WORKING PAPER 539

INTERNAL MIGRATION PROJECTION IN ENGLAND: THE OPCS/DOE MODEL EXAMINED

Peter Boden and John Stillwell
School of Geography
University of Leeds
Leeds LS2 9JT

Paper presented at the Joint Session of the British Section of the Regional Science Association and the Population Geography Study Group of the Institute of British Geographers on Migration Modelling: Macro and Micro Perspectives, at the British Section of the Regional Science Association Annual Conference at the University of Liverpool, 7 September, 1990.

Abstract

The OPCS population projections for subnational areas in England contain a set of net migration assumptions generated by a migration submodel. This paper sets out the equations relating to each stage of the modelling framework and examines three particular features of the model using data from the National Health Service Central Register. The results indicate the need for a systematic specification of broad age group boundaries and a mechanism for updating assignment probabilities. A different methodology to that used by OPCS for grouping local authority areas is also proposed.

1 Introduction

The Office of Population Censuses and Surveys (OPCS) is responsible for the production of subnational population projections in England using a demographic component methodology (Armitage 1986). Base year population estimates are available for 5 year age groups which can be approximately disaggregated to single year of age using the national age profile. Age-specific mortality rates are applied to these population estimates to generate the number of persons surviving to be one year older. Fertility rates are then applied to the female population aged 15-44 during the year to calculate the number of live births. Finally, the number of persons entering and leaving each subnational area during the year are taken into account in order to allow for population change due to migration.

In this paper, we are concerned to outline the sub-model that is used by OPCS, in collaboration with the Department of the Environment (DOE) for generating estimates of internal migration between local authority areas. Prior to 1981, the net migration assumptions used within the projection procedure were derived in a hierarchical manner with initial totals for the standard regions being used to constrain the assumptions for local authority areas (metropolitan districts, non-metropolitan counties, London boroughs).

The DOE, with its close links with planning authorities, was responsible for the initial regional estimates. Lengthy consultation between the DOE and individual local authorities was then necessary in order to establish appropriate net totals for each local area, taking account of the possible effect of local factors upon net migration. Decision making was not made any easier by the dearth of available migration statistics. Once the DOE had decided the net figures, an age and sex structure was allocated to the migration flows. The whole projection process was deemed to be too costly in terms of time spent in consultation with the planning authorities and in the considerable effort required by OPCS to produce the age breakdown of net migration flows by sex for each individual local authority. OPCS and DOE consequently sought to improve the migration projections by making better use of existing data and producing a first estimate of net migration which could be input to the consultation process.

In 1981, Martin and Voorhees Associates and John Bates Services were contracted by the DOE to develop an improved model for generating the net migration assumptions for the 108 English local authority areas concerned; the 36 metropolitan districts, 39 shire counties and 33 London boroughs. The procedure which they proposed for the production of net migration estimates disaggregated by sex and single year of age is documented in Martin, Voorhees and Bates (1981), and has been used to produce sets of projections from base years 1981, 1983 and 1985 (OPCS, 1983; 1986; 1988). The methodology for producing projections for local authority areas (not the recent extension to health authority areas) is described in detail in Section 2 of the paper. In Section 3, certain features of the model are investigated using data from the National Health Service Central Register (NHSCR) on patient reregistrations between Family Practitioner Committee (FPC) areas in England and Wales. Firstly, a clustering procedure is used to show how the definition of the broad age groups adopted by OPCS in their modelling procedures might be improved. Secondly, changes in the distribution pattern of migration since 1980-81 are identified which call into question that part of the modelling procedure which uses probabilities based on 1981 Census data to assign projected outmigrants to destinations. Finally, a different method of grouping local authority areas is proposed which generates clusters that differ from those currently used in the OPCS/DOE methodology. Conclusions are drawn in the final section of the paper.

2 The OPCS/DOE migration projection methodology

The starting point for each round of projections is the set of mid-year population estimates for males and females by 5 year age group. These estimates, prepared by OPCS (1982, 1984, 1986) for each local authority area, include all persons usually resident in local government areas of England, together with Armed Forces personnel stationed in each area and students counted as residents at their term-time addresses.

There are four main stages in the methodology for generating internal migration estimates:

(i) the projection of migration flows out of each local authority area by age and sex;

(ii) the assignment of these outflows by broad age group to individual destinations;

(iii) the aggregation of these flows to provide area inmigration totals and their disaggregation by age and sex; and

(iv) the calculation of net migration assumptions from age-specific outflow and inflow totals for males and females.

We can explain the methodology in more detail by setting out the equations associated with each stage, following Rees and Willekens (1989).

Outmigration projection

The first stage of the migration projection procedure involves projecting age-specific outflows from each individual area for males and females for a given time period as the product of the mid-year population estimate or projection for the area, the area's gross migraproduction rate (gmr_i) and the proportion of this rate accounted for by a particular age and sex group. The model equation takes the form:

$$). \ om_I^{as} \ (r). \ P_i^{as} \ (t_0) \qquad (1)$$

where:

$M^{as}_{i*}(t)$ = the total number of outmigrations from area i by single year of age a and sex s for year t;

$gmr_i(t)$ = the gross migraproduction rate from area i in year t;

$om^{as}_I(r)$ = the proportion of the migraproduction rate in single year age group a and sex s (derived from modelled, standardized migration rates for area cluster I for a standard period r, 1980-81); and

$P^{as}_i(t_0)$ = the population of area i at single year of age a and sex s at t , the beginning of year t.

Current projections adopt 1981 Census gross migraproduction rates adjusted for underenumeration and modified in the light of changes evident from NHSCR movement data. The Census gmr for a local authority area is simply trended parallel to that of a NHSCR total outmigration rate using the slope of a linear regression through an annual time series of NHSCR patient re-registration data for each sex. The outmigration proportions have been derived from the age-specific profiles of clusters of areas. To reduce the data requirements of the model (i.e. to avoid having to use and store male and female age-specific outmigration rates for each area), a classification of area outmigration profiles was derived based on similarities between individual model migration schedules calibrated initially on 1971 Census data (Martin, Voorhees and Bates 1981, Bates and Bracken 1982, Bracken and Bates 1983) and subsequently on 1981 Census data (Bates 1984, Bates and Bracken 1987). Using the techniques developed by Rogers et al. (1978), model migration schedules were fitted to observed outmigration rate profiles for each area. The parameters describing the model schedules were used to compare areas and group them into categories based on profile similarities. A classification of seven groups emerged with the majority of English local authorities falling into one of three main groups containing respectively, most of the shire counties, the London boroughs and the metropolitan districts (Table 1). Subdivisions of groups 1 (shire counties) and 4 (home counties) were suggested in order to distinguish areas where a retirement effect was evident in the migration profile from areas where this effect was not present. The same grouping was used for females.

<u>Assignment</u>

The second stage of the projection procedure involves the assignment of the estimated outmigration totals to individual area destination areas. Because of the problem of handling matrices containing large numbers of small or zero flows, single year of age outmigrants are grouped into three broad age bands (0-16/29-59, 17-28 and 60+) which correspond with different types of migratory movement.

Table 1: Summary of English local authority groupings based on 1981 male outmigration profiles

Group	Area type	Number of areas
1A	Shire counties (most)	27
1B	Shire counties (remainder)	4
2	London boroughs (most)	28
3	Metropolitan districts (most)	34
4A	Home counties A	4
4B	Home counties B (remainder)	3
5	Outer London boroughs (remainder)	4
6	Lancashire and Bradford	2
7	Liverpool	1

Source: Bates (1984)

The first age group is composed of migrant families; the second involves moves about the time of entry to the labour force when young people frequently leave home; and the third group is associated with retirement and elderly migration. Assignment probabilities are then used to allocate outmigrants to destinations. The assignment submodel can be written as:

$$M_{ij}^{As}(t) = M_{i*}^{As}(t) \;.\; k_{ij}^{As}(r) \qquad (2)$$

where:

$M_{ij}^{As}(t)$ = migration flows from area i to area j in broad age group A and sex s during year t;

$$M_{i*}^{As}(t) = \sum_{a \epsilon A} M_{i*}^{as}(t) \qquad (3)$$

= total out migration from area i in broad age group A and sex s during year t; and

$k_{ij}^{As}(r)$ = the proportion of outmigrants from area i in broad age group A and of sex s which moved to destination area j in a standard period, r.

1981 Census data are currently used as the basis for assigning outflows to destinations and so the period r refers to the 12 months before the Census (1980-81). No updating of the inter-area probability information has been undertaken and therefore the distribution patterns of migration are assumed to have remained constant since 1980-81.

Inmigration projection

The product of the second stage in the procedure is a series of migration inflows to individual areas disaggregated by broad age group and sex. The third component of the projection model aggregates these flows to produce total migration and then disaggregates the totals to provide flows by single year of age and sex. The inflow model can be written as:

$$M_{*j}^{as}(t) = \sum_{A}\sum_{i} M_{ij}^{As}(t) . \; im_{J}^{as}(r) \tag{4}$$

where:

$M_{*j}^{as}(t)$ = the total number of inmigrants to area j by single year of age a and sex s in year t;

$\sum_{A}\sum_{i} M_{ij}^{As}(t)$ = the aggregation of inter-area flows by broad age group and sex to provide total inmigration to area by sex in year t; and

$im_{J}^{as}(r)$ = the proportion of inmigration gmr in single year age group a and sex s (derived from modelled, standardized migration rates for area cluster J for a standard period, r).

