Combining Census and Registration Data to Analyse Ethnic Migration Patterns in England from 1991 to 2007

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

7 July 2008 (Draft)

Contact Author: James Raymer

Address: Division of Social Statistics
School of Social Sciences
University of Southampton
Southampton SO17 1BJ
United Kingdom

Email: raymer@soton.ac.uk

Telephone: +44 (0)23 8059 2935

Acknowledgements: This paper was prepared for presentation at the 2008 *European Population Conference* in Barcelona, Spain. Support for this research came from the Economic and Social Research Council (RES-000-22-2501). The data were provided by the Office for National Statistics and are Crown Copyright. The authors take full responsibility for the analyses and interpretations.

ABSTRACT

At present, the publicly available internal migration data in England are limited due to differences in sources, availability, quality and measurement. In this paper, we present a statistical model for combining (incomplete) annual National Health Service registration data with (auxiliary) decennial census data to predict interregional ethnic migration flows by age and sex from 1991 to 2007. Annual flows of migration by origin, destination, age and sex can be obtained from the National Health Service registers. The ethnic detail can only be obtained from the decennial censuses. More detailed and current information on migration flows are needed so that local governments have the means to improve their planning policies directed at supplying particular social services or at influencing levels of migration. Also, these flows are needed for understanding ethnic population redistribution in relation to areas with, for example, high unemployment, high costs of living or high immigrant concentrations.

CONTENTS

1. INTRODUCTION

The study of migration is often hindered by data availability and data quality. Detailed migration data at high levels of spatial disaggregation are generally only available from censuses, which only occur every five or ten years and are often outdated by the time they come out. Surveys, though frequently carried out, are usually not large enough to capture migration movements, as most people remain in a particular locality within a given time period. In other cases, data are inadequate because of non-response or data suppression for confidentiality reasons. This is particularly relevant for analyses at relatively high levels of spatial disaggregation.

In England, there are two main sources of internal migration data. Annual flows of migration by origin, destination, age and sex can be obtained from the National Health Service Central Register (NHSCR). More detailed data, such as by ethnicity, can be obtained from the decennial censuses. Our research extends a methodology developed by Raymer et al. (2007) to combine these two sources of information for the purpose of producing reliable and detailed estimates of internal migration over time. By detailed, we refer to migration cross-classified by five variables: origin, destination, age and sex, and ethnicity. These estimates can then be used to improve our understanding of such migration behaviours, including the forecasting of future patterns.

The advantages to having a consistent and reliable set of migration flows are numerous. Detailed estimates of migration flows are needed so that local governments have the means to improve their planning policies directed at supplying particular social services or at influencing levels of migration. This is important because migration is currently (and increasingly) the major factor contributing to population change at sub-national levels in many countries throughout the world, including England. Furthermore, our understanding of how or why populations change requires detailed information about migrants. Without these, the ability to predict, control or understand that change is limited.

1.1 Motivation

The reasons for internal migration are many. People move for employment, family reunion or amenity reasons. Reported statistics on these flows, on the other hand, are relatively confusing or nonexistent (Bell et al. 2002). There are two main reasons. First, no consensus exists on what exactly is a 'migration'. Second, the event of migration is rarely measured directly. More often it is inferred by a comparison of places of residence at two points in time or as a change in residence recorded by a population registration system. The challenge is compounded because countries use different methods of data collection. Migration statistics may come from administrative data, decennial population censuses or surveys. Harmonization of data collection processes and the data they generate is not even close to being realized. So, how does one overcome these obstacles to obtain an overall and consistent picture of the migration patterns occurring, say, within a specific country? One possibility is to have a methodology for combining existing migration data that accounts for the various strengths of the different data sources.

Inadequate, missing or inconsistent migration data makes analysing the patterns of, for example, Whites and non-Whites, young and elderly, first and second generation immigrants, skilled and unskilled, and employed and unemployed over time very difficult or incomplete. Detailed migration data are usually only available from censuses, which only occur every ten years and are published three to four years after the census date. General purpose surveys often collect migration data but, because of relatively small sample sizes, they are usually inadequate below the national or broad regional levels. Population registers may be used to track migration flows, however, these sources often do not contain much demographic, socioeconomic or spatial detail. Also, because migration data are often collected from sources that have other purposes, the questions underlying the patterns may not fit a particular research question of interest, e.g., measuring migrant status tells us little about migration frequency. There may also be situations in which the required data are available but cannot be considered reliable due to, for example, age misreporting. Missing data is usually caused by suppression of data or by non-response.

In order to include migration data from different sources in a study, one has to first account for the differences in measurement (see Bell et al. 2002; Long and Boertlein 1990; Morrison et al. 2004; Rogers et al. 2003a; Rogerson 1990; United Nations 1992). For example, migration events, which can occur multiple times within a one year time period, are captured by population registration systems while changes in residential status (or transitions) from one point in time to another are captured by censuses (and surveys). These two data collection systems capture two different types of migration data, i.e., 'migrations' and 'migrants' (Rees and Willekens 1986). This work combines these two types of migration data by focusing on the underlying structures, which are similar to each other --- allowing us to produce a synthetic data base of detailed migration patterns that can be used to study, for example, ethnic migration and its relationships with areas of high unemployment, high costs of living or high immigrant concentrations.

1.2 Background

In the United Kingdom, there have been many studies that have examined or modelled internal migration flows (e.g., Bates and Bracken 1982, 1987; Bell and Rees 2006; Champion 1996;

Dixon 2003; Kalogirou 2005; Stillwell 1994). Other studies have examined the determinants of internal migration (Fotheringham et al. 2004) and the description of social change caused by international migration (Dorling and Rees 2003; Rees and Butt 2004), including the linkages between immigration and internal migration (Hatton and Tani 2005; Stillwell and Duke-Williams 2005). These studies have all relied on available data. They have not attempted to combine the various internal migration data sources available in the United Kingdom. Our research does. It allows for both intercensal and post-census estimates of detailed migration flows. These estimates have the possibility to increase our understanding of population redistribution at various levels. The ethnic population in England has grown substantially since 1991. By having internal migration data by ethnicity over time, we can detect when and how certain ethnic groups in the population have become more spread out or more concentrated as a result of internal migration, which can then be compared with other studies that focus on immigration or population change as measured by the decennial censuses (e.g., Rees and Butt 2004). Finally, this study does not seek to determine the factors underlying the migration patterns, such as in Fotheringham et al. (2004) but instead focuses on the development of a model that can effectively incorporate migration data from multiple sources to produce reliable and detailed estimates over time.

2. BUILDING A DATA SET TO STUDY ETHNIC MIGRATION

2.1 Sources of Internal Migration

In England, the most reliable internal migration data come from the decennial censuses and the National Health Service Central Register (NHSCR). The information obtained from the census contains much of the information needed for detailed analyses, but are only collected every ten years and may contain problems of incomparability between censuses for certain variables (see Stillwell and Duke-Williams 2007 for recent discussion). Migration data from the NHSCR 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).

Conceptually, there are several different ways data can be collected on the relocation of persons from one permanent address to another, all of which can yield different counts for the same flow (Rees and Willekens 1986). Registration data captures movements or events of migration. Census data captures migration transitions or changes in residential status. This naturally creates higher counts in the registration data as multiple moves within a one-year period can take place, including return movements. Although the census only captures place of residence at two points in time, Boden et al. (1992) found high levels of correlation between the in-migration, out-migration and net migration totals captured by these two data sources. More recently, Raymer et al. (2007), in analysing migration data in England and Wales, found that the overall levels of elderly internal migration obtained from the 2000-2001 National Health Service (NHS) register were substantially higher than those obtained from the 2001 Census, whereas the underlying marginal structures were very similar. This led them to create a model to combine the two sources of data to estimate post-census flows of elderly migration flows by health status.

