THE STATISTICAL ANALYSIS AND MODELLING OF INTERNAL MIGRATION FLOWS WITHIN ENGLAND AND WALES

by

Stamatis Kalogirou

M.Sc., Leicester University, 2001

A THESIS SUBMITTED IN PARTIAL FULFILMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY
in the School of
Geography, Politics and Sociology

University of Newcastle

July 2003

To my parents

Abstract

Migration flows in recent decades suggest that Britain is a nation of people *on the move*. The combination of information technology and structured data collection allow today a very close and detailed examination of trends in internal migration in England and Wales. However, there has been a relative dearth in the analysis of migration data using quantitative techniques.

In this thesis I attempt to analyse migration patterns and to model migration moves in order to explain the main factors affecting individuals' migration decisions. I try to link this work to existing research in this research field by revising and applying recently developed quantitative methods. My main aim in this thesis is to provide empirical evidence that the effects of many socio-economic factors on individuals' migration decisions are non-stationary across space.

More specifically, there are four sections of data analysis in this thesis: the exploration of migration flows using data visualization and local statistics; the analysis of the effects of socio-economic factors on out-migration rates; the analysis of the attraction of migrants from areas with varying socio-economic profiles via the construction of global and local models; and the examination of model residuals. The main objectives of this work are twofold: the first is to provide a thorough investigation of internal migration using a rich dataset on annual migration during the 1980s and the 1990s; the second is to remove the inaccuracy traditional global models introduce by assuming that the processes being examined are stationary over space. I do this through the use of newly developed local statistical methods.

Former attempts made to provide local forms of statistical analysis have limitations. *Geographically Weighted Regression* is used in this thesis to allow for local modelling. This method provides not only a technique for best model fit but also for the evaluation of the results using modern goodness of fit statistics such as the Akaike Information Criterion.

The temporal dimension of my data allows the examination of the stability in migration flows over time. It also provides a means of checking the consistency of the significance of the spatial variation of local parameter estimates derived from each annual set of migration data. The migration data themselves are disaggregated in 14 sex/age groups. The age disaggregation reflects the stage of life individuals are at (e.g. people forming a family are young adults aged 25 – 29 years old).

In order to facilitate the examination of migratory moves over time, I introduce a new way of visualising in-, out- and net migration rates, the *heat map*. The results for the migration models show that the parameter estimates of some of the migration determinants

exhibit significant spatial variation. This suggests that the effect of some determinants on migration decisions in both origin and destination of a migratory move vary across space. The spatial patterns of the local parameter estimates usually show a North-South or a Northwest – Southeast divide.

When out-migration models are concerned, there is strong evidence for a spatially variable effect of employment rate for all migrant groups and percentage non-white population along with percentage long-distance commuters for mature male adults. When destination choice models are concerned, there is strong evidence for a spatially variable effect of destination accessibility, house prices, listed buildings, vacant and derelict dwellings, distance and total population.

These new findings on local migration modelling are of high interest and potential benefit to policy makers. The spatial and temporal migration trends confirm the continuation of the counterurbanisation phenomenon in England and Wales. The local out-migration models suggest the effect of some ecological conditions on out-migration is associated with the location of the origin. The local destination choice models suggest that there are differences on what determines short migration moves and longer moves. They also suggest that the behaviour of those leaving an area is not stationary for all England and Wales. Finally, similarly to out-migration, there are instances where the effect of some ecological conditions on destination choice is associated with the location of the destination.

This thesis also presents an attempt for constructing more robust migration models, signalling the need for additional migration determinants.

Acknowledgements

Getting a PhD after three-year's hard work is a highly personal achievement, but it would never be possible without the precious support of several people whose help I would now like to acknowledge. These three years were for me not only a time to form and examine some research questions and to develop a personal research agenda and academic style, but more importantly to understand and express myself through research in human geography and communicate my creativity through publishable research material. My PhD time was not free of moments of loneliness and depression but I was lucky to have personal support that helped me get over such moments. There were also many periods of satisfaction for research outcomes and relaxing nights full of humour and fun in Newcastle, Durham, and elsewhere.

First of all, I would like to thank my supervisor, Stewart Fotheringham, who has been a good teacher, tutor and mentor to me. I also thank him for his trust in my skills and personal values and his offer of opportunities for consultancy and teaching assistance. Stewart has been more that supportive throughout my PhD studies. Not only has he taught me some quantitative geography, but also many times he has guided me on how to survive in the competitive academic world and has helped me in my career development. His personal academic achievements are a source of inspiration and guidance to me and I wish him to continue being an important member of our academic community for many more years. I would also like to thank all the members of the Spatial Analysis Research Group at the University of Newcastle, especially Seraphim Alvanides, Chris Brunsdon and Martin Charlton for their support in academic and personal issues. Seraphim has been a close friend to me and his support in many work and life matters has been very precious. Tony Champion has been also supportive throughout my PhD research, providing valuable advice on population-related research issues. I would like to thank him for these as well as for his valuable support in my future career development.

I would like to express my continuing gratitude to the School of Geography, Politics and Sociology for the honour of awarding me a three-year departmental studentship. Many thanks to all individuals involved in this award and later during my studies. Special thanks to Mrs Judith Houston for her advice and support in the early days, Nicola Kelly for her excellent support with financial and other administrative issues, and George Kania for his help with technical computer issues. Also, Andy Pike for his support as a postgraduate director, especially his monitoring of my progress.

It is necessary to acknowledge the provision of the original dataset for this research which was part of the early phase of a research project for the Office of Deputy Prime Minister (ODPM), formerly the Department of Transport, Local Government and the Regions (DTLR) under contract RADS 5/9/22. The map of the UK is the Crown copyright.

I would like to thank my student colleagues for their advice in everyday postgraduate issues, for sharing their academic experiences as well as their enthusiasm for doing research. I would also like to thank my housemates Konstantinos and Nikolaos for their patience at moments of stress and for making a pleasant and friendly environment at home. I wish them all the best for their studies and future life.

Many friends, here and back home, have kindly offered me their emotional support. They encouraged me in continuing my postgraduate studies, several times they reminded me what was the purpose of doing a PhD and helped me not to lose my self-confidence. I would like to express my appreciation and love to all of them. I would also like to thank my special friend Nancy for her emotional support in recent months that gave me the strength to work long nights. She has inspired me to overcome difficulties and she helped me to appreciate more some human values, which I sometimes forget during intensive work.

Finally, I thank my immediate family in Hellas, especially my mother, for being supportive and believing in my efforts all these years. They have been the light in this long journey in the world of academic knowledge and the understanding of human behaviour. I love and miss them and wish them the best in their lives.

Table of Contents

List of Figures

List of Tables

Chapter 1
Introduction

Migration is the *permanent or semi-permanent change of residence by an individual or group of people* (Johnston et al., 2000, p. 504). Zelinsky (1971) explains that answering who is a migrant is not a simple question and he suggests, *genuine migration obviously means a perceptible and simultaneous shift in both spatial and social locus, so that the student cannot realistically measure one kind of movement while he ignores the other* (Zelinsky, 1971, p. 224). There is a distinction between migration and commuting. Migration involves the change of main residence for a relatively long period of time, whereas commuting involves a journey-to-work for a limited period of time, usually a few hours or days.

The collection of statistical data on migration requires a boundary of some sort to have been crossed and a certain length of time to have been spent over that boundary in the new area of residence (Johnston et al., 2000, p. 504). Thus, there are several kinds of migration in terms of the geography of areas and their boundaries. A few examples are: internal migration; international migration (emigration and immigration); inter-regional migration (between regions); urban-rural migration (urban and rural areas interaction); and intra-urban migration (within an urban area or city, referred to also as residential mobility). Internal migration involves a change of residence within a country, whereas international migration involves a change of residence between countries (going abroad or coming from abroad).

Undoubtedly, internal migration is an important process with many implications for local communities. It is a field with an agenda in many disciplines related to human studies, quintessential economic development and policy making, being concerned with movements over space, but it remains a geographical problem. Migration is a broad term: here I look at migrants who change their residence and probably their employment in a major career move; teenagers changing residence for higher education or work; highly educated professionals moving to places with better working and living opportunities; families with children moving to places with better housing, lower crime rate and educational facilities of high standards; and retired people moving to more pleasant environments.

This work focuses on the geographical elements of internal migration. However, it is hard to separate the various determinants of migration decisions. I examine a large dataset on migration within England and Wales, starting with identifying the main flows and population interactions across the country and culminating in a model that can predict the number of migrants from any location in England and Wales and distribute these migrants to any set of locations within England and Wales.

Although one can argue that migration is a well-researched topic, all previous work in the literature contains limitations in explaining migration decisions. For example, previous empirical work in modelling migration assumes that the role of migration determinants in producing or attracting migrants is spatially stationary. It is possible to remove this assumption by **investigating the existence of spatial non-stationarity in the role of migration determinants.**

The development of a statistical technique for spatially disaggregated modelling (Geographically Weighted Regression) allows the investigation of the existence of spatial non-stationarity in migration processes. It is also important to note that the power of recent GIS and statistical software allows more efficient visualisation of the data and model results. Visual forms of large volumes of data and results help in the better understanding and modelling of trends in migration processes.

1.1 Motivations and Innovations

This research will identify, explore and explain trends in internal migration within England and Wales. This includes spatial and temporal aspects of population mobility as well as identifying the factors that contribute to the production and attraction of migrants. In addressing this overall objective, the project's aims are:

- To evaluate empirical work in migration modelling.
- To identify and explain the spatial and temporal trends of out- and in- and net migration in FHSAs in England and Wales between 1984 and 1998.
- To understand and model the factors influencing the production of migrants at an origin (departure decision-making process)
- To understand and model the factors influencing the attraction of migrants from a destination (destination choice process).
- To examine if significant non-stationarity in the parameter estimates of migration models exists.
- To examine the presence of large residuals in estimated migration rates and to provide possible explanations for these.

The motivation for this work is the opportunity to explore new migration data recently made available and to introduce a new way of modelling migration using these data. For the former I introduce a new technique for visualising migration data which I call ***heat maps*** whereas for the latter, the development of a new technique called Geographically Weighted Regression (GWR), (Brunsdon et al., 1996; 1998a; 1998b; 1999a; 1999b; and Fotheringham et al., 1996; 1997a; 1997b; 1998; 2000; 2002a), allows the comparison between global and local migration models (both out-migration and destination choice models). The latter allow the examination of the research question of ***whether parameter estimates in a local model exhibit a significant spatial variation***. This research question is based on the assumption that location matters to the way migration determinants affect an individual's decision to migrate and to his/her destination choice.

1.2 Scope of the thesis

Human migration is a very broad term. There are a few thousand studies in the literature and several text books written on it. It is a social phenomenon as old as the existence of human beings. In terms of geographical scale, migration can be a move from one side of a city to another or from one continent to another. In terms of methodology there are two main streams of investigation: qualitative and quantitative. Both modes of analysis have many variations based on the theoretical underpinnings of each specific method. In terms of disciplines, migration can be seen from the point of view of an economist, a sociologist, a geographer, a demographer, a statistician, a politician, a planner, a historian, an archaeologist, a psychologist, to list only a few. In terms of the profile of the interaction areas (social, economical, cultural, political), there are several types of migration streams such as rural-urban migration, migration from less developed areas to more developed areas, migration from areas of conflict (religious, war, cultural) to more relaxed areas, and migration from countries/regions that human rights are not respected to regions where they are. In terms of population groups, migration data can be disaggregated along many different lines such as sex, age, occupation, marital status, ethnic origin, and social class.

Here, human migration is looked at internally within a country, at a middle level of geographical aggregation (Family Health Service Areas: a combination of districts and counties in England and Wales). In terms of the methodology used, quantitative methods have been applied including mainly contemporary statistical techniques along with advanced goodness-of-fit measures to check the quality and robustness of the results. Migration is seen from a geo-computational point of view, with focus being given on the spatial and temporal elements of a statistical modelling of the factors that affect out-migration and destination choice.

1.3 Organisation of the thesis

Chapter 2 contains a review of previous work focusing on migration trends in the 1980's and 1990's in developed countries and mainly in England and Wales; a discussion of the significance of the inclusion of socio-economic and other explanatory variables in the models (Chapters 6 and 7); a listing of the findings of previous empirical work in migration modelling; and a discussion of some technical issues (statistical and computational).

A discussion about data quality issues as well as a presentation of the dataset used here follows in Chapter 3. Some techniques used to calculate some of the variables, such as the Principal Component Analysis (PCA), are also discussed in this chapter.

In Chapter 4 an analytical presentation of the methodology used here is made. There are also some methodological issues discussed to defend the methodology used. Migration modelling techniques and goodness-of-fit measurement techniques are mainly discussed.

Chapters 5 to 7 are the "Analysis and Results" part of this thesis, although Chapter 8 contains sample analysis on model residuals. There are three parts; Chapter 5 contains the presentation and explanation of out-, in- and net migration rates as well as migration flows; Chapters 6 and 7 discuss the two stages of migration modelling, i.e. global and local models of out-migration and destination choice, respectively; Chapter 6 also contains a discussion on model residuals and ways of reducing them. Chapter 8 is a summary of the thesis and some conclusions of this work. The cited references follow at the end of this thesis.

1.4 Summary

There is an on-going interest in understanding the determinants of migration within England and Wales. In my research, I extend on existing work in two directions: migration data visualisation and local forms of migration modelling.

In order to provide a better means of exploring migration data (in-, out- and net migration) over time I introduce here a new means of data visualisation: the *heat map*.

I calibrate local models of migration for both stages of migration decisions: out-migration and destination choice. As a result, I find that migrants' responses to ecological variables that determine migration decisions are not stationary across space as is often assumed.

In the remainder of this thesis, I put this research within a theoretical framework and I present the methodologies and the analysis necessary to provide empirical evidence for my findings. I make use of data visualization (graphs and maps) to communicate many of my results. I also link my findings with existing findings in the migration literature.

In this first chapter, I defined migration and I discussed the kind of migration I am going to look at. I introduced my research questions and I explained why I believe this research is innovative. I listed the aims of this study, which I am going to address in the following chapters. I also discussed the scope of my thesis in order to clarify the dimensions in which I study migration. Finally, I provided an outline of the thesis.

I now review migration trends, methodologies and empirical findings of migration models and some theoretical underpinnings of my research.

Chapter 2
Background

In this chapter a review of existing work in internal migration is presented. There is a multitude of publications concerning models, trends and empirical work in internal migration for different countries, geographical scales and time periods. The earliest examples that can be called *scientific studies of internal migration* are Ravenstein's (1885; 1889) papers. A history of the early contributions to the scientific study of migration is provided in Greenwood and Hunt (2003) covering the period between the 1890s and the 1940s. Here I focus on migration studies in developed countries since the 1950s.

Inter-state (US) or interregional (Europe) flows are those typically studied in the early works in exploring and explaining internal migration. However, the geographical scale used in these studies often can hide important migration trends. This is because of the averaging of migration and its determinants over large populations can remove significant variation in the data. This smoothing effect is less noticeable when migration processes are investigated at finer geographical scales.

Selecting the geographical scale of areas (geographical units) across which migration trends should be studied and migration models calibrated is important for the interpretation of the results and their value for policymaking. For aggregated data analysis, there are several geographical frameworks dividing the UK into small area units. The census geography (e.g., wards, districts) which defines local authority administration areas is the most frequently used in migration studies. For practical reasons, usually, the data availability necessitates the use of a specific geography and geographical scale. Here migration data from the National Health Service Central Registrar (NHSCR) are available. The smallest geographical units for which these data are available are the Family Health Service Areas (FHSAs). More information about the construction of FHSAs follows.

Theoretically, smaller geographical units have the advantage of having distinct socio-economic profiles and independence within an interaction system. In the postal geography, the smallest geographical unit is that defined by a postcode whereas in the census geography it is the enumeration district (ED). Disadvantages of small geographical units are: limitations in data availability both in migration flow data and migration determinants data; restrictions in analysis because of confidentiality issues; and practical problems in modelling because of small and zero migration flows.

The confidentiality issues concern the identification of individual migrants within the data during analysis and after publication. To avoid such identification, offices publish aggregated data at certain geographical scales. At small geographical scales (e.g., EDs, wards), especially in rural areas it is more likely that there are no migrants leaving or coming to an area, and many of the inter-area flows are zero, causing problems in statistical analysis. There are two problems associated with zero or low flows.

One problem is that low volumes of migration are more subject to idiosyncratic decisions, which we do not want to capture in terms of identifying what *environmental* variables (e.g. labour market variables) etc. affect migration behaviour.

The second problem occurs because some methods of statistical analysis used here (e.g. regression methods) require that the dependent variable (here migration flows) needs to be logged during the calibration process. A zero flow cannot be logged, and thus, it is impossible to model zero flows. Some methods (e.g., Poisson) overcome this problem but at a cost of fewer diagnostics and less obvious interpretations of parameter estimates.

The problems associated with very low flows are eliminated at coarser geographical scales. However, at coarser geographical scales the effects of some of the variables determining migration (e.g. house prices, crime rate) are averaged over large areas and thus may be quite misleading.

Champion (1989, p. 84) suggests that for studying the population deconcentration phenomenon in the UK, which matches one aim of this study (exploring migration in England and Wales), *the principal alternatives are between functionally defined urban regions, physically defined urban areas and administratively defined local authority (municipality units).* Coombes et al. (1982) and Champion et al. (1984, 1987) have strongly argued against the use of local authority areas favouring the use of the Functional Regions family of areas (Champion, 1989). The new (1970s) local government areas (districts), for which census data are reported, *have been defined for administrative purposes and do not necessarily represent meaningful geographical units like towns and cities* (Champion et al., 1984, p. 187), whereas Functional Region zones, constructed from EDs, do. In line with the former, Bell et al. (2002, p. 439) note that the administration units for which migration data are made available by their providers *rarely have any functional basis, bearing little relationship to the underlying distribution of socioeconomic variables.* Champion (1989, p.85) reports that *in practice, data problems make it necessary to use a combination of geographical frameworks in order to build up a full picture of counterurbanization in Britain. The Functional Regions family forms the most comprehensive and up-to-date framework for studying urban change in Britain.* Champion (1989) argues that a *key advantage* of his study is the use of Local Labour Market

Areas (LLMAs). These are based on journey-to-work flows to employment centres. For Great Britain, there are 280 LLMAs.

Although it is beyond the scope of this work to provide empirical evidence for or against the use of LLMAs for studying population trends in the UK, I believe that the geographical scale of Local Authorities (Districts in England and Wales Census) is appropriate for analysis of such trends (based on aggregated migration data). This is because at this scale, most of the areas are population structures (towns or cities) that are expected to have their own cultural identity. The residents of a single town or city are also expected to have a degree of common behaviour. The effects of various socioeconomic factors on people's migration decisions are expected to be more stationary within city limits than between different cities. This would not necessarily be the case at coarser geographical scales (e.g., county, region). The data problems discussed above (zero flows) also apply at this geographical scale. Additionally, there are not many socioeconomic data available on an annual basis to allow temporal analysis. In summary, there are data constraints that do not allow robust statistical analysis (migration modelling) at most of the finer geographical scales discussed above.

Unfortunately, depending on the geographical scale of spatial data analysis, problems of inaccuracy and misspecification may arise. Fotheringham and Brunsdon (1999, p. 347) suggest that *modelling spatial behaviour (here migration) at the individual level is prone to the atomistic fallacy, missing the context in which individual behaviour occurs (Alker, 1969), whereas modelling behaviour at the aggregate level is prone to the ecological fallacy, that the results may not apply to individual behaviour (Robinson, 1950).* Fotheringham and Rogerson (1993) discuss issues concerning aggregate and disaggregate models. They state, *behavioural theorists argue that it is impossible to learn anything about individual behaviour if aggregate models are used, while modellers of aggregate phenomena argue that behavioural models do not give sufficient insight into systemwide behaviour* (Fotheringham and Rogerson, 1993, p. 15).

Ecological, in behavioural studies, is a term that refers to a group of people: for example, *ecological correlation* is the correlation between variables that are measured with aggregated data (e.g., migration and percentage illiterate). On the other hand an *individual correlation* is a correlation between variables that are measurements of individual characteristics (e.g., decision to migrate and family status). Robinson (1950) concludes that ecological correlations cannot validly be used as substitutes for individual correlations. Thus, it is false to argue about the behaviour of specific individuals based on the results of aggregate data analysis. Alker (1969) re-examines Robinson's findings and discusses a typology of ecological fallacies. From his work, which contains the necessary mathematical reasoning (covariance theorems) for his arguments, the ecological fallacy and the individualistic fallacy are relevant to Fotheringham and Brunsdon's (1999, p. 347) argument discussed above. Alker (1969) suggests that it is not possible to generalise conclusions made at one level of analysis (e.g. aggregate or individual) to another level (individual or aggregate) unless such inference is proved. However, Amrhein and Flowerdew (1992) in their analysis of Canadian migration at different scales provide empirical evidence that the behaviour of migration determinants is scale independent. In this thesis, ecological correlations are discussed, thus the conclusions refer to groups of persons. However, the fact that the aggregation of the data is such that the people in each group will have similar characteristics, makes these conclusions of significant interest.

The NHSCR data used here are only available for the 98 FHSAs in England and Wales. Some of the FHSAs have the same boundaries as counties, such as the shire counties. In highly populated areas, such as the metropolitan areas of Newcastle, Leeds, Manchester, and Birmingham, the FHSA boundaries match the district boundaries. For example, Leicestershire is a single FHSA, while Tyne and Wear has five FHSAs: Newcastle; Gateshead; North Tyneside; South Tyneside and Sunderland. The boundaries of the 16 FHSAs in London are groupings of the 32 London Boroughs. Figure 3.2 in Chapter 3 is a map showing the FHSAs in London, while Figure 3.1 shows all FHSAs in England and Wales. It is necessary to recognise that the fact some of the FHSAs match the county geography is a limitation of this study in terms of geographical scale.

An important issue concerns the sex-age disaggregation of the migration data. Many of the studies to date are based on aggregated migrant flows (Sommers and Suits, 1973; Miller, 1973; Congdon, 1989; Fik et al., 1992; Boyle and Flowerdew, 1997). One of the common arguments is that the sex-age disaggregation has the disadvantage of producing some very small values in some of the origin-destination flows. The disaggregated flow matrices might consist of large numbers of zero flows, that make statistical analysis problematic as discussed above. However, modelling total number of migrants, rather than the age-sex disaggregated values, may result in misinterpretation of the parameter estimates. The reason for this is that different sex-age groups have varying behaviour in their decision to migrate; they value differently the factors affecting this decision.

Finally, many of the previous studies in migration modelling include only population, distance and competition variables to explain migration. They provide evidence that these alone can explain most of the variation in migration flows. However, it is important to include more factors such as economic, housing, and environmental conditions in the origins and destinations. The additional variables will explain some effects population and distance fail to do and will also reduce the chances that the estimated population and distance parameters act as surrogates for unknown effects.

2.1 Trends in Migration

In this section trends in migration flows (out- and in- migration, origin destination flows) in the developed countries are presented. Trends include the relation between age and migration flows; temporal variation of migration flows; and spatial variation of migration flows. A more detailed discussion on trends in England and Wales during the last three decades follows.

2.1.1 Age-specific migration rates

In recent years, a strong attempt has been made to explore and explain age-disaggregated migration flows (Rogers et al., 2002). Rogers et al. (1978) introduced the *construction of model migration schedules.* Their aim was to capture the regularities exhibited by age patterns in observed migration rates (or age profile of migration schedules), as these *seem to be repeated, with only minor differences, in virtually all developed and developing nations of the globe* (Rogers et al., 1978, p. 475). They refer to Long's (1973) age-specific annual migration rates graph (Figure 2.1), which seems to be an interpolation of the histogram of single year total migration rates from 0 to 70 years old people. The description of the empirically obtained age-specific migration curve, which Rogers et al. (1978) modelled in their work follows.

Migration, viewed as an event, is highly selective with regard to age, with young adults generally being the most mobile group in any population. Levels of migration are also high among children, varying from a peak during the first year of life (the initial peak) to a low point around the age of sixteen. The migration age profile then turns sharply upward until it reaches a second peak (the high peak) in the neighborhood of 22 years, after which it declines regularly with age, except for a slight hump (the retirement peak) around the ages of 62 to 65 (Rogers et al., 1978, p. 476).

Rogers et al. (1978) suggest two alternative approaches to model the age-specific migration curve: the mortality approach, where regression coefficients are calculated for 5-year interval age groups; and the fertility approach, which incorporates a curve fitting. The age-specific migration curve can be decomposed into three curves: a single negative-exponential and two skewed unimodal bell-shaped functions. Bates and Bracken (1982) and

Bracken and Bates (1983) revisited Rogers' et al. (1978) work using data from England and Wales. A new approach (logit models) on modelling the age and spatial structures of migration is provided by Rogers et al. (2002) along the lines of the mortality approach.

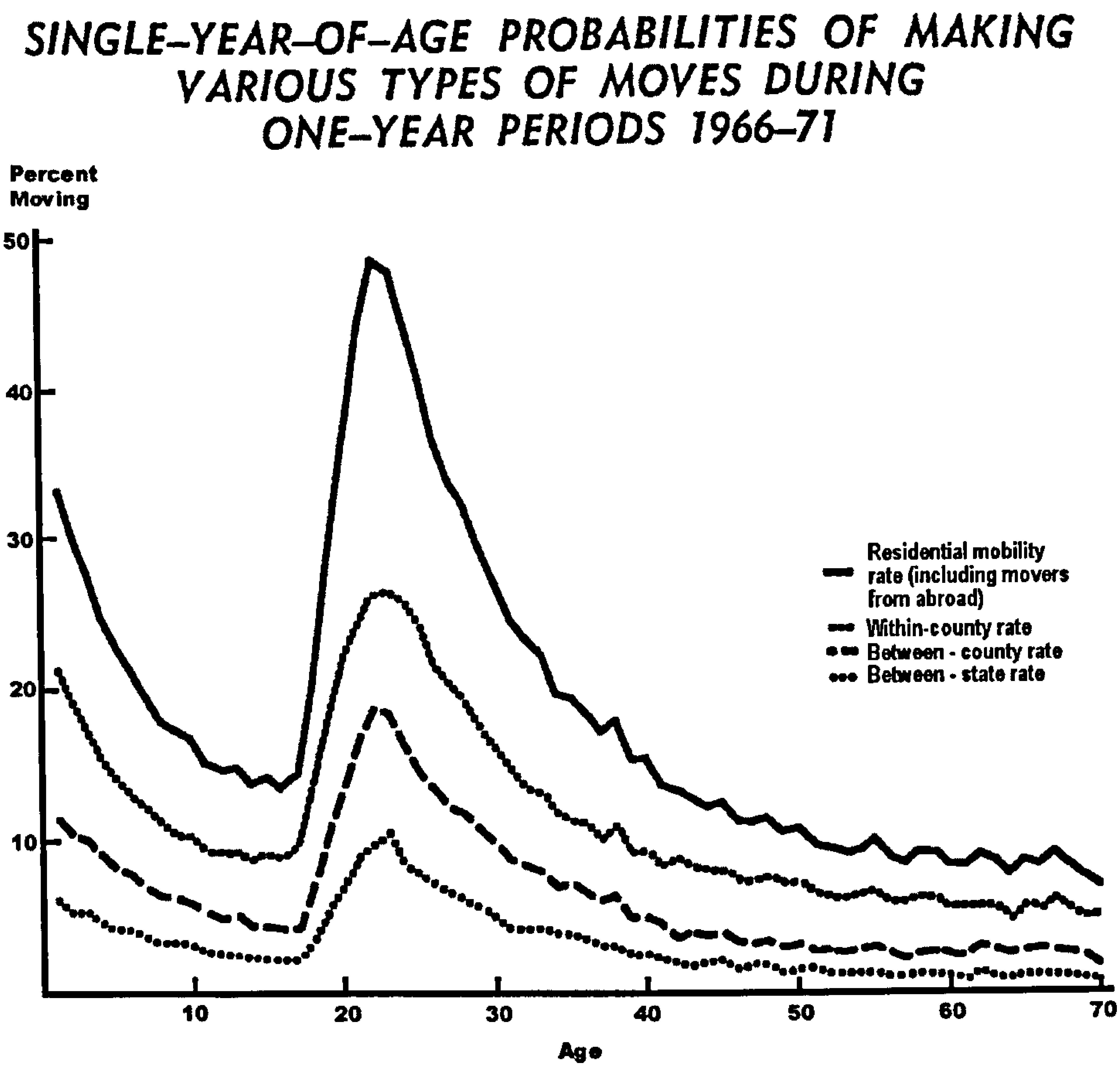


Figure 2.1. Age-specific migration rates of total US population
(Source: Long, 1973, p. 38)

2.1.2 Migration Flows in England and Wales, 1970's – 1990's

In this section a review of publications on trends in internal migration in the UK over the last two decades is provided. Spatial and temporal trends are described for total population as well as for sex and age groups. The temporal trends discussed in the literature are usually based on Census Special Migration Statistics analysis, and less frequently based on the NHSCR migration data. For the former data, measures take place every decade, and for the latter quarterly.

2.1.2.1 Census of Population Special Migration Statistics

The general findings for the 1981 and 1991 Census of Populations Special Migration Statistics (SMS) suggest a persistent counterurbanisation trend that had started during the 1970s. Atkins et al., 1996, conclude that this trend *is widespread and set to continue, despite* *slowing down in more recent years for most cities and going into reverse in Inner London as far as overall population change is concerned* (Atkins et al., 1996, p. 9).

Counterurbanisation refers to the trend of population deconcentration away from large urban settlements towards more rural areas. In his comprehensive review of this trend, Champion (1989, Chapter 2) discusses Berry's counterurbanisation thesis. According to Berry, *Counterurbanization is a process of population deconcentration; it implies a movement from a state of more concentration to a state of less concentration* (Berry, 1976, p. 17; 1980, p. 21; Champion, 1989, p. 20)

National Economic Indicators for England suggest 1990-91 was a period of economic recession. Thus, it is expected that migratory moves in this period were lower compared to those in years of economic growth. Indeed the 1991 Census SMS Statistics show lower numbers of migrants compared to the mid-1980s and mid-1990s NHSCR migration statistics. Nevertheless, the 1991 Census of population in England alone recorded 4.6 million migrants, which is 9.8% of all English residents enumerated by the same census. More than half of these migrants did not cross the boundary of their local authority district and 56% of the total migrants moved to an address less than 5km from their old one (Atkins et al., 1996).

Atkins et al. (1996) also find that the trend of total migrants derived from the 1991 Census for England confirms the counterurbanisation effect. In terms of mobility, those aged less than 16 and over 45 years old are the least mobile, those aged 30 – 44 are averagely mobile and those aged 16 – 29 are the most mobile residents. In terms of spatial patterns, urban areas are net losers of all people with the exception of young people (16 – 29) from outer London Boroughs. Areas with traditions of mining and heavy industry are also net migration losers (Champion et al., 1996). Net gainers of population are urban-rural districts, remoter mainly rural districts and resort, port and retirement districts. However, in the work by Champion et al. (1996) the age group 16 – 29 is a mixture of people in different life stages, thus the trends they suggest are potentially misleading. This is demonstrated in Chapter 5, where it is suggested that urban areas with universities are net gainers of student populations and net losers of university graduates (except London).

The counterurbanisation phenomenon in Britain during the 1970s and the 1980s is also confirmed by an analysis based on the LLMA framework (Champion, 1989) concluding a very general and widespread process of deconcentration in Britain. The latter study identified the following trends: longer-distance deconcentration has been underway for most of the postwar period and was particularly noticeable in the 1970s; the contribution of retirement migration was fairly modest in the 1970s and the 1980s; and generally, counterurbanisation in Britain appears to have involved all ages in fairly equal proportions (Champion, 1989).

Similar findings are reported in Champion (1994) along with comparisons of the internal migration in Britain with that in the US and some western European countries.

2.1.2.2 National Health Service Central Registrar (NHSCR) migration data

Since the NHSCR data are available for this study (for the period 1983 to 1998) it is interesting to focus on reviewing previous research using the same source of data. Stillwell et al. (1995) examines migration trends at three levels of geography (North-South, Regional, FHSA level) and with two groups of population (total population; sex and age groups). Their work discusses data between 1976 and 1992. There is a period of 9 years of data that is common to the work reported here.

Stillwell et al. (1995), Figure 2.2, demonstrate a fall of national migration propensities during the second half of the 1970s, a progressive increase from a low in 1981-82 to a peak in 1987-88, a rapid decline to 1990-91 and an increase after then. The examination of the age/sex specific groups (5-year age groups for males and females) suggests various trends. To demonstrate the time trend for the disaggregated groups they plot indexes where the migration rate of year 1980-81 is 100 and the migration rate before and after is a percentage of this base year rate. There are 32 sex/age groups, each experienced different temporal trends. Stillwell and the others (1995) identify a relative decline in the rates of migration of the 0-9 year olds between 1976 and 1992 and a dramatic increase in migration rates of those aged over 75 years after 1986. Teenage migration rates (10-19 year olds) are stable during the mid-70s to the mid-1980s, they fall during the second half of 1980s and increase after 1990. The migration rates of the older age groups experience a dramatic increase during mid-1980s, but there are differences in the scale of this trend. For example, those aged 35-39 display much greater variations in migration rates than those aged 30-34. There is a consistent decline in migration rates of all adult age groups after 1987-88 but most emphatic for those aged over 60 years. In the younger age groups there are similar trends between males and females, but for those aged more than 55 years, the peak in male migration rate observed in 1987-88 is not matched by that of females. After 1990, migration rates increased with those of 15-19 year olds experiencing the most dramatic increase (Stillwell et al., 1995).

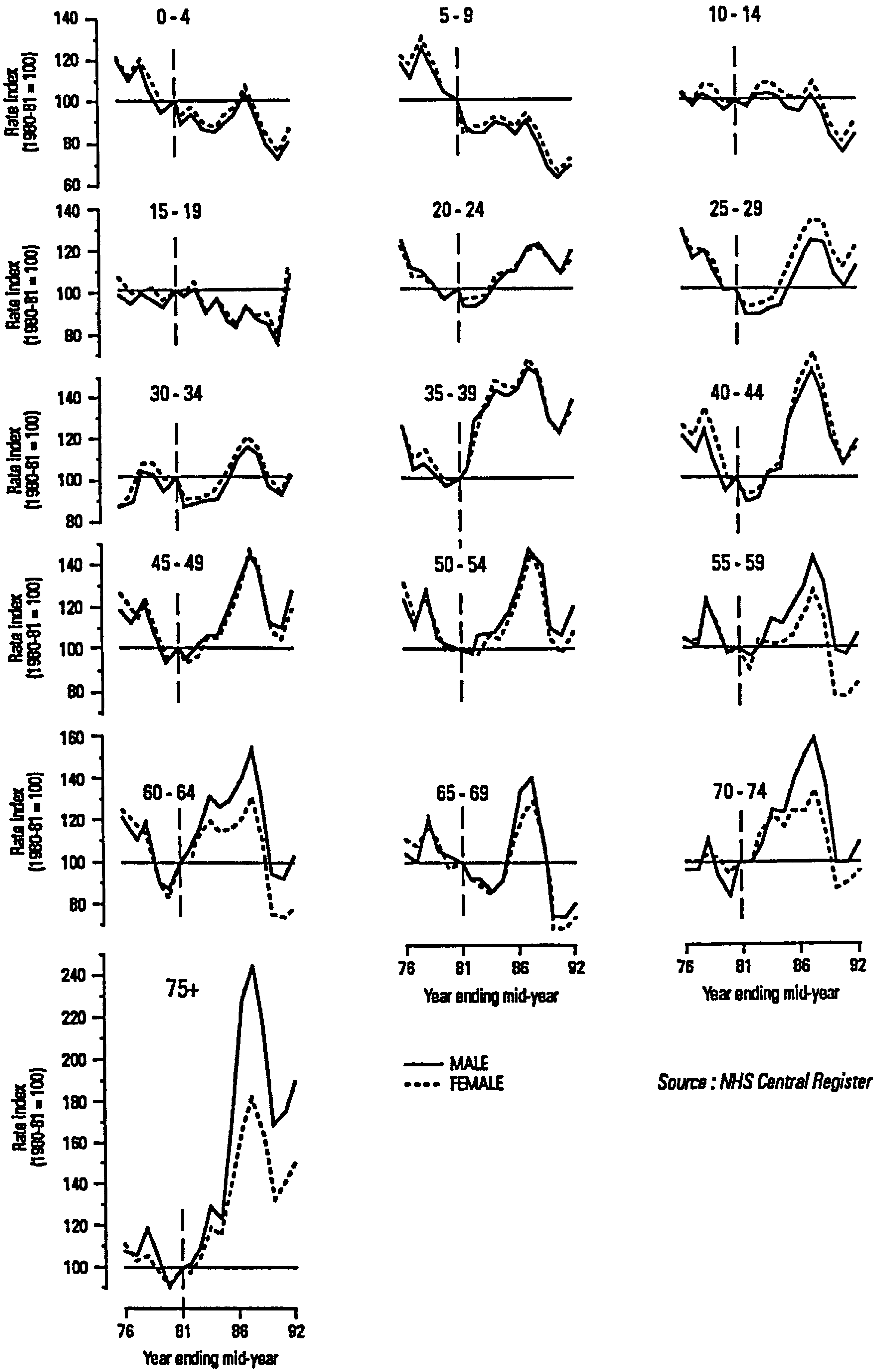


Figure 2.2. Age- and Gender- specific Migration Rates Between FHSAs Time Series Indices, 1975-92[a]

([a] Source: NHS Central Register); Stillwell et al., 1995, p. 348

The low migration propensities in 1981-82 (Stillwell et al., 1995) conclude a period of continuous decline since 1971 (Devis, 1984). During the period 1975-82, there is a decline in net migration loss for Greater London and a decline in both in-migration and out-migration for most FHSAs in England and Wales. Over the same period, net migration gain for those aged 16 – 24 in Greater London increased because of stable in-migration and a declining out-migration (Devis, 1984). The latter is a very interesting result that has not clearly reported in other migration studies. This trend of London being a net migration loser in most age groups except young people observed in 1975 still applies (1996/97).

Champion et al. (1998) discuss NHSCR aggregated migration data at the regional, county and FHSA level. From their review most important is the identification of *three main dimensions of net population redistribution produced by internal migration: north-south drift, urban-rural shift and local urban decentralisation.* They found that during the 1980s and the 1990s the southward drift of population has continued, metropolitan areas and city districts registered a steady net migration loss to more rural areas, and there is a continuing suburbanisation and local decentralisation taking place. These findings are confirmed in Champion (1996) and Stillwell et al. (1990). A rarely reported finding in migration trends during the 1980s is that FHSAs containing major universities are net teenager population (aged 15–19) gainers and excluding London and some FHSAs in the southeast, net young adult population (aged 20 -24) losers (Stillwell, 1994). The reason for these trends is student and graduate migration. These trends continued in the 1990s (see Chapter 5).

Internal migration trends using NHSCR data have also been reported in Green (1994), Stillwell and Boden (1989); Stillwell et al., (1992; 1996); and Stillwell (1985; 1986; 1994).

Generally, the trends of census data should be very similar to those of the NHSCR data. The differences are because these two measures record or do not record specific population groups (see Chapter 3). The census data are expected to be the most accurate data available. A detailed comparison between the census migration data and the NHSCR data is presented in Boden et al. (1988).

2.2 The Lowry Debate: Are socioeconomic variables significant?

Lowry's (Lowry, 1966) observation that the economic conditions of the origin are not significant components of out-migration motivated a debate in the literature that is still on-going (Vias, 2001). Most of the research provides evidence that there is some relationship between economic conditions and out-migration but the same conditions are *more* significant in explaining in-migration rates.

However, the debate has mainly concerned migration studies in the US and not in Europe and particularly in the UK. The aim here is to examine Lowry's Hypothesis using time scale data in the UK context. I focus on the following issues: the significance of out- and in-migration to net migration; the significance of explanatory variables acting as pull or push factors in the same system of interacting zones; and the stability of parameter estimates of specific variables in models with different structures (different set of variables). The latter issue rises from Miller's (Miller, 1973) argument that the presence of employment growth in the model along with family income does change the sign and the magnitude of the parameter estimate of family income. He suggests this may result in misinterpretations for some of the migration determinants' parameter estimates, such as the family income parameter estimate.

In the following sections an extensive review of the literature recounts research and empirical evidence for and against Lowry's Hypothesis. The focus is on the variables included in the model, how significant they are and the main trends. The relationship between in- and out-migration, and net migration is examined. This examination includes models in the literature where gross migration is a component of net migration.

Data issues in terms of variable selection are then described. The methodology section is concerned with technical details of models often used in migration studies. When empirical evidence from the UK is provided in a study I discuss the results. Some conclusions are drawn by comparing what has been suggested in the literature and attention is drawn to gaps in the literature.

2.2.1 The significance of push factors in modelling out-migration

In the literature there are two streams of evidence on whether out-migration is affected by economic conditions at the origin: those who believe that there is a significant relationship between economic conditions and out-migration rates and those who do not. The latter group of authors (such as Alonso, 1972, 1973; Lansing and Mueller, 1967; Lowry, 1966; Morrison, 1975; and Morrison and Relles, 1975) constitute a minority. Lowry is one of the first researchers to argue that labour market conditions at the origin zone are irrelevant to the determination of migratory outflow. Alonso also finds similar results. However, it is important to note the time and type of their analyses. Lowry studied inter-metropolitan migration in the US in the 1960's. With origin and destination zones having positive growth, good employment opportunities and healthy economies, Lowry's findings do make sense; his argument is that good economic conditions do not motivate migration but uneven out-migration rates are probably produced by other factors such as housing conditions, service provision and cultural attractions. However, a complete interaction system should include rural areas. In a complete interaction system, one would expect greater variance in migration determinants than in a system with similar types of zones. It is also necessary to study the trends over different times and in different systems, in order to examine whether Lowry's findings in the 1960s apply in the 1980s and 1990s; and to investigate if such findings apply to datasets with more variation in economic conditions.

An interesting paper by Miller (1973) examines not only whether out-migration is affected by economic conditions but how different variables in a model perform and how the presence of some variables can affect the significance and behaviour of others. He primarily examines if out-migration rate is reduced by high wages, high employment growth rates and warm winters and if it is increased by high unemployment rates. He supports the argument that an important out-migration determinant from an origin is the proportion of people in that origin who have migrated before (Goldstein, 1954; 1964; Land, 1969; Morrison, 1967; 1971; Myers et al., 1967; Rogers, 1969). To measure this determinant he uses two variables: the fraction of the population living outside their state of birth (US) and the in-migration rate (people who lived less than five years in an area). He also introduces to his model a measure of higher education, the median family income, employment and unemployment rates, mean January temperature, employment growth rate, and the natural logarithm of the total population. Then he examines several models of out-migration using different combinations of these variables. Miller's (1973) findings contradict the Lowry Hypothesis. He finds employment growth to be the primary economic determinant of out-migration rates, high values of which deter migrants from leaving an area. Family income, total population and winter temperature all have a negative sign, meaning that the higher their value the lower out-migration rate becomes whereas in-migration rates, college attendance rates and percentage of people living outside their state of birth have positive signs, denoting that when the values of these variables increase, out-migration rate increases, ceteris paribus. The former variables suggest that out-migration rates are lower in economically healthy areas and that a high proportion of mobile people increases out-migration rates. These observations are interesting and should be tested in the UK context. What Miller did not include in his model are the cost of living in an area (house prices); and quality-of-life variables such as crime rates, pollution and aesthetics.

Further evidence on the relationship between income and out-migration has been provided by Feder (1982). He reviews previous research and findings that suggest income may not always be significant and negative in the origin as traditional economic theory suggests. Greenwood (1975) suggests the income coefficient is usually smaller at the origin than at the destination of a migration trip. The origin income coefficient is sometimes not significantly different from zero and in some cases it is significantly positive (Greenwood, 1971; Greenwood and Ladman, 1978). These findings motivated Feder to conduct further analysis. He found that the relationship between average income at the origin and the rate of out-migration is not necessarily monotonic. *Rather, it is appropriate to expect that the relationship will be positive at the range of low average incomes and negative at the range of high average incomes... therefore [we should] utilize a polynomial specification (a quadratic or other invented U-shaped functions) to account for the impact of origin income on migration* (Feder, 1982).

Evidence from the UK is generally against Lowry's Hypothesis. Cordey-Hayes (Cordey-Hayes, 1975; Cordey-Hayes and Gleave, 1973) tested the relationship between in- and out-migration based on city-region data from England and Wales and found a strong direct correlation between in- and out-migration rates. This contradicts the *empirically derived hypothesis that out-migration is independent of the economic characteristics of the area (and is therefore unrelated to in-migration, which is dependent on areal characteristics)* (Cordey-Hayes, 1975, p. 806). A more recent study using inter-county migration data from the National Health Service Central Register (NHSCR) suggests a significant relationship between economic conditions and out-migration (Millington, 2000). That paper focusses on the role of housing and labour market variables on migration decision-making. Millington (2000) finds significant differences in the effects migration determinants have on different age groups. In his empirical work he uses several variables as both push and pull factors and calibrates separate models for five age groups. However, he does not examine sex disaggregated data. His results are very interesting and detailed and motivate a study to examine if trends observed in 1987-88 remain the same ten years after.

More work using disaggregated data for inter-borough migration by professional and managerial workers in 1981 in London provides evidence that unemployment (a representative economic variable) is not significantly correlated with out-migration whereas it is strongly correlated with in-migration (Congdon, 1989). However, it is not clear what unemployment really measures in Congdon's models: unemployment of the professional and managerial workers or general unemployment? It is important that variables used for a disaggregated population group reflect the labour market conditions of the specific group, and not the general economic conditions.

Another important issue is whether migration is related to housing factors and quality of life as one would expect in London. It is not clear that a migrant who moves from one London Borough to another changes employment as commuting to the existing working location is possible. Therefore, labour force conditions would appear to be much less significant factors in both origins and destinations for inter-borough migrations. If such variables are found to be significant in a model, they possibly act as surrogates for other variables such as service provision usually found to a higher standard in more prosperous areas of London. These issues concern Congdon (1989) who examines migration models for different age groups (15-29; 30-44; 45-retirement) and different status groups (economically active household heads, professional and managerial workers, non manual workers, skilled manual workers etc.) in order to investigate differences in the effects of housing and labour factors to in- and out-migration across these disaggregated population groups. His findings, especially those of economic conditions (unemployment and economic growth) are not consistent. He does not clearly support unemployment and economic growth being significant push factors. He argues they are for one group (household heads, economically active) and they are not for others (socio-economic groups, age groups) based on his empirical findings from models which are lacking important explanatory variables. However, he addresses several issues on migration modelling. Thus, it is worth re-examining his findings with the new data available here since the geography matches and comparisons can be made. In fact he encourages that himself in the conclusions of his 1988 paper.

2.2.2 The relationship between in- , out- and net migration

Several authors (Alonso, 1972, 1973; Lansing and Mueller, 1967; Lowry, 1966; Morrison, 1975; and Morrison and Relles, 1975) find that there is an apparent absence of any relationship between the economic conditions prevalent in an area and the propensity of its inhabitants to migrate. They also find out-migration to have little influence on variations in net migration, the major determinant which is the rate of in-migration. Kriesberg and Vining (1978) conclude that a reason these studies result in "no-push" findings might be because they look only at inter-metropolitan migration flows and exclude rural areas from their analyses.

Beale (1969) analyses the behaviour of gross (in- and out-) migration to determine if these rates vary depending on the net migration rate. He uses aggregated data for metro and non-metro areas in the US. He finds in-migration being highly correlated with net migration in growing regions (positive net migration) and out-migration highly correlated with net migration in declining regions (negative net migration) whereas out- and in-migration respectively do not have a significant impact on net migration in areas with little change in their population. Kriesberg and Vining (1978) confirm Beale's findings. They provide evidence using time series analysis for Japanese prefectures. From the time series Japanese data, they find that most of the change in net migration in peripheral prefectures (largely rural areas) is due to shifts in out-migration flows, whereas most of the change in net migration in central prefectures (largely urban areas) is due to shifts in in-migration flows.

New data from the US allow Vias (2001) to re-examine Beale's finding using more disaggregated data. He finds that in larger non-metro counties close to metropolitan areas the relationship Beale found still applies. In all other cases, where applicable, Beale's findings do not apply. For a large number of remote agricultural counties in the central part of the US there is little variability in gross migration rates, even though some counties were growing where others were declining. Consequently, there is no strong relationship between gross and net migration. Vias (2001) finds that there is much more diversity in the patterns of non-metro migration than previously noted.

A more flexible approach in this debate is provided by Plane et al. (1984). They interpret Lowry's results in a different way. They argue that in fact Lowry observed an apparent asymmetric effect of economic conditions on in- and out-migration and they disagree with authors (e.g. Wertheimer 1967) who interpret Lowry's findings as out-migration not being affected by economic conditions at all. Plane et al. (1984) provide evidence *that in the absence of any "push-pull" response to differential regional attractiveness, in-migration should be more variable across all the regions than out-migration.* They also show that *through the inclusion of regional attractiveness measures, the magnitude of the greater variation of in- than out-migration can be enhanced by the typically skewed distributions of regional desirability, size, economic opportunity and interregional distance* (Plane et al., 1984, p. 173). In commenting on Beale's (1969) arguments they agree that policy-making based on experimental results should be attempted with extreme caution.

Instead of dividing migration moves in two stages, the production of migrants at an origin and their distribution to potential destinations, Sommers and others tried to model net migration to an area using socio-economic properties of that area (Sommers and Suits, 1973, Sommers 1981, Meyer et al. 2001). In the 1970's Sommers and Suits (1973) examined the effects of economic and regional factors influencing net migration patterns of black and white families in the US for the decades 1950-60 and 1960-70 at the US state level. They found different results between the ethnic groups and over the two decades. Income, for example, was significant (positive in all cases) for white families in both periods but it was not significant for black families in the early period turning to a significant factor in the later period. Unemployment was negative but not significant in both periods for black families and negative and significant for white families in 1960-70 but non-significantly positive 10 years earlier. Welfare parameter estimates in terms of net migration were negative for white families and positive for black families in both periods with a strong significance for black families and non-significance for white families in the early period and vice versa in the later period (Sommers and Suits, 1973). Sommers and Suits explain the positive effect of welfare in black families' net migration due to the benefit these families got from moving to northern areas (US) where the welfare per family was higher than in the southern areas. Thus, the higher the welfare in an area, the higher the net migration of black families to this area was (this positive net migration of black families was due to high in-migration motivated by higher welfare). The increasing proportion of black families in an area may encourage white families to migrate, thus the negative effect of welfare to net migration of white families could be explained; welfare for the latter group acts as surrogate for a cultural effect. The regional effect was measured using dummy variables for 8 major regions in the US. Their effect was found not to be significantly different from zero in the early period, but a definite regional pattern emerges in the following decade.

The re-examination of the economic and regional effects in mid-1990s (Meyer et al., 2001) showed a stability of their effects on net migration over time. However, the level of employment opportunities appears to be the most important factor in the mid-1990s whereas income is the most important factor in the 1970s. Income in the mid-1990s is positive but not significant, welfare has a significant negative effect and the newly introduced health insurance appears to have a negative non-significant effect. Furthermore, regional dummy variables are important to the analysis as they capture significant migration flows between certain geographic regions. The study of Meyer et al. (2001) is a good example of how important it is to model net migration as an alternative to the traditional two-stage model. The results of net migration models could be very useful to policy makers for controlling the population of an area. For the methodology of the Meyer et al. (2001) paper to be applied in the UK, it is necessary to model net migration at the geographical scale of districts in order to ensure that an area is a town and its suburbs and not a combination of urban and rural areas. However, modelling net migration is not without problems and limitations. Perhaps the major disadvantage is the exclusion of information about the in-coming and out-going populations, and the reason for these. The latter are more important to policy makers in order to control population shifts in favour of the population cohort and the nation's goals.

Several studies in the US use various regional dummy variables to capture a potential regional effect. Here I am concerned with their importance in the UK context. During preliminary analysis, a dummy used to capture a potential capital effect (dummy for London) was found not to be significant in many of the time periods, and was thus rejected from the models. Local modelling is used to capture any local effect, thus the regional dummy variables are not appropriate for this study.

2.3 Quantitative Research in Migration

In this section a review of migration studies using quantitative methods is provided. A brief discussion of the geography, data and research focus is presented with the aim of identifying what researchers have achieved to date and what questions remain or need to be revisited. The first subsection is a brief review of global and local forms of migration modelling and the second subsection is a summary of previous findings on the effect of several migration determinants on migration decisions.

2.3.1 Modelling

2.3.1.1 Introduction

A review of previous work in migration modelling is presented here. Although various types of migration models will be mentioned below, the technical details of the major model to be used in this study will be presented in Chapter 4.

According to Stillwell and Congdon (1991), after theoretical assumptions and propositions have been formulated (for the application of a model), it is necessary to construct hypotheses, which can be tested for their significance. Migration models are the conceptual, mathematical or statistical expressions of the hypotheses in question, which frequently involved attempting to identify the factors that explain certain types of migration behaviour. Typically a model is made operational by selecting a specific measure of migration as the dependent or predicted variable; choosing a mathematical or statistical function with which to relate the migration variable to one or more independent variables; and adopting a suitable method of calibrating the model equation and a set of statistics with which to assess how close the predictions generated by the model are to the observed information.

2.3.1.2 The three main categories of migration models

Weeden (1973) in his study on interregional migration in Great Britain in the 1960s, provides also a good review of statistical models of migration found in the literature. He suggests that from a variety of migration models found in the literature, three main categories can be identified: *ad hoc* models, gravity models (both groups estimated by regression or analysis of covariance methods) and Markov chain models (Weeden, 1973). The category of gravity models can be renamed here, using the more general term *spatial interaction models*. Weeden also examines the differences between the models and tries to determine which is the best model.

Weeden uses the title *ad hoc* 'to identify a group of models which attempt to discover whether net migration is associated with various economic (or other explanatory) factors, ..., which are specifications of a linear regression model, ..., and their results are evaluated in terms of the conventional criteria of a good R^2 and significant coefficients with the correct signs' (Weeden, 1973, pp. 53-54). He refers to Oliver (1964), who used analysis of covariance, testing not only for regional differences in intercepts but also for differences in slopes for the model. A simple example of an *ad hoc* model, one of the three specifications of Oliver (1964) of net migration flows over unemployment, follows (Weeden, 1973, p. 52):

$$\frac{N_{it}}{P_{it}} = A_i + B_i X_{it} + V_{it} \tag{2.1}$$

where N = net or gross (in or out) migration flow,

P = total population,

A = intercept,

B = parameter of the unemployment percentage to be estimated,

X = unemployment percentage,

i = refers to the ith region,

t = refers to the tth period of time, and

V = an additive error term.

There is an extensive literature on spatial interaction models and their application in migration modelling. A good review of these models (Fotheringham et al., 2000, ch. 9), divides the history of spatial interaction models over the last 150 years into four phases in terms of their theoretical framework. Spatial interaction modelling has its roots in the family of gravity models, although there are several phases of this methodological approach. *In chronological order, these phases are: (a) spatial interactions as* *<u>social physics</u>* (1860 - 1970); *(b) spatial interaction as* *<u>statistical mechanics</u>* (1970-80); *(c) spatial interaction as* *<u>aspatial information processing</u>* (1980 - 90); *(d) and spatial interaction as* *<u>spatial information processing</u>* (1990 - onwards), (Fotheringham et al., 2000, p. 215). A more detailed description of the family of gravity models is given in Fotheringham and O'Kelly (1989) and Haynes and Fotheringham (1984) along with applications in economics, retail and marketing as well as migration. Another review of spatial interaction models is provided in Plane and Rogerson (1994).

Weeden (1973), suggests that *gravity models are at a different level of aggregation* (than *ad hoc* models) *and use a log-linear not a linear form. Several specifications, all of which could be called gravitational, are available in the literature. A simple version is*:

$$M_{ij} = A\left(\frac{P_i P_j}{D_{ij}}\right)^{\beta_1} \bullet \frac{X_i^{\beta_2}}{X_j^{\beta_3}} \bullet v_{ij} \tag{2.2}$$

where M_{ij} = migration from the *i*th to the *j*th region,

A = intercept* (* it is not defined in Weeden, 1973),

P_i = population of the *i*th region (P_i and P_j represent the 'masses' at *i* and *j*),

D_{ij} = distance between region *i* and *j*,

X_i = the value of an explanatory variable X for region i, taken to be relevant to its attractiveness or unattractiveness,

X_j = the value of X for region j,

$\beta_1, \beta_2, \beta_3$ = parameters to be estimated, and

v_{ij} = a multiplicative error term.

This model is easily generalized to include more explanatory variables, or regional dummies may be used to estimate 'push' and 'pull' factors peculiar to particular regions (Weeden, 1973, p. 52).

Markov chain models are based on the probability of someone migrating. For a better understanding of these models, assume a migration flow matrix M_3 (Table 2.1), with its diagonal elements equal to intra-regional migrants plus non-migrants. The simplest Markov chain model assumes that the transition proportions (migration probabilities), $q_{ij} = M_{ij}/P_i$ are stable or change in a predictable way over time (like input-output analysis). The full matrix of these probabilities, Q, can be used in conjunction with models of births, deaths for population projection. However, *the predictive power of the projection depends entirely on the validity of the initial assumption that q_{ij} is constant* (or predictable) (Weeden, 1973, pp. 52-53). Figure 2.3 presents the migration flow matrix. The three variations of this matrix depending on the values of the diagonal of the matrix are presented on Table 2.1. Plane and Rogerson (1994) also discuss (*fixed-transition-probability*) Markov chains models; they present several specifications such as the *Feeny's Model*, and the *Destination Population Weighted (DPW) Model.*

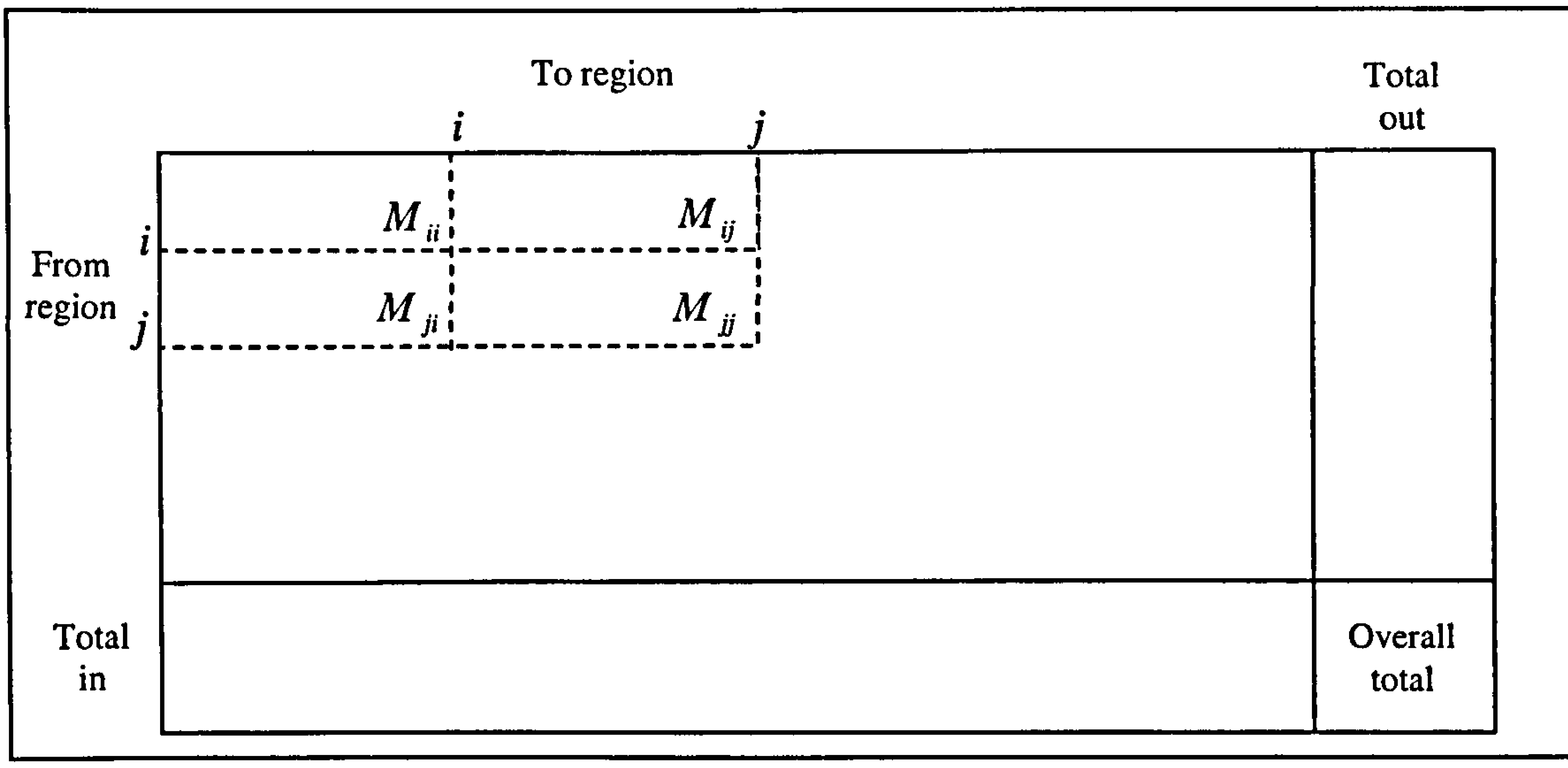


Figure 2.3. Migration flow matrix.
(Source: Chart 1, Weeden, 1973, p. 44)

Table 2.1. Migration flow matrices

Matrix	M_{ii}	Row and column totals	Overall Total
M_1	Zero	Emigrants and immigrants for each region	Total interregional migrants
M_2	Intra-regional migrants	'Movers'[a], in and out	Total 'movers'[a]
M_3	Non-migrants plus intra-regional migrants	Regional populations, at start and end of period	Total population[b]
[a] Persons changing address	[b] Assumed constant		

(Source: Table 1, Weeden, 1973, p. 45)

2.3.1.3 Spatial Interaction Models

A general form of the destination choice model (production constrained gravity model) provided by Fotheringham (1991), follows.

The general methodology for obtaining information on the sensitivity of migrants' destination choices to various destination attributes is to calibrate a spatial interaction model of the following form:

$$M_{ij} = \frac{O_i W_{1j}^{\alpha_1} W_{2j}^{\alpha_2} \cdots W_{kj}^{\alpha_k} d_{ij}^{\beta}}{\sum_j W_{1j}^{\alpha_1} W_{2j}^{\alpha_2} \cdots W_{kj}^{\alpha_k} d_{ij}^{\beta}} \qquad (2.3)$$

where M_{ij} represents the number of migrants between origin i and destination j; O_i represents the total number of migrants leaving origin i; W_{1j} represents attribute 1 of destination j which affects its overall attraction to migrants and there are k such attributes;

the a parameters reflect the sensitivity of a migrant's destination decision to changes in the respective attribute; d_{ij} *represents the spatial separation between i and j and is usually measured by distance; and the parameter* β *represents the sensitivity of a migrant's destination choice to distance and is commonly referred to as a distance-decay parameter* (Fotheringham, 1991, p. 58).

The equation (2.3) can be rewritten based on Stillwell's (1991) review of gravity models, to

$$M_{ij} = A_i O_i W_{1j}^{a_1} W_{2j}^{a_2} \cdots W_{kj}^{a_k} d_{ij}^{\beta} \tag{2.4}$$

where, $A_i = 1/\sum_j W_{1j}^{a_1} W_{2j}^{a_2} \cdots W_{kj}^{a_k} d_{ij}^{\beta}$ is *a balancing factor derived endogenously to ensure that the total migration from origin area i is equal to the sum of migrations arriving at all destinations from area i:* $O_i = \sum_j M_{ij}$ (Stillwell, 1991, p. 37).

Fotheringham (1991) strongly suggests that equation (2.3) should be calibrated separately for each origin in the system, because there are some factors such as housing costs and employment that are important for the migrant's destination choice decision but might depend on the comparison between origin and destination and thus their effect become origin-specific. This origin-specific form of the model is written as:

$$M_{ij} = \frac{O_i W_{1j}^{a_{1i}} W_{2j}^{a_{2i}} \cdots W_{kj}^{a_{ki}} d_{ij}^{\beta_i}}{\sum_j W_{1j}^{a_{1i}} W_{2j}^{a_{2i}} \cdots W_{kj}^{a_{ki}} d_{ij}^{\beta_i}} \tag{2.5}$$

In equation 2.5 all the parameters have a subscript *i* denoting the origin for which the model is calibrated. An extended example of the calibration of an origin-specific migration model using UK migration data and eleven destination attributes is provided by Fotheringham and O'Kelly (1989, pp. 98-106). Fotheringham suggests that if further data are available, the model of the equation (2.5) can be disaggregated by both origin and person-type. For example, equation (2.6) is the age/sex disaggregated model, used in this thesis. Note that all the parameters have a superscript *as* denoting the age/sex population group for which the model is calibrated.

$$M_{ij}^{as} = \frac{O_i^{as} W_{1j}^{a_{1i}^{as}} W_{2j}^{a_{2i}^{as}} \cdots W_{kj}^{a_{ki}^{as}} d_{ij}^{\beta_i^{as}}}{\sum_j W_{1j}^{a_{1i}^{as}} W_{2j}^{a_{2i}^{as}} \cdots W_{kj}^{a_{ki}^{as}} d_{ij}^{\beta_i^{as}}} \tag{2.6}$$

An evolution of spatial interaction models is that of hierarchical choice models. A specification of the latter is the Competing Destinations Model (CDM) introduced by Fotheringham (1983, 1991, 2000). It is based on the idea that in spatial decision-making, humans seem to use a hierarchical classification of the destinations. It is very common for the human brain to group some destinations, which later compete with each other. In fact, most migrants do not have all the necessary information to compare all possible destinations and thus they do not compare all the destination alternatives for their decision at once. The model is the evolution of the simple gravity model by introducing a new variable to measure destination competitiveness. For an origin-destination (A-B) flow one way of defining this variable is as the sum of the distance weighted population for all alternative destinations X_i of origin A. All competing destinations could be considered with destinations in close proximity being weighted more heavily. In which case, the formula for this new variable is:

$$A_j = \sum_{m \neq j} W_m / d_{jm}$$

where A_j is the potential accessibility of destination j to all other potential destinations m, W_m is a weight generally measured by population, and d_{jm} is the distance between j and m. The incorporation of this variable into equation (2.5) yields the competing destinations model:

$$M_{ij} = \frac{O_i W_{1j}^{\alpha_{1i}} W_{2j}^{\alpha_{2i}} \cdots W_{kj}^{\alpha_{ki}} d_{ij}^{\beta_i} A^{\gamma_i}}{\sum_j W_{1j}^{\alpha_{1i}} W_{2j}^{\alpha_{2i}} \cdots W_{kj}^{\alpha_{ki}} d_{ij}^{\beta_i} A^{\gamma_i}} \tag{2.7}$$

where the parameter γ_i reflects the relationship between migration and a destination's centrality (Fotheringham, 1991, p. 67).

Fotheringham (1991) suggests that one of the problems that has persisted in the mathematical modelling of migration, is that of the spatial variation of parameter estimates referred to as *the spatial structure effect* or *context dependency.* A simple solution of this model involves the addition of a single variable to a classic migration destination choice model to produce what has become known as *the competing destinations model. The competing destinations model is shown to provide a potentially useful breakthrough in understanding the so-called spatial structure effect in spatial interaction modelling and may be a key to unravelling the persistent geographic mystery of why estimated distance-decay parameters appear to exhibit unexpected spatial variation*" (Fotheringham, 1991).

Fik and Mulligan (1990) introduce their variation of the competing central places model. They evaluate Fotheringham's CDM and suggest a stronger hierarchical approach to enhance spatial competitions. They provide empirical evidence based on 1980 domestic airplane passenger traffic amongst cities in the U.S.. They conclude that the model they introduced is a generalization of the competing destinations model and that it overcomes the deficiencies of simple gravity-type approaches. A US state-to-state labour migration study enhances the CDM by introducing the intervening opportunities variable (Fik et al., 1992).

The new model by Fik et al. is called *competing and intervening destinations (CID) model.* They use a rather simple model without any socio-economic explanatory variables.

Pellegrini and Fotheringham (1999) support the superiority of CDM and provide empirical evidence for young adult inter-metropolitan movements in the US. They applied the method in age-disaggregated data (groups 25-34 and 35-44 years old) and model flows from 10 origins to 20 destinations (all metropolitan areas in mainly US cities). Migration rates are between 0 and 43 per thousand population of the origin and for those areas counts are a few thousands in most cases. The latter suggests that the aggregation level is still high. Thus, one would expect most of the flows to be explained by just population. This study is interesting because it provides evidence that the competition variable is significant in a model that includes a satisfactory number and variety of migration determinants. The importance of this study also is that it uses disaggregated data and a multinomial (MNL) discrete choice model.

Ferguson and Kanaroglou (1997) present their empirical evaluation of the aggregated spatial choice model, a specification of discrete choice models. They reject the ordinary multinomial choice model because it does not represent subaggregate heterogeneity in the aggregate data. They implement empirical work for interregional migration in Canada in 1990. However, they use a rather simple model in terms of measuring destination attractiveness, most of which is based on dummy variables.

Evers (1989) discusses the theory of the interdependencies between labour migration and commuting flows. He introduces two ways of dealing with those interdependencies: a macroeconomic model and a microeconomic approach. He concludes that economic rather than demographic components are dominant in modelling regional labour supply. To defend that he provides empirical evidence based on observed migration and commuting flows in the Netherlands during the 1980's. He also identifies that in the late 1980's volumes of commuting flows increased and those of migration flows decreased. He recognises that the data available were not sufficient for econometric analysis and safe conclusions towards the trend that recently there is a significant number of relatively longer distance commuters than in the past. That is a very interesting area to investigate in the UK. It is not only the identification of whether such a trend exists, it is more important to test if the volume of commuting flows is negatively related to migration flows. Unfortunately, there is no detailed data on commuting flows available here (there is only an explanatory variable: percentage of long distance commuters, see Chapter 3) to test the above hypothesis. If statistical evidence for the existence of the trend that workers are prepared to commute longer distances to their work rather than migrate can be provided, it will be necessary to revise the way models represent distance-decay and spatial structure.

2.3.1.4 Comparison between the models

A very interesting and hard to answer question is 'what is the best model?'. Clearly there is no straight answer. The evaluation of different models can be made in terms of the theoretical framework of each model or in terms of the empirical findings of its application to specific data. For the latter, the goodness-of-fit statistics, the residuals, the significance of the parameter estimates, are some means of evaluation. It is also important to select the appropriate model based on the data and level of aggregation available.

Weeden (1973) attempts to compare the three categories of models he reviewed. He groups *ad hoc* and gravity models to form *regression models*, which he compares with *fixed-proportions models* (Markov chain models). He argues that *in general a fixed-proportions model is appropriate for forecasting future migration (it requires data on migration and population alone), on the other hand a regression model is analytical in nature and intended to isolate the factors influencing migration and the size of their effects* (Weeden, 1973, p. 64). Kelley and Weiss (1969) attempt to evaluate the Markov process as description of migration by comparing its predictions with those of two economic models of migration (a linear and a log-linear form of model). They conclude that *as a predictor of interregional and interstate migration, the Markov process will consistently understate the ultimate adjustment in population, given the parameters that seemed to exist at present* (Kelley and Weiss, 1969, p. 280). Plane and Rogerson (1994, p. 211) based Plane's (1993) arguments on the theoretical comparison of Markov chain and gravity models, suggest that *fixed transition probabilities do not represent a correct behavioral representation of a migration system. As the gravity model suggests, the changing distribution of population should, itself, be a determinant of future patterns of destination choice.* Plane (1993, p. 221) argues that *the fixed-transition rate assumption inherent in much previous migration research makes little sense from a behavioural perspective. The results of this form of model, although seductive because of the long-term stability properties of Markovian structures, do not stem from sound behavioural assumptions inherent to most single-period conceptualisations of migration process. A role should be assigned in dynamic models of migration to the changing populations of both origin and destination regions.*

As this thesis is interested in exploring and explaining migration, not only the spatial patterns of migration flows, but also the reasons affecting out-migration and destination choice (push and pull factors) it is most appropriate to use a regression model, and more specifically a log-linear regression model.

2.3.1.5 Local forms of modelling

All the modelling techniques discussed above assume that spatial processes and behaviour are stationary across space. However, it is necessary to examine whether spatial processes and behaviour vary across space. If such non-stationarity exists, then the results of the models discussed above give limited and thus, not very useful, information about spatial processes and behaviour. In order to test for non-stationarity it is necessary to conduct local forms of modelling.

Several studies (Fotheringham, 1997; Fotheringham and Brunsdon, 1999; Fotheringham et al., 2002a) provide a review of attempts made for conducting spatially disaggregated modelling and local multivariate methods for spatial data analysis. Local versions of regression analysis are relevant here. Advances in these include the development of the expansion method (Casetti, 1972, 1982, 1997; Casetti and Jones, 1983; Jones and Casetti, 1992), multilevel modelling (Goldstein, 1987; Jones, 1991a; 1991b), locally weighted regression (Cleveland, 1979; Cleveland and Devlin, 1988), moving window regression (Hagerstrand, 1965; Martin, 1989; Fotheringham et al., 1996) and geographically weighted regression (Brunsdon et al., 1996; 1998a; 1998b; 1999b; and Fotheringham et al., 1996; 1997a; 1997b; 1998; 2000; 2002a).

The expansion method is an attempt to produce a more realistic model specification. Casetti (1972) suggests that it is possible for some of the independent variables of a model to be functions of other variables: *The expansion method is a procedure whereby a terminal model is generated from an initial one by making the parameters of the latter functions of some variables* (Casetti, 1972, p. 82). However, he provides examples of aspatial expansions of models that often appear in data analysis in social studies. In a more recent publication, Casetti (1997) provides a comprehensive summary of the expansion method. He also discusses the incorporation of space in the expansion method in his section on spatial modelling and spatial econometrics (Casetti, 1997). He suggests two approaches of spatial modelling: *spatial polynomial expansions designed to identify and display the statistically significant spatial variation of a model; ... and ... expansions based on many indices of spatial differentiation, or based on factors extracted from these indices* (Casetti, 1997, p. 27).

Fotheringham and Brunsdon (1999) recognise the importance of Casetti's expansion method: *the expansion method has been extremely important in highlighting the concept that relationships may vary over space and that the parameters of regression models applied to spatial data might exhibit spatial nonstationarity* (Fotheringham and Brunsdon, 1999, p. 346). However, they identify three main limitations in the expansion method: *one is that the* *technique is restricted to displaying trends in relationships over space with the complexity of the measured trends being dependent upon the complexity of the expansion equations...; a second is that the form of the expansion equations needs to be assumed a priori, although more flexible functional forms than the linear could be used; a third, and most problematic, is that the expansion equations must be assumed to be deterministic in order to remove problems of estimation in the terminal model* (Fotheringham and Brunsdon, 1999, p. 346). It is not clear whether the third limitation applies or not. Casetti (1997) argues that a deterministic model can be converted into an econometric, stochastic model by introducing random variable(s), namely, error terms. It is not necessary that the variance(s) of these variable(s) need to be zero. Unfortunately, Casetti (1972; 1997) does not provide empirical evidence to defend his arguments.

Casetti (1982) introduces a Drift Analysis of Regression Parameters (DARP), a heuristic technique that explores *local* variations in the parameter estimates. For example, he discusses a weighted regression of fertility (decline in crude birth rates) over energy consumption (logarithm of per capita energy consumption) with a weighted function based on an attribute space rather than a geographical space (Casetti, 1982). Casetti and Jones (1983) provide some empirical evidence for the spatial application of DARP and the Expansion Method. Several applications of the Expansion Method are discussed in Jones and Casetti (1992), among which a destination-choice migration modelling in Ecuador (Ellis and Odland, 1992) is relevant here. Further evidence on the significance of the expansion method comes from Brown and Jones (1985), Eldridge and Jones (1991), and Fotheringham and Pitts (1995) on migration modelling with focus on the spatial variation of the distance-decay parameter estimates. These empirical examples of the application of the expansion method can be linked with the local modelling of migration presented in Chapters 6 and 7. In this section more details on the work of Eldridge and Jones (1991) is presented.

Eldridge and Jones (1991) compare global and local distance-decay parameter estimates. They suggest the following gravity model

$$I_j = kM_j^{a_1} W_j^{a_2} U_j^{a_3} A_j^{a_4} d_j^{b} \tag{2.8}$$

where I_j is the migration flow from a given origin to a destination j, M_j is the population at the destination j, W_j is the mean family income at the destination j, U_j is the unemployment at j and A_j is the competing destinations variable, that controls for the effect of the spatial structure of destination at interaction (Fotheringham, 1983, 1984). In order to allow for distance-decay parameter to vary spatially they redefine b as a function of the spatial coordinates (of the destinations), yielding the expansion equation

$$b = b_0 + b_1 x + b_2 y + b_3 xy + b_4 x^2 + b_5 y^2 \tag{2.9}$$

The linearized version of the gravity model produces the following terminal model

$$\begin{aligned}\ln(I_j) &= \ln(k) + a_1 \ln(M_j) + a_2 \ln(W_j) + a_3 \ln(U_j) + a_4 \ln A_j \\ &+ b_0 \ln(d_j) + b_1 x \ln(d_j) + b_2 y \ln(d_j) + b_3 xy \ln(d_j) + b_4 x^2 \ln(d_j) + b_5 y^2 \ln(d_j)\end{aligned} \tag{2.10}$$

that can be calibrated using ordinary least squares regression. As soon as the parameters b_0-b_5 are estimated they can produce a surface for the parameter b showing its spatial variation. The results are very interesting and are discussed in the next section. Eldridge and Jones (1991) argue that this procedure not only allows for the identification of uneven effects of distance, but also the assessment of the statistical significance.

However, the procedure by Eldridge and Jones (1991) contains the limitations of Casetti's expansion method discussed above. It also does not allow for the parameters of all other variables in the model to have spatial variation (although this is possible by re-expressing each parameter by an expansion equation in terms of location) which may introduce misspecification bias. The Geographically Weighted Regression discussed below overcomes these limitations.

Multi-level modelling is an attempt to remove the ecological fallacy models of aggregate data contain by including the individual's characteristics (micro-level) and the regional characteristics (macro-level) simultaneously within the model (Jones, 1991a). Thus, multi-level modelling is more appropriate to explore and explain hierarchical structures of social or other phenomena. For example (Jones, 1991a) in order to model house prices it is necessary to include housing characteristics as well as characteristics of the districts these houses are located. Because of the nature of the model, OLS estimates need to be replaced by *shrinkage estimators*, which according to Jones (1991a) are found to use information in a highly efficient manner. More details on 2-level and multi-level modelling is provided by Goldstein (1987). Unfortunately, there are no individual data available in this work, thus, no comparisons can be made.

Cleveland (1979) presents a methodology for locally weighted regression as well as robust locally weighted regression. He bases his work on Macauley's (1931) smoothing of time series plots by fitting local polynomial, which he extends by replacing the temporal element with a spatial element. The result is a technique for locally weighted regression. Cleveland (1979) incorporates a technique of robust estimation (Beaton and Tukey, 1974; Andrews, 1974) that is an adaptation of iterated least squares. In his locally weighted regression, Cleveland (1979) suggests a Weighted Least Squares Regression fitted for an observation that includes the number of nearest neighbours and uses a spatial proximity based weighting function. His technique refers to univariate data analysis.

Cleveland and Devlin (1988) extend Cleveland's (1979) technique. This extension is also called Locally Weighted Regression (or *loess*). They argue that the latter is *a way of estimating a regression surface through a multivariate smoothing procedure, fitting a function of the independent variables locally and in a moving fashion analogous to how a moving average is computed for a time series* (Cleveland and Devlin, 1988, p. 596). They also recognise some restrictions in their methodology: one is the assumption of normality and constant variance of the errors, the other is that it can be used for studies in which the relevance of each independent variable in explaining the dependent variable has already been ascertained (Cleveland and Devlin, 1988, p. 608).

