Forthcoming in *Journal of the Royal Statistical Society Series A*

Combining available migration data in England to study economic activity flows over time

Peter W.F. Smith, James Raymer[1] and Corrado Giulietti
Southampton Statistical Sciences Research Institute, University of Southampton

28 October 2009

Information about internal migration in England may come from decennial censuses, surveys or population (health) registers. In this paper, we propose a methodology that allows us to combine aspects from multiple data sources to provide a time series of detailed migration flows. By detailed, we refer to a migration flow table cross-classified by origin, destination, age, sex and economic activity (e.g., employees, retirees or students). Our results can be used to analyse the movements of various population groups between counties in England over time.

Keywords: combining data, economic activity, England, internal migration, log-linear models

1. INTRODUCTION

In order to understand how a population responds to different economic situations, we need to have detailed information on migration patterns at frequent intervals over time. For example, can we identify any associations between the changes in the behaviours of employee and unemployed migrants, and economic recessions or booms? How have the origin-destination movements of students changed over time? Can this be linked to educational policies or changing preferences for education? What about the age-specific migration patterns of current retirees? How have they changed in recent years? These questions are relevant, particularly in times such as now, when the economic situation in England is very uncertain. The aim of this paper is to provide a methodology to estimate the migration flows that could help address these questions and others of this type. Furthermore, because our model includes age, sex and time, we are able to also examine the migration patterns of people at various stages in their life course.

The model developed in this paper allows the combination of multiple sources of internal migration data for the purpose of studying how the migration patterns of six economic activity groups in England have changed over time. To do this, we use three sources of data: the Patient Register Data System (PRDS), which provides annual National Health Service (NHS) registration flows of migration by origin, destination, age and sex from 1999 to 2007; the 2001 Census, which provides the economic activity detail; and the Labour Force Survey (LFS), which

[1] Corresponding author: Southampton Statistical Sciences Research Institute, School of Social Sciences, University of Southampton, Southampton, SO17 1BJ; telephone: +44 (0)23 8059 2935; email: raymer@soton.ac.uk.

contains quarterly information about the age, sex and economic activity of migrants but at very low levels of geographical detail. The result is a synthetic data set of migration flows, cross-classified by county of origin, county of destination, age, sex and economic activity from 1999 to 2007.

The methodology for combining migration data presented in this paper represents an extension of the frameworks provided by Raymer *et al.* (2007) and Raymer *et al.* (2009), who combined census and registration data to study (i) elderly retirement and return migration across twelve area groups in England and Wales and (ii) ethnic migration across nine regions in England, respectively. Our study adds a third data source of migration flows (i.e., the LFS) and increases the number of areas considered (i.e., 47 counties) to study the migration patterns of six economic activity groups, by age and sex, from 1999 to 2007. The idea is similar to that described in Willekens (1994, pp. 17-20), where census, survey and administrative data sets were combined to estimate internal migration flows in the United States during the 1980s.

The study of migration by economic activity groups (e.g., students, employees and retirees) is important because they have distinct patterns of migration and socio-economic needs. With the estimates produced in this paper, local governments can improve their planning policies directed at supplying particular social services or at influencing the levels of migration for these groups. This is also important because migration is currently, and increasingly, the major factor contributing to population change at sub-national levels in England. However, because we do not have detailed and updated information about migrants over time, our understanding of how or why populations change, is (usually) limited to the most recent census, which only captures the detailed migration patterns for a one-year period. Furthermore, this information is often out-dated. For example, at the time of writing, England's most recent census occurred over eight years ago. The methodology set out in this paper offers a pragmatic solution, one that makes the best use of available data. Finally, we show the reasonableness of our results, and the levels of detail available for analysis or population planning.

To combine migration data from different sources, one first has to account for the differences in measurement (see Bell *et al*, 2002; Long and Boertlein, 1990; Morrison *et al.*, 2004; Rogers *et al.*, 2003; Rogerson, 1990; Willekens, 1994). For example, population registration systems tend to capture migration events, which can occur multiple times within a one-year time period, whereas censuses or surveys tend to capture changes in residential status (or transitions) from one point in time to another. These two data collection systems represent two different types of migration data, i.e., 'migrations' and 'migrants' (Rees and Willekens, 1986) and, depending on the time intervals being compared, substantial differences can occur. With time intervals of similar length, however, the patterns exhibited by the different data sources are often similar. For example, Boden *et al.* (1992) analysed migration data obtained from the National Health Service Central Register (NHSCR) and the 1991 Census in England and Wales. They found high levels of correlation between the annual in-migration, out-migration and net migration totals. More recently, Raymer *et al.* (2007), in analysing elderly internal migration, found that that the main differences between the 2001 PRDS flows and the 2001 Census flows were in the levels. The spatial patterns, on the other hand, were very similar once the levels were controlled for. The knowledge that censuses and population health registers in England exhibit similarities in their underlying structures, after controlling for differences in levels, allows us to develop a methodology for combining these two data sources together. In this paper, we also test whether data from the LFS can be brought in to further improve the estimation of detailed migration patterns.

There have been many studies that have analysed internal migration flows in the United Kingdom (e.g., Bates and Bracken, 1982, 1987; Bell *et al.*, 2002; Bell and Rees, 2006; Bruegel, 2000; Champion, 1996; Dixon, 2003; Finney and Simpson, 2008; Kalogirou, 2005; Stillwell, 1994). Others have developed regression models to estimate various patterns (see, e.g., Fotheringham *et al.*, 2004 and the recent review by Rees *et al.*, 2009). The main difference between our work and the above studies is that our work identifies the key structures in the cross-classified data that are influencing the patterns, and then uses this information to combine multiple data sets for a more complete picture of the patterns over time. The other studies do not combine data sets but rather describe or model them in isolation.

2. AVAILABLE DATA

In England, internal migration data are available in several sources, such as the decennial censuses, the PRDS and the LFS. Censuses contain much of the detail needed for analyses, but are only available every ten years and have problems with compatibility over time for certain variables (Stillwell and Duke-Williams, 2007). Migration data from the PRDS are available annually, but with minimal information on migrant behaviour (i.e., only origin, destination, age and sex are available) and with a tendency to miss important population groups, such as young adult males, who are known to be less inclined to register (Fotheringham *et al.*, 2004). However, the registration data constitute a good up-to-date source of internal migration as nearly all residents in England are patients of a general practitioner employed by the National Health Service, including those who may also have private healthcare provision. Furthermore, the average delay between moving house and registering with a new general practitioner is about one month (Office for National Statistics, 2002). Note, the Office for National Statistics (ONS) produces two migration data sets based on health service registers (Office for National Statistics, 2005; Rees *et al.*, 2009). The first, capturing events of migration between at health authority geographies, is the NHSCR data. The second is the PRDS, which measures transitions in the NHS register and is available at the local authority district level (which can be aggregated into counties). These data are available from 1999 and onwards and, more importantly, are consistent with the geographical units of the 2001 Census data.

The LFS provides quarterly migration data with a rich detail of socioeconomic and demographic characteristics. These include, for example, ethnic group, country of birth, occupation and wages. The major disadvantage of the LFS for the purpose of studying internal migration is that the sample size is relatively small, which means, for example, that the highest level of spatial detail available for analysis is regions. Also, the migration data represent persons who changed their address in the past year, mixing residential movers and migrants. For this study, this appears not to be a major issue. We compared the proportions of economic activity migrants by sex in the 2001 LFS with those for inter-county migration in the 2001 Census. The patterns were found to be very similar, with the only exceptions being the proportions of female employee and inactive migrants (see Figure 1).

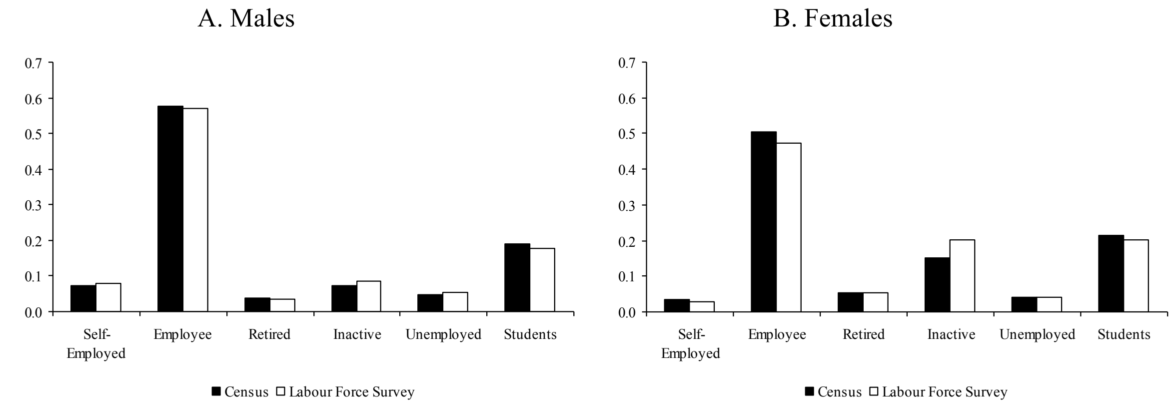


Figure 1. Proportion of migrants in each economic activity group by sex: A comparison of data from the Census (inter-county moves) and the Labour Force Survey (all moves), 2001

For this study, we estimate the 1999 to 2007 annual migration flows between the 47 counties of England as defined in the 1991 Census (see Figure 2) and for twelve five-year age groups (15-19, 20-24, ..., 70-74 years), two sexes and six economic activity groups (self-employed, employee, retired, inactive, unemployed and student). The Census and PRDS data were obtained from ONS by request. The 2001 Census migration data are contained in 'Table MG105'. The 1999-2007 annual PRDS data are included in tables 'Tab2aLA' (origin by destination) and 'Tab1LAQ' (origin by age by sex and destination by age by sex). Both the Census and PRDS data were tabulated at the local authority level. We used the '2001OA to 1991 Wards Lookup v1.0', provided by ONS on a CD-ROM, to aggregate these data to county level. Finally, the 1999-2007 LFS data were downloaded from the Economic and Social Data Service website (http://esds.ac.uk/) and pooled to form yearly data, where each year begins in April, the same month as the Census.

