COMBINING MIGRATION DATA IN ENGLAND AND WALES

James Raymer
Peter W.F. Smith
Corrado Giulietti

University of Southampton

15 October 2007

The following is taken from our recently successful ESRC grant proposal, which starts at the beginning of January 2008. We plan to present the model and some preliminary results from this research project.

The study of migration is hindered by data availability and data quality. Overcoming data issues is a central problem in migration research. Surveys are usually not big 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 of data suppression for confidentiality reasons. This is particularly relevant for analyses at relatively high levels of spatial disaggregation. Also, inconsistencies arise due to different migrant definitions caused by various migrant data collection systems (e.g., surveys, censuses or population registers). Finally, 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disaggregation are generally only available from censuses, which only occur every five or ten years and are often outdated by the time they come out. In England and Wales, 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 registers. More detailed data can be obtained from the decennial censuses. Our research will develop methodologies to combine these two sources of information for the purpose of producing reliable and detailed estimates of migration. Methods for combining a third data source, such as survey information will also be explored. The outcome will consist of three time series of detailed estimates of internal migration cross-classified by five variables: origin, destination, age and sex, and ethnicity, education or employment. These estimates can then be used to improve our understanding of migration behaviour or even the forecasting of future migration flows.

At present, the internal migration data in England and Wales are limited due to differences in sources, availability, quality and measurement. In this project, we will provide some solutions to deal with the above problems by developing a statistical model that combines different data types and information to predict detailed flows and we will demonstrate the usefulness of these estimates by analysing them over time from 1991 to 2005. 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 England and Wales. 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.

Aims and objectives

The overall aim of this research is to improve the estimation of migration flows that can be used in population planning or policy. We will utilise and extend a recently developed statistical model for estimating detailed migration flows over time. This log-linear model facilitates the combining of *incomplete* registration data with *auxiliary* census and survey data. More precisely, the objectives are:

1. To apply the statistical model to produce estimates of detailed migration flows over time in England and Wales at the local authority level by origin, destination, age, and sex, and ethnicity, education or employment.

2. To extend the statistical model to allow for multiple sources of auxiliary data and to include more detailed registration data.

3. To analyse the estimated flows to obtain a better understanding of internal migration in England and Wales with the view to aiding policy makers and migration researchers.

4. To disseminate the results of this research by publishing in a substantive journal and a methodological journal, and through presentations at conferences and academic events.

Background

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. There are two reasons. First, no consensus exists on what exactly is a "migration". Therefore, comparative analyses suffer from differing national views concerning who is a migrant. 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 within England and Wales? One possibility is to have a methodology for combining existing migration data that accounts for the various strengths in the multiple migration data sources.

The data needed to understand population redistribution and migration behaviour over time and across groups are often inadequate, missing or inconsistent. This 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 by migrants.

The measurement of 'migration' is not consistent across various data sources. 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. 2003; 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). In analysing migration data in England and Wales, Raymer et al. (2007) 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.

Methodology

The starting point for this research is a recent paper written by the two investigators of this proposal (Raymer et al. 2007) that 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. More research is needed to test the applicability of the approach to a wider set of migration patterns. This project will extend the methodology by including more auxiliary information (e.g., two censuses or a census and survey) and by producing an annual time series of estimates from 1991 to 2005, which can then be used for analysis. Note we expect additional problems to surface when expanding to a wider set of migration patterns, such as the occurrence of non-response by young adult males in both main sources of data. 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. Ways to deal with non-response in the methodology will be explored, for example, by including adjustment factors based on other age-specific patterns of males or based on the patterns of young adult females.

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; Rees et al. 1996; 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 proposed project does. It will allow for both intercensal and post-census estimates of detailed migration flows. These estimates have the possibility to increase our understanding of population change. For example, by having internal migration data by ethnicity over time, we will be able to see when and how the population has become more spread out or more concentrated, which we will compare with other studies that just focus on the decennial census data (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). The issues this study focuses on are those concerning inconsistencies between sources of migration data and those arising in the development of a model that can effectively incorporate migration data from multiple sources to produce reliable and detailed estimates over time.