Geographically Weighted Regression (GWR) extends the traditional regression framework (linear regression) by allowing local rather than global parameters to be estimated (Fotheringham and Brunsdon, 1999). It is an evolution of Casetti's Expansion Method (Fotheringham et al., 1998). The global and local regression models are presented in Equations 2.11 and 2.12 respectively.

$$y_i = a_0 + \sum_k a_k x_{ik} + \varepsilon_i \tag{2.11}$$

$$y_i = a_{0i} + \sum_k a_{ki} x_{ik} + \varepsilon_i \tag{2.12}$$

where a_{ki} represents the value of a_k at point *i*.

In the calibration of the GWR model it is assumed that observed data near to point i have more of an influence in the estimation of the a_{ki}*s than do data located farther from i. ... In GWR an observation is weighted in accordance with its proximity to point i so that the weighting of an observation is no longer constant in the calibration but varies with i. Data from observations close to i are weighted more than data from observations farther away. ... The variation of the weights with i distinguishes GWR from traditional weighted least squares where the weighting matrix is constant. Typically, the weights are defined as continuous functions of distance...* (Fotheringham and Brunsdon, 1999, pp. 348-349).

More details on the GWR modelling technique along with recent evolutions of it are presented in Chapter 4. Applications of GWR include: examining relationships between population density and attributes of the physical landscape (Fotheringham et al., 1996); relating car ownership to social class and male unemployment (Brunsdon et al., 1996), the distribution of limiting long-term illness in the northeast of England (Fotheringham et al., 1998; 2000), a simulation experiment to study the role of the competing destinations spatial interaction model in capturing the effects of hierarchical destination choice (Fotheringham et al., 2001), the re-examination of the relationship between annual rainfall total and gauge elevation over Great-Britain (Brunsdon et al., 2001), local spatial interaction modelling based on the GWR approach (Nakaya, 2001), and hedonic model of house prices (Fotheringham et al., 2002a) to name a few. All these applications provide empirical evidence for the superiority of GWR over other global or local modelling techniques, in producing more informative results regarding parameter variation over space.

Two other attempts at local modelling are the adaptive filtering (Foster and Gorr, 1986) and the Random Coefficient Models (Swamy, 1971). *The Spatial Adaptive Filter (SAF) uses generalised damped negative feedback to estimate spatially-varying parameters for multivariate models* (Foster and Gorr, 1986, p. 878). From this work, it is important to mention two ideas: the idea of including observations that are located within a circle of fixed distance centred on a point in space for which calculation (feedback signal) is made; and the employment of Monte Carlo theory to assess the efficiency of SAF. Both these two ideas have been implemented in GWR.

In the literature there are many other attempts at local modelling. Local univariate methods for spatial data analysis (Fotheringham et al., 2002a) which include: local forms of point pattern analysis (Getis and Boots, 1978; Boots and Getis, 1988), local filters, and local measures of spatial dependency (Getis and Ord, 1992; Ord and Getis, 1995; 2001).

Among others, the most relevant here are the methods for identifying spatial clusters for a given variable. Methods examining spatial autocorrelation or spatial dependency can be applied to identify spatial clusters in in-, out- and net migration rates. These are more sophisticated than the tradition aspatial clustering, such as k-means clustering. Examples of global and local statistics relevant here are the Moran's I, the Getis' G and G^*, and the Geary's c. A review of these statistics and their application are discussed in Chapter 5.

It is beyond the scope of this work to provide empirical evidence for the superiority of one technique over another. This author adapts the conclusion for the superiority of GWR (Fotheringham and Brunsdon, 1999; Fotheringham et al., 2002a) to conduct local spatial data analysis. The three main reasons for this choice are: innovation, completeness, support. Geographically Weighted Regression is a new and emerging technique that addresses several problems of the previous attempts at local spatial data analysis (e.g. goodness of fit statistics). It is a well documented method including several journal articles and two books (Fotheringham et al., 2000; 2002a), and there is also a growing literature of GWR applications. There is software available (a user friendly interface has been developed) that makes the application of the GWR technique an easy task. Finally, there is the availability of support on queries about GWR as well as periodical updates to the method and the software.

2.3.2 Migration Determinants – Variable Selection

One of the most time-consuming exercises in migration modelling research is the selection of the appropriate variables that determine the production and attraction of migrants. In the literature there is no consistency in terms of variable inclusion for certain migration sex/age groups. It is obvious that researchers use variables available at a time, and do not in general spend a great deal of time and effort to collect additional data. Many migration studies include census data and only a few include economic data such as household income and housing cost.

In this section, migration determinants included in empirical work found in the literature along with their interpretation by the corresponding authors are presented. The reason for this is to position the work presented here in the framework of existing empirical studies of migration and to review existing findings in the association between migration and its determinants. The variables included below are solely those used in migration models of this work and a direct link between previous findings and this empirical work should be made. The associated text to each variable includes a discussion for the reasons for the inclusion of this variable, its effect on out-migration models and destination choice models, where appropriate. A section with similar structure in the following chapter presents the sources of these variables, data quality issues and variable construction issues where appropriate.

In the reminder of this section, the first two subsections refer to those variables included in the analysis of this thesis (Chapters 6 and 7), with special attention given to the distance-decay parameter. The last subsection reviews the use of dummy variables in previous work.

2.3.2.1 The effect of determinants on out-migration and destination choice

- **Air Quality**

Air quality is an important factor affecting the health of all residents especially children and elderly. The importance of air quality in residents' quality of life can be demonstrated by the increasing interest in Britain for issues associated with air quality (Mitchell and Dorling, 2003). Those who suffer the greatest exposure to poor air quality are those who have the least ability to move away from polluted areas, namely children and the poor (Mitchel and Dorling, 2003). This suggests that air quality is associated with migration decisions. Generally, people would prefer to move out of areas with poor air quality to areas of good air quality.

Out-migration rates in England and Wales are generally higher from areas of poorer air quality and lower from areas with good air quality (Fotheringham et al., 2002b; Fotheringham et al., 2003). This effect is expected to be stronger for children and pensioners and weaker for teenagers and young adults, as the former population groups are more sensitive to environmental conditions. Air quality is not a common variable in migration models and has only been reported as a push factor.

- **Climate Index**

The climate factor, while significant in countries the size of the US, is expected to be less significant in the UK where the climate changes are not very dramatic across the country. The general conditions suggest the western part of the country to be wetter than the eastern and the southern part to be warmer than the northern. However, the differences are minor compared to those found in larger countries such as the US.

Climatic factors are perhaps of more importance in the destination choice of migrants than in influencing out-migration decisions (Long, 1988; Pellegrini and Fotheringham, 1999). It is expected that people prefer to live in warmer and drier climates. Thus, a warm and dry climate is expected to have a negative effect on out-migration and a positive effect on destination choice.

From previous studies there is some evidence that climate influences out-migration. Mean July temperature was found to have a weak negative effect on out-migration rates both in the US (Miller, 1973) and Britain (Millington, 2000). Coldness was found to have a positive effect on out-migration rates in Canada (Liaw, 1990) and the US (Liaw et al., 2002). Evidence from England and Wales (Fotheringham et al., 2002b; Fotheringham et al., 2003) suggests that higher out-migration rates are associated with wetter and colder areas.

Pellegrini and Fotheringham (1999) provide empirical evidence for the significant positive effect of climate on destination choice in the US. Mean temperature also has a significantly positive effect on migrants' destination choice in both Canada and Britain (Ferguson and Kanaroglou, 1997; Millington, 2000). In the US, there is weak evidence for such an effect (Fotheringham and Pitts, 1995). A strong negative effect of coldness on destination choice has also been reported for young Canadian migrants (Liaw, 1990) and the elderly in the US (Liaw et al., 2002).

- **Percentage of long distance commuters**

Long-distance commuting is associated with the decision to leave an area because of people's need to reduce commuting time by moving to a residency closer to their work location. Data for commuting flows suggest an increase in recent years of distances and times people are prepared to commute. However, long-distance commuters are more likely to migrate than are shorter distance commuters. Commuting has probably a stronger effect on out-migration from areas with large populations that are far apart from other areas with large populations. Such areas include FHSAs in the Northeast (Newcastle), the Southwest (Avon) and South Wales (South Glamorgan). A weaker effect should be expected for FHSAs in London, West Midlands, Greater Manchester and West and South Yorkshire, because these are areas close and well connected to other areas of large populations.

This is a variable used only in out-migration models. Generally, it has a positive effect on out-migration rates (Fotheringham et al., 2002b; Fotheringham et al., 2003).

- **Contiguity**

This destination choice variable is used in order to remove the portion of migration that is a result of very short distance moves. Very short distance residential moves within an FHSA are not recorded as migration. However, there are cases where such moves involve a crossing of an FHSA boundary. These moves increase in- and out-migration for the corresponding FHSAs. Contiguity is included to capture such moves and to allow the distance decay variable to have a more real effect.

The nearest-neighbour dummy (Pellegrini and Fotheringham, 1999) is another name for contiguity which is expected to have a positive effect on destination choice. However, because of its binary nature (it gets the values 0 or 1) it can cause problems in local statistical techniques, thus, it is omitted from them. In global models it is reported to have a significant positive effect on destination choice explaining the high volumes/rates of short distance migration (Weeden, 1973; Fotheringham and O'Kelly, 1989; Boyle and Flowerdew, 1993; 1997; Liaw and Kawabe, 1994; Pellegrini and Fotheringham, 1999).

- **Crime Index**

Crime is an important migration determinant for families with children and elderly people. High crime rates are expected to have a positive effect on out-migration (people would leave areas of high crime rates) and a negative effect on destination choice (people would be attracted to low crime areas). The opposite effects are expected for low crime rates. Negative values of crime index denote a relatively lower crime rate than the England and

Wales average and positive values a higher crime rate. Thus, a positive relationship should be expected between crime index and out-migration and a negative relationship between crime index and destination choice. However, it is possible that crime rates are correlated with urbanity and thus may have a positive effect on young people's destination choice.

Crime is a serious problem mainly observed in deprived urban areas. It is connected to several variables such as level of education, deprivation, age, gender, religion, culture, police cover, ethnic minorities and others. Crime as a problem is a very sensitive issue and requires careful interpretation. Here, the interest is concentrated on the possibility that crime can be correlated with other variables in the model (non-white); or it may explain some variation of variables that are not in the model (deprivation index). In order to understand and interpret crime as a factor affecting migration decisions, it is necessary to understand its roots and its effects on people's everyday life.

Empirical evidence suggests a weak positive effect on out-migration and a significant negative effect on destination choice for mature and older adults (Millington, 2000; Fotheringham et al., 2002b). A strong positive effect of violent crime rate on out-migration of elderly in the US provides further evidence for the above (Liaw et al., 2002). A weak positive effect on destination choice of young adults has also been reported (Millington, 2000). However, there is strong evidence for a negative effect of crime rate on out-migration rate (Fotheringham et al., 2002b; Fotheringham et al., 2003). Areas with high crime generate fewer migrants probably because people living in these areas cannot afford moving to locations that are more desirable.

- **Council tax**

This is an economic variable partly accounting for cost of living in an area. It should therefore have a stronger effect on low-income people such as pensioners. It is necessary to note that dwellings occupied by people less than 18 or students are exempted from council tax payment. Council tax is only examined here for its effect on migrants' destination choices. It has been reported (Fotheringham et al., 2002b) to have a negative effect on migrants' destination choices for those aged 45 and over but a positive effect for those aged 16 – 19. A possible reason for this discrepancy is that a high council tax makes living costs higher thus acting as a deterrent but many teenagers are students and thus do not pay council tax. The latter are attracted to urban areas where council taxes are often higher.

- **Destination Accessibility**

Destination accessibility or centrality or competition is a variable that when added to a destination choice model forms the competing destinations choice model. *If the migration destination choice is consistent with a hierarchical information-processing strategy, the parameter associated with this variable will be negative* (Pellegrini and Fotheringham, 1999, p. 1099). Empirical work supports the hypothesis that destination choice results from hierarchical information processing (Fotheringham and Williams, 1983; Eldridge and Jones, 1991; Fik et al., 1992; Fotheringham and Curtis, 1992; Atkins and Fotheringham, 1999; Pellegrini and Fotheringham, 1999). Destination accessibility parameter estimates are typically significantly negative and vary across origins (Fotheringham and Pitts, 1995).

There is also empirical evidence (Nakaya, 2001) that in some situations there is significant spatial variation in local parameter estimates of destination accessibility, which can be positive or negative across destinations.

- **Employment Growth**

Employment growth at the FHSA level is possibly an indicator of a growing or stagnating local economy. This is because increase in employment rates is linked with investments for new, expanding or modernised businesses. The latter will also require a new, highly qualified labour force. Thus, high employment growth is expected to play a significant role in migration decisions.

Miller (1973) suggests that employment growth is the most important economic determinant of out-migration rates; high employment rates produce fewer out-migrants. This is because there is little incentive to leave an economically growing area. This effect is confirmed in other studies (Liaw, 1990; Liaw and Kawabe, 1994; Millington, 2000). However, there is some weak evidence of a positive effect on out-migration (Congdon, 1989) suggesting that economically growing areas may increase the turnover of population. A strong positive effect on out-migration rates has also been reported (Fotheringham et al., 2002b; Fotheringham et al., 2003). This is because of a high turnover effect on population caused by high employment growth.

Empirical work (Congdon, 1989; Fotheringham and O'Kelly, 1989; Liaw and Kawabe, 1994; Pellegrini and Fotheringham, 1999; Millington, 2000; Liaw et al., 2002) supports the hypothesis that destinations with high employment growth are attractive to migrants.

However, Weeden (1973) finds that employment growth loses its explanatory power when an unemployment rate variable is included in the model, as these two are highly correlated variables. This is also apparent in other models (Liaw, 1990). Recent work (Fotheringham et al., 2002b; Fotheringham et al., 2003) suggests lack of empirical evidence for a consistent effect of employment growth on migrants' destination choice.

- **Employment Rate**

Employment rate is a major economic variable. Together with unemployment rate it indicates the health of the local economy. It is expected to have a negative effect on out-migration (people prefer to stay in areas with more employment opportunities) and a positive effect on destination choice (people are attracted to areas with more employment opportunities). It is also possible that high employment rates are connected with more mobile workers and higher turnout suggesting the opposite effects of those described above.

There is evidence for a strong positive relationship between employment rates and out-migration rates for mature and older adults (aged 30 – 59), but a lack of a relationship for younger age groups (Fotheringham et al., 2002b; Fotheringham et al., 2003). A strong positive relationship possibly suggests the ability of residents with a good job to afford moving to areas that are more attractive. This should be the case for less attractive areas in England and Wales such as those with poor air quality, high crime rates, low cultural opportunities and negative reputation. However, the relationship could be negative in the cases of more attractive areas in England and Wales. Unfortunately, global regression models fail to capture potential spatial variations of the effect of the variable. However, local models applied here address this issue.

Mueller (1982) found employment rate to have a non-significant negative effect on labour migrants' destination choice. Further empirical evidence suggests a lack of a serious effect of employment rates on destination choice (Fotheringham et al., 2002b; Fotheringham et al., 2003).

- **Household Income**

Average household income is a measure of economic prosperity (or deprivation) of an area and it is thus expected to have a negative effect on out-migration and positive effect on destination choice. It is also a proxy for wage levels.

Empirical findings of the effect of income on out-migration vary. Lowry (1966) suggests a lack of effect. Miller (1973) finds that there is a negative effect occasionally significant, but with employment growth in the model the income parameter estimates were found to be non-significantly positive. Weeden (1973, p.94) suggests *there is evidence (from datasets) that the level of out-migration from high-income regions is higher than the level of* *in-migration from low-income regions.* More evidence for a positive (Liaw and Kawabe, 1994; Fotheringham and Pitts, 1995) and a strong positive effect (Fotheringham et al., 2002b; Fotheringham et al., 2003) comes from recent studies. The interpretation for such an effect is that higher income residents have the resources to afford to migrate. A significant negative effect of wage levels on out-migration has also been reported (Millington, 2000).

Some empirical findings show a weak explanatory power of the income variable in destination choice (Pellegrini and Fotheringham, 1999). Occasionally, significantly positive parameter estimates have been reported (Lowry, 1966; Liaw, 1990; Eldridge and Jones, 1991; Pellegrini and Fotheringham, 1999; Cannari et al., 2000; Liaw et al., 2002; Fotheringham et al., 2002b). Liaw and Kawabe (1994) found income at the destination to be highly correlated to employment growth and population size, thus it has insignificant positive parameter estimates in a full model. The latter, however, should not be interpreted as lack of exploratory power of income, which is found to be significant in a model from which employment growth is omitted. Millington (2000) finds an unexpected strong and significant negative effect of real wages on destination choice. A negative effect of household income on destination-choice for those aged 45 years old and over has also been reported (Fotheringham et al., 2002b).

- **House Prices**

House prices play a significant role in determining migration at both origins and destinations. Their effect is expected to be stronger for people with children and pensioners for different reasons. It is expected to have a positive effect on the production of migrants and a negative effect on the attraction of migrants. However, in gender, age and occupationally disaggregated data these effects are expected to be variable.

Empirical evidence suggests the migration-type variations discussed above and confirms the expected positive effect of house prices on out-migration (Congdon, 1989) as well as a negative effect on destination choice (Congdon, 1989; Cannari et al., 2000). The reverse effects are reported elsewhere: a significant negative relationship between house prices and out-migration rates along with possible explanations (Fotheringham et al., 2002b; Fotheringham et al., 2003), and a generally positive relationship between house prices and destination choice (Fotheringham and O'Kelly, 1989; Fotheringham and Pitts, 1995; Atkins and Fotheringham, 1999; Fotheringham et al., 2002b).

- **Listed Buildings**

This variable accounts for the attractiveness of a destination connected to its history, culture and beauty. It is found to have a positive effect on migrants' destination choices which increases with the age of the migrants (Fotheringham et al., 2002b).

- **Percentage of Students at Parental Domicile**

The aim of this variable is to capture students leaving home for university at the origin and to capture students returning to their parents' house after graduating at the destination. Empirical evidence suggest a positive effect on out-migration rates for males aged 16 -19, a negative effect on destination choice for those aged 20 – 24 and a positive effect on destination choice for those aged 25 – 29 (Fotheringham et al., 2002b; Fotheringham et al., 2003).

- **New housing on former urban land**

This variable is the proportion of new housing on recycled urban land. It is associated with regeneration projects that improve the housing quality and the appeal of an urban area. Such projects are usually found in formerly industrial cities and perhaps old towns. New housing on former urban land is an indicator of new housing availability usually in central city locations. Typically, areas with high quality housing availability generate fewer migrants. Such areas are also more attractive migrant destinations. The reason for the former is that there is less pressure in terms of housing unavailability to leave an area.

Empirical work suggests a generally non-linear negative relationship between new housing on former urban land and out-migration rates in England and Wales (Fotheringham et al., 2002b; Fotheringham et al., 2003). Surprisingly, a negative relationship is evident between this variable and destination choice, with a possible explanation being this variable acting as surrogate for high deprivation levels (Fotheringham et al., 2002b).

- **Percentage of net re-lets in social sector**

This variable is an indicator of social sector housing turnover. High proportions of net re-lets indicate a more unsettled population which is prone to have an increased number of migrants. Fotheringham et al. (2003) provide empirical evidence for an association of higher out-migration rates with higher levels of net re-lets in the social sector. A marginal effect on destination choice has been reported (Fotheringham et al., 2002b).

- **Percentage non-white**

This variable is used here as a push factor. Usually, non-white residents tend to be more unsettled in England and Wales. This is mainly because many non-white people are international migrants or new generations of such migrants and are less likely to have social ties with the places they live. Thus, high proportions of non-white populations should be associated with high out-migration rates.

Empirical evidence suggests a strong positive non-linear relationship between out-migration rates and percentage of non-white population with another explanation being white population leaving areas of mixed race (Fotheringham et al., 2002b; Fotheringham et al., 2003).

- **Occupational migration index**

The occupational Migration Index is a measurement of occupational structure of the origin and it has higher values in areas where employment is dominated by professional and managerial workers. Thus, it is expected to have a strong positive effect on out-migration rates.

Empirical findings show mixed results for sex/age disaggregated migrant groups in England and Wales suggesting *there is little conclusive evidence of any consistence relationship between out-migration rates and occupational structure* (Fotheringham et al., 2003, p. 32).

- **Total population**

In most of the empirical findings (Lowry, 1966; Boyle and Flowerdew, 1993; 1997) total population of the origin has a significant positive effect on out-migration. Here, out-migration rates per thousand people are modelled, thus, origin population is essentially forced to have a parameter estimate equal to 1.

Population at the destination is used to measure the effect of the size of cities on attracting migrants. Pellegrini and Fotheringham (1999) suggest that a destination's population, to some extent, measures employment opportunities, amenities and the probability of having friends and relatives there. Thus, it is expected to have a positive effect in destination choice. Indeed, in their analysis they find a significant positive effect of destination population on destination choice of migrants aged 25-29 and 35-44 in the US. Further evidence for a significant positive effect comes from a plethora of studies (Lowry, 1966; Fik et al.; 1992; Boyle and Flowerdew, 1993; 1997; Fotheringham and Pitts, 1995; Atkins and Fotheringham, 1999; Fotheringham et al., 2002b). The parameter estimates for most origins and model configurations generally vary from 0.5 to 0.9 but sometimes exceeds 1. Yano et al. (2003, p. 5) suggest that values of the population parameter estimates exceeding 1.0 *indicate strong urbanisation trends* (the largest cities' population grows more rapidly through net migration than that of smaller cities) whereas those less than 1.0 *indicate a counter-urbanisation trend* (smaller cities have higher percentages of population gains through net migration).

A regional measure of total population has also been included in the out-migration models. This is to account for the effect of nearby populated areas in motivating people to leave an origin.

- **Percentage of New Build Completions in Private Sector**

Percentage of New Build Completions in Private Sector is the proportion of private owned dwellings that are newly built. This measure indicates the existence of modern housing in possibly newly designated urban areas. Thus, it should be associated with migration decisions of higher income residents, as new private housing would possibly be of high standards and thus expensive to rent or buy. It should also be associated with the appeal of the area and its economic growth potential as the private sector would generally invest in areas that allow high profits in the longer term.

However, this variable has rarely been used in migration models. Congdon (1989) provides empirical evidence for a significant negative effect of new private housing on out-migration and a significant positive effect on destination choice. These findings suggest that housing availability plays an important role in migration, especially for those migrants who are looking to improve their housing quality. However, these relationships are not confirmed with empirical work elsewhere (Fotheringham et al., 2002b; Fotheringham et al., 2003) which finds the opposite effects and a rather weak association between new private housing and migration decisions. Fotheringham et al. (2003) suggest that a positive association between out-migration and new private housing may be because the latter may attract relatively mobile migrants that have increased possibility to move on to another area.

- **Percentage of New Build Completions in Social Sector**

Similar to the previous variable, Percentage of New Build Completions in Social Sector is the proportion of public sector dwellings that are newly built. However, social housing is usually associated with lower income population. Also, investments of local authorities in the housing sector are more likely to take place in areas in which there is a lack of interest from the private sector to invest in residential properties. Thus, although new housing would possibly lower out-migration and encourage in-migration, new social housing may act as a surrogate for areas with low appeal. In general, an association of this variable to migration decisions is likely, but the interpretation of such a relationship is not straightforward.

Fotheringham et al. (2003) find a positive relationship between out-migration rates and the percentage of newly built housing in the social sector for most of the age/sex migrant groups of their analysis. They suggest that areas where there is a lot of public sector housing being built may be undergoing social upheaval.

There is empirical evidence (Millington, 2000) for a significant positive effect of new housing availability at the destination in attracting migrants. This is also confirmed in Fotheringham et al. (2002b) suggesting that high rates of new public housing are associated with high numbers of in-migrants.

- **Percentage of vacant dwellings in all sectors**

This variable is the proportion of all dwellings that are vacant and indicates the overall housing availability in an area. The availability of housing is an important migration determinant at both origins and destinations. This is because an important reason for internal migration is a change of housing condition for families and individuals. This can be driven by improving quality of life reasons or economic reasons. Therefore housing availability in an area allows more people to satisfy their housing needs, thus, reducing the probability of leaving the area for better housing opportunities (negative relationship with out-migration). A good supply of housing is also expected to make a destination more attractive.

Empirical evidence suggests a negative non-linear relationship between out-migration rates and the proportion of vacant dwellings at an area (Fotheringham et al., 2002b; Fotheringham et al., 2003). No conclusions for the association of this variable with migrants' destination choices can be drawn (Fotheringham et al., 2002b). This is because the parameter estimates of this variable are occasionally positive or negative across different origins in origin specific destination-choice models of each one migrant group. However, Mueller (1982) provides evidence for a significant positive effect of the percentage rental housing vacant on labour migrants' destination choices.

- **Index of the private stock in poor condition and Index of the Local Authority stock in poor condition**

These two variables measure the proportion of housing stock in poor condition in the private sector and local authority sector respectively. Thus, these are variables measuring an origin's deprivation and a destination's attractiveness connected with the appeal of an area, as poor housing is associated with the latter. Empirical work suggests weak evidence for an effect of these variables on people's migration decisions (Fotheringham et al., 2002b).

- **Percentage of students at term time address**

This variable is meant to measure the proportion of students living at a term-time address and thus likely to leave an area after graduation (origin variable) and the attraction of university places at a potential destination (destination variable). Empirical evidence suggests a significant positive effect of this variable on both out-migration rates and destination choice (Fotheringham et al., 2002b; Fotheringham et al., 2003).

- **Age and sex specific unemployment rate**

Unemployment rate is a main measure of employment opportunities. It is expected to have a positive effect on out-migration and a negative effect on destination choice. This is because high unemployment is associated with low or no job availability and thus encourages higher proportions of people to move out in seek for employment opportunities. A destination with low employment opportunities is possibly unattractive.

There is an on-going debate on whether a significant relationship between unemployment rates and out-migration rates exists, with many researchers providing empirical evidence for a lack of such relationship (Lowry, 1966; Miller, 1973; Millington, 2000). However, other researchers suggest there is some evidence for a weak positive effect of unemployment rate on out-migration rates but not a clear-cut one (Flowerdew and Lovett, 1988) and others suggest there is some evidence for a weak negative effect (Liaw, 1990). Fotheringham et al. (2003) find this variable to be insignificant for most population groups except females aged 30 – 44 and those aged 45 – 59 (positive effect).

Several studies (Lowry, 1966; Weeden, 1973; Flowerdew and Lovett, 1988; Liaw, 1990; Cannari et al., 2000; Fotheringham et al., 2002b) find unemployment to be significantly negative in destination choice, whereas others find that unemployment rate has a disappointing performance for understanding destination choice with non-significant negative parameter estimates (Eldridge and Jones, 1991; Pellegrini and Fotheringham, 1999). Using origin-specific destination choice models in the US, Fotheringham and Pitts (1995) find unemployment to have a significant impact on destination choice in only 4 out of 48 contiguous US states. A surprising positive effect of unemployment on destination choice has also been reported for UK migration (Atkins and Fotheringham, 1999).

- **All vacant and derelict dwellings**

This variable is a measure of the proportion of total dwellings that are vacant and derelict. Generally, high proportions of vacant and derelict dwellings are associated with low appeal, high crime or abandoned areas. Thus, it is expected to have a negative effect on migrants' destination choice. Empirical work in England and Wales confirms this effect (Fotheringham et al., 2002b).

2.3.2.2 The role of distance in gravity models

Fotheringham and Curtis (1999) suggest that the role of distance between an origin and a potential destination has to do with the information about the destination. The longer the distance a potential destination is apart from an origin, the less likely are migrants living in that origin to have information about the destination, and are thus less likely to select that destination. The above hypothesis is based on the idea that the more the uncertainty (lack of information) about a destination, the less attractive a destination is. As a result, the distance to a destination is often the most important factor in explaining migration flows.

Fotheringham (1981) reviews the concept of distance-decay in spatial interaction modelling in order to examine whether a relationship between spatial structure and estimated distance parameters exists. He concludes that a statistical relationship exists but he provides directions for future research for a further investigation in the theoretical framework of this issue. Later studies (Fotheringham, 1991; Pellegrini and Fotheringham, 2002; Fotheringham et al., 2001) show that destination accessibility rather than distance is associated with spatial structure. *Destination accessibility removes the spatial structure effect from the model which results in misspecification bias, particular in the estimated distance-decay parameters* (Yano et al., 2003, p. 4).

An interesting work in modelling migration in the UK is that of Boyle and Flowerdew (1997) who introduce a new method of measuring distance between origin and destination; this is the migration-weighted distance. In their study they provide empirical evidence for the model improvement when a migration-weighted distance replaces a population-weighted centroid distance in a simple gravity model. They argue that migration-weighted distance measure gives a more realistic representation of migration movements rather than population distribution. They conclude that the migration-weighted distance provides better migration flow estimates for small distances and for distances 150 - 299 miles for the UK inter-county migration flows.

As well as being a proxy for information about the destination, distance may also measure the monetary and psychic costs of moving (Pellegrini and Fotheringham, 1999). Distance is expected to yield negative parameter estimates. However, the magnitude of these estimates varies across studies. *Highly negative values indicate that interaction is locally confined, whereas less negative values are associated with broader pattern of spatial interaction* (Eldridge and Jones, 1991, p. 501). In Pellegrini and Fotheringham's (1999) empirical work in the US intermetropolitan migration in 1990, distance has a significant negative effect and the parameter estimate varies from -0.25 to -1.7 depending on the origin and migrant group under study. Consistently in the destination choice models of migration, distance parameter estimates are significantly negative.

Eldridge and Jones (1991) examine not only the variation of the distance parameter estimates across models, but also within a model across space. They suggest an application of the expansion method to allow for local variation in the distance-decay parameter. Their findings led them to reject the assumption that the effect of distance is stationary over space. More evidence on this is provided in this thesis.

Physical distance parameter estimates were found to be insignificant in Ferguson and Kanaroglou (1997). Instead social distance, defined by social barriers such as spoken language, was found to play a role in destination attractiveness in Canada.

There is some empirical evidence from migration studies in the US (Fotheringham and Pitts, 1995) and Japan (Nakaya, 2001) for a significant spatial variation of the distance decay parameter of an origin-specific spatial interaction model across destinations. This suggests that migrants leaving an origin cognize its distance to potential destinations differently across destinations.

2.3.2.3 The role of dummy variables

Ferguson and Kanaroglou (1997) demonstrate the importance of dummy variables to capture non-quantifiable effects and to act as masks for groups of migrants that are attracted by certain attributes that do not attract other groups. Empirical findings (through regional dummy variables) show evidence of regional distinctions in parameter estimates and region-specific trends (Fik et al., 1992). Applying Geographical Weighted Regression removes the need for dummy variables that capture regional differences.

Dummy variables have also been used to capture non-spatial effects. Liaw and Kawabe (1994) used several dummy variables to model personal factors of migrants as well as ecological variables such as linguistic similarities. They found some of the dummy variables to have very significant effects on migration, especially the personal factors on out-migration.

2.3.3 Spatial Cognition

Another interesting literature includes studies of spatial cognition and spatial learning, which have been analysed by both geographers and psychologists (Lloyd, 1997; 2000; Golledge and Timmermans, 1990; Golledge, 1993; Golledge and Stimson, 1997). Fotheringham and Curtis (1999) provide a good starting point from the view of geographers with an application to migration studies. In fact, in order to state an explanation of this spatial choice, the researcher must involve the findings of spatial cognition in psychology, ways of getting information about destination and various other parameters. It is not the aim of this research to look at that. However it would be interesting to look at modelling spatial cognition and perception, not only as the common human reaction, but how that is correlated with an individual's socio-economic profile.

Blaut (1999) discusses the meaning of the word space in geography and psychology: *the first meaning is the idea of space as scale, as the size of geographical or environmental places and processes; the second is the idea of pure space, spatial structure, ... as geometry.* He believes that geographers only look at space and social processes on it as geometry, which in fact is not the case in quantitative research where we see human interaction in a spatiotemporal macro scale context as the psychologists do. He recognizes that spatial cognition is a result of experience and learning of humans mostly during childhood, and the framework for studying this way of learning is the theory of empiricism; the scientific way. He also rejects the Piagetian tradition and constructivist development theories that assume that the source and nature (in a Neo-Kantian manner) of spatial knowledge is something internal to the mind and prior the experience (Piaget and Inhelder 1956, 1969; Piaget 1971).

2.4 Qualitative vs. Quantitative Methods in migration

I believe that quantitative analysis on migration data generally yields more objective results and conclusions than qualitative analysis usually based on the views of few selected migrants. The aim here is to discuss the value of quantitative research in the decision making of migrants within a developed country and not to a priori reject alternative methods (qualitative). The remainder of this section includes my personal views as well as some ideas drawn from the existing literature.

The main reasons for the superiority of quantitative research in exploring and explaining migration concern both the nature of the data used and the methods employed. Secondary data (census, large surveys) are well defined and are independent of their collectors with identified limitations that allow corrections whereas primary data (e.g. interviews, participant observation) used in qualitative research tend to be very limited, depend on their collector and there is no way they can be verified and corrected. Statistical methods are also independent of the researcher and there is little flexibility in the interpretation of the results; the results are provided to the reader as empirical evidence. In contrast, primary data are subject to the condition of the sample (e.g., the mood of the interviewee during the interview), and their interpretation depends heavily on the researcher's point of view. The original data are rarely if ever made available to the reader. Quantitative analysis can be tested by other researchers whereas qualitative research practically cannot.

I believe that for national and local government policy making concerning control of population shifts because of internal migration it is necessary to conduct quantitative analysis. The role of qualitative analysis on migration research for policymaking purposes should be additional to the quantitative analysis and by no means substitutional. For example, random tests using qualitative techniques could check reported trends (of quantitative research) for those individual policy-makers that are not convinced by the latter. Qualitative methods are also useful for examining some very specific and of local interest questions on sensitive issues (such as the social acceptance and integration of newcomers in local communities related to their culture) that cannot be asked to the entire population of a nation.

My personal view is that is it misleading and dangerous to generalise conclusions based on the view of few migrants to important research questions such as what are the factors generating migrants and attracting migrants within national boundaries. It is therefore surprising quantitative methods have been attacked and abandoned in recent years. I hope this work strengthens the trust to the value of quantitative methods, especially the cutting edge techniques (such as GWR) and motivates more researchers to incorporate geocomputational techniques in human geography and more specifically migration studies from the geographical perspective.

The decline in the real share of intellectual activity of quantitative analysis in human geography continues in the 1990s (Longley, 2000) even though data accuracy and new methods have a potential for more robust analysis. Particularly, the use of GIS and spatial analysis is a multibillion industry in the private sector in the North America and Northwest Europe, however, it has little share among human geographers in the UK. Although the ESRC and other funding bodies have publicised their willingness to promote quantitative analysis, practically proposals for such research fail to get funding in favour of cultural geography projects.

There is an ongoing debate for the use of multi-method analysis in human geography and particularly in migration studies (Findlay and Li, 1999; McKendrick, 1999). Findlay and Li (1999) recognise that researchers choose their methods restricted by epistemology and their philosophical perspectives (for a review see Kitchin and Tate, 2000). Thus, there are two poles of methods: quantitative supported by positivism and behavioralism, and qualitative supported by all the remaining philosophical schools (e.g. feminism, postmodernism, poststructuralism). The supporters of multi-method research suggest methods should be independent of the epistemology of the research and particularly in order for migration studies to be complete, both quantitative and qualitative methods have to be used. However, there are practical difficulties for such a research methodology. First, the training and skills of researchers are either qualitative or quantitative. Second, the collaboration of researchers from different poles is restricted by 'communication barriers' because of differences in their research cultures and perspectives.

An interesting work is Casetti's (1999) explanation for the superiority of the mathematical mode of inquiry in human geography. He also calls for a stronger voice in the discipline of those geographers committed to the mathematical mode of inquiry. The debate is ongoing.

2.5 Summary

In this chapter, I provide a review of the related to this study literature. I do this in order to create a theoretical framework and to position this study within existing research. At the same time, I discuss what are the previous methods and findings in order to set a starting point for my research and to demonstrate why the methods I am going to use and the expected findings make my work innovative. There are four parts of this literature: trends in migration, the Lowry debate, migration modelling methods and migration determinants.

The discussion of migration trends concerns migration data recorded either by the Census of Population for England and Wales or by the NHSCR in the last three decades. This literature suggests the continuation of a counterurbanisation phenomenon which is examined further in Chapter 5. I also report migration trends resulted from the NHSCR data that have not been previously reported.

The Lowry debate is an ongoing debate in the literature. It relates to the significance of economic conditions affecting out-migration. Here, I review the most important findings related to this debate. I do this in order to be able to compare my findings with previous work and to take a side on this debate.

I briefly reviewed relevant quantitative research in migration modelling in order to identify the most appropriate techniques for my purposes. I demonstrated that the most appropriate technique for global modelling is multivariate regression and for local modelling it is Geographically Weighted Regression.

I made a thorough review of previous findings of the effects of migration determinants on migration decisions in order to justify the criteria for the selection of these determinants for my models (through providing empirical evidence for their importance in affecting migration). I reported these findings in a form that direct links can be made to the findings of my analysis.

Finally, I comment on the difference between quantitative and qualitative methods in migration, another debate in the discipline. I do this in order to demonstrate my point of view as a researcher when I analyse migration data and decisions.

Overall, this chapter addresses the aim of evaluating empirical work in migration modelling. I now present alternative sources of migration data and a detailed description of the data set I used in this study.

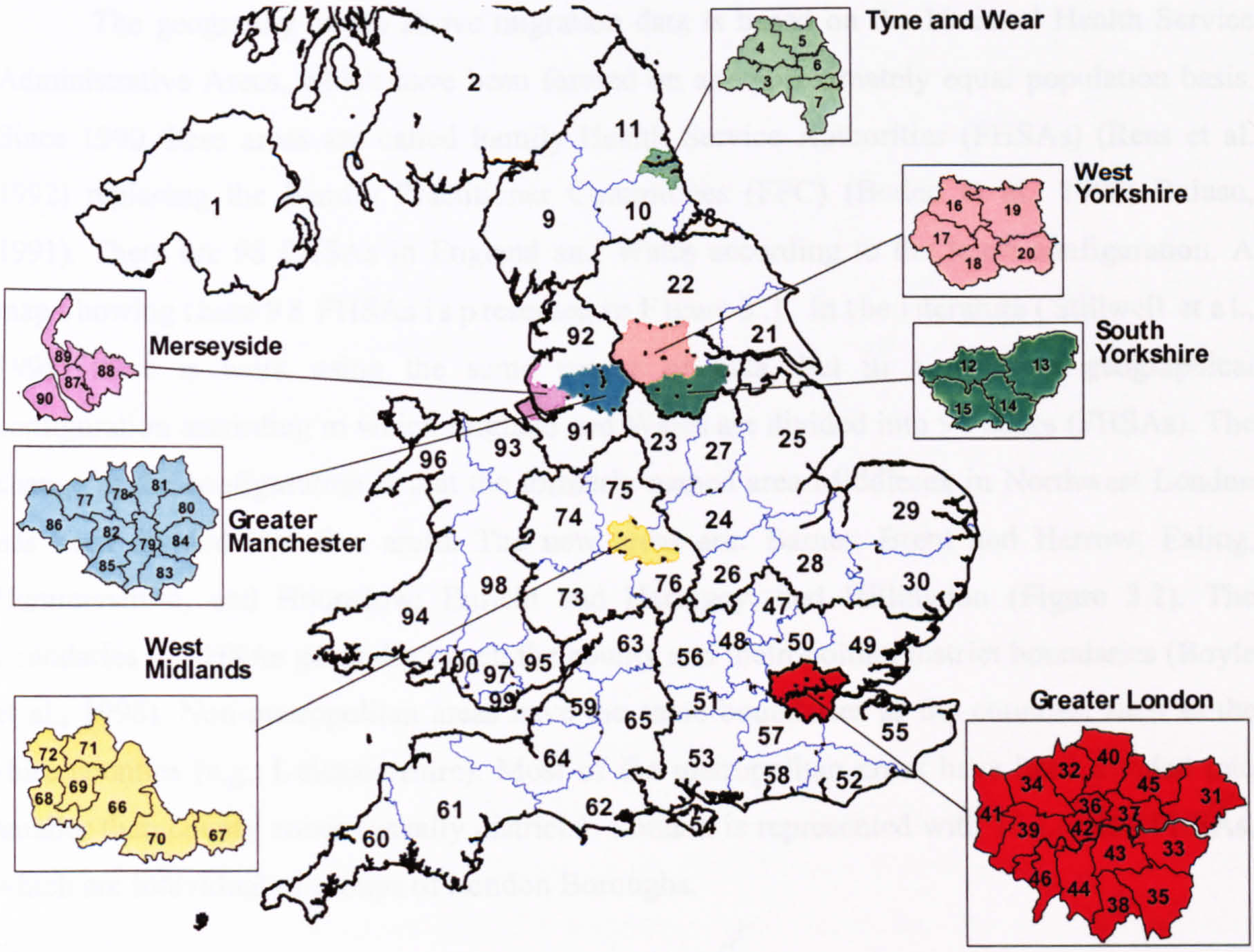
Chapter 3
Data Issues

In this chapter details on migration data and migration determinants available for this thesis are presented. It is necessary to acknowledge that the dataset used here is the result of a research project funded by the Office of Deputy Prime Minister (ODPM, formerly Department of Transport, Local Government and the Regions – DTLR, DETR) which will be referred as the 'ODPM Project'. To date, the ODPM Project has two phases: Phase One is the "Development of a Migration Model for Great Britain" (DETR); and Phase Two is the "Development of a Migration Model: Analytical and Practical Enhancements" (DTLR). Phase One of the ODPM Project involved data collection and modelling, whereas Phase Two involved quality control on data and more robust modelling (Fotheringham et al., 2002b). Here, only the data used in the thesis are discussed.

There are two main sections in this chapter. In section one a full description of migration data is presented. A discussion about migration determinants used here follows in section two. The source of the variables and the details of the raw data are contained in the technical appendixes of the ODPM Project (Fotheringham et al., 2002b), where further acknowledgments and details can be found. In Fotheringham et al. (2003) an updated discussion of the out-migration determinants is presented.

3.1 Migration Data

The migration data used here include out- and in-migration for each zone as well as migration flows from each zone to all others. Their source is the National Health Service Central Register (NHSCR). Taking into account the sensitivity of data confidentiality, NHSCR publishes every quarter data for 100 Family Health Service Authorities (FHSAs) in the UK (one for Scotland and one for Northern Ireland). An interface (TIMMIG) created in Leeds (Stillwell, 1994) allows the extraction of age and sex disaggregated data. In this thesis, annual out-migration data are available for the time period 1983/84 – 1997/98, and annual flow data are only available for the time period 1990/91 – 1996/97. The period covered is determined by the needs of the ODPM Project, and not by the original NHSCR data. The age disaggregation selected for the ODPM Project and also used in this thesis is based on migrants' life stage. Age groups include: children aged 0-15, teenagers aged 16-19, young adults aged 20-24, adults aged 25-29, mature adults aged 30-44, older adults aged 45-59 and pensioners aged 60 and over. Adding the sex disaggregation, 14 sex/age population groups are formed.



1 Northern Ireland

2 Scotland

I North
3 Gateshead
4 Newcastle
5 North Tyneside
6 South Tyneside
7 Sunderland
8 Cleveland
9 Cumbria
10 Durham
11 Northumberland

II YORKS & HUMBERSIDE
12 Barnsley
13 Doncaster
14 Rotherham
15 Sheffield
16 Bradford
17 Calderdale
18 Kirklees
19 Leeds
20 Wakefield
21 Humberside
22 North Yorkshire

III NORTH WEST
77 Bolton
78 Bury
79 Manchester
80 Oldham
81 Rochdale
82 Salford
83 Stockport
84 Tameside
85 Trafford
86 Wigan
87 Liverpool
88 St. Helens & Knowsley
89 Sefton
90 Wirral
91 Cheshire
92 Lancashire

IV EAST MIDLANDS
23 Derbyshire
24 Leicestershire
25 Lincolnshire
26 Northamptonshire
27 Nottinghamshire
28 Cambridgeshire
29 Norfolk
30 Suffolk

V WEST MIDLANDS
66 Birmingham
67 Coventry
68 Dudley
69 Sandwell
70 Solihull
71 Walsall
72 Wolverhampton
73 Hereford and Worcester
74 Shropshire
75 Staffordshire
76 Warwickshire

VI SOUTH WEST
59 Avon
60 Cornwall
61 Devon
62 Dorset
63 Gloucestershire
64 Somerset
65 Wiltshire

IX WALES
93 Clwyd
94 Dyfed
95 Gwent
96 Gwynedd
97 Mid Glamorgan
98 Powys
99 South Glamorgan
100 West Glamorgan

VII GREATER LONDON
31 Barking, Havering
32 Barnet
33 Bexley, Greenwich
34 Brent, Harrow
35 Bromley
36 Camden, Islington
37 City, Hackney, Newham, Tower Hamlets
38 Croydon
39 Ealing, Hammersmith, Hounslow
40 Enfield, Haringey
41 Hillingdon
42 Kensington & Chelsea, Westminster
43 Lambeth, Southwark, Lewisham
44 Merton, Sutton, Wandsworth
45 Redbridge, Waltham Forest
46 Richmond, Kingston

VII REST OF SOUTH EAST
47 Bedfordshire
48 Buckinghamshire
49 Essex
50 Hertfordshire
51 Berkshire
52 East Sussex
53 Hampshire
54 Isle of Wight
55 Kent
56 Oxfordshire
57 Surrey
58 West Sussex

Figure 3.1. Map of the 100 FHSAs in the United Kingdom

(remapped based on Stillwell et al., 1995, p. 344)

The geography of the above migration data is based on the National Health Service Administrative Areas, which have been formed on an approximately equal population basis. Since 1990 these areas are called Family Health Service Authorities (FHSAs) (Rees et al. 1992) replacing the Family Practitioner Committees (FPC) (Boden et al., 1992; Bulusu, 1991). There are 98 FHSAs in England and Wales according to the latest configuration. A map showing these 98 FHSAs is presented in Figure 3.1. In the literature (Stillwell et al., 1995) there is work using the same source of data, but in a different geographical configuration according to which England and Wales are divided into 94 zones (FHSAs). The change in the configuration is that the formerly named area Middlesex in Northwest London has been divided into five areas. The new areas are: Barnet; Brent and Harrow; Ealing, Hammersmith, and Hounslow; Enfield and Haringey; and Hillingdon (Figure 3.2). The boundaries of FHSAs generally match the county and metropolitan district boundaries (Boyle et al., 1998). Non-metropolitan areas have the same boundaries as the counties, such as the shire counties (e.g.: Leicestershire). Most of the metropolitan areas have been divided into smaller than county zones (usually districts). London is represented with 16 London FHSAs, which are individual or groups of London Boroughs.

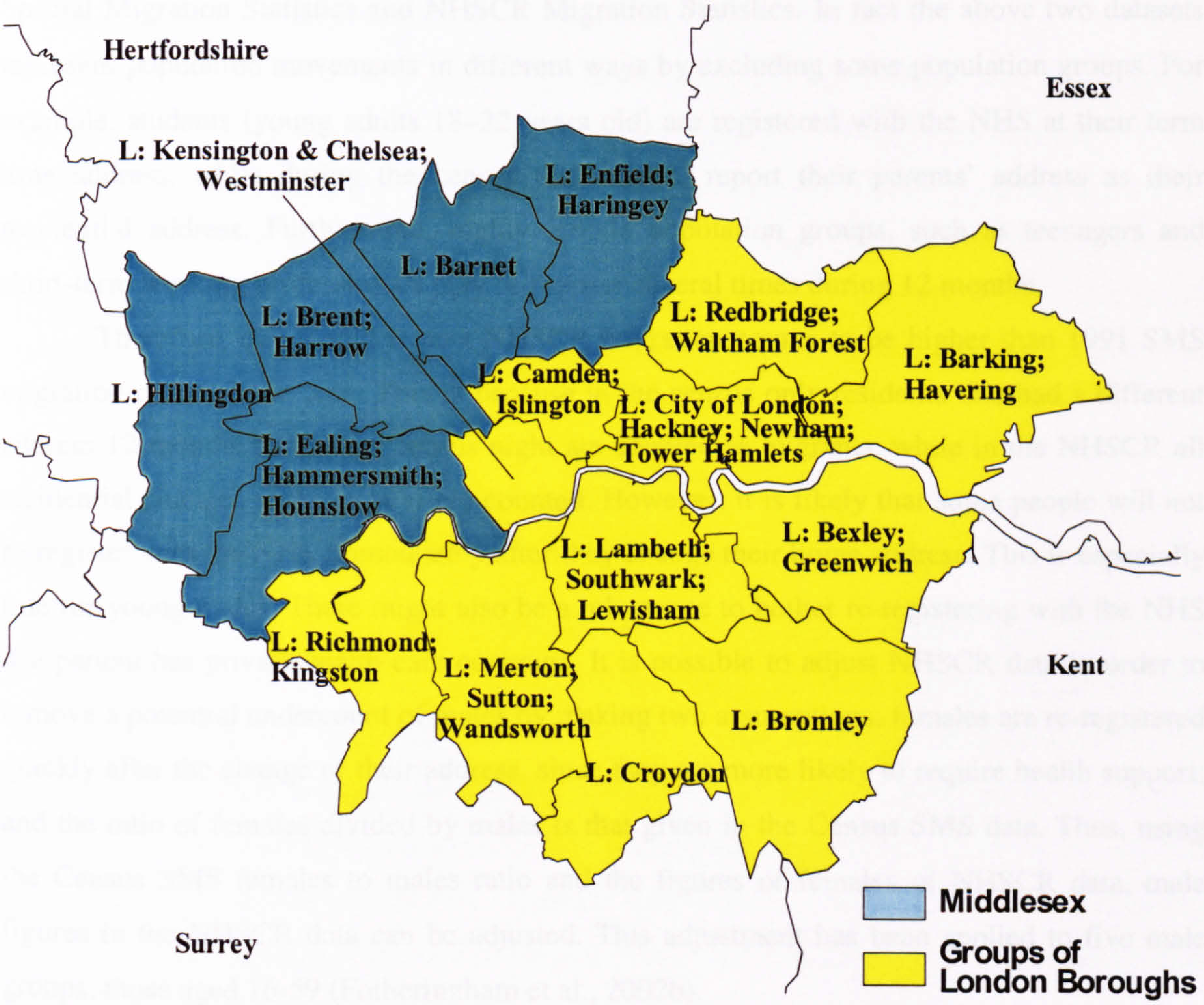


Figure 3.2. FHSAs in London

The NHSCR migration data is derived from individuals' medical records. *Most people in Britain are registered with a medical doctor known as general practitioner (GP). Family health service authorities (FHSAs) administer the payment of GPs and this is based, almost entirely, upon the number of patients that are registered with them. ... When a patient migrates between FHSAs and registers with a new GP, their medical records are passed between the origin and destination FHSAs through the National Health Service Central Register (NHSCR), where a record of the move is made. From this source, it is possible to derive a continuous series of inter-FHSA migration data, broken down by age and sex.* (Boyle et al., 1998, p. 41).

3.1.2 The quality of NHSCR data

In the recent literature the quality of NHSCR data and whether they represent the real world in good detail (Rees et al., 1992; Stillwell et al., 1996) is discussed. Devis and Mills (1986), Boden et al. (1992), and Stillwell et al. (1996) make a comparison between 1991 Special Migration Statistics and NHSCR Migration Statistics. In fact the above two datasets represent population movements in different ways by excluding some population groups. For example, students (young adults 18–22 years old) are registered with the NHS at their term time address, while during the census they should report their parents' address as their residential address. Furthermore, highly mobile population groups, such as teenagers and short-term working professionals may re-register several times during 12 months.

Therefore, one would expect NHSCR migration counts to be higher than 1991 SMS migration counts in an area. This is because in the census only residents that had a different address 12 months before the census night are counted as migrants, while in the NHSCR all residential changes within a year are counted. However, it is likely that some people will not re-register with the NHS immediately after they change their home address. This is especially true for young males. There might also be a reluctance to bother re-registering with the NHS if a patient has private health care coverage. It is possible to adjust NHSCR data in order to remove a potential undercount of males by making two assumptions: females are re-registered quickly after the change of their address, since they are more likely to require health support; and the ratio of females divided by males is that given in the Census SMS data. Thus, using the Census SMS females to males ratio and the figures of females of NHSCR data, male figures in the NHSCR data can be adjusted. This adjustment has been applied to five male groups, those aged 16-59 (Fotheringham et al., 2002b).

The above issues are also discussed in earlier studies using NHSCR data (Ogilvy, 1979; 1980; 1982). Ogilvy (1982) suggests that an appraisal of NHSCR migration figures showed that the figures provide a useful measure of population movement at the regional scale; however, they must be carefully interpreted because of measurement differences with Census data. She confirms that the NHSCR migration figures are higher than those of the Census: *in 1971 the NHSCR recorded about 20% more inter-regional moves than did the Census.* There are two main reasons for this: *the entries in the NHSCR include multiple and return moves made within a year, as well as moves made by students and by children less than one year old; both groups are excluded from the census definition of migrants* (Ogilvy, 1982, p. 66). However, there are also two main reasons for a potential undercount of migrants in the NHSCR data: migrants who do not re-register and *short distance moves made by people who do not change their doctor even though their moves have crossed an administrative boundary* (Ogilvy, 1982, p. 66). A full description of limitations in the NCHSR data is provided in an appraisal of the NHSCR as a data source by Ogilvy (1980).

Another issue reported by Ogilvy (1982, p. 66) is that *there is often a time lag between change of address and transfer to a new doctor, but NHSCR figures allow for this by assuming the average delay to be three months.*

The reason that NHSCR data have been selected for this thesis is that they provide migration statistics for many years and allow temporal analysis, which is not possible with the Census SMS data. Ogilvy (1982, p. 66) suggests that *when the NHSCR figures are used as time-series, it is reasonable to assume that any bias or other difference in the figures has remained generally constant over time; thus one may take the variation in the series as an index to relative changes in migration flows.* The latter has been implemented by Stillwell et al. (1995).

3.1.3 Other sources of migration data

In this section a comparison between the NHSCR data and other main sources of migration data as well as a review of the latter is presented. The main source of migration data other than the NHSCR Migration Estimates is the Census of Population Special Migration Statistics (SMS) (now available through the Web-based interface to Census Interaction Data – WICID; Stillwell and Duke-Williams, 2003). There are several studies on NHSCR data that also provide information about other migration data sources. The most detailed presentation of migration data sources is the one of Bulusu (1991). He examines actual and potential data sources for internal as well as international migration. However, his review is out-of-date.

Table 3.1. Comparison of migration measures and advantages and disadvantages of main sources of internal migration data

	Measure of Migration	Advantages	Disadvantages
Census	Change of usual address over one year period.	• Population coverage *Usually resident population* • Age-gender details for migrants *Age 1 and above* • Smallest area migration estimates *Wards*	• Frequency of measurement *Only every ten years (and data not available until at least one year after the Census)* • Missed migrants *Infants, persons who move shortly before death or leaving the country* • Other problems e.g. 1991 Census *Under-enumeration of population – migrants disproportionately represented* *An additional 7 percent of Census respondents who stated that they lived at a different address a year ago, did not provide information on previous address.* *Adjustment required to transfer students from home address to term-time address*
National Health Service Central Register (NHSCR)	All patient re-registration with doctors in different FHSAs. Multiple moves in one year possible.	• Frequency of measurement *Continuous* • Age-gender details for migrants *All ages*	• Population coverage *NHS patients. Excluded: persons not registered with the NHS, armed forces personnel, long-term prisoners, patients in long-stay hospitals* • Smallest area migration estimates *FHSAs*
Electoral Registers	Relative change over one year period in the number of people eligible to vote.	• Frequency of measurement *Annual* • Smallest area migration estimates *Wards*	• Population coverage *Exclude usual residents who are ineligible to vote (e.g. non-Commonwealth) and children* *Some double counting (dual registration is legal although voting in more than one place is not). Unlikely that every eligible person is on the register particularly those aged under 25.* • Age-gender details for migrants *No* • Other problems *Affected by variations between areas and over time in the propensity of people to register* *Affected by variations between areas in the maintenance and updating of registers*
ONS FHSA Postcoded Data System (FPDS)	Change of usual address (postcode) over one year period.	• Frequency of measurement *Annual* • Age-gender details for migrants *Age 1 and above* • Smallest area migration estimates *Unit postcode*	• Population coverage *NHS patients. Excluded: persons not registered with the NHS, armed forces personnel, long-term prisoners, patients in long-stay hospitals* • Missed migrants *Infants, persons who move shortly before leaving and those who move soon after joining or rejoining the NHS*

Source: Scott and Kilbey, 1999, Table 2, p. 49

Scott and Kilbey (1999) discuss the FHSA Registers data, also coming from NHS, as an improved dataset of migration data compared to the NHSCR migration estimates. The following tasks constitute these improvements.

The ONS processing system – The FHSA Postcoded Data System (FPDS) derives migration estimates by:

- *Removing information on temporary patients (as these people can be present on more than one register);*
- *Eliminating duplicate registrations (i.e. those with the same NHS number);*
- *Matching individual's records (using NHS number) for consecutive years;*
- *Isolating cases with changes in postcodes in consecutive years (ignoring those changes in postcode that are re-classifications by Royal Mail);*

- *Tabulating inflows and outflows (by gender and age) for the geographical areas of interest.* (Scott and Kilbey, 1999, p. 45).

Table 3.1 shows a comparison of this new dataset with NHSCR migration data, the 1991 Census of population Migration Statistics and the Electoral Registers migration estimates. These four sources are the main sources of secondary data on migration. Other sources of migration data are the Samples of Anonymised Records (SARs) also from the Census of Population and the Longitudinal Study (LS) which monitors a sample of individuals every decade, taking place the same year as the Census of Population. The latter two sources provide micro-data (Stillwell, 1994). Table 3.2 shows a summary of the migration data sources other than those in Table 3.1. The table rows in grey background show potential and not actual sources of migration data.

Table 3.2. Other Sources and potential sources of migration data

Data Source	Frequency of measurements	Geography	Institution	Population Cover	Reference
Samples of Anonymised Records	Every 10 years/ Since 1961		ONS: Census		Stillwell, 1994
Longitudinal Study	Every 10 years / Since 1971		OPCS / ONS	Continuing study of 1% of the population	Stillwell, 1994; Bulusu, 1991
General Household Survey (GHS)	Annual / Migration questions since 1991	Postcode		20,000 individuals / 12,500 private households	Boyle et al., 1998; Bulusu, 1991
Labour Force Survey (LFS)	Quarterly (5 times per year)	Region	OPCS: Social Survey Division	60,000 – 100,000 households	Bulusu, 1991
Community Charge Registers	Annual	Building-brick areas (aggregated from electoral wards)	Community Charge Registers Officers (Local Authorities)	Migration data by change of address	Bulusu, 1991
Departmental Central Index		Address / Postcode	Department of Social Security	All persons with a National Insurance Number	Bulusu, 1991
Inland Revenue Data	Annual	Address / Postcode	Inland Revenue	Tax payers who send a tax return form	Bulusu, 1991
Driver Licensing		Address / Postcode	Driver and Vehicle Licensing Agency (DVLA)		Bulusu, 1991

There are many other potential sources of migration data (Bulusu, 1991) but with poor data quality (e.g., TV licensing), thus, inappropriate for analysis. In recent years there have been several problems with secondary data. Two main issues are confidentiality and licensing. The former concerns restriction in the aggregation level of data available for analysis, to ensure that no individuals can be identified using such data. The latter is that a licensing scheme is necessary to provide revenue to the data holders for providing the data. Thus, many data cannot be accessed from the academic community or the public.

Finally, even though many Local and National Government Departments hold computerised information on people's addresses, and thus, have a potential for extracting migration data, no policies for such action exist. In contrary, many of the existing secondary data on migration flows are not fully explored because of the lack of interest in statistical analysis of the latter among the human geographers in recent years.

3.2 Migration Determinants

Migration determinants are defined as the variables that measure a characteristic of a zone (FHSA) in which migration has been counted and are used within models to help understand the causes of migration. In the literature, there is no consistent set of variables used in migration models. In most cases, researchers use several migration determinants (other than population, distance and measures of centrality) when data are available, recognising the importance of the inclusion of such determinants for better model specifications. Thus, one would argue that the more variables are included in a model, the better estimates a model can produce.

However, there are three issues concerning the above argument: one is the limitation of data availability; a second concerns the limitations of the observations to be modelled (many variables will reduce the degrees of freedom); and thirdly, the possibility of the correlation between some of the variables.

The first issue has been partly addressed here. For the research presented in this thesis, there are 140 variables available to explain out-migration and 60 variables to explain destination choice. Although the quality of the measures of some of these variables is debatable, they cover most of the possible factors affecting migration. What they do not cover are issues such as voting patterns, religion, quality of educational amenities and general cost of living at the FHSA level.

The second issue needs to be taken more into account. A migration model with high degrees of freedom provides stronger evidence for the effects of migration determinants on migration flows. The degrees of freedom increase when the number of observations increases and decrease when the number of variable increases, ceteris paribus. For each model of the age/sex disaggregated data, there are 98 observations available for 14 years of data. In previous work (Fotheringham et al., 2002b), all these data are combined resulting in 1372 observations. In this work, model specifications included approximately 45 variables and a same number of their quadratic terms. These models had reasonably high value of degrees of freedom. However, the methodology used here allows only 98 observations at a time. Thus, to allow a model to have substantially high degrees of freedom, I suggest the number of variables should be limited to 14 – 18, the most important ones.

The final issue has been addressed by auxiliary regression exercises. It is easy to identify the pairs of variables that are highly correlated between each other by studying the Variance Inflation Factor (VIF) of each variable in a model and removing the appropriate variables. More details on this issue are presented in Section 4.1.5.

It is not necessary to discuss every single variable of the dataset when it has not been used in the analysis discussed below. The migration determinants have been divided into two groups: those explaining out-migration and those explaining destination choice. The observed migration data available here include 14 years of observations for out-migration determinants and 7 years of observations for destination-choice. Thus, determinants are required for the equivalent time periods where possible. Obviously, some of the migration determinants are common to both phases of the migration modelling. An out-migration determinant can be also a destination-choice determinant, since every origin can be a destination in the migration zone system. However, some of the variables are appropriate and available only for one phase of the migration modelling.

The variables available are time, age/sex, or sex-specific. Many of the variables are derived from the 1991 Census of Population data and thus are available only for one year. These variables are called cross-sectional variables. In the out-migration modelling, all the time-specific variables have been lagged by one year. The theoretical explanation for the lagging of some variables is based on the assumption that a decision to migrate is affected by the conditions that occurred at some time (a year) before the actual migration.

Based on their nature, the variables available can be classified into eight types: spatial structure (e.g., contiguity); demography (e.g., non-white persons); economic (e.g., household income); employment (e.g., unemployment rate); housing (e.g., vacant dwellings); social (e.g., high social class); environmental (e.g., mean July temperature); and access to services, amenities and miscellaneous variables (e.g., council tax). The selection of the variables has been informed by the findings of the review *The Determinants of Migration Flows in England* (Champion et al., 1998).

3.2.1 Out-Migration Determinants

In this section a discussion of out-migration determinants is presented. A detailed description of each variable used in the out-migration models (Chapter 6) is given. This includes a brief description of the variable, details on its construction (e.g., the numerator and denominator if it is a ratio), information about its source, the methodology of its calculation when applicable (e.g., indexes), and occasionally some notes on its quality. More information (including the range of its values) is provided for indexes derived from principal component analysis and for regional variables (explanation follows). The Principal Component Analysis method is discussed just before the out-migration variables.

As mentioned above, it was necessary to select the 14-18 most important variables for the models. In order to do this, two issues were taken into account: one was the empirical evidence for their significance in measuring migration found in the literature and the second was the inclusion of a representative of all eight types of variables discussed above. Thus, comparisons with previous work can be made. Because each model is age group specific, it has a different configuration for each age group and any particular model will not include all the 16 variables that are discussed below. More details on the models can be found in the modelling/analysis chapter (Chapter 6).

Each of the following 16 variables has a name and a code (presented in brackets). Within the code there is information about the nature of the variable. Some of the variables are time-series in nature. The suffix 'L' in the code denotes a variable that is time-specific and has been lagged by one year. Three variables are indexes, each one derived from a principal component analysis of variables measuring similar attributes. The values of these variables are negative for some FHSAs and positive for others. Thus, because of the non-positive values, these variables cannot be logged in the log-log OLS model. These are denoted with the suffix 'UNLG'. Finally, there is one regional variable. Regional variables are meant to capture the possible pull effects on out-migration caused by conditions elsewhere in the country (Fotheringham et al., 2002b). The suffix 'Y' denotes that a variable is a regional variable. More explanation on regional variables is provided below.

It is necessary to acknowledge that information about the variables, their sources and robustness has been discussed in the Technical Appendixes of the ODPM Project (Fotheringham et al., 2002b)

Principal Component Analysis

The SPSS (Statistical Package for Social Sciences), version 10, Principal Component Analysis technique was used here. Some notes from the online help of this software follow. Principal Component Analysis for SPSS is a Data Reduction/Factor Analysis function. *Factor analysis attempts to identify underlying variables, or factors, that explain the pattern of correlations within a set of observed variables. Factor analysis is often used in data reduction to identify a small number of factors that explain most of the variance observed in a much larger number of manifest variables. Factor analysis can also be used to generate hypotheses* *regarding causal mechanisms or to screen variables for subsequent analysis (for example, to identify collinearity prior to performing a linear regression analysis).*

In this study the *variables* are the out-migration determinants that measure similar phenomena and the *principal component* is the resulting out-migration determinant after the principal component analysis has been applied to the former. Principal Component Analysis will result in at least three components, the first of which is called *principal component* and explains most of the variance in the variables. Principal Component Analysis results in three main output tables (among others): the *Correlation Matrix* which gives information about the collinearity between the variables, the *Total Variance Explained* which gives information about the variance explained in each component (eigenvalues), and the *Component Matrix* which gives the parameter estimates of each variable selected for the principal component.

For example for Crime Index (more details below) these tables are Table 3.3, Table 3.4 and Table 3.5 respectively.

Table 3.3. Crime Index PCA: Correlation Matrix

		RTOFF	ILCHIP	SEHCR
Correlation	RTOFF	1	0.47385	0.48623
	ILCHIP	0.47385	1	0.59328
	SEHCR	0.48623	0.59328	1

Table 3.4. Crime Index PCA: Total Variance Explained

	Initial Eigenvalues			Extraction Sums of Squared Loadings		
Component	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	2.037513	67.91709	67.91709	2.037513	67.91709	67.91709
2	0.55614	18.538	86.45509			
3	0.406347	13.54491	100			
Extraction Method: Principal Component Analysis.						

Table 3.5. Crime Index PCA: Component Matrix [a]

	Component
	1
RTOFF	0.781581
ILCHIP	0.841612
SEHCR	0.847546
Extraction Method: Principal Component Analysis.	

[a] 1 components extracted.

For a better understanding of PCA it is necessary to present some more technical details. Some concepts and definitions follow. *Total variation* in the data with regard to the variables X_1, X_2, ..., X_k is mathematically defined to be the sum of the sample variances of the k variables:

$$\text{total variation} = S_1^2 + S_2^2 + \cdots + S_k^2 \tag{3.1}$$

where S_j^2 is the sample variation of $X_j, j = 1, 2, \ldots, k$. *The purpose of principal-components is to explain as much of the total variation in the data as possible with as few factors (i.e., principal components) as possible.* The *first principal component*, PC(1), is the weighted linear combination of the variables that accounts for the largest amount of the total variation in the data.

$$\mathrm{PC}(1) = w_{(1)1}X_1 + w_{(1)2}X_2 + \cdots + w_{(1)k}X_k \tag{3.2}$$

where the weights $w_{(1)j}$, $j = 1, 2, \ldots, k$, have been chosen to maximise the quantity $\frac{\text{variance of PC(1)}}{\text{total variation}}$, and to satisfy the restriction $\sum_{j=1}^{k} w_{(1)j}^2 = 1$ so that the variance of PC(1) will not exceed the total variation" (Kleinbaum et al., 1988, pp. 615 – 616). Each additional principal component can be similarly defined ($\mathrm{PC}(i) = w_{(i)1}X_1 + w_{(i)2}X_2 + \cdots + w_{(i)k}X_k$). The *correlation matrix* shows the relationships between a set of variables and has the following form

$$\mathrm{R} = \begin{array}{c|cccc} & X_1 & X_2 & \cdots & X_k \\ \hline X_1 & 1 & r_{12} & \cdots & r_{1k} \\ X_2 & r_{12} & 1 & \cdots & r_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ X_k & r_{1k} & r_{2k} & \cdots & 1 \end{array} \tag{3.3}$$

where $r_{ij} : i, j = 1,2,\ldots,k, \text{and}\, i \neq j$, is the correlation between X_i and X_j (r is the correlation coefficient). Table 3.3 is the correlation matrix for the components of Crime Index PC(1). Table 3.4 shows how much of the total variance Crime Index PCs explain. In this case PC(1) explains 67.92% of the total variance, PC(2) explains 18.54% of the total variance, and PC(3) explains 13.54% of the total variance. Note that the number of principal components cannot exceed the number of components in a PCA. Thus, here only three PCs can be calculated. Finally, Table 3.5 shows the weights of the three components of Crime Index PCA. In the case of Crime Index: $k = 3$, X_1 = Offences recorded by police (RTOFF), X_2 = Household insurance premiums (ILCHIP), X_3 = Crime as a serious problem (SEHCR), $w_{(1)1}$ = 0.782, $w_{(1)2}$ = 0.842, and $w_{(1)3}$ = 0.848.

Air Index (AIR_UNLG)

Air Index is the first principal component derived from the variables NO_2, and Ozone. It is actually their difference: AIR $_j$ = 0.926 * $NO_{2\,j}$ - 0.926 * OZONE $_j$. More information about NO_2 and Ozone is presented in Table 3.6.

Table 3.6. NO_2 and Ozone in a nutshell

Variable Name	NO_2
Description	Annual mean of NO_2 level
Source	1996 (AEA Technology)
Notes	FHSA values are each taken from the location in the dataset nearest to their population-weighted centroid

Variable Name	Ozone
Description	Number of days above 50ppb hour mean
Source	1995 (AEA Technology)
Notes	FHSA values are each taken from the location in the dataset nearest to their population-weighted centroid; high ozone levels are a form of poor air quality *but* may not be seen as such by potential migrants

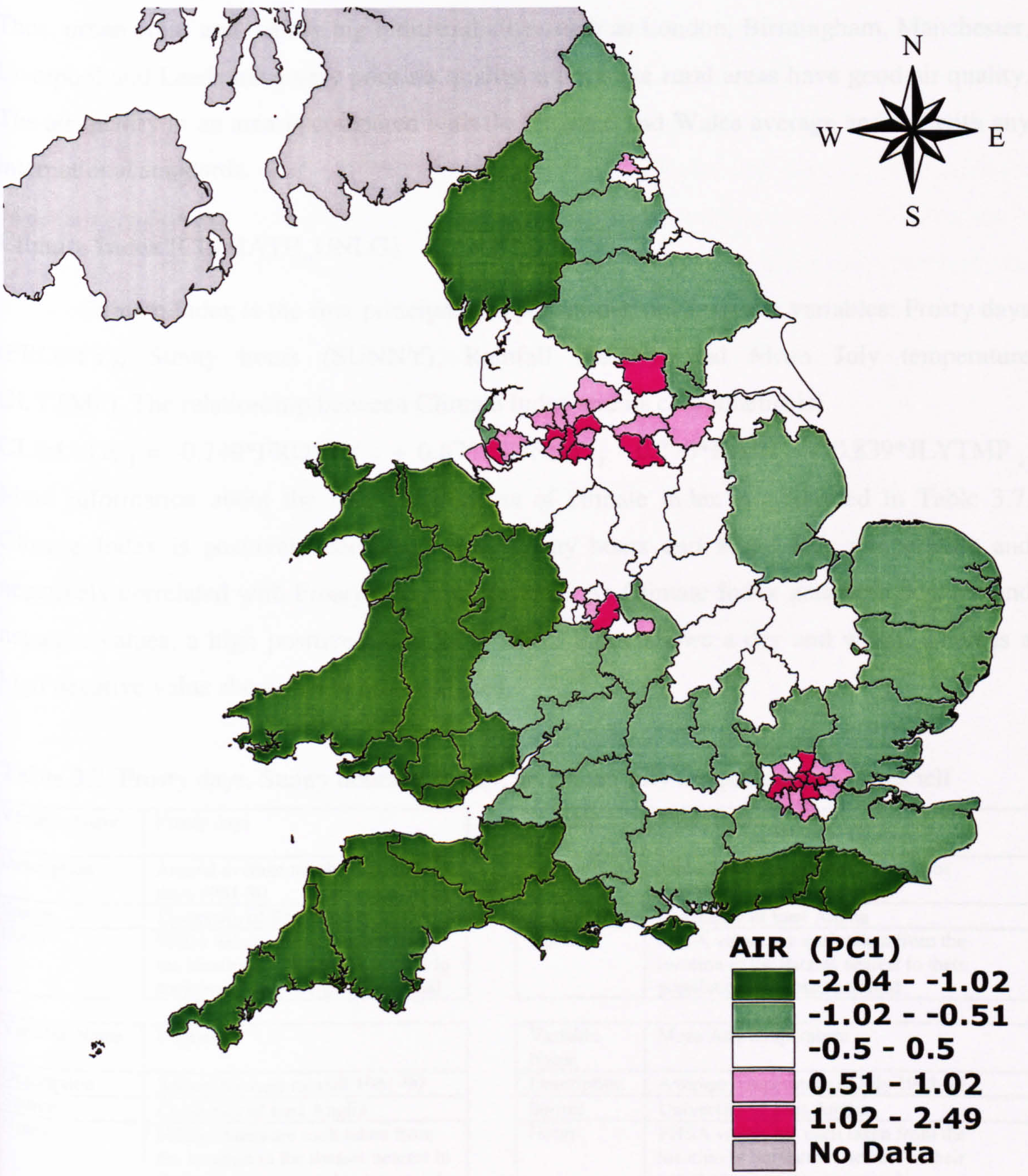


Figure 3.3. Map of Air Index for England and Wales

Air Index has negative values where the air quality is good and positive values where the air quality is poor. Figure 3.3 shows a map of the Air Index for England and Wales. As might be expected, rural areas have relatively better air quality than urban areas. FHSAs in Wales, Cumbria, Shropshire, FHSAs in the South West (Cornwall; Devon; Somerset; Dorset), East and West Sussex; and the Isle of Wight enjoy relatively good air quality whereas Leeds, Liverpool, Salford, Manchester, Oldham, Stockport, Sheffield, Birmingham and some of the FHSAs in London suffer from poor air quality. Cornwall enjoys the best air quality (index = -2.04), and Kensington & Chelsea and Westminster suffer the worst air quality (index = 2.49). Thus, urban areas and mainly big industrial cities such as London, Birmingham, Manchester, Liverpool and Leeds have very poor air quality and remote rural areas have good air quality. The air quality in an area is compared with the England and Wales average and not with any international standards.

Climate Index (CLIMATE_UNLG)

Climate Index is the first principal component derived from the variables: Frosty days (FROSTY), Sunny hours (SUNNY), Rainfall (RAIN), and Mean July temperature (JLYTMP). The relationship between Climate Index and its components is:

$CLIMATE_j = -0.749*FROSTY_j + 0.875*SUNNY_j - 0.770*RAIN_j + 0.839*JLYTMP_j$.

More information about the four components of climate index is presented in Table 3.7. Climate Index is positively correlated with Sunny hours and Mean July temperature and negatively correlated with Frosty days and Rainfall. As Climate Index gets both positive and negative values, a high positive value for Climate Index shows a dry and warm, whereas a high negative value shows a wet and cold area.

Table 3.7. Frosty days, Sunny hours, Rainfall, and mean July temperature in a nutshell

Variable Name	Frosty days
Description	Annual average number of frosty days 1961-90
Source	University of East Anglia
Notes	FHSA values are each taken from the location in the dataset nearest to their population-weighted centroid

Variable Name	Sunny hours
Description	Annual average number of hours of sunshine 1961-90
Source	University of East Anglia
Notes	FHSA values are each taken from the location in the dataset nearest to their population-weighted centroid

Variable Name	Rainfall
Description	Annual average rainfall 1961-90
Source	University of East Anglia
Notes	FHSA values are each taken from the location in the dataset nearest to their population-weighted centroid

Variable Name	Mean July temperature
Description	Average temperature in July 1961-90
Source	University of East Anglia
Notes	FHSA values are each taken from the location in the dataset nearest to their population-weighted centroid

One would expect Climate Index for the west part of England and Wales to take negative values and for the South and Southeast England to take highly positive values. A map showing the spatial variation of Climate Index is presented in Figure 3.4. There are three clear zones of different climate. Climate Index for FHSAs in the North and Northwest England, Yorks and Humberside, West Midlands and Wales has negative values, for FHSAs in the Southwest and Southeast England (only those FHSAs by the sea) and London it has positive values, and in the rest of England (FSHAs in Southwest and Southeast England and East Midlands) has close to zero (average) values.

There are four FHSAs for which the Climate Index has unexpected values. These are South Tyneside, Sunderland, Sefton and Wirrall. In Table 3.8 the values of Climate Index and its components are shown for the above FHSAs, for a neighbour of the above FHSAs for comparison and for averages of the FHSAs in England and Wales. The high Climate Index for South Tyneside and Sunderland (they are actually the same) compared to that of Newcastle is due to fewer Frosty days and Rainfall and higher Sunny hours and Mean July Temperature. The same applies when Sefton and Wirrall is compared to Liverpool. It is important to note that the map in Figure 3.4 is sensitive to changes in the range of each Climate Index value interval.