The im values are derived from area clusters based on inmigration data from the 1981 Census. The classification of areas on the basis of their inmigration profiles in 1980-81 is more complex than the outmigration grouping illustrated earlier: eleven groupings with distinctive profiles are identified by Bates (1984). Since each area is associated with a particular group profile, area inflow totals are disaggregated into single years of age according to proportions evident from the standardized profiles of the cluster to which that area belongs.

Net migration calculation

At this stage of the procedure, gross outflows and inflows disaggregated by single year of age and sex have been projected for each area. From these it is possible to compute net migration assumptions as:

$$N_i^{as}(t) = M_{*i}^{as}(t) - M_{i*}^{as}(t) \tag{5}$$

where:

$N_i^{as}(t)$ = net migration for area i by single year of age a and sex s in year t;

$M_{*i}^{as}(t)$ = projected gross inmigration to area i by age a and sex s in year t; and

$M_{i*}^{as}(t)$ = projected gross outmigration from area i by age a and sex s in year t.

These net migration assumptions form part of the OPCS model to produce subnational population projections. This model has certain facilities to amend the gross migraproduction rate of any area and to input maximum or minimum values of net migration as constraints for a few specified individual areas or groups of areas.

Consultation

The generation of projections by OPCS is followed by an exercise in which the DOE consults the local authorities over the in-, out- and net migration projections and local authorities are able to respond by stating the factors that they believe will result in figures different from the trend-based assumptions (e.g. house building or clearance programmes). As a result of these consultations, changes are made to the migration assumptions and the population projections before the procedure is finalized.

3 Three features of the model examined

There is little doubt that the procedure developed by Martin, Voorhees and Bates has provided a much improved framework for generating projections. The derivation of migration assumptions is based essentially on outmigration and inmigration age profiles and inter-area assignment matrices for the 1980-81 pre-Census period, with the Census gross migraproduction rate trended using more recent NHSCR mid-year to mid-year movement data. However, as the length of time since 1981 increases, there is inevitably a growing concern about the changes in age structure and spatial distribution of migration that may have occurred during the last decade. There is no mechanism for their incorporation in the existing OPCS/DOE model structure, other than through the gross migraproduction rate.

In the remainder of this paper we therefore use data from the NHSCR to examine three selected questions relating to the OPCS/DOE model. It should be recognized that the FPC areas do not fully correspond with the local authority areas for which the OPCS makes its projections although this inconsistency is only important in Greater London. Furthermore, despite OPCS projections only being generated for England, the ready availability of NHSCR data on movements involving Welsh FPC areas suggests that Wales be included in the analysis, as it was in the earlier work of Martin Voorhees and Bates (1981), Bates and Bracken (1982) and Bracken and Bates (1983). The reliability of the NHSCR data and its precise relationship with Census transition data for 1980-81 is discussed in Devis and Mills (1986) and in Boden, et al. (1987; 1988).

The first of three questions concerns the validity of the broad age groups defined for use in the model. Whilst the broad age groups that are used in the assignment stage are deemed to encapsulate the major components of migration: family movement, entry to the labour force and retirement migration, little justification is given by OPCS/DOE for adopting these precise age group boundaries. An alternative broad age group classification is therefore derived which is based on age-specific inter-area movement occurring between FPC areas in England and Wales between 1984 and 1986.

The second question we address is concerned with what changes in the spatial patterns of migration have taken place since 1980-81, given the implicit assumption of the OPCS/DOE model that the distribution of migrants has remained constant. Mid-year to mid-year NHSCR data for 1985-86 is used to explore the extent of change since 1980-81 in gross and net migration for aggregate and selected age-specific groups of movers.

The third question focuses on the classification of local authority areas into groupings on the basis of their age profiles. An alternative categorization of outmigration and inmigration areas to that of Bates and Bracken (1987) is proposed which uses a clustering technique based on each area's profile of observed age-specific rates rather than the parameters of each area's model migration schedule.

Systematic derivation of broad age groups

The three broad age categories used in the assignment stage of the OPCS/DOE model have been chosen to reflect the major components of migration. No systematic methodology is provided for the precise definition of the broad age group boundaries. However, cluster analysis procedures are available in several statistical packages which can be used to derive optimum groupings from age-disaggregated information. In this context, a clustering procedure in SPSSX (Norusis 1985) is used to generate optimum broad age groups. The analysis should be conducted, ideally, using single year of age data, but the size of the array presents severe storage problems and consequently five year age group inter-FPC area movement matrices for 1984-85 and 1985-86 are used.

The initial step is to compute 'distances' between the values in corresponding cells of the origin-destination matrices for each age group. In order to remove the effect of the level of migration upon the clustering process, the flows are standardized by expressing each value as a proportion of the total movement between the two 'areas'. The Squared Euclidian Distance (SED) between two 'cases' (age groups a and b) is defined as:

$$SED^{ab} = \sum_{t}\sum_{i}\sum_{j}\left(M^{a}_{ij}(t)/M^{*}_{ij}(t) - M^{b}_{ij}(t)/M^{*}_{ij}(t)\right)^{2} \qquad (6)$$

where:

$M^{a}_{ij}(t)$, $M^{b}_{ij}(t)$ = the number of moves between origin area i and destination area j in age group a (b) in year t; and

$M^{*}_{ij}(t)$ = the total number of moves between origin area i and destination area j in year t.

The computation of the distance array is a precursor to the actual clustering procedure which is based on the 'average linkage between groups' method. The distance between two clusters is the average of the distances between all pairs of age groups in which one member of the pair is from each of the clusters. Thus the average linkage value or the 'distance coefficient' (D) between clusters A and

B is defined as:

$$D^{AB} = \sum_{a\epsilon A} \sum_{b\epsilon B} SED^{ab} / n^{A} n^{B} \tag{7}$$

where n^A, n^B = the number of age groups in clusters A and B respectively.

The process is agglomerative, commencing with the maximum number of clusters (16 age groups), and then combining those cases where the distance coefficient is smallest on each iteration, until all cases are classified in one cluster. The agglomeration schedules for the two mid-year to mid-year periods (Figure 1) indicate how the distance coefficient increases as the number of clusters decreases. The D value increases steadily until the point at which four clusters become three (stage 13), whereupon the distance coefficient increases more significantly. In Table 2, the three and four cluster solutions, which are the same for both time periods, are compared with the broad age categories used by OPCS/DOE. Although this comparison is complicated by the use of 5 year age groups and the variation by age which occurs in the ratio of Census to NHSCR migration data as reported in Boden et al. (1988), two important differences can be identified. Firstly, the results show that the inclusion of moves by persons in their late 20s with those in their late teens and early 20s is not appropriate. Although it is not feasible to suggest that 15 is a better break point than 17, the results do indicate that 29 is too

Figure 1: Agglomeration schedules from the clustering of five year age groups based on NHSCR inter-area migration data

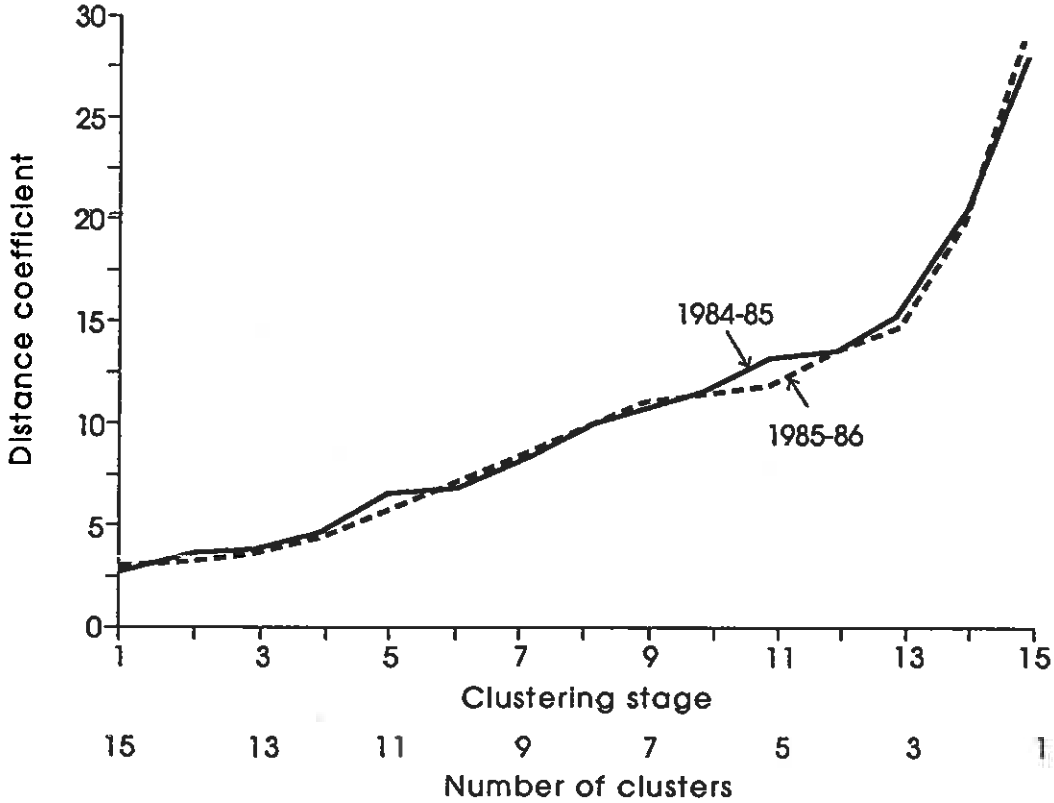


Table 2: OPCS/DOE broad age groups compared with classifications derived from clustering of five year age groups

OPCS/DOE groups	3-cluster solution	4-cluster solution
0-16/29-59	0-14/25-54	0-14/25-54
17-28	15-24	15-24
60+	55+	55-69
		70+

high a break point for the category defined by OPCS to include moves around the time of entry to the labour force, when young people frequently leave the parental home. The 25-29 age group combines readily (ie. at the second stage) with the 30-34 age group in the clustering procedure and therefore 24 is a more appropriate upper age limit for this most highly mobile group. Secondly, the inclusion of moves by persons in their late 50s with retirement moves and post-retirement moves is more appropriate than combining them with the 'family' moves of those aged 25-54. The 55-59 age group combines most readily with the 60-69 age group and then subsequently, with the 70+ age groups at the next stage. More research is clearly required in this context using migration data disaggregated by single year of age and by sex to establish the boundaries precisely.