The proposed 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 2005; Willekens 1977, 1980, 1982, 1983, 1999). The log-linear model version of the spatial interaction model (Willekens 1980, 1983) is of particular importance for this study. 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).

Migration flow data are commonly presented in a square table, with off-diagonal entries containing the number of people migrating from origin i to destination j. Origins and destinations can represent specific places or regions or aggregations of places. We will use the log-linear model developed by Raymer et al. (2007) as a starting point to estimate detailed migration flows over time. By detailed migration flows, we mean origin by destination flow tables disaggregated by age, sex, and at least one other characteristic, such as ethnicity, education or employment. The model in Raymer et al. (2007) combined 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 was imposed on the registration (incomplete) data. (See Attachment 1 for more details.) This model will be used to produce various sets of annual detailed estimates of migration flows (Objective 1) and extended in three ways (Objective 2). First, we will test the effectiveness of the model to a wider set of detailed migration flows. Second, we will extend the model to take advantage of the origin-destination two-way associations in the incomplete data and develop methods for correcting any noticeable problems with the incomplete data (e.g., under-registration of young adult males). Third, we will extend the model to take advantage of multiple sources of auxiliary data, such as 1991 and 2001 censuses, by allowing the association structure imposed on the registration data to evolve over time.

Once the model has been extended and tested, we will describe and analyse detailed migration estimates over time by ethnicity, education and employment (Objective 3). For example, consider migration flows between local authorities by ethnicity, which are currently only available from the decennial censuses. An estimated annual time series of these data would allow one to identify the changing sources of growth due to internal migration, which can then be compared with other sources of growth, such as fertility or immigration. Furthermore, because the estimated data will consist of multi-way tables, methods for analysing these patterns and presenting the results will be examined. Here the multiplicative component approach will be used. This approach basically disaggregates the flows into an overall level, main effect (proportions) and interaction effects. These were used by Raymer et al. (2006) to analyse age-specific interregional migration patterns in Italy from 1970 to 2000. Another approach to be pursued is to focus on particular origin-destination-specific flows that are of interest to policy makers. For instance, Raymer et al. (2007) analysed age-specific elderly migration flows between Centres of Industry and Coastal and Countryside in England and Wales. This allowed them to estimate and analyse patterns of elderly retirement and return migration. In this project, we will compare age and sex-specific flows from and to areas of foreign population concentrations by ethnicity, and between geographical areas by education and employment.

Data

Internal migration data are typically obtained from national-level surveys or registration systems. In England and Wales, migration data can be obtained from the decennial censuses, the General Household Survey (GHS) and the National Health Service Central Register (NHSCR). The information obtained from the census contains much of the information needed but are only collected every ten years. The GHS asks reasonably detailed questions about migration behaviour every year, but the sample size is not big enough for sub-national analyses. The NHSCR organises large national registers but with minimal information on migrant behaviour (i.e., origin, destination, age and sex) and with a tendency to miss important population groups, such as young adult males, who are known to be less likely to register.

Much of the data required for this study have already been prepared for analysis as part of a current pilot project to study the relationships between immigration and internal migration funded by the University of Southampton. The prepared migration data consist of flows between local authorities and twelve ONS groups from (1) the 2001 Census (ONS 2004); (2) the 2001 Samples of Anonymised Records (SAR); (3) the 1991 Census; and (4) the 1991-2005 NHSCR flows. The fourth data set required reconciling the 1991-2000 geography with the post-2000 geography. (See Attachment 2 for more details on the NHSCR and 2001 census data.) The migration data from the GHS will have to be prepared for analysis as part of this project. These data allow us to analyse aggregate migration patterns between local authorities, counties, regions or ONS classification groups (which are constructed from data at the local authority level) over time and across subgroups (in 1991 and 2001). To analyse subgroup patterns over time, we will first focus on combining post-2000 NHSCR data with the 2001 census data. Later, we will extend the methods to include detailed patterns over time and to incorporate survey data (i.e., from 2001 SAR or from GHS). The methodology will be applied to estimate patterns at the local authorities, county and ONS classification group levels. To simplify the modelling and analysis at the local authority level, we will focus on patterns within particular geographic regions (e.g., migration between local authorities in the South East region).

