Working Paper 234

REGIONAL POPULATION PROJECTION MODELS AND ACCOUNTING METHODS

Philip H. Rees

Paper presented at a meeting of the Royal Statistical Society, School of Hygiene and Tropical Medicine, University of London, Tuesday, 17th October, 1978.

School of Geography
University of Leeds
Leeds LS2 9JT

October, 1978.

Contents

Abstract

The paper outlines how population accounting methods may be used in a variety of ways in regional population projection. Modes of use of regional projection models are reviewed. A multiregional population projection model is then built using an accounting framework, and this is then applied to a four region case study. The final sections of the paper discuss how accounting methods can help in the calibration of regional projection models.

Acknowledgments

My thanks are due to two of my students who have just left the University of Leeds for pastures new. I am grateful to John Jenkins for his help in clarifying many of the issues involved in regional projection while collaborating on an earlier paper (Jenkins and Rees, 1977). The computer program described in that paper has proved an invaluable learning tool. I must also thank Dil Thomas for his help in developing and testing out ideas about accounts calibration with Soviet data, the results of which are to be reported in Thomas and Rees, 1978.

"But we haven't just pulled figures out of the sky. Well, not often."

McEvedy and Jones (1978, p.11).

"On the debit side, however, the construction of a full scale account-based model will frequently require far greater computer and data-resources than simpler traditional models, and it may not always be certain that the gain in conceptual correctness and relative accuracy will be sufficient to justify the extra effort, expense and time taken in the production of forecasts."

Baxter and Williams (1978, p.43).

Part One: the framework and materials for discussion

1. The field of interest: some definitions and points of view

Western society has long been characterized by a concern for the future. Individuals plan for their own future through investment in education now in order to gain career advancement or an enhanced income stream in the future. Sociologists have taken this future orientation to be one of the salient features that distinguish for middle classes of society from the lower classes. When future orientation is coupled with the Protestant ethic the issue is planning in all its many guises, private or public.

In order to invest effectively, to plan urban infra-structure, and to organize welfare systems (such as pension schemes) knowledge of the demands likely to be made by people in the future is essential. So the practice of population projection arose, at first to meet the needs of national governments in working out national insurance and pension schemes, and more recently to meet the needs of regional or local governments.

The methodology of population projection at the national level was developed in the earlier part of this century - in part through the work of biologists (Lotka, 1907; Leslie, 1945), in part by demographers (Whelpton, 1928; Whelpton, 1936). The discrete components-of-growth cohort survival model that stems from this earlier work is now the standard model for carrying out national population projections.

The methods used in projecting the population of sub-national territorial units have been developed more recently, particularly since 1960. The methods have been reviewed recently in monographs both in the U.S.A. (Pittenger, 1976) and in the U.K. (Baxter and Williams, 1978). Demographic methods have been used; methods based on deriving population from housing supply have been applied; population has been projected as the labour supply consequence of an economic model or a transport-land use model (Batty, 1976). Simpler trend or extrapolation techniques have tended to go out of fashion, except in the distribution to small areas of regional totals projected by a more sophisticated model. To date, no consensus has been arrived at as to the most appropriate method of population projection at the regional or local level, as the second quotation that heads this section indicates.

The intention in this paper is to review some of the problems involved in defining projection methods for regional populations in the remainder of Part One of the paper. In the second part of the paper an attempt is made to fill a gap in present methodology for projection, that of calibration. Calibration is the process of finding variable and/or parameter values for the projection model that fit the observed situation. This step is a key one in the modelling of most systems but has been curiously neglected in the population field.

The focus of concern will be rather narrower than that of Pittenger or of Baxter and Williams. Attention will be concentrated on the demographic component model and therefore on larger scale regions rather than smaller areas. Only a small part of the projection modelling map presented in Figure 1 will be explored. Figure 1 has been drawn up as a map in which latitude refers to position on a pure-applied scale and longitude to position on a modelling sequence scale (for demographic component type models, that is). Topics are located as places on the map and the places are connected by a transport network along which information flows. All places on the network are connected to each other, but in general only single links will be considered.

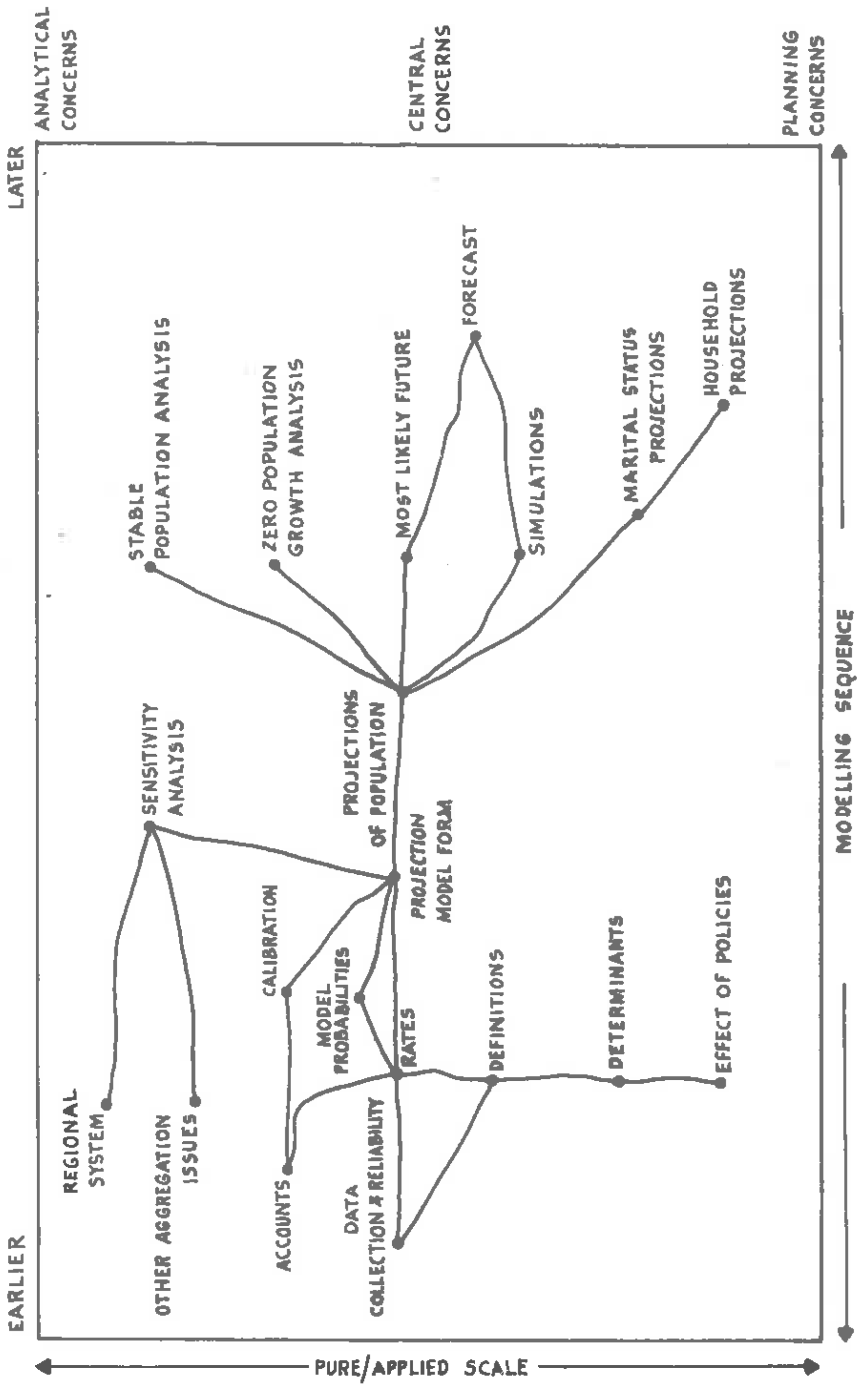

Figure 1. A 'map' of topics of concern in regional population projection

Of central concern to all involved in population projection are the collection of data on population stocks and flows, the proper definition of rates that enter the model, projection model form, projections of the population. The analyst will be more interested, for example, in the sensitivity of the projections to model form, to regional definitions and to the age/sex disaggregations. Practitioners of projection will be more concerned with having a model that works for an area of direct planning concern.

It is appropriate to conclude this section of the paper with more precise definitions of some of the terms used in Figure 1 that will be frequently employed in the remainder of the paper.

By <u>projection</u> is meant the evaluation of the consequences of a given future trajectory of demographic rates in terms of the numbers of people likely to be living at future points in time. If the trajectory is a particularly variable one the term <u>simulation</u> is sometimes used. There are clearly as many projections as there are rate trajectories, and from these the practitioner has to select the one that represents the <u>most likely future</u> (as in Bonsall, Champernowne, Mason and Wilson, 1977) or the single projection to be adopted as the basis for planning and investment, <u>the forecast</u>. The term <u>population estimate</u> is confined to a figure for a regional population stock estimated for some time in the past. Thus, the population graphs in McEvedy and Jones (1978) consist mostly of <u>estimates</u> (up to 1975) and partially of projected figures (1976-2000).

2. <u>A sketch of the various modes of use of regional projection models</u>

Although there is probably broad agreement about the verbal meaning of the terms described in the last paragraph, there are a number of points of detail that need clarification in the context of regional projection. This is best done in the form of a series of graphs (Figure 2).