Table 3.8. List of Climate Index and values of its components for selected FHSAs

FHSA	FROSTY	SUNNY	RAIN	JLYTMP	Climate Index
Newcastle	40.129	40.752	21.390	42.580	-0.108
South Tyneside	27.667	43.182	17.286	56.440	0.619
Sunderland	27.667	43.182	17.286	56.440	0.619
Sefton	40.652	55.276	25.878	72.280	0.760
Wirral	44.487	49.489	24.083	74.260	0.629
Liverpool	47.973	45.640	27.738	72.280	0.356
England and Wales Mean	54.647	48.173	25.128	58.800	

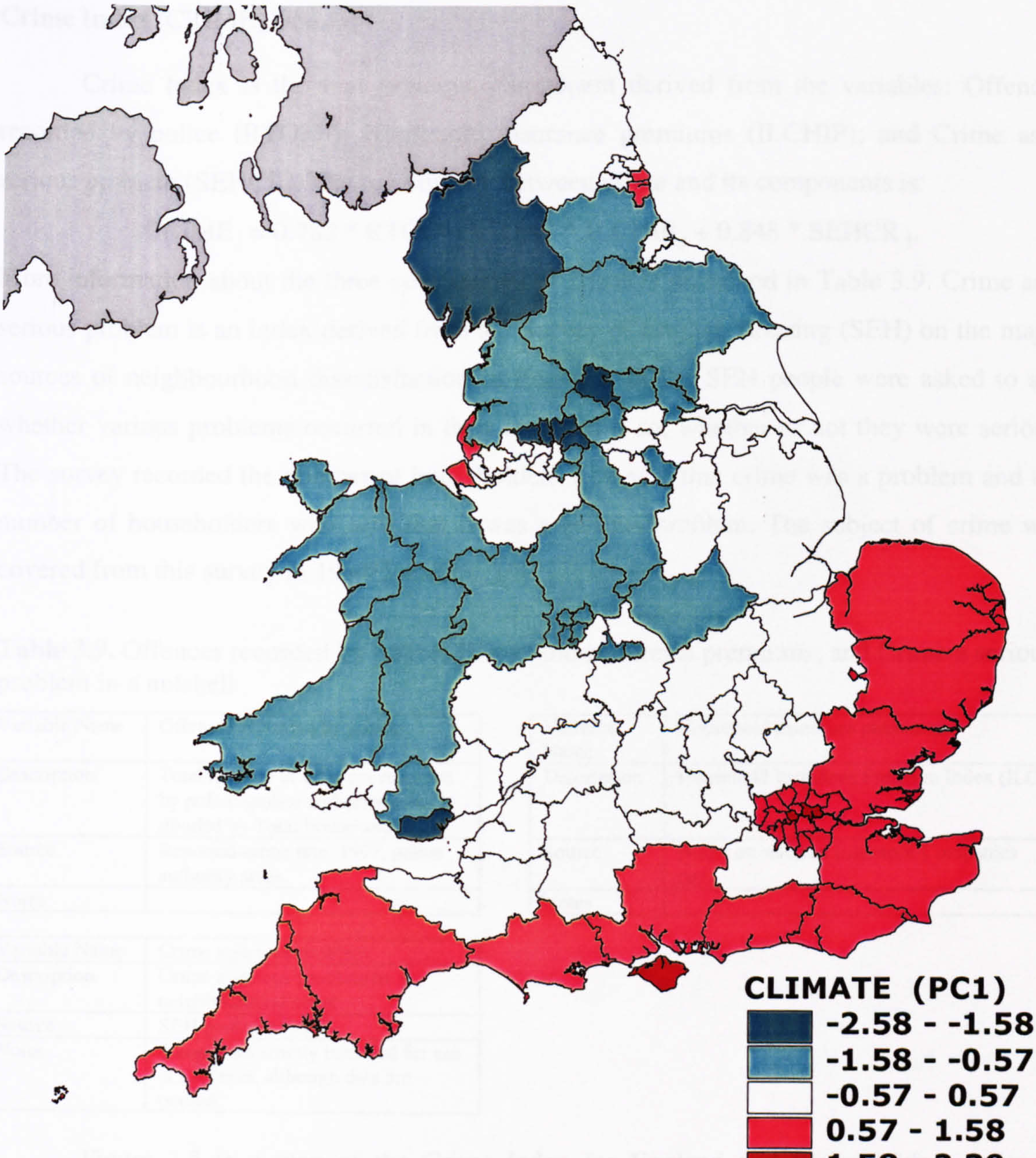


Figure 3.4. Map of Climate Index for England and Wales

Percentage of long distance commuters (COMMUT)

Percentage of long distance commuters is the number of commuters living 10 kilometres or more from their workplace divided by the number of economically active people in the same area. The numerator (commuters living 10 kilometres or more from their workplace) is 10% sample data, but this is perfectly adequate. The source of this data is the 1991 Census of Population.

Crime Index (CRIME_UNLG)

Crime Index is the first principal component derived from the variables: Offences recorded by police (RTOFF); Household insurance premiums (ILCHIP); and Crime as a serious problem (SEHCR). The relationship between crime and its components is:

$$CRIME_j = 0.782 * RTOFF_j + 0.842 * ILCHIP_j + 0.848 * SEHCR_j.$$

More information about the three components of crime is presented in Table 3.9. Crime as a serious problem is an index derived from the Survey of English Housing (SEH) on the major sources of neighbourhood dissatisfaction in England. In the SEH people were asked to say whether various problems occurred in their area and if so, whether or not they were serious. The survey recorded the number of householders who said that crime was a problem and the number of householders who said that it was a serious problem. The subject of crime was covered from this survey in 1994/5 and 1997/8.

Table 3.9. Offences recorded by police; Household insurance premiums; and Crime a serious problem in a nutshell

Variable Name	Offences recorded by police
Description	Total number of offences recorded by police (police authority area) divided by Total households
Source	Reported crime rate, 1997, police authority areas.
Notes	

Variable Name	Household insurance premiums
Description	Household Insurance Premium Index (ILC)
Source	Based on selected Insurance Companies rates
Notes	

Variable Name	Crime a serious problem
Description	Crime a serious problem in neighbourhood score
Source	SEH pooled 1995-8
Notes	Sample not strictly intended for use at LA level, although data are pooled

Figure 3.5 is a map of the Crime Index for England and Wales. Crime Index is positively correlated with its components. The Crime Index contains both positive and negative values. The negative values denote a relatively lower crime rate than the England and Wales average and the positive values a higher crime rate. One would expect high crime to be an urban characteristic, which it clearly is with the higher rates being in Manchester (2.4), Salford (2.2), Trafford (1.7), FHSAs in London including the City of London with Hackney, Newham and Tower Hamlets FHSA (1.5), Bolton (1.33), Newcastle (1.32), Lambeth with Southwark and Lewisham FHSA (1.31), and Tameside (1.29). In contrast, Suffolk (-2.1), Cornwall (-2.0), Wiltshire (-2.0), Isle of White (-1.8), Devon (-1.7), North Yorkshire (-1.56), Lincolnshire (-1.55), Dorset (-1.47) and Oxfordshire (-1.40) are the areas with the lowest Crime Index in England and Wales.

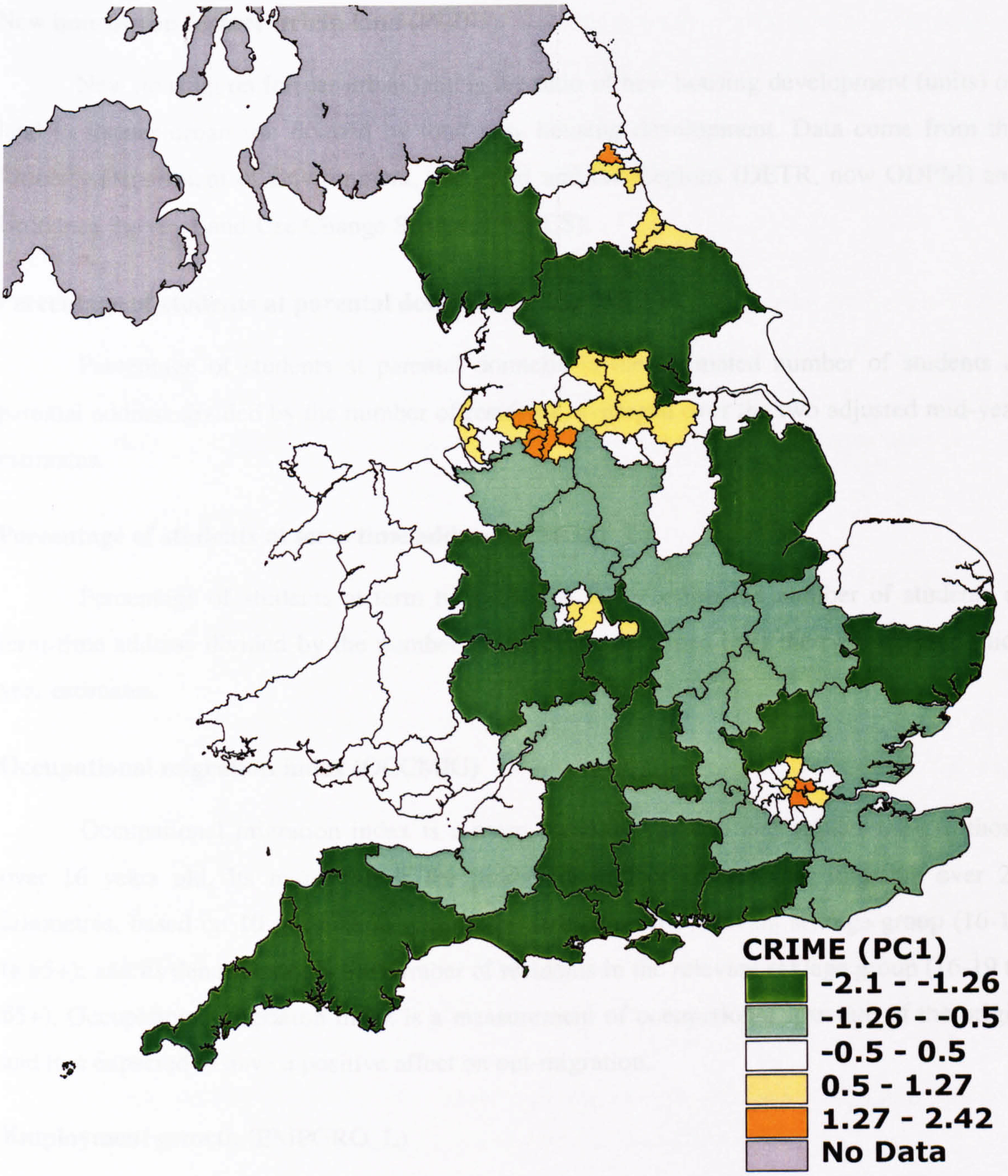


Figure 3.5. Map of Crime Index for England and Wales

Percentage non-white (NONWH)

Percentage non-white is calculated as the total non-white population divided by total population. Its source is the 1991 Census. The number of non-white population at Census date 1991 was extracted from ECPOP (European Community POPulation projection model; Rees, 1996, p. 336) for districts and then aggregated to FHSAs. The percentage of non-white population was then calculated using the total usually resident population of the zone.

New housing on former urban land (PNBU)

New housing on former urban land is the ratio of new housing development (units) on land in former urban use divided by total new housing development. Data come from the formerly Department of Environment, Transport and the Regions (DETR, now ODPM) and Ordnance Survey Land Use Change Statistics (LUCS).

Percentage of students at parental domicile (PARDOM_L)

Percentage of students at parental domicile is the estimated number of students at parental address divided by the number of residents, averaged over the two adjusted mid-year estimates.

Percentage of students at term time address (TERMT_L)

Percentage of students at term time address is the estimated number of students at term-time address divided by the number of residents, averaged over the two adjusted mid-year estimates.

Occupational migration index (OCCMIG)

Occupational migration index is a more complex variable that applies only to those over 16 years old. Its numerator is the predicted number of one-year migrants over 20 kilometres, based on 10 occupation group propensities, in the relevant sex/age group (16-19 to 65+); and its denominator is the number of residents in the relevant sex/age group (16-19 to 65+). Occupational migration index is a measurement of occupational structure of the origin and it is expected to have a positive affect on out-migration.

Employment growth (EMPGRO_L)

Employment growth is the number of full-time equivalent employees by workplace (from the Census/Survey of Employment) divided by full-time equivalent employees by workplace three years earlier. There are substantial sampling errors, especially after 1994. The values have been uplifted 2% for the years before 1993 for comparability of coverage. This variable does not cover self-employed.

Employment rate (EMPR_L)

Employment rate is the ratio of full-time equivalent employees by workplace (from Census/Survey of Employment) divided by working-age population by residence. The percentage full-time equivalent (FTE) employment was calculated using the sex disaggregated FTE number of jobs and the working age population of the zone. This can give a value of over 100% because jobs are based on the workplace whereas population is based on residence zone. There are also substantial sampling errors, especially after 1994. 2% uplift for the years before 1995 took place for comparability of coverage reasons.

Household income (HHINC_L)

Household income is the gross weekly household income divided by the number of households. It is a cross-sectional estimate of mean gross household income derived from proxy based model (see Bramley and Smart, 1996; Bramley, 1998). It is time series based on Office for National Statistics (ONS) Personal Disposable Income per head series (RDPI).

House prices (HPRICE_L)

House prices are the average house price for each FHSA. This is an important housing variable. It is available for more than 14 years, but the data come from different sources:

- For the period 1980 to 1991 it is the standardised average house price from the Nationwide Building Society database of transactions. For its calculation the four most discriminating/common house types are selected. The sum of house prices of each house type was divided by the number of transactions of each house type, weighted. Fixed weights for each house type were used.
- For the years 1992-1996 it is again the standardised average house price from the Nationwide Building Society database of transactions. Here the price of a post-war semidetached house of 900 square feet was used as the standard, rescaled to link to the series 1980 to 1991.
- For the year 1997 it is the standardised average house price from the Land Registry. Again the data has been rescaled to link to the series 1980 to 1996.
- For the year 1998 it is again the standardised average house price from the Nationwide Building Society database of transactions.

Percentage of net re-lets in social sector (PNRL_L)

Percentage of net re-lets in social sector is the rate of the net re-lets in social sector housing divided by the number of social rented sector dwellings. The sources are the Local Authorities (LA) and the Registered Social Landlord (RSL). It is a measure of social sector turnover.

Percentage of vacant dwellings in all sectors (PVAC_L)

Percentage of vacant dwellings in all sectors is the rate of the vacant housing of all tenures divided by the total number of dwellings of all tenures. Its source is the Housing Investment Programme (HIP1).

Regional variable of the total population (TPOPN_Y_L)

The Regional variable of the total population is calculated as an index that compares the total population in a zone with the total population of the surrounding zones weighted by a second power of distance. It is used to capture a pull effect produced when an origin is surrounded by very populous zones that draw migrants from the origin. The methodology of calculating a regional variable is as follows (Fotheringham et al., 2002b; Fotheringham et al., 2003).

Regional variables are meant to capture the possible pull effects on out-migration caused by conditions elsewhere in the country. In this study a new formula for the regional variables has been used:

$$Y_i = \frac{\sum_{j,j\neq i}(X_j / X_i) * d_{ij}^{\beta}}{\sum_{j,j\neq i} d_{ij}^{\beta}} \tag{3.4}$$

where X_i is the value of X at location i and X_j represents the value of X at one of the other FHSAs. The formula thus produces a distance-weighted average ratio of X_j to X_i where nearby locations are weighted more heavily in the calculation than more distant ones. The value of β was taken as –2 which, from experience, gives a reasonably differentiated surface of Y values. Values of β less negative than this give a surface which is smoother; values of β more negative will give a spikier surface.

Values of $Y_i > 1$ indicate that X_i is generally **smaller** than its neighbours

Values of $Y_i = 1$ indicate that X_i is generally very similar to its neighbours

Values of $Y_i < 1$ indicate that X_i is generally **larger** than its neighbours.

This formula has been applied to population (TPOPN); the results of which are mapped in Figure 3.6. The dark values indicate FHSAs where the Y variable is much greater than 1 – that is, where the population is generally much less than in neighbouring FHSAs. These areas tend to be in close proximity to major urban areas or else have relatively low rural populations. The very light values on the map indicate FHSAs where the Y variable is much less than 1 – that is, where the population is much greater than in surrounding FHSAs. These tend to be areas having large populations or else are close to FHSAs with very low populations.

In summary there are 16 out-migration determinants of which six are cross-sectional (air, climate, commut, crime, nonwh, and pnbu), five are time-specific (hhinc, hprice, pnrl, pvac, and tpopn_y), four are time- and sex-specific (empgro, empr, termt, and pardom), and only one is time-, age- and sex-specific (occmig). The main categories are included by the variables: tpopn_y for spatial structure; nonwh, pardom, and termt for demography; hhinc for economic; empgro, empr, and commut for employment; hprice, pnrl, and pvac for housing; occmig and crime for social; and air, climate, and pnbu for environmental.

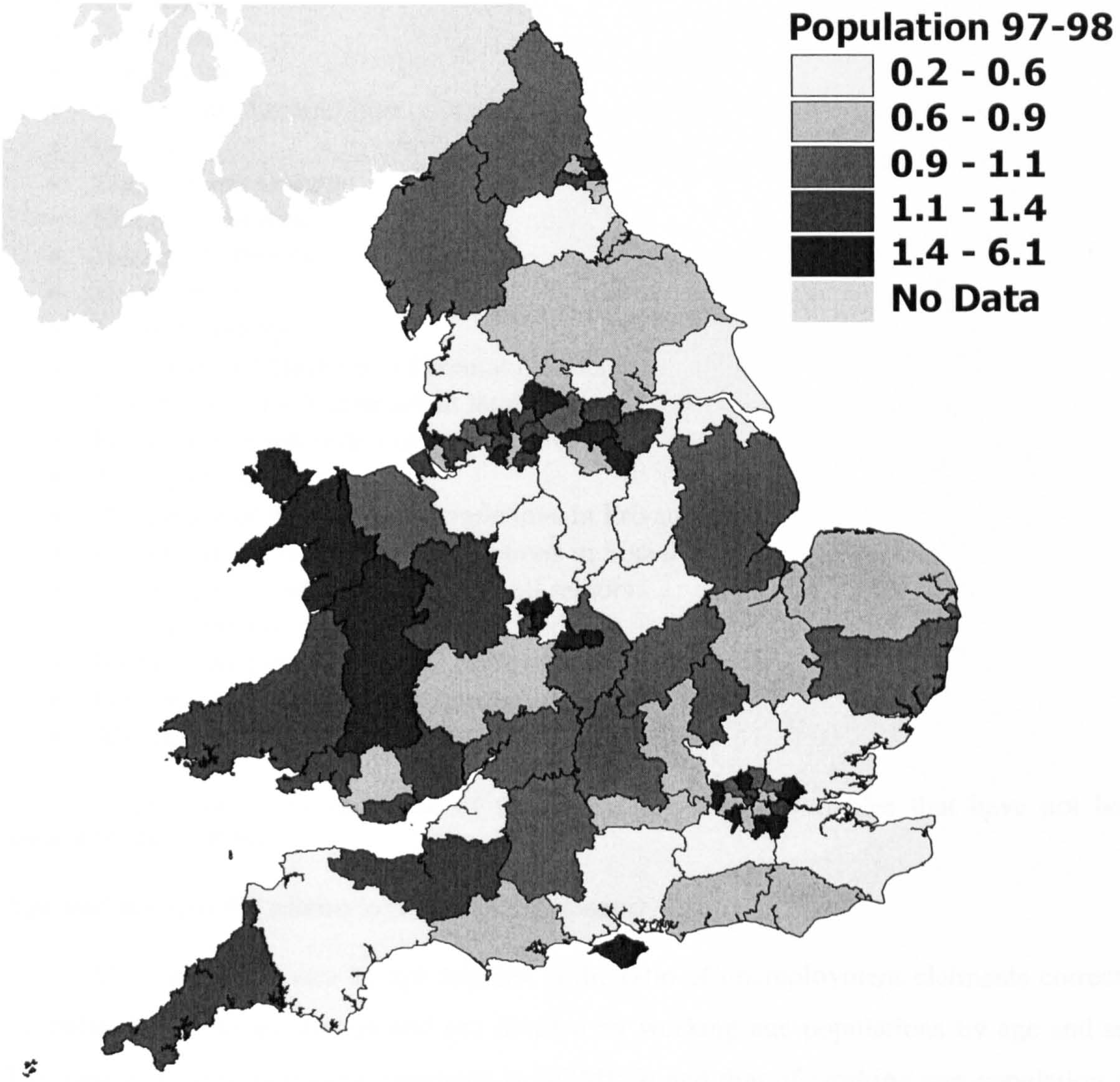


Figure 3.6. Regional variable for total population in mid year 1997-98.

3.2.2 Destination Choice Determinants

Continuing from the previous sections here a description of variables exclusively used in destination choice modelling is presented. Again the source of information is previous work (Champion et al., 1998; Fotheringham et al., 2002b). In the destination choice models 23 variables have been used. These are:

- Age and sex specific unemployment rate
- Climate Index
- Contiguity
- Crime Index
- Council tax
- Destination Accessibility
- Distance
- Employment Growth
- Employment Rate
- Household Income
- House Prices
- Listed Buildings
- Percentage of Students at Parental Domicile
- New housing on former urban land
- Percentage of net re-lets in social sector
- Total population
- Percentage of New Build Completions in Private Sector
- Percentage of New Build Completions in Social Sector
- Percentage of vacant dwellings in all sectors
- Index of the private stock in poor condition
- Index of the Local Authority stock in poor condition
- Percentage of students at term time address
- All vacant and derelict dwellings

A description for each one of the destination choice variables that have not been presented above follows.

Age and sex specific unemployment rate (asunem)

Unemployment rate by age and sex is the ratio of unemployment claimants corrected for definitional change by age and sex divided by working age populations by age and sex. The source of unemployment claimants is NOMIS® and that of working age population the mid-year ONS population estimates.

Contiguity (contig)

Contiguity is a dummy variable that equals one for pairs of zones (FHSAs) sharing a boundary and zero in all other cases. It is used to explain part of the high volumes of short distance migration flows, some of which is just crossing the zone boundary. Boyle and

Flowerdew (1997) provide empirical evidence for the improvement the introduction of a contiguity dummy to a migration flows model does.

Council tax (ctax)

Council tax is the average Council Tax rate for Band D household. Information was extracted from the formerly DETR (now ODPM) Local Government Website.

Destination Accessibility (destacc)

Destination Accessibility is an important variable in the destination choice model. When this variable is introduced in the spatial interaction model it becomes the Competing Destinations Model (Fotheringham, 1991). The destination accessibility (or destination centrality) is calculated as follows:

$$A_j = \sum_{m \neq j} W_m / d_{jm} \tag{3.5}$$

where A_j is the potential accessibility of destination j to all other potential destinations m, W_m is a weight generally measured by population, and d_{jm} is the distance between j and m (Fotheringham, 1991, p. 67). Here W_m is destination populations and d_{jm} is the network distance between j and m. More details on the network distance follow.

Distance (dist)

The full name for this variable is network weighted distance. It was calculated in order to provide a more realistic measure of the separation between the 100 zones (98 FHSAs in England and Wales, Scotland and Northern Ireland) of the migration model than straight-line distances. The distance was calculated based on the topology of districts in the UK. A detailed procedure of is calculation is presented in Appendix 2 within Fotheringham et al. (2002b, pp. 163–164).

Listed Buildings (listed)

Listed buildings variable is a ratio of the number of listed buildings in 1999/2000 (provided by national built heritage organisations) divided by dwellings in 1998 (from the national Council Tax registers). A problem affecting the quality of this variable is that decisions on the listing of buildings are unlikely to be consistent between (or even perhaps within) countries.

Total population (popn)

Total population is the number of residents in an FHSA. It is an average of mid-year estimates. The source of the data is the ONS (Office for National Statistics) mid-year population estimates

Percentage of New Build Completions in Private Sector (pqpr)

Percentage of New Build Completions in Private Sector is the ration of the Private New Build Completions (from Local Housing Statistics returns) divided by the number of private sector dwellings (from HIP1)

Percentage of New Build Completions in Social Sector (pqsr)

Percentage of New Build Completions in Social Sector is the ratio of the Social New Build Completions divided by number of social rented sector dwellings. Information comes from several sources: Local Authorities (LA), Registered Social Landlord (RSL), Local Housing Statistics (LHS) returns and Housing Investment Programme (HIP1).

Index of the private stock in poor condition (rlapsc)

Index of the private stock in poor condition is simply the private sector housing stock in poor condition. Data comes from the Housing Needs Index (HNI). The index used in HNI system is based on the English Housing Condition Survey (EHCS) and proxies. Data are available for England only. Data for Wales have been estimated and are equal across FHSAs in Wales.

Index of the Local Authority stock in poor condition (rlasc)

Similarly, the Index of the Local Authority stock in poor condition is just the LA housing stock in poor condition from the Generalised Needs Index (GNI). The index used in GNI system by DETR (now ODPM) is based on stock profile and the EHCS. Data are available for England only. Data for Wales have been estimated and are equal across FHSAs in Wales.

All vacant and derelict dwellings (vacdrl)

All vacant and derelict dwellings variable is the ratio of vacant and derelict land and buildings area divided by the total dwellings in 1998. The source of the former data is the

National Land Use Database (NLUD) and for the latter the HIP1. This data covers England only. Missing values imputed with average values.

3.3 Summary

In this Chapter, I discussed some issues concerning migration data in general and I presented the dataset I used here consisting of migration data and determinants. First, I explained the origin of my dataset (ODPM funded project) and I described the geography, time scale and disaggregation level of the migration data (NHSCR). I then discussed their quality; I presented alternative data sources and I explained why the NHSCR data allow me to address my research questions. Finally, I presented the variables I included in my analysis. I provided information on the measure and construction of each variable. In the case of indexes and regional variables I provided the methodology used in their calculation and the range of the values they take.

In the previous Chapter, I demonstrated the relevance of each migration determinant in explaining migration decisions, whereas here I described its nature and quality, in order to identify potential weaknesses in the interpretation of my empirical findings. A good understanding of what each variable is meant to measure results in a more transparent interpretation of empirical findings for its effect on out-migration. I now discuss some modelling issues and I provide the exact equations of my models.

Chapter 4
Methodological Issues: Modelling Migration

A detailed discussion of methodology issues is presented in this chapter. It is mainly concerned with the modelling issues of out-migration and destination-choice. Minor methodological issues are not discussed here, but in the appropriate chapter. For example, in Chapter 3, Principal Component Analysis and the Regional Variables construction are discussed as they have been used during the data preparation. In Chapter 5, the clustering algorithm k-means is discussed.

The first section of this chapter discusses regression techniques for global and local models. This discussion includes different calibration methods, statistical inference, goodness-of-fit statistics and model selection. The remainder of this chapter contains two other sections, one for out-migration modelling and one for destination choice modelling.

4.1 Modelling issues

In this section an overview of the main multivariate regression techniques is presented. These regression techniques refer to modelling at both stages of migration (out-migration and destination choice). Specific migration models are presented in the following sections referring to each of the analysis chapters.

4.1.1 Introduction, the model selection exercise

Most of the literature on migration modelling is focused on destination choice models (Lowry, 1966; Congdon, 1988; Fik and Mulligan, 1990; Pellegrini and Fotheringham, 1999; Boyle and Flowerdew, 1993, 1997) where the gravity model has typically been applied. Here, migration is modelled in two stages: Stage 1 concerns the modelling of out-migration rates with a set of push factors. Stage 2 concerns the modelling of destination choice (migration counts) with a set of pull factors. A power function model has been used to model out-migration rates and the competing destinations choice model to model destination choice.

Both these models are non-linear models and thus are expected to give better results than simple linear models. Fotheringham and O'Kelly (1989) show that a power function of a variable (they refer to distance) results in parameter estimates independent of the scale the analysis is conducted. A power form of equation can be easily become linear by logging both sides of the equation. However, a data issue arises here: variables having non-positive values cannot be logged. To overcome this problem, exponential instead of power functions of such variables used. The corresponding equations follow (Sections 4.2.1 and 4.3.1).

4.1.2 Calibration Techniques (OLS, WLS, ML, Poisson)

A migration model (either out-migration or destination choice) can be calibrated using different methods. Amongst others, the most important are: Ordinary Least Square (OLS), Weighted Least Square (WLS), Maximum Likelihood (ML) and Poisson Regression.

The simplest linear model is a line:

$$y = b_0 + b_1 x \tag{4.1a}$$

where x is the independent variable, y is the dependent variable, b_0 is the constant and b_1 is the parameter showing the relationship between x and y. Assuming that there are n observations of x, y, a line can be fitted and the parameters b_0 and b_1 can be estimated. The fitted line will cross the Y-axis at a point b_0 (the intercept) and will have a slope, b_1 (Davis, 2002, Chapters 4 and 6). If all y can be exactly estimated then the model is called deterministic. However, in a real world problem it is impossible to achieve this. Thus, it is necessary to add a random variable (e) in the right hand side of equation 4.1a, which is the so-called *error term*. It is a priori assumed that the mean value of this random variable is 0. Equation 4.1a becomes:

$$y_i = b_0 + b_1 x_i + e_i \tag{4.1b}$$

The new model is called *stochastic* or *probabilistic*, since it allows for a non-perfect fit that has error attached to the predicted values. The equation of the estimated line will be

$$\hat{y}_i = b_0 + b_1 x_i \tag{4.2}$$

where $\hat{y}_i$ is the estimated value of y at specific values of x for is each (i) of the n observations.

One method of estimating b_0 and b_1 is Ordinary Least Squares (OLS) Regression according to which the sum of the squares of the differences between the observed and the estimated values of the dependent variable (y) is minimised (Equation 4.3).

$$\sum_{i=1}^{n} (\hat{y}_i - y_i)^2 = \text{minimum} \tag{4.3}$$

The result of the model fit is two estimated values one for the intercept (b_0) and one for the parameter of the independent variable (b_1); in this case these are

$$b_1 = \frac{\sum_{i=1}^{n} x_i y_i - \left(\sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i\right) \Big/ n}{\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2 \Big/ n} \quad \text{and} \quad b_0 = \frac{\sum_{i=1}^{n} y_i}{n} - b_1 \frac{\sum_{i=1}^{n} x_i}{n} \tag{4.4}$$

If many independent variables are going to be included in the model, then the multivariate linear model can be constructed as follows:

$$y = b_0 + \sum_{k=1}^{m} a_k x_k \tag{4.5}$$

where m is the number of independent variables. Similarly Equation 4.2 becomes Equation 4.6 and again the total sum of squares (Equation 4.3) has to be minimised.

$$\hat{y}_i = b_0 + \sum_{k=1}^{m} b_k x_{ki} \tag{4.6}$$

It is necessary here to discuss the statistical inference for linear models. *Standard statistical inference for least-squares simple regression analysis is based on the statistical model* $y_i = b_0 + b_1 x_i + e_i$. *The key assumptions of the model concern the behavior of the errors* e_i*: (1) Linearity,* $E(e_i) = 0$*; (2) constant variance,* $V(e_i) = \sigma_e^2$*; (3) normality,* $e_i \sim N(0, \sigma_e^2)$*; (4) independence,* e_i*,* e_j *are independent for* $i \neq j$*; and (5) the x-values are fixed, or if random, are independent of the errors.* (Fox, 1997, p. 114). Equivalently to the above assumptions, the dependent variable must be a linear function of the independent variable, it also needs to have a constant variance and a normal distribution. The above assumptions are the same for the multivariate linear regression model.

Under the assumptions of the regression model, the least-squares coefficients have certain desirable properties as estimators of the population regression coefficients. The least-squares coefficients are: linear functions of the data and therefore have simple sampling distributions; unbiased estimators of the population regression coefficients; the most efficient unbiased estimators of the population regression coefficients; maximum-likelihood estimators; and normally distributed. (Fox, 1997, p. 116).

In practice, the error variance σ_e^2 is never known, thus it is necessary to construct confidence intervals and to test hypotheses of whether the estimators b_0 and b_1 are statistically significant. There are several statistical techniques to check the statistical significance of the parameter estimates (the t-test in the case of OLS), and the accuracy of the model's fit also called *goodness-of-fit statistics* (for example the r-squared). These are discussed below.

Weighted Least Squares (WLS) is a modification of OLS where the dependent variable is multiplied with another variable called *weight*. Equations 4.2 and 4.3 can be rewritten as follows

$$w_i\hat{y}_i = w_i(b_0 + b_1 x_i) \qquad (4.7)$$

$$\sum_{i=1}^{n} w_i(\hat{y}_i - y_i)^2 = \text{minimum} \qquad (4.8)$$

Maximum Likelihood (ML) is a more generic estimation procedure than OLS. It allows the construction of a likelihood function the maximisation of which is the way to establish the estimators. *The ML method produces estimators whose properties are optimal for large samples (under certain conditions of mathematical regularity) when the assumed likelihood function is correct. ML estimators are said to be asymptotically optimal in the sense that desirable properties such as unbiasedness, minimum variance, and normality hold exactly in the limit only as the amount of data becomes infinitely large. In practice, this means that it is reasonable to assume for large datasets that an ML estimator will be essentially unbiased, have a small variance, and be approximately normally distributed when the appropriate maximum likelihood function is being used* (Kleinbaum et al., 1988, p. 489).

In order to fit the model of Equation 4.1b by ML, first a maximum likelihood function must be specified. Under the assumptions that y_i (dependent variable) is normally distributed with a mean $\mu_i = E(y_i)$ and with variance $\mathrm{Var}(y_i) = \sigma^2$, x_i is non-stochastic (x_i is measured without error) and $\{y_i\}_{i=1}^{n}$ is mutually independent, the distribution of y_i is

$$f_{Y_i}(y_i;b_0,b_1,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}[y_i-(b_0+b_1x_i)]^2} \qquad (4.9)$$

and the likelihood function is

$$L(\mathrm{y};b_0,b_1,\sigma^2) = \prod_{i=1}^{n} f_{Y_i}(y_i;b_0,b_1,\sigma^2) = f_{Y_i}(y_i;b_0,b_1,\sigma^2) = \frac{1}{(2\pi\sigma^2)^{n/2}} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}[y_i-(b_0+b_1x_i)]^2} \qquad (4.10)$$

where $-\infty < y_i < +\infty$, $i = 1,2, \ldots, n$. The ML estimators of b_0, b_1 and σ^2 will be those values of b_0, b_1 and σ^2, denoted $\hat{b}_0$, $\hat{b}_1$ and $\hat{\sigma}^2$ respectively, for which $L(\mathrm{y};b_0,b_1,\sigma^2)$ attains its *maximum value* as a function of b_0, b_1 and σ^2. Using calculus, the specific values $\hat{b}_0$, $\hat{b}_1$ and $\hat{\sigma}^2$ of b_0, b_1 and σ^2, respectively, that maximise the function L, can be found by setting the derivatives of Equation 4.10 with respect to b_0, b_1 and σ^2 equal to 0 and then solving the resulting (ML) equations for $\hat{b}_0$, $\hat{b}_1$ and $\hat{\sigma}^2$, respectively. By solving simultaneously the three ML equations

$$\frac{\partial}{\partial b_0}[\ln L(\mathrm{y};b_0,b_1,\sigma^2)] = 0, \quad \frac{\partial}{\partial b_1}[\ln L(\mathrm{y};b_0,b_1,\sigma^2)] = 0, \text{ and } \quad \frac{\partial}{\partial \sigma^2}[\ln L(\mathrm{y};b_0,b_1,\sigma^2)] = 0$$

one can show that the ML estimators $\hat{b}_0$, $\hat{b}_1$ and $\hat{\sigma}^2$, are, respectively,

$$\hat{b}_0 = \bar{y} - \hat{b}_1 \bar{x}, \qquad \hat{b}_1 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}, \tag{4.11}$$

and

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} [y_i - (\hat{b}_0 + \hat{b}_1 x_i)]^2 = \frac{SSE}{n} \tag{4.12}$$

where *SSE* is the sum of squares of residuals about the fitted straight line.

The only real conceptual difference between Poisson regression and standard multiple regression is the former involves a Poisson distribution and the latter the normal distribution. The Poisson probability distribution with parameter μ is given by the formula

$$p_Y(y;\mu) = pr(Y = y;\mu) = \frac{\mu^y e^{-\mu}}{y!} \qquad y = 0,1,\ldots,\infty \tag{4.13}$$

It can be theoretically shown that $E(Y)$ = Var (Y) = μ. The Poisson regression analysis is a ML-based procedure that has a more complex likelihood function. A general form of the likelihood function is

$$\begin{aligned} L(y;\beta) &= \prod_{i=1}^{n} p_{Y_i}(y_i;\beta) \\ &= \prod_{i=1}^{n} \left\{ \frac{[\ell_i \lambda(x_i,\beta)]^{y_i} e^{-\ell_i \lambda(x_i,\beta)}}{y_i!} \right\} \\ &= \frac{\left\{ \prod_{i=1}^{n} [\ell_i \lambda(x_i,\beta)]^{y_i} \right\} \exp\left[-\sum_{i=1}^{n} \ell_i \lambda(x_i,\beta) \right]}{\prod_{i=1}^{n} y_i!} \end{aligned} \tag{4.14}$$

where $E(Y_i) = \mu_i = \ell_i \lambda(x_i,\beta), i = 1,2,\ldots,n$ (Y is the dependent-random variable).

In practice, a particular form of the function for the rate function $\lambda(x_i,\beta)$ needs to be specified. An example of λ is

$$\lambda(x_i,\beta) = \exp(\beta_0 + \sum_{j=1}^{k} \beta_j x_{ij}), \text{where } \beta_0 + \sum_{j=1}^{k} \beta_j x_{ij} > 0 \tag{4.15}$$

(Kleinbaum et al., 1988). In the migration literature, a simple Poisson regression model can be written also as

$$y_i = \exp(\sum_{j=1}^{k} b_j x_{ij}) + b_0 \tag{4.16}$$

assuming that the random variable y_i has a Poisson distribution (Boyle and Flowerdew, 1997).

The ML estimators $\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_k$ of $\beta_0, \beta_1, \ldots, \beta_k$ are obtained from (4.14) as the solution of the $k+1$ equations

$$\frac{\partial}{\partial \beta_i}[\ln L(y;\beta)] = 0 \qquad j = 0,1,\ldots,k \tag{4.17}$$

A discussion of the appropriateness of each model for modelling migration follows.

4.1.3 Least Squares vs. Maximum-Likelihood

In the previous section a discussion of regression techniques included only those consistently used in migration literature. The maximum-likelihood and Poisson approaches are more appropriate when migration data are modelled as flows, whereas OLS is more appropriate for modelling migration rates. This is because, for example, a Poisson regression is better if the dependent variable is a count (Fox, 1997) of independent individuals (here migrants who take their decision to migrate independently from each other) and especially if this count is relatively small compared to the population size (Kleinbaum et al., 1988).

Flowerdew and Aitkin (1982) compare two methods of fitting a gravity model; the Poisson and the OLS (log-normal model). They provide empirical evidence for the superiority of Poisson regression in modelling migration flows. This is because Poisson regression addresses four specific problems of the log-linear model: *the bias in estimated flows introduced by fitting the model in logarithmic form, the failure of the assumption that the error terms are normally distributed, unequal variance in the error terms and the sensitivity of model results to the treatment of zero flows* (Flowerdew and Aitkin, 1982, p. 201). Flowerdew (1982) suggests an iterative weighted method to overcome the problem of heteroscedasticity (when the common variance assumption in a set of random variables is not true) in log-linear models. Although this approach is an improvement of the log-linear gravity model, Poisson is still a superior method for modelling migration flows (Poisson also overcomes the problem of heteroscedasticity). Fotheringham and Williams (1983) identified some problems in Flowerdew's comparisons between the log-normal and the Poisson approach of modelling migration flows, and provide further investigation and discussion. They also conclude that the Poisson model is more appropriate for modelling migration flows (especially for matrices of interaction data with low counts and many 0s) than any improved log-linear model.

Poisson regression, however, assumes that the movements of individuals are independent, and they have a Poisson distribution. These are approximations that are not necessarily true with particular datasets such as migration (Fotheringham and Williams, 1983). Flowerdew and Aitkin (1982), Flowerdew and Lovett (1988, 1989) conclude that although Poisson works better than OLS, the fit of Poisson models is not always satisfactory. They believe that one of the reasons might be the violation of the assumption of individuality of migrants. This may be because people in practice migrate as household. To overcome this problem Flowerdew and Lovett (1989) introduce compound and generalised Poisson models. Another solution is the household-size model (Flowerdew and Boyle, 1995) that accounts for different household sizes. However, the results of the latter need careful interpretation.

Empirical studies of migration using Poisson models showed the problem of overdispersion: *given a standard exponential family generalised linear model (GLM) with a specific variance/mean relationship, we observe on fitting the model that the variance is greater than that predicted by the mean, observable in a large residual deviance or Pearson* X^2, *with some large individual standardized (Pearson) residuals. Other sources of variations are present in the data which have not been included in the regression model* (Aitkin, 1996, p. 251). This occurs when migration flows are regressed only over measures of populations and distance (Flowerdew and Aitkin, 1982; Flowerdew and Lovett, 1988). Flowerdew and Boyle (1995) try to provide some explanation for this. Aitkin (1996) presents a more general discussion on overdispersion in generalised linear models. In contrast, the problem of underdispersion in Poisson regression models also exists. Boyle and Flowerdew (1993) and Flowerdew and Boyle (1995) identify it in their analysis of inter-ward migration within the counties of Hereford and Worcester, and they suggest a solution through simulation.

In their empirical work, Flowerdew and Lovett (1988, 1989) illustrate that the simple Poisson regression does not satisfactorily fit with inter-urban migration data from 1971 Census of Population. As well as fitting a generalised Poisson model to account for the household distribution, the introduction of additional explanatory variables does improve goodness of fit (Flowerdew and Lovett, 1988).

In recent years, households in the UK have become smaller; thus, the decision to migrate is increasingly a decision made by an individual or a couple. In the case of a family with children, the decision of the parents will affect the migration of their children; the Poisson assumption is violated in such an instance. In the sex/age disaggregated data used here, for most of the sex/age groups (excluding children 0 – 15) the Poisson assumption of individuality is not violated. This is because the members of a single family will belong to different migrant groups; the parents will be modelled separately (because of sex disaggregation independently of their age) as well as their children (those aged 0 – 15 and 16 – 19) of different sex and/or age. The same age/sex children will belong in the same group, but they will not be the decision makers in a moving family anyway. This overcomes one the problems discussed above and removes the need for a generalised Poisson model. The availability of many explanatory variables also is expected to result in better model fits than those reported in earlier studies.

4.1.4 Local Modelling: Geographically Weighted Regression

Geographically Weighted Regression (GWR) is a recent technique (first paper published in 1996 by Fotheringham, Charlton and Brunsdon); which allows the examination of local variations in spatial processes. There is a growing number of applications of GWR, many of which are in press and will appear soon in the literature. This is not only because of the recognition of GWR as a powerful tool to identify spatially varying relationships in spatial data, but also because of its support. The latter consists of the provision and user support of software (GWR 2.2) that allows the calibration of models using the GWR theoretical framework, as well as detailed documentation. There is also support in the form of a book about GWR (Fotheringham et al., 2002a) that provides the necessary theoretical underpinnings and the mechanics of the method. It also provides a user guide to the software (GWR 2.0), which is available through the authors of the book. The user interface of GWR 2.0 is as friendly as possible since its developer extensively uses it for his own research presented here.

The software for GWR is evolving and new versions will appear before the completion of this thesis, however, GWR 2.0 has been widely deployed and publications are expected to have based their analysis on it. Therefore, the current version of GWR 2.0 was used to calibrate the models of Chapter 6. Below, the general statistics of GWR are presented, and the exact specifications used here are explained.

Equation 4.18 is stochastic version of (4.5); it is a global regression model (traditional in the literature), and its algebraic solution is (4.19)

$$y_i = b_0 + \sum_{k=1}^{m} a_k x_{ik} + \varepsilon_i \qquad (4.18)$$

$$\mathbf{b} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y} \qquad (4.19)$$

where **b** represents the vector of global parameters to be estimated (estimates of a_k), X is a matrix of independent variables with the element of the first column set to 1, and y represents a vector of observations on the dependent variable. GWR is a technique that allows the calibration of a local model (it allows for local rather than global parameters to be estimated). This is possible by calibrating a model around a point i in space including all or some of n

observations in the dataset weighted by a weighting scheme (usually a distance function). The local model and its solution (estimator) follow.

$$y_i = b_0(u_i, v_i) + \sum_{k=1}^{m} a_k(u_i, v_i)x_{ik} + \varepsilon_i \tag{4.20}$$

$$\mathbf{b}(u_i, v_i) = (\mathbf{X}^T\mathbf{W}(u_i, v_i)\mathbf{X})^{-1}\mathbf{X}^T\,\mathbf{W}(u_i, v_i)\mathbf{y} \tag{4.21}$$

where $\mathbf{W}(u_i, v_i)$ is an n by n matrix whose off-diagonal elements are zero and whose diagonal elements denote the geographical weighting of observed data *for point i*. That is,

$$\mathbf{W}(u_i, v_i) = \begin{pmatrix} w_{i1} & 0 & 0 & \cdots & 0 \\ 0 & w_{i2} & 0 & \cdots & 0 \\ 0 & 0 & w_{i3} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & w_{in} \end{pmatrix} \tag{4.22}$$

where w_{in} denotes the weight of the data in point n on the calibration of the model around point i. These weights depend on the location of i which is not the case in WLS (Fotheringham et al., 2000).

$\mathbf{W}(u_i, v_i)$ is a weighting scheme based on the proximity of the regression point i to the data points around i (Fotheringham et al., 2002a). *Data from observations close to i are weighted more than data from observations farther away. This is shown in Figure 4.2 where a* <u>*spatial kernel*</u> *is placed over each calibration point and the data around that point are weighted according to the distance-decay curve displayed by the kernel* (Fotheringham et al., 2000, p. 108).

The method of fitting a spatial kernel to the data can be graphically demonstrated (Figure 4.1). Numerous weighting schemes (kernels) can be used in GWR depending on the distance function that defines each of them. There are two main categories of kernels: fixed or adaptive. A significant component of a kernel is its *bandwidth*: this determines the radius around point i that defines the area in space (around point i) that observations will be weighted and included in the regression. In the case of a fixed kernel it is constant across the area of study, whereas in the case of the adaptive kernel it is variable. In the case of the adaptive kernel the *number of nearest neighbours* needs to be defined. This is the number of observations around point i that need to be included in the regression.

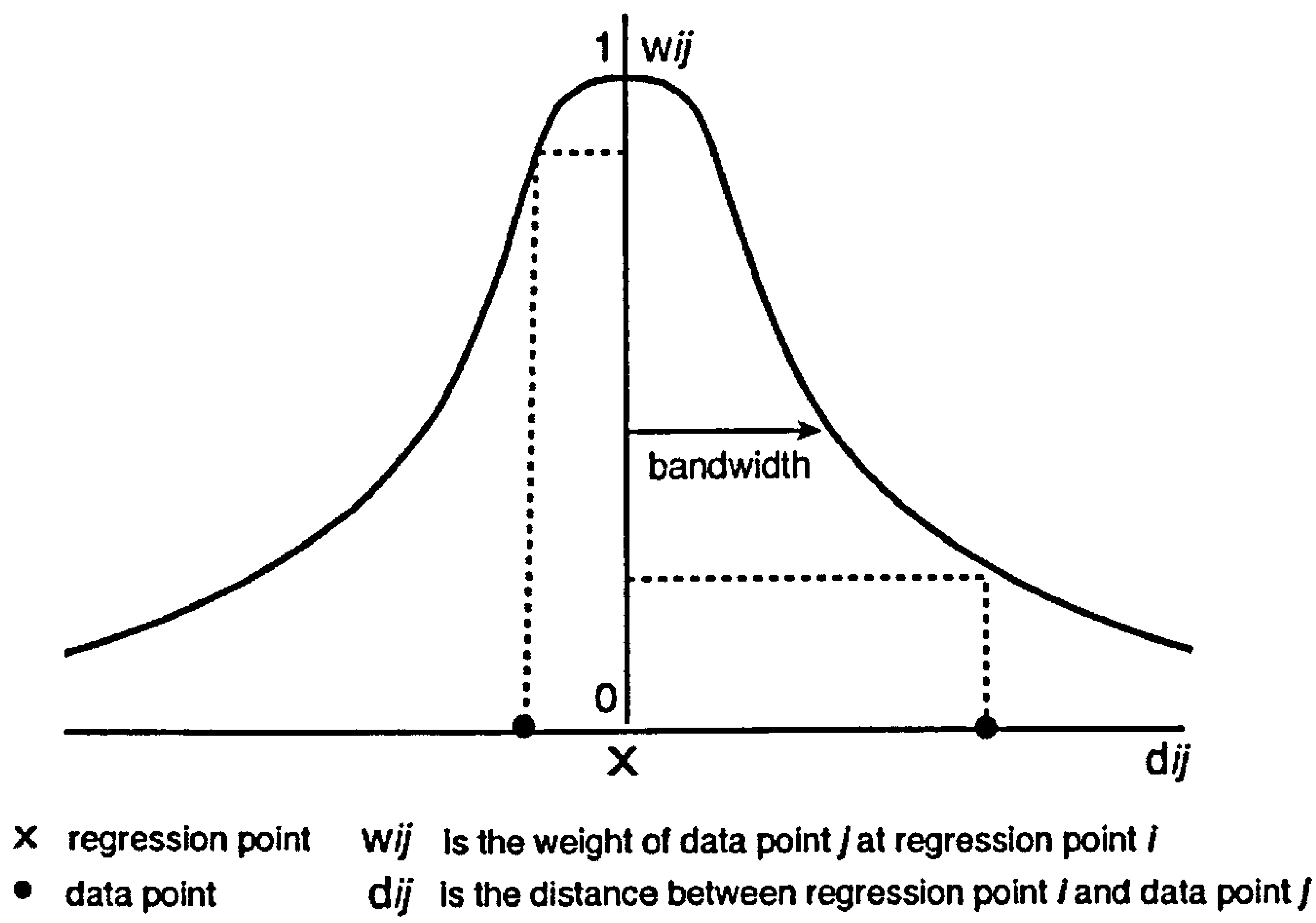


Figure 4.1. A spatial kernel
Source: Fotheringham et al. (2002a), p. 44

Below examples of kernels are discussed. The weighting function is each case is w_{ij} where j is a specific point is space at which data are observed and i is any point in space for which parameters are estimated. A global model (OLS) is defined when w_{ij}=1 for all i and j. The simplest local model follows

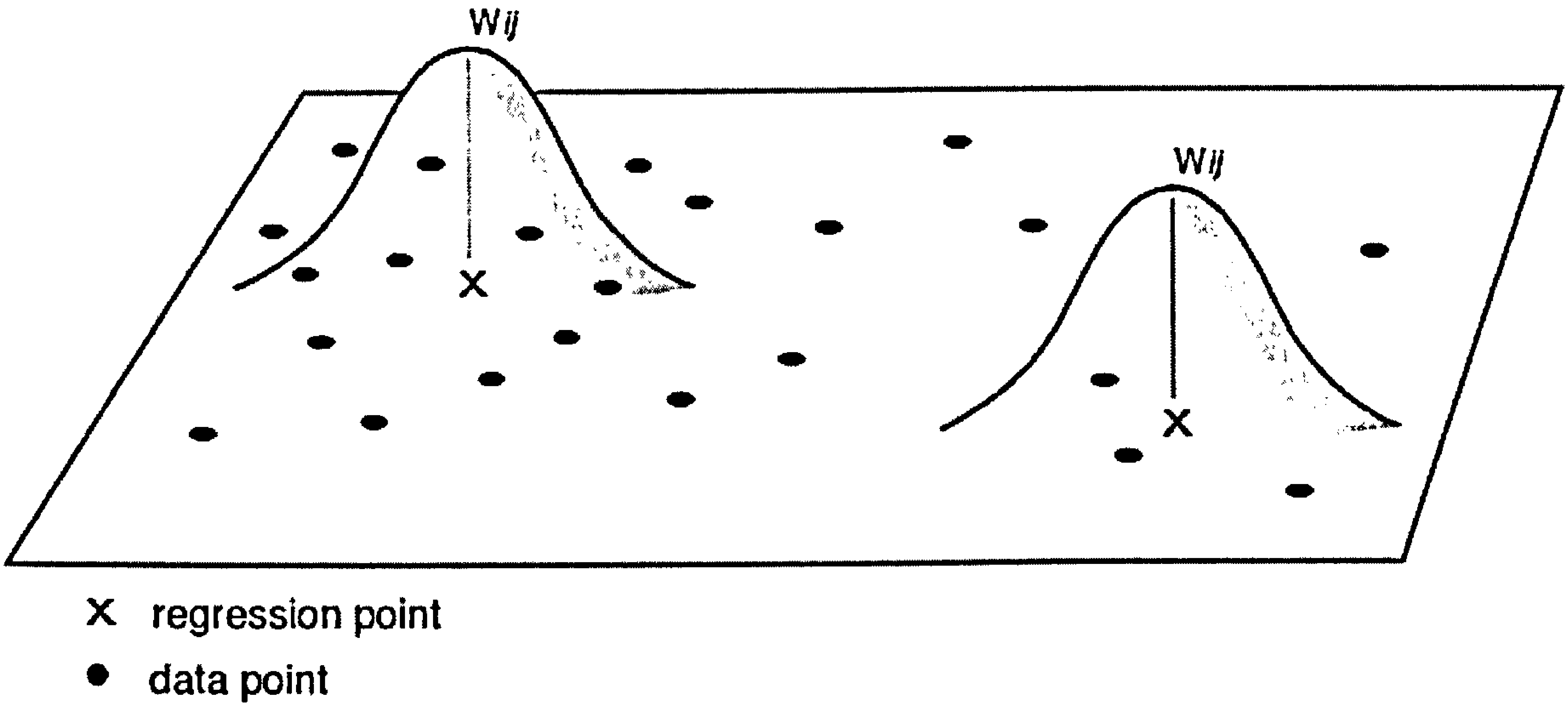
$$w_{ij} = \begin{cases} 1 \text{ if } d_{ij} < d \\ 0 \text{ otherwise} \end{cases} \tag{4.23}$$

where d is a fixed distance (bandwidth) that defines the inclusion area for observations in the model calibration. Two examples of a fixed weighting scheme are the Gaussian function (Equation 4.24) and the bi-square function (Equation 4.25)

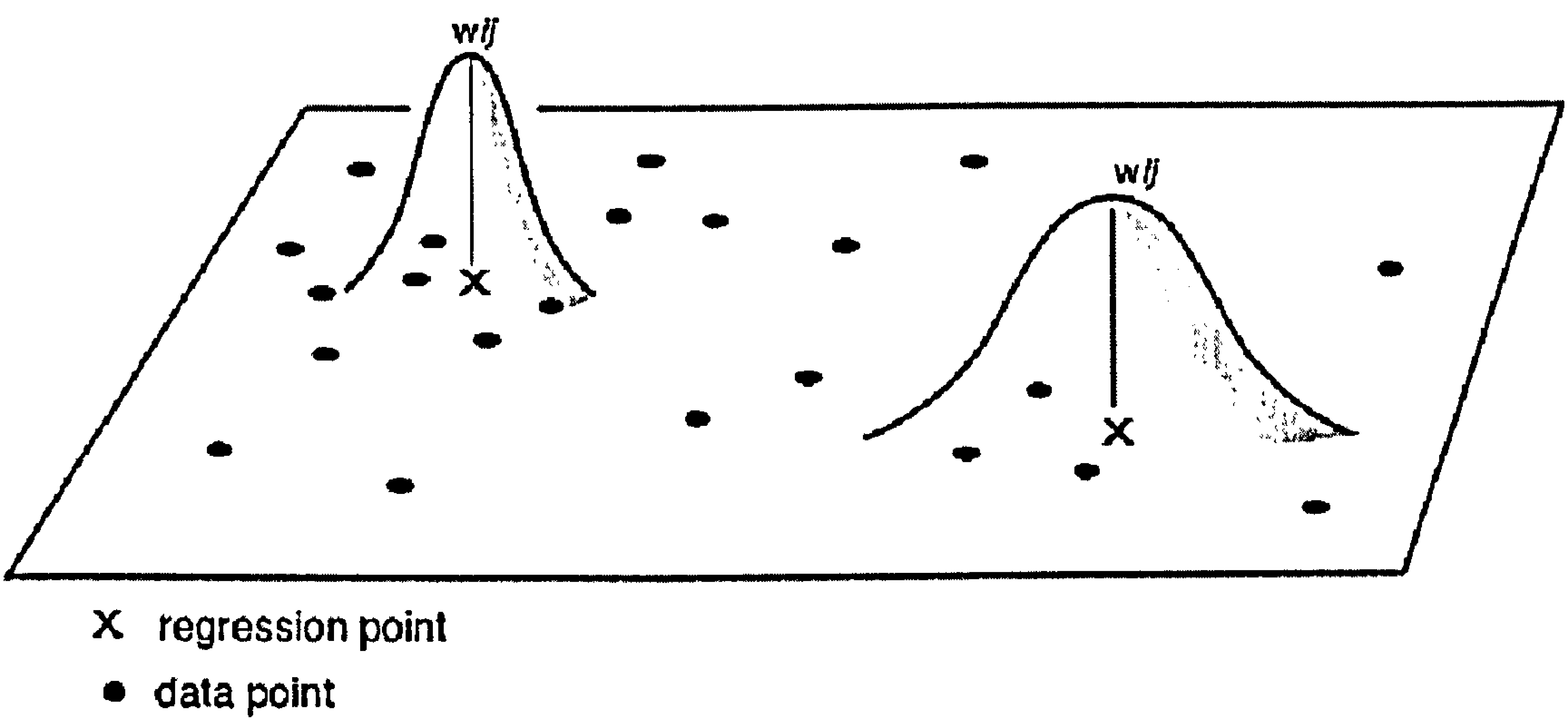
$$w_{ij} = \begin{cases} e^{-\frac{1}{2}(d_{ij}/h)^2} \text{ if } d_{ij} < h \\ 0 \text{ otherwise} \end{cases} \tag{4.24}$$

$$w_{ij} = \begin{cases} [1-(d_{ij}/h)^2]^2 \text{ if } d_{ij} < h \\ 0 \text{ otherwise} \end{cases} \tag{4.25}$$

where h is the bandwidth. GWR 2.0 uses the Gaussian function as the weighting scheme of the fixed kernel.



a. Fixed kernel, fixed distance around point **x**



b. Adaptive kernel, varying distance around point **x**, set number of neighbours

Figure 4.2. Examples of fixed and adaptive kernels in GWR
Source: Fotheringham et al. (2002a), pp. 45 - 47

Fotheringham et al. (2000; 2002a) discussed three weighting functions for an adaptive kernel: the first is the bi-square function basest on nearest neighbours (Equation 4.26); the second is based on ranked distances (Equation 4.27); and the third constrains the sum of squares for any calibration point to be constant (Equation 4.28).

$$w_{ij} = \begin{cases} [1-(d_{ij}/h_i)^2]^2 \text{ if } d_{ij} \le h_i \\ 0 \text{ otherwise} \end{cases} \tag{4.26}$$

where h_i is the Nth nearest neighbour distance from i.

$$w_{ij} = e^{-R_{ij}/h} \tag{4.27}$$

where R_{ij} is the rank of the distance data point j is from calibration point i.

$$\sum_j w_{ij} = C \quad \text{for all } i \tag{4.28}$$

GWR 2.0 uses the bi-squared function as the weighting scheme of the adaptive kernel. This kernel was used to calibrate the local out-migration models because the spatial distribution of the observed data used here is not homogeneous. More discussion on the latter follows in Chapter 6.

The choice of the bandwidth has a large impact on the results obtained from GWR. It is possible to think of the bandwidth as a <u>smoothing</u> parameter, with larger bandwidths causing greater smoothing. An oversmoothed model will produce parameters that are similar in value across the study area and an undersmoothed model will produce parameters with so much local variation that is difficult to determine whether there are any patterns at all. The 'best' bandwidth is that which provides a happy medium between these two extremes. GWR 2.0 allows the user to choose on of three methods of bandwidth selection:

1. *providing a user supplied bandwidth;*
2. *selecting the bandwidth that minimises a cross-validation function;*
3. *selecting the bandwidth that minimises the Akaike Information Criterion (AIC)*

(Fotheringham et al., 2002a, p. 211).

Obviously, specifying the bandwidth mainly refers to the use of a fixed kernel. In the case of the adaptive kernel one should specify the number of nearest neighbours, currently not supported by GWR 2.0. Another option is the inclusion of all or a subset of data in the calculation of the optimum bandwidth. This is more appropriate for large datasets in order to reduce the time of estimating the bandwidth. However, here there are no reasons to pre-define a specific bandwidth nor to use a subset of the observations.

GWR 2.0 supports two methods for selecting the most appropriate bandwidth when it is unknown to the analyst. As mentioned above these methods are: the Cross-Validation (CV) score minimisation and the Akaike Information Criterion (AIC) minimisation. Cross-Validation is a technique (suggested for local regression by Cleveland, 1979) in which the optimal bandwidth is that which minimises the following score:

$$CV = \sum_{i=1}^{n} (y_i - \hat{y}_{\neq i})^2 \tag{4.29}$$

where n is the number of data points and $\hat{y}_{\neq i}$ is the fitted value of y_i with the observations for data point i omitted from the calibration process (Fotheringham et al., 2000).

Adjusting the bandwidth changes the degrees of freedom in the model. ... The AIC takes into account the different number of degrees of freedom in different models so that their relative performances can be compared more accurately. A model with a lower AIC than another (the rule of thumb is difference of 3 and more) is held to be a better model. The AIC used in GWR is computed as:

$$AIC_C = 2n\ln(\hat{\sigma}) + n\ln(2\pi) + n\left\{\frac{n+\mathrm{tr(S)}}{n-2-\mathrm{tr(S)}}\right\} \tag{4.30}$$

where n is the number of observations, $\hat{\sigma}^2$ *= (Residual Sum of Squares) / (n – k); k is the number of parameters in the model and* $\hat{\sigma}$ *is the estimated standard deviation of the error term, and tr(S) is the trace of the hat matrix S (which is an n by n matrix that transforms the raw dependent values* y_i *to the fitted values* $\hat{y}_i$ *in the following manner:* $\hat{Y} = SY$*) which is a function of the bandwidth* (Fotheringham et al., 2002a, pp. 55, 61, 91–92, 212).

The advantages of AIC over CV is that it accounts for degrees of freedom and it a more general statistic. Apart for being a technique of calculating the optimal bandwidth for GWR models, it is also a goodness-of-fit statistic that can be calculated for the global model and for both linear and Poisson regression methods. It can thus used to assess the superiority of a local over a global model and vice-versa (Fotheringham et al., 2002a).

Nakaya's (2001) application of GWR is interesting to this work because is the first attempt to locally model migration behaviour. Although, Nakaya recognises the importance of GWR, he uses alternative approaches to those traditionally used in GWR models. These include the weighting scheme, the bandwidth selection and the goodness-of-fit statistics. He uses a Cauchy function to define a fixed kernel:

$$w_{ij} = \frac{1}{(1 + d_{ij}^2 / h^2)^2} \tag{4.31}$$

where h is the bandwidth. He suggests the deviance (a statistic used by Flowerdew in Poisson Regression) as a goodness-of-fit statistic and the Bayesian Information Criterion (BIC) instead of AIC. Fotheringham et al. (2002a, pp. 61 – 62) comment on the latter but do not clearly conclude if BIC is a better statistic than AIC.

Although it is beyond the scope of this thesis to present the underlying mathematics of the AIC, it is necessary to provide some more details on it and to present a few references for the advanced reader. The idea of AIC was introduced by Hirotogu Akaike (Akaike, 1973). He suggested that based on the principle of maximum likelihood it is possible to construct a general information theoretic criterion to aid answering many practical problems of statistical model fitting.

For historical reasons the AIC is defined as

AIC = – 2(maximum log likelihood) + 2(number of parameters[1]) ⇔

$$\mathrm{AIC} = -2l(\hat{\theta}) + 2k \tag{4.32}$$

[1] Burnham and Anderson (1998) suggest this is the *number of estimable parameters in the approximating model*

where $l(\hat{\theta}) = \sum_{i=1}^{n} \log f(x_i \mid \hat{\theta})$ *is the maximum log likelihood function,* $\hat{\theta}$ *is the maximum likelihood estimator* (Sakamoto et al., 1986; Sakamoto, 1991; Burnham and Anderson, 1998). *The AIC was originally designed for parametric models as an approximately unbiased estimate of the expected Kullback-Leibler (K-L) information* (Hurvich et al., 1998, p. 274). *The derivation of AIC shows that the minimisation of AIC is an approximate minimization of the K-L information quantity, that is, the maximisation of the entropy* (Sakamoto, 1991, p. 15). *Sugiura (1978) derived a second-order variant of AIC that he called c-AIC. Hurvich and Tsai (1989) further studied this small-sample (second-order) bias adjustment, which led to a criterion that is called* AIC_c:

$$\mathrm{AIC_C} = -2l(\hat{\theta}) + 2k(\frac{n}{n-k-1}) \tag{4.33}$$

where the penalty term (2k) is multiplied by the correction factor $n/(n-k-1)$; n is the number of observations and k the number of independent variables in the regression. *This can be rewritten as*

$$\mathrm{AIC_C} = -2l(\hat{\theta}) + 2k + \frac{2k(k+1)}{n-k-1} \tag{4.34}$$

or, equivalently,

$$\mathrm{AIC_C} = \mathrm{AIC} + \frac{2k(k+1)}{n-k-1} \tag{4.35}$$

(Burnham and Anderson, 1998, p.51). $\mathrm{AIC_c}$ is a corrected version of AIC which was found to be less biased than AIC (Hurvich et al., 1998). The above $\mathrm{AIC_c}$ is for the case of maximum-likelihood estimation. In the case of least squares estimation (OLS) with normally distributed errors, and apart from an arbitrary additive constant, AIC can be expressed as

$$AIC = n\ln(\hat{\sigma}^2) + 2k \tag{4.36}$$

where

$$\hat{\sigma}^2 = \frac{\sum \hat{\varepsilon}_i^2}{n} (\text{the MLE of } \sigma^2) \tag{4.37}$$

and $\hat{\varepsilon}$ are the estimated residuals for a particular candidate model (Burnham and Anderson, 1998). The exact formula given by Hurvich and Tsai (1989) for parametric linear regression and autoregressive time series is

$$AIC_C = \log(\hat{\sigma}^2) + \frac{1+p/n}{1-(p+2)/n} = \log(\hat{\sigma}^2) + 1 + \frac{2(p+1)}{n-p-2} \tag{4.38}$$

where $\hat{\sigma}^2$ is the estimated error variance and p is the number of regression parameters in the model. For smoothing parameter selection p=tr(H) i.e. the trace of the *hat matrix*. This is the basis for the AIC$_C$ used here. The version of AIC$_C$ used here is presented in Equation 4.30 above.

4.1.5 Modelling Diagnostics

Goodness of fit refers to the accuracy with which a model replicates some known data. ... Experimental significance testing procedures involve the computation of a statistic for many different random drawings of data and the comparison of these values with the value obtained from the real data (Fotheringham and Rogerson, 1993, p. 12). Here the statistical tests examining the statistical significance of the parameter estimates in migration models (presented in later chapters) as well as the overall fit of these models are presented. First are the goodness-of-fit statistics for multiple regression models: *r-squared*, *F*-statistic, *ANOVA*, *deviance*, *chi-squared* and *psi*. Second are the t test for the statistical significance of the coefficients, and the variance inflection factors for testing for multicollinearity problems. Third are tests for examining the significance of local models; a Monte Carlo test for the significance of spatial variation in the local coefficients. The Akaike Information Criterion (AIC) used to measure both global and local model goodness-of-fit has been discussed above.

A comprehensive review of goodness-of-fit statistics for spatial interaction models is presented in Knudsen and Fotheringham (1986). They divide these statistics in three types: information-based statistics (including among others the Kullback and Leibler's (1951) *information gain statistic*, the *phi statistic* and the *psi statistic*), general distance statistics (such as the *standardised root mean square error* – SRMSE), and the traditional statistics (including the *r-squared statistic* and the *chi-squared statistic*).

The simplest way of getting a feeling of the explanatory power of a linear model is the coefficient of determination, also known as the square of the multiple correlation coefficient or r-squared (R^2), a statistic that measures how much of the variance of the dependent variable the independent variables have explained. R^2 varies from 0 to 1. A value of 0 denotes that there is no association between the explanatory variables (X) and the dependent variable (y), and a value of 1 denotes a perfect fit. In practice, R^2 is always less than 1. The closer R^2 to 1 is, the stronger the linear association between y and X is. The formula to calculate this statistic is

$$\text{r}-\text{squared} = \frac{\text{variation due to regression}}{\text{total unexplained variation}} \Leftrightarrow R^2 = \frac{\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \qquad (4.39)$$

where n is the number of observations, y_i is the observed values of the dependent variable (y), $\bar{y}$ is the mean of y, and $\hat{y}_i$ is the estimated values of y (Davis, 2002).

However, careful use of R^2 is necessary. This is because *there are two common misconceptions about R^2 that occasionally lead a researcher to make spurious interpretations about the relationship between X and y: One is that R^2 is not a measure of the magnitude of the slope of the regression line, and the second is that R^2 is not a measure of the appropriateness of the straight-line model* (Kleinbaum et al., 1988, p. 87). A better goodness-of-fit statistic for the linear model is the F statistic.

$$F\text{ statistic} = \frac{\text{mean squares of regression}}{\text{mean squares of residual}} = \frac{\text{sum of squares of regression/number of variables}}{\text{sum of squares of residual/degrees of freedom}} \Leftrightarrow$$

$$F = \frac{\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2 / k}{\sum_{i=1}^{n}(y_i - \bar{y})^2 / (n-k-1)} \Leftrightarrow F = \frac{R^2 / k}{(1-R^2)/(n-k-1)} \qquad (4.40)$$

Table 4.1. ANOVA for multiple linear regression.

Source of Variation	Degrees of Freedom (df)	Sum of Squares (SS)	Mean Squares (MS)	F-Test	Coefficient of Determination (R^2)
Linear Regression	k	$SS_R = \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2$	$MS_R = SS_R / k = \frac{\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}{k}$	$MS_R / MS_E \Leftrightarrow F = \frac{R^2 / k}{(1-R^2)/(n-k-1)}$	$R^2 = \frac{\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$
Deviation (Residual)	$n-k-1$	$SS_E = \sum_{i=1}^{n}(\hat{y}_i - y_i)^2$	$MS_E = SS_E / (n-k-1) = \frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{n-k-1}$		
Total Variation	$n-1$	$SS_T = \sum_{i=1}^{n}(y_i - \bar{y})^2$			

In order to find the critical value of F for hypothesis testing, there are tables of the F distribution that provide critical values by degrees of freedom and confidence level (e.g., Kleinbaum et al., 1988, pp. 649 – 655).

A summary of model goodness of fit statistics can be presented in a tabular form, the ANOVA statistics table. ANOVA stands for Analysis of Variance. Table 4.1 shows ANOVA for multiple linear regression.

An alternative to the R^2 and F statistics is the *deviance*, a statistic more appropriate for Poisson and Logit regression, the χ^2 statistic and the ψ statistic. These statistics are more appropriate for models of migration flows. Several formulas for calculating the deviance appear in the literature. Fox (1997) suggests that the deviance for the logit model is analogous to the residual sum of squares for a linear model and is defined as $G^2 = -2 \times$ the maximised log likelihood. Flowerdew and Lovett (1989) suggest the deviance D for the Poisson model as an equivalent to the log likelihood-ration statistic (G^2) which is calculated as

$$D = 2\sum_{i}\sum_{\substack{j \\ i \neq j}} n_{ij} \ln\left(\frac{n_{ij}}{\hat{\lambda}_{ij}}\right) \tag{4.41}$$

where n_{ij} is the observed migration flow from i to j and $\hat{\lambda}_{ij}$ is the estimated flow produced by the model. This formula is also used in Boyle and Flowerdew (1993; 1997).

Nakaya (2001) suggests a formula for the deviance calculated separately for each origin i, which is equivalent to (4.37):

$$DEV_i = 2\sum_{j} Y_{ij} \ln\left(\frac{Y_{ij}}{\hat{Y}_{ij}}\right) \tag{4.42}$$

and he suggests that if the calibrated model is correct, the distribution of the deviance is asymptotically chi-squared.

Fotheringham and Williams (1983) suggest a formula for calculating *a deviance statistic*:

$$D = \frac{\sum_{i}\sum_{j}\left|T_{ij} - \hat{T}_{ij}\right|}{\sum_{i}\sum_{j} T_{ij}} \times 100 \tag{4.43}$$

which is not equivalent to the *deviance* used in studies presented above.

Flowerdew and Aitkin (1982) suggest a chi-squared statistic as an alternative measure of goodness-of-fit:

$$\chi^2 = \sum_{i=1}^{n}\sum_{\substack{j=1 \\ j \neq i}}^{n} \frac{\left(n_{ij} - \hat{\mu}_{ij}\right)^2}{\hat{\mu}_{ij}} \tag{4.44}$$

Finally, the modified *psi* statistic (Knudsen and Fotheringham, 1986) can be used to measure goodness of fit. This is defined as

$$\psi = \sum_j p_{ij} \left| \ln\left(\frac{p_{ij}}{s_{ij}}\right) \right| + \sum_j q_{ij} \left| \ln\left(\frac{q_{ij}}{s_{ij}}\right) \right| \tag{4.45}$$

where p_{ij} is the observed migration from i to j divided by the total out-migration from i; q_{ij} is the predicted migration from i to j divided by the total out-migration from i; and $s_{ij}=(p_{ij}+q_{ij})/2$ (Fotheringham et al., 2002b). For all latter three statistics (deviance, chi-squared, psi) the lower their value is, the better the model fit is.

In order to evaluate the statistical significance of the parameter estimates, the simple (Student's) t test is employed. The *Student's t* distribution curve is similar to the normal distribution curve (symmetrical around 0). The t test focuses of testing whether a parameter estimate (estimated coefficient $\hat{\beta}$ of a variable x in a regression model) is significantly different from 0. The null hypothesis is H_0: $\hat{\beta}=0$ and the equivalent statistic for testing this null hypothesis is $T=\frac{\hat{\beta}}{S_{\hat{\beta}}}$, where $S_{\hat{\beta}}$ is the estimate of the standard error of $\hat{\beta}$. Both $\hat{\beta}$ and $S_{\hat{\beta}}$ are printed by standard regression programs. In performing this test, H_0: $\hat{\beta}=0$ is rejected if

$$\begin{cases} |T| > t_{n-k-2,1-a/2} & \text{(two-sided test; } H_A : \hat{\beta} \neq 0) \\ T > t_{n-k-2,1-a} & \text{(upper one-sided test; } H_A : \hat{\beta} > 0) \\ T < -t_{n-k-2,1-a} & \text{(lower one-sided test; } H_A : \hat{\beta} < 0) \end{cases} \tag{4.46}$$

where n is the number of observations and k the number of variables in the regression model, $n-k-2$ is the degrees of freedom, a is the significance level, and t is the critical value (from t distribution tables) (Kleinbaum et al., 1988).

Typically, the two-sided test is used; a parameter estimate of a variable in a multiple linear regression model is statistically significant when its T value exceeds the critical value t in the 95% *confidence interval* (a=0.05). This confidence interval is typically used in social sciences. The above tests tell us about the power of the explanatory variables and the model to explain and estimate the dependent variable.

Another problem of regression models is multicollinearity effects in the independent variables. This is the degree of correlation between two independent variables. If this correlation is high, then the model is not efficient, the coefficients are biased and the interpretation of the effect an independent variable has on the dependent variable is not clear since the independent variable in the model may capture other effects that it is meant to. To overcome this problem it is necessary to conduct auxiliary regression; this is a regression of each independent variable of the model on the remaining independent variables. Standard regression programs (e.g., SPSS) calculate during the model calibration the *variance inflection factors* (VIFs). The VIF is computed as

$$VIF_j = \frac{1}{1-R_j^2}, j = 1,2,...,k \tag{4.47}$$

where R_j^2 is the coefficient of determination obtained from the auxiliary regression of independent variable j on the remaining independent variables. The larger a variable's VIF is the more troublesome the variable is. *As a rule of thumb, a VIF greater than 10 (equivalent to R_j^2 being greater than 0.9) indicates a multicollinearity problem that will lead to problems with interpretation of the final model if it is not corrected* (Fotheringham et al., 2002b, p. 95). The set of variables used in the models here (Chapters 6 and 7) is a subset of the variables used in the models presented in previous work (Fotheringham et al., 2002b; Fotheringham et al., 2003) where multicollinearity problems have already been addressed. In some sample tests for multicollinearity all VIFs were small (2-5) and occasionally higher (7-9) for some variables, but less than 10 in all cases.

In the case of local forms of regression (GWR) it is necessary to test for the existence of significant spatial variation of the local coefficients (parameter estimates). In order to do this there are two methods supported by the software used here (GWR2.0), a Monte Carlo significance test or Hope test (Hope, 1968) and a test attributed to Leung et al. (2000a; 2000b).

The aim of a nonstationary significance test is to examine if the location of the observations i (x_i, y_i) in the local model is such that the model calibration results in local coefficients that would not been resulted in if these observations where differently allocated in space. *Under the null hypothesis, any permutation of (x_i, y_i) pairs among the geographical sampling points i are equally likely to occur.*" (Brunsdon et al., 1996, p. 288). "*Thus, the observed values of s_k (the standard deviation of n local parameter estimates of a variable in the local model) could be compared with the values obtained from randomly rearranging the data in space and repeating the GWR procedure. The comparison between the observed s_k value and those obtained from a large number (99 in this case) of randomisation distributions forms the basis of the significance test. Making use of the Monte Carlo approach, it is also the case that selecting a subset of random permutations of (x_i, y_i) pairs amongst i and s_k computing will also give a significance test when compared with the observed statistics.* (Fotheringham et al., 1998, p. 1912). More details on Monte Carlo test and Leung's test are discussed in Fotheringham et al. (2002a) in *Statistical Inference and GWR* (Chapter 4) and in *Software for GWR* (Chapter 9). Here, the following rule of thumb was used for the Monte

Carlo test: when the test value is equal to or less than 0.05 the local coefficient exhibits significant spatial variation.

4.2 Out-migration models

This section discusses the global and local forms of out-migration models. The results of the analysis are presented in Chapter 6. This section also mentions alternative calibration techniques which were tested during preliminary analysis but not used in the final model.

The aim of modelling here is to identify the effect several out-migration determinants (independent variables) have on out-migration rates (dependent variable). The simplest way to regress out-migration rates on a set of out-migration determinants is through linear regression. However, assuming that out-migration determinants are linearly correlated with out-migration rates is prone to misspecification bias. Thus, it is safer to construct a non-linear model. In the literature there are many variations of such models: the simple linear model (Miller, 1973; Sommers and Suits, 1973; Meyer et al., 2001), the power model (Fotheringham et al., 2002b), and the logit model (Ferguson and Kanaroglou, 1997; Cannari et al., 2000) to name only a few. In those cases (of the above studies) that a linear model was used, out-migration or net migration instead of out-migration rates was used as the dependent variable.

The out-migration model specification follows the form of that presented in Fotheringham et al. (2002b). Out-migration rate is expressed as the product of power and exponential functions of out-migration determinants. The out-migration rate has been calculated as the number of out-migrants per 1000 residents. Thus, population in the origin is forced in the model to have a linear relationship with out-migration.

In the following two sections, the global and local forms of the log-linear out-migration model are discussed.

4.2.1 The log-log OLS out-migration model

The linear out-migration model is presented in Equation 4.48. In this model, the dependent variable is out-migration rates per thousand population (M). This is regressed on a set of k out-migration determinants (x_j independent variables). The regression results in the intercept ($\hat{a}_0$) and k parameter estimates ($\hat{a}_j$, $j=1,2,...,k$), estimates of the out-migration rates ($\hat{M}$) and the residuals ($\hat{e}_i$) as shown in Equation 4.49.

$$M_i = a_0 + \sum_{j=1...k} a_j x_{ij} + e_i \tag{4.48}$$

$$\hat{M}_i = \hat{a}_0 + \sum_{j=1...k} \hat{a}_j x_{ij} + \hat{e}_i \tag{4.49}$$

It was explained above that a log-linear model is more appropriate for out-migration. An exponential form of the model (4.48) can be formed (Equation 4.50).

$$M_i = a_0 \times \prod_{j=1...k} x_{ij}^{a_j} \times e_i \tag{4.50}$$

The model (4.50) can be logged and the result is a linear model relatively easy to calibrate in standard computer programs (such as SPSS). However, there are two problems with the model (4.50). One is that there are variables that have non-positive values and thus cannot be logged; the other is that statistical bias is introduced to the intercept of the log-log regression (Heien, 1968). These issues are addressed in model (4.51) where the set of variables with non-positive values are separated and an adjustment factor for the intercept is introduced. Model (4.51) also includes subscripts denoting the sex/age disaggregation of the migration data used here. The migration rate (*M*) of model (4.48) is presented as a ratio the numerator of which is the total out-migrants in each area (*i*) for each sex/age group (*as*) and the denominator is the number of residents in thousands for the appropriate sex/age group.

$$O_{ias} \big/ (P_{ias} / 1000) = e^{K_{as}} \times A_{as} \times \prod_{1 \le k \le N_1} V_{ki}^{a_k^{as}} \times \prod_{N_1 \le m \le N_2} e^{a_m^{as} \times V_{mi}} \tag{4.51}$$

where $e^{K_{as}}$ in the intercept in an exponential form, A_{as} is an adjustment factor to ensure the total estimated out-migrants equals the total observed out-migrants. N_1 is the number of the variables that can be logged and N_2 is the total number of variables. There are N_2–N_1 variables that cannot be logged because of zero and negative values. Both sides of equation 4.51 can be logged resulting equation 4.52.

$$\ln(Migr_{ias}) = K_{as} + \ln(A_{as}) + \sum_{1 \le k \le N_1} a_k^{as} \ln(V_{ki}) + \sum_{N_1 \le m \le N_2} a_m^{as} \times V_{mi} \tag{4.52}$$

Model (4.51) was introduced in previous work (Fotheringham et al., 2002b, p. 168) and is used here to allow comparisons. However, recent work proposed a quadratic model for out-migration (Fotheringham et al., 2002b, Chapter 11; Fotheringham et al., 2003).