Changing migration patterns

In the assignment stage of the OPCS/DOE model, it is assumed that the distribution pattern of migration by broad age group in 1980-81 has been continued in subsequent years. The validity of this assumption can be assessed by comparing the NHSCR matrix of moves between FPC areas in England and Wales for the 12 month period which most closely approximates to the year before the 1981 Census (1 April, 1980 to 31 March, 1981) with the corresponding matrix for mid-year, 1985 to mid-year 1986. Two goodness-of-fit statistics, the index of dissimilarity and the correlation coefficient, provide measures of association between the patterns of inter-FPC area migration for six age groups (Table 3) in addition to the aggregate flows.

Dissimilarity is most significant and correlation weakest for the post-retirement age group, suggesting a quite substantial alteration in their pattern of movement. The highest rates of gross inmigration for the 70+ group are found in the FPC areas which constitute the South East (outside Greater London), the South West and East Anglia, whereas the highest gross outmigration rates are found to occur from Greater London boroughs. The observed rates for this age group are relatively small in comparison with other age groups and so their percentage change between 1980-81 and 1985-86 is more marked. The map of changing inmigration rates (Figure 2) shows the majority of FPC areas experiencing increased rates during the period. A number of metropolitan FPC areas in Greater Manchester, Merseyside and the West Midlands showed appreciably higher inmigration rates in 1985-86 but the more important increases occurred in the non-metropolitan counties such as Devon, Cornwall, Bedfordshire, Kent, Dyfed and Powys. A number of Greater London boroughs experienced declines in both the rate of inmigration and outmigration (Figure 3), but the most significant decreases in outmigration rates occurred from more rural areas including Hampshire, East Sussex, Kent, Lincolnshire, Lancashire, Durham, Scotland and the two Welsh FPC areas of Mid and South Glamorgan. As with inmigration, most FPC areas experienced outmigration rate increases.

Table 3: Goodness-of-fit statistics comparing inter-FPC area movement by age group, 1980-81 with 1985-86

Age group	Index of dissimilarity	Correlation coefficient
0-14	20.1	0.950
15-19	24.7	0.908
20-24	19.1	0.953
25-54	14.0	0.977
55-69	27.2	0.926
70+	33.5	0.891
All ages	10.5	0.982

Note: The index of dissimilarity compares the two distributions by calculating the sum of the deviations between cell proportions in the two matrices

Figure 2: Percentage change in inmigration rates of 70+ age group, 1980-81 to 1985-86

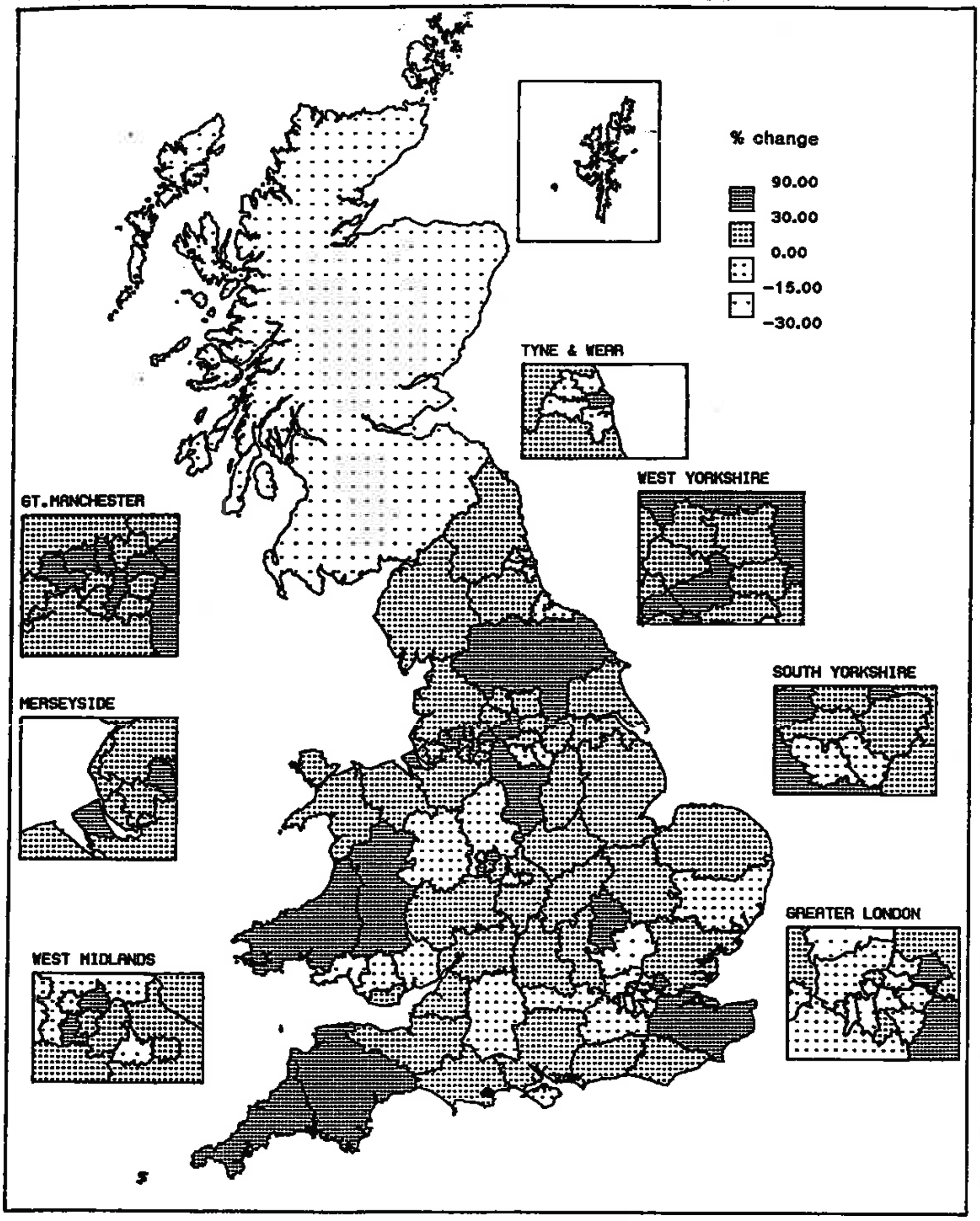
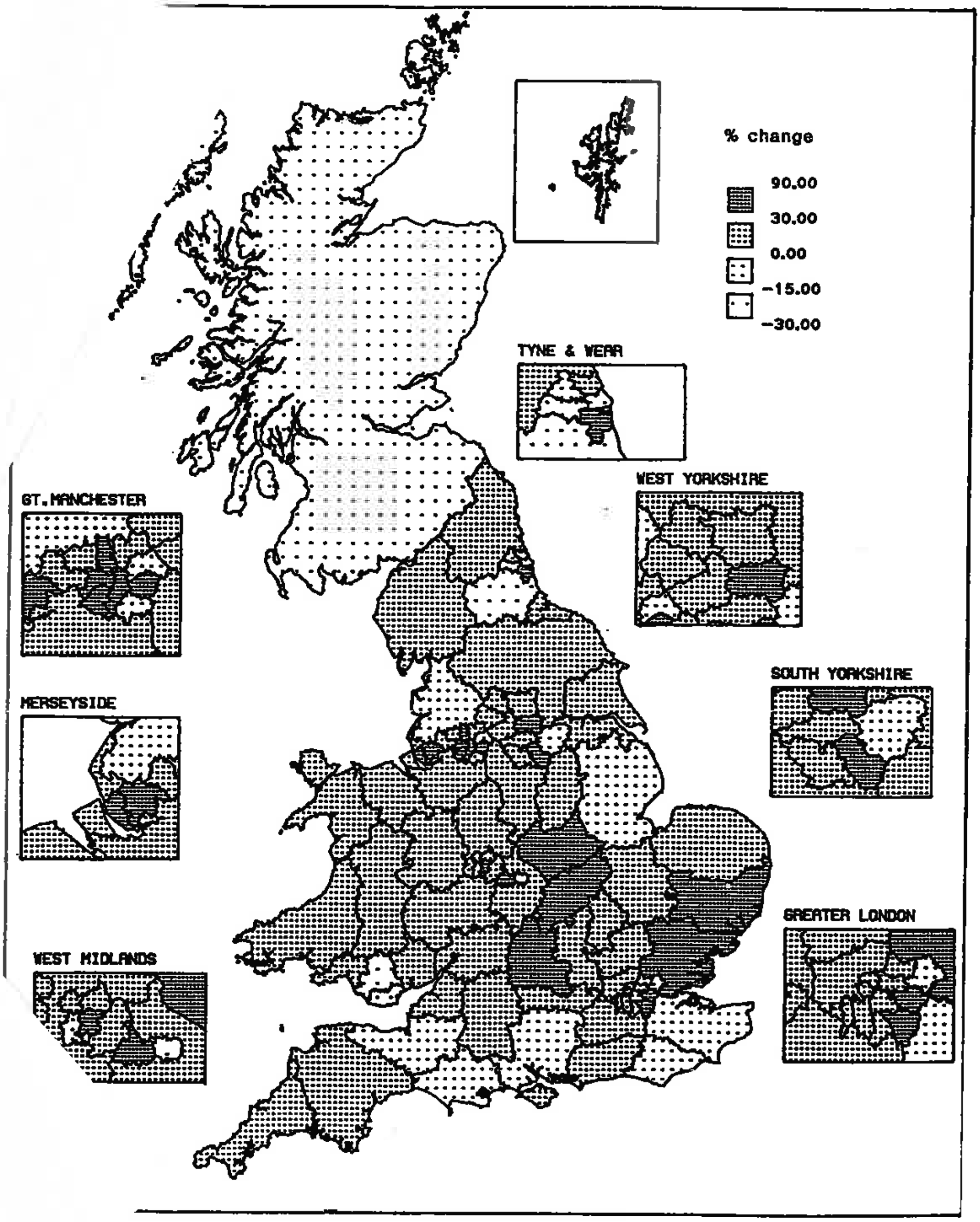


