# Chapter 15
# Analysing Structures of Interregional Migration in England

**James Raymer**
*University of Southampton, UK*

**Corrado Giulietti**
*University of Southampton, UK*

## ABSTRACT

*In this chapter, we explore the age and ethnic structures of interregional migration in England, as measured by the 1991 and 2001 Censuses. In doing so, we first analyse the main effect and two-way interaction components of migration flow tables cross-classified by (1) origin, destination and age and (2) origin, destination and ethnicity. Second, we test the significance of three-way interaction terms over time by comparing various unsaturated log-linear model fits. The aim is to identify the key structures in the migration flow tables and how they have changed over time. This is important for understanding the mechanisms underlying the more general patterns of migration. These analyses could also be used to inform the estimation or projection of migration flows. Our findings are that, despite a large increase in the levels of interregional migration, migration structures in England have remained fairly stable over time. The main changes have to do with the increases in the relative levels of ethnic migration over time, which has been unequal across space.*

## INTRODUCTION

A methodology for identifying key structures in migration flow tables is presented in this chapter. The tables contain flows of persons cross-classified by origin, destination and some other characteristic, such as age or ethnicity. This work is important for both understanding aggregate migration patterns in general and for estimating flows in the context of incomplete migration data. The research contained in this chapter represents an important first step in larger project funded by the Economic and Social Research Council on combining migration data in England and Wales.

In order to combine data, we need to understand the comparability between different data sources, as well as the underlying structures that drive the migration patterns. For example, is an increase in the migration from origin $i$ to destination $j$ caused by an overall increase in the level of mobility within a country, a relative increase in the numbers leaving the origin, a relative increase in the attractiveness of the destination, an increase in the connectivity between the two, or some combination of any or all of these factors (or 'structures')? The approach taken in this study follows from Baydar (1984) and Willekens and Baydar (1986) and, more recently, from Raymer *et al.* (2006) and Raymer and Rogers (2007), who disaggregated migration flows into multiplicative components for analysis and estimation.

DOI: 10.4018/978-1-61520-755-8.ch015





We begin the chapter with a description of a general modelling framework for describing and analyzing various structures of interregional migration flows. The approach decomposes observed flows of migration into their multiplicative structures. Two applications follow. First, we illustrate the multiplicative decomposition of the 1991 and 2001 age-specific and ethnic-specific interregional migration flows in England. Second, we identify the key structures contained in the observed 1991 and 2001 age-specific and ethnic-specific interregional migration flows by fitting several unsaturated log-linear models to them. The chapter ends with a summary of key findings and a discussion of how these results may be used.

## ANALYSING MIGRATION STRUCTURES

A sequence of recent papers have set out an analytical framework for describing and estimating the age and spatial structures of internal migration, represented by a multiplicative log-linear model (Raymer *et al.*, 2006; Raymer and Rogers, 2007; Rogers *et al.*, 2002; Rogers *et al.*, 2001; 2002b; 2003). This chapter adds to that research by in two ways. First, the structures of migration are examined for interregional migration in England over time to identify both continuity and change. Second, the analytical framework is extended to include migration by age and ethnicity, providing a better understanding of population change and redistribution. Studies of age and ethnic differences in the migration patterns of England have been explored in previous studies (see, for example, Bates and Bracken, 1982; 1987, for age-specific migration and Finney and Simpson, 2008; Stillwell and Duke-Williams, 2005, for ethnic-specific migration patterns). We extend this work by analysing the underlying (multiplicative) structures of these patterns.

### Multiplicative Component Framework

Two-way tables of interregional migration flows (i.e., origin by destination) can be disaggregated into four separate structures or components (Rogers *et al.*, 2002): an *overall* component representing the level of migration, an *origin* component representing the relative 'pushes' from each region, a *destination* component representing the relative 'pulls' to each region, and a two-way *origin-destination interaction* component representing the impacts of physical or social distance between places not explained by the level or main effects of origin and destination. This breakdown is multiplicative, such that

$$n_{ij} = (T)(O_i)(D_j)(OD_{ij}) \qquad (1)$$

where $n_{ij}$ is an observed flow of migration from region $i$ to region $j$, $T$ is the total number of migrants (i.e., $n_{++}$), $O_i$ is the proportion of all migrants leaving from region $i$ (i.e., $n_{i+} / n_{++}$), and $D_j$ is the proportion of all migrants moving to region $j$ (i.e., $n_{+j} / n_{++}$). The interaction component $OD_{ij}$ is defined as $n_{ij} / \left[(T)(O_i)(D_j)\right]$ or the ratio of observed migration to expected migration (for the case of no interaction). This general type of model is called a *multiplicative component model.*

Next, consider three-way tables of interregional migration (i.e., origin by destination by some other variable). The multiplicative component model for this general case is specified as:

$$n_{ijk} = (T)(O_i)(D_j)(X_k)(OD_{ij})(OX_{ik})(DX_{jk})(ODX_{ijk}) \quad (2)$$

where $X_k$ is the proportion of all migrants in group $k$ of variable $X$, such as males and females if $X$ is sex and different age groups if $X$ is age. This model is more complicated because there are now three two-way interaction components and a single three-way interaction component between the origin, destination, and the variable $X$. However, the interpretations of the parameters remain relatively simple and follow the same format as presented for the two-way table (e.g., Raymer *et al.*, 2006).