In England, a person must register with a local doctor in order to receive services. The NHS maintains two registers: the Central Register, which records moves between health authorities, and the Patient Register (since 2000), which tracks migration between local authorities. Data are periodically transmitted to the Office for National Statistics (ONS) which provides annual estimates on a quarterly base for the Central Register and once a year for the Patient Register. 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 NHS, 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 (ONS Migration Statistics Unit 2002).

2.2 Description of Data Collected

The migration data used in this paper come from censuses carried out in England in 1991 and 2001 and from the NHSCR tables from 1991 to 2007. The 1991 census tables were obtained from the Special Migration Statistics (SMS) dataset called 'SMSGAPS' available on the Centre for Interaction Data Estimation and Research (CIDER) website.[1] The 2001 census data were obtained from the SMS CD-ROM provided by the Office for National Statistics (ONS 2004). The NHSCR Table 1 and Table 2a from 1991 to 2007 were obtained from CD-ROMs provided by ONS. Data differ in their coverage, geographical detail and type of migration flows recorded. A description of census and NHSCR data is set out in Table 1. Note, in this paper, we focus on interregional migration flows in England. Future work will extend the analysis to smaller geographical levels, such as counties and local authority districts.

[1] http://cider.census.ac.uk/cider/.

Table 1. Internal migration data in England: Available tables from the 1991 and 2001 censuses and from the 1991-2007 NHSCR

Source	Tables	Geography	Type of flows
Census 1991	Table 5	Local authority districts	Origin, destination and ethnicity (ODE)
Census 2001	MG103	Local authority districts	Origin, destination, sex and ethnicity (ODSE)
NHSCR 1991-2007	Table 1	Family Health Service Areas (1991-2000) and Health Areas (2001-2007)	Origin, age and sex (OAS) and Destination, age and sex (DAS)
NHSCR 1991-2007	Table 2a	Family Health Service Areas (1991-2000) and Health Areas (2001-2007)	Origin and destination (OD)

The entire data set organised for this study consists of migration flows between the nine Government Office Regions (GOR), sixteen five-year age groups (i.e., 0-4, 5-9, ..., 75+ years), two sexes (i.e., male and female) and four ethnic groups (i.e., White, South Asian,[2] Black[3] and Chinese & Other[4]). The GOR regions consist of the North East, North West, Yorkshire and the Humberland, East Midlands, West Midlands, East of England, South East, South West and London.[5]

2.3 Key Structures

For the purposes of this section and the remainder of the paper, we denote cross-classified tables by letters. For example, an OD table is a two-way origin by destination table of migration flows. Furthermore, age is denoted by A, sex by S and ethnicity by E.

Once we collected the data, the next step was to identify the key structures that we wanted to include from each data source and year. This was undertaken by comparing various

[2] South Asian = Indian, Pakistani and Other South Asian
[3] Black = African, Carribbean and Other Black
[4] Includes Mixed Ethnicity. This classification was not included in the 1991 census (see also Stillwell and Duke-Williams 2007).
[5] A map of the nine regions produced by the Office for National Statistics can be found at: http://www.statistics.gov.uk/geography/downloads/GB_GOR98_A4.pdf.

unsaturated log-linear models (see, e.g., Agresti 2002) of two four-way tables, i.e., ODAS and ODSE, obtained from the 2001 census. The complete five-way table ODASE was not publicly available. However, our assumption is that the relationship between age of migration and ethnicity is relatively weak and therefore not very important in the overall patterns of migration. In the future, the complete ODASE table should be examined to test this assumption.

In our analysis of the unsaturated log-linear models for the ODAS table, we found that the best model, penalised for complexity, is the one that includes interactions between origin and destination (OD table), origin and age (OA table), destination and age (DA table) and age and sex (AS table). The two-way interactions between sex and origin (OS table) and sex and destination (DS table) and three-way interactions between origin, destination and sex (ODS table), origin, age and sex (OAS) and destination, age and sex (DAS table) did not contribute significantly to the overall model fit. The three-way interactions between origin, destination and age (ODA table) improved the overall model fit slightly but not enough to justify it its inclusion considering the large number of parameters required. This supports the work by Raymer et al. (2006) and Raymer and Rogers (2007), who found that the three-way interaction term between origin, destination and age does not affect the overall prediction by much, except in extreme origin-destination-specific flow cases, e.g., with a retirement peak.

The above analyses provide us with some direction on how to proceed with the combining of migration flow data. First, we have ODAS tables from the NHSCR, however, we do not need to include the complete data to produce accurate results. In fact, there are advantages to having a simplified model. The two-way interaction model that includes the OD, OA, DA and AS information, for example, produces estimates that are nearly indistinguishable from the observed values in the complete ODAS table. Also, this model has the advantage of producing smoother estimates, particularly over age. Hence, for the NHS data, we only need information from the OD, OA, DA and AS tables to produce good results. Second, we have ODE tables from the 1991 and 2001 censuses that can be used as auxiliary information. Here, we would like to include all the information contained in these tables, as it is likely that the spatial patterns of migration are different for the four ethnic groups.

2.4 Adjustment of NHSCR data

The algorithm developed in the following section to estimate the migration flows requires consistency in the marginal distributions of the incomplete data, namely of the OD, OA, DA and AS tables, extracted from NHSCR. Unfortunately, the marginal distributions of these tables were not consistent. The reason for this is that the OA, DA and AS tables also included the flows from and to Wales, Scotland and Northern Ireland. For the OD table, this was not a problem, as these flows could be separated out. To force the OA, DA and AS tables to match the OD table, we applied an iterative proportional fitting procedure similar to that used in Raymer et al. (2007), assuming that the age and sex proportions of migration for England are same as those for the nation as a whole.

3. A LOG-LINEAR MODEL FOR COMBINING DATA

3.1 Background

This study draws from a long history of modelling internal migration flows (Cadwallader 1992; Fotheringham et al. 2000; Plane 1981, 1982; Raymer and Rogers 2007; Raymer et al. 2006; Raymer et al. 2007; Rogers et al. 2002, 2003; Stillwell forthcoming; Willekens 1977, 1980, 1982, 1983, 1999). The log-linear model version of the spatial interaction model (Willekens

1980, 1983) is of particular importance. The advantage of the log-linear model over the general spatial interaction model is that it has a well-formed theory and methods, associated in the framework of categorical data analysis (e.g., Agresti 2002) and missing data analysis (e.g., Little and Rubin 2002).

More specifically, the methodology developed for this work is motivated by Raymer et al. (2007), which combined census (auxiliary) and registration (incomplete) data to estimate detailed elderly migration flows in England and Wales. This work was a first attempt at developing detailed estimates of migration flows. The methodology is extended by including more auxiliary information (i.e., from two censuses), by producing an annual time series of estimates from 1991 to 2007 and by increasing the dimensions from three-way tables to five-way tables over time.

3.2 Model Specification

Our objective for this project is to estimate migration flows for an ODASE table for each year from 1991 to 2007. The diagonals of the OD partial tables (i.e., the within-region flows) are excluded. However, the methodology could also be used to estimate these flows if needed. The basic idea is to supplement information from the NHS register with more detailed information from the censuses. The log-linear model developed by Raymer et al. (2007) is used as a starting point. This model combines one-way marginal information available in the incomplete registration data with complete (but outdated) census data. In essence, the association structure of the census (auxiliary) data is imposed on the registration (incomplete) data.