The cross-classified tables of 2001 Census, PRDS and LFS migration obtained for this study are summarised in Table 1. The available tables are denoted by letters. For example, ODAS is a four-way (origin by destination by age by sex) table of migration flows and AET a three-way (age by economic activity by time) table of migration flows. Our aim is to estimate a six-way ODASET table of migration flows by using pertinent information contained in the available tables. Note, in the next section, we use the same notation to refer to association structures. For example, an OD table of migration can also be thought of as a partial table or a structure within the ODASET table.

Source: boundaries extracted from http://edina.ac.uk/ukborders/

Figure 2. Map of counties in England, 1991 definition

Table 1. The sources of available migration data used to estimate the ODASET table for England, 1999-2007

Source	Available	Used
2001 Census	ODAS, ODSE	ODE
1999-2007 Patient Register Data System	ODT, OAST, DAST	ODT, OAT, DAT, AST
1999-2007 Labour Force Survey	ASET	AET, SET

Note: O = origin, D = destination, A = age, S = sex, E = economic activity and T = time period.

3. LOG-LINEAR MODELS FOR COMBINING DATA

3.1 Identifying Key Structures

Our methodology begins by specifying a model for the migration flow table of interest. This first involves a review of the general migration literature and the context in which the migration is occurring. Where possible, the hypothesised model should then be tested against empirical data, which may be obtained from multiple sources. The testing can be undertaken by comparing various unsaturated hierarchical log-linear models fitted to the available migration flow tables.

Based on the model selected by Raymer et al. (2009) and our general knowledge of migration by economic activity, we believe a good model for the five-way ODASE table (ignoring time for now) would contain the following key structures: ODE, ASE, OA and DA. This implies that for migration, the origin-destination patterns and age and sex distributions differ by economic activity, as well as the age patterns by county of origin or destination. To confirm our hypotheses, we assessed the importance of these structures in three of the available tables for the year 2001: the Census's ODAS and ODSE tables and the Labour Force Survey's ASE table (see Table 2). Our goodness-of-fit measure divides the likelihood ratio statistic, G^2, by the residual degrees of freedom. This measure allows us to compare the models by controlling for their relative complexities, which is useful for identifying the best model in terms of overall fit and simplicity. The residual degrees of freedom represent the number of parameters 'not used' to predict the flows. To calculate the residual degrees of freedom, we simply subtract the number of parameters in the unsaturated model from the number of parameters in the saturated model. Note, structural zeros are used where origin i = destination j, i.e., non-migrants and intra-county migrants are excluded from the analyses in Table 2 (and throughout the paper).

Table 2. Selected hierarchical log-linear model fits of the Census's ODAS and ODSE tables of inter-county migration and the Labour Force Survey's ASE table of all moves, 2001

Model	G^2	P	G^2 / rdf
A. 2001 Census ODAS table			
OD, A, S	409,902	2,982	6.0
OD, OA, DA, AS	235,639	4,289	3.5
OD, OA, OS, DA, DS, AS	233,824	4,397	3.5
ODA, AS	95,986	35,760	2.7
ODAS	0	71,280	0.0
B. 2001 Census ODSE table			
OD, OE, DE, SE	199,821	2,633	8.6
OD, OS, OE, DS, DE, SE	197,814	2,725	8.5
ODE, SE	60,222	12,748	4.6
ODSE	0	25,944	0.0
C. 2001 Labour Force Survey ASE table			
A, S, E	16,518	18	131.1
AS, E	16,351	29	142.2
AE, S	2,116	73	29.8
SE, A	15,066	23	124.5
AE, SE	664	78	10.1
AS, AE, SE	379	89	6.9
ASE	0	144	0.0

Notes: G^2 = likelihood ratio statistic versus the satuarated model; residual degrees of freedom (rdf) = number of parameters (P) in the saturated model less the number of parameters in the unsaturated model; O = origin, D = destination, A = age, S = sex and E = economic activity; structural zeros are included for origin *i* = destination *j*.

For an ODAS table of migration representing flows between nine regions in England, Raymer *et al.* (2009) selected a hierarchical model with the OD, OA, DA and AS structures in preference to a more complicated model with a three-way ODA interaction term that fitted slightly better. We repeated the analysis but for migration flows between 47 counties in England, and came to the same conclusion (for additional examples of log-linear models applied to age-specific patterns of migration, see Raymer and Rogers, 2007; van Wissen *et al.*, 2008; Willekens, 1994). A selection of the log-linear model fits are listed in Table 2A. The model with the three-way ODA interaction term slightly improved the fit over the two-way interaction models, but at the expense of an additional large number of parameters. The simpler two-way interaction model implies that the age patterns of origin-destination-specific migration are captured by the age patterns exhibited by the total in-migration and out-migration flows. Notice that the OS and DS structures do not contribute substantially to the overall model fit.

For the ODSE table, we also compared the relative fits of log-linear models. A selection of the model fits are presented in Table 2B. We conclude that a hierarchical log-linear model with the ODE structure and the SE structure provided a good fit in terms of G^2 and simplicity, which implies that the six economic activity groups have different spatial and sex patterns of migration. The other structures, such as OS, DS or OES, did not substantially influence the model fit.

Although we were unable to formally analyse the significance of ASE in the Census data, we believe our model should have this structure included. However, when analysing this table for the 2001 LFS, we found that we only needed AS, AE and SE interactions (Table 2C). Therefore, we dropped the more complicated three-way ASE interaction from our model and instead considered the ODE, OA, DA, AS, AE, SE model as a good representation of the five-way ODASE table. The final set of tables used in the estimation of migration flows are set out in the final column of Table 1. Where possible, we allowed the interactions to evolve over time.

Finally, it should be noted that one could have simply used all of the association structures in the available data to estimate the detailed migration flows over time, i.e., for this study, the tables listed in the second column of Table 1. However, we believe our strategy, which imposes conditional independencies between some of the variables, results in better estimates over time because only the structures identified as being important, both empirically and theoretically, are included. It also smoothes over possible sample variations and natural year-to-year variability, thus preventing inaccuracies or biases in the higher order structures being unnecessarily imposed on the estimated flows.

3.2 Model Specification

Our objective is to estimate migration flows for an ODASE table for each year from 1999 to 2007 (i.e., the ODASET table) with the diagonals of the OD partial tables (i.e., the within-county flows) excluded. The basic idea is to supplement information from the PRDS with more detailed information from the Census and LFS. The log-linear models developed by Raymer *et al.* (2007) and Raymer *et al.* (2009) are used as a starting point. These models combine marginal information available in the incomplete registration data (e.g., OD, OAS, DAS), with complete (but outdated) census data (e.g., ODSE). In essence, the association structures of the census data are imposed on the registration data. We extend this model to also include association structures from a second auxiliary data source, i.e., the LFS.

Spatial interaction models can be thought of as log-linear models (Willekens, 1983), and are commonly used to model origin-destination-specific migration flow data. Overviews of these models and frameworks can be found in Fotheringham *et al.* (2000, pages 213-235), Stillwell (2008) and Willekens (1999). A simplistic version of the spatial interaction model to estimate the number of migrations in an incomplete data set, from origin *i* to destination *j* during a unit interval may be applied as in Willekens (1999):

$$\mu_{ij}^{OD} = \tau_i^O \tau_j^D m_{ij}^{OD}, \quad (1)$$

where μ_{ij}^{OD} is the expected number of migration flows during the respective time interval and $i, j = 1, 2, \ldots, R$ for R origins and destinations. The τ_i^O and τ_j^D parameters represent background factors related to the characteristics of the origin and destination, respectively. The m_{ij}^{OD} factor is the auxiliary information on migration flows. This is additional data relating to migration between the same origins and destinations as in the incomplete data but is not a parameter in the model. As a result, the associations between origins and destinations in the auxiliary data are replicated in the estimated table of flows.

The above model focuses on estimating migration flows between two dimensions, origin and destination. Raymer *et al.* (2007) extended this model to include a third variable of interest not available in the incomplete migration data. For example, an origin by destination by economic activity table can be modelled by using the following log-linear with offset form of the spatial interaction model:

$$\log \mu_{ijz}^{ODE} = \lambda_i^O + \lambda_j^D + \log m_{ijz}^{ODE}, \qquad (2)$$

where μ_{ijz}^{ODE} is the expected flows from origin *i* to destination *j* for level *z* of the third variable. The λ_i^O and λ_j^D parameters are related to the characteristics of the origin and destination, respectively, and m_{ijz}^{ODE} is the auxiliary information on migration flows. Note, there are no parameters corresponding to the dimension indexed by *z*. Here, we rely on the auxiliary data to provide the missing margin and association structures not contained in the incomplete data.

Raymer *et al.* (2007) assumed the three-way auxiliary interaction structure remained constant over time. Raymer *et al.* (2009) allowed this structure to vary over time from 1991 to 2007 by using information from two censuses to geometrically interpolate the counts from 1992 to 2000 and to geometrically extrapolate forward from 2001. We, on the other hand, use the annual LFS data, to allow the AE and SE association structures to evolve over time (but keep the ODE association structure from the Census fixed).

