WORKING PAPER 531

OLD MODEL FACES NEW CHALLENGES: A REVIEW OF THE STATE OF THE ART IN MULTISTATE POPULATION MODELLING

Philip Rees

School of Geography
University of Leeds
Leeds LS2 9JT

September 1989

Paper presented at the Annual Conference of the Institute of British Geographers, Population Geography Study Group on "Recent research and future priorities in population geography", September 11-13th, 1989, University of Newcastle-upon Tyne.

CONTENTS

List of tables

List of figures

Abstract

1. Old model faces new challenges

2. The multistate model: a thumbnail sketch

3. The different ways of measuring migration: the concept challenge

4. External migration: the "glasnost" challenge

5. The age dimension: the parsimony challenge (1)

6. The spatial dimension: the parsimony challenge (2)

7. Some are more equal than others: the diversity challenge

8. New causes of mortality: the AIDS challenge

9. Improved presentation: the GIS challenge

10. The multistate model is dead. Long live the multistate model!

References

Tables

Figures

LIST OF TABLES

1. A classification of subnational population projection models.
2. Illustrative computations of a multiregional projection model.
3. Type of migration measure and survival rate estimation method.
4. Alternative methods for incorporating external migration in the multistate projection model.
5. Projected populations of selected UK regions in 2006 based on alternative methods for incorporating external migration.
6. A family of multiexponential models of the variation of component rates with age.

LIST OF FIGURES

1. A schema for modelling component rate schedules by age.

ABSTRACT

The paper reviews the state-of-the-art in multistate population modelling. After a thumbnail sketch of the projection version of the model, a series of challenges to the model structure and utility are considered. The challenges to incorporate different migration concepts and external migration have resulted in improvements in model specification. The challenge to replace the large number of input variables with a more parsimonious model is described for age and spatial dimensions. The multistate model's cousin, the dynamic microsimulation model, is then analyzed as a response to the challenge to enrich model outputs. Finally, the paper discusses two new challenges to users of multistate modellers: to deal with the AIDS pandemic, and to respond to the GIS challenge. Le modele est mort. Vive le modele.

> "The people, and the people alone, are the motive force in the making of world history." Mao Tse-Tung, Quotations from Chairman Mao Tse-Tung (1966), 11.

> Parsimony, n. sparingness in the spending of money: praiseworthy economy in use of means to an end: avoidance of excess: frugality: niggardliness. [L.parsimonia - parcere, to spare.] Chambers Twentieth Century Dictionary, 1972.

1. OLD MODEL FACES NEW CHALLENGES

Models can be defined as simplified representations of the real world. Many population models involve specification of the processes of change over time in total numbers of people and their characteristics. The models are then used to predict the future fate of the population. We do this to reduce our uncertainty about the future and to guide our actions (collective and private) so as to improve the future. Model predictions are being constantly faced with real world outcomes, are found wanting and are replaced by new ones.

Demographic models are, however, extraordinarily persistent compared with economic or social models. The life table model of a stationary population, invented by John Graunt in the middle of the seventeenth century (Graunt 1662), is in continuous use today in demography, in the insurance industry and in public health work, though, of course, in much modified form. Similarly, the component-cohort survival model was developed for projecting the United States population broken down by age and sex by Thompson and Whelpton (cited in Stockwell 1976) in the 1930s and has been in world wide use ever since.

These classic models deal with single populations largely closed to exchanges with other populations, except for the addition of net migration to the cohort survival model. In the middle 1960s an American professor of Russian extraction, Andrei Rogers, wrote a series of papers that re-specified the cohort survival model as a model of many regional populations exchanging migrants (see Rogers 1968 for a summary). He continued this work in the 1970s, with a series of student collaborators (Ledent, Willekens, Castro), by showing how the single population life table could be generalized as a multiregional life table, which predicts how the expectations of remaining life will be spread over interacting regions (see Rogers 1975 for a summary and Willekens and Rogers 1978 for a clear exemplification). Parallel to these developments was the derivation of multiregional population projection models by a group of British researchers (Wilson, Rees, Plessis-Fraissard, Stillwell) drawing on a social accounting tradition (see Rees and Wilson 1977 and United Nations 1975 for summaries).

In the second half of the 1970s and early 1980s, these new models were applied to the regional populations of some seventeen countries in a major cross-national comparative study (summarized in Rogers and Willekens 1986). In the course of this project, several difficulties of application of the multiregional model using available data were faced and these stimulated the research team to develop new methods and insights. These include methods of data infilling (Willekens, Por and Raquillet 1979, 1981), of model migration schedule construction (Rogers, Raquillet and Castro 1978; Rogers and Castro 1981), of introducing different kinds of migration (Ledent 1980; Ledent and Rees 1986; Rees and Willekens 1986), of relaxing the Markovian assumptions of the multiregional model (Ledent 1981), of opening the multiregional model to the outside world (Willekens and Drewe 1984; Rees 1984), of extending schedule parameteriztaion to all aspects of the model (Rogers and Planck 1984; Rogers 1986), and of developing effective software for variable rates population projection (MUDEA - ProGamma 1988; DIALOG - Scherbov and Grechucha 1988). The term multiregional model was superseded by the term multistate model when it was appreciated that all the techniques developed for geographic regions could be applied to any other mutually exclusive set of population states.

The multistate population model, originally proposed in the 1960s, has faced several important challenges. The aims of this paper are to describe how those challenges have been confronted and overcome (putting flesh on the skeletal review above), and to examine some new challenges that demographic and other developments in the 1980s have posed the model. In dong this the paper draws on two recent sources - the draft of a manual for subnational population projection (United Nations 1989) and one of the latest in a long line of papers by the model's inventors (Rogers 1989).

Section 2 of the paper provides a brief sketch of the multistate model in a general form. Section 3 then exposes the differences between the movement and transition versions of the model in their solutions to the problem of measuring survival rates. Section 4 reviews the alternatives for connecting the many population states in one country to the outside world, and their long run consequences under constant input assumptions. Section 5 summarizes progress achieved in reducing the model input variables by representing the age dimension as a set of functions. Section 6 speculates on how much equivalent parsimony could be achieved vis a vis the spatial dimension by modelling migration as a spatial interaction and linking it to a wider set of processes collectively known as development. Section 7 confronts the challenge of providing much more information about people in the model. Section 8 asks the question: can the multistate model be used to investigate the consequences of the AIDS pandemic? Finally, section 9 looks briefly at the challenge posed by developments in the new field of geographical information systems.

2. THE MULTISTATE MODEL: A THUMBNAIL SKETCH

Population projection models can be classified according to the way the spatial and age dimensions are treated into four categories (Rogers 1985 and 1989): (1) models that project single populations without age or locational disaggregation, (2) models that project a single population with age detail but no locational, (3) models that project the populations of many locations without age detail, and (4) models that project the populations of many locations with age detail. Table 1 shows these four categories and how they relate to the family of models discussed in United Nations (1989), which further distinguish between models by the number and type of components included and their accounting treatment. Models 6 and 7 in this framework are multiregional population projection tools, which differ according to the nature of migration data input (moves or transitions).

The basic idea of the multiregional projection model can be explained quite simply. To the populations of the regions of interest are applied rates or probabilities of migration and survival which produce estimates of the number of out-migrants. The end of interval populations of the regions consist of the survivors who stay within the region plus surviving in-migrants. New infant members are added to the regional populations by applying fertility rates to the appropriate populations at risk, and by applying migration and survival rates to the newly born to see them through to the end of the time interval.

Formally, the model can be defined as follows. It applies to all subnational units. Let

$P^{i\cdot}_{ag}(t)$ = the population of subnational unit i in period-cohort a and gender g at time t, the start of the projection time interval

$P^{\cdot i}_{ag}(t+u)$ = the population of subnational unit i in period-cohort a and gender g at time t+u, the end of the projection time interval

s^{ij}_{ag} = the rate at which persons in period-cohort a and gender g resident in subnational unit i at the start of the time interval, survive in region j at the end of the time interval

f^{i}_{a} = the rate at which females of period-cohort a and gender g give birth in subnational unit i over the time interval (normally only female fertility rates are non-zero)

x^{i}_{g} = the proportion of births in subnational unit i which are of gender g

a = period-cohort label that refers to persons moving from one age group to another in a time interval e.g. 5 to 6 years of age or 0-4 to 5-9 years of age

A = the last period-cohort, which is open ended (e.g. 85 and over to 90 and over)

For all period-cohorts from the first to the last, the end of time interval population is given by

$$P^{\cdot i}_{ag}(t+u) = \Sigma_j\, s^{ji}_{ag}\, P^{j\cdot}_{ag}(t) \quad \text{for } 1 <= a <= A, \text{ for } g = m,f \quad (1)$$

Where the initial and final locations in s^{ij}_{ag} are different, the rate is one of migration and survival, where they are the same (s^{ii}_{ag}), the rate is one of staying and survival.

The first period-cohort population is derived by projecting the number of births by applying fertility rates to suitable populations at risk, and then sexing and surviving the infants. The most convenient definition of the population at risk is the average population of females in a period-cohort over the time interval.:

$$PAR^{i}_{af} = 0.5\ (P^{i\cdot}_{af} + P^{\cdot i}_{af}) \text{ for } a_1 <= a <= a_2 \quad (2)$$

where a_1 and a_2 are the youngest and oldest cohorts for which fertility rates are reported. The necessary final populations have already been computed using equation (1). Births are then projected and sexed:

$$B^{i}_{g} = x^{i}_{g} \sum_{a=a_1}^{a_2} f^{i}_{a}\ PAR^{i}_{af} \quad \text{for } g = m,f \quad (3)$$

where B^{i}_{g} are the births in subnational unit i of gender g in the time interval. These births are then subjected to the migration and survival process:

$$P^{\cdot i}_{0g}(t+u) = \sum_{j} s^{ji}_{0g} B^{j}_{g} \quad \text{for } g = m,f \quad (4),$$

where the period-cohort labelled zero refers to persons born in the time interval.