Spatial interaction models are commonly used to model origin-destination-specific migration flow data. Overviews of these models and frameworks can be found in Fotheringham et al. (2000:211-235), Stillwell (forthcoming) and Willekens (1983, 1999). A simplistic version of the spatial interaction model to estimate the number of migrations, n_{ij}^{OD}, in an incomplete data set, from origin *i* to destination *j* during a unit interval may be applied as in Willekens (1999):

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

where μ_{ij}^{OD} is the expected number of migration flows from origin *i* to destination *j* during the respective time interval and $i = 1,2,\ldots,R; j = 1,2,\ldots,R$ 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 (NHS) migration data. For example, an origin by destination by ethnicity table, with counts n_{ijz}^{ODE} 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.

If information on two-way or higher associations exists in the incomplete data, the model can be extended to include this. Furthermore, we may not wish to impose the higher-way interactions from the auxiliary data. For example, as discussed in Section 2.2, we wish to use the OD, OA, DA and AS tables from the NHS data and impose the three-way associations from the ODE census table. This is achieved by using the following log-linear model for n_{ijxyz}^{ODASE}, the counts in the five-way ODASE table:

$$\log \mu_{ijxyz}^{ODASE} = \lambda_{ij}^{OD} + \lambda_{ix}^{OA} + \lambda_{jx}^{DA} + \lambda_{xy}^{AS} + \log m_{ijz}^{ODE} . \qquad (3)$$

Should a different model for the flows be thought appropriate, then Model (3) can be modified by adding or removing interaction parameters, or by changing the offset term, provided the pertinent information is available in the incomplete or auxiliary data, respectively.

Models (1) to (3) can be fitted using maximum likelihood estimation. It is straightforward to derive and solve, using an iterative procedure, the likelihood equations for these models to obtain estimates of the λ-parameters and flows. Raymer et al. (2007) did this for Models (1) and (2). However, since our interest is primarily in the estimation of the flows, we just apply an iterative proportional fitting algorithm to obtain them instead (see, e.g., Agresti 2002, Section 8.7.2). The initial values are given by the counts in the ODE table from the census: $\mu_{ijxyz}^{ODASE(0)} = m_{ijz}^{ODE}$ for all *x* and *y*. They are then successively multiplied by adjustment factors so that the marginal tables match the counts in the NHSCR OD table, then the NHSCR OA table, then the NHSCR DA table and finally the NHSCR AS table. This is repeated until the marginal tables of estimated flows simultaneously match all of the counts contained in the four NHSCR tables.

Finally, Raymer et al. (2007) assumed the three-way auxiliary interaction structure remained constant over time. We, on the other hand, allow this structure to vary over time from 1991 to 2007. We do this by geometrically interpolating the counts from 1992 to 2000 and by geometrically extrapolating from 2002 to 2007. The 1991 and 2001 census values are used as benchmarks. Model (3) is then run for each year with these auxiliary structures used as offsets.

Once the models were run, we then checked the results for their reasonableness. In doing so, we identified an important problem with the NHS data relating to the age structure of migration by sex (i.e., the AS table). Here, it was found that females had higher levels of migration (52.3 percent on average) than males (47.7 percent on average), with the gap between the two sexes slightly widening over time (Figure 1). The corresponding age patterns obtained from the 1991 and 2001 censuses, however, showed a different pattern with males representing 50.8 percent in both years. The reason for this difference has primarily to do with males being less likely to register with the NHS register, particularly in young adult years (see Fotheringham et al. 2004:1637-1640 for discussion). In Raymer et al. (2007), this was not an issue as they only examined migration patterns of elderly persons, a group less likely to be missed in a health service population register. These differences are illustrated in Figure 2 for 1991 and 2007. The corresponding patterns reported by the 1991 and 2001 censuses are very different in that the age-sex patterns are nearly identical, except in the last age group of 75+ years, where females have higher levels of migration (associated with their higher population numbers in these years).

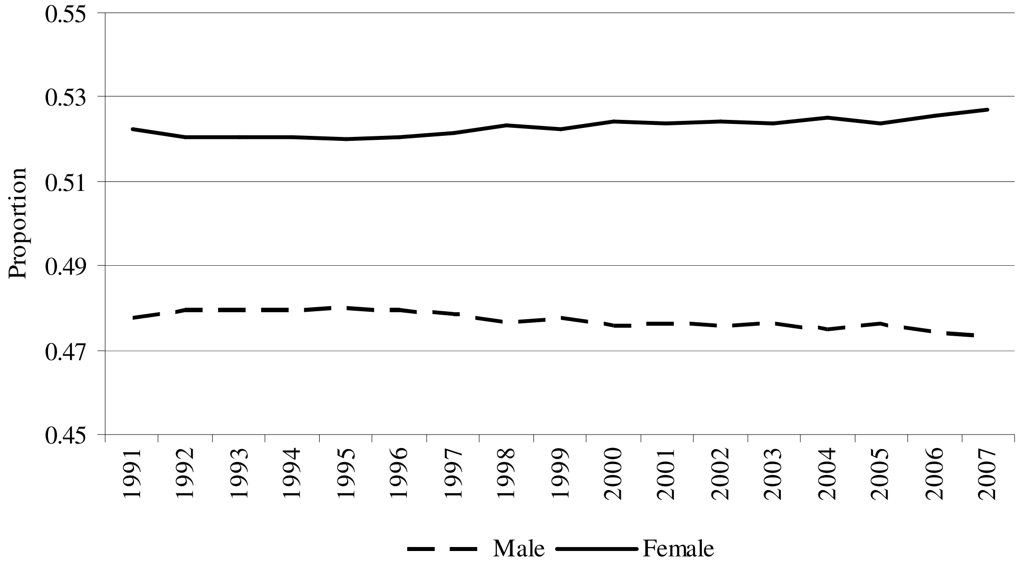


Figure 1. Interregional migration in England by sex, 1991-2007

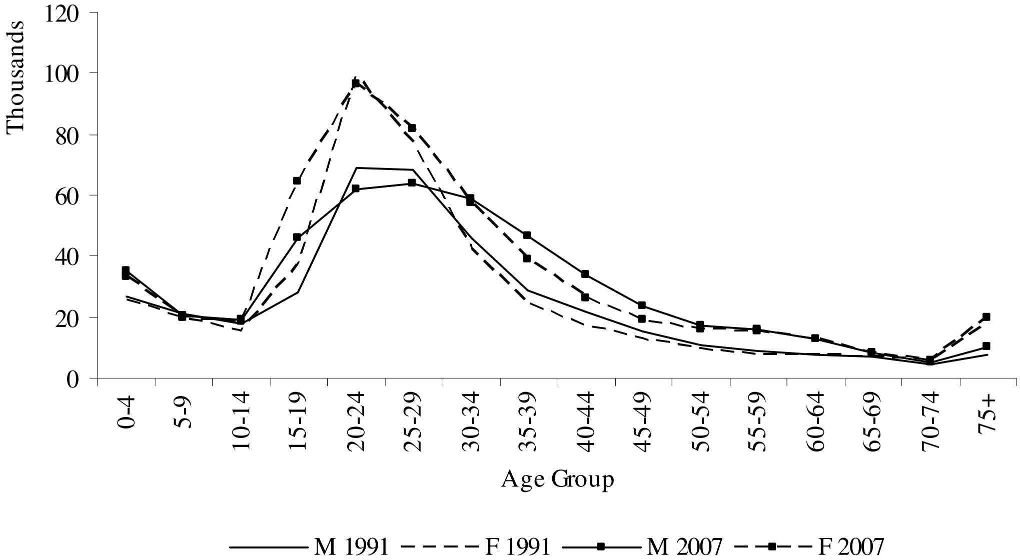


Figure 2. Age patterns of NHS interregional migration in England by sex, 1991 and 2007