Attachment 1: Statistical methodology

A general methodology that allows the combination of registration and census migration flow data is described in this section, with the aim of providing detailed estimates between census dates. We assume that observations in both the *auxiliary* census data and the *incomplete* registration data are samples from the same underlying population. For our particular study, we require annual migration between groups by age, sex and some other variable (e.g., ethnicity, employment or skill level). The NHS only provides annual migration data by age and sex. Information regarding the other variable and origin-destination-specific flows has to be obtained from the most recent census.

Migration flow data is commonly presented in a square table, with off-diagonal entries containing the number of people migrating from origin i to destination j. Origins and destinations can represent specific places or regions or aggregations of places. The diagonal information in the migration flow table can vary: sometimes it is omitted, sometimes it contains migrants within a region or aggregation of places and sometimes it contains both non-migrants and migrants within a region or aggregation of places.

The modelling of migration flow data is commonly undertaken using a spatial interaction models (refer to Fotheringham et al., 2000, pages 211-235 and Willekens, 1983, 1999 for a discussion of the models and a review of the literature). A simplistic version of the spatial interaction model to estimate the number of migrations, n_{ij}, in our incomplete data set, from origin i to destination j during a unit interval may be applied as in Willekens (1999):

$$\lambda_{ij} = \alpha_i \beta_j m_{ij} , \qquad (1)$$

where λ_{ij} 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 α_i and β_j parameters represent background factors related to the characteristics of the origin and destination, respectively. The m_{ij} factor represents auxiliary information on migration flows, which imposes its interaction structure onto the estimated flow matrix. This is typically additional data relating to migration between the same origins and destinations as in the incomplete data and is not a parameter in the spatial interaction model. As a result the associations between origins and destinations in the auxiliary data will be replicated in the estimates for the incomplete data. Combining the incomplete and auxiliary data provides up-to-date estimates of the expected number of migration flows by origin and destination.

As described by Willekens (1999), the estimation of α_i and β_j can be performed by re-expressing the spatial interaction model (1) in terms of the log-linear model

$$\log \lambda_{ij} = \log \alpha_i + \log \beta_j + \log m_{ij} , \qquad (2)$$

where the final term is commonly referred to as an offset and, unlike standard log-linear models, no intercept term is included. For estimation, a single constraint must be placed on the α_i and β_j.

The probability of observing n_{ij} migrations during a unit interval is given by the Poisson distribution function:

$$P\left(N_{ij}=n_{ij}\right)=\frac{\lambda_{ij}^{n_{ij}}}{n_{ij}!}\exp\left(-\lambda_{ij}\right). \tag{3}$$

Hence, the likelihood function for $\boldsymbol{\lambda}=\left\{\lambda_{ij}, i, j=1,\ldots,R\right\}$ given $\mathbf{n}=\left\{n_{ij}, i, j=1,\ldots,R\right\}$ migrations, provided migrations are independent, is

$$L(\boldsymbol{\lambda};\mathbf{n})=P\left(N_{11}=n_{11}, N_{12}=n_{12},\ldots,N_{RR}=n_{RR}\right)=\prod_{ij}\frac{\lambda_{ij}^{n_{ij}}}{n_{ij}!}\exp\left(-\lambda_{ij}\right). \tag{4}$$