## Log-Linear Models

As outlined in Raymer and Rogers (2007), the multiplicative component descriptive model set out in equation 2 can be expressed as a *saturated* log-linear statistical model:

$$\ln(n_{ijk}) = \lambda + \lambda_i^O + \lambda_j^D + \lambda_k^X + \lambda_{ij}^{OD} + \lambda_{ik}^{OX} + \lambda_{jk}^{DX} + \lambda_{ijk}^{ODX} \quad (3)$$

where the $\lambda$ s are simply the natural logarithms of the variables appearing in equation 2. In multiplicative form, this model is expressed as:

$$n_{ijk} = \tau\tau_i^O\tau_j^D\tau_k^X\tau_{ij}^{OD}\tau_{ik}^{OX}\tau_{jk}^{DX}\tau_{ijk}^{ODX} \quad (4)$$

where the $\tau$ s denote the model's multiplicative parameters or 'effects'. We use this form to be consistent with the multiplicative component model. The saturated model is expressed as ($ODX$), using the notation set out in Agresti (2002, p. 320). The parameters of the log-linear model can be analyzed using standard statistical techniques for categorical data analysis to identify key structures in the data. Simplified versions of the model set out in equations 3 and 4 are called *unsaturated* models. For example, the model that only includes the *main effects* of origin, destination, and variable $X$ is specified as

$$\hat{n}_{ijk} = \tau\tau_i^O\tau_j^D\tau_k^X \quad (5)$$

and is known as the *mutual independence model* (Agresti, 2002 p. 318). This model assumes independence between each of the categories of origin, destination, and variable $X$ and is designated ($O, D, X$). A model that includes the interaction between origin and destination plus all of the main effects is designated as ($OD, X$) and corresponds to the following model:

$$\hat{n}_{ijk} = \tau\tau_i^O\tau_j^D\tau_k^X\tau_{ij}^{OD} \quad (6)$$

Such notations are used because these models are hierarchical, that is, for two-way interactions, the main effect parameters must be included, and for three-way interaction terms all the main effects and two-way interactions must be included. In this chapter, we exclude non-migrants or intraregional migrants by incorporating structural zeros (Willekens, 1983).

Various unsaturated models in this chapter are evaluated by using the likelihood-ratio statistic ($G^2$) and Akaike Information Criteria ($AIC$). $G^2$ is defined as:

$$G^2 = 2\sum n_{ijk}\ln(n_{ijk} / \hat{n}_{ijk}) \quad (7)$$

where $\hat{n}_{ijk}$ denotes the predicted age-specific migration flows, and where values of $G^2$ closest to zero are associated with 'good' fits. Furthermore, we adjust for model complexity by dividing $G^2$ by the residual degrees of freedom, obtained by subtracting the number of parameters used in the estimation from the total number of observations. This basically penalises the model for complexity.





$AIC$ is defined as:

$$AIC = -2l + 2p \quad (8)$$

where $l$ is the log-likelihood for the Poisson distribution

$$l = \sum\left(n_{ijk}\ln(\hat{n}_{ijk}) - \hat{n}_{ijk} - \ln(n_{ijk}!)\right) \quad (9)$$

and $p$ denotes the number of parameters. Similar to the correction applied to the $G^2$ values, the $AIC$ penalises for model complexity.

## ENGLAND'S MIGRATION FLOW DATA

### Description of Data

The migration data used in this chapter come from censuses carried out in England in 1991 and 2001. The 1991 tables were obtained from the Special Migration Statistics (SMS) dataset called SMS-GAPS available on the Centre For Interaction Data Estimation And Research (CIDER) website (http://cider.census.ac.uk/cider). For 2001, the data were obtained from the SMS CD-ROM provided by the Office for National Statistics (ONS, 2004). Both data sets contained origin-destination-age and origin-destination-ethnicity tables (but not origin-destination-age-ethnicity tables) at the local authority level.

The geography of England changed from 'districts' to 'local authorities and unitary authority districts' between 1991 and 2001. To make the data consistent over time, we used the 'CIDS 1991/2001 common geography' filter for Great Britain available on the CIDER website. The harmonised to 2001 geography consists of 349 common local authority districts. These data were then aggregated to Government Office Regions (GOR) consisting of the North East, North West, Yorkshire and the Humberside, East Midlands, West Midlands, East of England, South East, South West and London. For the purposes of this paper, migration flows within regions are excluded. The age variable contains sixteen five-year age groups, ranging from 0-4 years to 75+ years. The ethnicity variable contains four ethnic groups, as classified in the 1991 data: White, South Asian, Black and Chinese and Other.

### Comparability of 1991 and 2001 Censuses

The comparison of the 1991 and 2001 Census migration data is not straightforward because of change in the definitions of ethnicity and usual residence of students. Compared to 2001, there is less ethnic detail available in 1991 migration flows. In 1991, only White, Black, Indian, Pakistani and Bangladeshi (which for our purposes we have renamed 'South Asian') and Chinese and Other flows were made available. In 2001, White, Indian, Pakistani and other South Asian, Chinese, Black Caribbean, Black African and any other Black background, Mixed and Other ethnic groups were made available. This limited our analyses to the 1991 ethnic group classifications. The 2001 ethnic groups were aggregated as follows: White and Black remain the same; Indian, Pakistani and other South Asian represent South Asian; Chinese, Mixed and other represent Chinese and Other. Note, the main inconsistency with this aggregation is with the Mixed ethnic group. In 1991, the Mixed ethnic population may exist in any ethnic group category, whereas in 2001, this group has been merged with Chinese and Other. Thus, the comparison of migration over time for this group is likely to be overstated. Although there are techniques to compare 1991 and 2001 Census populations – which for example redistribute the Mixed population across other groups (see Rees, 2004) – these seems to be inappropriate in the case of migration data, in that migrants are a non-random subset of the total population. The second issue concerns the treatment of student migration. In 1991, the addresses of students corresponded with their parents' usual residence, while in 2001, the addresses were located with their place of study. The effect of this is likely to inflate the number of 2001 student-aged migrants relative to 1991 (see Stillwell and Duke Williams, 2007 for analysis and discussion).