As illustrated in Figure 2, nearly all the differences in the age patterns of male and female migration as reported in the NHS data occur in the 15-19 year, 20-24 year and 25-29 year age groups.[6] To correct for the differences in the age-sex patterns, there are two options. The first is to impose the interactions contained in the census AS table instead of the NHS table. Here, Model (3) can be rewritten as follows:

$$\log \mu_{ijxyz}^{ODASE} = \lambda_{ij}^{OD} + \lambda_{ix}^{OA} + \lambda_{jx}^{DA} + \log\left(m_{ijz}^{ODE} m_{xy}^{AS}\right). \qquad (4)$$

This model maintains all of the above associations but with the age-sex structure from the censuses. The problem with this model, however, is that it does not correct for the undercounting of males. The overall levels of migration would remain the same, which means that the levels of female migration would have to be lowered to make the age-sex differences correspond with the census patterns. We assume that females are counted accurately in the NHS data.

The second option is to weight the estimates from Model (3) 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 ethnicity. This approach maintains all of the associations implied by Model (3). The weights applied to the male migrants in the three age groups are set out in Table 2, along with the resulting adjustment ratios for the all males (12 to 15 percent increase) and males plus females (6 to 7 percent increase).

[6] Fotheringham et al. (2004) also found differences across spatial units. Our analysis at the regional level did not find differences to be significant.

Table 2. Adjustment ratios for NHSCR migration data, 1991-2007

	Age Group				Both
Year	15-19	20-24	25-29	Males	Sexes
1991	1.315	1.436	1.131	1.123	1.059
1992	1.287	1.411	1.128	1.117	1.056
1993	1.300	1.409	1.131	1.120	1.058
1994	1.294	1.398	1.136	1.116	1.056
1995	1.322	1.385	1.121	1.117	1.056
1996	1.334	1.403	1.130	1.122	1.058
1997	1.339	1.408	1.136	1.119	1.057
1998	1.348	1.441	1.156	1.125	1.059
1999	1.344	1.433	1.157	1.122	1.058
2000	1.343	1.454	1.176	1.127	1.060
2001	1.388	1.473	1.161	1.131	1.062
2002	1.393	1.482	1.177	1.130	1.062
2003	1.399	1.487	1.188	1.132	1.063
2004	1.400	1.514	1.217	1.137	1.065
2005	1.390	1.490	1.223	1.138	1.066
2006	1.399	1.523	1.255	1.145	1.069
2007	1.400	1.553	1.281	1.148	1.070

4. PATTERNS OF ETHNIC MIGRATION

In this section, the estimated interregional migration flows by age, sex and ethnicity are presented. These flows represent the re-weighted estimates from Model (3) discussed in the previous section. First, we describe the patterns over time and across space and then by age and sex.

4.1 Over Time

The overall levels of interregional migration by ethnic group and over time are set out in Figure 3. Here, we see that the total levels of migration have only increased slightly from just less than 900 thousand to around one million persons per year and that the vast majority of the flows are comprised of Whites. In 1991, Whites represented about 94 percent of the flows. In 2007, the estimated share is about 85 percent. The increasing levels of South Asian, Black and Chinese &

Other migration are more clearly visible in Figure 4. Here, we see that the flows of all three groups increased substantially over time, from around 48 thousand in 1991 to around 156 thousand in 2007. The relative shares of non-White ethnic migration remained pretty much the same over time with South Asians representing around 45 percent, Blacks around 22 percent and Chinese & Other around 33 percent.

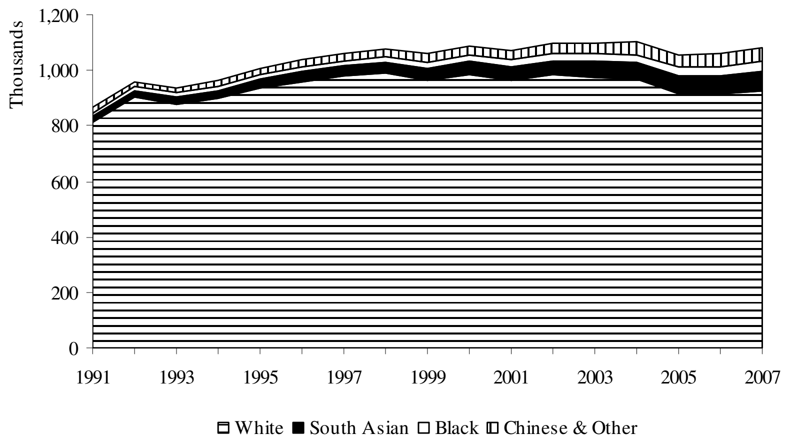


Figure 3. The levels of interregional migration in England by ethnicity, 1991-2007

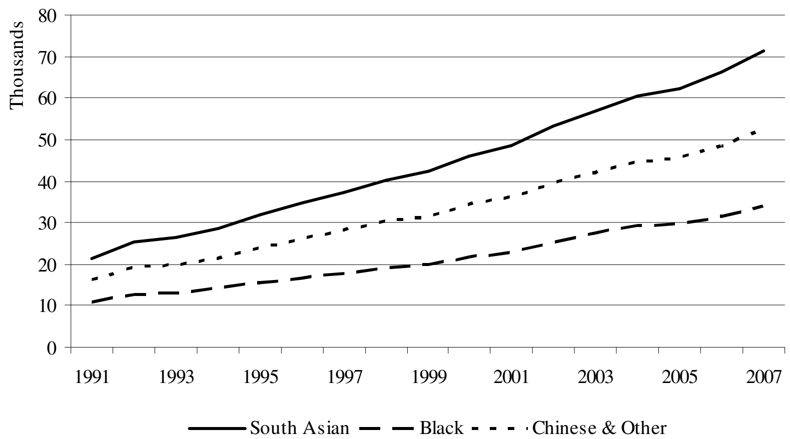


Figure 4. The levels of South Asian, Black and Chinese & Other interregional migration in England, 1991-2007

4.2 Spatial Patterns

Two examples of origin-destination-specific flows are set out to illustrate the differences between White, South Asian, Black and Chinese & Other migration. These represent migration flows from London (Figure 5) and from the South East (Figure 6), the two largest sources of interregional migration. (Note, the y-axis scales are different for White migrants.) For migration from London (Figure 5), the top two destinations for all ethnic groups are the South East and East of England, for which the levels have been increasing steadily over time. Interestingly, Black migrants have the same migration levels going to both regions, whereas for the other ethnic groups, the South East is the preferred destination. Larger differences in the migration patterns appear when the third choice of destination is considered. For Whites, the South West comes third in terms of destination choice, whereas it is West Midlands for South Asians and Blacks. There is not much difference in remaining destination choices for the Chinese & Other ethnic group. For migration from South East (Figure 6), the top destination for all three non-White ethnic groups is London. For White migration, the patterns are more spread out and relatively level over time. Here, the top three destinations are London, South West and East of England.