Entering the log-linear spatial interaction model (2) into the expression (4) and taking logarithms gives the log-likelihood function:

$$\begin{aligned} l(\boldsymbol{\alpha},\boldsymbol{\beta};\mathbf{n}) &= \sum_{ij}\left\{n_{ij}\log\left(\alpha_i\beta_j m_{ij}\right)-\alpha_i\beta_j m_{ij}-\log n_{ij}!\right\} \\ &= \sum_i n_{i+}\log\alpha_i+\sum_j n_{+j}\log\beta_j-\sum_{ij}\alpha_i\beta_j m_{ij}+c, \end{aligned} \tag{5}$$

where $\boldsymbol{\alpha}=\left\{\alpha_i, i=1,\ldots,R\right\}$, $\boldsymbol{\beta}=\left\{\beta_j, j=1,\ldots,R\right\}$ and $c=\sum_{ij}n_{ij}\log m_{ij}-\sum_{ij}\log n_{ij}!$. Here, $n_{i+}=\sum_j n_{ij}$ and $n_{+j}=\sum_i n_{ij}$ are the marginal totals and from (5) are sufficient statistics for α_i and β_j, respectively.

The maximum-likelihood estimates of α_i and β_j are obtained by maximising the log-likelihood function (5). The term *c*, which does not involve the parameters, may be ignored for this purpose. Hence, conveniently, only the marginal totals for the incomplete data and the auxiliary data, m_{ij}, are required to estimate the spatial interaction model.

Differentiation of the likelihood function with respect to each parameter gives the likelihood equations:

$$\frac{\partial l}{\partial\alpha_i}=\frac{n_{i+}}{\alpha_i}-\sum_j\beta_j m_{ij}=0 \tag{6}$$

and

$$\frac{\partial l}{\partial\beta_j}=\frac{n_{+j}}{\beta_j}-\sum_i\alpha_i m_{ij}=0. \tag{7}$$

These mean that the maximum likelihood estimators $\hat{\alpha}_i$ and $\hat{\beta}_j$ can be written

$$\hat{\alpha}_i=\frac{n_{i+}}{\sum_j\hat{\beta}_j m_{ij}} \tag{8}$$

and

$$\hat{\beta}_j=\frac{n_{+j}}{\sum_i\hat{\alpha}_i m_{ij}}. \tag{9}$$

Direct estimates of α_i and β_j cannot be obtained, since there are no closed-form expressions for the solution of equation (9) and (10). However, as described in Willekens (1999, pages 258-259), an iterative procedure can be used to obtained indirect estimates. Given initial estimates of β_j, equation (9) is used to obtain initial estimates of α_i. Equation (9) is then used to update the estimates of β_j. This process is repeated until convergence. This is conditional maximisation, also called stepwise ascent. Maximum likelihood estimates of λ_{ij} can be obtained using equation (1). Willekens (1999, page 259) discussed how this procedure is a special case of the iterative proportional fitting algorithm and the Expectation-Maximisation (EM) algorithm.

The above model focuses on estimating migration flows between two dimensions, origin and destination. We can expand this approach to include a third variable of interest not available, such as health status in the NHS migration data. Therefore, we consider a log-linear with offset form of the spatial interaction model

$$\log \lambda_{ijk} = \log \alpha_i + \log \beta_j + \log m_{ijk}\,, \tag{10}$$

where λ_{ijk} is the expected flows from origin *i* to destination *j* for level *k* of the third variable. The α_i and β_j parameters are related to the characteristics of the origin and destination respectively and m_{ijk} is the auxiliary information on migration flows. Note there are no parameters corresponding to the dimension indexed by *k* since it is assumed here that there is no information in the incomplete data set to estimate them. Instead we rely on the auxiliary data to provide the missing margin and association structure not contained in the incomplete data.

