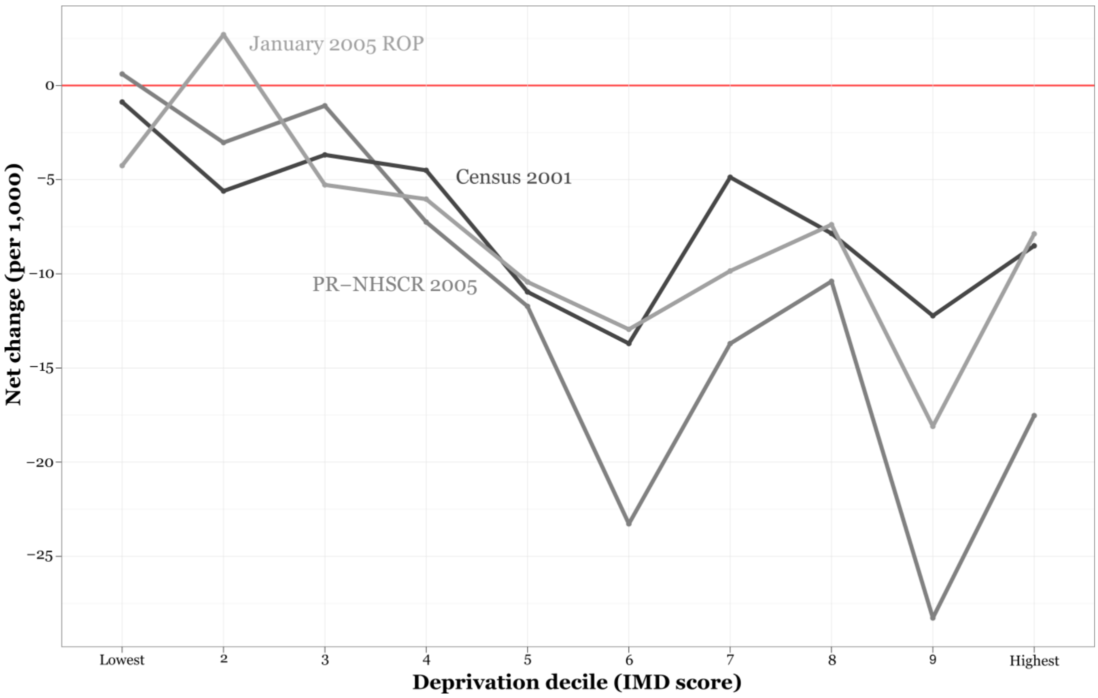


Figure 5.18. Net migration change by IMD decile at regional level: London

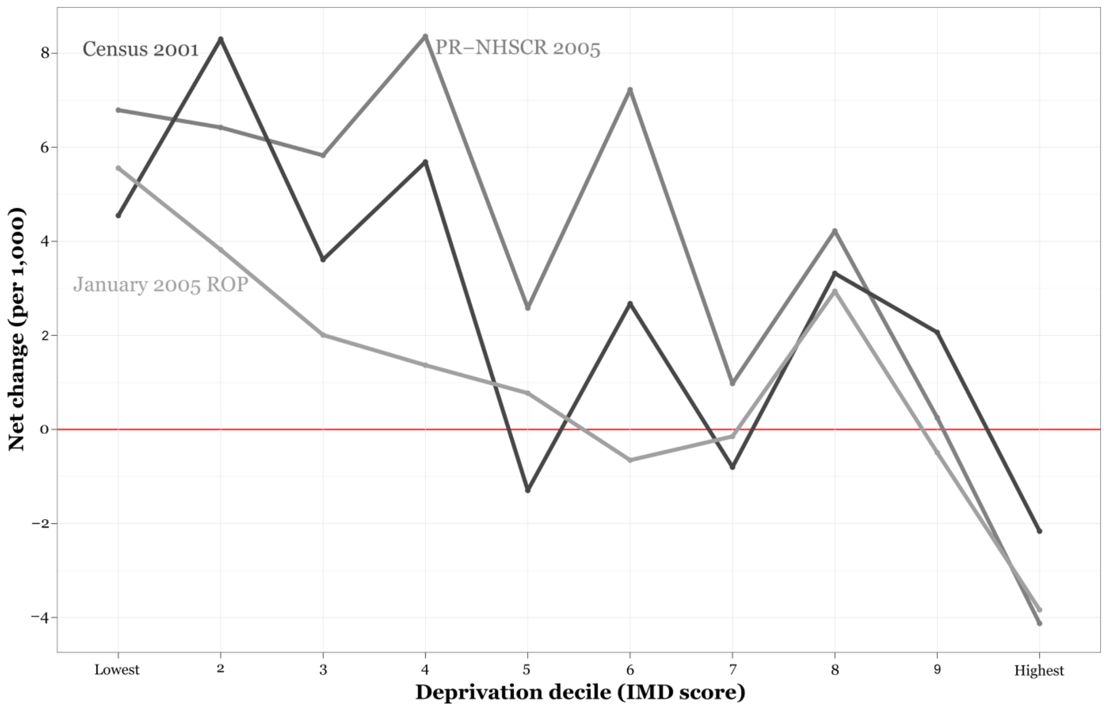


Figure 5.19. Net migration change by IMD decile at regional level: Midlands

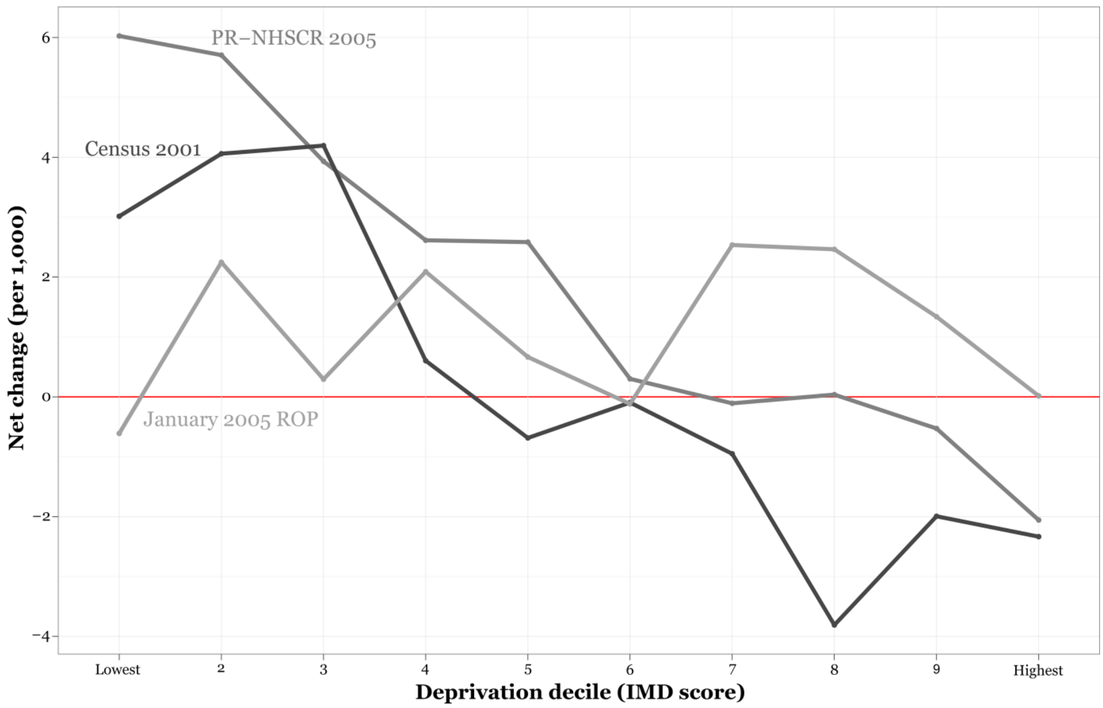


Figure 5.20. Net migration change by IMD decile at regional level: North

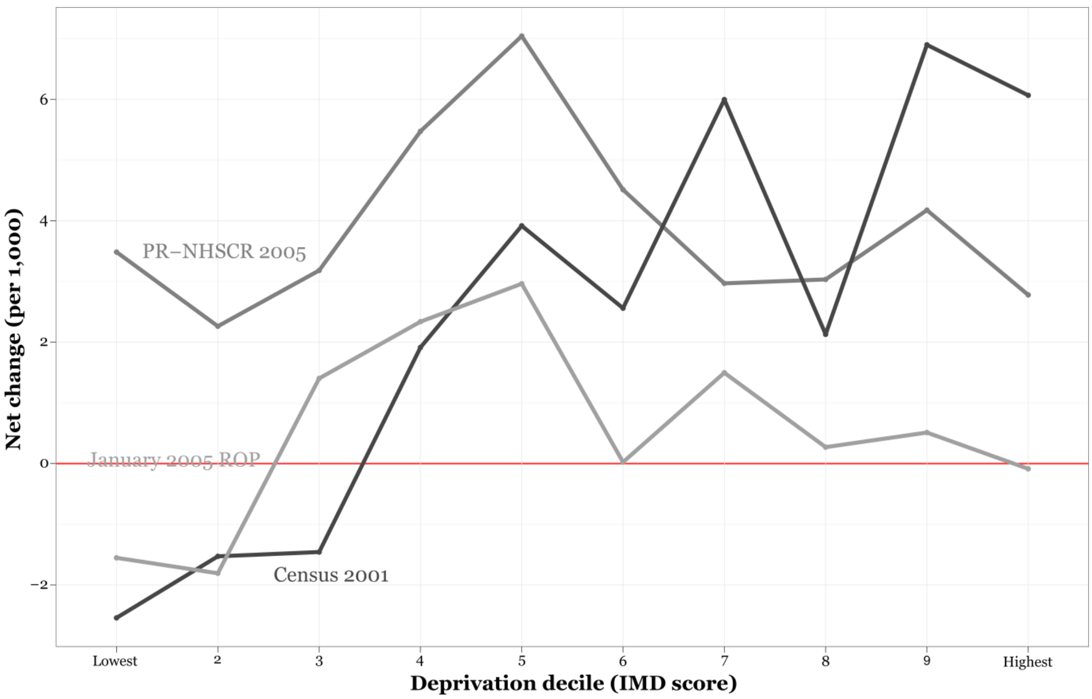


Figure 5.21. Net migration change by IMD decile at regional level: South (rest of)

Focussing on some of the other migration measures (Tables 5.12-5.14), across all three data sources, the gross components of the flows suggest that, on the whole, an increase in total migration counts develops as the deprivation deciles increase, although this is not the case for London. In terms of migration effectiveness (Tables 5.12-5.14), all data sources suggest that the most deprived deciles have negative ratios suggesting that the majority of migrants leave as opposed to move into these areas. For the Midlands and London, the values are large suggesting that migration is working to significantly redistribute the population in these regions, thus producing a large net effect relative to the volume of migrants. Conversely, the South has relatively low effectiveness values across the deciles suggesting that migration is inefficient as a mechanism for population redistribution in this region, a phenomenon that could again be potentially associated with the tight housing markets in this region.

Thus, from the figures and tables displayed in this subsection, the raw January 2005 ROP appears reasonably successful in picking up the general directional flows apparent in the PR-NHSCR 2005 and Census 2001 data. Moreover, in so far as the 2001 Census can be considered as the optimum point of reference for those interested in population statistics and small area analysis (Raymer *et al.*, 2012), it is encouraging to see the raw ROP sample reflect many of the patterns observed in the census (including those across different measures of migration). In fact, in some cases the raw ROP appears to be more successful than the PR-NHSCR in matching the patterns in Census 2001, for instance in the reversal of the general migration shift (net outflows in the least deprived and net inflows in the most deprived) for the South.

Finally, moving beyond the focus on deprivation, Figures 5.22 and 5.23 are maps of inter-district net migration rates (per 1,000) in Great Britain. While there are variations between the two maps, again the sample biases will be playing a role, Figure 5.22 appears to support a clear and persistent pattern observed in many previous analyses on internal migration in Great Britain, namely that of urban/rural shift/counter-urbanisation (Rees and Stillwell, 1992; Champion, 2005b; Dennett and Stillwell, 2008). The vast majority of urban districts in GB are characterised by net losses, especially those that represent the major metropolitan districts and the districts of London, while at the same time we can observe net gains for the more rural districts of GB, for instance those in East Anglia and Lincolnshire.

Table 5.12. Measures of migration (average) by region and deprivation decile: Census 2001

North	**Lowest**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**Highest**
In-migration count	8982.3	11117.7	12182.7	10932.6	11641.6	31905.1	19798.1	17275.1	18296.4	25858.5
Out-migration count	8849.3	10558.9	11891.1	10840.0	11700.1	31696.1	19932.6	17927.1	18621.8	26182.3
Gross-migration count	17831.7	21676.6	24073.8	21772.6	23341.7	63601.2	39730.8	35202.2	36918.2	52040.8
Net-migration count	133.0	558.8	291.6	92.6	-58.6	209.0	-134.5	-652.0	-325.3	-323.8
Migration effectiveness	0.7	2.6	1.2	0.4	-0.3	0.3	-0.3	-1.9	-0.9	-0.6
Midlands	**Lowest**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**Highest**
In-migration count	7320.5	8977.1	10821.4	8764.7	9148.7	8061.8	10923.3	12011.8	16996.4	36407.9
Out-migration count	6969.5	8263.9	10360.1	8247.0	9305.6	7863.5	10987.4	11708.0	17028.3	37361.4
Gross-migration count	14290.0	17241.0	21181.5	17011.7	18454.3	15925.3	21910.7	23719.8	34024.7	73769.3
Net-migration count	351.0	713.3	461.3	517.7	-156.9	198.3	-64.1	303.8	-31.9	-953.6
Migration effectiveness	2.5	4.1	2.2	3.0	-0.8	1.2	-0.3	1.3	-0.1	-1.3
London	**Lowest**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**Highest**
In-migration count	20421.8	13068.0	21502.7	29868.5	22325.0	27414.0	21634.5	30705.3	27127.3	22877.0
Out-migration count	20655.0	14348.0	22474.7	31116.8	24552.7	30772.0	22732.0	32188.0	29989.3	24471.7
Gross-migration count	41076.8	27416.0	43977.3	60985.3	46877.7	58186.0	44366.5	62893.3	57116.7	47348.7
Net-migration count	-233.3	-1280.0	-972.0	-1248.3	-2227.7	-3358.0	-1097.5	-1482.7	-2862.0	-1594.7
Migration effectiveness	-0.6	-4.7	-2.2	-2.0	-4.8	-5.8	-2.5	-2.4	-5.0	-3.4
South (rest of)	**Lowest**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**Highest**
In-migration count	10422.8	11727.3	11180.5	10040.8	10307.5	12817.7	10614.2	13790.1	11766.4	20725.4
Out-migration count	10643.9	11886.2	11317.8	9884.3	9917.4	12459.8	10127.1	13565.9	11118.0	19834.0
Gross-migration count	21066.7	23613.5	22498.3	19925.1	20224.9	25277.5	20741.2	27355.9	22884.4	40559.4
Net-migration count	-221.1	-158.9	-137.3	156.5	390.1	357.9	487.1	224.2	648.4	891.4
Migration effectiveness	-1.0	-0.7	-0.6	0.8	1.9	1.4	2.3	0.8	2.8	2.2

Table 5.13. Measures of migration (average) by region and deprivation decile: PR-NHSCR 2005

North	**Lowest**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**Highest**
In-migration count	3562.2	4540.0	4672.2	3793.8	3661.1	9407.8	5228.8	4716.7	5296.7	8516.3
Out-migration count	3078.9	3875.6	4288.9	3505.0	3368.9	9420.0	5176.3	4930.0	5494.4	9273.8
Gross-migration count	6641.1	8415.6	8961.1	7298.8	7030.0	18827.8	10405.0	9646.7	10791.1	17790.0
Net-migration count	483.3	664.4	383.3	288.8	292.2	-12.2	52.5	-213.3	-197.8	-757.5
Migration effectiveness	7.3	7.9	4.3	4.0	4.2	-0.1	0.5	-2.2	-1.8	-4.3
Midlands	**Lowest**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**Highest**
In-migration count	4041.3	4358.6	4785.0	3684.3	4448.6	3317.5	4341.4	4258.8	6070.0	13137.1
Out-migration count	3522.5	3757.1	4163.8	2927.1	4231.4	2742.5	4280	3838.8	6350	15111.4
Gross-migration count	7563.8	8115.7	8948.8	6611.4	8680	6060	8621.4	8097.5	12420	28248.6
Net-migration count	518.8	601.4	621.3	757.1	217.1	575	61.4	420	-280	-1974.3
Migration effectiveness	6.9	7.4	6.9	11.5	2.5	9.5	0.7	5.2	-2.3	-7.0
London	**Lowest**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**Highest**
In-migration count	11502.5	8463.3	11956.7	16967.5	12046.7	14330.0	12770	18766.7	15336.7	13503.3
Out-migration count	11262.5	9293.3	12463.3	19082.5	14590	20186.7	15720	21423.3	22206.7	17090.0
Gross-migration count	22765.0	17756.7	24420	36050	26636.7	34516.7	28490	40190	37543.3	30593.3
Net-migration count	240.0	-830	-506.7	-2115	-2543.3	-5856.7	-2950	-2656.7	-6870	-3586.7
Migration effectiveness	1.1	-4.7	-2.1	-5.9	-9.5	-17	-10.4	-6.6	-18.3	-11.7
South (rest of)	**Lowest**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**Highest**
In-migration count	6006.9	6184.4	5876.3	4975.6	4991.3	6135.6	4604.1	5881.3	4690.0	8027.5
Out-migration count	5643.1	5925.0	5553.8	4434.4	4360.6	5506.3	3995.3	5670.7	4399.4	7795.6
Gross-migration count	11650.0	12109.4	11430.0	9410.0	9351.9	11641.9	8599.4	11552.0	9089.4	15823.1
Net-migration count	363.8	259.4	322.5	541.3	630.6	629.4	608.8	210.7	290.6	231.9
Migration effectiveness	3.1	2.1	2.8	5.8	6.7	5.4	7.1	1.8	3.2	1.5

Table 5.14. Measures of migration (average) by region and deprivation decile: Raw January 2005 ROP

North	**Lowest**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**Highest**
In-migration count	16.0	25.3	21.9	23.4	35.6	64.4	46.9	36.3	50.9	38.5
Out-migration count	16.4	22.8	22.4	22.4	34.6	65.0	44.6	35.2	49.6	39.4
Gross-migration count	32.4	48.1	44.3	45.8	70.1	129.4	91.5	71.6	100.4	77.9
Net-migration count	-0.4	2.6	-0.6	1.0	1.0	-0.6	2.3	1.1	1.3	-0.9
Migration effectiveness	-1.4	5.3	-1.3	2.2	1.4	-0.4	2.5	1.6	1.3	-1.1
Midlands	**Lowest**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**Highest**
In-migration count	14.1	20.3	27.4	22.0	19.3	16.9	19.4	30.1	37.7	55.9
Out-migration count	12.0	17.9	25.6	21.9	19.0	16.6	19.3	28.0	37.3	67.7
Gross-migration count	26.1	38.1	53.0	43.9	38.3	33.5	38.7	58.1	75.0	123.6
Net-migration count	2.1	2.4	1.8	0.1	0.3	0.3	0.1	2.1	0.4	-11.9
Migration effectiveness	8.1	6.4	3.3	0.3	0.7	0.7	0.4	3.7	0.6	-9.6
London	**Lowest**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**Highest**
In-migration count	31.3	19.7	32.7	42.3	23.3	28.3	39.0	27.3	30.0	25.3
Out-migration count	35.5	26.3	40.7	50.8	31.0	37.3	51.5	34.0	47.3	31.0
Gross-migration count	66.8	46.0	73.3	93.0	54.3	65.7	90.5	61.3	77.3	56.3
Net-migration count	-4.3	-6.7	-8.0	-8.5	-7.7	-9.0	-12.5	-6.7	-17.3	-5.7
Migration effectiveness	-6.4	-14.5	-10.9	-9.1	-14.1	-13.7	-13.8	-10.9	-22.4	-10.1
South (rest of)	**Lowest**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**Highest**
In-migration count	18.1	22.1	26.2	23.3	22.1	26.3	24.4	34.3	33.3	43.1
Out-migration count	19.1	23.6	25.6	21.6	19.8	25.6	22.9	34.0	31.4	43.4
Gross-migration count	37.2	45.6	51.8	44.9	41.9	51.9	47.4	68.3	64.7	86.5
Net-migration count	-0.9	-1.5	0.6	1.7	2.3	0.6	1.5	0.3	1.8	-0.3
Migration effectiveness	-2.5	-3.3	1.2	3.8	5.5	1.2	3.1	0.4	2.8	-0.3

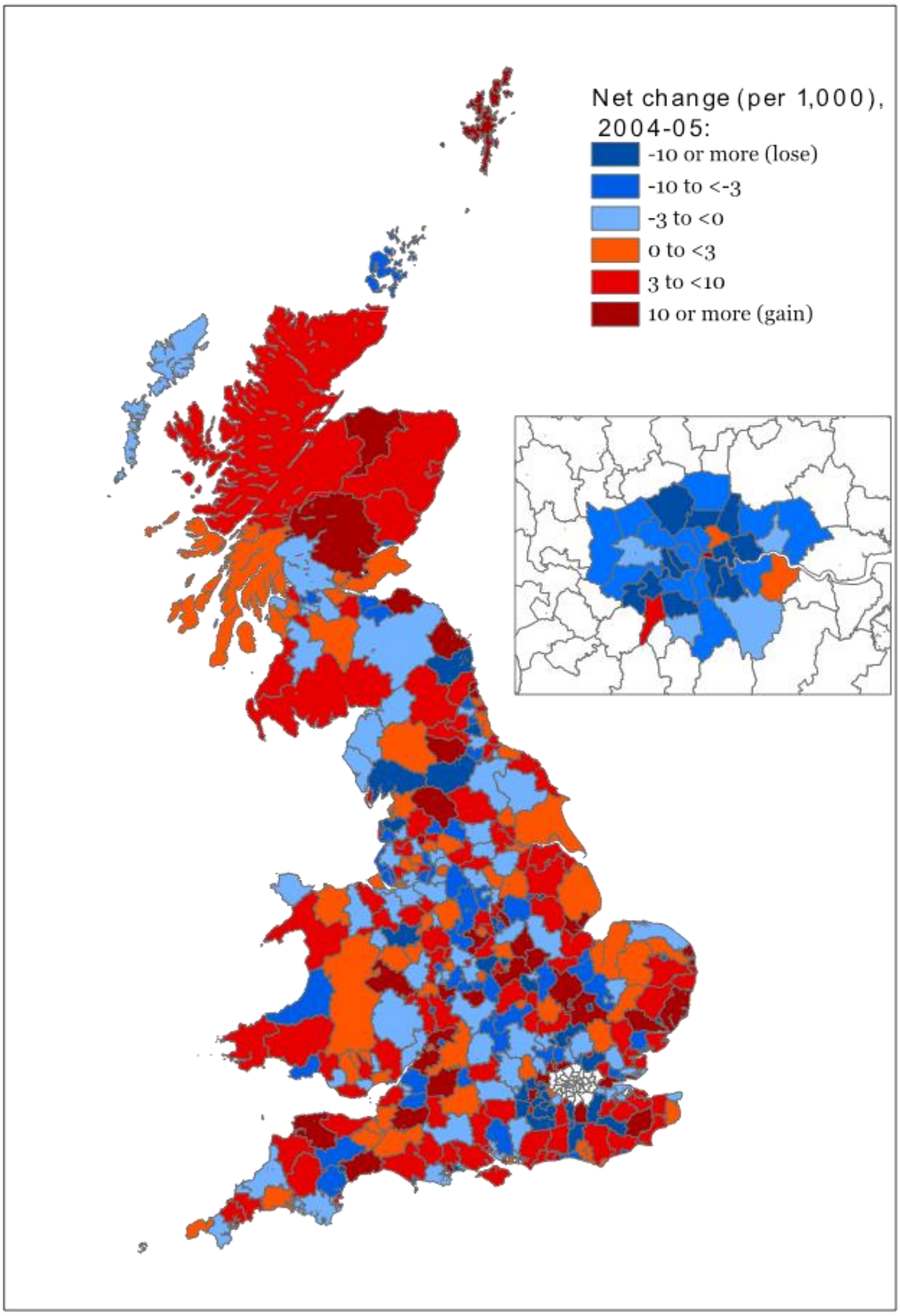


Figure 5.22. District net migration rates for GB: Raw January 2005 ROP

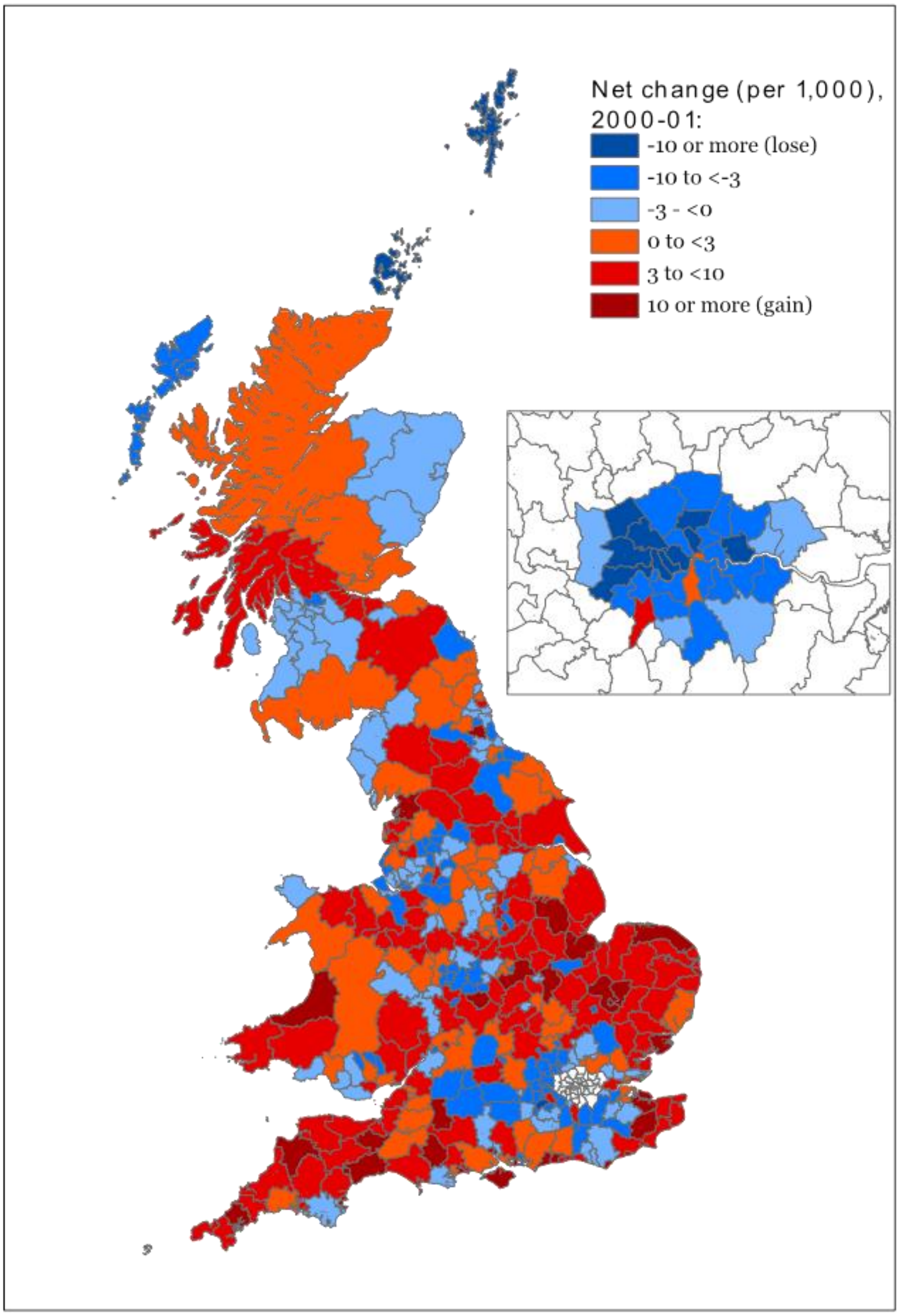


Figure 5.23. District net migration rates for GB: Census 2001

5.6 Summary and conclusions

This chapter has been concerned with outlining some data quality issues, including the data preparation and cleaning exercises, and reporting on the initial descriptive-based benchmarking exercises employed on the raw ROP cross-sections, using the raw January 2005 ROP as an example. Overall, the ROP provides rather mixed results in terms of its value for empirical/descriptive-based population migration analysis. Indeed, as has been shown in previous work by Thompson *et al.* (2010), there are concerns around the sample distributions for certain key variables including age, sex and mover status. Moreover, perhaps linked to the socio-demographic bias within the raw samples, the under sampling and over sampling of different geographical areas is also evidenced by the outliers presented in the aggregate benchmarking exercises comparing inter district migration flows. Unfortunately, if more simple descriptive-based empirical analysis is desired, the ROP will require significant sample adjustments before it can be relied upon to produce useful insights into population mobility in GB. However, the future application of sample adjustment techniques including spatial microsimulation (Harland *et al.*, 2012) may be valuable for allowing novel descriptive-based research to be undertaken, for instance the exploration of patterns in migration flows within GB for specific policy relevant population subgroups (e.g. young and highly educated adults or the long-term unemployed). However, there are positives to be drawn from the exercises reported above. For instance, the aggregate level benchmarking showed there to be significant positive correlation between the ROP inflow counts and those of the alternative population data sources. Similarly, the analyses at the micro level suggests that, despite the raw sample bias concerns, the overall patterns found in the official sources, and documented in previous research, are picked up by the ROP sample. For instance, the broad patterns to the life-course in the age specific mobility rates, the raised mobility rates for ethnic minority groups as compared to the white majority population and the significantly raised propensities for movement in the renting tenure groups, are all clearly reflected in the raw ROP. However, as was mentioned in the introduction to this chapter, the analytical focus of this project is model-based, exploring the different micro and contextual variations in residential mobility across GB. As a result, whilst the descriptive-based benchmarking provides some indication of the basic distributional distortions held within the raw ROP cross-sections, the validation of the ROP for model-based analysis requires alternative, and necessarily more complex, validation and benchmarking procedures and techniques. This is the focus of the next chapter, Chapter 6.

Chapter 6
Data validation: Model-based benchmarking

6.1 Introduction

Building on the descriptive-based benchmarking exercises of the previous chapter, where instances of clear sample bias in certain subpopulation groups of the raw ROP were revealed, this chapter seeks to explore the reliability of the ROP samples for model-based analysis. Indeed, the overall project aims are deemed to be most suitably addressed through the application of different model-based procedures, described in Chapter 4, on the detailed geo-referenced microdata held in the ROP. However, for the findings of the proposed model-based analyses to hold weight, it is important that the estimates derived are reasonably robust to the known distortions contained within the sample distributions of the ROP. With this in mind, this chapter compares estimates derived from like-for-like weighted and unweighted binary logistic regression models, where the dependent variable is move/stay. The rationale, design and application of this comparative sampling weight strategy is provided in sections 6.3 and 6.4. Moreover, in keeping with the emphasis on benchmarking the ROP with official statistics, additional models are calibrated to compare the ROP estimates with those of the Census 2001 Individual SAR.

All models are calibrated so as to allow for the analysis of variations in the associational patterns of demographic, socio-economic and behavioural/lifestyle characteristics for movers as compared to non-movers. However, the substantive analytical discussion contained within this chapter is deliberately very brief. By restricting the substantive analysis of the models, the attention can be focussed more specifically on the primary focus of this chapter, that is, the assessment of the reliability of model-based estimates through the comparison of (un)weighted model estimates and Census SAR benchmarking. Chapter 7 builds on what is revealed in this chapter, and therefore attempts to take the analytical focus of the models a stage further by exploring how the intricate, and interlinked, micro-level behaviours and characteristics of movers and non-movers vary according to broad life-course stages.

6.2 Item and unit nonresponse in the ROP

Whilst the size of the ROP has advantages, it is clear from empirical benchmarking that the data do not come free of problems. Indeed, the raw ROP samples contain inherent individual- and area-level biases on a number of important characteristics including: age, sex, geography, ethnic group, income group and mover/stayer status (see Chapter 5; also Thompson *et al.*, 2010). Such biases can be expected to be driven, to a large extent, by survey (unit) non-response and errors in the sampling frame. Unfortunately, as was noted in Chapter 3, due to commercial sensitivity, we do not know basic survey response rates; nor is it possible to obtain information on the addresses of those who failed to provide a response. Moreover, excluding the responding household's current postcode address, the ROP microdata are delivered in raw format, where concerns surrounding missing and/or 'impossible' values are left for the end user to evaluate and deal with. Regrettably, whilst multiple imputation techniques (Rubin, 1987; 1996) may present a theoretically and statistically sound method for dealing with question (item) non-response, the nature of the ROP, both in terms of its size and the magnitude of missingness in some variables (e.g. highest qualification includes 163,923, or 40 per cent, non-responses in the raw January 2005 ROP, excluding Northern Ireland) (see Table 6.1), means that this approach is not computationally feasible given the requirement for multiple imputed datasets of sizes ≈ 400,000 records. Moreover, more simple single imputation methods, such as hot deck imputation (Andridge and Little, 2010), are avoided due to their potential for introducing further bias into the sample such as distributional peaking at the mean/modal value of heavily imputed variables, their failure to account for the multivariate associations within the data, something that is particularly important for regression-based analysis (Bethlehem *et al.*, 2011), and their tendency to underestimate the uncertainty of the imputed/introduced data when calculating estimates (where imputed values are taken to be true, and thus variances are not appropriately inflated) (Little, 2008).

Table 6.1. Rates of item nonresponse in selected variables of the raw January 2005 ROP

Variable	Item Missing	Nonresponse %
Highest qualification	163,923	40.01
Annual gross household income	119,691	29.22
Like your neighbourhood	107,419	26.22
Ethnic group	80,084	19.55
Household size	74,988	18.30
Type of house	61,005	14.89
Occupation	47,997	11.72
Housing tenure	46,884	11.44
Marital status	13,426	3.28
Age	3,947	0.96
Sex	0	0.00

Consequently, following extensive efforts to clean and retain as much of the raw ROP data as possible (Chapter 5), given the advantage of the inherent size of the samples and the lack of any suitably superior alternative options, list-wise deletion (synonymous with complete case analysis) is used on each cross-section, thus removing records that fail to provide usable values for the key variables of interest. Given the scale of missing and/or 'impossible' values in the raw ROP data, the cleaned complete case samples for the January 2005-07 ROPs, whilst still large in comparison to conventional government surveys, have been reduced to approximately a third of the size of their raw equivalents. A comparison of the raw sample and complete case sample for each ROP cross-section, including the respective migrant subsamples, is provided later in the chapter (Table 6.6). In addition, Table 6.7 provides a useful summary and reminder of the different (sub)samples used within the thesis, with pointers to the relevant chapters included.

6.3 A practicable strategy for nonresponse adjustment: Survey raking

Whilst the ROP is able to generate a sample of suitable size, coverage and detail to make it attractive for use in an analysis of population mobility, the combination of its undocumented approach to data collection and the raw state in which the data are delivered does make the task of benchmarking and validating this alternative commercial data source a rather challenging prospect. However, beyond the initial descriptive-based benchmarking of Chapter 5, it is possible, through different applications of auxiliary population data, to further extend the practicable approach to validation and explore the value of the ROP for use in model-based analyses of population mobility. Indeed, aside from the benchmarking of model-based estimates against those drawn from the Census 2001 SAR, auxiliary population data can additionally serve a purpose in the adjustment of certain key ROP sample distributions for which inconsistencies have been found (e.g. age, sex, geography, mover status); that is, where ROP sample distributions are adjusted so as to be aligned with the relevant GB population distributions. Sample raking, also known as raking ratio estimation (Kalton, 1983) or iterative proportional fitting (IPF) (Deming and Stephan, 1940; Deming, 1943), is a technique that repeatedly adjusts sampling weights in an attempt to rebalance the sample response counts to known population totals. The derived sampling weights can be used to provide a degree of protection against potential distortions in model-based parameters by accounting for the unequal probabilities of selection within the ROP samples, a particularly useful trait when acknowledging that little prior information on the sample design is publically known. By employing the bespoke weights within a comparative framework of like-for-like weighted and unweighted models, detailing the relative differences in the estimated coefficients, it is possible to uncover the stability and robustness of results drawn from the now cleaned complete case ROP cross-sections. The raking procedure is explained in the following subsection, before the weighted and unweighted binary logistic regression models, calibrated for the analysis of mover/stayer characteristics, are specified.

6.3.1 The survey raking procedure

The ideal scenario would be to construct a complete multi-way cross-tabulation of relevant variables, wherein a multi-dimensional table is created with known population counts for each cell value before rebalancing the survey values to these population counts. However, if it was deemed necessary to reweight the sample by post-stratifying according to say age (15 categories), sex (2 categories), region (10 categories) and migrant status (2 categories) a multi-dimensional population table with ($15_{age}*2_{sex}*10_{region}*2_{migrant}$) or 600 known population cells would be required. Such a level of detail could be problematic, if not impossible, given the lack of available/sufficient population data and the likelihood that some demographic and geographical sub-groups (i.e. particular combinations of variable categories) do not exist empirically in the ROP sample(s).

Raking, on the other hand, can be thought of as broadly similar to fitting a log linear model for the probability of being observed in a particular cell of the complete multi-dimensional cross-tabulation of variable categories, given the probabilities for the known marginal distributions (Little and Wu, 1991). Therefore, continuing the example above, we would only require a marginal adjustment table with ($15_{age}+2_{sex}+10_{region}+2_{migrant}$) or 29 marginal counts; however, the limitations associated with the available demographic subgroups in the sample still restrict the number of population margins used. Raking is practically very useful as it allows for the use of marginal counts from different data sources; for instance, the ONS mid-year population estimates can be used to derive timely and accurate GB population estimates of age, sex and geographical region for those aged 18 and over, and the APS can be used to derive 12 month residential mover counts also for the GB population aged 18 and over.

For the decision on which variables to use in marginal adjustment, Lumley (2010: 153) cites an experiment by Keeter *et al*. (2000) which compared two identical telephone surveys, one of which paid serious attention to reducing non-response (response rate 60%) and the other less so (response rate 36%). Indeed, the published results show that, before any reweighting, the differences in demographic variables for the respondents were far larger than the differences in their political and social attitudes, the latter being the chosen outcome variables. This pattern is common, with nonresponse rates in postal surveys often closely reflected in respondents' demographic characteristics, with higher rates of nonresponse often observed for young adults, men and ethnic minorities (Bethlehem *et al.*, 2011). Yet beyond this, from a practical point of view, it is the basic demographic attributes of the population that are most routinely and reliably collected by national statistics agencies, and thus the characteristics that are most commonly published and available for use. Details on the marginal adjustment variables used in the validation exercise of this chapter are given in section 6.4 and include marginal population counts for age, sex, Government Office Region (GOR), and mover/non-mover status.

With access to detailed documentation of the ROP sampling strategy not available, we must, and can only, have the initial assumption that the complete case ROP cross-section is equally weighted (i.e. each individual within the sample carries the same weight). Therefore, in the case of the unweighted data, the individual weights w_i, where $i = 1, \ldots, n$, are equal to 1, and thus $w_i = 1$ for each individual i. With this initial vector of equal weights, modification can begin, using the iterative raking algorithm to reflect the unequal probabilities of selection in the ROP sample. Once the final weights are generated, they can be used within model-based analyses to provide a degree of protection, through the incorporation of known population data, against potential unequal nonresponse distortions and, once compared with unweighted equivalents, allow for inconsistencies in parameter estimates to be exposed.

Drawing on previous examples (Deming and Stephan, 1940; Bishop *et al.*, 1975; Simpson and Tranmer, 2005; and Battaglia *et al.*, 2009), the raking algorithm can now be defined. With the requirement to rake on a number of ROP variables, one can imagine a multi-dimensional table where the sum of the initial w_i in cell θ is defined as w_θ with a set of levels $q = 1, \ldots, s$ varying for each of the known population control totals T, with $T_{\theta q}$ corresponding to cell θ at level q. The algorithm proceeds iteratively, modifying the initial weights w_θ and thus producing new multidimensional totals m_θ that are superscripted with the number of the step. The first step of the first iteration uses the initial sample cell totals and fits these to the initial marginal levels (marginal subtotals) in order to derive our first modified estimates:

$$m_\theta^{(1)} = w_\theta^{(0)} \frac{T_\theta}{w_{\theta 1}^{(0)}} \qquad (6.1)$$

This process is repeated for all of the q levels where the first cycle (r) of the required s steps is completed:

$$m_\theta^{(s)} = m_\theta^{(s-1)} \frac{T_{\theta s}}{m_{\theta s-1}^{(s-1)}} \qquad (6.2)$$

In general, at the tth step, where $t - q$ is a multiple of s, the modified estimate is defined as:

$$m_{\theta}^{(t)} = m_{\theta}^{(t-1)} \frac{T_{\theta_q}}{m_{\theta_q}^{(t-1)}} \tag{6.3}$$

Iteration occurs until the rth cycle, where $t = rs$, and where the estimate $m_{\theta}^{(rs)}$ satisfies a predetermined convergence criterion δ^r, for example 0.1 or 0.0001, at which point a further complete r cycle fails to modify any cell by more than this pre-specified criterion (Bishop *et al.*, 1975: 85), thus:

$$|m_{\theta}^{(rs)} - m_{\theta}^{(rs-s)}| < \delta^r \tag{6.4}$$

With the desired level of accuracy achieved, the final modified sampling weights are obtained, ready for use within the necessary analyses.

6.3.2 A worked example of the raking procedure

To further aid understanding of the process, a simple two-dimensional example of the procedure, using real data, can now be worked through. The two variables used in the example are gross annual household income and household tenure. The marginal population totals for gross annual household income are weighted estimates derived from the 2006-2007 Survey of English Housing with the marginal totals for household tenure coming from the 2006 General Household Survey. The totals were adjusted so that, when summed, they agreed with the ONS Mid-2005 Population Estimates for individuals aged 18+ in Great Britain ($N =$ 45,775,200). The sample data used are from the complete case pooled ROP (Table 6.6, $n =$ 348,953) (combining all cases from the January 2005, 2006, and 2007 ROPs) where each individual is equally weighted (i.e. each individual has a weight equal to 1, $w_i = 1$ for each i). In the initial two-dimensional table (Table 6.2), the row totals refer to the marginal population control totals for income while the column totals refer to the marginal population control totals for tenure. Each cell value (θ) is the sum of the sampled individuals (i), where $w_i = 1$, whose characteristics match the corresponding margins.

Table 6.2. Two-dimensional example of raking (IPF) procedure: Initial values

Income ↓	**Tenure →**	**Owns home**	**Council rent**	**Housing association rent**	**Private rent**
		32,972,701	*4,829,504*	*3,342,199*	*4,630,796*
Up to £9,999	*3,432,360*	29,912	21,103	9,685	10,714
£10,000-£19,999	*9,111,355*	59,701	15,183	7,946	11,584
£20,000-£29,999	*8,420,083*	55,734	5,771	3,456	7,538
£30,000-£39,999	*8,813,724*	42,506	2,049	1,319	4,421
£40,000-£49,999	*6,891,122*	25,719	685	373	2,281
£50,000 plus	*9,106,556*	28,740	257	184	2,092

N.B. Italicised control totals indicate population control totals (or agreement with population control totals).

The first step (s) of the first cycle (r) is described in Equation 6.1 and involves fitting the initial cell totals (w_θ) to the corresponding marginal (row) population income totals (T_θ) (Table 6.3).

Table 6.3. Two-dimensional example of raking (IPF) procedure: Fitting to marginal population income totals (cycle 1, step 1)

Income ↓	**Tenure →**	**Owns home**	**Council rent**	**Housing association rent**	**Private rent**
		35,589,051.62	3,746,095.84	2,006,873.27	4,433,179.27
Up to £9,999	*3,432,360.00*	1,437,655.81	1,014,270.21	465,488.65	514,945.32
£10,000-£19,999	*9,111,355.00*	5,761,401.96	1,465,224.47	766,823.00	1,117,905.57
£20,000-£29,999	*8,420,083.00*	6,472,984.54	670,247.85	401,382.18	875,468.43
£30,000-£39,999	*8,813,724.00*	7,448,775.27	359,067.91	231,142.30	774,738.52
£40,000-£49,999	*6,891,122.00*	6,099,276.16	162,448.16	88,457.17	540,940.51
£50,000 plus	*9,106,556.00*	8,368,957.87	74,837.24	53,579.97	609,180.93

At the end of the first step, the counts in each cell will sum to the known control totals for income but will not sum to the column control totals for tenure. It follows therefore that the second and step of the first cycle is to fit the now modified cell totals (m_θ) to the corresponding marginal population totals for tenure (Table 6.4).