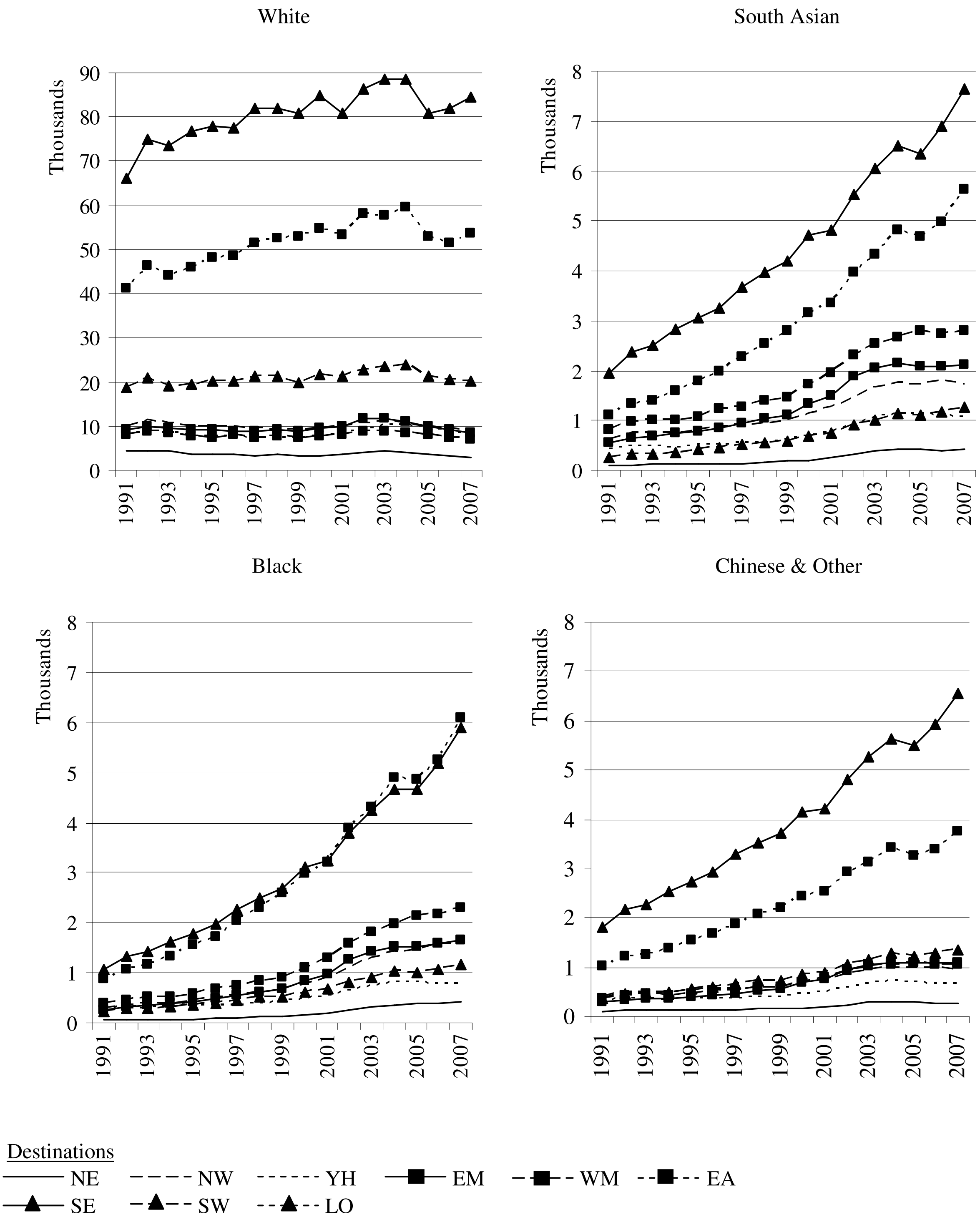


Figure 5. Interregional migration from London by ethnicity, 1991-2007

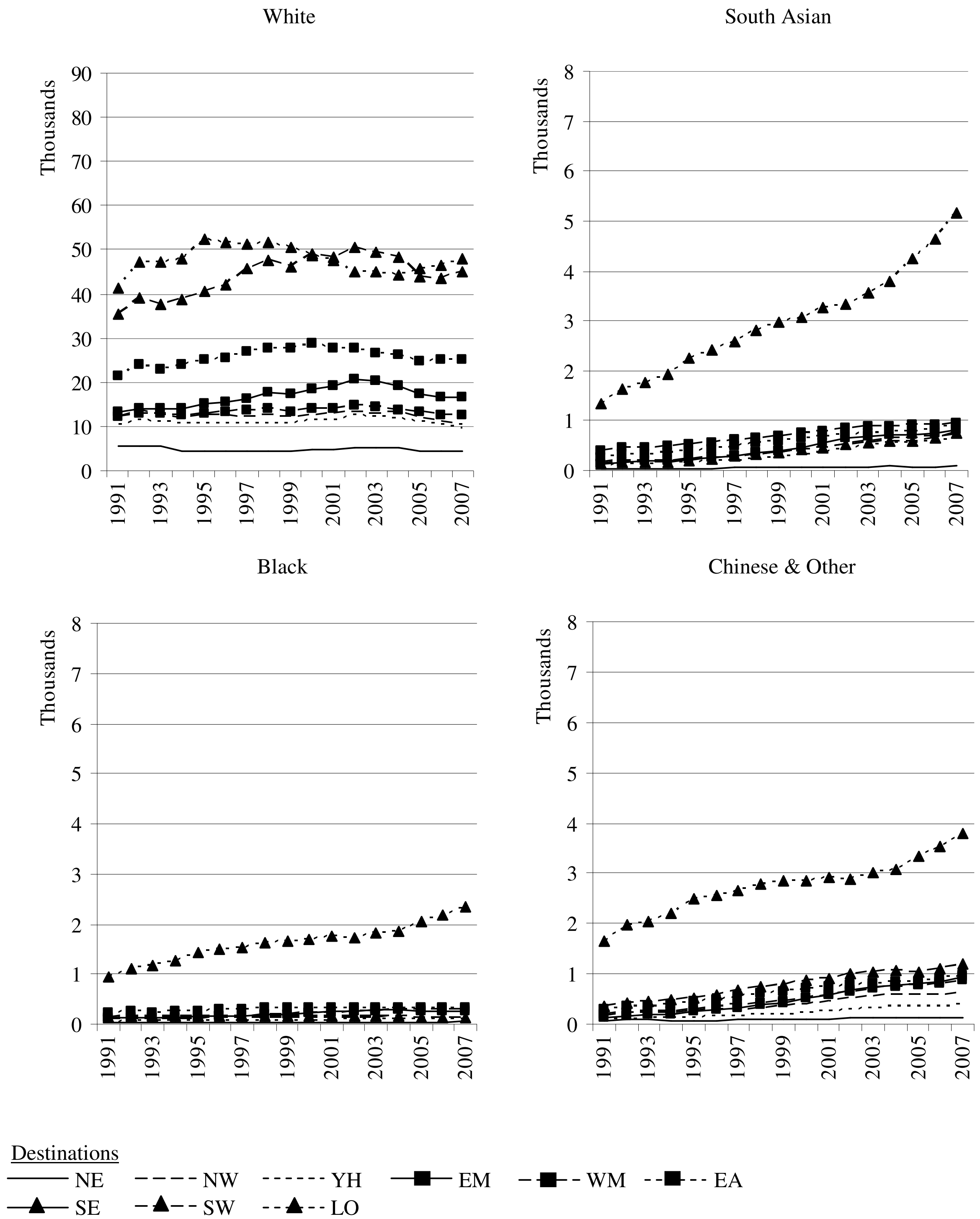


Figure 6. Interregional migration from South East by ethnicity, 1991-2007

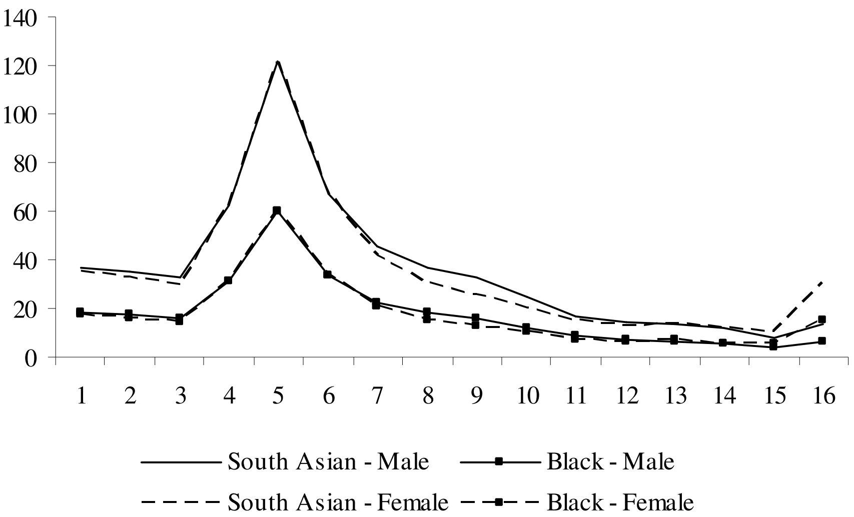
4.2 By Age and Sex

Next, consider the estimated age- and sex-specific interregional migration flows. For illustration purposes, we first compare the 1991 differences of South Asian and Black migration between London and West Midlands (Figure 7), London and South West (Figure 8) and London and South East (Figure 9) flows. Second, we compare the age patterns of female South Asian migration over time (i.e., 1991, 1999 and 2007) between London and West Midlands (Figure 10) and London and South East (Figure 11).

In Figures 7, 8 and 9, we see that the adjustment factors have resulted in very similar age patterns for males and females. The only major difference exists in the last age group, where females are known to contain a much larger share of the population. Also, by design, the age patterns of all ethnic groups have the same origin-destination-specific shapes. The three figures show (1) how