To fit our overall model identified in Section 3.1, we use the ODT, OAT, DAT and AST tables from the PRDS data (see Table 1) and impose the three-way ODE associations from the Census and the AET and SET associations (and levels) from the LFS. This is achieved by first fitting the following log-linear model:

$$\log\left(\mu_{ijxyzt}^{*}\right) = \lambda_{xzt}^{AET} + \lambda_{yzt}^{SET} + \log\left(m_{ijz}^{ODE}\right). \qquad (3)$$

This model provides an estimate of the counts in the ODASET table, $\hat{\mu}_{ijxyzt}^{*}$, that have the same AET and SET associations as the LFS and the same ODE associations as the Census. The logarithms of these counts are then used as an offset in the following model:

$$\log\left(\mu_{ijxyzt}^{ODASET}\right) = \lambda_{ijt}^{ODT} + \lambda_{ixt}^{OAT} + \lambda_{jxt}^{DAT} + \lambda_{xyt}^{AST} + \log\left(\hat{\mu}_{ijxyzt}^{*}\right). \qquad (4)$$

This model combines the ODE, AET and SET association structures resulting from model (3) with the three-way ODT, OAT, DAT and AST association structures from the PRDS. The resulting estimated flows now have all the required association structures, as identified in Section 3.1, and are benchmarked to the levels in the PRDS tables. Should a different model for the flows be thought appropriate, for example one with the ASET interaction or one without ODE interaction, then Models (3) and (4) can be modified by adding or removing interaction parameters, or by changing the offset terms, provided the pertinent information is available in one of the data sources.

Since our interest is primarily in the flows rather than the λ-parameters, we follow Raymer *et al.* (2009) and apply an iterative proportional fitting algorithm to obtain the maximum likelihood estimates of the flows. For Model (3), the initial values are given by the counts in the ODE table from the census: $\mu_{ijxyzt}^{*(0)} = m_{ijz}^{ODE}$ for all *x, y* and *t*. They are then successively multiplied by adjustment factors so that the marginal tables match the counts in the AET table and then the SET table, both from the LFS. This is repeated until the marginal tables of estimated flows simultaneously match all of the counts contained in the two LFS tables. Model (4) repeats this process, using as initial values the $\hat{\mu}_{ijxyzt}^{*}$ resulting from Model (3) and matching with the four PRDS three-way tables.

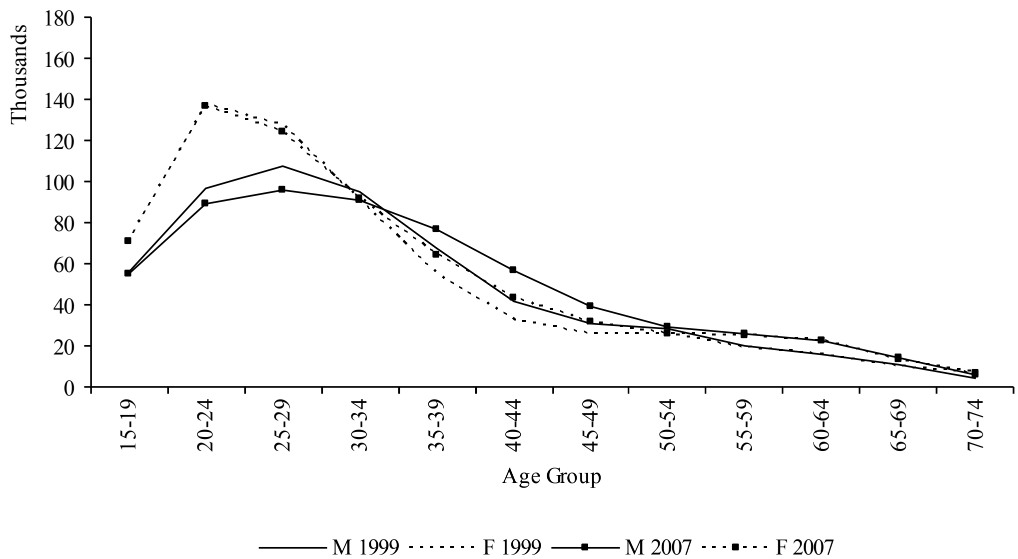
3.3 Model Implementation

The algorithm to fit Model (4) requires consistency in the marginal distributions of the incomplete data, namely of the OD, OA, DA and AS tables for each year. Ideally, these would have come from a single four-way table. However, when we extracted the one-way margins from the OD, OAS and DAS tables provided by ONS, they did not match because the OAS and DAS tables included within county migration and migration to and from Wales, and the OD table included migrants aged 0-14 years and migrants aged 75+ years. To make them consistent, for each year, we used the following procedure. We started with the OD table and removed the diagonal elements and the rows and columns corresponding to areas in Wales. We then scaled the OA table so that its O margin matched that in the reduced OD table. Likewise, we scaled the DA table so that the D margin matched that in the reduced OD table. Next, we subtracted four age groups (i.e., 0-4, 5-9, 10-14 and 75+) from the scaled OA and DA tables, which were then used to scale the reduced OD and AS tables so that their O, D and A margins matched. Hence, all four tables required for modelling had the same one-way margins as required and had the necessary age groups for the study.

The PRDS migration data have a problem relating to the undercounting of young males, as seen in Figure 3A for 1999 and 2007. However, the 1991 and 2001 censuses indicated that the proportions of young adult male and female migrants were approximately equal (Figure 3B). The reason for this difference has primarily to do with males being less likely to register, particularly in their young adult years (see Fotheringham *et al.*, 2004, pages 1637-1640). Note, this was not an issue in Raymer *et al.* (2007) because they only examined migration patterns of elderly persons, a group less likely to be missed in a health service population register.

As illustrated in Figure 3A, nearly all the differences in the age patterns of male and female migration as reported in the PRDS data occur in the 15-19 year, 20-24 year and 25-29 year age groups. To correct for the differences in the age-sex patterns, we follow the procedure used in Raymer *et al.* (2009), which assumes that females are counted accurately in the PRDS data and weights the estimates from Model (4) to account for the age-sex differences. The weights represent ratios of female to male migration for the 15-19, 20-24 and 25-29 age groups, marginalising over origin, destination and economic activity. This approach maintains all of the associations implied by Model (4). The weights applied to the male migrants in the three age groups are set out in Table 3, along with the resulting adjustment ratios for all males (a 13 to 15 percent increase) and both sexes together (a 6 to 7 percent increase).

A. Patient Register Data System, 1999 and 2007



B. Census, 1991 and 2001

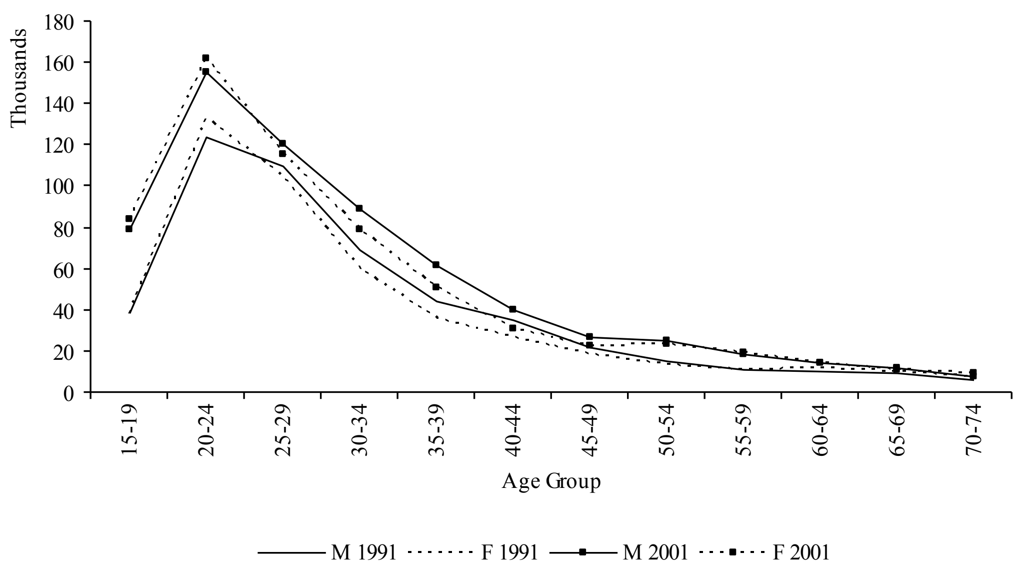


Figure 3. Age patterns of inter-county migration in England by sex: Patient Register Data System (1999 and 2007) and Census (1991 and 2001)

Table 3. Adjustment ratios for migration flows from the Patient Register Data System, 1999-2007

Year	Age Group (Males) 15-19	20-24	25-29	All Males	Both Sexes
1999	1.264	1.422	1.182	1.131	1.063
2000	1.263	1.430	1.193	1.131	1.063
2001	1.277	1.441	1.187	1.134	1.064
2002	1.270	1.454	1.197	1.133	1.063
2003	1.289	1.468	1.211	1.137	1.066
2004	1.295	1.479	1.233	1.140	1.067
2005	1.287	1.454	1.231	1.140	1.067
2006	1.311	1.491	1.259	1.150	1.071
2007	1.289	1.523	1.295	1.153	1.072

4. RESULTS

In this section, the estimated interregional migration flows by age, sex and economic activity group are presented. These estimates have been adjusted to correct for the male undercount in three age groups as discussed in the previous section. The full set of estimates produced by Models (3) and (4) consists of more than 2.8 million cells. The evolution of patterns over time is first described, followed by a focus on the age, sex and spatial patterns of employee, retired, inactive and student migrants from the Greater Manchester and Hampshire counties. These two counties and four economic activity groups were selected out of the set of estimates for illustration purposes only.