## INTERREGIONAL MIGRATION: ANALYSIS OF MAIN EFFECTS AND TWO WAY INTERACTIONS

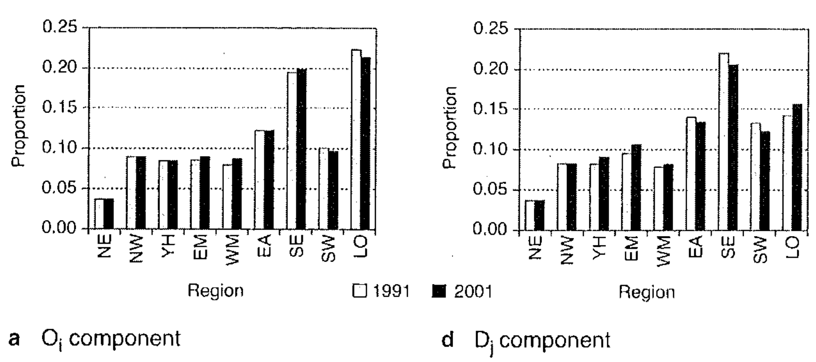
In this section, the interregional migration flows are analysed by describing the main effect and two-way interaction components for three-way tables that include age or ethnicity. The aim is to identify regularities or changes in these components over time, as well as the main contributors to the migration patterns. Driving much of the change in patterns over time is the increase in the overall levels (i.e., $n_{+++}$): interregional migration increased from 630 thousand in 1991 to 920 thousand in 2001.

### Interregional Migration

Common to both the age and ethnic migration tables are the aggregate interregional migration flows. The proportions of migration from and to each region are set out in Figure 1. These represent the origin and destination main effect components ($O_i$ and $D_j$, respectively) described earlier. For the most part, there were no major changes over time, with the minor exceptions of migration from London (decrease) and migration to East Midlands (increase), South East (decrease), South West (decrease) and London (increase). The South East and London sent out about 20% of all migrants each. The South East received the highest number of migrants (over 20%), followed by East of England, South West and London (which received about the same levels). The North East stood out in both cases as the region that sent and received the fewest amount of migrants.

The origin-destination interaction components for 1991 and 2001 are set out in Figure 2. These ratios capture the connectivity between regions by comparing the observed level with the expected one. If the ratio is greater than one, the connection between two regions is considered to be relatively strong, whereas the opposite is true for ratios less than one. From the information in Figure 2, we find that the connectivity is stronger between neighbouring regions, for example, between Yorkshire and the Humber and North East which exhibited ratios of 2.8 in 1991 and 3.2 in 2001. This means that

*Figure 1. The origin and destination main effects of interregional migration in England, 1991 and 2001*




NE: North East; NW: North West; YH: Yorkshire and the Humber; EM: East Midlands; WM: West Midlands; EA: East of England SE: South East; SW: South West; LO: London





*Figure 2. Table of origin-destination interaction components of interregional migration in England, 1991 and 2001*

| | Destination | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Origin | NE | NW | YH | EM | WM | EA | SE | SW | LO |
| A. 1991 | | | | | | | | | |
| NE | | 1.822 | 2.963 | 0.928 | 0.875 | 0.700 | 0.723 | 0.585 | 0.876 |
| NW | 1.742 | | 2.130 | 1.236 | 1.757 | 0.680 | 0.770 | 0.919 | 0.920 |
| YH | 2.783 | 2.147 | | 2.130 | 1.011 | 0.734 | 0.663 | 0.693 | 0.763 |
| EM | 0.986 | 1.286 | 2.091 | | 1.844 | 1.235 | 0.800 | 0.792 | 0.693 |
| WM | 0.833 | 1.726 | 1.075 | 1.835 | | 0.680 | 0.813 | 1.409 | 0.789 |
| EA | 0.801 | 0.795 | 0.835 | 1.378 | 0.861 | | 1.171 | 0.979 | 1.826 |
| SE | 0.851 | 0.892 | 0.768 | 0.919 | 0.986 | 1.043 | | 1.875 | 1.998 |
| SW | 0.771 | 1.029 | 0.802 | 0.803 | 1.566 | 0.794 | 1.693 | | 1.032 |
| LO | 0.607 | 0.655 | 0.539 | 0.534 | 0.587 | 1.895 | 1.881 | 0.839 | |
| B. 2001 | | | | | | | | | |
| NE | | 2.051 | 2.830 | 0.957 | 0.869 | 0.669 | 0.636 | 0.494 | 0.858 |
| NW | 1.981 | | 2.487 | 1.139 | 1.653 | 0.614 | 0.685 | 0.774 | 0.940 |
| YH | 3.189 | 2.425 | | 1.919 | 1.011 | 0.633 | 0.636 | 0.632 | 0.819 |
| EM | 0.996 | 1.343 | 2.085 | | 1.940 | 1.064 | 0.770 | 0.774 | 0.803 |
| WM | 0.843 | 1.677 | 1.060 | 1.783 | | 0.640 | 0.778 | 1.461 | 0.872 |
| EA | 0.765 | 0.748 | 0.870 | 1.483 | 0.885 | | 1.178 | 0.909 | 1.684 |
| SE | 0.663 | 0.795 | 0.693 | 0.970 | 0.976 | 1.070 | | 1.999 | 1.862 |
| SW | 0.740 | 0.945 | 0.725 | 0.764 | 1.528 | 0.744 | 1.752 | | 1.144 |
| LO | 0.530 | 0.618 | 0.478 | 0.546 | 0.612 | 2.098 | 2.008 | 0.805 | |
| C. Ratio 2001 to 1991 | | | | | | | | | |
| NE | | 1.126 | 0.955 | 1.031 | 0.993 | 0.956 | 0.879 | 0.844 | 0.979 |
| NW | 1.137 | | 1.168 | 0.921 | 0.941 | 0.903 | 0.891 | 0.843 | 1.021 |
| YH | 1.146 | 1.130 | | 0.901 | 1.001 | 0.862 | 0.958 | 0.912 | 1.072 |
| EM | 1.010 | 1.045 | 0.997 | | 1.052 | 0.862 | 0.962 | 0.978 | 1.159 |
| WM | 1.012 | 0.972 | 0.986 | 0.972 | | 0.940 | 0.956 | 1.037 | 1.106 |
| EA | 0.955 | 0.940 | 1.042 | 1.076 | 1.029 | | 1.006 | 0.929 | 0.922 |
| SE | 0.779 | 0.891 | 0.902 | 1.055 | 0.990 | 1.026 | | 1.066 | 0.932 |
| SW | 0.959 | 0.918 | 0.905 | 0.951 | 0.976 | 0.938 | 1.034 | | 1.108 |
| LO | 0.872 | 0.944 | 0.887 | 1.021 | 1.044 | 1.107 | 1.068 | 0.960 | |

Notes: NE = North East, NW = North West, YH = Yorkshire and the Humber, EM = East Midlands, WM = West Midlands, EA = East of England, SE = South East, SW = South West and LO = London.