differences in levels are represented (i.e., greatest in the young adult age groups, Figure 7), (2) that the regularities are maintained even for very small flows (Figure 8) and (3) the different shapes than can emerge for different flows (i.e., narrow labour force peaks in Figure 7 versus more wide labour force peaks in Figure 9).

The increasing levels of migration by age for South Asian female migration between London and West Midlands and London and South East are illustrated in Figure 10 and Figure 11. In both cases, the levels of migration have increased over time. In Figure 10, we see that the shape of the labour force peak has changed over time by attracting relatively more 15-19 year olds, whereas the shape has remained relatively constant over time in Figure 11.

A. West Midlands to London



B. London to West Midlands

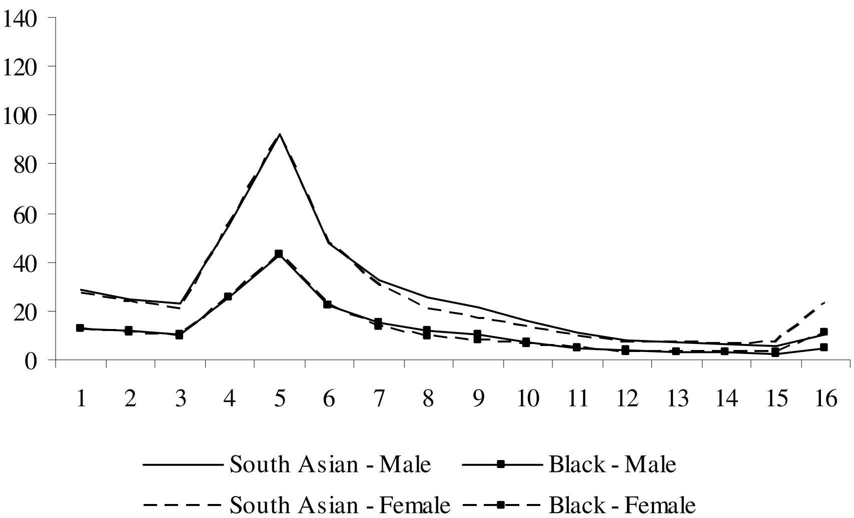
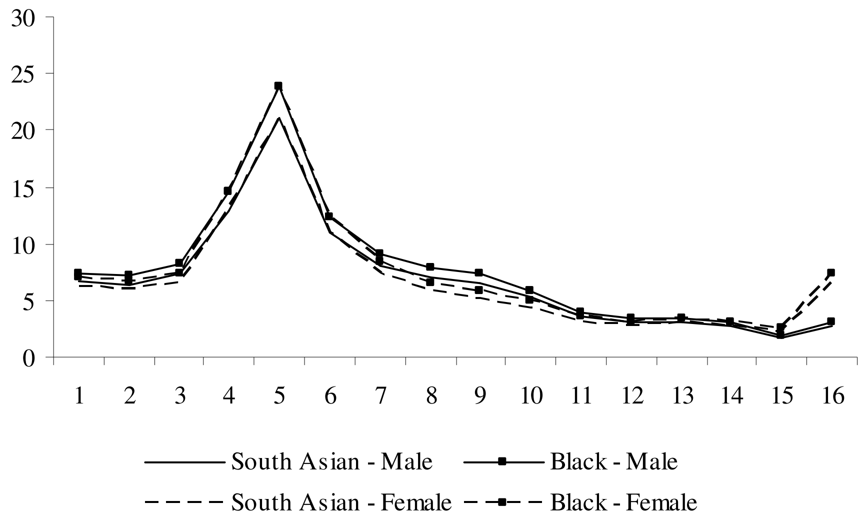


Figure 7. Age- and sex-specific migration of South Asians and Blacks between London and West Midlands, 1991

A. South West to London



B. London to South West

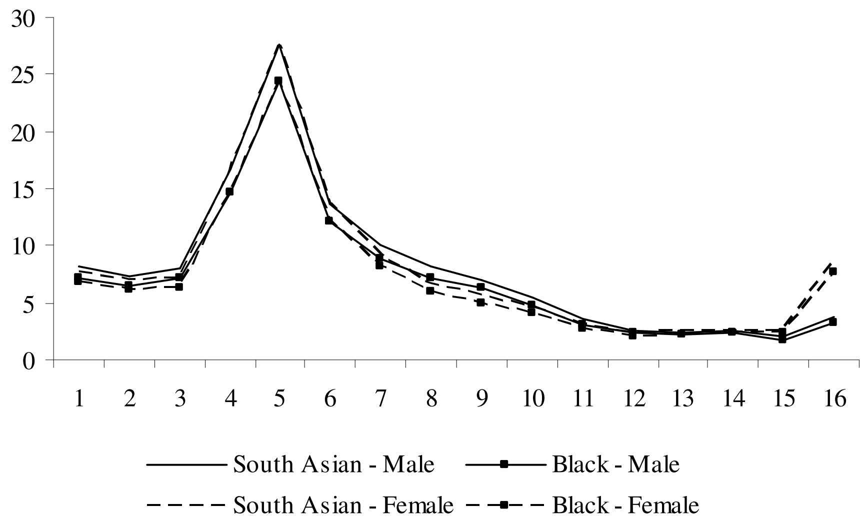
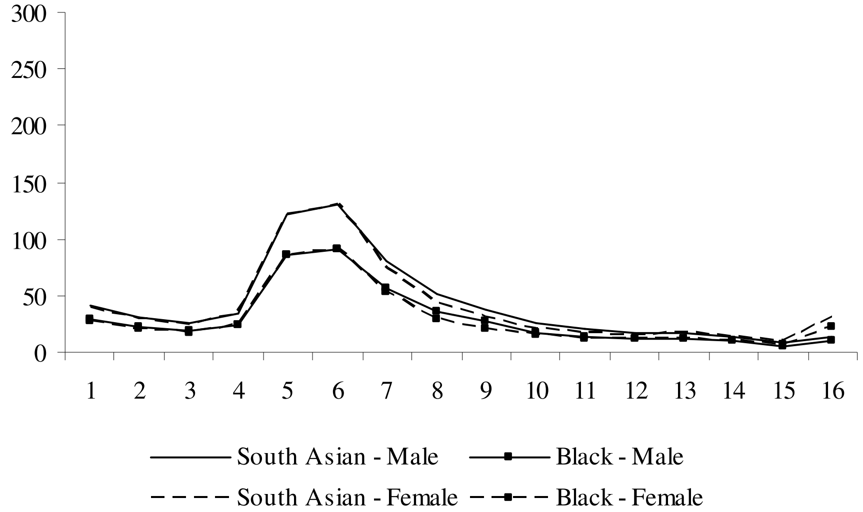