It is important to point out that the patterns described below are of economic activity groups, as measured at the time of the census or survey. There is no information on the economic activity status of these migrants prior to the move. Therefore, it is not possible to attach any causal relationships between economic activity and migration. For example, we do not know if an employed migrant was, say, employed, unemployed or a student prior to migrating. However, the patterns that we describe are informative about the movements of these persons prior to their current economic activity status and could be used to forecast future trends in economic activity by origin, destination, age and sex.

4.1 The Evolution of Economic Activity Over Time

The overall levels of inter-county migration by economic activity are set out in Figure 4 for the years 1999 to 2007. During this time, the total levels of migration increased from about 1.20 million to around 1.26 million. Employees make up the largest share of the total migrants. In 1999, this group represented about 50 percent of the flows. In 2007, this share was about 52 percent. Similarly, the other economic activity groups exhibited relatively small changes in their shares over time. Students and inactive persons represent the two other large flows. Students are counted at their term time address, rather than their home address, in the 2001 Census. The inactive population consists of all individuals out of the labour force, such as those who take care of household members or those that are permanently sick or disabled. Flows of self-employed migrants are relatively small because this group only makes up a small portion of the population and possibly because their economic activity makes them less mobile than other groups. The share of unemployed as a total of the working age population (i.e., excluding male migrants aged over 64 and female migrants over 60) is on average five percent, in line with the national average unemployment rate (according to information obtained from the ONS's Nomis website, www.nomisweb.co.uk). Finally, retired migrants (up to 74 years old) represent the smallest group.

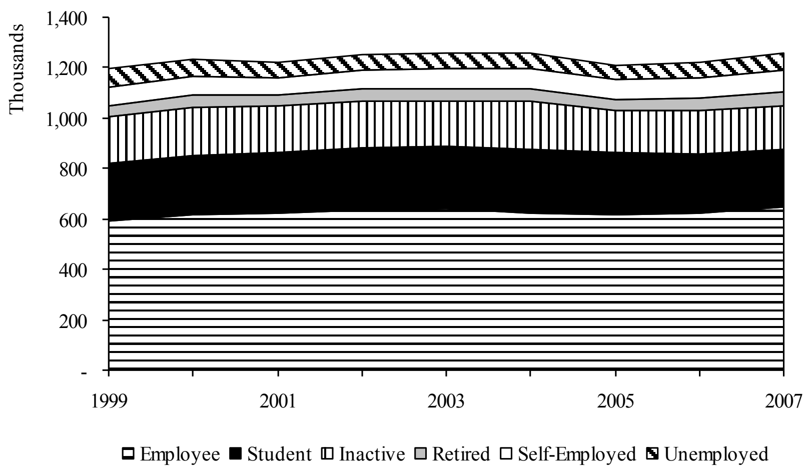


Figure 4. The levels of inter-county migration in England by economic activity group, 1999-2007

4.2 Spatial Patterns of Migrants from Greater Manchester and Hampshire Counties

The flows from Greater Manchester and Hampshire to the top ten destinations for employees, retired, inactive and students in 1999 are presented in Figures 5 and 6, respectively. The size of each arrow is proportional to the share of migrants (within each group) that move to a particular destination (refer to Figure 2 for names of counties).

The top left panel in Figure 5 presents the top ten flows of employee migrants from Greater Manchester. The two main destinations of Lancashire and Cheshire receive roughly 30 percent of all employee migrants. The other nearby destinations are Merseyside, West Yorkshire and Derbyshire. Not all employee migrants move to neighbouring counties, as demonstrated by the relatively sizeable flows towards Inner and Outer London and by the flows to Berkshire and Hampshire. The second panel illustrates the out-migration of retired migrants. Here, the largest share is to Lancashire, a well-known destination for retirees in the North of England. The remaining flows are distributed across several destinations in the North (e.g., Cumbria and North Yorkshire) and in the South (Cornwall & Isles of Scilly and Devon). The bottom left panel depicts migration of inactive migrants. The rather diverse type of destinations is indicative of the heterogeneity of this group; although the top four receiving counties (including London) are the same as those for employee migrants, other flows are directed to destinations that are more common for retired migrants, such as Cumbria and Somerset. The last panel presents student migrants. A large share of students had migrated to neighbouring counties that include university towns, such as West Yorkshire, Lancashire and Merseyside. The remaining flows are distributed across other counties where large institutions are located, such as Tyne and Wear, London and West Midlands.

Figure 6 presents the migration from Hampshire to the top ten destinations for the same four economic activity groups as Figure 5. Here, employee migrants are more focused in terms of destination choice than those from Greater Manchester. This can be explained partly by Hampshire's relative proximity to London, which attracts about 30 percent of the flows of employee migrants. A relatively large share of employee migrants also moves to Surrey. The remaining flows are to neighbouring counties, such as Berkshire and West Sussex in the South East and Dorset and Wiltshire in the South West. The pattern of retired migrants is presented in the top right panel. The majority of retired migrants have moved to counties that are well-known tourist destinations in the South West region (Dorset, Devon and Cornwall & Isles of Scilly). The remaining top destinations are neighbouring counties, with the exception of Lincolnshire and Norfolk, which are relatively distant locations. Inactive migrants are represented in the bottom left panel. Similarly to the case of Greater Manchester, inactive migrants tend to move partly to destinations that are similar to employees (e.g., Surrey and Outer London) and partly to counties where retired migrants live (e.g., Dorset and Devon). The large majority of student migrants from Hampshire have moved to neighbouring counties, such as Surrey and Inner and Outer London, where there are many university towns in these locations. A minor share of student migrants also moves to relatively distant locations, such as the West Midlands and Devon.

It is useful to investigate how the above spatial patterns evolved over time. In Figures 7 and 8, the migration patterns to the top five destinations for employee, retired, inactive and student migrants are presented for the 1999 to 2007 years for flows from Greater Manchester and Hampshire, respectively. The flow of employee migrants from Greater Manchester increased in all the receiving counties, except Outer London, where the levels remained constant over time. The patterns for retired migrants were irregular for the top two destinations (Lancashire and Merseyside), whereas they were pretty constant for the remaining counties. The patterns for inactive and student migrants were also fairly constant over time.

All top five flows of employee migrants from Hampshire increased over time, except the one towards Surrey. The top five flows of retired migrants, interestingly, exhibited decreases from 2000 to 2005 (except to West Sussex) but had similar levels in 1999 and 2007. The flows of inactive migrants from Hampshire were relatively constant over time, as were the flows of student migrants.

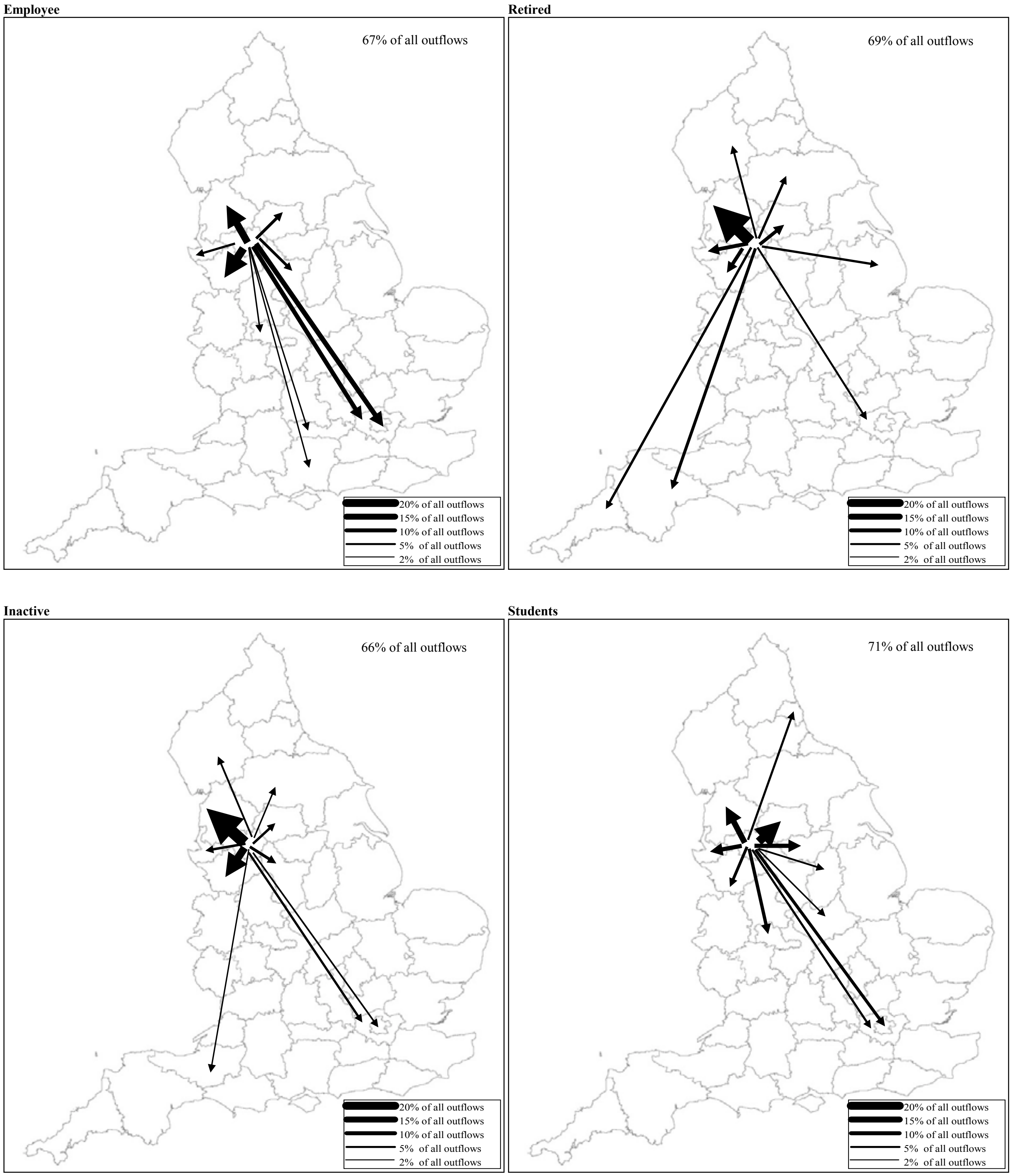


Figure 5. Top 10 destinations from Greater Manchester for employee, retired, inactive and student migrants, 1999