Before repeating the computation for the next time interval, the final populations of the current interval become the initial population of the next

$$P^{i\cdot}_{a+1g}(t+u) = P^{\cdot i}_{ag}(t+u) \quad \text{for } 0 <= a <= A-1, \text{ for } g = m,f \quad (5)$$

except that the last two are added together

$$P^{i\cdot}_{Ag}(t+u) = P^{\cdot i}_{A-1g}(t+u) + P^{\cdot i}_{Ag}(t+u) \quad \text{for } g = m,f \quad (6)$$

in order to maintain a fixed, open-ended last period-cohort. The model is applied to all subnational units within a nation.

Table 2 illustrates the computations for females in one region, East Anglia, in a three region system that partitions Great Britain (East Anglia, South East and the Rest of Britain). The first panel (A) of the table shows the survival calculations for all period-cohorts of the East Anglian population. For example, the projected population of 47,699 women aged 40-44 in 1971 is made up of 42,756 surviving the 1966-71 period in East Anglia (a rate of .909934 times a population of 46,988) plus 2,904 in-migrating from the South East (.005483 x 529,649) and 2,039 from the Rest of Britain (.00188 x 1,079,708). The middle panel (B) illustrates the computation of population at risk (equation 2) and births (equation 3). Panel C of Table 2 applies the survival rates to births in all regions, using equation (4), to obtain a projection of the number of 0-4 year old girls in the East Anglian population in 1971 (62,811).

For many purposes (though not computation), it is convenient to arrange the populations and rates involved in a multistate, cohort-survival model in matrix form. The survival rates for each period-cohort are gathered together in a matrix, $\underline{H}_{ag}$, for each period-cohort and gender

$$\underline{H}_{ag} = \begin{vmatrix} s^{11}_{ag} & s^{12}_{ag} & \dots & s^{1N}_{ag} \\ s^{12}_{ag} & s^{22}_{ag} & \dots & s^{2N}_{ag} \\ : & : & & : \\ s^{N1}_{ag} & s^{N2}_{ag} & \dots & s^{NN}_{ag} \end{vmatrix} \quad (6)$$

and the initial populations in a row vector $\underline{P}_{ag}(t)$

$$\underline{P}_{ag}(t) = [\ P^{1\cdot}_{ag} \quad P^{2\cdot}_{ag} \quad \dots \quad P^{N\cdot}_{ag}\ ] \quad (7)$$

The terms referring to the infant period-cohort are gathered together in a matrix $\underline{B}_{ag}$ for each fertile period-cohort and gender

$$\underline{B}_{ag} = \begin{vmatrix} b^{11}_{ag} & b^{12}_{ag} & \dots & b^{1N}_{ag} \\ b^{12}_{ag} & b^{22}_{ag} & \dots & b^{2N}_{ag} \\ : & : & & : \\ b^{N1}_{ag} & b^{N2}_{ag} & \dots & b^{NN}_{ag} \end{vmatrix} \quad (8)$$

The individual "birth rate" terms in this submatrix are a product of infant survival rates, gender proportions, fertility rates and female survival rates in the maternal ages and the population at risk definition:

b^{ij}_{ag} = rate at which females in period-cohort a in subnational unit i give birth to infants of gender g who survive in subnational unit j at the end of the interval

which, in terms of variables already defined is given by

$$b^{ij}_{ag} = s^{ij}_{og}\ x^{i}_{g}\ 0.5\ f^{i}_{a}\ (1 + s^{ii}_{af}) + \Sigma_k\ s^{kj}_{og}\ x^{k}_{g}\ 0.5\ f^{k}_{a}\ s^{ik}_{af} \quad (9).$$

These submatrices exhibit rates of transition from row states to column states, but Rogers and others have used the convention of transition from column to row states, so that it is necessary to define transposed versions.

Let

$$\underline{S}_{ag} = \underline{H}_{ag}{}^{*} \qquad (10)$$

$$\underline{F}_{ag} = \underline{B}_{ag}'$$

$$\underline{K}_{ag} = \underline{P}_{ag}'$$

These submatrices can be assembled in a large matrix, $\underline{G}$, which, when multiplied by the initial population vector, yields the projected population

$$\left|\begin{array}{c} \underline{K}_{1m}(t+u) \\ \underline{K}_{2m}(t+u) \\ \vdots \\ \underline{K}_{am}(t+u) \\ \vdots \\ \underline{K}_{Am}(t+u) \\ \hline \underline{K}_{1f}(t+u) \\ \underline{K}_{2f}(t+u) \\ \vdots \\ \underline{K}_{af}(t+u) \\ \vdots \\ \underline{K}_{Af}(t+u) \end{array}\right| \begin{array}{c} = \\ \\ \\ \\ \\ \\ \\ = \\ \\ \\ \\ \\ \\ \end{array} \left|\begin{array}{cccc|ccccccc} 0 & \dots & & 0 & 0 & \dots & \underline{F}_{a1m} & \dots & \underline{F}_{a2m} & \dots & 0 \\ \underline{S}_{1m} & \dots & & 0 & 0 & \dots & 0 & \dots & 0 & \dots & 0 \\ \vdots & & & \vdots & \vdots & & \vdots & & \vdots & & \vdots \\ 0 & \underline{S}_{am} & & 0 & 0 & \dots & 0 & \dots & 0 & \dots & 0 \\ \vdots & & . & 0 & \vdots & & \vdots & & \vdots & & \vdots \\ 0 & \dots & \underline{S}_{A-1m} & \underline{S}_{Am} & 0 & \dots & 0 & \dots & 0 & \dots & 0 \\ \hline 0 & \dots & & 0 & 0 & \dots & \underline{F}_{a1f} & \dots & \underline{F}_{a2f} & \dots & 0 \\ 0 & \dots & & 0 & \underline{S}_{1f} & & 0 & \dots & 0 & \dots & 0 \\ \vdots & & & \vdots & \vdots & . & \vdots & & \vdots & & \vdots \\ 0 & \dots & & 0 & 0 & & \underline{S}_{af} & \dots & 0 & \dots & 0 \\ \vdots & & & \vdots & \vdots & & \vdots & & . & & \vdots \\ 0 & \dots & & 0 & 0 & \dots & 0 & \dots & 0 & \underline{S}_{A-1f} & \underline{S}_{Af} \end{array}\right| \times \left|\begin{array}{c} \underline{K}_{1m}(t) \\ \underline{K}2_{m}(t) \\ \vdots \\ \underline{K}_{am}(t) \\ \vdots \\ \underline{K}_{Am}(t) \\ \hline \underline{K}_{1f}(t) \\ \underline{K}_{2f}(t) \\ \vdots \\ \underline{K}_{af}(t) \\ \vdots \\ \underline{K}_{Af}(t) \end{array}\right| \qquad (11)$$

or in summary form

$$\underline{K}(t+u) = \underline{G}\ \underline{K}(t) \qquad (12).$$

The survival rate submatrices are placed on the subdiagonal one below the principal, except for the submatrix for the final period-cohort which is placed in the principal diagonal. Note that the male birth rates are placed in the top right quadrant of the $\underline{G}$ matrix so that they are applied, in the matrix multiplication, to female initial populations. Equation (11) is a two-sex, "female dominant" version of the multistate population projection model.

3. THE DIFFERENT WAYS OF MEASURING MIGRATION: THE CONCEPT CHALLENGE

The multistate model specification described above was in place by the mid-1970s (Rogers 1975; Rees and Wilson 1977, both which synthesize earlier journal articles). In applying the model to projection (and in a related form to life table construction) in a variety of countries involved in the Migration and Settlement project (Rogers and Willekens 1986), the question of how to measure the rates that enter the model was posed. Hitherto the approach was to employ any available migration data and compute the rates by dividng by a convenient population denominator for the origin subnational unit:

$$s^{ij}_{ag} = F^{ij}_{ag} / P^{i}_{a} \qquad (13)$$

where F^{ij}_{ag} is defined as "any-old" migration flow from origin i to destination j over the benchmark period. Two developments posed a challenge to the multistate model, which could not be ignored. The first was the existence of a body of theory about the measurement of transition intensities and probabilities in conventional demographic theory that required generalization to many, interacting populations. The second was the realization that some countries used retrospective questions in censuses to measure migration, and some used population registers to record migrations via change of address forms.

Table 3 crossclassifies the type of migration measure with the methods suggested for converting the different measures into rates usable in multistate projection models. The methods are divided into transition and movement approaches. Transition approaches use tabulations of persons who migrate (migrants) either within a fixed period (T1 and T2) or within variable periods (T3 and T4) depending on the exact nature of the question about migration. Movement approaches (M) use migration events counted by registration authorities as their raw material.

The first transition approach, T1, is straightforward. The survival rates used are simply derived by dividing the surviving migrant flows in census tabulations by the population at the start of time interval, taking care with age group and time interval definitions

$$s^{ij}_{ag} = P^{ei,sj}_{ag} / P^{i\cdot}_{ag}(t) \qquad (14)$$

where $P^{ei,sj}_{ag}$ represents the number of persons in existence in subnational unit i at the start of the time interval who survive in subnational unit j at the end. This term is divided by the initial population in the period-cohort and gender group. The infant period-cohort survival rates are defined as

$$s^{ij}_{0g} = P^{bi,sj}_{0g} / B^{i}_{g} \qquad (15)$$

where $P^{bi,sj}_{0g}$ are the persons born in subnational unit i who survive at the end of the time inetrval. These migration flows are sometimes tabulated by census authorities along with the migration data for the other period-cohorts (e.g. Statistics Canada); sometimes separate place of birth by place of residence by age tables are produced from which infant migration flows can be extracted (e.g. U.S. Bureau of the Census); sometimes they are they simply ignored (e.g. Office of Population Censuses and Surveys, UK) and the researcher must estimate their magnitude.