Figure 8. Age- and sex-specific migration of South Asians and Blacks between London and South West, 1991

A. South East to London



B. London to South East

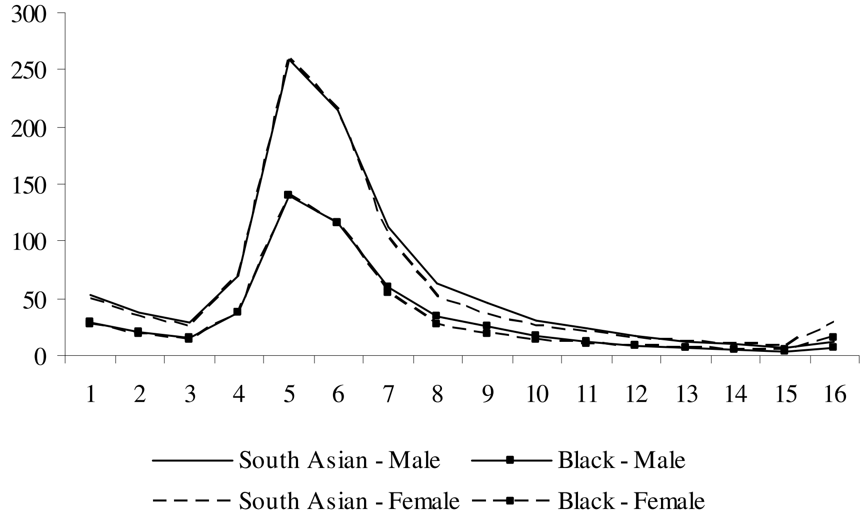
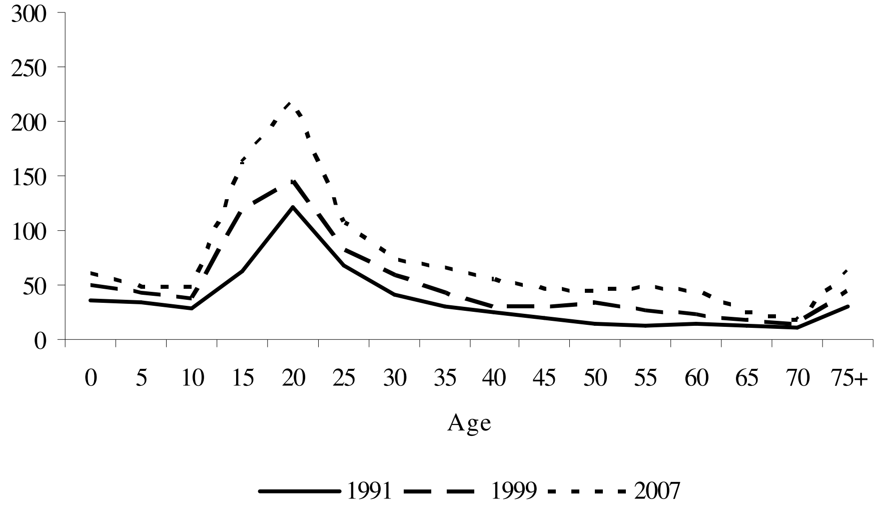


Figure 9. Age- and sex-specific migration of South Asians and Blacks between London and South East, 1991

A. West Midlands to London



B. London to West Midlands

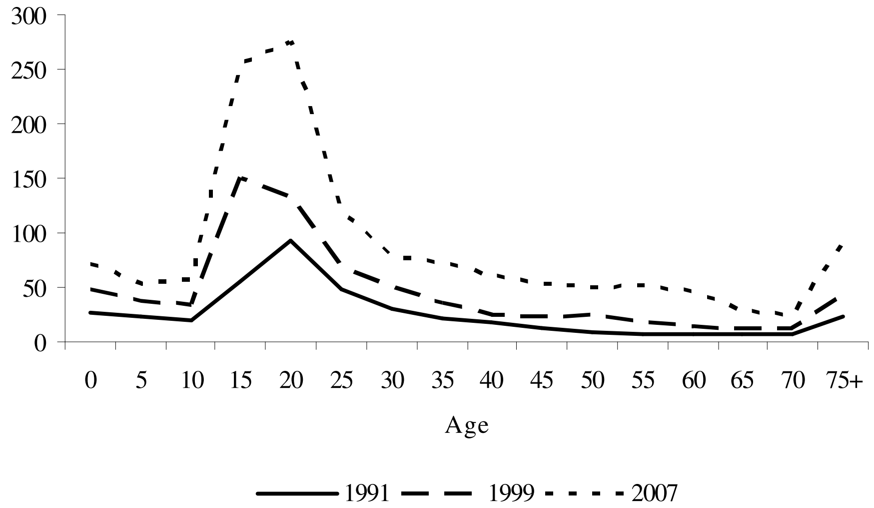
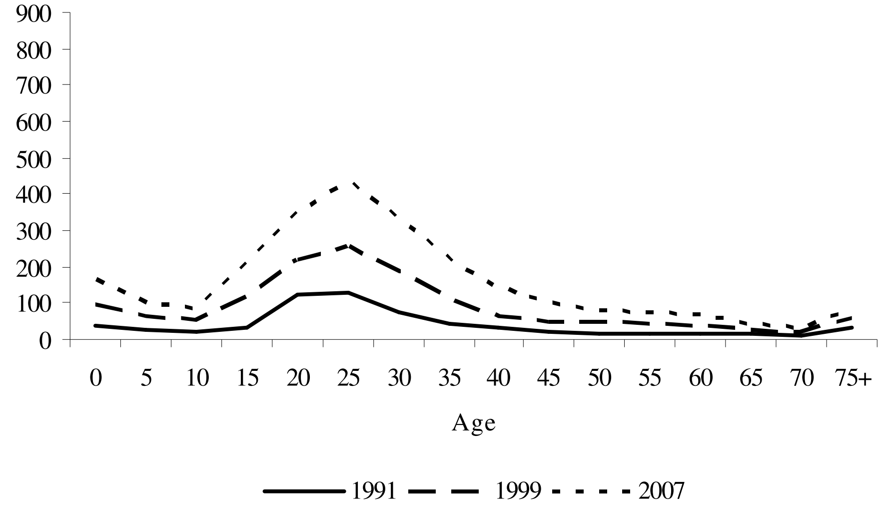