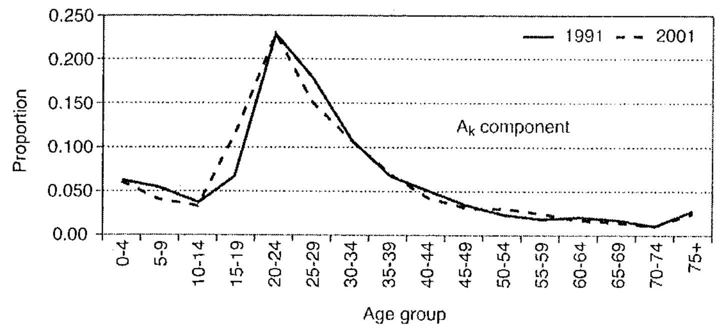
the levels of migration between these two regions were about three times the expected value, which is based on the total number leaving Yorkshire and the Humber and the total number going to the North East, and that the connection between these two places increased over time. The connection between London and Yorkshire and the Humber, on the other hand, was relatively weak with ratios of 0.54 in 1991 and 0.48 in 2001. Here, there were roughly one half the numbers of migrants than expected, with the connection between these two places getting slightly weaker over time.

Finally, in Figure 2, we find that the connections between the southern regions and the northern regions have gotten weaker over time, while the connections between northern regions got notably stronger. Also, the connections between the Midlands and South West with London got considerably stronger (over 10% increase).

### Interregional Migration by Age

The proportions of migration in each five-year age group are set out in Figure 2 for the 1991 and 2001 flows. Thus the area under each curve sums to unity. These represent the age main effects of the multiplicative component models for the origin by destination by age tables. Specifically,

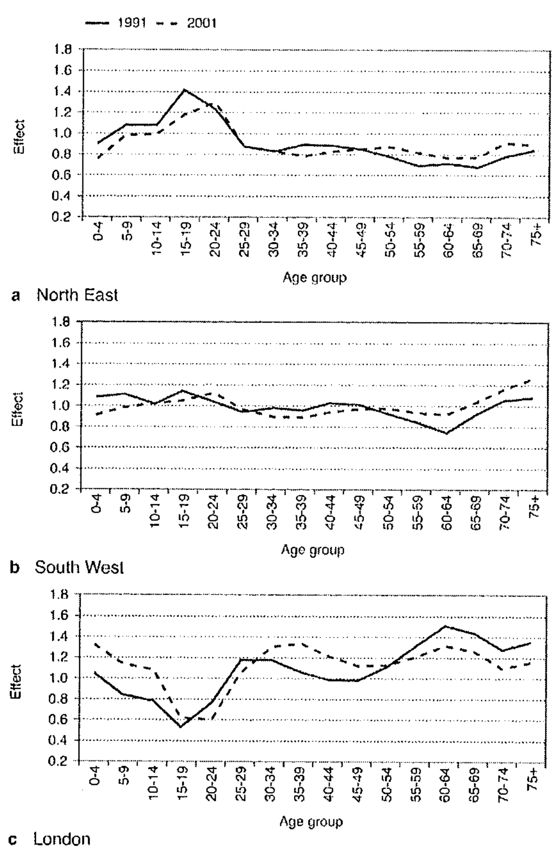
*Figure 3. The age main effect components of interregional migration in England, 1991 and 2001*




the proportions represent the $A_x$ term in the following model:

$$n_{ijx} = (T)(O_i)(D_j)(A_x)(OD_{ij})(OA_{ix})(DA_{jx})(ODA_{ijx}) \quad (10)$$

*Figure 4. The origin-age interaction effects of interregional migration in England, 1991 and 2001, north east, south west and London*




About 70% of migrants are aged 15-44 years (see Chapter 8 for a more detailed analysis of age-specific rates in 2001). The patterns stayed the same over time for the most part. The exceptions were migrants in the 15-19 year age group which increased and migrants in the 5-9 year and 25-29 year age groups which declined. These changes may have been caused by the differences in the measurement of students in the two censuses.

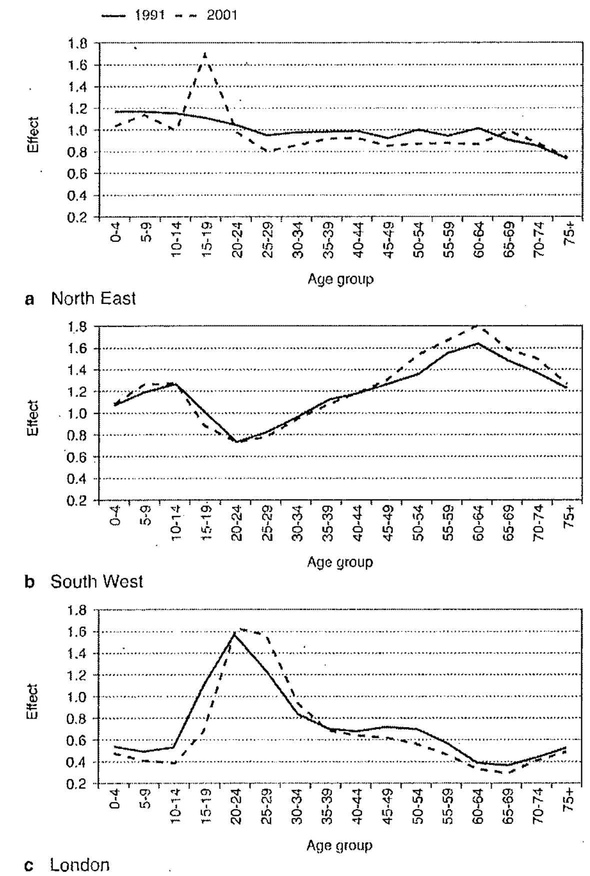
Note: Y axes = proportions; X axis = age group.