Surviving stayer rates have been treated in three different ways. In the first (T1) they have been estimated by subtracting from population data of the recording census the total number of in-migrants, when not directly tabulated (the intra-region term in migrants tables is usually reserved to record migrants who have moved exclusively within the same subnational unit):

$$P^{ei,si}_{ag} = P^{\cdot i}_{ag}(t+u) - \Sigma_j P^{ej,si}_{ag} \quad (16).$$

The second method, T2, involves estimating surviving stayers as a residual by subtracting from the initial population the total of out-migrants and non-survivors

$$P^{ei,si}_{ag} = P^{i\cdot}_{a}(t) - \Sigma_{j\neq i} P^{ei,sj}_{ag} - P^{ei,d\cdot}_{ag} \quad (17)$$

where $P^{ei,d\cdot}_{ag}$ refers to the total of persons of period-cohort a and gender g originating in subnational unit i at the start of the time interval who die (d) somewhere. The out-migrant total is available from the migration table of the census. The non-survivors term is not available directly but may by approximated from mortality data

$$P^{ei,d\cdot}_{ag} = D^{i}_{ag} \quad (18)$$

where D^{i}_{ag} are the observed deaths in subnational unit i to persons in period-cohort a and gender g.

The third method for estimating surviving stayers, T1.3, is to improve the approximation of equation (18) through the development of an iterative, accounts based model. Surviving stayers are estimated using equation (17) but better estimates of non-survivors are obtained using the iterative model. Full details of the model, rather too detailed for presentation here, are given in Rees and Wilson 1977 and a computer algorithm is provided in Rees 1981.

Approach T1 to the estimation of survival rate or transition probabilities depends on reliable estimates of the initial populations in the benchmark, measurement period and a regular time interval for the migration question ("where were you living x years ago?") that matches the age disaggregation o the projection model (normally one or five year detail). Often such conditions are not met. Ledent (1978) and Ledent and Rees (1980, 1986) developed a method to handle such a situation, as experienced in the French case study of Ledent with Courgeau 1982.

Firstly, migration rates conditional on survival are computed from the census migration data (using method T1.1 to obtain the surviving stayer terms)

$$s_{ag}(j|ei,s.) = p^{ei,sj}_{ag} / \Sigma_j \, p^{ei,sj}_{ag} \quad (19)$$

where $s_{ag}(j|ei,s.)$ is the probability of location in subnational unit j given origin in unit i and survival over the benchmark period. The denominator term on the RH side is the sum of all surviving migrants originating in subnational unit i plus surviving stayers.

Secondly, survival rates for period-cohorts are derived using conventional life table methods applied to each subnational unit:

$$s^{i\cdot}_{ag} = L^{i}_{x(a)+ug} / L^{i}_{x(a)g} \quad (20)$$

where $s^{i\cdot}_{ag}$ is the survival rate or persons in subnational unit i, period-cohort a and gender g over the unit time/age interval, the L variables refer to life table stationary populations specific to subnational unit i and gender g for successive ages x and x+u, associated with period-cohort a.

Thirdly, the migration and survival rates are estimated by multiplying the two rates specified in equations (19) and (20):

$$s^{ij}_{ag} = s_{ag}(j|ei,s.) \, s^{i\cdot}_{ag} \quad (21).$$

Note that equation (20)'s estimate of the rate of non-survival makes the same approximation as the T1.2 method in equation (18).

Many censuses use migration questions alternative to the fixed period question. "Where was your last residence?" and "where were you born?" have had widespread use. Researchers have struggled long and hard to define methods for using migration data based on these questions in projection models. Courgeau (1980) is pessimistic about their utility; United Nations (1989, Chapter IV) summarizes what can be done.

The second set of approaches to survival rate estimation is applicable if register based data on moves as events are employed as measures of migration. Movement approach M1 involves the generaliztion of the method used in single population demography for converting occurrence-exposure rates into survival probabilties, achieved by Ledent and Rogers (1976) in the life table context, and extended and used in projection context by Willekens and Rogers (1978) and Willekens and Drewe (1984). The method uses the assumption that population change and change in migration intensities over short time and age intervals respectively can be represented by a linear function. This assumption suffices for almost all practical purposes.

An occurrence-exposure rate is the ratio of demographic occurrences or events to the population exposed to the risk of the event. For events occurring in a subnational unit, the initial population of the unit is not an adequate estimate of the population at risk because some members leave and have reduced risks, and because members of other subnational unit populations may enter and be exposed to risk of the event. The linear assumption leads to the following population at risk definition

$PAR^{i}_{ag} = 0.5 (P^{i\cdot}_{ag}(t) + P^{\cdot i}_{ag}(t+u))$ (22).

Migration rates are defined using origin region populations at risk

$m^{ij}_{ag} = M^{ij}_{ag} / PAR^{i}_{ag}$ (23)

where M^{ij}_{ag} is the count of moves between subnational unit i and subnational j for persons of gender g in period-cohort a at the time of the move and death rates are given as

$d^{i}_{ag} = D^{i}_{ag} / PAR^{i}_{ag}$ (24).

Note that the intra-region rate must be defined, in the movement approach, as a residual (cf. transition approach T1.2)

$r^{i}_{ag} = 1 - \Sigma_{j\neq i} m^{ij}_{ag} - d^{i}_{ag}$ (25).

What we now need to do is to take the occurrence-exposure rates and transform them into survival rates for use in the multistate model.

Initial and final populations in a period-cohort are linked by the following projection equation

$$P^{\cdot i}{}_{ag}(t+u) = P^{i\cdot}{}_{ag}(t) - \Sigma_{j\neq i}\ m^{ij}{}_{ag}\ PAR^{i}{}_{ag} - d^{i}{}_{ag}\ PAR^{i}{}_{ag}$$
$$+ \Sigma_{j\neq i}\ m^{ji}{}_{ag}\ PAR^{j}{}_{ag} \quad (26).$$

Substituting the RH side of (22) into the RH side of (26) we obtain

$$\begin{aligned} P^{\cdot i}{}_{ag}(t+u) = P^{i\cdot}{}_{ag}(t) &- \Sigma_{j\neq i}\ m^{ij}{}_{ag}\ 0.5(P^{i\cdot}{}_{ag}(t) + P^{\cdot i}{}_{ag}(t+u)) \\ &- d^{i}{}_{ag}\ 0.5(P^{i\cdot}{}_{ag}(t) + P^{\cdot i}{}_{ag}(t+u)) \\ &+ \Sigma_{j\neq i}\ m^{ji}{}_{ag}\ 0.5(P^{j\cdot}{}_{ag}(t) + P^{\cdot j}{}_{ag}(t+u)) \end{aligned}$$
$$(27).$$

This projection equation needs to be arranged so that all $P^{\cdot i}{}_{ag}(t+u)$ and $P^{\cdot j}{}_{ag}(t+u)$ terms are on the LH side, with the RH side rearranged as a rate expression multiplying the $P^{i\cdot}{}_{ag}(t)$ term. Because the projection equation for unit i contains population terms for other units (j), the equations for all subnational units must be solved together. The problem can be solved by adopting a matrix method.

A simplification of the notation will help in deriving the method from first principles: the period-cohort and gender labels are dropped as the method applies to all groups, and the notation for the population is simplified. The projection equations for n subnational units are as follows:

$$\begin{aligned} P^{\cdot 1} = {} & (P^{1\cdot} - (\Sigma_{j\neq 1}\ m^{1j} + d^{1})\ 0.5(P^{1\cdot} + P^{\cdot 1})) \\ & + m^{21}\ 0.5(P^{2\cdot} + P^{\cdot 2}) \\ & + \ldots + m^{n1}\ 0.5(P^{n\cdot} + P^{\cdot n}) \end{aligned}$$

$$\begin{aligned} P^{\cdot 2} = {} & m^{12}\ 0.5(P^{1\cdot} + P^{\cdot 1}) \\ & + (P^{2\cdot} - (\Sigma_{j\neq 2}\ m^{2j} + d^{2})\ 0.5(P^{2\cdot} + P^{\cdot 2})) \\ & + \ldots + m^{n2}\ 0.5(P^{n\cdot} + P^{\cdot n}) \end{aligned}$$

$$\vdots \qquad \vdots \qquad \vdots \qquad \vdots$$

$$\begin{aligned} P^{\cdot n} = {} & m^{1n}\ 0.5(P^{1\cdot} + P^{\cdot 1}) \\ & + m^{2n}\ 0.5(P^{2\cdot} + P^{\cdot 2}) + \ldots \\ & + (P^{n\cdot} - (\Sigma_{j\neq n}\ m^{nj} + d^{n})\ 0.5(P^{n\cdot} + P^{\cdot n})) \end{aligned}$$
$$(28).$$

Let us rearrange these terms in vectors for population and the other terms in a matrix

$$\underline{P}(t+u) = \begin{vmatrix} P^{\cdot 1} \\ P^{\cdot 2} \\ \vdots \\ P^{\cdot n} \end{vmatrix} \qquad \underline{P}(t) = \begin{vmatrix} P^{1\cdot} \\ P^{2\cdot} \\ \vdots \\ P^{n\cdot} \end{vmatrix}$$

$$\underline{M} = \begin{vmatrix} (\Sigma_{j\neq 1}\ m^{1j} + d^{1}) & -m^{21} & \cdots & m^{n1} \\ -m^{12} & (\Sigma_{j\neq 2}\ m^{2j} + d^{2}) & \cdots & m^{n2} \\ -m^{1n} & -m^{2n} & \cdots & (\Sigma_{j\neq n}\ m^{nj} + d^{n}) \end{vmatrix}$$

Then the equation set (28) can be re-expressed as the matrix equation

$$\underline{P}(t+u) = \underline{P}(t) - 0.5\underline{M}\ (\underline{P}(t) + \underline{P}(t+u))$$

Multiplying through this becomes

$$\underline{P}(t+u) = \underline{P}(t) - 0.5\underline{M}\ \underline{P}(t) - 0.5\underline{M}\ \underline{P}(t+u)$$

so that, after rearranging

$$\underline{P}(t+u) + 0.5\underline{M}\ \underline{P}(t+u) = \underline{P}(t) - 0.5\underline{M}\ \underline{P}(t)$$

and taking out common vectors on both sides, we obtain

$$[\underline{I} + 0.5\underline{M}]\ \underline{P}(t+u) = [\underline{I} - 0.5\underline{M}]\ \underline{P}(t)$$

where $\underline{I}$ is the identity matrix. Multiplying both sides by the inverse of $[\underline{I} + 0.5\underline{M}]$ yields

$$\underline{P}(t+u) = [\underline{I} + 0.5\underline{M}]^{-1}\ [\underline{I} - 0.5\underline{M}]\ \underline{P}(t)$$

so that the survival rates matrix is generally defined as

$$\underline{S} = [\underline{I} + 0.5\underline{M}]^{-1}\ [\underline{I} - 0.5\underline{M}] \qquad (29),$$

Adding back the missing period-cohort and gender labels gives us

$$\underline{S}_{ag} = [\underline{I} + 0.5\underline{M}_{ag}]^{-1}\ [\underline{I} - 0.5\underline{M}_{ag}]\ (30).$$

Various versions of (29) and (30) have been used: they vary depending on the age-time period plan used for measuring migration and mortality, and whether the survival rates are to be used in a projection model (as here) or in a life table computation. Note that implementation of this estimating equation involves use of a matrix inversion algorithm (see Rogers 1971 or Willekens and Rogers 1978 for suitable programs).