Figure 3: Percentage change in outmigration rates of 70+ age group, 1980-81 and 1985-86



The level and direction of migration by those age 55-69 and 15-19 also appear to have undergone quite substantial change between 1980-81 and 1985-86, according to the goodness-of-fit statistics, with age groups 20-24 and 25-54 exhibiting most consistency. It is, however, important to acknowledge the relative volumes of the flows involved. The 20-24 age group comprised almost 20% of total inter-area migration in both periods. The changing distribution of migration for this age group is summarized in the net migration patterns illustrated in Figure 4. By 1985-86, most of the FPC areas outside the South East and East Anglia had a negative rate of net migration with only a limited number of exceptions (in Greater Manchester, for example). As in 1980-81, most of the counties in the South West experienced rates of net loss, but the pattern elsewhere in the South was one of net migration gains in this age group, in Greater London in particular. Thus, the general movement of this most mobile age group has become increasingly directed towards the high density areas of the South at the expense of the remainder of the UK.

A classification of FPC areas into high, medium-high, medium-low and low density groups (see Stillwell, et al. 1990) provides a framework for summarizing the changes in migration rates by age group to and from the high density areas of the South. The rate of movement from these areas to high density areas of the North increased between 1980-81 and 1985-86 for all but the 0-14 age group (Figure 5). In

Figure 4: Net migration rates of 20-24 age group for 1980-81 and 1985-86

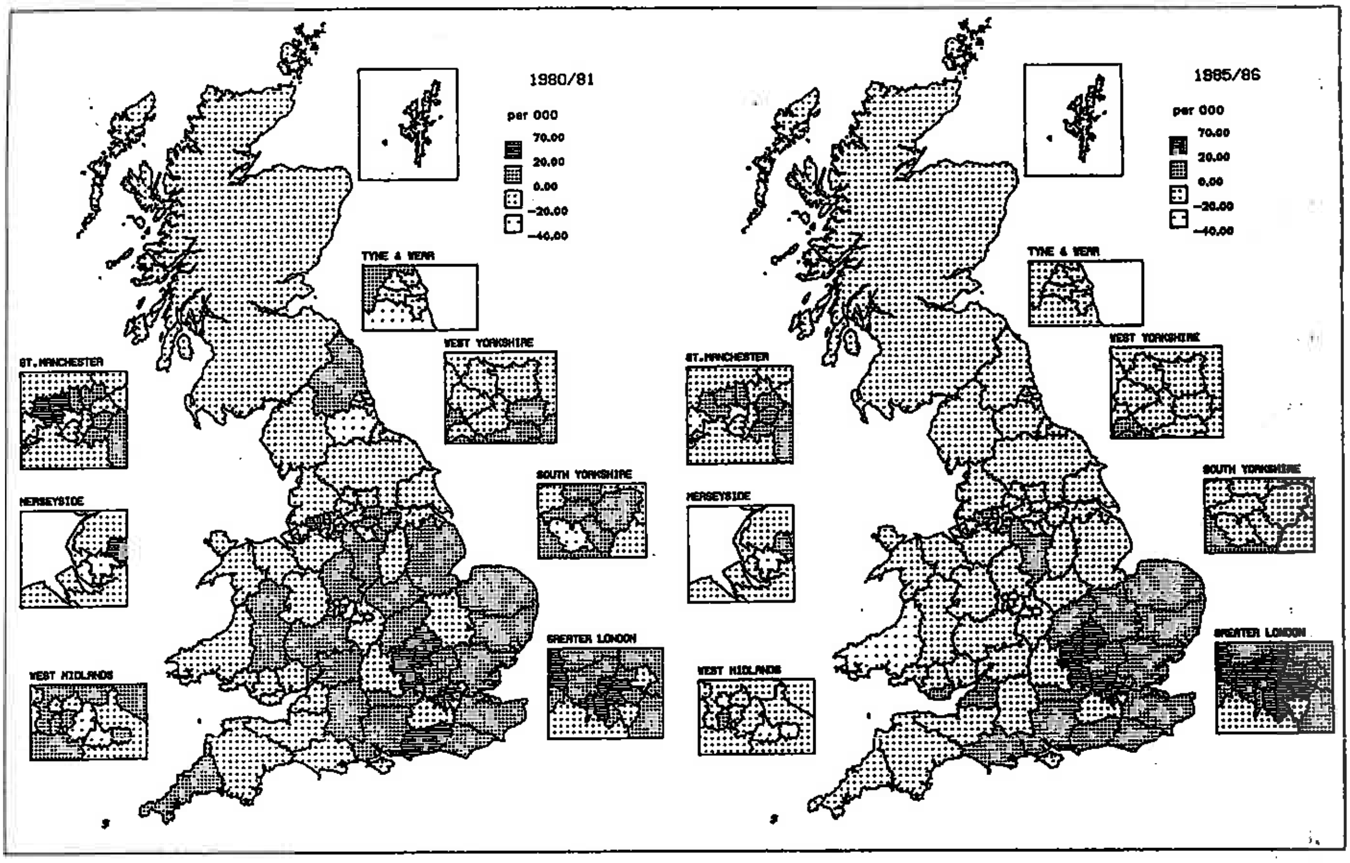
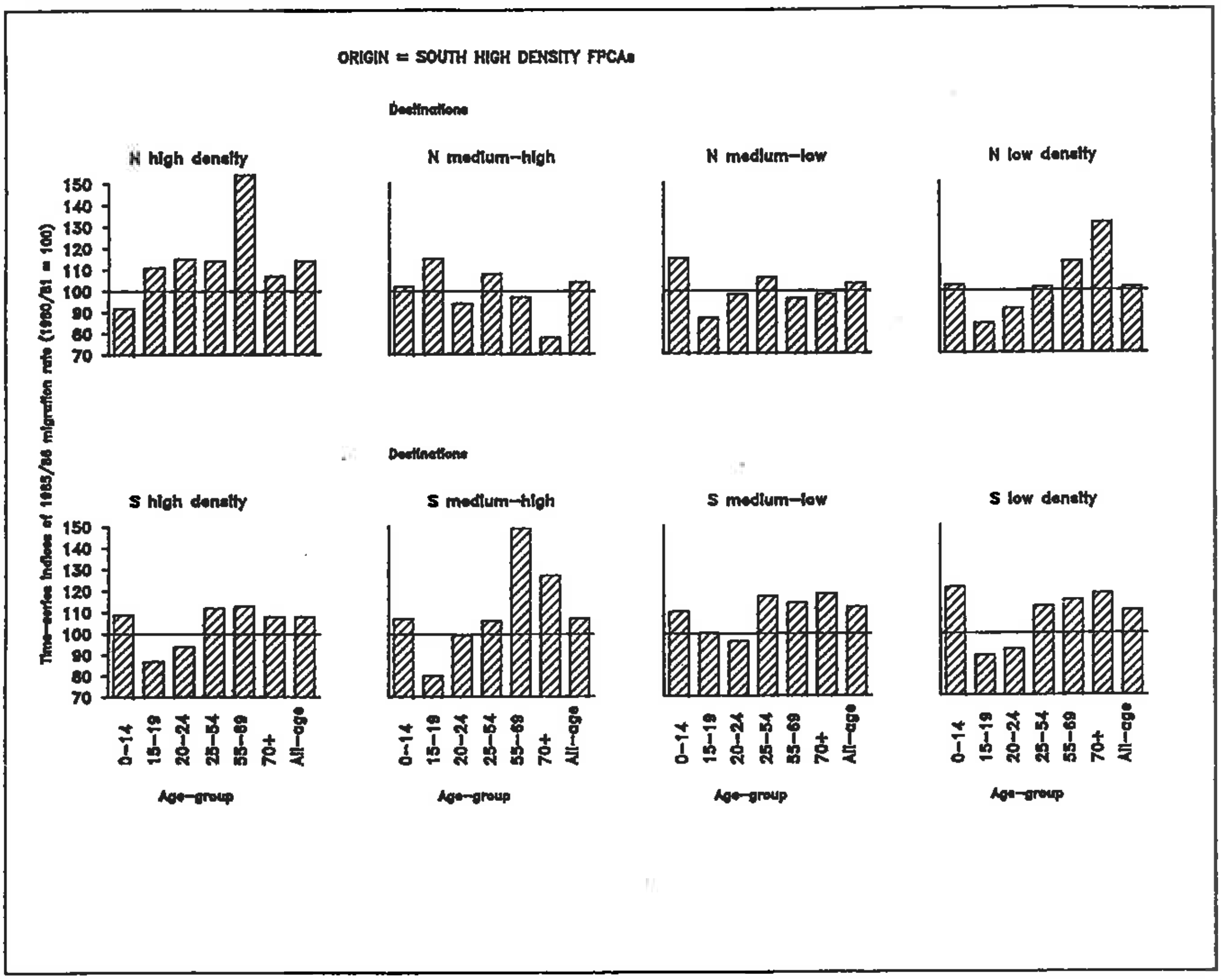
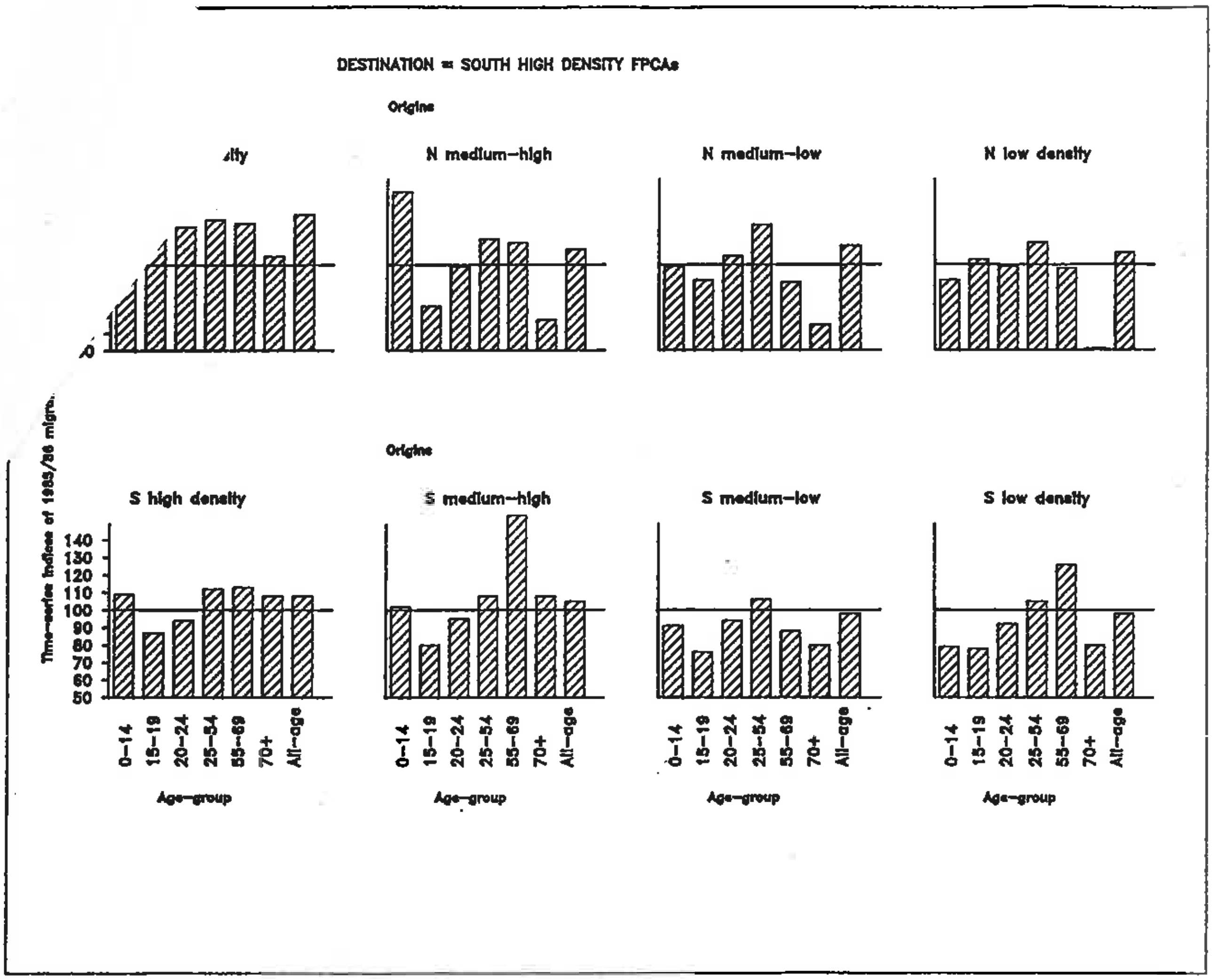