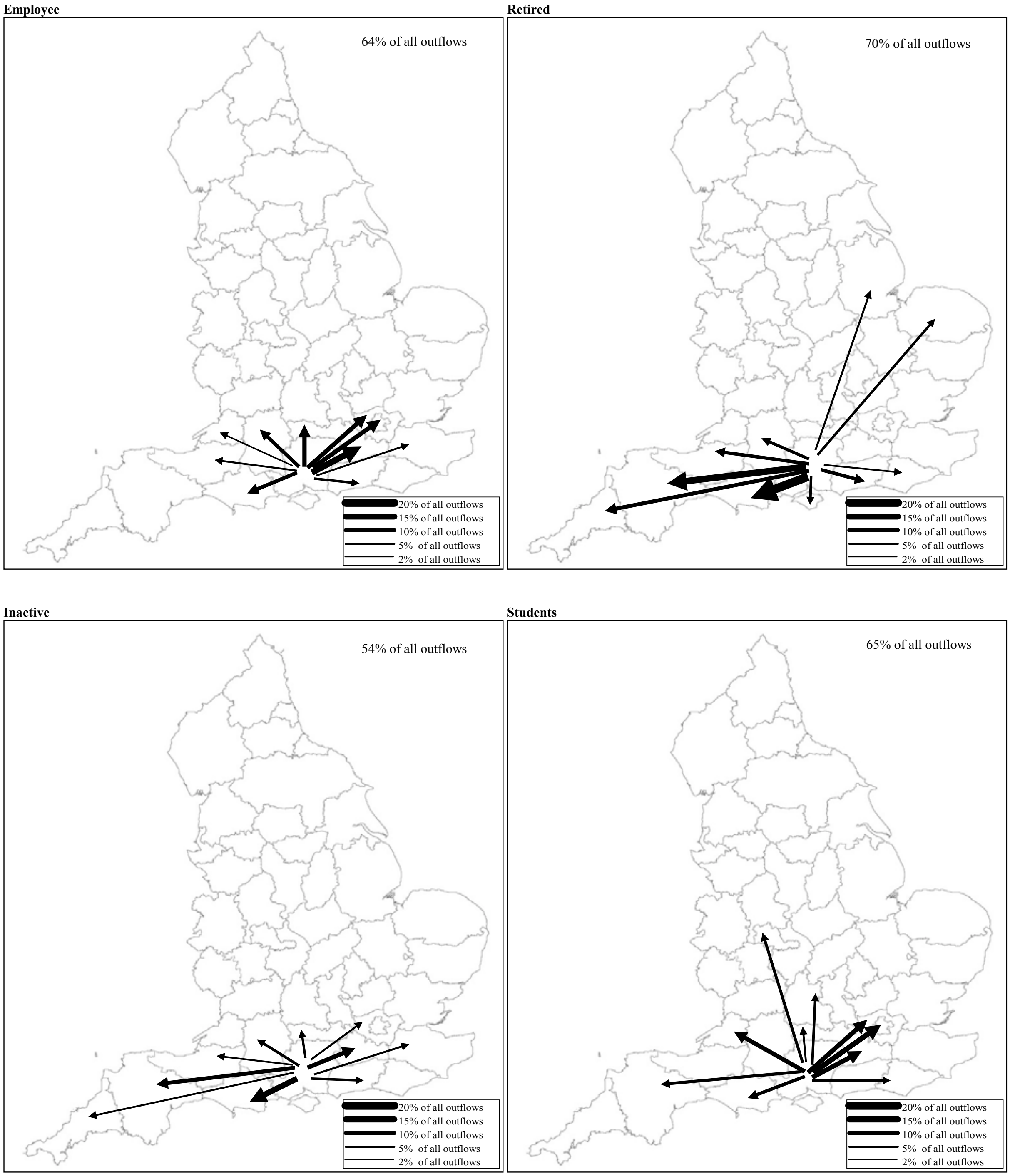


Figure 6. Top 10 destinations from Hampshire for employee, retired, inactive and student migrants, 1999

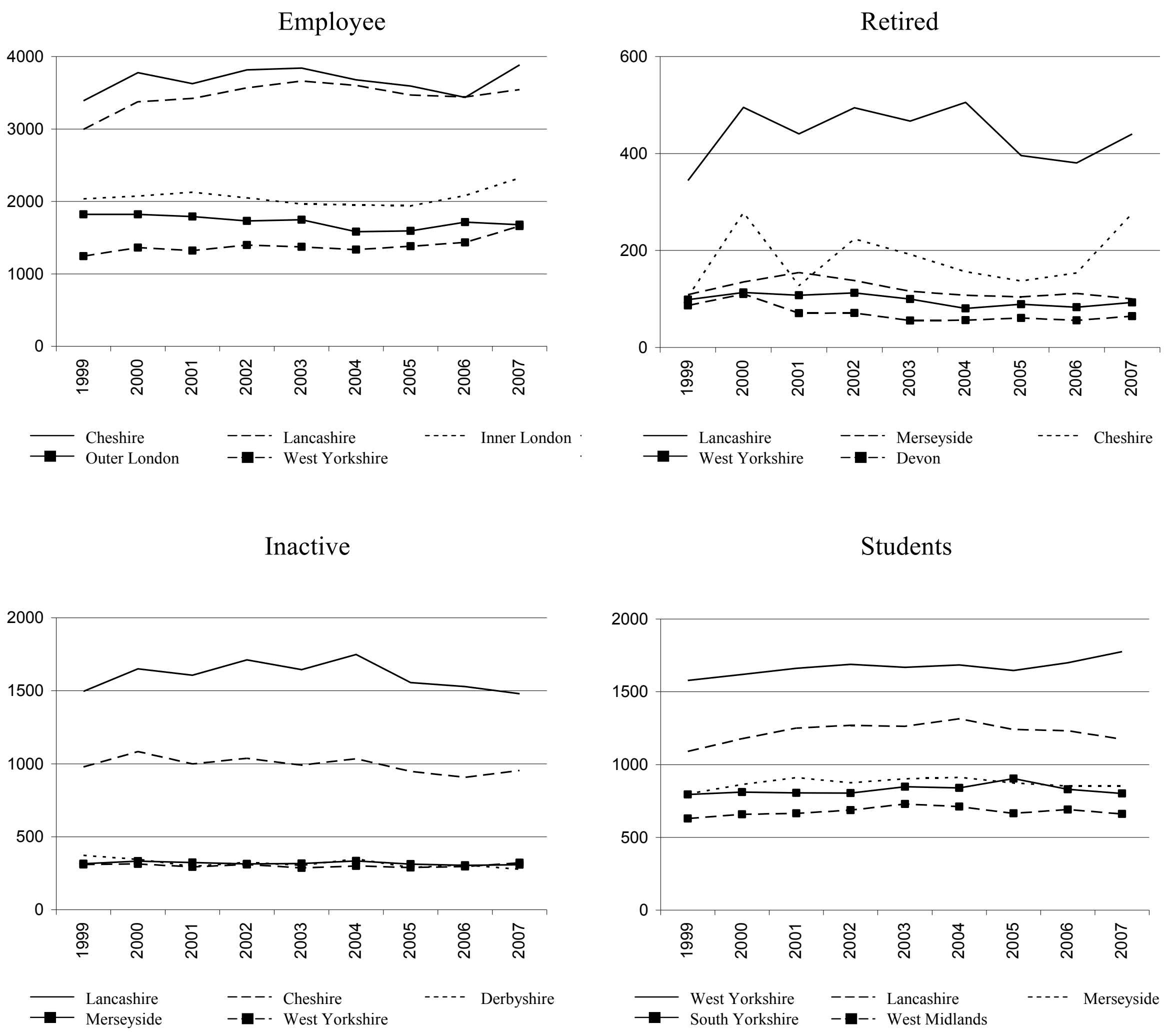


Figure 7. Top five destinations from Greater Manchester for employee, retired, inactive and student migrants, 1999-2007