The final method listed in Table 3 is not really a method of estimating survival rates but rather an alternative method of solving the projection equations in the movement case through iteration.

In Step (1), the population at risk for the current iteration, I, is computed

$$PAR^{i}_{ag}(I) = 0.5(P^{i\cdot}_{ag}(t) + P^{\cdot i}_{ag}(t+u)) \text{ for all } i \qquad (31)$$

In Step (2), this is compared with the values in the previous iteration and if

$$\left| PAR^{i}_{ag}(I) - PAR^{i}_{ag}(I-1) \right| < \text{critical value} \qquad (32)$$

for all subnational units, the procedure is stopped. For the first iteration

$$PAR^{i}_{ag}(1) = P^{i}_{ag}(t) \qquad (33)$$

and the comparison is skipped. If inequality (32) is not true for all i, the procedure continues.

In Step (3), the final population is projected by implementing equation (26) for all subnational units. The iteration number is increased by 1 and the procedure returns to Step (1).

The iterative model is implemented in a computer algorithm in Rees (1984), which also provides the user with the flexibility to substitute alternative projections of the M^{ij}_{ag} terms derived from an exogeneous model.

4. EXTERNAL MIGRATION: THE "GLASNOST" CHALLENGE

In the early specifications of the multistate model, the role of external migration was ignored. It is only recently that Rogers has himself included external migration flows in the model specification (Rogers 1989), although alternative specifications based on accounting principles (Stone 1971; Rees and Wilson 1977) included external migration ab initio. Clearly, the projected population of any subnational unit will depend on future international flows of people across its borders as well as future intranational and vital flows.

The many ways of adapting the multistate projection model to this openness challenge are listed in Table 4.

The simplest is the world model (Table 4A) which closes the system by adding the Rest of the World as a unit to the national system of interest. The multistate model remains unaltered. Populations and vital rates for the Rest of the World have to be estimated, but this is a relatively simple matter of subtracting national population, births and deaths figures from figures for the world population estimated by the United Nations in their World Population Prospects series (e.g. United Nations 1985). The emigration and immigration flows are estimated from national sources.

The second method, the net rates model, also involves relatively little alteration of the multistate projection model. Rates of net immigration (positive or negative) are added to the intra-unit survival rates which become

$$s^{ii}_{ag} + n^{i}_{ag}$$

where n^{i}_{ag} is the rate at which immigrants are added to the population of subnational unit i in period-cohort a and gender g. The rate is defined, using transition data, as

$$n^{i}_{ag} = (P^{eo,si}_{ag} - P^{ei,so}_{ag}) \,/\, P^{i\cdot}_{ag}(t) \quad (34)$$

where the o superscript refers to outside world. This rate does not enjoy a happy theoretical existance, and could be replaced by an admission rate:

$$n^{i}_{ag} = (P^{eo,si}_{ag} - P^{ei,so}_{ag}) \,/\, P^{\cdot i}_{ag}(t+u) \quad (35)$$

but this would necessitate alterations to the model form which are not pursued here.

Neither method A nor method B alters the long-run behaviour of the multistate projection model subject to a constant rates input. That behaviour is evolution to a stable growing (or declining) population with a fixed region-age-gender distribution.

Because external migration in most countries is subject to restriction and control (particularly immigration but also emigration as a result of immigration controls in destination countries), the world and net rates models have been regarded as unsatisfactory. More popular has been a flows model (Method C in Table 4) in which a net immigration flow vector, $\underline{N}$, is added in each projection period

$$\underline{P}(t) = \underline{G}\ \underline{P}(t) + \underline{N} \quad (36)$$

If fertility rates are below replacement, in the long run the population will achieve a stationary level dependent on the size of net external balance. Esphenshade, Bouvier and Arthur 1982, Esphenshade 1987 and Rogers 1989 explore the nature of this model and its long run implication for national populations. One criticism of this net flows model is that it is not the net flow which is subject to national regulation but the gross immigration flow. Emigration is not controlled in most non-totalitarian states. A more appropriate model (D in Table 4) might be one in which emigration is modelled by using rates multiplied by origin region populations at risk together with a vector of immigration flows subject to control. Willekens and Drewe (1984) developed such a model for the Netherlands, using movement data from the Dutch population register. Using the matrix notation defined in section 3, the model for a typical period-cohort, gender group is:-

$$\underline{P}_{ag}(t+u) = \underline{S}_{ag}\ \underline{P}_{ag}(t) + \underline{F}_{ag}\ \underline{I}_{ag} \quad (37)$$

where $\underline{P}_{ag}(t+u)$, $\underline{P}_{ag}(t)$ and $\underline{S}_{ag}$ have already been defined; $\underline{F}_{ag}$ is a matrix of survival probabilities for immigration and $\underline{I}_{ag}$ is a vector of immigrations for period-cohort a and gender g. The $\underline{F}_{ag}$ matrix is defined as follows in terms of occurrence-exposure rates:

$$\underline{F}_{ag} = [\underline{I} + 0.5\underline{M}_{ag}]^{-1} \quad (38)$$

(see Willekens and Drewe 1984, pp.323-324 for the derivation which extends that for $\underline{S}_{ag}$ given above).

The final model for incorporating external migration in multistate projections (Method E in Table 4) involves using external flows and immigration rates. This model would be applicable to a country that subjected its citizens to strict emigration control, but placed no restrictions in immigration. No applications have, to my knowledge, been made with this model.

Table 5 reports some selected results from a projection of UK regional populations using Methods A to D, subject to a regime of late 1970s rates and/or flows. As fertility was depressed in this period and the net external balance heavily negative, population decline sets in later in the 20th century when methods B, C and D are used. The results of method A, where population does not decline, suggest an obvious policy option to combat the decrease: relax immigration controls!

5. THE AGE DIMENSION: THE PARSIMONY CHALLENGE (1)

A criticism frequently levelled at the multistate model is that it involves too many input variables, particularly if each has to be forecast into the future. The number of variables (rates) in the models rises in proportion to the square of the number of regions employed. When single year of age populations are demanded by the projection user, the rates become difficult to estimate reliably and computer algorithms run into storage problems.

One answer is to substitute sub-models for particular aspects of the multistate model. Models have been designed to represent the schedules of mortality, fertility and migration by age. Rogers and Planck (1984) and Rogers (1986) have integrated these component models into a projection model and Rogers and Gard (1989) have reviewed mortality schedule models. Models of component rate schedules have the general form

$$r(x) = f_1(x) + f_2(x) + \ldots + f_n(x) \qquad (39)$$

where the demographic rate at age x is a sum of a series of functions of age. The number and type of functions depends on the component represented and the subnational unit being studied.

What do such sub-models achieve? Firstly, they achieve parsimony if they are fewer parameters in the functions on the RH side of equation (39) than there are particular values required of the LH side rates. Secondly, it may be possible to link trends in the model parameters to social or economic developments (and so improve the performance of the multistate projection model in the face of surprises). Thirdly, using a model of the rate schedule forces the analyst to separate the level of a component from its distribution across the ages.

In practice, not a lot is gained in terms of number of input variables from employing model parameters compared with using observed rates when the projection model uses five year age and time intervals. For example, specfication of some migration schedules may require use of 13 parameters compared with observed rates for 18 period-cohorts. However, if single year age and time intervals are used in the projection, up to 100 observed rates may be involved and parsimony is essential. One important application of such schedule models is to help in interpolating (graduating) for single years of age from five year age group data.

Many suggestions have been made as to the functions to employ. Here the suggestions of Rogers and colleagues (Rogers, Raquillet and Castro 1978, Rogers and Castro 1981, Rogers and Planck 1983, 1984, Rogers 1986 and Rogers and Gard 1989) are generalized. Table 6 gathers together the specifications that have been proposed based on single or double exponential functions and presents them as one family.

The migration schedule is the most complex (Figure 1). A function is needed to represent the fall off of migration from infancy to the middle teens: a_1 represents the weight that the early-life rates have in the overall schedule and $alpha_1$ the steepness of the decline. A second function is needed to represent the rapid rise of migration rates as young people enter adult life (higher education, work, armed forces, marriage, cohabitation) and slower fall off as they establish themselves in suitable niches. The a_2 parameter weights this middle life function, $alpha_2$ measures the steepness of the decline slope and $lambda_2$ measures the ascent gradient. In some migration schedules, the $alpha_1$ and $alpha_2$ parameters may be approximately equal, where the migration of families (as opposed to individuals) dominates the parental years. A third function represents the local peak of migration associated with retirement from work, and a fourth function represents the rise in migration at post-retirement ages as spouses die and people move to locations with better care. Finally, there is usually a small constant term required.

It is rare for all four functions to be required. A majority of migration stream schedules will be represented by

$$m(x) = f_1(x) + f_2(x) + c \qquad (40).$$

Streams from metropolitan regions to retirement areas in countries where migration on retirement is common (USA, UK, Australia, Canada) are best modelled by

$$m(x) = f_1(x) + f_2(x) + f_3(x) + c \qquad (41)$$

while in other countries (Netherlands, Belgium, Poland) a post-retirement function provides a better description

$$m(x) = f_1(x) + f_2(x) + f_4(x) + c \qquad (42).$$

Rogers and Gard (1989) have shown that the multiexponential model (Table 6) gives results comparable with the more widely used Heligman-Pollard (1980) model (this also has three functions), but that both require fixing of the first death rate (of those born in an interval). Fertility schedules require only a single function for the fertile ages.