Table 6.4. Two-dimensional example of raking (IPF) procedure: Fitting to marginal population income totals (cycle 1, step 2)

	Tenure →	**Owns home**	**Council rent**	**Housing association rent**	**Private rent**
Income ↓		*32,972,701.00*	*4,829,504.00*	*3,342,199.00*	*4,630,796.00*
Up to £9,999	3,952,686.56	1,331,965.68	1,307,607.24	775,213.73	537,899.91
£10,000-£19,999	9,671,617.92	5,337,849.02	1,888,981.95	1,277,048.78	1,167,738.17
£20,000-£29,999	8,444,155.67	5,997,119.17	864,090.19	668,452.34	914,493.97
£30,000-£39,999	8,558,300.71	6,901,174.06	462,913.92	384,938.89	809,273.84
£40,000-£49,999	6,572,682.32	5,650,884.19	209,429.78	147,314.47	565,053.88
£50,000 plus	8,575,756.82	7,753,708.88	96,480.91	89,230.80	636,336.23

With the second step completed, the cell values have been modified so as to match the tenure margins. However, as is clear in Table 6.4, they now no longer match with the population margins for income (Table 6.2). As is described in Equation 6.3, this process continues, raking on each dimension, until we reach the rth cycle and the estimate ($m_{\theta}^{(rs)}$) satisfies the convergence criterion (δ^{r}) of 0.001in this example. After 14 cycles, the desired level of accuracy was achieved with the results shown in Table 6.5.

For this worked example, the final modified sampling weights for each sampled individual can be obtained through a simple calculation: dividing the cell total w_{θ} (the sum of the sampled individuals (i), where the original sampling weights are specified as equal, $w_i = 1$, whose characteristics match of the given cell θ) (Table 6.2), by the final modified cell total $m_{\theta}^{(rs)}$ (Table 6.5).

Table 6.5. Two-dimensional example of raking (IPF) procedure: Convergence criterion satisfied (cycle 14, step 2)

	Tenure →	**Owns home**	**Council rent**	**Housing association rent**	**Private rent**
Income ↓		*32,972,701.00*	*4,829,504.00*	*3,342,199.00*	*4,630,796.00*
Up to £9,999	*3,432,360.000*	1,104,293.73	1,176,796.27	691,839.82	459,430.18
£10,000-£19,999	*9,111,355.000*	4,880,074.75	1,874,650.84	1,256,782.74	1,099,846.67
£20,000-£29,999	*8,420,083.000*	5,873,861.85	918,697.47	704,764.51	922,759.18
£30,000-£39,999	*8,813,724.000*	7,030,366.74	511,902.95	422,123.29	849,331.01
£40,000-£49,999	*6,891,122.000*	5,883,273.34	236,686.48	165,097.52	606,064.66
£50,000 plus	*9,106,556.000*	8,200,830.59	110,769.99	101,591.10	693,364.31

We are effectively dividing the now modified cell frequency between its members in the sample. In this example, a homeowner with a gross annual household income of £30,000-£39,000 has a sampling weight approximately equal to 165.397 (7,030,366,743 ÷ 42,506 = 165.397), and therefore is estimated to represent 165.397 individuals in the 18+ GB population[6].

6.4 Data and measures

In keeping with the desire to use this unique source of data for the analysis of population movement in GB, binary logistic regression models are employed but with adjustments that take into account the sampling weights (Section 6.5). The binary response is non-mover (0) and mover (1) with the selected covariates reflecting some of the key demographic and socio-economic characteristics that previous studies have shown to be important in explaining the likelihood of population migration. However, beyond this, the ability to explore some of the more subjective/personal and seemingly understudied characteristics of movers and non-movers, for instance their neighbourhood (output area) characteristics (OAC), neighbourhood satisfaction, household income and plans for a future move, allows for additional dimensions to this analysis, and the relevant variables are therefore included in the models. The OAC (Vickers and Rees, 2006, 2007) is a hierarchical

[6] If necessary, the probability of selection for each sampled individual can be calculated as the reciprocal of the sampling weight (e.g. $1/165.397 = 0.006046$).

geodemographic classification of small areas into groups based on the similarity of the demographic, socio-economic and housing profile of their residents; all of which are factors raised in the literature as being potentially important factors for influencing neighbourhood attractiveness and more general residential satisfaction. Defined at the 2001 Census OA level of geography, for which there are 175,434 in England and Wales with each comprising on average a population of 297 individuals and 124 households (Martin, 2002a; 2002b), the OAC provides us with an independent census based measure of the immediate neighbourhood context. Drawn from the OAC's three-level hierarchy (7, 21, 52 clusters respectively), this analysis employs the second level which contains 21 geodemographic groups ranging, for instance, from OAs defined as 'Terraced blue collar' and 'Public housing' to those categorised as 'Accessible countryside', 'Senior communities', and 'Prospering younger families'. The rationale behind the choice of the reference category used for each explanatory variable varies; for ordinal categorical variables, the median value is used; while for nominal variables, the modal values in the sample and, occasionally, the most typical in the population, are used.

As was mentioned in Chapter 3, whilst five separate ROP cross-sections are available, issues of consistency in the questions asked and the requirement for the inclusion of certain key demographic, socio-economic, lifestyle, mobility and address information in the analysis, the results presented here are based on the January 2005, January 2006 and January 2007 cross-sections only. Replicate microdata models are calibrated for the different ROP samples –January 2005 (n = 125,945), January 2006 (n = 50,686), and January 2007 (n = 172,322) – as well as on the pooled data (n = 348,953) so as to explore data/model consistency across the separate samples. There are a number of apparent advantages to the increased sample size associated with the pooling of the ROP data, including: the potential for greater precision in the estimates; an increase in the migrant subsample; and the reduced risk of sparsity, wherein there are small numbers within certain sampled sub-groups. Given the small (two-year) temporal variation in the sample, it is necessary to incorporate dummy variables (indicating which sample the respondent is member of) within the models to control for any unwanted influence associated with this variation. Table 6.6 provides a breakdown of the numbers of movers and non-movers in each sample as well as the percentage that moved. Movers are specified as individuals who have changed address in the 12 months prior to survey completion, providing full address details of their previous residence, with non-movers making up the remainder of the cases.

Table 6.6. Residential mobility status for the selected ROP data sets

Residential mobility status	January 2005	January 2006	January 2007	Pooled
Non-mover	121,551	49,711	168,337	339,599
Mover	4,394	975	3,985	9,354
% movers	3.49	1.96	2.37	2.68
Complete case *n*	125,945	50,686	172,322	348,953
Raw *n*	405,794	198,026	346,838	950,658

N.B. Mobility status totals refer to the complete case samples.

The numbers presented in Table 6.6 refer to the complete case analytical samples (number 3, Table 6.7), which contain records that provided usable answers to all the variables obtained for the proceeding analyses, and the raw *n* refers to the raw samples after cleaning but before list-wise deletion (number 2, Table 6.7), therefore excluding those from Northern Ireland and those who failed to provide even the very basic indicators of age and sex. Whilst movers as a percentage are clearly underrepresented in the sample (Dennett and Stillwell, 2010), a relatively large subsample of movers in absolute terms is still retained, particularly when the data are pooled.

Table 6.7. Selected ROP sample hierarchy and corresponding chapters

Sample hierarchy	January 2005	January 2006	January 2007	Pooled	Corresponding chapter(s)
1. Raw delivered: UK including Northern Ireland	411,325	314,580	349,588	1,075,493	Chapter 5
2. Raw cleaned - GB with usable PC, age & sex data (subsample of 1)	405,794	198,026	346,838	950,658	Chapter 6
3. GB: Analytical complete case (subsample of 2)	125,945	50,686	172,322	348,953	Chapters 6 & 7
4. England & Wales: Duration of residence < 20 years (subsample of 3)	75,979	32,240	115,945	224,164	Chapter 8
5. England & Wales: Moved in previous 3 years with full origin and destination postcode address (subsample of 3)	14,685	3,372	8,631	26,688	Chapter 9

N.B. PC = refers to the respondent's current postcode address

The January 2005, January 2007 and Pooled weighted models presented here use sampling weights that have been adjusted according to marginal population totals for age, sex, Government Office Region (GOR), and mover/non-mover status. Due to the relatively small sample size in the January 2006 ROP (especially for the mover sub-group, Table 6.6), the sampling weights designed for the January 2006 weighted model are limited to the use of population totals for age, sex and mover/non-mover status only. The inclusion of geography, even at the regional level, is not possible due to the nonexistence of sampled individuals in certain cells of the required multi-dimensional adjustment table[7]. Theoretically, we can rake on as many variables as we have population data for; however, the size of the sample limits us to a select few in practice. Tables 6.8-6.12 provide details on the sources of the population data and a full breakdown of the population counts for each marginal population total. All subtotals are constrained before the raking procedure to meet the 18+ Mid-2005 Population Estimates for Great Britain ($N = 45{,}775{,}200$) which themselves reflect ONS revisions due to improved migration measures.

Table 6.8. Government Office Region (GOR) population totals

GOR	Population
North East A	2,074,000
North West B	5,503,900
Yorkshire D	4,124,800
East Midlands E	3,503,600
West Midlands F	4,282,800
East of England G	4,472,800
London H	6,046,000
South East J	6,591,200
South West K	4,158,400
Wales W	2,384,500
Scotland X	4,165,800
GB total (16+)	47,307,800

Source: Table 8 of the Mid-2005 Population Estimates: Selected age groups for local authorities in the United Kingdom; estimated resident population.

7 There are 484 cells in the multi-dimensional adjustment table for age (11), sex (2), geography (11), and mover/non-mover status (2) and only 44 cells in the adjustment table used for the January 2006 ROP sample.

Table 6.9. Age group population totals

Age	Population
18-24	5,345,300
25-29	3,651,700
30-34	4,051,100
35-39	4,511,800
40-44	4,475,300
45-49	3,926,300
50-54	3,566,800
55-59	3,812,400
60-64	3,030,100
65-69	2,641,800
70+	6,762,600
GB total (18+)	45,775,200

Source: Table 2 of the Mid-2005 Population Estimates: Great Britain; estimated resident population by single year of age and sex; reflecting revisions due to improved migration. Office for National Statistics, General Register Office for Scotland.

Table 6.10. Sex group population totals

Sex	Population
Male	22,118,300
Female	23,656,600
GB total (18+)	45,774,900

Source: Table 2 of the Mid-2005 Population Estimates: Great Britain; estimated resident population by single year of age and sex; reflecting revisions due to improved migration. Office for National Statistics, General Register Office for Scotland.

Table 6.11. Mover/non-mover group totals

Length of residence	Population
Less than 12 months	4,032,346
More than 12 months	39,344,060
GB total (18+)	43,376,406

Source: Quarterly Labour Force Survey Household Dataset, April - June, 2005. Weight: Person household weight. Crown copyright material is reproduced with the permission of the Controller of HMSO and the Queen's Printer for Scotland.

6.5 Model specification

As was shown in Chapter 4, the binary logistic regression model with multiple predictor variables $x_1, x_2, \ldots, x_k$ can be written, following Heeringa *et al.* (2010), as:

$$logit[\pi_i(X_i)] = \ln\left(\frac{\pi_i(X_i)}{1-\pi_i(X_i)}\right) = \beta_0 + \beta_1 x_{1i} + \cdots + \beta_k x_{ki} \qquad (6.5)$$

where, in this case, $\pi_i(X_i)$ is the conditional probability of y occurring ($y = 1$ (in this case, having changed residence)) for individual i, given the vector of observed predictor variables, X for individual i. In the models presented here, where all variables are categorical in nature, β_0 represents the constant term, which contains all of the reference categories associated with each predictor variable. $\beta_1, \ldots, \beta_k$ are the logistic regression coefficients, where β_k gives the change in the log odds of $y = 1$ for a given category k within a predictor variable when compared to the odds that $y = 1$ for the reference category within that variable. Once the model is fitted, $\pi_i(X_i)$ can be recovered from the log scale through the following function:

$$\hat{\pi}_i(X_i) = \frac{\exp(\hat{\beta}_0 + \hat{\beta}_1 x_{1i} + \cdots + \hat{\beta}_k x_{ki})}{1 + \exp(\hat{\beta}_0 + \hat{\beta}_1 x_{1i} + \cdots + \hat{\beta}_k x_{ki})} \quad (6.6)$$

By exponentiating the estimated parameters, $\hat{\beta}$, a more meaningful interpretation is provided where, for the variables modelled here, $\exp(\hat{\beta})$ (the odds ratio) represents the change in the estimated ratio of the odds of $y = 1$ for a given category within a predictor variable, when compared to the odds that $y = 1$ for the reference category. For a simple random sample, the binary logistic regression coefficients and standard errors are estimated using maximum likelihood based on the binomial distribution (Agresti, 2002). The likelihood function for logistic regression with a binomial dependent variable can be written as:

$$L(\beta|X) = \prod_{i=1}^{N} \pi(X_i)^{y_i} \, [1 - \pi(X_i)]^{1-y_i} \quad (6.7)$$

where:

$$\pi(X_i) = \frac{\exp(X_i\beta)}{[1 + \exp(X_i\beta)]} \quad (6.8)$$

However, when sampling weights are included, the use of maximum likelihood estimation is no longer possible because the probabilities of selection for the sample observations are no longer equal (Heeringa *et al.*, 2010). Consequently, an alternative method of pseudo-maximum likelihood estimation (Binder, 1981; 1983) can be used which allows for complex sample characteristics to be modelled correctly by making use of the sampling weights (w_i), the observed sample values

(y_i) and the modelled $\pi(X_i)$ values (Heeringa *et al*., 2010). Therefore, the weighted pseudo-likelihood function for logistic regression with a binomial dependent variable is defined as:

$$PL(B|X) = \prod_{i=1}^{n} \{\pi(X_i)^{y_i} \cdot [1 - \pi(X_i)]^{1-y_i}\}^{w_i} \tag{6.9}$$

where:

$$\pi(X_i) = \frac{\exp(X_i B)}{[1 + \exp(X_i B)]} \tag{6.10}$$

In line with Heeringa *et al*. (2010), the parameters β are changed to B and now represent finite population parameters, which are the weighted function of the observed sample values (y_i) and the estimated $\pi(X_i)$ values. Therefore, the weighted pseudo-likelihood function for logistic regression with a binomial dependent variable (Equation 6.9) is used in the weighted models with maximum likelihood based on the binomial distribution (Equation 6.7) being used in the unweighted models presented below. Finally, in terms of evaluating model goodness-of-fit (GOF), a number of statistics discussed in Chapter 4 are provided at the bottom of Tables 6.12-6.15.

6.6 Comparing unweighted and weighted regression model results

The results of the unweighted and weighted main effects models for each ROP sample can be seen in Tables 6.12-6.15 and Figures 6.1-6.4. For each tabular comparison (Tables 6.12-6.15), the relative difference in the odds ratios (in percentage terms) are provided in order for us to assess the extent to which the weighted and unweighted models diverge. It should be noted that the estimated odds ratio for the constant has no real substantive analytical value; however, for comparative purposes, in terms of measuring the relative difference, it is included in Tables 6.12-6.15. The plotting of the results in Figures 6.1-6.4 greatly helps in assessing not only the (dis)similarities in the directional patterns, but also in comparing the size of effects and therefore the relative substantive importance, above and beyond the simple statistical significance, that certain characteristics may have over others in terms of their associated relationship with residential (im)mobility in GB. To be clear, an estimated coefficient ($\hat{\beta}$) that falls to the right of the dashed line (marking zero – i.e. no difference) suggests that individuals with this characteristic are, *ceteris paribus*, more likely to have moved than those with the reference characteristic of a given categorical predictor. Estimated coefficients that fall to the left of the line, therefore, suggest a move is less likely than it is for the reference.

Table 6.12. January 2005 ROP: Main effects comparison and relative difference

Predictor	**January 2005 unweighted**			**January 2005 weighted**			**Relative difference**
	Beta	**S.E.**	**Odds**	**Beta**	**S.E.**	**Odds**	**(%)**
Constant	-4.495*	0.103	0.011	-3.885*	0.125	0.021	-83.903
Age (ref: 45-49)							
18-19	2.435*	0.178	11.418	2.370*	0.190	10.700	6.287
20-24	2.094*	0.081	8.120	2.096*	0.102	8.136	-0.197
25-29	1.601*	0.075	4.958	1.616*	0.085	5.033	-1.503
30-34	1.181*	0.073	3.257	1.188*	0.081	3.281	-0.747
35-39	0.704*	0.074	2.022	0.706*	0.081	2.025	-0.138
40-44	0.293*	0.077	1.340	0.305*	0.083	1.357	-1.227
50-54	-0.161*	0.086	0.851	-0.183*	0.091	0.833	2.117
55-59	-0.228*	0.086	0.796	-0.271*	0.091	0.762	4.273
60-64	-0.409*	0.097	0.664	-0.384*	0.103	0.681	-2.481
65-69	-0.410*	0.106	0.664	-0.340*	0.113	0.712	-7.224
70-74	-0.421*	0.117	0.656	-0.393*	0.125	0.675	-2.831
75-79	-0.693*	0.144	0.500	-0.683*	0.151	0.505	-0.952
80+	-0.903*	0.178	0.405	-0.826*	0.186	0.438	-8.024
Gender (ref: Female)							
Male	-0.157*	0.036	0.854	-0.135*	0.047	0.874	-2.264
Ethnic group (ref: white)							
Asian	0.249*	0.113	1.283	0.062	0.134	1.063	17.118
Black	0.560*	0.131	1.751	0.334*	0.164	1.396	20.263
Other	-0.077	0.112	0.926	-0.162	0.150	0.851	8.112
Marital status (ref: single)							
Married	0.010	0.050	1.010	0.157*	0.058	1.170	-15.911
Living with partner	0.450*	0.051	1.568	0.558*	0.059	1.748	-11.478
Divorced/separated	0.543*	0.057	1.721	0.562*	0.064	1.755	-1.956
Widowed	0.240*	0.099	1.271	0.170	0.110	1.185	6.780
Occupation (ref: Higher managerial administrative and professional occupations)							
Not economically active	0.003	0.035	1.003	0.031	0.041	1.032	-2.846
Routine and manual occupations	0.061	0.040	1.063	0.118*	0.047	1.126	-5.888
Intermediate occupations	0.024	0.039	1.024	-0.066	0.047	0.936	8.592

Table 6.12. (continued)

	January 2005 unweighted			January 2005 weighted			Relative difference
	Beta	S.E	Odds	Beta	S.E.	Odds	(%)
Annual gross household income (ref: £20,000-£29,999)							
Up to £9,999	0.115*	0.058	1.122	-0.004	0.067	0.996	11.217
£10,000-£19,999	0.064	0.047	1.066	0.004	0.055	1.004	5.835
£30,000-£39,999	-0.089*	0.053	0.915	0.052	0.062	1.054	-15.156
£40,000-£49,999	0.109*	0.047	1.115	0.024	0.056	1.025	8.073
£50,000 plus	0.022	0.039	1.022	0.069	0.047	1.072	-4.820
Highest qualification (ref: 5 or more GCSEs)							
No formal qualifications	0.152*	0.034	1.165	0.179*	0.040	1.196	-2.719
2+ 'A' levels	0.144*	0.035	1.154	0.143*	0.042	1.153	0.088
First degree and higher	-0.099*	0.039	0.906	-0.131*	0.047	0.877	3.174
Tenure (ref: Own home)							
Council rent	0.039*	0.057	1.173	0.168*	0.067	1.183	-0.846
Housing association rent	0.281*	0.068	1.324	0.236*	0.082	1.266	4.363
Private rent	0.752*	0.045	2.122	0.732*	0.054	2.080	1.978
Type of home (ref: Semi-detached)							
Detached	0.324*	0.055	1.383	0.219*	0.065	1.245	9.962
Terraced	0.089*	0.044	1.094	0.126*	0.053	1.134	-3.725
Bungalow	0.695*	0.069	2.004	0.518*	0.083	1.678	16.261
Maisonette	0.169	0.111	1.185	0.209	0.133	1.233	-4.077
Flat	0.520*	0.054	1.682	0.512*	0.067	1.669	0.728
OAC Super-group level (ref: Typical traits)							
Blue collar communities	-0.133*	0.051	0.875	-0.117*	0.061	0.889	-1.613
City living	-0.172*	0.082	0.842	-0.090	0.102	0.914	-8.562
Countryside	-0.021	0.061	0.980	-0.005	0.072	0.995	-1.579
Prospering Suburbs	-0.117*	0.055	0.890	-0.115*	0.067	0.892	-0.222
Constrained by circumstances	-0.036	0.056	0.965	-0.020	0.067	0.980	-1.577
Multicultural	-0.491*	0.076	0.612	-0.429*	0.094	0.651	-6.359
Plan to move in next 12 months (ref: No)							
Yes	-0.040	0.047	0.961	-0.075	0.056	0.927	3.503
Like your neighbourhood (ref: No)							
Yes	0.441*	0.060	1.555	0.389*	0.074	1.476	5.067

Null deviance	38122 on 125944 *df*
Residual deviance	33639 on 124896 *df*
Improvement (X^2)	4482.644*, *df* = 48
AIC	33737

*N.B. n = 125,945. * indicates parameter is significant at the 95 % level. The GOF summary measures relate to the unweighted model only, such statistics are currently not incorporated in the R 'survey' (Lumley, 2012) package software for complex sample survey data analysis.*

The modelled results for the January 2005 ROP (Table 6.12 and Figure 6.1) are reassuring with the similarity in the direction and magnitude of the weighted and unweighted estimates immediately apparent. Moreover, beyond the simple similarities, the coefficients of both models suggest relationships commonly cited in the literature (Chapter 2). Indeed, it appears that age (stage in life course) is, as we would expect, a very significant influence on the propensity to move, with the younger age groups having higher propensities to move than those in the older age categories. Other findings that suggest a substantively important relationship with mover/non-mover status can be found for marital status, with the likelihood of moving being far greater for those living with a partner and those that are divorced/separated than those that are single; and tenure, with renters having a far greater likelihood of moving than home owners. The OAC functional geographies suggest varying propensities to move, however, in substantive terms, those living in multicultural neighbourhoods tend to be characterised by greater immobility than those living in areas that reflect more typical traits. Finally, it appears that greater neighbourhood satisfaction is associated with recent movers. Somewhat of a surprise, here and in the following comparative models, is the relative unimportance of occupational class, household income and educational attainment, for which conventional theories would suggest are important selective characteristics. However, as will is discussed in particular detail in Sections 7.4.2 and 7.5, these findings are highly likely to be a relic of the analytical framework than the data used within it.

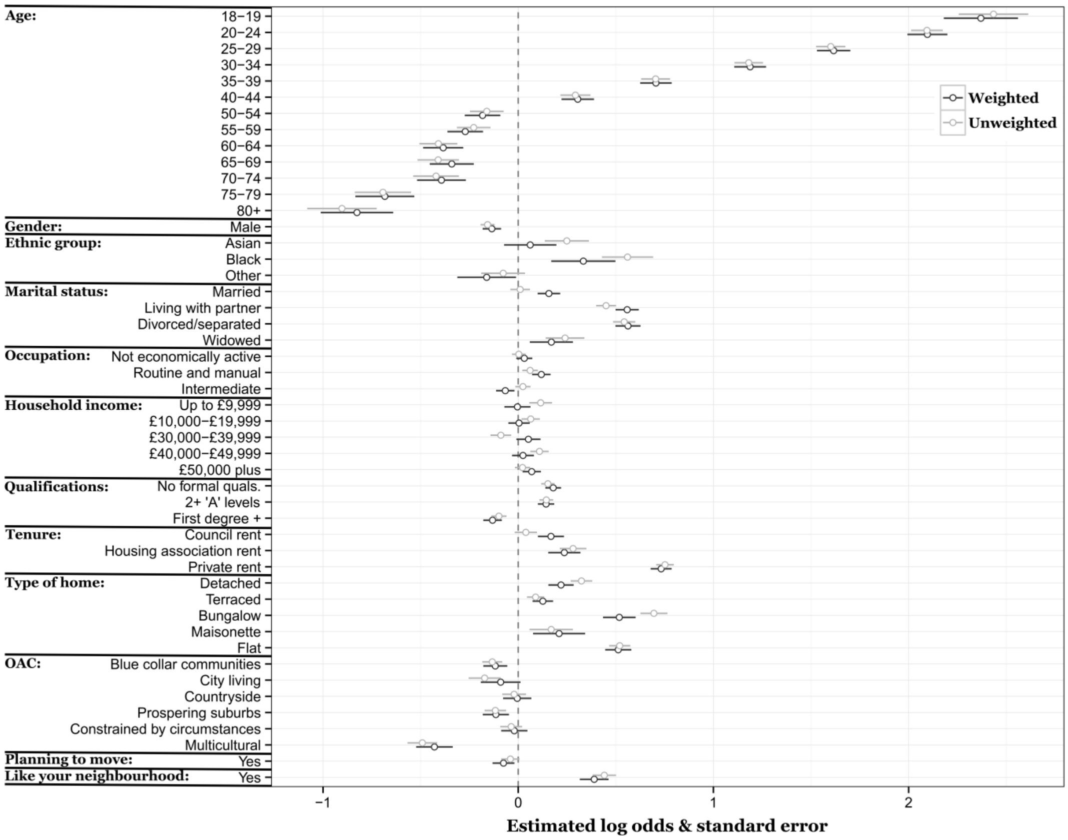


Figure 6.1. January 2005 ROP weighted and unweighted model estimates

In terms of the stability between the model estimates, there are only two cases (the constant and Black ethnic group) where the relative difference in the estimated coefficient odds ratio has exceeded the 20 per cent level. However, for both the constant and Black ethnic groups, the directional patterns (+/-) remain in agreement[8]. The models do present contradictory estimates, where one model suggests a positive/negative associational pattern in contrast to the other. These additional contradictory estimates are the household income groups “up to £9,999” and “£30,000-£39,999”, yet in both cases, the contradictory estimates are statistically non-significant in the weighted model with the size of the standard errors suggesting that both estimates could easily have pointed to the same directional association suggested by the unweighted model.

---

8 By definition the application of weights should change the intercept due to adjustments in the proportion of respondents with $Y = 1$.

Table 6.13. January 2006 ROP: Main effects comparison and relative difference

Predictor	January 2006 unweighted			January 2006 weighted			Relative difference
	Beta	S.E.	Odds	Beta	S.E.	Odds	(%)
Constant	-5.329*	0.221	0.005	-3.672*	0.257	0.025	-424.384
Age (ref: 45-49)							
18-19	1.523*	0.316	4.585	1.428*	0.338	4.169	9.062
20-24	1.480*	0.167	4.394	1.362*	0.219	3.905	11.129
25-29	1.245*	0.149	3.474	1.101*	0.179	3.008	13.403
30-34	1.023*	0.137	2.782	0.926*	0.156	2.524	9.280
35-39	0.600*	0.132	1.822	0.560*	0.144	1.751	3.909
40-44	-0.012	0.144	0.988	-0.042	0.151	0.958	2.982
50-54	-0.506*	0.173	0.603	-0.476*	0.181	0.621	-3.066
55-59	-0.329*	0.165	0.720	-0.335*	0.173	0.715	0.616
60-64	-0.382*	0.183	0.682	-0.384*	0.191	0.681	0.199
65-69	-0.629*	0.222	0.533	-0.666*	0.228	0.514	3.674
70-74	-0.469*	0.225	0.626	-0.473*	0.235	0.623	0.418
75-79	-0.954*	0.295	0.385	-0.886*	0.306	0.412	-7.090
80+	-0.541*	0.279	0.582	-0.474	0.298	0.623	-6.944
Gender (ref: Female)							
Male	-0.161*	0.075	0.851	-0.269*	0.091	0.764	10.194
Ethnic group (ref: white)							
Asian	0.248	0.218	1.281	0.085	0.278	1.089	14.982
Black	-0.042	0.350	0.959	-0.156	0.400	0.855	10.788
Other	0.530*	0.227	1.698	0.226	0.287	1.253	26.185
Marital status (ref: single)							
Married	0.151	0.110	1.163	0.125	0.127	1.133	2.593
Living with partner	0.795*	0.112	2.214	0.846*	0.131	2.331	-5.302
Divorced/separated	0.413*	0.126	1.511	0.165	0.138	1.180	21.945
Widowed	0.358*	0.195	1.430	0.041	0.211	1.042	27.152
Occupation (ref: Higher managerial administrative and professional occupations)							
Not economically active	0.050	0.071	1.051	0.069	0.084	1.072	-1.984
Routine and manual occupations	0.177*	0.083	1.194	0.204*	0.096	1.226	-2.697
Intermediate occupations	-0.107	0.078	0.898	-0.129	0.094	0.879	2.160
Annual gross household income (ref: £20,000-£29,999)							
Up to £9,999	0.049	0.107	1.050	-0.081	0.122	0.923	12.116
£10,000-£19,999	0.006	0.088	1.006	-0.119	0.099	0.888	11.771
£30,000-£39,999	-0.132	0.099	0.876	-0.032	0.113	0.969	-10.565
£40,000-£49,999	-0.115	0.089	0.892	-0.240*	0.104	0.787	11.802
£50,000 plus	0.007	0.079	1.007	-0.004	0.092	0.996	1.058
Highest qualification (ref: 5 or more GCSEs)							
No formal qualifications	0.275*	0.066	1.317	0.381*	0.077	1.464	-11.162
2+ 'A' levels	0.216*	0.073	1.241	0.270*	0.087	1.310	-5.515
First degree and higher	-0.104	0.083	0.901	-0.055	0.101	0.947	-5.037
Tenure (ref: Own home)							
Council rent	0.025	0.130	1.026	-0.074	0.154	0.928	9.506

Table 6.13. (continued)

	January 2006 unweighted			January 2006 weighted			Relative difference
	Beta	S.E.	Odds	Beta	S.E.	Odds	(%)
Housing association rent	0.467*	0.129	1.596	0.311*	0.150	1.364	14.514
Private rent	0.685*	0.100	1.983	0.515*	0.122	1.674	15.604
Type of home (ref: Semi-detached)							
Detached	0.165	0.105	1.179	-0.030	0.124	0.971	17.670
Terraced	0.081	0.095	1.085	0.076	0.113	1.079	0.530
Bungalow	0.483*	0.129	1.621	0.313*	0.149	1.368	15.580
Maisonette	0.549*	0.227	1.732	0.508*	0.267	1.662	4.035
Flat	0.780*	0.118	2.182	0.762*	0.149	2.142	1.846
OAC Super-group level (ref: Typical traits)							
Blue collar communities	-0.122	0.112	0.886	-0.066	0.134	0.936	-5.707
City living	-0.239	0.178	0.788	-0.164	0.216	0.849	-7.817
Countryside	0.176	0.118	1.193	0.229	0.141	1.257	-5.397
Prospering Suburbs	0.063	0.107	1.065	0.084	0.127	1.087	-2.088
Constrained by circumstances	-0.152	0.125	0.859	-0.101	0.157	0.904	-5.278
Multicultural	-0.575*	0.169	0.563	-0.613*	0.201	0.542	3.737
Plan to move in next 12 months (ref: No)							
Yes	-0.162	0.121	0.851	0.006	0.138	1.006	-18.293
Like your neighbourhood (ref: No)							
Yes	0.848*	0.146	2.335	0.679*	0.164	1.972	15.518
Null deviance	9635.5 on 50685 *df*						
Residual deviance	8752.7 on 50637 *df*						
Improvement (X^2)	882.834*, *df* = 48						
AIC	8850.7						

*N.B. n = 50,686. * indicates parameter is significant at the 95 % level. The GOF summary measures relate to the unweighted model only, such statistics are currently not incorporated in the R 'survey' (Lumley, 2012) package software for complex sample survey data analysis.*

The model results for the 2006 ROP (Table 6.13 and Figure 6.2) suggest that the comparability between the weighted and unweighted models is somewhat less impressive. However, this is not unexpected given the substantial (approx. 60 per cent) reduction in the sample size relative to the 2005 ROP. The general directional associations and patterns depicted in Figure 6.2 suggest that the substantive findings again appear to be fairly well reflected in both. As with the 2005 results, there is strong evidence of the important role that age (stage in life course) plays on the likelihood of moving or staying, with the younger age groups being generally more likely to move than those in more elderly age groups. Again, as with the 2005 results, the likelihood of moving is found to be far greater for those living with a partner than those who are single. Additionally, those living in flats as well as those who rent privately or from a housing association, are on average, significantly more likely to have moved in the 12 months prior to the survey than those who live in semi-detached accommodation and those who own their property. As before, we also associate greater neighbourhood satisfaction with those who move residence as opposed to those who do not.

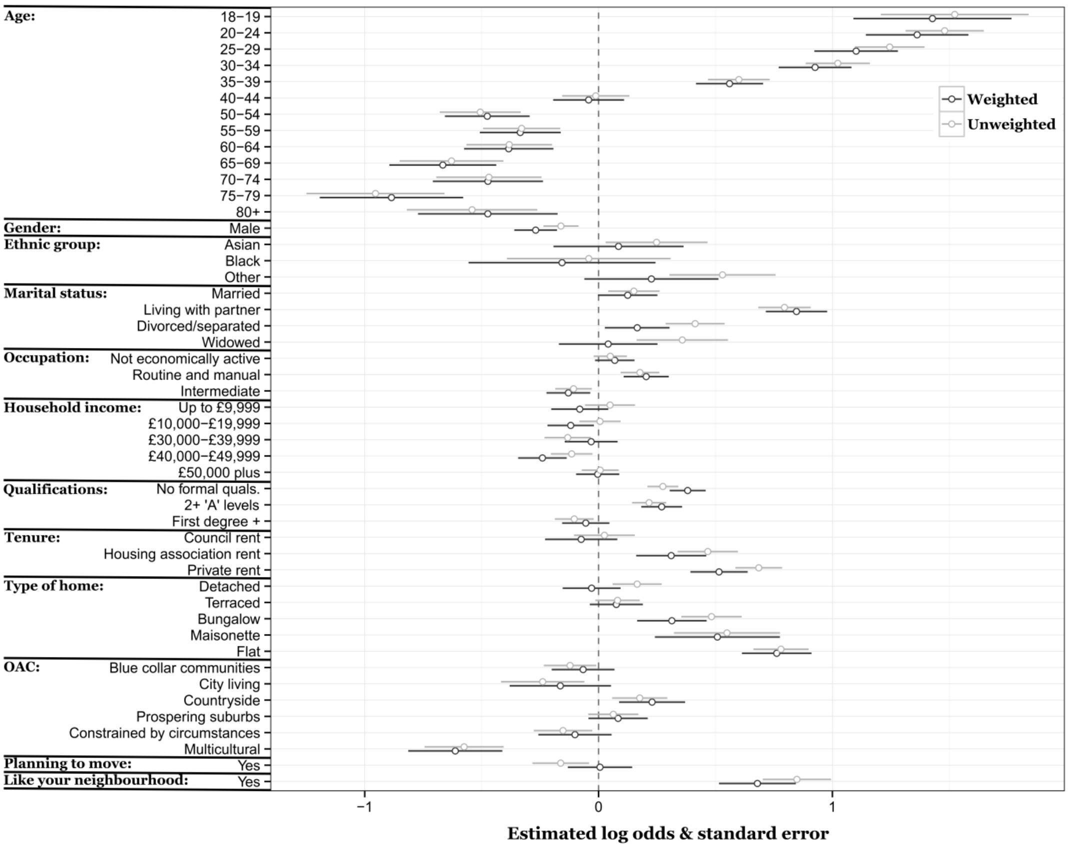


Figure 6.2. January 2006 ROP weighted and unweighted model estimates

When thinking about the stability in the estimated odds ratios, and while accepting that the comparability between the estimates is less impressive than the January 2005 ROP, none of the observed contradictions should be considered particularly problematic. For the 2006 analysis, there are four cases where the relative difference in the estimated coefficient odds ratio exceeds the ±20 per cent point (the constant, Other ethnic group, divorced/separated and widowed) but again the relative differences do not result in a disagreement with the direction (+/-) of the associations. There are contradictions in the models' estimates, however, in all cases (detached housing; council rent; income up to £9,999, £10,000-£19,999, £50,000 plus; and planning to move), the substantive effects are very small and statistically non-significant in both models.

Table 6.14. January 2007 ROP: Main effects comparison and relative difference

Predictor	January 2007 unweighted			January 2007 weighted			Relative difference
	Beta	S.E.	Odds	Beta	S.E.	Odds	(%)
Constant	-5.061*	0.109	0.006	-3.686*	0.124	0.025	-295.792
Age (ref: 45-49)							
18-19	1.254*	0.168	3.505	1.304*	0.181	3.685	-5.148
20-24	1.448*	0.085	4.255	1.491*	0.106	4.441	-4.358
25-29	1.204*	0.075	3.333	1.251*	0.086	3.494	-4.842
30-34	0.829*	0.074	2.291	0.850*	0.083	2.339	-2.130
35-39	0.583*	0.073	1.792	0.609*	0.080	1.838	-2.595
40-44	0.207*	0.076	1.230	0.234*	0.083	1.263	-2.731
50-54	-0.040	0.084	0.961	-0.051	0.090	0.951	1.040
55-59	-0.093	0.086	0.911	-0.097	0.092	0.907	0.457
60-64	-0.045	0.090	0.956	-0.043	0.096	0.958	-0.238
65-69	-0.150	0.107	0.861	-0.133	0.113	0.875	-1.696
70-74	-0.246*	0.125	0.782	-0.255*	0.132	0.775	0.949
75-79	-0.521*	0.153	0.594	-0.478*	0.166	0.620	-4.415
80+	-0.853*	0.189	0.426	-0.789*	0.199	0.455	-6.686
Gender (ref: Female)							
Male	0.011	0.035	1.012	0.017	0.042	1.017	-0.530
Ethnic group (ref: white)							
Asian	-0.235*	0.116	0.791	-0.326*	0.131	0.722	8.667
Black	-0.484*	0.167	0.616	-0.506*	0.198	0.603	2.166
Other	-0.230*	0.139	0.794	-0.353*	0.152	0.702	11.586
Marital status (ref: single)							
Married	0.058	0.054	1.060	0.129*	0.061	1.138	-7.361
Living with partner	0.545*	0.054	1.724	0.606*	0.060	1.833	-6.325
Divorced/separated	0.443*	0.064	1.557	0.454*	0.071	1.575	-1.173
Widowed	0.348*	0.101	1.417	0.363*	0.110	1.437	-1.436
Occupation (ref: Higher managerial administrative and professional occupations)							
Not economically active	0.170*	0.034	1.185	0.187*	0.037	1.206	-1.711
Routine and manual occupations	0.019	0.036	1.019	0.026	0.039	1.026	-0.670
Intermediate occupations	0.031	0.038	1.031	0.067*	0.040	1.069	-3.681
Annual gross household income (ref: £20,000-£29,999)							
Up to £9,999	0.068	0.050	1.070	0.000	0.054	1.000	6.607
£10,000-£19,999	0.042	0.041	1.043	0.014	0.045	1.014	2.767
£30,000-£39,999	-0.045	0.049	0.956	-0.021	0.054	0.980	-2.507
£40,000-£49,999	0.070	0.045	1.073	0.053	0.050	1.054	1.741
£50,000 plus	0.071*	0.039	1.074	0.079*	0.043	1.082	-0.755
Highest qualification (ref: 5 or more GCSEs)							
No formal qualifications	0.149*	0.033	1.160	0.181*	0.036	1.198	-3.242
2+ 'A' levels	0.074*	0.036	1.076	0.065*	0.039	1.068	0.819
First degree and higher	-0.129*	0.041	0.879	-0.194*	0.045	0.823	6.309
Tenure (ref: Own home)							
Council rent	-0.281*	0.069	0.755	-0.291*	0.077	0.748	1.033

Table 6.14. (continued)

	January 2007 unweighted			January 2007 weighted			Relative difference
	Beta	S.E.	Odds	Beta	S.E.	Odds	Beta
Housing association rent	-0.129*	0.077	0.879	-0.169*	0.087	0.844	3.878
Private rent	0.159*	0.050	1.172	-0.029	0.061	0.972	17.112
Type of home (ref: Semi-detached)							
Detached	0.266*	0.053	1.305	0.181*	0.057	1.199	8.144
Terraced	0.151*	0.047	1.163	0.179*	0.053	1.197	-2.849
Bungalow	0.869*	0.060	2.386	0.812*	0.066	2.251	5.628
Maisonette	0.276*	0.122	1.318	0.240*	0.134	1.271	3.535
Flat	0.790*	0.057	2.204	0.839*	0.066	2.313	-4.975
OAC Super-group level (ref: Typical traits)							
Blue collar communities	-0.259*	0.056	0.772	-0.229*	0.062	0.796	-3.058
City living	-0.255*	0.083	0.775	-0.222*	0.096	0.801	-3.340
Countryside	0.197*	0.059	1.218	0.278*	0.065	1.320	-8.381
Prospering Suburbs	0.084	0.052	1.087	0.154*	0.058	1.166	-7.251
Constrained by circumstances	-0.178*	0.062	0.837	-0.201*	0.070	0.818	2.344
Multicultural	-0.271*	0.076	0.762	-0.288*	0.088	0.750	1.634
Plan to move in next 12 months (ref: No)							
Yes	-0.345*	0.054	0.708	-0.317*	0.059	0.728	-2.807
Like your neighbourhood (ref: No)							
Yes	0.631*	0.069	1.880	0.563*	0.078	1.756	6.593
Null deviance	37899 on 172321 *df*						
Residual deviance	35770 on 172273 *df*						
Improvement (X^2)	2129.008*, *df* = 48						
AIC	35868						

*N.B. n = 172,322. * indicates parameter is significant at the 95 % level. The GOF summary measures relate to the unweighted model only, such statistics are currently not incorporated in the R 'survey' (Lumley, 2012) package software for complex sample survey data analysis.*

The results for the weighted and unweighted models using January 2007 ROP data (Table 6.14 and Figure 6.3) are more consistent than both of the previous data sets. The substantive patterns seen in the 2005 and 2006 ROPs reappear, with the greatest likelihood of mobility found for the youngest age groups and the greatest immobility in the eldest age groups. The importance of the type of accommodation is reemphasised with those living in flats or bungalows characterised by having greater mobility rates, on average, than those who live in semi-detached accommodation. Marital status is also found to have a statistically significant and reasonably large effect on propensities to move with those living with their partner being particularly more likely to move than those who are single. Greater immobility is observed for those in Asian, Black and Other ethnic groups, when compared to those from White ethnic backgrounds. Again, as with the 2005 ROP findings, individuals living in 'multicultural' neighbourhoods tend to be characterised by greater immobility than those living in areas characterised by more 'typical traits', with those living in 'blue collar communities' and areas 'constrained by circumstances' also characterised by particularly greater immobility. Greater satisfaction with their neighbourhood and a lower likelihood of planning for a future move are also significantly associated with movers when compared to stayers.