To maximise the log-likelihood of the spatial interaction model with respect to α_i and β_j we obtain the likelihood equations

$$\frac{\partial l}{\partial \alpha_i} = \frac{n_{i++}}{\alpha_i} - \sum_{jk} \beta_j m_{ijk} = 0 \tag{11}$$

and

$$\frac{\partial l}{\partial \beta_j} = \frac{n_{+j+}}{\beta_j} - \sum_{ik} \alpha_i m_{ijk} = 0\,, \tag{12}$$

which only require the marginal totals, n_{i++} and n_{+j+} of the incomplete data and the auxiliary m_{ijk}. The likelihood equations can be written as

$$\hat{\alpha}_i = \frac{n_{i++}}{\sum_{jk} \hat{\beta}_j m_{ijk}} \tag{13}$$

and

$$\hat{\beta}_j = \frac{n_{+j+}}{\sum_{ik} \hat{\alpha}_i m_{ijk}}\,, \tag{14}$$

which can be solved by iteration. Again this is a conditional maximisation of the likelihood function and converges to give estimates of α_i and β_j. Note that the choice of the initial values of β_j in (15) implicitly specifies the constraint required for parameter identification. Maximum likelihood estimates of λ_{ijk} can be obtained using equation (10).

Model (10) is the start point for our study. For example, when data on origin-destination two-way associations are available, a model of the form

$$\log \lambda_{ijk} = \log \alpha_{ij} + \log m_{ijk} \quad (15)$$

will be considered. Furthermore, we will also include auxiliary data available at two or more points in time to model the m_{ijk}, which will allow the association structure to evolve over time.

Attachment 2: Internal Migration Data from the NHS and 2001 Census

There are two main sources of internal migration data in England and Wales: (1) annual NHS data from the Central Register (NHSCR) and the Patient Register Data (PRD); and (2) decennial censuses. Both sources of data are organized and produced by the Office of National Statistics (ONS) and collect information on people's movements by origin, destination, age and sex at the local authority district level. The applicants have already access to the 2001 Census migration data, to NHSCR Tables (from 1991 to 2005) and to PRD Tables (from 2001 to 2005).

In England and Wales, a person must register with a local doctor in order to receive services. This implies that whenever a move takes place, the individual must register. 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 ONS which provides annual estimates on a quarterly base for the Central Register and once a year for the Patient Register. Estimates at local authority level are actually obtained by combining both sources. The registration data constitute a good up-to-date source of internal migration as nearly all residents in England and Wales 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). However, due to the sparseness found in the migration flow table, rounding adjustments are made to maintain confidentiality. Since errors due to rounding are compounded when aggregated to other levels, we will use the unadjusted inflows and outflows by age and sex readily available from the ONS website (www.statistics.gov.uk). For this study, the NHS migration data are lacking in two respects: reliable origin-destination-specific flows and other characteristics such as ethnicity, employment or skill level.

The most recent census in England and Wales took place on 29 April 2001. The census migration flows represent place of residence at the time of the census by place of residence one-year earlier. In particular, we will use the data contained in Table MG103 provided by ONS (2004), which includes flows between local authorities by sex and ethnic group.

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). NHS 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 NHS 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, implying efficiency in their comparability.

Both the census and NHS registration data have their weaknesses. Census estimates may be affected by under-enumeration, especially for younger age groups. Responses may be unrecorded because of either a non-statement of origin or incorrect completion of forms and misrepresentation of those usually resident on census night. The NHS database only records members of the population who have registered with a doctor. Migration patterns for certain groups that rarely visit a general practitioner, such as young male adults, may be misrepresented. On the other hand, it is very likely to capture movements of the elderly population, who are more dependent on health care. The greatest contrast between the data sources is the frequency of data collection. The census estimates provide a very fine level of detail but only for the year leading up to census night. The NHS data provides a continuous time series of migration information but without the detail available from the census. Note, however, the geography of these data between 1991 and 2000 is inconsistent with the post-2000 geography.

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.

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

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.
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 (2003) 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 (2003) 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 JCH (2004) Monitoring intercensal migration in the United Kingdom. *Environment and Planning A* 26(11):1711-1730.
Stillwell J (2005) Inter-regional migration modelling: A review and assessment. Paper presented at the 45th Congress of the European Regional Science Association, Vrije Universiteit, Amsterdam, 23-27 August.
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.
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.