The most extensive use of model schedules in a multistate projection model is that carried out by the Office of Population Censuses and Surveys using methods developed by Martin and Voorhees Associates and John Bates Associates (1981) and Bates and Bracken (1982, 1987). However, they take the data reduction process one step further by using model migration parameters for about a dozen clusters of areas rather than for all 108 subnational units involved in the projection. The migration schedule models are applied to total out-migration flows and total in-migration and not to individual origin-destination flows. this spatial dimension of the multistate projection is now considered.

6. THE SPATIAL DIMENSION: THE PARSIMONY CHALLENGE

One obvious route to the reduction of the number of input variables in a multistate projection is to replace the N^2 interaction variables (gross migraproduction rates if model migration schedule parameters are available) by an alternative migration model. A large body of literature on such spatial models exists but has been largely ignored in the context of migration. Martin Voorhees Associates and John Bates Services (1981) considered the development of a gravity model but felt that it would not yield sufficiently accurate predictions. Other researchers have avoided integration of such models into a projection context because new kinds of long run behaviour of the multistate model are implied. Such views need to be revised through a more careful analysis of the type of spatial interaction model (SIM) that might be employed, and through a more thorough specification of the problem of migration rate projection. This is now attempted.

The problem is to produce a time series of projected migration rates for input to the projection of subnational populations. Virtually no attention has been paid to this problem. Analysts have explored the consequences of using alternative migration data sets from past censuses and surveys or registers (Frey 1983; Termote 1980; Rees 1988; Kupiszewski 1988) but no equivalents have been developed to the small range of high, middle and low mortality and fertility scenarios as a matter of course used in national population projections (e.g. OPCS 1987). Two exceptions are the work of P.A. Stone and associates in the 1960s (Stone 1970) and the recent development of a flexible, microcomputer version of the multistate model by Scherbov and Grechucha (1988). The Stone work was based on guessing migration scenarios for British regions using a net migration model. While of empirical interest, the work has no lasting theoretical significance. The Scherbov and Grechucha program asks the user to design scenarios by supplying new level variables for each subnational unit: for fertility gross reproduction is used, for mortality, life expectancy at age zero is employed, and for migration the gross migraproduction rate is utilized. Either the age distribution of rates in the benchmark period is re-used with the new level variables or a new schedule can be introduced. In the latter case no gain in parsimony has been made. Just a convenient method of entering data has been designed.

The problem needs more careful thought. Here ideas from the work of Willekens and Baydar (1986), who investigated time series of migration rates in the Netherlands, are used to decompose the variable array into the product of a more manageable set of variables. Some results from an investigation of a time series of migration rates and flows in the UK by Boden (1989) are also used, together with modified ideas from the official English subnational projections (Martin Voorhees Associates and John Bates Services 1981).

Let $F^{ij}_{ag}(p)$ be the count of the migration flow from subnational unit j for persons in period-cohort a and gender g in period p and $f^{ij}_{ag}(p)$ is the corresponding rate. The letter F is used to indicate that the analysis applies equally to moves (M) or transitions (T), though the population at risk used in the denominator of the rate equation would be different. The system consists of N units, A age groups, G genders and P periods. A typical projection system might consist of all the counties or county equivalents in the UK (65), for single years of age (0 ... 90+), for 2 genders and for 25 annual periods. The array of migration rates to be projected would have the size

$$(N^2 - N) \times A \times G \times P$$

$= (65^2 - 65) \times 91 \times 2 \times 25 = 18.928$ million!

Clearly, parsimony is essential.

The age dimension can be reduced in size by using the schedule models described in section 5, and by using cluster averages for sets of similar profiles. Bates and Bracken (1982, 1987) cluster origin and destination profiles but not origin-destination flows, per se.

Formally, the rates would be modelled thus

$$f^{ij}_{ag}(p) = GMR^{ij}_{g}(p)\ m^{k}_{ag} \qquad (43)$$

where GMR^{ij}_{a} is the gross migraproduction rate for the migration stream from subnational unit i to subnational j for gender g and m^{k}_{ag} is the model generated migration rate for cluster k, to which origin i belongs, the underlying model having between 6 and 13 parameters, say an average of 8. The form of the model schedule would be assumed constant, and no interaction would be assumed between origin-destination on the one hand and age-gender on the other. Assuming that about a dozen clusters (K) of similar migration profiles for each gender was needed, the number of input variables would reduce to

$$((N^2 - N) \times G \times P) + (K \times AP \times G)$$

$$= ((65^2 - 65) \times 2 \times 25) + (12 \times 8 \times 2) = 208,192$$

where AP is the average number of parameters in a cluster.

The next simplification would be to assume that the ratio of male to total GMR and female to total GMR for any out flow, r^{i}_{g}, remained fairly constant (they do not differ much anyway for most flows). The migration rates model becomes

$$f^{ij}_{ag}(p) = GMR^{ij}(p)\ r^{i}_{g}\ m^{k}_{ag} \qquad (44)$$

so that the number of input variables reduces to

$$((N^2 - N) \times P) + (N \times G) + (K \times AP \times G)$$

$$= ((65^2 - 65) \times 25) + (65 \times 2) + (12 \times 8 \times 2) = 104,322$$

Let us now consider how to decompose the interaction term $GMR^{ij}(p)$ by drawing on standard theory about spatial interaction models or SIMs (Wilson 1971, 1975; Stillwell 1978, 1986). This interaction term can be conveniently represented by a production-constrained SIM:

$$GMR^{ij}(p) = GMR^{i}(p)\ a^{ij}(p) \qquad (45)$$

where $a^{ij}(p)$ is the probability that a migrant/migration from origin unit i has destination j in period p, and $GMR^{i}(p)$ is the gross migraproduction rate for all out-migration from unit i in period p. The distribution probability a^{ij} may be modelled as

$$a^{ij}(p) = O^{i}(p)\ W^{j}(p)\ f(c^{ij})\ /\Sigma_{i}\ O^{i}(p)\ W^{j}(p)\ f(c^{ij}) \qquad (46)$$

where $O^{i}(p)$ are the total origins or out-migrations from unit i in period p, $W^{j}(p)$ is the attractiveness of unit j in period p to migrants and $f(c^{ij})$ is a function (usually negative) of the cost of interaction between unit i and j, most simply represented as an inverse power function of distance. The numerator expression on the RH side of equation (46) predicts the migration flow from unit i to unit j, and the denominator totals the flows from all origins to destination j. It is, in effect, a "competing migrants" formulation (Stouffer 1960).

The total number of origins from unit i are computed in course of the projection as

$$O^{i}(p) = \Sigma_{a}\ \Sigma_{g}\ GMR^{i}(p)\ r^{i}{}_{g}(p)\ m_{ag}\ PAR_{ag}(p) \qquad (47),$$

the cost function is represented as

$$f(c^{ij}) = (d^{ij})^{-beta} \qquad (48)$$

and the attractiveness variables $W^{j}(p)$ can, at the simplest, be set equal to trended total in-migration flows or populations, but could be a function of other variables that have been found important in migration (employment changes, amenity levels, climatic factors). The effect of new developments can be studied by developing scenarios for this vector of variables. The other important change that is being made is the use of a fixed interaction matrix, that of impedance, which no longer varies by period. The impedance function here depends on distance, but equally it could be made a function of the last observed migration matrix.

A final step would be to represent the origin unit GMRs as a time series equation such as

$$GMR^{i}(p) = GMR^{i}(0) + b^{i}_{1}\ p \qquad (49)$$

but with an upper or lower bound L, to prevent unreasonable extrapolation. The migration rates are now predicted by the following model

$$\begin{aligned} f^{ij}_{ag}(p) = {} & (GMR^{i}(0) + b^{i}_{1}\ p) \\ & \times (\Sigma_{a} \Sigma_{g}\ GMR^{i}(p)\ r^{i}_{g}(p)\ m^{k}_{ag}\ PAR^{i}_{ag}(p)\ W^{j}(p)\ (d^{ij})^{-beta}) \\ & / (\Sigma_{i} \Sigma_{a} \Sigma_{g}\ GMR^{i}(p)\ r^{i}_{g}(p)\ m^{k}_{ag}\ PAR^{i}_{ag}(p)\ W^{j}(p)\ (d^{ij})^{-beta}) \\ & \times r^{i}_{g} \\ & \times m^{k}_{ag} \qquad (50). \end{aligned}$$

The input requirements of this model are

(N x 3) elements for $GMR^{i}(0) + b^{i}_{1}\ p$

(N x P) elements for $W^{j}(p)$

(N^2 + 1) elements for $(d^{ij})^{-beta}$

(N x 2) elements for r^{i}_{g}

(K x AP x 2) elements for m^{k}_{ag}

= 6,368.

The reduction of the number of input variables could proceed further. Instead of N parameters b^{i}_{1} for each origin apply just one national parameter and national upper and lower limits, yielding (N+3) variables. Instead of N x P attractiveness parameters use the projected final populations of the previous period, so that this variable becomes endogeneous to the model, although major new developments can still be introduced if planned. Instead of N^2 distance elements, input N pairs of

X,Y coordinates from which distance could be computed yielding a reduction to 2N variables. As a result the number of input variables required would be

$$(N+3) + q(N \times P) + (N \times 2) + (N \times 2) + (K \times AP \times 2)$$

where q is the fraction of the attractiveness factors per period that need alteration (say 10%), so that the total number of input variables would be reduced, in our example, to just 683!

This account of how a very large array of migration rates might be modelled serves to illustrate how spatial as well age parsimony could be achieved. Of course, as the number of input variables is reduced, some accuracy in prediction is sacrificed. The particular set of decisions about what to hold constant over time and space and what to allow to vary as set out in equations (42) to (49) may not be the best. A programme of empirical investigation of an existing time series of migration rates would be needed. In the case of Britain, Rees (1979, 1989, Chapter 6) shows that there is far more variation in migration profile by age across migration streams than across out-migration totals (e.g. the retirement peak is particular to flows from metropolitan areas to high amenity areas, and is very muted in the profiles for total origins). The assumption that one interaction pattern will do for all age groups can be challenged; in the official subnational projection model interaction matrices for three broad age groups, 0-16/29-59, 17-28 and 60+ are used (without much justification) whereas the analysis of Boden (1989) suggests five interaction matrices for ages 0-14/25-54, 15-19, 20-24, 55-69 and 70+ would capture the different spatial patterns of family, student, yuppie, retirement and elderly age groups, to which different sets of attractiveness criteria apply, and for which (Stillwell 1986) different distance friction parameters are relevant.