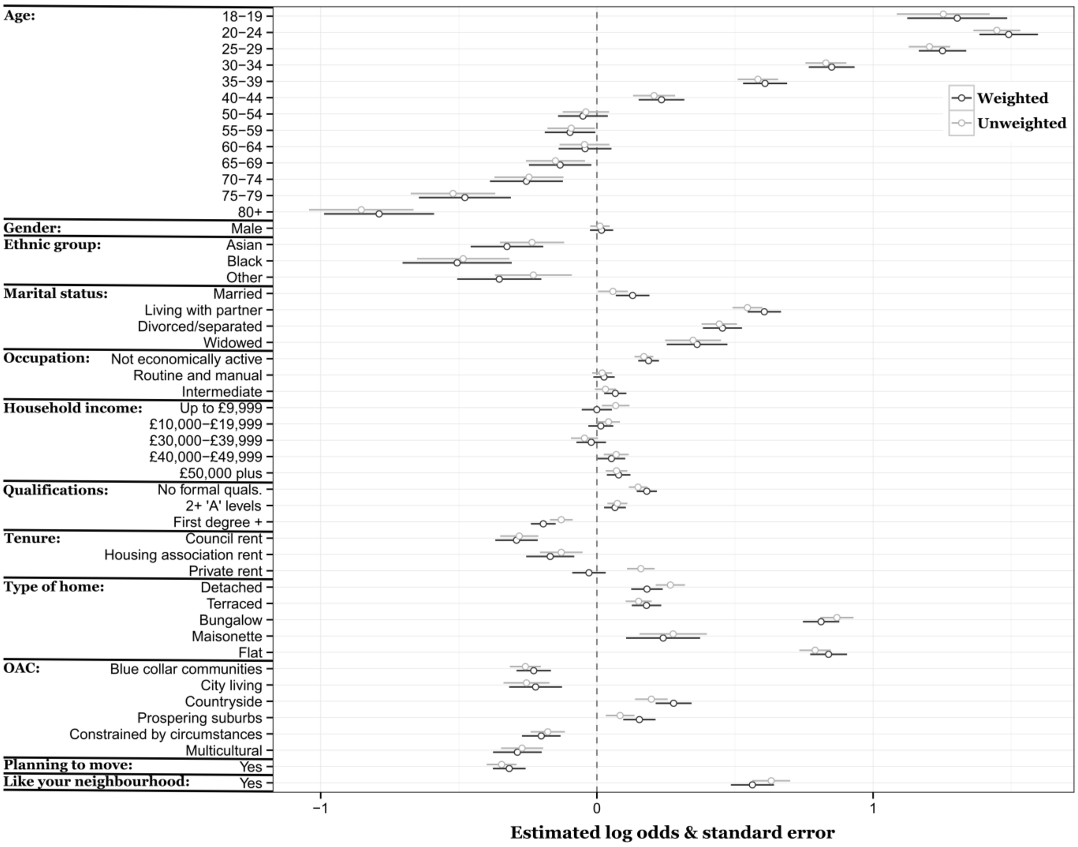


Figure 6.3. January 2007 ROP weighted and unweighted model estimates

In terms of consistency in the model estimates, only the constant has a relative difference in the estimated coefficient odds ratio that exceeds the ±20 per cent mark. Moreover, the only example of a contradictory estimate is for private rent; however, the effects are very small in both models and the standard error in the weighted model crosses zero.

Table 6.15. Pooled (January 2005-07) ROP: Main effects comparison and relative difference

Predictor	Pooled unweighted			Pooled weighted			Relative difference
	Beta	S.E.	Odds	Beta	S.E.	Odds	(%)
Constant	-4.455*	0.071	0.012	-3.262*	0.081	0.038	-229.592
Age (ref: 45-49)							
18-19	1.610*	0.111	5.001	1.592*	0.117	4.914	1.732
20-24	1.726*	0.054	5.616	1.724*	0.068	5.607	0.159
25-29	1.374*	0.050	3.950	1.385*	0.057	3.996	-1.150
30-34	1.009*	0.049	2.742	1.013*	0.054	2.754	-0.446
35-39	0.644*	0.048	1.904	0.651*	0.053	1.917	-0.683
40-44	0.220*	0.051	1.246	0.228*	0.055	1.256	-0.801
50-54	-0.156*	0.056	0.856	-0.178*	0.060	0.837	2.258
55-59	-0.205*	0.057	0.815	-0.244*	0.061	0.784	3.798
60-64	-0.268*	0.061	0.765	-0.294*	0.065	0.746	2.513
65-69	-0.372*	0.069	0.689	-0.397*	0.073	0.672	2.438
70-74	-0.433*	0.078	0.649	-0.498*	0.082	0.608	6.274
75-79	-0.732*	0.097	0.481	-0.767*	0.103	0.464	3.513
80+	-0.887*	0.116	0.412	-0.891*	0.122	0.410	0.445
Gender (ref: Female)							
Male	-0.089*	0.023	0.915	-0.082*	0.029	0.922	-0.700
Ethnic group (ref: white)							
Asian	-0.005	0.076	0.995	-0.149*	0.087	0.862	13.406
Black	0.025	0.097	1.025	-0.107	0.113	0.898	12.351
Other	-0.082	0.081	0.922	-0.156	0.097	0.855	7.175
Marital status (ref: single)							
Married	0.066*	0.034	1.069	0.148*	0.039	1.159	-8.488
Living with partner	0.549*	0.034	1.732	0.636*	0.039	1.890	-9.128
Divorced/separated	0.492*	0.040	1.635	0.495*	0.044	1.640	-0.268
Widowed	0.319*	0.066	1.376	0.291*	0.071	1.337	2.827
Occupation (ref: Higher managerial administrative and professional occupations)							
Not economically active	0.139*	0.022	1.149	0.165*	0.024	1.180	-2.677
Routine and manual occupations	-0.015	0.024	0.986	-0.017	0.027	0.984	0.203
Intermediate occupations	-0.103*	0.023	0.902	-0.121*	0.026	0.886	1.779
Annual gross household income (ref: £20,000-£29,999)							
Up to £9,999	0.085*	0.035	1.088	-0.002	0.038	0.998	8.334
£10,000-£19,999	0.051*	0.029	1.052	0.007	0.032	1.007	4.324
£30,000-£39,999	-0.034	0.033	0.966	0.046	0.036	1.047	-8.360
£40,000-£49,999	0.043	0.030	1.044	-0.014	0.034	0.986	5.535
£50,000 plus	0.051*	0.026	1.052	0.077*	0.029	1.080	-2.648
Highest qualification (ref: 5 or more GCSEs)							
No formal qualifications	0.183*	0.022	1.200	0.224*	0.024	1.251	-4.203
2+ 'A' levels	0.134*	0.023	1.143	0.137*	0.026	1.147	-0.346
First degree and higher	-0.123*	0.026	0.884	-0.170*	0.030	0.844	4.536
Tenure (ref: Own home)							
Council rent	-0.016	0.041	0.984	-0.051	0.045	0.950	3.425

Table 6.15. (continued)

	Pooled unweighted			Pooled weighted			Relative difference
	Beta	S.E.	Odds	Beta	S.E.	Odds	(%)
Housing association rent	0.134*	0.047	1.144	0.058	0.054	1.060	7.311
Private rent	0.428*	0.031	1.534	0.276*	0.037	1.317	14.127
Type of home (ref: Semi-detached)							
Detached	0.279*	0.036	1.321	0.176*	0.040	1.192	9.759
Terraced	0.125*	0.030	1.133	0.159*	0.035	1.172	-3.467
Bungalow	0.785*	0.042	2.193	0.695*	0.048	2.003	8.674
Maisonette	0.261*	0.077	1.299	0.262*	0.087	1.299	-0.030
Flat	0.676*	0.037	1.966	0.704*	0.044	2.021	-2.813
OAC Super-group level (ref: Typical traits)							
Blue collar communities	-0.180*	0.036	0.835	-0.153*	0.040	0.858	-2.768
City living	-0.228*	0.055	0.796	-0.179*	0.065	0.836	-4.980
Countryside	0.109*	0.040	1.115	0.176*	0.045	1.192	-6.906
Prospering Suburbs	0.002	0.035	1.002	0.056	0.041	1.057	-5.482
Constrained by circumstances	-0.103*	0.039	0.902	-0.106*	0.044	0.900	0.305
Multicultural	-0.406*	0.051	0.667	-0.401*	0.059	0.670	-0.435
Plan to move in next 12 months (ref: No)							
Yes	-0.152*	0.033	0.859	-0.139*	0.037	0.870	-1.302
Like your neighbourhood (ref: No)							
Yes	0.560*	0.043	1.750	0.489*	0.049	1.631	6.791
Data set (ref: January 2005)							
January 2006	-0.583*	0.036	0.558	-0.610*	0.041	0.543	2.677
January 2007	-0.662*	0.023	0.516	-0.768*	0.027	0.464	10.061
Null deviance	86162 on 348952 *df*						
Residual deviance	78866 on 348902 *df*						
Improvement (X^2)	7295.825*, *df* = 50						
AIC	78968						

N.B. n *= 348,953. * indicates parameter is significant at the 95 % level. The GOF summary measures relate to the unweighted model only, such statistics are currently not incorporated in the R 'survey' (Lumley, 2012) package software for complex sample survey data analysis.*

The comparisons between the weighted and unweighted models for the January 2005, 2006, and 2007 ROP samples suggest reasonable levels of reliability. Impressive levels of comparability are also observed, in terms of the direction and magnitude of the associational patterns, across the different survey cross-section for: life course, gender, marital status, tenure, type of home, occupational class, and neighbourhood satisfaction. Subsequently, a similar investigation of the pooled data (combining all cases from the January 2005, 2006, and 2007 ROPs) is performed in order to determine its reliability for further, and more sophisticated, analyses in

Chapters 7, 8 and 9. Given the relatively small (two-year) temporal variation, the changes in residential mobility frequencies and overall sample sizes (Table 6.6), and the small but observable analytical variations between the ROP samples, it is deemed useful to incorporate dummy terms indicating for which sample the respondents are members of. The inclusion of the dummy terms is designed to help to control for some of the unwanted influence associated with this inter-sample variation.

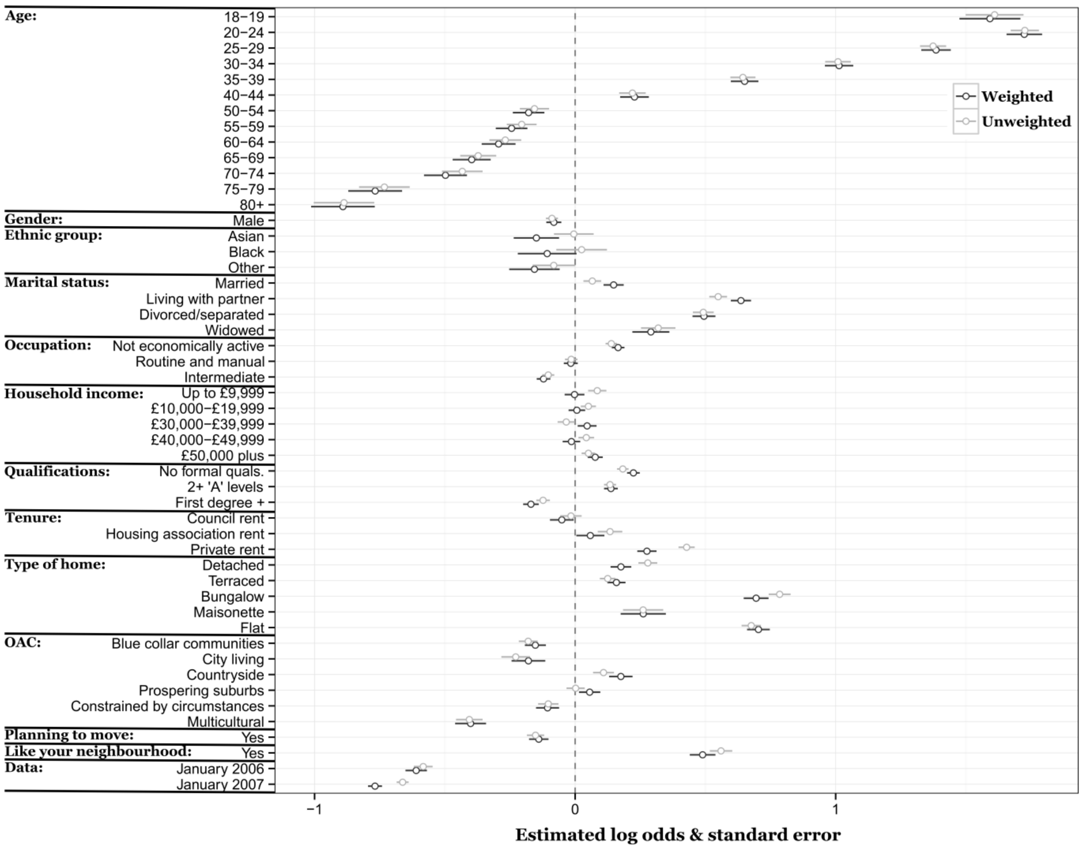


Figure 6.4. Pooled (January 2005-07) ROP weighted and unweighted model estimates

The results from the pooled models (Table 6.15 and Figure 6.4) suggest an impressive level of agreement with only the constant exceeding the ±20 per cent level of relative difference in the estimated coefficient odds ratio. Moreover, where there are directional relationship disagreements in the models (i.e. Black ethnic groups and up to £9,999, £30,000-£39,999, £40,000-£49,999 income groups), the effects are found to be substantively small and statistically non-significant (with the standard errors crossing the zero, in most cases) in at least one of the comparative models. In terms of the most influential characteristics, the prominence of age (stage in life-course) for the propensity to move/stay is striking, with the common patterns associated with marital status, home type, neighbourhood satisfaction, neighbourhood type, and plans for a future move also revealed. It is also clear that the inclusion of the (nuisance) dummy indicators for each of the ROP samples is justified given that they are both statistically significant and have relatively large effect sizes.

While the influence of nonresponse bias in the unweighted model results cannot be discounted, a reasonable degree of stability is observed both across and between the eight models. Furthermore, from an analytical point of view, the major associational patterns to do with the demographic, socio-economic and behavioural/lifestyle characteristics of movers/non-movers are repeated across each model. Finally, it is thought useful to provide a brief comparison of broadly similar logistic models using pooled ROP analytical sample data and Census 2001 Individual SAR data ($n \approx$ 1.2m), where the focus is again on mover/stayer propensities ($y = 1$ moved in previous 12 months). Of course, within this comparison, inconsistencies in the definitions of variables and the sampling frame are important to consider, for instance, the SAR includes those aged 16 and 17 in its sample and the SAR definition of single is a legal definition and not a cohabitation measure as in the ROP. Perhaps the key difference relates to the fact that the ROP refers to a household representative whereas the SAR refers to individuals. Aside from these points, the broad comparisons are presented in Figures 6.5 (Pooled ROP) and 6.6 (2001 SAR).

Again, in spite of the obvious definitional discrepancies and general difficulties in direct comparison of the two different microdata sources, encouraging comparability is observed. Age and house type are both seen to have the largest effects on mobility with very similar patterns emerging from both data sources Where the small disagreements in the age pattern are found (youngest age group), it is most likely the result of the definitional discrepancies (inclusion of 16 and 17 year olds in the SAR). Similar patterns and/or magnitudes are also found for gender, ethnic group and marital status. The main variations are largely related to the educational attainment and occupation classification variables; however, in both cases the SAR includes a fourth coefficient that measures those who are not applicable, something that is not included in the ROP and therefore removes the opportunity for direct comparability. The council and housing association renter groups are also found to be contradictory when compared in the different models; however, there are definitional variations that could explain some of these discrepancies. For instance, shared ownership schemes are recorded as homeowners in the SAR (in the ROP these are likely to be defined as housing association) and individuals who live rent free are classified as private renters in the SAR (in the ROP these individuals would provide the tenure type of the household, which could be any of the possible tenure categories). Broadly speaking, where there are substantively significant effects and where the definitions of variables are fairly comparable, the ROP and SAR show a good level of agreement adding further encouragement to the argument that model-based results drawn from the ROP can be genuinely useful for the analysis of residential movement in GB.

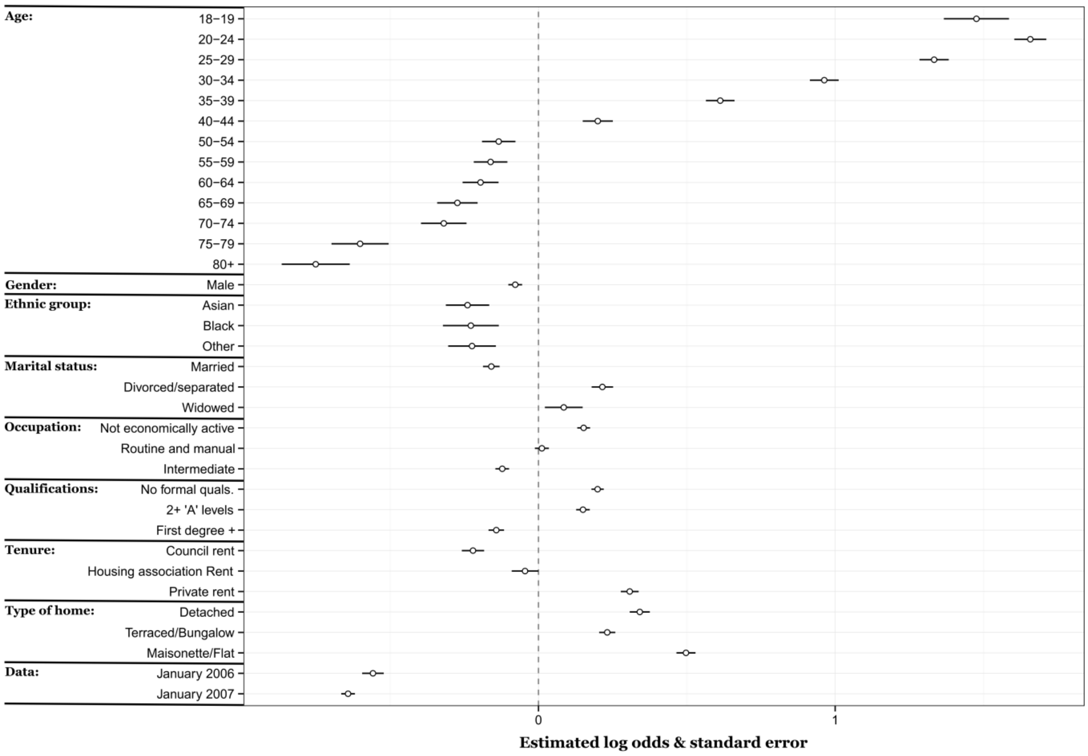


Figure 6.5. Census 2001 Individual SAR benchmarking: Pooled ROP model estimates

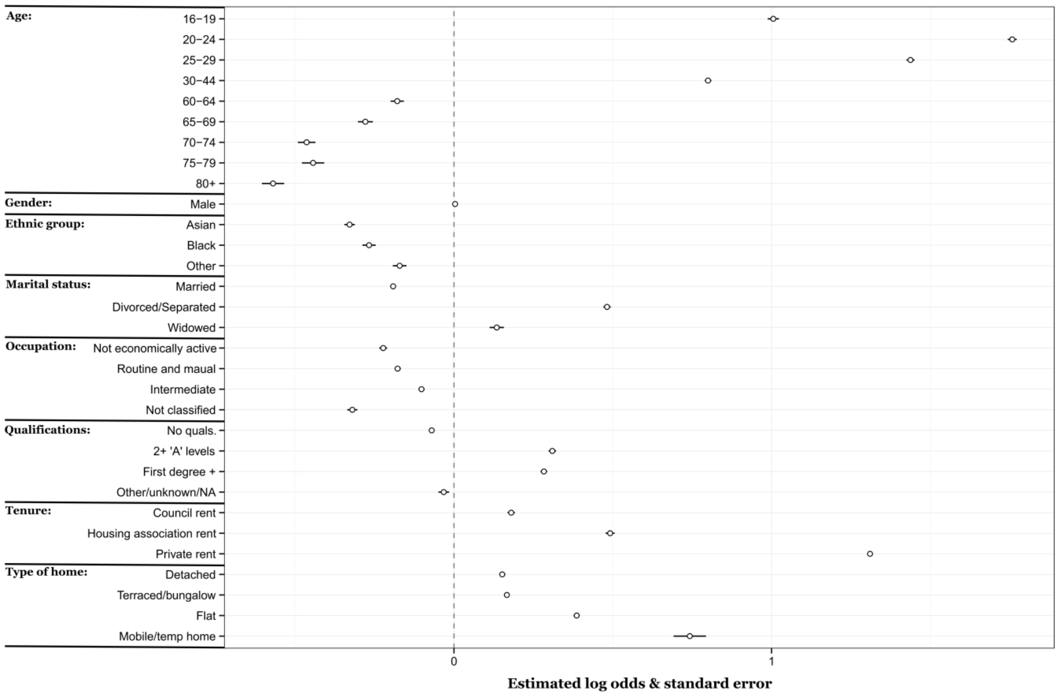


Figure 6.6. Census 2001 Individual SAR benchmarking: 2001 Individual SAR model estimates

6.7 Summary conclusions

Given the results presented above, it is argued that the ROP is a valuable source of data for the model-based exploration and analysis of population mobility in GB. Benefitting from a large geo-referenced sample, rich variable detail and an inherent flexibility, the ROP undoubtedly holds serious potential for analyses aimed at improving our understanding of how various facets of population mobility are conditioned by characteristics operating at both the individual and area (origin and destination) level. However, whilst this commercial data presents us with the opportunity to look at dimensions of population mobility previously restricted, it also makes the task of initial data management and more general validation a difficult one. Indeed, the lack of detailed knowledge on the sampling strategy and the degree of missingness associated with certain variables are two issues that require careful attention when planning analyses on the data. Therefore, the validation of the ROP data, and perhaps other sources of alternative ‘big data’, requires the researcher to be thorough as well as practicable and pragmatic in their approach. Consequently, building on the descriptive-based benchmarking of Chapter 5, this chapter has applied a sample raking technique that allows for the generation of sampling weights that incorporate known population distributions for a selection of key variables; with the purpose of uncovering some of the potential effects that unequal probabilities of selection, which are known to exist within the ROP, may have on the estimation of model-based associational relationships. With the analytical focus of the chapter being concerned with uncovering substantial variations in weighted and unweighted estimates of associational patterns related to the various demographic, socio-economic and behavioural/lifestyle characteristics of movers *vis-à-vis* stayers, a comparative framework of like-for-like weighted and unweighted binary logistic regression models was devised.

In terms of the validation of this commercial data source, it can certainly be argued that the results presented here are very encouraging. Whilst the effects of non-response bias cannot be entirely discounted, for instance contradictory relationships are observed although none are found to be substantively or statistically significant, the consistency observed across and between the weighted and unweighted model estimates is useful in showing the robustness of the model findings to the unequal probabilities of selection. That is, the covariates included in the models appear to work as suitable adjustment confounders, in controlling for sample distortions in associations between the predictors and the response, without the need for sampling weights (Lumley, 2010). Furthermore, the substantively important associational patterns found for many of the modelled characteristics conform to much of the existing empirical and theoretical literature. Indeed, further model-based benchmarking with the Census 2001 Individual SAR highlighted the consistency and comparability of the major associational relationships. Thus, the ROP is a source of data with great potential for application within the analysis of population migration in GB, and particularly for the exploration of various processes, patterns and factors for which most conventional sources of data fail to allow. In the context of ongoing discussions of alternative sources by the ONS (see Beyond 2011 programme, Chapter 3), the validation and analysis of alternative sources such as the ROP is of clear importance. Indeed, socio-demographic microdata are essential not only for academic analysis, but also for the planning and delivery of essential services now and in the future. If valuable analytical results can be obtained from detailed geo-referenced commercial data like the ROP, then stakeholders, including for instance national statistical agencies, must think seriously about working in partnership with the commercial sector to obtain and utilise data resources such as these.

In a more immediate context, the relative confidence in the data for model-based analysis of population mobility makes it particularly suitable for use in attempting to address the overall project aims set out in Chapter 1 of this thesis. Thus, building on the analytical findings shown here, Chapter 7 presents a micro-level analysis of variations in the associational behaviours and characteristics of movers and non-movers across the life course. Following this, however, the next analytical challenge is to further investigate subsamples of the pooled analytical data, exploring the relationships between the individual- and area-level variables using multilevel hierarchical models and cross-classified models. In particular, attempts are made to explore the relative contextual contributions of the origin and destination on postcode-postcode distance moved (Chapter 9), and differences between places in terms of the duration of residence effects and future mobility propensities (Chapter 8).

Chapter 7
Modelling mover/stayer characteristics across the life course

7.1 Introduction

Residential mobility is a key mechanism in the evolution of both the size and structure of local populations and is of importance to policy makers tasked to provide resources and services. Whilst there exists fairly extensive knowledge of the broad demographic and socio-economic characteristics of individuals that determine the basic propensity to migrate, further analysis of these and the other more personal/subjective characteristics (e.g. neighbourhood satisfaction and plans for future moves) of movers and non-movers, across the life course, is essential if we are to better understand the processes and behavioural mechanisms that underpin residential mobility and immobility. Consequently, this chapter builds directly on Chapter 6 and exploits the pooled ROP analytical sample for a more thorough substantive analysis of variations in residential mobility behaviour across the broad stages of the life course. In doing so, it uncovers some interesting associational patterns specifically related to some of the characteristics of movers *vis-à-vis* stayers that have, until very recently, been seriously understudied due to the lack of suitable data.

7.2 Motivations for residential mobility and immobility

As was outlined in Chapter 2, residential mobility is something that will affect almost all of us at some point in our lifetimes. Of the three demographic processes (i.e. fertility, mortality and population migration), household migration within the country usually has the largest impact on local area population size and composition (Bogue, 1969; Nam *et al.*, 1990; Rees *et al.*, 2009). Moreover, beyond the simple change in numbers, residential mobility operates to transform the demographic character and structure of populations, in some cases affecting real change to the social, cultural, physical and economic characteristics of an area. With this in mind, it is clear that the measurement and analysis of movers and non-movers, and their respective behaviours and characteristics, is a hugely important task. Research exploring the decision-making processes and experiences of movers stretches right back to seminal works by Thomas (1938) and Rossi (1955). Whilst the theoretical and empirical analyses presented in these early pioneering works have been tested, rethought and developed, time and time again, the fundamental study of mobility and immobility, in equal measure, remains essential to the sub-disciplines of demography and population geography (Courgeau and Lelievre, 2006; Cooke, 2011).

In Chapter 2, it was suggested that the decision to move or remain *in situ* can be understood to be motivated by the utility-maximising behaviours of supposed rational individuals/households. Thus, with the aim of maximising expected welfare, for some future period of time, the expected benefits and costs of moving are weighed up against the same parameters for staying (Bartel, 1979; Cushing and Poot, 2005). Moreover, this cost-benefit calculation was said to be inextricably tied to transitions through the life course, transitions that can be expected to recondition the evaluative framework through which the dynamic motivations to move/stay are defined. Given that the focus of this chapter is on the differences in mover/stayer characteristics across the life course, it is thought useful to briefly reemphasise the importance of life-course events and transitions for mobility behaviours and outcomes.

Rossi (1955) provided a very early depiction of the interrelationship between the family life course and residential mobility. His work detailed the traditional sequence of family life transitions that, by and large, remain relevant to the majority of people today. For instance, we can think of life-course transitions into adulthood associated with either a move from school to university or directly into employment, or into employment following higher education – each of which may necessitate a change of residence (Champion, 2005a; Smith, 2009). After this stage, the subsequent years, for those aged in their early 30s to mid 40s, are commonly characterised by relatively sharp reductions in mobility and are generally considered the years of family formation and child rearing. The decline is then reduced somewhat, for those aged 45-64, with more recent research linking this reduction with a transition from parenthood to ‘empty nesting’, prompting the desire, at least for some, to change residence in order to downsize (Wulff *et al.*, 2010). For the following transition into retirement and old age, the picture is more mixed, with some small but noticeable recoveries in the mobility rate associated with the exit from the labour market, but with greater immobility as older age increases (Fielding, 2012). Finally, the mobility rate is observed to increase again, to some extent, for those in the eldest age groups, commonly linked with a need for closer proximity to family members and social/health services (Evandrou *et al.*, 2010). However, whilst this normative generalisation of the life course into certain follow-on stages, each working to increase/decrease the likelihood of moving, continues to be supported and reflected in empirical analyses of census data (Figure 2.1; Duke-Williams and Stillwell, 2010), there is a growing acceptance that an increasing portion of the population do not follow a sequential trajectory. Indeed, drawing on the specific benefits of longitudinal panel data, more recent approaches to mobility analysis have attempted to emphasise the diversity in individual and interdependent life-course trajectories and events (Clark and Dieleman, 1996; Clark, 2013; Mulder and Wagner, 2012). Whether it is unemployment, pregnancy or the birth of a child, union formation/dissolution or occupational promotion, certain life-course events can occur that, whether positive or negative, expected or otherwise, operate as the causal motive behind observed residential mobility, where again mobility is understood as the rational utility-maximising outcome which is itself defined according to the selective constraints of the financial and social context within which the individual/household in question find themselves. Of course, as mentioned in Chapter 2, the availability of greater resources (e.g. income, asset wealth, education) is essential for enabling the individual/household to act on any desired move.

Thus drawing on the above, this chapter seeks to disentangle the age effect from what are the real influences behind residential mobility, that is, by uncovering the associational relationships of characteristics thought to be of importance for informing mobility outcomes at different stages of the life course. As Clark (2013: 327) makes clear, "*it is not age per se that is creating the mobility process but rather the events that occur within the ageing process*". Moreover, with the growing availability of detailed large-scale microdata sets like the ROP, there is more potential than ever before for uncovering innovative insights into differing mover/stayer characteristics. For instance, whilst there is a reasonably detailed literature on the role of several selective demographic and socio-economic factors on mobility, the availability of variables in the ROP detailing some of the more personal/subjective characteristics of individuals, such as neighbourhood satisfaction and/or plans for future moves, provides this analysis with the potential for disentangling the age effects and for making important insights into what are particularly understudied factors. The advantages of the ROP's large sample size and attribute detail are to some extent balanced by certain limitations pertaining to the cross-sectional nature of the data. Indeed, whilst it is possible to uncover the associational relationships of certain key variables, it is not possible to observe any change in the characteristics of the individual/household before or after a potential move and thus certainly not possible to make any causal inference.

7.3 Modelling framework and analysis

With this chapter building on the confidence of the model-based validation exercise in Chapter 6, the same pooled analytical sample and measures are used. However, given the substantive interest in variations across the life course, four standard binomial logistic regression models are specified (Equation 6.5), each with the purpose of exploring variations in the associational patterns of demographic, socio-economic and behavioural/lifestyle characteristics of movers when compared to non-movers for four major life-course stages: 18-29, the transition into adulthood with the associated high levels of mobility (Model 1); 30-44, traditionally the stage of family formation and reductions in mobility (Model 2); 45-64, a stage of reduced decline in mobility (Model 3); and finally 65+, the transition into retirement and old age and relatively low propensities to move (Model 4). Each model is designed to accommodate the potential differential effects of age at smaller intervals within these broader life-course groupings.

The rationale behind initially using four separate models, instead of a single all-embracing model, is related to the modelling of interaction effects. By separating the models by stage in the life course, it is possible to more easily and efficiently model interactions that may be specific to a single stage, thus avoiding the need to model others that may be irrelevant to it, but relevant to another stage for explaining variations in mobility behaviour. The use of an all-embracing model removes this ability and would therefore require a greater number of model interaction terms, thus greatly increasing the complexity and risk of sparsity within the model.

With all predictor variables being categorical in type, the reference groups are specified as the median value for ordinal covariates and the modal value for nominal covariates. Grouped parameter Wald tests are used in order to assess the contribution of sets of parameters, while holding others fixed, in the fitted multivariate model (e.g. testing the contribution of all of the dummy terms associated with a categorical predictor variable together) (Heeringa *et al.,* 2010). Finally, as before, to test and compare overall model fit, the Akaike Information Criterion (AIC) is used (Agresti, 2007).

7.4 Model results

In order to aid with the interpretation and presentation of the model results, the four life-course models are broken down according to four covariate themes (Table 7.2, socio-demographic characteristics; Table 7.3, labour market characteristics; Table 7.4, housing market characteristics; Table 7.5, subjective/evaluative characteristics) with the overall model fit statistics, constant and dummy indicator variable for year of survey completion given in Table 7.1. To briefly summarise Table 7.1, the model fit statistics suggest that the models are a statistically significant improvement on more simple models, where [a] (Models 1-3) suggests an improvement on the main effects only model and [b] (Model 4) suggests an improvement on the null (empty) model. Moreover, it is clear from the effect size and the associated statistical significance that the inclusion of the survey indicator variable is justified.

Table 7.1. Residential mobility across the broad stages of life-course: Overall model fit statistics, constant and year of survey indicator

Predictor	Model 1: Ages 18-29			Model 2: Ages 30-44			Model 3: Ages 45- 64			Model 4: Ages 65+		
	Beta	S.E.	Odds	Beta	S.E.	Odds	Beta	S.E.	Odds	Beta	S.E.	Odds
Constant	-2.806*	0.185		-4.161*	0.151		-5.081*	0.179	0.006	-5.763*	0.349	
Data set (ref: January 2005)												
January 2006	-0.727*	0.078	0.483	-0.516*	0.057	0.597	-0.451*	0.071	0.637	0.580*	0.117	0.560
January 2007	-1.124*	0.045	0.325	-0.708*	0.038	0.493	-0.228*	0.045	0.796	-0.139	0.163	0.871
Overall model fit statistics												
Null deviance	18557 on 32367 *df*			30252 on 103902 *df*			24187 on 142864 *df*			9060.6 on 69816 *df*		
Residual deviance	17233 on 32315 *df*			28771 on 103854 *df*			23326 on 142821 *df*			8458.9 on 69776 *df*		
Improvement (χ^2)	61.110*, *df* = 13[a]			74.479*, *df* = 9[a]			10.673*, *df* = 3[a]			601.633, *df* = 40[b]		
AIC	17339			28869			23414			8540.9		

*N.B. Model 1 n = 32,368; Model 2 n = 103,903; Model 3 n = 142,865; Model 4 n = 69,817. * indicates parameter is significant at the 95 per cent level. [a] Improvement on main effects only model, [b] improvement on null model.*

7.4.1 Socio-demographic characteristics

Whilst the models (1-4) are themselves broken down according to rather broad life-course stages, each stand-alone model was designed to accommodate the potential effects of age at the smaller intervals found *within* the specific life-course groupings. The results are presented in Table 7.2 and provide evidence that marked differences according to age within these broad stages of life are apparent. For instance, the greatest mobility within the early adulthood stage is associated with those in the 18-19 age group, that conventionally is associated with moves away from the parental home to higher education (Champion, 2005a; Duke-Williams, 2009; Smith, 2009), whilst at the opposite end of the life course, there is significantly greater immobility for those in their 70s compared to individuals in the immediate years following retirement. Of course, beyond the expected increase in immobility for more elderly cohorts, we have come to expect the ages associated with retirement, as with those associated with moves to university, to reflect a greater propensity to move (Evandrou *et al*., 2010), relative to other broad age groups.

Table 7.2. Residential mobility across the broad stages of life-course: Socio-demographic characteristics

Predictor	Model 1: Ages 18-29			Model 2: Ages 30-44			Model 3: Ages 45-64			Model 4: Ages 65+		
	Beta	S.E.	Odds	Beta	S.E.	Odds	Beta	S.E.	Odds	Beta	S.E.	Odds
Age												
Model 1 (ref: 18-19)												
20-24	-0.087	0.203	0.917									
25-29	-0.262*	0.126	0.770									
Model 2 (ref: 30-34)												
35-39				-0.725*	0.040	0.484						
40-44				0.001	0.037	1.001						
Model 3 (ref: 45-49)												
50-54							-0.198*	0.046	0.820			
55-59							0.029	0.042	1.029			
60-64							-0.031	0.042	0.970			
Model 4 (ref: 65-69)												
70-74										-0.461*	0.086	0.631
75-79										-0.089	0.079	0.915
80+										0.086	0.076	1.090
Gender (ref: Female)												
Male	-0.258*	0.082	0.772	-0.001	0.038	0.999	-0.185*	0.044	0.831	-0.189*	0.073	0.827
Ethnic group (ref: White)												
Asian	-0.342*	0.135	0.710	0.181	0.108	1.199	0.227	0.187	1.255	0.118	0.419	1.125
Black	-0.298	0.191	0.743	0.123	0.139	1.131	0.390	0.200	1.477	0.290	0.596	1.337
Other	-0.246	0.142	0.782	0.123	0.121	1.130	0.008	0.176	1.008	-1.993*	1.001	0.136
Marital status (ref: Single)												
Married	0.141	0.072	1.151	-0.139*	0.054	0.870	-0.063	0.077	0.939	0.255	0.160	1.291
Living with partner	0.493*	0.057	1.637	0.326*	0.059	1.385	0.399*	0.097	1.490	0.933*	0.238	2.542
Divorced/ separated	-0.046	0.165	0.955	0.405*	0.062	1.500	0.395*	0.077	1.484	0.301	0.178	1.351
Widowed	-0.351	0.432	0.704	-0.918*	0.359	0.399	0.249*	0.114	1.282	0.496*	0.165	1.643
Gender x Marital status												
Male, married	0.272	0.140	1.313									
Male, living with partner	0.475*	0.110	1.609									
Male, divorced/ separated	0.504	0.395	1.655									
Male, Widowed	-10.699	101.537	0.000									

*N.B. 95% confidence intervals can be calculated as: coefficient (Beta) minus 1.96 * SE (lower boundary) and coefficient (B) plus 1.96 * SE (upper boundary) where SE is the standard error. *indicates parameter is significant at the 95 per cent level.*

As has been shown in previous analysis (Duke-Williams and Stillwell, 2010), a greater likelihood of mobility is observed for women of all stages of the life course apart from those in their 30s and early 40s, when compared to men. The absence of a differential pattern for the 30-44 age groups is an interesting empirical observation. However, given the common theme of family formation and childbearing at this life stage, it is perhaps not so unexpected. After all, the relative plateauing of the female mobility lead can be thought of as linked to the ways in which the social and cultural norms associated with such household and family based phenomena affect mobility behaviours and propensities differently according to gender (Boyle *et al.*, 2001; Magdol, 2002; Boyle *et al.*, 2009).

According to research by Stillwell and Hussain (2010) and Finney and Simpson (2008), almost all ethnic minority groups in Britain (bar certain Asian groups) are characterised by higher rates of residential mobility than the White-British majority. However, this is to a large extent tied to the fact that the White-British majority is, on average, an older population and therefore a seemingly less mobile one (Stillwell and Hussain, 2010). With this in mind, the results in Table 7.2 are useful in showing the remaining effect of the individual's ethnic background once it is sufficiently disentangled from their age/stage in life course. The findings suggest that there are clear patterns in mobility and immobility according to ethnicity which vary through the life course, with particularly interesting results associated with those in early adulthood. Indeed, a greater likelihood of mobility for individuals from the White majority background than those in the non-White groups is revealed, with a particularly strong, and statistically significant, reduction in mobility found for individuals from Asian ethnic backgrounds. However, this relationship reverses as we move through the stages of the life course with those from White ethnic backgrounds in the 30-44, 45-64, and 65+ age groups seen to be less mobile than those in the other ethnic groups. The exception to this rule is for those who are classified as 'Other' in the post-retirement/elderly (aged 65+) stages, where a substantial level of immobility is evident when compared to the White reference group. However, the size of the standard error would suggest that this estimate is open to a particularly wide degree of uncertainty and so should be treated with a good deal of caution.

Moving beyond the typical demographic characteristics uncovers further patterns. For instance, whilst a change in marital status cannot be inferred, given the cross-sectional nature of the ROP data, a measure of marital status does provide a proxy for family formation, cohabitation and the concept of linked moves. That is, for cohabiting couples, decision making is expected to be made collectively, informed by a bargaining process, weighing up the positives/negatives of movement for each partner, which can be particularly complex for duel career households (Abraham *et al.*, 2010). However, a focus on the current marital status of movers and non-movers does reveal some patterns that appear to vary across the life-course. When focussing on those in early adulthood, the sole substantive and statistically significant difference is found between individuals who live with a partner and individuals who class themselves as single, with the former suggesting greater mobility than the latter. Given that by its very nature, living with a partner suggests cohabitation, we can expect a change of residence to be necessary for at least one, and possibly both, of the partners. Moreover, given that these results are for the youngest age group, there is an increased likelihood that the partnership formation is relatively recent and therefore the move could well be a response to this. Applying Wald tests to the model parameters suggests that the interaction of gender and marital status, at least at this stage in the life course, significantly contributes to the multivariate model (Wald $X^2 = 19.0;\ df = 4; p < 0.01$) and, as a result, should be included. With the added gender-marital status interaction term, we can observe that this relationship is further amplified for men; in other words, there is a positive and additional effect for men who live with their partners when compared to women who live with theirs[9]. Therefore, men living with their partners are 2.03 ($\exp^{0.71}$) times more likely to have undertaken a residential move within the last 12 months than the reference group, women who are single[10]. This compares to women living with their partners who are 1.64 times more likely to have moved than single women. Given that cohabitation would necessitate at least one individual changing residence, these findings perhaps suggest a slightly greater propensity for men to do the moving in. Interestingly, this interaction is not found to be significant for any of the later stages in the life course.

The significance of marital status increases somewhat in the more stable family forming/childrearing stages of life (Model 2). Married people, perhaps reflecting this apparent stability, are found to be 0.87 times as likely to move as those who are single. However, those living with their partners experience higher rates of mobility than singles (odds ratio, 1.39). Divorced/separated people also have greater mobility than single people, where, as with family/household formation, the breakdown of

---

[9] The main effect for marital status is interpreted to be the effect for women (the reference category in the gender variable) while the interaction terms reflect the additional effect of being male.

[10] The total effect for men living with a partner in this model is: $-0.258*1 + 0.493*1 + 0.475*(1*1) = 0.71$.

relationships will in most cases also necessitate the move of one, and possibly both, of the individuals (Geist and McManus, 2008; Mulder and Wagner, 2010). Being widowed in this group is also found to have a substantial effect with widowers having far greater levels of immobility when compared to singles; however, again the magnitude of the standard error calls into question the reliability of this estimate. The relationship roughly follows the same pattern in the later stages of the life course, with the exception being the rather unsurprising increase in mobility associated with widowhood, something known to influence a greater likelihood of residential mobility (Chevan, 2005; Evandrou *et al.*, 2010).

7.4.2 Labour market characteristics

The literature suggests that occupational class, household income and educational attainment all play important selective roles in residential mobility (Borjas *et al.*, 1992; Fielding, 1992; 1998; 2007). However, once we control for the additional demographic, socio-economic and lifestyle/behavioural characteristics of the individual, a substantively important relationship between the various occupational, income and qualification groups and residential mobility/immobility is lacking. For instance, whilst the appearance of greater mobility for the intermediate occupational groups in the 18-29 and 30-44 age groups, when compared to the higher level occupations, is statistically significant, the magnitude of the effect is comparatively small, with odds ratios of 1.15 and 1.12 respectively (Table 7.3). Likewise, those with routine and manual occupations between the ages of 45 and 64 also experience a statistically significant, yet seemingly small, increase in mobility when compared to the highest occupational groups (odds ratio, 1.14). Whilst it remains relatively trivial compared to the other characteristics included in the life-course models, the income dimension is perhaps a little more interesting. For instance, for those in early adulthood, there is some evidence of a relatively linear relationship, with greater household income associated with greater mobility. This is a commonly theorised relationship with greater financial resources, indicated by a higher income, leading to improved choice within the housing market as well as an increased ability to cover the financial costs associated with changing residence. Yet for those in the 30-44 and 65+ age groups, we see this admittedly slight association shift into more of a U-shaped relationship with small increases in mobility for those in the lower and upper income groups, when compared to the middling income levels. It should be said that other studies focussed on specific stages in the life course have also suggested the relative irrelevance of household income on residential mobility/immobility patterns; for instance, the study of the mid-life stage by Wulff *et al*. (2010) and the analysis of migration in later life by Evandrou *et al.* (2010).

Generally speaking, these findings contradict the conventional theories which suggest that we should expect residential mobility to increase with occupational class, household income and educational attainment. Yet, whilst this may be so, it is important to keep this study in context. Indeed, the analysis concentrates on variations in the associational patterns of demographic, socio-economic and lifestyle/behavioural characteristics for all movers, as opposed to non-movers, with no differentiation for the distance moved; for which the average across all residential movers modelled here, is assumed to be relatively short given the well-known frictional effect of distance on mobility (Stillwell, 1991). If residential movers were to be modelled separately as short-distance movers, which are typically thought to be more strongly associated with the economics of housing markets, and longer-distance migrants, which are again theorised to be more closely tied to the economics of the labour market, the expectation might be to find the latter group varying considerably, in terms of income and occupation, from those in the former short-distance group (Gordon, 1982). Certainly, the multilevel analysis of distances moved, in Chapter 9, would support this assumption.