The lines on the graphs represent the path taken by the population of a single region through time (past, present and future) but they can be understood to trace the path between states of the whole system (all

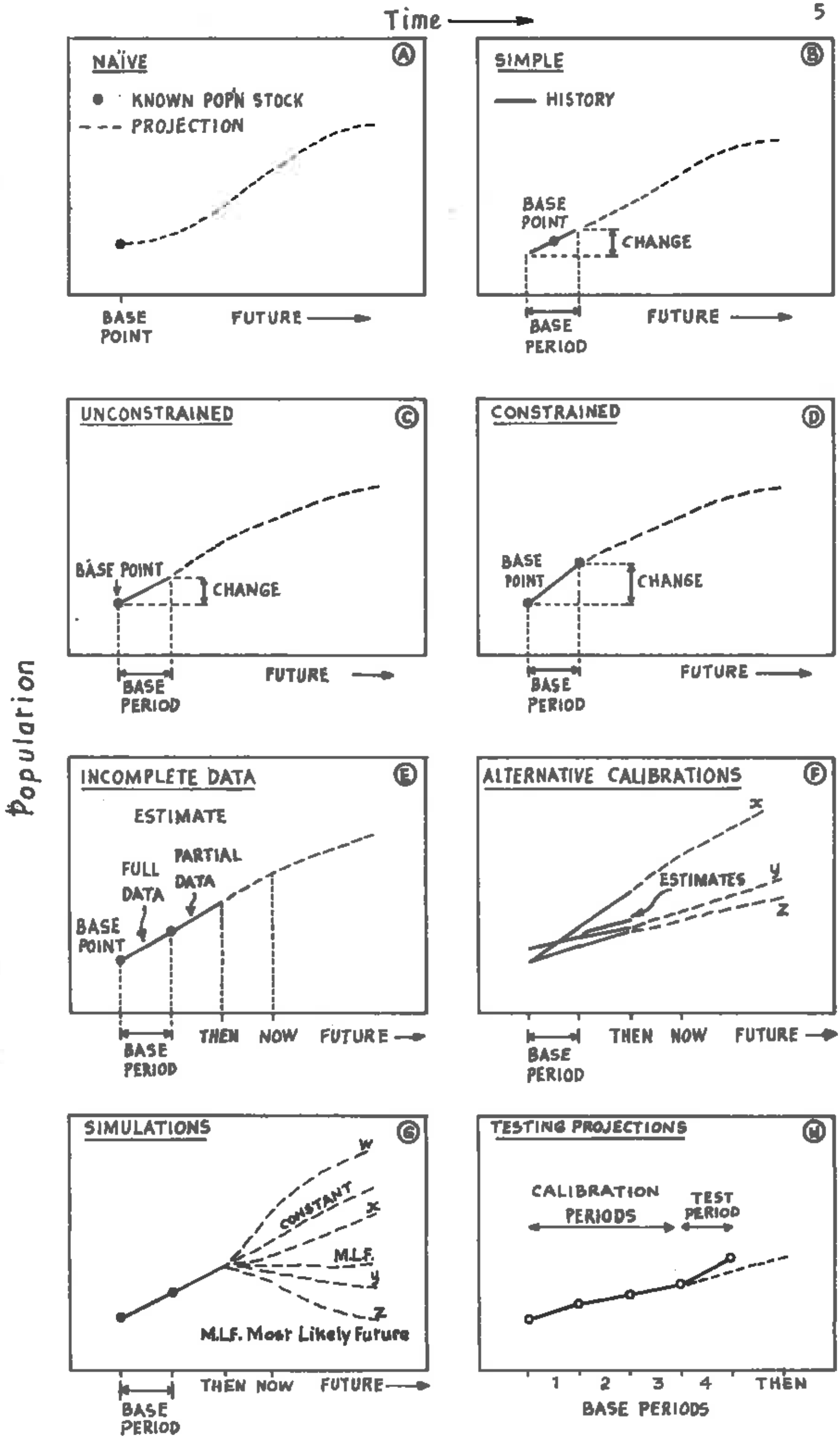

Figure 2 Different views of the projection process

regional population stocks and flows) through time. In the first graph (Figure 2.A) the situation is one where the regional population is known for a base point in time and is projected forward from then. In practice, this view is naïve and must give way to the recognition (in Figure 2.B) that the change rates entering a projection model are measured (at least initially) for a base period and that the population stock at the base point is usually observed in the middle of the base period. A mismatch of the base period and base point can then occur if care is not taken in model design and the alternative of placing the base point at the beginning of the base period (Figure 2.C) is often to be prefered. If this is done it is immediately clear that we may have two possible base points - at the start of the base period and at the end of the base period (Figure 2.D). Or, to put it another way, there now exists a check that the population flows measured in the base period do convert the initial population into the final. This may not be the case and the values of the population flows may have to be adjusted or constrained so that they do reproduce the observed change in population stocks.

In practice the data for input to the projection model may not come neatly packaged and available for the latest calendar year. It may be that all the required data is available for some base period in the recent past, and that partial data is available for some portion of the time between the end of the base period and the present. This point in time is labelled 'then' is Figure 2.E. This situation is faced when a multiregional population projection model is constructed for British standard regions. Now in 1978 the last period for which all requisite data is fully available in published form is 1970-71 (the year prior to the 1971 Census) because the census is the only full source for internal interregional migrant data. Information on regional births, deaths and external migration is published in full for the years up to 1975, and reasonable estimates can be made for 1976 and 1977. Ways in which this gap in internal migration data can be filled will be described later in the paper but the length of time between the end of the base period and 'then' is characteristically greater for projection models embodying interregional migration then those using only net migration.

This period of partial data can, in certain circumstances, be used to advantage. Alternative methods of adjusting the base period information to ensure consistent picture of population change can be compared. Alternative "calibrations" of the base period projection model rates are prepared and projections are carried out for the period between the end of the base period and 'then' using the partial data that are known. The 'projected' populations can be compared with the current estimate series and the base period calibration which gives the best fit can be selected. In this limited sense the projection can be tested, though the test will be dependent on the reliability of the population estimate series. Thus, of the X, Y and Z projections in Figure 2.F, projection Y would be selected because the populations projected between the end of the base period and 'then' approximated most closely those of the estimate series.

The next graph in the Figure 2 series (Figure 2.G) illustrates the methodology of simulation or variant projections in which whole suites of alternative scenarios for the rates can be examined and compared against either the results of the continuation of the latest available rates or of the results of the most likely future trajectories of rates.

The final graph in Figure 2 (Figure 2.H) reveals what sort of data are required in order to test population projection models. At least three periods are required to develop an idea of trends in the model rates and an additional period is needed for testing. Such data are readily available for the fertility and mortality components of demographic models but less frequently available for the migration component. Had there been a 1976 Census and had there been questions about residence five years and one year prior to census night then in the U.K. we would have had three base periods giving five year internal migrant data (1961-66, 1966-71, 1971-76) and four base periods for one year migrant data (1960-61, 1965-66, 1970-71 and 1975-76). Even if the 1976 Census had been carried out there would still have been the problem of converting the earlier data to current regional and local definitions. As it is proper testing of a multiregional population projection model will have to await the results of the 1981 Census. This perhaps explains why population modellers have generally failed to give proper attention to issues of calibration and testing which are treated as a matter of course in many other fields (Wilson, 1974).

3. Notation, the concept of population accounts and a model example

In order to explore some of the issues raised in discussing the contents of Figure 2, a suitable notation needs to be defined for building multiregional population projection models, and for defining the population accounts that underlie them and from which the rates used in the projection model are derived. Here a very abbreviated version is presented and more details are given in Rees and Wilson (1977).

Let $\underline{K}$ represent the population accounts matrix, the rows of which represent the states in which the population originates in a period - either existence (subscript E or superscript ε) in a region or birth (subscript B or superscript β) there - and the columns of which represent the states in which the population ends up in a period - either surviving in a region at the end of the period (subscript S or superscript σ) or dying there before the end of the period (subscript D or superscript δ). The accounts matrix $\underline{K}$ consists of 4 submatrices

$$\underline{K} = \left[\begin{array}{c|c} \underline{K}_{ES} & \underline{K}_{ED} \\ \hline \underline{K}_{BS} & \underline{K}_{BD} \end{array}\right] \tag{1}$$

Each quadrant has the same structure of sub-submatrices one for each interregional transition. Thus the $\underline{K}_{ES}$ submatrix is composed as follows

$$\underline{K}_{ES} = \begin{bmatrix} \underline{K}^{\varepsilon(1)\sigma(1)} & \underline{K}^{\varepsilon(1)\sigma(2)} & \cdots & \underline{K}^{\varepsilon(1)\sigma(j)} & \cdots & \underline{K}^{\varepsilon(1)\sigma(N)} \\ \underline{K}^{\varepsilon(2)\sigma(1)} & \underline{K}^{\varepsilon(2)\sigma(2)} & \cdots & \underline{K}^{\varepsilon(2)\sigma(j)} & \cdots & \underline{K}^{\varepsilon(2)\sigma(N)} \\ \vdots & \vdots & & \vdots & & \vdots \\ \underline{K}^{\varepsilon(i)\sigma(1)} & \underline{K}^{\varepsilon(i)\sigma(2)} & \cdots & K^{\varepsilon(i)\sigma(j)} & \cdots & \underline{K}^{\varepsilon(i)\sigma(N)} \\ \vdots & \vdots & & \vdots & & \vdots \\ K^{\varepsilon(N)\sigma(1)} & \underline{K}^{\varepsilon(N)\sigma(2)} & \cdots & K^{\varepsilon(N)\sigma(j)} & \cdots & \underline{K}^{\varepsilon(N)\sigma(N)} \end{bmatrix} \tag{2}$$

The typical $\underline{K}^{\varepsilon(i)\sigma(j)}$ sub-submatrix in the $\underline{K}_{ES}$ quadrant of the accounts has equivalent $\underline{K}^{\beta(i)\sigma(j)}$ in the $\underline{K}_{BS}$ quadrant, $\underline{K}^{\varepsilon(i)\delta(j)}$ in the $\underline{K}_{ED}$ quadrant and $\underline{K}^{\beta(i)\delta(j)}$ in the $\underline{K}_{BD}$ quadrant. Each sub-submatrix is disaggregated by the age group associated with the origin state and with the destination state. If the age group interval is equal to the time period length the $\underline{K}^{\varepsilon(i)\sigma(j)}$ sub-submatrix has the structure of a transposed Leslie matrix

$$\underline{K}^{\varepsilon(i)\sigma(j)} = \begin{bmatrix} 0 & K_{12}^{\varepsilon(i)\sigma(j)} & 0 & \dots 0 & \dots 0 \\ 0 & 0 & K_{23}^{\varepsilon(i)\sigma(j)} & \dots 0 & \dots 0 \\ 0 & 0 & 0 \quad \ddots & \dots K_{r-1r}^{\varepsilon(i)\sigma(j)} & \dots 0 \\ 0 & 0 & 0 & \dots 0 \quad \ddots & \dots K_{R-1R}^{\varepsilon(i)\sigma(j)} \\ 0 & 0 & 0 & \dots 0 & \dots K_{RR}^{\varepsilon(i)\sigma(j)} \end{bmatrix} \quad (3)$$

where r refers to a discrete age group of which there are R. Transitions occur between one age group and the next - people aged 20-24 at time t will be aged 25-29 at time t+5, assuming survival.

The $\underline{K}^{\varepsilon(i)\delta(j)}$ sub-submatrix has the same structure as specified in Equation (3) with a change of one superscript, σ to δ, if by $K_{r-1r}^{\varepsilon(i)\delta(j)}$ is meant all persons alive in region i migrating to region j and dying there who are in the cohort that is initially aged r-1 and which ages into age group r, death occurring in either age group. If, on the other hand, such a cohort classification of deaths used in Rees (1978) is not available, and only deaths by age group at death are published, then a pair of terms $K_{r-1r-1}^{\varepsilon(i)\delta(j)}$ and $K_{r-1r}^{\varepsilon(i)\delta(j)}$ will be present with the former being located on the principal diagonal of $\underline{K}^{\varepsilon(i)\delta(j)}$.