Having shown that the state-space of a multistate projection model can be reduced to a manageable size, in the next section it is suggested that users of population projections require that it be massively expanded, with challenging consequences.

7. SOME ARE MORE EQUAL THAN OTHERS: THE DIVERSITY CHALLENGE

Projections normally provide the following characteristics of the population: age, gender and geographical location or age, gender and marital status or age, gender and labour force status. The numbers of households and household characteristics are generated by the cross-sectional application of headship and related rates (Corner 1987). Sometimes multistate projection models are applied to separate population groups (e.g. Rees and Ram 1987 on Indians in Bradford), but the range of population characteristics used is very limited. Both private companies for marketing purposes and public agencies for planning purposes require much more information on household composition, income and benefits, on employment characteristics, on the attributes of the housing occupied as well as on all the demographic attributes simultaneously. It is necessary to diversify the outputs of the model, while retaining the processes of birth, death and migration, fundamental to any model of population dynamics.

One solution to this problem has been around for a long while now - microsimulation (Hagerstrand 1967, Orcutt et al, 1962). In microsimulation models a population of individuals acquires characteristics (states) and undergoes transition between states as a result of random sampling of a sequence of probability distributions. The product of the simulation for each period is a list of individuals in the population for which all characteristics are known and counts of all relevant events and transitions. This list can be treated as if was the raw data from a periodic census, and crosstabulations of any kind produced from it.

The sequence of probability distributions, referred to as a probability chain, is drawn up by the constructors of the microsimulation model based partly on the logical sequence of dependence among the characteristics being modelled and partly on the availability of the necessary tabulations. The probability chain might have the form

$$Pr(a,b,c,d,e,f,g) = Pr(a)\ Pr(b|a)\ Pr(c|a,b) \times Pr(d|a)\ Pr(e|c,d)\ Pr(f,g|a,b,c) \qquad (51)$$

where Pr stands for probability and a to g are a set of attributes of interest. The full joint probability array is never generated (this would trigger off the same state space problems as exemplified in section 6). Instead, a rectangular array, NK in size, is generated of the K attributes of the population with N members, which is very much smaller than the full joint probability matrix with $n_1 \times n_2 \times \ldots \times n_k \times \ldots n_K$ elements, where n_k is the number of categories for characteristic k, once K is more than a handful. Note that the problem of constructing a suitable version of equation (50) in any applied situation resembles quite closely the problem of decomposing the migration rates array in section 6.

There is no doubt that the fields of multistate modelling with its macro-concentration on the prediction of aggregate numbers and of microsimulation with its micro-concern for simulating individuals have a lot to learn from each other. Clarke and Holm (1987) provide a recent review of the later field, and Clarke (1986), Duley, Rees and Clarke (1988) provide accounts of microsimulation models incorporating birth, death and migration processes.

8. NEW CAUSES OF MORTALITY: THE AIDS CHALLENGE

In 1988 the cumulative total of cases of Acquired Immune Deficiency Syndrome (AIDS) reported to the World Health Organization reached 250,000 and between five and ten millions are already infected by the AIDS virus (Mann et al 1989). Already AIDS is the leading cause of death among adult males aged 20 to 40 in New York City and San Francisco. The epidemic has been manifest since 1980 (patients dying of AIDS related disease) although the infection with human immunodeficiency virus (HIV) must have occurred earlier given recent estimates of the mean time between infection and onset of full blown AIDS symptoms of at least seven years. It was only in 1988 that the first cognizance of AIDS was taken in the UK in the preparation of national population projections. The Cox Report (Department of Health and Welsh Office 1988) reported on "Short term prediction of HIV infection and AIDS in England and Wales". By short term was meant to 1992 only, although one of the Working Party had been asked by the insurance industry to take projections much further into the future (Institute of Actuaries 1987, 1988a, 1988b and Wilkie 1988). In 1989 the mortality assumptions of the national projections for England and Wales were altered to incorporate AIDS deaths from Wilkie's F scenario (OPCS 1989).

The method adopted was to convert the AIDS case predictions into death rates by age (for men only) and to add these to the conventionally extrapolated (declining) mortality rates. The Government Actuary's projection experts report:

> "The assumed annual number of deaths due to AIDS would rise to about 13 thousand by the late 1990s and the total assumed number of deaths from AIDS in the next 30 years would be about 200 thousand, of which about 100 thousand would be by the end of the century. However, in the context of overall mortality, the effect of AIDS on the scale assumed would be cause only a small reduction in the expectation of life at birth of a male, before long term improvement resumed." (OPCS 1989, p.2).

Neither the Cox Report predictions nor the OPCS projections fill me with confidence. The former are myopic in their time horizon, restricted in their view of who will suffer from the disease (gays), and crude in their reliance on extrapolative methods. Only Wilkie begins to define a model with the correct states and transitions, and even he makes the assumption that only homosexual men are going to die of AIDS. The female drug addicts and their babies in New York City and the whores in East Africa's cities would have something to say about that assumption. A brief glance at WHO reports listing AIDS cases by at risk groups shows that the UK is unusual in its concentration of cases (84%) among male homosexuals.

A more realistic scenario for the spread of AIDS in the UK was incorporated in the subnational projections reported in Rees (1988). There it is assumed, following Knox (1986) that an equilibrium level of 30,600 deaths from AIDS would be reached by the end of the century, that both men and women would die, and that the epidemic would diffuse out from the current capital concentration to the cities of the Rest of the UK. As a result of these projections, the population of the UK is lower by 2026 by 1 million (AIDS deaths plus births foregone) than in an equivalent forecast without AIDS death rates added. However, the underlying methodology in no way matches the sophisticated ways in which the HIV virus is spreading.

The HIV-AIDS epidemic constitutes a considerable threat to the future of our peoples, which demands the attention of population researchers. Suitable multistate projection and microsimulation models need to be urgently developed. This is the challenge of AIDS.

9. IMPROVED PRESENTATION: THE GIS CHALLENGE

The final challenge to multistate population modelling to be discussed here arises from the new and rapidly expanding field of geographical information systems (GISs). Geographical Information Systems are computer software systems of organizing, analyzing and presenting large volumes of data about geographic locations (points, lines, areas) which are constructed to provide information for research or decision making purposes. GISs typically incorporate data management, graphic display and computer cartography, but are deficient in analytical capability.

Multistate projection and microsimulation models may generate enormous quantities of information which needs to be swiftly examined by the researcher or decision maker. The poor organization of much of this output from existing packages on paper has been criticized (Rees and Willekens 1989). If multistate models are embedded in GISs then much wider varieties of output can be provided - population and component pyramids for every unit for every time interval, time series grpahs of all rates, summary indices and composition indicators, maps of population distributions, migration flows, vital rates, block diagrams of population age evolution and many more. Note that most of these are not generated in suitable form by existing general purpose graphics packages. The GIS is not designed to produce a mountain of paper to fill the researcher's office but rather offers the potential to generate any one of thousands of different presentations as the need arises by means of a suitable algorithm linked to a chosen piece of the generated data base about the future.

Such a GIS is being gradually added to existing multistate projection packages (DIALOG - Scherbov and Grechucha 1988; MUDEA - ProGamma 1988), but an enormous potential for further intelligent display exists. The writing of such software will require considerable and expensive investments.

10. THE MULTISTATE MODEL IS DEAD. LONG LIVE THE MULTISTATE MODEL!

Models which incorporate kernels of truth about the history of people ten to have long lives, even if the assumptions they embody prove to untenable. The multistate projection model, for example, embodies many insights from previous unistate models (cohort survival, transition processes, the importance of age) and generalizes them for a set of interacting units holding populations (section 2). However, as a result, new and uncomfortable inconsistencies and confusions occurred because the new phenomenon incorporated, migration, could be measured in many different ways. These conceptual problems were investigated and the clarity of the model improved as a result (section 3). An early simplification of the model, that it applied only to a closed national system, proved unnecessary and the model was successfully opened up to the world (section 4).

However, when it came to apply the multistate model to large scale, real world systems, the large number of input variables involved proved both difficult to handle computationally and statistically doubtful. The drive towards re-introducing parsimony has concentrated, with much success, on the age dimension of the model (section 5). Much work remains to be done to achieve the same results for the spatial dimension (section 6). The customers of projection models demand, as well as simplicity, a much more varied description of the population. This could only be achieved, it was suggested, by converting the multistate macro-model of populations into a related microsimulation model of individuals.

Mention was made of two challenges to the efficacy of the multistate model. Could it be adapted to provide a simultaneous demographic projection and AIDS diffusion model? The answer to this question is still awaited. The final challenge was that of presentation: could the myriad outputs of the multistate model be made conveniently accessible to researchers and users? The answer is certainly yes if resources are devoted to the task. The multistate model has proved to be an adaptable beast and is likely to live on into the 1990s.

REFERENCES

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, 14, 889-900.

Boden, P. (1989) The analysis of internal migration in the United Kingdom using census and National Health Service Central Register data. Unpublished Ph.D. thesis, School of Geography, University of Leeds, Leeds, UK.

Clarke, M. and Holm, E. (1987) Microsimulation methods in spatial analysis and planning. Geografiska Annaler, 69B, 2, 145-164.

Corner, I.E. (1987) Developing the DOE household projection method. Building Research Establishment, Department of the Environment, Note N/83/87. Watford, Herts.

Courgeau, D. (1980) Analyse quantitative des migrations humaines. Masson, Paris.

Department of Health and Welsh Office (1988) Short-term prediction of HIV infection and AIDS in England and Wales. HMSO, London.

Duley, C., Rees, P. and Clarke, M. (1988) A microsimulation model for updating households in small areas between censuses. Working Paper 515, School of Geography, University of Leeds, Leeds, UK.

Espenshade, T.J. (1987) Population dynamics with immigration and low fertility. In Davis, K., Bernstam, M.S. and Ricardo-Campbell, R. (eds.) Below-replacement fertility in industrial societies. Cambridge University Press, Cambridge. Pp.248-261.