Table 7.3. Residential mobility across the broad stages of life-course: Labour-market characteristics

Predictor	Model 1: Ages 18-29			Model 2: Ages 30-44			Model 3: Ages 45-64			Model 4: Ages 65+		
	Beta	S.E.	Odds	Beta	S.E.	Odds	Beta	S.E.	Odds	Beta	S.E.	Odds
Occupation (ref: Higher managerial administrative and professional occupations)												
Not economically active	0.013	0.056	1.013	0.041	0.049	1.042	0.023	0.058	1.023	-0.268	0.191	0.765
Routine and manual occupations	0.078	0.074	1.081	-0.005	0.058	0.995	0.132*	0.064	1.141	-0.141	0.217	0.868
Intermediate occupations	0.141*	0.056	1.152	0.110*	0.045	1.117	0.095	0.060	1.100	-0.164	0.262	0.849
Annual gross household income (ref: £20,000-£29,999)												
Up to £9,999	-0.042	0.069	0.959	0.227*	0.057	1.255	0.049	0.069	1.050	0.164	0.179	1.179
£10,000-£19,999	-0.020	0.059	0.980	0.141*	0.044	1.151	0.001	0.057	1.001	0.169	0.154	1.184
£30,000-£39,999	0.098	0.067	1.103	-0.129*	0.054	0.879	-0.018	0.063	0.982	-0.320*	0.162	0.726
£40,000-£49,999	0.046	0.059	1.047	0.088	0.049	1.092	-0.023	0.058	0.977	0.227	0.148	1.254
£50,000 +	0.154*	0.049	1.166	0.043	0.042	1.044	-0.038	0.051	0.963	0.004	0.117	1.004
Highest qualification (ref: 5 or more GCSEs)												
No formal qualifications	0.288*	0.044	1.334	0.165*	0.035	1.179	0.119*	0.043	1.126	0.072	0.081	1.074
2+ 'A' levels	0.163*	0.050	1.177	0.143*	0.042	1.154	0.149*	0.046	1.161	0.060	0.085	1.062
First degree and higher	-0.134*	0.058	0.874	-0.149*	0.048	0.862	-0.093	0.050	0.911	-0.203*	0.088	0.816

*N.B. *indicates parameter is significant at the 95 per cent level.*

7.4.3 Housing market characteristics

Following Gordon's (1982) suggestions, if the proposed effects of the more labour-market relevant variables are suppressed in these models, due to the greater likelihood of movers being short-distance migrants, we can be forgiven for supposing that the effects of the housing-market orientated characteristics will be amplified. The findings from the models presented in Table 7.4 do, to a large extent, encourage this assertion. Tenure for example, regardless of the stage in the life course, is found to be one of the most substantively important and highly significant characteristics. Across the board, from those in the stages of early adulthood right through to the post-retirement stages of life, there appears to be greater mobility for individuals who rent their accommodation than those who own it, an observation that is by no means new (Rossi and Shlay, 1982; Boyle, 1995; Champion *et al.*, 1998; van Ham and Feijten, 2008). Indeed, homeownership is a particularly inflexible tenure type where financial costs (e.g. high transaction costs, transfer taxes and mortgage costs) and ownership benefits (e.g. security of tenure and protection against eviction) work to reduce regular residential movements. Conversely, private renting is seen to be the most flexible tenure type reflecting lower movement costs, short-term contract durations and, for some, insecurity of tenure, which all work to encourage greater movement propensities (Mulder, 2013).

Consequently, the greatest disparities are observed for private renters and homeowners. Private renters are found to be almost twice as likely to move compared with homeowners in the early stages of adulthood, with the magnitude of the relationship increasing in the 30s and early 40s (3.5 times more likely), and again in the middle-age/pre-retirement stage where the likelihood of moving is almost four times greater for private renters. The extent of the greater likelihood of mobility observed for private renters depreciates somewhat (odds ratio 2.46) in the final stage of post-retirement and old age, but remains strongly predictive of greater mobility. Increased mobility is also observed for those who rent from the council, with the non-significant exception of individuals aged 45-64, and those who rent from housing associations. Interestingly, Wald tests suggest that the mobility rates associated with private renters and council tenants significantly vary according to age within the broad stages of the life-course, but only for those associated with early adulthood and, more specifically for this stage only council tenants (Model 1) (Wald $X^2 = 29.5;\ df = 6; p < 0.01$), and those in the family forming/childrearing stage (Model 2) (Wald $X^2 = 61.4;\ df = 6; p < 0.01$).

Table 7.4. Residential mobility across the broad stages of life-course: Housing-market characteristics

Predictor	Model 1: Ages 18-29			Model 2: Ages 30-44			Model 3: Ages 45-64			Model 4: Ages 65+		
	Beta	S.E.	Odds	Beta	S.E.	Odds	Beta	S.E.	Odds	Beta	S.E.	Odds
Tenure (ref: Own home)												
Council rent	0.518*	0.210	1.678	0.425*	0.197	1.530	-0.161	0.282	0.852	0.298*	0.132	1.347
Housing association rent	0.464	0.259	1.590	0.479*	0.230	1.614	0.553	0.282	1.738	0.617*	0.141	1.853
Private rent	0.669*	0.198	1.952	1.266*	0.188	3.545	1.362*	0.223	3.902	0.900*	0.115	2.460
Type of home (ref: Semi-detached)												
Detached	-0.182*	0.090	0.833	0.437*	0.053	1.549	0.278*	0.068	1.320	0.772*	0.141	2.164
Terraced	0.213*	0.054	1.238	-0.033	0.048	0.967	0.163*	0.064	1.177	0.198	0.159	1.219
Bungalow	0.038	0.142	1.039	0.434*	0.090	1.544	0.995*	0.069	2.705	1.484*	0.122	4.409
Maisonette	0.318*	0.124	1.374	-0.010	0.136	0.990	0.324*	0.162	1.382	0.755*	0.327	2.127
Flat	0.642*	0.063	1.900	0.301*	0.067	1.351	0.708*	0.077	2.030	1.595*	0.143	4.927
OAC Super-group level (ref: Typical traits)												
Blue collar communities	-0.098	0.065	0.907	-0.159*	0.057	0.853	-0.276*	0.075	0.759	-0.308*	0.142	0.735
City living	-0.172	0.096	0.842	-0.346*	0.103	0.707	-0.135	0.110	0.874	-0.121	0.158	0.886
Countryside	0.197*	0.087	1.218	0.055	0.066	1.056	0.103	0.073	1.108	0.064	0.119	1.066
Prospering suburbs	0.191*	0.072	1.210	0.016	0.056	1.016	-0.046	0.069	0.955	-0.268*	0.118	0.765
Constrained by circumstances	-0.163*	0.072	0.849	-0.043	0.065	0.958	-0.066	0.077	0.936	-0.330*	0.133	0.719
Multicultural	-0.306*	0.088	0.737	-0.315*	0.082	0.730	-0.483*	0.109	0.617	-0.737*	0.220	0.478
Age x Tenure												
20–24, council rent	-0.919*	0.251	0.399									
25–29, council rent	0.315	0.166	1.371									
20–24, rent housing association	-0.501	0.334	0.606									
25–29, rent housing association	0.021	0.219	1.021									
20–24, rent private	-0.034	0.229	0.966									
25–29, rent private	0.038	0.146	1.039									
35–39, council rent				0.334*	0.106	1.396						
40–44, council rent				0.078	0.104	1.082						
35–39, rent housing association				0.117	0.133	1.124						
40–44, rent housing association				-0.011	0.126	0.989						
35–39, rent private				0.591*	0.081	1.806						
40–44, rent private				-0.103	0.078	0.902						

*N.B. *indicates parameter is significant at the 95 per cent level.*

Given the inclusion of the interaction terms, the main effects of tenure for those in the 18-29 and 30-44 groups should be interpreted as the effects for individuals in the reference age brackets, 18-19 in Model 1 and 30-34 in Model 2. With this being the case, it should be noted that those who record themselves as homeowners at the age of 18-19 are quite probably living in their parents (owned) home. Looking at these finer age group variations, council tenants aged 18-19 are estimated to be 1.68 times more likely to have moved than the reference group, homeowners aged 18-19, whereas council tenants aged 20-24 actually buck the general trend with the likelihood of having moved estimated to be 0.61 times that of the reference group. Conversely, council tenants in the 30-44 stage are found to have the same directional associations, with greater mobility found when compared to homeowners, although the magnitude of the relationship is significantly weaker for those aged 35-39 who are shown to be only 1.13 times more likely to have moved than those in the reference group, homeowners aged 30-34. This pattern for individuals aged 30-34 is also significant for private renters where again, *ceteris paribus*, we see them being slightly less likely to have moved than private renters aged 30-34, when compared to homeowners of the same age. In terms of the bigger picture, the greater mobility for council tenants is particularly interesting as they have traditionally been associated with lower rates of mobility, although more specifically at the inter-regional level, partly linked to the rather rigid allocation system employed in Britain (Hughes and McCormick, 2000). However, such structural restrictions are greatly reduced for localised moves and therefore, given the likelihood that most of the recorded moves will be short distance in nature, the higher mobility associated with council tenants, in comparison to homeowners, is not as unexpected as perhaps first thought.

Continuing the housing related trend, house type is also found to be highly influential for patterns of mobility/immobility, although the type-specific relationships vary depending on the stage of life course. For the youngest stage (early adulthood), mobility is significantly higher for those in flats (odds ratio, 1.90), maisonettes (odds ratio, 1.37) and terraced housing (odds ratio, 1.24) and significantly lower for those in detached housing (odds ratio, 0.83), when compared to those in semi-detached housing. Given that we are talking about people at the start of their housing/occupational careers, it is perhaps unsurprising that individuals in the housing types we generally associate with lower transaction costs reflect a greater likelihood of moving. The picture becomes a little more mixed in the middle stages of life (Models 2 and 3) with individuals from detached accommodation now reflecting, on average, a greater propensity for residential mobility than those in semi-detached housing. This relative increase in mobility associated with detached housing, and the relative decrease in the mobility witnessed for those in flats when compared to semi-detached accommodation, is likely to reflect the importance of family formation, especially for those aged 30-45, and the necessary housing adjustments that changes in family composition are known to entail. Indeed, whilst there is no direct measure of dependent children in the household, pregnancy and/or the birth of a child (or children), often synonymous with this stage in the life course, is known to greatly alter housing preferences, with issues of space, quality, safety and security growing in significance (Mulder, 2013). For those in the final stages of the life-course, the substantive importance of housing-type increases still further with rather pronounced rates of mobility associated with bungalows (odds ratio, 4.41) and flats, the latter suggestive of a mobility propensity almost five times greater than that of the reference category, semi-detached. Indeed, whilst change to family composition, through family formation, can be thought to influence the increased mobility rates observed for the larger accommodation types, the increase in the substantive importance of the smaller accommodation types, for this stage in the life course, can also be understood to reflect such factors. For instance, it might be assumed that the housing needs for retired and elderly individuals, in terms of space, are somewhat reduced when compared to individuals in earlier stages of life. Moreover, given the onset of old age and the physical problems that this can bring, it is of no surprise that a rather substantial shift towards single-level accommodation types is apparent.

The effect of the individual’s current neighbourhood type can, to a certain extent, be seen to further condition the likelihood of undertaking a residential move. All things being equal, and irrespective of stage in the life course, individuals living in multicultural areas are found, on average, to have the lowest levels of mobility. Similarly, individuals living in ‘blue collar communities’, excluding those in early adulthood, can also be seen to have significantly reduced rates of mobility, when compared to individuals living in areas classified as ‘typical traits’. However, aside from these rather consistent findings, the remaining effects associated with neighbourhood type, as observed in previous studies (Kearns and Parkes, 2003; van Ham and Clark, 2009; Rabe and Taylor, 2010), are fairly trivial when compared to the individual’s demographic, socio-economic and behavioural/lifestyle characteristics. Yet it is possible that the technical and analytical limitations associated with the inclusion of neighbourhood type in the manner presented here, as a series of fixed effects dummy term variables within a single-level modelling framework, are working to obscure substantively interesting neighbourhood characteristic/context influences on residential mobility/immobility. With this in mind, debates about the substantive relevance of contextual influences on mobility decision-making and mobility outcomes are discussed in particular detail in the following analytical chapters (Chapters 8 and 9), where more suitable multilevel methodologies are employed to test contextual influences operating at different levels of aggregation and for different mobility outcomes.

7.4.4 Subjective/evaluative characteristics

Finally, we are left with the seemingly more nuanced characteristics of movers and non-movers, namely those associated with greater conjecture and subjectivity. Individuals' moving desires, expectations and plans are of clear importance to the study of residential mobility and immobility. However, from an empirical perspective, the focus on such factors remains surprisingly lacklustre. That said, research in this area is increasing, with key contributions focussing on the interrelationship between pre-move desires and subsequent moving behaviour (Lu, 1998; Kley and Mulder, 2010; Kley, 2011; Coulter *et al.*, 2011; 2012). Unfortunately, the nature of the ROP makes it impossible to study the relationship between pre-move desires and subsequent mobility. However, in spite of the lack of longitudinal data, we are able to uncover whether individuals who have moved within the last 12 months are more/less likely to be planning a further move within the next 12 months.

Table 7.5. Residential mobility across the broad stages of life-course: Subjective/evaluative characteristics

Predictor	Model 1: Ages 18-29			Model 2: Ages 30-44			Model 3: Ages 45-64			Model 4: Ages 65+		
	Beta	SE	Odds	Beta	SE	Odds	Beta	SE	Odds	Beta	SE	Odds
Plan to move in next 12 months (ref: No)												
Yes	-0.211*	0.052	0.810	-0.109*	0.055	0.896	0.043	0.081	1.044	-0.130	0.207	0.878
Like your neighbourhood (ref: No)												
Yes	0.745*	0.144	2.107	1.019*	0.136	2.769	0.792*	0.152	2.208	0.745*	0.220	2.106
Tenure x Like your neighbourhood												
Council rent, likes neighbourhood	-0.417*	0.193	0.659	-0.502*	0.205	0.605	0.265	0.288	1.303			
Rent housing association, likes neighbourhood	-0.458*	0.233	0.633	-0.427	0.241	0.652	-0.304	0.294	0.738			
Rent private, likes neighbourhood	-0.554*	0.176	0.574	-0.656*	0.192	0.519	-0.598*	0.229	0.550			

*N.B. *indicates parameter is significant at the 95 per cent level.*

Looking at the results from the life-course models in Table 7.5, the directional relationships, aside from those in the 45-64 stage, appear to suggest that individuals are less likely to be planning a future move if they have already recently moved. This observation is particularly significant, and statistically more stable, for those in the early adulthood phase, where individuals planning to move are on average, 0.81 times as likely to have already moved in the 12 months prior to the survey. At first sight, this appears to contradict the cumulative inertia hypothesis, wherein individuals with the shortest durations of residence are thought to be the most likely to move again, a theory that has been important in explaining the high correlation between out-migration and in-migration rates at the aggregate levels (Cordey-Hayes and Gleave, 1974), and that is the subject of particular interest in Chapter 8.

However, as is made clear in Chapter 8, micro-level studies, with their notable inclusion of important covariates such as age, have shown that the relationship between residence duration and the likelihood of considering a future move does not follow a simple monotonic relationship, that is, with probabilities of moving decreasing as duration increases. For instance, micro-level analysis by Gordon and Molho (1995: 1970) suggests that the likelihood of considering a move is lower for those with the shortest durations (e.g. within the first 12 months), given that they are in a residential environment that only a year or so earlier suited their residential preferences and encouraged their movement to it. Again, the results in Chapter 8 add further weight to this argument. Consequently, it could be argued that the residential moves already performed by individuals, particularly in the early adulthood stage, are to a certain extent successful in fulfilling the factors that motivated their move in the first place. At this stage in the life course for instance, interrelated events such as leaving the parental home, going to university, starting a career and forming relationships resulting in cohabitation, are all factors that stimulate residential mobility. And it follows, therefore, that they are all factors that can be satisfied, to varying degrees, by residential mobility. Additionally, given that a residential migrant would, by definition, have lived at the address for fewer than 12 months, the financial requirements of a further move, within such a short timeframe, would undeniably weigh heavy on any plan for a further move. Of course, planning to move is a more definitive statement than simply desiring a move and would suggest that more serious practical considerations of the residential move, such as the financial implications, had been made (Lu, 1998; Coulter *et al.*, 2011).

The importance of the neighbourhood, in terms of subjective measures of satisfaction, has become an increasingly interesting area within the residential mobility literature (Clark and Ledwith, 2006; Feijten and van Ham, 2009; Permentier *et al.*, 2009; Hedman, 2011). The analyses presented in this literature suggest that, aside from household needs and preferences, (dis)satisfaction with the wider neighbourhood is fundamental in motivating a decision to move/stay, with greater neighbourhood satisfaction tied closely to a greater likelihood to remain in place. However, the processes behind neighbourhood satisfaction are clearly complex and dynamic in nature; with variations likely to be driven by differences operating at the level of the individual as well as the household (Parks *et al.*, 2001). Therefore, it is perhaps not surprising that the relationship between neighbourhood satisfaction and residential mobility is found to vary significantly according to tenure type, although only for those aged 18-29 (Wald $X^2 = 10.2;\ df = 3; p < 0.05$), 30-44 (Wald $X^2 = 12.6;\ df = 3; p < 0.01$) and 45-64 (Wald $X^2 = 10.8;\ df = 3; p < 0.05$). Overall, greater neighbourhood satisfaction is found to be consistently and rather strongly associated with residential mobility. Across the various stages of the life course, people who are satisfied with their neighbourhood are more likely to have recently moved than not. However, allowing for this relationship to vary according to tenure uncovers further interesting findings. All things being equal, for the relationship between neighbourhood satisfaction and residential mobility, there is a positive additional effect associated with homeowners and conversely a negative additional effect for renters (be they council, housing association or private). In other words, the higher level of neighbourhood satisfaction associated with residential movers is lessened somewhat if their tenure type is renter, as opposed to homeowner. Such findings are perhaps to be expected given that movers who own their home are more likely to have invested for the long-term, and subsequently, one would imagine, are more likely to have chosen an area/neighbourhood that fits their housing, lifestyle and consumption desires more comfortably. After all, the difference is particularly pronounced when comparing homeowners to private renters, the latter being the tenure group most closely associated with short-term residential durations (Bailey and Livingston, 2005).

7.5 Summary and conclusions

This initial substantive analytical chapter aimed to explore how the complex and interlinked micro-level characteristics of movers and non-movers vary according to broad life-course stages. Separating the life course into four major phases – ages 18-29, 30-44, 45-64 and 65+ – has uncovered some interesting patterns, some of which varied across the life-course (for instance, the effects of ethnic background) and others of which remained constant throughout (for instance, the effects of neighbourhood type). One important conclusion to be drawn from the life-course models is the relative unimportance of what can be thought of as the labour market characteristics of individuals. However, as mentioned above, it is important to think carefully about what is being measured here. The analysis presented explores the variations between movers and non-movers, measured explicitly as two homogenous groups. In reality, these broad categorisations are problematic in that they do not allow for the representation of what will be substantial within-group variation. For instance, various theoretical and empirical studies have detailed the motivational difference between long-distance migrants and short-distance movers; with short-distance mobility theorised to be driven by the economics of the housing market and long-distance migration being thought to be motivated by the economics of the labour market (e.g. Gordon, 1982). Thus, where these models are, through the frictional effect of distance on mobility (Stillwell, 1991), more accurately measuring variations between 'short-distance movers' and 'stayers', it is not surprising that a relative marginality is found for individual/household labour market characteristics while the residential factors suggest great substantive relevance. This topic is taken up as the focus of Chapter 9, which models individual and place variations in origin-destination distance moved.

Accepting that these findings could be influenced by aforementioned issues, a focus on the more subjective behaviours/characteristics of movers and non-movers has uncovered results worthy of further discussion. Shifting to the dynamic role of neighbourhood satisfaction for mobility and immobility, some rather interesting (and to the knowledge of the author) previously unobserved findings are found. Indeed, the role of neighbourhood satisfaction is found to be a complex one, wherein it would appear to be linked rather strongly to the individual's housing tenure. Primarily, across the various stages of the life course, people who are satisfied with their neighbourhood are more likely to have recently moved than remained *in situ*. Yet, all things being equal, a positive additional effect is associated with homeowners and a negative additional effect for renters regardless of type. In other words, the higher level of neighbourhood satisfaction associated with residential movers is lessened somewhat if their tenure type is renter, be it council, housing association, or private, as opposed to homeowner. It is thus suggested that movers who own their home are, for varying reasons, more likely to have chosen a neighbourhood that more closely fits their housing, lifestyle and consumption desires. Yet, in order to get a better handle on the causal nature of such phenomena, a useful direction for future research would be to explore these complex inter-relationships over time through the use of appropriate longitudinal data, with measures of pre- and post-move characteristics, and methods.

Future plans to move are found to be negatively associated with mobility, especially for those in their early adulthood. It is suggested therefore that individuals who undertook a residential move within the 12 months prior to the survey were largely successful in fulfilling the factors that motivated their move in the first place, be it university, cohabitation, lifestyle or career driven. However, beyond this, it is also highly likely that very recent movers are comparatively less likely to plan a further move given the various forms of additional investment (in terms of time, emotion, and finance) that would be required, an issue that would be likely to increase if we were to reduce the timeframe between the last move and the proposed future move still further. It was also suggested that the definition of planning a move was more definitive as a statement than, for instance, desiring a move would be. As a result, it is thought highly likely that individuals who are planning to move within the next 12 months have taken these more practical, investment related considerations into account. However, with the availability of duration data and measures of future mobility propensities, the ROP provides us with the opportunity to explore the duration-of-residence effects in far more detail, incorporating theories of residential satisfaction and duration dependence; indeed it is this that is the focus of the next chapter.

Whilst this initial substantive analysis has provided some useful preliminary results, the focus in the following analytical chapters is on exploring specific areas of relatively under explored research, areas for which the ROP can be considered to be particularly well suited. Indeed, as has been noted in the previous chapters, traditional survey sources have limited the simultaneous analysis of individual *and* place variations in residential mobility. With that in mind, Chapters 8 and 9 incorporate measures of both micro and macro-contextual characteristics thought important for informing first the decision to move (Chapter 8), and second, once the decision to move has been made, the distance moved (Chapter 9).

Chapter 8
Modelling the duration-of-residence and plans for future residential mobility: A multilevel analysis

8.1 Introduction

As was detailed in Chapter 3, the ROP holds a number of unique attributes that make it particularly attractive for novel analyses of individual and place variations in population mobility. Consequently, this chapter seeks to exploit some of these attributes for the analysis of a topic that has been of great interest and debate for many decades; that is, the functional form of the relationship between duration-of-residence and mobility propensity. Indeed, whilst the concept of ‘cumulative inertia’ has long been a mainstay in population mobility research literatures, there is a scarcity of empirical evidence supporting the existence of such a relationship. The theory suggests that as the length of residence increases, stronger social and economic ties to an area are developed, a process that cumulatively reduces the likelihood of a future move. Equally plausible contradictory theories also exist, most notably in the discussions of cumulative stress (Speare *et al.,* 1975; Huff and Clark, 1978; Clark *et al.*, 1979), where longer residential durations are thought to promote an increasing residential dissatisfaction as our needs, preferences and desires evolve with the movement through, and experience of, different stages and events pertaining to the life course (see Chapter 7).

Therefore, by utilising the relative advantages of the ROP, this chapter presents a series of hierarchical multilevel models that aim to explore the relationship between the probability of planning a residential move within the next 12 months and the length of stay (duration) at the current address. The chapter builds on previous analyses of this kind by incorporating both individual and area-level characteristics into a multilevel statistical framework. Indeed, as was detailed in Chapter 4, multilevel modelling makes it possible to correctly and efficiently estimate and explore potentially important cross-level interactions (between variable measured at different levels) as well as complex higher-level (contextual) heterogeneity in the duration-of-residence effect.

Given the theory of cumulative inertia, a cross-level interaction that is of particular interest relates to whether individual residential durations vary according to different levels of neighbourhood population (in)stability, measured via a population churn statistic. The existing literature suggests that variations in neighbourhood population (in)stability will condition the duration relationship differently due to changing opportunities for the development of strong area-based social attachments. It is also argued that the housing stock and demographic profile that generally characterises neighbourhoods of high population churn are also influential in terms of attracting individuals who need/desire residential flexibility, and thus relatively short residential durations. However, beyond the potential importance of measurable individual/household and neighbourhood dynamics, it is also argued that the more subtle, indirect, and harder to measure effects of differential neighbourhood socialisation, relational networks, institutional resources and routines and other social and economic place-based processes and practices will greatly condition the effects of duration and the decision/ability to be planning a residential move. Furthermore, it can be expected that these more subtle place-based neighbourhood influences require a certain amount of exposure time (residential duration) before they are able to exert any appreciable effects on variations in individual propensities for residential movement (Tienda, 1991; Hedman, 2011; van Ham *et al.*, 2014).

Consequently, by adopting a multilevel modelling framework employing a random intercepts and random coefficients, this analysis seeks to incorporate the aforementioned complexities by allowing for the potential of different duration-of-residence effects for different neighbourhoods and districts of England and Wales. The strategy also provides an opportunity to observe whether longer durations are associated with greater variability in the propensity to change residence than shorter durations, where perhaps the appreciable effects of neighbourhood externalities (positive and negative) have not had sufficient time to form. The chapter begins with a detailed review of the relevant duration-of-residence literature before describing the sample, measures, analysis and modelling strategy used to explore the effects of residential duration on plans for future residential mobility. Finally, the results of particular substantive value are interpreted and summarised before conclusions are drawn and suggestions for future research are made.

8.2 Duration-of-residence and residential mobility: Theories of cumulative inertia and cumulative stress

The functional form of the relationship between the length of stay (duration) at a residence and individual probabilities of movement from one residential location to another has long been the subject of interest and debate in residential mobility research. Perhaps the oldest and most commonly theorised relationship is that of *cumulative inertia*, where one's propensity to change residence is thought to decrease monotonically as the length of stay at the current residence increases. As McGinnis (1968: 716), an early proponent of such thinking, explains "*not all elements in state s_1 at time t are governed by a single law of mobility. In particular, those who have been there longer have a greater probability of remaining than do relative newcomers*". At the micro-level, the theory implies that as residential duration increases, stronger social and economic ties to the place of residence (household and/or area) are developed, a sort of socio-economic integration that increases the 'costs', and therefore cumulatively reduces probabilities of movement away. As was mentioned in the last chapter, at an aggregate level, the theory of cumulative inertia has been attractive due to its relevance as a possible explanatory factor for the high correlation found between measured migrant inflows and outflows for spatial units at various aggregate scales (Cordey-Hayes and Gleave, 1974). However, empirical demonstrations of this supposed functional form are noticeably lacking in analyses at the micro-level, though early examples by Land (1969) and Morrison (1971) are noted.

A somewhat lesser known though equally plausible counter-theory for the relationship between residential duration and future residential movement is that of *cumulative stress*, wherein it is expected that individuals/households become progressively dissatisfied with their housing and/or areas their needs, aspirations and desires change over time (Speare *et al.*, 1975; Huff and Clark, 1978; Clark *et al.*, 1979). If, under normal circumstances, we can expect an individual/household who is new to a residence to have selected the accommodation and area which, to at least some extent, fulfils their current housing, lifestyle and consumption preferences, under the assumption of cumulative stress, we should expect preferences to continue to evolve such that the longer the duration at the residence, the greater the mismatch between the desired and current residential characteristics. If indeed this is the case, the functional form of the relationship between the length of stay at a residence and the probability of future residential movement should reflect a monotonic increase.

Yet whilst theories of cumulative inertia and cumulative stress propose quite strict, and indeed contradictory, linear relationships between duration of stay and propensities for future mobility, an important contribution by Gordon and Molho (1995) suggests a more complex, unified, and intuitively more realistic, functional form. The appeal of Gordon and Molho's (1995) theoretical model lies in its ability to succinctly synthesise both concepts of cumulative stress and inertia, that is, "*recognising that, although longer residence in an area may generally increase the 'costs' of any subsequent move, the passage of time will always lead a minority (at least) of the population to reevaluate their original preferences in favour of some other area, job, or house*" (p. 1972). This theorisation, which is supported by empirical findings from their analysis of the 1983 General Household Survey, implies a nonlinear associational relationship that is characterised by a rise to an initial peak followed by a gradual tailing-off in movement probabilities as duration increases. According to Gordon and Molho (1995), a key factor behind their empirical demonstration of a non-monotonic duration relationship is the development and incorporation of suitably rigorous controls for important additional sources of heterogeneity in the response variable. That is, controls designed with the purpose of helping to separate out independent duration-of-residence effects[11]. Consequently, whilst in studies of this type we can never be confident that all relevant sources of heterogeneity have been covered in the final model (Feijten and van Ham, 2009); it is thought important to remind the reader of what are said to be some of the most important characteristics, behaviours and phenomena that influence the probability to change residence.

Indeed, as has been mentioned in the preceding chapters, critical to our understanding of the complexity found in patterns of residential mobility are the known influences pertaining to life-course transitions and the associated shifts in

---

11 As in early studies that incorporated few additional covariates, a negative monotonic duration-of-residence relationship was observed by Gordon and Molho (1995) when a simple bivariate analysis was run. However, once controls were included, the relationship changed to the initial rise then falling-off described above.

household structure, housing tenure, and income, occupational and educational attainment (Fielding, 2007; Boyle *et al.*, 2008; Mulder and Wagner, 2010); all of which have been observed to be of great importance in terms of shaping individuals' residential preferences and indeed their ability to act upon them. As the UK population census has shown for many decades (Figure 2.1), mobility rates are highest for those who are in the 18-25 age bracket, with moves in this group often motivated by the pursuit of early career educational and occupational opportunities, before a relatively sharp decline, commonly associated with career stability, family formation and child rearing, sets in. Following this, mobility propensities are observed to reflect a lower rate with some recoveries thought to be tied to the transition from parenthood to 'empty nesting', as well as the transition to retirement and the exit from the labour market. Of course, whilst this generalisation of the life course into certain follow-on 'stages' can be useful when attempting to interpret the well-known and general patterns, characteristics and trends of residential mobility, it is important to remember that certain disruptions, expected or otherwise, can also emerge. These disruptions can alter preferences, in some cases exacerbating the residential mismatch, and greatly increase/decrease the probability of a mobility event occurring. Personal events such as unemployment (Clark and Davies Withers, 1999; Böheim and Taylor, 2002; Fielding, 2012), pregnancy and the birth of children (Kulu and Steel, 2013), union dissolution (Boyle *et al.*, 2008; Mulder and Wagner, 2010), and widowhood (Chevan, 2005; Evandrou *et al.*, 2010) are clear cases in point.

Yet, implicit in the notions of cumulative stress as well as inertia is the idea that our evaluations of residential satisfaction extend beyond the individual and household. Whilst mixed empirical results have led to much debate about the relative importance of the neighbourhood context on residential mobility (see Kearns and Parkes, 2005; Clark and Ledwith, 2006; Rabe and Taylor, 2010), an increasing volume of work does suggest that certain neighbourhood characteristics are relevant (van Ham and Clark, 2009; Feijten and van Ham, 2009; Hedman *et al.*, 2011). The socioeconomic status of the neighbourhood is thought to be important in terms of general social cohesion (Taylor *et al.*, 2010; Sturgis *et al.*, 2013); however, more specifically in terms of mobility intentions, social norms and discourses surrounding social status and neighbourhood desirability are said to motivate individuals to leave neighbourhoods where their neighbours are *assumed* to be reflecting of lower socioeconomic status than themselves (Harris, 1999). Indeed, empirical analysis of UK census data by Bailey and Livingston (2008) does suggest a process of neighbourhood sorting tied to such individual and neighbourhood differentials, showing individuals in higher socio-economic brackets to be more likely to move away from areas of increasing deprivation. Moreover, in a similar way to the socioeconomic status of neighbourhoods, the relevance of the ethnic heterogeneity of the neighbourhood population on mobility decisions has also been of central and recurrent interest. Again, whilst the literature reveals mixed empirical findings, some analysts have suggested that greater ethnic heterogeneity be associated with greater residential dissatisfaction, and resultant mobility, amongst the majority population; the mainly US centred ‘white flight’ hypothesis being the common theme here (Ellen, 2000; Crowder, 2000).

Conversely, for minority groups, the opposite relationship between ethnic heterogeneity and residential mobility has been suggested, with more diverse neighbourhoods being more attractive. As Bailey and Livingston (2005: 17) explain “*some minority groups have a strong propensity to co-locate, for positive reasons (the importance of extended-family ties or access to particular amenities) as well as negative ones (a defensive reaction to racism or more limited options in housing or labour market terms)*”. Yet, further studies suggest that once important confounding factors are controlled for, most noticeably that of the socioeconomic status of the neighbourhood, the effects of neighbourhood ethnic composition on mobility intentions, as well as more general features such as social cohesion and trust (Sturgis *et al.*, 2011; Sturgis *et al*., 2013), should disappear or be very minimal at most (Harris, 1999). Yet whilst both socioeconomic status and ethnic heterogeneity have featured prominently in the literature on wider residential context and mobility behaviour, a third dimension, the degree of population (in)stability in the neighbourhood, has also been noted as an important dynamic and, given the focus of this chapter, is deemed deserving of particular attention.

Indeed, population (in)stability, the intensity of movement into and out of an area, can be expected to have a great deal of influence on a multitude of individual/household and neighbourhood dynamics relevant to mobility behaviour. Much of the literature exploring the influence of wider neighbourhood population instability has suggested it be associated with broadly negative residential externalities, reflected, for instance, in greater fears and occurrences of violence and crime (Taylor and Covington, 1993; Sampson *et al.,* 1997), and generally lower residential attractiveness (Andersson and Bråmå, 2004). However, of particular interest to the investigation of duration-of-residence effects, the degree of neighbourhood population (in)stability can also be thought to influence the opportunity and potential for residents to form meaningful community and place-based social interactions and attachments. Of course, a central tenant to the theory of duration dependence is the notion that with longer residential durations stronger social ties and attachments are formed, increasing the 'costs' and reducing the probability of a residential movement away. Thus, whilst stable residential populations may be conducive to the formation of stronger social ties and networks, high population instability in the neighbourhood, where a large proportion of neighbours tend to come in and move out in rapid succession, can be expected to disrupt their formation and maintenance (Hedman, 2011; Hedman*et al.*, 2011).

By and large, in the UK context, the highest levels of population instability are found in the more dynamic urban areas, for instance areas of city living and/or high student populations (Dennett and Stillwell, 2008). Such neighbourhoods are characterised by high proportions of privately rented dwellings and, closely linked to this, young and typically single adults (Bailey and Livingston, 2007; van Ham and Clark, 2009). Consequently, whilst areas of high population instability may be detrimental to the forging of stronger place-based social ties, given the housing stock and demographic profile of these areas, it is equally likely that those living in and moving to them are less concerned by such matters given their assumed desire/need for residential flexibility (short residential durations). Indeed, if moves into areas of high instability are generally made with the pre-understanding that residency will be short-term, it would be fair to expect the residents of such areas, principally those with short durations, to have particularly inflated probabilities for further movement, when compared to similar residents in areas of greater population stability. Similar reasoning underpins the theory behind the strong correlation between rates of in-migration and out-migration at aggregate levels noted above. Figure 8.1 is a map of population churn (per 1,000) (described in Section 8.3 below) for England and Wales as well as three major metropolitan areas, with the patterns reflecting those just discussed (e.g. high churn in dynamic urban areas, student areas and areas of city living, though London, given its unique urban geography, provides a rather more mixed picture).

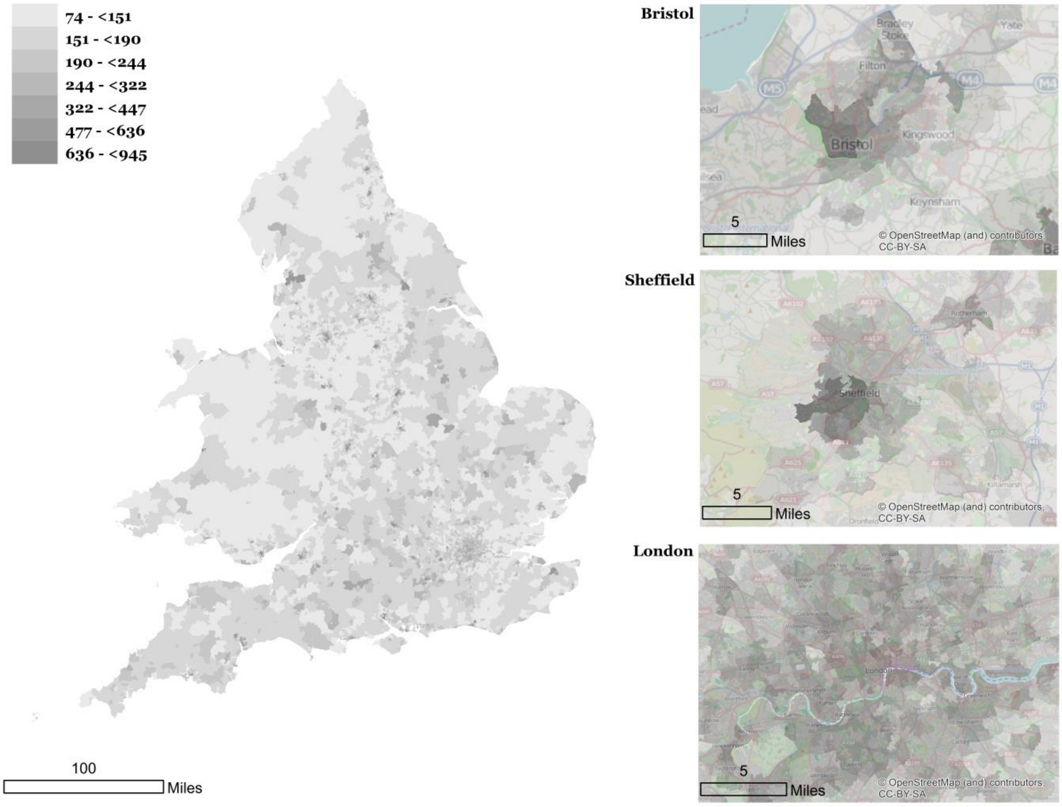


Figure 8.1. Map of population churn (per 1,000) by MSOA in England and Wales. Source: 2001 Census SMS.

As a final consideration, it should be noted that residential duration may also be important in mediating the extent to which macro-level influences can inform micro-level mobility behaviours. Indeed, whilst the composition and characteristics of the neighbourhood population are argued to be relevant for evaluations of one's wider residential milieu and associated mobility outcomes, the extent to which other more subtle, and harder to measure, neighbourhood influences are realised may depend on the resident remaining in place for a critical period of (exposure) time (Tienda, 1991; Hedman, 2011). As Sampson *et al.*'s (2002) seminal review of the neighbourhood effects literature has suggested, a plethora of additional complex and multifaceted factors associated with social processes and institutional mechanisms can be expected to contribute to our evaluation of and commitment to places. Certainly, between-neighbourhood variations in the more subtle dynamics linked to the opportunity to develop, for instance, strong social ties, familiarity and interactions, mutual trust and collective efficacy, and an attachment to local institutional resources and routines (e.g. organised social and recreational activities) should be expected. However, it may take a certain duration in the locality (exposure) for such 'effects' to develop and thus have influence over individual behavioural patterns.

Thus, given our focus on duration-of-residence effects, we can perhaps predict that individual duration-dependence will vary in strength from neighbourhood to neighbourhood, and possibly district to district, as a result of the development of these more subtle neighbourhood influences and externalities. Moreover, given that their functioning is thought to be closely tied to exposure times, we could additionally predict that the importance of these subtle place-based dynamics, be they positive or negative in their influence, will grow as duration increases. Indeed, whilst the relevance of wider residential environment effects have been questioned due to previous empirical findings, it may well be the case that analysts have not sufficiently accounted for the importance of critical periods of residential exposure within their work.

8.3 Data and Measures

The individual-level data used are a subsample drawn from the pooled analytical sample used in the previous two chapters. Indeed, the ROP has a number of advantages over alternative data sources for the analysis of duration dependence and residential mobility. First, for the years covered by the pooled sample, the ROP included the questions *"When did you move to this address?"* and *"Are you planning to move (in the next 12 months)?".* From these questions it is possible to both calculate the duration-of-residence (by year) for respondents as well as their propensity to move in the 12 months following survey completion. Second, with the data being pooled across the three-year period for which the relevant questions are asked, the ROP produces a large and spatially extensive (non-clustered) sample of individual household respondents, and is thus favourable for detailed geographical analysis. Third, each individual respondent has a full unit postcode address identifier (Raper *et al.*, 1992), allowing for a great deal of flexibility in the decisions of how to operationalise the wider residential contexts discussed above.

The binary dependent variable is whether the individual (household respondent) is planning a move in the next 12 months; yes (1) or no (0). Whilst the cross-sectional nature of the data again prevents any examination of actual mobility behaviour, planning a move would suggest that serious practical considerations for a residential move, such as financial costs and likelihood of success, had been made and can thus be expected to correlate closely with actual movement outcomes (Lu, 1998; Coulter *et al*., 2011). In keeping with similar studies of duration effects, the independent variable of interest, duration-of-residence, is measured from the time of arrival at the current residence (for adult movers) or the time immediately after reaching adulthood (where movement intentions are assumed to be more independent), here defined as 18 years of age (Gordon and Molho, 1995). The range of durations is limited to 20 years (Table 8.1) so as to avoid problems with sparsity in the sample and, whilst the year of arrival at the current residence should be a generally memorable characteristic and thus well recorded, reduce the potential for recall bias which can be expected to be particularly severe for those with very long durations.

Table 8.1. Sample size and percentage of total sample size for each year of duration recorded

Duration (years)	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
N	19,819	17,277	18,389	18,908	18,032	17,293	16,066	13,278	11,045	9,420
%	8.8	7.7	8.2	8.4	8	7.7	7.2	5.9	4.9	4.2
Duration (years)	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**
N	9,379	7,041	6,811	5,974	6,140	5,617	6,084	5,708	6,049	5,834
%	4.2	3.1	3	2.7	2.7	2.5	2.7	2.5	2.7	2.6

N.B. Refers to Sample 4 in Table 6.7 - n = 224,164, England and Wales, January 2005-07.

Drawing on the literature reviewed above, a battery of micro-level (individual and household) covariates are collected in an attempt to control for important sources of heterogeneity in the dependent variable. These controls include: age, gender, ethnicity, occupation, gross household income, educational attainment, housing tenure and marital status. Again, an additional control is included to account for the fact that the analytical sample is made up of three separate ROP cross-sections (January 2005-07).

Moreover, given that individual evaluations of residential satisfaction are assumed to extend beyond the individual and household, a multilevel model is thought most appropriate, with individuals being nested within neighbourhoods that are themselves nested within higher level spatial units; the technical and substantive benefits of this approach are outlined in detail in Chapter 4, but are again briefly noted below. Drawing on the spatial coverage and detail of the ROP, the local neighbourhood is defined using the census Middle Super Output Area (MSOA) geography for England and Wales. In England and Wales, these geographic units are designed to be stable over time, similar in terms of their constituent population size ($n \approx 3{,}000$ households), and take into account conceptual definitions as well as physical features (e.g. major roads and topological features) in the construction of their boundaries (Martin, 2002a; 2002b).

An additional benefit of the MSOA geography relates to the fact that it nests perfectly into the Local Authority Districts (LADs), the level of local government operation and resource allocation. LADs can themselves be aggregated into functional geographical city regions. City regions are spatial units designed to provide a manageable set of regions based on metropolitan cores and their 'tributary' hinterland areas (metro rest, near, coast and country areas) (Stillwell *et al.*, 2000; 2001). They are particularly useful for mobility analysis as they provide useful approximations for the urban hierarchy and wider spatial economic system in England and Wales. A map of the 33 macro-geographical regions, based on the major metropolitan centres of England and Wales (Birmingham, Bristol, Cardiff, Leeds, Liverpool, London, Manchester, Newcastle and Sheffield), is given in Figure 8.2.

Figure 8.2. Map of the city regions and their component parts in England and Wales. Source: Stillwell *et al.*, 2000.

The nesting of MSOAs into these higher level geographies is important for a number of reasons. First, as hinted at in the choice of higher level units, we can expect there to be certain sources of variation in mobility propensities that operate at levels beyond the neighbourhood such as macro regional and district level variations in property markets, labour markets, wealth, urbanicity and the environment (Champion *et al.*, 1998; Sampson *et al*., 2002; Bailey and Livingston, 2008; Fielding, 2012). Second, as Brunton-Smith and Sturgis (2011: 342) have pointed out, the reliance on fixed MSOA boundaries means our definition of the neighbourhood is somewhat arbitrary; particularly for those who live on the edges of areas and are therefore highly likely to see their 'neighbourhood' incorporate influences and characteristics of adjacent areas. By nesting MSOAs into the higher units in a multilevel model, we acknowledge this conceptual dependency and clustering of not only individuals, but also of nearby areas; of course there are technical benefits to this too given the standard assumption of the independence of observations (both micro and areal) in regression modelling (Chapter 4).