Figure 5: Changes in migration rates from Southern high density areas, 1980-81 to 1985-86



contrast, the rate of retirement and post-retirement movement to the two medium density area categories of the North decreased, whilst, at the same time, low density areas in the North become more attractive to migrants from Greater London in these two age groups. The importance of family moves and retirement and post-retirement moves out of Greater London to the more rural areas of the South East, East Anglia, South West and East Midlands increased, emphasizing the preference of these groups for residence away from Greater London. This contrasts with the patterns of the 15-19 and 20-24 age groups whose rate of movement out of high density to lower density FPC areas decreased significantly up to 1985-86. Figure 6 illustrates the changes in rates of outmigration from other areas to the high density FPC areas of Greater London. Movement rates from Northern high density areas increased in all age groups: there was a significant reduction in the rate of movement of those aged 70+ from Northern medium and low density areas; a decline also occurred in the rate of movement of 15-19 year olds to the capital from medium density areas of the North; the rate of movement in the early age groups from lower density areas in the South declined whereas inmovement by 25-54 year olds from low density areas increased marginally and inmovement by 55-69 year olds, paradoxically, increased from both the medium-high and the low density areas. These illustrations suggest that some marked changes have occurred in migration distribution patterns since 1980-81. The OPCS/DOE model would benefit from a method which incorporated

Figure 6: Changes in migration rates to Southern high density areas, 1980-81 to 1985-86



a procedure for updating or trending the assignment probabilities.

Migration profile classification

The groupings of local authority areas in stages (i) and (iii) of the OPCS/DOE model outlined in Section 2 are determined on the basis of the similarity of the parameters of model schedules. The methodology requires the calibration of model migration schedules for the standardized outmigration and inmigration rate profiles of each area in 1980-81 (Bracken and Bates 1987), and therefore requires prior selection of a model for each area either with or without a retirement component. An alternative classification method is suggested here which derives groupings of areas on the basis of observed single year of age migration rate schedules. Using the observed schedules removes the difficulty of identifying whether or not a retirement component is present, and retains more information than that based on smoothed profiles. Age-specific rates within each cluster are then aggregated and a single cluster model migration schedule is calibrated.

The clustering technique used to group the standardized migration rates is similar in principle to that outlined in a previous section. NHSCR data on gross outmigration and inmigration by single year of age for FPC areas in England and Wales in 1985-86 are used, and an age range of 1-79 is adopted because small numbers beyond age 79 lead to large fluctuations in rates. The initial step in the process is the computation of the Squared Euclidian Distance (SED) between two FPC areas i and j by direction d (out or in) as:

$$SED^d_{ij} = \sum_a (m_i^{ad} - m_j^{ad})^2 \tag{8}$$

where m_i^{ad}, m_j^{ad} = the standardized out or inmigration rates for areas i and j by single year of age a.

The 'average linkage between groups' method is again used where the distance (D) between two clusters, I and J, is the average distance between all pairs of FPC areas in which one member of the pair is from each of the clusters:

$$D_{IJ} = \sum_{i \in I} \sum_{j \in J} SED_{ij}/n_I n_J \tag{9}$$

where n_I and n_J = the number of areas in clusters I and J respectively.

The agglomeration schedules produced in the clustering of outmigration and inmigration profiles (Figure 7) indicate that the increase in the distance coefficient is fairly constant until around stages 75-80 of the procedure (15-20 clusters). After this point, increases become larger and less uniform indicating that the association of areas becomes more irregular. A 'break' in the series can be identified as a point at which an optimum number of clusters has been reached and Tables 4 and 5 outline the FPC area composition of 15 clusters for inmigration and outmigration respectively. The age-specific migration

Figure 7: Agglomeration schedules from clustering out- and inmigration rate schedules, 1985-86