Next, we consider the $OA_{ix}$ and $DA_{jx}$ interaction components. These values are obtained by dividing the observed age profile of migration from each region by the overall age profile of migration, i.e., $(n_{i+x}/n_{i++})/(n_{++x}/n_{+++})$ for $OA_{ix}$ and by dividing the observed age profile of migration to each region by the overall age profile of migration, i.e., $(n_{+jx}/n_{+j+})/(n_{++x}/n_{+++})$ for $DA_{jx}$. Here, the overall age profile of migration is the expected one, which is set out in Figure 3. Examples of $OA_{ix}$ are set out in Figure 4 for migration from the North East, South West and London. The patterns over time are generally stable. For migration from the North East, the age profile resembles the expected one, with the exception of the relatively higher levels exhibited in the 15-19 and 20-24 year age groups. For migration from the South West, the patterns are very similar to the expected one (Figure 3). For London, however, we see that the patterns differed substantially from the overall age profiles. Here, migrants over the age of 30 years (and their children) were more likely to leave whereas persons 15-24 years were much less likely to leave. Also, the ratios shifted slightly over time from 1991 to 2001 with lower levels in the 60+ age groups and higher levels in the 0-14 and 30-49 age groups.





*Figure 5. The destination-age interaction effects of interregional migration in England, 1991 and 2001, north east, south west and London*




Examples of $DA_{jx}$ are set out in Figure 5. Migration to the North East generally resembled the overall age pattern found in Figure 3. Migration to the South West and to London, on the other hand, exhibited age patterns of migration that were quite different than the overall age profile. Here, we find that 15-29 year olds were less likely to choose South West and more likely to choose London, while the remaining age groups were less likely to choose London and more likely to choose South West. As for patterns over time, we see a large increase in the ratio for 15-19 year olds migrating to the North East, a slight increase in elderly migration to the South West and a slight shift to the right in the labour force peak for migration to London.

## Interregional Migration by Ethnicity

The proportions of migration by each ethnic group are set out in Figure 6 for the 1991 and 2001 flows. These represent the ethnic main effects of the multiplicative component models for the origin by destination by ethnicity tables. Specifically, the proportions represent the $E_z$ term in the following model:

$$n_{ijz} = (T)(O_i)(D_j)(E_z)(OD_{ij})(OE_{iz})(DE_{jz})(ODE_{ijz}) \quad (11)$$

The proportion of White migrants decreased substantially from 95% in 1991 to 90% in 2001. The three ethnic minority groups all experienced substantial increases in their proportions of overall migration, with roughly equal increases across groups (about 80%).

The $OE_{iz}$ and $DE_{jz}$ interaction components are set out in Figure 7 for the four ethnic groups. These values are obtained by dividing the observed proportion of ethnic migration from each region by the overall proportion of ethnic migration, i.e., $(n_{i+z}/n_{i++})/(n_{++z}/n_{+++})$ for $OE_{iz}$ and by dividing the proportion of ethnic migration to each region by the overall ethnic proportion of migration, i.e., $(n_{+jz}/n_{+j+})/(n_{++z}/n_{+++})$ for $DE_{jz}$. Here, the overall proportions of ethnic migration are the expected ones set out in Figure 6.

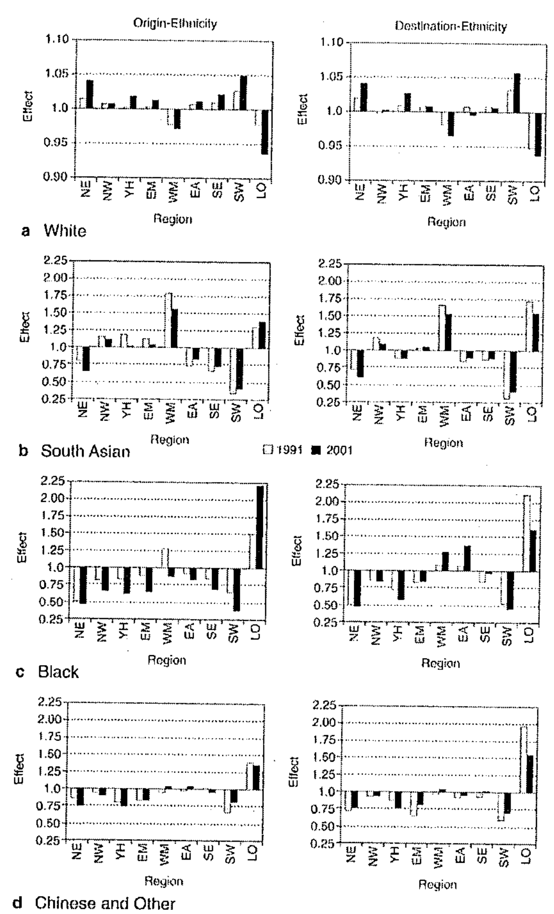
The $OE_{iz}$ and $DE_{jz}$ patterns over time show some interesting and notable differences between Whites, South Asians, Blacks and Chinese and Other ethnic groups. Note, because the $OE_{iz}$ and $DE_{jz}$ components for White migrants are on average smaller than for the other ethnic groups,