Unfortunately, MSOAs are not available for Scotland and whilst intermediate zones (equivalent to MSOAs) were designed in 2005, the lack of availability/consistency of relevant neighbourhood characteristics measured at this level in Scotland means that this analysis is restricted to England and Wales only. The neighbourhood characteristics used are informed by the literature discussed above and are derived from a mixture of 2001 Census aggregate population data (England and Wales) and ONS model-based estimates[12]. In line with other studies of this type, the Herfindahl Concentration Index (Sturgis *et al.*, 2011; 2013) was applied using 2001 aggregate ethnic group census data to calculate the degree of ethnic heterogeneity in each MSOA. Furthermore, an ONS model-based estimate (Fry, 2011) of the proportion of households in poverty (defined as below 60 per cent of the UK median net equalised household income after housing costs) for each MSOA (2007/08) is used to get a handle on the levels of relative deprivation and income poverty in the wider residential neighbourhood. Finally, differential levels of neighbourhood population (in)stability are measured via a population churn statistic using data from the 2001 Special Migration Statistics Level 3, MSOA level migration data for total migrants from the 2001 Census SMS. The churn statistic (CH) for area i is defined as:

$$CH_i = \left(\frac{D_i + O_i + W_i}{P_i}\right)1000 \qquad (8.1)$$

where D_i is the inflow of individuals to MSOAi, O_i is the outflow of individuals from MSOA i, W_i is the count of individuals moving within MSOA i, and P_i is the population in MSOA i at census date, 2001. The inclusion of within area moves is important for reducing the potential influence of applying fixed boundaries of

12 Potentially important macro level characteristics, including measures of median house price and job density at the LAD and city region levels, were also collected but proved empirically insignificant when tested in the modelling framework outlined below.

varying geographical size, the Modifiable Areal Unit Problem (Openshaw and Taylor, 1981), by measuring all moves regardless of whether they cross these predefined boundaries. Moreover, a failure to incorporate internal movement relative to the population size can lead to a situation where two areas of similar turnover, but with drastically different internal mobility, are treated as similar residential contexts when in reality the stability of the neighbourhood populations are very different (Dennett and Stillwell, 2008).

8.4 Analysis and modelling strategy

Using the definitions outlined above, a substantial analytical sample of 224,164 individuals in England and Wales (25,978 planning to move, 11.6%) is incorporated into a multilevel statistical framework with 7,192 (level-2) MSOAs (containing a mean average of 31.2 respondents), 346 (level-3) LADs and 33 (level-4) city regions. As was noted in Chapter 4, multilevel modelling allows us to efficiently and simultaneously model the effect of individual and area level characteristics, as well as any cross-level interactions of potential substantive interest, on the propensity to be planning a move. Moreover, by nesting individuals into neighbourhoods and neighbourhoods into higher-level units, the multilevel framework handles the dependency and clustering of individuals and of nearby areas and allows for the separation and exploration of the relative contribution of each level to the total variation in the response.

A full multilevel logistic regression model with random intercepts and random coefficients is specified. Having randomly varying intercepts allows us to uncover the between-region, within-region, between-LAD and within-LAD, between-MSOA residual differences in the propensity to be planning a move whilst randomly varying coefficients provide the opportunity to test whether certain slope terms vary across higher level units. Indeed, given what was discussed in the review of the literature, the coefficient for duration-of-residence is allowed to vary across neighbourhoods (level-2) and districts (level-3). Equation 8.2 shows a simplified form of the full random intercepts and slopes logit model incorporating a single individual-level variable, a single neighbourhood-level variable and a cross-level interaction between the two variables:

$$\ln\left(\frac{\pi_{ijkl}}{1-\pi_{ijkl}}\right) = \beta_0 + \beta_{1jk}x_{1ijkl} + \beta_2 x_{2jkl} + \beta_3 x_{1ijkl}x_{2jkl} + f_{0l} + v_{0kl} + v_{1kl}x_{1ijkl} + u_{0jkl} + u_{1jkl}x_{1ijkl}, \quad (8.2)$$

where

$\ln\left(\frac{\pi_{ijkl}}{1-\pi_{ijkl}}\right)$	is the log-odds that individual i(level-1) in neighbourhood j (level-2), district k(level-3) and region l(level-4) is planning a move in the next 12 months (i.e. $y = 1$);
x_{1ijkl}	is a level-1 predictor variable (e.g. duration at residence);
x_{2jkl}	is a level-2 predictor variable (e.g. neighbourhood churn);
β_0	is the overall intercept and represents the log-odds that $y = 1$ across all $i, j, k,$ and l units when all predictors are held at their reference (i.e. $x = 0$ and $f = 0, v = 0, u = 0$);
β_{1jk}	is the estimated slope term associated with the level-1 predictor variable, the jk subscripts denote that this term is allowed to vary at level-2 and level-3;
β_2	is the estimated slope term associated with the level-2 predictor variable;
β_3	is the estimated slope term associated with the cross-level interaction between the level-1 and level-2 predictor variables;
f_{0l}	is the conditional random differential intercepts term for city regions (level-4);
v_{0kl}, v_{1kl}	are the within-region between-district conditional random differential intercepts term and random coefficient term (level-3);
u_{0jkl}, u_{1jkl}	are the within-district between-neighbourhood conditional random differential intercepts term and random coefficient term (level-2).

Due to the binary (0-1) outcome, the level-1 variance is assumed to come from the Bernoulli distribution with mean π_{ijkl} and a variance $\pi_{ijkl}(1-\pi_{ijkl})$. The random effects in equation 8.2 f_{0l}, v_{0kl}, v_{1kl}, u_{0jkl} and u_{1jkl} are on the logit scale and are assumed to follow normal distributions with zero means, variances σ^2_{f0}, σ^2_{v0}, σ^2_{v1}, σ^2_{u0} and σ^2_{u1} respectively, and covariances σ_{v01} and σ_{u01} reflecting the covariance between the intercepts and slopes at level-3 and level-2 respectively. All level-1 and level-2 fixed-part predictor variables have been centred at their mean (or typical) value so as to aid interpretation of the random part. Gross annual household income and residential duration are both measured using orthogonal polynomials, a parsimonious parameter coding system that allows for the maintenance and measurement of order within a categorical variable that is itself on an ordinal scale (Rasbash *et al.*, 2012).

Due to the binary nature of the response variable, MCMC estimation is used, providing a more efficient and robust estimation to the maximum likelihood based alternatives (Chapter 4). All models are estimated using the MLwiN 2.29 software (Rasbash *et al.*, 2013). Initial parameter starting values are based on maximum likelihood methods and, for the fully specified model, a burn-in of 5,000 iterations is followed by a monitoring chain of 800,000 simulations with model convergence assessed following the good-practice recommendations of Draper (2006) and Jones and Subramanian (2013). To aid with the mixing of MCMC parameter chains, hierarchical centring and orthogonal parameterisation techniques are used (Browne *et al.*, 2009; Browne, 2012).

The modelling strategy involves specifying a series of three models. Model 1, a null model (variance components model) with random intercepts only, gives an idea of how the total variability in the propensity to be planning a move is partitioned across individuals, neighbourhoods, districts and regions, before compositional differences between individuals and neighbourhoods are accounted for. With the aim of uncovering the conditional effect of residential duration on mobility propensities, Model 2 includes all the level-1 and level-2 variables described above as well as the theoretically informed cross-level interactions between them. Model 3 (Equation 8.2) extends on Model 2 by allowing the effect of residential duration to vary across the different neighbourhoods and districts of England and Wales. By adding the random slope terms, it is possible to assess the extent to which remaining (residual) between-neighbourhood and between-district differentials depend on duration. Moreover, the opportunity to test the importance of exposure times for the development of distinct residual areal externalities and differentials is also provided. The following section briefly describes the variance components and model fit statistics, before a detailed analysis and discussion of duration-of-residence effects is provided. However, whilst the substantive focus remains with duration effects, other covariates that show particularly interesting patterns and relationships are also interpreted and discussed.

8.5 Model results

Table 8.2 presents the results from the three models. In order to facilitate an interpretation of the magnitude of non-individual variance, the between individual variance is assumed to follow a standard logistic distribution of 3.29 (Snijders and Bosker, 2012). Through the use of this standard assumption, the null model (Model 1) estimates that 4 per cent of the variance in individuals' plans for future residential movement is attributable to contextual, non-individual, variation ([0.03 + 0.02 + 0.078] / [0.03 + 0.02 + 0.078 + 3.29]). Whilst this value may initially appear rather small, it closely reflects the findings of similar analyses by Feijten and van Ham (2009). Indeed, given that the micro-level (individual/household) is the level where ultimately the decision/ability to change residence is made, we should expect the differential characteristics at this level to be dominant, although there is evidence of contextual variation in the probability to be planning a move. Thus, building on this, the theoretically informed level-1 and level-2 covariates and (micro-level and cross-level) interactions are included (Model 2), leading to a substantial improvement in model fit (DIC is 13,561 units lower than in Model 1) before, finally (Model 3), the estimated duration effects are allowed to vary across neighbourhoods (level-2) and districts (level-3), which again leads to a significant reduction in the DIC statistic (DIC is 66 units lower than in Model 2).

Table 8.2. Multilevel logit model results - planning a residential move

	Model 1 Null: Beta (S.E.)	Model 2 Intermediate: Beta (S.E.)	Model 3 Full: Beta (S.E.)
Fixed Effects			
Constant	-2.115 (0.034)*	-2.660 (0.027)*	-2.682 (0.028)*
Age (centred at 46)		-0.022 (0.001)*	-0.022 (0.001)*
Age^2		0.000 (0.000)*	0.000 (0.000)*
Gender (ref = Female)		0.073 (0.015)*	0.073 (0.015)*
Ethnic group (ref = White)			
Asian		0.222 (0.063)*	0.221 (0.063)*
Other		0.352 (0.051)*	0.352 (0.051)*
Black		0.342 (0.074)*	0.338 (0.075)*
Marital status (ref = Married)			
Divorced/separated		0.272 (0.023)*	0.271 (0.023)*
Single		0.125 (0.021)*	0.125 (0.021)*
Living With Partner		0.262 (0.021)*	0.262 (0.021)*
Widowed		0.027 (0.051)	0.027 (0.051)
Gross household income (linear polynomial)		0.385 (0.035)*	0.386 (0.035)*
Occupation group (ref = Intermediate)			
Retired		-0.021 (0.038)	-0.020 (0.038)
Homemaker		0.025 (0.027)	0.026 (0.027)
Higher managerial administrative & professional		0.051 (0.020)*	0.052 (0.020)*
Routine & manual		-0.075 (0.026)*	-0.075 (0.026)*
Unemployed		0.189 (0.034)*	0.188 (0.034)*
Student		-0.048 (0.039)	-0.047 (0.039)
*Educational attainment (*ref = 5+ GCSEs)			
None		-0.051 (0.021)*	-0.051 (0.021)*
2+ ‘A’ levels		0.046 (0.020)*	0.046 (0.020)*
Degree		0.100 (0.019)*	0.101 (0.020)*
Housing tenure (ref = Home owner)			
Rent private		1.063 (0.027)*	1.068 (0.027)*
Rent housing association		0.227 (0.038)*	0.227 (0.038)*
Rent council		0.159 (0.032)*	0.163 (0.032)*
*Housing tenure*Age*			
Rent private, Age		-0.022 (0.002)*	-0.022 (0.002)*
Rent private, Age^2		-0.000 (0.000)*	-0.000 (0.000)*
Rent housing association, Age		-0.028 (0.002)*	-0.028 (0.002)*
Rent housing association, Age^2		0.000 (0.000)	0.000 (0.000)
Rent council, Age		-0.025 (0.002)*	-0.025 (0.002)*
Rent council, Age^2		0.000 (0.000)*	0.000 (0.000)*
Residential duration			
Linear polynomial		-0.356 (0.045)*	-0.455 (0.056)*
Quadratic polynomial		-0.071 (0.041)*	-0.135 (0.043)*
Cubic polynomial		0.147 (0.040)*	0.143 (0.041)*
Quartic polynomial		-0.116 (0.035)*	-0.117 (0.035)*
Data set (ref = 2007)			
2005		0.037 (0.015)*	0.036 (0.015)*
2006		-0.356 (0.024)*	-0.357 (0.024)*
Neighbourhood churn (gm-centred)		0.001 (0.000)*	0.001 (0.000)*
*Duration*Neighbourhood churn*		-0.003 (0.000)*	-0.003 (0.001)*
Neighbourhood income poverty (gm-centred)		0.008 (0.001)*	0.008 (0.001)*
*Gross household income*Neighbourhood income poverty*		0.011 (0.003)*	0.012 (0.003)*
Neighbourhood ethnic heterogeneity (gm-centred)		0.474 (0.083)*	0.462 (0.081)*

Table 8.2. (continued)

	Model 1 Null: Beta (S.E.)	Model 2 Intermediate: Beta (S.E.)	Model 3 Full: Beta (S.E.)
*Ethnic group*Neighbourhood ethnic heterogeneity*			
Asian, Ethnic heterogeneity		-0.518 (0.195)*	-0.507 (0.196)*
Other, Ethnic heterogeneity		-0.898 (0.203)*	-0.886 (0.203)*
Black, Ethnic heterogeneity		-0.178 (0.213)	-0.160 (0.215)
Random Effects			
Level-4 City Region:			
σ^2_{f0} (Intercept variance)	0.030 (0.011)	0.004 (0.002)	0.003 (0.002)
Level-3 District:			
σ^2_{v0} (Intercept variance)	0.020 (0.003)	0.003 (0.002)	0.008 (0.003)
σ^2_{v1} (Duration slope variance)			0.070 (0.029)
σ_{v01} (Intercept-duration covariance)			0.021 (0.008)
Level-2 Neighbourhood:			
σ^2_{u0} (Intercept variance)	0.078 (0.007)	0.026 (0.006)	0.043 (0.010)
σ^2_{u1} (Duration slope variance)			0.717 (0.162)
σ_{u01} (Intercept-duration covariance)			0.113 (0.035)
Level-1 Individual:			
Variance (Residual)	3.29	3.29	3.29
DIC:	159345.474	145784.156	145718.504

*N.B. Estimated coefficients (Beta) are logits; * indicates parameter is significant at the 95 per cent level; gm-centred denotes variable is centred on its grand mean value – ethnic heterogeneity (centred at 0.102), income poverty (centred at 21.946) and churn (centred at 184.501).*

8.5.1 Additional substantive observations

Before moving to the main analytical focus of this chapter, i.e. the relationship between duration-of-residence and plans for future residential mobility, it is thought important to briefly discuss some additional observations that are observed to be of particular relevance for predicting the probability to be planning a future residential move. Thus, largely conforming to the expected patterns described in the literature, Figure 8.3 presents the additional individual/household and neighbourhood characteristics that are found to reflect the greatest differentials in the probability to be planning a residential move. These being: (a) marital status; (b) occupational status; (c) educational attainment; (d) the interaction between age and housing tenure; and (e) the cross-level interactions between ethnicity and neighbourhood ethnic heterogeneity and (f) household income and neighbourhood deprivation. The rather wide credible intervals relate to the fact that the predicted probabilities are for an otherwise typical person (level 1) in the typical MSOA, typical LAD and the typical City Region, where in the latter case there are only 33 spatial units.

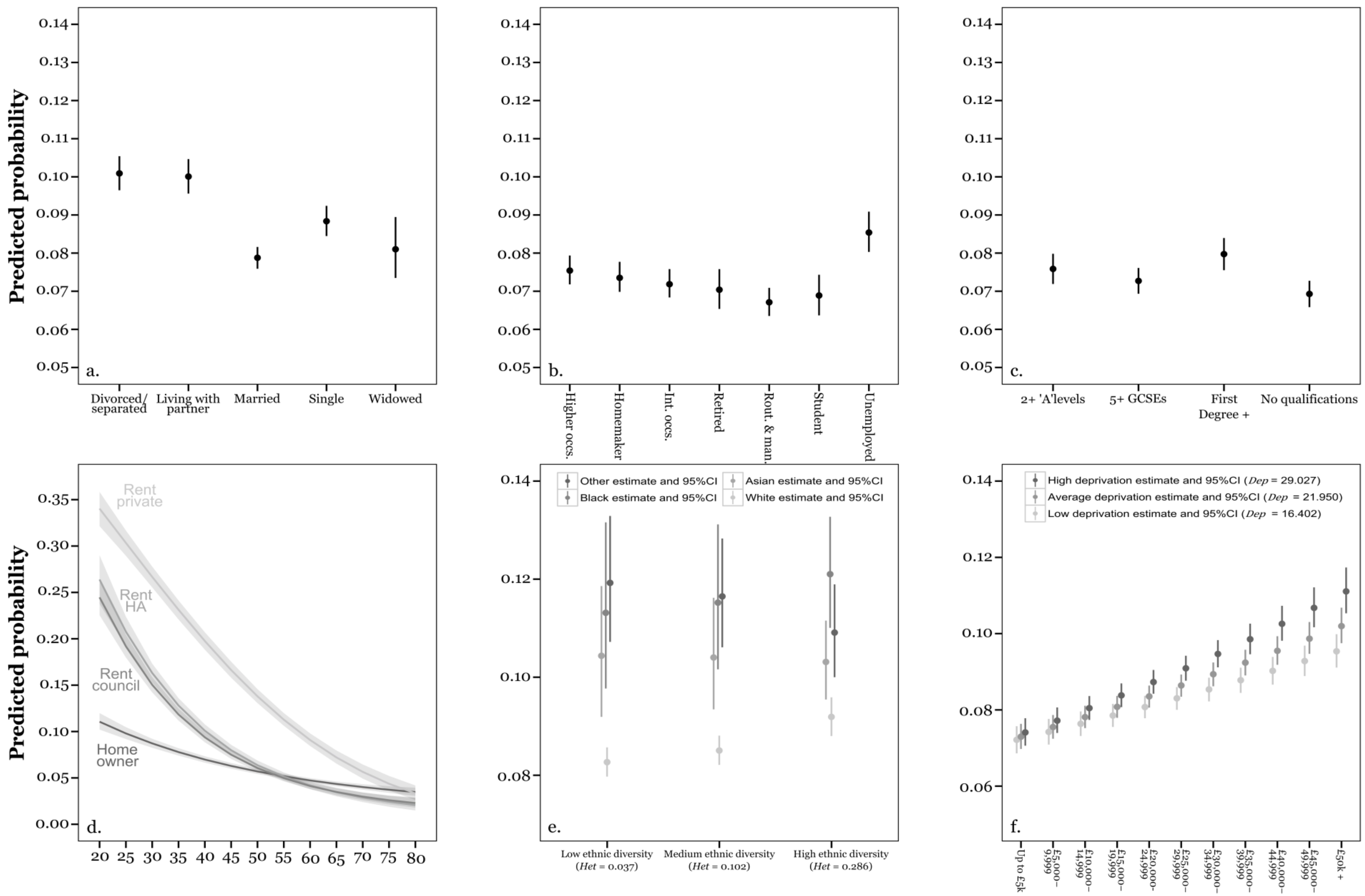


Figure 8.3. Predicted probability of planning a residential move by selected covariates, estimates and 95% credible intervals

As with all of the analyses presented in this thesis, marital status is used here as a rather crude proxy for co-residency, household structure and the identification of tied decision-making processes. Indeed, a particular limitation of relevance is the unfortunate omission of additional family relevant covariates detailing the effects of pregnancy and childbirth, which are known to raise the probability for residential mobility (Clark and Huang, 2003; Rabe and Taylor, 2010), and the presence of school-age dependent children in the household, where parental desires to avoid disrupting children's education is thought to lower mobility propensities (Fielding, 2012). Whilst the lack of a longitudinal perspective and certain relevant covariates in the data removes the opportunity to explore potentially important household events and transitions (Boyle *et al.*, 2008; Mulder and Wagner, 2010), differences between the crude marital statuses are found to be relatively small with the 95% credible intervals overlapping in most instances (Figure 8.3a). Although, it does appear that, holding all else constant, married individuals are the least likely to be planning a change of residence, perhaps reflecting the relative stability of this partnership formation when compared to others.

With regards occupational status, quite small differentials between the groups are observed, with the exception being the unemployed who, whether self-determined or socially/economically imposed, are seen to have particularly raised probabilities of planning a future residential mobility (Figure 8.3b). In terms of educational qualifications, the common pattern emerges wherein those with higher educational attainment are generally associated with increased probabilities of planning a residential move, those with the highest educational attainment are often expected to be in careers that require a degree of spatial flexibility for progression (van Ham *et al.*, 2001) and are additionally expected to have access to sufficient resources so as to make a change of residence a viable option if desired (Champion *et al.*, 1998; Bailey and Livingston, 2005; Fielding, 2012).

By far the largest differential effects are found when age and household tenure are interacted (Figure 8.3d). As expected, across all tenure groups the probability of planning a move decreases with age; however, the extent of each slope is quite different, particularly for private renters as compared with home owners. Indeed, it appears that across the age groups, owner occupiers generally reflect comparatively low movement propensities. Again, as Mulder (2013) has noted, home ownership is comparatively the least flexible tenure type, the high transaction costs of home ownership, the transfer tax, and the mortgage costs as well as the security of tenure and protection against eviction that ownership provides, are all factors that can be expected to reduce the desire/enactment of residential movement, when compared to other tenure groups. This is contrasted with private renting where greater flexibility, or for some insecurity, is associated with a mix of lower movement costs, short-term residential durations and higher propensities for residential mobility. Private renting can be expected to reflect very different social and economic circumstances for individuals depending on their stage in the life course, which itself may explain the particularly sharp decrease in movement propensities with age for this group. Whilst private renting is a common and often desired tenure type for those in younger age groups, considering the greater space, quality and security often afforded by home ownership as well as the strong social norms prescribing home ownership as the desirable/successful way of living (Lauster, 2010), private renting can, in some cases, be expected to reflect a relatively disadvantaged social and economic position for those in their middle and later years. In terms of the cross-level interactions, both appear to support the theorised relationships explained in the literature.

Whilst the white ethnic majority have lower propensities for movement than the non-white ethnic minorities, their propensity to move increases as the neighbourhood diversity grows (Figure 8.3e). Moreover, in terms of household income and neighbourhood deprivation (Figure 8.3f), the expected pattern emerges wherein higher levels of neighbourhood deprivation are seen to be particularly important for encouraging plans for residential movement. Furthermore, those with access to the greatest household incomes are more likely still to be planning a move. Given the negative externalities associated with high neighbourhood deprivation, those with the highest incomes can be expected to be better able to act on the associated residential stresses and plan a move away, whereas those with access to the lowest incomes are unlikely to be in a position to approach a stage of serious planning, even if a move away is indeed desired.

8.5.2 Duration-of-residence effects

Having controlled for a wide range of theoretically informed characteristics, including those discussed above, the estimates from Model 3 (Table 8.2) reveal a statistically significant nonlinear duration-of-residence effect. Indeed, as shown by the thick line in Figure 8.4, the functional form for the average residential duration relationship is reflective of neither simple cumulative inertia nor cumulative stress; rather it more closely echoes that of Gordon and Molho's (1995) analysis, wherein we observe an initial rise before a gradual tailing-off in movement probabilities as duration increases. As Gordon and Molho (1995: 1972) argued in their conclusions, whilst longer residential durations may indeed increase the costs of a residential move, changes brought on by the passage of time will often lead to a situation where original residential preferences are re-evaluated in favour of alternatives. Thus whilst residential mobility is relatively low for the very shortest (1 year) durations, the non-stationarity of residential preferences can lead to residential mismatches that, in some cases, will necessitate a change of residence.

Figure 8.4 does support the importance of variations in neighbourhood population (in)stability in conditioning the duration effect. As was predicted, the probability of planning a move is raised for those residing in neighbourhoods of higher population instability (churn). Moreover, as was also suggested, the differential effect is found to be most pronounced when the duration-of-residence is relatively short. Thus, whilst it is possible that population instability creates an environment that is detrimental to the forging and maintenance of stronger place-based social ties, given the raised peaking for short durations and the apparent insignificance of population instability as a mediating factor for those with longer durations, an alternative explanation is more appropriate. Indeed, as was discussed above, neighbourhoods with the highest population instability are generally those that contain high proportions of privately rented dwellings and student and young unattached populations. Thus, given the housing stock and demographic profile of these neighbourhoods, it is perhaps more probable that moves to such areas are made with the pre-understanding and preference that residency will be short-term.

As was briefly mentioned in Section 8.2, previous empirical findings revealing cumulative inertia effects have been said to more accurately portray the results of the unwanted correlation between the duration dependence variable and the residual, a result of the failure to sufficiently account for sample heterogeneity. That is, the cumulative inertia relationship is a spurious effect of selection, wherein those with low propensity to move over and above the effects of the explanatory variables, perhaps due to unmeasured behavioural predispositions, will tend to have longer than expected durations while those with a high propensity to move will tend to have shorter than expected duration-of-residence (Davies, 1991). Thus, duration dependence may take on a spurious negative relationship with the propensity to move, because those with longer durations are also those with a predisposition to stay put, and thus those with the lowest chances of planning a future move. Whilst this is something that must be considered, the results presented here do not reflect the relationship that would be expected if this were the case, i.e. a cumulative negative duration dependence on movement probabilities. Indeed, the inclusion of a large number of theoretically informed individual, household and neighbourhood controls appears to have been successful in capturing the sample heterogeneity which has, in previous analyses, led to spurious negative duration dependence[13]. Consequently, given the consistency in the pattern here and observed by Gordon and Molho (1995), it can be argued with reasonable confidence that the pattern revealed is one more accurately reflecting genuine duration effects as opposed to simple selection effects.

[13] A simple bivariate model, before controls, did reveal a simple (monotonic) negative duration dependence pattern.

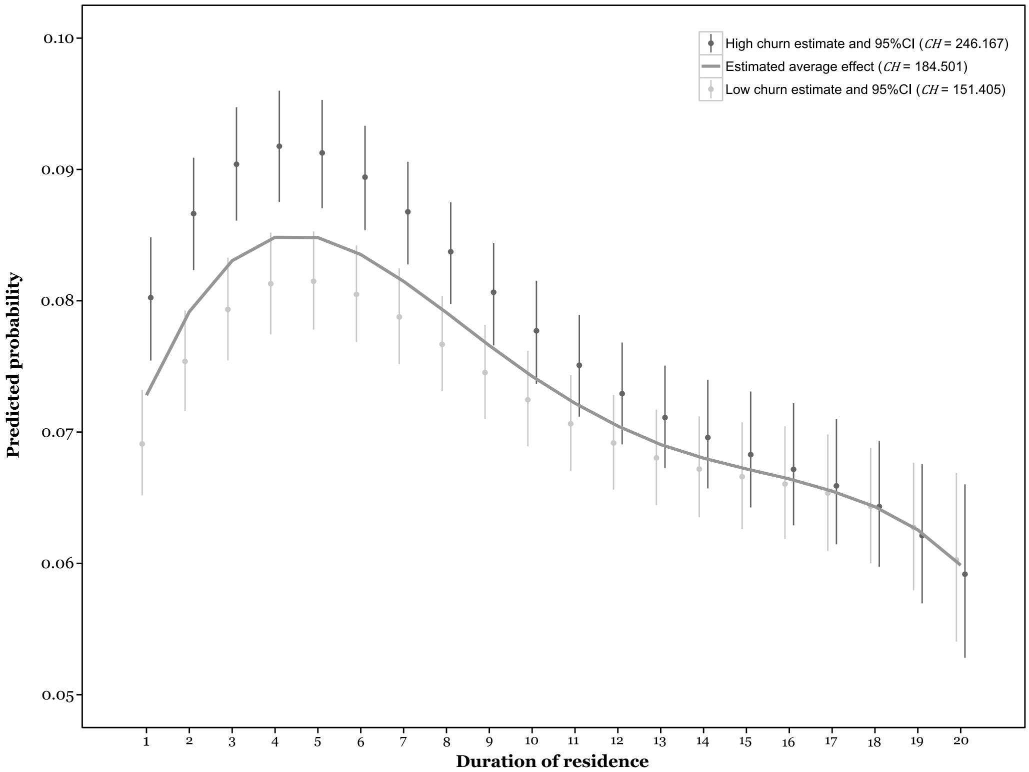


Figure 8.4. Probability of planning a residential move by residential duration and neighbourhood population (in)stability (measured by population churn)

Whilst the average duration effect is of clear interest, given the expected importance of social and economic processes, practices and mechanisms that operate within the wider residential locale, there is strong reason to expect the effect of residential duration on movement propensity to vary geographically, from neighbourhood to neighbourhood, and even district to district. Indeed, it is possible that certain residential environments engender greater residential dissatisfaction whereas others encourage greater stability and thus lower probabilities of movement with time. Moreover, closely linked to this is the expectation that greater exposure times to wider residential environments are important in allowing for neighbourhood effects/influences to manifest themselves, and thus influence individual mobility decision-making and outcome behaviours. With this in mind, the effect of residential duration is allowed to vary across neighbourhoods (level-2) and districts (level-3), the results of which are presented in Figures 8.5-8.9.

The positive covariance terms for levels 2 (σ_{u01}) and 3 (σ_{v01}) in Table 8.2 suggest, that there is evidence for a quadratic growth in contextual variation as duration increases. Whilst the between-district variation remains relatively small across the duration scale (Figure 8.5), the between-neighbourhood variation[14] reflects a pattern that is consistent with what would be expected if longer exposure times are indeed relevant for the emergence of substantively important neighbourhood influences and externalities. In other words, whilst the existence of omitted variables operating at the individual and contextual levels makes the definitive confirmation of neighbourhood effects problematic, the pattern revealed in Figure 8.5 appears consistent with the argument that the residential environment grows in significance as the duration of residence (exposure) increases.

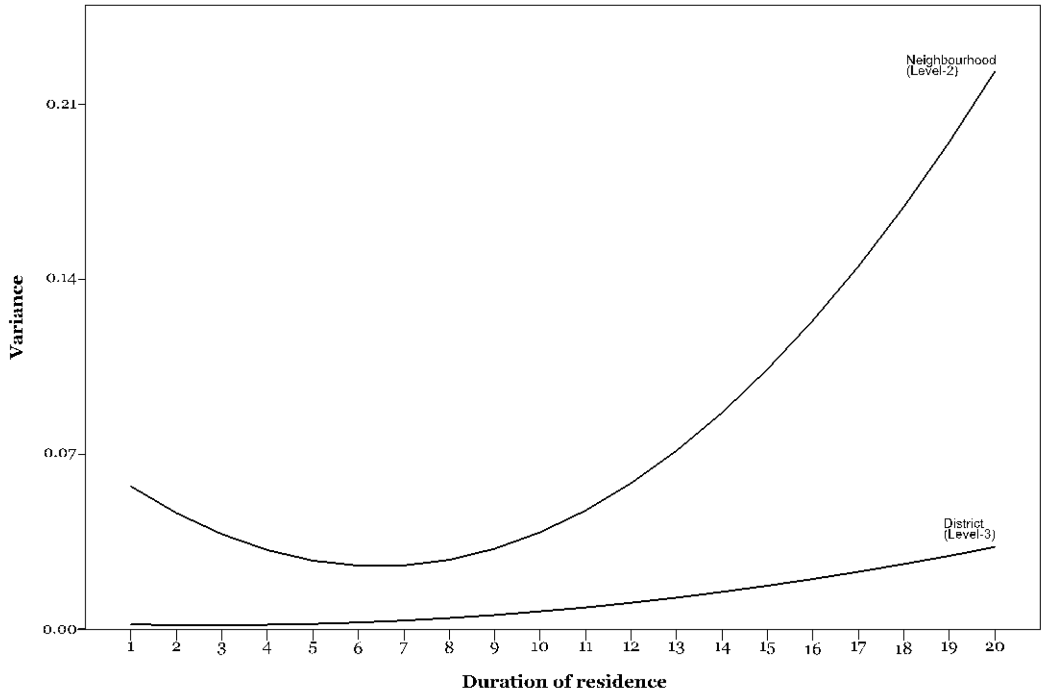


Figure 8.5. Between-neighbourhood and between-district variance (as a function of duration) in the log-odds of planning a residential move

A more specific exploration of the between-neighbourhood variation is provided for in Figures 8.6-8.9, where it is clear that quite substantial differences exist between neighbourhoods in both the strength and direction of their respective slope terms. For instance, Figure 8.6 presents a sample of 30 neighbourhoods which are

---

14 Using the terms from Equation 8.2, the between-neighbourhood variance for Model 3 with a random slope for residential duration and a quadratic specification is:

$$var(u_{0jkl}x_0 + u_{1jkl}x_{1ijkl}) = \sigma^2_{u0}x^2_0 + 2\sigma_{u01}x_0x_{1ijkl} + \sigma^2_{u1}x^2_{1ijkl}$$

characterised by negative coefficients wherein, on average, the probability of planning a residential move is found to decrease with duration (i.e. consistent with cumulative inertia). It may well be the case that these neighbourhoods are particularly conducive to the creation of greater social and institutional capital discussed by Sampson *et al.* (2002).

Conversely, Figure 8.7 presents a similar sized sample of neighbourhoods with flat duration relationships, that is, the length of duration in these neighbourhoods does not appear to be important for informing individual propensities for residential mobility. Figure 8.8 shows neighbourhoods with patterns reflecting those expected under cumulative stress, wherein longer duration-of-residence is associated with a greater probability to be planning a move. Again, these neighbourhoods may well engender particular unmeasured externalities that work to cumulatively encourage movement away.

Finally, Figure 8.9 presents the random slopes for all of the neighbourhoods included in the analysis (n = 7,192), revealing the extent of appreciable neighbourhood heterogeneity across England and Wales in the probability to be planning a move. Indeed, the heterogeneity reflects differing duration-of-residence effects and appears to offer some support to the theorised relevance of greater exposure times (duration) for the effects of wider contextual externalities and differences to emerge. Indeed, for an otherwise typical person, it is suggested that those with longer durations at an address will see their probability of planning a move noticeably vary according to a constellation of unmeasured factors, and potentially including the unmeasured contextual differences associated with the residential environment in which they live. In an attempt to uncover any geographical patterning to the higher/lower probabilities of movement with duration, the neighbourhood (MSOA) slope coefficients were visualised using a GIS, where the resulting maps suggested no clear evidence of any systematic spatial patterning or clustering.

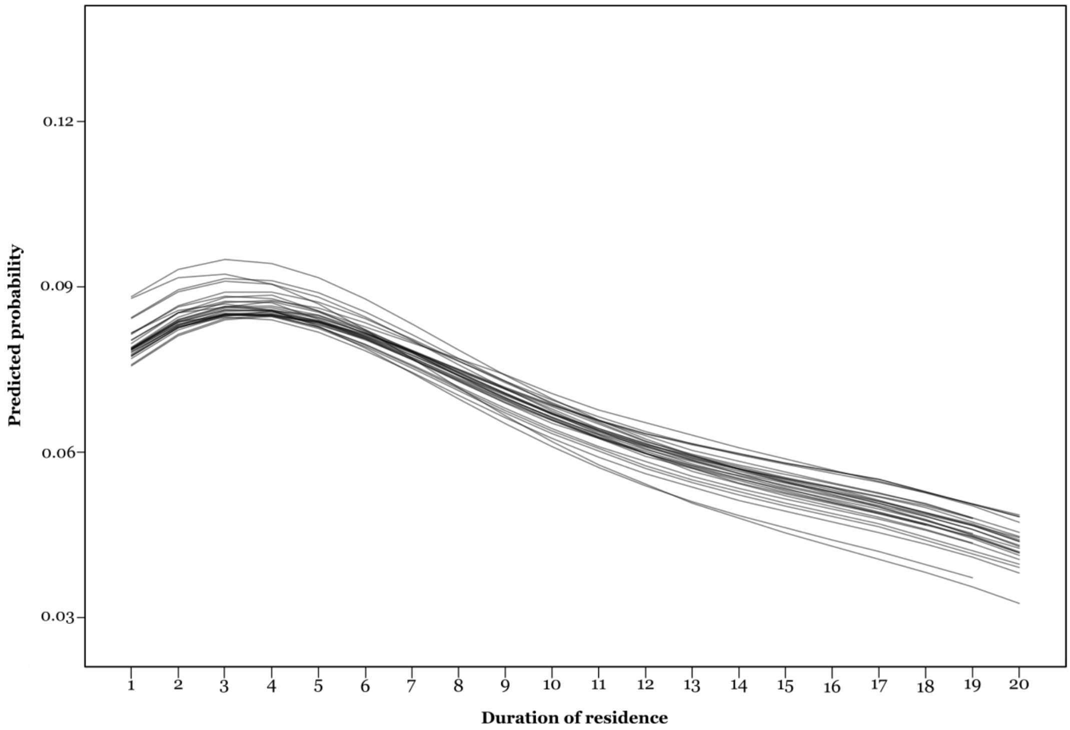


Figure 8.6. Predicted probability of planning a residential move by residential duration (years) across selected (MSOA) neighbourhoods: Inertia neighbourhoods

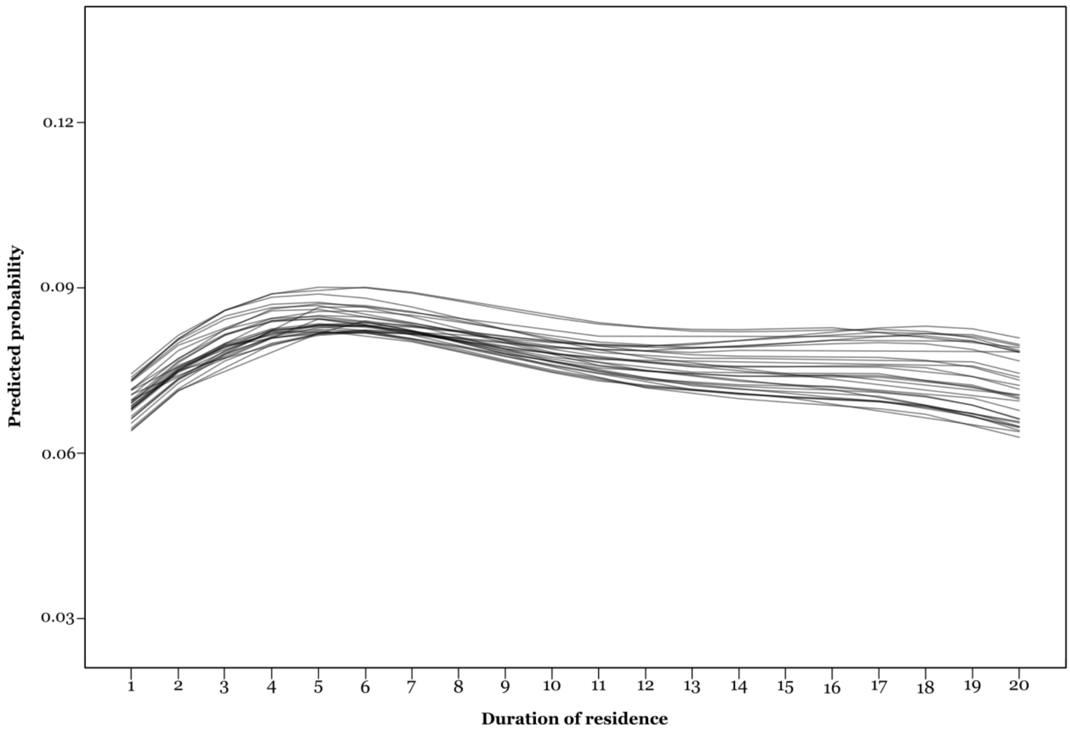


Figure 8.7. Predicted probability of planning a residential move by residential duration (years) across selected (MSOA) neighbourhoods: Flat neighbourhoods

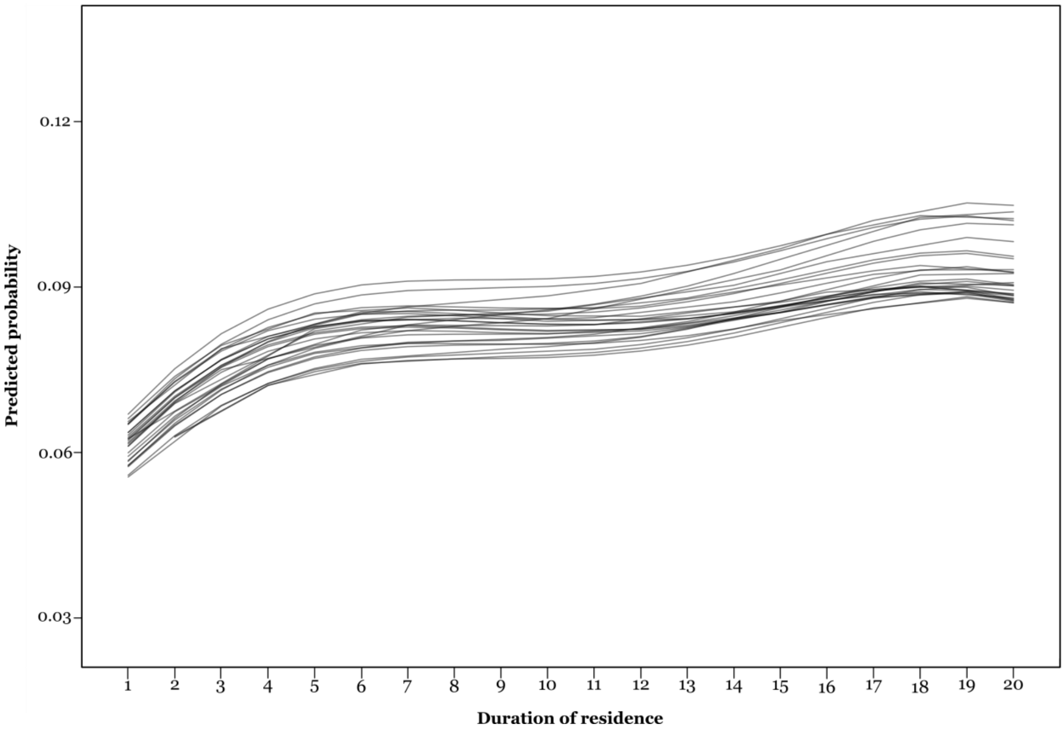


Figure 8.8. Predicted probability of planning a residential move by residential duration (years) across selected (MSOA) neighbourhoods: Stress neighbourhoods

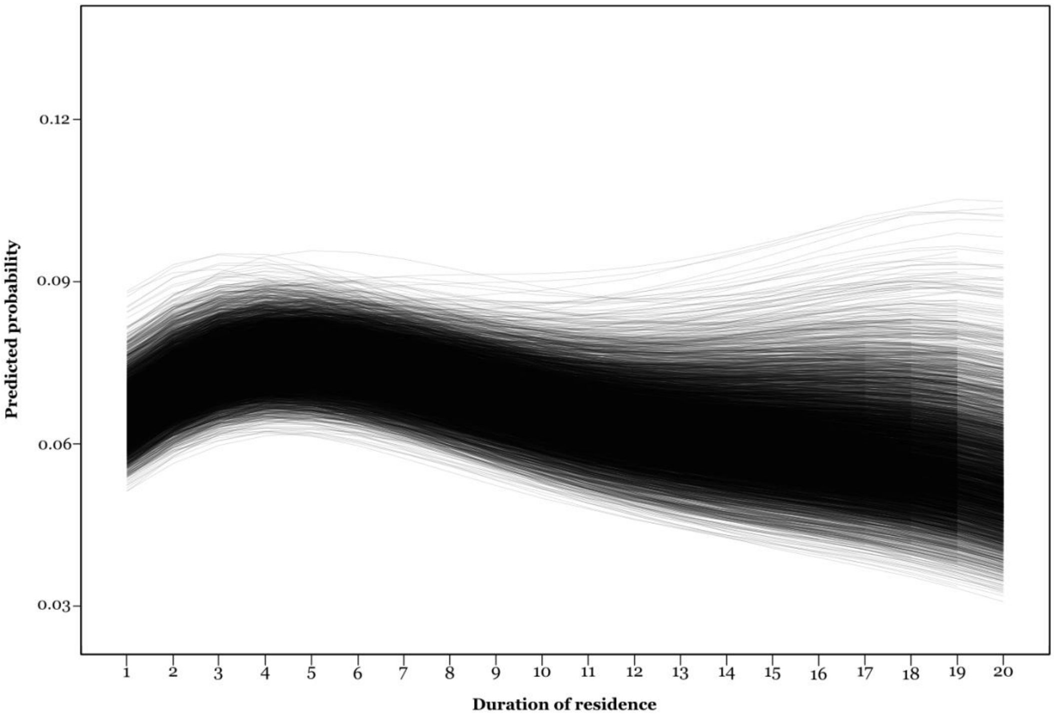


Figure 8.9. Predicted probability of planning a residential move by residential duration(years) across all neighbourhoods (MSOAs)

8.6 Summary and conclusions

This chapter has attempted to explore the functional form of the relationship between duration-of-residence and future residential mobility behaviour. Through the combined use of the ROP, with its large, attribute rich and geographically extensive sample, and appropriate statistical methods, it has been possible to build on previous empirical analyses in uncovering the extent to which various multilevel factors associated with residential duration work to influence propensities for future residential movement. Firstly, it has been shown that the simple monotonic laws of cumulative inertia and cumulative stress fail to reflect the complexity of the phenomenon at hand. Rather, with the inclusion of a number of suitable controls for additional sources heterogeneity in the response, the empirical analysis suggests that a functional form consistent with that first put forward by Gordon and Molho (1995) is more appropriate. That is, the predicted probability to be planning a move is observed to first rise in the early years of duration before gradually tailing-off as duration extends. Thus, as an average effect it would appear fair to agree with Gordon and Molho (1995) in suggesting that, whilst longer durations of residence may increase the costs of a residential move through the forging of stronger place-based social and economic ties, different events and changes brought on by the passage of time will inevitably lead to a situation where some individuals/households reevaluate their original residential preferences in favour of alternative possibilities in an area or house elsewhere.