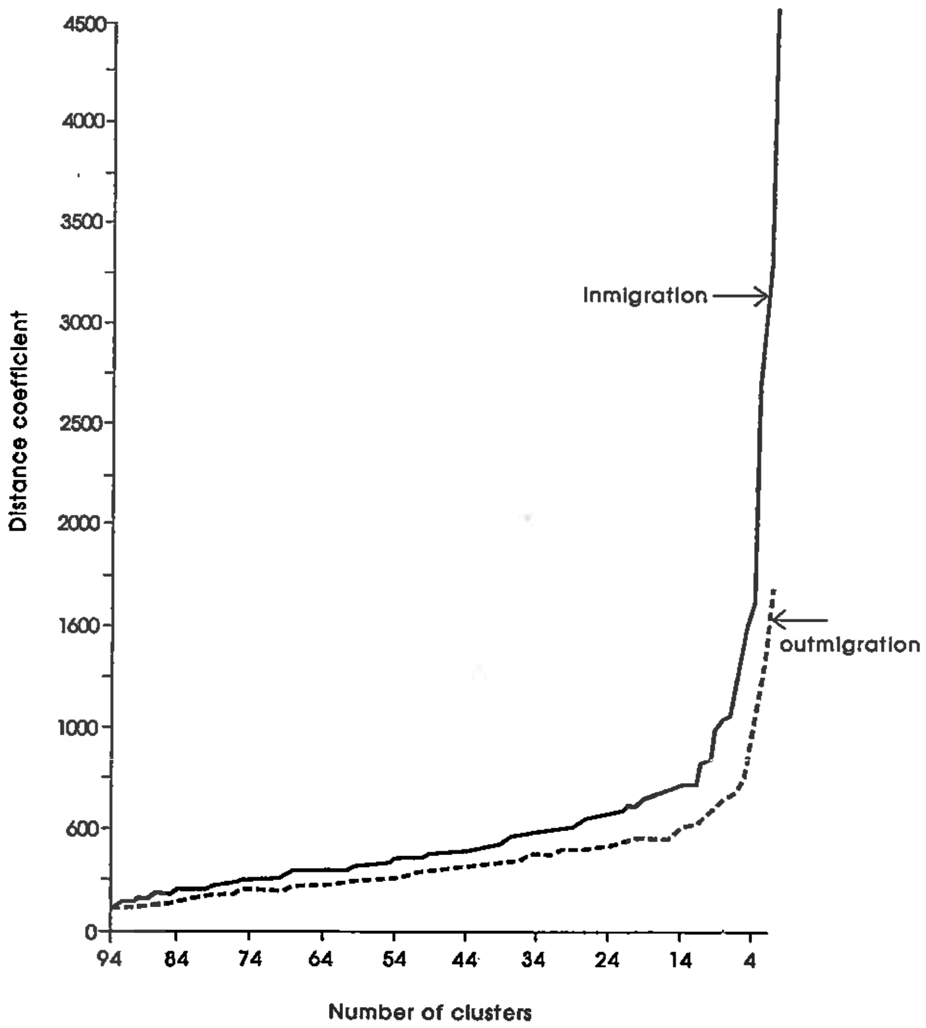


Table 4: 15-cluster stage of FPC area classification process for 1985-86 immigration

1. Camden and Islington; Kensington, Chelsea and Westminster

2. Redbridge and Waltham Forest; Croydon; Greenwich and Bexley

3. City and East London; Kingston and Richmond; Merton, Sutton and Wandsworth; Lambeth, Southwark and Lewisham; Middlesex

4. Lincolnshire; Suffolk; Isle of Wight; W. Sussex; Cornwall; Dorset; Somerset; Hereford; Salop; Clwyd; Powys

5. Humberside; Norfolk; E. Sussex; Devon; Lancashire; Dyfed; Gwynedd

6. Cumbria; Northumberland; Rotherham; Calderdale; Wakefield; Derbyshire; Northamptonshire; Buckinghamshire; Essex; Kent; Gloucestershire; Wiltshire; Warwickhire; Bolton; Oldham; Rochdale; Tameside; Sefton; Wirral; Cheshire; Gwent

7. Gateshead; N. Tyneside; Barnsley; Bedfordshire; Hertfordshire; Berkshire; Barking and Havering; Bromley; Dudley; Sandwell; Walsall; Bury; Stockport; Trafford; Wigan; St. Helens

8. Durham; Leicestershire; Nottinghamshire; Oxfordshire; Avon; Birmingham; Wolverhampton; Manchester; Salford; S. Glamorgan

9. Newcastle; Sheffield; Coventry

10. Leeds; Liverpool

11. Sunderland; Cleveland; Bradford; Kirklees; Cambridgeshire; Staffordshire; Mid-Glamorgan

12. Scotland; Doncaster; N. Yorkshire; Hampshire

13. S. Tyneside

14. Solihull

15. W. Glamorgan

Note: Middlesex represents a group of five FPC areas

Table 5: 15-cluster stage of FPC area classification process for 1985-86 outmigration

1.	City and East-London; Redbridge and Watham Forest; Barking and Havering; Merton, Sutton and Wandsworth; Croydon; Lambeth, Southwark and Lewisham; Bromley; Greenwich and Bexley; Middlesex
2.	Newcastle; Sheffield; Leeds; Coventry; Wolverhampton; S. Glamorgan
3.	Cambridgeshire; Oxfordshire; Surrey; Kingston and Richmond; Birmingham; Manchester; Salford; Liverpool; E. Sussex
4.	Bradford; Bedfordshire; Buckinghamshire; Essex; Hertfordshire; Berkshire; Hampshire; Rochdale; Tameside
5.	Durham; Leicestershire; Nottinghamshire; Avon; Devon; Dyfed; W. Glamorgan
6.	Northumberland; Barnsley; Docaster; Rotherham; Calderdale; Wakefield; Lincolnshire; Northants; Suffolk; W. Sussex; Dudley Solihull; Warwickshire; Boldton; Bury; Oldham; Stockport; Trafford; Wigan; Wiltshire
7.	Cleveland; Cumbria; Kirklees; N. Yorkshire; Derbyshire; Norfolk; Cornwall; Dorset; Gloucestershire; Somerset; Hereford; Salop; Staffordshire; Sefton; Wirral; Cheshire; Lancashire; Clwyd; Gwent; Mid-Glamorgan
8.	N. Tyneside; Walsall; St. Helens; Sunderland
9.	Scotland; Gateshead
10.	Camden and Islington; Kensington, Chelsea and Westminster
11.	S. Tyneside
12.	Isle of Wight
13.	Sandwell
14.	Gwynedd
15.	Powys

Note: Middlesex represents a group of 5 FPC areas rates for areas within each cluster are aggregated to form a 'cluster profile'. Finally, characteristic features of profiles can be quantified by calibrating a model migration schedule for each cluster using a version of the MODEL package developed by Rogers and Planck (1984) and operationalized at Leeds by Stillwell et al. (1987). The MODEL package requires pre-selection of a schedule for each cluster with either (i) a retirement peak, (ii) a retirement slope or (iii) no retirement component. The full model schedule with a retirement component is defined as:

$$\begin{aligned} m_I^a = \; & b_1 \exp(-\alpha_1 a) \\ & + b_2 \exp\{-\alpha_2 (a - \mu_2) - \exp(-\lambda_2 (a - \mu_2))\} \\ & + b_3 \exp\{-\alpha_3 (a - \mu_3) - \exp(-\lambda_3 (a - \mu_3))\} \\ & + c \end{aligned} \qquad (10)$$

where m_I^a is the rate of migration into or out of one cluster I of those aged a, and where the profile of the schedule is defined by seven of the eleven parameters (α_1, α_2, μ_2, λ_2, α_3, μ_3, λ_3) and the level of the schedule is determined by the remaining parameters (b_1, b_2, b_3, c). This model is similar to that used by OPCS. A complete model schedule was only fitted to certain profiles (Figures 8 and 9) after careful inspection of the observed rates. A seven parameter model was fitted in other cases. The goodness-of-fit of each model schedule is measured using the E statistic following Rogers and Castro