Espenshade, T.J., Bouvier, L. and Arthur, W.B. (1982) Immigration and the stable population model. Demography, 19, 125-133.

Graunt, J. (1662) Natural and political observations made upon the bills of mortality. Reprinted in Laslett, P. (ed.) (1973) The earliest classics: pioneers of demography. Gregg International, Farnborough.

Heligman, L. and Pollard, J.H. (1980) The age pattern of mortality. Journal of the Institute of Actuaries, 107, 49-80.

Hagerstrand, T. (1967) On Monte Carlo simulation of diffusion. In Garrison, W. and Marble, P.F. (eds.) Quantitative geography. Northwestern University Press, Evanston, Illinois.

Institute of Actuaries, AIDS Working Party (1987) Report. AIDS Bulletin No.2, Institute of Actuaries, London.

Institute of Actuaries, AIDS Working Party (1988a)
The implications of AIDS for life assurance companies.
Supplement to AIDS Bulletin No.2, Institute of Actuaries, London.

Institute of Actuaries, AIDS Working Party (1988b) Report.
AIDS Bulletin No.3, Institute of Actuaries, London.

Knox, E. (1986) A transmission model for AIDS.
European Journal of Epidemiology, 2, 165-177.

Kupiszewski, M. (1988) Application of two types of migration data to multiregional demographic projections.
Geographia Polonica, 54, 44-61.

Ledent, J. (1978) Some methodological and empirical considerations in the construction of increment-decrement life tables.
Research Memorandum RM-78-25. International Institute for Applied Systems Analysis, Laxenburg, Austria.

Ledent, J. (1980) Multistate life tables: movement versus transition perspectives. Environment and Planning A, 12, 533-562.

Ledent, J. (1981) Constructing multiregional life tables using place of birth specific migration data. In Rogers, A. (ed.)
Advances in multiregional mathematical demography.
Research Report RR-81-6. International Institute for Applied Systems Research, Laxenburg, Austria. Pp.35-49.

Ledent, J. with Courgeau, D. (1982) Migration and settlement: 15. France. Research Report RR-82-28, International Institute for Applied Systems Analysis, Laxenburg, Austria.

Ledent, J. and Rees, P.H. (1980) Choices in the construction of multiregional life tables. Working Paper 289, School of Geography, University of Leeds, Leeds/Working Paper WP-80-173, International Institute for Applied Systems Analysis, Laxenburg (Austria).

Ledent, J. and Rees, P.H. (1986) Life tables. Chapter 10 in Rogers, A. and Willekens, F. (eds.) Migration and settlement: a multiregional comparative study. Reidel, Dordrecht. Pp.385-418.

Liaw, K-L. (1979) Analysis and projections of the Canadian interregional population system. Research Report to the Social Science Research Council of Canada. Department of Geography, McMaster University, Hamilton, Ontario.

Mann, J., Chin, J., Piot, P. and Quinn, T.(1989) The international epidemiology of AIDS. In The science of AIDS, readings from Scientific American magazine, W.H. Freeman, New York. Pp.51-61.

Martin and Voorhees Associates and John Bates Services (1981) Developing the migration component of the official subnational population projections. Final Report to the Department of the

Environment.

OPCS (Office of Population Censuses and Surveys) (1987) Population projections. 1985-2025. Series PP2, No.15. Her Majesty's Stationery Office, London.

OPCS (Office of Population Censuses and Surveys) (1989) Population projections: mid-1987 based. OPCS Monitor, PP2 89/1, Information Branch, OPCS, St.Catherines House, 10 Kingsway, London WC2B 6JP.

Orcutt, G.H., Greenburger, M., Rubin, A. and Korbel, J. (1962) Micro-analysis of socioeconomic systems: a simulation study. Harper and Row, New York.

ProGamma (1988) MUDEA: a program for multiregional demographic analysis. ProGamma, Groningen, Netherlands.

Rees, P.H. (1977) The future population of East Anglia and its constituent counties (Cambridge, Norfolk and Suffolk). A report prepared for the East Anglian Economic Planning Council.

Rees, P.H. (1979) Migration and settlement: 1. United Kingdom. Research Report RR-79-3, International Institute for Applied Systems Analysis, Laxenburg, Austria.

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, Leeds/Working Paper No.50, Netherlands Interuniversity Demographic Institute, The Hague.

Rees, P.H. (1986) Choices in the construction of regional populaton projections. Chapter 7 in Woods, R.I. and Rees, P.H. (eds.) Population structures and models. Allen and Unwin, London. Pp.126-159.

Rees, P.H. (1988) A tale of two cities: the prospects for future aging of the populations of London and Harare. Working Paper 512, School of Geography, University of Leeds, UK.

Rees, P.H. (1989) Britain's population: a geographical analysis. Cambridge University Press, Cambridge. In preparation.

Rees, P.H. and Ram, S. (1987) Projections of the residential distribution of an ethnic group: Indians in Bradford. Environment and Planning A, 19, 1323-1358.

Rees, P.H. and Willekens, F. (1986) Data bases and accounts. Chapter 2 in Rogers, A. and Willekens, F. (eds.) Migration and settlement: a multiregional comparative study. Reidel, Dordrecht. Pp.19-58.

Rees, P.H. and Willekens, F. (1989) How the Dutch and English adopted multiregional models for subnational population projection. In Stillwell, J.C.H. and Schoelten, H. (eds.) The comparative population geography of the UK and the Netherlands.

Kluwer, Amsterdam. In press.

Rees, P.H. and Wilson, A.G. (1977) Spatial population analysis. Edward Arnold, London.

Rogers, A. (1968) Matrix analysis of interregional population growth and distribution. University of California Press, Berkeley and Los Angeles.

Rogers, A. (1971) Matrix methods in urban and regional analysis. Holden-Day, San Francisco.

Rogers, A. (1975) Introduction to multiregional mathematical demography. Wiley, New York.

Rogers, A. (1985) Regional population projection models. Sage Publications, Beverly Hills.

Rogers, A. (1986) Parameterized multistate population dynamics and projections. Journal of the American Statistical Association, 81, 48-61.

Rogers, A. (1989) The multistate stable population model with immigration. Working Paper WP-89-4, Population Program, Institute of Behavioral Science, University of Colorado, Boulder, Colorado.

Rogers, A. and Castro, L.J. (1981a) Model migration schedules. Research Report RR-81-30, International Institute for Applied Systems Analysis, Laxenburg, Austria.

Rogers, A. and Castro, L.J. (1981b) Age patterns of migration: cause-specific profiles. In Rogers, A. (ed.) Advances in multiregional mathematical demography. Research Report RR-81-6. International Institute for Applied Systems Research, Laxenburg, Austria. Pp.125-159.

Rogers, A. and Gard, K. (1989) Applications of the Heligman-Pollard model migration mortality schedule. Working Paper WP-89-1, Population Program, Institute of Behavioral Science, University of Colorado, Boulder, Colorado.

Rogers, A. and Ledent, J. (1976) Increment-decrement life tables: a comment. Demography, 13, 287-290.

Rogers, A. and Planck, F. (1983) MODEL: a general program for estimating parameterized schedules of fertility, mortality, migration, marital and labour force status transitions. Working Paper WP-83-102, International Institute for Applied Systems Analysis, Laxenburg, Austria.

Rogers, A. and Planck, F. (1984) Parameterized multistate population dynamics and projections. Working Paper WP-84-1, Population Program, Institute of Behavioral Science, University of Colorado, Boulder, Colorado.

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 Willekens, F. (eds.) (1986) Migration and settlement: a multiregional comparative study. Reidel, Dordrecht.

Scherbov, S. and Grechucha, V. (1988) DIAL - a system of modeling multidimensional demographic processes. Working Paper WP-88-36, International Institute for Applied Systems Analysis, Laxenburg, Austria.

Stillwell, J.C.H. (1986) The analysis and projection of interregional migration in the United Kingdom. Chapter 8 in Woods, R.I. and Rees. P.H. (eds.) Population structures and models. Allen and Unwin, London. Pp.160-202.

Stockwell, E.G. (ed.) (1976) The methods and materials of demography: condensed edition. Academic Press, New York.

Stone, P.A. (1970) Urban development in Britain. Vol.1: population trends and housing. Cambridge University Press, Cambridge.

Stone, R. (1971) Demographic accounting and model building. Organization for Economic Cooperation and Development, Paris.

Termote, M. (1980) Migration and settlement: 6. Canada. Research Report RR-80-29, International Institute of Applied Systems Analysis, Laxenburg, Austria.

United Nations (1970) Manual VI: methods of measuring internal migration. ST/SOA/Series a/47. United Nations, New York.

United Nations (1975) Towards a system of social and demographic statistics. Department of Economic and Social Affairs, Studies in Methods, Series F, No.18. United Nations, New York.

United Nations (1985) World population prospects: estimates and projections as assessed in 1982. Department of International Economic and Social Affairs, Population Studies No.86. United Nations, New York.

United Nations (1989) A manual for subnational population projections. Population Division, Department of International Economic and Social Affairs, New York. In preparation.

Wilkie, D. (1988) Forecasting AIDS using an actuarial model. Appendix 10 in Department of Health and Welsh Office (1988) Short-term prediction of HIV infection and AIDS in England and Wales. HMSO, London. Pp. 65-69.

Willekens, F. and Baydar, N. (1986) Forecasting place-to-place migration with generalized linear models. Chapter 9 in Woods, R.I. and Rees, P.H. (eds.) Population structures and models. Allen and Unwin, London.

Willekens, F. and Drewe, P. (1984) A multiregional model for regional demographic projection. In ter Heide, H. and Willekens, F. (eds.) Demographic research and spatial policy. Academic Press, London. Pp.309-334.

Willekens, F., Por, A. and Raquillet, R. (1979) Entropy, multiproportional and quadratic techniques for inferring detailed migration patterns from aggregate data. Working Paper WP-79-88, International Institute for Applied Systems Analysis, Laxenburg, Austria.