Beyond this, the importance of factors relevant to the wider residential context has also been revealed. As was expected given the review of the relevant literature, the degree of neighbourhood population (in)stability is found to play an important mediating role in the duration-of-residence effects, though its influence is largely restricted to those with relatively short durations. Whilst high neighbourhood population instability may indeed create an environment that is problematic for the creation and maintenance of strong place-based social networks and ties, given that the major differentials are found to be between those with short durations, it is suggested that the effects are perhaps more an artefact of the differing housing and demographic profiles of the neighbourhoods than much else. With high-churn neighbourhoods generally observed to be the more dynamic urban areas of England and Wales (Figure 8.1), those with high proportions of young single adults, students, and privately rented dwellings, we can perhaps expect individuals who recently moved to these areas to have made the decision with the pre-understanding and preference that residency would be highly flexible and thus short-term.

In addition to the basic characteristics and composition of the neighbourhood, it is also argued that an attempt to identify the potential influences of more subtle, and harder to measure, neighbourhood phenomena should also be made. Indeed, opportunities to develop strong place-based habitual practices, social ties and interactions, mutual trust, an appreciation of the collective efficacy of one's neighbourhood and an attachment to local institutional resources, such as organised social and recreational activities, can all be expected to take time; whilst their influences on individual evaluations of residential satisfaction and mobility behaviour can also be expected to necessitate a critical period of cumulative exposure (Tienda, 1991; Sampson *et al.*, 2002; Hedman, 2011; van Ham *et al.*, 2014).

With the specification of a random coefficients model, the analysis presented here allowed not only for a substantive focus on how the duration-of-residence effects are far from homogenous in their form, indeed they are found to vary quite substantially in both direction and effect across the different neighbourhoods of England and Wales, it also made possible the identification of a pattern that suggests that greater heterogeneity in the propensity to be planning a move exists when residential durations are longer, as compared to when they are shorter. Whilst we should always remain cautious of the potential influence of omitted variable bias in multilevel analysis of neighbourhood effects (van Ham and Clark, 2009), such findings can be thought to lend support to the notion that greater exposure times are important for the detection and observation of appreciable residential neighbourhood effects, and perhaps most significantly the more subtle hard to measure externalities, on individuals' evaluations of their residential milieu and associated movement behaviours.

Future research should attempt to build on the work presented here and incorporate a longitudinal dimension to the analysis of duration effects on movement propensities. For instance, whilst this research has been useful in exploring the relative importance of geographical context, through data limitations it lacks a detailed longitudinal context/perspective that would allow for the exploration and interaction of important life-course events and transitions, expected or otherwise (e.g. unemployment, pregnancy or the birth of a child, union formation/dissolution, widowhood, etc.), on mediating the effects of residential duration and mobility, and *vice versa*. Beyond this, an acceptance for the changing nature of neighbourhoods through time also needs greater attention; indeed, in this analysis neighbourhoods have been treated as if they are static in their composition and characteristics. Whilst measures of neighbourhood churn provide us with some concept of neighbourhood change, a more explicit incorporation of the temporally dynamic nature of neighbourhood characteristics and phenomena through time may complement the work presented here and further our understanding of duration effects, wider residential evaluations and movement propensities. Moreover, as Gordon and Molho (1995) suggested in their conclusions almost 20 years ago, further work is still required to establish the prevalence of duration effects in relation to the length of duration in a locality rather than that in a single household. After all, whilst duration dependencies at the level of the housing unit are found not to be consistent with cumulative inertia, there may be evidence for the process at a more aggregate geographical level (neighbourhood/district/region), where people move house but stay ‘local’. Again the data used here did not allow for this, but with the development of increasingly rich large-scale geo-coded longitudinal datasets (e.g. Understanding Society), there is certainly some potential for future analyses of this kind.

Chapter 9
Modelling micro, meso and macro variations in origin to destination distance moved

9.1 Introduction

The two previous analytical chapters have focussed on the basic decision to move, incorporating and testing various individual, household and place based characteristics considered important for conditioning the desires, constraints and likelihood of people moving as opposed to remaining *in situ*. This chapter shifts the analytical focus from variations in the propensity to move, towards the variations in the distance of migration, once the decision to move has been made. Implicit in the definition of residential mobility is the physical relocation from one place of usual residence to another where the origin and destination may be in close proximity or separated by long distance. The theoretical literature relating to variations in the distance over which residential movement takes place emphasises the importance and complexity of influences that operate simultaneously at the origin and the destination, in addition to the role of the distance or opportunities intervening between them (Lee, 1966). Explanatory factors are likely to embrace those that relate to variables impacting at various levels from the circumstances of the individual and the household in which the migrant resides to the local neighbourhood in which the migrant's household is located through to the region, nation and indeed beyond, perhaps even to the global level. Individual migration behaviour in the UK in the second half of the 2000s has been influenced by what, in effect, has been the impact of global recession.

However, in practice, much empirical work on residential mobility falls short in terms of recognising these realistic complexities by focussing exclusively on one level, be it the micro- or the macro-level distinguished in Chapter 4, and therefore failing to account for potentially important influences operating at other levels that are omitted. Moreover, on the rare occasions when realistic multilevel structures/influences have been analysed (see, for instance, Boyle and Shen, 1997), a failure to accommodate influences operating at both the origin *and* the destination is apparent. With the aim of rectifying this partiality, the intention of this analytical chapter is to develop a theoretically informed modelling approach that captures the effects at different levels which impact on the distance over which individuals change residence. The chapter proceeds in the following manner: first, the relevant theoretical and empirical literature is reviewed with a key focus placed on drawing out the major processes, patterns and characteristics that operate at the micro- (individual/household), meso- (neighbourhood context) and macrogeographical- (structural region) levels (Section 9.2). Following this, the data and measures used for the analysis are described in detail (Section 9.3) leading to the outline of an analytical framework and modelling strategy that is designed with the purpose of accommodating the necessary levels of complexity for the exploration of multilevel variations in distance moved (Section 9.4). The results of the multilevel analysis are presented and discussed in Section 9.5, before the summary and conclusions in Section 9.6.

9.2 A multilevel theory of variations in origin to destination distance moved

As was briefly mentioned in Chapter 2, one of the most important contributions to an all-embracing multilevel origin to destination theory of population movement was given by Everett S. Lee (1966) in his seminal paper *"A theory of migration"*. Central to the paper are four headings which Lee outlines as being essential for informing the *"decision to migrate and the process of migration"* (Lee, 1966: 49); these are: *factors associated with the area of origin*, *factors associated with the area of destination*, *intervening obstacles*, and *personal factors*. Fundamentally, it is assumed that such factors influence the decision to move by informing the evaluation of a balance between the degree of satisfaction with one's current residence and the strength of the desire, need and indeed ability to seek alternative residency (Quigley and Weinberg, 1978; Clark and Dieleman, 1996; Clark and Ledwith, 2006).

The patterns, processes and characteristics of residential mobility are thought therefore to be driven by certain 'push' and 'pull' dynamics that are conditioned (encouraged or discouraged) by a constellation of factors operating at different levels at both the origin and the destination (Rossi, 1955; Massey, 1990; Fielding, 2012).

For example, the decision to change residence can be influenced by 'pulls' to a new residence, for instance driven by the potential for new or improved employment and/or lifestyle possibilities, as well as 'pushes' at the current residence, enacted in some cases, for instance, by a sudden change in household composition (e.g. birth, death or cohabitation) or a gradual shift in lifestyle and consumption preferences away from those currently on offer. Yet, whilst Lee's is a considerably more general theory of mobility in its broadest sense, as an overarching theory it can be thought to hold great relevance and potential for the more specific examination and explanation of variations in the distance moved between the origin and the destination.

The influence of intervening obstacles and the selective dimensions (behaviours and characteristics) that operate at the individual/household level are well rehearsed within the existing literature on residential mobility (see, for example, Rossi, 1955; Champion *et al.*, 1998; Bailey and Livingston, 2005; Fielding, 2012). However, given the focus of this chapter, it is important to reemphasise the centrality of distance itself as an important obstacle. Indeed, intervening distance, when operating in parallel with additional selective dimensions, functions so as to make residential movements over long distances largely the preserve of a relative economic and social elite. The increasing distance of a residential move is thought to be tied to increasing restrictions and costs associated with, for instance, the relinquishing of ties to locality-specific social networks and amenities (Brown, 2002); the likely change in employment and/or the workplace (Owen, 1992); the financial costs and implications associated with the search and of the move itself (Flowerdew, 1976); and the requirement and acquisition of information on opportunities available in places far afield (Flowerdew, 1982). Thus, if a long-distance move is the desired outcome, be it for push and/or pull factors working at the origin and/or the destination respectively, functioning in combination, these costs and restrictions can be understood to intervene in the process by filtering those individuals/households with sufficient resources and motivation to ultimately satisfy the desire to migrate to destinations further afield.

The understanding of the strong selective nature of the micro-level dynamics behind variations in distance moved is supported by much empirical work demonstrating how certain individual/household characteristics are associated with short-distance moves while others are more closely aligned with moves over longer distances. For instance, the average distance moved is often found to increase in a linear manner with the level of educational attainment and household income (Fielding, 2007; 2012; Poston and Bouvier, 2010). Individuals with higher educational attainment and associated occupations are known to search over a much wider labour market and are seen to have a much greater spatial flexibility associated with, and driven by, career progression (van Ham *et al.*, 2001). This compares to other groups, particularly the more routine and manual occupations, who are generally more spatially restricted or tied to certain locales and local labour markets (van Ham *et al.*, 2001; Fielding, 2012).

Moreover, as mentioned above, those with greater educational and occupational attainment typically have access to greater financial resources, thus allowing individuals/households to mitigate the increased costs associated with longer-distance moves. Two important subgroups, who are somewhat separated from the underlying influence of the labour market, include those recently retired and university students. Whilst motivated by different things, both students and retirees are observed to form parts of distinctive migration streams commonly associated with moves over long distances and between particular types of origin and destination, be they university towns for students or amenity-rich environments for retirees (Champion *et al*., 1998; Smith, 2009; Fielding, 2012).

An additional selective factor that has been observed to further mediate the distance moved is that of household tenure. Most notably in the British context, attention has been paid to the restrictive nature of social housing provision where, through stringent local access rules, tenants of social housing find themselves particularly restricted in making moves between local authority districts and thus over longer distances (Boyle, 1995; Boyle and Shen, 1997; Champion *et al.*, 1998; Hughes and McCormick, 2000).

Similarly, though enacted through somewhat more subtle means, strong variations in distance have been observed when comparing different ethnic groups. Indeed, whether motivated by positive (e.g. maintaining familial ties or access to cultural amenities) or negative factors (e.g. reacting to discrimination or restricted opportunities), non-white ethnic groups tend to be more spatially concentrated in specific geographic locations, particularly in London but also in certain other large urban centres, than is the case for the more spatially dispersed majority white group

(Bailey and Livingston, 2005; Simpson and Finney, 2009; Stillwell and Hussain, 2010; Stillwell, 2010). All things equal, such variations in concentration and distribution can be expected to promote the variations in distance commonly observed for different ethnic groups. Whilst the examples given here are far from an exhaustive list of the selective individual/household characteristics observed to have influence on variations in the distance moved, these examples are useful in outlining important intervening obstacles and selective dimensions operating at the micro-level.

Of course, whilst such micro-level influences are of great importance, an ignorance of context, including factors that operate at the origin and the destination, leaves the analyst open to accusations of atomistic error as well as a failure to accommodate substantively important complexity (Chapter 4; Lee, 1966; Massey, 1990; Courgeau and Baccaini, 1998). Indeed, multilevel theories and aggregate level empirical research on migration certainly do suggest that simultaneous origin *and* destination residential contexts work to influence our ability and desire to move shorter or longer distances. Perhaps the most difficult task is first outlining what is meant by context, and second, what the *a priori* expectations about the role of specific elements of such contexts are.

In an important contribution to a multilevel theorisation of appropriate social and spatial context, Kearns and Parkinson (2001) define three broad spatial levels as central to what they would understand as a relevant milieu; running from what is termed the *home area* of familiarity and community, through to the *locality,* a wider area associated with everyday residential activities, and finally up to the *urban district or region* which is theorised to be the landscape of social and economic opportunities, operationalized through employment connections, leisure interests and social networks (Kearns and Parkinson, 2001; 2104). A general understanding of social and spatial context in this way, as a multilevel phenomenon, is certainly very useful when attempting to conceptualise how an areal push-pull theory operates in practice.

As was argued in the previous chapter, intertwined in the subjective assessment of one's residential satisfaction, the neighbourhood context (reflecting the home area and locality) has been identified as a potentially important predictor of mobility outcomes (Boehm and Ihlanfeldt, 1986; Lee *et al.,* 1994). Whilst in practice the evidence is rather mixed (Kearns and Parkes, 2003; Clark and Ledwith, 2006; Rabe and Taylor, 2010), the characteristics of one's neighbourhood are thought to play some role in conditioning both the desire to move, the ability to move and the decision of where to move to. For instance, levels of deprivation, ethnic heterogeneity and population stability have been noted as important drivers of neighbourhood desirability given their perceived role in influencing levels of social cohesion, crime, the physical environment and positive/negative social externalities (Galster and Killen, 1995; Harris, 1999; Feijten and van Ham, 2009; van Ham and Clark, 2009; Chapter 8).

In this way, the profile of the neighbourhood population can be expected to both push, particularly if it exacerbates the degree of residential dissatisfaction, or pull individuals/households where it offers enhanced opportunities to correct for residential dissatisfaction. Of course, individuals/households who have access to sufficient resources can act on such forces and do tend to move to neighbourhoods that reflect what are generally considered to be desirable living conditions (Clark and Dieleman, 1996). However, as with individual/household characteristics, the neighbourhood is also thought to act as a selective mechanism where, particularly for the most deprived neighbourhoods, those without sufficient resources are restricted in their opportunities to act on mobility desires and particularly to move over sufficient distances in order to reach the more desirable neighbourhoods (Galster and Killen, 1995), neighbourhoods, in the British context, that are often spatially segregated (Dorling and Rees, 2003).

Beyond the neighbourhood, important factors are thought to operate at the broader regional (macro) level; regional economic robustness and differential lifestyle opportunities are said to influence the attractiveness of different locations, and are thus used to explain many of the clear and persistent patterns of residential mobility at the macro-level. For instance, the pivotal role of London in the national migration system is well documented (Fielding, 1992; Champion, 2008; Duke-Williams and Stillwell, 2010). Whilst the capital tends to attract young and usually well-educated adults from across the country, largely for employment but also lifestyle reasons, it is by far the largest net loser to internal residential movement. Whilst London has continued to grow over the last decade or so, much of this observed growth has been driven by a combination of strong natural increase and significant net immigration from outside the UK (Champion, 2008). However, London is not alone in losing significant numbers of people to other parts of the country. Indeed, over recent decades the dominant characteristic of internal residential movement has been that of urban-rural shift and counterurbanisation (Rees and Stillwell, 1992; Champion, 2005b; Dennett and Stillwell, 2008), a phenomenon that has been recognised by many to be driven by place-based preferences, an improvement in the ease of commuting, a growing proportion of pleasure-seeking retirees, and a widespread attachment to the supposed 'rural idyll' (Champion *et al.*, 1998; Mitchell, 2004; Fielding, 2012). As Champion *et al.* (1998) suggest, "*mythical or otherwise, the 'rural idyll'* […] *would seem to be providing the cognitive framework within which many people are, consciously or subconsciously, making their decisions to join the urban exodus"*. Of course, whilst they are significantly smaller in their scale, there are important counter-streams with, as alluded to above, a persistent movement of young people away from smaller towns and rural areas towards the cities (Stockdale, 2004) and, in particular, increasingly large student flows into university towns and cities (Champion, 2005a; Smith, 2009).

In summary, then, the key theoretical and empirical work suggests that factors operating simultaneously at the origin and the destination, from the micro through to the macro, combine to produce multilevel variations in origin-destination distance. With this in mind, the data and measures used in the analysis are now considered, before a suitable modelling framework appropriate for dealing with such complexities is defined.

9.3 Data and measures

As in the last chapter, for the analysis presented here a subsample drawn from the pooled analytical sample introduced in Chapter 6 is used. With the defining parameters discussed below, the analysis of distance moved is based on an analytical sample size of 26,688 individual residential migrants in England and Wales[15]. A migrant is defined as an individual who has moved to his/her current postcode

---

[15] This migrant subsample represents 7.65% of the pooled (England, Wales & Scotland) analytical sample (n = 348,953) used in the previous model based benchmarking exercises in Chapter 6.

address (destination) within the three years prior to survey completion and who has additionally provided a full postcode address for his/her previous residence (origin).

The benefit of having detailed postcode identifiers is twofold: firstly the area of origin and destination can be defined in far greater detail than is allowed for in alternative sources such as the 2001 Census I-SAR (where only Government Office Region geography is provided at the origin); and second, it is possible to calculate straight-line distance as an unbanded continuous variable, measured directly from origin postcode grid reference to destination postcode grid reference. By limiting the migration interval to three years, the potential for distortions associated with time-varying characteristics is reduced while at the same time the generation of a large sub-sample with good geographic coverage is made possible, the latter being of particular importance given the focus on spatial distribution and context.

However, it should be noted that certain peoples' characteristics may well change more rapidly than others over this three year period; for instance, young people when compared to the more settled older population, and therefore measurement error pertaining to non-stationarity at the micro-level, is likely to be greater for the former. Moreover, the ability to make certain micro-level inferences is limited somewhat given that all such characteristics are measured at the time of the survey and thus the destination only; unfortunately, therefore, it is not possible to explore relationships between the individual/household at the beginning of the move and at the end of the move. Finally, the omission of migrants to and from Scotland is motivated by concerns for sparsity in the sample for particular regions.

The micro-level characteristics obtained for analysis are motivated by discussions here and in previous chapters and reflect those that are deemed to be the most important predictors of variation in distances moved. Again, measured at the time of survey completion only (i.e. the destination), these are: age, sex, ethnicity, marital status, annual gross household income, household tenure, occupational class and educational attainment. Indicator variables to adjust for potential confounding effects associated with the small temporal variations, these being the differences in duration at the current address and the year of survey completion, are also included.

Based on the discussions above, it is suggested that the specially designed measure of small area profiles, the 2001 Output Area Classification (OAC) (Vickers and

Rees, 2006; 2007), provide the best option for operationalising the immediate neighbourhood context. As was briefly mentioned in Chapter 6, the OAC is a hierarchical geodemographic classification of small areas into groups based on the similarity of the demographic, socio-economic and housing profile of their residents; all of which are factors raised in the literature as being potentially important factors for influencing neighbourhood attractivenesss and more general residential satisfaction. To represent the macro-regional level, the system of city regions is used again. Through the employment of city regions at the macro-level, it is possible to get a direct measure of the spatial distribution of migrants' origins and destinations and, more specifically, to explore this in relation to important macro processes linked to population density (the urban/rural component) and the spatial economic system, for which the geography of city regions was designed to represent.

9.4 Modelling framework and analysis

Given the inherent substantive interest in a multilevel theory and analysis of variations in origin to destination distance migrated, a more advanced cross-classified multilevel framework is chosen (Chapter 4). Building on the models of the previous chapter, the cross-classified model allows for the observation of not only the micro-level drivers of variation in distance moved, but also the remaining meso/macro contextual variations in distance moved at the origin and the destination, having controlled for the micro-level composition. All things being equal, if there are remaining contextual effects at the origin and the destination, a degree of spatial heterogeneity can be expected to be observed, wherein certain areas send/receive (push/pull) migrants over longer/shorter distances than others.

Moreover, from a statistical modelling perspective, if both origin *and* destination factors are found to contribute significantly to variations in the outcome, the modelling of only one such context/classification, the origin *or* the destination, would fail to account for possible confounding effects associated with an underspecified model (Fielding and Goldstein, 2006). For example, if one only includes the multilevel context of the destination in the model, there is a risk of overstating the importance of destination as a source of variation at the expense of the origin; that is, you fail to disentangle variation between different destination contexts from that which may be more accurately estimated as variation between different origin contexts.

Therefore, drawing on the classification notation of Browne *et al*. (2001), the cross-classified model that forms the basis of the analysis presented here can be specified as follows:

$$\begin{aligned}
&y_i = (X\beta)_i + u^{(5)}_{orig\ region\ (i)} + u^{(4)}_{orig\ neighbourhood\ (i)} + u^{(3)}_{dest\ region\ (i)} + u^{(2)}_{dest\ neighbourhood\ (i)} + e_i,\\
&orig\ region(i) \in (1, \dots, J_5), \quad orig\ neighbourhood(i) \in (1, \dots, J_4),\\
&dest\ region(i) \in (1, \dots, J_3), \quad dest\ neighbourhood(i) \in (1, \dots, J_2),\\
&u^{(5)}_{orig\ region\ (i)} \sim N(0, \sigma^2_{u(5)}), \ u^{(4)}_{orig\ neighbourhood\ (i)} \sim N(0, \sigma^2_{u(4)}), \ u^{(3)}_{dest\ region\ (i)} \sim N(0, \sigma^2_{u(3)}),\\
&u^{(2)}_{dest\ neighbourhood\ (i)} \sim N(0, \sigma^2_{u(2)}), \ e_i \sim N(0, \sigma^2_{e_i}), \ i = 1, \dots, N,
\end{aligned} \tag{9.1}$$

where y_i is the natural logarithm of origin to destination distance in kilometres (km) for the ith migrant of N migrants in total, itself a function of $(X\beta)_i$ which represents the fixed part of the model, a vector of X explanatory variables whose parameters, β, are again referred to as 'fixed parameters' and, for this analysis, are all measured at the migrant level. Within this vector the first element, the constant (β_0), takes a value of one for each migrant i and, when all explanatory variables are held at their base (i.e. 0), provides the estimated mean distance migrated from origin to destination across all origin and destination neighbourhood types and regions. For the random part of the model, $u^{(5)}_{orig\,region\ (i)}$ is the additional effect of migrant i's region at origin, $u^{(4)}_{orig\,neighbourhood\ (i)}$ is the additional effect of migrant i's neighbourhood at origin, $u^{(3)}_{dest\ region\ (i)}$ is the additional effect of migrant i's region at destination, $u^{(2)}_{dest\ neighbourhood\ (i)}$ is the additional effect of migrant i's neighbourhood at destination with e_i representing the remaining migrant level residual error.

As in the more traditional strictly hierarchical approaches, all parameters in the random part of the model are assumed to follow a normal distribution with a mean of zero and a constant variance and, additionally, are assumed to be independent across classifications. To aid interpretation of the model design, a classification diagram (Figure 9.1) is included; the classification notation does not make clear the multilevel structure of the data. Thus, for the purpose of clarification, each box in

Figure 9.1 represents a set of units and each arrow suggests the nesting of one set of units into the other, such that individual migrants are simultaneously nested within origin and destination hierarchies.

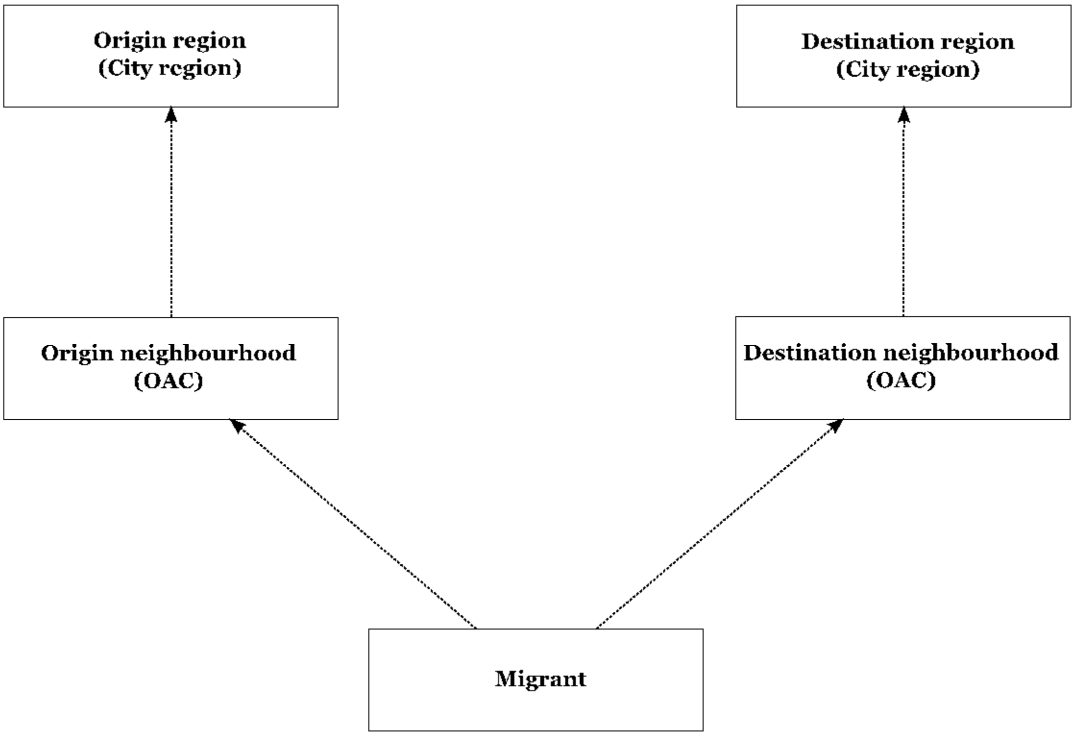


Figure 9.1. Classification diagram of the origin and destination cross-classified multilevel model

Because of the complex structure of the cross-classified model and the relatively small number of city region units, MCMC estimation is used, again providing a more efficient and robust estimation to the maximum likelihood based alternatives (Browne and Draper, 2006; Browne, 2012). All models are estimated using the MLwiN 2.28 software (Rasbash *et al.*, 2013). As with Chapter 8, initial parameter starting values are based on maximum likelihood methods with model convergence assessed following the good-practice recommendations of Draper (2006) and Jones and Subramanian (2013). For the fully specified cross-classified model, a burn-in of 500 iterations is followed by a monitoring chain of 55,000 simulations. To aid with the mixing of MCMC parameter chains, the parameter expansion method of hierarchical centring is used (Browne *et al*., 2009; Browne, 2012).

In terms of the modelling strategy, three initial null (constant only) models with random intercepts are specified, Model 1 with neighbourhood (level-2) and regional (level-3) contexts defined at the origin; Model 2 with neighbourhood (level-2) and regional (level-3) contexts defined at the destination; and finally, Model 3 where the individual (level-1) is nested within the two simultaneous hierarchies, an origin (level-2 & level-3) and destination (level-2 & level-3) neighbourhood/city region cross-classification. Specifying the three null models allows for the partitioning of the total variability in distance across the different levels/classifications. For instance, before accounting for the compositional differences between areas, the null models can be used to inspect whether there is indeed any evidence for variation in distance attributable to differences between city regions and/or differences between geodemographic neighbourhood types within city regions. As mentioned, this can be done for the origin and destination separately and as a cross-classification of the two, where, in the latter case, there is the advantage of being able to explore the relative contribution of the multilevel contexts at the origin net of the relative contribution of multilevel contexts at the destination, or *vice versa*. Following this, compositional differences between areas are accounted for by introducing the individual/household level covariates into the fixed part of the cross-classified model. Of course, whilst the influence of micro-level covariates on variations in origin to destination distance is of interest in itself, having controls for the compositional effects is additionally beneficial in that one is better able to reliably identify which areas send/receive (attract/repulse) migrants over longer or shorter distances.

9.5 Model results

9.5.1 Null model results

Table 9.1 shows the results of the three null models for migrants nested within their origin hierarchy (Model 1), migrants nested within their destination hierarchy (Model 2) and migrants nested within a unified cross-classification of their origin and destination hierarchies (Model 3). For the strictly hierarchical models, the majority of variation is found between individuals, as we would expect; however, there is some evidence of non-individual variation. Indeed, the within-city-region-between-neighbourhood variation is estimated to account for around 4% of the total variation in distance migrated[16] in both the origin and the destination models, with the between-city-region differences observed to account for around 2% of the total variation in each hierarchical model. However, as has been argued above, the casting of the model as a strict hierarchy has serious statistical and substantive analytical limitations, both of which can be expected to have serious implications for the reliability of the modelled results and subsequent substantive interpretations.

Table 9.1. Null models for migrant origin to destination distance (log km)

	Model 1: Null origin		Model 2: Null destination		Model 3: Null cross-classified	
	Beta	**S.E.**	**Beta**	**S.E.**	**Beta**	**S.E.**
Fixed Part						
Constant	1.254	0.052	1.339	0.055	1.233	0.208
Random Part						
$\sigma^2_{u(5)}$ Origin city region variance	0.069	0.023			0.728	0.205
$\sigma^2_{u(4)}$ Origin neighbourhood variance	0.151	0.018			0.129	0.017
$\sigma^2_{u(3)}$ Destination city region variance			0.081	0.028	0.672	0.187
$\sigma^2_{u(2)}$ Destination neighbourhood variance			0.155	0.019	0.067	0.011
σ^2_e Individual migrant variance	3.468	0.031	3.498	0.031	3.187	0.028
DIC	109228.061		109459.595		107187.676	
d.o.f	302.608		305.670		520.564	
Units: Origin city region	33				33	
Units: Origin neighbourhood	621				621	
Units: Destination city region			33		33	
Units: Destination neighbourhood			621		621	
Units: Individual migrant	26,688		26,688		26,688	

When the model is specified as a cross-classification of origin and destination context model fit is considerably improved (the DIC in Model 3 is more than 2,000 units smaller than in Models 1 and 2), while the change in the way in which total variation is partitioned between the different classifications is equally noticeable. The between-individual differences remain as the primary source of total variation (67%), however, the total macrogeographical variation, that is, the total macro origin and destination contexts combined, is now estimated to account for a substantial

16 The origin value, for example, is calculated as: $\sigma^2_{u(4)}/(\sigma^2_{u(5)} + \sigma^2_{u(4)} + \sigma^2_e)$

29% of the total variation in distance migrated[17] (where 15% is sourced at the origin and 14% at the destination).

However, before any exploration of potential patterning to the observed macro-level variation is made, it is important consider the micro-level factors (individual and household) and, in doing so, allow for the socio-demographic composition of such areas to be taken into account. Indeed, without allowing for their composition, it is impossible to conclude that the quite substantial variations found at the macro-level are the result of place-based differences, as opposed to a mere reflection of simple variations in their differential composition.

9.5.2 Fully specified cross-classified model

As expected, the introduction of the micro-level covariates into the fixed part of the cross-classified model (Model 4, Table 9.2) is reflected by a further, and again very considerable, reduction in the DIC. The estimated grand mean distance moved $(\exp(\beta_0))$, that is the distance of the typical migrant across all neighbourhoods and all regions, is predicted to be 3.34 km, matching closely with estimates based on both previous census data and recent residential estate agency records (Boyle and Shen, 1997; Hamptons International Ltd., 2013). An inspection of the random part of Model 4 suggests that the inclusion of the micro-level covariates has helped to reduce the unexplained variation at the migrant level, the migrant level residual error, by approximately 3.4% while at the same time their ability to control for the composition of areal units has dramatically reduced what were already very marginal effects for neighbourhood type (4% in Model 3), by 42.6% and 44.8% at the origin and destination respectively. Yet even after controlling for micro-level factors and neighbourhood type, at both the origin and the destination, considerable differences between the city regions remain evident (28% of the remaining residual variation in Model 4 lies at the combined macro-level). However, before a more in-depth exploration and interpretation of the macro-level variation is attempted, it is perhaps important to first summarise the results of the fixed part.

---

[17] Calculated as: $(\sigma^2_{u(5)} + \sigma^2_{u(4)})/(\sigma^2_{u(5)} + \sigma^2_{u(4)} + \sigma^2_{u(3)} + \sigma^2_{u(2)} + \sigma^2_e)$

9.5.2.1 Fixed part results

Broadly speaking, the results from the fixed part of the model (Table 9.2; Figure 9.2) reflect the importance of many of the micro-level associations posited above. Taken as a whole, it would seem fair to agree with the assertion made earlier, namely, that residential movements over longer distances are largely the preserve of a group whose social and economic characteristics are indicative of relative affluence. For instance, of the various individual/household factors that were taken into account, many of the largest differentials in distance can be found to relate to specific variations in migrants' socio-economic status, including for example differences in: educational attainment, occupation, annual household income and housing tenure. Beyond this, however, certain additional socio-demographic differences can be seen to play some role in predicting variations in origin to destination distance; although, aside from one or two examples, their influence is less pronounced when compared to the socio-economic factors. Nevertheless, for a more extensive and better revealing insight of the micro-level dynamics, it is important to provide a detailed breakdown of some of the key individual/household covariates shown in Table 9.2 and Figure 9.2, the latter of which has had its axes scaled to allow for a better comparison of the relative size of the effects associated with each fixed part covariate. Again, as was mentioned in relation to the 95% credible intervals for the probabilities in Figure 8.3, estimates in Figure 9.2 are based on predictions for an otherwise typical individual in the typical origin and destination neighbourhood and city region.

In terms of the ethnic group differences, there is very little separating the average distance travelled by the Black and Other ethnic groups from that travelled by the reference group, the White majority. However, there does appear to be a statistically significant and substantively rather interesting pattern for the Asian ethnic group (Indian, Pakistani and Bangladeshi), wherein the average distance migrated for this group is considerably shorter than that of the others. This pattern has been observed in previous analyses of census data (for instance, Finney and Simpson, 2008) and is perhaps reflective of the concentrated spatial distribution of particular Asian minority ethnic groups in particular parts of England and Wales' metropolitan centres (Simpson and Finney, 2009; Stillwell, 2010).

The effects of differing marital status, which for lack of any better alternative is used here as a rather crude proxy for relational dependency and cohabitation, does not suggest any particularly striking influence over variations in distance migrated. That said, those recorded as currently divorced/separated are estimated to have migrated marginally shorter distances, on average, than those in the reference category, married. Yet whilst there is no measure of whether individual migrants have dependent children, or whether the measured migration follows their relationship dissolution, previous research by Feijten and van Ham (2007) does suggest that the separated are likely to stay relatively 'local' so as to maintain their location-specific capital and social networks, and, perhaps most importantly, the relationship with any dependent children they may have.

With respect to the migrant's age, a rather complex relationship is at play, a relationship that is itself inextricably linked to one of the key socio-economic characteristics outlined above. Indeed, when measured as a main effect, that is, free of any interaction effects, an increase in migrant age is found to have a positive linear relationship with the distance migrated (Figure 9.2). However, when the migrant's age is interacted with their housing tenure type (the main effects of which are also given in Figure 9.2), a far more interesting and substantively revealing relationship is displayed. Where the estimate for age in Table 9.2 now represents the estimate for age when the migrant is a homeowner, the direction of the relationship between age and distance migrated is found to be very different depending on which tenure group the migrant is a member.

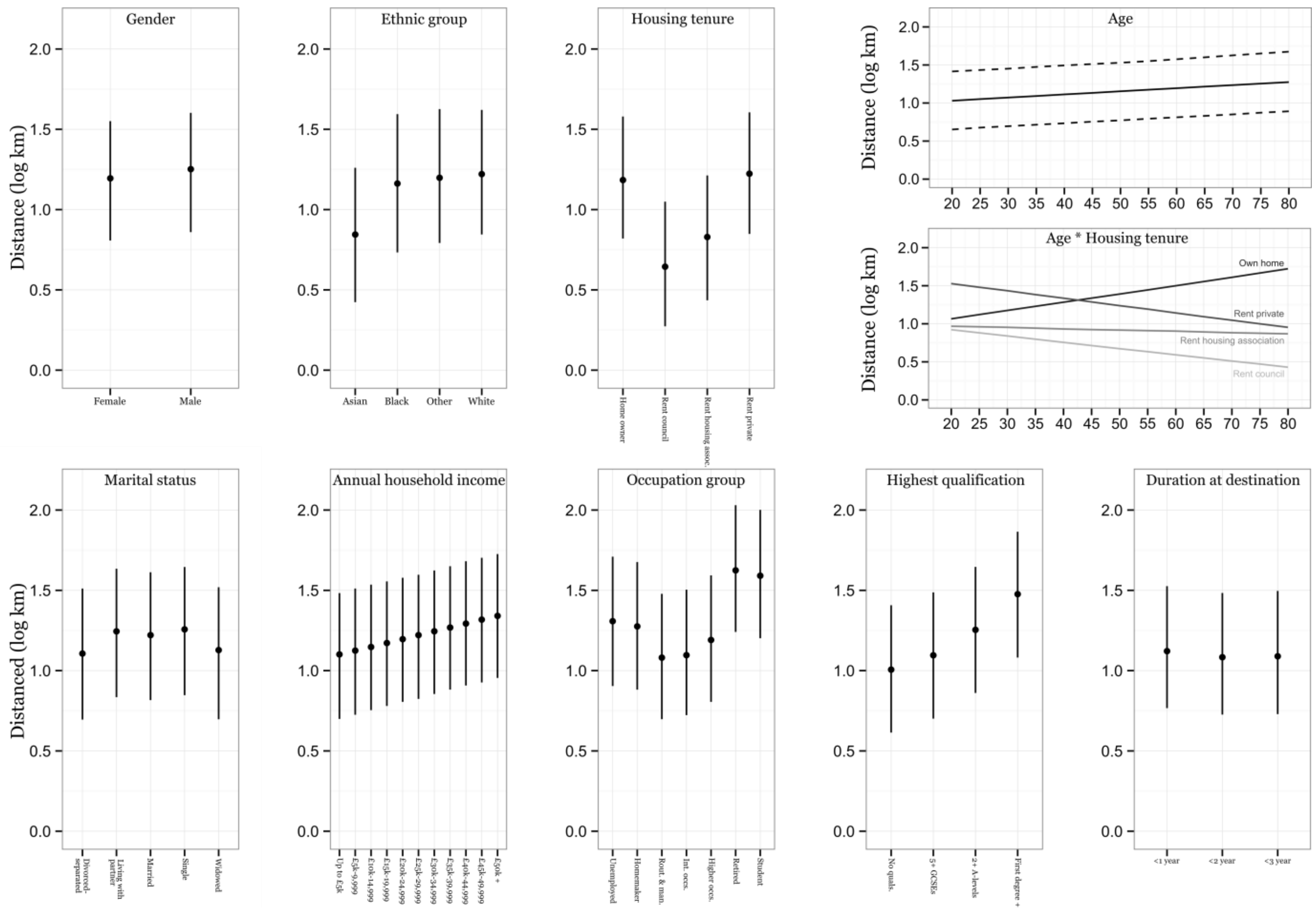


Figure 9.2. Model 4 fixed part predictions and 95% credible intervals

Table 9.2. Multilevel cross-classified model estimates for origin to destination distance (log km)

	Model 4: Full cross-classified			
	Beta	**S.E.**	**CI (2.5%)**	**CI (97.5%)**
Fixed Part				
Constant	1.208	0.198	0.817	1.599
Age (centred at 40)	0.011	0.001	0.008	0.014
Gender (1 = Male)	0.057	0.024	0.010	0.104
Ethnic group (ref = White)				
Asian	-0.380	0.076	-0.530	-0.231
Other	-0.024	0.074	-0.169	0.122
Black	-0.059	0.091	-0.236	0.120
Marital status (ref = Married)				
Single	0.037	0.035	-0.031	0.106
Living with partner	0.025	0.031	-0.036	0.085
Divorced/separated	-0.111	0.037	-0.184	-0.038
Widowed	-0.093	0.066	-0.222	0.036
Highest qualification				
Linear polynomial	0.349	0.028	0.294	0.404
Quadratic polynomial	0.068	0.024	0.020	0.115
Annual household income (linear polynomial)	0.255	0.056	0.145	0.365
Occupation group (ref = Intermediate)				
Retired	0.528	0.052	0.428	0.629
Student	0.498	0.063	0.373	0.622
Homemaker	0.177	0.042	0.094	0.259
Unemployed	0.209	0.061	0.089	0.328
Routine & manual	-0.017	0.042	-0.099	0.066
Higher managerial administrative & professional	0.091	0.031	0.030	0.152
Housing tenure (ref = Home owner)				
Rent private	0.052	0.032	-0.012	0.116
Rent council	-0.525	0.041	-0.605	-0.445
Rent housing association	-0.347	0.047	-0.440	-0.254
Duration at destination (ref = <1 year)				
<2 years	-0.038	0.027	-0.090	0.014
<3 years	-0.033	0.027	-0.085	0.019
Data set (ref = January 2005)				
January 2006	-0.130	0.034	-0.197	-0.062
January 2007	-0.108	0.025	-0.157	-0.060
*Housing tenure*Age*				
Rent private, Age(40)	-0.021	0.002	-0.025	-0.017
Rent council, Age(40)	-0.019	0.002	-0.024	-0.015
Rent housing association, Age(40)	-0.013	0.003	-0.018	-0.007
Random Part				
$\sigma^2_{u(5)}$ Origin city region variance	0.657	0.183	0.387	1.093
$\sigma^2_{u(4)}$ Origin neighbourhood variance	0.074	0.012	0.052	0.099
$\sigma^2_{u(3)}$ Destination city region variance	0.605	0.168	0.357	1.010
$\sigma^2_{u(2)}$ Destination neighbourhood variance	0.037	0.008	0.023	0.054
σ^2_e Individual migrant variance	3.080	0.027	3.027	3.134
DIC	106201.116			
d.o.f	444.019			
Units: Origin city region	33			
Units: Origin neighbourhood	621			
Units: Destination city region	33			
Units: Destination neighbourhood	621			
Units: Individual migrant	26,688			

Contrary to the relationship shown by the simple main effect, *ceteris paribus*, a single unit increase in age is actually found to be negatively associated with distance moved for those migrants who are recorded as being renters at the destination; this is likely to relate to a broader socio-economic dimension where private renting during your early adulthood is generally the norm, and is less restricting of mobility than other tenure groups, whereas in older age, private renting and the associated insecurities can more accurately reflect a degree of relative deprivation. As expected, individuals/households who are renting from a local authority, and to a lesser extent those renting from a housing association, are associated with moves over shorter distances (Boyle and Shen, 1997; Hughes and McCormick, 2000). This suggests that the barriers often mentioned with regards to social housing are still relevant factors in restricting the distances travelled by migrants in England and Wales. On the flipside, private renters and homeowners are associated with relatively longer-distance moves; however, of the two, the tenure type associated with the longest distances varies with age. Whilst private renters are found, on average, to be the migrants moving over the longest distances in the younger age groups, the propensity for longer-distance moves reduces year-on-year until, at approximately 40-45 years of age, home owning migrants take over as the group most likely to migrate over relatively longer distances. Whilst those in the older age groups are more likely to be free from occupational and familial (dependent-child) constraints, homeowners in the older age groups are also likely to be relatively more (asset-) affluent, at least when compared to other tenure groups. Consequently, if a long-distance move is the desired outcome, perhaps for reasons linked to retirement and the pursuit of residential milieu that better reflect their lifestyle and consumption desires, a combination of such factors could be expected to make this group particularly able when attempting to overcome the intervening obstacles commonly associated with longer-distance migrations.

Beyond the housing tenure type of the migrant, other micro-level socio-economic characteristics are found to be deserving of more detailed attention. Estimates associated with the migrant’s annual household income and educational attainment (highest qualification) present the directional relationships found in many previous theoretical and empirical analyses. Both variables are measured using orthogonal polynomials. Making use of this parameterisation, it is clear that greater levels of household income are positively, and linearly, associated with greater distance. Moreover, greater levels of educational attainment are also found to be positively associated with greater distance. However, this time the relationship is curvilinear suggesting that the magnitude of this association increases as we move up the scale. As such, in common with the previous findings outlined above, individuals with access to high household income and higher levels of education (particularly degree level and above) are significantly more likely to have migrated over longer distances than those in the lower income brackets and those with poorer educational attainment.