There are some issues that need to be clarified to allow comparisons between this work and previous work (Fotheringham et al., 2002b; Fotheringham et al., 2003). The first issue refers to the mathematical form of the out-migration models: Fotheringham et al. (2002b) Phase II and Fotheringham et al. (2003) use the quadratic model whereas here the log-log model is used (also used in Fotheringham et al., 2002b, Phase I). The second issue refers to the number of observations included in a single model. In the quadratic model, 1372 observations were included in a single regression combining 14 years of data. However, here, 98 observations were included in a single regression, but there are 14 separate regressions for the equivalent years of data. The third issue is that different variable configurations were applied; previous models (Fotheringham et al., 2002b; Fotheringham et al., 2003) included a time trend, several regional variables and some national time-trend variables additionally to cross-sectional variables. Here, a few cross-sectional variables and one regional variable (regional population) were included (see Chapter 3 for more details).

The model configuration varies across age disaggregated data. This is because some of the cross-sectional variables are inappropriate for all age groups. The log-log out-migration model is a global model opposed to its local version discussed below. The results of the global out-migration models are presented in Chapter 6.

4.2.2 A Geographically Weighted version of the log-log OLS model

The geographical weighted version of the log-log out-migration model presented above is:

$$\ln(Migr_{ias}) = K_{as}(u_i,v_i) + \ln(A_{as})(u_i,v_i) + \sum_{1\leq k\leq N_1} a_k^{as}(u_i,v_i)\ln(V_{ki}) + \sum_{N_1\leq m\leq N_2} a_m^{as}(u_i,v_i)\times V_{mi} \quad (4.53)$$

After calibration in GWR 2.0 this model results in 98 local parameter estimates for each of the N_2 variables as well as 98 local intercepts. Models such as (4.53) have been calibrated for 14 sex/age disaggregated migration groups (see Chapter 3) for each of the 14 time periods complete data are available for. The number of observations is 98 (FHSAs in England and Wales) and the number of variables varies between 13 and 15 depending on the disaggregated migration group. The results are presented and discussed in Chapter 6.

In order to present the results of the local out-migration models, it was necessary to summarise and visualise these results. This was a challenging task given the big number of results. To automate this task, it was necessary to write code in several software packages (see below). GWR 2.0 outputs are in two forms: a text file containing all diagnostics, regression results, and goodness-of-fit statistics of both global and local regressions; and a structured file (several options are available) containing the parameter estimates, residuals and diagnostics for each local model (data point). The latter can be mapped using appropriate software, here ESRI ArcView.

The boxplots presented in Chapter 6 have been created using script in the statistical package called *R*, a freeware software package that is of growing use among academics (Ihaka and Gentleman, 1996; CRAN, 2003). For the maps to be converted from ArcInfo output format to ArcView shapefile format a two-step procedure was used. First the ArcInfo output file (.e00) was imported using the ArcView utility *Import71* and then code in Visual Basic using ESRI MapObjects allowed the transformation to shape files.

4.3 Destination-choice models

Destination-choice is the second stage of a migration decision process. At this stage, models try to explain the effect of characteristics of potential destinations on attracting migrants. For reasons that have been discussed above, the origin-constrained gravity model is the most appropriate way to analyse the dataset available here. The calibration technique selected for this is the Poisson regression, because at this stage, migration flows rather than rates are the values of the dependent variable.

The modelling technique presented below was applied to sex/age disaggregated data for migrants leaving an FHSA in North England (Newcastle) and an FHSA in South England (Camden and Islington). The reason the modelling did not include all 98 potential origins in the system was because of the very big number of results that makes it impossible to include them in this thesis. I believe that by examining and comparing trends on the way pull factors locally attract migrants originated from either North or South England, my aim of testing trends in destination-choice migration models is satisfied. In each model there are 97 observations, referring to the 97 potential destinations (FHSAs) of an origin. The data available here are for seven time periods, from 1990-91 to 1996-97. The time varying variables here have not been lagged by one year, as in the case of the out-migration rates models. This is because although there is an accumulative decision to migrate out of an area, the choice of the destination is based on its attractiveness at the time the choice is to be made. An issue here is the low degrees of freedom, as the number of variables is over 20. Because it is important to test for effects of a set of variables that found to be of some significance in previous research (Fotheringham et al., 2002b), the initial models included all variables used in previous work. However, model configurations including only variables found to have a significant effect on destination choice are also presented. The results are presented in Chapter 7.

4.3.1 The Competing Destinations Model

As discussed above, there is empirical evidence that migrants chose their destinations hierarchically. Thus, it is necessary to account for this in the model. This is possible with the inclusion of a variable called *destination accessibility* in the production-constrained gravity model. The result is the *competing destinations model (CDM)* (Fotheringham, 1983; 1991; 2000).

Standard statistical package algorithms (such as Poisson in R) require linear model forms. Thus, for the power model applied here it is required that independent variables are logged. However, some of the variables (such as Climate Index and Employment Growth) cannot be logged because they have non-positive values. Thus, they are included un-logged in the model. The general form of the competing destinations model (2.7) for the disaggregated data used here is

$$M_{ij}^{as} = \frac{O_i^{as} \prod_k W_{jk}^{a_k^{as}} d_{ij}^{\beta^{as}} A_j^{\gamma^{as}}}{\sum_j \prod_k W_{jk}^{a_k^{as}} d_{ij}^{\beta^{as}} A_j^{\gamma^{as}}} \tag{4.54}$$

where, M_{ij}^{as} is the number of individuals belonging to migrant group *as* who migrate from origin i to destination j; O_i^{as} is the known out-migration volume of migrants of group *as* from FHSA i; W is one of the k characteristics of destination j affecting the choice of j by migrants of group *as* from i; d_{ij} is the distance between i and j; A_j is the destination accessibility of j; and a_k^{as}, β, and γ are their parameter estimates, respectively. When the model is calibrated for a specific origin the rate $O_i^{as} / \sum_j \prod_k W_{jk}^{a_k^{as}} d_{ij}^{\beta^{as}} A_j^{\gamma^{as}}$ is constant (k_i^{as}), thus (4.54) becomes

$$M_{ij}^{as} = k_i^{as} \prod_k W_{jk}^{a_{ik}^{as}} d_{ij}^{\beta_i^{as}} A_j^{\gamma_i^{as}} \tag{4.55}$$

To be more consistent with the Poisson regression as well as the inclusion of un-logged variables the correct form of the destination-choice model for a specific origin i and a specific migration group *as* is

$$M_j = \exp[k + \sum_{1 \le l \le N_1} a_l \ln(W_{lj}) + \sum_{N_1 \le m \le N_2 - 2} a_m W_{mj} + \beta \ln(d_{ij}) + \gamma \ln(A_j)] \tag{4.56}$$

where, j is one of the 97 destinations, l is one of the N_1 variables that can be logged (among which is total population of the destination), m is one of the $N_2 - N_1 - 2$ variables that cannot be logged, and N_2 is the total number of variables in the model.

4.3.2 A Geographically Weighted version of The Competing Destinations Model

The aim for a local destination-choice model is to examine spatial non-stationarity of the attraction effects of migration determinants on migrants. With the exception of Nakaya's (2001) model, traditional destination choice models assume a stationary effect of a destination characteristic across potential destinations. For example, it is assumed that total population of a potential destination has the same effect on attracting migrants leaving a given origin regardless of the location of the destination. This is not to be confused with the different behaviour of migration determinants on destination choice across different origins.

Here I examine only the existence of spatial non-stationarity of migration determinants on the destination-choice decision by migrants leaving Newcastle as well as Camden and Islington FHSAs. The Geographically Weighted version of model (4.56) is

$$M_j = \exp[k(u_j,v_j) + \sum_{1 \le l \le N_1} a_l(u_j,v_j)\ln(W_{lj}) + \sum_{N_1 \le m \le N_2 - 2} a_m(u_j,v_j)W_{mj} + \beta(u_j,v_j)\ln(d_{ij}) + \gamma(u_j,v_j)\ln(A_j)] \tag{4.57}$$

where, (u_j,v_j) are the x, y coordinates of the centroid j of each of the 97 FHSAs.

4.4 Summary

The aim of this chapter is to provide a detailed discussion of the methodology used here. Thus, I described the most common multivariate regression techniques and I justified the selection of the appropriate techniques for my modelling (log-log OLS for out-migration and Poisson regression for destination choice models). I also discussed the statistical inference for these techniques. I presented the modelling diagnostics I am going to use in order to evaluate the goodness-of-fit for my models and the statistical significance of my results. I provided a quick review of the methodology for local regression (GWR) because this is used for the first time in the migration literature. Finally, I presented the definitions of global and local models of out-migration and destination choice by providing the exact equations.

To this point, I discussed the background, the dataset and the methodology for my study. These are necessary in order to justify the various choices I made out of the variety of data sources and methodologies available for migration studies. They are also necessary in order to position this work in the spectrum of migration studies and to make it relevant to the corresponding disciplinary research stream. The reminder of this thesis includes my analysis, results and conclusions.

I now discuss temporal and spatial trends in migration in England and Wales in the 1980's and the 1990's. One of the innovative parts of the following chapter is the use of ***heat maps*** as a new means for visualising migration rates over time.

Chapter 5
Spatial and Temporal trends in migration

In this chapter, spatial and temporal trends in migration in England and Wales are examined. The aim is to identify areas of low, average, and high gross (in-, out-) and net migration rates, to examine the temporal stability of these rates and to examine trends as well as the temporal variation of the migration flows between pairs of places. For the latter I examine trends for Newcastle upon Tyne FHSA as well as London (16 FHSAs in London aggregated).

Because of the large number of migration data available, it is necessary to employ several visualisation techniques. Furthermore, contemporary techniques for exploring spatial data are tested for their usability in exploratory univariate data analysis. In order to classify areas of similar out-migration rates, single-variable k-means clustering was initially applied. The results are then plotted to allow the identification of any spatial clusters. The temporal stability of out-migration rates across all years can be examined in two ways: by k-means clustering of the sex/age disaggregated data, or by plotting the out-migration rates for each FHSA of the age disaggregated data.

Additionally to the k-means clustering, statistics measuring spatial dependence and spatial autocorrelation have been applied to sample data. These are contemporary techniques that may have a lot of potential in data exploration. These include Moran's I, Geary's c and Getis G statistics. Furthermore, the calculation of the local geographically weighted mean has been applied to some data. All these approaches aim to shed more light in unveiling spatial patterns of migration involving spatial analysis, since k-means is an aspatial approach. The first section of the chapter is a discussion of the k-means clustering algorithm and the second describes the selection of the appropriate number of clusters for a k-means cluster analysis. Five sections follow. These discuss:

Section 3: Age disaggregated out-migration rates

Section 4: Sex/age disaggregated out-migration rates

Section 5: Exploratory spatial data analysis and local statistics

Section 6: In-migration rates, out-migration rates, and net migration rates

Section 7: Migration flows (Newcastle FHSA and London as origins).

5.1 K-means clustering

The k-means clustering algorithm is described in detail by Hartigan (1975). The k-means used here is an efficient version of the algorithm presented in Hartigan and Wong (1979). *The aim of the K-means algorithm is to divide M points in N dimensions into K clusters so that the within-cluster sum of squares is minimized. It is not practical to require that the solution has minimal sum of squares against all partitions, except when M,N are small and K=2. We seek instead "local" optima, solutions such that no movement of a point from one cluster to another will reduce the within-cluster sum of squares* (Hartigan and Wong, 1979, p. 100). Here N=1 and M=98; the algorithm is allowed to have ten iterations maximum and the initial cluster centres are not introduced, but selected from the algorithm. The steps of the algorithm along with the cluster centres initialization process follow.

The algorithm requires as input a matrix of *M* points in *N* dimensions, and a matrix of *K* initial cluster centres in *N* dimensions. The number of points in cluster L is denoted by *NC(L)*. *D(I,L)* is the Euclidean distance between point *I* and cluster *L*.

Step 1. For each point *I* (*I* = 1,2, ..., *M*), find its closest and second closest cluster centres, *IC*1(*I*) and *IC*2(*I*) respectively.

Step 2. Update the cluster centres to be the averages of points contained within them.

Step 3. Initially, all clusters belong to the live set.

Step 4. This is the optimal-transfer (*OPTRA*) stage: Consider each point *I* (*I* = 1,2, ..., *M*) in turn. If cluster *L* (*L* = 1,2, ..., *K*) is updated in the last quick transfer (*QTRAN*) stage, then it belongs to the live set throughout this stage. Otherwise, at each step, it is not in the live set if it has not been updated in the last *M* optimal-transfer steps. Let point *I* be in cluster *L*1. If *L*1 is in the live set, do *Step* 4a; otherwise do *Step* 4b.

Step 4a. Compute the minimum of the quantity, $R2 = [NC(L)*D(I,L)^2]/[NC(L)+1]$, over all clusters *L* ($L \neq L1$, $L = 1,2, \ldots, K$). Let *L*2 be the cluster with the smallest *R*2. If this value is greater than or equal to $[NC(L1)*D(I,L1)^2]/[NC(L1) - 1]$, no reallocation is necessary and *L*2 is the new *IC*2(*I*). (Note that the value $[NC(L1)*D(I,L1)^2]/[NC(L1) - 1]$ is remembered and will remain the same for point *I* until cluster *L*1 is updated.) Otherwise, point *I* is allocated to cluster *L*2 and *L*1 is the new *IC*2(*I*). Cluster centres are updated to be the means of points assigned to them if reallocation has taken place. The two clusters that are involved in the transfer of point *I* at this particular step are now in the live set.

Step 4b. This step is the same as *Step* 4a, except that the minimum *R*2 is computed only over clusters in the live set.

Step 5. Stop if the live set is empty. Otherwise, go to *Step* 6 after one pass through the data set.

Step 6. This is the quick transfer (*QTRAN*) stage: Consider each point *I* (*I* = 1,2, ..., *M*) in turn. Let *L*1 = *IC*1(*I*) and *L*2 = *IC*2(*I*). It is not necessary to check the point *I* if both the clusters *L*1 and L2 have not changed in the last M steps. Compute the values

$R1=[NC(L1)*D(I,L1)^2]/[NC(L1) - 1]$ and $R2=[NC(L2)*D(I,L2)^2]/[NC(L2) + 1]$.

(As noted earlier, *R*1 is remembered and will remain the same until cluster *L*1 is updated.) If *R*1 is less than *R*2, point *I* remains in cluster *L*1. Otherwise, switch *IC*1(*I*) and *IC*2(*I*) and update the centres of clusters *L*1 and *L*2. The two clusters are also noted for their involvement in a transfer at this step.

Step 7. If no transfer took place in the last *M* steps, go to *Step* 4. Otherwise go to *Step* 6.

One way of obtaining the initial cluster centres is: the points are first ordered by their distances to the overall mean of the sample. Then, for cluster *L* (*L* = 1,2, ..., *K*), the $\{1+(L-1)*[M/K]\}$th point is chosen to be its initial cluster centre.

Source: Hartigan and Wong, 1979, pp. 100-103.

The k-means procedure attempts to identify relatively homogeneous groups of cases based on selected characteristics, using an algorithm that can handle large numbers of cases. The procedure tries to form groups that do differ. The reason for a k-means cluster analysis is that it allows the grouping of FHSAs into categories of similar out-migration rates. It is a quick algorithm the results of which can be mapped. The large number of sex/age disaggregated data requires a level of visualization so that the data exploration is easier than employing tables of figures or even graphs.

5.2 Selecting the number of clusters

The number of clusters is subjective and the only way to select the most appropriate is to test with different numbers of clusters and study the differences. K-means with two, three, four and five clusters are applied to the out-migration rates for all 14 sex/age groups and for 14 years of data for 98 FHSAs in England and Wales. The selection of the number of clusters is based on sample analysis for the male age group 30 – 44. This is because this is the largest migrant group and its behaviour is important for the policy decision makers. Figure 5.1 shows the clusters and their centres for males 30-44 in 1997-98.

When two clusters are selected, most of the FHSAs in London (except Bexley and Barking) and Manchester are classified as the high out-migration cluster and the rest of England and Wales as the low out-migration cluster. Increasing the number of clusters by one results in a split of the lower cluster into two new clusters: one that includes populous areas such as Newcastle, Liverpool, Manchester and neighbour FHSAs and a spatial cluster that includes Birmingham, Cambridgeshire, the FHSAs in the North and Southwest London as well as Bexley, Barking, and Croydon; and the rest FHSAs in England and Wales. The only change of the high out-migration cluster is the removal of Croydon.

When four clusters are selected several changes take place in the cluster membership. While the highest out-migration cluster loses two of its members (Hillingdon and Redbridge), there are significant changes to the rest of the clusters and their centres. The lowest cluster now includes rural and remote areas and surprisingly some of the neighbouring FHSAs to Manchester and Leeds. The second higher cluster includes Newcastle, Salford, Trafford, Stockport, Birmingham, Solihull, Oxfordshire, Buckinghamshire, Hertfordshire, Berkshire, Surrey and the FHSAs in London that are not in the highest cluster.

In the case of five clusters k-means, the cluster of the highest out-migration rates remains the same as in the case of four cluster k-means. However, changes take place in the cluster membership for the remaining FHSAs. These are now distributed over four clusters. There are two clusters, those with centres 34.8953 and 43.0284, which are not very distinct and could be unified. There are also cases (other migrant groups) where one of the five clusters (that of the highest out-migration rates) has few members (2-4). Actually, such a cluster separates the extreme out-migration rates out of a particular data set. Thus, the five cluster k-means does not overall improve the identification of spatial clusters of out-migration compared to the four cluster k-means.

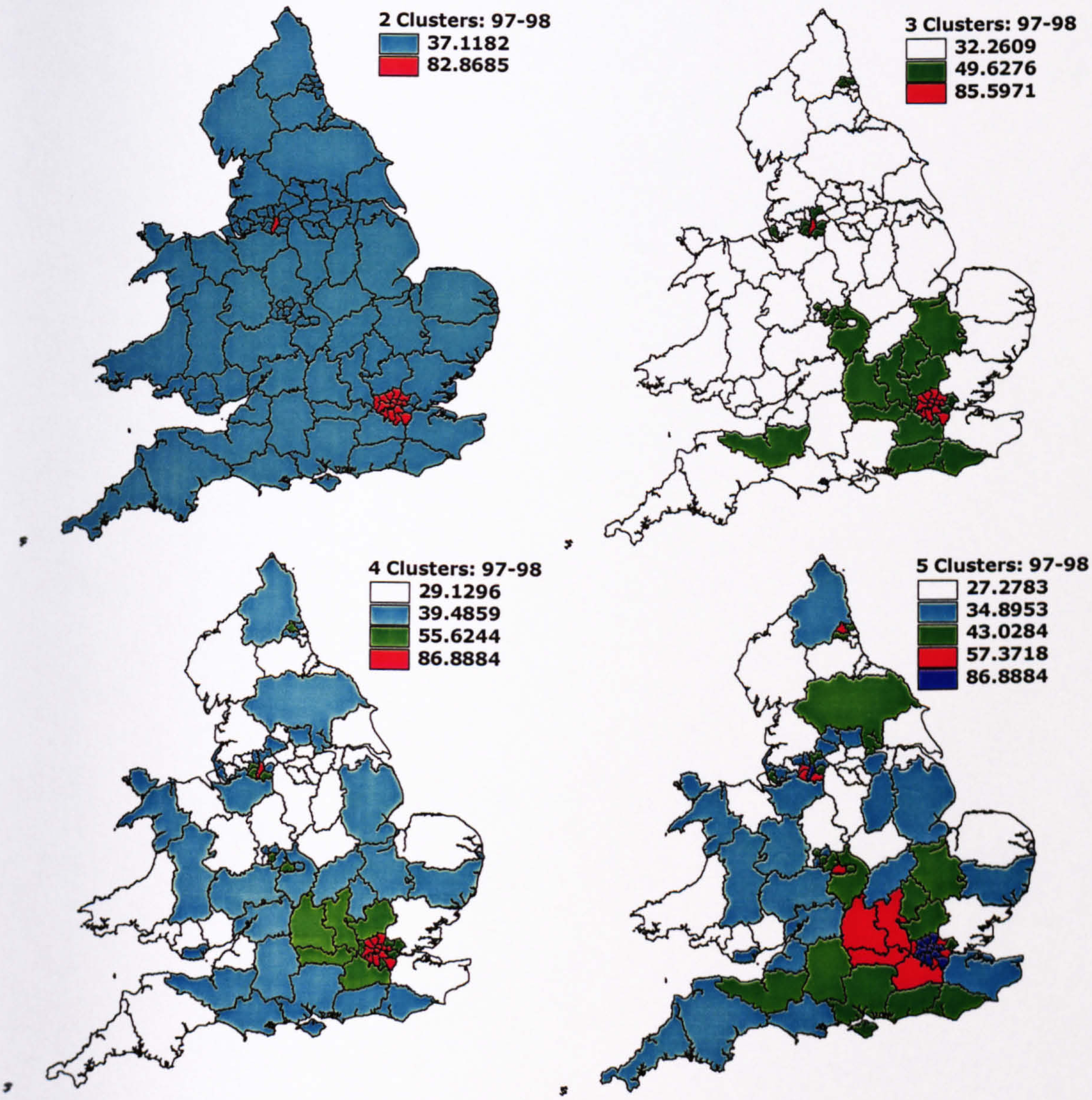


Figure 5.1. Clusters and their centres for males aged 30-44 in 1997-98.

The aim here is to identify clusters of very low and very high out-migration rates, and to ensure that clusters do not have too few members (two or three, for example). The two cluster k-means gives little information about low and high out-migration rates. The five cluster k-means generally has too few members in the highest cluster. Thus, the decision is to be made mainly between three and four cluster k-means, but five cluster k-means could also be considered. The three cluster approach will divide the observations into low, average and high out-migration rates. The introduction of the fourth cluster divides the average rates into two groups, but at the same time removes the observations with the highest rates from the lowest rates cluster and the observations with the lowest rates from the highest rates cluster. The latter removes the averaging effect in the centre of the two extreme clusters (lowest and highest out-migration rates). The five cluster approach divides the low out-migration rates from two into three clusters, which does not help making the picture of spatial patterns identification clearer. Thus, I believe four cluster k-means is the best way to satisfy the above stated aims of this method. Four cluster k-means has been selected as the most appropriate for the out-migration data and is applied henceforth. The resulting clusters are presented in the following section.

5.3 Age disaggregated out-migration rates

Before any spatial analysis of out-migration rates is undertaken, it is necessary to present a basic image of out-migration rates in different areas and time. Generally, age disaggregated out-migration rates (out-migrants per 1000 population of the corresponding age group) of FHSAs in England and Wales are low and stable over time for children (aged 0-15), mature adults (aged 30-44), older adults (aged 45-59) and pensioners (aged 60 and over). In most of the FHSAs out-migration rates of these four groups form four parallel lines when plotted on a time series chart (see Figures 5.2a-d). In an ascending order of out-migration rates, lowest are those of pensioners, followed by those of older adults and children, and highest are the rates for mature adults. The remaining age groups (teenagers, young adults and adults) exhibit a significant temporal variation, and a ranking variation.

In Figures 5.2a to 5.2d charts of out-migration rates in four FHSAs, 7 age groups and 15 time periods are presented. South Tyneside, Hampshire, Newcastle, and Lambeth (with Southwark and Lewisham) FHSAs are representatives of their corresponding clusters in a four cluster k-means analysis (Section 5.2). The selection has been made on the basis that the out-migration rate of each FHSA has the smallest difference with the corresponding cluster centre. The clusters and their centres are presented in Section 5.2 (see Figure 5.1).

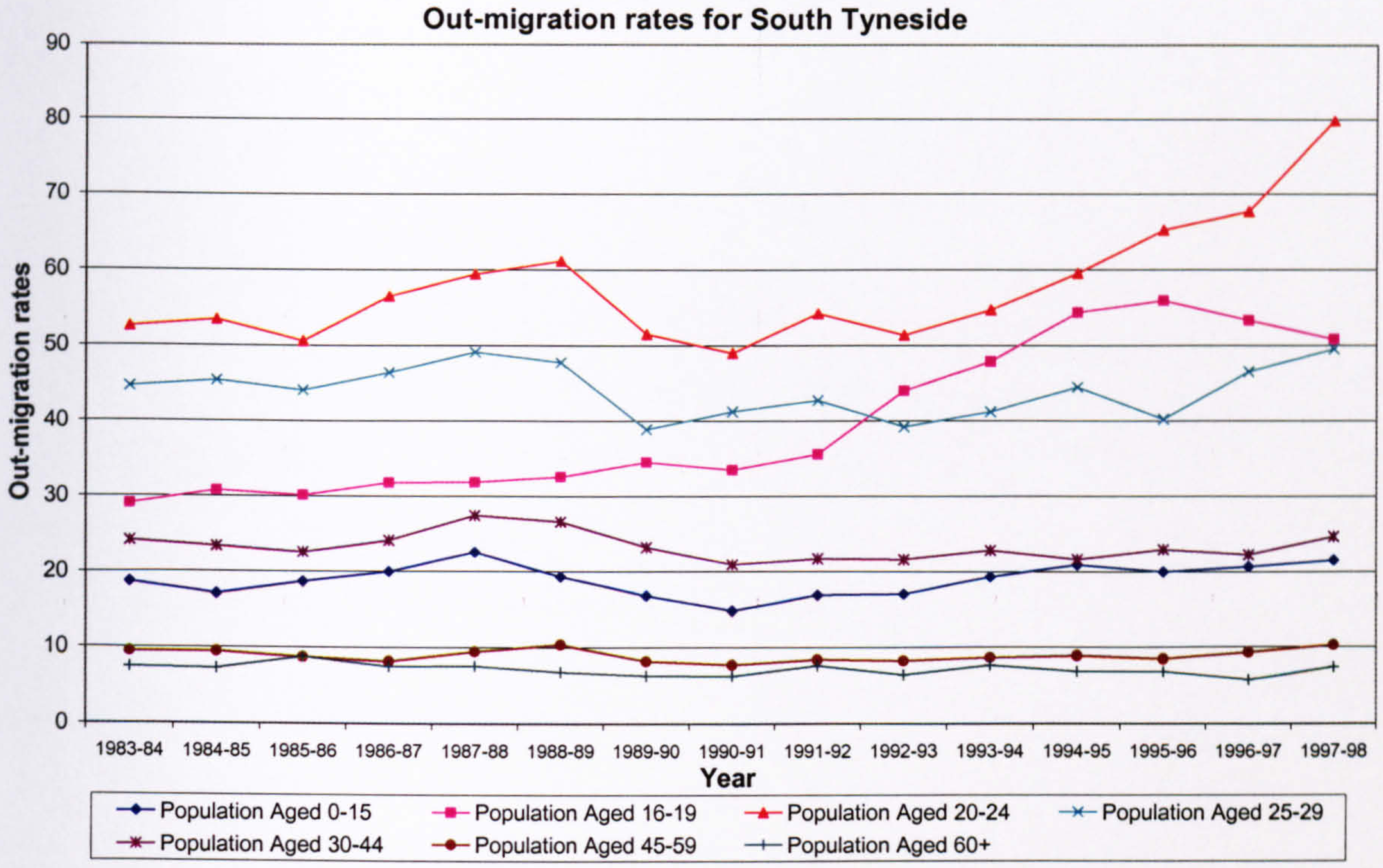


Figure 5.2a. Out-migration rates in South Tyneside FHSA for the period 1983-1998.

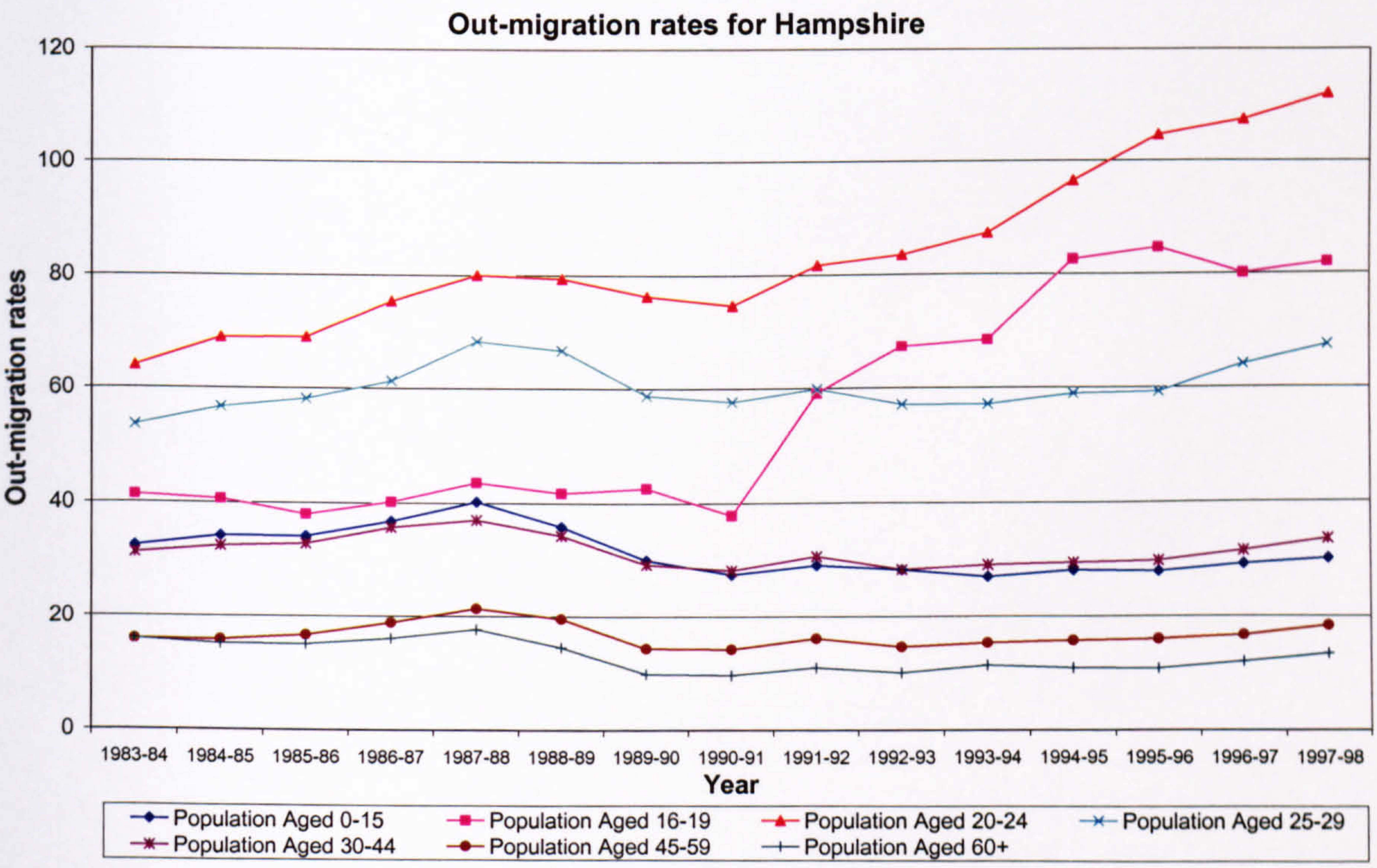


Figure 5.2b. Out-migration rates in Hampshire FHSA for the period 1983-1998.

South Tyneside represents the areas that exhibit the lowest out-migration rates. The data in Figure 5.2a confirm the general trends of out-migration for age groups (Rogers et al., 1978) and time. The lowest out-migration rates are for pensioners and older adults (less than

1%) which are very stable over time. The difference in rates for mature adults and (their) children is low. The rates for the latter two age groups exhibit slight temporal variation with a high peak in 1987-88 and a trough in 1990-91. Out-migration rates for teenagers are higher than the previous four groups and they exhibit a consecutive increase from mid-1983 to mid-1995 and a slight decrease in the last two years of the study period (mid-1983 to mid-1997). The increase in the rates for the latter group is very rapid between mid-1991 and mid-1995 a trend which has been observed in all four FHSAs presented here. The out-migration rates for young adults and older adults have an inconsistent temporal variation; they increase and decrease over the 15 years. The most significant trend is the rapid increase in the rates for young adults after the mid-1992. The latter age group experiences the highest out-migration rates among all age groups in this cluster of FHSAs (lowest out-migration rates).

Trends in out-migration rates for Hampshire (a representative of the second lowest out-migration clusters in England and Wales) are similar to those of South Tyneside (Figures 5.2a-b). Out-migration rates for children; mature and older adults; and pensioners exhibit a slight increase during the mid-1980s (high peak in mid-1987) followed by a rapid decrease with a trough in 1990-91 and a recovery to the mid-1983 rates in mid-1997. In Hampshire, the rates for children and their parents (mature adults) are very similar; the former are higher during the first half of the study period and lower during the second half. Out-migration rates for teenagers have almost no variation from mid-1983 to mid-1990, but they dramatically increase (over 100%) over the following four years. The increase of out-migration rates for young adults is over 50% for the 15 years study period and most of this increase takes place during the 1990s. Out-migration rates for adults increase in the long term, they follow waves of increase and decrease similar to that of the older age groups.

For South Tyneside and Hampshire rates for teenagers are the highest among all age groups; rates for adults are higher than those for young adults from mid-1983 to mid-1991 when this trend reverses.

There are two interesting trends in out-migration rates in Newcastle (Figure 5.2c). One is that the rates for teenagers are as low as the rates for children between 1983-84 and 1990-91, but they rise dramatically by more than 200% during the 1990s and they become higher that the rates for mature adults. The second trend is the doubling of out-migration rates for adults from 1992-93 to 1997-98 when they become the highest rates among all age groups. Rates for teenagers are stable during the 1980s, they increase by less than 50% in early 1990s and then they slightly decrease. The remaining age groups exhibit similar trends to the previous FHSAs with the only difference being that there is a slight increase in the rates in the long term.

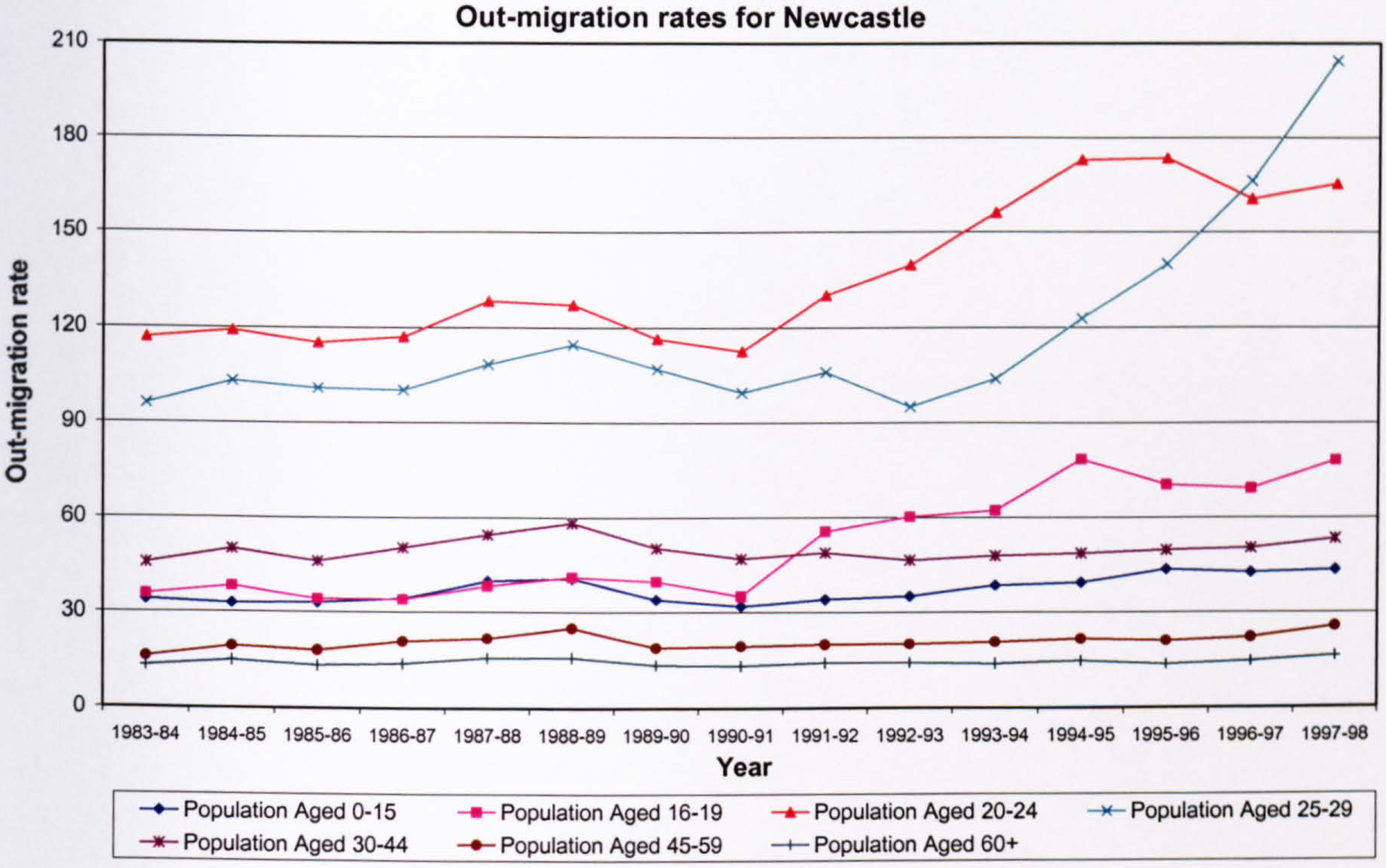


Figure 5.2c. Out-migration rates in Newcastle FHSA for the period 1983-1998.

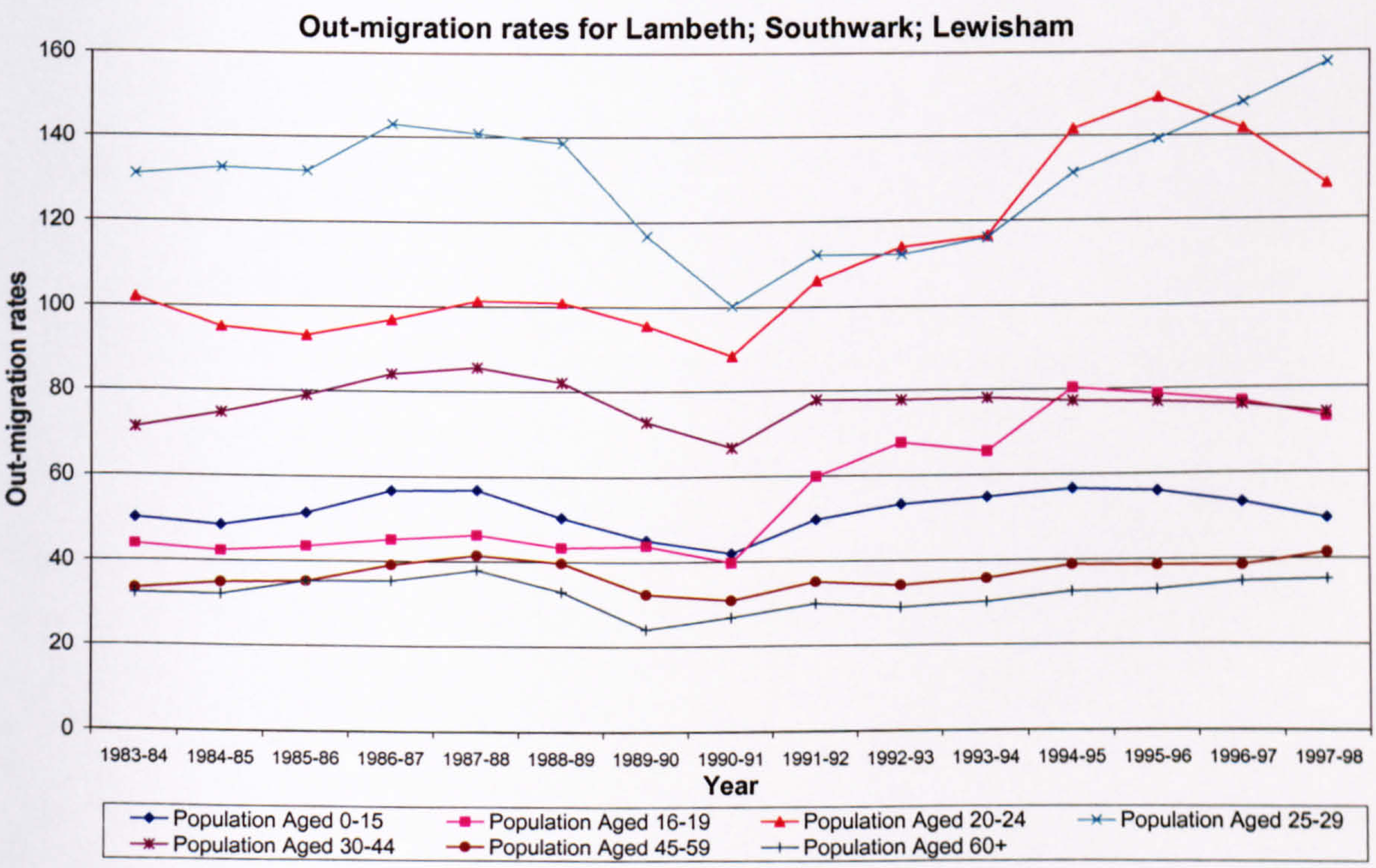


Figure 5.2d. Out-migration rates in Lambeth; Southwark; Lewisham FHSA for the period 1983-1998.

Lastly, the temporal trends for areas of high out-migration, such as Manchester and FHSAs in London are presented in Figure 5.2d. The representative FHSA for this cluster is Lambeth (with Southwark and Lewisham). Here all age groups have a different behaviour compared to each other, but general trends compared to the previous FHSAs are fairly consistent. The out-migration rates for pensioners and older adults (groups with the lowest rates) slightly increase during mid-1983 and mid-1987; then they decrease more rapidly than they increased with a trough in mid-1989 for pensioners and mid-1990 for older adults. The trends for children and mature adults in Figure 5.2d are parallel, but the gap between them is relatively wide, compared to the difference between these rates in other types of FHSA. Out-migration rates for mature adults in Lambeth are 50% greater than those for children, an observation that can be expected on the basis that in big cities there are more mature adults without children and families are smaller (fewer children per household) compared to the less populous areas. In the long term out-migration rates for the latter two groups are stable; a high peak in mid-1987 and a trough in mid-1990 have been observed similarly in the previous FHSA types. Trends in out-migration rates for teenagers in Lambeth are similar to those in Newcastle; they doubled between mid-1990 and mid-1994. In Lambeth out-migration rates for adults are the highest among all age groups (except the period between mid-1992 and mid-1995); they fluctuate during the 1980s (trough in mid-1990) and they increase during the 1990s. Finally, out-migration rates for young adults are stable during the 1980s and they increase during the 1990s (peak in mid-1995).

In summary, young people (16-29) tend to be more mobile compared to families with children and older adults, and their mobility trends vary over time. Teenagers tend to have high out-migration rates in rural and less populous areas, and lower rates in rich counties and big cities. A significant increase in the out-migration rates for people aged 16-29 is observed during the 1990s and is probably connected to the increase of the rates of participation in higher education. Between 1990-91 and 1997-98, the total number of students in higher education (from data on parental domicile) in England and Wales increased from 1,509,203 to 2,681,391.

5.4 Out-migration clusters: Sex/Age disaggregation and temporal stability

In this section, spatial and temporal trends of sex/age disaggregated out-migration rates are discussed. This includes a presentation of spatial out-migration patterns for the 14 sex/age groups along with an examination of the temporal stability of the spatial patterns exhibited by these groups.

For this section, 14 sets of 14 maps were prepared for presenting k-means clusters of sex/age disaggregated out-migration data. Each set represents one of the 14 sex/age migrant groups and each map is a four-cluster classification of out-migration rates in a single year of observations (there are 14 years of observations). From these maps many pieces of information can be extracted: the centre of each cluster shows the temporal variation of the rates for each sex/age group; the colour denotes the spatial variation of out-migration rates; the map sets show how the spatial distribution changes over time. Comparison between sexes in the same age group can be also made.

However, one could argue that there is some unnecessary information presented here. This is because there are no dramatic changes over time. Furthermore, there are too many maps (196) and the reader is abstracted from the plethora of information. Thus, complete sets of maps are presented only for migrants aged 30 – 44, and one representative map (out-migration rates in 1997-98) for each of the remaining 12 sex/age groups. The motivation for this presentation was based on the lack of similar graphical presentation of sex/age disaggregated data elsewhere, as well as the importance for a reference to the modelling presented in the following chapter (Chapter 6).

5.4.1 Female and Male Children (aged 0 – 15)

Figure 5.3 shows out-migration rates for children across England and Wales in 1997-98. Actually, here we are measuring families with children; it is not likely for children to migrate without their parents. The FHSAs with the higher out-migration rates are mainly observed in central London (such as Kensington). During the 1980s the cluster centres for female children are slightly higher than those of male while the reverse is the case during the 1990s.

Although the use of the k-means algorithm results in different clusters for males and females the broad patterns of similarity exist. The spatial patterns are rather as expected. The highest out-migration rates are observed in the urban FHSAs in London, Manchester, Newcastle, and Birmingham (and some neighbour FHSAs). The lowest rates are observed in FHSAs in Wales, East England, East Midlands, North England, and South West England. There is no obvious temporal variation in the pattern of out-migration rates looking at the centres of the clusters and the spatial patterns. Average out-migration rates (these are the two middle clusters) are stable over time and together include the majority of the FHSAs.

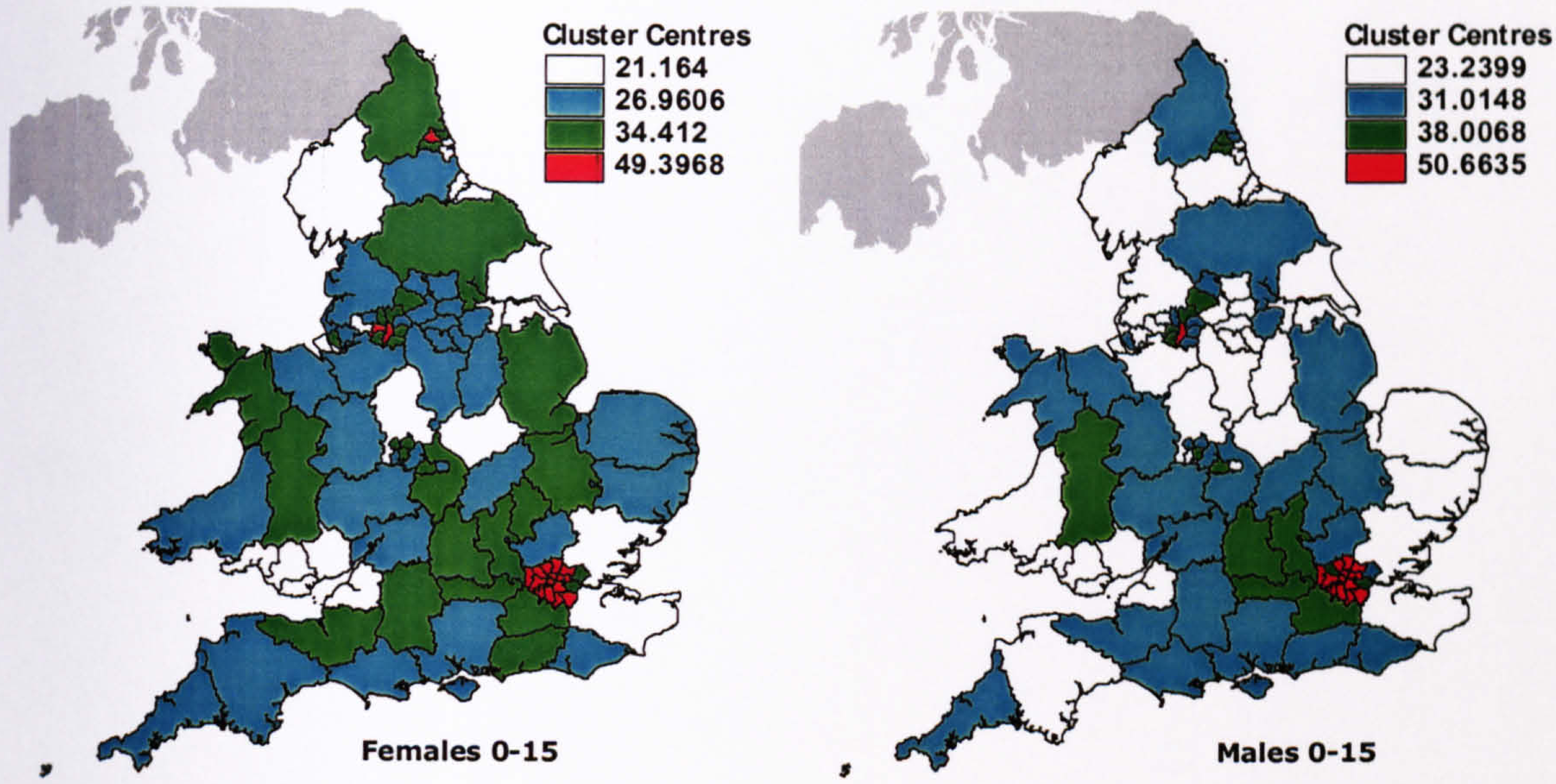


Figure 5.3. Spatial Patterns of out-migration rates for females and males aged 0–15 (1997/98)

5.4.2 Female and Male Teenagers 16 – 19

Out-migration rates for teenagers in 1997-98 are shown in Figure 5.4. For teenagers there are two apparent trends: the dissimilarities between the two sexes and the temporal variation. Out-migration rates for female teenagers are consistently higher than those of male teenagers for all time periods. During the 1980's the centres of the k-means clusters for teenagers are stable, but during the 1990's they increase by almost 100% (mid-1994). After a test comparison between male and female teenager rates, in 80% of the FHSAs in mid-1986 females have higher out-migration rates than males.

The spatial patterns for teenagers are very interesting. The highest out-migration rates for female teenagers are observed in FHSAs in rural areas of Wales and North England, as well as in some urban districts of West Midlands, –shire counties Northwest of London, the South, and some of the outer FHSAs in London. In 1997-98, the highest out-migration rates for male teenagers are observed in Somerset, Berkshire, Lincolnshire, Shropshire, Powys, and Wiltshire. The highest out-migration rates for female teenagers appear in Somerset, Powys, Surrey, Richmond and Kingston, Lincolnshire, and North Yorkshire where male teenager rates are not always the highest. It is interesting that the out-migration rates for female teenagers in Bromley (London) in 1997-98 is 12.83% of the population whereas for males it is 7.98%. This suggests that females are more likely to leave London while going for studies, work or forming a family, compared to males, even if their city offers working and educational opportunities.

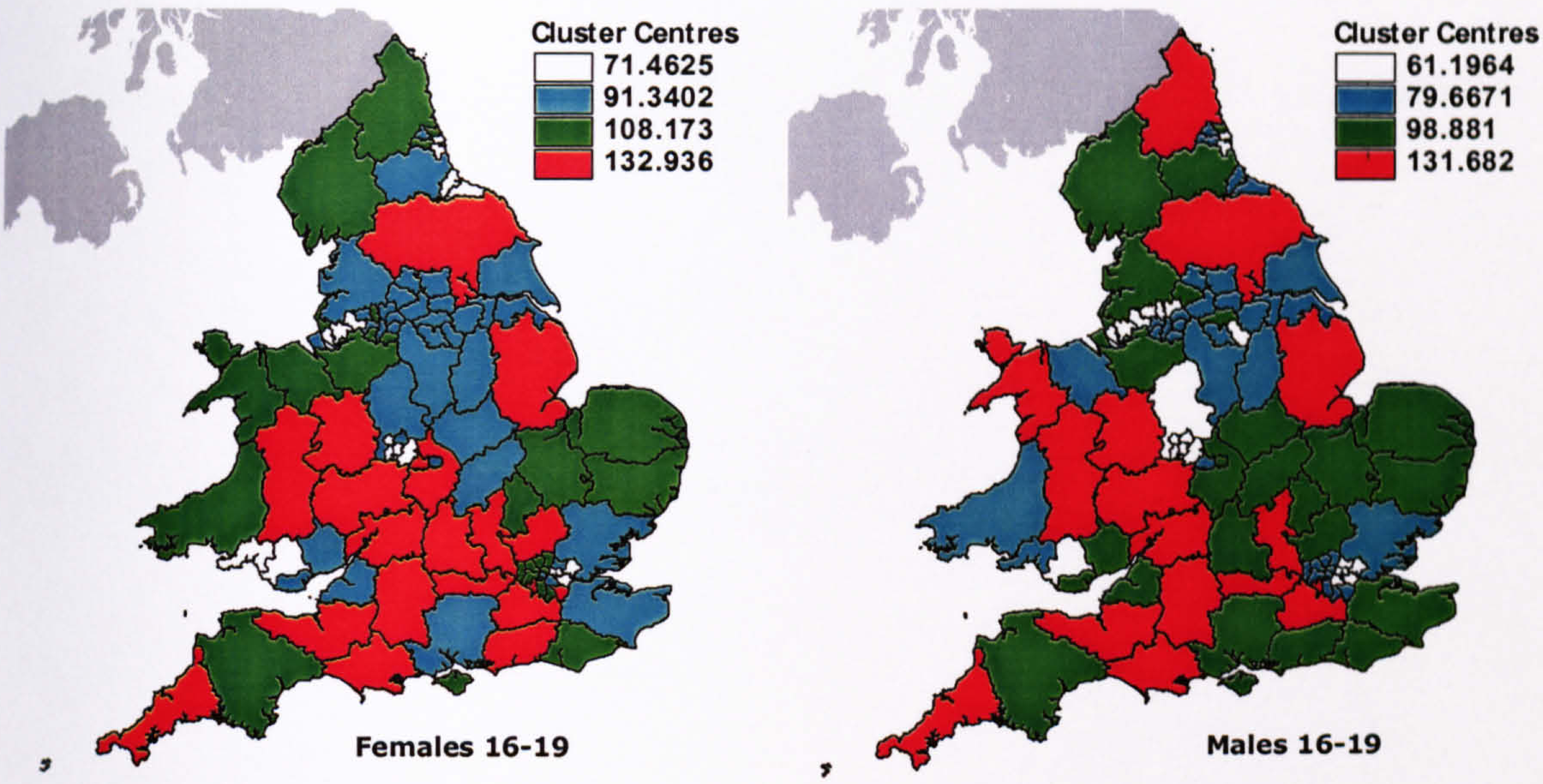


Figure 5.4. Spatial Patterns of out-migration rates for females and males aged 16–19 (1997/98)

Out-migration rates in the second high centre cluster (coloured green) for female teenagers are mainly observed in FHSAs in London, Manchester and Newcastle Metropolitan areas (excluding Manchester and Newcastle cities), East Wales, Cumbria, Northumberland, Devon, Isle of Wight, East Sussex and some FHSAs in East Midlands. Trends are different for male teenagers. In the latter group, high out-migration rates are observed in North Tyneside, Cumbria, Durham, Avon, Coventry, FHSAs in the Southeast (excluding London) and East Midlands, South Glamorgan and Gwent. FHSAs with lower out-migration rates for male teenagers include London, West Midlands, Leeds and neighbour FHSAs, and FHSAs in Southwest Wales. Similar is the spatial pattern for female teenagers with lower out-migration rates observed in Northeast and North West England, West Midlands and South Wales, but not in London.

5.4.3 Young Female and Male Adults 20 – 24

The next age group under investigation is young adults for whom out-migration rates in 1997-98 are presented in Figure 5.5. An immediate observation is that there are similarities between the sexes in terms of spatial patterns, but the out-migration rates for females are higher than those for males. The rates increase in the long term, but spatial clusters are rather stable over time.

Indeed, in 91% of the FHSAs out-migration rates for young female adults are higher than those of young male adults in mid-1993. Note that the rates for males have been adjusted, and are higher than those counted from the NHSCR. Areas where the highest out-migration rates have been observed (in 1997-98) for young female adults include FHSAs in London (e.g., Camden and Islington), Newcastle, Manchester, Dyfed, Surrey, Oxfordshire, South Glamorgan, North Yorkshire, Salford, Sheffield, Coventry, Liverpool and Leeds; in descending order. The highest out-migration rates for young adult males are observed in some FHSAs in London (e.g., Camden and Islington), Newcastle, Oxfordshire, Surrey, North Yorkshire, Coventry, Manchester, Gwyndd, Salford, Buckinghamshire, South Glamorgan, Sheffield, Dyfed and Liverpool; in descending order. Low out-migration rates for both sexes are mainly observed in FHSAs in Manchester-Leeds Metropolitan areas, Birmingham Metropolitan area, Gwent, Mid Glamorgan, Wiltshire, and East England (Norfolk, Suffolk, Essex and Kent).

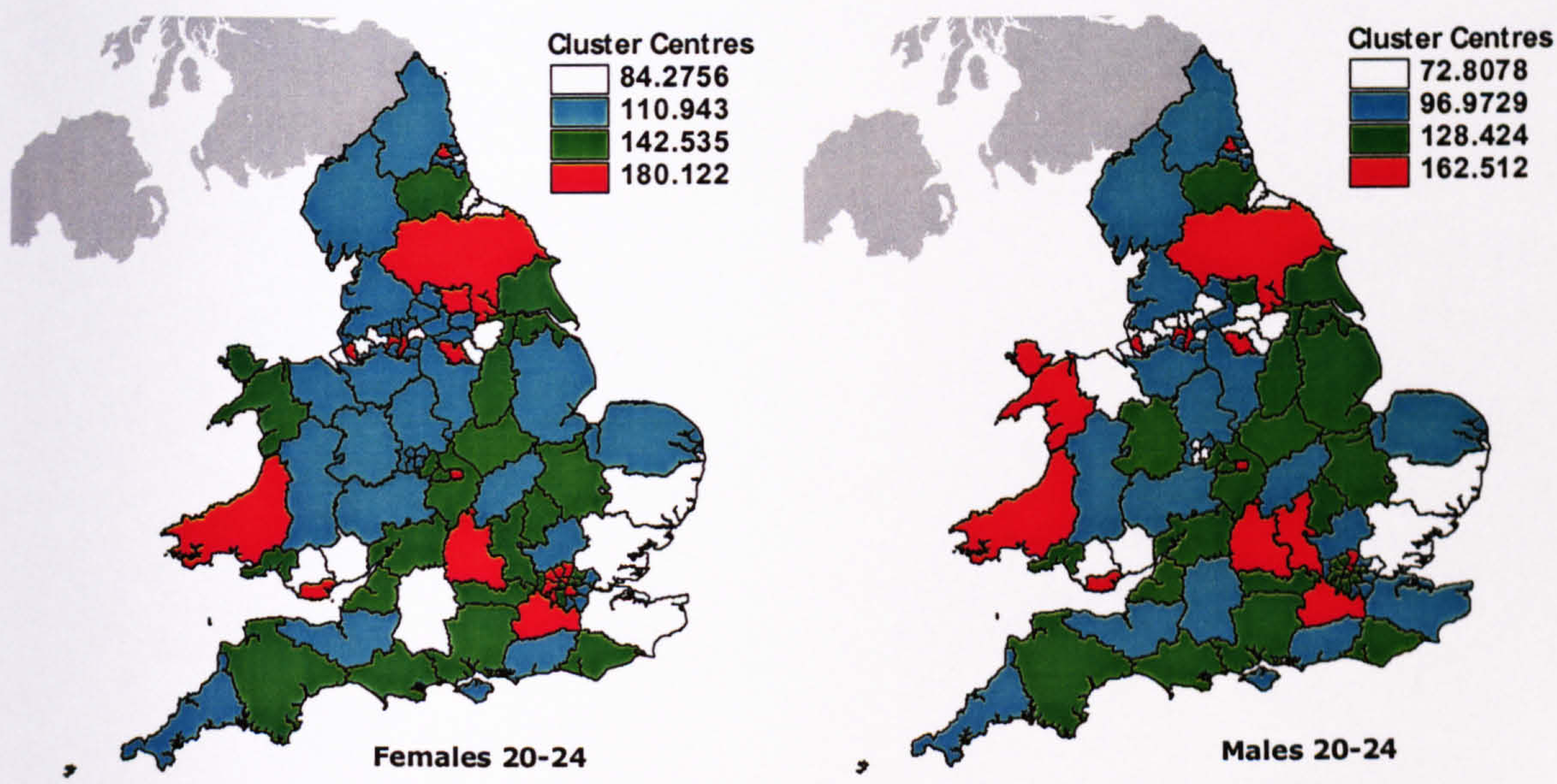


Figure 5.5. Spatial Patterns of out-migration rates for females and males aged 20–24 (1997/98)

5.4.4 Female and Male Adults 25 – 29

Out-migration rates for adults are high compared to the other age groups. However, when clustered they form big clusters with low rates and small clusters with high rates, a trend that distinguishes this age group with the remaining (Figure 5.6). There are no great differences between the sexes in terms of spatial patterns; however, rates for males are higher than those for females, something that is not clear from the maps. In 1991-92 for example in only 33% of the FHSAs rates for females are higher than those of males. The two clusters of low out-migration rates include most of the 98 FHSAs and the largest cluster is the one with the lowest centre. Clusters of high rates include few areas and the cluster with the highest centre includes areas with the extreme rates.

In 1997-98 FHSAS with the highest out-migration rates for adults include most FHSAs in London, Newcastle and Manchester. High rates are observed in Surrey, Coventry, Trafford, Solihull, S alford, Liverpool, Oxfordshire, South Glamorgan, Wolverhampton and Solihull. FHSAS with the lowest rates are observed in FHSAs in the North England, in all the east coast and surprisingly the Greater Metropolitan areas in a line from Liverpool to Doncaster excluding the inner cities (Manchester, Leeds, Trafford and Coventry).

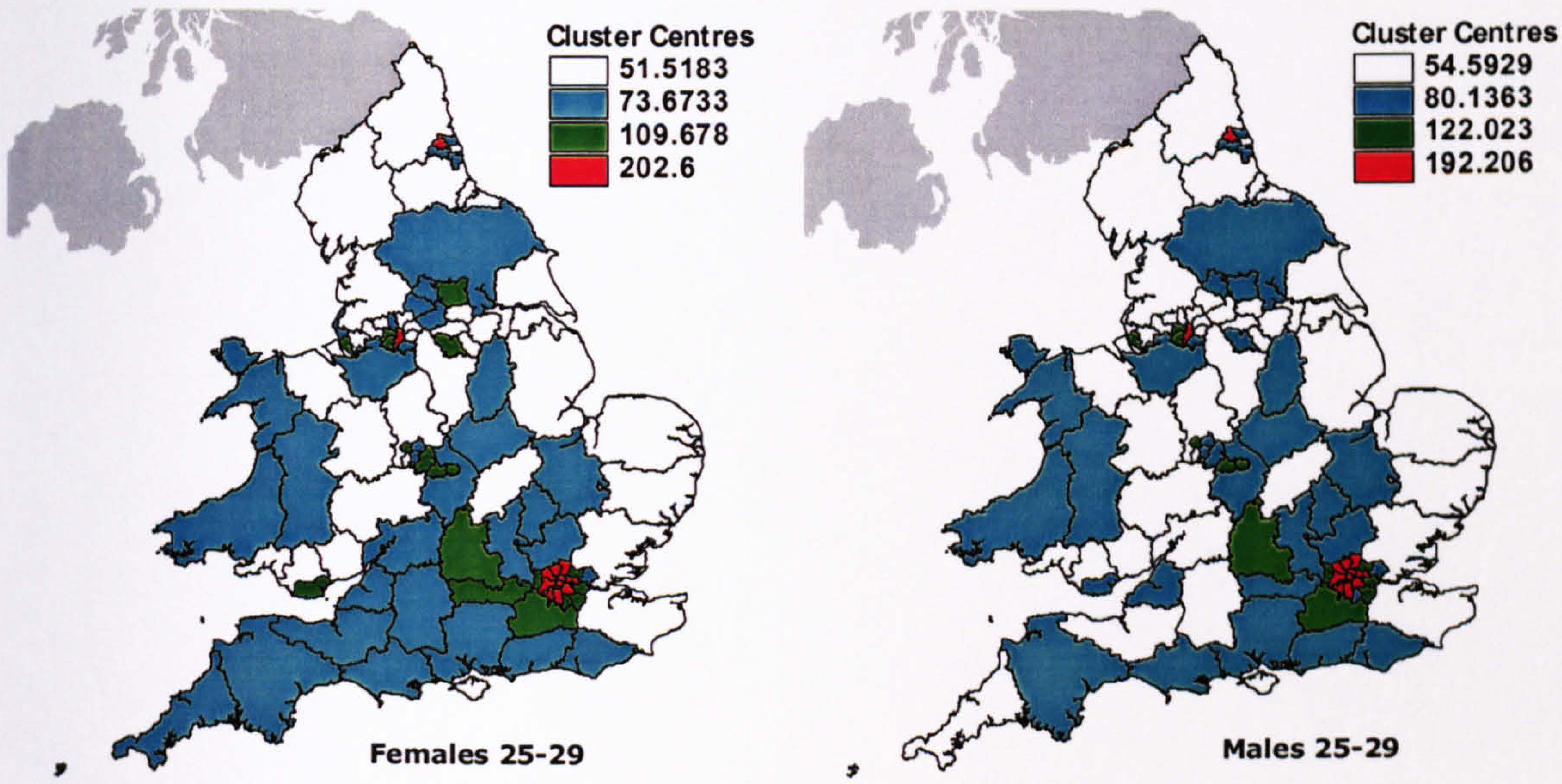


Figure 5.6. Spatial Patterns of out-migration rates for females and males aged 25–29 (1997/98)

5.4.5 Mature Female and Male Adults 30 – 44

Mature adults (aged 30 – 44) is the largest group, in terms of migration counts. Most of the mature adults are in a life stage of raising a family. Mature adults and children (aged 0-15) are likely to share similar trends in out-migration rates. This is because the children migrate with their parents. However, the spatial and temporal patterns, Figures 5.7i-ii, suggest that this is not the case. Similarly to adults, the two clusters, those including FHSAs with the lowest and lower out-migration rates are the largest (have many members), while the two clusters of FHSAs with high and the highest out-migration rates are the smallest (have few members).

Figure 5.7i. Spatial Patterns of out-migration rates for females aged 30–44 (1984/85 – 1997/98)

There is a clear North South divide during the mid-1980s which somewhat changes later. In the mid-1980s, high out-migration rates are observed only in FHSAs in London and Manchester. During the late 1980s and 1990s, Newcastle, FHSAs in West England and West Midlands are added to FHSAs with high out-migration rates. Thus, the North-South division becomes less clear. It is important to note that like children, there is no great temporal variation in both spatial patterns and cluster centres. There is little, if any, variation across sexes in terms of spatial clusters. However, in mid-1993 in all 98 FHSAs, out-migration rates for males are higher than those for females. The latter can be observed also by looking at the cluster centres in the two sets of maps.

Figure 5.7ii. Spatial Patterns of out-migration rates for males aged 30–44 (1984/85 – 1997/98)

5.4.6 Older Female and Male Adults 45 – 59

Out-migration rates for older adults are presented in Figure 5.8. There are similarities with mature adults both in terms of the relationship between sexes and the spatial patterns of out-migration rates. Rates for males are overall higher than rates for females (in mid-1987 in 96 FHSAs male rates are higher) and there is little temporal variation. However, there are two distinctive trends: one is that out-migration rates are low compared to other age groups; and the FHSA membership in each cluster is balanced. There is a North-South divide, which is clearer during the 1990s. The highest out-migration rates for older adults are observed in FHSAs in London and in Manchester City. High out-migration rates are also observed in Newcastle, Trafford, Solihull, West Sussex, Surrey, Berkshire, Buckinghamshire and Hertfordshire. There is a nice spatial pattern in the South skewed to the Southwest, which looks like a set of common-centre circular buffers around central London, in which FHSAs further from London have lower out-migration rates. The latter pattern is more obvious in late 1980s and early 1990s. The low out-migration rates are mainly observed in rural and remote areas as well as the Northern England FHSAs. It is obvious that for older people (aged 30 and over) there is a tendency to leave mainly big cities and not rural areas.

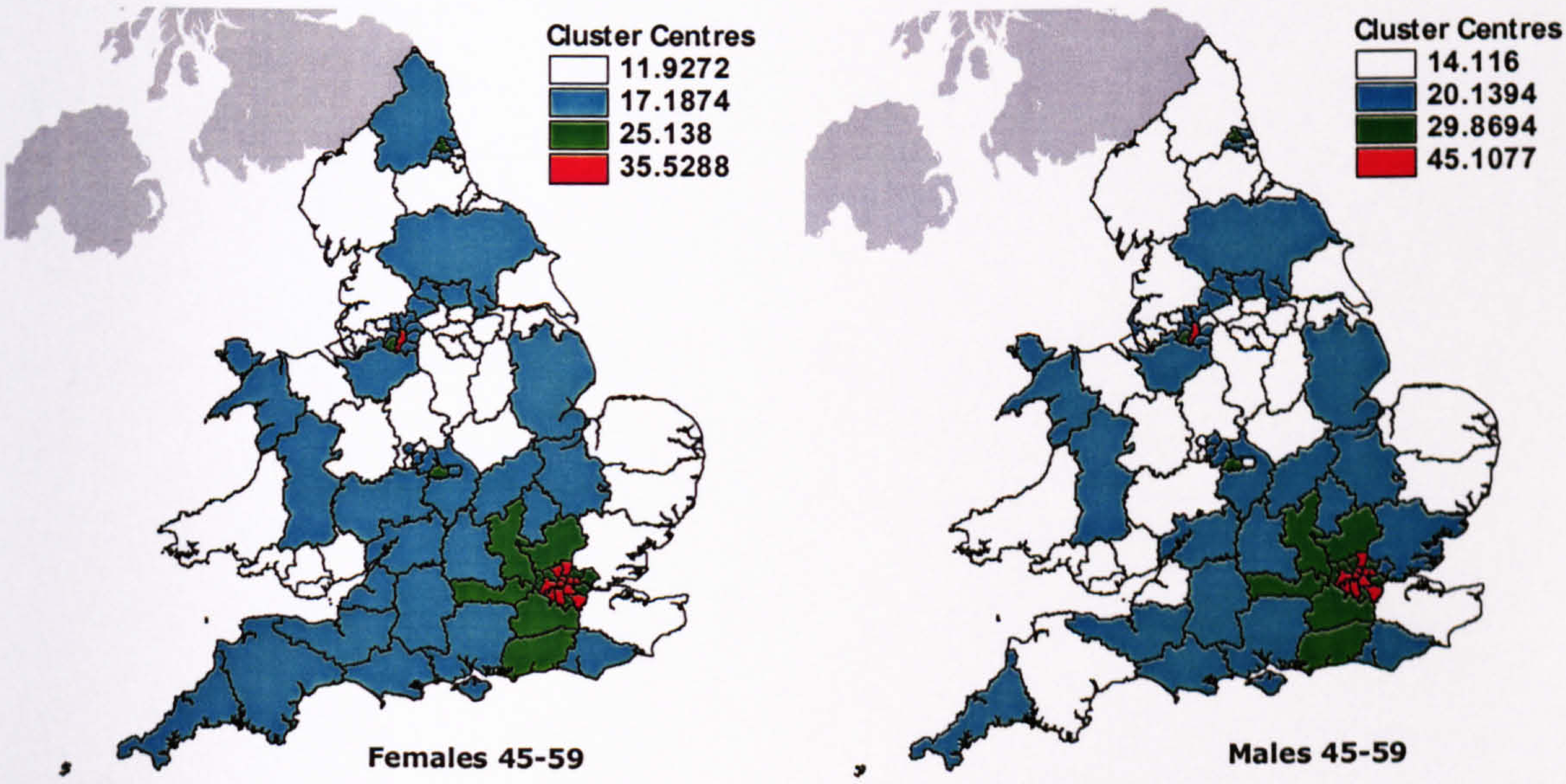


Figure 5.8. Spatial Patterns of out-migration rates for females and males aged 45–59 (1997/98)

5.4.7 Female and Male Pensioners 60 and over

The last age group under examination is that of people aged 60 and over (mainly retired). There are great similarities to the previous age group in terms of temporal variation, spatial patterns but not in terms of differences between sexes. One reason for out-migration rates for females to be higher than those of males (in 68% of the cases in mid-1997) is that the latter has not been adjusted as in the other age groups (males aged 16-59). It is also possible that female pensioners will look for care after their husbands die, thus will move to be nearer their family or into a care house increasing the migration volumes of their group. There is a clear North-South divide, mainly observed during the 1990s and there is a tendency for elderly to leave big cities (Newcastle, Manchester, Birmingham, London).

The spatial pattern of buffers around London discussed for older adults has also been observed in the case of elderly, and it is smoother. Another clear trend is the temporal variation. The spatial patterns suggest a fall during the late 1980s and the early 1990s and a recovery of out-migration rates in the late 1990s. Such a trend exists, but is less obvious in the charts provided in the beginning of Section 5.2.

Figure 5.9 shows the out-migration rates of pensioners in 1997-98. The highest out-migration rates in a descending order are in FHSAs in London, Manchester and Surrey. High out-migration rates are observed in many FHSAs but the image may be misleading because of a low cluster mean. The lowest out-migration rates are in FHSAs in North England, Midlands and Wales. The smallest rate for female pensioners is in Cleveland (6.51 per thousand population) and for male pensioners in West Glamorgan (5.85 per thousand population).

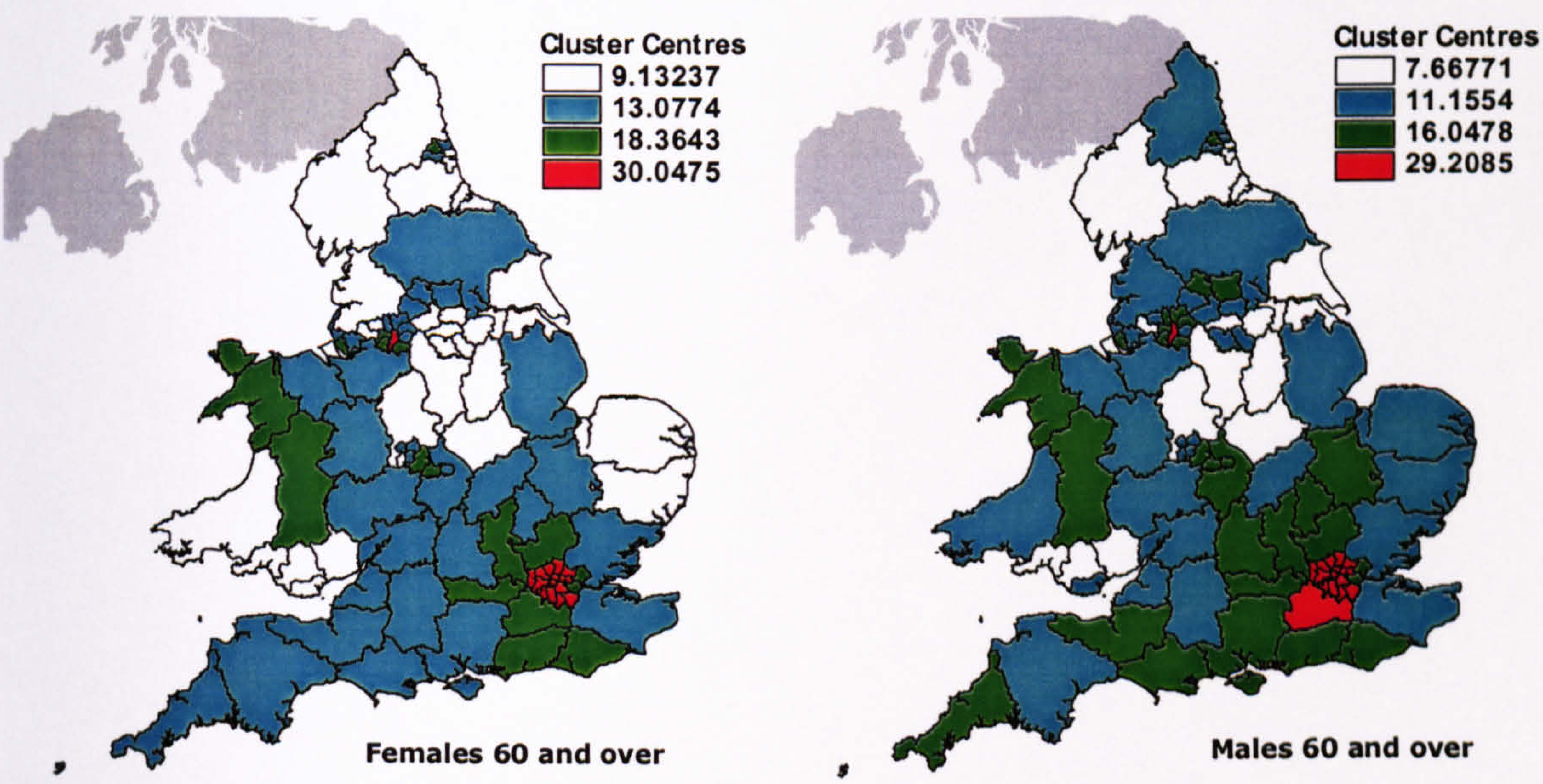


Figure 5.9. Spatial Patterns of out-migration rates for females and males aged 60 and over (1997/98)

5.5 Exploratory spatial data analysis (ESDA) and local statistics

In this section, some methods that examine spatial dependence or spatial autocorrelation as well as a geographically weighted local mean are presented. This includes a quick review of the available methods and a more analytical discussion of global and local (where appropriate) versions of Moran's I, Geary's c and Getis' G statistics. The geographically weighted local mean (using GWR) is presented last. These methods have been applied to some of the migration data available here and are presented in the following sections of this chapter.

In the previous sections a simple classification approach was used to help in identifying spatial clusters in out-migration rates. However, this is a very simplistic method that does not incorporate any use of space, i.e. coordinates, distances and contiguity. There is an extensive literature on more sophisticated methods of identifying such spatial patterns. Most of these methods have been developed to be applied in large datasets, however, they can be also applied in the migration data available here.

These methods belong to broad research areas which can be identified in the literature terminology as point-pattern analysis, spatial autocorrelation, univariate analysis, analysis of spatial association, local indicators of spatial association (LISA), analysis of spatial dependence only to name the dominant terms. Reviews of global and local methods of spatial dependency / autocorrelation include Getis (1991), Fotheringham and Rogerson (1993), Getis and Ord (1996), Anselin (1995; 1998), Fotheringham and Brunsdon (1999), and Fotheringham et al. (2002a).

5.5.1 Moran's I

Moran's I is one of the oldest statistics used to examine spatial autocorrelation. Cliff and Ord (1973, 1981) present a comprehensive work on spatial autocorrelation. They present their version of Moran's *I* based on Moran's (1948) calculation of the *moments* and Moran's (1950) first coefficient. Moran's first version of *I* is:

$$I = \frac{n}{2A} \frac{\sum_i \sum_j \delta_{ij} z_i z_j}{\sum_{i=1}^{n} z_i^2} \tag{5.1}$$

where $z_i = x_i - \bar{x}$, $\bar{x}$ is the mean of x, A is the total number of joins in the system and δ_{ij} is the weight. The formula of Getis and Ord (1973, 1981) to calculate Moran's *I* is:

$$I = \frac{n\sum_i \sum_j w_{ij} z_i z_j}{W\sum_{i=1}^{n} z_i^2} \tag{5.2}$$

where $z_i = x_i - \bar{x}$, $\bar{x}$ is the mean of *x*, and $W = \sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}$. This statistics can be interpreted by the evaluation of the *standard normal deviate* that is computed as Z = [*I* – E(*I*)]/σ(*I*), where E(*I*) is the *expected I* and σ(*I*) is the variance of *I* (Cliff and Ord, 1973).