Willekens, F., Por, A. and Raquillet, R. (1981) Entropy, multiproportional and quadratic techniques for inferring patterns of migration from aggregate data. In Rogers, A. (ed.) Advances in multiregional demography. Research Report RR-81-6, International Institute for Applied Systems Analysis, Laxenburg, Austria. Pp.83-124.

Willekens, F. and Rogers, A. (1978) Spatial population analysis: methods and computer programs. Research Report RR-78-18, International Institute for Applied Systems Analysis, Laxenburg, Austria.

TABLE 1. A classification of subnational population projection models

CLASSIFICATION OF ROGERS	SPATIAL AND AGE DIMENSIONS	COMPONENT DIMENSION			
		Total change	Natural change and net migration	Natural change and gross migration	
				ACCOUNTING DIMENSION	
			Movement	Movement (M)	Transition (T)
	Single unit				
No age and no location	All ages	MODEL 1	MODEL 2	MODEL 3M	MODEL 3T
Age and no location	Age detail		MODEL 4	MODEL 5M	MODEL 5T
	Multiunit				
Location and no age	All ages			MODEL 6M	MODEL 6T
Age and location	Age detail			MODEL 7M	MODEL 7T

Sources: Table II.12 in United Nations (1989), Rogers (1985) and Rogers (1989), Figure A.1.

TABLE 2. Illustrative computations for a multiregional projection model

A. SURVIVAL COMPUTATIONS FOR THE EAST ANGLIAN POPULATION (EQ. 1)

Period-cohort		East Anglia	Staying in East Anglia		South East	Migration to East Anglia		Rest of Britain	Migration to East Anglia		Total popn.
1966	1971	Popn.	Rate	Flow	Popn.	Rate	Flow	Popn.	Rate	Flow	
		1	2	3	4	5	6	7	8	9	10
0-4	5-9	61,339	.854937	52,441	690,681	.008059	5,566	1,484,932	.002993	4,445	62,452
5-9	10-14	54,822	.886797	48,616	605,573	.006179	3,742	1,349,948	.002179	2,941	55,299
10-14	15-19	52,048	.898113	46,745	549,814	.004898	2,693	1,225,832	.002003	2,455	51,893
15-19	20-24	57,528	.808267	46,498	649,508	.010910	7,086	1,343,169	.004222	5,671	59,255
20-24	25-29	48,563	.797500	38,729	589,862	.012813	7,558	1,122,686	.004236	4,756	51,043
25-29	30-34	43,869	.851991	37,376	526,277	.008292	4,364	1,026,140	.003353	3,441	45,181
30-34	35-39	44,117	.884444	39,019	503,437	.006978	3,513	1,008,803	.002215	2,234	44,766
35-39	40-44	46,988	.909934	42,756	529,649	.005483	2,904	1,079,708	.001888	2,039	47,699
40-44	45-49	49,972	.924818	46,215	567,230	.004802	2,724	1,172,384	.003548	4,160	53,099
45-49	50-54	46,676	.934077	43,599	539,789	.004485	2,421	1,102,766	.003428	3,780	49,800
50-54	55-59	49,044	.941991	46,199	573,830	.004496	2,580	1,149,173	.003415	3,924	52,703
55-59	60-64	50,204	.926341	46,506	563,692	.006486	3,656	1,145,995	.000906	1,038	51,200
60-64	65-69	46,674	.898016	41,914	506,120	.003047	1,542	1,053,987	.001529	1,612	45,068
65-69	70-74	41,448	.854806	35,430	428,726	.004726	2,026	910,230	.002326	2,117	39,573
70-74	75-79	33,363	.772772	25,782	349,234	.004201	1,467	705,316	.002280	1,608	28,857
75+	80+	51,445	.559355	28,776	547,793	.002990	1,638	943,241	.001738	1,639	32,053
0+	5+	778,100	.856704	666,601	8,721,215	.006361	55,480	17,824,310	.002685	47,860	769,941

Notes: Col.3 = Col.1 x Col.2; Col.6 = Col.4 x Col.5;
Col.9 = Col.7 x Col.8; Col.10 = Col.3 + Col.6 + Col.9.

B. FERTILITY COMPUTATIONS FOR EAST ANGLIA (EQ.s 2,3)

Period-cohort		Initial popn. 1966	Final popn. 1971	Average popn. 1966-71	Fertility rate	Births	Sex proportion	Births, female
1966	1971	1	2	3	4	5	6	7
10-14	15-19	52,048	51,893	51,971	.070289	3,653		
15-19	20-24	57,528	59,255	58,392	.555093	32,413		
20-24	25-29	48,563	51,043	49,803	.901572	44,901		
25-29	30-34	43,869	45,181	44,525	.593195	26,412		
30-34	35-39	44,117	44,766	44,442	.266617	11,849		
35-39	40-44	46,988	47,699	47,344	.097626	4,622		
40-44	45-49	49,972	53,099	51,536	.017037	878		
45-49	50-54	46,676	49,800	48,238	.000933	45		
Totals					2.502362	124,773	.484047	60,396

C. SURVIVAL COMPUTATIONS FOR INFANT PERIOD-COHORT (EQ. 4)

Period-cohort		East Anglia	Staying in East Anglia		South East	Migration to East Anglia		Rest of Britain	Migration to East Anglia		Total popn.
	1971	Births	Rate	Flow	Births	Rate	Flow	Births	Rate	Flow	
		1	2	3	4	5	6	7	8	9	10
Birth	0-4	60,396	.896285	54,132	656,985	.006382	4,193	1,440,960	.003113	4,486	62,811

Notes: as for Panel A. Source: Rees, 1977.

TABLE 3. Type of migration measure and survival rate estimation method

Source	Migration measure	Methods of estimation of rates used in the multistate population projection model
		Transition approaches
Census	Transitions/ Migrants	T1. Divide migrant flows by start of period populations
		T1.1 Estimate surviving stayers from migrant flows and end of period populations
		T1.2 Estimate surviving stayers from migrant flows, start of period populations and mortality data directly (Rogers 1968)
		T1.3 Estimate surviving stayers from migrant flows, start of period populations and mortality in an iterative model (Rees and Wilson 1977)
		T2. Multiply survival rates by rates of migration conditional on survival (Ledent 1978; Ledent and Rees 1986)
Census	Last migration/ migrants	T3. No effective method developed but see Courgeau (1980) and United Nations (1989) for reviews
Census	Lifetime migration	T4. No effective method developed but see United Nations (1970) and (1989) for reviews
		Movement approaches
Register	Moves/ Migrations	M1. Use occurrence-exposure rates in multistate version of probability estimation equation of single state demography (Rogers and Ledent 1976)
		M2. Use occurrence-exposure rates directly in an iterative projection model (Rees 1984, 1986)

TABLE 4. Alternative methods for incorporating external migration in the multistate projection model

Method	Component		Application
	Immigration	Emigration	
A. World model	Migration and survival rates from the Rest of the World	Emigration rates	Rees (1986)
B. Net rates model	(Immigration (admission rates ((Net immigration (rates	Emigration rates	Liaw (1979) Rees (1986)
C. Net flows model	(Immigration (flows ((Net immigration (flows	Emigration flows	Rees (1986) Rogers (1989)
D. In-flows, out-rates model	Immigration flows	Emigration rates	Willekens and Drewe (1984) Rees (1986)
E. In-rates, out-flows model	Immigration rates	Emigration flows	-

TABLE 5. Projected populations of selected UK regions in 2006 based on alternative methods for incorporating external migration

Method	Regions				
	Greater London	Rest of the South	Industrial Heartland	Periphery	United Kingdom
A. Closed	6.510	22.305	16.062	12.231	57.106
B. Net rates	6.055	21.765	15.683	12.010	55.515
C. Net flows	6.075	21.802	15.689	12.001	55.569
D. Imm. flows Em. rates	6.120	21.708	15.702	12.019	55.548

Source: computed from Table 7.10 in Rees (1986)
Notes:
1. Regional definitions are as follows in terms of standard regions:
   Greater London = former county
   Rest of the South = Rest of South East, East Anglia, East Midlands and South West
   Industrial Heartland = West Midlands, North West, Yorkshire and and Humberside
   Periphery = Wales, North, Scotland and Northern Ireland
2. Populations are in millions.
   1981 population = 56.289; 1987 population estimate = 56.930 millions.
3. The projections employ constant 1976-81 rates or flows. Since 1981 fertility rates have been higher, mortality rates lower and net emigration has become slight net immigration, so that the table's projections are underestimates.

TABLE 6. A family of multiexponential models of the variation of component rates with age

Component / Age range	Function of age
Migration	$m(x) = f_1(x) + f_2(x) + f_3(x) + f_4(x) + f_5(x)$
Early-life	$f_1(x) = a_1 \exp(-alpha_1 x)$
Middle-life	$f_2(x) = a_2 \exp(-alpha_2(x-mu_2)-\exp(-lambda_2(x-mu_2)))$
Retirement	$f_3(x) = a_3 \exp(-alpha_3(x-mu_3)-\exp(-lambda_3(x-mu_3)))$
Post-retirement	$f_4(x) = a_4 \exp(alpha_4 x)$
Constant	$f_5(x) = c$
Mortality	$d(x) = f_1(x) + f_2(x) + f_4(x)$
Infant	$f_0(x) = a_0 = d(0)$ observed for age 0 only
Early-life	$f_1(x) = a_1 \exp(-alpha_1 x)$
Middle-life	$f_2(x) = a_2 \exp(-alpha_2(x-mu_2)-\exp(-lambda_2(x-mu_2)))$
Late-life	$f_4(x) = a_4 \exp(alpha_4 x)$
Fertility	$f(x) = f_2(x)$
Middle-life	$f_2(x) = a_2 \exp(-alpha_2(x-mu_2)-\exp(-lambda_2(x-mu_2)))$
Widowhood	$w(x) = f_2(x) + f_4(x)$
Middle-life	$f_2(x) = a_2 \exp(-alpha_2(x-mu_2)-\exp(-lambda_2(x-mu_2)))$
Late-life	$f_4(x) = a_4 \exp(alpha_4 x)$

Source: adapted from Rogers (1986) and Rogers and Gard (1989).
Notes: the fertility model applies also to marriage and divorce.

FIGURE 1. A scheme for modelling component rate schedules by age