Whilst levels of household income and educational attainment are found to be very significant factors in determining variations in distance moved, the greatest effects are found amongst the different occupational groups. For those in paid employment, there is little difference in the mean distance travelled, although for what small differences do appear, the trend of increasing distance being linked to higher occupational groups is visible (Figure 9.2). Moreover, there is some evidence of increased distance being associated with those who are currently unemployed and those who describe themselves as homemakers. That said, however, the single largest estimated effects are found for the retired and student groups. As mentioned above, both groups have been observed to form well-known and distinctive migration streams which often entail residential moves over longer distances (see Section 9.2).

Finally, the inclusion of the indicator for the year of survey completion appears to be somewhat justified with a relatively small, yet statistically significant, differential effect detected. However, the indicator for duration at the destination is found to be of very little substantive or statistical relevance. Both indicators were included due to concerns surrounding the potential for distortions associated with the small temporal variations in the analytical sample.

9.5.2.2 Random part results

Each random part classification is found to have a statistically significant contribution to the residual variation in origin to destination distance (Table 9.2). However, from a substantive point of view, the remaining within-city-region-between-neighbourhood-type variation is found to be quite minor. Instead, the place-based differentials of noticeable size and interest are found to operate at the macrogeographical level, where 28% of the remaining variation is located. Having controlled for the compositional influences at the micro- (individual/household) and meso- (neighbourhood type) levels, there appears clear evidence of systemic spatial heterogeneity in place based attractiveness, wherein certain macrogeographical areas send/receive (attract/repel) migrants over significantly longer or shorter distances than would otherwise be expected.

Indeed, the conditional 95% coverage interval for the origin macro regions[18] suggests that city regions which lie at the 97.5th percentile of the distribution send the typical migrant a distance of 16.40km whereas for an origin region at the bottom 2.5th percentile of the 'sending' distribution, that same migrant is estimated to move just 0.68km. Similarly, for the 'receiving' (destination macro regions) distribution, the typical migrant whose destination is at the top 2.5% is estimated to have moved a distance of 15.37km while those whose destination is at the bottom 2.5% are found to have moved 0.73km. Yet whilst such statistics are useful in demonstrating the existence of considerable macro heterogeneity, they are of little help when attempting to draw out any underlying patterns to the variation. Consequently, where the dashed lines represent the estimated grand mean distance (β_0), i.e. the average distance moved across all residential migrants, all neighbourhood types and all regions, Figure 9.3 plots the modelled origin and destination city region residuals (differentials) against one another and in doing so uncovers the types of macrogeographic regions that lay at the extremes.

Indeed, drawing on Figure 9.3, a clear systemic pattern to the heterogeneity emerges, one that closely reflects a process of urban-rural shift and counterurbanisation observed in previous aggregate-level studies of the UK. As a general trend it is apparent that the major metropolitan cores (particularly London core), and to a certain extent their surrounding satellite towns and cities (i.e. metropolitan rest), send migrants over longer distances and attract migrants over shorter distances than the national average. Conversely, for the macro regions described as "coast and country" the opposite pattern is observed, with such regions being seen to pull migrants over longer distances and send them considerably shorter distances. Therefore, having controlled for individual and neighbourhood composition within the city regions, a persistent pattern of strong urban repulsion, with urban cores pushing migrants over considerably longer distances, and an equally strong rural/coastal attraction, where such areas are seen to pull migrants over significantly longer distances, is observed when compared to the national average.

---

18 Calculated as: $(-1.96\sigma_{u(5)}, +1.96\sigma_{u(5)}) = (-1.96\sqrt{0.657}, +1.96\sqrt{0.657}) = (-1.59, +1.59)$

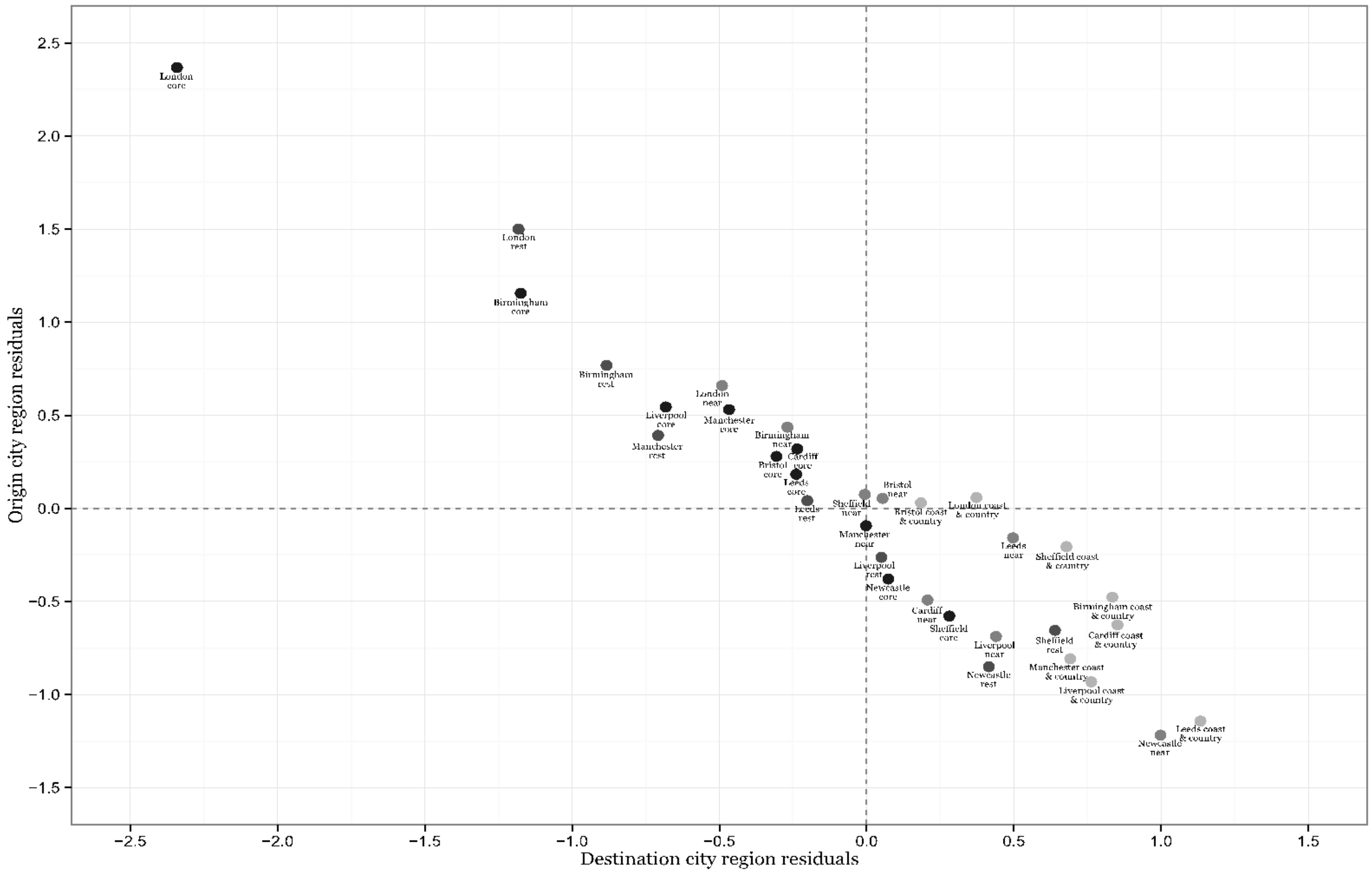


Figure 9.3. Model 4 origin and destination city region residuals (log km scale) *N.B grayscale used to differentiate the urban-hierarchy*

Whilst longstanding neo-classical economic theories would suggest a pull towards the major metropolitan cores, for employment/labour market reasons (see for instance, Sjaastad, 1962), a growing volume of evidence presents place-based attractiveness to be increasingly driven by desires for improved lifestyle and consumption opportunities, and therefore towards the more rural/coastal amenity-rich destinations (Boyle and Halfacree, 1998; Champion, 2005b; Stockdale, 2010; Morrison and Clark, 2011). Indeed, beyond the significant contribution associated with the major economic restructuring of the 1970s, itself an important driver of (uneven) decentralisation of employment opportunities away from the old metropolitan cores and towards new nodes of economic growth (for instance, the M4 and M11 motorway corridors) (Dunford and Fielding, 1997; Fielding, 2012), an improvement in the ease of travel and communications has enabled an increasing disconnect between one's place of work and one's place of residence to emerge.

Empirical work has shown recent (working-age) in-migrants to the surrounding peri-urban and rural regions to be, on average, more likely to commute over significantly longer distances and durations (Boyle *et al.*, 2001; Axisa *et al.*, 2011). Moreover, in a comparative analysis of commute data from the 1991 and 2001 Censuses, Nielsen and Hovgesen (2007) suggest a strong growth in longer-distance commuting to have occurred, a growth which, they argue, is explained by a combination of the deconcentration of populations and jobs as well as a general socio-cultural preference for rural living. Of course, as has been alluded to above, place-based attraction and repulsion, and the ability to act on these things, are different for different people. For example, in contrast to the dominant theme of counterurbanisation in the UK (Rees and Stillwell, 1992; Champion, 2005b; Dennett and Stillwell, 2008), students and young professionals, for a variety of largely labour market and career relevant factors, are known to form a significant counter-stream *towards* the larger urban centres, and particularly London (Fielding, 1992; Fielding, 2007). However, when focussing on the residential mobility system as a whole, it would appear fair to agree with Morrison and Clark (2011: 1949) in suggesting that, whilst continued employment is of paramount importance for the majority of working-age migrants, in countries where employment opportunities are relatively abundant both spatially and in absolute terms, "*migration to enhance employment gives way to movement to enhance other goals*".

9.6 Summary and conclusions

This chapter has presented an analysis of variations in the distance over which migrants travel when moving to new residential locations in a manner that better reflects the realistic multilevel complexity associated with such a phenomenon. Whilst major theoretical contributions to explaining residential movement have emphasised the importance of processes and characteristics that operate simultaneously across different levels, at both the origin and the destination, the majority of existing research has struggled to confirm this empirically. However, drawing on a series of multilevel statistical models, it is suggested that the analysis presented in this chapter goes some way to addressing this shortfall.

As with all models, through practical necessity, it is only possible to gain partial insights into the true reality of migration behaviour. To exemplify, in this research we are inherently restricted to focussing on the direct and independent (additive) effects of context at the origin and destination. However, theoretically we can expect the evaluation of (pull) factors at the destination area to interact with, and indeed be conditioned by, the migrant's further interpretation of (push) factors at the area of origin and *vice versa*. Methodologically, it is possible, to some extent, to account for this aggregate inter-relationship within a multilevel statistical framework through the addition of a random interaction classification (Goldstein, 2011). Unfortunately, for the analysis of variations in origin to destination distance, the spatial nature of the problem makes the addition of a random interaction classification, between a geographic place of origin and a geographic place of destination, a self-fulfilling prophecy. That is, the interaction between the migrant's place-specific origin and destination will, by its very nature, perfectly predict their distance travelled. However, a useful direction for future research might be to explore a random interaction classification, but for origins and destinations that are defined purely on (geodemographic) area type.

The findings for England and Wales suggest that the inclusion of micro-level influences as well as wider origin and destination contextual settings are necessary for a more statistically robust and substantively complete understanding of variations in origin-destination distance, and particularly the role of place-based attractiveness. As expected, residential moves over longer distances are found to be strongly associated with individuals/households who have access to particularly high levels resources, both social and economic. Thus, relatively speaking, those moving the longest distances tend to be those who are highly educated, have access to greater annual household income, are older homeowners and, free from the spatial constraints of employment, are retired. It follows therefore that, *ceteris paribus*, migrants typically moving the shortest distances tend to be low paid, have very basic educational attainment, are member to an Asian ethnic minority group, and rent from a local authority or housing association.

Whilst the micro-level determinants are of clear substantive and empirical relevance, significant spatial heterogeneity, particularly at the macrogeographic level, is observed. When cast as a cross-classified origin and destination model, a clear pattern of urban-rural shift emerges, wherein, on average, a typical residential migrant is pulled over significantly longer distances towards rural/coastal (amenity-rich) city region destinations and, at the same time, is pushed significantly longer distances if the origin city region happens to be a metropolitan core (or metropolitan rest). Thus, by incorporating measures for residential context at the area of origin and destination, it is possible to get a handle on the relative importance of additional place-based attractiveness for enacting variations in the distance over which people move. As such, with the dominant pattern of counter-urbanisation apparent, the findings would appear to add further weight to the argument that residential movement is becoming increasingly a means through which people attempt to satisfy their leisure, lifestyle and consumption desires, a situation which has driven, and apparently continues to drive, the quite significant redistribution of the population towards the amenity-rich environments of England and Wales' coast and countryside (Champion *et al.*, 1998; Fotheringham *et al*., 2000; Morrison and Clark, 2011; Fielding, 2012).

Chapter 10
Conclusions

10.1 Introduction

This thesis set out to address a set of detailed research questions (Subsection 1.2) that provided the rationale for the following overall project aims:

a.) to investigate individual and place variations in residential mobility and immobility in Great Britain using commercial data and official statistics;

b.) to explore the effects of duration of residence, and additional cross-level interactions, on the propensity for future residential moves; and

c.) to examine the potential variations in migrant origin to destination distance according to individual and place-based characteristics.

Following the completion of the nine preceding chapters, it is argued here that the thesis has been successful in addressing all three aims through the combined use of detailed and geographically extensive microdata, appropriate statistical methods, and well-informed micro and macro theory. This chapter concludes the thesis by first summarising the research findings (Subsection 10.2), with a specific focus on identifying how the six specific research objectives set out in Table 1.1 have been achieved and in turn, the overall project aims have been met and the research questions answered. Some final reflections on the general approach used and the potential for future research are offered in Subsection 10.3.

10.2 Summary of research findings

In Chapter 1 it was argued that the overall research aims would be best met through addressing a series of specific research objectives. Consequently, this subsection will spell out the six research objectives in turn and demonstrate how each was met through the work presented in the prior chapters of this thesis.

I. To explore and review the existing literature associated with individual and area demographic, socio-economic and behavioural/lifestyle dimensions of population movement in GB and provide the theoretical and empirical context for the analyses herein

Through the combination of reviews in Chapters 2, 7, 8 and 9, the thesis provides a necessary theoretical and empirical background to individual and place variations in population movement in GB. The analysis of residential movement, defined here in its most general terms as a move from one location to another regardless of the distance travelled (Subsection 2.2), is argued to be of huge social and economic importance as a phenomenon with the potential to transform the character and structure of populations, and in some cases affect real change to the social, cultural, physical and economic characteristics of an area. Indeed, at the micro and macrogeographical levels, the measurement, analysis and understanding of what drives the flows of different people between different places is key for informing policy development, resource allocation and service delivery at the local and national scales (Rees *et al.*, 2009). However, as has been argued throughout the substantive chapters, the traditional dichotomy between micro and macro approaches to population migration analysis has often resulted in a limited empirical interrogation of many longstanding theories that are more appropriately defined as multilevel in nature.

Indeed, since the early work of Thomas (1938) and Rossi (1955), the decision to change residence has been widely accepted to be a utility-maximising behaviour performed, within the context of relative social and economic constraints (see Chapter 2), by individuals, either independently or collectively within a household, reacting to disequilibrium between the current residential environment and a perceived environment elsewhere (Bartel, 1979; Clark and Dieleman, 1996; Clark, 2013). Furthermore, this evaluation of the current and perceived residential environment in different places, and following this, the desire to move and the decision of where to move to, is considered to not only be influenced by factors operating at the individual/household level, for instance those multiple factors associated with life-course transitions and events discussed in Chapters 7 and 8, but also by factors more appropriately defined as contextual in type (Lee, 1966; Massey, 1990; Kearns and Parkinson, 2001; Sampson *et al.*, 2002; Courgeau and Lelievre, 2006; Hedman, 2011; Fielding, 2012). Further still, the complex and multifaceted influences of geographical context and place are also argued to operate at and across different levels of aggregation, from the neighbourhood context (reflecting the home area and locality) to the wider macro-geographical region (reflecting the wider landscape of social, economic and environmental constraints and opportunities). As is discussed in Chapters 2 and 8, various characteristics including the relative deprivation and socioeconomic status of the neighbourhood, the demographic and housing profile of the neighbourhood, and the relative stability of the neighbourhood population have all been the subject of empirical and theoretical interrogation with regards their potential influence on individual residential evaluations and selective mobility outcomes.

Yet beyond these measurable neighbourhood dynamics, it is also noted that the more subtle, indirect, and harder to measure effects of differential neighbourhood socialisation, relational networks, institutional resources and routines and other social and economic place-based processes and practices may be important for conditioning the decision/ability of individuals to change residence (see Chapter, 8; Tienda, 1991; Hedman, 2011; van Ham *et al.,* 2014). Operating at levels beyond the neighbourhood are a wide variety of additional macrogeographical influences linked, for instance, to the underlying geography of wealth and power, the associated spatial division of labour, the degree of medium- and short-term regional economic robustness, and the differing lifestyle and environmental opportunities afforded for in different macro-geographic areas (Fielding, 1992; Massey, 1995; Champion, 2008; Fielding, 2012). It was noted that these factors are important for informing the decision to move, but are perhaps of more importance when the decision to move is motivated by particular factors, most notably those pertaining to education, the labour market and/or the environment, which are often thought to encourage moves over longer distances and between different regions. For instance, Chapter 9 in particular, provides a detailed discussion on the importance of the differential attractiveness of different origin and destination contexts, as measured in terms of push and pull factors, for engendering the patterns, propensities and trends to the migration flows observed in the macro migratory system of GB.

Thus, what is immediately apparent from the discussions of the existing theoretical and empirical literature is that population movement is an overtly heterogeneous and complex phenomenon, characterised by marked differences, at all levels, from the various patterns, propensities and trends observed at the interregional level, through to the complex micro processes of residential evaluation and satisfaction, differential selectivity and resultant mobility behaviours and outcomes at the individual and household level. Indeed, as is outlined in particular in Chapters 2, 8 and 9, the apparent importance of both the micro and the macro, and the interaction between the two, is widely discussed and supported in the theoretical literature. Whether tied to the apparent dichotomy in micro and macro approaches to migration analysis, or the more likely result of a longstanding dearth in suitably detailed large sample microdata, there remains surprisingly little empirical work recognising the simultaneous effects of different micro processes and contextual effects on the movement behaviours and outcomes of individuals in GB. It is this ongoing situation that justifies the substantive focus of the thesis.

II. To critically evaluate the existing sources of secondary data (aggregate and micro) for the analysis of population movement in Great Britain

In order to contextualise and justify the combined use of commercial data and official statistics, Chapter 3 provided a thorough examination of the current data landscape in GB, arguing that researchers interested in population mobility often find themselves in a situation where they must utilise a variety of data sources each with different strengths and weaknesses, and each characterised by varying degrees of coverage, detail and accuracy. Indeed, as is argued throughout the thesis, limitations to the existing data landscape are often the key factor behind the apparent scarcity in the empirical demonstration of important multilevel phenomena relevant to population migration.

For instance, it was noted that census statistics are the most comprehensive and reliable of all, however, in their current guise, they are quickly outdated and, in terms of microdata such as the SAR and SAM, are often deliberately limited in geographical detail in order to protect respondent confidentiality. The administrative sources reviewed were shown to be useful in terms of their timeliness and geographic coverage; however, they are severely restricted by the population coverage and variable detail they contain. Surveys were given particular attention in Chapter 3 since they are by far the most varied, dynamic and understudied of all the potential sources of migration data. In general, surveys are found to produce regular outputs of highly detailed socio-demographic and socio-economic data, but generally speaking, they reflect serious limitations for mobility analysis relating to their relatively small sample sizes and restricted geographical detail/coverage.

However, further to the critical evaluation of the existing data sources, all of which are observed to contain certain attributes relevant for the analysis of population movement, Chapter 3 introduced the ROP, a source of detailed geo-demographic data with unique strengths for the analysis of both individual and place variations in residential moves, but also certain weaknesses that necessitate careful consideration. Indeed, as is detailed in Chapter 3, and to some extent in Chapters 5 and 6, the ROP's strengths lie in its ability to generate a sample of suitable size, geographical coverage and attribute detail to make it attractive for use in the simultaneous analysis of micro and macro contextual variations in population mobility behaviours and outcomes. However, the same chapters also present its relative weaknesses, including consistency issues, undocumented sampling strategies, sample bias and the raw nature of the microdata at delivery. Whilst Chapters 5 and 6 present a practicable approach to evaluating and validating the ROP as a reliable source of population migration microdata, the review in Chapter 3 additionally sought to locate it within the wider context of the ONS 'Beyond 2011' programme. Indeed, with ONS still searching for the optimal blend of methods and data sources for use in the future provision of population and socio-demographic statistics in the UK, a question remains as to whether the ONS should seriously consider the ROP as a valuable source of migration data. Whilst its limitations are far from trivial, it provides a biannual sample size that is far larger than any comparative government survey source. Moreover, with postcode level geo-identifiers and detailed lifestyle and socio-demographic variables, the samples contain many attributes that are of undoubted value to academics as well as service providers. Thus, within this context, it is argued that the research presented in the thesis should be of potential interest to ONS, given the attempts to benchmark, validate and integrate a hitherto unused source of commercial geo-demographic microdata with official statistics.

III. To review the current methodological approaches to the quantitative analysis of population movement at the macro, micro and multilevel scales in GB

Somewhat interlinked with the previous two objectives, which focus on theory and data, the need to review the various methodological approaches to the quantitative analysis of population movement was essential in providing the necessary context and justification for the analytical approach used in the thesis. As Chapter 4 makes clear, a longstanding dichotomy exists between the micro-level and macro-level approaches to the analysis of population migration (Stillwell and Congdon, 1991). Where the former is noted as being concerned with methods that analyse the behaviour of an individual migrant (or household), the influences on the decision-making process and the consequences of migration as far as the micro unit is concerned, the latter approaches are noted for their attempts to analyse aggregate migrant flows of people and identify the significance of macro explanatory variables (population size, employment rate or environmental factors) at either or both origin and destination, together with distance moved.

Indeed, with a tradition that can be dated back to Zipf (1946), macro migration modelling employs aggregate data in an attempt to understand aggregate population dynamics and the evolution of population structures and composition at different aggregate spatial levels. However, they cannot be used to explore the various micro (individual/household) characteristics, behavioural mechanisms or micro motives that are central to the variations in observed movement propensities, patterns and outcomes. On the other hand, micro approaches to modelling migration draw on highly detailed longitudinal and cross-sectional microdata sources. Through various generalised linear modelling techniques (though rule-based ABMs are also noted), micro approaches have been used to test and explore hypothesised relationships involving different personal characteristics and situations and their associated movement behaviours and outcomes. However, as is made clear in Subsection 4.3, there is a danger in micro-level modelling of only considering the characteristics of the individual and/or household, when theory would suggest that consideration of the wider residential context is also essential. As the thesis makes clear, mobility behaviours and outcomes most likely depend on the simultaneous combination of both micro-level characteristics and (perceptions of) macro variables, the latter of which can operate at and across various levels of aggregation.

Consequently, following a review of the traditional micro and macro approaches to migration modelling, Chapter 4 focussed in detail on multilevel modelling, an inherently flexible approach to modelling that has many of the necessary attributes required to justify its selection as the most appropriate methodology for maximising the utility of commercial and official statistics data in the analysis of individual and place variations in residential moves (see Chapters 8 and 9).

IV. To benchmark and validate the Acxiom Ltd. Research Opinion Poll, as a source of population migration microdata in GB, using official statistics (census, administrative and survey).

As was noted in Chapter 3, with its large geo-referenced sample, rich variable detail and extensive geographical coverage, the ROP microdata undoubtedly hold serious potential for the analysis of population migration. However, whilst the nature of this commercial data presents us with a novel opportunity to look at both individual and place variations in population mobility, it also makes the task of initial data management, cleaning and more general validation a difficult one. As mentioned above, cross-sample consistency issues, undocumented sampling strategies, sample bias and the raw nature of the microdata at delivery are all issues that require careful attention when planning analyses on the data. Consequently, in the context of these preimposed features of the ROP microdata, Chapters 5 and 6 employed a practical approach to its benchmarking and validation.

Chapter 5 began by introducing the extensive data preparation and cleaning exercises required for getting the raw ROP samples into a usable format whilst at the same time retaining as much of the raw data as possible. Following this, Chapter 5 reported on the initial descriptive-base benchmarking exercises employed on all of the raw ROP cross-sections for which the key variables of interest were recorded (i.e. January 2005, January 2006 and January 2007), though using the January 2005 ROP as an example. Whilst the thesis employs a model-based approach to the analysis of the ROP, the descriptive-based benchmarking was useful in uncovering bias in the different sub-sample distributions of the raw ROP samples, and informing an assessment of how successful the raw samples are in reflecting certain micro, aggregate and spatial mobility patterns found in alternative population data sources (Census 2001, PR-NHSCR, APS, and Acxiom Ltd. Aggregate Data).

Overall, the results of the descriptive-based benchmarking were mixed. Building on the findings of previous work by Thompson *et al.* (2010), the raw ROP samples are found to reflect bias in the distributions of certain key variables including age, sex ethnic group and mover status. Moreover, under/over sampling of certain geographical areas is also evident. Such findings mean that descriptive-based empirical analyses of the ROP should be avoided, at least until a method of sample adjustment has been employed. Indeed, the future application of sample adjustment techniques such as spatial microsimulation (Harland *et al.*, 2012) may well provide the platform for valid descriptive-based research to be undertaken using the ROP. For instance, the detailed geo-identifiers at both origin and destination, coupled with the variable detail included, are attractive features when considering the dearth in analyses of important population subgroup migration flows within GB. The example subgroups given in Chapter 5 included the young and highly educated adult population and the long-term unemployed.

However, whilst there were clear discrepancies between the raw sample distributions and those of official population statistics sources, there were positives to be drawn from the descriptive-based benchmarking exercises. Aggregate comparisons of inflow counts between the raw ROP and the alternative population data sources did suggest significant positive correlations. Moreover, despite concerned over certain sample distributions in raw ROP, micro-level comparisons presented reassuringly similar substantive patterns to the official statistics in terms of age specific, ethnic group and housing tenure mobility rates.

Whilst the initial cleaning and descriptive-based benchmarking exercises proved very useful in terms of retaining as much of the data as possible and evaluating the basic distributional distortions contained within the cleaned raw ROP cross-sections, the analytical approach used in the project is model-based in nature and thus further model-based approaches to validation and benchmarking, described and employed in Chapter 6, were required. Indeed, as was noted in Chapter 6, for the findings of the model-based analyses of the thesis to hold weight, it is important that the estimates derived are reasonably robust with respect to the known distortions contained within the sample distributions of the ROP. The initial focus of Chapter 6 (Subsection 6.2) was on detailing the issue of nonresponse in the ROP samples and justifying the choice of list-wise deletion (complete case analysis) as the most practical and pragmatic approach dealing with item nonresponse in the ROP. Whilst alternative imputation methods do exist, Subsection 6.2 suggested that single imputation methods should be avoided due to their potential for increasing sample bias, whereas multiple imputation methods were deemed computationally infeasible when applied to ROP datasets.

Thus, using the analytical complete case ROP samples (Table 6.7), Chapter 6 continued the practical approach that is deemed necessary for benchmarking and validating a commercial data source for which little to no information on sample design is available. This time, however, the focus was on the relative usefulness of the ROP in the model-based analysis of population mobility. As such, a method of sample reweighting, based on the use of auxiliary population data, was employed (Subsection 6.3); designed with the purpose of adjusting the sampling distributions of key variables in the ROP and checking the effects of the sample adjustments on the estimated model coefficients, as compared to unweighted model coefficients. Further to this, a brief model-based benchmarking exercise against data drawn from the 2001 Census was also presented in an attempt to further assess the relative value of the ROP for use in model-based analyses of population movement in GB.

Broadly speaking, the results of the model-based validation exercises were very encouraging. Whilst the effects of non-response bias cannot be entirely discounted, for instance contradictory relationships are observed although none are found to be substantively or statistically significant, the consistency observed across and between the weighted and unweighted model estimates is useful in showing the robustness of the model findings to the known sample discrepancies. As argued in Chapter 6, the covariates included in the models appear to work as suitable adjustment confounders, in controlling for sample distortions in associations between the predictors and the response, without the need for sampling weights (Lumley, 2010). Moreover, whilst the chapter did not focus on any serious substantive interpretation of the models, the major associational patterns that were revealed by the models did conform to much of the existing empirical and theoretical literature described in Chapter 2. Furthermore, the model-based benchmarking against the Census 2001 Individual SAR again highlighted the consistency and comparability of the major associational relationships.

Whilst Chapter 5 contributed significantly to achieving objective IV, it did suggest that the application of sample adjustment methods, such as spatial microsimulation, are required if the ROP is to be used as a source of population migration microdata for descriptive-based empirical analyses. However, in the context of the immediate aims to model individual and place variations in residential moves, Chapter 6 proves essential in demonstrating the relative reliability of results drawn from the cleaned complete case analytical samples, and particularly the pooled analytical sample (Subsection 6.6). Yet, in addition to the dedicated validation chapters (Chapters 5 and 6), it is also worth noting that the major analytical findings presented in the substantive chapters (Chapters 7, 8 and 9) also reflect many of the patterns, propensities and trends that the existing empirical and/or theoretical literature suggest should occur. For instance, the observed variations in the associational behaviours and characteristics of movers and stayers across the broad life-course stages (Chapter 7); the non-monotonic relationship between duration-of-residence and propensities for future residential mobility (Chapter 8); and the observed variations in the distance travelled according to individual/household characteristics as well as the macro geographical context of the origin and destination city region (Chapter 9). Indeed, in answer to Research Question 1 (Subsection 1.2), such findings add further weight to the idea that very reasonable, and indeed valuable, results can be drawn from the ROP for the analysis of individual and place variations in residential moves in GB.

V. To determine and quantify any individual and/or contextual variations in residential mobility with an initial detailed focus on micro-level demographic, socio-economic and lifestyle influences before allowing for, and modelling, variance heterogeneity where possible in a multilevel framework

Building on the positive findings of the previous validation chapters, the substantive analytical chapters (Chapters 7, 8 and 9) collectively addressed the fifth objective. Chapter 7 emerged from the validation models used in Chapter 6 in a way that allowed for the micro-level analysis of variations in the demographic, socio-economic and lifestyle characteristics of movers and stayers across broad life-course stages. This initial micro-level analysis was a required component of Research Question 2 and Objective V, and uncovered interesting associational patterns specifically related to some of the micro-level characteristics of movers *vis-à-vis* stayers that had, to this point, been seriously understudied due to the lack of suitably detailed microdata. Moreover, by separating the life course into four general stages – ages 18-29, 30-44, 45-64 and 65+ – the analysis also made possible the examination of consistency and dynamism in some of the key variables associated with movement propensities across the broad life-course stages. For instance, the propensity to move by ethnic group was found to vary depending on the life-course stage, whereas the effect of living in certain neighbourhood types was seen to be particularly consistent.

As detailed in Chapter 7, there were two key observations of this micro-level that, at face value, appeared to contradict certain longstanding theoretical assumptions. The first was related to the apparent insignificance of labour market characteristics on movement propensities, a finding that is said to be closely tied to the frictional effect of distance on mobility, whilst the second was related to an apparent contradiction of the widely theorised negatively monotonic duration-dependence relationship, where shorter durations correlate with greater propensities to change residence. Usefully, these observations could be explored in further detail in Chapters 8 and 9, where each formed the principal areas of substantive interest.

Chapter 8 employed a random intercepts and random slopes multilevel model, described in Chapter 4, in order to explore the functional form of the relationship between duration-of-residence and future residential mobility propensities. Drawing on the geographical coverage and rich attribute detail of an ROP subsample supplemented with official statistics, the modelling approach used made it possible to answer Research Question 3 and, in doing so, build on previous empirical analyses. Indeed, by explicitly modelling variance heterogeneity, the chapter was able to not only reveal an average non-monotonic relationship, matching that previously put forward by Gordon and Molho (1995), but also the extent to which the duration-of-residence relationship varied quite substantially in both direction and effect across the different neighbourhoods of England and Wales. Moreover, the between-neighbourhood variation in the propensity to be planning a move was observed to increase with individual durations of residence at the current address, a finding which appears to support the notion that a critical period of exposure is necessary for appreciable (unmeasurable) residential externalities to influence individual residential evaluations and movement behaviours.

The appropriate modelling of variance heterogeneity was also central to the analyses reported in Chapter 9, where the focus extended beyond multilevel variations in the ability/decision to change residence; instead, to the individual and place variations in the distance of move, once a movement event had taken place. Benefiting from longstanding multilevel theories and the availability of detailed geo-referenced origin and destination address data in the ROP, Chapter 9 employed a cross-classified multilevel model which presented, for the first time empirically, the importance of simultaneous individual and place-based variations, at both the origin and destination, in the distance moved by residential migrants in GB. Indeed, the findings confirmed the importance of micro-level variations in distance according to household income, educational attainment and housing tenure whilst simultaneously revealing the significance of macrogeographic variations, wherein a typical migrant was found to be pulled over significantly longer distances towards rural/coastal (amenity-rich) destination environments and, at the same time, pushed over significantly longer distances from urban-core origins. As a result, this chapter was successful in addressing the final research question, Research Question 4.

VI. To summarise the findings of the aforementioned objectives with a focus on answering the overall research aims

Whilst undeniably a topic of broad interest and importance, for a variety of reasons discussed throughout this thesis, the simultaneous measurement, analysis and understanding of individual and place variations in residential mobility and immobility, as well as the distances moved, has been limited in GB. However, through the combined use of detailed and geographically extensive commercial microdata, appropriate statistical methods, and well-informed multilevel theory, this project has been successful in answering the research objectives, questions and overall project aims set out in Chapter 1. In doing so, it has contributed to the substantive literature, providing some unique insights through a detailed and simultaneous analysis of various micro, macro and cross-level processes, characteristics and trends; many of which have often been well theorised but often hard to demonstrate using traditional data sources and/or methodological approaches. As discussed above, within the context of analysing individual and place variations in population movement in GB, this thesis has demonstrated the relative reliability of data drawn from the ROP. Indeed, model-based analyses in particular have been observed to produce findings that strongly reflect longstanding theoretical expectations and, where comparisons are possible, previous empirical demonstrations. Thus, following extensive reviews of the substantive literature, the methodological approaches taken, the existing data landscape and the management, cleaning and validation of the ROP, the work presented here, and in particular in Chapters 7, 8 and 9, has been successful in meeting the detailed research questions and the overall project aims set out in the introductory chapter. Yet, whilst this may be so, as has been alluded to in various parts of the thesis, there are a number of directions through which the research presented here can be taken forward.

10.3 Reflections on the approach taken and the potential for future research

In this penultimate section of the thesis, a number of reflections on particular issues confronted in the thesis are presented and some suggestions are made for future research. Indeed, as has been pointed out at certain points throughout the thesis, when using observational data, the regression based analyses of individual and place variations cannot be free from the concerns of omitted variables bias; that is, where an omitted variable is related both to the response and the included predictor variable. In addition, the omission of certain levels, the most notable here being the omission of the household as a separate level, can also be considered as potentially limiting not only from a substantive analytical point of view, but also though the possibility that the omission may attenuate the magnitude of certain other observed fixed- and random-part model findings. Moreover, in a similar manner, the open-ended choice of what constitutes a meaningful residential context means that different analytical observations may be found when different operational definitions are used.

Yet whilst such concerns are present, and indeed discussed, throughout the analyses chapters, they are not restricted to this project alone. Indeed, they are common concerns for which all researchers interested in individual and place variations must be aware of when employing regression-based methods on observational data (Subramanian, 2004b; Jones, 2010). Whilst the further inclusion of relevant micro and macro variables and contexts would almost always be useful within an empirical analysis, there are perpetual restrictions pertaining to data availability and, somewhat related to this, the extent to which all important predictors, of movement behaviours and outcomes for instance, can be reasonably quantified. However, as has been suggested in the analyses presented here, such issues need not be considered insurmountable to researchers. Indeed, whilst being aware of the potential pitfalls, it is possible for the development and interpretation of new and interesting insights to be made, so long as the research is thoroughly grounded in well thought through micro/macro theory, a sensible application and operationalization of measures and methods, and the careful/critical interpreting of model results.

Whilst every effort has been made to assuage the aforementioned limitations in the approach taken, other limitations exist that require alternative approaches and, ultimately, alternative data. For instance, whilst the ROP provides a unique opportunity to test contextual variations at particularly detailed geographical scales, it is somewhat limited by its design as cross-sectional survey. Indeed, as is clear from the discussions in Chapters 7, 8 and 9, the ROP restricts certain interpretive opportunities by failing to provide a measure for the timing of critical life-course events and transitions as well as wider dynamic processes operating at the household, neighbourhood and regional levels. However, with the development of increasingly rich large-scale geo-coded longitudinal datasets, the most notable example being the UKHLS - Understanding Society (Subsection 3.4.8), there does appear to be a good deal of potential in future analyses to focus on the necessary combination of both temporal and geographical context in the analysis of population movement in GB.

In the coming years, and once sufficient waves have been published, the utilisation of the UKHLS, within a suitably adjusted multilevel modelling approach, could well prove valuable for addressing some of the limitations here and more generally in extending the overall knowledge base relating to individual and place variations in residential moves. One such modelling approach could be the use of a repeated measures multiple membership model (Goldstein, 2011: 258-9), for instance with repeated measurements (at level 1) nested within individuals (level 2) who in turn are nested into households (level 3) and finally neighbourhoods (level 4). In this case, the model (realistically) allows for the movement of individuals into different household and neighbourhood contexts, thus making the outcome a modelled function of the changing characteristics of the individual, as well as a weighted function of the current and past household and neighbourhood characteristics. In this example, the level 3 and level 4 weights reflect the amount of time that each individual has spent in his/her current and previous household(s) and neighbourhood(s) respectively. Not only would it be possible to explore the impact and timing of certain life-course transitions and events, as Subramanian (2004b: 1964) has argued, creative multilevel structures such as this one, *"should allow an estimation of changing neighbourhood effects, [whilst] controlling for the changing population composition"*.

10.4 Concluding remark

Through the combined use of a detailed and geographically extensive commercial microdata and official statistics, appropriate statistical methods and well informed theory, this thesis has explored individual and place variations in residential moves in GB. In doing so it has offered unique insights into various patterns, propensities and characteristics of residential mobility that, whilst long theorised, have often been difficult to demonstrate empirically due to a scarcity in access to both appropriate data and methods. However, there are of course areas of research still to be improved, estimates to be updated, data to be gathered, and techniques to be honed. Yet, with the emergence of new geo-referenced longitudinal data sets and the ongoing development of realistically complex methodological approaches, the coming years look set to be an exciting time for the quantitative analysis of population migration.

Bibliography

Abraham, M., Ausburg, K. & Hinz, T. (2010). Migration decisions within dual-earner partnerships: A test of bargaining theory, *Journal of Marriage and Family*, 72: 876-892.

Acxiom Ltd. (2007). *Research Opinion Poll: January 2007*. London: Acxiom Ltd.

Acxiom Ltd. (2011). *PersonicX® Geo: Giving you a fresh perspective on your customers*. Online at: http://personicx.co.uk/UserDocumentation/Personicx%20Geo%20Brochure.pdf (last accessed: 25/06/2014).

Acxiom Ltd..(2014a). *About Acxiom*. Online at: http://www.acxiom.co.uk/about-acxiom/ (last accessed: 25/06/2014).

Acxiom Ltd..(2014b*). Marketing Data Solutions*. Online at: http://www.acxiom.co.uk/data-solutions/ (last accessed: 25/06/2014).

Agresti, A. (2002). *Categorical Data Analysis (2nd edition)*. Wiley: New Jersey.

Agresti, A. (2007). *An Introduction to Categorical Data Analysis (2nd edition)*. Wiley: New Jersey.

Alker, H.A. Jr. (1969). A typology of ecological fallacies, in Dogan, M. & Rokkan, S. (eds.) *Quantitative Ecological Analysis*. Cambridge, Massachusetts: Massachusetts Institute of Technology. pp. 69–86.

Andersson, R. & Bråmå, A. (2004). Selective migration in Swedish distressed neighbourhoods: Can area-based urban policies counteract segregation processes?, *Housing Studies*, 19: 517-539.

Andridge, R. & Little, R. (2010). A Review of Hot Deck Imputation for Survey Non-response, *International Statistical Review*, 78: 40-64.

Axisa, J., Scott, D.M. & Newbold, K.B. (2011). Factors influencing commute distance: A case study of Toronto's commuter shed, *Journal of Transport Geography*, 24: 123-129.

Bailey, A.J. (2009). Population geography: Lifecourse matters, *Progress in Human Geography*, 33: 407–418.

Bailey, N. & Livingston, M. (2005). Determinants of individual migration: an analysis of SARs data, *SCRSJ Working Paper No. 3*, Scottish Centre for Research on Social Justice, University of Glasgow, Glasgow.

Bailey, N. & Livingston, M. (2007).*Population Turnover and Area Deprivation*. Bristol: Policy Press.

Bailey, N. & Livingston, M. (2008). Selective migration and neighbourhood deprivation: Evidence from 2001 Census migration data for England and Scotland, *Urban Studies*, 45, 943-961.

Bartel, A. (1979). The migration decision: What role does job mobility play?, *American Economic Review*, 69: 775-786.

Battaglia, M.P., Izrael, D., Hoaglin D.C. & Frankel, M.R. (2009). Practical considerations in raking survey data, *Survey Practice*, (June). http://surveypractice.org/2009/06/29/raking-survey-data.

Bethlehem, J., Cobben, F. & Schouten, B. (2011). *Handbook of Nonresponse in Household Surveys*. New Jersey: Wiley.

Bishop, Y.M., Fienberg, S.E. & Holland, P.W. (1975). *Discrete Multivariate Analysis: Theory and Practice*. Cambridge, MA: MIT press.

Bogue D.J. (1969). *Principles of Demography*. Wiley: New York.

Bondi, L. & Bradford, M. (1990). Applications of multilevel modelling to geography, *Area*, 22: 256-63.

Boehm, T.P. & Ihlanfeldt, K.R. (1986). Residential mobility and neighborhood quality, *Journal of Regional Science*, 26: 411–424.

Böheim, R. & 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: 369-392.

Borjas, G., Bronars, S. G. & Trejo, S. J. (1992). Self-selection and internal migration in the United States, *Journal of Urban Economics*, 32: 159-85.

Boyle, P.J. (1993). Modelling the relationship between tenure and migration in England and Wales, *Transactions of the Institute of British Geographers*, 18: 359-76.

Boyle, P.J. (1995). Public housing as a barrier to long-distance migration, *International Journal of Population Geography*, 1: 147–64.

Boyle, P.J. & Halfacree, P. (1998). *Migration into Rural Areas*. Chichester: Wiley.

Boyle, P.J. & Shen, J. (1997). Public housing and migration: A multi-level modelling approach, *International Journal of Population Geography*, 3, 227-242.

Boyle, P., Feng, Z. & Gayle, V. (2009). A new look at family migration and women's employment status, *Journal of Marriage and Family*, 71: 417-431.

Boyle, P., Cooke, T. J., Halfacree, K. H. & Smith, D. (2001). A cross-national comparison of the impact of family migration on women's employment status, *Demography*, 38: 201-213.

Boyle, P.J., Kulu, H., Cooke, T.J., Gayle, V. & Mulder, C.H. (2008). Moving and union dissolution, *Demography*, 45: 209-222.

Brown, D. (2002). Migration and community: Social networks in a multilevel world, *Rural Sociology*, 67: 1-23.

Browne, W.J. (2012). *MCMC estimation in MLwiN, v2.26*. Centre for Multilevel Modelling, University of Bristol. Online at: http://www.bristol.ac.uk/cmm/software/mlwin/download/manuals.html.

Browne, W.J. & Draper, D. (2000). Implementation and performance issues in the Bayesian and likelihood fitting of multilevel models, *Computational Statistics*, 15:391-420.

Browne, W.J. & Draper, D. (2006). A comparison of Bayesian and likelihood-based methods for fitting multilevel models, *Bayesian Analysis*, 1: 473-550.

Browne, W.J., Goldstein, H. & Rasbash, J. (2001). Multiple membership multiple classification (MMMC) models, *Statistical Modelling*, 1: 103-124.

Browne, W.J., Steele, F., Golalizadeh, M. & Green, M.J. (2009). The use of simple reparameterizations to improve the efficiency of Markov chain Monte Carlo estimation for multilevel models with applications to discrete time survival models, *Journal of the Royal Statistical Society: Series A*, 172: 579-598.