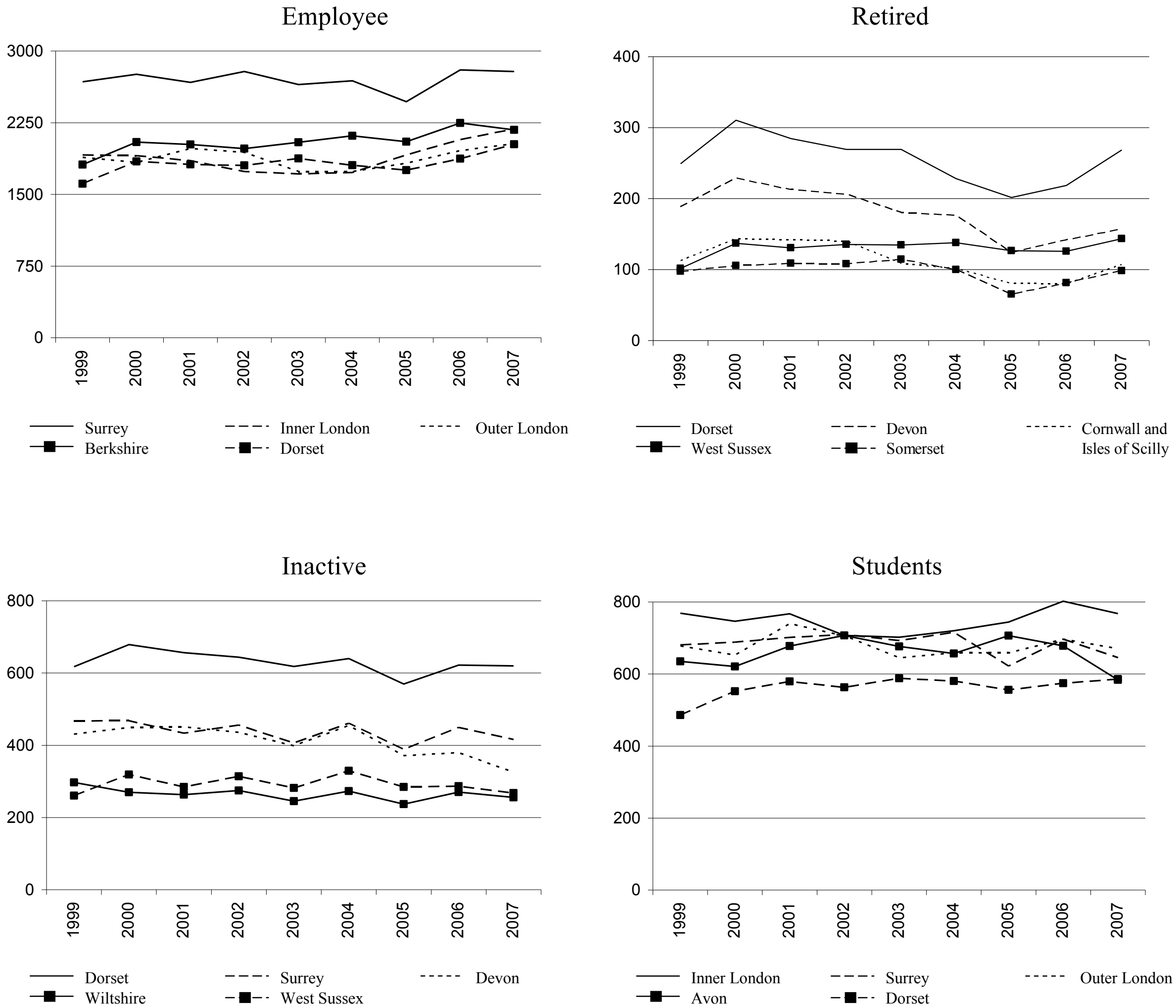


Figure 8. Top five destinations from Hampshire for employee, retired, inactive and student migrants, 1999-2007

4.3 Age- and Sex-Specific Patterns of Migrants from Greater Manchester and Hampshire Counties

In this section, the age- and sex-specific inter-county migration flows for employee, retired, inactive and student migrants are presented. For each of these economic activity groups, we first present the age profiles of male and female migrants to the top destination in 1999. Second, we present the corresponding female age profiles for 1999, 2003 and 2007 to examine changes over time.

The shapes of the employee migrant age profiles from Greater Manchester to Cheshire are very similar for males and females (see Figure 9), albeit with males exhibiting larger numbers in all age groups (except the oldest one). We expect male employees to have higher levels of migration as they generally have higher levels of labour participation. The age profiles of these migrants exhibit a peak in the 25-29 year old age group, followed by a smooth decline towards the older age groups. The flows of retired migrants to Lancashire do not differ between sexes, except for the last age group, where there is a larger number of females, related to their lower mortality and higher population numbers. Flows of inactive female migrants to Lancashire are higher than for males. This is consistent with females being less likely to be in the labour force and more likely to be at home taking care of children. Flows of student migrants to West Yorkshire are represented in the bottom right panel. Levels for females are slightly higher than for males, both with a peak in the 'college' years. The corresponding age profiles of migrants from Hampshire to top destinations are set out in Figure 10. These age-specific patterns are very similar to those from Greater Manchester (Figure 9), with the exception of student migrants to Inner London, who were estimated to have relatively lower numbers of migrants in the 15-19 year old age group.

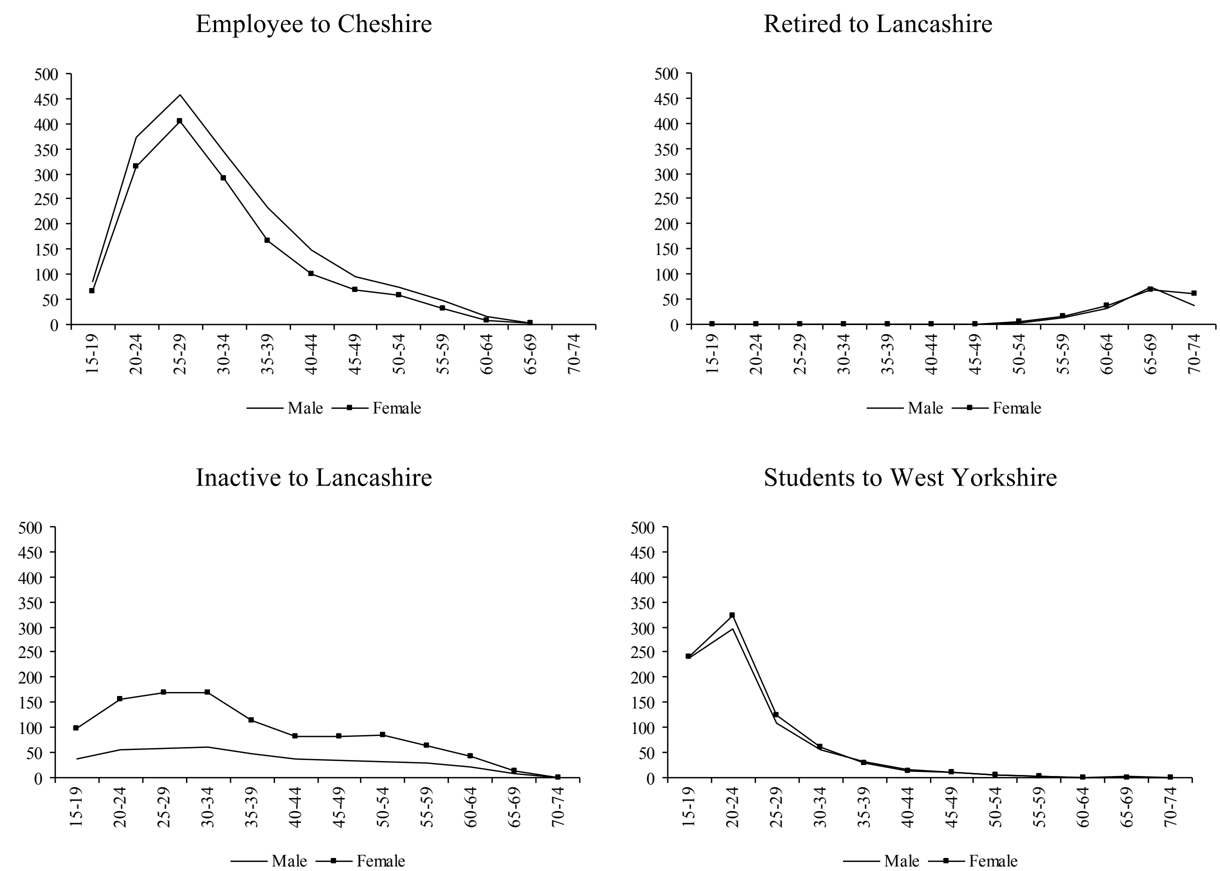


Figure 9. Age and sex patterns of migration from Greater Manchester: The top destinations for employee, retired, inactive and student migrants, 1999