The origin state age classification of the $\underline{K}_{BS}$ and $\underline{K}_{BD}$ quadrants is that of age group of mother (at time of maternity or at the start of the period) and only the first column of each $\underline{K}^{\beta(i)\sigma(j)}$ and $\underline{K}^{\beta(i)\delta(j)}$ sub-submatrix is occupied thus:

$$\underline{K}^{\beta(i)\sigma(j)} = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \underline{K}_{31}^{\beta(i)\sigma(j)} & 0 & \cdots & 0 \\ \vdots & \vdots & & \\ \underline{K}_{r1}^{\beta(i)\sigma(j)} & 0 & \cdots & 0 \\ \vdots & \vdots & & \\ \underline{K}_{10\,1}^{\beta(i)\sigma(j)} & 0 & \cdots & 0 \\ \vdots & \vdots & & \\ 0 & 0 & \cdots & 0 \end{bmatrix} \tag{4}$$

The elements of the accounts matrix are the counts of people falling in those categories, the raw data used in constructing rates for input to the projection model. Accounts matrices differ from simple matrices of population transitions in always being closed (Stone, 1971): a state is always included that captures all exits from the regional system of interest and supplies all exogenous entries. Stone (1971) calls this state "the outside world" and Rees and Wilson (1977) call it "the rest of the world". Inclusion of the rest of the world as a state means that the row and column sums have additional meaning. The expanded accounts matrix, usually known as an accounts table, called here $\underline{KT}$ has the following form

$$\underline{KT} = \left[\begin{array}{cc:c} \underline{K}_{ES} & \underline{K}_{ED} & \underline{k}_{E} \\ \underline{K}_{BS} & \underline{K}_{BD} & \underline{k}_{B} \\ \hdashline \underline{k}_{S} & \underline{k}_{D} & \underline{k} \end{array}\right] \tag{5}$$

where $\underline{k}_E$ is a column vector of initial regional populations disaggregated by age, $\underline{k}_B$ is a column vector of regional births broken down by age group of mother, $\underline{k}_S$ is a row vector of final regional populations broken down by age, $\underline{k}_D$ is a row vector of regional deaths disaggregated by age, and $\underline{k}$ is a scalar, the total number of persons moving through the system over the period. Usually the rest of the world (region N) terms are set to zero so that the row totals refer to immigrant sums to the system and the column totals to emigrant sums from the regional set.

It is useful to break up the accounts matrix a little further than in Equation (5) in order to distinguish states within the regional set of particular interest and elements outside it. The labels I and 0 will be used to distinguish <u>i</u>nside and <u>o</u>utside (the system of interest) sets of regions and will be applied to relevant submatrices. Thus, $\underline{K}$ becomes

$$\underline{K} = \left[\begin{array}{cc:cc} \underline{K}_{ES}^{II} & \underline{K}_{ES}^{IO} & \underline{K}_{ED}^{II} & \underline{K}_{ED}^{IO} \\ \underline{K}_{ES}^{OI} & \underline{K}_{ES}^{OO} & \underline{K}_{ED}^{OI} & \underline{K}_{ED}^{OO} \\ \hdashline \underline{K}_{BS}^{II} & \underline{K}_{BS}^{IO} & \underline{K}_{BD}^{II} & \underline{K}_{BD}^{IO} \\ \underline{K}_{BS}^{OI} & \underline{K}_{BS}^{OO} & \underline{K}_{BD}^{OI} & \underline{K}_{BD}^{OO} \end{array}\right] \qquad (6)$$

The submatrix $\underline{K}_{ES}^{II}$ contains only exist-survive population transitions between regions inside the system of interest, the $\underline{K}_{ES}^{IO}$ submatrix contains only transitions from regions inside the system to those outside, the $\underline{K}_{ES}^{OI}$ submatrix holds only population flows from outside regions to inside, and $\underline{K}_{ES}^{OO}$ contains only flows between regions outside the regional set of interest. Similar classifications hold for the other quadrants. If this latter submatrix is set to zero and totals are added after the fashion of definition - Equation (5), the result is

$$\underline{KT} = \left[\begin{array}{cc:cc:c} \underline{K}_{ES}^{II} & \underline{K}_{ES}^{IO} & \underline{K}_{ED}^{II} & \underline{K}_{ED}^{IO} & \underline{k}_{E}^{I} \\ \underline{K}_{ES}^{OI} & \underline{0} & \underline{K}_{ED}^{OI} & \underline{0} & \underline{k}_{E}^{O} \\ \hdashline \underline{K}_{BS}^{II} & \underline{K}_{BS}^{IO} & \underline{K}_{BD}^{II} & \underline{K}_{BD}^{IO} & \underline{k}_{B}^{I} \\ \underline{K}_{BS}^{OI} & \underline{0} & \underline{K}_{BD}^{OI} & \underline{0} & \underline{k}_{B}^{O} \\ \hdashline \underline{k}_{S}^{I} & \underline{k}_{S}^{O} & \underline{k}_{D}^{I} & \underline{k}_{D}^{O} & \underline{k} \end{array}\right] \qquad (7)$$

An example of an accounts table with this structure is given in Table 1 in order to give rather more substance to the notation so far presented. The figures refer to all the population irrespective of age. An age disaggregated example is given in Rees and Wilson (1977, Figure 13.32, pp. 208-9).

Where do the figures in Table 1 derive from? They are estimated from available data on regional population stocks, birth totals, death totals, internal surviving migrants, surviving immigrants and surviving emigrants using one of a variety of estimation models (called accounts based models). More details of the procedures used are given in Sections 6 and 7 of the paper and in Rees and Wilson (1977), Rees (1978a) and Jenkins and Rees (1977). For the moment let us assume that an accounts matrix for a base period can be constructed. A second question then presents itself. How can we construct a regional projection model following the framework of the accounts and using the information contained therein?

There are many answers to that question in the form of alternative models but before these are explored a word is necessary about the number and nature of regions for which the models are built. By their very nature accounts matrices are multiregional. Even Stone's (1971) educational accounts contain two "regions" - "this country" and "the outside world". However, there is no requirement that multiregional accounts must be highly disaggregated into many regions. Two will suffice, the region of interest and the rest of the world, and sets of three regions consisting of the region of interest, the rest of the country and the rest of the world are often very convenient from the point of view of available data. Such triregional aggregations applied separately to each region in turn should yield projections fairly close to those of a fully explicit multiregional population projection according to the "shrinking experiments" of Rogers (1976). Rogers was looking at biregional aggregations in a single county as a closed system but his conclusions probably apply equally to the triregional aggregations proposed here. Thus, the criticism levelled at accounting methods by Baxter and Williams (1978, p. 43, quoted at the start of the paper) that such methods require <u>far</u> greater computing and data resources is really not fair.

Table 1 An example accounts table for British regions, 1974-5

Initial states		Final states	SURVIVAL AT M.Y. 1975					DEATH M.Y. 1974-M.Y. 1975					Totals
			Inside Set I				Outside Set 0	Inside Set I				Outside Set 0	
			N	YH	...	S	RW	N	YH	...	S	RW	
EXISTENCE AT M.Y. 1974	Inside Set 0	North	3027310	10477	...	5069	16127	38127	64	...	31	100	3128699
		Yorks. and Humberside	8306	4732987	...	3753	32769	52	58410	...	23	199	4897401
		⋮	⋮	⋮		⋮	⋮	⋮	⋮		⋮	⋮	⋮
		Scotland	4961	4883	...	5053247	47091	31	30	...	62901	289	5216600
	0	Rest of world	7389	24218	...	26670	0	46	147	...	164	0	347099
BIRTH M.Y. 1974-M.Y. 1975	Inside Set I	North	38158	83	...	41	100	239	0	...	0	0	38868
		Yorks. and Humberside	70	60650	...	31	207	0	371	...	0	1	61840
		⋮	⋮	⋮		⋮	⋮	⋮	⋮		⋮	⋮	
		Scotland	41	40	...	67791	312	0	0	...	419	1	68958
	0	Rest of world	46	153	...	176	0	0	0	...	1	0	2181
Totals			3116401	4894896	...	5201215	444382	38681	59392	...	63811	2628	55456542

In the models presented below the regional set I may contain from 1 to N_I regions and the regional set O may contain from 1 to N_O regions. In the Table 1 example N_I equalled ten and N_O one.

To construct population projection models rates of transition are required. These can be defined by dividing the elements in the accounts by the appropriate row totals: the matrix of transition rates, which we call $\underline{H}$, are derived thus

$$\underline{H} = \underline{\tilde{k}}\,\underline{K} \tag{8}$$

where $\underline{\tilde{k}}$ is a matrix with reciprocals of values in the $\underline{k}_E$ and $\underline{k}_B$ row sum vectors arranged along the principal diagonal. The $\underline{H}$ matrix can be partitioned in the same way as the $\underline{K}$ matrix

$$\underline{H} = \left[\begin{array}{c|c} \underline{H}_{ES} & \underline{H}_{ED} \\ \hline \underline{H}_{BS} & \underline{H}_{BD} \end{array}\right] \tag{9}$$

and recognizing the two regional sets I and O

$$\underline{H} = \left[\begin{array}{cc|cc} \underline{H}_{ES}^{II} & \underline{H}_{ES}^{IO} & \underline{H}_{ED}^{II} & \underline{H}_{ED}^{IO} \\ \underline{H}_{ES}^{OI} & \underline{0} & \underline{H}_{ED}^{OI} & \underline{0} \\ \hline \underline{H}_{BS}^{II} & \underline{H}_{BS}^{IO} & \underline{H}_{BD}^{II} & \underline{H}_{BD}^{IO} \\ \underline{H}_{BS}^{OI} & \underline{0} & \underline{H}_{BD}^{OI} & \underline{0} \end{array}\right] \tag{10}$$

The first row of submatrices on the R.H. side of Equation (10) contains rates formed by dividing each element by the corresponding origin state stock, and the third row contains rates formed by dividing interregional infant flows by the births total of the region of birth. These, therefore, are fairly normal transition rates with all the properties of probabilities in a Markov chain. The infant rates will not show quite the behaviour expected of them from conventional rate analysis because the events in the denominator occupy only a triangle in the Lexis diagram rather than a full parallelogram (as will those involved in the first row of submatrices).

The second and fourth row of submatrices contain transition rates formed by dividing immigrant flows by the relevant total immigration. They allocate this external inflow among the regions of the internal set.

Multiregional cohort survival matrix models are normally arranged with rates in a form in which the transpose of the $\underline{H}$ matrix is used:

$$\underline{C} = \underline{H}' = \left[\begin{array}{c|c} \underline{H}'_{ES} & \underline{H}'_{BS} \\ \hline \underline{H}'_{ED} & \underline{H}'_{BD} \end{array}\right] = \left[\begin{array}{c|c} \underline{C}_{SE} & \underline{C}_{SB} \\ \hline \underline{C}_{DE} & \underline{C}_{DB} \end{array}\right] \tag{11}$$

where the transition is from column state to row state, rather than vice versa (as in Table 1).