Figure 8: Observed and estimated schedules for inmigration clusters, 1986-86

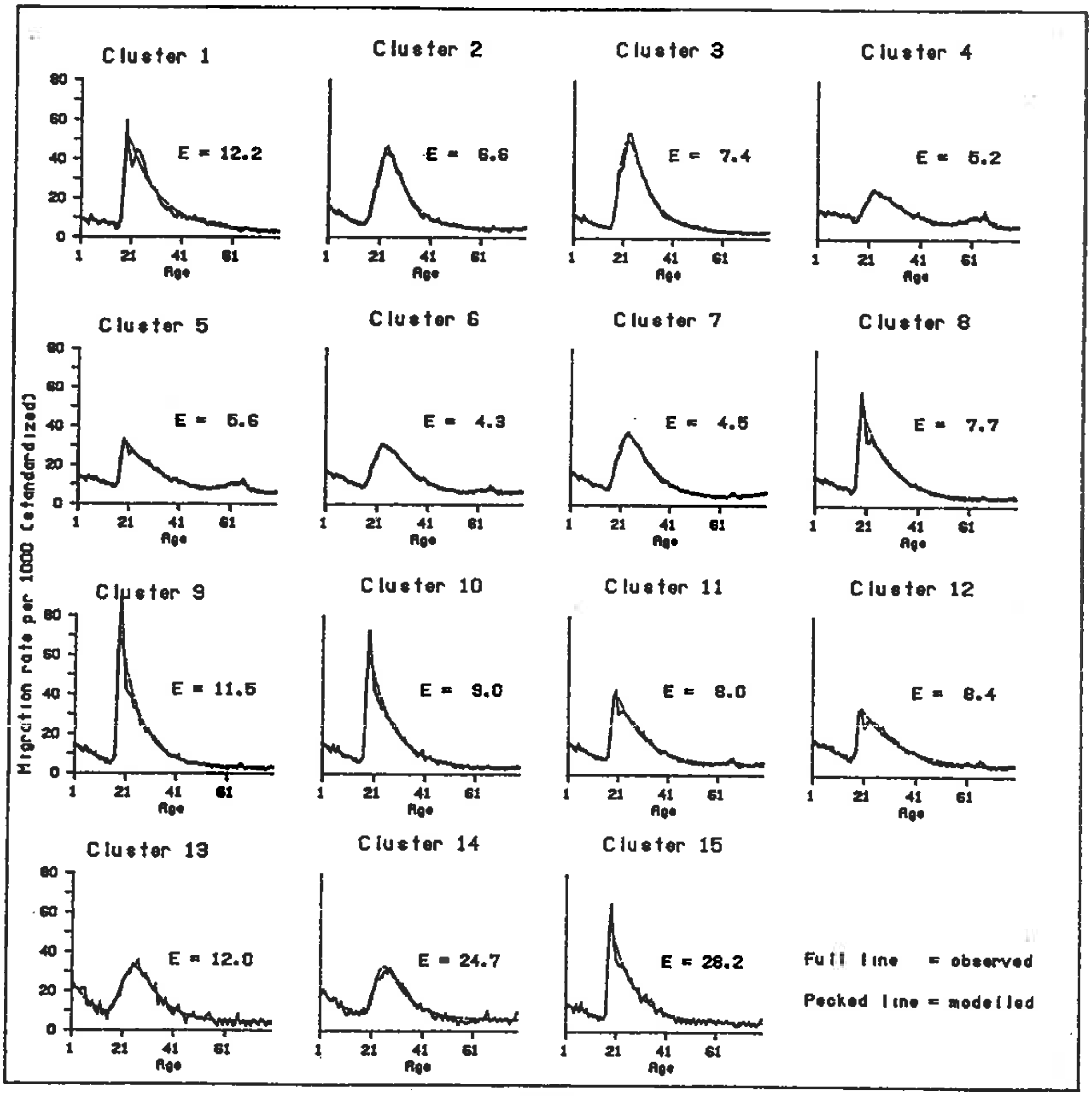
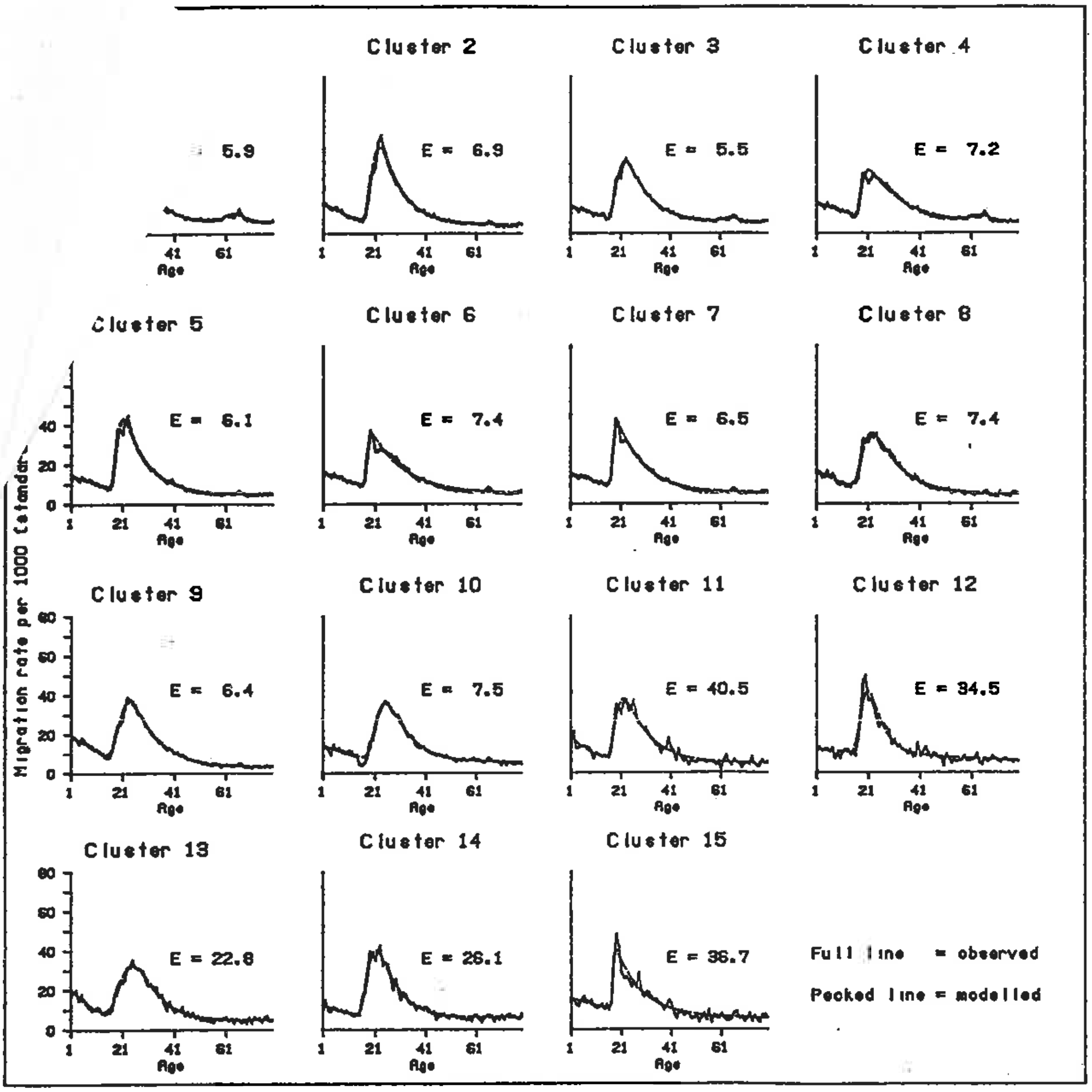


Figure 9: Observed and estimated schedules for outmigration clusters, 1985-86



(1981) which is the mean of the absolute difference between the predicted and observed rates expressed as a percentage of the observed mean, or

$$E = 100\left(\sum_a |m_I^a(\text{mod}) - m_I^a(\text{obs})|/n \quad / \sum_a m_I^a(\text{obs})/n\right) \tag{11}$$

where $m_I^a(\text{mod})$ = the predicted migration rate for age a; and

$m_I^a(\text{obs})$ = the observed migration rate for age a;

and n = the number of age-specific rates.

Apart from where model schedules are fitted to single areas, the E values fall below 10.0 in most cases. The parameter values for each of the 15 cluster profiles are presented in Tables 6 and 7 together with three further measures:

(i) the child dependency index

$$\gamma = b_1/b_2 \tag{12}$$

Table 6: Parameters and parameter ratios for immigration clusters

PARAMETER	CLUSTER 1	2	3	4	5	6	7	8
b_1	0.007	0.013	0.010	0.010	0.010	0.014	0.016	0.012
α_1	0.039	0.108	0.091	0.031	0.046	0.066	0.058	0.059
b_2	0.055	0.087	0.088	0.026	0.027	0.043	0.058	0.050
μ_2	18.319	22.028	20.434	20.899	18.186	20.949	21.024	18.129
α_2	0.087	0.148	0.123	0.096	0.071	0.096	0.104	0.098
λ_2	1.514	0.289	0.404	0.403	1.580	0.352	0.335	1.978
b_3	.	.	.	0.000	0.000	0.004	0.002	.
μ_3	.	.	.	81.157	80.985	60.257	0.000	.
α_3	.	.	.	0.545	0.557	0.047	0.051	.
λ_3	.	.	.	0.099	0.101	0.384	0.000	.
c	0.003	0.005	0.003	0.005	0.005	0.004	0.002	0.004
γ	0.132	0.155	0.119	0.383	0.389	0.332	0.280	0.246
β	0.447	0.729	0.735	0.325	0.650	0.686	0.553	0.610
ε	17.492	1.960	3.275	4.195	22.114	3.655	3.214	20.280

PARAMETER	CLUSTER 9	10	11	12	13	14	15
b_1	0.014	0.014	0.013	0.015	0.024	0.018	0.011
α_1	0.091	0.081	0.077	0.064	0.105	0.095	0.078
b_2	0.076	0.064	0.042	0.029	0.067	0.050	0.054
μ_2	17.949	18.081	18.539	18.771	24.210	23.300	17.900
α_2	0.121	0.107	0.089	0.070	0.137	0.121	0.105
λ_2	2.745	2.159	1.304	2.512	0.210	0.400	3.060
b_3	.	.	0.003	0.005	.	.	.
μ_3	.	.	70.000	86.825	.	.	.
α_3	.	.	0.500	0.500	.	.	.
λ_3	.	.	0.227	0.623	.	.	.
c	0.003	0.003	0.005	0.005	0.004	0.006	0.005
γ	0.181	0.224	0.306	0.523	0.358	0.360	0.204
β	0.752	0.756	0.863	0.917	0.766	0.785	0.743
ε	22.747	20.106	14.594	36.123	1.533	3.306	29.143

Table 7: Parameters and parameter ratios for outmigration clusters

PARAMETER	CLUSTER 1	2	3	4	5	6	7	8
b_1	0.012	0.013	0.010	0.012	0.011	0.012	0.011	0.014
α_1	0.125	0.096	0.074	0.059	0.066	0.058	0.055	0.074
b_2	0.063	0.069	0.054	0.033	0.056	0.032	0.041	0.052
μ_2	23.254	20.055	20.227	18.804	19.046	18.219	18.227	19.524
α_2	0.142	0.119	0.114	0.078	0.113	0.077	0.096	0.099
λ_2	0.247	0.497	0.438	0.781	0.715	1.457	1.353	0.398
b_3	0.000	.	0.000	0.000	.	.	.	.
μ_3	79.905	.	76.001	77.829	.	.	.	.
α_3	0.638	.	0.821	0.640				
λ_3	0.113	.	0.157	0.126	.	.	.	.
c	0.006	0.004	0.005	0.005	0.005	0.005	0.005	0.004
γ	0.192	0.196	0.178	0.368	0.205	0.389	0.282	0.272
β	0.882	0.805	0.647	0.755	0.581	0.749	0.572	0.743
ε	1.748	4.166	3.839	10.047	6.316	18.875	14.052	4.010

PARAMETER	CLUSTER 9	10	11	12	13	14	15
b_1	0.019	0.011	0.016	0.012	0.019	0.008	0.011
α_1	0.070	0.083	0.105	0.008	0.111	0.104	0.045
b_2	0.059	0.058	0.053	0.076	0.070	0.067	0.036
μ_2	20.893	22.386	19.200	19.800	24.010	19.100	17.900
α_2	0.108	0.116	0.099	0.215	0.142	0.134	0.102
λ_2	0.331	0.351	0.500	0.400	0.190	0.410	2.340
c	0.003	0.006	0.005	0.005	0.005	0.006	0.005
γ	0.316	0.183	0.302	0.158	0.271	0.119	0.306
β	0.650	0.717	1.061	0.037	0.782	0.776	0.441
ε	3.063	3.038	5.051	1.861	1.338	3.060	22.941

which measures the rate at which children migrate with their parents by comparing the level parameters of the pre-labour force and labour force components;

(ii) the parental-shift regularity

$$\beta = \alpha_1/\alpha_2 \tag{13}$$

which measures the ratio between the rates of descent of the pre-labour force and labour force curves; and

(iii) the labour asymmetry

$$\epsilon = \lambda_2/\alpha_2 \tag{14}$$

which involves the relationship between the rates of ascent and descent of the labour force curve.