Figure 10. Age-specific migration of female South Asians between London and West Midlands: 1991, 1999 and 2007

A. South East to London



B. London to South East

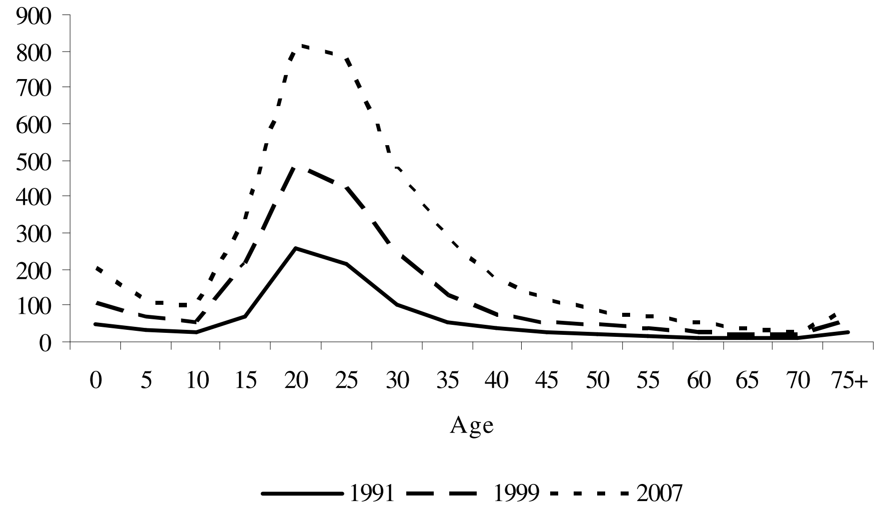


Figure 11. Age-specific migration of female South Asians between London and South East: 1991, 1999 and 2007

5. CONCLUSION

The above analysis has demonstrated the results of the estimated time series of ethnic interregional migration flows in England by age and sex. The differences in the levels and spatial patterns are driven by the changes in the ethnic migration patterns found in the 1991 and 2001 censuses. We assumed that age and sex patterns were similar across ethnic groups. The results show that the patterns of migration differ substantially across ethnic groups with the share of White migration decreasing substantially from 94 percent in 1991 to 85 percent in 2007. Although flows of non-White ethnic groups represent a relatively small share of the total level, their increase over time has been relatively fast. Finally, the shapes of the age patterns of migration and the sex ratios of migration have not changed much over time.

In conclusion, a model and framework for combining migration data from registration systems with more detailed information obtained from censuses has been presented. The outcome is a synthetic database that can be used for analysing the current or future evolution of ethnic-specific migration patterns. Furthermore, the combining data framework could be readily extended to a more general context, for example, to other migration groups of interest (e.g., education or economic activity) or to more categories in the analysis (e.g., regions or ethnic groups). Future work includes extending the application of the model developed in this paper to a more detailed level of spatial units (i.e., counties and local authority districts), the investigation of the missing age-ethnicity information and including auxiliary information from a third source, such as survey data.

REFERENCES

Agresti A (2002) *Categorical data analysis*. Hoboken: Wiley-Interscience.

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 PH, Stillwell J and Hugo G (2002) Cross-national comparison of internal migration: Issues and measures. *Journal of the Royal Statistical Society Series A* 165(3):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* (ed. J. Stillwell, P. Rees, and P. Boden), pp. 13-27. London: Belhaven Press.

Cadwallader M (1992) *Migration and residential mobility: Macro and micro approaches*. Madison: The University of Wisconsin Press.

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.

Dorling D and Rees P (2003) A nation still dividing: The British census and social polarisation 1971-2001. *Environment and Planning A* 35:1287-1313.

Fotheringham AS, Champion T, Wymer C and Coombes M (2000) Measuring destination attractivity: A migration example. *International Journal of Population Geography* 6:391-421.

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

Hatton TJ and Tani M (2005) Immigration and inter-regional mobility in the U.K., 1982-2000. *The Economic Journal* 115(November):F342-F358.

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

Little RJA, Rubin DB (2002) *Statistical analysis with missing data*, 2nd edition. Hoboken: John Wiley & Sons.

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

Morrison PA, Bryan TM and Swanson DA (2004) Internal migration and short-distance mobility. In *The methods and materials of demography*, J. S. Siegel and D. A. Swanson, eds., pp. 493-521. San Diego: Elsevier Academic Press.

ONS (2004) 2001 Census origin-destination statistics: Local Authorities (CD-ROM). Office of National Statistics.

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

Plane DA (1981) Estimation of place-to-place migration flows from net migration totals: A minimum information approach. *International Regional Science Review* 6(1):33-51.

Plane DA (1982) An information theoretic approach to the estimation of migration flows. *Journal of Regional Science* 22(4):441-56.

Raymer J, Abel G and Smith PWF (2007) Combining census and registration data to estimate detailed elderly migration flows in England and Wales. *Journal of the Royal Statistical Society A* 170(4).

Raymer J, Bonaguidi A and Valentini A (2006) Describing and projecting the age and spatial structures of interregional migration in Italy. *Population, Space and Place* 12:371-388.

Raymer J and Rogers A (2007) Using age and spatial flow structures in the indirect estimation of migration streams. *Demography* 44(2):199-223.

Rees P and Butt F (2004) Ethnic change and diversity in England, 1981-2001. *Area* 36(2):174-186.

Rees P and Willekens F (1986) Data and accounts. In *Migration and settlement: A multiregional comparative study*, Rogers and Willekens, eds., pp. 19-58. Dordrecht: D. Reidel Publishing Company.

Rogers A, Raymer J and Newbold KB (2003a) Reconciling and translating migration data collected over time intervals of differing widths. *The Annals of Regional Science* 37(4):581-601.

Rogers A, Willekens FJ and Raymer J (2002) Capturing the age and spatial structures of migration. *Environment and Planning A* 34:341-359.

Rogers A, Willekens FJ and Raymer J (2003b) Imposing age and spatial structures on inadequate migration-flow datasets. *The Professional Geographer* 55(1):56-69.

Rogerson PA (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(11):1711-1730.

Stillwell J (forthcoming) Inter-regional migration modelling: A review. In *Migration and human capital: Regional and global perspectives*, Poot J, Waldorf B and van Wissen L, eds., Chapter 2. Cheltenham: Edward Elgar.

Stillwell J and Duke-Williams O (2005) Ethnic population distribution, immigration and internal migration in Britain: What evidence of linkage at the district scale? Paper presented at the British Society for Population Studies Annual Conference, University of Kent, Canterbury, 12-14 September.

Stillwell J and Duke-Williams O (2007) Understanding the 2001 UK census migration and commuting data: the effect of small cell adjustment and problems of comparison with 1991. *Journal of the Royal Statistical Society Series A* (Statistics in Society) 170(2):1-21.

United Nations (1992) Preparing migration data for subnational population projections. Department of International Economic and Social Affairs, New York.

Willekens FJ (1977) The recovery of detailed migration patterns from aggregate data: An entropy maximizing approach. RM-77-58, International Institute for Applied Systems Analysis, Laxenburg, Austria.

Willekens FJ (1980) Entropy, multiproportional adjustment and the analysis of contingency tables. *Systemi Urbani* 2/3:171-201.

Willekens FJ (1982) Multidimensional population analysis with incomplete data. In *Multidimensional mathematical demography*, Land K and Rogers A, eds., pp. 43-111. New York, Academic Press.

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

Willekens FJ (1999) Modeling approaches to the indirect estimation of migration flows: From entropy to EM. *Mathematical Population Studies* 7(3):239-278.