A discussion of local *I* is provided in Anselin (1995) (Getis and Ord, 1996). The local *I* is defined as follows:

$$I_i = \frac{z_i}{s^2}\sum_{j=1}^{n} w_{ij} z_j, j \neq i \tag{5.3}$$

where z_i and z_j are deviations from the mean ($z_i = x_i - \bar{x}$; $z_j = x_j - \bar{x}$) and $s^2 = \sum_{k=1}^{n}(x_k - \bar{x})^2 / n$. The number of non-zero weights equals the number of neighbours within a selected distance *d* around point *i* in space. The evaluation of local Moran's *I* (I_i) is analogous to that of the global Moran's *I* presented above, i.e. the evaluation of the $Z(I_i) = (I_i - E[I_i]) / \sqrt{\mathrm{Var}[I_i]}$. *For the I_i, a positive value indicates spatial clustering of similar values (either high or low), and negative values a clustering of dissimilar values (for example a location with high values surrounded by neighbors with low values), as in the interpretation of the global Moran's I* (Anselin, 1995, pp. 102 – 103).

Table 5.1. A comparison of various spatial models and the cross-product statistic

Model	W_{ij}	Y_{ij}	Restrictions W_{ij}	Restrictions Y_{ij}	Scale
Cross-product statistics					
$\Gamma = \sum\sum W_{ij} Y_{ij}$	W_{ij}	Y_{ij}	none	none	none
$\Gamma = \sum\sum W_{ij} Y_i$	W_{ij}	Y_i	none	none	none
Spatial autocorrelation models					
Join count					
$BB = \frac{1}{2}\sum\sum W_{ij} x_i x_j$	W_{ij}	$x_i x_j$	0/1	0/1	½
$BW = \frac{1}{2}\sum\sum W_{ij} (x_i - x_j)^2$	W_{ij}	$(x_i - x_j)^2$	0/1	0/1	½
$WW = \frac{1}{2}\sum\sum W_{ij} (1 - x_i)(1 - x_j)$	W_{ij}	$(1 - x_i)(1 - x_j)$	0/1	0/1	½
Moran's					
$I = \frac{n\sum\sum W_{ij}(x_i - \bar{x})(x_j - \bar{x})}{W\sum(x_i - \bar{x})^2}$	W_{ij}	$(x_i - \bar{x})(x_j - \bar{x})$	none	none	$\frac{n}{W\sum(x_i - \bar{x})^2}$
Geary's					
$c = \frac{(n-1)\sum\sum W_{ij}(x_i - x_j)^2}{2W\sum(x_i - \bar{x})^2}$	W_{ij}	$(x_i - x_j)^2$	none	none	$\frac{(n-1)}{2W\sum(x_i - \bar{x})^2}$
Semi-variance					
$\gamma = \frac{1}{2}\sum_{i=h}^{n-h}\sum_{j=1+h}^{n} W_{ij}(x_i - x_j)^2$	W_{ij}	$(x_i - x_j)^2$	1	none	½
Second-order					
$K(d) = \frac{\sum\sum W_{ij}(d) x_i x_j}{(\sum x_i)^2 - \sum x_i^2}$	$W_{ij}(d)$	$x_i x_j$	0/1	positive	$\frac{1}{(\sum x_i)^2 - \sum x_i^2}$
Getis model					
$G_i(d) = \frac{\sum_j W_{ij}(d) x_i x_j}{\sum_j x_i x_j}$	$W_{ij}(d)$	$x_i x_j$	0/1	positive	$\frac{1}{\sum_j x_i x_j}$
Spatial interaction models					
General gravity					
$T_{ij} = k x_i^{\alpha} x_j^{\tau} W_{ij}^{-\beta}$	$W_{ij}^{-\beta}$	$x_i^{\alpha} x_j^{\tau}$	none	positive	k
Origin-specific, production constrained					
$T_{ij} = \frac{x_i x_j^{\alpha} W_{ij}^{-\beta}}{\sum_j x_j W_{ij}^{-\beta}}$	$W_{ij}^{-\beta}$	$x_i x_j^{\alpha}$	none	positive	$\frac{1}{\sum_j x_j W_{ij}^{-\beta}}$
General spatial models					
i-to-all-*j* model					
$G_i = \frac{\sum_j x_i x_j W_{ij}^{-\beta}}{\sum_j x_i x_j}$	$W_{ij}^{-\beta}$	$x_i x_j$	none	positive	$\frac{1}{\sum_j x_i x_j}$
i-to-*j* model					
$G_{ij} = \frac{x_i x_j W_{ij}^{-\beta}}{x_i x_j}$	$W_{ij}^{-\beta}$	$x_i x_j$	none	positive	$\frac{1}{x_i x_j}$

Note: BB, black-black joins; BW, black-white joins; WW, white-white joins; In Geary's c: $W = \sum\sum W_{ij}$

Source: Getis, 1991, p. 1271.

5.5.2 Geary's c

Geary's c statistic was introduced by Geary (1954) as the *contiguity ratio c* is originally defined as follows:

$$c = \frac{(n-1)}{2K_1} \frac{\sum'_{t \neq t'} (z_t - z_{t'})^2}{\sum_t (z_t - \bar{z})^2} \qquad (5.4)$$

where n is the number of counties in a administrative boundary map, z_t is the measure of the tth county, with a number of connections k_t, $K_1 = \Sigma k_t$, Σ is the sum over all counties and Σ' is the sum over contiguous counties. A more recent form of the above formula is:

$$c = \frac{(n-1)}{2\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}} \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(x_i - x_j)^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \qquad (5.5)$$

where w_{ij} is the weight that will be non-zero within distance d of each point (county centroid) i in the system, and thus, will account for the contiguity. A local version of c (c_i) presented in Anselin (1995). For each observation i

$$c_i = \frac{n}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \sum_{j=1}^{n} w_{ij}(x_i - x_j)^2 \qquad (5.6)$$

In the literature, global and local Geary's c is discussed and used along with global and local Moran's I most of the time (Cliff and Ord, 1973; 1981; Anselin, 1995; Getis, 1991; Getis and Ord, 1996).

5.5.3 Getis G

A family of statistics, G, introduced in Getis and Ord (1992) provides means for measuring spatial association in spatially distributed variables. There are two main statistics and their variations; the global or general G statistic:

$$G(d) = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(d) x_i x_j}{\sum_{i=1}^{n}\sum_{j=1}^{n} x_i x_j}, j \neq i \qquad (5.7)$$

and the local (point specific) G statistic:

$$G_i(d) = \frac{\sum_{j=1}^{n} w_{ij}(d)x_j}{\sum_{j=1}^{n} x_j}, j \neq i \tag{5.8}$$

where d is the distance that defines a neighbourhood area around each point i, w_{ij} is the weight for each point j in space other than i (the weight is usually *1* for points within the neighbourhood area and *0* for the remaining points) and x_i, x_j is the value of the spatially distributed variable X at point i, j respectively. This variable (X) needs to have a natural origin and positive values, a criterion that is satisfied in the case of gross and net migration and migration rates. The characteristics of the G_i statistics are shown below (Table 5.2).

Table 5.2. Characteristics of G_i Statistics

	j not equal to i	j may equal i
Statistic	$G_i(d)$	$G_i^*(d)$
Expression	$\frac{\sum_j w_{ij}(d)x_j}{\sum_j x_j}$	$\frac{\sum_j w_{ij}(d)x_j}{\sum_j x_j}$
	$W_i = \sum_j w_{ij}(d)$	$W_i^* = \sum_j w_{ij}(d)$
Definitions	$Y_{i1} = \frac{\sum_j x_j}{(n-1)}$	$Y_{i1}^* = \frac{\sum_j x_j}{n}$
	$Y_{i2} = \frac{\sum_j x_j^2}{(n-1)} - Y_{i1}^2$	$Y_{i2}^* = \frac{\sum\sum_{ij}(x_i x_j)^2}{n} - (Y_{i1}^*)^2$
Expectation	$W_i/(n-1)$	W_i^*/n
Variance: Var $G_i(d)$	$\frac{W_i(n-1-W_i)Y_{i2}}{(n-1)^2(n-2)Y_{i1}^2}$	$\frac{W_i^*(n-W_i^*)Y_{i2}^*}{n^2(n-1)(Y_{i1}^*)^2}$

Source: Getis and Ord, 1992, p. 192.

Advances of these first attempts have been made by removing the restrictions in the variable (natural origin and positive values), by allowing nonbinary weighting schemes to be used and by improving the statistical inference and testing of significance (Ord and Getis, 1995). Furthermore, Ord and Getis (2001) suggest a new statistic (O_i) that tests for local spatial autocorrelation in the presence of global autocorrelation.

The $G_i(d)$ statistic measures the degree of association that results from the concentration of weighted points (or area represented by a weighted point) and all other weighted points included within a radius of distance d from the original weighted point (Getis and Ord, 1992, p. 190). If point i is included in the calculation of the above statistics the result is the $G^*(d)$ and the $G^*_i\ (d)$ statistic respectively. There are also significance tests for these statistics.

In order to identify clusters of high or low values it is necessary to calculate a statistic $Z_i = \{G_i(d) - E[G_i(d)]\} / \sqrt{\operatorname{Var} G_i(d)}$ from the local $G_i(d)$. *In typical circumstances, the null hypothesis is that the set of x values within d of location i is a random sample drawn without replacement from the set of all x values. The estimated $G_i(d)$ is computed from equation (5.8) using the observed x_j values. Assuming that $G_i(d)$ is approximately normally distributed, when Z_i is positively or negatively greater than some specified level of significance, then we say that positive or negative spatial association obtains. A large positive Z_i implies large values of x_j (values above the mean x_j) are within d of point i. A large negative Z_i means that small values of x_j are within d of point i* (Getis and Ord, 1992, p. 192). The global *G* statistic gives a feeling of the existence of a spatial cluster, but not where this cluster is. In the recent literature (Getis and Ord, 1996) the definition of the $G_i(d)$ statistic matches the Z_i statistic and the statistic of equation (5.2) the Getis' version of local mean.

The G statistics have been applied to raster data as well as to vector data. Examples of the former include spatial analysis of remotely sensed images (Getis, 1994) and of the latter analysis of variables (socioeconomic, health) using administrative boundary geography (Getis and Ord, 1992; Ord and Getis, 2001).

The *G* statistics are well connected with the Moran's *I* statistics and they are and could be used together. Getis and Ord (1996) suggest that if the number of observations is relatively small, as few as eight neighbours could be used to calculate local means for G without serious inferential error unless the underline distribution is very skewed.

As with many of such statistics it is necessary to see if there are implementations than make them easy to calculate and use. This author is aware of three implementations: A commercial version called SpaceStat provided by TerraSeer (SpaceStat, 2003); an SPSS macro available from Michael Tiefelsdorf's personal webpage (GeoStat, 2003); and a package of statistics implemented in R, called spdep (CRAN, 2003).

The former is available after a license is paid and is accompanied with a free ESRI ArcView extension to allow an easy management of spatial data. Tiefelsdorf's GlobalLocalMoran SPSS Macro is available with some sample data license free from his webpage. The software used here is the package spdep within R. The version 0.1-8 of spdep released on the 20th January 2003 along with its documentations was applied for calculating Geary's c as well as the global and local Moran's *I* and Getis *G* statistics.

The spdep Package is entitled: "Spatial dependence: weighting schemes, statistics and models" and its author is Roger Bivand (Roger.Bivand@nhh.no), with contributions by Nicholas Lewin-Koh (kohnicho@comp.nus.edu.sg), Michael Tiefelsdorf (tiefelsdorf.1@osu.edu) and Hisaji ONO (hi-ono@mn.xdsl.ne.jp). A description of the package as it is presented in its documentation follows. *A collection of functions to create spatial weights matrix objects from polygon contiguities, from point patterns by distance and tesselations, for summarising these objects, and for permitting their use in spatial data analysis; a collection of tests for spatial autocorrelation, including global Moran's I, Geary's C, Hubert/Mantel general cross product statistic, Empirical Bayes Index, and Getis/Ord G, local Moran's I and Getis/Ord G, saddlepoint approximations for global and local Moran's I; and functions for estimating spatial simultaneous autoregressive (SAR) models* (The spdep Package, documentation of version 0.1-8, 24-Jan-2003).

The package requires the copyright agreements for non-profit academic use as well as R, and it is license free for research within higher education. The reason this is used here, is not only that it is freely available and well supported, but also because the R software is a very powerful and convenient tool for analysing and visualising data using statistical techniques.

Anselin (1998) in his discussion of ESDA techniques also presents three practical implementations for visualisation of the results of such analysis: the ArcView/XGobi/XploRe link (Symanzik et al., 1997), the S+ArcView (Bao and Martin, 1997) link and the SpaceStat extension for ArcView (Anselin and Smirnov, 1998) mentioned above.

5.5.4 Local mean (GWR)

A recent development of techniques that can be used in exploratory data analysis are Geographically Weighted Local Statistics (Fotheringham et al., 2002a). These local descriptive statistics share the same idea as the Geographically Weighted Regression discussed above. The idea of a fixed and adaptive kernel, the bandwidth and the number of nearest neighbours are the same as in GWR. The point in space for which local descriptive statistics are calculated can be named as the *summary point*. What is interesting for this analysis is the geographically weighted mean (GWM) defined as

$$\bar{x}_i = \frac{\sum_j x_j w_{ij}}{\sum_j w_{ij}} \tag{5.9}$$

where the weights w_{ij} are calculated using a distance weighted scheme such as the bi-square function of Equation 4.22. More details on the methodology and alternative Geographically Weighted Local Statistics (GWLS) can be found in Fotheringham et al. (2002a) in Chapter 7. To date there are plans for implementing these to GWR software, but this new version has not been released and, thus, can not be used here. Alternatively, an R version of GWR and GWLS provided by Chris Brunsdon is used for calculating and visualising the GWM (see below).

5.6 Out-, In- and Net migration

In this section, I apply the statistics described in the previous section to out-migration rates of older male adults[2] for a more advanced exploratory analysis. I also explore out-, in- and net migration over time for Newcastle FHSA and London (16 FHSAs combined). For the latter, I have developed a new means of visualising migration rates, which I call a *heat map*. Finally, I identify FHSAs that are net population gainers or losers over time (1990/91 – 1996/97). In all the empirical examples below the weight used is scaled so that $\sum_j w_{ij} = 1$ and $W = \sum_i \sum_j w_{ij} = n$.

5.6.1 Out-migration of males 45 – 59 years old

Here the distribution of out-migration rates of males 45 – 59 years old is examined and the local statistics discussed above are calculated. The results have been visualised to allow a better interpretation. Table 5.3 shows basic descriptive statistics of out-migration rates of males aged 45–59 from 1984-85 to 1997-98. The skewness and kurtosis give a feeling of the normality of the distribution. Figure 5.10 shows the histograms of out-migration rates in 1987-88 and 1997-98 along with the normal curve.

Table 5.3. Some descriptive statistics of out-migration rates of males aged 45–59

Statistics / Year	84-85	85-86	86-87	87-88	88-89	89-90	90-91	91-92	92-93	93-94	94-95	95-96	96-97	97-98
Mean	18.737	20.108	22.165	24.739	23.145	17.802	17.154	18.449	17.623	18.477	19.109	19.212	20.552	21.619
Std. Deviation	8.389	9.316	10.461	11.431	10.357	8.235	8.176	8.980	8.743	9.333	9.706	9.493	9.183	9.887
Skewness	2.191	1.566	1.468	1.125	1.413	2.213	1.856	1.895	2.011	2.102	1.812	1.998	1.505	1.502
Kurtosis	7.771	3.326	2.893	0.632	2.229	6.719	3.857	4.088	4.615	5.704	3.350	4.710	2.141	1.868

The out-migration rates are positively skewed in all cases. The skewness values observed for these data suggest that the normality assumption is questionable, since skewness is 0 for any symmetric distribution (such as the normal distribution). The data for 1987-88 suggest the best possible normality whereas those for 1989-90 the worse. To get a better feeling for the distribution the kurtosis values have been calculated. Kurtosis is 0 for normal distributions. A positive kurtosis suggests a heavy tailed distribution (Kleinbaum et al., 1988). Here, the kurtosis is positive, but not very high to suggest a very heavy tailed distribution. Again, the data for 1987-88 is as symmetric as possible, and should give more reliable statistics that assume normality in the distribution (as most of the statistics discussed here do).

[2] This sex/age group was randomly selected as an example for the use of alternative methods of spatial analysis for identifying spatial clusters of out-migration rates.

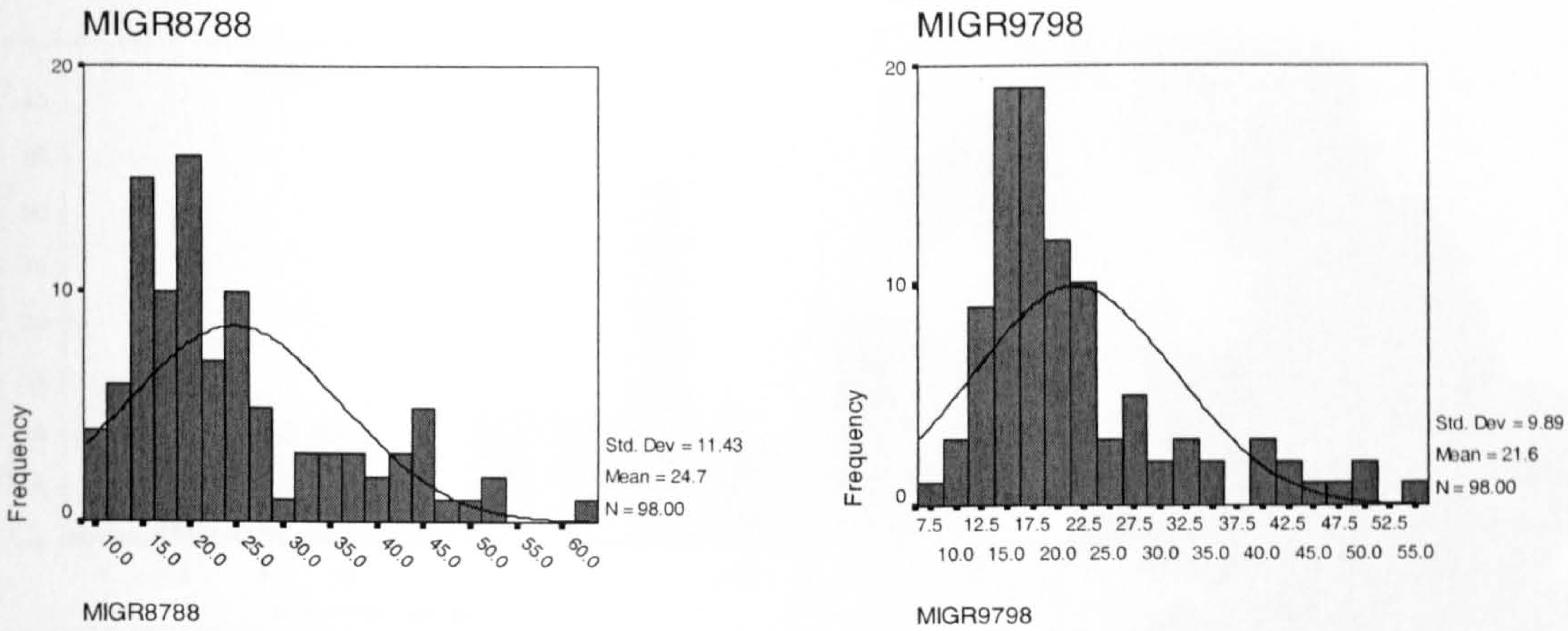


Figure 5.10. Histogram of out-migration rates for males aged 45–59 in 1987-88 and 1997-98

The first local statistic is the Local G. Using R (CRAN, 2003) the following code was used to calculate Local *G* for out-migration rates for males aged 45–59 in 1997-98 using a distance of 100 kilometres.

```
#Read Data, find the number of neighbours and calculate local G using default options
mydata <- read.csv("C:/WORK/PHD/lstat/g/out-mgr-m4559.csv")
xycoords <- cbind(mydata$x,mydata$y)
nb10 <- dnearneigh(xycoords, 0, 10)
G10 <- localG(spNamedVec("adjMigr9798",mydata),nb2listw(nb10))

#Write results to local disk
lg10 <-cbind(G10)
out.data <- data.frame(ID=mydata$MODZONECOD,LocalG9798=lg10)
write.table(out.data,file="localg10.csv",sep=',',row.names=F)
```

In order for the software to run properly, the coordinates, measured in meters, were scaled down by a factor of 10,000; thus, a distance of 1 is a real distance of 10 km. The function *dnearneigh* finds the neighbours that are between 0 and 100km from each point in the sample data (the x, y coordinates for each FHSA are those of its centroid). The histogram of the number of neighbours for all 98 FHSAs is presented in Figure 5.11. There is no FHSA without any neighbour; 15 FHSAs have 8 or less neighbours and 10 FHSAs have more than 30 neighbours. In the same figure, the resulted Local *G* is also presented. Two clusters can be clearly identified: a cluster of high positive values of G_i in London and the South East and a cluster of high negative values of G_i in the North West and Yorks and Humberside. The former suggests a spatial cluster of high out-migration rates and the latter a spatial cluster of low out-migration rates.

The cluster of high out-migration rates can be confirmed with k-means clustering (Figure 5.8), but not the cluster of low out-migration rates. A reason for the latter is that Local *G* with a distance of 100km is a smoothing function and the high out-migration rates in Manchester are averaged because of very low out-migration rates in nearby FHSAs.

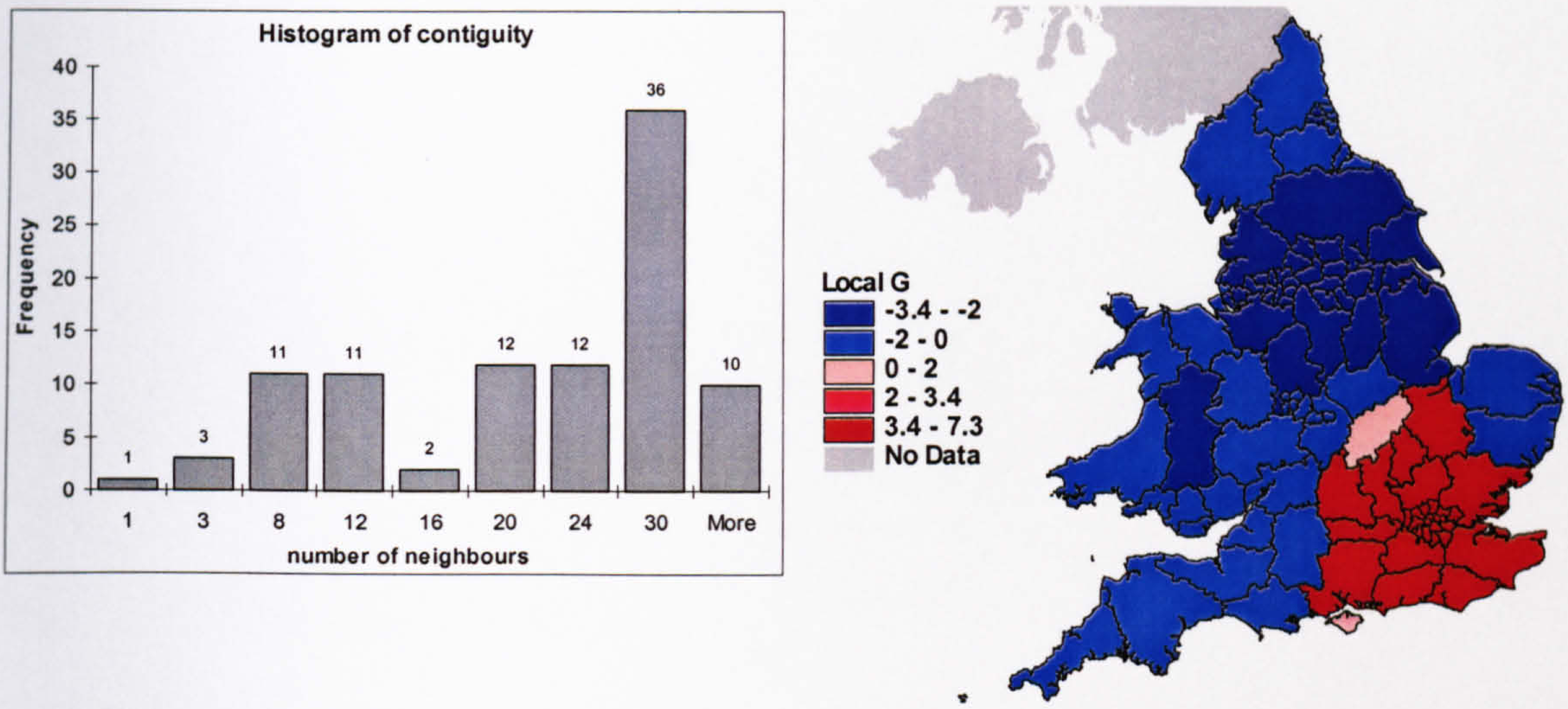


Figure 5.11. Histogram of contiguity and Map of $G_i(10)$ of Local G analysis (males aged 45 – 59 in 1997-98)

To get a better feeling of where and what kind of spatial cluster may be for this dataset, the local Moran's I was also calculated using the same neighbours as in the case of Local *G*. The results are shown in Figure 5.12. The high positive values of local I suggest a high similarity of out-migration rates in the corresponding FHSAs. Thus, the cluster of high out-migration rates in the South East is confirmed. There is also some confirmation of a cluster of similar values in Yorks and Humberside. However, what the local Moran's I identifies are some FHSAs with negative, in some cases high negative, values suggesting heterogeneity in the out-migration in the corresponding areas. Such examples include Manchester, Kent and Essex. The latter have very dissimilar out-migration rates with their neighbours.

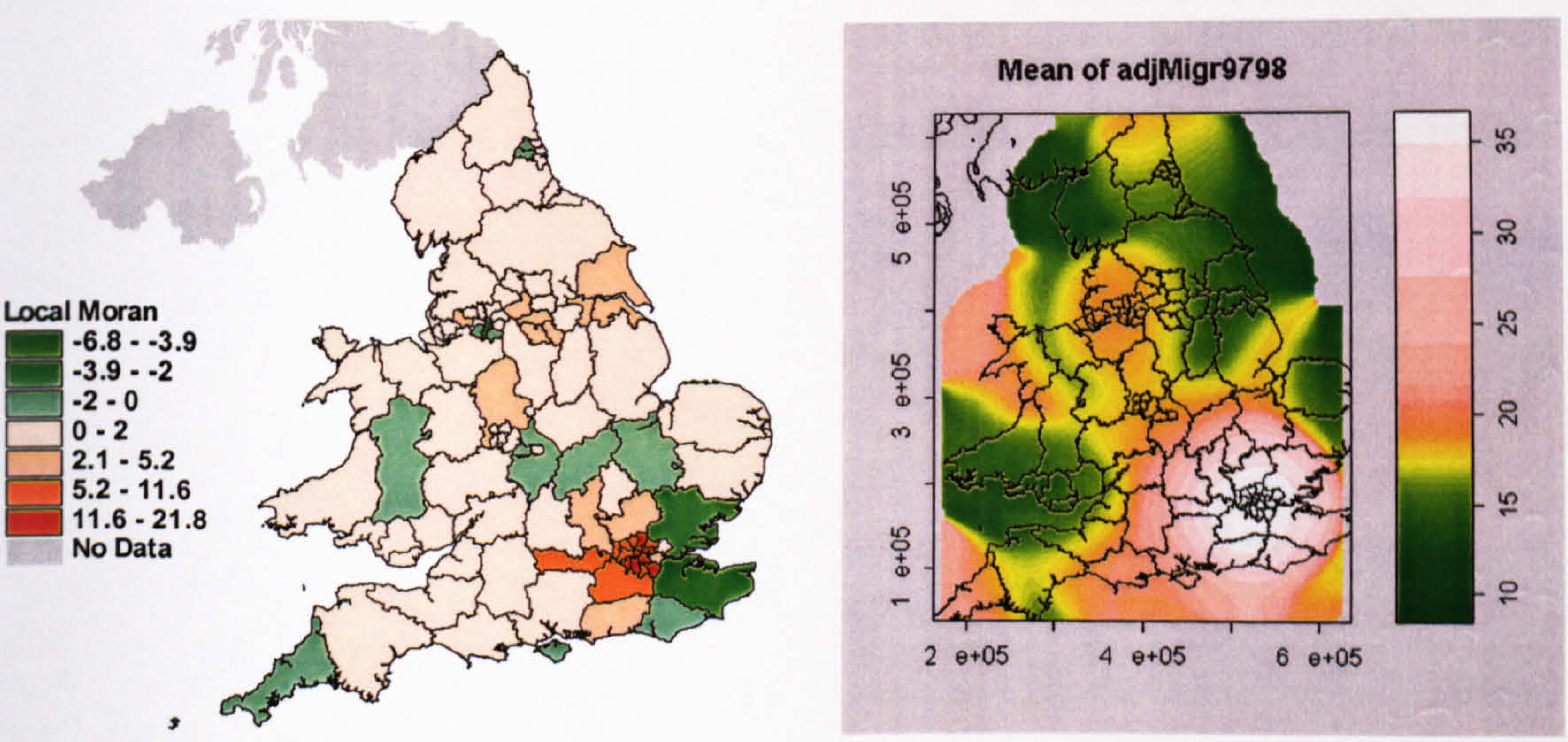


Figure 5.12. Map of $I_i(10)$ of Local Moran's I analysis and GWR local mean (males aged 45 – 59 in 1997-98)

Finally, the local mean using GWR was calculated for the same dataset and is also presented in Figure 5.12. The calculation of the local mean using GWR is possible through an R extension called GWR Local Statistics, available along with the GWR core from Chris Brunsdon. The extension allows the calculation of the local mean by selecting a fixed or an adaptive kernel. Here, for comparability reasons a fixed kernel with 100 km bandwidth is selected. The calculation for the local mean was made for the cell of a 100 x 100 grid that matches the geography of the FHSAs in England and Wales. The plotting of the results is possible though an integrated plot function within the extension. The result map (Figure 5.12) is a grid showing the local mean intersected by a FHSA boundary map. In the x, y axis there are the coordinates of the map, whereas the colours meaning is shown by a legend on the right of the map.

In the local mean map, there is an apparent spatial cluster of high out-migration rates in the South East and a smaller in the North West (Manchester). There are also some clusters of low out-migration rates; one in South Wales and on in Yorks and Humberside.

Although the statistics presented here are intended to use in large spatial datasets, their usefulness is also apparent in this dataset. The example presented above suggests that k-means, although a very simple and aspatial algorithm, is adequate for exploratory data analysis of this geographical scale and detail. What k-means fails to capture is the extremes because of the selected number of clusters. It is also not sure that k-means will work well with any dataset, whereas the ESDA statistics presented here are designed to do so.

5.6.2 Out-, in- and net migration for Newcastle FHSA and London (1991-1997)

In this section I focus on the FHSA of Newcastle as well as London (16 FHSAs combined) and I present what I call *heat maps*. These are tables in which the cells represent migration rates for a single area disaggregated by age and sex over a period of time. The cells contain the migration rate and are coloured based on a quintile classification of all data in the table. The result is a grid shown in Figures 5.13 – 5.15 for out-, in- and net migration rates respectively for Newcastle FHSA and Figures 5.16 – 5.18 for London.

Each figure consists of two parts: the heat map and its legend. The former is a table the rows of which refer to the seven years of observed migration flows and the columns of which refer to the 14 age/sex disaggregated groups. These data are summaries of the migration flow matrices. London is treated here as a single area with the flow data of its 16 FHSAs combined. The colour scheme used for in- and out-migration rates is common; the darker the colour the higher the rate. A different colour scheme was used in the net migration rates heat map. Here, when the rate is negative the higher the absolute value of the rate, the darker the colour (blue). All positive net migration rates were coloured in light red. There are some low negative rates also coloured red for Newcastle (Figure 5.15) and some positive rates coloured turquoise (blue) for London (Figure 5.18) because of the automatically selected intervals. The intervals result from a quintile classification of all data performed using SPSS descriptive statistics (Frequencies).

Out-migration rates for Newcastle		Age and Gender Migrant Groups													
		0-15		16-19		20-24		25-29		30-44		45-59		60+	
		F	M	F	M	F	M	F	M	F	M	F	M	F	M
Year	1990-91	31.16	28.63	40.39	25.86	127.12	88.09	96.84	91.80	39.20	49.81	15.68	21.38	13.05	12.55
	1991-92	31.95	31.80	63.36	38.69	149.45	95.89	102.50	97.27	39.98	51.41	16.68	20.66	13.96	12.90
	1992-93	34.23	32.38	64.74	47.62	157.49	105.67	93.79	86.22	39.26	47.93	17.08	21.32	13.97	13.11
	1993-94	35.16	37.34	70.65	46.19	174.77	119.39	97.69	95.01	41.46	48.35	16.73	22.94	13.93	12.09
	1994-95	38.12	36.04	86.55	59.26	192.84	136.91	115.28	112.93	40.77	51.01	18.12	23.10	13.77	14.11
	1995-96	42.05	41.11	75.62	54.35	197.64	134.23	135.41	128.86	43.47	51.54	16.89	23.54	12.97	13.35
	1996-97	39.19	42.02	75.56	53.11	176.42	128.47	152.80	159.17	43.70	52.71	18.81	24.48	14.53	13.92

a. *Heat map*

12.080 – 17.912	
17.912 – 38.462	
38.462 – 51.462	
51.462 – 98.652	
98.652 – 197.650	

b. Legend

Figure 5.13. Heat map of out-migration rates at Newcastle FHSA (1990-1997)

Although out-migration rates were discussed above, here the data presented are slightly different, because of non-adjustment in the male migration rates. The adjustment is necessary for the out-migration model robustness and is applied only to out-migration time series data. From the heat maps it is possible to extract two pieces of information, the stability of migration rates over time and sex/age group and the overall population change in terms of the migration factor (other factors of population change include immigration, emigration, fertility and mortality). Figure 5.13 shows similar trends as Figure 5.2c, but the visualisation of the data is different and it communicates the differences in out-migration rates among sex/age groups.

Figure 5.14 shows in-migration rates to Newcastle. The heat map of in-migration rates is similar of that of out-migration rates in its left and right parts, but not in its middle part (ages 16-29). In-migration rates are very high for teenagers and young female adults and less high for the young male adults and adults. For example, for every 100 female teenagers living in Newcastle there are 21-25 more coming to Newcastle during the 1990s. Small proportion of older adults and pensioners compared to the local population are coming to Newcastle.

In-migration rates for Newcastle		Age and Gender Migrant Groups													
		0-15		16-19		20-24		25-29		30-44		45-59		60+	
		F	M	F	M	F	M	F	M	F	M	F	M	F	M
Year	1990-91	23.35	25.99	96.23	65.86	128.63	90.94	72.22	64.62	26.43	35.50	10.63	15.75	6.44	6.60
	1991-92	24.35	23.89	212.00	169.78	134.85	106.91	70.79	68.42	29.37	40.03	10.88	17.91	7.38	8.80
	1992-93	27.47	27.83	218.35	175.92	144.66	118.61	78.34	71.39	30.05	39.78	11.78	18.37	7.23	8.56
	1993-94	25.68	24.31	232.03	179.08	138.72	120.22	73.35	74.51	28.91	38.13	12.02	16.62	7.39	7.45
	1994-95	25.34	25.65	254.68	188.92	122.94	102.93	79.42	74.10	26.89	34.89	10.70	17.38	9.32	9.14
	1995-96	29.03	29.89	242.73	177.37	126.01	98.62	94.91	89.56	29.03	37.98	11.59	17.00	7.53	8.51
	1996-97	28.10	28.96	237.51	176.32	121.88	99.72	104.71	118.35	27.88	37.84	12.42	16.23	8.27	9.01

a. *Heat map*

6.430 – 11.972

11.972 – 27.686

27.686 – 69.368

69.368 – 120.552

120.552 – 254.690

b. Legend

Figure 5.14. Heat map of in-migration rates at Newcastle FHSA (1990-1997)

Net migration rates (Figure 5.15) show changes in the population of Newcastle in terms of migration. Overall, Newcastle is declining in terms of population. The only age group Newcastle benefits from other part of England and Wales is that of 16–19 year olds. A major part of this group are college and university students, since Newcastle University and also Northumbria University offer tens of thousands of student positions each year. Currently, there are about 14,000 students in Newcastle University and 24,000 students in Northumbria University. There are also 30,000 students attending Newcastle College.

There is a consistent decrease over time of older adults and pensioner populations in Newcastle. The reduction of mature adults slowed down in 1992-93, but increased in later years. The decrease of adults follows a reverse trend of the increase of teenagers suggesting that most of teenagers coming in to Newcastle leave some years later. There is a dramatic drop in net migration rates of this group in 1992-3 as in the case of mature adults. In the early 1990's Newcastle was a net gainer in terms of younger male adults but became a net loser in the mid-1990's. There are three years of university studies usually followed by a years master's or job contract. Thus, it is expected that people coming to Newcastle for studies will stay for 3-4 years and then they will leave. This is what is happening in the case of young adults. The net migration of teenagers in 1990-91 is 47.92 per thousand people (average of males and females), but is tripled a year after. The low net in-migration for teenagers in 1990-91 results in no or low net out-migration for young adults in 1993-4. The dramatic change in net migration for young male adults since 1994-95 (net out-migration) is because of the dramatic increase in net in-migration for male teenagers since 1991-92. The 3-years gap suggests that most of the English and Welsh students are coming to Newcastle for their bachelor's degree and go to other parts of the country after their graduation. A substantial number of postgraduate students in the Universities of Newcastle are overseas and are likely to return to their home countries after graduation.

Net migration rates for Newcastle		**Age and Gender Migrant Groups**													
		0-15		16-19		20-24		25-29		30-44		45-59		60+	
		F	M	F	M	F	M	F	M	F	M	F	M	F	M
Year	1990-91	-7.81	-2.64	55.84	40.00	1.51	2.86	-24.62	-27.18	-12.77	-14.31	-5.05	-5.64	-6.61	-5.94
	1991-92	-7.60	-7.91	148.64	131.08	-14.60	11.03	-31.71	-28.85	-10.61	-11.38	-5.80	-2.74	-6.59	-4.11
	1992-93	-6.76	-4.55	153.61	128.31	-12.83	12.94	-15.45	-14.83	-9.21	-8.15	-5.30	-2.96	-6.74	-4.55
	1993-94	-9.48	-13.03	161.37	132.89	-36.05	0.83	-24.34	-20.50	-12.55	-10.21	-4.71	-6.32	-6.55	-4.63
	1994-95	-12.78	-10.40	168.13	129.66	-69.90	-33.99	-35.86	-38.83	-13.88	-16.12	-7.42	-5.72	-4.44	-4.97
	1995-96	-13.03	-11.22	167.11	123.01	-71.63	-35.61	-40.50	-39.30	-14.45	-13.56	-5.29	-6.54	-5.44	-4.83
	1996-97	-11.09	-13.06	161.95	123.20	-54.53	-28.75	-48.08	-40.81	-15.83	-14.87	-6.39	-8.25	-6.26	-4.91

a. *Heat map*

-71.640 – -16.996

-16.996 – -10.802

-10.802 – -6.132

-6.132 – -1.946

-1.946 – 168.140

b. Legend

Figure 5.15. Heat map of net migration rates at Newcastle FHSA (1990-1997)

Similar trends apply in the case of female teenagers and young female adults. The students dominate these groups and this is reflected in the extreme high net in-migration rates for female teenagers and net out-migration rates for young female adults.

Finally, as is expected, the net migration rates of children share the same trends as those of mature adults. The fact that rates of children are less than those of mature adults suggests that the mature adults group is a combination of family members with children as well as individuals without children. Some of these trends match those of the 1991 Census migration statistics (Champion et al., 1996). The 1991 census data suggest big-city and metropolitan areas (such as Newcastle) gain much of the migrants aged 16-29 years old whereas they lose much of older ages. However, this is misleading. The net gain for those 16-29 in metropolitan districts is because of the large number of in-migrants aged 16–19. Metropolitan districts, except London, are net losers for those 20 – 29, something that is missed in Champion et al. (1996) because of their age grouping strategy.

The heat map of out-migration rates for London (Figure 5.16) has many similarities with that for Newcastle in terms of temporal stability for children and older age groups. Here out-migration rates are higher for all sex/age groups except young adults. This suggests higher mobility rates, independent of sex and age for those living in London compared to those living in Newcastle. For young adults it suggests there are fewer returning graduates leaving London than Newcastle, mainly because of the working and living opportunities in the former.

In London, out-migration rates have increased after 1990/91 for all migrant groups. There is a surprising temporal stability in out-migration rates since 1991-92 for children and those aged 30 and over. For the remaining migrant groups, the changes in out-migration rates over time are not as great as in the case of Newcastle. Although out-migration rates for teenagers and young adults peaked in the mid-1990s and fell in 1996-97, those for adults increase every year since 1992/93.

Another interesting trend is that the out-migration rates of males are somewhat more than half those of females in the ages 16 – 29. This suggests that young females living in London are far more mobile than their male counterparts. It is difficult to know why this is so but possible reasons are: females being more likely to return to their parental domicile after graduation; forming families in earlier ages than males; and being more flexible in following their partner who is likely to be of an older age and thus might belong to an older age group. It may also be that females living in London are more adventurous than males. This gender division trend could also be connected with labour market conditions. Working opportunities for young professionals (high share in working opportunities in London) are still more available for males than females, possibly forcing the latter to select alternative destinations. Certainly, further investigation is necessary in order to understand this gap in out-migration rates between males and females in London.

One may argue that this gap may be due to an undercount of male migrants in the NHSCR data. A solution suggested in previous research (Fotheringham et al., 2002b) is to adjust the figures for males based on those of females and given the male to female ratio of the 1991 Special Migration Statistics (1991 Census). This solution, applied to out-migration figures for males aged 16 – 59, may be good for the NHSCR data in and around 1990. However, given the different net migration rates between males and females (Figure 5.18) this approach does not guarantee that the same ratio for adjustment applies in later years especially in 1996/97. Anyway, the undercount for Newcastle young male adults is 24% whereas for London young male adults is 28%, but the difference in the gap between out-migration rates for young male and female adults is bigger in the case of London. In 1996-97, out-migration rates for young female adults are 82% higher than those for young male adults in the case of London, but only 37% higher in the case of Newcastle.

Out-migration rates for London		Age and Gender Migrant Groups													
		0-15		16-19		20-24		25-29		30-44		45-59		60+	
		F	M	F	M	F	M	F	M	F	M	F	M	F	M
Year	1990-91	38.58	39.04	50.37	32.47	120.68	70.39	123.80	89.38	57.73	63.68	25.66	29.68	22.88	23.17
	1991-92	45.66	45.25	75.59	51.55	141.05	82.34	136.29	99.04	65.08	72.40	28.14	33.95	25.35	25.34
	1992-93	45.60	45.89	85.12	56.12	145.11	83.79	133.27	98.11	64.20	70.49	26.92	32.69	24.49	24.45
	1993-94	46.64	46.78	85.41	57.00	144.81	85.51	134.58	99.06	66.92	72.09	28.55	34.19	26.09	26.59
	1994-95	48.39	48.93	101.76	66.22	166.64	97.13	144.27	109.05	65.77	71.67	30.00	35.71	26.84	26.46
	1995-96	47.28	47.62	98.77	65.68	172.15	97.77	149.87	112.29	66.15	71.44	29.24	35.12	26.82	26.41
	1996-97	47.70	47.75	97.62	62.02	165.12	90.85	157.71	114.69	67.28	71.74	30.01	35.06	27.96	27.11

a. *Heat map*

22.867 – 29.594

29.594 – 47.664

47.664 – 70.434

70.434 – 99.043

99.043 – 172.162

b. Legend

Figure 5.16. Heat map of out-migration rates at London (1990-1997)

Figure 5.17 shows the heat map for in-migration rates in London. There is a temporal stability in these rates for children and those aged 30 and older since 1991-92. The rates for teenagers peaked in 1994-95 and have fallen since. Those for young adults peaked in 1995-96 and those for adults have increased since 1992-93 following similar temporal patterns with out-migration rates.

In London, in-migration rates for young adults are higher than those for teenagers, a reverse trend compared to Newcastle. This suggests that fewer teenagers come to London for universities or work from elsewhere in England and Wales and more young adults are returning graduates or job hunters compared to those going to Newcastle. The former is because of the higher living costs in London which possibly deter potential students, and the latter is because of the career development opportunities in London. In-migration rates for all age groups except teenagers are higher in London than in Newcastle.

In-migration rates for London		Age and Gender Migrant Groups													
		0-15		16-19		20-24		25-29		30-44		45-59		60+	
		F	M	F	M	F	M	F	M	F	M	F	M	F	M
Year	1990-91	28.49	28.35	58.38	35.69	141.15	84.36	116.55	91.54	43.94	51.87	18.17	22.80	11.59	12.25
	1991-92	34.33	33.45	84.01	53.82	160.40	98.08	127.44	100.72	50.24	59.09	19.81	25.89	12.93	13.12
	1992-93	34.59	34.85	91.69	57.66	167.66	101.09	126.02	100.46	50.36	58.77	18.73	24.90	12.29	12.46
	1993-94	35.64	35.12	90.36	58.20	172.78	104.49	129.86	103.31	52.67	60.45	19.42	25.77	13.16	13.28
	1994-95	37.02	37.38	108.23	65.73	207.01	129.20	140.57	117.06	52.42	60.60	20.55	27.28	14.41	14.39
	1995-96	35.61	35.26	104.13	63.80	219.32	131.99	147.74	122.29	52.22	60.54	19.97	26.44	14.55	14.81
	1996-97	34.17	33.58	98.30	57.67	212.23	124.85	155.66	126.34	51.93	58.98	19.75	25.10	15.10	14.50

a. *Heat map*

11.576 – 19.941

19.941 – 35.624

35.624 – 59.635

59.635 – 116.649

116.649 – 219.330

b. Legend

Figure 5.17. Heat map of in-migration rates at London (1990-1997)

Similar to the gap between out-migration rates for young male and female adults, there is an apparent gap between in-migration rates for young male and female adults. This is a bit surprising because it seems that the reasons motivating out-migration of young female adults do not discourage in-migration of the same group in London. The benefit for London is the stability in its female population.

It is apparent that migration flows from and to London are very high compared to Newcastle. It is interesting to see the picture of net migration. Here there are great differences between London and Newcastle. London is a net population gainer for female teenagers, young adults and male adults (Figure 5.18), whereas Newcastle is a net gainer only for male and female teenagers. On the other hand, London has higher net out-migration rates for children, older adults and pensioners than Newcastle. The picture is more mixed for mature adults, where the net migration rates are similar. Both areas are net losers for female adults but the rates for Newcastle are 24 times higher than those for London in 1996-97.

In London, there is a temporal stability in net migration rates for children and those aged 30 and over. The positive net migration rates are decreasing for female teenagers and increasing for young adults and female adults since 1991-92. The negative net migration rates are decreasing for young female adults since 1991-92 to -2 per 1000 population in 1996-97. The net migration rates for male teenagers which are positive in 1990-91 decreased until 1993-94 and became negative the following year.

Net migration rates for London		Age and Gender Migrant Groups													
		0-15		16-19		20-24		25-29		30-44		45-59		60+	
		F	M	F	M	F	M	F	M	F	M	F	M	F	M
Year	1990-91	-10.09	-10.69	8.01	3.23	20.47	13.96	-7.25	2.17	-13.79	-11.81	-7.49	-6.88	-11.29	-10.92
	1991-92	-11.32	-11.80	8.42	2.27	19.35	15.74	-8.85	1.68	-14.84	-13.31	-8.33	-8.06	-12.42	-12.22
	1992-93	-11.01	-11.04	6.57	1.54	22.55	17.30	-7.25	2.35	-13.84	-11.73	-8.20	-7.79	-12.20	-11.99
	1993-94	-11.01	-11.66	4.95	1.19	27.97	18.98	-4.72	4.25	-14.25	-11.64	-9.13	-8.42	-12.93	-13.31
	1994-95	-11.37	-11.55	6.47	-0.50	40.37	32.07	-3.70	8.00	-13.35	-11.06	-9.45	-8.42	-12.43	-12.07
	1995-96	-11.67	-12.36	5.36	-1.89	47.17	34.23	-2.14	10.00	-13.92	-10.90	-9.27	-8.69	-12.27	-11.60
	1996-97	-13.52	-14.17	0.67	-4.34	47.11	34.00	-2.05	11.65	-15.35	-12.76	-10.26	-9.96	-12.87	-12.61

a. *Heat map*

-15.363 – -12.230

-12.230 – -10.911

-10.911 – -6.017

-6.017 – 6.856

6.856 – 47.178

b. Legend

Figure 5.18. Heat map of net migration rates at London (1990-1997)

5.6.3 Net population gainers and losers

In the previous section migration rates for Newcastle FHSA were examined suggesting that Newcastle is a net gainer in teenagers during the 1990s and this is partly because the city offers a substantial number of university places. Further evidence for the argument that FHSAs with universities (such as Newcastle) are net gainers of teenagers comes from the comparison of migration rates for teenagers in a FHSA with and without big universities.

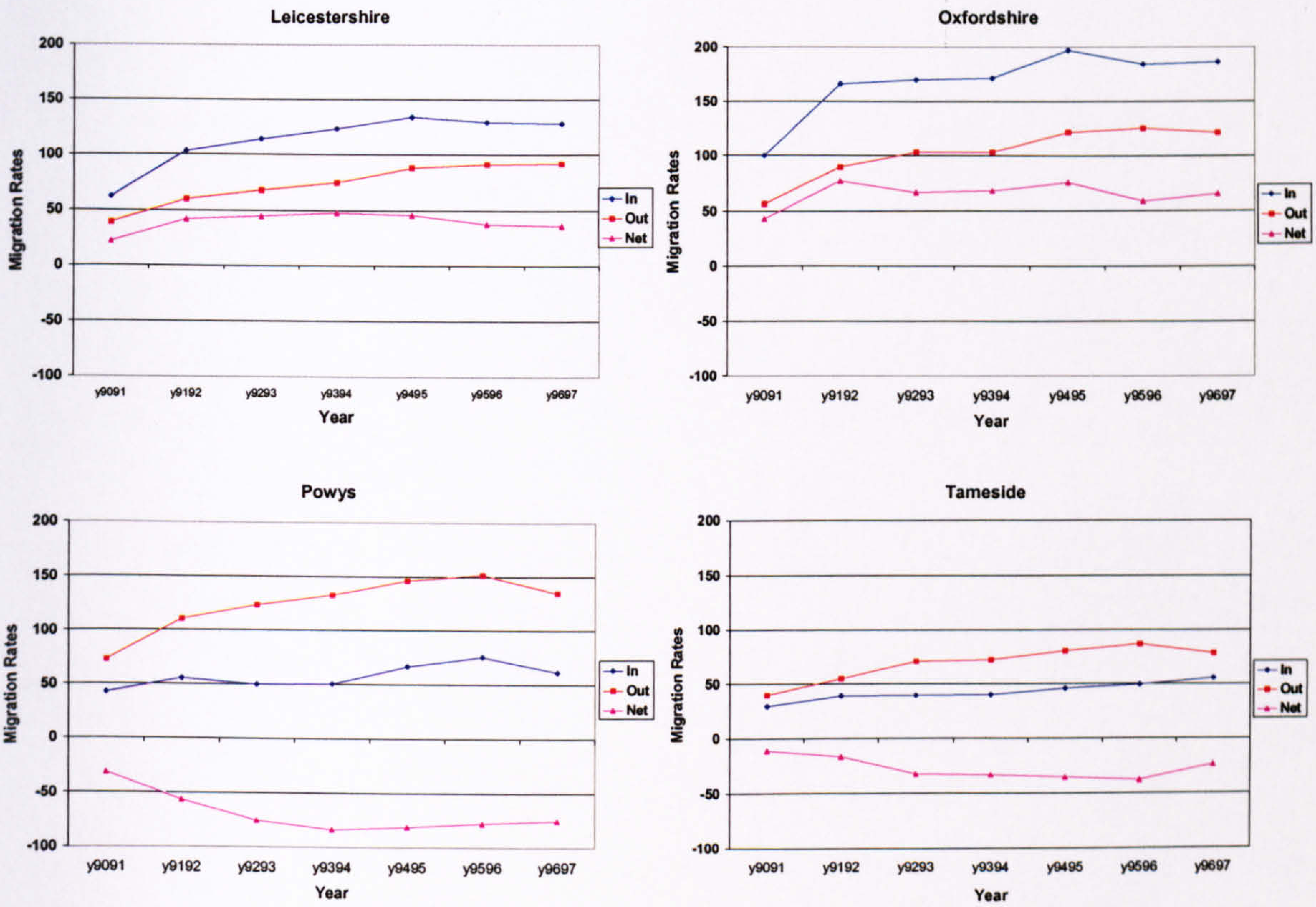


Figure 5.19. Female teenagers in-, out-, and net migration rates in 1990s.

Figure 5.19 shows graphs of female teenagers in-, out-, and net migration rates for four FHSAs: Leicestershire, Oxfordshire, Powys and Tameside. Both the former two FHSAs, which offer tens of thousands of university places, are net gainers of female teenager population whereas the latter two FHSAs, which are lacking universities, are net female teenager population losers. These observations are in line with the analysis by Stillwell (1994) concerning net migration trends in the 1990's. He also suggest FHSAs with major universities are net migration gainers of teenagers and net losers of young adults.

Table 5.4 shows FHSAs with the lowest and highest net migration rates in 1996/97. The data are presented for the 14 sex/age migrant groups. FHSAs in metropolitan areas (London and Manchester) are net population losers for all ages except teenagers and young adults, where they are net gainers if universities are present. Rural and remote areas (Powys, Isle Of Wight) on the other hand, are net population losers for teenagers and young adults, but are net gainers for children, older adults and pensioners. Although mature adults seem to abandon the metropolitan FHSAs, they are preferably concentrated in prosperous South East FHSAs. The champion net gainer for this migrant group is West Sussex. FHSAs in London are net gainers for young adults and adults. One reason for this is the level of job availability in the capital, especially for graduates.

Table 5.4. Highest net population gainers and losers FHSAs in 1996/97

Sex/Age Group	Net population loser FHSA	rate*	Net population gainer FHSA	rate*
Females 0-15	Kensington & Chelsea, Westminster	-29.38	Isle Of Wight	14.55
Males 0-15	Kensington & Chelsea, Westminster	-33.35	Solihull	17.45
Females 16-19	Powys	-74.33	Manchester	174.09
Males 16-19	Somerset	-66.22	Sheffield	134.68
Females 20-24	Dyfed	-82.06	Merton, Sutton, Wandsworth	87.62
Males 20-24	Coventry	-57.26	Merton, Sutton, Wandsworth	68.32
Females 25-29	Manchester	-56.38	Richmond, Kingston	30.51
Males 25-29	Manchester	-42.29	Richmond, Kingston	34.43
Females 30-44	Kensington & Chelsea, Westminster	-35.89	West Sussex	16.94
Males 30-44	Manchester	-25.94	West Sussex	16.81
Females 45-59	Lambeth, Southwark, Lewisham	-15.20	Isle Of Wight	18.05
Males 45-59	City of London, Hackney, Newham, Tower Hamlets	-16.72	Cornwall	18.15
Females 60+	City of London, Hackney, Newham, Tower Hamlets	-26.94	Dorset	8.81
Males 60+	Lambeth, Southwark, Lewisham	-22.10	Dorset	12.95

* net migration rates per thousand population

5.7 Migration Flows (Newcastle FHSA and London as origin and destination)

It is interesting to identify FHSA pairs between which there are high migrant flows in order to investigate if there are any persistent spatial patterns in these locations. In the literature, such attempts have been made by Fielding (1993), and Engels and Healy (1981). Fielding suggests a statistical measure called *migration velocities*, which are migration rates standardised for the population sizes at both origin and destination. This statistic can be calculated as $mv = k\ (M_{ij}/P_iP_j)$ where k is a scale factor, M_{ij} is the migration flow from origin i to destination j and P_i, P_j are the population of origin i and destination j, respectively. Migration velocities are actually the probabilities of the observed flows out of all possible flows. In the literature I am aware of, nobody else has used this method to analyse migration flows suggesting it has not become very popular. Migration velocities calculated for the disaggregated data matrices available here are very small numbers and they do not offer a powerful means of identifying hidden trends. Another way of analysing flow data, more appropriate for time series is the calculation of indexes of dissimilarities (Engels and Healy, 1981). However, these fail to extract significant information here because the data analysed have a substantial temporal stability especially those referring to children and those aged over 29 years old.

Nevertheless, the expectation is that most of the pairs with high interaction migration will be neighbour FHSAs and large populated FHSAs (such as metropolitan FHSAs) with close spatial proximity. Here, migration flows from and to Newcastle FHSA as well as London (combined FHSAs in London) are examined. Figure 5.20 shows the distribution of total migrants leaving (or coming to) Newcastle in 1996/97. The numbers shown are migrants going to (or coming from) a particular FHSA as a percentage of the total number of migrants leaving (or coming to) Newcastle FHSA. Very similar trends apply to the corresponding migration flows in 1990/91. This empirical example verifies the expected trends discussed above.

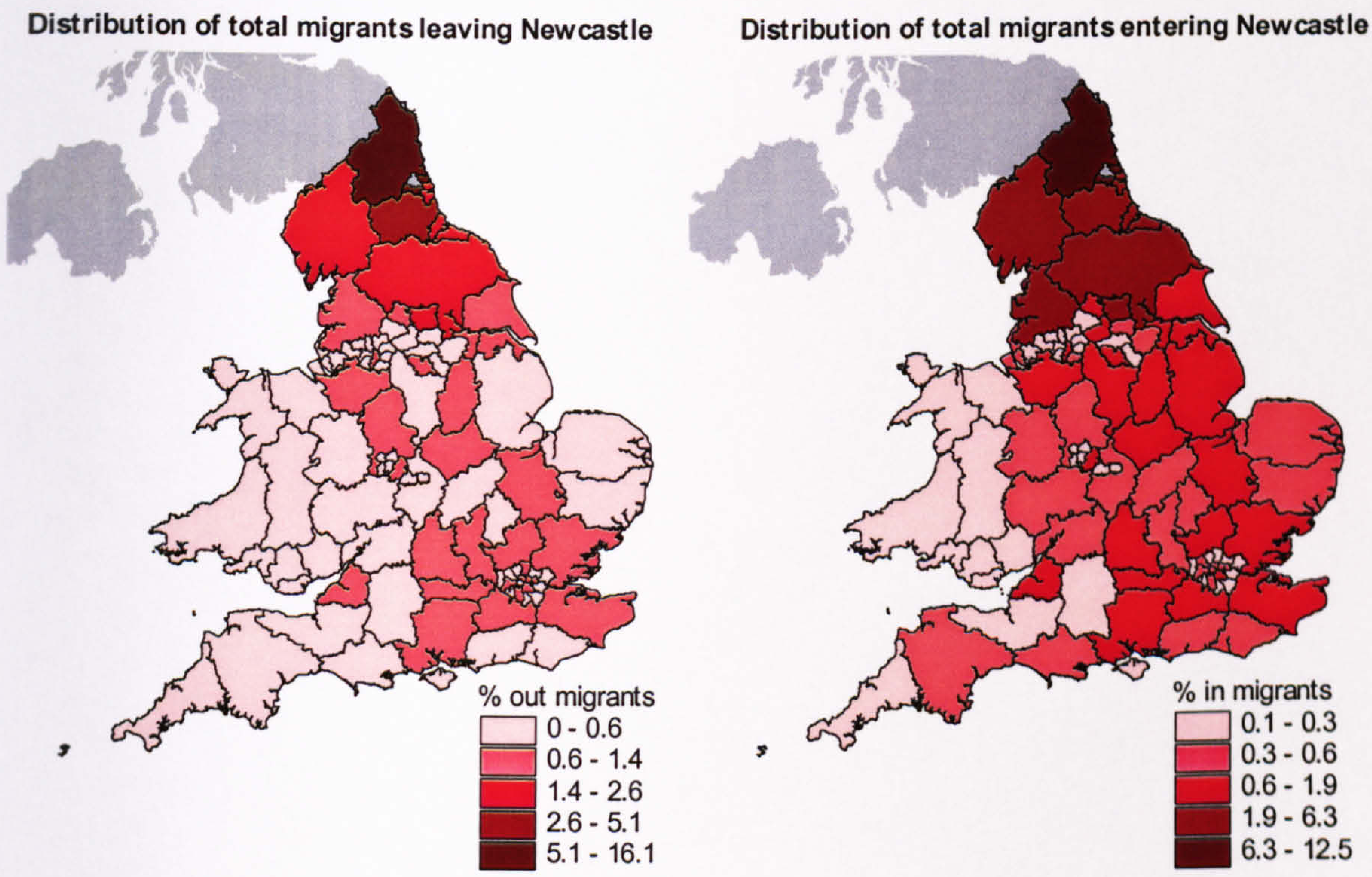


Figure 5.20. Maps of migrants leaving and arriving to Newcastle in 1996/97

In 1996/97, 15,001 people left Newcastle migrating to the rest of England and Wales. Half of these migrants went to the neighbour FHSAs; North Tyneside (14.85%), Northumberland (14.01%) and Gateshead (9.88%); and populated cities in the North; Durham (4.95%), North Yorkshire (3.15%) and Leeds (2.39%). In the same year, 13,019 migrants came to Newcastle from England and Wales. Almost half of those came from the same areas as above: neighbouring FHSAs including North Tyneside (12.48%), Northumberland

(10.25%) and Gateshead (9.09%); and populated cities in the North including Durham (6.3%), Sunderland (3.77%) and North Yorkshire (3.33%).

Based on these observations it is clear that the desirability of an area also plays a big role for people to move. For example, there are smaller proportions of migrants going from Newcastle to Sunderland than those coming to Newcastle from Sunderland. However, most of the migration trips are short distance trips, some of which just cross the administrative border of an FHSA. This image however changes when disaggregated data flows are analysed. Table 5.5 shows the destinations of migrants leaving Newcastle for all 14 sex/age groups in 1990/91. The rule for FHSAs to be included in the list of each migrant group is that more than 2% of the total out-migrants select that FHSA.

People leaving Newcastle mainly go to North Tyneside, Gateshead and Northumberland and to a much lesser extent to South Tyneside and Sunderland that are also very close to Newcastle. It is also apparent that most of the older people will almost exclusively go to the former three FHSAs (North Tyneside, Gateshead and Northumberland), for example 60% in the case of older adults and pensioners. Those aged 16 – 29 select many more destinations when they leave Newcastle than the other age groups. This is also expected as young people are more mobile and more willing to explore the 'unknown' by going to all parts of England and Wales. If distance is seen as fear for the unknown and loss of social ties, then young people are the bravest in their destination choice. They are also likely to have visited more places than family people and older people, thus, have more information about potential destinations.

Table 5.6 is similar to Table 5.5 and refers to people coming to Newcastle. Although most of the migrants arriving in Newcastle are from neighbour FHSAs (North Tyneside, Gateshead, Northumberland, Durham, Sunderland), there are other substantial areas whose people are attracted to Newcastle. These include the more deprived and rural areas of North England (Cumbria and Cleveland), areas from the North West England (Lancashire, Cheshire) and Midlands (Derbyshire, Leicestershire, Nottinghamshire) and surprisingly from the South East England (Kent, Bedfordshire, Hampshire). It can be argued here, that people coming to Newcastle come from far more FHSAs than Newcastle people are going to. Especially in the case of young people only an eighth of teenagers and a sixth of young adults coming to Newcastle are from the rest of Tyne and Wear whereas a third of teenagers and a fifth of young adults are going from Newcastle to the rest of the Tyne and Wear (North Tyneside, South Tyneside, Gateshead, and Sunderland).

Table 5.5. Distribution of migration flows from Newcastle to the remaining 97 FHSAs in 1990/91

	FHSA	%		FHSA	%
F 0-15	North Tyneside	18.77	M 0-15	North Tyneside	18.55
	Northumberland	17.13		Northumberland	16.99
	Gateshead	12.47		Gateshead	14.40
	Durham	6.05		Durham	6.61
	South Tyneside	3.90		Cumbria	3.50
	Sunderland	3.53		Sunderland	3.11
	Birmingham	2.27		South Tyneside	2.33
	Leeds	2.14			
	Cumbria	2.02			
F 16-19	Gateshead	14.05	M 16-19	North Tyneside	14.29
	North Tyneside	12.42		Northumberland	9.85
	Northumberland	9.48		Durham	8.87
	Durham	4.58		Gateshead	8.37
	Leeds	4.25		Humberside	3.94
	South Tyneside	3.92		Sheffield	3.45
	Manchester	2.94		Somerset	2.46
	Nottinghamshire	2.29		Liverpool	2.46
	Cumbria	2.29			
F 20-24	North Tyneside	9.72	M 20-24	Gateshead	8.62
	Gateshead	9.61		North Tyneside	7.07
	Northumberland	4.41		North Yorkshire	3.98
	North Yorkshire	4.30		Northumberland	3.46
	Durham	4.13		Durham	3.39
	Leeds	2.66		Cleveland	3.32
	Cheshire	2.26		Leeds	2.73
	Lambeth, Southwark, Lewisham	2.20		Ealing, Hammersmith, Hounslow	2.51
	Cleveland	2.20		Leicestershire	2.36
	Sunderland	2.04		Humberside	2.28
				Cheshire	2.28
F 25-29	Gateshead	15.46	M 25-29	North Tyneside	12.97
	North Tyneside	14.95		Gateshead	11.42
	Northumberland	8.36		Northumberland	6.77
	Durham	6.25		Durham	4.49
	Sunderland	3.13		Cleveland	2.53
	Leeds	2.36		Leeds	2.53
	South Tyneside	2.11		Lambeth, Southwark, Lewisham	2.28
				North Yorkshire	2.28
				Sunderland	2.28
				Hampshire	2.04
F 30-44	North Tyneside	19.61	M 30-44	North Tyneside	18.94
	Northumberland	17.74		Gateshead	14.30
	Gateshead	13.54		Northumberland	12.22
	Durham	5.42		Durham	5.29
	Sunderland	2.24		South Tyneside	2.79
	South Tyneside	2.05		Sunderland	2.64
F 45-59	North Tyneside	26.01	M 45-59	North Tyneside	25.41
	Northumberland	21.36		Northumberland	19.29
	Gateshead	12.38		Gateshead	14.59
	Durham	4.02		Sunderland	4.47
	Sunderland	3.72		Durham	3.53
	South Tyneside	3.41			
F 60+	North Tyneside	32.62	M 60+	North Tyneside	29.93
	Northumberland	16.95		Northumberland	16.45
	Gateshead	13.73		Gateshead	13.49
	Durham	5.58		Durham	8.55
	Sunderland	3.00		Sunderland	3.29
				South Tyneside	2.96
				North Yorkshire	2.30

The above measurements and trends refer to migration flows in 1990/91. Table 5.7 shows destinations of migrants leaving Newcastle for all 14 sex/age groups in 1996/97. The general trends in the most recent data remain. Some changes are discussed below.

Table 5.6. Distribution of migration flows to Newcastle from the remaining 97 FHSAs in 1990/91

	FHSA	%		FHSA	%
F 0-15	North Tyneside	20.71	M 0-15	North Tyneside	19.88
	Northumberland	14.78		Northumberland	16.57
	Gateshead	12.01		Gateshead	12.07
	Durham	9.10		Durham	7.46
	Sunderland	4.62		Sunderland	5.33
	South Tyneside	4.35		South Tyneside	4.26
				Cleveland	2.96
F 16-19	North Yorkshire	6.51	M 16-19	North Yorkshire	7.02
	North Tyneside	5.92		Northumberland	4.71
	Durham	5.32		North Tyneside	4.20
	Northumberland	4.40		Cumbria	4.20
	Cleveland	4.18		Durham	3.84
	Gateshead	3.91		Lancashire	3.62
	Lancashire	3.69		Humberside	3.62
	Cumbria	3.47		Leeds	3.33
	Humberside	3.37		Cleveland	3.19
	Leeds	3.26		Gateshead	3.04
	Leicestershire	2.50		Cheshire	2.90
	Derbyshire	2.28		Lincolnshire	2.46
	Cheshire	2.28		Derbyshire	2.24
	Nottinghamshire	2.23		Berkshire	2.24
	Surrey	2.06			
F 20-24	North Tyneside	10.15	M 20-24	Northumberland	8.34
	Northumberland	7.55		Gateshead	7.52
	Durham	6.75		North Tyneside	7.31
	Gateshead	6.31		Durham	5.66
	Sunderland	4.95		Sunderland	3.86
	Leeds	3.34		North Yorkshire	3.66
	Cumbria	3.28		Cumbria	3.45
	North Yorkshire	3.09		Leeds	3.24
	Cleveland	2.41		Lancashire	3.03
	Lancashire	2.41		Cleveland	2.97
	South Tyneside	2.04		Humberside	2.76
				Cambridgeshire	2.48
				Nottinghamshire	2.28
				Leicestershire	2.14
				South Tyneside	2.07
F 25-29	North Tyneside	14.97	M 25-29	Gateshead	12.61
	Gateshead	11.22		North Tyneside	11.33
	Northumberland	11.22		Northumberland	10.15
	Durham	6.92		Durham	7.09
	Sunderland	4.65		Sunderland	4.53
	South Tyneside	3.63		Cumbria	2.96
	Leeds	2.38		South Tyneside	2.86
	North Yorkshire	2.27		Lancashire	2.46
	Cleveland	2.04		North Yorkshire	2.36
				Cleveland	2.17
F 30-44	North Tyneside	19.20	M 30-44	North Tyneside	18.28
	Northumberland	15.63		Gateshead	13.15
	Gateshead	12.30		Northumberland	11.79
	Durham	5.75		Durham	6.98
	Sunderland	3.79		Sunderland	5.13
	Cleveland	2.53		South Tyneside	3.37
	South Tyneside	2.30		Cleveland	2.41
				Cumbria	2.33
F 45-59	North Tyneside	18.98	M 45-59	North Tyneside	19.83
	Northumberland	17.88		Northumberland	15.64
	Gateshead	17.52		Gateshead	14.53
	Durham	6.93		Durham	8.38
	Sunderland	3.65		Sunderland	4.47
	Leeds	2.55		South Tyneside	3.35
	South Tyneside	2.55			
F 60+	North Tyneside	22.06	M 60+	Northumberland	21.53
	Northumberland	21.69		North Tyneside	18.18
	Gateshead	12.87		Gateshead	15.79
	Sunderland	9.56		Durham	7.66
	Durham	4.04		Sunderland	3.83
	Cumbria	2.57		South Tyneside	2.39
				Cleveland	2.39
				Cumbria	2.39

Table 5.7. Distribution of migration flows from Newcastle to the remaining 97 FHSAs in 1996/97

	FHSA	%		FHSA	%
F 0-15	Northumberland	21.19	M 0-15	Northumberland	21.29
	North Tyneside	19.68		North Tyneside	20.47
	Gateshead	13.06		Gateshead	11.34
	Durham	5.96		Durham	6.85
	South Tyneside	2.65		Cleveland	2.77
	North Yorkshire	2.18		South Tyneside	2.77
				Cumbria	2.61
				North Yorkshire	2.45
				Lancashire	2.04
F 16-19	North Tyneside	11.60	M 16-19	Northumberland	10.82
	Gateshead	11.26		North Tyneside	10.10
	Northumberland	10.24		Gateshead	8.41
	Leeds	6.83		Durham	6.25
	Durham	6.31		Nottinghamshire	5.53
	North Yorkshire	3.58		Leeds	5.29
	Sheffield	3.41		North Yorkshire	5.05
	Manchester	2.90		Manchester	3.37
	Lancashire	2.56		Sheffield	3.13
	Cumbria	2.05		Sunderland	2.16
F 20-24	North Tyneside	6.80	M 20-24	North Tyneside	5.30
	Gateshead	5.64		Northumberland	5.03
	North Yorkshire	4.66		North Yorkshire	5.03
	Northumberland	4.45		Gateshead	4.87
	Durham	3.85		Leeds	3.43
	Leeds	3.76		Lancashire	3.37
	Nottinghamshire	2.57		Durham	2.78
	Lancashire	2.52		Cheshire	2.68
	Cleveland	2.48		Cleveland	2.57
	Ealing, Hammersmith, Hounslow	2.44		Humberside	2.52
				Hampshire	2.41
				Ealing, Hammersmith, Hounslow	2.36
				Nottinghamshire	2.09
				Cumbria	2.03
F 25-29	North Tyneside	15.70	M 25-29	Gateshead	11.50
	Northumberland	10.49		North Tyneside	11.50
	Gateshead	10.33		Northumberland	8.42
	Durham	4.82		Durham	4.76
	Sunderland	2.87		Lambeth, Southwark, Lewisham	2.56
	Cleveland	2.80		Leeds	2.56
	North Yorkshire	2.72		North Yorkshire	2.56
	Ealing, Hammersmith, Hounslow	2.18		Merton, Sutton, Wandsworth	2.34
				Cleveland	2.05
F 30-44	Northumberland	21.92	M 30-44	North Tyneside	19.06
	North Tyneside	20.89		Northumberland	17.16
	Gateshead	11.66		Gateshead	13.01
	Durham	5.65		Durham	5.81
	South Tyneside	2.49		Sunderland	2.25
	North Yorkshire	2.13		North Yorkshire	2.25
				South Tyneside	2.07
F 45-59	Northumberland	29.64	M 45-59	Northumberland	26.48
	North Tyneside	20.00		North Tyneside	21.30
	Gateshead	13.25		Gateshead	13.89
	Durham	4.82		Durham	5.19
	Sunderland	3.13		Sunderland	2.59
	North Yorkshire	2.41			
	Cumbria	2.17			
F 60+	North Tyneside	29.08	M 60+	North Tyneside	27.24
	Northumberland	24.06		Northumberland	26.63
	Gateshead	10.04		Gateshead	8.67
	Durham	5.02		Durham	4.02
	North Yorkshire	2.30		South Tyneside	3.41
	Sunderland	2.09		North Yorkshire	2.48
				Sunderland	2.17

The percentage of short distance moves in 1996/97 compared to 1990/91 has been increased for children, mature and older adults; decreased for teenagers, young adults and adults; and slightly decreased for pensioners. An interesting observation is that in 1996/97 a greater proportion of adults left Newcastle to go to London FHSAs (Ealing with Hammersmith & Hounslow, Merton with Sutton & Wandsworth, Lambeth with Southwark & Lewisham) than in 1990/91. Another observation is that the variety of destinations chosen by Newcastle migrants in most of the sex/age migrant groups has increased in recent years. The latter has also been reported in Findlay and Rogerson (1993, p. 33) who suggest that *...migration patterns are becoming more complex and migrants are responding to a much more varied set of stimuli than in the past, leading to greater selectivity in the types of places people chose to move to. In particular, people are giving more attention to quality of life considerations....*

To conclude, more than half of the people leaving Newcastle will move to another place in North England, except for young adults where only one third will stay in North England. Not surprisingly more that three quarters of pensioners will move from Newcastle to another location within the North East.

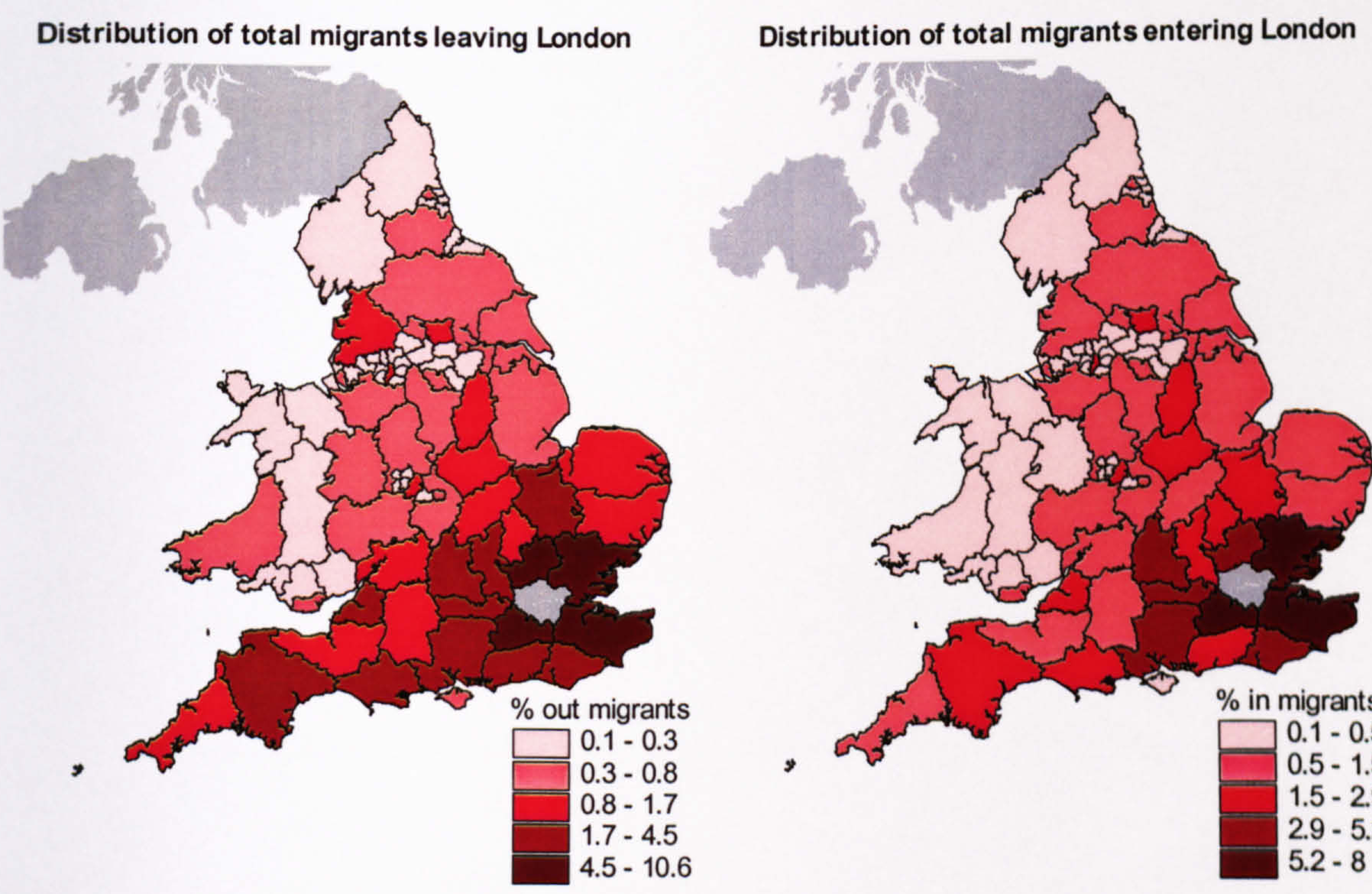


Figure 5.21. Maps of migrants leaving and arriving to London in 1996/97

Figure 5.21 shown above is similar to Figure 5.20. It presents the distribution of migrants leaving and entering (percentages of total migrants leaving and entering, respectively) London in 1996/97. It is apparent that migrants leaving London seem to select more destinations across England and Wales than those leaving Newcastle. The former are also more evenly distributed across destinations than the latter. Higher proportions of long distance migrants select London than Newcastle, but this is expected as Newcastle is a city of 280,923 residents whereas London has 7,088,899 residents (according to the mid-1996 population estimate). Both London and Newcastle attract people from more areas than those they supply.