*Figure 6. Table of the ethnicity main effects of interregional migration in England, 1991 and 2001*

| Year | White | South Asian | Black | Chinese & Other |
|---|---|---|---|---|
| 1991 | 0.945 | 0.024 | 0.012 | 0.019 |
| 2001 | 0.899 | 0.045 | 0.021 | 0.034 |

the proportions in the Y-axis are represented with a different scale. The proportions of White migrants were lower than expected from/to London and West Midlands, while they were higher than expected from/to North East and the South West. Over time, we find increases in the ratios for North East and South West and decreases in the ratios for London. The proportions of South Asian migrants were substantially higher than expected for migration from/to West Midlands and London but substantially lower than expected for migration from/to South West. Over time, the ratios decreased for West Midlands and London (destination only). The proportions of Black migration were much higher than expected from/to London and much lower than expected for most other regions. Interestingly, most of the ratios changed considerably over time, with London losing its attractiveness for these migrants. Finally, Chinese and Other migrants exhibited a different set of patterns, with London being the only origin and destination with higher than expected proportions.

*Figure 7. The origin-ethnicity and destination-ethnicity interaction effects of interregional migration in England, 1991 and 2001, four ethnic groups*




NE: North East; NW: North West; YH: Yorkshire and the Humber; EM: East Midlands; WM: West Midlands; EA: East of England SE: South East; SW: South West; LO: London

In summary, the analysis of the main effect and two-way interaction effects of interregional migration in England by age and ethnicity over time provides several interesting findings. First, connectivity between regions is not uniform across origin and destinations of England, although it is rather stable over time. Second, the age profile of migration has exhibited strong regularities over time, but there are some differences in the age patterns of migrants leaving from or going to each region. Third, the main effects and two-way interactions of ethnicity differ greatly according to each ethnic group. The importance of these structures, however, has not been assessed. Also, we have not examined the three-way interaction terms. This is carried out in the next section by comparing unsaturated log-linear model fits of the patterns over time with the corresponding observed values.





*Figure 8. Table of unsaturated log-linear model fits of migration tables cross-classified by origin, destination, age and time*

| MODEL | $G^2$ | df | $G^2$/df | AIC |
|---|---|---|---|---|
| O,D,A,T | 444,287 | 2,271 | 196 | 461,829 |
| OD,A,T | 144,000 | 2,216 | 65 | 161,653 |
| OD,OA,DA,T | 39,297 | 1,976 | 20 | 57,429 |
| OD,OA,DA,AT | 23,938 | 1,961 | 12 | 42,100 |
| ODA,T | 30,912 | 1,151 | 27 | 50,694 |
| ODA,AT | 15,553 | 1,136 | 14 | 35,365 |
| ODT,A | 139,370 | 2,145 | 65 | 157,164 |
| ODT,AT | 124,011 | 2,130 | 58 | 141,835 |

## LOG-LINEAR ANALYSES OF MIGRATION FLOW TABLES: 1991 AND 2001 CENSUSES

In this section, we analyse unsaturated log-linear models of two four-way tables. These tables represent migration by origin, destination, age and time and migration by origin, destination, ethnicity and time. There are many possible models for four-way tables. We only compare eight selected models, starting with a very simple mutual independence model:

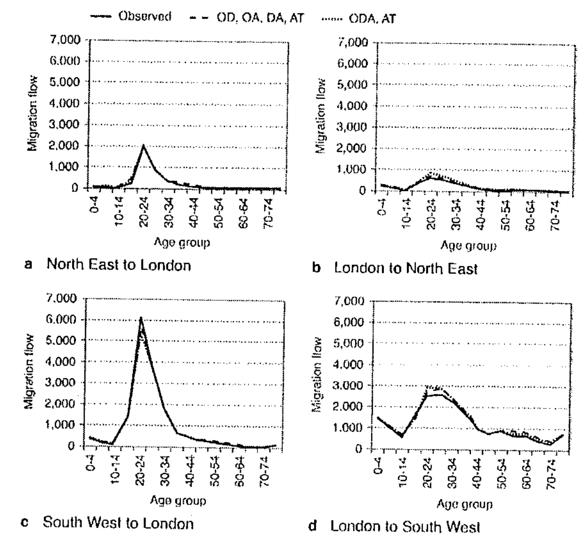
1. *O, D, X, T* Our simplest model that assumes mutual independence between variables.
2. *OD, X, T* Model including only the interaction between origin and destination.
3. *OD, OX, DX, T* Model with the interactions between origin and variable *X* and destination and variable *X* added to the model above.
4. *OD, OX, DX, XT* Model with the interaction between the variable *X* and time added to the model above.
5. *ODX, T* Model with the three-way interaction between origin, destination and variable *X*.
6. *ODX, XT* Model with two-way interaction between variable *X* and time added to the model above.
7. *ODT, X* Model with three-way interaction between origin, destination and time.
8. *ODT, XT* Model with two-way interaction between *X* and time added to the model above.

The comparison of these models allows us to test the importance of various structures in the tables. For example, we can test whether including a three-way interaction term produces better estimates than a model that only includes two-way interaction terms.

### Spatial and Age Structures over Time

The unsaturated log-linear models compared in this section focus on the age structures of migration. The results of the eight model fits are set out in Figure 8. These models can be compared by examining the $G^2$ and *AIC* goodness of fit statistics (see equations 7 and 8, respectively), where values closer to zero implies better fits. These measures are meant to be interpreted relatively. Here, we see that the worst fitting model is the [*O, D, A, T*] and the best fitting model is [*ODA, AT*]. However, these statistics do not take into account model complexity. We do this by dividing the $G^2$ statistic by the residual degrees of freedom. For the models set out in Figure 8, there are 2304 observations (i.e., 9 origins * 9 destinations * 16 age groups * 2 time points, less the 288 diagonal elements between origin and destination). So, for example, the [*O, D, A, T*] model is the simplest with only 33 parameters. This model, not surprisingly, fits poorly. The [*ODA, AT*] model performs well but requires more than twice the number of parameters than the simplest model. A good compromise appears to be the [*OD, OA, DA, AT*] model. The *AIC* does

*Figure 9. Estimated age patterns of migration between north east and London and south west and London, 2001, a comparison of two unsaturated log-linear models*




not appear to penalise as much as the $G^2$ divided by the residual degrees of freedom.