These parameter ratios assist in describing differences between the model schedules for inmigration and outmigration. The London FPC areas are divided into three distinct groups on the basis of the inmigration data. Camden, Islington, Kensington, Chelsea and Westminster, with a double peaking evident in the observed labour force curve, form the first group. The model schedule smoothes the curve and produces a relatively low μ_2 parameter, illustrating the importance of inmigration of those in their late teens. The high ϵ value emphasises the sharp increase in the rate of migration at age 18. The other two groups of London FPC areas have profiles with a high but later labour force peak. The labour force curves are rather more symmetrical with a less emphatic jump in the rate of migration on the upward slope. Clusters 4, 5 and 6 contain many of the rural counties and all show some evidence of a small peak in retirement inmigration. The labour force peaks for these groups are generally at a lower level than for the London groups, and significantly, these clusters exhibit a considerably higher child dependency index than the London groups. In cluster 7, a number of constituent areas show clear evidence of an upward retirement slope: Bedfordshire, Hertfordshire, Berkshire, Surrey, St. Helens, Trafford and Barking/Havering.

The schedules of clusters 8, 9 and 10 are dominated by very high peaks in their labour force curves at an early age and no retirement component. The sharp increase in the rate of inmovement around age 18 is emphasized by the relatively high λ_2 values giving significant asymmetry. The student factor may be important in determining the shape of the schedule for these three cluster which are all major university cities or counties. Clusters 11 and 12 contain rather unusual combinations of areas. Cambridgeshire stands out as having a unique profile which contains a labour-force curve with a sharp peak at age 19 (student factor again) and quite a significant retirement component. Areas of group 12 show evidence of a relatively flat labour force curve but Scotland, North Yorkshire and Hampshire have a localized peak at age 19. The only three areas failing to combine were South Tyneside, which had a very late labour force peak; Solihull, having a unique profile shape for a metropolitan area with a rounded labour force curve and a retirement peak; and West Glamorgan.

At the 15 cluster stage of the outmigration classification, 10 distinct groups were formed with five individual areas failing to combine. The majority of London FPC areas make up the first cluster. The peak in outmigration is much lower and later than in clusters 2 and 3, and there is a significant retirement component. The labour force curve is more symmetrical and the correspondence between child and parent migration is high. The high density FPC areas in London of Camden, Islington, Kensington, Chelsea and Westminster make up cluster 10 where the labour force peak is later and more skewed and where migration activity around retirement age is less significant. Clusters 2 and 3 contain the majority of high density metropolitan areas outside Greater London, together with in cluster 3, a number of counties in the South. Both schedules have a peak at age 20 with that of cluster 2 at a higher level. The β value is greatest for the metropolitan districts in cluster 2, indicating the importance of family movement away from these high density areas. Clusters 3 and 4 have a significant retirement component although cluster 4 is distinguished by a double peak in its labour force curve. Cluster 5 is also non-metropolitan in area composition but has no retirement component. It has a slightly later modelled peak in labour force movement but the observed schedule has a double peak indicating higher levels of outmovement around the age of leaving school and after higher education.

Clusters 6 and 7 are two much larger groups and their model schedules stand out in Figure 9 as having the greatest degree of labour asymmetry. The peak in labour force migration occurs around age 18 and although observed rates do rise to age 65 in each case, the cluster schedules were modelled without a retirement component. Cluster 8 contains four metropolitan areas whose schedules are characterized by a labour force curve with a relatively late peak, whereas cluster 9 combines Scotland with the metropolitan district of Gateshead. Those areas failing to combine at this 15 cluster stage are Gwynedd; Sandwell; South Tyneside with a multi-peaked labour force component; Isle of Wight with a high labour force peak but large fluctuations in age-specific rates after age 40; and Powys, with its labour force decline curve interrupted by a small peak at age 30.

4 Conclusion

This paper has attempted to outline the framework of the OPCS/DOE migration projection methodology, to set out the equations associated with different stages in a transparent form, and to examine three selected features of the model in more detail using data from the NHSCR. Three recommendations emerge. Firstly, a careful and systematic specification of the broad age group boundaries of the migrant streams defined in the assignment stage is required. Secondly, changes have taken place in the age-specific distributions of migration since the beginning of the 1980s which cast doubts on the assumption of constant 1981 Census-based assignment probabilities. It would therefore be appropriate to consider a mechanism for trending or updating the assignment probabilities based on the information provided by the NHSCR data for more recent years, particularly as the interval between the Census and the base period of projection increases. Such a mechanism would need to accommodate the conceptual and definitional differences in the two measures of migration identified by Devis and Mills (1986) and by Boden et al. (1987; 1988). An alternative solution, given the availability of the continuous time series of NHSCR data, would be to adopt a movement-based migration projection model (Rees 1984) as suggested in Boden (1989), rather than attempting to modify further the existing Census-based transition model. Our third and final recommendation is to adopt a systematic procedure for grouping local authority areas which is based on observed rates rather than model parameters. Model migration schedules can then be used to identify the different characteristics of particular groups of areas. Once again, there is the opportunity of using NHSCR Primary Unit Data to take into account the shifts occurring in the level and age-structure of migration.

Acknowledgement

The authors are very grateful for comments on an earlier version of this paper by colleagues from OPCS. We remain, nevertheless, entirely responsible for the contents.

References

Armitage, R. (1986) Population projections for English local authority areas, Population Trends, 43, 37-40

Bates, J.J. (1984) Migration age profiles from the 1981 Census Report prepared for RPD2 Division of the DOE, London

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

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

Boden, P. (1989) The analysis of internal migration in the United Kingdom using Census and National Health Service Central Register data, Unpublished PhD Thesis, School of Geography, University of Leeds

Boden, P., Stillwell, J.C.H. and Rees, P.H. (1987) Migration data from the National Health Service Central Register and the 1981 Census: further comparative analysis, Working Paper 495, School of Geography, University of Leeds

Boden, P., Stillwell, J.C.H. and Rees, P.H. (1988) Linking Census and NHSCR migration data, Working Paper 511, School of Geography, University of Leeds

Bracken, I. and Bates, J.J. (1983) Analysis of gross migration profiles in England and Wales: some developments in classification, Environment and Planning A, 15, 343-355

Devis, T.L.F. and Mills, I. (1986) A comparison of migration data from the National Health Service Central Register and the 1981 Census, OPCS Occasional Paper 35, OPCS, London

Martin and Voorhees Associates and John Bates Services (1981) Developing the migration component of the official sub-national population projections, Final Report prepared for DPRP3, DOE, London

Norusis, M. (1985) SPSSX Advanced Statistics Guide, McGraw Hill

OPCS (1982, 1984, 1986) OPCS Monitor PP1, mid-year population estimates for local government and health authority areas of England and Wales, OPCS, London

OPCS (1983, 1986, 1988) Population projections: area population projections by age and sex for standard regions, counties, London boroughs and metropolitan districts of England from mid-1981, mid-1983 and mid-1985 Series PP3, HMSO, London

Rees, P.H. (1984) Spatial population analysis using movement data and accounting methods: theory, models, the MOVE program and examples, Working Paper 404, School of Geography, University of Leeds

Rees, P.H. and Willekens, F. (1989) Population projection: Dutch and English multiregional methods, Chapter 2 in Stillwell, J.C.H. and Scholten, H.J. (eds) Contemporary Research in Population Geography. A Comparison of the United Kingdom and the Netherlands, Kluwer, Dordrecht, pp 19-37

Rogers, A. and Castro, L.J. (1981) Model Migration Schedules RR-81-30, IIASA, Laxenburg, Austria.

Rogers, A., Raquillet, R. and Castro, L.J. (1978) Model migration schedules and their applications, Environment and Planning A, 10, 475-502

Rogers, A. and Planck, F. (1984) MODEL: a general program for estimating parametrized model schedules of fertility, mortality, migration and marital status and labour force status transitions, Working Paper 83-102, IIASA, Laxenburg, Austria

Stillwell, J.C.H., Boden, P. and Rees, P.H. (1987) Migration schedule construction using MODEL and GIMMS, Computer Manual 29, School of Geography, University of Leeds

Stillwell, J.C.H., Boden, P. and Rees, P.H. (1990) Trends in internal net migration in the UK: 1975 to 1986, Area, 22(1), 57-65