In 1996/97, 208,976 people left London migrating to the rest of England and Wales. More than half of these migrants went to the Rest of South East; mainly Essex (10.6%), Surrey (10.5%) and Kent (9.17%). The more desirable destinations outside the South East are Devon (2.2%) and equally Avon, Dorset and Cambridgeshire (each 2.1%). In the same year, only 159,005 migrants came to London from the rest of England and Wales. Almost half of those came from the rest of the South East, mainly from neighbouring FHSAs including Surrey (7.95%), Kent (7.54%) and Essex (6.89%). From outside the South East, London was selected from those leaving Avon (2.85%), Buckinghamshire (2.72%), Cambridgeshire (2.64%), Birmingham (2.42%), Devon (2.29%), West Sussex (2.22%), Leicestershire (2.06%) and Manchester (1.93%).

These figures suggest London is a population loser for the benefit of the rest of the South East and South West. It is also apparent that London does not equally benefit in terms of population from these areas. Instead, it attracts migrants from all over the rest of England and Wales as well as international migrants (not included here). These observations provide some evidence for the continuation of the counterurbanisation phenomenon (Champion, 1989) in the mid-1990s.

The distribution of migration flows from London to the remaining 82 FHSAs in England and Wales in 1990/91 and 1996/97 are shown in Table 5.8 and Table 5.9, respectively. These tables have the same structure as Table 5.5 and Table 5.7. However, in the former only destination FHSAs selected with more than 3% of the total out-migrants were included. This was necessary for the table to fit onto a single page.

Table 5.8. Distribution of migration flows from London to the remaining 82 FHSAs in 1990/91

	FHSA	%
F 0-15	Essex	9.58
	Surrey	9.56
	Kent	9.07
	Hertfordshire	7.08
	Hampshire	4.40
	East Sussex	3.86
	Buckinghamshire	3.80
	Berkshire	3.40
F 16-19	Surrey	7.28
	Kent	7.01
	Essex	6.79
	Hampshire	4.70
	Hertfordshire	4.47
	Leeds	3.78
	East Sussex	3.73
	Avon	3.71
	Cambridgeshire	3.39
	Oxfordshire	3.08
F 20-24	Surrey	7.67
	Essex	7.64
	Kent	7.08
	Hertfordshire	6.61
	Hampshire	4.36
	East Sussex	3.65
	Berkshire	3.14
F 25-29	Surrey	10.93
	Hertfordshire	8.59
	Essex	8.32
	Kent	7.59
	Hampshire	3.91
	Berkshire	3.89
	Buckinghamshire	3.54
	East Sussex	3.04
F 30-44	Surrey	11.36
	Essex	8.55
	Hertfordshire	7.97
	Kent	7.38
	Hampshire	4.12
	East Sussex	3.94
	Buckinghamshire	3.90
	Berkshire	3.66
F 45-59	Essex	10.27
	Surrey	9.38
	Kent	8.62
	Hertfordshire	6.73
	East Sussex	5.27
	Hampshire	4.08
	West Sussex	3.65
	Dorset	3.21
	Norfolk	3.17
	Devon	3.06
	Buckinghamshire	3.03
F 60+	Essex	11.64
	Kent	9.90
	Surrey	9.00
	East Sussex	7.02
	Hertfordshire	6.84
	West Sussex	5.54
	Hampshire	4.60
	Dorset	3.70
	Buckinghamshire	3.23

	FHSA	%
M 0-15	Surrey	9.41
	Essex	9.25
	Kent	8.69
	Hertfordshire	7.08
	Hampshire	4.44
	East Sussex	4.10
	Berkshire	3.66
	Buckinghamshire	3.55
	West Sussex	3.32
M 16-19	Surrey	7.42
	Essex	6.76
	Kent	6.33
	Hampshire	5.18
	Avon	4.62
	Hertfordshire	4.31
	East Sussex	3.97
	Oxfordshire	3.79
	Cambridgeshire	3.35
M 20-24	Kent	7.17
	Essex	7.15
	Surrey	6.83
	Hertfordshire	6.06
	Hampshire	4.28
	East Sussex	3.93
	Avon	3.42
	Cambridgeshire	3.02
M 25-29	Surrey	10.66
	Hertfordshire	8.92
	Essex	8.85
	Kent	7.86
	Berkshire	3.95
	Hampshire	3.67
	East Sussex	3.09
	Buckinghamshire	3.07
M 30-44	Surrey	11.66
	Essex	9.38
	Hertfordshire	8.06
	Kent	7.87
	Buckinghamshire	4.01
	East Sussex	3.86
	Berkshire	3.76
	Hampshire	3.69
M 45-59	Essex	10.27
	Surrey	10.17
	Kent	8.98
	Hertfordshire	7.52
	East Sussex	4.62
	Hampshire	3.67
	West Sussex	3.27
	Buckinghamshire	3.17
	Dorset	3.11
M 60+	Essex	10.87
	Kent	9.14
	East Sussex	7.63
	Surrey	7.52
	Hertfordshire	5.82
	West Sussex	5.72
	Hampshire	4.51
	Dorset	4.20
	Norfolk	3.71
	Devon	3.16

The sex/age disaggregated data show that migrants leaving London chose from a larger variety of destinations than migrants leaving Newcastle did. For example, in 1996/97 more than half of the pensioners leaving Newcastle moved either to North Tyneside or to Northumberland, whereas half those leaving London moved to Essex, Kent, Surrey, Hertfordshire, West Sussex and East Sussex. Another apparent trend is the presence of many zero flows between Newcastle and the remaining FHSAs in both directions, which is not the case in any of the sex/groups for London. Less desirable FHSAs for people leaving Newcastle are the Isle of Wight, Sandwell and Walsall, whereas people leaving London mostly avoid South Tyneside and Rotherham.

Between 1990/91 and 1996/97 there are no significant changes in the way migrants leaving London select destinations. The most recent data suggest that higher percentages of migrants (except teenagers) select the three most attractive destinations (Essex, Kent and Surrey). Teenagers leaving London chose from a bigger range of destinations, usually cities with big universities, in 1996/97 than in 1990/91.

In 1996/97, about two thirds of all migrants from London chose the Rest of the South East (except teenagers for which this proportion drops to four tenths). However, when short distance moves are concerned (FHSAs sharing a boundary with London), the above proportions are circa 50% for children and mature adults, 45% for adults, older adults and pensioners, 35% for young adults and only 25% for teenagers. Generally, males move shorter distances than females.

Table 5.9. Distribution of migration flows from London to the remaining 82 FHSAs in 1996/97

	FHSA	%
F 0-15	Essex	12.51
	Surrey	11.16
	Kent	11.03
	Hertfordshire	9.42
	Buckinghamshire	4.21
	East Sussex	4.04
	West Sussex	3.69
	Berkshire	3.37
	Hampshire	3.27
F 16-19	Kent	6.62
	Hampshire	5.84
	Surrey	5.83
	East Sussex	4.92
	Essex	4.73
	Hertfordshire	4.34
	Avon	4.17
	Manchester	4.03
	Cambridgeshire	3.86
	Oxfordshire	3.82
	Birmingham	3.43
	Leeds	3.22
F 20-24	Essex	8.51
	Surrey	8.30
	Hertfordshire	7.33
	Kent	7.13
	Hampshire	4.29
	East Sussex	4.11
	Berkshire	3.69
	Avon	3.01
F 25-29	Surrey	11.91
	Hertfordshire	10.76
	Essex	10.11
	Kent	7.92
	Berkshire	4.12
	East Sussex	3.76
	Hampshire	3.72
	Buckinghamshire	3.66
F 30-44	Surrey	13.29
	Essex	10.56
	Hertfordshire	10.20
	Kent	8.94
	Buckinghamshire	4.63
	East Sussex	4.22
	Berkshire	3.58
	Hampshire	3.50
	West Sussex	3.39
F 45-59	Essex	12.77
	Kent	10.67
	Surrey	10.46
	Hertfordshire	7.27
	East Sussex	5.88
	West Sussex	4.50
	Hampshire	3.58
	Devon	3.39
	Dorset	3.35
	Norfolk	3.15
F 60+	Essex	14.11
	Kent	10.91
	Surrey	9.16
	Hertfordshire	7.81
	West Sussex	5.93
	East Sussex	5.80
	Hampshire	4.17
	Dorset	3.98
	Buckinghamshire	3.13

	FHSA	%
M 0-15	Essex	12.23
	Surrey	11.15
	Kent	10.78
	Hertfordshire	9.18
	East Sussex	4.29
	Buckinghamshire	4.15
	West Sussex	3.86
	Hampshire	3.66
	Berkshire	3.47
M 16-19	Hampshire	6.67
	Kent	6.12
	Surrey	5.37
	Oxfordshire	5.22
	Cambridgeshire	4.68
	Essex	4.61
	Avon	4.58
	Manchester	4.07
	Hertfordshire	3.70
	East Sussex	3.58
	Birmingham	3.58
	Leicestershire	3.50
	Nottinghamshire	3.10
	Leeds	3.04
M 20-24	Kent	7.53
	Essex	7.51
	Surrey	7.13
	Hertfordshire	5.83
	Hampshire	4.78
	East Sussex	4.49
	Berkshire	3.40
M 25-29	Surrey	10.74
	Essex	10.16
	Hertfordshire	10.04
	Kent	8.63
	Berkshire	4.03
	East Sussex	3.74
	Hampshire	3.62
	Buckinghamshire	3.33
M 30-44	Surrey	12.94
	Essex	11.01
	Hertfordshire	9.99
	Kent	9.04
	Buckinghamshire	4.20
	East Sussex	4.03
	Berkshire	3.74
	Hampshire	3.35
	West Sussex	3.32
M 45-59	Essex	11.75
	Kent	10.97
	Surrey	10.26
	Hertfordshire	7.45
	East Sussex	5.76
	West Sussex	3.82
	Hampshire	3.44
	Buckinghamshire	3.41
	Devon	3.18
M 60+	Essex	13.55
	Kent	11.14
	Surrey	7.83
	East Sussex	7.18
	Hertfordshire	6.49
	West Sussex	5.94
	Dorset	4.26
	Hampshire	3.94
	Norfolk	3.49
	Devon	3.12

5.8 Summary

This is the first chapter of data analysis, which concerns spatial and temporal aspects of population mobility. This chapter identified and explained the spatial and temporal trends of out-, in- and net migration in FHSAs in England and Wales between 1984 and 1998. Thus, this chapter extended what is known from existing literature on trends of annual migration moves in England and Wales. It also identified areas, pairs of areas and spatial clusters of low and high in-, out- or net migration. This makes the following two chapters (migration modelling) more relevant because they will try to explain why these trends occur by looking at the socioeconomic profiles of the corresponding areas.

In the beginning of this chapter, the k-means clustering algorithm was presented which is used later in the chapter to help identifying clusters of low and high out-migration rates. Then, general temporal trends of age disaggregated out-migration rates were presented followed by the results of k-means clustering. In order to further investigate spatial clusters of low or high out-migration rates, more sophisticated methods (explanatory spatial data analysis and geographically weighted local statistics) were introduced and applied in a sample of the dataset. The above analysis concerned the out-migration data available for the period 1984-1998. The second part of the analysis was based on flow data available for the period 1991-1997. The latter allowed a more detailed investigation of flow trends with a special attention given to migratory moves of students and graduates. It also allowed the study of the contribution of either out- or in- migration rates to net migration rates. The latter allowed the identification of net population losers or gainers due to migration.

It is apparent that the NHSCR data suggest a continuation of the counter-urbanisation phenomenon in England and Wales. Several existing techniques were used to visualise the data and allow an easier identification of the above trends. Additionally, the use of ***heat maps*** as a new means for visualising migration rates over time demonstrated the importance of introducing new ways of analysing and presenting migration data.

I now examine the effects of several ecological variables on migration decisions. The following chapter not only explains what motivates people to migrate, but it also demonstrates the power of local modelling in identifying local variations in the determinants of out-migration. I also present some preliminary analysis and ideas for further investigation on improving the estimates of out-migration models. This is perhaps possible by identifying trends in the model residuals and trying to account for these trends by introducing new variables in the models or by applying robust geographically weighted regression techniques.

Chapter 6
Global and Local Models of Out-migration

In this chapter, the results of global and local out-migration models are presented. The out-migration rates of the 14 sex/age migrant groups are presented in Chapter 5.

The aim here is to examine the existence of spatial variation of the parameter estimates of the local models as well as the temporal stability in the parameter estimates of all out-migration models. The existence of spatial variation suggests that the decision to migrate is affected in different ways by the same migration determinants depending on the migrant's location in space. The existence of temporal variation in the parameter estimates suggests that the effect of a migration determinant on the decision to migrate changes over time. Another aim here is to examine if local models improve the residuals of the global models.

In order to achieve the above, global and local out-migration models are calibrated. The log-log Ordinary Least Squares (log-log OLS) method was used for the global models, whereas a Geographically Weighted version of log-log OLS (GW log-log OLS) was used for the local out-migration models. The software used for the latter is GWR 2.0 (Fotheringham et al., 2002a), which supports the calibration of the models as well as a rich set of goodness of fit statistics to evaluate the performance of these models. A detailed discussion about what GWR is and how it was applied here is presented in Chapter 4.

Global out-migration models result in a single parameter estimate for each variable in the model whereas local out-migration models result in a number (here equal to the number of observations) of parameter estimates for each variable in the model. Separate models of out-migration rates for the 14 sex/age groups and 14 years of data were calibrated. Hence, 196 sets of local models and an equivalent number of global models were calibrated over a set of 98 observations (FHSAs in England and Wales). Each model includes 13-15 independent variables. The analysis presented here results in 2,800 global and 274,400 local parameter estimates, as well as 196 global and 19,208 local intercepts.

The existence of spatial variation in the local parameter estimates of a migration determinant can be evaluated using GWR. Significant spatial variation exists when the Monte Carlo test (Hope, 1968) for a given migration determinant is equal or less than 0.05. This means that this migration determinant exhibits significant spatial variation at the 95% level; this level of significance is a standard in statistical analysis in social sciences. A description of the Monte Carlo test is presented in Chapter 4. The temporal variation can be observed by looking at the global parameter estimates as well as the mean and range of a set of local parameter estimates of a migration determinant over time.

The existence of such a big number of values motivates the need for visual representation of the results. Below (Section 6.2), two tables are presented, one summarising the statistical significance of the global parameter estimates and a second summarising the significance of spatial variation of the local parameter estimates. Both global and local results are presented for each variable used in the models along with a discussion and interpretation of the findings.

In Section 6.1 a model selection exercise is discussed, since Geographically Weighted Regression allows for different ways of local model calibration (see Chapter 4). In Section 6.2 summarized goodness of fit statistics (t-tests for global parameter estimates and Monte Carlo test for local parameter estimates) are presented and discussed. In the same section general observations on all the results are discussed. A more detailed discussion on the parameter estimates for each variable follows in Section 6.3. The significant findings are presented in more detail using tables, box-plots, graphs and maps. During the discussion, links with previous findings in the literature are made. Finally, in Section 6.3 some preliminary analysis and ideas for an attempt to reduce the model residuals is presented.

6.1 Choosing the appropriate local model

Before calibrating the local models for all data, it is necessary to select the most appropriate technique. The software GWR 2.0, a current implementation of Geographically Weighted Regression (GWR) allows four different options which are combinations of two types of kernel and two methods of bandwidth selection (Fotheringham et al., 2002a, Chapter 2).

The two types of kernel are the *fixed kernel* and the *adaptive kernel*. The FHSA centroids (the geographical reference data are attached to) have variable spatial distribution, i.e. there are areas such as London where the FHSA centroids are dense, whereas in areas such as Wales the FHSA centroids are sparse. Because it is important to ensure that a sufficient number of observations is included in each local model, the adaptive kernel is more appropriate for the data used here.

The two types of bandwidth selection are *Cross Validation* (CV) and *Akaike Information Criterion* (AIC). The bandwidth selection in the case of the fixed kernel is the selection of a *fixed distance* and in the case of the adaptive kernel is the selection of a *number of nearest neighbours*. The selection is made by the software GWR 2.0 using the sample data (out-migration and its determinants). The software supports the selection of a bandwidth using all the sample data or a subset of the sample data increasing the number of model choices. However, the sample data are only 98 cases here, therefore all the data arc uscd in determining the bandwidth selection. Figures 6.1 and 6.2 contain a cross-tabulation comparison of different forms of GWR (GW log-log OLS) for males aged 16–19 and for males aged 30–44 both in mid-year 1997/98.

```
GLOBAL REGRESSION PARAMETERS

Residual sum of squares.........  2.28721775
Effective number of parameters..  16.
Sigma...........................  0.167011674
Akaike Information Criterion.... -48.4859205
Coefficient of Determination....  0.660500568

The number of locations to fit model in all cases is 98
```

Fixed — AIC

```
** Convergence: Bandwidth= 325444.00038

GWR ESTIMATION
Bandwidth (in data units)..... 325444.
Residual sum of squares....... 1.84474078
Effective number of parameters 21.5701497
Sigma......................... 0.15535891
Akaike Information Criterion..-51.7705109
Coefficient of Determination.. 0.726178914

ANOVA
Source            SS   DF     MS       F
OLS Residuals     2.3  16.00
GWR Improvement   0.4  5.57   0.0794
GWR Residuals     1.8  76.43  0.0241   3.2912

Tests based on the Monte Carlo sign. test
Parameter  P-value
Intercept  0.74000 air_unlg   0.36000
climate_   0.92000 commut     0.10000
crime_un   0.83000 nonwh      0.70000
pnbu       0.29000 Occmig     0.06000
pardom_l   0.52000 empgro_l   0.84000
empr_l     0.12000 hhinc_l    0.31000
hprice_l   0.54000 pnrl_l     0.32000
pvac_l     0.11000 tpopn_y_   0.29000
```

Fixed — CV

```
** Convergence: Bandwidth= 131932.11343

GWR ESTIMATION
Bandwidth (in data units)..... 131932.113
Residual sum of squares....... 0.787612807
Effective number of parameters 43.8884281
Sigma......................... 0.120645551
Akaike Information Criterion.. -25.7796355
Coefficient of Determination.. 0.883091979

ANOVA
Source            SS   DF     MS       F
OLS Residuals     2.3  16.00
GWR Improvement   1.5  27.89  0.0538
GWR Residuals     0.8  54.11  0.0146   3.6943

Tests based on the Monte Carlo sign. test
Parameter  P-value
Intercept  0.19000 air_unlg   0.01000
climate_   0.51000 commut     0.18000
crime_un   0.10000 nonwh      0.74000
pnbu       0.37000 Occmig     0.27000
pardom_l   0.20000 empgro_l   0.52000
empr_l     0.16000 hhinc_l    0.47000
hprice_l   0.07000 pnrl_l     0.50000
pvac_l     0.18000 tpopn_y_   0.37000
```

Adaptive — AIC

```
Convergence: NNN=     97

GWR ESTIMATION
Number of nearest neighbours.. 97
Residual sum of squares....... 1.57928464
Effective number of parameters 25.934563
Sigma......................... 0.148035779
Akaike Information Criterion.. -51.0852645
Coefficient of Determination.. 0.765581463

ANOVA
Source            SS   DF     MS       F
OLS Residuals     2.3  16.00
GWR Improvement   0.7  9.93   0.0713
GWR Residuals     1.6  72.07  0.0219   3.2517

Tests based on the Monte Carlo sign. test
Parameter  P-value
Intercept  0.68000 air_unlg   0.31000
climate_   0.70000 commut     0.15000
crime_un   0.68000 nonwh      0.77000
pnbu       0.34000 Occmig     0.08000
pardom_l   0.53000 empgro_l   0.89000
empr_l     0.15000 hhinc_l    0.37000
hprice_l   0.57000 pnrl_l     0.45000
pvac_l     0.08000 tpopn_y_   0.27000
```

Adaptive — CV

```
** Convergence: NNN=     79

GWR ESTIMATION
Number of nearest neighbours.. 79
Residual sum of squares....... 1.01314179
Effective number of parameters 37.6560228
Sigma......................... 0.129574085
Akaike Information Criterion.. -40.0748128
Coefficient of Determination.. 0.849615952

ANOVA
Source            SS   DF     MS       F
OLS Residuals     2.3  16.00
GWR Improvement   1.3  21.66  0.0588
GWR Residuals     1.0  60.34  0.0168   3.5041

Tests based on the Monte Carlo sign. test
Parameter  P-value
Intercept  0.41000 air_unlg   0.01000
climate_   0.45000 commut     0.46000
crime_un   0.44000 nonwh      0.80000
pnbu       0.34000 Occmig     0.12000
pardom_l   0.22000 empgro_l   0.84000
empr_l     0.10000 hhinc_l    0.47000
hprice_l   0.15000 pnrl_l     0.77000
pvac_l     0.17000 tpopn_y_   0.29000
```

Figure 6.1. Model comparison for males aged 16 – 19 in 1997-98

```
GLOBAL REGRESSION PARAMETERS

Residual sum of squares......... 2.45787158
Effective number of parameters.. 15.
Sigma........................... 0.172084022
Akaike Information Criterion.... -44.367826
Coefficient of Determination.... 0.819909411

The number of locations to fit model in all cases is 98
```

Fixed — AIC

```
** Convergence: Bandwidth = 230882.25053

GWR ESTIMATION
Bandwidth (in data units)..... 230882.251
Residual sum of squares....... 1.56347322
Effective number of parameters 25.0611956
Sigma......................... 0.146408369
Akaike Information Criterion.. -55.4120586
Coefficient of Determination.. 0.88544283

ANOVA
Source           SS    DF     MS       F
OLS Residuals    2.5   15.00
GWR Improvement  0.9   10.06  0.0889
GWR Residuals    1.6   72.94  0.0214   4.1471

Tests based on the Monte Carlo sign. test
Parameter  P-value
Intercept  0.79000 air_unlg    0.72000
climate_   0.25000 commut      0.09000
crime_un   0.45000 nonwh       0.14000
pnbu       0.09000 Occmig      0.20000
empgro_l   0.03000 empr_l      0.00000
hhinc_l    0.38000 hprice_l    0.21000
pnrl_l     0.28000 pvac_l      0.72000
tpopn_y_   0.04000
```

Fixed — CV

```
** Convergence: Bandwidth= 171507.83097

GWR ESTIMATION
Bandwidth (in data units)..... 171507.831
Residual sum of squares....... 1.22211619
Effective number of parameters 32.2919127
Sigma......................... 0.136378791
Akaike Information Criterion.. -49.133984
Coefficient of Determination.. 0.910454385

ANOVA
Source           SS    DF     MS       F
OLS Residuals    2.5   15.00
GWR Improvement  1.2   17.29  0.0715
GWR Residuals    1.2   65.71  0.0186   3.8423

Tests based on the Monte Carlo sign. test
Parameter  P-value
Intercept  0.63000 air_unlg    0.81000
climate_   0.07000 commut      0.13000
crime_un   0.56000 nonwh       0.07000
pnbu       0.16000 Occmig      0.20000
empgro_l   0.03000 empr_l      0.00000
hhinc_l    0.28000 hprice_l    0.12000
pnrl_l     0.32000 pvac_l      0.53000
tpopn_y_   0.05000
```

Adaptive — AIC

```
** Convergence: NNN=     91

GWR ESTIMATION
Number of nearest neighbours.. 91
Residual sum of squares....... 1.36021996
Effective number of parameters 27.597254
Sigma......................... 0.138998391
Akaike Information Criterion.. -59.1234531
Coefficient of Determination.. 0.90033539

ANOVA
Source           SS    DF     MS       F
OLS Residuals    2.5   15.00
GWR Improvement  1.1   12.60  0.0871
GWR Residuals    1.4   70.40  0.0193   4.5099

Tests based on the Monte Carlo sign. test
Parameter  P-value
Intercept  0.86000 air_unlg    0.67000
climate_   0.10000 commut      0.18000
crime_un   0.49000 nonwh       0.04000
pnbu       0.14000 Occmig      0.30000
empgro_l   0.04000 empr_l      0.00000
hhinc_l    0.46000 hprice_l    0.24000
pnrl_l     0.28000 pvac_l      0.66000
tpopn_y_   0.03000
```

Adaptive — CV

```
** Convergence: NNN=     78

GWR ESTIMATION
Number of nearest neighbours.. 78
Residual sum of squares....... 1.11436468
Effective number of parameters 35.3006191
Sigma......................... 0.133315929
Akaike Information Criterion.. -43.387269
Coefficient of Determination.. 0.91834944

ANOVA
Source           SS    DF     MS       F
OLS Residuals    2.5   15.00
GWR Improvement  1.3   20.30  0.0662
GWR Residuals    1.1   62.70  0.0178 3.7236

Tests based on the Monte Carlo sign. test
Parameter  P-value
Intercept  0.65000 air_unlg    0.74000
climate_   0.06000 commut      0.31000
crime_un   0.60000 nonwh       0.03000
pnbu       0.17000 Occmig      0.24000
empgro_l   0.04000 empr_l      0.00000
hhinc_l    0.44000 hprice_l    0.12000
pnrl_l     0.34000 pvac_l      0.41000
tpopn_y_   0.05000
```

Figure 6.2. Model comparison for males aged 30 – 44 in 1997-98

The selection of the appropriate calibration technique can be made theoretically or empirically. Theoretically, as explained above, an adaptive kernel is more appropriate for the data used here. In terms of the bandwidth selection, AIC is a more sophisticated test than cross validation because it takes into account both the number of observations and the number of variables, to evaluate the models goodness of fit.

To serve the empirical selection of the appropriate technique, Figures 6.1 and 6.2 show the global and local goodness of fit statistics. These include the *coefficient of determination* (r-squared), the *residual sum of squares*, the *AIC score*, and the *effective number of parameters* (a measure for degrees of freedom) for both the global and local models. For the local models Anova Statistics and Monte Carlo tests for the variables have been included. The rules of thumb for the above goodness of fit statistics are:

- the higher the *coefficient of determination* (closer to 1) the better the model fit
- the lower the *residual sum of squares* the better the model fit
- the lower the *AIC score* the better the model fit

If the difference in the AIC score between two models is equal to or more than 3, then the model with the lower AIC fit is substantially better, whereas if this difference is less than 3 there is not enough evidence to argue that the one model fits better than the other (Burnham and Anderson, 2002)

Concerning males aged 16–19, a group traditionally difficult to model, the best fit is the fixed kernel AIC method of bandwidth selection approach. It is the model with the lowest AIC value (-51.77), but has also the lowest coefficient of determination. The adaptive kernel AIC method of bandwidth selection approach has also low AIC value (-51.09), but higher coefficient of determination than the latter approach. Because the rule of thumb for AIC suggests there is no significant difference between the two models (difference in AIC is 0.68 < 3) and taking into account that theoretically an adaptive kernel is more appropriate the fixed kernel AIC method of bandwidth selection approach is selected in the case of male teenagers.

In the case of mature adult males (30 – 44) the selection based on AIC is easier, since the AIC value (-59,123) for the adaptive kernel is the lowest and its difference with the second lowest AIC value (-55,412), fixed kernel AIC method of bandwidth selection approach, is more than 3. The coefficient of determination for the former approach is also higher than that for the latter, providing more evidence for the superiority of the former approach.

The fixed kernel CV method of bandwidth selection approach in both tests resulted in a very low bandwidth (132 and 172 kilometres), which will result in a low number of observations being included in some of the local calibrations (remote FHSAs). Thus, this approach (fixed kernel) has to be rejected even though it has higher coefficient of determination compared to the two approaches with AIC method of bandwidth selection.

The AIC method of bandwidth selection, independently of the kernel type, provides a clear improvement in the local model fit compared to the global model fit, which is not the case when the CV method of bandwidth selection is applied. However, the coefficient of determination is higher in all local models compared to the global models, and the Anova statistics show an improvement in the local model fit. The Monte Carlo values change in each of the four local modelling approaches (see Figures 6.1 and 6.2). However, those variables found to exhibit a significant spatial variation in their parameter estimates (Monte Carlo test should be equal or less than 0.05) remain the same in all cases (except Air Index for male teenagers).

In conclusion, based on both theoretical and empirical evidence, the most appropriate modelling technique for the out-migration data calibrated here is GWR with an adaptive kernel and an Akaike Information Criterion method of bandwidth selection.

6.2 The performance of global and local models

In this section an attempt to identify the significant findings of both local and global models is made. There are two tables, one showing the frequency of parameter estimates found to be statistically significant in the global models (Table 6.1) and a second showing the frequency of parameter estimates exhibiting significant spatial variation in the local models (Table 6.2). Although there are similarities in the organisation of the two tables, no direct connection or comparison can or should be made.

Each table has 16 columns and 20 rows. The first column is the name of each variable in the model, the next 14 columns refer to the 14 sex/age groups, and the last column is the summary frequencies across each row. The first row is the age group, the second row is the gender (sex) division of each age group, and the next 17 rows refer to the intercept of each model and the 16 variables used in the models. As explained above, the number in each cell of the table within the 14x17 matrix of figures is the frequency of a variable's parameter estimate found to be statistically significant over the 14 year time periods. Thus, each figure in the cell ranges between 0 and 14. The higher this figure the more robust is the evidence that that variable plays an important role in determining out-migration rates for that sex/age group.

The tests (t-test, Monte Carlo test) are carried out at the 0.05 significance level. The null hypothesis that a global parameter estimate is statistically significant is accepted when the t-test value is equal or over 1.99. The null hypothesis that a local parameter estimate exhibits significant spatial variation is accepted when the Monte Carlo test value is equal or less than 0.05. However, carrying out a test at the 0.05 significance level means there is still a probability of 0.05 or less that accepting the null hypothesis is false (Type II error). The 0.05 probability of finding a non-significant parameter estimate to be significant, suggests that careful analysis of the results is necessary. Thus, it can be argued that a frequency less than 3 suggests little evidence; a frequency between 3-9 suggests relatively weak evidence; and a frequency between 10-14 suggests strong evidence for a variable having systematic effect on out-migration rates. Of course, it could be that a variable has an important real effect on only one or two migrant groups and so we must tread cautiously in interpreting these results. It is important to mention again here that in Table 6.1 the significance refers to the statistical significance of a variable's parameter estimate in the global model (t-test) and in Table 6.2 the significance refers to the significant spatial variation a variable's local parameter estimates exhibit in the local model (Monte Carlo test).

The row summaries (last column in the table) provide an indication of the importance of a variable over all sex/age groups.

6.2.1 Global modelling summary

From Table 6.1, it appears that regional total population, percentage of students at term time address, occupational migration, house prices, percentage non-white and employment rate are the variables that have a significant impact on out-migration rates across most sex/age groups of migrants, whereas available new housing and percentage of vacant dwellings have no effect on out-migration rates.

There is weak evidence for air quality and climate affecting out-migration rates for children and older age groups (30+); percentage long commuters affecting out-migration rates for female teenagers, young male adults and female pensioners; employment growth affecting out-migration rates for male children and males aged 25 and over; household income affecting out-migration rates for males 16-44; percentage net re-lets affecting out-migration rates for male teenagers; percentage of students under parental domicile affecting out-migration rates for teenagers.

There is also strong evidence that percentage long commuters affects out-migration rates for male adults; crime affects out-migration rates for teenagers and young male adults; household income affects out-migration rates for young male adults.

The differences in variable significance between gender is strong in the cases of percentage of long commuting (16 – 29; 60+), crime (20 – 24), employment growth (all ages), employment rate (20 – 29, 60+), household income (16 – 44), house prices (16 – 29), and percentage of net re-lets in social sector (16 – 19). This provides some certification for gender disaggregation in modelling out-migration rates.

House prices significantly affect out-migration rates of family members (children and mature adults) and people close to retirement age. This is expected as in the case of family members there is probably a need for improving housing conditions. People close to retirement age try to profit from differences in house prices between areas in England and Wales in order to make some savings for their retirement.

Table 6.1. Frequency of statistically significant parameter estimates in the global models

	0 – 15		16 – 19		20 – 24		25 – 29		30 – 44		45 – 59		60+		
	f	m	f	m	f	m	f	m	f	m	f	m	f	m	Sum
Intercept	10	10	2	0	4	3	10	0	6	4	10	9	12	14	94
Air Index	5	5	1	0	0	0	0	0	4	6	6	4	5	6	42
Climate Index	6	7	1	1	0	2	0	0	7	8	7	9	7	7	62
% Long Commuters	0	0	6	0	0	6	0	12	0	0	2	2	6	1	35
Crime Index	0	0	10	11	2	13	0	0	0	4	0	3	0	1	44
% Non-white	0	1	7	11	1	2	5	4	10	10	11	10	9	8	89
New Housing	0	0	0	0	0	0	0	0	0	2	0	1	1	0	4
Occupational Migration	-	-	10	9	14	14	14	14	14	14	9	11	0	0	123
Employment Growth	2	6	3	1	0	0	2	5	2	6	3	6	2	4	42
Employment Rate	8	8	2	1	0	3	0	12	8	11	11	10	6	0	80
Household Income	0	2	0	5	3	7	1	11	0	5	0	1	0	2	37
House Prices	9	12	5	1	7	0	9	5	6	8	10	10	11	13	106
% Net Re-lets	2	1	0	9	0	2	1	1	1	1	2	1	2	0	23
% Vacant Dwellings	2	4	0	1	0	0	0	0	0	0	0	0	1	2	10
Regional Total Population	14	14	14	11	11	13	14	14	14	14	14	13	14	14	188
% Term Time Address	-	-	-	-	14	14	12	12	-	-	-	-	-	-	52
% Parental Domicile	-	-	5	5	-	-	-	-	-	-	-	-	-	-	10

6.2.2 Local modelling summary

Each time a reference to a local out-migration model is made in this section it refers to the set of 98 calibrated models for one year and one sex/age group (there are 14x14 = 196 of them). Table 6.2 shows the frequency of years the local parameter estimates of each variable in each local out-migration model exhibits significant spatial variation. An immediate conclusion is that most of the parameter estimates in the local models do not exhibit significant spatial variation. However, some do. These are the local parameter estimates of employment rate for all age groups; those of Climate Index for children, and migrants aged 30 years older and over; the local parameter estimates of percentage long commuters for children, adults and mature adults; the local parameter estimates of percentage non-white for mature adults; the local parameter estimates of house prices for female adults; and the local parameter estimates of regional total population for mature male adults.

Males 30–44, females 20–24, children (0 – 15) and females 30 – 44 are the sex/age migrant groups that appear to have much greater spatial variation in the local parameter estimates of the determinants of their out-migration rates compared to the remaining sex/age migrant groups. It is also interesting to see the differences in significant spatial variation in the local parameter estimates of variables among males and females of a single age group. For example, the local parameter estimates of Climate Index exhibit significant spatial variation for female children, those of regional total population for mature male adults, and those of house prices for female adults only.

Table 6.2. Frequency of parameter estimates exhibiting significant spatial variation in the local models

	0 – 15		16 – 19		20 – 24		25 – 29		30 – 44		45 – 59		60+		
	f	m	f	m	f	m	f	m	f	m	f	m	f	m	Sum
Intercept	0	0	2	1	1	0	0	0	1	0	0	0	1	0	6
Air Index	0	0	0	1	0	0	1	0	0	0	0	0	0	0	2
Climate Index	8	3	2	1	0	0	1	0	0	6	5	3	6	2	37
% Long Commuters	5	8	0	1	0	0	7	3	3	9	0	1	2	1	40
Crime Index	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
% Non-white	0	1	1	0	0	1	0	3	11	10	1	0	3	1	32
New Housing	0	0	2	0	0	1	4	3	2	0	1	2	0	0	15
Occupational Migration	-	-	2	3	0	0	1	0	0	0	1	0	1	0	8
Employment Growth	1	3	2	3	0	0	2	0	3	2	2	4	2	2	26
Employment Rate	13	11	6	3	5	6	13	13	12	12	12	9	10	5	130
Household Income	3	1	0	0	2	0	4	2	3	3	0	3	0	0	21
House Prices	4	3	1	0	2	1	6	1	1	0	1	2	1	2	25
% Net Re-lets	1	2	2	3	1	2	0	1	1	1	1	4	0	0	19
% Vacant Dwellings	2	0	0	0	1	4	0	0	0	0	0	0	2	1	10
Regional Total Population	0	0	5	0	1	0	0	2	0	9	2	3	2	1	25
% Term Time Address	-	-	-	-	1	5	0	0	-	-	-	-	-	-	6
% Parental Domicile	-	-	0	0	-	-	-	-	-	-	-	-	-	-	0

More details on the significance of the local parameter estimates of each variable are provided in the following section (Section 6.3)

6.3 Out-Migration Determinants

In this section, the global and local parameter estimates for all out-migration determinants are presented. A discussion concerns mainly the significant findings. The local parameter estimates are too many and the inclusion of non-significant findings is unnecessary. The complete set of global parameter estimates is presented.

Air Quality (AIR_UNLG)

Air Quality is a composite variable that has both positive (poor air quality) and negative (good air quality) values. A positive parameter estimate indicates that poor air quality increases the number of out-migrants and good air quality decreases the number of out-migrants, *ceteris paribus*. A negative parameter estimate suggests the opposite effect, i.e. poor air quality deters out-migration and good air quality encourages out-migration. The latter is not an expected observation as not many people would want to live in an area with poor air quality, *ceteris paribus*. For some people, especially young adults, air quality is not a serious concern. However, in order to explain potential negative parameter estimates, it is necessary to give other explanations for this observation.

Poor air quality is a characteristic of large cities and industrial cities. These places are attractive to teenagers and young adults for other reasons than air quality (e.g. working and living opportunities). Thus, it can be argued that air quality can explain some of the urban effect. Originally there was a variable to measure urbanity of an area (urban index), but this was omitted from the models during preliminary analysis because it was highly correlated with other variables.

Table 6.3. Air Quality global parameter estimates

	84-85	85-86	86-87	87-88	88-89	89-90	90-91	91-92	92-93	93-94	94-95	95-96	96-97	97-98
Females 0-15	**0.084**	**0.092**	0.058	**0.076**	**0.065**	0.059	0.049	0.025	0.014	**0.071**	0.052	0.051	0.064	0.067
Males 0-15	**0.080**	**0.096**	0.052	**0.080**	**0.066**	**0.067**	0.045	0.012	0.009	0.063	0.059	0.048	0.065	0.066
Females 16-19	0.046	0.043	0.025	0.045	**0.052**	0.044	0.046	0.032	0.007	0.020	0.017	0.026	0.033	0.036
Males 16-19	0.001	-0.001	-0.026	-0.004	0.009	-0.004	-0.005	-0.017	-0.040	-0.024	-0.028	-0.006	0.000	-0.010
Females 20-24	-0.005	-0.007	-0.035	-0.018	-0.018	-0.022	-0.004	0.024	0.008	0.017	-0.017	-0.039	-0.026	-0.015
Males 20-24	0.009	0.009	-0.016	0.007	0.004	-0.003	0.006	0.025	0.026	0.036	0.010	0.019	0.027	0.040
Females 25-29	0.023	0.037	0.022	0.055	0.047	0.040	0.020	0.012	-0.003	0.021	0.022	0.002	-0.012	-0.008
Males 25-29	0.026	0.038	0.028	0.052	0.045	0.042	0.022	0.006	-0.009	0.017	0.016	0.007	0.003	0.024
Females 30-44	**0.051**	**0.076**	0.055	0.059	**0.079**	0.062	0.050	0.037	0.013	0.057	0.054	0.034	0.047	**0.067**
Males 30-44	**0.056**	**0.084**	**0.056**	**0.068**	**0.086**	**0.063**	0.047	0.034	0.011	0.049	0.043	0.020	0.044	0.062
Females 45-59	**0.071**	**0.108**	**0.076**	**0.087**	**0.093**	**0.084**	0.053	0.033	0.019	0.033	0.038	0.049	0.052	0.073
Males 45-59	0.060	**0.103**	**0.061**	**0.076**	**0.087**	0.068	0.031	-0.006	-0.015	-0.001	0.007	0.023	0.030	0.047
Females 60+	0.056	**0.079**	0.067	**0.112**	**0.082**	**0.098**	**0.085**	0.049	0.041	0.057	0.045	0.043	0.048	0.064
Males 60+	**0.070**	**0.097**	**0.082**	**0.091**	**0.104**	**0.085**	0.072	0.034	0.032	0.037	0.023	0.024	0.033	0.061

Table 6.3 shows the global parameter estimates of Air Quality for all years of study and for each of the 14 sex/age migration groups. The figures with bold characters are the statistically significant parameter estimates. This variable has significant parameter estimates for children, mature adults, older adults and pensioners in the 1980s, suggesting that families with children and elderly people are more sensitive to air quality. Many of the parameter estimates of Air Quality are close to 0 suggesting little, if any, effect on out-migration rates. It is apparent that for male teenagers and young female adults Air Quality has non-significant negative parameter estimates for most of the 14 year time periods, whereas in most of the

remaining groups the parameter estimates are positive. There is no obvious temporal pattern, although there are temporal variations (ups and downs) in the parameter estimates for most of the sex/age groups. These general findings match with previous analysis (Fotheringham et al., 2002b; Fotheringham et al., 2003).

The parameter estimates are very small for male teenagers and young adults, which is an expected finding. Surprisingly, the parameter estimates for female teenagers are not as similar to those for male teenagers as one would expect. Air Quality parameter estimates for female teenagers are positive and similar to those for mature adults. This is some evidence that female teenagers are less independent from their families than male teenagers. However, the latter requires further investigation.

Generally, the local parameter estimates for Air Quality do not exhibit significant spatial variation. The two out of the 196 possible cases the local parameter estimates for Air Quality exhibit significant spatial variation are more likely to have occurred by chance.

Climate Index (CLIMATE_UNLG)

Climate Index is a composite variable that has both positive (dry and warm climate) and negative (wet and cold climate) values. Its relationship with out-migration rates in the model is linear. Thus, a positive parameter estimate indicates that a dry and warm climate increases the number of out-migrants and a wet and cold climate decreases the number of out-migrants, *ceteris paribus*. A negative parameter estimate suggests the opposite effect, i.e. a dry and warm climate deters out-migration and a wet and cold climate encourages out-migration. The latter is the expected observation as one would want to live at an area with dry and warm climate, *ceteris paribus*.

Table 6.4 shows the global parameter estimates of Climate Index for all years of study and all 14 sex/age migration groups. The figures with bold characters are the statistically significant parameter estimates. This variable has positive parameter estimates for children, mature adults, older adults and pensioners, which are non-significant in the 1980s but significant in the 1990s. Climate Index for teenagers, young adults, and adults has non-significant negative global parameter estimates for most of the 14 year time periods, whereas for most of the remaining groups the global parameter estimates are positive. There is a temporal pattern, suggesting global parameter estimates for children and those 30 years old and over increase in the 1990s (reaching a peak in the mid-1990s) compared to the estimates in the 1980s.

The significant positive effect of Climate Index on out-migration rates for children, the insignificant effect on out-migration rates for mature adults or little, if any effect on out-migration rates for those aged 30 and over previously found (Fotheringham et al., 2002b; Fotheringham et al., 2003) are not confirmed here; the opposite is evident in some cases.

Table 6.4 Climate Index global parameter estimates

	84-85	85-86	86-87	87-88	88-89	89-90	90-91	91-92	92-93	93-94	94-95	95-96	96-97	97-98
Females 0-15	0.002	0.020	0.045	0.023	0.012	0.019	0.046	0.095	0.067	0.071	0.068	0.081	0.067	0.041
Males 0-15	0.000	0.020	0.065	0.036	0.002	0.022	0.028	0.070	0.067	0.072	0.070	0.071	0.071	0.045
Females 16-19	-0.053	-0.041	-0.023	-0.036	-0.034	-0.036	-0.024	-0.010	-0.020	-0.025	-0.006	0.002	0.005	-0.024
Males 16-19	-0.058	-0.035	-0.018	-0.012	-0.022	-0.013	-0.009	-0.018	-0.034	-0.043	-0.027	-0.022	-0.009	-0.012
Females 20-24	-0.015	-0.018	0.005	-0.017	-0.021	-0.049	-0.039	0.002	-0.006	-0.030	0.001	0.003	-0.015	-0.036
Males 20-24	-0.026	-0.019	0.003	-0.007	-0.021	-0.033	-0.029	-0.018	-0.027	-0.051	-0.042	-0.029	-0.053	-0.061
Females 25-29	0.008	-0.005	0.025	-0.010	-0.025	-0.023	-0.010	0.031	0.023	0.019	0.029	0.028	0.026	0.007
Males 25-29	-0.009	-0.011	0.010	-0.004	-0.042	-0.051	-0.042	-0.009	-0.007	-0.011	-0.012	-0.014	-0.029	-0.035
Females 30-44	0.021	0.034	0.053	0.016	0.006	0.026	0.042	0.079	0.071	0.077	0.072	0.076	0.078	0.053
Males 30-44	0.027	0.034	0.055	0.022	-0.005	0.010	0.026	0.065	0.076	0.084	0.080	0.079	0.076	0.057
Females 45-59	0.007	0.038	0.041	0.004	-0.005	0.027	0.040	0.083	0.078	0.088	0.097	0.100	0.089	0.065
Males 45-59	0.030	0.060	0.061	0.042	0.017	0.050	0.044	0.089	0.100	0.110	0.121	0.124	0.106	0.085
Females 60+	0.039	0.043	0.067	0.014	-0.031	-0.011	0.040	0.081	0.077	0.111	0.119	0.113	0.106	0.068
Males 60+	0.027	0.054	0.083	0.064	-0.028	0.008	0.035	0.095	0.068	0.120	0.125	0.116	0.103	0.064

Negative parameter estimates for Climate Index suggests that young people (16-29) prefer to remain in a dry and warm climate but tend to leave a wet and cold climate. The positive parameter estimates analyses for children and mature and older adults (30+) suggest these migrant groups prefer leaving areas of dry and warm climate and remaining in areas of a wet and cold climate, a rather odd finding. Climate has regional variations, in the South and South East England the climate is dry and warm whereas in Wales, West Midlands and North England it is mainly wet and cold. Thus, Climate Index may capture a regional effect on out-migration suggesting increased out-migration rates of young people from FHSAs in Wales, West Midlands and North England and decreased out-migration rates in FHSAs in the South and South East England; the reverse trend appears to apply for children and adults aged 30 years old and over.

It is interesting that there are no variations on the effect of Climate Index in out-migration rates between males and females. Climate Index has similar effect on out-migration rates for children and mature adults. This is evidence for the argument that children and mature adults are more likely to migrate in groups (families) and thus similarities between the two groups are expected for most of the variables.

A positive effect of coldness (equivalent to a negative effect of Climate index) is reported for young Canadian adults (Liaw, 1990) is confirmed here only in the case of young male adults, but there is no evidence for significant effects. However, there is relatively little variation in climate across England and Wales compared to Canada. The significantly negative effect of mean temperatures at the origins (Millington, 2000) for older adults and pensioners also contradicts the findings present above.

In some of the age groups (female children, mature male adults, older female adults, and female pensioners) for almost half of the 14 time periods the local parameter estimates for Climate Index exhibit significant spatial variation. However, there is no strong evidence for a significant spatial variation of the local parameter estimates of this variable. The findings of local models for Climate Index provide some empirical evidence that it has different effect on out-migration rates across the 98 FHSAs in England and Wales. Thus, the odd findings in the global models can be understood.

Percentage of long distance commuters (COMMUT)

The relationship between out-migration rates and percentage of long distance commuters is a power function. The values of this variable range from a minimum of 7.256 to a maximum of 49.058; the average is 26.033. A 0.213 exponent on the average percentage of long distance commuters can double the out-migration rate and a -0.213 exponent can half it, *ceteris paribus*. Since all values for the percentage of long distance commuters are positive, a positive parameter estimate suggests a positive effect on out-migration rates, whereas a negative parameter estimate suggests a negative effect on out-migration rates.

FHSAs with large numbers of long-distance commuters tend to generate greater volumes of out-migration. These are people who are more likely to move because their jobs are located well away from their residences (Fotheringham et al., 2002b).

Table 6.5. Percentage of long distance commuters global parameter estimates

	84-85	85-86	86-87	87-88	88-89	89-90	90-91	91-92	92-93	93-94	94-95	95-96	96-97	97-98
Females 0-15	-0.044	0.130	0.037	0.057	0.034	0.011	0.101	0.098	0.085	0.044	-0.038	-0.012	-0.012	0.033
Males 0-15	-0.089	0.025	-0.027	0.052	-0.021	-0.011	0.044	0.088	0.065	-0.035	-0.084	-0.074	-0.022	-0.038
Females 16-19	0.095	**0.210**	0.121	**0.175**	**0.199**	0.143	0.154	**0.198**	**0.173**	**0.162**	0.108	0.083	0.025	0.132
Males 16-19	-0.002	0.069	-0.005	0.050	0.089	0.041	0.029	0.083	0.047	-0.013	-0.105	-0.097	-0.078	-0.037
Females 20-24	-0.094	0.031	-0.063	-0.023	0.044	0.080	0.120	0.135	0.134	0.057	-0.011	-0.028	0.029	0.007
Males 20-24	0.095	**0.178**	0.080	0.109	**0.233**	**0.274**	**0.280**	**0.240**	**0.246**	0.148	0.126	0.128	0.205	0.141
Females 25-29	-0.077	0.052	0.003	0.061	0.032	0.100	0.127	0.170	0.147	0.067	0.000	0.051	0.043	0.005
Males 25-29	0.104	**0.223**	**0.222**	**0.244**	**0.229**	**0.310**	**0.302**	**0.337**	**0.356**	**0.283**	0.210	**0.297**	**0.333**	**0.293**
Females 30-44	-0.076	0.040	-0.027	0.043	-0.014	-0.010	0.046	0.056	0.010	-0.005	-0.046	-0.003	-0.027	0.004
Males 30-44	-0.060	0.061	0.012	0.067	0.009	0.009	0.037	0.074	0.041	0.020	-0.041	0.005	0.067	0.043
Females 45-59	0.027	**0.275**	0.142	**0.256**	0.175	0.083	0.138	0.206	0.109	0.124	0.082	0.042	-0.018	0.149
Males 45-59	-0.041	**0.206**	0.147	**0.257**	0.168	0.065	0.091	0.138	0.105	0.083	0.014	0.023	-0.039	0.099
Females 60+	0.075	**0.337**	**0.246**	**0.400**	**0.284**	0.209	0.229	**0.279**	**0.235**	0.099	0.135	0.114	0.083	0.189
Males 60+	-0.005	0.193	0.173	**0.343**	0.229	0.114	0.111	0.104	0.168	-0.009	0.051	-0.011	0.044	0.113

Table 6.5 shows the global parameter estimates of percentage of long distance commuters for all years of study and all 14 sex/age migration groups. The figures with bold characters are the statistically significant global parameter estimates. The global parameter estimates for this variable are negative (but occasionally positive) for male teenagers and mature female adults, and positive for female teenagers, young male adults, and male adults.

For some of the sex/age migration groups the global parameter estimates are mainly positive, but occasionally negative. These groups are children, young female adults, female adults and those aged 45 and over.

Most of the global parameter estimates are not significant; however, there is some evidence that percentage of long distance commuters has a strong positive effect on out-migration rates for female teenagers, young male adults, male adults and female pensioners confirming previous findings (Fotheringham et al., 2002b; Fotheringham et al., 2003). For example, in 1993-94 the global parameter estimate for male adults is 0.356. In FHSAs such as Bromley, Essex, Barking with Havering, Bexley with Greenwich, and Richmond with Kingston, where percentage of long distance commuters ranges from 40 to 49, the factor of this determinant ranges from 3.72 to 4, a rather strong positive effect on out-migration rates.

It is interesting to study the local parameter estimates in order to investigate where local variation exists, and what are the spatial trends of this variation. Figure 6.3 shows the local parameter estimates of percentage of long distance commuters for all years and sex/age groups. The local parameter estimates for most of the sex/age migration groups exhibit some spatial variation. This is significant for children, female adults and mature male adults. The local parameter estimates get both positive and negative values for most of the 14 sex/age migration groups (except female teenagers, young male adults, male adults, and female pensioners). The significant findings for children should be compared with those for mature adults, since the rates of those ages 30 – 44 were used in models of migrants aged 0 – 15.

Figure 6.4 shows the local parameter estimates of percentage long distance commuters for mature male adults in 1994-95 as well as a choropleth map of the values of this variable. These two pieces of information allow a better interpretation of the effect of the percentage long distance commuters on out-migration rates. This effect varies across space. There is an obvious North-South divide: in the North of England the effect is positive, in Midlands and North Wales the effect is low occasionally positive or negative and in South Wales and South England (especially the Southeast England) the effect is negative.

The above findings suggest that in FHSAs of North England the higher the percentage of long distance commuters the higher the out-migration rates, whereas in FHSAs in South England the higher the percentage of long distance commuters the lower the out-migration rates. Thus, people living in the North are more likely to move out because their jobs are not located close to their residence. However, people living in the South tend not to change their residence if they commute long distances to get to their jobs, probably because they already live in a suitable residential area outside London.

Figure 6.3. Boxplots of the local parameter estimates for percentage of long distance commuters

The global estimator for the local parameter estimates for mature male adults in 1994-95 is -0.041, suggesting a weak, non-significant negative effect stationary across all FHSAs in England and Wales, a rather misleading finding. In outer London and Home Counties, most of the families live in better housing conditions than the families in inner London. Much of the high population in these areas is the result of counterurbanization (Champion, 1989). These people moved there for better housing or because of housing availability and they are prepared to commute long distances, mainly to inner London. However, in North England, there is more housing availability. Thus, long distance commuters are more likely to migrate closer to their work location. This explains the positive affect of percentage of long distance commuters on out-migration rates in these areas.

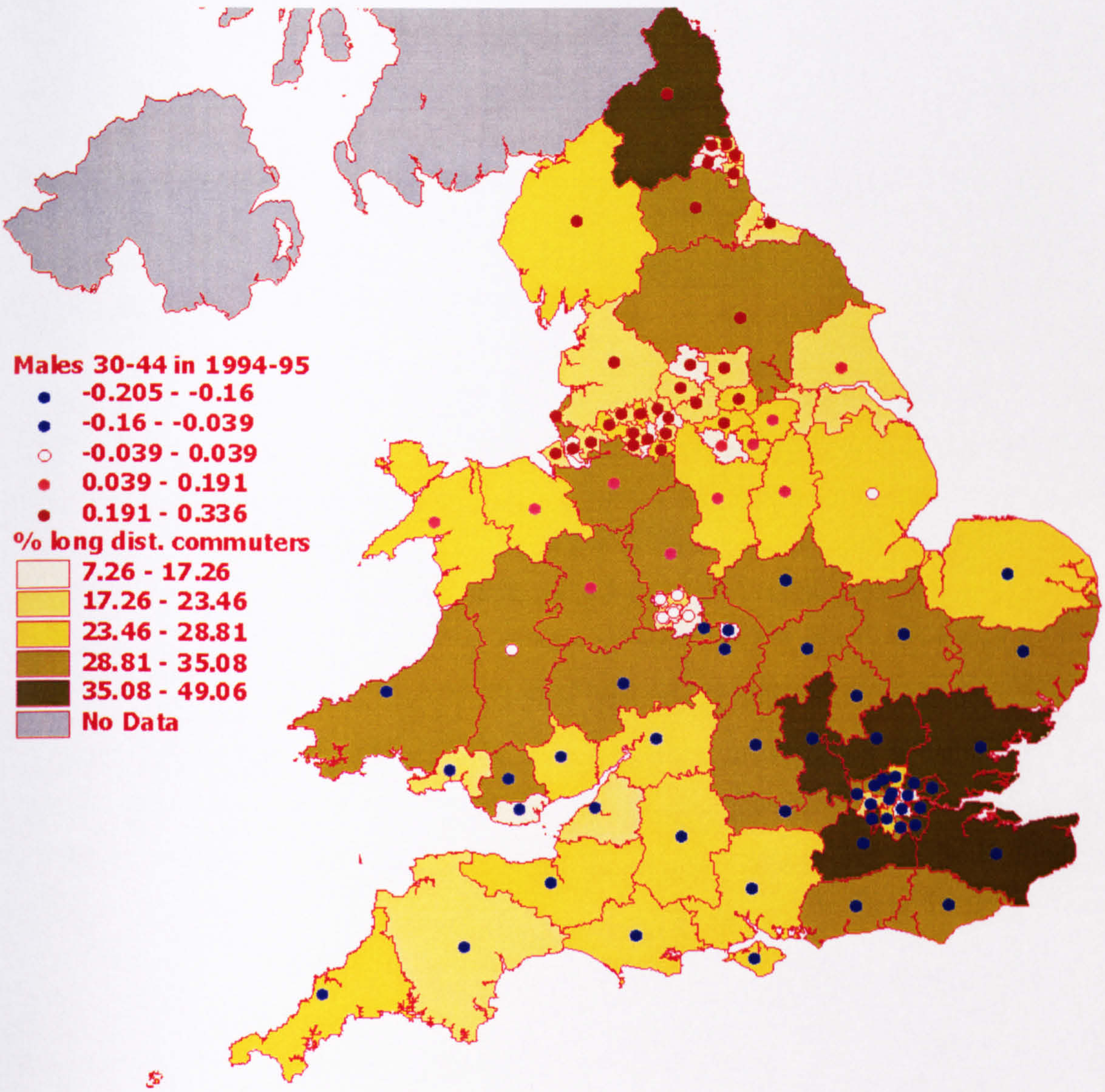


Figure 6.4. Map of percentage long distance commuters and its local parameter estimates for mature male adults in 1994-95

Crime Index (CRIME_UNLG)

Crime Index is a composite variable that has both positive (high crime) and negative (low crime) values. Its relationship with out-migration rates in the models is exponential. Thus, a positive parameter estimate indicates that high crime increases out-migration rates and low crime decreases out-migration rates, *ceteris paribus*. A negative parameter estimate suggests the opposite effect, i.e. high crime deters out-migration and low crime encourages out-migration. The latter is not an expected observation as nobody would want to live in a high crime area, *ceteris paribus*. Crime is connected with deprivation, thus, it is more likely that people living in deprived (high crime) areas cannot afford to migrate, even if they would like to do so. Crime is a serious concern for families raising children and thus, it is expected to be more significant for children, adults and mature adults.

The global parameter estimates of this variable are significantly negative for teenagers and young male adults and positive for mature male adults and older male adults (significant for some time periods). Many of the global parameter estimates of Crime Index are small suggesting a little effect of Crime Index on out-migration rates, especially for children and adults. The significantly negative global parameter estimates for young people (15 – 24) provide some evidence for Crime Index to measure other effects such as deprivation. The positive global parameter estimates for mature and older male adults suggest that family decision makers tend to avoid areas with high crime rates. The non-significance of the parameter estimates of this variable for mature and older female adults suggest either that these migration groups are not effected by crime or that it is the male partner whose sensitivities to migration determinants have more weight in the couple's/family's decision to migrate.

The negative parameters estimates are in line with previous research using the same dataset (Fotheringham et al., 2002b; Fotheringham et al., 2003). However, the positive parameter estimates are in line with the expected affect of crime on out-migration. The positive parameter estimates for those aged 25 years old and over match the findings of Millington (2000) of significantly positive parameter estimates of crime at the origin in a doubly constrained gravity model.

The local parameter estimates for Crime Index do not exhibit significant spatial variation. For most of the sex/age groups the local parameter estimates are distributed around zero, suggesting a weak effect of Crime Index on out-migration rates.

Percentage non-white (NONWH)

The percentage non-white is a variable that has generally a positive effect on out-migration rates. There are two possible reasons for this: white population leaves areas with mixed ethnic population; and/or the non-white population tends to be more mobile (Fotheringham et al., 2002b; Fotheringham, 2003).

A positive parameter estimate of percentage non-white suggests the higher the proportion of non-white population is, the higher the out-migration rate is. A negative parameter estimate suggests that the higher the proportion of non-white population is, the lower the out-migration rate is, *ceteris paribus*. The values of this variable range from 0.48 (Cumbria) to 38.74 (City); the mean across England and Wales is 6.65.

Table 6.6 shows the global parameter estimates of percentage non-white. These are negative for most of the time periods for teenagers and young adults and positive for the remaining migration groups. These global parameter estimates are significant for teenagers and people 30 years old and over. The effect of percentage non-white on out-migration rates tends to be stronger in later time periods suggesting a temporal variation that could be important.

Table 6.6. Percentage non-white global parameter estimates

	84-85	85-86	86-87	87-88	88-89	89-90	90-91	91-92	92-93	93-94	94-95	95-96	96-97	97-98
Females 0-15	-0.034	0.007	0.032	0.020	-0.016	-0.019	0.020	0.048	0.046	0.055	0.050	0.040	0.047	0.023
Males 0-15	**-0.044**	-0.016	0.020	0.018	-0.022	-0.032	0.005	0.026	0.029	0.047	0.042	0.048	0.036	0.018
Females 16-19	**-0.061**	**-0.049**	-0.015	-0.036	-0.040	**-0.068**	**-0.061**	**-0.064**	**-0.053**	-0.031	-0.014	-0.018	-0.014	**-0.046**
Males 16-19	**-0.098**	**-0.077**	-0.038	-0.051	-0.054	**-0.089**	**-0.077**	**-0.114**	**-0.113**	**-0.094**	**-0.073**	**-0.094**	**-0.088**	**-0.110**
Females 20-24	-0.016	-0.002	0.024	0.005	-0.001	-0.015	-0.017	-0.008	-0.020	-0.001	0.039	**0.062**	0.050	0.023
Males 20-24	**-0.049**	-0.031	-0.007	-0.017	-0.017	-0.025	-0.030	-0.033	-0.043	**-0.061**	-0.028	-0.018	-0.021	-0.042
Females 25-29	0.012	0.036	0.053	0.034	0.010	0.007	0.033	**0.051**	0.035	0.050	**0.071**	**0.081**	**0.086**	**0.079**
Males 25-29	-0.002	0.024	0.035	0.031	0.011	0.002	0.022	0.043	0.038	0.044	**0.071**	**0.074**	**0.072**	**0.063**
Females 30-44	0.011	0.040	**0.071**	**0.059**	0.011	0.044	**0.084**	**0.094**	**0.079**	**0.098**	**0.089**	**0.105**	**0.085**	**0.071**
Males 30-44	0.006	0.027	**0.057**	**0.055**	0.004	0.028	**0.069**	**0.081**	**0.075**	**0.105**	**0.099**	**0.113**	**0.085**	**0.072**
Females 45-59	0.041	**0.067**	**0.090**	**0.077**	0.037	0.047	**0.085**	**0.113**	**0.099**	**0.120**	**0.118**	**0.125**	**0.098**	**0.081**
Males 45-59	0.037	**0.062**	**0.066**	**0.076**	0.047	0.057	**0.079**	**0.096**	**0.093**	**0.113**	**0.128**	**0.146**	**0.091**	0.072
Females 60+	0.038	0.054	**0.090**	**0.068**	0.009	0.009	0.051	**0.098**	**0.082**	**0.100**	**0.131**	**0.124**	**0.111**	**0.086**
Males 60+	0.035	0.055	**0.099**	**0.090**	0.022	0.011	0.052	**0.076**	**0.067**	**0.096**	**0.112**	**0.093**	**0.100**	0.055

Most of the non-white population is concentrated in metropolitan areas, mainly in Bradford, Manchester, Midlands (Leicestershire, Birmingham, Coventry, Sandwell, and Wolverhampton) and London. Thus, the negative effect of percentage non-white on out-migration rates of young people (16 – 24) may capture an urban effect; out-migration rates of young people are low from cities. This negative effect could also suggest that young people prefer to live in a multi-cultural environment (offered by non-white population) rather than in a traditional British-dominated cultural environment. Previous findings (Fotheringham et al., 2002b; Fotheringham et al., 2003) suggest a significantly positive effect for most migrant groups (except teenagers), which is not fully confirmed here.

The local parameter estimates for percentage non-white exhibit no significant spatial variation in most sex/age migrant groups. However, there is strong evidence that the local parameter estimates for percentage non-white exhibit significant spatial variation in the cases of mature male and female adults. This is an interesting finding suggesting that the effect of percentage non-white population varies across England and Wales. Recently the phenomenon of xenophobia in European Union increases the importance of the role non-home residents play in local communities. In terms of temporal variation, the effect is stronger in the late time periods.

Figure 6.5 shows two sample sets of boxplots of the local parameter estimates of percentage non-white. The first set of boxplots (females 20 – 24) is an example of local parameter estimates exhibiting no significant spatial variation whereas the second set of boxplots is an example (males 30 – 44) of significant spatial variation in the local parameter estimates of this variable.

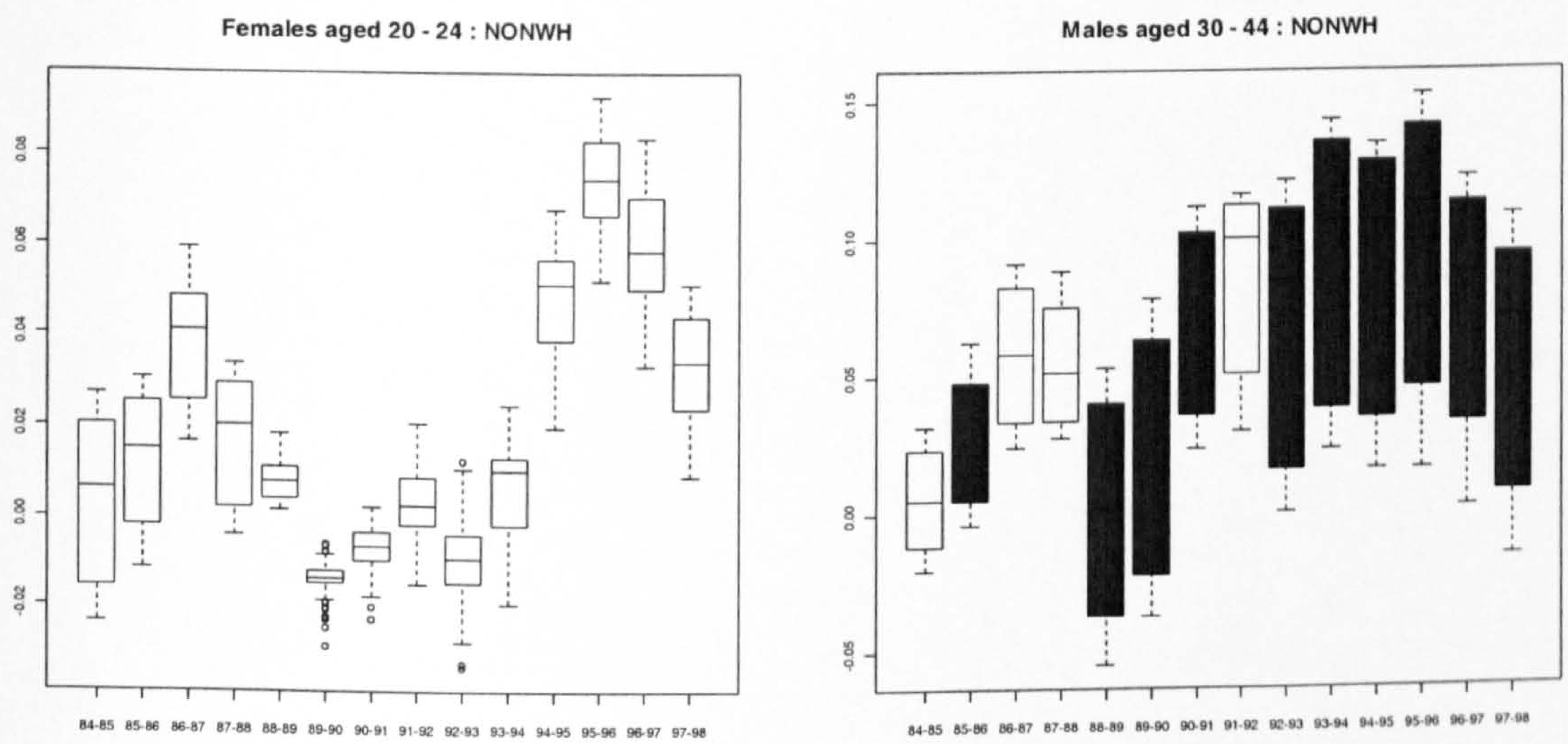


Figure 6.5. Local parameter estimates of percentage non-white for females 20–24 and males 30–44.

The global parameter estimate for males 30-44 in 1980-90 suggests a weak, non-significant positive effect of percentage non-white on out-migration rates, whereas the local parameter estimates suggest a significantly variable effect that can be negative in some locations and positive in some others. In the 1990s, these local parameter estimates range from 0 to 0.15 whereas the global parameter estimates are stationary across England and Wales ranging from 0.07 to 0.11. It is interesting to study maps in order to identify the spatial distribution of the local parameter estimates.

Figure 6.6 shows three maps: a choropleth map of the percentage non-white population, and two maps presenting the local parameter estimates for mature male adults in 1988-8 and 1996-97. The first map shows a clear spatial pattern located in the centre of the English territory with an ellipsoidal shape the southern boundary of which is London and the northern is Lancashire. The combination of percentage non-white and its local parameter estimate in each FHSA show the contribution of this determinant in out-migration rates. In the remaining two maps, the spatial pattern is a clear Southeast to Northwest divide. In 1988-89 the percentage non-white has a negative effect on out-migration rates of mature male adults in North Wales, North and Northwest England. It could be that white and non-white population integrate harmonically and there is not tension to leave the area, or that the less established and less paid non-white people cannot afford moving out of these areas compared to the Southeast where economic conditions are better. It may also be the fact that there are very low proportions of non-white people in northern areas. North England is more likely to be a final destination of immigrants from non-white populations, thus their integration level will be higher.

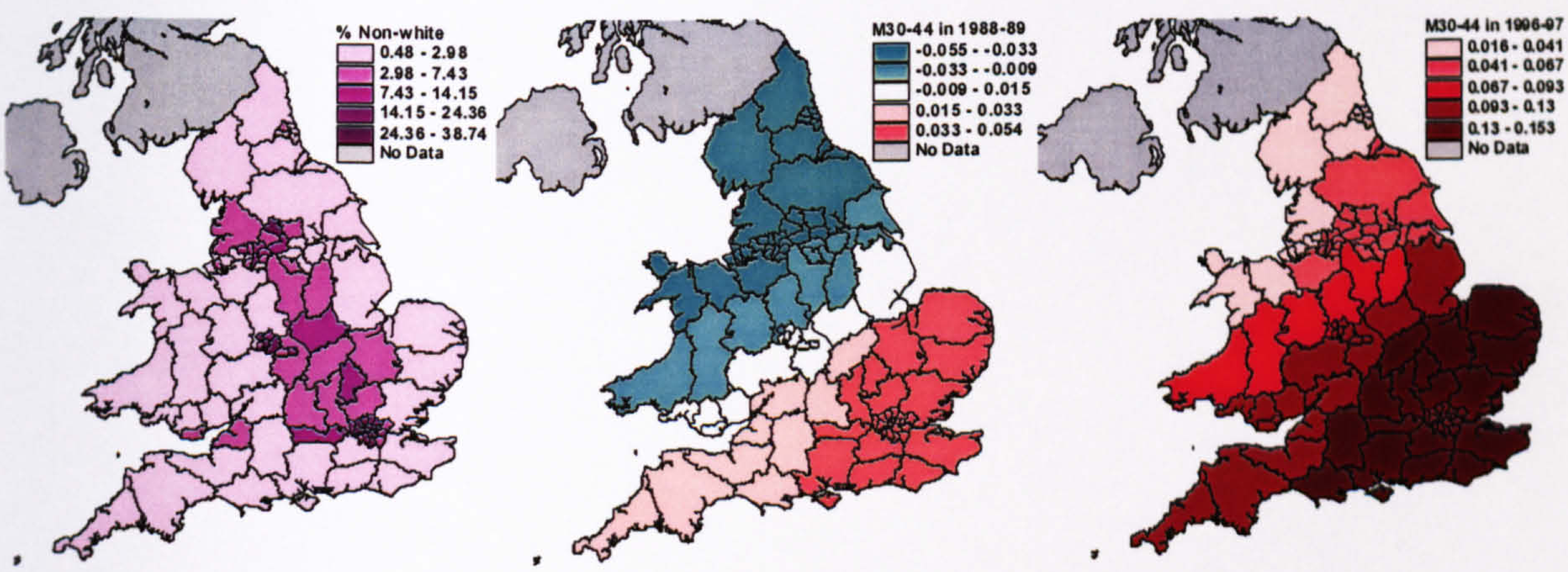


Figure 6.6. Maps of percentage non-white population and its local parameter estimates for mature male adults in 1988-8 and 1996-97.

In Southeast England there is a strong positive effect of percentage non-white on out-migration rates of mature male adults mainly due to the high mobility of non-white people as well as the tension of white people to avoid areas with high non-white population. In London, more than one fourth of the population is non-white whereas in Suffolk, Essex and Kent it is less than 3%. However, in the latter FHSAs the local parameter estimates of percentage non-white are the highest. This is a finding not easy to interpret. It is a regional relationship that cannot be connected with the spatial distribution of the percentage non-white in the Southeast. It could be locally acting as a surrogate for something else.

New housing on former urban land (PNBU)

The new housing on former urban land is a variable trying to capture the role new housing availability plays in people's decisions to migrate and it is expected to have a negative effect on out-migration rates in areas of population growth and a positive effect in areas of population decline. It is expected that the more houses available in an extensively population growth area, such as FHSAs within the South England, the more people can benefit from the housing availability and remain in the area, thus lowering the out-migration rate. In areas of population decline such as rural and remote FHSAs in North Wales and North England, new housing availability might be because of city centre regeneration projects. The latter indicates a presence of deprivation or unattractiveness in these areas in the near past which might be associated with an increase in the tendency of people to leave the area.

Previous findings (Fotheringham et al., 2002b; Fotheringham et al., 2003) suggest that this variable has generally a negative effect on out-migration rates. However, the findings here suggest there is no evidence new housing on former urban land affects out-migration rates. The global parameter estimates (Table 6.1) are not significantly in all cases and the local parameter estimates (Table 6.2) do not exhibit significant spatial variation in the majority of the cases.

Percentage of students at parental domicile (PARDOM_L)

Percentage of students at parental domicile has only been included in models for teenagers. It is expected to have a positive effect on out-migration rates because these people are more likely to migrate for university studies. The global parameter estimates for this variable are presented in Table 6.7. They are positive for both male and female teenagers. These global parameter estimates are significant in the time periods 1988-1993. In the early 1990s the effect of percentage of students at parental domicile is stronger compared to previous years. It is also significant and strong in 1997-98 for male teenagers. It is important to comment that in 1991 the Polytechnics were upgraded to Universities in England. This resulted in a significant increase in the student population because it allowed more teenagers with lower performance in their A-level exams to be able to continue their studies in higher education.

Table 6.7. Percentage of students at parental domicile global parameter estimates

	84-85	85-86	86-87	87-88	88-89	89-90	90-91	91-92	92-93	93-94	94-95	95-96	96-97	97-98
Females 16-19	0.079	0.030	0.066	0.073	**0.104**	**0.140**	**0.122**	**0.254**	**0.245**	0.102	0.052	0.081	0.105	0.097
Males 16-19	0.068	0.020	0.024	0.043	0.091	**0.136**	**0.171**	**0.306**	**0.305**	0.165	0.051	0.174	0.166	**0.269**

The findings for the effect of this variable confirm recent empirical work (Fotheringham et al., 2002b; Fotheringham et al., 2003) except that here, global parameter estimates are significant for both males and females.

The local parameter estimates for the percentage of students at parental domicile do not exhibit significant spatial variation. In some cases there is some variation, which could be presented here without statistical evidence. For example, the local parameter estimates for male teenagers in 1996-97 range from -0.139 to 0.354. Negative local parameter estimates are observed in FHSAs in Wales, Cornwell and Devon; low positive local parameter estimates in FHSAs in North England and high positive local parameter estimates in FHSAs in the Greater South East. The latter could be because of the stronger economic power of families in the South East, thus, the higher likelihood for their children go to universities compared with families living in North England and Wales. It could also be because of cultural differences between England and Wales; teenagers in Wales may stay longer with their families.

Percentage of students at term time address (TERMT_L)

Percentage of students at term time address has only been included in models for young adults and adults. It explains the effect of student population on out-migration rates for these age groups. Table 6.8 shows the global parameter estimates of this variable, which found to be significantly positive across all time periods for young adults and most time periods for adults confirming previous findings (Fotheringham et al., 2002b; Fotheringham et al., 2003). A positive parameter estimate of percentage of students at term time address suggests the higher the proportion of student population at term time address is, the higher the out-migration rate is, *ceteris paribus*. This is because many students move out of their term time address after graduation.

Table 6.8. Percentage of students at term time address global parameter estimates

	84-85	85-86	86-87	87-88	88-89	89-90	90-91	91-92	92-93	93-94	94-95	95-96	96-97	97-98
Females 20-24	0.105	0.085	0.100	0.088	0.097	0.108	0.097	0.392	0.472	0.436	0.352	0.318	0.291	0.215
Males 20-24	0.186	0.179	0.193	0.187	0.196	0.218	0.217	0.390	0.447	0.532	0.463	0.513	0.447	0.401
Females 25-29	0.055	0.048	0.073	0.066	0.067	0.093	0.071	0.166	0.167	0.113	0.094	0.142	0.184	0.155
Males 25-29	0.080	0.077	0.110	0.097	0.097	0.113	0.089	0.117	0.102	0.133	0.106	0.195	0.228	0.267

The global parameter estimates for percentage of students at term time address show an interesting temporal variation similar to that of the global parameter estimates for percentage of students at parental domicile: they are high in the early and late 1990s. In the case of young female adults the peak is in 1992-93, in the case of young male adults the peak is in 1993-94, in the case of female adults the peak is in 1996-97, and in the case of male adults the peak is in 1997-98.

The local parameter estimates exhibit some spatial variation and there is weak evidence that this is significant for young male adults only. The temporal variation of the global parameter estimates of this variable is also observed in the case of local parameter estimates. Figure 6.7 shows a set of boxplots for the local parameter estimates for young male adults and a map with the local parameter estimates of this migration group in 1994-95. The map also shows the percentage of students at term time address across all FHSAs in England and Wales, which is high in cities with universities (Newcastle, Liverpool, Manchester, Sheffield, Coventry, Devon, and part of London north of river Thames).

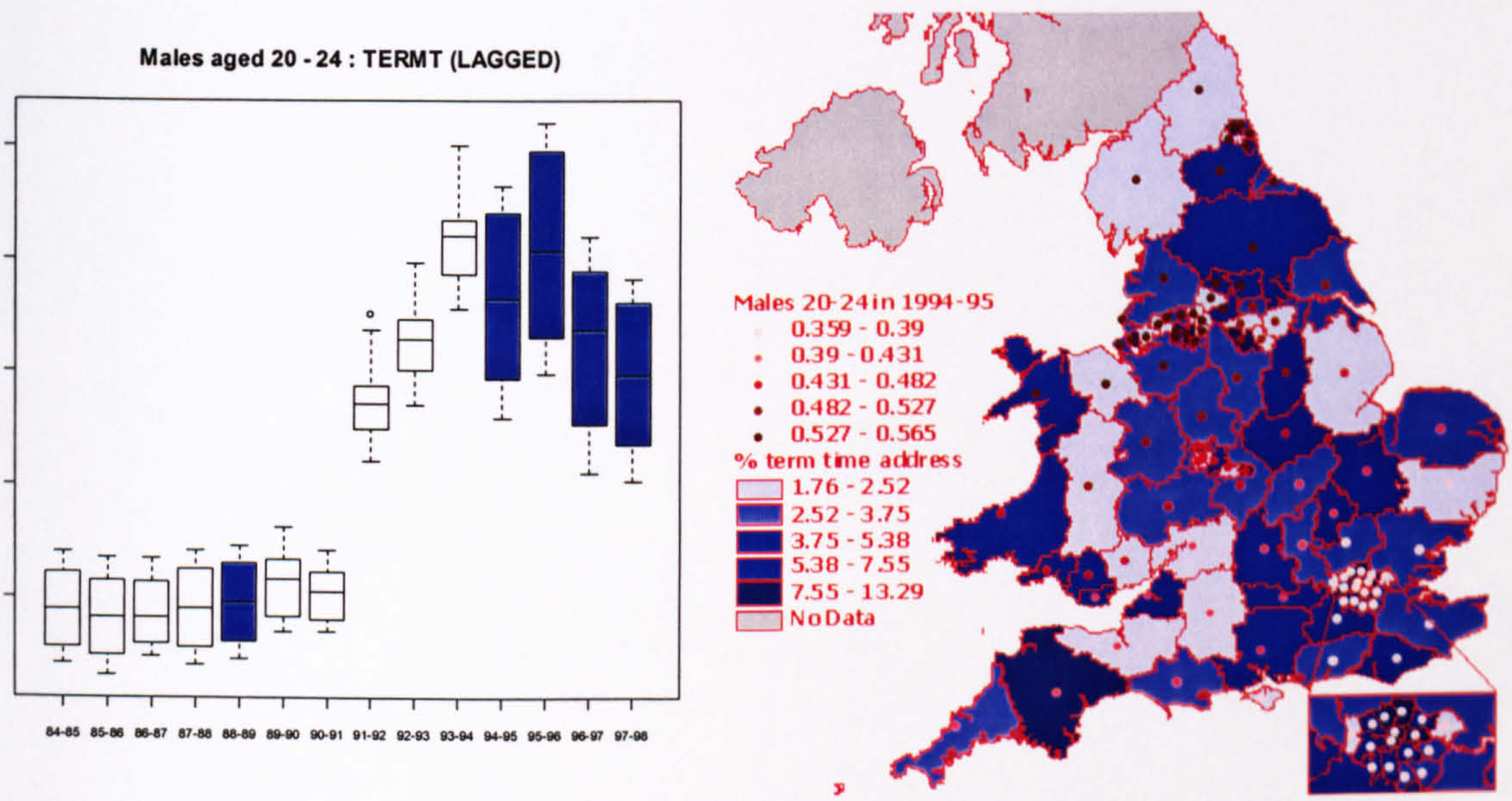


Figure 6.7. Local parameter estimates of percentage of students at term time address for males 20 – 24.