To get a better sense of the differences between the two best models, [*OD, OA, DA, AT*] and [*ODA, AT*], we compare the predicted values in Figure 9. Here, we see that the differences between the two are negligible. This means that for estimation purposes, one could use the simpler two-interaction model and expect good results. Finally, the reason why [*ODT, T*] and [*ODT, AT*] do not perform well is that they do not include the interactions between origin and age or destination and age. The interaction between age and time appears to be important.

### Spatial and Ethnic Structures over Time

The results of the eight unsaturated log-linear models that include ethnicity are set out in Figure 10. Similar to the models for age, it appears that the two best models are [*OD,OE,DE,ET*] and [*ODE,ET*]. The predicted values for South

*Figure 10. Unsaturated log-linear model fits of migration tables cross-classified by origin, destination, ethnicity and time*

| MODEL | $G^2$ | df | $G^2$/df | AIC |
|---|---|---|---|---|
| O,D,E,T | 364,379 | 555 | 657 | 368,885 |
| OD,E,T | 65,242 | 500 | 130 | 69,857 |
| OD,OE,DE,T | 19,726 | 452 | 44 | 24,438 |
| OD,OE,DE,ET | 8,782 | 449 | 20 | 13,500 |
| ODE,T | 16,929 | 287 | 59 | 21,971 |
| ODE,ET | 5,985 | 284 | 21 | 11,033 |
| ODT,E | 60,726 | 429 | 142 | 65,483 |
| ODT,ET | 49,781 | 426 | 117 | 54,545 |

Asian migration from North East, South West and London are set out in Figure 11. Again, both models appear to predict similar values and one could probably get away with the simpler two-way interaction model for estimation purposes. Notice that the totals do not match up with the observed. This is because the interaction between origin and time was not included in either model. The reason for not including this term was based on the comparison of the main effects set out in Figure 1. The inclusion of the interaction between ethnicity and time improves the model fit considerably.

## CONCLUSION

We have explored the age and ethnic structures of interregional migration in England, as measured by the 1991 and 2001 censuses. Despite a large increase in the level of interregional migration, we find that the structures have remained fairly stable over time with the important exceptions of the main effects and two-way interaction terms involving ethnicity. The analyses of these ethnic structures reveal the large differences existing between the four main groups, in terms of levels and spatial patterns. The most important change that has occurred has been the substantial decrease in the proportion of White interregional migration from 95% in 1991 to 90% in 2001. As for the other structures in the three-way migration tables, we find some broad changes over time, notably the increasing connectivity between Northern regions





*Figure 11. Estimated south Asian migration flows from the north east, south west and London, 1991 and 2001: A comparison of two unsaturated log-linear models*

| | | | Destination | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Origin | Year | Model | NE | NW | YH | EM | WM | EA | SE | SW | LO | Total |
| NE | 1991 | OD,OE,DE,ET | | 57 | 75 | 32 | 42 | 21 | 33 | 7 | 94 | 362 |
| | | ODE,ET | | 63 | 48 | 26 | 57 | 27 | 33 | 14 | 92 | 362 |
| | | Observed | | 68 | 66 | 30 | 61 | 39 | 64 | 15 | 84 | 427 |
| | 2001 | OD,OE,DE,ET | | 159 | 209 | 89 | 116 | 59 | 92 | 21 | 262 | 1,007 |
| | | ODE,ET | | 176 | 135 | 74 | 159 | 75 | 92 | 40 | 258 | 1,007 |
| | | Observed | | 171 | 117 | 70 | 155 | 63 | 61 | 39 | 266 | 942 |
| SW | 1991 | OD,OE,DE,ET | 7 | 40 | 28 | 39 | 107 | 34 | 119 | | 172 | 546 |
| | | ODE,ET | 3 | 39 | 35 | 46 | 70 | 46 | 133 | | 174 | 546 |
| | | Observed | 5 | 57 | 21 | 46 | 81 | 44 | 82 | | 142 | 478 |
| | 2001 | OD,OE,DE,ET | 20 | 112 | 78 | 108 | 297 | 94 | 331 | | 480 | 1,521 |
| | | ODE,ET | 8 | 110 | 96 | 127 | 194 | 130 | 372 | | 484 | 1,521 |
| | | Observed | 6 | 92 | 110 | 127 | 183 | 132 | 423 | | 516 | 1,589 |
| LO | 1991 | OD,OE,DE,ET | 69 | 316 | 227 | 323 | 466 | 1,095 | 1,676 | 206 | | 4,378 |
| | | ODE,ET | 68 | 363 | 242 | 433 | 551 | 998 | 1,503 | 219 | | 4,378 |
| | | Observed | 61 | 399 | 274 | 387 | 530 | 901 | 1,583 | 200 | | 4,335 |
| | 2001 | OD,OE,DE,ET | 191 | 880 | 632 | 901 | 1,297 | 3,050 | 4,669 | 575 | | 12,196 |
| | | ODE,ET | 191 | 1,011 | 674 | 1,207 | 1,535 | 2,780 | 4,187 | 612 | | 12,196 |
| | | Observed | 198 | 975 | 642 | 1,253 | 1,556 | 2,877 | 4,107 | 631 | | 12,239 |

and the decreasing connectivity between Northern and Southern regions. The analysis of the age components found some important differences in the patterns across regions, as captured by the two-way interactions with origin and destination. The three way interaction terms, however, did not contribute much additional information. For estimation purposes, these terms could be dropped.