The matrix population model outlined by Rogers (1968, Chapter 2; 1975, Chapter 2) can now be constructed using these building blocks.

First, rates of birth, transition and survival are defined

$$\underline{B}_S = \tilde{\underline{k}}_E \, \underline{H}_{BS} \tag{12}$$

where $\tilde{\underline{k}}_E$ is a diagonalized matrix with terms $1/K_{r*}^{\varepsilon(i)*(*)}$ in the principal diagonal, and these are transposed and added to the transition rates of the exist-survive quadrant of the accounts

$$\underline{G}_S = \underline{C}_{SE} + \underline{B}'_{BS} \tag{13}$$

This matrix $\underline{G}_S$ of growth rates survives, migrates, the already existing population and adds a contribution of survived infants. The population projection model then is

$$\underline{k}'_S = \underline{G}_{SE} \, \underline{k}_E \tag{14}$$

where $\underline{k}_E$ is a column vector of regional populations disaggregated by age at the start of the base period and $(\underline{k}'_S)$ is a column vector of regional populations disaggregated by age. If the label t is assigned to an initial point in time, t+T to the point in time at the end of the period, then the projection rolls forward as a simple process of matrix multiplication

$$\underline{k}'_S\ (t+T) = \underline{G}_{SE}\ \underline{k}_E\ (t) \tag{15}$$

$$\underline{k}_E\ (t+T) = \underline{k}'_S\ (t+T) \tag{16}$$

$$\underline{k}'_S\ (t+2T) = \underline{G}_{SE}\ \underline{k}_E\ (t+T) \tag{17}$$

or $$\underline{k}'_S\ (t+2T) = \underline{G}^2_{SE}\ \underline{k}_E\ (t) \tag{18}$$

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

$$\underline{k}'_S\ (t+nT) = \underline{G}^n_{SE}\ \underline{k}_E\ (t) \tag{19}$$

This simple model is now treated as an "Aunt Sally". Criticisms are made and improved models suggested in the next and subsequent sections. The improvements are designed to match multiregional projection models to the consistency of the accounts framework and to build in features common in single region projection models.

4. Key issues in model design and improved models

4.1 Some issues and some principles

Some immediate problems must be faced with the model outlined in Equations (12) through (19). The model fails to allow for interaction between the regional set of interest and the environment (other regions). In order words Equation (14) could have been written

$$\underline{k}^{I'}_S = \underline{G}^{II}_{SE}\ \underline{k}^I_E \tag{20}$$

This equation is only correct in the base period if $\underline{K}^{IO}_{ES}$ and $\underline{K}^{OI}_{ES}$, and $\underline{K}^{IO}_{BS}$ and $\underline{K}^{OI}_{BS}$ have equal and opposite effects on $\underline{k}^I_S$.

The model also assumes that the base period rates continue indefinitely. This is useful for the investigation of the stable population situation: that is, the overall growth rate and stable distribution to which the system is tending. However, most practitioners have some idea of how the rates of birth, death and migration are trending and wish to use this knowledge.

There is probably a strong case for modelling the component processes of survival and of fertility (at least) separately. It is easier to modify and model subprocesses. In addition, in the Equation (12) through (19) model there is a subtle but real mismatch between the infant flows and the population at risk of giving birth ($\underline{k}_E$) assumed in the model.

As a response to these observations four general principles may be put forward to guide in the design of multiregional projection models:

(i) The system being studied should be closed.

(ii) The model constructed must allow input of rates variable over time.

(iii) The component processes embodied in the model should be dealt with separately.

(iv) Populations at risk in the model and those used in the rates input to the model should be closely matched.

These principles can be called those of closure, of temporal flexibility, of component separability and of proper rate/population at risk definition. These principles are now applied.

4.2 Improvements to the model

4.2.1 Effective closure

Closure of the model can be effected in a number of different ways, some of which experience has shown are better than others.

The conventional method is to add either a vector of net immigrant flows to the projection or a vector of net immigration rates. In the former case (cf. Stone, 1971) the model becomes

$$\underline{k}_S^{I'} = \underline{G}_{SE}^{II}\ \underline{k}_E^{I} + \underline{n}_S^{IO} \tag{21}$$

where $\underline{n}_S^{IO}$ is a vector of net surviving migrants from all the set O regions to the set I regions, made up of infants and non-infants

$$\underline{n}_S^{IO} = \underline{n}_{SE}^{IO} + \underline{n}_{SB}^{IO} \tag{22}$$

If the model rates have been derived from a set of base period accounts then

$$\underline{n}_{SE}^{IO} = \left(\left[\underline{i}\ (\underline{K}_{ES}^{OI}) - (\underline{K}_{ES}^{IO})'\right]\right)' \tag{23}$$

where $\underline{i}$ is a row vector of ones. However, there will now be an inconsistency since the $\underline{G}_{SE}^{II}\ \underline{k}_{E}^{I}$ term in the equation will yield only survivors internal to the system

$$(\underline{K}_{ES}^{II} + \underline{K}_{BS}^{II})' = \underline{G}_{SE}^{II}\ \underline{\tilde{k}}_{E}^{I} \tag{24}$$

where $\underline{\tilde{k}}_{E}^{I}$ is a diagonalized matrix with the elements of $\underline{k}_{E}^{I}$ in the principal diagonal and zeros elsewhere, and

$$\underline{i}'(\underline{K}_{ES}^{II} + \underline{K}_{BS}^{II})' + \underline{n}_{S}^{IO} \neq k_{S}^{I} \tag{25}$$

Instead of the $\underline{n}_{S}^{IO}$ term should be $\underline{i}'(\underline{K}_{ES}^{OI} + \underline{K}_{BS}^{OI})$ for the proper accounting relations to hold.

However, the situation is normally compensated for by an incorrect estimation of the rates of staying and survival in the $\underline{H}_{ES}^{II}$ and $\underline{G}_{SE}^{II}$ submatrices. Conventionally,

$$\begin{aligned} \text{rate of staying and survival} &= 1 - \text{death rate} - \text{total internal out-migration rate} \\ &= \text{true rate of staying and survival} + \text{external out-migration rate} \end{aligned}$$

Hence, use of a vector of net immigrants or of net immigration rates compensates for this error (see Rees, 1978A,pp. 27-28, for an elaboration of this point).

Such a method of closure may be adequate at the country level but Rogers (1976) suggests that it is wholly inadequate at the regional level. Similarly, Rees (1977) has argued that the reasons for the differences between official projections of the population of East Anglia and his own multiregional projections was that the former assumed a constant net inmigration from the South East whereas the latter assumed a constant migration rate from the South East to East Anglia with a fall in the actual volume due to decreasing population in the South East under the replacement and lower fertility scenarios.

An alternative solution to the closure problem used by Rees (1976) was to include the rest of the world in the inside set of regions (the standard regions of Britain). The population of the rest of the world continued to grow fast as that of Britain exhibited little growth, and a constant rate of migration from the rest of the world to Britain applied to a growing rest of the world resulted in substantial increases in the volume of inflows. In effect, the scenario of no immigration controls or quotas was traced out in the model.

It is more satisfactory to model external migration flows explicitly. The choices available can be envisaged by examining the $\underline{K}_{ES}$ quadrant of the accounts matrix: each model will project the contents of this submatrix (and the $\underline{K}_{BS}$) although on most occasions only the column sums will be output. The population flows in the $\underline{K}_{ES}$ matrix have been divided into <u>three</u> categories in Figure 3:

(i) surviving internal stayers, that is, all $\underline{K}^{\varepsilon(i)\sigma(j)}$ sub-submatrices where i=j and all $i,j\varepsilon I$; and surviving internal migrants, that is, all $K^{\varepsilon(i)\sigma(j)}$ sub-submatrices where $i \neq j$ and all $i,j\varepsilon I$;

(ii) surviving external outmigrants (emigrants), that is, all $\underline{K}^{\varepsilon(i)\sigma(j)}$ sub-submatrices for $i\varepsilon I$, $j\varepsilon O$; and

(iii) surviving external inmigrants (immigrants), that is, all $\underline{K}^{\varepsilon(i)\sigma(j)}$ sub-submatrices for $i\varepsilon O$, $j\varepsilon I$.

<u>Three</u> methods of projection are available:

(a) The terms could be input as flows. They would be derived from other submodels or from knowledge of legislative restrictions and quotas (in the case of immigrants).

(b) The terms could be modelled by multiplication of an initial region population by a transition rate:

$$K^{\varepsilon(i)\sigma(j)} = h^{\varepsilon(i)\sigma(j)} K^{\varepsilon(i)*(*)} \tag{25}$$

where $h^{\varepsilon(i)\sigma(j)}$ is a transition rate between regions i and j defined as

$$h^{\varepsilon(i)\sigma(j)} = K^{\varepsilon(i)\sigma(j)} / K^{\varepsilon(i)*(*)} \tag{27}$$

and $K^{\varepsilon(i)*(*)}$ is the initial population of region i.

(c) The terms could be modelled by multiplication of an initial or final region population by an admission rate:

$$_{K}\varepsilon(i)\sigma(j) = {}_{a}\varepsilon(i)\sigma(j)\ {}_{K}\varepsilon(j)*(*) \tag{28}$$

where

$$_{a}\varepsilon(i)\sigma(j) = {}_{K}\varepsilon(i)\sigma(j) / {}_{K}\varepsilon(j)*(*) \tag{29}$$

or

$$_{K}\varepsilon(i)\sigma(j) = {}_{a}\varepsilon(i)\sigma(j)\ {}_{K}*(*)\sigma(j) \tag{30}$$

where

$$_{a}\varepsilon(i)\sigma(j) = {}_{K}\varepsilon(i)\sigma(j) / {}_{K}*(*)\sigma(j) \tag{31}$$

Equations (28) and (29) suffer from a mismatch of the flow and the population at risk, though this is probably not serious empirically, and Equations (30) and (31) involve iteration in the projection model since $K^{*(*)\sigma(j)}$ is initially unknown.

With three types of flow and three modelling choices, there are 3^3 or 27 permutations available. Many of these can be eliminated. Use of admission rates for internal and emigrant flows is probably not valid in the context of a "push-type" demographic model so the number of alternatives is narrowed to 12 (2x2x3). Of these the mixture of flows input for the internal terms and transition rates for the emigrant terms is unlikely and a further 3 choices can be eliminated together with the permutation of flows for internal and emigrant terms and transition rates for the immigrant terms. Eight choices remain with one choice having two varieties thus yielding the nine options displayed in Figure 3.