The spatial pattern of local parameter estimates is very interesting. It shows a stronger effect of percentage of students at term time address on out-migration rates in North England and a weaker effect in London and the South East. This is because most of young adults are more likely to stay for work in London and South East after their studies, because of the good working opportunities, and only small percentage will be returning migration. The majority of students in these areas are local or coming from short distance areas. In Northern England, however, most of the students come from other parts of the country and are more likely to return after the end of their studies.

Occupational migration index (OCCMIG)

Occupational migration index is expected to have a positive effect on out-migration rates. It applies only in those aged 16 and over. The global parameter estimates shown in Table 6.9 confirm the expected findings. These results provide strong statistical evidence for the significance of the global parameter estimates for those aged 16 to 59 years old. The relationship between out-migration rates and occupational migration index is exponential. Occupational migration index has values in the range 0.64 to 8.49, thus a high parameter estimate combined with a high value of this variable can contribute a lot to out-migration rates of an area. The strongest relationship is observed in the results for mature female adults in 1995-96; the global estimate is 3.836. If the latter is combined with the range of values occupational migration index has (1.74 to 2.3) then the factor on the right hand side of the equation ranges from 8.37 to 24.4, suggesting a strong positive effect.

The global parameter estimates of occupational migration index are non-significantly positive for pensioners, and surprisingly non-significantly negative for male pensioners between 1993-94 and 1997-8.

Table 6.9. Occupational migration index global parameter estimates

	84-85	85-86	86-87	87-88	88-89	89-90	90-91	91-92	92-93	93-94	94-95	95-96	96-97	97-98
Females 0-15														
Males 0-15														
Females 16-19	1.532	1.559	1.459	1.420	1.048	0.992	0.684	1.467	1.164	0.995	1.247	1.329	1.359	0.648
Males 16-19	0.867	1.098	1.163	1.110	1.051	0.873	0.933	0.778	0.626	0.579	0.848	0.537	0.350	0.089
Females 20-24	1.373	1.825	1.657	1.582	1.876	2.098	1.918	1.065	0.906	1.123	1.755	1.728	1.623	1.462
Males 20-24	1.429	1.668	1.307	1.241	1.468	1.498	1.501	1.125	1.113	0.944	1.395	1.115	1.296	1.066
Females 25-29	1.785	1.832	2.148	1.603	2.069	2.695	2.276	2.474	2.001	2.426	3.031	3.165	3.361	3.496
Males 25-29	1.696	1.640	1.758	1.444	1.765	2.150	1.771	2.016	2.065	2.087	2.452	2.371	2.668	2.452
Females 30-44	1.562	1.646	1.937	1.549	2.374	3.190	3.311	3.261	3.162	3.020	3.197	3.836	2.712	2.614
Males 30-44	1.274	1.141	1.181	1.092	1.493	1.833	1.701	1.916	1.867	1.929	1.925	2.318	1.871	1.753
Females 45-59	1.131	1.199	0.827	0.508	1.655	1.918	1.653	1.487	1.320	1.242	1.125	1.713	0.661	1.359
Males 45-59	1.341	1.672	1.143	0.960	1.892	1.989	1.530	1.345	1.127	0.842	1.089	1.846	0.466	0.983
Females 60+	0.614	0.783	1.322	1.235	0.855	1.194	0.824	0.838	0.526	0.901	1.126	1.204	0.997	0.709
Males 60+	0.071	0.145	0.642	0.256	0.627	0.570	0.334	0.266	0.191	-0.116	-0.042	-0.156	-0.078	-0.555

The local parameter estimates exhibit no significant spatial variation in most of the sex/age migration groups and across almost all years of study.

Employment growth (EMPGRO_L)

Employment growth is an indicator of the economical health of an area. Theoretically, an area with high employment growth should be a desirable place to live possibly because of low unemployment and high wages. Thus, employment growth is expected to have a negative effect on out-migration rates. However, the opposite could also be true. An area with employment growth could encourage the mobility of people so increasing out-migration rates. Many researchers (Miller, 1973; Liaw, 1990; Liaw and Kawabe, 1994; Millington, 2000) found employment growth to have a negative effect on out-migration rates, while Lowry (1966) suggests economic variables do not significantly affect out-migration rates. There is also weak evidence of a positive effect (Congdon, 1989; for 1981 census data).

The global parameter estimates of employment growth are shown in Table 6.10. There are two general observations: firstly that the effect of this variable on out-migration rate is very weak with many of the estimates less than ±0.01; and secondly that there are no consistency in these findings concerning the sign of the global parameter estimates and their statistical significance. These disappointing results may be partly because of the inclusion of other economic variables in the model (employment rate, family income) and that this variable is correlated with those other variables to some extent.

During the mid-1980s the global parameter estimates are positive for all sex/age migrant groups and significant for all groups except male teenagers, and young adults. These results confirm to some extent previous findings (Fotheringham et al., 2002b; Fotheringham et al., 2003) and suggest a consistency with findings for 1981 Census data (Congdon, 1989). During the 1990s, the global parameter estimates of children, adults and mature adults are negative and occasionally significant. Thus, in the 1980s high values of employment growth resulted in higher out-migration rates, whereas in the 1990s high values of employment growth resulted in lower out-migration rates. This temporal change is interesting and should be further investigated with the use of the 2001 Census data.

Table 6.10. Employment growth global parameter estimates

	84-85	85-86	86-87	87-88	88-89	89-90	90-91	91-92	92-93	93-94	94-95	95-96	96-97	97-98
Females 0-15	**0.031**	**0.026**	0.026	0.000	-0.021	-0.026	-0.009	-0.002	-0.002	-0.002	-0.002	-0.009	0.004	-0.009
Males 0-15	**0.036**	**0.037**	**0.051**	**0.035**	0.000	-0.009	**-0.032**	**-0.038**	-0.018	-0.005	-0.014	-0.019	-0.004	0.001
Females 16-19	**0.021**	**0.030**	**0.035**	0.024	0.004	0.014	0.000	0.012	0.010	0.007	0.008	0.006	0.020	0.000
Males 16-19	0.009	0.018	0.024	**0.033**	0.008	0.023	0.013	0.002	0.000	-0.001	-0.006	-0.012	-0.002	0.000
Females 20-24	0.018	0.017	0.016	-0.004	-0.006	0.001	-0.006	0.005	0.008	0.014	0.022	0.016	0.020	0.000
Males 20-24	0.012	0.012	0.020	0.020	0.017	0.023	0.010	-0.005	-0.001	-0.015	-0.006	-0.008	-0.009	-0.001
Females 25-29	0.018	**0.020**	0.022	-0.001	-0.009	-0.016	**-0.027**	-0.013	-0.006	-0.001	-0.002	-0.006	0.013	-0.006
Males 25-29	**0.028**	**0.023**	**0.033**	0.016	0.000	-0.002	**-0.030**	-0.024	-0.015	-0.015	**-0.019**	-0.015	-0.005	-0.002
Females 30-44	**0.021**	**0.020**	0.023	-0.006	-0.014	-0.009	-0.010	-0.007	-0.007	-0.011	-0.018	-0.022	0.008	-0.010
Males 30-44	**0.022**	**0.023**	0.026	0.015	-0.012	-0.021	**-0.039**	**-0.041**	-0.028	-0.014	**-0.023**	**-0.024**	-0.010	-0.002
Females 45-59	**0.027**	**0.034**	**0.053**	0.009	-0.019	-0.008	-0.009	0.010	-0.003	-0.009	-0.011	-0.011	0.013	-0.016
Males 45-59	**0.027**	**0.037**	**0.066**	**0.047**	0.011	-0.002	-0.029	**-0.037**	-0.013	-0.007	**-0.023**	-0.017	0.005	0.002
Females 60+	**0.023**	0.022	0.030	0.003	-0.031	-0.008	-0.011	0.020	0.014	0.000	-0.003	-0.005	0.012	**-0.020**
Males 60+	**0.036**	**0.036**	**0.063**	**0.075**	0.016	0.018	-0.019	-0.027	0.003	-0.007	-0.015	-0.015	0.005	-0.011

There is no evidence for a significant spatial variation of the local parameter estimates.

Employment rate (EMPR_L)

Employment rate is a major economic factor that is expected to be examined for its significance in out-migration rates. Fotheringham et al. (2003) suggest employment rate may be acting as a deprivation indicator. This means areas with low employment rates are deprived areas, and thus, have lower out-migration rates. However, areas of low employment rates are also likely to be associated with high out-migration rates. This is because people would be attracted to other where employment rates (and thus opportunities) are higher.

The global parameter estimates of this variable shown in Table 6.11 are positive for all migrant groups and across all periods of time. They are significant only for children, male adults, mature and older adults, and female pensioners. A positive global parameter estimate means that the higher the employment rate at an area the higher the out-migration rate is. This relationship suggests that areas with good economy generate more migrants. This could be because people living in economically good areas have strong expenditure power, thus, it is easier for them to move to a more desired place to live. It also could be that the good economy is because of highly qualified labour and invested capital, which result in a more mobile population and thus higher out-migration rates.

Table 6.11. Employment rate global parameter estimates

	84-85	**85-86**	**86-87**	**87-88**	**88-89**	**89-90**	**90-91**	**91-92**	**92-93**	**93-94**	**94-95**	**95-96**	**96-97**	**97-98**
Females 0-15	**0.323**	**0.354**	**0.318**	**0.339**	**0.353**	**0.361**	**0.351**	**0.346**	0.243	0.193	0.098	0.208	0.064	0.115
Males 0-15	**0.236**	**0.226**	**0.203**	**0.299**	**0.276**	**0.313**	**0.267**	**0.324**	0.213	0.105	0.099	0.130	0.083	0.050
Females 16-19	0.062	0.071	-0.011	0.040	0.072	0.096	**0.199**	**0.227**	0.176	0.120	0.075	0.083	-0.073	0.071
Males 16-19	0.100	0.054	0.008	0.046	0.080	0.135	0.144	**0.257**	0.174	0.068	0.019	0.085	0.017	0.064
Females 20-24	0.108	0.124	0.045	0.081	0.081	0.063	0.117	0.073	0.042	0.017	0.043	0.078	0.129	0.117
Males 20-24	0.147	0.124	0.065	0.069	0.149	**0.194**	**0.221**	0.201	0.186	0.149	0.194	0.221	**0.274**	0.204
Females 25-29	0.106	0.137	0.091	0.195	0.135	0.093	0.107	0.107	0.077	0.061	0.041	0.166	0.117	0.072
Males 25-29	0.159	**0.204**	0.194	**0.243**	**0.208**	**0.213**	**0.211**	**0.260**	**0.295**	**0.251**	**0.221**	**0.331**	**0.335**	**0.296**
Females 30-44	**0.306**	**0.274**	0.217	**0.293**	**0.243**	**0.248**	**0.255**	**0.256**	0.180	0.204	0.175	**0.313**	0.149	0.139
Males 30-44	**0.300**	**0.283**	**0.265**	**0.287**	**0.257**	**0.263**	**0.223**	**0.272**	**0.237**	0.216	0.183	**0.266**	**0.245**	0.186
Females 45-59	**0.261**	**0.362**	0.237	**0.301**	**0.320**	**0.320**	**0.317**	**0.417**	**0.332**	**0.291**	**0.309**	**0.339**	0.085	0.268
Males 45-59	0.163	**0.251**	**0.256**	**0.286**	**0.279**	**0.291**	**0.262**	**0.379**	**0.347**	0.239	**0.258**	**0.298**	0.103	0.229
Females 60+	0.197	**0.301**	0.202	**0.330**	**0.293**	**0.322**	**0.321**	**0.365**	0.232	0.200	0.210	0.232	0.120	0.191
Males 60+	0.093	0.122	0.083	0.182	0.180	0.240	0.194	0.224	0.180	0.005	0.107	0.104	0.031	0.030

This positive effect of employment rates on out-migration rates is stronger for female children, older female adults and female pensioners. Employment rates have no significant effect on out-migration rates for teenagers, young adults and female adults confirming previous findings (Fotheringham et al., 2002b). This could be because people in these ages are less sensitive to the economic conditions of an area probably because there are still in education.

Figure 6.8. Boxplots of the local parameter estimates of employment rate

The local parameter estimates of employment rate provide strong empirical evidence for the importance of local modelling. Sets of boxplots are presented for all migrant groups in Figure 6.8. The local parameter estimates exhibit significant spatial variation for most of the time periods in the case of children, adults, mature and older adults, and female pensioners. These are the same migrant groups for which the global parameter estimates were significant.

It is interesting to note that the local parameter estimates are negative for some of the FHSAs. It is important to study maps of the local parameter estimates in order to see where negative and positive local parameter estimates are located.

Sample local parameter estimates from all sex/age groups over random periods of time are mapped. They all show the same spatial patterns: there is an apparent Northwest – Southeast divide. Figure 6.9 shows two representative maps, one with both negative and positive local parameter estimates (referring to male children in 1993-94) and one with only positive local parameter estimates (referring to mature male adults in 1996-97).

The local parameter estimates suggest employment rate has a strong positive effect on out-migration rates in FHSAs in the North and Northwest England and North Wales. The effect weakens moving towards the South East England. When negative local parameters are estimated, these are for FHSAs in Southeast Midlands, and the Greater South East (Figure 6.9).

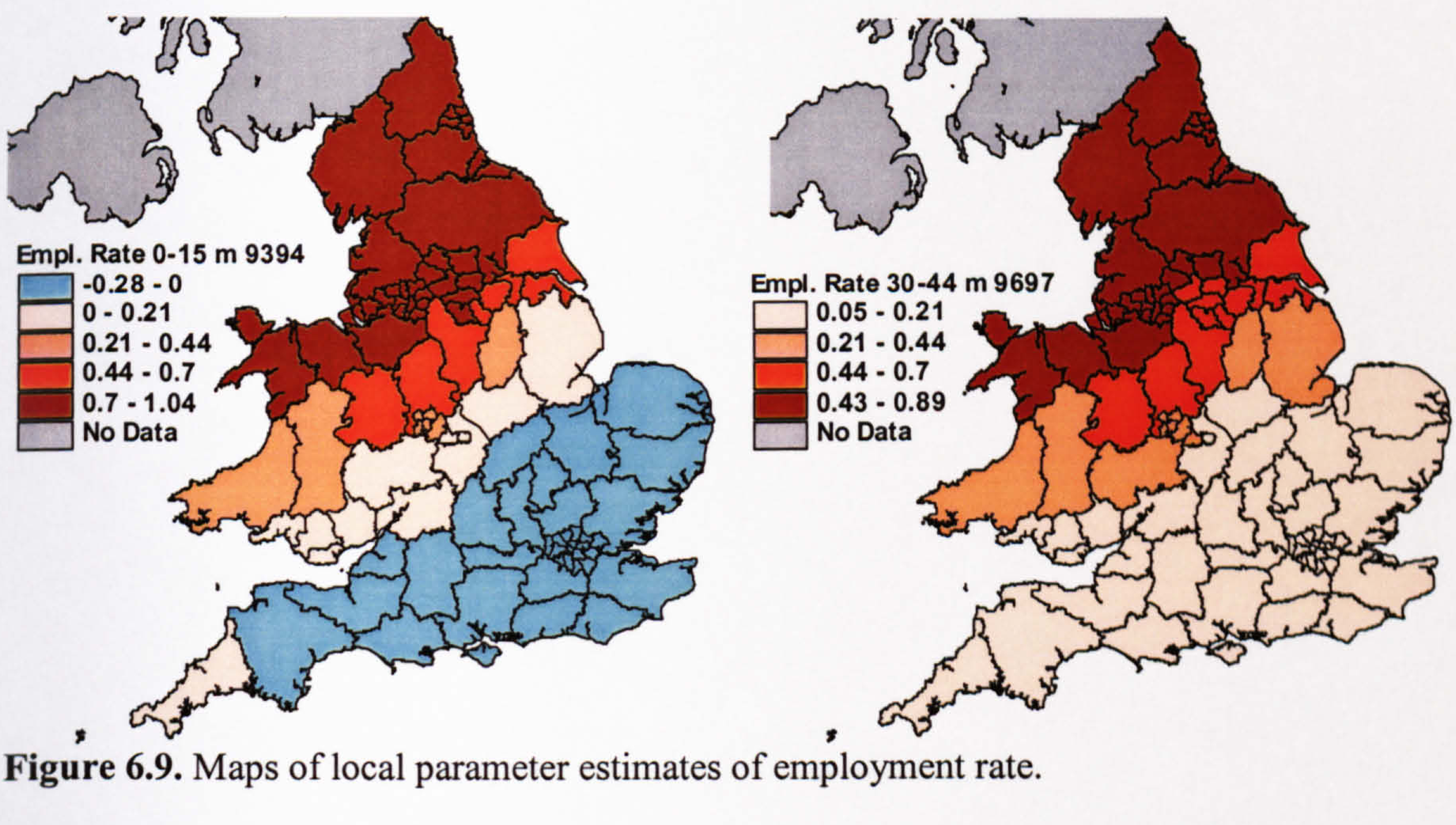


Figure 6.9. Maps of local parameter estimates of employment rate.

The spatial patterns of the local parameter estimates suggest that healthy local economies in the North and Northwest England stimulate higher proportions of the population to move somewhere else in the country. In the Southeast England there are cases that higher employment rates produce less migrants (negative local parameter estimates) and cases that the positive effect of employment rates on out-migration rates is weak. The above may prove employment rate acts as a deprivation indicator suggesting that people living in the North England are more desperate to leave their residence when the economic conditions allow it, compared with those living in Southeast England.

Household income (HHINC_L)

Household income is another economic variable that is expected to have a key role in the individual's decision to migrate. Theoretically, high household income should deter the creation of migrants. This is because satisfactory income discourages one to change his/her work and thus, change his/her residence. Of course, migration is also motivated by other factors. Household income may also act as surrogate of deprivation as well as occupation.

Highly skilled professionals and business owners should be among those having high household income. These people tend to be more mobile and thus high household income could encourage migration. It is also true that lower income people are less able to migrate. The above suggest that either negative and positive parameter estimates could be expected. Weeden (1973) gives three possible reasons for a positive effect: people with high income can afford the financial costs of a migration move; income may pick up the occupational composition of migrants (another variable accounts for this here); and high labour turnover may be associated with income rather than unemployment.

Table 6.12. Household income global parameter estimates

	84-85	85-86	86-87	87-88	88-89	89-90	90-91	91-92	92-93	93-94	94-95	95-96	96-97	97-98
Females 0-15	0.170	0.029	0.304	0.167	0.152	0.268	0.118	0.223	0.076	0.239	0.410	0.450	0.143	0.213
Males 0-15	0.100	0.027	0.217	0.126	0.149	0.293	0.117	0.222	0.068	0.246	0.580	0.526	0.207	0.209
Females 16-19	0.116	0.160	0.196	0.085	0.080	0.176	0.001	0.102	0.183	0.247	0.289	0.286	0.331	0.303
Males 16-19	0.073	0.167	0.189	0.146	0.161	0.256	0.153	0.160	0.238	0.416	0.459	0.550	0.532	0.724
Females 20-24	-0.234	-0.417	-0.137	-0.256	-0.337	-0.441	-0.380	-0.181	-0.192	-0.247	-0.125	-0.131	-0.151	-0.037
Males 20-24	-0.467	-0.549	-0.327	-0.448	-0.583	-0.582	-0.427	-0.401	-0.338	-0.288	-0.165	-0.062	-0.083	-0.049
Females 25-29	-0.183	-0.365	-0.053	-0.136	-0.269	-0.177	-0.313	-0.160	-0.282	-0.171	-0.005	-0.173	-0.327	-0.427
Males 25-29	-0.479	-0.592	-0.431	-0.436	-0.586	-0.539	-0.555	-0.439	-0.578	-0.396	-0.247	-0.348	-0.502	-0.605
Females 30-44	-0.084	-0.053	-0.002	-0.116	-0.172	-0.062	-0.227	-0.039	-0.134	0.097	0.291	0.322	0.115	0.149
Males 30-44	-0.361	-0.368	-0.328	-0.462	-0.499	-0.308	-0.358	-0.275	-0.348	-0.176	0.040	0.077	-0.092	-0.058
Females 45-59	-0.092	-0.124	-0.060	-0.063	-0.191	0.038	-0.101	-0.026	0.042	0.132	0.411	0.340	0.325	0.014
Males 45-59	-0.208	-0.342	-0.344	-0.290	-0.511	-0.102	-0.115	-0.080	0.019	0.166	0.442	0.250	0.381	0.045
Females 60+	-0.086	-0.172	-0.024	-0.322	-0.339	-0.124	-0.029	-0.011	-0.046	0.125	0.233	0.270	0.115	-0.127
Males 60+	-0.076	-0.126	0.000	-0.183	-0.418	0.063	0.244	0.247	0.122	0.420	0.595	0.601	0.235	0.072

Table 6.12 presents the global parameter estimates of this variable. These provide some evidence for the negative effect of household income on out-migration rates for young male adults, male adults and mature male adults. The latter effect is strong in some of the early time periods. The global parameter estimates are non-significantly positive for children and teenagers across all years of study and also for people aged 30 and over only after 1992-93. Focusing only on the significant findings it can be concluded that areas with high

household income will produce lower out-migration rates for males aged 20 – 44. This negative effect on out-migration contradicts findings of previous work (Fotheringham et al., 2002b; Fotheringham et al., 2003). This variable seems to have a significant effect on out-migration rates for males but not for females.

The local parameter estimates provide no strong evidence of significant spatial variation.

House prices (HPRICE_L)

House prices is an important housing variable; not only does it indicate the affordability of a location in terms of house purchase but areas with high house prices are more likely to have higher rents and higher living costs (e.g., more expensive parking, services, transport costs, entertainment, prices of goods in local shops). Thus, this variable not only accounts for housing cost but also for the general cost of living. This variable could also act as a surrogate for deprivation. However, some of the economic variables discussed above already capture deprivation effects. Thus, here house prices can explain a clear relationship with out-migration.

Table 6.13. House prices global parameter estimates

	84-85	85-86	86-87	87-88	88-89	89-90	90-91	91-92	92-93	93-94	94-95	95-96	96-97	97-98
Females 0-15	0.713	0.635	0.351	0.475	0.475	0.336	0.337	0.151	0.455	0.304	0.209	0.015	0.283	0.326
Males 0-15	0.784	0.750	0.373	0.449	0.530	0.394	0.471	0.256	0.583	0.396	0.198	0.076	0.313	0.373
Females 16-19	0.359	0.241	0.126	0.176	0.155	0.139	0.237	0.191	0.300	0.227	0.198	0.063	0.045	0.192
Males 16-19	0.383	0.185	0.040	-0.022	0.041	0.045	0.073	0.143	0.299	0.172	0.211	0.026	0.059	-0.006
Females 20-24	0.536	0.424	0.196	0.332	0.267	0.306	0.272	0.156	0.287	0.207	0.022	-0.071	-0.118	-0.115
Males 20-24	0.313	0.144	0.047	0.162	0.131	0.153	0.073	0.116	0.092	0.107	-0.011	-0.064	-0.123	-0.089
Females 25-29	0.677	0.778	0.406	0.586	0.519	0.330	0.448	0.150	0.537	0.333	0.076	0.049	0.081	0.287
Males 25-29	0.426	0.495	0.254	0.393	0.388	0.252	0.376	0.112	0.255	0.076	-0.125	-0.096	-0.081	0.105
Females 30-44	0.677	0.592	0.373	0.528	0.458	0.188	0.214	0.099	0.364	0.150	0.010	-0.253	0.162	0.254
Males 30-44	0.607	0.663	0.469	0.578	0.546	0.297	0.329	0.125	0.277	0.055	-0.061	-0.317	0.017	0.049
Females 45-59	0.959	0.682	0.595	0.670	0.573	0.274	0.417	0.248	0.448	0.361	0.158	-0.013	0.335	0.454
Males 45-59	0.936	0.677	0.644	0.603	0.599	0.283	0.497	0.351	0.509	0.491	0.189	0.005	0.395	0.462
Females 60+	1.056	0.933	0.528	0.765	0.801	0.494	0.451	0.232	0.557	0.489	0.303	0.147	0.411	0.692
Males 60+	1.282	1.119	0.686	0.742	0.976	0.586	0.605	0.376	0.697	0.632	0.416	0.283	0.584	0.887

Table 6.13 shows the global parameter estimates for house prices for all sex/age migrant groups over all periods of time. These are generally significantly positive for most of the sex/age migrant groups, but occasionally non-significantly negative (young adults, male adults, and mature adults after 1993-4). A strong significantly positive relationship between house prices and out-migration rates applies for the older migrant groups: older adults and pensioners. One explanation for this is that people close to their pension age, because of the low income their pensions provide, tend to profit from differences in house prices. To achieve this older people sell their house, which is more likely to be a big house, and buy another

house that is cheaper and smaller. This transaction usually results in a migration, in some cases long distance. There is also a South to North move of the elderly, because there are big house price differences between these parts of England.

House prices have also a significantly positive affect on out-migration rates for children, young female adults, adults and mature adults; however, the statistical evidence is not very strong. House prices have a non-significant affect on out-migration rates for teenagers and young male adults. This could be because of high percentage of migrants in these age groups are not overly concerned about the housing market. The temporal trends suggest a decline in the effect of house prices on out-migration rates over time.

The positive parameter estimates are in line with previous findings (Congdon, 1989; Millington, 2000) but contradict findings of Fotheringham et al. (2002b) and Fotheringham et al. (2003). Overall, house prices seem to be of a high importance in determining out-migration.

The local parameter estimates exhibit very little spatial variation in most of the cases.

Percentage of net re-lets in social sector (PNRL_L)

Percentage of net re-lets in social sector is a housing variable trying to capture the mobility in social sector housing. Some of the social sector housing, especial cheap city council housing is the preferable accommodation for immigrants, young adults, pensioners and people on lower incomes. For some of the above (pensioners, people receiving benefits) this is the only accommodation they can afford, thus, they are expected to have lower out-migration rates because of being more constrained in their housing choice. However, immigrants and young adults are very mobile; council housing is their short term accommodation until they improve their economical situation and then possibly move on.

Table 6.14. Percentage of net re-lets in social sector global parameter estimates

	84-85	**85-86**	**86-87**	**87-88**	**88-89**	**89-90**	**90-91**	**91-92**	**92-93**	**93-94**	**94-95**	**95-96**	**96-97**	**97-98**
Females 0-15	**0.127**	**0.112**	0.037	0.027	0.086	0.111	-0.020	0.168	0.017	-0.055	-0.040	-0.029	0.034	0.025
Males 0-15	0.071	**0.107**	0.040	0.027	0.071	0.110	-0.021	0.190	0.077	-0.069	-0.039	-0.036	0.059	0.057
Females 16-19	0.018	0.060	0.024	0.003	0.042	0.077	0.011	0.106	0.031	0.048	0.040	0.013	0.054	0.093
Males 16-19	0.108	**0.148**	**0.058**	0.041	**0.155**	**0.240**	0.132	**0.225**	**0.150**	**0.155**	0.126	0.104	**0.153**	**0.295**
Females 20-24	0.101	0.096	0.033	0.020	0.024	0.021	-0.040	0.078	0.033	0.059	0.034	-0.026	-0.005	-0.051
Males 20-24	**0.153**	**0.124**	0.044	0.024	0.036	0.087	0.064	0.097	0.022	0.094	0.091	0.037	0.013	-0.008
Females 25-29	0.015	0.048	0.013	0.007	-0.010	-0.067	**-0.138**	0.090	0.056	-0.055	-0.123	-0.074	0.027	-0.020
Males 25-29	0.009	0.019	-0.006	-0.007	-0.045	-0.101	**-0.142**	0.020	0.064	-0.052	-0.093	-0.045	-0.013	-0.078
Females 30-44	0.106	0.082	0.022	0.025	0.055	-0.064	**-0.150**	0.099	0.048	-0.086	-0.058	-0.069	0.061	-0.039
Males 30-44	0.067	0.068	0.028	0.027	0.056	-0.025	**-0.129**	0.061	0.019	-0.100	-0.061	-0.053	0.039	-0.108
Females 45-59	**0.129**	**0.131**	0.031	0.019	0.064	0.010	-0.100	0.147	0.026	-0.029	-0.031	-0.048	0.052	0.017
Males 45-59	0.065	0.081	0.038	0.025	0.021	0.000	-0.092	**0.203**	0.060	0.006	-0.045	-0.025	0.022	-0.032
Females 60+	**0.158**	**0.138**	0.037	0.028	0.082	0.060	-0.079	0.023	-0.003	-0.092	-0.012	-0.066	0.056	-0.076
Males 60+	0.065	0.081	0.032	0.043	0.082	0.141	-0.037	0.091	-0.005	-0.043	0.001	-0.009	0.054	-0.056

The global parameter estimates of this variable (Table 6.14) are insignificant for all sex/age migrant groups except male teenagers. There is some evidence that there is a positive effect of percentage of net re-lets in social sector on out-migration rates of male teenagers suggesting that the higher the percentage of net re-lets is the higher the out-migration rate of male teenagers is. There is no consistent effect of this variable on out-migration rates; in most of the migrant groups there is weak positive or negative effect. These results are in line with the lack of relationship previously reported (Fotheringham et al., 2002b; Fotheringham et al., 2003).

The local parameter estimates show only very weak evidence of any spatial variation of the parameter estimates.

Percentage of vacant dwellings in all sectors (PVAC_L)

Percentage of vacant dwellings in all sectors is a variable that measures the role of housing availability in an area. A negative effect of this variable on out-migration rate could be because in areas where there is little vacant dwellings people leave to gain housing elsewhere. A positive effect could be because of the feeling of deprivation in areas where there is a relatively high percentage of vacant dwellings.

Table 6.15. Percentage of vacant dwellings in all sectors global parameter estimates

	84-85	85-86	86-87	87-88	88-89	89-90	90-91	91-92	92-93	93-94	94-95	95-96	96-97	97-98
Females 0-15	0.036	0.051	0.064	0.041	0.094	0.147	0.152	0.170	**0.263**	0.185	**0.211**	0.089	0.116	0.097
Males 0-15	0.055	0.056	0.060	0.024	0.090	0.149	**0.174**	0.180	**0.260**	**0.233**	**0.203**	0.134	0.112	0.101
Females 16-19	0.000	0.026	0.018	0.017	0.034	0.034	0.012	0.003	0.074	0.081	0.025	-0.034	0.015	0.116
Males 16-19	0.030	0.061	0.066	0.055	0.078	0.077	0.052	0.060	0.120	0.146	0.043	0.015	0.126	**0.195**
Females 20-24	-0.036	-0.032	-0.043	-0.083	-0.074	-0.025	0.009	-0.065	0.025	0.041	-0.009	-0.060	-0.064	0.061
Males 20-24	-0.012	0.000	-0.019	-0.069	-0.059	-0.021	0.021	-0.046	0.042	-0.014	-0.021	-0.070	-0.009	0.080
Females 25-29	0.019	0.035	0.048	0.023	0.028	0.091	0.088	0.017	0.070	0.045	0.066	-0.044	-0.074	0.023
Males 25-29	0.026	0.041	0.036	-0.001	0.001	0.039	0.044	-0.021	-0.005	-0.024	0.006	-0.061	-0.094	-0.075
Females 30-44	-0.008	0.026	0.010	0.008	0.018	0.070	0.025	0.070	0.129	0.034	0.033	-0.059	0.012	0.039
Males 30-44	-0.012	0.030	0.018	0.000	0.033	0.117	0.091	0.072	0.108	0.042	0.020	-0.011	0.035	0.080
Females 45-59	0.046	0.023	0.022	-0.009	0.053	0.109	0.134	0.103	0.146	0.126	0.083	0.007	0.122	0.025
Males 45-59	0.019	-0.011	-0.017	-0.065	-0.027	0.050	0.119	0.035	0.125	0.138	0.044	-0.049	0.098	0.052
Females 60+	0.033	0.075	0.044	0.034	0.109	0.128	0.064	0.106	**0.224**	0.062	0.108	0.082	0.154	0.118
Males 60+	0.041	0.051	0.030	0.017	0.044	0.099	0.145	0.095	**0.293**	0.159	0.157	0.167	**0.190**	0.154

Both global (Table 6.15) and local parameter estimates suggest no relationship between percentage of vacant dwellings in all sectors and out-migration rates. The parameter estimates are close to zero, positive or negative and occasionally significant. However, there is no evidence to support a significant relationship.

Regional variable of the total population (TPOPN_Y_L)

The regional variable of the total population is a pull factor accounting for the effect of the population in surrounding areas on out-migration rates. Large population centres in the neighbouring areas of a particular FHSA attract migrants from this FHSA. Thus, the regional variable of the total population is expected to have a positive effect on out-migration rates: the higher the populations of the surrounding areas, the higher the out-migration rates, *ceteris paribus.*

The global parameter estimates (Table 6.16) provide strong evidence for the significant positive effect this variable has on out-migration rates for all sex/age migrant groups. The effect is stronger on out-migration rates for children and adults and weaker on those for male teenagers and young adults. In terms of temporal variation, the global parameter estimates increase during the 1980s, have a peak in 1991-92 and decline for the rest of the time periods.

Table 6.16. Regional variable of the total population global parameter estimates

	84-85	85-86	86-87	87-88	88-89	89-90	90-91	91-92	92-93	93-94	94-95	95-96	96-97	97-98
Females 0-15	0.099	0.128	0.142	0.149	0.119	0.152	0.153	0.208	0.172	0.172	0.164	0.148	0.159	0.146
Males 0-15	0.096	0.109	0.125	0.145	0.136	0.135	0.165	0.198	0.172	0.149	0.152	0.132	0.154	0.157
Females 16-19	0.136	0.148	0.151	0.161	0.154	0.158	0.154	0.164	0.150	0.124	0.131	0.118	0.124	0.141
Males 16-19	0.082	0.090	0.090	0.099	0.096	0.100	0.091	0.124	0.109	0.054	0.060	0.064	0.083	0.101
Females 20-24	0.087	0.108	0.113	0.091	0.121	0.117	0.132	0.172	0.155	0.114	0.090	0.063	0.066	0.044
Males 20-24	0.095	0.116	0.115	0.094	0.133	0.128	0.153	0.174	0.159	0.130	0.122	0.106	0.096	0.069
Females 25-29	0.114	0.125	0.144	0.151	0.124	0.140	0.144	0.201	0.187	0.169	0.138	0.144	0.126	0.096
Males 25-29	0.137	0.151	0.166	0.171	0.148	0.155	0.164	0.193	0.198	0.193	0.156	0.191	0.167	0.147
Females 30-44	0.099	0.101	0.119	0.142	0.111	0.107	0.111	0.146	0.142	0.133	0.130	0.149	0.132	0.127
Males 30-44	0.093	0.097	0.113	0.137	0.113	0.105	0.118	0.135	0.127	0.119	0.101	0.132	0.129	0.121
Females 45-59	0.102	0.118	0.148	0.162	0.135	0.115	0.141	0.168	0.150	0.148	0.158	0.141	0.147	0.102
Males 45-59	0.085	0.098	0.122	0.135	0.118	0.093	0.130	0.154	0.137	0.131	0.129	0.124	0.123	0.088
Females 60+	0.107	0.113	0.114	0.142	0.116	0.134	0.131	0.160	0.151	0.123	0.157	0.131	0.152	0.133
Males 60+	0.116	0.109	0.109	0.124	0.134	0.127	0.149	0.170	0.153	0.137	0.151	0.163	0.149	0.150

Figure 6.10 shows sets of boxplots for all sex/age migrant groups across all periods of time. In most cases local parameter estimates exhibit spatial variation; however, there is evidence that this is mainly significant for mature male adults (aged 30 – 44). All the local parameter estimates are positive.

Figure 6.10. Boxplots of the local parameter estimates of regional variable of the total population

Figure 6.11 presents a map of the local parameter estimates for mature male adults in 1996-97. There is a clear North – South divide in the effect this variable has on out-migration rates. The effect in North England is almost two times as strong as in South England, suggesting that northern English residents are more attracted to neighbouring cities.

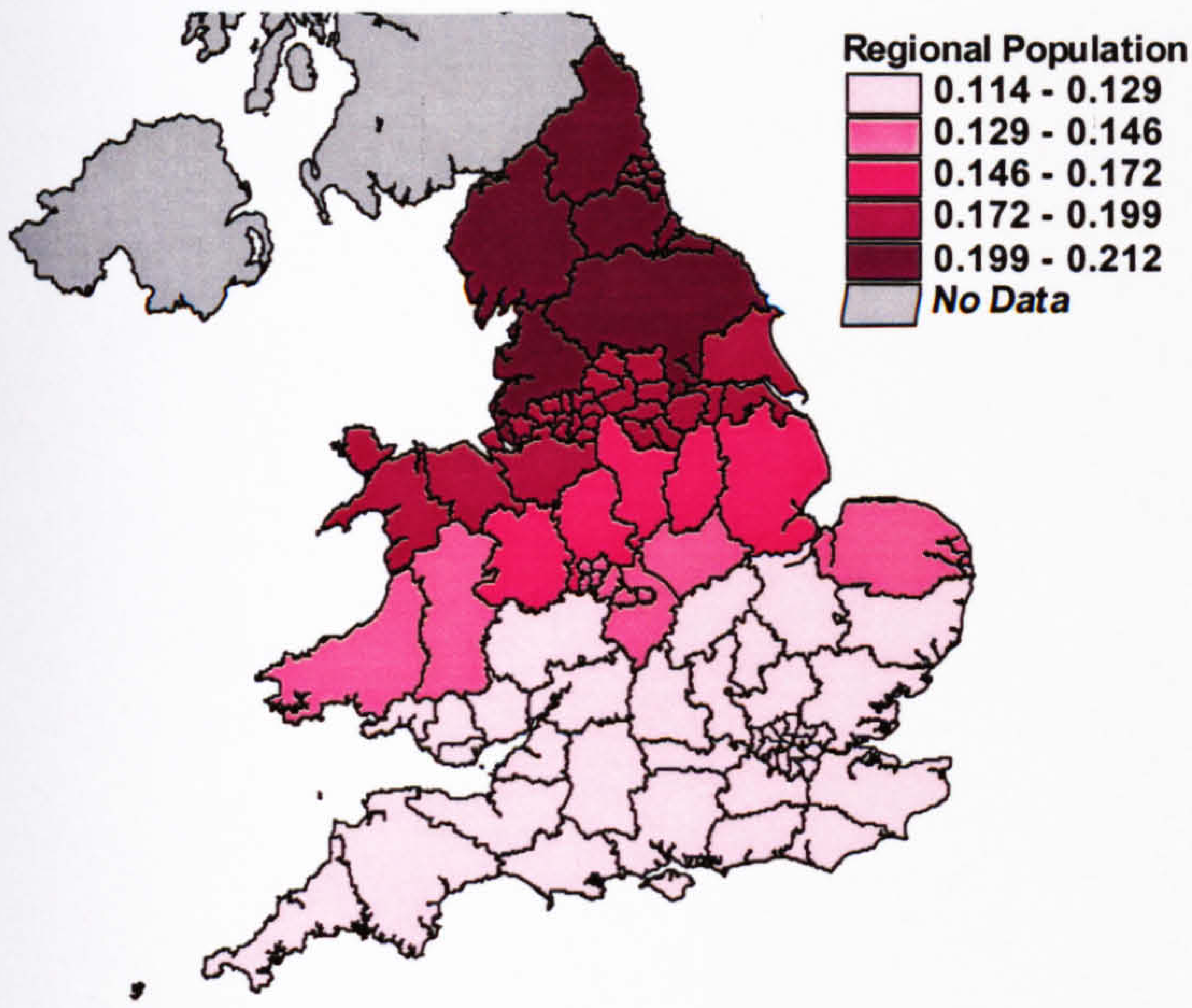


Figure 6.11. Map of local parameter estimates of regional variable of the total population.

6.4 Studying the model residuals

One of the issues concerning the performance of any model is the residuals. Even though models fit well with significant parameter estimates and high degrees of freedom (Fotheringham et al., 2002b), there are residuals, and in some cases these can be large. Here I try to study the magnitude and spatial distribution of the residuals from the out-migration model for males 30 – 44 in 1997 – 98. As shown in previous chapters, local models fit better than global models. There are several ways to improve the parameter estimates, such as increasing the number of explanatory variables and observations. It is apparent though that the residuals are not there because of generally poor model fit, but because of the failure to capture some local anomalies in migration flows at this geographical level of analysis.

In the literature, there are examples of researchers using dummy variables to capture regional differences or phenomena that cannot be quantified, such as linguistic and cultural barriers. One of the limitations of the work presented here is that in the model calibration only ecological variables were used.

Through test analysis, I found that leaving only statistically significant variables in a regression model worsens rather that improves the residuals. Here I present the parameter estimates and goodness of fit statistics for four out-migration models for males 30–44 (1997 – 98). The reason for selecting migrants aged 30-44 for my analysis here is that the models for this migrant group have the best performance. Table 8.1 presents three global log-log OLS models. The basic model (*OLS*) has 14 explanatory variables, a second model is configured by adding a London Dummy variable (*OLS (LD)*) and a third model is a variation of the second by adding the area of each FHSA (*OLS (LD-Area)*). I try to examine if the inclusion of the London Dummy and FHSA area will improve the model residuals. The fourth model is a GWR log-log OLS (*GWR*) the local parameter estimates of which are shown in Figure 8.1.

Table 8.2 shows out-migration, total population and residuals of three models (*OLS, OLS (LD)* and *GWR*) for the 98 FHSAs in England and Wales. From Table 8.2 it is evident that when a London dummy is included in the log-log OLS model the residuals improve.

Table 6.17. Out-migration models for males 30 – 44 in 1997-98: OLS (with London dummy and polygon area variables) and GWR.

	OLS (LD-Area)		OLS (LD)		OLS		GWR	
	B	*t*	*B*	*t*	*B*	*t*	*Mean t*	*Monte Carlo*
Constant	1.8910	1.18	1.8872	1.23	0.7111	0.46	0.23	0.86
AIR_UNLG	**0.0652**	2.07	**0.0652**	2.15	0.0619	1.96	2.22	0.67
CLIMATE_	0.0187	0.69	0.0188	0.70	**0.0572**	2.39	2.02	0.10
COMMUT	0.0209	0.19	0.0206	0.19	0.0430	0.38	0.14	0.18
CRIME_UN	0.0063	0.18	0.0065	0.22	0.0358	1.23	1.23	0.49
NONWH	0.0396	1.34	0.0397	1.42	**0.0724**	2.73	2.21	**0.04**
PNBU	0.1300	1.71	0.1299	1.85	0.1078	1.49	1.65	0.14
OCCMIG	**1.8530**	4.39	**1.8527**	4.45	**1.7531**	4.06	3.46	0.30
EMPGRO_L	0.0021	0.21	0.0021	0.22	-0.0022	-0.23	-0.12	**0.04**
EMPR_L	0.2340	1.98	**0.2334**	2.08	0.1857	1.61	2.55	**0.00**
HHINC_L	-0.1360	-0.58	-0.1356	-0.58	-0.0579	-0.24	0.53	0.46
HPRICE_L	-0.0495	-0.28	-0.0495	-0.28	0.0489	0.27	-0.37	0.24
PNRL_L	-0.0545	-0.63	-0.0547	-0.65	-0.1081	-1.26	-1.57	0.28
PVAC_L	0.0437	0.51	0.0435	0.53	0.0796	0.95	1.56	0.66
TPOPN_Y_	**0.0973**	2.30	**0.0975**	2.61	**0.1214**	3.20	3.54	**0.03**
LONDONDU	**0.2630**	2.78	**0.2631**	2.83				
AREA	-0.0002	-0.01						
R Square	0.836		0.836		0.820		0.900	
Adjusted R Square	0.804		0.806		0.790			
AIC					-44.36		-59.12	

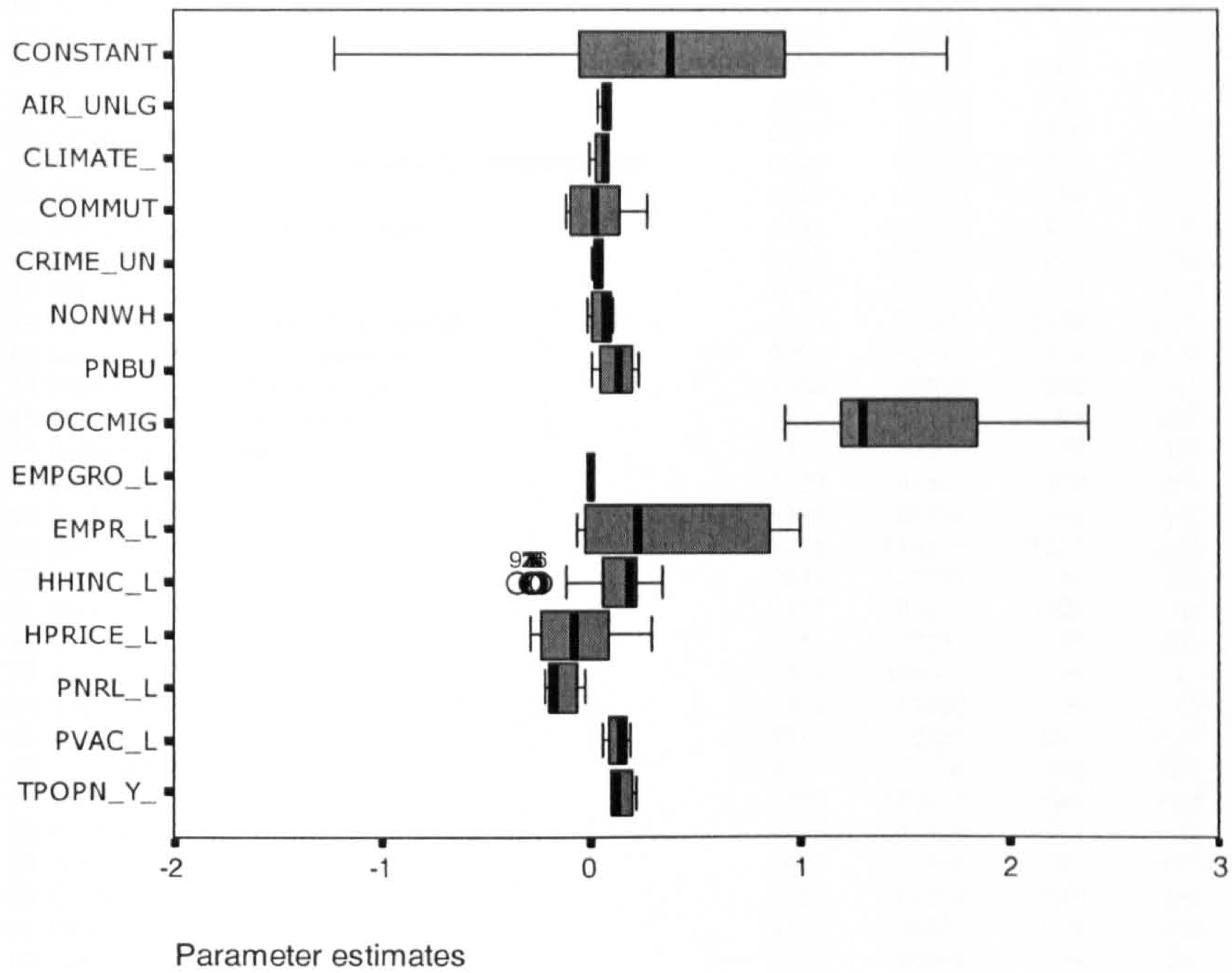


Figure 6.12. Parameter estimate of local out-migration model for males 30 – 44 in 1997-98

The local model, in some FHSAs, mainly in South England, fitted very well. Residuals of the local models are in most of the cases lower than those of any of the global models. The large residuals are also fewer and smaller than in global models. This can be seen better in the boxplots of Figure 8.2. The mean residual is 0 in all cases, however the quartiles and the range of the residuals varies with a ranking from best to worse being *GWR*, *OLS (LD)* and *OLS*

Table 6.28. Residuals of out-migration rates regressions for males 30 – 44 in 1997-98

				Residuals			% Residuals		
ID	FHSA	Migration	Population	GWR	OLS	OLS (LD)	GWR	OLS	OLS (LD)
3	Gateshead	922	23011	-47	-232	-180	-5	-25	-20
4	Newcastle	1847	32969	-110	-383	-374	-6	-21	-20
5	North Tyneside	907	21155	-119	-117	-131	-13	-13	-14
6	South Tyneside	506	17408	23	60	46	5	12	9
7	Sunderland	883	32177	-63	46	52	-7	5	6
8	Cleveland	1206	61188	563	642	635	47	53	53
9	Cumbria	1428	53154	-113	-304	-177	-8	-21	-12
10	Durham	1790	67933	67	-67	76	4	-4	4
11	Northumberland	1192	32673	-212	-273	-225	-18	-23	-19
12	Barnsley	667	26424	79	96	101	12	14	15
13	Doncaster	1077	32574	-38	-11	-9	-4	-1	-1
14	Rotherham	785	29190	156	177	169	20	22	22
15	Sheffield	1976	64503	555	771	837	28	39	42
16	Bradford	1868	55836	-18	238	191	-1	13	10
17	Calderdale	824	21622	31	-17	23	4	-2	3
18	Kirklees	1303	43346	304	474	372	23	36	29
19	Leeds	3217	87102	451	530	470	14	16	15
20	Wakefield	1131	37220	44	18	-10	4	2	-1
21	Humberside	2449	95988	248	350	375	10	14	15
22	North Yorkshire	3200	79646	-561	-696	-478	-18	-22	-15
23	Derbyshire	3206	110986	-222	0	131	-7	0	4
24	Leicestershire	2969	105299	768	904	871	26	30	29
25	Lincolnshire	2271	63484	-439	-305	-218	-19	-13	-10
26	Northamptonshire	2418	69648	0	-25	-41	0	-1	-2
27	Nottinghamshire	3760	115631	176	310	323	5	8	9
28	Cambridgeshire	3680	82940	-55	58	-75	-1	2	-2
29	Norfolk	2372	79724	156	327	187	7	14	8
30	Suffolk	2409	69432	210	473	394	9	20	16

31 Barking, Havering	2016	42148	-78	-116	97	-4	-6	5
32 Barnet	3084	41507	446	260	395	14	8	13
33 Bexley, Greenwich	2918	51338	70	-103	32	2	-4	1
34 Brent, Harrow	4578	59923	-109	-466	-336	-2	-10	-7
35 Bromley	2529	33673	-21	-92	-46	-1	-4	-2
36 Camden, Islington	6117	50985	-539	-375	11	-9	-6	0
37 City of London, Hackney, Newham, Tower Hamlets	6690	80133	-619	-287	136	-9	-4	2
38 Croydon	2750	42383	-50	113	233	-2	4	8
39 Ealing, Hammersmith, Hounslow	7981	90888	-823	-1286	-969	-10	-16	-12
40 Enfield, Haringey	5243	65741	-280	-796	-694	-5	-15	-13
41 Hillingdon	2048	31324	93	228	374	5	11	18
42 Kensington & Chelsea, Westminster	5525	55066	1670	2122	2396	30	38	43
43 Lambeth, Southwark, Lewisham	9262	106961	-816	-1535	-1045	-9	-17	-11
44 Merton, Sutton, Wandsworth	7586	89036	-840	-1519	-1064	-11	-20	-14
45 Redbridge, Waltham Forest	3954	56404	81	-358	-248	2	-9	-6
46 Richmond, Kingston	3737	41094	-74	-141	102	-2	-4	3
47 Bedfordshire	2830	67121	187	206	-62	7	7	-2
48 Buckinghamshire	4344	82760	-351	-440	-721	-8	-10	-17
49 Essex	5271	174254	1221	1335	909	23	25	17
50 Hertfordshire	5922	122478	42	162	-184	1	3	-3
51 Berkshire	5415	101627	-104	-98	-620	-2	-2	-11
52 East Sussex	3281	77197	30	343	96	1	10	3
53 Hampshire	7505	190573	50	-62	-161	1	-1	-2
54 Isle of Wight	463	11760	38	83	57	8	18	12
55 Kent	5311	169260	963	1114	636	18	21	12
56 Oxfordshire	3900	72708	-315	-354	-610	-8	-9	-16
57 Surrey	7050	125187	-661	-401	-832	-9	-6	-12
58 West Sussex	3583	79070	-268	-138	-284	-7	-4	-8
59 Avon	3842	115546	722	635	343	19	17	9
60 Cornwall	1595	47594	-148	-259	-285	-9	-16	-18
61 Devon	3719	108662	-358	-477	-322	-10	-13	-9
62 Dorset	2662	69601	39	114	73	1	4	3
63 Gloucestershire	2180	62263	139	179	156	6	8	7
64 Somerset	2057	48278	-505	-492	-489	-25	-24	-24
65 Wiltshire	2804	70488	-151	-235	-159	-5	-8	-6
66 Birmingham	4854	115159	-136	-153	-334	-3	-3	-7
67 Coventry	1379	33963	8	14	-20	1	1	-1
68 Dudley	1106	35199	162	184	192	15	17	17
69 Sandwell	1542	32143	-247	-405	-440	-16	-26	-29
70 Solihull	1137	21925	79	118	94	7	10	8
71 Walsall	1055	28742	16	34	5	2	3	0
72 Wolverhampton	1064	25102	-47	-43	-92	-4	-4	-9
73 Hereford and Worcester	2715	76154	-290	-253	-177	-11	-9	-7
74 Shropshire	1446	46208	74	30	59	5	2	4
75 Staffordshire	3243	119140	26	54	289	1	2	9
76 Warwickshire	2423	56372	-220	-236	-253	-9	-10	-10
77 Bolton	988	29561	133	265	224	13	27	23
78 Bury	854	22109	-85	27	32	-10	3	4
79 Manchester	3977	48052	-731	-1244	-1363	-18	-31	-34
80 Oldham	811	24333	140	133	155	17	16	19
81 Rochdale	1059	23330	-218	-184	-178	-21	-17	-17
82 Salford	1532	25995	-213	-356	-413	-14	-23	-27
83 Stockport	1649	32226	78	15	-42	5	1	-3
84 Tameside	1133	26443	-122	-145	-152	-11	-13	-13
85 Trafford	1350	24584	237	231	55	18	17	4
86 Wigan	903	35972	-57	101	79	-6	11	9
87 Liverpool	2350	52851	-87	156	-27	-4	7	-1
88 St. Helens & Knowsley	1237	37352	-60	-102	-102	-5	-8	-8
89 Sefton	1072	29230	-20	230	144	-2	21	13
90 Wirral	963	34296	190	490	462	20	51	48
91 Cheshire	3788	109607	347	46	192	9	1	5
92 Lancashire	4070	152937	190	595	824	5	15	20
93 Clwyd	1340	43260	75	-74	-30	6	-5	-2
94 Dyfed	944	34413	47	92	79	5	10	8
95 Gwent	1271	48823	143	188	182	11	15	14
96 Gwynedd	821	22848	-63	-50	-73	-8	-6	-9
97 Mid Glamorgan	1353	57860	14	33	104	1	2	8
98 Powys	478	12788	-44	-10	-9	-9	-2	-2
99 South Glamorgan	1810	48165	-3	136	123	0	7	7
100 West Glamorgan	948	38547	169	174	163	18	18	17

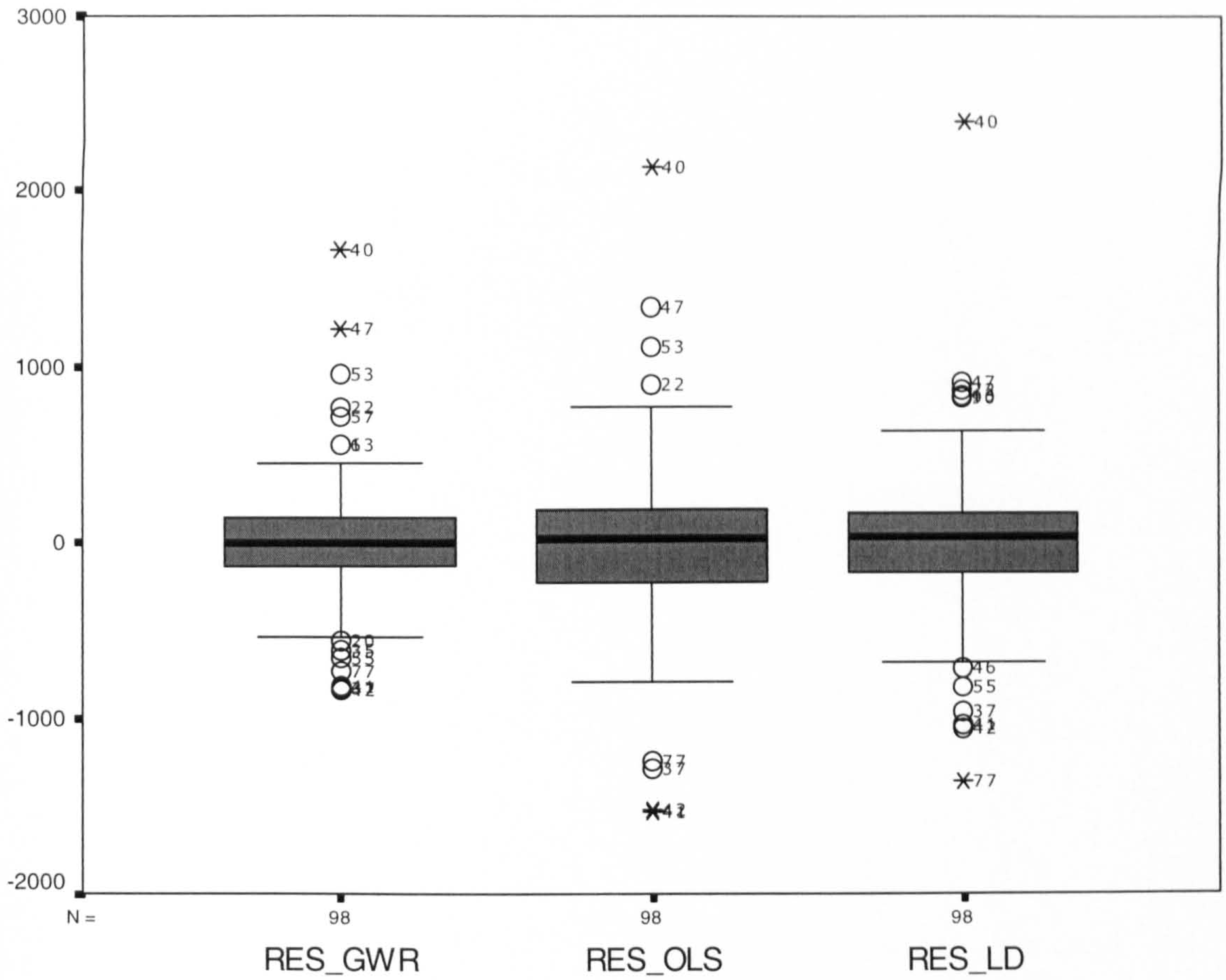


Figure 6.23. Residuals of the local model and the global *OLS* and *OLS (LD)* models respectively.

In the improved global model (*OLS (LD)*) there are 12 FHSAs with residuals over 20% (the estimated migration is 20% higher or lower of the observed migration); this figure is 8 in the case of the local model. In the case of the local model, the FHSAs where out-migration was overestimated more than 20% are, in descending order: Cleveland (47%), Kensington with Chelsea and Westminster (30%), Sheffield (28%), Leicestershire (26%), Kirklees (23%) and Essex (23%), whereas those where out-migration was underestimated more than 20% are Rochdale (-21%) and Somerset (-25%). A map of the percentages of all residuals in the local model follows in Figure 8.3.

From the map it is difficult to suggest clear spatial clusters of high or low residuals. This is rather expected since local modelling accounts for spatial variations, thus the residuals should be randomly distributed across space.

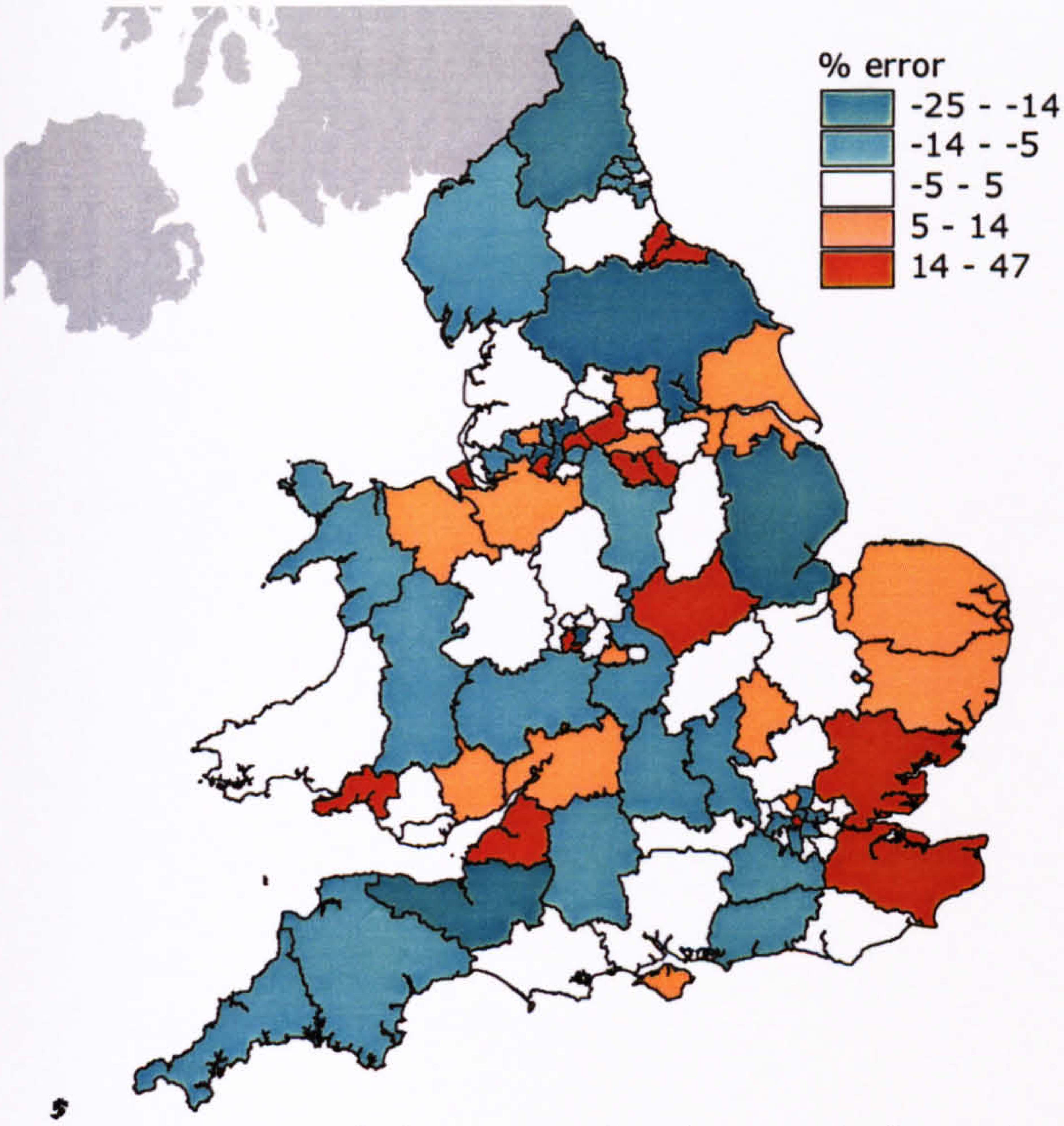


Figure 6.34. Population normalised residuals for males 30 – 44 in 1997 – 98; local model.

6.5 Summary

This is the second chapter of data analysis and the first chapter of migration modelling. Its aims were the understanding and modelling of the factors influencing the production of migrants at an origin (departure decision-making process); the examination of whether significant non-stationarity in the parameter estimates (of these factors) exists; the investigation for the presence of large residuals in estimated migration rates and the provide possible explanations for these.

To address these aims, I first conduct preliminary analysis in order to select the appropriate model configuration. I then calibrate global and local models for all data available. I present a summary of the results of these models in order to identify variables found to have a consistent significant effect. This is followed by a presentation of global and local parameter estimates through tables, boxplots and maps. Attention is given to the interesting results which are discussed in detail. Most of the insignificant results have been excluded. The discussion concerns the interpretation of the effect of each variable on out-migration rates and a possible reason for this. This is because this study focuses on explaining the behaviour of migrants and not on the predictive power of the models. However, the latter is also important in this scientific field. Thus, a section looking at the model residuals along with a discussion for their improvement is concluding this chapter.

This Chapter described the effects of several variables on people's decision to migrate (out-migration). It provided empirical evidence for a significant effect of regional total population, occupational migration, house prices and employment rate on out-migration for all migrants groups. A significant effect of percentage non-white on out-migration for those 30 and over, of long commuting on out-migration for mature male adults and of the percentage in term time address for those 20 -29 was also found.

However, the innovative part of this research is the existence of spatial non-stationarity in the local parameter estimates of some of the out-migration determinants. These include employment rate for all migrant groups, percentage non-white for mature adults and long commuting and regional total population for mature male adults. The spatial patterns of the local parameter estimates suggest a North South divide. For example, high employment rates in FHSAs in North England are strongly associated with high out-migration rates of mature male adults in 1996/97. A much weaker association is evident in South England.

Finally, it is clear that local models improve the residuals compared to global models. However, there are still large residuals signalling the need for more or other variables to be included in the model. Such variables could account for cultural variations and general appeal of an area, the measures of which are not straightforward.

I now examine the effect of several ecological variables on migrant's destination choice.

Chapter 7
Local Models of Destination Choice

In this chapter, I present the empirical findings of global and local models of migrants' destination choices. I focus my analysis on two origins: one from North-East England (Newcastle) and one from South-East England (Camden & Islington). The main reason for focusing on only two origins is that previous work (Fotheringham et al., 2002b) provided empirical findings of variations in the parameter estimates of destination choice determinants across origins in 1996/97. However, here I am interested in examining the existence of temporal variations in these parameter estimates (1990/91 – 19996/97) for a single origin. Additionally, I examine the existence of spatial non-stationarity in the local parameter estimates of destination choice determinants.

I also focus on one migrant group: mature male adults. This is because this is the largest migrant group. Many of the migrants aged 30 – 44 are household heads and thus are the major decision-makers of the family's destination selection. They also respond more directly to socioeconomic conditions of an area than other migrant groups. Thus, the migration behaviour of males 30 – 44 is particularly important for policy makers interested in understanding migration flows. The findings presented below may be useful to improve our understanding of which characteristics attract male adults (30 – 44) and their dependants to certain areas.

The remainder of this chapter includes some details on destination choice model configurations and the most important empirical findings of these models.

7.1 Introducing local spatial interaction models

Spatial Interaction Models generally try to model a flow of some kind between an origin and a destination to attributes of both locations and their distance or separation (Yano et al., 2003). When spatial interaction models of migration are concerned, this modelling takes place in two steps: relating out-migration to a set of attributes of origins (previous chapter) and relating destination choice to a set of attributes of destinations. This split in the modelling process has a theoretical explanation provided in Rees at all. (2003).

Yano et al. (2003, p. 419) provide a nice description of what migration modelling such as is presented in this chapter is meant to achieve: *measure migration behaviour in terms of* *elasticities of migration responses to various aspects of destination attractiveness.... Origin-specific migration destination choice models are calibrated to obtain this information.*

Although there is already a space disaggregation element in Spatial Interaction Models in general and migration destination choice models more specifically, the term *local forms* of these models introduced here adds a second dimension of special disaggregation: they account for the location of the destination. This results in *origin- and destination-specific migration destination choice models* or simply *local migration destination choice models.*

In praxis, when a destination choice model is calibrated for a single origin, it is assumed that the effect of a destination attribute is stationary across destinations. This assumption makes such a model *global* in terms of the terminology used in this thesis. Here I examine whether the effects of destination attributes on destination choice are not stationary over space. Allowing for such variation makes a model *local.* The latter is calibrated using GWR.

It is now transparent that the synthesis of all the results of local models will produce a two dimensional spatial disaggregation across origins and across destinations. However, it is difficult to visualise the results of such a synthesis. Thus, to keep things simple, I calibrated models for two single origins here and below I present the results only across destination for each of these origin-specific models.

This chapter demonstrates a way of extracting more information from an interaction dataset as well as removing the inaccuracy of traditional global models by accounting for spatial non-stationarity in the parameter estimates of destination choice.

7.2 Destination choice modelling mechanics

Here I discuss some of the technical details concerning my analysis on destination choice models. Overall, a set of global and a set of local models were calibrated using Poisson regression. Traditional global destination choice models for migrants leaving Newcastle were calibrated for all seven time periods (1990/91 – 1996/97) and 14 migrant groups. The same variables were used as in Fotheringham et al. (2002b), except for Destination Accessibility, which I had to recalculate. An equivalent number of local models were also calibrated. This is possible using GWR Poisson in R, which was made available to me by Chris Brunsdon. Finally, two sets of models (global and local) equivalent to the above were calibrated for those leaving Camden and Islington, but this time only for one migrant group (males 30 – 44). Summaries and significant findings of these models are presented and discussed below.

The reason for applying a Poisson regression model is because it is more appropriate for regressing counts of individuals than OLS regression and it handles the existence of zero migration flows without further calculations (in OLS for example to overcome the problem of zero flows it is necessary to add 0.5 to all flows and then calculate an adjustment factor to ensure the total estimated flows match the total observed).

Destination Accessibility was recalculated in order to provide time series data. An available measure included Northern Ireland (NI) and Scotland in its calculations. However, I believe it was more appropriate to remove NI and Scotland from the calculations for this variable.

A new version of the GWR software used in the analysis presented in Chapter 6 is now available that supports Poisson regression. However, it did not perform well with my dataset. Instead, the R code for GWR Poisson performed well. An additional advantage of R is that it is possible to calibrate all global and local Poisson models at the same time using loops.

For the local model calibrations, I selected an adaptive kernel and 78 nearest neighbours. The decisions for these were based on findings of preliminary analysis. An indicator for the overall fit of a global model is the *psi* statistic (equation 4.45). The calculation of the AICc allows the comparison of the fit between two models. The statistical significance of the global parameter estimates was evaluated using the standard *t* test. In order to get a feeling of whether the local parameter estimates exhibit significant spatial variation I calculated the *l*-statistic (Fotheringham et al., 2002a):

$$l = \frac{\text{standard error of the global parameter estimate}}{\text{standard deviation of the local parameter estimate}}$$

There is no standard rule of thumb; however, a value of *l* more than 1.5 would suggest potential interesting results and a value of *l* more than 3.0 would be a good indicator that the local parameter estimates of a variable exhibit significant spatial variation.

A summary of the findings of these models follows.

7.3 The performance of global and local models

In this section, I present the general findings of global and local destination choice models. Table 7.1 (equivalent to Table 6.1) shows the frequency of parameter estimates found to be statistically significant in the global destination choice models for those leaving Newcastle. Some general findings of the local models follow.

Table 7.1. Frequency of statistically significant parameter estimates in the destination choice global models for Newcastle (as origin)

	0 – 15		16 – 19		20 – 24		25 – 29		30 – 44		45 – 59		60+		
	f	m	f	m	f	m	f	m	f	M	f	m	f	m	Sum
Intercept	3	1	5	4	7	7	6	7	5	7	1	5	2	0	60
Climate Index	2	2	3	3	1	1	1	1	2	1	1	1	1	0	20
Crime Index	2	2	0	3	4	2	3	4	3	4	0	0	1	1	29
Council Tax	1	2	2	1	4	6	1	2	1	1	0	0	0	0	21
Destination Accessibility	7	5	2	1	1	2	7	3	7	6	6	5	6	5	63
Household Income	0	2	0	1	3	0	4	4	1	2	0	0	1	0	18
House Prices	7	7	4	4	7	7	7	6	7	7	6	7	6	5	87
Listed buildings	6	7	7	7	7	5	6	5	7	7	6	6	4	4	84
New Housing	0	3	0	0	1	4	1	0	1	1	0	0	2	2	15
% Net Re-lets	2	4	3	4	2	3	3	1	2	2	0	1	1	0	28
Total Population	7	7	7	7	7	7	7	7	7	7	7	7	7	7	98
New buildings (Private)	3	3	3	2	1	2	2	0	1	1	1	1	3	1	24
New buildings (Social)	4	5	3	1	5	4	3	5	1	3	2	0	1	1	38
% Vacant Dwellings	3	4	1	2	4	3	1	0	4	1	2	1	1	0	27
Poor condition (Private)	1	0	1	1	5	4	1	0	2	1	0	0	1	0	17
Poor condition (LA)	3	4	2	2	4	1	2	3	5	5	2	4	0	1	38
Vacant and derelict	5	5	0	1	1	1	5	5	6	7	7	6	4	5	58
Employment Growth	5	4	0	2	0	5	1	4	3	5	3	2	2	1	37
Employment Rate	1	2	1	1	2	1	3	1	4	2	1	2	1	1	23
Distance	7	7	7	7	7	7	7	7	7	7	7	7	7	7	98
Age/Sex Unemployment Rate	-	-	2	2	3	3	4	1	2	2	1	1	-	1	22
Contiguity	7	7	7	7	6	6	6	7	7	7	6	7	6	5	91
Term time address	-	-	7	7	7	2	-	-	-	-	-	-	-	-	23
Parental domicile	-	-	-	-	3	4	6	7	-	-	-	-	-	-	20

In Table 7.1 each cell represents the frequency a variable found to be significant in the time-specific global Poisson models. Because there are seven time periods in total (1990/91 – 1996/97), the frequency can range from zero to seven. It can be argued that a frequency less than 2 suggests little evidence; a frequency between 3-4 suggests relatively weak evidence; and a frequency between 5-7 suggests strong evidence for a variable having a systematic effect on destination choice. It could be that a variable has an important real effect on only one or two migrant groups and so we must tread cautiously in interpreting these results.

Based on the above assumptions the global models provide strong evidence for a systematic effect of population, distance and contiguity on destination choice for all migrant groups. There is also strong evidence for a significant effect on migrants' destination choices of house prices (all migrant groups except for teenagers); listed buildings (all migrant groups except for pensioners); destination accessibility and vacant and derelict dwellings (all migrant groups except for teenagers and young adults); and local authority housing in poor condition (mature adults). Finally, the percentage of students at term time address has a systematic effect on the destination choices of teenagers and young female adults as well as the percentage of students at parental domicile has on destination choices of adults.

There are some variables that seem to have little or no significant affect on migrants' destination choices. These include climate conditions, council tax (except young adults), household income (except adults), new housing on former urban land, new building completions in private sector, private housing in poor condition (except young adults), employment rates and age/sex specific unemployment rates.

The results of the local models I calibrated here suggest local parameter estimates of most variables exhibit a degree of spatial variation (judging from the *l* statistic). Two examples of local parameter estimates are shown in Tables 7.3 and 7.5. Usually, this variation is significant for the variables the global parameter estimates of which are significant. These include distance, vacant and derelict dwellings, destination accessibility, house prices and listed buildings. However, there are two variables, generally not significant in the global models, that have local parameter estimates exhibiting significant spatial variation in the local models. These are all vacant dwellings and private housing stock in poor condition.

The above findings refer to migrants leaving Newcastle. More details on models for males 30 – 44 are presented in the following section.

7.4 Newcastle

In this section, a full description of specific global and local models is presented. These refer to males 30 – 44 leaving Newcastle. The temporal variation of the global parameter estimates and the spatial variation of the local parameter estimates are examined in detail.

7.4.1 Time trends in destination choice determinants

Table 7.2 shows the global parameter estimates of Poisson regressions for males 30 – 44 in the 1990s. The numbers in bold fonts are those found to be significant at the 95% confidence level. At the bottom of the table, the *psi* statistic and the AIC and AICc for each of the seven models are also presented.

Table 7.2. Parameter estimates of the destination choice global models for males 30 – 44 leaving Newcastle during 1990/91 – 1996/97

	1990/91	1991/92	1992/93	1993/94	1994/95	1995/96	1996/97
Intercept	**-10.540**	**-22.117**	**-13.557**	**-18.714**	**-17.771**	**-27.021**	**-25.150**
Climate Index	-0.010	0.097	-0.024	0.075	0.058	0.043	**-0.135**
Crime Index	**0.342**	0.132	0.085	**0.296**	**0.303**	-0.001	**0.174**
Council Tax	0.277	**0.822**	-0.090	0.540	0.086	0.623	-0.065
Destination Accessibility	**-1.251**	**-1.214**	**-1.139**	**-0.813**	**-0.784**	-0.435	**-1.191**
Household Income	1.004	-0.454	-0.687	**1.288**	0.952	0.665	**1.755**
House Prices	**1.119**	**2.664**	**2.694**	**1.034**	**1.262**	**1.807**	**2.182**
Listed buildings	**0.449**	**0.327**	**0.274**	**0.333**	**0.412**	**0.284**	**0.283**
New Housing	**0.426**	-0.107	-0.129	0.116	0.144	-0.160	-0.112
% Net Re-lets	0.218	**0.647**	0.250	0.125	**0.596**	0.112	0.161
Total Population	**1.058**	**0.975**	**0.781**	**1.074**	**1.051**	**1.095**	**1.266**
New buildings (Private)	**0.225**	0.089	0.096	-0.207	-0.047	-0.141	0.077
New buildings (Social)	**0.272**	0.037	-0.147	0.059	**-0.127**	0.072	**-0.145**
% Vacant Dwellings	-0.009	0.338	-0.123	-0.157	-0.163	**0.470**	0.082
Poor condition (Private)	-0.135	-0.190	0.010	0.031	0.061	-0.003	**0.372**
Poor condition (LA)	**0.241**	**0.384**	**0.436**	**0.283**	0.150	**0.300**	0.117
Vacant and derelict	**-0.368**	**-0.278**	**-0.288**	**-0.149**	**-0.274**	**-0.209**	**-0.238**
Employment Growth	**-0.095**	-0.046	**-0.141**	**-0.087**	**-0.107**	-0.038	**0.073**
Employment Rate	-0.303	-0.039	-0.122	-0.140	0.020	**-0.515**	**-0.765**
Distance	**-1.165**	**-1.136**	**-1.111**	**-0.952**	**-1.117**	**-1.208**	**-1.216**
Age/sex Specific Unemployment Rate	0.239	-0.217	0.227	-0.083	0.211	**0.491**	**0.929**
Contiguity	**1.194**	**0.921**	**0.928**	**1.637**	**1.068**	**0.895**	**1.366**
Psi Poisson	0.2609	0.2219	0.2117	0.1982	0.1988	0.1913	0.1989
AIC	580	587	576	533	565	574	578
AICc	594	601	590	547	579	588	592

Several variables are statistically significant. An indicator for the overall model fit is the *psi* statistic. The closer to zero *psi* is the better the model fit. The *psi* statistic is zero when the model predictions are perfect. It is not straightforward to judge how good is the model fit based on the values of the *psi* statistic. It is important to note that one reason for a poorer model fit could be the fact that many of the values of the dependent variable are either zero (flows to 4 destinations in 1996/97) or very low (flows to 44 destinations are five or fewer migrants in 1996/97). However, there are still some interesting conclusions to be made out of these models.