Brunton-Smith, I. & Sturgis, P. (2011). Do neighbourhoods generate fear of crime? An empirical test using the British Crime Survey, *Criminology*, 49: 311-369.

Bryan, M. (2011). Measuring Work: Prospects for labour market research in understanding Society, in McFall, S.L. & Garrington, C. (Eds.), *Early findings from the first wave of the UK's household longitudinal study.* Colchester: Institute for Social and Economic Research, University of Essex.

Burton, J., Laurie, H. & Lynn, P. (2011).Appendix: Understanding Society Design Overview', in McFall, S.L. & Garrington, C. (Eds.), *Early findings from the first wave of the UK's household longitudinal study.*Colchester: Institute for Social and Economic Research, University of Essex.

Cadwallader, M. (1989). A conceptual framework for analysing migration behaviour in the developed world, *Progress in Human Geography*, 13: 494-511.

Calder, A. & Swan, J. (2011).*Beyond 2011: The future for population statistics?*, presentation at the Beyond 2011 User Requirements Workshop, The Carriageworks, Leeds, 03 November.

Cangiano, A. (2010). *UK data sources on international migration and the migrant population: A review and appraisal*, COMPAS research resource paper. Oxford: University of Oxford.

Champion, T. (2005a). Population movement within the UK, in R. Chappell (ed.), *Focus on people and migration*. Basingstoke: Palgrave Macmillan. pp. 92-114.

Champion, T. (2005b). The counterurbanization cascade in England and Wales since 1991: The evidence of a new migration dataset, *BELGEO*, 1-2: 85-101.

Champion, T. (2008). Population change in England since 1981: Is an 'urban renaissance' really underway?, *Géocarrefour*, 83: 77-86.

Champion, T., Fotheringham, A.S., Rees, P.H., Boyle, P.J. and Stillwell, J. (1998). *The Determinants of Migration Flows in England: A Review of Existing Data and Evidence*, Report for the Department of the Environment, Transport and the Regions, University of Newcastle upon Tyne.

Chevan, A. (2005). Holding on and letting go: Residential mobility during widowhood, *Research on Aging*, 17: 278-302.

Chi, G. & Voss, P. (2005). Migration decision-making: A hierarchical regression approach, *Regional Analysis & Policy*, 35: 11-22.

Clark, W.A.V. (2013). Life course events and residential change: Unpacking age effects on the probability of moving, *Journal of Population Research*, 30: 319-334.

Clark, W.A.V. & Davies Withers, S. (2009). Family formation, mobility and labor force participation: A study in synchronicity, *Population Space and Place*, 15: 305–321.

Clark, W.A.V. & Dieleman, F.M. (1996). *Households and Housing: Choice and Outcomes in the Housing Market*. New Brunswick: Centre for Urban Policy Research.

Clark, W.A.V. & Huang, Y. (2003).The life course and residential mobility in British housing markets. *Environment and Planning A*, 35: 323-339.

Clark, W.A.V. & Ledwith, V. (2006). Mobility, housing stress, and neighborhood contexts: evidence from Los Angeles, *Environment and Planning A*, 38: 1077–93.

Clark, W.A.V, Huff, J.O. & Burt, J.E. (1979). Calibrating a model of the decision to move, *Environment and Planning A*, 11: 689-704.

Congdon, P. (1991). An application of general linear modelling to migration in London and the South East, in Stillwell, J. & Congdon, P. (eds.) *Migration Models: Macro and Micro Perspectives*. London: Belhaven Press. pp. 113-136.

Congdon, P. (2010). *Applied Bayesian Hierarchical Methods*. Boca Ranton, FL: Chapman & Hall/CRC.

Cooke, T.J. (2011). It is not just the economy: Declining migration and the rise of secular rootedness, *Population, Space and Place*, 17: 193–203.

Cordey-Hayes, M. & Gleave, D. (1974). Migration movements and the differential growth of city regions in England and Wales, *Papers of the Regional Science Association*, 33: 99-123.

Coulter, R., van Ham, M. & Feijten, P. (2011). A longitudinal analysis of moving desires, expectations and actual moving behaviour, *Environment and Planning A*, 43: 2742-2760.

Coulter, R., van Ham, M. & Feijten, P. (2012). Partner (dis)agreement on moving desires and the subsequent moving behaviour of couples, *Population, Space and Place*, 18: 16-30.

Courgeau, D. (1995). From the group to the individual: What can be learned from migratory behaviour, *Population: An English Selection*, 1: 145-162.

Courgeau, D. & Baccaini, B. (1998). Multilevel analysis in the social sciences, *Population: An English Selection*, 10: 39-71.

Courgeau, D. & Lelievre, E. (2006). Individual and social motivations for migration, in Caselli, G., Vallin, J. & Wunsch, G. (eds.), *Demography: Analysis and Synthesis (Vol. II)*. Oxford: Elsevier Inc. pp. 345-357.

Crockett, A., Afkhami, R., Rafferty, A., Higgins, V. & Marshall, A. (2011). *Weighting the Social Surveys: ESDS Government*. Colchester: ESDS.

Crooks, A., Castle, C. & Batty, M. (2008). Key challenges in agent-based modelling for geo-spatial simulation, *Computers, Environment and Urban Systems*, 32: 417-430.

Crowder, K. (2000). The racial context of white mobility: an individual-level assessment of the white flight hypothesis, *Social Science Research*, 29: 223-257.

Cushing, B. & Poot, J. (2004). Crossing boundaries and borders: Regional science advances in migration modelling, *Papers in Regional Science*, 83: 317-338.

Dale, A., Creeser, R., Dodgeon, B., Gleave, S. & Filakti, H. (1993). An introduction to the OCPS Longitudinal Study, *Environment and Planning A*, 25: 1387-98.

Dale, A., Holdsworth, C. & Fieldhouse, E. (2000). *The Analysis of Census Microdata*. London: Edward Arnold.

Davies, R.B. (1991). The analysis of housing and migration careers, in Stillwell, J. & Congdon, P. (eds.) *Migration Models: Macro and Micro Approaches*. London: Belhaven. pp. 207-227.

DCLG. (2013). *English Housing Survey technical advice note: Survey overview and methodology 2011-12 update*. London: DCLG.

Deming, W.E. (1943). *Statistical Adjustment of Data*. New York: Wiley.

Deming, W.E. & Stephan, F.F. (1940). On a least squares adjustment of a sampled frequency table when expected marginal totals are known, *Annals of Mathematical Statistics*, 11: 427-44.

Dennett, A, Duke-Williams, O. & Stillwell, J. (2007).Interaction datasets in the UK: An audit', *Working Paper 07/4*. Leeds: School of Geography, University of Leeds.

Dennett, A. & 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., & Stillwell, J. (2010). Internal migration in Britain, 2000-01, examined through an area classification framework, Population Space and Place, 16: 517-538.

Dennett, A. & Wilson, A. (2013). A multilevel spatial interaction modelling framework for estimating interregional migration in Europe, *Environment and Planning A*, 45: 1491–1507.

Diez Roux, A.V. (2002). A glossary for multilevel analysis, *Journal of Epidemiology & Community Health*, 56: 588-594.

Dorling, D. & Rees, P. (2003). A nation still dividing: The British census and social polarisation 1971-2001, *Environment and Planning A*, 35: 1287-313.

Draper, D. (2006). *Bayesian Multilevel Analysis and MCMC*. New York: Springer.

Duke-Williams, O. (2009). The geographies of student migration in the UK, *Environment & planning A*, 41: 1826-1848.

Duke-Williams, O. (2011). The role of questions about migration in UK censuses: A simple matter of counting, or a means of exerting power?, *Geoforum*, 42: 615-629.

Duke-Williams, O. & Stillwell, J. (2007). Investigating the potential effects of small cell adjustments on interaction data from the 2001 Census, *Environment and Planning A*, 39: 1079-1100.

Duke-Williams, O. & Stillwell, J. (2010). Temporal and spatial consistency, in Stillwell, J., Duke-Williams, O. & Dennett, A. (eds.), *Technologies for Migration and Commuting Analysis: Spatial Interaction Data* Applications. Hershey: IGI Global. pp. 89-110.

Duncan, C., Jones, K. & Moon, G. (1998). Context, composition and heterogeneity: Using multilevel models in health research, *Social Science and Medicine*, 46, 1: 97-117.

Dunford, M. & Fielding A.J. (1997). Greater London, the Southeast Region and wider Britain: Metropolitan polarization, uneven development and inter-regional migration, in Blotevogal, H.H. & Fielding, A.J. (eds.), *People, jobs and mobility in the new Europe*. Chichester: Wiley. pp. 247-276.

Dunstan, S. (2011). *General Lifestyle Survey Overview: A Report on the 2009 General Lifestyle Survey*. London: ONS.

DWP. (2011). *Family Resources Survey: United Kingdom, 2009-10*, May 2011. London: DWP.

Dykstra, P.A. & van Wissen, L.J.G. (1999). Introduction: The life course approach as an interdisciplinary framework for population studies, in van Wissen, L.J.G. & Dykstra, P.A. (eds.), *Population Studies: An Interdisciplinary Focus*. New York: Plenum Press. pp. 1-22.

Elder, G.H. (1994). Time, human agency, and social change: Perspectives on the life course, *Social Psychology Quarterly*, 57: 4-15.

Ellen, I. (2000). *Sharing America's Neighbourhoods*. Cambridge, MA: Harvard University Press.

Evandrou, M., Falkingham, J. & Green, M. (2010). Migration in later life: Evidence from the British Household Panel Study, *Population Trends*, nr-141: 77-94.

Feijten, P. & Mulder, C.H. (2002). The timing of household events and housing events in the Netherlands: A longitudinal perspective, *Housing Studies*, 17: 773-792.

Feijten, P. & van Ham, M. (2007). Residential mobility and migration of the divorced and separated, *Demographic Research*, 17: 623-654.

Feijten, P. & van Ham, M. (2009). Neighbourhood change… reason to leave?, *Urban Studies*, 46: 2103-2122.

Field, A., Miles, J. & Field, Z. (2012). *Discovering Statistics Using R*. London: Sage.

Fielding, A. (2004). The role of the Hausman test and whether higher level effects should be treated as random or fixed, *Multilevel Modelling Newsletter*, 16: 3–9.

Fielding, A.J. (1989). Inter-regional migration and social change: A study of South East England based on data from the Longitudinal Study, *Institute of British Geographers Transactions,* 14: 24-36.

Fielding, A.J. (1992). Migration and social mobility: South East England as an escalator region, *Regional Studies*, 26: 1-15.

Fielding, A.J. (1995). Migration and middle class formation in England and Wales, 1981-91, in Butler, T. & Savage, M. (eds.), *Social Change and the Middle Classes*. London: UCL Press. pp. 169-189.

Fielding, A.J. (1998). Gender, class and region in England and Wales: A longitudinal analysis, *Ritsumeikan Chirigaku*, 10: 1-22.

Fielding, A.J. (2007). Migration and social mobility in urban systems: national and international trends, in Geyer, H.S. (ed.), *International Handbook of Urban Policy*. Cheltenham: Edward Elgar. pp. 107-137.

Fielding, A.J. (2012). *Migration in Britain: Paradoxes of the Present, Prospects for the Future*. Cheltenham: Edward Elgar.

Fielding, A. & Goldstein, H. (2006). Cross-classified and multiple membership structures in multilevel models: An introduction and review. *Research Report RR791*. London: Department for Education and Skills.

Finney, N. & Simpson, L. (2008). Internal migration and ethnic groups: evidence for Britain from the 2001 Census, *Population, Space and Place*, 14: 63-083.

Flowerdew, R. (1976). Search strategies and stopping rules in residential mobility, *Transactions of the Institute of British Geographers*, 1: 47-57.

Flowerdew, R. (1982). Institutional effects on internal migration, in Flowerdew, R. (ed.), *Institutions and Geographical Patterns*. New York: St Martin's Press. pp. 209-227.

Flowerdew, R. (1991). Poisson regression modelling of migration, in Stillwell, J. & Congdon, P. (eds.), *Migration Models: Macro and Micro Perspectives*. London: Belhaven Press. pp. 92-112.

Flowerdew, R. & Green, A. (1993).Migration, transport and workplace statistics from the 1991 Census, in Dale, A., & Marsh, C. (eds.), *The 1991 Census User's Guide*. London: HMSO. pp. 269-294.

Flowerdew, R. (2010). Modelling migration with Poisson regression, in Stillwell, J., Duke-Williams, O. & Dennett, A. (eds.), *Technologies for Migration and Commuting Analysis: Spatial Interaction Data Applications*. Hershey: IGI Global.

Fotheringham, A.S. (1983). A new set of spatial interaction models: the theory of competing destinations, *Environment and Planning A*, 15: 15-36.

Fotheringham, A.S., Champion, T., Wymer, C. & Coombes, M. (2000). Measuring destination attractivity: A migration example, *International Journal of Population Geography*, 6: 391-421.

Fotheringham, A.S., Nakaya, T., Yano, K., Openshaw, S. & Ishikawa, Y. (2001). Hierarchical destination choice and spatial interaction modelling: a simulation experiment, *Environment and Planning A*, 33: 901-920.

Fotheringham, A.S., Brunsdon, C. & Charlton, M. (2002). *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*. Chichester: Wiley.

Fotheringham, A.S, Rees, P., Champion, T., Kaglogirou, S. &Tremayne, A.R. (2004). The development of a new migration model for England and Wales: Overview and modelling out-migration, *Environment and Planning A*, 36: 1633-1672.

Frey, W.H. (2003). Internal migration, in Demeny, P. & McNicoll, G. (eds.), *Encyclopaedia of Population (volume II)*. New York: Macmillan Reference. pp. 545-548.

Fry, R. (2011). Understanding household income poverty at small area level, *Regional Trends*, 43: 87-102.

Galster, G.C. & Killen, S.P. (1995). The geography of metropolitan opportunity: A reconnaissance and conceptual framework, *Housing Policy Debate*, 6:7–43.

Geist, C. & McManus, P.A. (2008). Geographical mobility over the life-course: motivations and implications, *Population, Space and Place*, 14: 283-303.

Gelman, A., Carlin, J.B., Stern, H.S. & Rubin, D.B. (2004). *Bayesian Data Analysis (2nd edition)*. Boca Ranton, FL: Chapman & Hall/CRC.

Goldstein, H. (1986). Multilevel mixed linear model analysis using iterative generalised least squares, *Biometrika*, 73: 43-56.

Goldstein, H. (1989). Restricted unbiased iterative generalised least squares estimation, *Biometrika*, 76: 622-623.

Goldstein, H. (2011). *Multilevel Statistical Models (4th edition)*. Chichester: Wiley.

Goldstein, H., Browne, W.J. & Rasbash, J. (2002). Partitioning variation in multilevel models. *Understanding Statistics*, 1: 223-232.

Gordon, I. (1982). The analysis of motivation-specific migration streams, *Environment and Planning A*, 14: 5-20.

Gordon, I. & Molho, I. (1995), Duration dependence in migration behaviour: Cumulative inertia versus stochastic change, *Environment and Planning A*, 27: 1961-75.

Gould, M. & Jones, K. (1996). Analyzing perceived limiting long-term illness using U.K. Census Microdata. *Social Science and Medicine*, 42: 857-869.

Hägerstrand, T. (1975). Space, time and human conditions, in Karlqvist, A., Lundqvist, L. & Snickars, F. (eds.) *Dynamic Allocation of Urban Space*. Lexington, MA: Lexington Books. pp. 3–14.

Hamnett, C. & Randolph, B. (1987). The OPCS Longitudinal Study: A new tool for social research in England and Wales, *Area*, 19: 69-73.

Hamptons International Ltd. (2013). 32 Years Old and Living in London? Chances are You’re About to Escape, but Only by 26 Miles, Press release Monday 2 December 2013. London: Hamptons International Ltd.

Harland, K. & Stillwell, J. (2007).Using PLASC data to identify patterns of commuting to school, residential migration and movement between schools in Leeds, *Working Paper 07/03*. Leeds: School of Geography, University of Leeds.

Harland, K., Heppenstall, A., Smith, D. & Birkin, M. (2012). Creating realistic synthetic populations at varying spatial scales: A comparative critique of population synthesis techniques, *Journal of Artificial Societies and Social Simulation*, 15: 1-15.

Harris, D.R. (1999). Poverty values drop when blacks move in, because …: Racial and socioeconomic determinants of neighbourhood desirability, *American Sociological Review*, 64(3), 461-479.

Hattersley, L. & Boyle, P. (2007). The Scottish Longitudinal Study: An introduction, *Longitudinal Studies Centre Scotland Working Paper 1.0*. St Andrews: LSCS.

Hedman, L. (2011). The impact of residential mobility on measurements of neighbourhood effects, *Housing Studies*, 26: 501-519.

Hedman, L., van Ham, M. & Manley, D. (2011). Neighbourhood choice and neighbourhood reproduction, *Environment and Planning A*, 43: 1381-1399.

Heeringa, S.G., West, B.T. & Berglund, P.A. (2010). *Applied Survey Data Analysis*. London: Chapman & Hall/CRC.

HESA (2011) *CEO's Report for HESA Board, 26th May 2011*. Online at: http://www.hesa.ac.uk/index.php/content/view/2321/292/ (last accessed 25/06/2014).

Howe, T. (2010). *Life Opportunities Survey: Technical Report 2009/10*. London: ONS.

Huff, J.O. & Clark, W.A.V. (1978). Cumulative stress and cumulative inertia: a behavioural model of the decision to move, *Environment and Planning A*, 10: 1101-1119.

Hughes G. & McCormick, B. (1985). Migration intentions in the UK: Which households want to migrate and which succeed?, *Economic Journal, Conference Supplement*, 95: 113-123.

Hughes, G. & McCormick, B. (2000). *Housing Policy and Labour Market Performance*. London: ODPM.

Jackman, S.D. (2009). *Bayesian Analysis for the Social Sciences*. New York: Wiley.

Jefferies, J., Horsfield, G., Newman, J. & Stokes, K. (2003). *Report on Research into Revising Internal Migration Estimates*. London: ONS.

Jones, C. (2011). *The latest developments on the Integrated Household Survey – Information for LFS users*, Presentation to the Labour Force Survey / Annual Population Survey User Meeting, Royal Statistical Society, London, 6 December.

Jones, K. (1991a). Specifying and estimating multilevel models for geographic research, *Transactions of the Institute of British Geographers,* 16: 148–160.

Jones, K. (1991b). *Multilevel models for geographical research: Concepts and techniques in modern geography*. Norwich: Environmental Publications.

Jones, K. (2010). An introduction to statistical modelling, in Somekh, B. & Lewin, C. (eds.), *Research Methods in the Social Sciences*. London: Sage. pp. 241-255.

Jones, K. & N, Bullen. (1994). Contextual models of urban house prices: a comparison of fixed- and random-coefficient models developed by expansion, *Economic Geography*, 70: 252-272.

Jones, K. & Duncan, C. (1995). Individuals and their ecologies: analysing the geography of chronic illness within a multilevel framework, *Health and Place*, 1: 27-40.

Jones, K. & Duncan, C. (1996). People and places: The multilevel model as a general framework for the quantitative analysis of geographical data, in Longley, P. & Batty, M. (eds.), *Spatial Analysis: Modelling in a GIS Environment*. Cambridge: Geoinformation Group, Ely. pp. 79-104.

Jones, K. & Subramanian, V.S. (2013). *Developing Multilevel Models for Analysing Contextuality, Heterogeneity and Change (Volume 1)*. Bristol: Centre for Multilevel Modelling, University of Bristol. Online at: http://www.mendeley.com/profiles/kelvyn-jones/.

Jones, K., Gould, M. & Watt, R. (1998). Multiple contexts as cross-classified models: The Labour vote in the British general election of 1992, *Geographical Analysis*, 30: 65-93.

Jordan, R., Birkin, M. & Evans, A.J. (2011). Agent-based simulation modelling of housing choice and urban regeneration policy, in Bosse, T., Geller, A. & Jonker, C. (eds.) *Multi-Agent-Based Simulation XI*. Berlin: Springer.

Kalton, G. (1983). *Compensating for Missing Survey Data*. Michigan: Survey Research Center, Institute for Social Research, University of Michigan.

Kawachi, I. & Subramanian, S.V. (2006). Measuring and modelling the social and geographic context of Trauma: A multilevel modelling approach, *Journal of Traumatic Stress*, 19: 195-203.

Kearns, A. & Parkes, A. (2003). Living in and leaving poor neighbourhood conditions in England, *Housing Studies*, 18: 827-851.

Kearns, A. & Parkinson, M. (2001). The significance of neighbourhood, *Urban Studies*, 38: 2103-2110.

Keeter, S., Miller, C., Kohut, A., Groves, R. & Presser, S. (2000). Consequences of reducing nonresponse in a national telephone survey, *Public Opinion Quarterly*, 64: 125-148.

King, R. (2012). Geography and migration studies: retrospect and prospect, *Population, Space and Place*, 18: 134–153.

Kley, S. & Mulder, C. (2010). Considering, planning, and realizing migration in early adulthood: The influence of life-course events and perceived opportunities on leaving the city in Germany, *Journal of Housing and the Built Environment*, 25: 73-94.

Kulu, H. & Steele, F. (2013). Interrelationships between childbearing and housing transitions in the family life course, *Demography*, 50: 1687-1714.

Land, K. (1969). Duration of residence and prospective migration: Further evidence, *Demography*, 6: 133-140.

Large, P. & Ghosh, K. (2006). Estimates of the population by ethnic group for areas within England, *Population Trends*, 124: 8-17.

Larsen, K., Petersen, J.H., Budtz-Jørgensen, E. & Endahl, L. (2000).Interpreting parameters in the logistic regression model with random effects, *Biometrics*, 56: 909-14.

Lauster, N.T. (2010). Housing and the proper performance of American motherhood, 1940-2005, *Housing Studies*, 25: 543-557.

Lee, E.S. (1966). A Theory of Migration, *Demography*, 3, 47-57.

Lee, B.A., Oropesa, R.S. & Kanan, J.W. (1994). Neighborhood context and residential-mobility, *Demography*, 31: 249–270.

Little, R.J. (2008). Weighting and prediction in sample surveys, *Calcutta Statistical Association Bulletin*, Vol. 60, September and December 2008, Nos. 239-240.

Little, R.J. & Wu, M.M. (1991). Models for contingency tables with known margins when target and sampled populations differ, *Journal of the American Statistical Association*, 86: 87-95.

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

Lu, M. (1998). Analyzing migration decision making: Relationships between residential satisfaction, mobility intentions, and moving behaviour, *Environment and Planning A*, 30: 1473-95.

Lumley, T. (2010). *Complex Surveys: A Guide to Analysis Using R*. New Jersey: Wiley.

Lumley, T. (2012). Survey: Analysis of complex survey samples. *R Package Version 3.28-2*.

Magdol, L. (2002). Is moving gendered? The effects of residential mobility on the psychological wellbeing of men and women, *Sex Roles*, 47: 553-560.

Malleson, N., Heppenstall, A. & See, L. (2008). Crime reduction through simulation: An agent-based model of burglary, *Computers, Environment and Urban Systems*, 31: 236-250.

Marquis, N. & Jivraj, S. (2009). Preparation of pupil level Annual School Census data for the analysis of internal migration, *CCSR Working Paper 2009-03*. Manchester: CCSR, University of Manchester.

Martin, D. (2002a). Geography for the 2001 Census in England and Wales, *Population Trends*, 108, 7-15.

Martin, D. (2002b). Output Areas for 2001, in Rees, P., Martin, D. & Williamson, P. (eds.), *The Census Data System*. Chichester: Wiley. pp. 37-46.

Massey, D.B. (1995). *Spatial Divisions of Labor: Social Structures and the Geography of Production (2nd edition)*. New York: Routledge.

Massey, D.S. (1990). Social structure, household strategies, and the cumulative causation of migration, *Population Index*, 56: 3-26.

McFall, S.L. (2011). *Understanding Society – The UK Household Longitudinal Study, Wave 1, 2009-2010, User Manual*. Colchester: University of Essex.

McGinnis, R. (1968). A stochastic model of social mobility, *American Sociological Review*, 33: 712-722.

Mitchell, C. (2004). Making sense of counterurbanization, *Journal of Rural Studies*, 20: 15-34.

Moon, G., Gould, M., & colleagues (2000). *Epidemiology for Nurses and the Caring Professions*. Milton Keynes: Open University Press.

Morrison, P.A. (1971). Chronic movers and the future redistribution of population: A longitudinal analysis, *Demography*, 8: 171-184.

Morrison, P.S. & Clark, W.A.V. (2011). Internal migration and employment: Macro flows and micro motives, *Environment and Planning A*, 43: 1948-1964.

Mulder, C.H. (1993). *Migration Dynamics: A Life Course Approach*. Amsterdam: Thesis Publishers.

Mulder, C.H. (2007). The family context and residential choice: A challenge for new research, *Population, Space and Place*, 13: 265–278.

Mulder, C.H. (2013). Family dynamics and housing: Concepts issues and empirical findings, *Demographic Research*, 29: 355-378.

Mulder, C.H. & Wagner, M. (1993). Migration and marriage in the life course: A method for studying synchronized events, *European Journal of Population*, 9: 55–76.

Mulder, C.H. & Wagner, M. (2010). Union dissolution and mobility: Who moves from the family home after separation?, *Journal of Marriage and Family*, 72: 1263-1273.

Mulder, C.H. & Wagner, M. (2012). Moving after separation: The role of location-specific capital, *Housing Studies*, 27: 839–852.

Nam, C.B., Serow, W.J. & Sly, D.F. (eds.) (1990). *International Handbook on Internal Migration*. New York: Greenwood Press.

Nielsen, T.A.S. & Hovgesen, H.H. (2008). Exploratory mapping of commuter flows in England and Wales, *Journal of Transport Geography*, 16: 90-99.

Norman, P. & Boyle, P. (2010).Using migration microdata from the Samples of Anonymised Records and the Longitudinal Studies, in Stillwell, J., Duke-Williams, O. & Dennett, A. (eds.), *Technologies for Migration and Commuting Analysis: Spatial Interaction Data Applications*. IGI Global: Hershey. pp. 133-152.

ODPM. (2004). *The English Indices of Deprivation 2004 (revised)*. Wetherby: ODPM Publications.

ONS. (2002) *Review of the Labour Force Survey. National Statistics Quality Review Series: Report No. 12*. London: ONS.

ONS. (2003a). *LS User Guide 20: Researching households and families using the ONS Longitudinal Study*. London: ONS.

ONS. (2003b). *Longitudinal Study 1971-2001: Completeness of census linkage*, Series LS no. 10. London: ONS.

ONS. (2010a). *Improved estimation of student migration within England and Wales*. London: ONS.

ONS. (2010b). *User Guide – Volume 1: IHS Background and Methodology 2010*. London: ONS.

ONS. (2011a). *Information Paper: Quality and Methodology Information – Internal Migration Estimates*. London: ONS.

ONS. (2011b). *Estimating Internal Migration – Customer Guidance Notes*. London: ONS.

ONS. (2011c). *Integrated Household Survey User Guide – Volume 3: Details of IHS variables 2011*. London: ONS.

ONS. (2011d). *Performance and Quality Monitoring Report – April to June 2011: Labour Force Survey*. London: ONS.

ONS. (2011e). *User Guide – Volume 1: LFS Background and Methodology 2011*. London: ONS.

ONS. (2011f). *Consultation: The Future of the General Lifestyle Survey – February 2011*. London: ONS.

ONS. (2011g). *Changes to the collection of EU-SILC data*. Online at: http://www.esds.ac.uk/government/frs/frseusilcchange.pdf (last accessed 25/06/2014).

ONS. (2011h). *Family Resources Survey: User Guide, 2009-2010 Version*. London: ONS.

ONS. (2011i). *English Housing Survey, 2009-2010: Household Dataset User Guide*. London: ONS.

ONS. (2011j). *English Housing Survey, 2009-2010: Household Data*. London: ONS.

ONS. (2011k). *Beyond 2011 Consultation on User Requirements: 17 October 2011-20 January 2012*. London: ONS.

ONS. (2011l). *News Release: ONS takes a fresh look 'Beyond 2011'*. London: ONS.

ONS. (2013). *Living Costs and Food Survey Volume A: Introduction, January – December 2012*. London: ONS.

ONS. (2014a). Social Survey Division and Northern Ireland Statistics and Research Agency. Central Survey Unit, *Quarterly Labour Force Survey, 1992-2013: Secure Access [computer file]. 4th Edition*. Colchester, Essex: UK Data Archive. SN: 6727 , http://dx.doi.org/10.5255/UKDA-SN-6727-5.

ONS. (2014b). *The Census and Future Provision of Population Statistics in England and Wales: Recommendation form the National Statistical and Chief Executive of the UK Statistics Authority*. London: ONS.

ONS. (2014c). *Beyond 2011: ONS response to recommendations from the Independent Review of Methodology*. London: ONS.

Openshaw, S. & Taylor, P.J. (1981). The modifiable areal unit problem, in Wrigley, N. & Bennett, R.J. (eds.), *Quantitative Geography: A British View*. London: Routhledge and Kegan Paul. pp. 60-70.

O'Reilly, D., Rosato, M., Catney, G., Johnston, F. & Brolly, M. (2012). Cohort description: The Northern Ireland Longitudinal Study (NILS), *International Journal of Epidemiology*, 41: 634-641.

Owen, D. (1992). Migration and employment, in Stillwell, J., Rees, P. & Boden, P. (eds.), *Migration processes and trends, (volume II): Population redistribution in the United Kingdom*. London: Belhaven Press. pp. 205-224.

Owen, D. & Green, A. (1992). Labour market experience and occupational change amongst ethnic groups in Great Britain, *New Community*, 19: 7-29.

Parks, A., Kearns, A. & Atkinson, R. (2002). What makes people dissatisfied with their neighbourhood?, *Urban Studies*, 39: 2413-2438.

Permentier, M., van Ham, M. & Bolt, G. (2009). Neighbourhood reputation and the intention to leave the neighbourhood, *Environment and Planning A*, 41: 2162–2180.

Phelps, M. (2011). *ONS Postcode Directory User Guide: 2011 Edition*. London: ONS.

Poston, D.L. & Bouvier, L.F. (2010). *Population and Society: An Introduction to Demography*. Cambridge: Cambridge University Press.

Quigley, J. & Weinberg, D. (1978). Intra-urban residential mobility: A review and synthesis, *International Regional Science Review*, 2: 41-66.

Rabe, B. & Taylor, M. (2010). Residential mobility, quality of neighbourhood and life-course events, *Journal of the Royal Statistical Society, Series A:* 173, 531–555.

Rafferty, A. & Acik-Toprak, N. (2011). *Introductory Guide to the Expenditure and Food Survey*. London: ESDS.

Raper, J., Rhind, D. & Shepherd, J. (1992). *Postcodes: The New Geography*. Harlow: Longman.

Rasbash, J. & Browne, W. (2001). Non hierarchical multilevel models, in Leyland, A. & Goldstein, H. (eds.), *Multilevel Modelling of Health Statistics*. Chichester: Wiley. pp. 93-105.

Rasbash, J., Charlton, C., Jones, K. & Pillinger, R. (2012). *Manual supplement to MLwiN Version 2.26*. Centre for Multilevel Modelling, University of Bristol.

Rasbash, J., Browne, W., Healy, M., Cameron, B. & Charlton, C. (2013). *MLwiN Version 2.29*. Centre for Multilevel Modelling, University of Bristol.

Raudenbush, S.W. & Bryk, A.S. (2002). *Hierarchical Linear Models: Applications and Data Analysis Methods (2nd edition)*. Thousand Oaks, CA: Sage.

Ravenstein, E.G. (1885). The laws of migration, *Journal of the Royal Statistical Society*, 48: 167-227.

Raymer, J., Rees, P., Blake, A., Brown, J. Disney, G., Lomax, N., Norman, P., Stillwell, J. & Boden, P. (2012). *A Conceptual Framework for UK Population and Migration Statistics, Commissioned Report for the Office for National Statistics*. London: ONS.

Rees, P.H. & Boden, P. (2006). *Estimating London's new migrant population: Stage 1 – Review of methodology*. London: Greater London Authority.

Rees, P.H. & Kupiszewski, M. (1999). *Internal Migration and Regional Population Dynamics in Europe*. Strasbourg: Council of Europe Publishing.

Rees, P.H. & Stillwell, J. (1992). Internal migration trends: an overview, in Stillwell, J., Rees, P. & Boden, P. (eds.), *Migration Processes and Trends, (volume II): Population Redistribution in the United Kingdom*. London: Belhaven Press. pp.28-55.

Rees, P.H., Fotheringham, A.S. & Champion, A.G. (2004). Modelling migration for policy analysis, in Stillwell, J. & Clarke, G. (eds.), *Applied GIS and Spatial Analysis.* Chichester: Wiley. pp. 259-296.

Rees, P.H., Stillwell, J., Boden, P. & Dennett, A. (2009). A review of migration statistics literature, in Rees, P.H., Stillwell, J., Boden, P. & Dennett, A. (eds.), *Migration Statistics: The Way Ahead?* London: UKSA.

Rodríguez, G. & Goldman, N. (2001). Improved estimation procedures for multilevel models with binary response: A case-study, *Journal of the Royal Statistical Society: Series A*, 164: 339-355.

Rossi, P. (1955). *Why Families Move: A Study in the Social Phycology of Urban Residential Mobility*. Glencoe, Illinois: Free Press.

Rossi, P.H. & Shlay, A.B. (1982). Residential mobility and public policy issues: "Why Families Move" revisited, *Journal of Social Issues,* 3: 21-34.

Rubin, D.B. (1987). *Multiple Imputation for Nonresponse in Surveys*. New York: Wiley.

Rubin, D.B. (1996). Multiple imputation after 18+ years, *Journal of the American Statistical Association*, 91: 473-89.

Sampson, R.J., Roudenbush, S. & Earls, F. (1997). Neighbourhoods and violent crime: A multilevel study of collective efficacy, *Science*, 277: 918-924.

Sampson, R.J., Morenoff, J.D. & Gannon-Rowley, T. (2002). Assessing “Neighbourhood Effects”: Social processes and new directions in research, *Annual Review of Sociology*, 28: 442-478.

Schelling, T.C. (1969). Models of segregation, *The American Economic Review*, 59: 488-493.

Schelling, T.C. (1971). Dynamic models of segregation, *Journal of Mathematical Sociology*, 1: 143-186.

Simpson, L. & Finney, N. (2009). Spatial patterns of internal migration: Evidence for ethnic groups in Britain, *Population, Space and Place*, 15: 37-56.

Simpson, L., & Tranmer, M. (2005). Combining sample and census data in small area estimates: Iterative proportional fitting with standard software, *Professional Geographer*, 57: 222-234.

Simpson, L., Jivraj, S. & Marquis, N. (2011). *Measuring internal and international migration from the National Pupil Database*, CCSR commissioned report for Department for Communities and Local Government. London: DCLG.

Sjaastad, L.A. (1962). The costs and returns of human migration, *Journal of Political Economy*, 70: 80-93.

Skinner, C.J., Holt, D. & Smith, T.F. (1989). *Analysis of data from complex surveys*. Chichester: John Wiley.

Skinner, C., Hollis, J. & Murphy, M. (2013). *Beyond 2011: Independent Review of Methodology*. London: LSE.

Smith D.P. (2009). Student geographies, *Environment & Planning A*, 41: 1795-1804.

Smith, P.W.F., Raymer, J. & Giuletti, C. (2010). Combining available migration data in England to study economic activity flows over time, *Journal of the Royal Statistical Society, Series A*, 173: 733-753.

Snijders, T.A.B. & Bosker, R.J. (2012). *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modelling (2nd edition)*. London: Sage.

Speare, A., Goldstein, S. & Frey, W. (1975). *Residential Mobility, Migration and Metropolitan Change*. Cambridge, MA: Ballinger.

Spiegelhalter, D.J., Best, N.G., Carlin, B.R. & van der Linde, A. (2002). Bayesian measures of model complexity and fit, *Journal of the Royal Statistical Society, Series B*, 64: 583-616.

Stegmueller, D. (2013). How many countries for multilevel modeling? A comparison of frequentist and Bayesian approaches. *American Journal of Political Science*, 57: 748–761.

Stillwell, J. (1978). Interzonal migration: some historical tests of spatial interaction models, *Environment and Planning A*, 10: 1187-1200.

Stillwell, J. (1991). Spatial interaction models and the propensity to migrate over distance, in Stillwell, J. & Congdon, P. (eds.), *Migration Models: Macro and Micro Approaches*. London: Belhaven Press. pp. 34-56.

Stillwell, J. (2008). Inter-regional migration modelling: a review, Chapter 2, in Poot, J., Waldorf, B. & van Wissen, L. (eds.) *Migration and Human Capital, New Horizons in Regional Science Series*. Cheltenham: Edward Elgar. pp. 29-48.

Stillwell, J. (2010). Ethnic population concentration and net migration in London, *Environment and Planning A*, 42: 1439-1456.

Stillwell, J. & Congdon, P. (eds.) (1991). *Migration Models: Macro and Micro Approaches*. London: Belhaven.

Stillwell, J. & 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*, 170: 425–445.

Stillwell, J. & Harland, H. (2010). Analysing interaction data, in Stillwell, J., Duke-Williams, O., and Dennett, A. (eds.), *Technologies for Migration and Commuting Analysis: Spatial Interaction Data Applications*. Hershey: IGI Global. pp. 69-88.

Stillwell, J. & Hussain, S. (2010). Exploring the ethnic dimension of internal migration in Great Britain using migration effectiveness and spatial connectivity, *Journal of Ethnic and Migration Studies*, 36: 1381-1403.

Stillwell, J., Boden, P. & Dennett, A. (2011). Monitoring who moves where: Information systems for internal and international migration, in Stillwell, J. & Clarke, M. (eds.), *Population Dynamics and Projection Methods, Understanding Population Trends and Processes (volume IV)*. Berlin: Springer. pp. 115-140.

Stillwell, J., Bell, M., Blake, M., Duke-Williams, O. & 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: 17-38.

Stillwell, J., Bell, M., Blake, M., Duke-Williams, O. & Rees, P. (2001). Net migration and migration effectiveness: A comparison between Australia and the United Kingdom, 1976–96 Part 2: Age-related patterns, *Journal of Population Research*, 18: 19-39.

Stockdale, A. (2004). Rural out-migration: Community consequences and individual migrant experiences, *Sociologia Ruralis*, 44: 149-176.

Stockdale, A. (2010). The diverse geographies of rural gentrification in Scotland, *Journal of Rural Studies*, 26: 31-40.

Sturgis, P. Brunton-Smith, I., Read, S. & Allum, N. (2011). Does ethnic diversity erode trust? Putnam's 'Hunkering Down' thesis reconsidered, *British Journal of Political Science*, 41: 57-82.

Sturgis, P., Brunton-Smith, I., Kuha, J. & Jackson, J. (2013). Ethnic diversity, segregation and social cohesion of neighbourhoods in London, *Ethnic and Racial Studies* (published online).

Subramanian, S. V. (2004a). Multilevel methods, theory and analysis, in Anderson, N. (ed.), *Encyclopaedia on Health and Behaviour*, Thousand Oaks, CA: Sage. pp. 602-608.

Subramanian, S.V. (2004b). The relevance of multilevel statistical methods for identifying causal neighbourhood effects, *Social Science & Medicine*, 58: 1961-1967.

Subramanian, S.V., Jones, K., Kaddour, A. & Krieger, N. (2009). Revisiting Robinson: The perils of individualistic and ecologic fallacy, *International Journal of Epidemiology*, 38: 342-360.

Taylor, R.B. & Covington, J. (1993). Community structural change and fear of crime, *Social Problems*, 40: 374-397.

Taylor, J., Tigg, L. & Mohan, J. (2010). Exploring the links between population heterogeneity and perceptions of social cohesion in England, in Stillwell, J. and van Ham, M. (eds.), *Ethnicity and Integration*. New York: Springer. pp. 269-288.

Thomas, D.S. (1938). *Research Memorandum of Migration Differentials*. New York: Social Science Research Council.

Thompson, C., Stillwell, J., Clarke, M. & Bradbrook, C. (2010). Understanding and validating Acxiom's Research Opinion Poll data, *Working Paper 10/06* Leeds: School of Geography, University of Leeds.

Tienda, M. (1991). Poor people and poor places: Deciphering neighbourhood effects on poverty outcomes, in Huber, J. (ed.), *Macro-Micro Linkages in Sociology*. Newburgh Park, CA: Sage. pp. 204-212.

Traynor, J. (2011). *Overview, Presentation at the 2011 Census Workshop Series on Specialist Output Products: Origin-Destination (Flow) Products*, Union Jack Club, London, 26 October.

van Der Linde A. (2005). DIC in variable selection. *Statistica Neerlandica*, 59: 45-56.

van Ham, M. & Clark, W.A.V. (2009). Neighbourhood mobility in context: household moves and changing neighbourhoods in the Netherlands, *Environment and Planning A*, 41: 1442-1459.

van Ham, M. & Feijten, P. (2008). Who wants to leave the neighbourhood? The effect of being different from the neighbourhood population on wishes to move, *Environment and Planning A*, 40: 1151–1170.

van Ham, M., Mulder, C.H. & Hooimeijer, P. (2001). Spatial flexibility in job mobility: Macro-level opportunities and micro-level restrictions, *Environment and Planning A*, 33: 921-940.

van Ham, M., Hedman, L., Manley, D., Coulter, R. & Östh, J. (2014). Intergenerational transmission of neighbourhood poverty: an analysis of neighbourhood histories of individuals, *Transactions of the Institute of British Geographers*, 39: 402-417.

Vickers, D. & Rees, P.H. (2006). Introducing the National Classification of Census Output Areas, *Population Trends*, 125: 15-29.

Vickers, D. & Rees, P.H. (2007). Creating the UK National Statistics 2001 output area classification, *Journal of the Royal Statistical Society, Series A*, 170: 379-403.

Walthery, P. (2011). *Introductory guide to the Integrated Household Survey*. London: ESDS.

Werner, B. (2006). *Reflections on Fifteen Years of Change in Using the Labour Force Survey*. London: ONS.

Wilson, A. (1970). *Entropy in Urban and Regional Modelling. Monographs in Spatial and Environmental Systems Analysis*. London: Pion.

Wilson, A. (1971). A family of spatial interaction models, and associated developments, *Environment and Planning A*, 3:1–32.

Wulff, M., Champion, A. & Lobo, M. (2010). Household diversity and migration in midlife: Understanding residential mobility among 45-64 year olds in Melbourne, Australia. *Population, Space and Place*, 16: 307-321.

Zipf, G. (1946). The P1 P2 / D hypothesis: On the intercity movement of persons, *American Sociological Review*, 11: 677-686.

List of Abbreviations

AIC	Akaike Information Criterion
APS	Annual Population Survey
BHPS	British Household Panel Study
CAS	Census Area Statistics
CAMS	Controlled Access Microdata Sample
DIC	Deviance Information Criterion
EHS	English Housing Survey
EU-SILC	EU Statistics on Income and Living Conditions
FRS	Family Resources Survey
GB	Great Britain
GLF	General Lifestyle Survey
GOF	Goodness of Fit
GOR	Government Office Region
GROS	General Register Office for Scotland
HA	Health Authority
HESA	Higher Education Statistics Agency
IGLS	Iterative Generalised Least Squares
IHS	Integrated Household Survey
IID	Independently and Identically Distributed
IMD	Index of Multiple Deprivation
LA	Local Authority
LAD	Local Authority District
LCF	Living Cost and Food Survey
LFS	Labour Force Survey
LOS	Life Opportunities Survey
LS	Longitudinal Studies
LSOA	Lower Super Output Area

MCMC	Markov Chain Monte Carlo
MSOA	Middle Super Output Area
NHS	National Health Service
NHSCR	National Health Service Central Register
NILS	Northern Ireland Longitudinal Study
NISRA	Northern Ireland Statistics and Research Agency
OA	Output Area
OAC	Census 2001 Output Area Classification
OLS	Ordinary Least Squares
ONS	Office for National Statistics
ONS-LS	ONS Longitudinal Study of England and Wales
PAF	Postal Address File
PRDS	Patient Register Data System
PR-NHSCR	Patient Register and National Health Service Central Register
RIGLS	Restricted Iterative Generalised Least Squares
ROP	Research Opinion Poll
SAM	Small Area Microdata
SAR	Samples of Anonymised Records
SLS	Scottish Longitudinal Study
SMS	Special Migration Statistics
UA	Unitary Authority
UK	United Kingdom
UKHLS	UK Household Longitudinal Study (Understanding Society)
VPC	Variance Partitioning Coefficient