The approach described in this chapter for analysing migration structures may be generally classified as demographic migration modelling (Stillwell, 2009). Here, the emphasis is on analysing the macro-level structures contained in the data rather than the development of an explanatory-type model (e.g., Fotheringham *et al.*, 2004). Willekens and Baydar (1986) and Rogers *et al.* (2001; 2002) provide good examples of the demographic modelling approach. In these works, the emphasis is on modelling the generation and distribution components of interregional migration. The generation component captures the relative levels of migration from each region; the distribution component captures the allocation of that level to the various destinations. Our work takes a slightly different approach by focusing on the main effects and interaction terms, where the interaction terms are expressed in terms of observed to expected ratios. This approach is more in line with standard analyses for categorical data (e.g., Agresti, 2002). In both the generation and distribution model and the multiplicative component model, macro structures exhibit strong stability over time. This means that the modelling of migration may be simplified to focus on the key structures that explain most of the patterns or that capture the most important changes (van Wissen *et al.*, 2009).

In conclusion, the results presented in this chapter are important for several reasons. First, this analysis may be used to better understand the factors that drive changes in the aggregate migration patterns over time. Second, these findings may be used to guide the estimation or projection of migration flows in the context of missing or inadequate migration flows. Finally, these findings may be used to develop models for combining migration obtained from different sources with different measurements (Raymer *et al.*, 2007; Raymer *et al.*, 2008).

## ACKNOWLEDGMENT

Support for this research came from the Economic and Social Research Council (RES-000-22-2501) and from the University of Southampton's Annual 'Adventures in Research' Grant, 2006-2007. 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

Agresti, A. (2002). *Categorical Data Analysis*. Hoboken, NJ: Wiley. doi:10.1002/0471249688

Bates, J., & Bracken, I. (1982). Estimation of migration profiles in England and Wales. *Environment & Planning A*, *14*, 889–900. doi:10.1068/a140889

Bates, J., & Bracken, I. (1987). Migration age profiles for local authority areas in England, 1971-1981. *Environment & Planning A*, *19*, 521–535. doi:10.1068/a190521

Baydar, N. (1984). *Issues in multiregional demographic forecasting*. Ph.D. Dissertation, Vrije Universiteit Brussel.

Congdon, P. (2005). *Bayesian Models for Categorical Data*. Chichester, UK: Wiley. doi:10.1002/0470092394

Finney, N., & Simpson, L. (2008). Internal migration and ethnic groups: evidence for Britain from the 2001 Census. *Population Space and Place*, *14*, 63–83. doi:10.1002/psp.481

Fotheringham, A. S., Rees, P., Champion, T., Kalogirou, S., & Tremayne, A. R. (2004). The development of a migration model for England and Wales: overview and modelling out-migration. *Environment & Planning A*, *36*, 1633–1672. doi:10.1068/a36136

ONS. (2004). *Origin-Destination Statistics: Local Authorities (CD-ROM)*. London: Office for National Statistics.

Raymer, J., Abel, G., & 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. Series A (General)*, *170*, 891–908.

Raymer, J., Bonaguidi, A., & Valentini, A. (2006). Describing and projecting the age and spatial structures of interregional migration in Italy. *Population Space and Place*, *12*, 371–388. doi:10.1002/psp.414

Raymer, J., & Rogers, A. (2007). Using age and spatial flow structures in the indirect estimation of migration streams. *Demography*, *44*(2), 199–223. doi:10.1353/dem.2007.0016

Raymer, J., Smith, P. W. F., & Giulietti, C. (2008). *Combining census and registration data to analyse ethnic migration patterns in England from 1991 to 2007*. Paper presented at the European Population Conference, Barcelona, 9-12 July.

Rees, P., & Butt, F. (2004). Ethnic change and diversity in England, 1981-2001. *Area*, *36*(2), 174–186. doi:10.1111/j.0004-0894.2004.00213.x

Rogers, A., Willekens, F. J., Little, J. S., & Raymer, J. (2002). Describing migration spatial structure. *Papers in Regional Science*, *81*, 29–48. doi:10.1007/s101100100090

Rogers, A., Willekens, F. J., & Raymer, J. (2001). Modeling interregional migration flows: Continuity and change. *Mathematical Population Studies*, *9*, 231–263. doi:10.1080/08898480109525506

Rogers, A., Willekens, F. J., & Raymer, J. (2002). Capturing the age and spatial structures of migration. *Environment & Planning A*, *34*, 341–359. doi:10.1068/a33226





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

Stillwell, J. (2009). Inter-regional migration modelling: a review. In Poot, J., Waldorf, B., & van Wissen, L. (Eds.), *Migration and Human Capital* (pp. 29–48). Cheltenham, UK: Edward Elgar.

Stillwell, J., & 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, University of Kent, Canterbury.

Stillwell, J., & Duke-Williams, O. (2007). Understanding the 2001 UK census migration and commuting data: the effect of small cell adjustment and problems of comparison with 1991 flow datasets. *Journal of the Royal Statistical Society A*, *170*(2), 425–445. doi:10.1111/j.1467-985X.2006.00458.x

Van Wissen, L., van der Gaag, N., Rees, P., & Stillwell, J. (2009). In search of a modelling strategy for projecting internal migration in European countries: Demographic versus economic-geographical approaches. In Poot, J., Waldorf, B., & van Wissen, L. (Eds.), *Migration and Human Capital* (pp. 49–74). Cheltenham, UK: Edward Elgar.

Willekens, F. J. (1983). Log-linear modelling of spatial interaction. *Papers / Regional Science Association. Regional Science Association. Meeting*, *52*, 187–205. doi:10.1007/BF01944102

Willekens, F. J., & Baydar, N. (1986). Forecasting place-to-place migration with generalized linear models. In Woods, R., & Rees, P. (Eds.), *Population structures and models: Developments in spatial demography* (pp. 203–245). London: Allen & Unwin.