Previous findings of destination choice models (Atkins and Fotheringham, 1999; Fotheringham and Pitts, 1995; Fotheringham et al., 2002b; Millington, 2000; Pellegrini and Fotheringham; 1999) are confirmed here: destination accessibility and distance have a significant negative effect on destination choice whereas total population and contiguity have a significant positive effect.

There is also strong evidence for a significant positive effect of house prices (confirming Fotheringham and O'Kelly, 1989; Fotheringham and Pitts, 1995; Atkins and

Fotheringham, 1999; Fotheringham et al., 2002b), listed buildings (confirming Fotheringham et al., 2002b) and local authority dwellings in poor condition as well as a negative effect of vacant and derelict dwellings (confirming Fotheringham et al., 2002b). There is also weak evidence for a significant negative effect of employment growth on destination choice.

The parameter estimates of the remaining variables do not show a systematic effect, although they are occasionally significant (for example employment rate and age/sex specific unemployment rate in recent time periods).

Generally, there are some temporal variations in the global parameter estimates. Those parameter estimates found to have a systematic effect are rather stable over time. Those parameters that change sign and value are more likely not to have a serious real effect on destination choice.

7.4.2 Spatial trends in destination choice determinants

Table 7.3 presents results from the two sets of models discussed above. These are a global model with a full set of variables (*Global Model*) and the corresponding local model (*Local Model*). These models are all calibrated with data on males 30 – 44 leaving Newcastle in 1996/97.

Table 7.3. Parameter estimates of the destination choice models (global and local) for males 30 – 44 leaving Newcastle in 1996/97

	Global Model	Local Model			
		Min		Max	*t*
Intercept	**-25.150**	-36.75	-	-1.96	2.76
Climate Index	**-0.135**	-0.11	-	0.54	2.24
Crime Index	**0.174**	-0.09	-	0.38	1.59
Council Tax	-0.065	-1.11	-	1.15	1.13
Destination Accessibility	**-1.191**	**-1.92**	-	**0.99**	4.47
Household Income	**1.755**	-1.43	-	2.16	1.83
House Prices	**2.182**	**-0.07**	-	**3.42**	3.65
Listed buildings	**0.283**	**-0.09**	-	**0.52**	3.24
New Housing	-0.112	-0.55	-	0.17	1.13
% Net Re-lets	0.161	-0.51	-	0.91	2.00
Total Population	**1.266**	0.60	-	1.43	2.18
New buildings (Private)	0.077	-0.22	-	0.51	1.57
New buildings (Social)	**-0.145**	-0.23	-	0.09	1.42
% Vacant Dwellings	0.082	**-1.04**	-	**0.86**	3.75
Poor condition (Private)	**0.372**	-0.19	-	0.56	1.68
Poor condition (LA)	0.117	**-0.36**	-	**0.54**	3.17
Vacant and derelict	**-0.238**	**-0.80**	-	**0.02**	5.39
Employment Growth	**0.073**	-0.02	-	0.06	0.46
Employment Rate	**-0.765**	-0.30	-	1.30	2.35
Distance	**-1.216**	**-2.10**	-	**0.17**	10.03
Age/sex Specific Unemployment Rate	**0.929**	-0.60	-	0.76	1.34
Contiguity	**1.366**				
Psi	0.199				
AIC	578				
AICc	592			615	

The AICc for the global model is lower than that for the local model (the difference is over 3) suggesting that the global model fits better than the local model. The local parameter estimates are very interesting. The last column of Table 7.3 shows the values of the *l*-statistic, an indicator of the significant spatial variation of the local parameter estimates. This statistic is high for distance, vacant and derelict dwellings, destination accessibility, percentage vacant dwellings, house prices, listed buildings and local authority stock in poor condition indicating a significant spatial variation of the local parameter estimates of these variables.

In Figure 7.1, the spatial distributions of local parameter estimates exhibiting significant spatial variation are presented for the following variables: destination accessibility, house prices, listed buildings, total population, vacant and derelict dwellings and distance. The reason the local parameter estimates for percentage vacant dwellings are not mapped here is that this variable has a very poor performance in the global models (Table 7.2). Instead, the local parameter estimates for total population are mapped, as this is an important variable to destination choice. A discussion of the effect of each variable follows.

Destination Accessibility

This variable has a significant negative effect on destination choice, suggesting an existence of hierarchical processing of migrants' destination choice. The global parameter estimate is -1.191 suggesting a non-linear relationship. However, the local parameter estimates (*Local Model*) range from -1.92 (Cumbria) to 0.99 (Northamptonshire). The spatial patterns of the latter are very interesting (upper left map in Figure 7.1).

The effect of destination accessibility for those leaving Newcastle is strongly negative in most of Northern England and Northern Wales. This suggests that those leaving Newcastle are less attracted by large urban areas in the North. It also indicates a more intensively hierarchical process of FHSA choice in these areas.

However, in most of the FHSAs in the South-East and East Midlands, destination accessibility has a positive effect on migrants' destination choices. The latter suggests that for those leaving Newcastle large urban areas in the South are more attractive than isolated population centres. High numbers of in-migrants are associated with areas of high accessibility.

The overall variation in the local parameter estimates of destination accessibility found here has some similarities to that reported for Japan (Nakaya, 2001). There are many cases where destination accessibility has a negative effect on the selection of a potential destination that is close to the origin and a positive effect if the potential destination is further apart. The findings discussed above suggest that destination accessibility explains much of the hierarchical processing in destination choice (Pellegrini and Fotheringham, 1999).

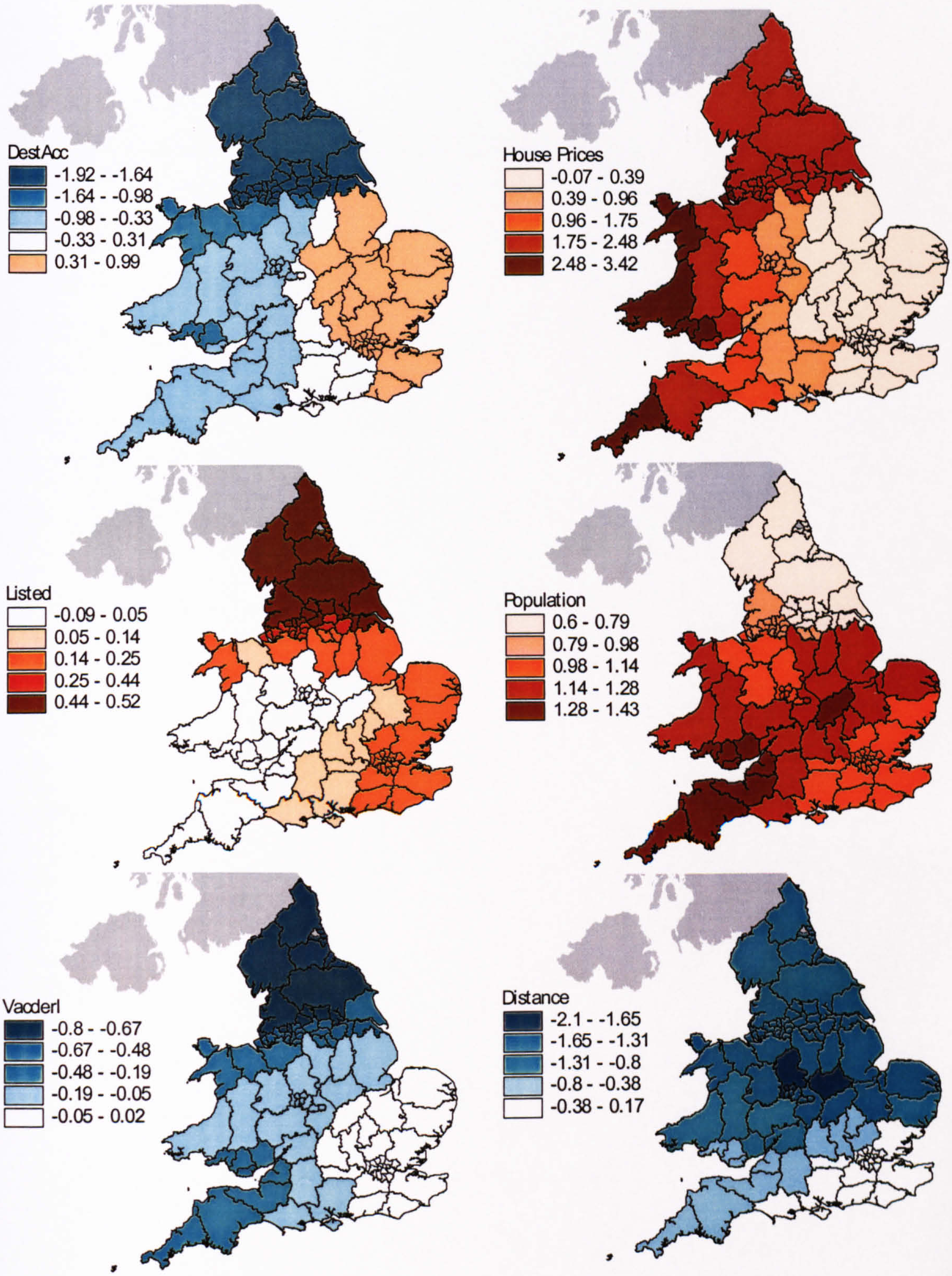


Figure 7.1. Maps of local parameter estimates for *Local Model* for Newcastle

House Prices

Both global and local parameter estimates suggest a positive effect of housing on migrants' destination choices. High prices make a destination more attractive to migrants, perhaps because house prices act as a surrogate for the economic conditions in an area. Thus, economically affluent areas attract more migrants than areas with poor local economic performance. These findings are in line with previous findings (Fotheringham and O'Kelly, 1989; Fotheringham and Pitts, 1995; Atkins and Fotheringham, 1999; Fotheringham et al., 2002b).

The global estimator is close to 2.0 suggesting a highly non-linear association of house prices and destination choice. The local parameter estimates vary from -0.07 (Essex) to 3.42 (West Glamorgan). The spatial distribution of the local parameter estimates (right upper map in Figure 7.1) suggests a South East – North West divide. For those leaving Newcastle, house prices play little role in their decision to move in an FHSA in South East England. However, high house prices are strongly associated with the selection of an FHSA in North, North West, Yorks and Humberside and Wales. These suggest that people are wishing to move to an affluent area if they can afford to. Furthermore, those leaving Newcastle are more favourable to affluent areas located in North England and Wales than those in South-East England.

Listed Buildings

This variable has attracted little attention by researchers for its importance in explaining migration moves. The lack of previous studies using this variable may be because of two reasons: lack of data availability or underestimation of a potential explanatory power of this variable. This variable is associated with the appeal of an area; areas with large number of listed building are more likely to be better known and thus, more likely to be selected as destinations. The empirical findings presented here suggest there is strong evidence for a systematic effect of this variable to destination choice (at least for those leaving Newcastle). Generally, the effect is positive, suggesting that areas with large volumes of listed buildings attract larger numbers of migrants.

It is important to note, that the existence of a relationship between the appeal of an area (also evident from a significant effect of vacant and derelict dwellings presented below) and its attractiveness leads to the conclusion that migrants are equally interested to move to areas with pleasant environments as well as areas offering employment and housing opportunities. This is very important for those wishing to influence population trends within the country.

The global parameter estimate suggests a non-linear positive relationship. The local parameter estimates range from -0.09 (Mid Glamorgan) to 0.52 (North Tyneside). The spatial pattern of the local parameter estimates is shown in Figure 7.1 (middle left map). This suggests there is little, if any, effect of this variable on the destination choice of FHSAs in Wales, West Midlands and South West England, a moderate effect on FHSAs located in South East England and East Midlands and a relatively strong effect in the remaining FHSAs (Northern England). These patterns suggest that for those leaving Newcastle, the appeal of an area is more important when shorter distance moves are concerned.

It is necessary to note that there is an issue concerning the data quality of this variable. The decisions on the listing of buildings are unlikely to be consistent between (or even perhaps within) FHSAs. Thus, there may be FHSAs that listed buildings have been over-recorded or under-recorded. Thus, further empirical evidence will be required to confirm if there are strong theoretical grounds for the findings presented here.

Population

This variable has a significant positive effect on destination choice, as expected. Destinations with large populations are more attractive to migrants for many reasons: more information is generally available for such destinations, the amenities, social services and cultural opportunities are higher and there are increased possibilities for the existence of social ties (e.g., a relative or friend may live at such a destination).

The global parameter estimate is 1.266. This indicates that the attraction of a place increases at an increasing rate as its population increases.

The local parameter estimates vary from 0.60 (Humberside) to 1.43 (Cornwall). However, the *l*-statistic suggests that there is only weak evidence that this variation is significant. The spatial pattern (middle right map, Figure 7.1) of the local parameter estimates suggests a stronger effect of population on destination choice for FHSAs in South Wales and South West England and a much weaker effect for FHSAs in North England. This suggests that migrants from Newcastle are attracted to larger urban areas in more distant (and rural) parts of the country such as South-West England and Wales but are relatively less attracted to rural areas in North England.

Vacant and derelict dwellings

This variable has a significant negative effect on destination choice. Areas with high proportions of vacant and derelict land are less attractive to migrants, probably because such areas are associated with deprivation and poor economic performance. The global parameter estimate is -0.238 suggesting a logarithmic relationship. The local parameter estimates range from -0.80 (Northumberland) to 0.02 (Essex). There is evidence that the local parameter estimates exhibit significant spatial variation. Their spatial pattern is very interesting (lower left map, Figure 7.1). When short distance moves are concerned, there is a strong negative effect, whereas little or no effect exists when long distance moves are concerned. This may be because longer distance moves are more likely to be associated with change of employment, whereas shorter distance moves are more likely to be associated with factors indicating quality of life. Therefore, this variable plays a more important role in short distance destination choices.

Distance

Traditional models suggest a strong negative effect of distance on destination choice in England and Wales. This is confirmed here; the global parameter estimate is -1.216 and the local parameter estimates range from -2.10 to 0.17. The local models suggest a spatially variable effect. The very high *l*-statistic for *Local Model* provides strong evidence for a significant spatial variation of the local parameter estimates of distance.

The localised distance decay parameter indicates how useful distance is in discriminating between choices of destinations near to a given destination *j*. Here, the local parameter estimates for distance indicate the ability of distance to discriminate migration trends in FHSAs close to Newcastle (in the northern half of England and Wales) but not further away (FHSA in the southern half of England).

One reason for distance being an important discriminating variable for short distance migration moves and unimportant discriminating variable for long distance migration moves is that migrants know much more about close areas (to where they live) than they do for areas further away. Another reason could be that people who move short distances for housing or quality of life improvement reasons are wishing to be as close as possible to their previous residence (for personal reasons). Thus, for short distance moves, a small difference in distances between an origin and two alternative destinations has a strong impact on destination choice (migrants are more likely to choose the closest destination), whereas for long distance moves (where the distance decay function is relatively flat) such a difference will have less or no impact on destination choice.

7.5 Camden and Islington

All the above findings concern those leaving Newcastle. I now present empirical findings of equivalent models for migrants from Camden & Islington, an FHSA with the highest out-migration rate (10%) for mature male adults in 1996/97 located in London.

7.5.1 Time trends in destination choice determinants

Table 7.4 shows the global parameter estimates of Poisson models for males 30 – 44 leaving Camden & Islington during 1990/91 – 1996/97. Similar to the findings presented in Table 7.2 for migrants from Newcastle, destination accessibility, house prices, listed buildings, total population, distance and contiguity have a systematic effect on destination choice. Additionally, there is evidence that climate, crime, council tax, new buildings in private sector and sex/age specific unemployment rates have significant effect on destination choice for Camden & Islington. However, vacant and derelict land was not significant here. The percentage of vacant dwellings is statistically significant, but it does not have a consistent effect on destination choice.

Table 7.4. Parameter estimates of the destination choice global models for males 30 – 44 leaving Camden & Islington during 1990/91 – 1996/97

	1990/91	1991/92	1992/93	1993/94	1994/95	1995/96	1996/97
Intercept	**-17.456**	**-19.561**	**-23.103**	-3.603	5.047	-4.887	**-13.241**
Climate Index	-0.038	**-0.175**	-0.062	**-0.129**	**-0.260**	**-0.243**	**-0.236**
Crime Index	**0.139**	**0.192**	**0.196**	**0.191**	**0.193**	**0.127**	-0.079
Council Tax	**1.257**	**0.982**	**0.692**	0.260	0.397	**1.408**	**1.367**
Destination Accessibility	**-0.902**	**-0.893**	**-0.501**	**-1.221**	**-1.295**	**-1.049**	**-0.907**
Household Income	0.029	-0.318	0.307	-0.078	**-1.014**	**-1.104**	**-1.533**
House Prices	**1.697**	**1.958**	**1.888**	**1.353**	**1.290**	**1.290**	**2.133**
Listed buildings	**0.267**	**0.226**	**0.399**	**0.143**	**0.292**	**0.310**	**0.290**
New Housing	0.056	0.137	0.200	**0.437**	**0.411**	**0.383**	0.174
% Net Re-lets	**0.427**	**-0.230**	**-0.240**	-0.033	**0.219**	**0.256**	-0.010
Total Population	**0.665**	**0.864**	**0.667**	**0.656**	**0.651**	**0.767**	**0.775**
New buildings (Private)	**0.400**	**0.377**	**0.303**	**0.277**	**0.247**	**0.112**	**-0.192**
New buildings (Social)	-0.044	0.037	**0.153**	**0.185**	**0.088**	0.067	0.037
% Vacant Dwellings	**0.386**	**0.409**	**-0.294**	**0.401**	**0.529**	**-0.718**	-0.025
Poor condition (Private)	-0.034	**0.174**	**0.186**	-0.047	0.020	**0.232**	0.060
Poor condition (LA)	**0.079**	-0.005	**0.115**	0.060	0.029	**-0.077**	**0.103**
Vacant and derelict	0.015	0.036	0.052	**0.071**	**0.058**	0.023	0.008
Employment Growth	**-0.082**	-0.012	-0.021	-0.012	-0.004	**-0.038**	**0.034**
Employment Rate	-0.254	-0.068	**-0.277**	0.211	-0.193	-0.050	**-0.358**
Distance	**-1.087**	**-1.158**	**-1.048**	**-1.305**	**-1.527**	**-1.525**	**-1.302**
Age/sex Specific Unemployment Rate	**0.435**	**0.700**	**1.122**	0.251	0.140	**0.389**	**0.527**
Contiguity	**-0.474**	**-0.451**	**-0.465**	**-0.520**	**-0.572**	**-0.487**	**-0.487**
Psi Poisson	0.1395	0.1407	0.1336	0.1621	0.1593	0.1506	0.1265
AIC	660	704	694	796	807	775	755
AICc	674	718	708	810	821	789	769

The *psi* statistics suggest an overall better fit of global destination choice models for Camden & Islington than for Newcastle. This probably indicates that migrants from London respond better to the destination choice determinants than migrants from Newcastle.

7.5.2 Spatial trends in destination choice determinants

In the case of Camden & Islington for most of the significant global parameter estimates, the corresponding p-value (showing their significance in the regression) is less than 0.001. The *Global Model* shows a significant negative effect of climate, destination accessibility, household income, new building completion in private sector, employment rate, distance and contiguity and a significant positive effect of council tax, house prices, listed buildings, total population and sex/age specific unemployment rates on destination choice. A significant effect of vacant and derelict land found for Newcastle migrants is not confirmed here. However, there is evidence for a significant effect of climate, council tax and sex/age specific unemployment rates not found for migrants from Newcastle.

Table 7.5. Parameter estimates of the destination choice models (global and local) for males 30 – 44 leaving Camden & Islington in 1996/97

	Global Model	Local Model		
		Min	Max	*t*
Intercept	-13.241	-28.589	19.007	2.33
Climate Index	-0.236	-0.647	0.472	9.06
Crime Index	-0.079	-0.153	0.313	1.91
Council Tax	1.367	-0.497	2.406	3.69
Destination Accessibility	-0.907	-2.868	-0.243	3.17
Household Income	-1.533	-2.828	0.822	3.58
House Prices	2.133	0.980	2.712	1.89
Listed buildings	0.290	0.105	0.387	1.89
New Housing	0.174	-0.725	0.792	3.23
% Net Re-lets	-0.010	-0.501	0.990	3.35
Total Population	0.775	0.357	1.145	4.53
New buildings (Private)	-0.192	-0.342	0.281	4.08
New buildings (Social)	0.037	-0.042	0.164	1.30
% Vacant Dwellings	-0.025	-0.767	0.179	1.37
Poor condition (Private)	0.060	-0.190	0.303	1.29
Poor condition (LA)	0.103	-0.438	0.161	7.77
Vacant and derelict	0.008	-0.037	0.156	3.07
Employment Growth	0.034	-0.075	0.140	4.55
Employment Rate	-0.358	-0.758	1.145	6.36
Distance	-1.302	-1.691	-1.108	1.59
Age/sex Specific Unemployment Rate	0.527	-0.765	1.194	4.15
Contiguity	-0.487			
Psi	0.127			
AIC	755			
AICc	769		732	

The AICc suggests that the *Local Model* is an improvement of the *Global Model* for Camden & Islington. Several of the local parameter estimates of the *Local Model* exhibit significant spatial variation, especially climate, local authority stock in poor condition, employment rate, employment growth, population and sex/age specific unemployment rate.

There are cases where the local parameter estimates of a variable exhibit significant spatial variation, however the corresponding global parameter estimate is not significant (e.g., new housing in former urban land, percentage new re-lets and vacant and derelict land). Perhaps, a different model construction strategy should be applied for the global and local models. A calibration algorithm with more robust significance tests and goodness-of-fit statistics is necessary to suggest such a strategy. At the moment, it is possible to calculate local *t* tests, however, the variable inclusion on the local model is based on its performance in the global model.

The local parameter estimates for some of the variables in the *Local Model* for Camden and Islington are mapped and presented in Figure 7.2. A brief discussion for the effect of each of these variables on destination choice follows.

Destination Accessibility

A clear-cut strong negative effect is evident here by both global and local parameter estimates of this variable. Isolated destinations are more attractive to migrants leaving Camden and Islington than centralised destinations. This is in line with the literature and supports the hypothesis for a spatial effect structure in migrants' destination choice. The spatial patterns of the local parameter estimates are very interesting. They show a stronger effect of this variable on selecting a destination of close proximity (short distance moves) and a weaker effect on selecting a destination that is more distant (to Camden and Islington FHSA). Here the counter-urbanisation effect is clearer, because those leaving Camden and Islington are strongly deterred from moving to any other urban area, especially if this is close to London.

House Prices

The empirical findings for the effect of house prices on destination choice of migrants from Camden and Islington are quite similar to those found for migrants from Newcastle. A general positive effect is evident, suggesting that high in-migration flows are associated with high house prices. Here there is a lower degree of spatial variation of the local parameter estimates than that found for destination accessibility for migrants from Newcastle. The spatial patterns here and in the case of migrants from Newcastle show that a very strong effect exists on selecting destinations located in Northern England, South West England and Wales. This perhaps suggests that the effect of house prices is destination-specific rather than origin-specific. As far as migratory moves to Northern England, South-West England or Wales are concerned, the role of house prices on destination choice is much stronger than in the case of migration flows to East Midlands and the South East.

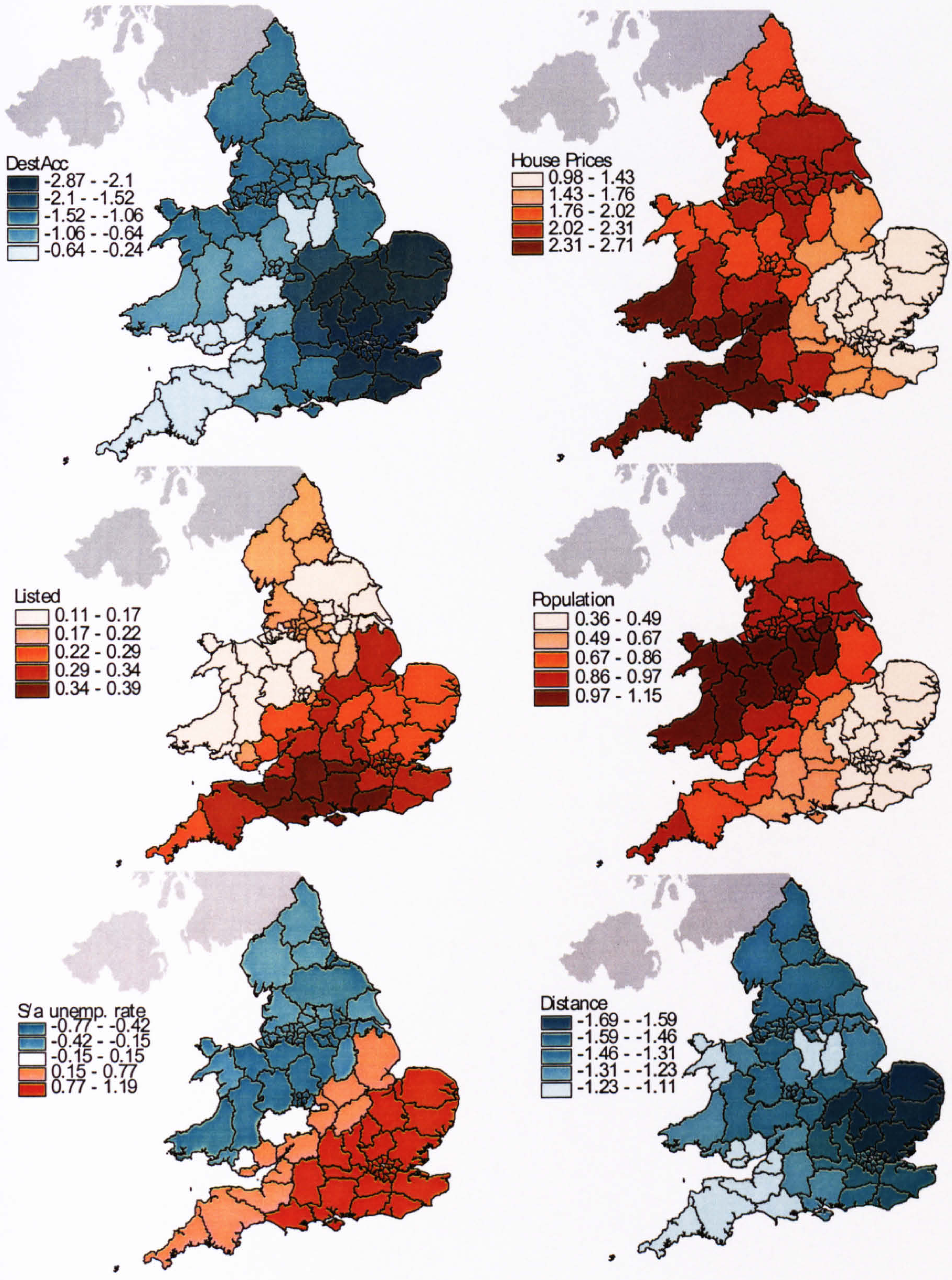


Figure 7.2. Maps of local parameter estimates for *Local Model* for Camden & Islington

Listed buildings

Here the effect of listed buildings is weaker than that reported for migrants from Newcastle. However, it is significant and positive providing more evidence that areas of high proportions of listed buildings attract more migrants. The spatial patterns of the local parameter estimates presented in Figure 7.2 show an interesting cluster where the effect of this variable is stronger. This cluster is formed by Dorset, Somerset, Wiltshire, Berkshire, Hampshire and West Sussex and the Isle of White. The effect of this variable on destination choice is stronger as far as short distance moves are concerned and weaker as far as long distance moves are concerned.

Population

Similar to the findings discussed for Newcastle, areas with large populations attract more migrants. However, here the effect of population is weaker, perhaps because migrants from Camden and Islington respond less to populated areas with large populations than migrants from Newcastle. Perhaps the latter is associated with the fact that mature male adults who live in high-populated areas are less attracted to such areas because of the problems associated with them (e.g. traffic). The spatial pattern of the local parameter estimates shows a similar trend to that found for migrants from Newcastle. This is that larger urban areas in more distant places are more attractive.

Distance

A strong negative effect is evident here. The further apart a destination is from an origin the less likely it is to be selected. Here, all local parameter estimates are negative. This along with findings discussed above suggest that migrants from Camden and Islington are less willing to migrate long distances than migrants from Newcastle. This may be connected with the fact that several opportunities (economic, cultural) available in the South East are not available anywhere else in England and Wales.

Sex/age specific unemployment rate

The global parameter estimates suggests a significant positive effect of sex/age specific unemployment rate on destination choice for migrants from Camden and Islington. This is also the case for migrants from Newcastle, and in line with empirical findings previously reported. Major findings in the migration literature (Lowry, 1966; Weeden, 1973; Flowerdew and Lovett, 1988; Liaw, 1990; Cannari et al., 2000; Fotheringham et al., 2002b) suggest a negative effect of unemployment rates on destination choice, however these refer to total unemployment rather than migrant groups specifically.

The local parameter estimates exhibit significant spatial variation. These are positive for destinations close to Camden and Islington and negative for destinations far apart. The interpretation of these trends is not straightforward. Perhaps further analysis is required to allow safer conclusions for the effect of this variable on destination choice.

The effect of some variables on destination choice for migrants from Camden and Islington have been discussed above. These variables were included to allow comparisons with trends found for migrants from Newcastle. Table 7.5 shows that there are more variables with a non-stationary effect on destination choice. These are council tax, household income, new housing in former urban land, percentage net re-lets, new building completions in private sector, employment growth and employment rate. However, for various reasons such as the lack of a significant effect of the variable in the global model or the lack of a clear effect consistent over time, the latter variables are not discussed in detail here.

7.6 Summary

This is the second chapter of migration modelling (destination choice) and the last chapter of data analysis. Within it I modelled the factors influencing the attraction a destination to migrants (destination choice process) and examined if significant non-stationarity in the parameter estimates (of these factors) exists. This was made possible through calibrated local and global models of destination choice.

Here, I focused my interest on specific origins and population groups. A thorough investigation of global destination choice models has already been presented (Fotheringham et al., 2002b) and a thorough investigation for local destination choice models is left for future work. However, the findings presented in this chapter are important and interesting.

It is the first time an attempt has been made to locally model migrants' destination choice in England and Wales. Previous local modelling for Japan (Nakaya, 2001) included only three variables: population, distance and destination accessibility. I believe that Nakaya's findings are interesting, however they are probably biased because of the lack of other explanatory variables in the models. Migrants' destination choices in England and Wales are affected not only by destination accessibility, population and distance, but also by house prices, measures associated with the culture, historical heritage and appeal of an area such as listed buildings and deprivation indicators such as proportions of vacant and derelict dwellings.

The above findings for local models suggest that there could be a significant spatial variation of the local parameter estimates of several explanatory variables in spatial interaction models. These refer to variables found to have a significant effect on destination choice (global models) for migrants in England and Wales summarised above. Locally varying parameter estimates remove the bias in traditional models that assume a stationary effect of explanatory variables across destinations. The effect of some destination choice determinants is associated with the location of the destination. For example, empirical findings presented above suggest that the effect of house prices is stronger when a destination is located in Wales, North England or South-West England than in South-East England or Midlands.

Migration decisions at the origin and destination have now been explored and explained, across not only sex and age groups, but also time and space. In many cases, in both this and the previous chapter, the results have been linked to the literature suggesting some consistency with previous findings. However, there are occasions that this work resulted in the need for more attention on variables connected with the general appeal of an area and less attention to its labour market conditions.

I now summarise this thesis and discuss the overall conclusions.

Chapter 8
Summary and Future Research

This is the final chapter of the thesis. Here I summarise the work presented above and I draw some general conclusions. I also discuss some limitations of this work. Finally, I close the thesis by discussing some of my general thoughts about migration studies.

Each of the previous chapters contributed in addressing the aims set out in Section 1.1. Chapter 2 aimed to evaluate empirical work in migration modelling. By reviewing the existing literature, Chapter 2 presented trends in migration flows in previous years and previous findings on the role of migration determinants in migration decisions. It also discussed some technical issues such as the geographical scale of the analysis and alternative methodologies. It was established that recently published (by NHSCR) annual data on migration were relatively unexplored. These have now been analysed and presented in Chapter 5 contributing to the literature update.

It is also clear that existing work on migration modelling assumed that the effects of socio-economic factors on individuals' migration decisions are stationary across space. This introduces inaccuracy in migration models. Chapters 6 and 7 demonstrate that it is possible to remove this inaccuracy by conducting local migration modelling. These chapters also provide empirical evidence that there are factors with a non-stationary effect. Thus, many of the existing findings may no longer be accurate if more detailed methods, such as local migration modelling, are applied to the corresponding data.

Chapters 3 and 4 helped in understanding the dataset and methodology used here. Chapter 3 discussed some data construction and quality issues to allow a careful interpretation of the results of migration modelling (Chapters 6-7). Chapter 4 demonstrated the superiority of Geographically Weighted Regression and the relevant statistical diagnostics in providing a tool for fitting and assessing local models of migration. The provision here of empirical evidence for this superiority (Chapters 6-7), suggests that advances in quantitative methodologies allow the extraction of more information from existing data and make this work relevant to the scientific field of Geographical Information Science.

Chapter 5 met its aim to identify and explain the spatial and temporal trends of out- and in- and net migration in FHSAs in England and Wales between 1984 and 1998 by the use of univariate data analysis (k-means clustering, ESDA, GWLM) and visualisation means (maps, graphs, and heat maps). It provided a comprehensive presentation of migration trends for several sex and age population groups.

Chapters 6 and 7 provided a better understanding and modelling of the factors influencing the production of migrants at an origin (departure decision-making process) and the attraction of migrants from a destination (destination choice process) by the use of a rich dataset. They also provided empirical evidence that significant non-stationarity in the parameter estimates for some of these factors exists (i.e. these parameter estimates exhibit significant spatial variation). The latter was the major aim of this thesis.

Finally, Chapter 6 found that although local migration models improve the model residuals (as well as they remove potential spatial autocorrelation), they do not remove the existence of large residuals. This suggests that there is still some unexplained variance in migration which requires more robust or correct measures of existing variables and perhaps new measures of area attributes not available here.

8.1 Concluding remarks

The results reported in this thesis support the idea that the modelling of migration decisions can be conducted meaningfully by applying localised regression methodologies on age and sex disaggregated data. It is also evident that the graphical representations of migration data are necessary to understand and communicate the spatial and temporal trends found in migratory moves. For the latter a new visualisation means introduced here (*heat maps*) seems to have some potential in representing disaggregated migration data over time.

In this thesis, I analysed patterns of in-, out- and net migration over space and time in order to identify interesting trends. I found that out-migration rates for teenagers, young adults and adults vary over time and have substantially increased in the 1990s. Out-migration for children, mature and older adults as well as pensioners are rather stable over time. There is an association between the increase of migration rates for young adults (i.e. 18–21 years old) and the increase of the number of university places in England and Wales. In terms of net migration rates, areas with big universities are net population gainers for migrants aged 16–19 but most of these areas (except London) suffer from a net out-migration of migrants aged 20–29. Overall, the temporal and spatial patterns of migration flows in England and Wales during the 1980s and the 1990s suggest the continuation of a counter-urbanisation phenomenon albeit at a reduced level than that reported for the 1970s.

I also tried several ways of examining the existence of spatial autocorrelation in out-migration rates. These included a simple k-means classification and some more complex statistics including Moran's I, Geary's c and Getis' G as well as Geographically Weighted local means. For the data analysed here, the more advanced statistics did not help in identifying spatial trends that could be missed from the simple k-means classification. Perhaps this is because the data here refer to only a few geographical areas (98).

One of the most important and innovative findings of this thesis is that there is empirical evidence that the effect of several ecological variables on migration decisions is not stationary across space.

In an investigation of local models of out-migration rates I found that the local parameter estimates for some migration determinants exhibit significant spatial variation. The most interesting finding is for the effect of employment rates on out-migration rates. Generally, I found that areas with high employment rates produce more migrants. This contradicts the migration literature, the Lowry hypothesis that suggests no effect of labour market conditions on out-migration rates and the theoretically expected negative effect of employment rates on out-migration. I also found that the effect of this variable is stronger for areas located in North England and weaker in areas located in the South, a clear North-South divide.

Generally, a positive relation between employment rates and out-migration rates suggests that people living in areas with high employment opportunities, possibly have higher family income and thus can afford to migrate in order to improve their quality of life. Areas with high employment rates perhaps have more recent migrants who are more likely to migrate. The spatial pattern of the local parameter estimates suggests that those living in Northern England are more prone to leave the area than those living in Southern England. Thus, employment opportunities in the North encourage population shifts, perhaps to the South, rather than keeping a balanced population distribution across England and Wales.

I also investigated local models of destination choice and I found that the local parameter estimates for some migration determinants exhibit a high degree of spatial variation. The statistical tests available at this time suggest there is a high potential for some interesting results, but cannot guarantee these are statistically significant. However, further investigation is needed to put the theoretical grounds for a spatial varying destination-specific effect of destination choice determinants.

The most interesting findings on destination choice models can be summarised as two general trends: the distinction between the effect of destination choice determinants on short-distance and longer-distance migration moves and the existence of a destination-specific effect independent of the origin of a migration trip.

It is evident for mature male adults, that destination choice determinants have a stronger effect on destination choice when short distance moves are concerned and a weaker effect for longer distance moves. This trend is evident from the empirical findings for destination accessibility, listed buildings and distance. Perhaps this is because short-distance moves are more prone to ecological effects whereas long-distance moves are rather based on idiosyncratic human behaviour.

The effect of some other destination choice determinants is associated with the location of the destination. Empirical findings presented in the previous chapter suggest that the effect of house prices is stronger when a destination is located in Northern England than in Southern England. Finally, the local parameter estimates for distance indicate the ability of distance to discriminate migration trends in close proximity FHSAs but not in FHSAs far apart.

I believe that some interesting findings have been presented here and there are several starting points for further research. I hope that the use of *heat maps* in representing time-varying migration rates will be adopted and used in the migration literature. I also hope that several policy makers in local and national government will be interested in these results. I believe that having presented significant findings after fitting local models for disaggregated migration data and after allowing for several explanatory variables in the models, the justification for traditional migration modelling is now weaker. The data and statistical tools are now available for researchers to move a step forward to more advanced migration modelling. At the time of this writing, discussions on advancing the Geographical Weighted Regression take place towards the direction of robust estimation. In the longer term, it is also possible to have a time-space weighted regression. All these advances will allow for further research and perhaps new interesting conclusions.

8.2 Limitations

The existence of large residuals in some of the out-migration models raise some questions regarding the extent to which migration determinants included in the models are adequate. Perhaps, some other factors could be added to the modelling framework for out-migration. For example, an index of cultural distance between the different parts on England and Wales, the length of residence in an area (e.g. the proportion of people lived less than 5 years in an area), and the percentage of people born outside an FHSA could be included if they were available. It could also be that high residuals are caused by the migration data being in error or some of the explanatory variables are not measured correctly.

The suggestion for the importance of such variables comes from previous findings reported in the migration literature (e.g. Miller, 1973): people who have recently migrated into an area are more likely to migrate than those who have lived there longer; people who were born outside an area and migrated in many years ago may have a tendency to return to the area from which they originated. As far as cultural distance is concerned most of the evidence comes from Canada and Japan where cultural differences, mainly constituted by religion, language spoken and family traditions, appear to affect out-migration and destination choice. It is not clear to what extent such differences exist within England and Wales but it is conceivable that there are North-South and England-Wales cultural differences not measured accurately in the current set of explanatory variables.

8.3 Epilogue

My three-year trip in the social sciences has been a fascinating experience. It is apparent that in our complex world the boundaries between disciplines are becoming less clear over time. I always believed that cross-disciplinary skills are required in order to achieve a high degree of innovation in research. For me research in the social sciences is a continuous road to understanding human behaviour. With my IT skills, I tried to use a more scientific approach to understanding this behaviour. In my analysis, I tried to shed some more light on understanding what determines migration decisions. Inevitably, the more complex world results in the need for more complex migration models.

I found it interesting that there are still several areas unexplored in the field of internal migration, although it has a history of over a century. Fortunately, the technical tools are available to conduct ever-better quantitative research. This work provides more empirical evidence for the potential of the Geographical Weighted Regression. I believe that the only limit in advancing research is the limit in the imagination of the researchers.

Although it is very important for individuals to have the ability to generate research ideas and to conduct analysis, it is equally important to report and communicate the findings of such analysis. I tried to make a good use of visual means to communicate my findings. For this purpose I introduced a new means of visualising migration rates, the *heat map*. I hope my work will contribute to the knowledge of our academic community.

References

Aitkin, M., 1996, A general maximum likelihood analysis of overdispersion in generalized linear models, *Statistics and Computing*, **6**, pp. 251 – 262.

Akaike, H., 1973, Information theory and an extension of the maximum likelihood principle, in *Proceedings of the Second International Symposium on Information Theory*, edited by B.N. Petrov and F. Csaki, (Budapest: Akademiai Kiado), pp. 267 – 281.

Alker, H. S., 1969, A Typology of Ecological Fallacies, in *Quantitative Ecological analysis*, edited by M. Dogan and S. Rokkan, (Boston, Mass.: MIT Press), pp. 69 – 86.

Alonso, W., 1972, The system of inter-metropolitan flows, in *Population Distribution and Policy*, edited by S. M. Mazie, (Washington, DC: US Government Printing Office), pp. 323 – 334.

Alonso, W., 1973, Urban zero population growth, in *The No-Growth Society*, edited by M. Olson and H. Landsberg, (New York: W. W. Norton), pp. 191 – 206.

Amrhein, C.G., and Flowerdew, R., 1992, The effect of data aggregation on a Poisson regression model of Canadian migration, *Environment and Planning A*, **24**, pp. 1381 – 1391.

Andrews, D.F., 1974, A Robust Method for Multiple Linear Regression, *Technometrics*, **16**, 4, pp. 523 – 531.

Anselin, L, 1995, Local indicators of spatial association, *Geographical Analysis*, **27**, pp. 93–115.

Anselin, L., 1998, Exploratory spatial data analysis in a geocomputational environment, in *Geocomputation: A Primer*, edited by P.A. Longley, S.M. Brooks, R. McDonnel and B. Macmillan (Chichester: John Wiley & Sons), pp. 77 – 94.

Anselin, L., and Smirnov, O., 1998, *The SpaceStat extension for ArcView 3.0*, MorganTown, Regional Research Institute, West Virginia University.

Atkins, D., Champion, T., Coombes, M., Dorling, D., and Woodward, R., 1996, *Urban Trends in England: Latest Evidence from the 1991 Census* (London: HMSO).

Atkins, D., and Fotheringham, S., 1999, Gender variations in migration destination choice, in *Migration and Gender in the Developed World*, edited by P. Boyle and K. Halfacree (London: Routledge), pp. 54 – 72.

Bao, S., and Martin, D., 1997, *User's Reference for the S+ArcView link*, Seattle, MathSoft Inc, Data Analysis Products Devision.

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

Beale, C, 1969, The relation of gross out-migration rates to net migration, *Paper presented at the annual meeting of the Population Association of America*, Atlantic City, N.J., April 10-12, US Department of Agriculture.

Beaton, A.E., and Tukey, J.W., 1974, The Fitting of Power Series, Meaning Polynomials, Illustrated on Band-Spectroscopic Data, *Technometrics*, **16**, pp. 147 – 185.

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

Berry, B.J.L., 1976, The counterurbanization process: urban America since 1970, in *Urbanization and counterurbanization*, edited by B.J.L. Berry, (Beverly Hills, California: Sage Publications), pp. 17 – 30.

Berry, B.J.L., 1980, Urbanization and counterurbanization in the United States, *Annals of the American Academy of the Political and Social Science*, **451**, pp. 13 – 20.

Blaut, J. M., 1999, Maps and Spaces, *The Professional geographer: the journal of the Association of American Geographers*, **51**, 4, pp. 510 – 515.

Boden, P., Stillwell, J., and Rees, P., 1988, Linking census and NHSCR migration, Working Paper 511, School of Geography, University of Leeds.

Boden, P., Stillwell, J., and Rees, P., 1992, How good are the NHSCR data?, in *Migration Processes and Patterns, Volume 2: Population Redistribution in the United Kingdom*, edited by J. Stillwell, P. Rees and P. Boden, (London: Belhaven), pp. 13 – 27.

Boots, B., and Getis., A., 1988, *Point pattern analysis*, (London: Sage).

Boyle, P., Halfacree, K., and Robinson, V., 1998, *Exploring Contemporary Migration*, (New York: Longman).

Boyle, P.J., and Flowerdew, R., 1993, Modelling sparse interaction matrices: interward migration in Hereford and Worcester, and the underdispersion problem, *Environment and Planning A*, **25**, pp. 1201 – 1209.

Boyle, P.J., and Flowerdew, R., 1997, Improving distance estimates between areal units in migration models, *Geographical Analysis*, **29**, 2, pp. 93 – 107.

Bracken, I., and Bates, J., 1983, Analysis of gross migration profiles in England and Wales: some developments in classification, *Environment and Planning A*, **15**, 3, pp. 343 – 355.

Bramley, G., 1998, Housing surpluses and housing need, in *Housing Abandonment in Britain: studies in the causes and effects of low demand housing*, edited by S. Lowe, S. Spencer and P. Keenan, (University of York: Centre for Housing Policy), pp. 9 – 25.

Bramley, G., and Smart, G., 1996, Modelling local income distributions in Britain, *Regional Studies*, **30**, 3, pp. 239 – 55.

Brown, L.A., and Jones III, J.P., 1985, Spatial Variation in Migration Processes and Development: A Costa Rican Example of Conventional Modeling Augmented by the Expansion Method, *Demography*, **22**, 3, pp. 327 – 352.

Brunsdon, C., Aitkin, M., Fotheringham, A.S., and Charlton, M.E., 1999a, A comparison of random coefficient modelling and geographically weighted regression for spatially non-stationary regression problems, *Geographical and Environmental Modelling*, **3**, 1, pp. 47 – 62.

Brunsdon, C., Fotheringham, A.S., and Charlton, M.E., 1996, Geographically Weighted Regression: A Method for Exploring Spatial Nonstationarity, *Geographical analysis*, **28**, 4, pp. 281 – 298.

Brunsdon, C., Fotheringham, A.S., and Charlton, M.E., 1998a, Spatial Non-Stationarity and Autoregressive Models, *Environment and Planning A*, **30**, pp. 957 – 973.

Brunsdon, C., Fotheringham, A.S., and Charlton, M.E., 1998b, Geographically Weighted Regression – Modelling Spatial Non-stationarity, *The Statistician*, **47**, pp. 431 – 443.

Brunsdon, C., Fotheringham, A.S., and Charlton, M.E., 1999b, Some notes on parametric significant tests for Geographically Weighted Regression, *Journal of Regional Science*, **39**, 3, pp. 497 – 524.

Brunsdon, C., McClatchey, J., and Unwin, D.J., 2001, Spatial Variations in the Average Rainfall-Altitude relationship in Great Britain: An Approach Using Geographically Weighted Regression, *International Journal of Climatology: a journal of the Royal Meteorological Society*, **21**, 4, pp. 455 – 466.

Bulusu, L., 1991, *Review of Migration Data Sources*, Occasional Paper 39, Population and Hospital Statistics Division, Office for Population Censuses and Surveys (London: OPCS).

Burnham, K.P., and Anderson, D.R., 1998, *Model selection and inference: a practical information-theoretic approach* (New York: Springer).

Burnham, K.P., and Anderson, D.R., 2002, *Model selection and multimodel inference: a practical information-theoretic approach*, 2nd Edition (New York: Springer).

Cannari, L., Nucci, F., and Sestino, P., 2000, Geographical labour mobility and the cost of housing: evidence from Italy, *Applied Economics*, **32**, pp. 1899 – 1906.

Casetti, E., 1972, Generating Models by the Expansion Method: Applications to Geographical Research, *Geographical Analysis*, **4**, 1, pp. 81 – 91.

Casetti, E., 1982, Drift Analysis of Regression Parameters: An Application to the Investigation of Fertility Development Relations, *Modeling and Simulation*, **13**, pp. 961 – 966.

Casetti, E., 1997, The Expansion Method: Mathematical Modeling, and Spatial Econometrics, *International Regional Science Review*, **20**, 1 & 2, pp. 9 – 33.

Casetti, E., 1999, The evolution of scientific disciplines, mathematical modeling, and human geography, *Geographical Analysis*, **31**, 4, pp. 332 – 339.

Casetti, E., and Jones, III, J.P., 1983, Regional Shifts in the Manufacturing Productivity Response to Output Growth: Sunbelt versus Snowbelt, *Urban Geography*, **4**, pp. 286 – 301.

Champion, A.G., 1989, *Counterurbanization: The Changing Pace and Nature of Population Deconcentration* (London: Edward Arnold).

Champion, A.G., 1994, Population change and migration in Britain since 1981: evidence for continuing deconcentration, *Environment and Planning A*, **26**, pp. 1501 – 1520.

Champion, A.G., 1996, Population Review: (3) Migration to, from and within the United Kingdom, *Population Trends*, **83**, pp. 5 – 16.

Champion, A.G., Coombes, M.G., and Openshaw, S., 1984, New regions for a new Britain, *Geographical Magazine*, **56**, pp. 187 – 190.

Champion, A.G., Green, A.E., Owen, D.W., Ellin, D.J., and Coombes, M.G., 1987, *Changing Places: Britain's demographic, economic and social complexion* (London: Edward Arnold).

Champion, A.G., Fotheringham, A.S., Rees, P., Boyle, P., and Stillwell, J., 1998, *The Determinants of Migration Flows in England: a Review of Existing Data and Evidence.* Department of Geography, University of Newcastle, for the Department of Environment, Transport and the Regions. ISBN 0 902155 39 3.

Champion, T., Wong, C., Rooke, A., Dorling, D., Coombes, M., and Brunsdon, C., 1996, *The Population of Britain in the 1990s: A Social and Economic Atlas* (Oxford: Clarendon Press).

Cleveland, W. S., 1979, Robust Locally Weighted Regression and Smoothing Scatterplots, *Journal of the American Statistical Association*, **74**, 368, pp. 829 – 836.

Cleveland, W. S., and Devlin, S. J., 1988, Locally Weighted Regression: An Approach to Regression Analysis by Local Fitting, *Journal of the American Statistical Association*, **83**, 403, pp. 596 – 610.

Cliff, A.D., and Ord, J.K., 1973, *Spatial autocorrelation* (London: Pion).

Cliff, A.D., and Ord, J.K., 1981, *Spatial processes: models and applications* (London: Pion).

Congdon, P., 1989, Modelling migration flows between areas: an analysis for London using the Census and OPCS longitudinal study, *Regional Studies*, **23**, pp. 87 – 103.

Coombes, M.G., Dixon, J.S., Goddard, J.B., Openshaw, S. and Taylor, P.J., 1982, Functional Regions for the Population Census of Great Britain, in *Geography and the urban environment: progress in research and applications 5*, edited by D.T. Herbert and R.J. Johnston (Chichester: Wiley), pp. 63 – 112.

Cordey-Hayes, M., 1975, Migration and the dynamics of multiregional population systems, *Environment and Planning A*, **7**, pp. 793 – 814.

Cordey-Hayes, M., and Gleave, D., 1973, *Migration movements and the differential growth of city regions in England and Wales*, RP-1, Centre for Environmental Studies, London, England. Also published in *Papers of the Regional Science Association*, **33**, pp. 99 – 123.

CRAN, 2003, *The Comprehensive R Archive Network*, http://www.stats.bris.ac.uk/R/, (Bristol, UK), last accessed on 14/02/2003.

Davis, J.C., 2002, *Statistics and Data Analysis in Geology: Third Edition* (New York: Jon Wiley & Sons).

Devis, T., 1984, Population movements measured by the NHS Central Register, *Population Trends*, **36**, pp. 18 – 24.

Devis, T., and Mills, I., 1986, *A comparison of migration data from the National Health Service Central Register and the 1981 Census*, Occasional Paper 35, Population Statistics Division, Office for Population Censuses and Surveys (London: OPCS).

Eldridge, J.D., and Jones III, J.P., 1991, Warped Space: A Geography of Distance Decay, *Professional Geographer*, **43**, 4, pp. 500 – 511.

Ellis, M., and Odland, J., 1992, Personal characteristics in models of migration decisions: an analysis of destination choice in Ecuador, in *Applications of the Expansion Method*, edited by J.P. Jones III and E. Casetti, (London: Routledge), pp. 115 – 132.

Engels, R.A, and Healy, M.K., 1981, Measuring interstate migration flows: an origin-destination network based on internal revenue service records, *Environment and Planning A*, **13**, pp. 1345 – 1360.

Evers, G.H.M., 1989, Migration, population and regional Labour supply, in *Advances in regional demography: information, forecasts, models*, edited by P. Congdon and P. Batey, (London; New York: Belhaven Press), pp. 229 – 245.

Feder, G, 1982, On the relation between origin income and migration, *Annals of Regional Science*, **16**, pp. 46 – 61.

Ferguson, M. R., and Kanaroglou, P. S., 1997, An Empirical Evaluation of the Aggregated Spatial Choice Model, *International regional science review*, **20**, 1 & 2, pp. 53 – 75.

Fielding, A.J., 1993, Migration and the metropolis: an empirical and theoretical analysis of inter-regional migration to and from South East England, *Progress in Planning*, **39**, pp. 71 – 166.

Fik, T. J., and Mulligan, G. F., 1990, Spatial Flows and competing central places: towards general theory of hierarchical interaction, *Environment and planning A*, **22**, pp. 527 – 549.

Fik, T. J., Amey, R. G., and Mulligan, G. F., 1992, Labor migration amongst hierarchically competing and intervening origins and destinations, *Environment & planning A*, **24**, 9, pp. 1271 – 1290.

Findlay, A., and Rogerson, R., 1993, Migration places and quality of life: voting with their feet?, in *Population Matters: The Local Dimension*, edited by T. Champion, (London: Paul Chapman Publishing), pp. 33 – 49.

Findlay, A.M., and Li, F.L.N., 1999, Methodological issues in researching migration, *Professional Geographer*, **51**, 1, pp. 50 – 59.

Flowerdew R., 1982, Fitting the Lognormall Gravity Model to Heteroscedastic Data, *Geographical Analysis*, **14**, 3, pp. 263 – 267.

Flowerdew, R., 1991, Poisson regression modelling of migration, in *Migration Models, Macro and Micro Approaches*, edited by J. Stillwell and P. Congdon, (London: Belhaven Press), pp. 57-91.

Flowerdew, R., and Aitkin, M., 1982, A method of fitting the gravity model based on the Poisson distribution, *Journal of Regional Science*, **22**, 2, pp. 191 – 202.

Flowerdew, R., and Boyle, P., 1995, Migration models incorporating interdependence of movers, *Environment and Planning A*, **27**, pp. 1493 – 1502.

Flowerdew, R., and Lovett, A., 1988, Fitting constrained Poisson regression models to inter-urban migration flows, *Geographical Analysis*, **20**, 4, pp. 297 – 307.

Flowerdew, R., and Lovett, A., 1989, Compound and generalised Poisson models for inter-urban migration, *Advances in regional demography: information, forecasts, models*, edited by P. Congdon and P. Batey, (London; New York: Belhaven Press), pp. 246 – 256.

Foster, S.A., and Gorr, W.L., 1986, An adaptive filter for estimating spatially-varying parameters: application to modelling police hours spent in response to calls for service, *Management Science*, **32**, 7, pp. 878 – 889.

Fotheringham, A.S., 1981, Spatial structure and distance-decay parameters, *Annals of the Association of American Geographers*, **71**, 3, pp. 425 – 436.

Fotheringham, A.S., 1983, A new set of spatial interaction models: the theory of competing destinations, *Environment and Planning A*, **15**, 1, pp. 15 – 36.

Fotheringham, A.S., 1984, Spatial flows and spatial patterns, *Environment and Planning A*, **16**, pp. 529 – 543.

Fotheringham, A.S., 1991, Migration and spatial structure: the development of the competing destinations model, in *Migration Models, Macro and Micro Approaches*, edited by J. Stillwell and P. Congdon, (London: Belhaven Press), pp. 57 – 72.

Fotheringham, A.S., 1997, Trends in quantitative methods I: stressing the local, *Progress in Human Geography*, **21**, 1, pp. 88 – 96.

Fotheringham A.S, and Brunsdon, C., 1999, Local Forms of Spatial Analysis, *Geographical Analysis*, **31**, 4, pp. 340 – 358.

Fotheringham, A.S, and Curtis, A., 1992, Encoding spatial information: the evidence for hierarchical processing, in *Theories and methods of spatio-temporal reasoning in geographic space*, edited by A.U. Frank, I. Campari and U. Formentini (Dortmund: Springer-Verlag), pp. 269 – 287.

Fotheringham, A.S, and Curtis, A., 1999, Regularities in spatial information processing: Implications for modelling destination choice, *The Professional Geographer*, **51**, 2, pp. 227 – 239.

Fotheringham, A.S., and O'Kelly, M. E., 1989, *Spatial Interaction models: formulations and applications* (Dordrecht: Kluwer).

Fotheringham, A.S., and Pitts, T.C., 1995, Directional variation in distance-decay, *Environment and Planning A*, **27**, pp. 715 – 729.

Fotheringham, A.S., and Rogerson, P.A., 1993, GIS and spatial analytical problems, *International Journal of Geographical Information Systems*, **7**, 1, pp. 3 – 19.

Fotheringham, A.S., and Williams, P.A., 1983, Further Discussion on the Poisson Interaction Model, *Geographical Analysis*, **15**, 4, pp. 343 – 347.

Fotheringham, A.S., Brunsdon, C., and Charlton, M.E, 2000, *Quantitative Geography* (London: Sage Publications).

Fotheringham, A.S., Brunsdon, C., and Charlton, M., 2002a, *Geographically Weighted Regression: the analysis of spatially varying relationships* (Chichester: John Wiley and Sons).

Fotheringham, A.S., Charlton, M.E., and Brunsdon, C., 1996, The Geography of Parameter Space: an Investigation into Spatial Non-Stationarity, *International Journal of Geographical Information systems*, **10**, pp. 605 – 627.

Fotheringham, A.S., Charlton, M.E., and Brunsdon, C., 1997a, Two Techniques for Exploring Non-stationarity in Geographical Data, *Geographical Systems*, **4**, pp. 59 – 82.

Fotheringham, A.S., Charlton, M.E., and Brunsdon, C., 1997b, Measuring Spatial Variations in Relationships with Geographically Weighted Regression, in *Recent Developments in Spatial Analysis: Spatial Statistics, Behavioral Modeling and Computational Intelligence*, Edited by M.M. Fisher and A. Getis (Berlin: Springer-Verlag), pp. 60 – 82.

Fotheringham, A.S., Charlton, M.E., and Brunsdon, C., 1998, Geographically Weighted Regression: A Natural Evolution of the Expansion Method for Spatial Data Analysis, *Environment and Planning A*, **30**, pp. 1905 – 1927.

Fotheringham, A.S., Nakaya, T.K.Y., Openshaw, S., and Ishikawa, Y., 2001, Hierarchical Destination Choice and Spatial Interaction Modelling: A Simulation Experiment, *Environment and Planning A*, **33**, pp. 901 – 920.

Fotheringham, A.S., Rees, P., Champion, T., Kalogirou, S., and Tremayne, A.R., 2003, The Development of a Migration Model for England and Wales I: Overview and

Modelling Out-migration, Manuscript available from the authors, *Environment and Planning A* (Accepted, in press).

Fotheringham, A.S., Barmby, T., Brunsdon, C., Champion, T., Charlton, M., Kalogirou, S., Tremayne, A., Rees, P., Eyre, H., Macgill, J., Stillwell, J., Bramley, G., and Hollis, J., 2002b, *Development of a Migration Model: Analytical and Practical Enhancements*, Office of the Deputy Prime Minister.

Fox, J., 1997, *Applied Regression Analysis, Linear Models, and Related Methods*, (London: Sage Publications).

Geary, R. C., 1954, The Contiguity Ratio and Statistical Mapping, *Incorporated statistician*, **5**, 3, pp. 115 – 127 & 129 – 145.

GeoStat, 2003, SPSS Macro GlobalLocalMoran, available at http://geog-www.sbs.ohio-state.edu/faculty/tiefelsdorf/GeoStat.htm, last accessed in March 2003.

Getis, A., 1991, Spatial interaction and spatial autocorrelation: a cross-product approach, *Environment and Planning A*, **23**, pp. 1269 – 1277.

Getis, A., 1994, Spatial dependence and heterogeneity and proximal databases, in *Spatial analysis and GIS*, edited by A.S. Fotheringham and P.A. Rogerson (London: Taylor and Francis), pp. 105 – 120.

Getis, A., and Boots, B., 1978, *Models of spatial processes: an approach to the study of point, line and area patterns*, (Cambridge: Cambridge University Press).

Getis. A., and Ord, J.K., 1992, The analysis of spatial association by use of distance statistics, *Geographical Analysis*, **24**, pp. 189 – 206.

Getis. A., and Ord, J.K., 1996, Local spatial statistics: an overview, in *Spatial Analysis: Modelling in a GIS Environment*, edited by P. Longley and M. Batty, (New York: John Willey and Sons), pp. 261 – 277.

Goldstein, S., 1954, Repeated migration as a factor in high mobility rates, *American Sociological Review*, **19**, 5, pp. 536 – 541.

Goldstein, S., 1964, The extend of repeated migration: an analysis based on the Danish Population Register, *Journal of the American Statistical Association*, **59**, 308, pp. 1121 – 1132.

Goldstein, H., 1987, *Multilevel Models in Educational and Social Research* (London: Charles Griffin).

Golledge, R.G., and Timmermans, H., 1990, Applications of behavioural research on spatial problems I: cognition, *Progress in human geography*, **14**, 1, pp. 57 – 99.

Golledge, R.G., 1993, Geographical Perspectives on Spatial Cognition, *Advances in psychology*, **96**, pp. 16 – 46.

Golledge, R.G., and Stimson, R. J., 1997, *Spatial behavior: a geographic perspective* (New York: Guilford Press).

Green, A.E., 1994, The role of migration in labour-market adjustment: the British experience in the 1980s, *Environment and Planning A*, **26**, pp. 1563 – 1577.

Greenwood, M.J., 1971, Regression analysis of migration to urban areas of a less developed country: the case of India, *Journal of Regional Science*, **11**, 2, pp. 253 – 262.

Greenwood, M.J., 1975, Research on internal migration in the United States: a survey, *Journal of Economic Literature*, **13**, 2, pp. 397 – 433.

Greenwood, M.J., and Hunt, G.L., 2003, The early history of migration research, *International Regional Science Review*, **26**, 1, pp. 3 – 37.

Greenwood, M.J., and Ladman, J.R., 1978, An economic analysis of migration in Mexico, *Annals of Regional Science*, **12**, 2, pp. 16 – 31.

Hagerstrand, T., 1965, A Monte Carlo approach to diffusion, *European Journal of Sociology*, **6**, pp. 43 – 67.

Hartigan, J.A., 1975, *Clustering Algorithms* (New York: Wiley).

Hartigan, J.A., and Wong, M.A., 1979, A K-means clustering algorithm, *Applied Statistics*, **28**, pp. 100 – 108.

Haynes, K. E., and Fotheringham, A. S., 1984, *Gravity and Spatial Interaction Models*, Vol.2, Sage Series in Scientific Geography (Beverly Hills, CA: Sage).

Heien, D.M., 1968, A note on log-linear regression, *Journal of the American Statistical Association*, **63**, 323, pp. 1034 – 1038.

Hope, A.C.A., 1968, A simplified Monte Carlo significance test procedure, *Journal of the Royal Statistical Society, Series B (methodological)*, **30**, 3, pp. 582 – 598.

Hurvich C.M, Simonoff, J.S., and Tsai, C.-L., 1998, Smoothing parameter selection in nonparametric regression using an improved Akaike Information Criterion, *Journal of the Royal Statistical Society Series B*, **60**, 2, pp. 271 – 293.

Hurvich C.M, and Tsai, C.-L., 1989, Regression and time series model selection in small samples, Biometrika, **76**, pp. 297 – 307.

Ihaka, R., and Gentleman, R., 1996, R: A language for data analysis and Graphics, *Journal of Computational and Graphical Statistics*, **5**, 3, pp. 299 – 314.

Johnston, R.J., Gregory, D., Pratt, G., and Watts, M., 2000, *The Dictionary of Human Geography*, 4th Edition, (Oxford: Blackwell Publishers Ltd.).

Jones, K., 1991a, Specifying and Estimating Multi-level Models for Geographical Research, *Transactions of the Institute of British Geographers*, **16**, pp. 148 – 159.

Jones, K., 1991b, *Multi-level Models for Geographical Research Concepts and Techniques in Geographical Research* (Norwich: Environmental Publications).

Jones, III, J.P., and Casetti, E., 1992, *Applications of the Expansion Method* (London: Routledge).

Kelley, A. C., and Weiss, L. W., 1969, Markov processes and economic analysis: the case of migration, *Econometrica*, **37**, 2, pp. 280 – 297.

Kitchin, R., and Tate, N. J., 2000, *Conducting Research in Human Geography: Theory, Methodology and Practice* (London: Prentice Hall).

Kleinbaum, D.G., Kupper, L.L., and Muller, K.E., 1988, *Applied Regression Analysis and Other Multivariable Methods*, 2nd Edition, (Boston: PWS-KENT).

Knudsen, D.C., and Fotheringham, A.S., 1986, Matrix Comparison, Goodness-of-Fit, and Spatial Interaction Modeling, *International Regional Science Review*, **10**, 2, pp. 127–147.

Kriesberg, E. M., and Vining, D. R., 1978, On the contribution of out-migration to changes in net migration: a time series confirmation of Beale's cross sectional results, *Annals of Regional Science*, **12**, pp. 1 – 11.

Kullback, S., and Leibler, R., 1951, On information and sufficiency, *Annals of Mathematical Statistics*, **22**, pp. 79 – 86.

Land, K. C., 1969, Duration of residence and prospective migration: further evidence, *Demography*, **6**, 2, pp. 133 – 140.

Lansing, J. B., and Mueller, E., 1967, *The geographical mobility of labor* (Lansing: Survey Research Center, Institute for Social Research).

Leung, Y., Mei, C.-L., and Zhang, W.-X., 2000a, Statistical tests for spatial nonstationarity based on the geographically weighted regression model, *Environment and Planning A*, **32**, pp. 9 – 32.

Leung, Y., Mei, C.-L., and Zhang, W.-X., 2000b, Testing for spatial autocorrelation among the residuals of the geographically weighted regression, *Environment and Planning A*, **32**, pp. 871 – 890.

Liaw, K.-L., 1990, Joint effects of personal factors and ecological variables on the interprovincial migration pattern of young adults in Canada: a nested logit analysis, *Geographical Analysis*, 22, 3, pp. 189 – 208.

Liaw, K.-L., and Kawabe, H., 1994, The dependence of marriage migrations in Japan in personal factors and ecological variables, *Mathematical Population Studies*, **4**, 4, pp. 235 – 258.

Liaw, K.-L., Frey, W.H., and Lin, J.-P., 2002, Location of adult children as an attraction for black and white elderly primary migrants in the United States, *Environment and Planning A*, **34**, pp. 191 – 216.

Lloyd, R., 1997, *Spatial Cognition* (Dordrecht: Kluwer Academic Publishers).

Lloyd, R., 2000, Self-Organized Cognitive Maps, *Professional Geographer*, **52**, 3, pp. 517 – 531.

Long, L.H., 1973, New estimates of migration expectancy in the United States, *Journal of the American Statistical Association*, **68**, 341, pp. 37 – 43.

Long, L., 1988, *Migration and residential mobility in the United States* (New York: Russell Sage Foundation).

Longley, P., 2000, Spatial Analysis in the New Millennium, *Annals of the Association of American Geographers*, **90**, 1, pp. 157 – 165.

Lowry, I. S., 1966, *Migration and metropolitan growth: two analytical models* (Sun Francisco: Chandler).

Macauley, F. R., 1931, *The Smoothing of Time Series* (New York: National Bureau of Economic Research).

Martin, D., 1989, Mapping population data from zone centroid locations, *Transactions of the Institute of British Geographers*, **14**, pp. 90 – 97.

McKendrick, J.H., 1999, Multi-method research: an introduction to its application in population geography, *Professional Geographer*, **50**, 1, pp. 40 – 50.

Meyer, D. A., Mathews, P. H., and Sommers, P. M., 2001, Net interstate migration revisited, *Applied Economic Letters*, **8**, pp. 131 – 134.

Miller, E., 1973, Is out-migration affected by economic conditions?, *Southern European Journal*, **39**, pp. 396 – 405.

Millington, J., 2000, Migration and Age: The Effect of Age on Sensitivity to Migration Stimuli, *Regional Studies*, **34**, 6, pp. 521 – 533.

Mitchell, G., and Dorling, D., 2003, An environmental justice analysis of British air quality, *Environment and Planning A*, **35**, pp. 909 – 929.

Moran, P.A.P., 1948, The interpretation of statistical maps, *Journal of the Royal Statistics Society, Series B (Methodological)*, **10**, 2, pp. 243 – 251.

Moran, P.A.P., 1950, Notes on continuous stochastic phenomena, *Biometrika*, **37**, pp. 17–23.

Morrison, P. A., 1967, Duration of residence and prospective migration: the evaluation of a stochastic model, *Demography*, **4**, 2, pp. 553 – 561.

Morrison, P. A., 1971, Chronic movers and the future redistribution of population: A Longitudinal Analysis, *Demography*, **8**, 2, pp. 171 – 184.

Morrison, P. A., 1975, Population movements and the shape of urban growth: implications for public policy, in *Regional Policy: Readings in Theory and Applications*, edited by J. Friedman and W. Alonso, (Cambridge: M.I.T. Press), pp. 221 – 243.

Morrison, P. A., and Relles, D., 1975, *Recent research insights into local migration flows* (Santa Monica: Rand Corporation).

Mueller, C.F., 1982, *The Economics of Labor Migration: a behavioural analysis* (London: Academic Press).

Myers, G. C., McGinnis, R., and Masnick, G., 1967, The duration of residence approach to a dynamic stochastic model of internal migration: a test of the axiom of cumulative inertia, *Eugenics Quarterly*, **14**, 2, pp. 121 – 126.

Nakaya, T., 2001, Local spatial interaction modelling based on the Geographically Weighted Regression approach, *GeoJournal*, **53**, pp. 347 – 358.

Ogilvy, A. A., 1979, Migration, the influence of economic change, *Futures*, **11**, 5, pp. 383 – 394.

Ogilvy, A. A., 1980, Inter-regional migration since 1971: an appraisal of data from the National Heath Service Central Register and Labour Force Surveys. Office of Population Censuses and Surveys, Occasional Paper 16.

Ogilvy, A. A., 1982, Population migration between the regions of Great Britain, 1971-79, *Regional Studies*, **16**, 1, pp. 65 – 73.

Oliver, R. F., 1964, Interregional Migration and unemployment: 1951-1961, *Journal of the Royal Statistical Society (series A)*, **127**, pp. 42 – 75.

Ord, J.K., and Getis, A., 1995, Local spatial autocorrelation statistics: distributional issues and an application, *Geographical Analysis*, **27**, pp. 286 – 306.

Ord, J.K., and Getis, A., 2001, Testing for local spatial autocorrelation in the presence of global autocorrelation, *Journal of Regional Science*, **41**, pp. 411 – 432.

Pellegrini, P.A., and Fotheringham, A.S., 1999, Intermetropolitan migration and hierarchical destination choice: a disaggregate analysis from the US Public Use Microdata Samples, *Environment and Planning A*, **31**, 6, pp. 1093 – 1118.

Pellegrini, P.A., and Fotheringham, A.S., 2002, Modelling special choice: a review and synthesis of in a migration context, *Progress in Human Geography*, **26**, pp. 487 – 510.

Piaget, J., 1971, *Psychology and Epistemology* (New York: Grossman).

Piaget, J., and Inhelder, B., 1956, *The child's conception of space* (London: Routledge and Kegan Paul).

Piaget, J., and Inhelder, B., 1969, *The psychology of the child* (New York: Basic Books).

Plane, D. A., 1993, Requiem for the Fixed-Transition-Probability Migrant, *Geographical Analysis*, **25**, 3, pp. 211 – 223.

Plane, D. A., and Rogerson, P. A., 1994, *The geographical analysis of population with applications to planning and business* (New York: J. Wiley & Sons).

Plane, D., Rogerson, P., and Rosen, A., 1984, The cross-regional variation of in-migration and out-migration, *Geographical Analysis*, **16**, 2, pp. 162 – 175.

Ravenstein, E.G., 1885, The laws of migration, *Journal of the Royal Statistical Society*, **48**, pp. 167 – 245.

Ravenstein, E.G., 1889, The laws of migration, *Journal of the Royal Statistical Society*, **52**, pp. 241 – 305.

Rees, P. H., 1996, Projecting the National and Regional Populations of the European Union using migration information, in *Population migration in the European Union*, edited by P. H. Rees, J. Stillwell, A. Convey and M. Kupiszewski, (Chichester: J. Wiley & Sons), pp. 331 – 365.

Rees, P., Fotheringham, A.S., and Champion, T., 2003, Modelling migration for policy analysis, in *Applied GIS and Spatial Analysis*, edited by G. Clarke and J. Stillwell (Chichester: J. Wiley), pp. 259 – 296.

Rees, P., Stillwell, J., and Boden, P., 1992, Internal Migration in the 1980s, in *Migration Processes and Patterns, Volume 2: Population Redistribution in the United Kingdom*, edited by J. Stillwell, P. Rees and P. Boden, (London: Belhaven), pp. 1 – 10.

Rees, P. H., Stillwell, J., Convey, A., and Kupiszewski, M., 1996, *Population migration in the European Union* (Chichester: J. Wiley & Sons).

Robinson, W. S., 1950, Ecological Correlations and the Behaviour of Individuals, *American Sociological Review*, **15**, pp. 351 – 357.

Rogers, T. W., 1969, Migration prediction on the basis of prior migratory behaviour: a methodological note, *International Migration*, **7**, pp. 13 – 22.

Rogers, A., Raquillet, R., and Castro, L. J., 1978, Model migration schedules and their applications, *Environment and Planning A*, **10**, pp. 475 – 502.

Rogers, A., Raymer, J., and Willekens, F., 2002, Capturing the age and spatial structures of migration, *Environment and Planning A*, **34**, pp. 341 – 359.

Sakamoto, Y., 1991, *Categorical Data Analysis by AIC* (Tokyo: KTK Scientific Publishers).

Sakamoto, Y., Ishiguro, M., and Kitagawa, G., 1986, *Akaike Information Criterion Statistics* (Tokyo: KTK Scientific Publishers).

Scott, A., and Kilbey, T., 1999, Can Patient Registers give and improved measure of internal migration in England and Wales?, *Population Trends*, **96**, pp. 44 – 55.

Sommers, P. M., 1981, Analysis of net interstate migration revisited, *Social Science Quarterly*, **62**, pp. 294 – 302.

Sommers, P. M., and Suits, D., B., 1973, Analysis of net interstate migration, *Southern Economic Journal*, **40**, pp. 193 – 201.

SpaceStat, 2003, Software for Spatial Data Analysis, available at http://www.terraseer.com/Spacestat.html, last accessed in March 2003.

Stillwell, J., 1985, Migration between metropolitan and non-metropolitan regions in the UK, in *Contemporary studies of Migration*, edited by P.E. White and A. G. van der Knaap (Norwich: Geo Books), pp. 7 – 25.
Stillwell, J., 1986, The analysis and projection of interregional migration in the United Kingdom, in *Population structures and models: developments in spatial demography*, edited by R. Woods and P. Rees (London: Allen and Unwin), pp. 160 – 202.

Stillwell, J., 1991, Spatial interaction models and the propensity to migrate over distance, in *Migration Models, Macro and Micro Approaches*, edited by J. Stillwell and P. Congdon, (London: Belhaven Press), pp. 34 – 56.

Stillwell, J., 1994, Monitoring intercensal migration in the United Kingdom, *Environment and Planning A*, **26**, pp. 1711 – 1730.

Stillwell, J., and Congdon, P., 1991, *Migration Models, Macro and Micro Approaches* (London: Belhaven Press).

Stillwell, J., and Boden, P., 1989, Internal migration: The United Kingdom, in *Contemporary research in population geography: a comparison of the United Kingdom and the Netherlands* edited by J. Stillwell and H.J. Scholten (London: Kluwer Academic Publishers), pp. 64 – 75.

Stillwell, J., and Duke-Williams, O., 2003, A new web-based interface to British census of population origin – destination statistics, *Environment and Planning A*, **35**, pp 113 – 132.

Stillwell, J., Boden, P., and Rees, P., 1988, Internal migration change in the UK: trends based on NHSCR movement data, 1975-6 to 1985-6, Working Paper 510, School of Geography, University of Leeds.

Stillwell, J., Boden, P., and Rees, P., 1990, Trends in internal net migration in the UK: 1975 to 1986, *Area*, **22**, 1, pp. 57 – 65.

Stillwell, J., Duke-Williams, O., and Rees, P., 1995, Time series migration in Britain: the context for 1991 Census analysis, *Papers in Regional Science*, **74**, 4, pp. 341 – 359.

Stillwell, J., Rees, P., and Boden, P., O., 1992, Internal migration tends: an overview, in *Migration Processes and Patterns, Volume 2: Population Redistribution in the United Kingdom*, edited by J. Stillwell, P. Rees and P. Boden, (London: Belhaven), pp. 28 – 55.

Stillwell, J., Rees, P., and Duke-Williams, O., 1996, Migration between NUTS Level 2 Regions in the United Kingdom, in *Population migration in the European Union*, edited by P. H. Rees, J. Stillwell, A. Convey and M. Kupiszewski, (Chichester: J. Wiley & Sons), pp. 275 – 307.

Sugiura, N., 1978, Further analysis of the data by Akaike's information criterion and the finite corrections, *Communication in statistics, Theory and Methods*, A7, pp. 13 – 26.

Swamy, P.A.V.B., 1971, *Statistical inference in random coefficient regression models* (Berlin: Springer).

Symanzik, J., Majure, J.J., Cook, D., and Megretskaia, I., 1997, *Linking ArcView 3.0 and XGobi: insight behind the front end*, Reprint 97-10, Department of Statistics, Iowa State University, Ames.

Vias, A. C., 2001, Reevaluating the relationship between in-, out-, and net migration for nonmetropolitan counties: An update on Beale's U-shaped curve, *Geographical Analysis*, **33**, 3, pp. 228 – 246.

Weeden, R., 1973, Interregional migration models and their application in Great Britain, in *National Institute of Economic and Social Research: Regional Papers II*, edited by P.C. Cheshire and R. Weeden, (London: Cambridge University Press), pp. 41 – 105.

Yano, K., Nakaya, T., Fotheringham, A.S., Openshaw, S., and Ishikawa, Y., 2003, A Comparison of Migration Behaviour in Japan and Britain Using Spatial Interaction Models, *International Journal of Population Geography*, **9**, pp. 419 – 431.

Zelinsky, W., 1971, The hypothesis of the mobility transition, *Geographical Review*, **61**, pp. 219 – 249.