The multiregional projection model can now be improved by selecting one or other of these options. Selection of option three yields the following model

$$\underline{k}_{S}^{I'} = \underline{G}_{SE}^{II}\ \underline{k}_{E}^{I} + (\underline{k}_{ES}^{OI'} + \underline{k}_{BS}^{OI'}) \tag{32}$$

in which column vectors of surviving immigrants, $\underline{k}_{ES}^{OI'}$, and surviving infant immigrants, $\underline{k}_{BS}^{OI'}$, are added to the product of the internal multiplication of the initial populations of the internal regions by the growth matrix. The emigrant flows do not appear explicitly in Equation (32) but could be generated as

Figure 3 The options for closing the system

The K_{ES} submatrix: types of flow

	I	O
I	Internal surviving stayers and migrants	surviving emigrants
O	surviving immigrants	0

Alternative inputs

F = population flows

TR = transition rates x par

AR = admission rates x par

par = population at risk

Options (selected)

$\boxed{1}$

	I	O
I	F	F
O	F	0

$\boxed{2}$

	I	O
I	TR	TR
O	TR^1	0

$\boxed{3}$

	I	O
I	TR	TR
O	F	0

$\boxed{4}$

	I	O
I	TR	F
O	TR	0

$\boxed{5}$

	I	O
I	TR	F
O	F	0

$\boxed{6}$

	I	O
I	TR	TR
O	TR^2	

$\boxed{7}$

	I	O
I	TR	TR
O	AR	0

$\boxed{8}$

	I	O
I	TR	F
O	AR	0

$\boxed{9}$

	I	O
I	F	F
O	AR	0

Notes

1. TR^1 – this option assumes that the transition rates "distribute and survive" an immigration total.

   TR^2 – this option assumes that regions formerly in set O have been included in set I and normal transition rates are used.

2. Options 1 through 6 are those available in the Jenkins and Rees (1977) computer program (Table 2, p. 30).

$$\underline{k}_S^{O'} = \underline{G}_{SE}^{IO} \underline{k}_E^{I} \qquad (33)$$

The effect of model choice has yet to be fully explored. The disadvantages of Option 6 (used in Rees, 1976) have already been mentioned but comparisons of Options 3 and 7 and of Options 5 and 8 need to be carried out. Admission rate treatment of the immigrant flows is included in the models of Liaw (1977, 1978).

4.2.2 Ensuring temporal flexibility

Temporal flexibility is easily introduced (Stone, 1971, Chapter 3) and algebraically it is simply a matter of adding the relevant period labels to the matrix of growth rates. When this is done to Equation (32) the following is obtained

$$\underline{k}_S^{I'}(t+T) = \underline{G}_{SE}^{II} (\theta_t^T) \underline{k}_E^{I}(t) + \underline{m}_S^{I} (\theta_t^T) \qquad (34)$$

where θ is a period label and θ_t^T refers to a period beginning with time t of T years duration and $\underline{m}_S^I(\theta_t^T)$ is a column vector of surviving immigrants in period θ_t^T. For the next period the process described by Equation (34) continues

$$\underline{k}_S^{I'}(t+2T) = \underline{G}_{SE}^{II}(\theta_{t+T}^T) \underline{k}_E^{I}(t+T) + \underline{m}_S^{I}(\theta_{t+T}^T) \qquad (35)$$

or

$$\begin{aligned} \underline{k}_S^{I'}(t+2T) = {} & \underline{G}_{SE}^{II}(\theta_{t+T}^T) \underline{G}_{SE}^{II}(\theta_t^T) \underline{k}_E^{I}(t) \\ & + \underline{G}_{SE}^{II}(\theta_{t+T}^T) \underline{m}_S^{I}(\theta_t^T) \\ & + \underline{m}_S^{I}(\theta_{t+T}^T) \qquad (36) \end{aligned}$$

which can be generalized to

$$\begin{aligned} \underline{k}_S^{I'}(t+nT) = {} & \prod_{a=0}^{n-1} \underline{G}_{SE}^{II}(\theta_{t+aT}^T) \underline{k}_E^{I}(t) \\ & + \prod_{a=0}^{n-2} \prod_{b=a}^{n-2} \underline{G}_{SE}^{II}(\theta_{t+(b+1)T}^T) \underline{m}_S^{I}(\theta_{t+aT}^T) \\ & + \underline{m}_S^{I}(\theta_{t+(n-1)T}^T) \qquad (37) \end{aligned}$$

The growth matrix is allowed to change with time so that with this model simulations such as those sketched in Figure 2.G are possible.

4.2.3 Separating the components

It is difficult to observe or to change the growth rates of the $\underline{G}$ matrix. It is easier to change the component rates of birth (fertility), of survival (and its complement, mortality), and of migration. To effect this separation it is necessary to convert the single equation model into a multi-equation and multi-step model.

At a minimum some five steps are necessary. These are:

(i) an equation or set of equations for the population flows in the exist-survive quadrant of the accounts matrix;

(ii) an equation or set of equations for ~~applying fertility rates~~ computing ~~to~~ the population at risk;

(iii) an equation or set of equations for applying fertility rates to the population at risk;

(iv) an equation or set of equations for surviving the infants born in the period (cf. step (i)); and finally

(v) an equation or set of equations that consolidate the results of steps (i) and (iv) in preparation for projecting in the next period.

For each one of these steps, bar the last, a variety of alternatives exist, of which a few are presented here, developing further the closure option three model of Equation (34).

Firstly, the vector of populations in the regions at the start of the projection period is survived over the time interval and the immigrants from outside the regional set of interest are added in:

$$\underset{N_I R \times 1}{\underline{k}_{SE}^{I'x}(t+T)} = \underset{(N_I R)\times(N_I R)}{\underline{C}_{SE}^{IIx}(\theta_t^T)}\ \underset{(N_I R)\times 1}{\underline{k}_{E}^{Ix}(t)} + \underset{(N_I R)\times 1}{\underline{m}_{SE}^{Ix}(\theta_t^T)}$$

$$\text{for } x=m,f \qquad (38)$$

The letters below the equation variables refer to the dimensions of the matrices and vectors involved: N_I is the number of regions in the internal set and R is the number of age groups. A superscript x referring to sex with categories m for male and f for female has been added to each term to recognize not only the differences between males and females in survival rates, which are well known, but also the differences between the sexes in migration propensity which have recently been noted by Rogers, Racquillet and Castro (1977). The $\underline{C}$ matrix contains just exist-survive transition rates for stayers and migrants (unlike the $\underline{G}$ matrix which includes birth/transition rates).

Then, the populations at risk of giving birth are worked out either by computing the average populations over the period

$$\underset{N_I R\times 1}{\hat{\underline{k}}^{If}\left(\begin{smallmatrix}T\\t\end{smallmatrix}\right)} = \left(\underset{N_I R\times 1}{\underline{k}_E^{If}(t)} + \underset{N_I R\times 1}{\underline{k}_G^{If'}(t+T)}\right)\underset{1\times 1}{\tfrac{1}{2}} \qquad (39).$$

The $\underline{k}$ vectors are so arranged as to yield age group populations at risk (the illustration is for the individual algebraic variables)

$$\hat{K}_{*r}^{B*if}(\theta_t^T) = (K_{r*}^{\varepsilon(i)W(W)f}(t) + K_{*r}^{*(*)\sigma(i)f}(t+T))\tfrac{1}{2} \qquad (40).$$

Or cohort populations at risk are worked out

$$\hat{K}_{r*}^{B*if}(\theta_t^T) = (K_{r*}^{\varepsilon(i)*(*)f}(t) + K_{*r+1}^{*(*)\sigma(i)f}(t+T)\tfrac{1}{2} \qquad (41)$$

depending on which classification by mother's age is given for the fertility rates used in the next step. More elaborate multiregional populations at risk can be defined and are used in the accounts-based projection model (Rees and Wilson, 1977, Chapters 4 and 10) but if these are used with conventionally derived fertility rates a mismatch of the population at risk used in the fertility step of the projection and that used in the definition of the fertility rates occurs.

The populations at risk are multiplied by an appropriate set of fertility rates and this multiplication yields the total number of births classified by age group of mother. Age-specific fertility rates are generally not classified by sex of the child so that the births must be sexed by multiplication by a vector containing the sex proportions for each region:

$$\underset{N_I x1}{\underline{k}_B^{Ix'}(\theta_t^T)} = \underset{N_I x N_I R}{\underline{\Sigma}^{Ix}} \quad \underset{N_I R x N_I R}{\underline{\tilde{F}}^{I}(\theta_t^T)} \quad \underset{N_I R x1}{\underline{\hat{k}}^{If}(\theta_t^T)} \tag{42}$$

where $\underline{\tilde{F}}$ is a matrix in which the regional fertility rates have been arranged along the principal diagonal with zeros elsewhere and $\underline{\Sigma}^{Ix}$ is a matrix of suitably arranged sex proportions

$$\underline{\Sigma}^{Ix} = \begin{bmatrix} \sigma_1^{1x} & \cdots & \sigma_R^{1x} & 0 & \cdots & 0 & \cdots \\ 0 & \cdots & 0 & \sigma_1^{2x} & \cdots & \sigma_R^{2x} & \cdots \\ \vdots & & \vdots & \vdots & & \vdots & \cdots \end{bmatrix} \tag{43}$$

Although Equations (39), (42) and (43) have been specified with an age dimension of R - say 18 five year age groups - in practice only L of these - say age groups 15-19 to 45-49 - will have non-zero fertility rates. In addition the sex proportions will normally be assumed not to vary by age of mother to any significant degree.

Then the infants born in the period are survived and relocated in a similar fashion to non-infants (cf. Equation (38)).

$$\underset{N_I x1}{\underline{k}_{SB}^{Ix'}(t+T)} = \underset{N_I x N_I}{\underline{C}_{SB}^{IIx'}(\theta_t^T)} \quad \underset{N_I x1}{\underline{k}_B^{Ix'}(\theta_t^T)} + \underset{N_I x1}{\underline{m}_{SB}^{Ix}(\theta_t^T)} \tag{44}$$

This equation survives and relocates the column vector of births by pre-multiplying it by a transposed matrix of transition rates for infants, and to the resulting vector is added a column vector of infant immigrants from outside the regional set of interest.

If the $\underline{k}_{SB}^{Ix'}(t+T)$ vector is expanded by adding redundant zeros for age groups above the first, the result can be added to the earlier $\underline{k}_{SE}^{Ix'}(t+T)$ vector to consolidate the projected populations

$$\underline{k}_S^{Ix'}(t+T) = \underline{k}_{SE}^{Ix'}(t+T) + \underline{k}_{SB}^{Ix'}(t+T)(\text{mod.}) \tag{45}$$

and then the full vector of projected populations becomes the initial population vector for the new period

$$\underline{k}_E^{Ix}(t+T) = \underline{k}_S^{Ix'}(t+T) \tag{46}$$

thus completing the fifth step of the improved multiregional cohort survival model.