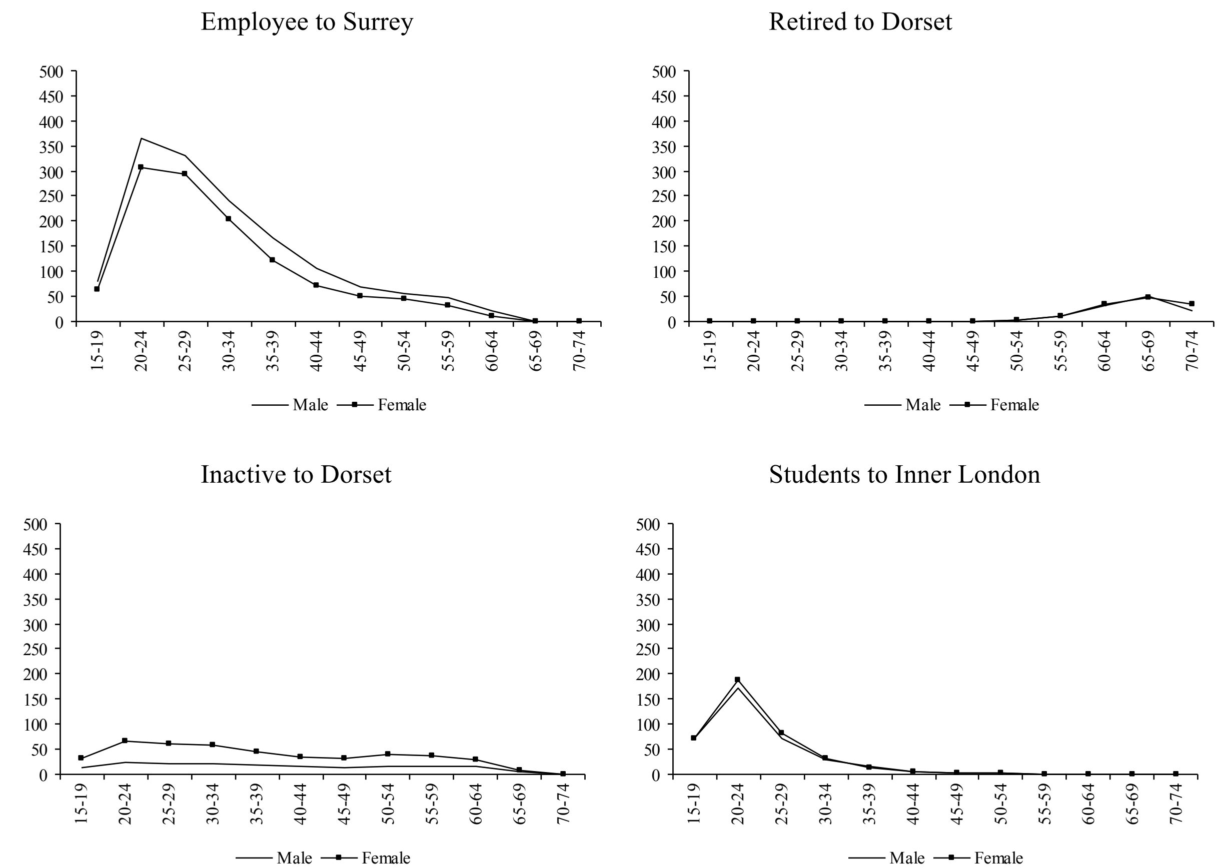


Figure 10. Age and sex patterns of migration from Hampshire: The top destinations for employee, retired, inactive and student migrants, 1999

To see how the predicted age patterns of migration changed over time, consider the female flows for the years 1999, 2003 and 2007 from Greater Manchester and Hampshire set out in Figures 11 and 12, respectively. In Figure 11, we see that the flow of employee migrants to Cheshire increased slightly over time, especially for age groups older than 30 years. The levels of retired migrants to Lancashire increased in the 60-64 year old age group, suggesting that females started retiring earlier. The pattern of inactive migrants to Lancashire is somewhat irregular over time, with some age groups gaining and others decreasing in numbers. On the other hand, there was a uniform increase in all the age groups for female student migrants to West Yorkshire. For the patterns set out in Figure 12, we see in the top left panel that female employee migrants to Surrey decreased in the age groups 20-24 and 25-29 years, but increased in the 35+ year old age groups. The migration of retired and inactive migrants to Dorset did not exhibit any major changes over time. Finally, there was a very slight increase in the flows of 25-29 year old student migrants to London in 2007 relative to the other two years.

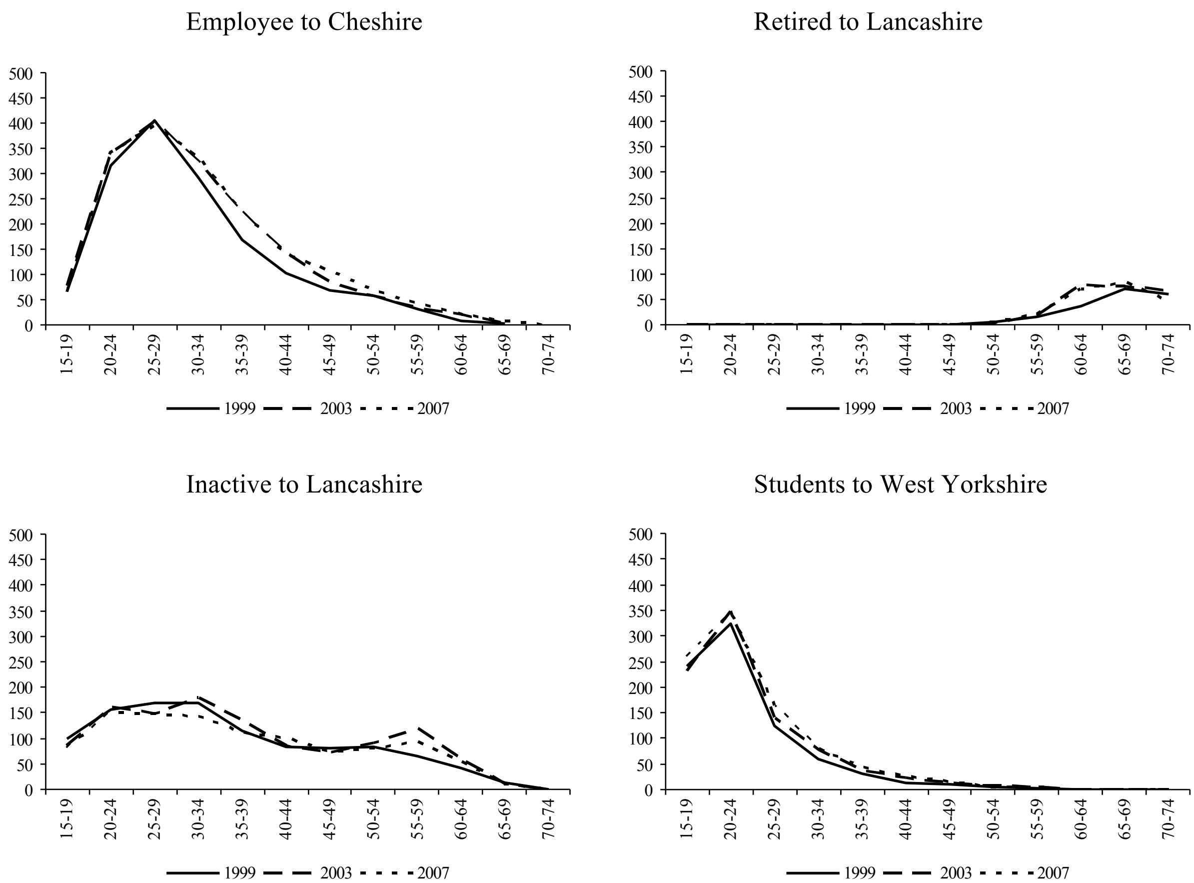


Figure 11. Age patterns of migration from Greater Manchester: The top destinations for female employee, retired, inactive and student migrants, 1999, 2003 and 2007

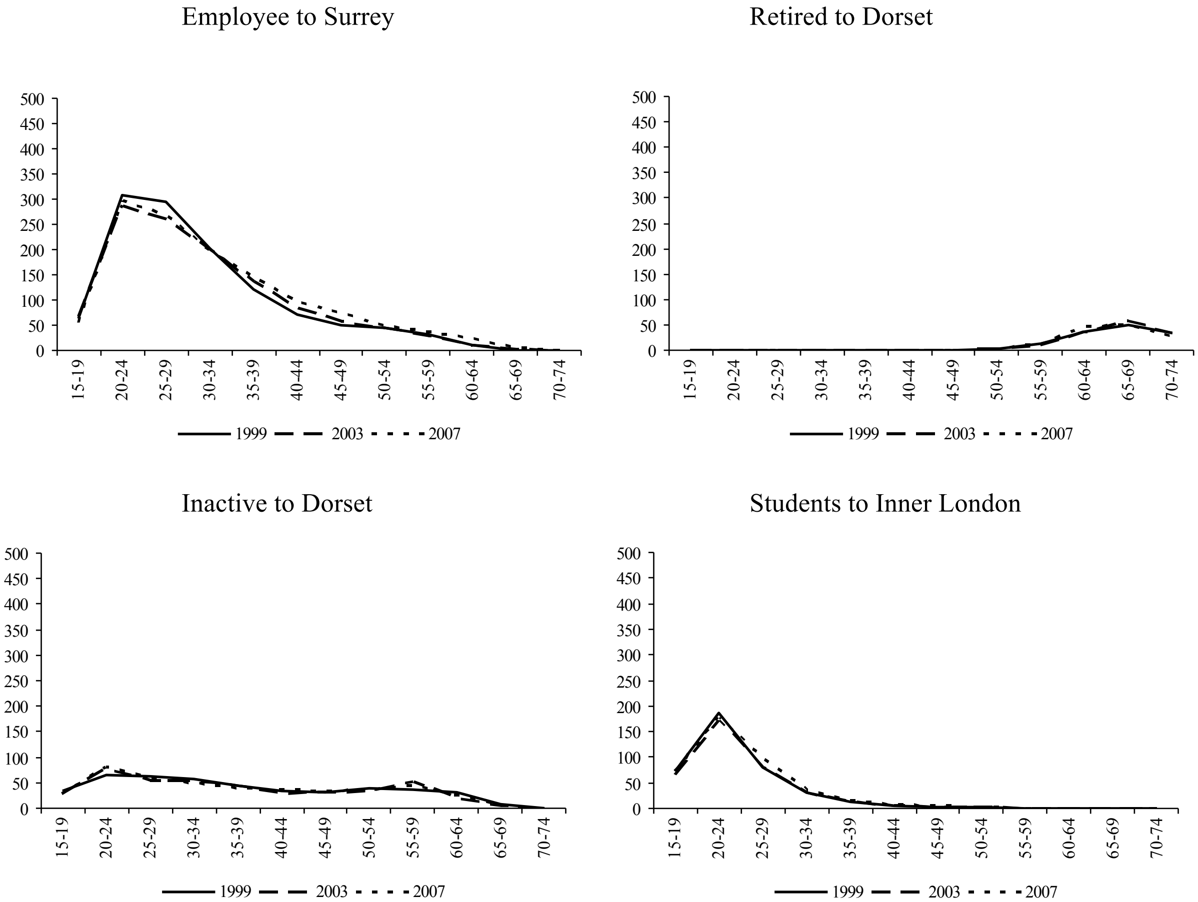


Figure 12. Age patterns of migration from Hampshire: The top destinations for female employee, retired, inactive and student migrants, 1999, 2003 and 2007

4.4 Summary

The above subsections have described a selection of estimated flows resulting from the fitting of Models (3) and (4). It is important to understand that these results have come about by combining information from the 1999-2007 PRDS, the 2001 Census and the 1999-2007 LFS. This synthetic data base provides some indication about how the migration patterns of specific economic activity groups have evolved over time. Earlier studies that have examined such detailed origin-destination-specific migration flows have mostly relied on census data, which occurs only once every ten years. The results presented in this section provide an additional eight years of data based on key structures available from the PRDS and LFS, thus furthering the possibilities for analyses and planning.

The synthetic data produced for this paper provide information on key migrant groups at different stages in their life courses. How these groups differ in their migration behaviours is important for understanding the overall patterns (Geist and McManus, 2008; Plane et al., 2005). We know that education and employment are important drivers of young adult migration (see, e.g., Mak and Moncur, 2003; Faggian *et al.*, 2006) and that retirees have their own separate destination preferences linked to places of birth and amenity (Rogers, 1992), but we do not know much about the patterns of the other economic activity groups or how these patterns affect local economies. We hope that our results will contribute to a fuller understanding of the mechanisms underlying migration.

5. CONCLUSION

Population and migration analysts require detailed and up-to-date information to inform policy and planning. This information, however, is often not available. To overcome this limitation, we have introduced a methodology for combining migration data based on available registration, census and survey data. It extends the earlier work by Raymer et al. (2007) and Raymer et al. (2009) by providing a time series of estimates that include additional information from a second auxiliary source and at a higher level of geography. This paper also demonstrates the generality of the methodology by applying it to a set of migrants with very different age, sex and spatial patterns of migration.

The methods described in this paper will help migration researchers and population planners make the best use of the data that are available to them, whether it comes from registrations, censuses or surveys. The above analysis has demonstrated the type of results that can be obtained from the estimated time series of economic activity migration flows in England by age and sex. It has allowed us to examine migration at various stages in the life course, demonstrating both the different age-specific shapes of migration according to economic activity, as well as the differences exhibited by males and females. Note, we plan to make the entire estimated data set available in the near future so that others may analyse patterns of particular interest to them.

Finally, while we have used observed proportions and association structures from the census and LFS to estimate detailed migration flows, the methodology could also be used to create various scenarios, based on hypothetical proportions and association structures. For example, one could increase the proportion of student migrants over time and assess how this affects the overall migration patterns, assuming the other structures remain unchanged. This could be used to explore the consequences of more people investing in their education during times of recession. Alternatively, we could allow the age distribution of migrants to evolve in line with a projected ageing population (Little and Rogers, 2007).

ACKNOWLEDGMENTS

The authors would like to thank the two anonymous reviewers for their valuable comments and suggestions, and the Economic and Social Research Council (RES-000-22-2501) for supporting this research. The data were provided by the Office for National Statistics and are Crown Copyright. The authors take full responsibility for the analyses and interpretations.

REFERENCES

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

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

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

Bell, M. and Rees P. (2006) Comparing migration in Britain and Australia: Harmonisation through use of age-time plans. *Environment and Planning A*, 38, 959-988.

Boden, P., Stillwell, J. and Rees, P. (1992) How good are the NHSCR data? In *Migration Processes & Patterns: Population Redistribution in the United Kingdom* (eds J. Stillwell, P. Rees, and P. Boden), pp. 13-27. London: Belhaven Press.

Bruegel, I. (2000) The restructuring of London's labour force: Migration and shifting opportunities, 1971-1991. *Area*, 32, 79-90.

Champion, T. (1996) Population review: (3) Migration to, from and within the United Kingdom. *Population Trends*, 83, 5-16.

Dixon, S. (2003) Migration within Britain for job reasons. *Labor Market Trends*, April, 191-201.

Faggian, A., McCann, P. and Sheppard, S. (2006) An analysis of ethnic differences in UK graduate migration behaviour. *The Annals of Regional Science*, 40, 461-471.

Finney, N. and Simpson, L. 2008. Internal migration and ethnic groups: evidence for Britain from the 2001 Census. *Population, Space and Place* **14** : 63-83.

Fotheringham, A.S., Brunsdon, C. and Charlton, M. (2000) *Quantitative Geography: Perspectives on Spatial Data Analysis*. London: Sage.

Fotheringham, A.S., Rees, P., Champion, T., Kalogirou, S. and Tremayne, A.R. (2004) The development of a migration model for England and Wales: Overview and modelling out-migration. *Environment and Planning A*, 36, 1633-1672.

Geist, C. and McManus, P.A. (2008) Geographical mobility over the life course: Motivations and implications. *Population, Space and Place*, 14, 283-303.

Kalogirou, S. (2005) Examining and presenting trends of internal migration flows within England and Wales. *Population, Space and Place*, 11, 283-297.

Little, J.S. and Rogers, A. (2007) What can the age composition of a population tell us about the age composition of its out-migrants? *Population, Space and Place*, 13, 23-39.

Long, J.F. and Boertlein, C.G. (1990) Comparing migration measures having different intervals. Current Population Reports No. 166, U.S. Census Bureau, Washington, DC.

Mak, J. and Moncur, J.E.T. (2003) Interstate migration of college freshmen. *The Annals of Regional Science*, 37, 603-612.

Morrison, P.A., Bryan, T.M. and Swanson, D.A. (2004) Internal migration and short-distance mobility. In *The Methods and Materials of Demography* (eds J.S. Siegel and D.A. Swanson), pp. 493-521. San Diego: Elsevier Academic Press.

Office for National Statistics (2002) Using patient registers to estimate internal migration: Customer guidance notes. Migration Statistics Unit, Office for National Statistics, Titchfield. (Available at http://www.statistics.gov.uk/statbase/expodata/commentary/WebGuidanceNotes.htm)

Office for National Statistics (2005) Estimating internal migration: Customer guidance notes. Migration Statistics Unit, Office for National Statistics.

Plane, D.A., Henrie, C.J. and Perry, M.J. (2005) Migration up and down the urban hierarchy and across the life course. *Proceedings of the National Academy of Science*, 102, 15313-15318.

Raymer, J., Abel, G. and Smith, P.W.F. (2007) Combining census and registration data to estimate detailed elderly migration flows in England and Wales. *Journal of the Royal Statistical Society A*, 170, 891-908.

Raymer, J. and Rogers, A. (2007) Using age and spatial flow structures in the indirect estimation of migration streams. *Demography*, 44, 199-223.

Raymer, J., Smith, P.W.F. and Giulietti, C. (2009) Combining census and registration data to analyse ethnic migration patterns in England from 1991 to 2007. *Population, Space and Place*, DOI: 10.1002/psp.565.

Rees, P., Stillwell, J., Boden, P. and Dennett, A. (2009) Part 2: A review of migration statistics literature. In Migration statistics: The way ahead?, Report 4, UK Statistics Authority, London. Available at http://www.statisticsauthority.gov.uk/assessment/ monitoring-reports/index.html.

Rees, P. and Willekens, F. (1986) Data and accounts. In *Migration and Settlement: A Multiregional Comparative Study* (eds A. Rogers and F. Willekens), pp. 19-58. Dordrecht: D. Reidel.

Rogers, A., ed. (1992) *Elderly Migration and Population Redistribution*. London: Belhaven Press.

Rogers, A., Raymer, J. and Newbold, K.B. (2003) Reconciling and translating migration data collected over time intervals of differing widths. *The Annals of Regional Science*, 37, 581-601.

Rogerson, P.A. (1990) Migration analysis using data with time intervals of differing widths. *Papers of the Regional Science Association*, 68, 97-106.

Stillwell, J. (1994) Monitoring intercensal migration in the United Kingdom. *Environment and Planning A*, 26, 1711-1730.

Stillwell, J. (2008) Inter-regional migration modelling: A review. In *Migration and Human Capital* (eds J. Poot, B. Waldorf and L. van Wissen), pp. 29-48. Cheltenham: Edward Elgar.

Stillwell, J. and Duke-Williams, O. (2007) Understanding the 2001 UK census migration and commuting data: The effect of small cell adjustment and problems of comparison with 1991. *Journal of the Royal Statistical Society Series A*, 170, 1-21.

van Wissen, L., van der Gaag, N., Rees, P. and Stillwell, J. (2008) In search of a modelling strategy for projecting internal migration in European countries. In *Migration and Human Capital* (eds J. Poot, B. Waldorf and L. van Wissen), pp. 49-74. Cheltenham: Edward Elgar.

Willekens, F. (1983) Log-linear modelling of spatial interaction. *Papers of the Regional Science Association*, 52, 187-205.

Willekens, F. (1994) Monitoring international migration flows in Europe: Towards a statistical data base combining data from different sources. *European Journal of Population*, 10, 1-42.

Willekens, F. (1999) Modeling approaches to the indirect estimation of migration flows: From entropy to EM. *Mathematical Population Studies*, 7, 239-278.