4.2.4 Proper rate and population at risk definition

Where the survival and birth components have been separated it is much easier to relate death, migration and fertility rate series to the models, and some earlier mismatches of the $\underline{G}$ matrix model are avoided.

Implicit in the Equation (20) model, for example, is a multiplication of a birth rate of the form b_{*r}^{*i} by a population at risk $K_{r*}^{\varepsilon(i)*(*)}$, where the i label in the birth rate normally refers to location of mother at time of maternity (and location of child at time of birth) and the r label refers to age group of mother at the time of maternity. Mothers contributing to such births will originate mainly in region i but in-migrant mothers will also contribute and be included in the population at risk. Equation (39) attempts to make allowances for this. The more detailed mathematics is discussed in Section 12.4 of Rees and Wilson (1977).

One further issue which warrants mention at this stage is the derivation of the surviving migrant and surviving stayer rates. This is still a matter of some conjecture vis à vis multiregional population projection models derived through the multiregional life table approach (Rogers, 1975; Ledent, 1978a, 1978b) but they are straightforwardly derived from accounts matrices via Equation (8).

In most multiregional projection models stayers are treated differently from migrants. Migrants in the accounts matrix are entered directly as flows, whereas stayers are calculated as residuals

$$\text{surviving stayers} = \text{population} - \text{outmigrants} - \text{deaths} \tag{47}$$

or

$$\begin{matrix}\text{rate of survival} \\ \text{and staying}\end{matrix} = 1 - \begin{matrix}\text{outmigration} \\ \text{rate}\end{matrix} - \begin{matrix}\text{death} \\ \text{rate}\end{matrix} \tag{48}$$

However, since the migration tables of the Census which yield the migrant and migration rate estimates simply classify persons by place of current and former residence it is relatively easy to extract numbers of persons living in the same place at Census date and one or five years earlier. So, stayers could alternatively be entered as flows as well. Which method provides the better estimate - the residual or the direct? This question will be examined when the issue of base period calibration is taken up in a later section. Before that it is probably useful to illustrate how the multiregional projection model proposed works and performs.

5. A case study in multiregional projection

The multiregional population projection model described in section 4 has recently been applied to a four region system consisting of East Anglia, the South East, the Rest of Britain and the Rest of the World in an investigation sponsored by the Department of the Environment on behalf of the East Anglian Economic Planning Council (Rees, 1977). The strategy followed in that study is briefly sketched out here and then some of the study results are considered.

The system of interest was defined to consist of the resident populations of the four regions listed above, disaggregated by sex and by quinquennial age groups. Population flows and events were considered over five year time periods in order that age group intervals were equal to the time period length. The region of interest was East Anglia and in one set of the projections carried out this region was further broken down into the three counties of Cambridge, Norfolk and Suffolk. In order to understand properly what was happening to the East Anglian population it was considered essential to include in the system of interest as a separate and identifiable region the South East, the most important demographic trading partner of East Anglia. The Rest of Britain was included in order to use the data contained in the 1971 Census migration tables most effectively, and the Rest of the World was included to close the system. In the 1966-71 quinquennium prior to the 1971 Census 41 per cent of East Anglia's gross migration was to and from the South East, compared with 37 per cent for the Rest of Britain and 22 per cent for the Rest of the World.

The base period, for which age-sex disaggregated population accounts were prepared (Appendix 3 in Rees, 1977 and Appendix 1 in Rees, 1978a), had to be the 1966-71 intercensal period. Population estimates, births, deaths and international migration data were available at the time that the projections were carried out (early 1977) for the years 1972 to 1976 but no published internal interregional migrant data were to hand for these years. Internal interregional migration data are collected only in the periodic censuses, the last of which was in 1971 (although there would seem to be no good reason for not gathering such information on a sample, yearly basis in the General Household Survey). Using the terminology of Figure 2E, the gap between the end of the base period and 'then' was approximately five years; the gap between 'then' and the 'now' of the date of the report (Rees, 1977) was about one year, and the gap between 'then' and the 'now' of this paper is about two and a quarter years. These gaps are highly unsatisfactory but can only be remedied by a much bolder use by the Office of Population Censuses and Surveys of the General Household Survey and the National Health Service Register and the publication of regional, county and local migration estimates.

In order to close this gap a set of updating techniques were employed. Aggregate population accounts were constructed for the four region system for the 1971-76 period using the births, deaths and international migration data for the period together with internal migration rates from aggregate accounts for 1966-71. These accounts were constrained to the mid-year 1976 population estimates, and a new estimate of the accounts matrix was obtained. A comparison was then made between the transition pattern characteristic of the 1966-71 accounts and those for the five years following. The transition rates of migrants were not themselves examined because the comparison would have been influenced by any differences in regional survival rates (and age compositions) between the two periods. Instead locational probabilities conditional on survival were defined thus

$$P(j \mid \epsilon, i, \sigma) \quad = \quad K^{\epsilon(i)\sigma(j)} \,/\, K^{\epsilon(i)\sigma(*)} \qquad (49)$$

where $P(j \mid \epsilon, i, \sigma)$ is the probability that a person will be located at the end of the five year period in region j given that he or she existed in region i at the start of the time interval, and survived the period.

The denominator on the R.H. side of the equation is the total number of survivors of the initial population of region i. Updating ratios were defined thus:

$$r^{\epsilon(i)\sigma(j)} = P(j|\epsilon, i, \sigma)(1971\text{-}76) \;/\; P(j|\epsilon, i, \sigma)(1966\text{-}71) \qquad (50)$$

and these were applied to the age-sex disaggregated locational probabilities from the 1966-71 age classified accounts

$$P(j, r+1|\epsilon, i, r, \sigma, x)(1971\text{-}76) = r^{\epsilon(i)\sigma(j)}\, P(j, r+1|\epsilon, i, r, \sigma, x) \qquad (51).$$

These locational probabilities could then be multiplied by survival probabilities appropriate to the projection period concerned to yield the transition rates required in the projection model

$$h_{rr+1}^{\epsilon(i)\sigma(j)x}(\theta_t^T) = P(j, r+1|\epsilon, i, r, \sigma, x)(\theta_t^T) \times h_{r*}^{\epsilon(i)\sigma(*)x}(\theta_t^T) \qquad (52)$$

where the h variables are the elements of the $\underline{H}$ matrix, defined in Equations (8), (9) and (10), quadrants of which are used in transposed form in the multiregional cohort survival model defined in section 4.2.3.

Mortality rates were investigated for trend using a negative exponential model. Decline parameters were estimated and these were applied to the 1966-71 death rates to yield death rates for each five year period to 1996-2001. The complements of the death rates, survival rates, were computed and used in Equation (32). The same survival rates were used irrespective of the other rate scenarios in the projections and a two to three year improvement in life expectancies were expected in the regions by 2001.

Both the 1966-71 and the 1971-76 patterns of migration were used as the basis of projections (the 'M' and 'N' series respectively) but there was little difference in the projected populations. It would have been possible to study any other altered migration pattern using the multiregional cohort survival model but the theoretical basis and rationale for doing so remain to be defined. Trend extrapolation of migration patterns might be possible in the fashion of mortality were not the equivalent, well defined time series not available; alteration of the attractiveness of regions to in-migrants using spatial interaction models (Stillwell, 1978) is another avenue to be explored. Despite the extensive literature on the modelling of migration virtually no attention has been paid to the use of such models in a predictive context.

This is not to deny that differences in migration inputs to regional projection models can effect substantial changes in the outcomes. The differences which do occur are the result primarily of choice of one or other of the migration options presented in Section 4.2.1 and Figure 3. Treatment of the same migration data using a net flow, gross flow or gross rates approach can result in substantially different projections.

Populations at two points in time were used as base populations in the projections: the Census resident populations of the regions at Census date 1971, and the home populations of the regions as estimated at mid-year 1976. Only small differences in the projected figures resulted.

The main dimension of variation introduced into the projections is that of fertility. The consequences of four alternate levels of national fertility are explored:

(i) The highest scenario, A, assumes a total fertility rate (TFR) level of 2.2 children per woman, adopted through most of the 1970s official projections as the level to which fertility would return in the 1980s.

(ii) The next scenario, B, assumes a total fertility rate level of 2.1 children per woman. This level was adopted in the 1976 and 1977 based official projections as the one to which the population would return in the late 1980s. In a low mortality population such as that in Britain a TFR of 2.1 currently corresponds to a net reproduction rate of 1, though with continued slight falls in mortality replacement level will probably drop below 2.1, particularly in the healthy East Anglia and South East regions.

(iii) The third scenario, C, assumes a national TFR of 1.8, the level of 1975 in England and Wales, below which it has subsequently dropped.

(iv) The fourth scenario, D, assumes a national TFR of 1.5, a level yet to be attained in Britain (the 1976 rate was 1.72, the 1977 rate 1.67 and the 1978 rate is estimated to be 1.69 (O.P.C.S., 1978)). This level has been attained in the Federal Republic of Germany.

Regional differentials in fertility are assumed to persist and transition to the long term levels (A, B, C, D) is assumed to occur by 1986-91. The scenarios are set out in Table 2. The South East has consistently experienced lower than average (G.B.) fertility in the last decade or so (Rees, 1978; Table 13), whereas East Anglia has moved from a position of substantially below average fertility in 1965, slightly below average fertility in 1970 and above average fertility in 1975. The rest of Britain experienced above average fertility in the first half of the 1970s.

The scenarios were built up as follows. It is assumed that the scenario fertility levels for each region are achieved in the quinquennium 1986-91. The 1971-76 levels are those observed for the period, those for 1976-81 are assumed to approximate the 1975-76 fertility levels with lower levels in 1977 and 1978 being compensated by higher fertility levels in 1979, 1980 and 1981. Fertility in 1981-86 is assumed to lie between the 1976-81 levels and the 1986-91 targets.

A selection of the results of the projections are gathered together in Table 3 for the end points of five year projection periods to 2001. Populations for males, females and persons in the three regions in our internal set are given for each of the four fertility scenarios using the 1971-76 pattern of migration rates and the downward trending mortality rates (the N76A, N76B, N76C and N76D projections respectively). Table 4 summarizes the growth projected for the three regions in the 1976-2001 period.

Under all scenarios the population of East Anglia grows substantially because the gain through a surplus of in-migrants over out-migrants is much bigger than the gains or losses through natural increase. In fact,

Table 2 Total fertility rate scenarios for East Anglia, the South East and the Rest of Britain, 1971-2001

Region	Scenario	Period					
		1971-76	1976-81	1981-86	1986-91	1991-96	96-2000
East Anglia	A	2.092	1.786	2.019	2.251	2.251	2.251
	B	2.092	1.786	1.967	2.148	2.148	2.148
	C	2.092	1.786	1.814	1.841	1.841	1.841
	D	2.092	1.786	1.661	1.535	1.535	1.535
South East	A	1.922	1.695	1.882	2.068	2.068	2.068
	B	1.922	1.695	1.835	1.974	1.974	1.974
	C	1.922	1.695	1.694	1.692	1.692	1.692
	D	1.922	1.695	1.553	1.410	1.410	1.410
Rest of Britain	A	2.103	1.815	2.040	2.264	2.264	2.264
	B	2.103	1.815	1.988	2.161	2.161	2.161
	C	2.103	1.815	1.834	1.852	1.852	1.852
	D	2.103	1.815	1.680	1.544	1.544	1.544

Source: Table 26 in Rees (1977).

Table 3 The projected populations of East Anglia, The South East and the Rest of Britain, 1976-2001, N76 series from Rees (1977)

EAST ANGLIA		Populations in 1000s			
SEX	Year (mid-point)	Fertility scenario (TFR for G.B.)			
		A (2.2)	B (2.1)	C (1.8)	D (1.5)
MALES	1976	893.6	893.6	893.6	893.6
	1981	930.7	930.7	930.7	930.7
	1986	973.0	971.2	965.9	960.6
	1991	1022.7	1016.3	1000.1	983.8
	1996	1068.0	1056.6	1030.1	1003.1
	2001	1106.5	1090.5	1054.2	1017.0
FEMALES	1976	909.3	909.3	909.3	909.3
	1981	968.5	968.5	968.5	968.5
	1986	1018.6	1016.9	1011.8	1006.7
	1991	1079.5	1073.3	1057.9	1042.3
	1996	1134.9	1123.9	1098.8	1073.0
	2001	1182.8	1167.3	1131.1	1097.7
PERSONS	1976	1802.9	1802.9	1802.9	1802.9
	1981	1893.2	1893.2	1893.2	1893.2
	1986	1991.6	1988.0	1977.7	1967.3
	1991	2102.2	2089.6	2058.0	2026.1
	1996	2202.8	2180.5	2128.9	2076.0
	2001	2289.3	2257.7	2187.4	2114.7
SOUTH EAST		Populations in 1000s			
SEX	Year (mid-point)	Fertility scenario (TFR for G.B.)			
		A (2.2)	B (2.1)	C (1.8)	D (1.5)
MALES	1976	8182.8	8182.8	8182.8	8182.8
	1981	8111.6	8111.6	8111.6	8111.6
	1986	8113.6	8099.7	8058.3	8016.8
	1991	8185.7	8103.3	8020.3	7896.9
	1996	8245.0	8101.6	7978.3	7778.7
	2001	8279.5	8083.1	7921.0	7651.9
FEMALES	1976	8710.5	8710.5	8710.5	8710.5
	1981	8614.7	8614.7	8614.7	8614.7
	1986	8594.9	8581.9	8542.7	8503.6
	1991	8643.2	8565.5	8487.8	8371.8
	1996	8680.5	8546.2	8431.2	8244.6
	2001	8699.4	8516.2	8365.2	8114.4
PERSONS	1976	16893.3	16893.3	16893.3	16893.3
	1981	16726.3	16726.3	16726.3	16726.3
	1986	16708.5	16681.6	16601.0	16520.5
	1991	16829.0	16668.9	16508.1	16268.6
	1996	16925.5	16647.8	16409.5	16023.3
	2001	16978.9	16559.2	16286.2	15766.3

Table 3/continued

SEX	REST OF BRITAIN Year (mid-point)	Populations in 1000s Fertility scenario (TFR for G.B.) A (2.2)	B (2.1)	C (1.8)	D (1.5)
MALES	1976	17394.1	17394.1	17394.1	17394.1
	1981	17432.6	17432.6	17432.6	17432.6
	1986	17635.6	17602.1	17502.9	17403.7
	1991	18013.6	17911.3	17610.8	17309.4
	1996	18350.3	18180.4	17690.1	17196.5
	2001	18602.2	18367.7	17701.8	17029.4
FEMALES	1976	18327.1	18327.1	18327.1	18327.1
	1981	18370.1	18370.1	18370.1	18370.1
	1986	18572.7	18541.2	18448.1	18354.9
	1991	18938.4	18842.8	18560.0	18276.9
	1996	19269.4	19110.4	18648.3	18183.8
	2001	19529.3	19309.3	18680.0	18045.7
PERSONS	1976	35271.2	35721.2	35721.2	35721.2
	1981	35802.7	35802.7	35802.7	35802.7
	1986	36208.3	36143.4	35951.0	35758.7
	1991	36952.0	36754.1	36170.8	35586.2
	1996	37619.7	37290.9	36338.3	35380.3
	2001	38131.5	37677.1	36381.8	35075.2

SEX	GREAT BRITAIN Year (mid-point)	Populations in 1000s Fertility scenario (TFR for G.B.) A (2.2)	B (2.1)	C (1.8)	D (1.5)
MALES	1976	26470.5	26470.5	26470.5	26470.5
	1981	26474.9	26474.9	26474.9	26474.9
	1986	26722.2	26673.0	26527.1	26381.1
	1991	27222.0	27030.9	26631.2	26190.1
	1996	27663.3	27338.6	26698.5	25978.3
	2001	27988.2	27541.3	26677.0	25698.3
FEMALES	1976	27946.9	27946.9	27946.9	27946.9
	1981	27953.3	27953.3	27953.3	27953.3
	1986	28186.2	28140.0	28002.6	27865.2
	1991	28661.1	28481.6	28105.7	27691.0
	1996	29084.8	28780.5	28178.3	27501.4
	2001	29411.5	28992.8	28178.3	27257.8
PERSONS	1976	54417.4	54417.4	54417.4	54417.4
	1981	54428.2	54428.2	54428.2	54428.2
	1986	54908.4	54813.0	54529.7	54246.3
	1991	55883.1	55512.5	54736.9	53881.1
	1996	56748.1	56119.1	54876.8	53479.7
	2001	57399.7	56534.1	54855.3	52956.1

Notes:

1. Source: For East Anglia, Tables A2.9,A2.10,A2.11 in Rees (1977). For the South East, and the Rest of Britain the computer output files :GEOPHRG.EAFORABB76,:GEOPHRG.EAFORBBB76, :GEOPHRG.EAFORCBB76, and :GEOPHRG.EAFORDBB76 on the University of Leeds ICL1906A computer.

2. T.F.R. = total fertility rate = sum of the five year cohort specific fertility rates.

Table 4 Relative growth to 2001

Region	Ratio of 2001 population to 1976 population			
	Fertility scenario			
	A (2.2)	B (2.1)	C (1.8)	D (1.5)
East Anglia	127	125	121	117
South East	101	98	96	93
Rest of Britain	107	105	102	98
Great Britain	105	104	101	97

Table 5 Official 1975-based regional projections

Sex	Year	Projected populations in 1000 s			
		East Anglia (Home popn.) (1)	South East (Home popn.) (2)	Rest of Britain (Residual) (3)	Great Britain (Total popn.) (4)
MALES	1975	879.5	8200.6	17477.9	26558
	1981	942.7	8120.2	17478.1	26541
	1986	1004.2	8159.7	17714.1	26878
	1991	1067.6	8255.5	18030.9	27354
	1996				27775
	2001				28017
FEMALES	1975	900.9	8735.5	18319.6	27956
	1981	965.2	8599.9	18286.9	27852
	1986	1024.1	8589.0	18461.9	28075
	1991	1082.3	8637.9	18696.8	28417
	1996				28698
	2001				28818
PERSONS	1975	1780.4	16936.1	35797.5	54514
	1981	1908.0	16720.1	35765.9	54394
	1986	2028.3	16748.7	36176.0	54953
	1991	2149.9	16893.4	36727.7	55771
	1996				56473
	2001				56835

Notes:

1. Estimated by subtraction of the Column (1) and (2) populations from those of Column A.
2. Sources: O.P.C.S. (1977) and O.P.C.S. (1978).

East Anglia only experiences net natural decrease under the D fertility scenario (-12.3 thousand over the 1976-2001 period). Conversely, the South East experiences losses of population at fertility levels of replacement (B scenario) and below because of sustained net out-migration, and begins to experience substantial decrease under the D scenario. The Rest of Britain behaves rather like the country as a whole and shows decreases only under the lowest fertility regime. Here the effect of the youthful age structure produced by the 1955-70 baby "boomlet" is to postpone the period when below replacement level fertility levels lead to a reduction in total population.

How do these projections compare with those of the official bodies? Table 5 reproduces figures from the latest (1975-based) English region projections (O.P.C.S., 1978), and the equivalent Great Britain projection (O.P.C.S. 1977). The long term fertility level in these projections is a TFR of 2.1 achieved in 1985. There is a slight mismatch of population concepts between the regional scale at which the home population concept is used and the national where the total population concept is employed so that the residual figures for the Rest of Britain in column (3) of Table 5 are probably too high at each date by circa 30,000. The results for both the multiregional projections and the single region official projections are summarized in graphical form in Figure 4.

At the Great Britain scale the results are fairly close with the B scenario results being the nearest. The O.P.C.S. projections are higher than the B scenario projections because of the different population concept used (total as opposed to home) and probably because the net emigration levels were assumed higher in the Rees (1977) projections than in the official projections. Greater differences are apparent for the East Anglia and South East projections with the projections of the multiregional model being substantially lower in each case.

It is difficult to be precise about the reasons for the differences, but some clues can be revealed by a careful examination of the components of growth (natural increase and net migration) associated with a variety of projections for East Anglia (Table 6) that were gathered together in an addendum to the main report on "The Future Population of East Anglia ..." (Rees,1977). Of the variety of projections reported in the table attention needs to be focused on the first set, the O.P.C.S. 74 based

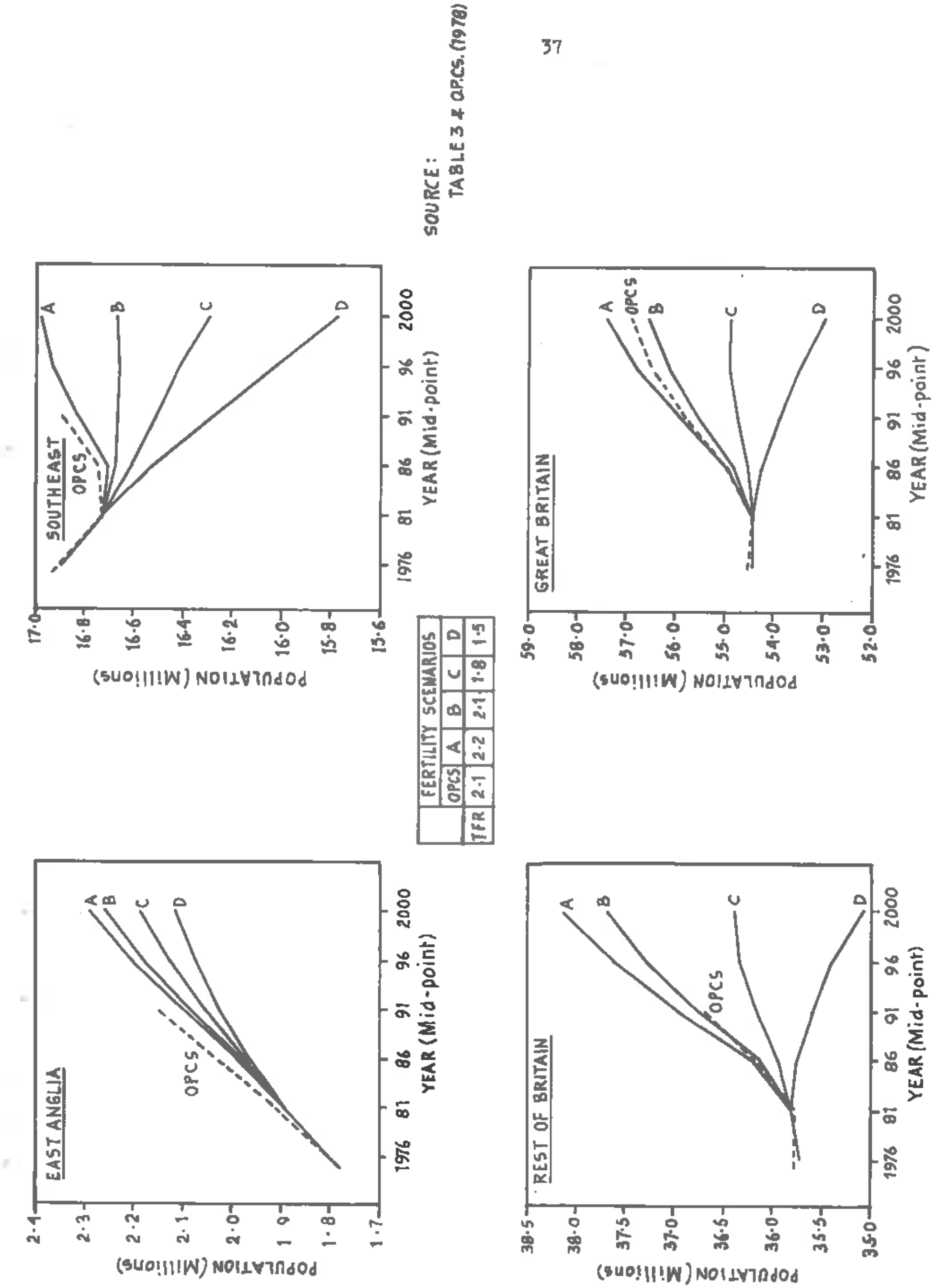

Figure 4. The projected population of East Anglia, the South East, the Rest of Britain and Great Britain, 1976 - 2001.

Projection	Initial Population P(t)	Natural Increase NI	Net Migration NM	Population Change ΔP	Final Population P(t+T)
1976-1991 period					
1. OPCS '74 based	1802.9	108.1	281.0	389.1	2192.0
2. ABM71-76 flows	1800.9	80.5	287.7	368.2	2169.0
3. ABM71-76 rates	1800.9	79.4	224.4	303.8	2104.7
4. ABM66-71 rates	1787.0	140.7	144.5	285.2	2072.4
5. MRCS M71B	1786.0	67.5	225.9	293.5	2079.7
6. MRCS N71B	1781.9	65.4	221.5	287.0	2068.9
7. MRCS N76B	1802.9	74.1	212.6	286.7	2089.6
8. MRCS N76A	1802.9	86.5	212.9	299.4	2102.2
9. MRCS N76B	1802.9	74.1	212.6	286.7	2089.6
10. MRCS N76C	1802.9	42.0	212.9	254.9	2058.0
11. MRCS N76D	1802.9	10.4	212.8	223.2	2026.1
1976-2001 period					
1. Extrapolated OPCS '74 based	1802.9	199.1	461.0	660.1	2463.0
2. ABM71-76 flows	1800.9	143.6	479.5	623.1	2423.8
3. ABM71-76 rates	1800.9	139.3	330.4	469.7	2270.6
4. ABM66-71 rates	1787.0	246.0	208.4	434.4	2241.5
5. MRCS M71B	1786.2	131.2	332.8	463.8	2250.0
6. MRCS N71B	1781.9	128.5	330.3	458.9	2240.8
7. MRCS N76B	1802.9	137.2	317.6	454.8	2257.7
8. MRCS N76A	1802.9	167.9	318.6	486.5	2289.3
9. MRCS N76B	1802.9	137.2	317.6	454.8	2257.7
10. MRCS N76C	1802.9	62.1	322.2	384.3	2187.4
11. MRCS N76D	1802.9	− 12.3	324.2	311.7	2114.7

Table 6. Summary components of growth, 1976-1991 and 1976-2001, for selected East Anglia Population projections.

OPCS - Office of Population Censuses and Surveys
ABM - Aggregate Accounts Based Model.
MRCS - Multi-Regional Cohort Survival Model.
M71B etc. - See Main Report. (Rees, 1977).

the latest official projection available at the time the time was prepared, the second and third sets, alternative versions of the projection version of the accounts based model (Rees and Wilson, 1977) and the eighth set, the multiregional cohort survival model.A scenario projections. The results of these projections are graphed in Figure 5.

For the 1976-2001 period, the first and second projections yield fairly comparable pictures in terms of component sizes despite major differences. The first is a single region cohort survival model with net migration added; the second is a multiregional aggregate (all age and sex) accounts based model, in which migration is modelled by option 1 of Figure 3 by the assumption that the migration flows matrix remains constant over time. The third and ninth projections also yield similar results despite the difference in aggregation degree (aggregate versions age-sex disaggregated) and scenario assumption (constant versus variable). The third projection uses an accounts based model in which option 3 of Figure 3 is employed, that is, transition rates for modelling the $\underline{K}^{II}$ and $\underline{K}^{IO}$ submatrices and flows for the $\underline{K}^{OI}$ submatrix.

A comparison of the accounts based model projections, numbers 2 and 3, can be used as a surrogate for the more difficult comparison of projections 1 and 8 (or 9) in which there are many more sources of variation. Projections 2 and 3 differ in only one respect, that of migration model. If a constant flows assumption is used (equivalent to the approximately constant net migration of projection 1) a net migration gain of 480 thousand is made by East Anglia by 2001. If a constant rates assumption is made, a gain of only 330 thousand is made. Model choice is thus revealed to be of critical importance in regional population projection.

This analysis cannot, of course, provide a definitive "best buy" in terms of projection since the comparisons have been made in a Figure 2.F situation rather than a Figure 2.H situation in which alternative methods of projection could be tested. The arguments for a multiregional population projection in which migration is interregional modelled in a fashion other than that of constant gross flows or net migration must remain theoretical for the time being. The simulation and sensitivity experiments could nevertheless be made more precise by running the different models with the same base data, and the same range of scenarios on the different models. Improvement and integration of the necessary

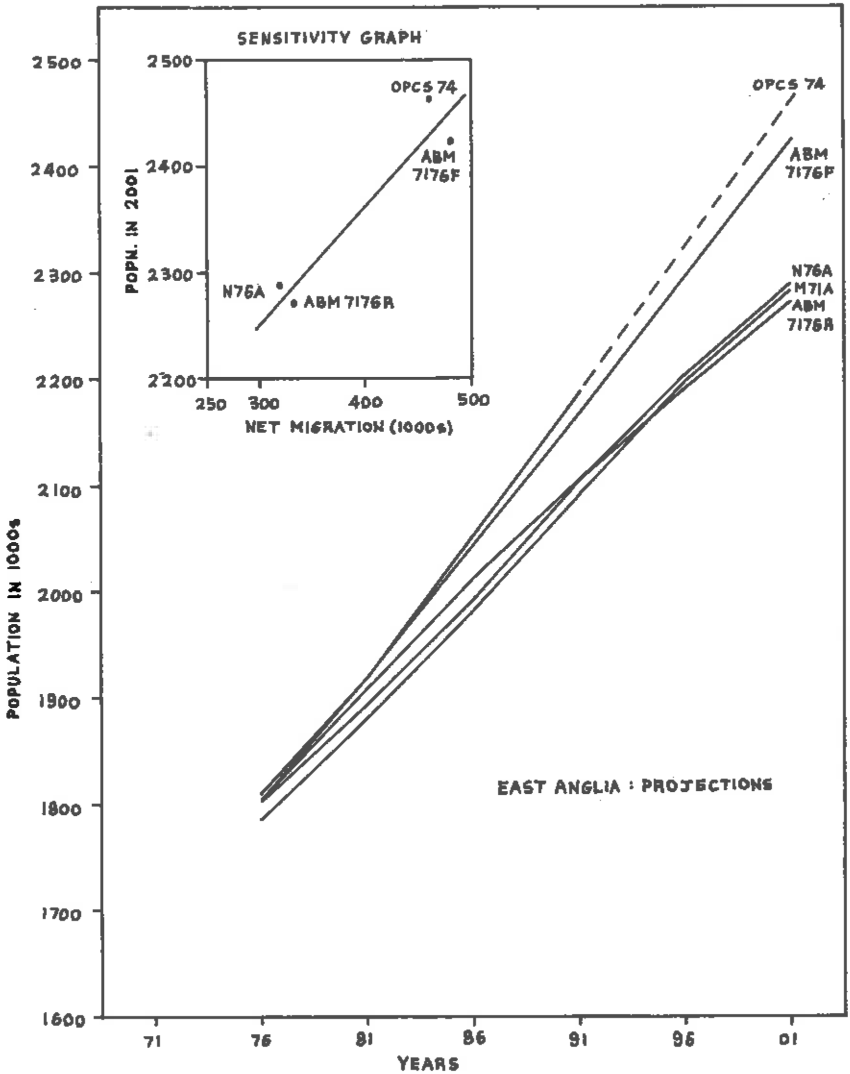

Figure 5 <u>East Anglia population projections with similar fertility levels (TFR=2.2) but incorporating different migration models</u>

computer programs would be a preliminary step, though one involving much effort.

Attention is now turned to the issue of calibration of the base period accounts - the situation of Figure 2.F.

6. Reproduction of observed change in a base period

In constructing population accounts the assumption is initially made that all the different data inputs are mutually consistent. Since they derive from different statistical sources this is rarely the case. Thomas and Rees (1978) reveal substantial differences between estimates of say net migration depending on which data source is regarded as the most reliable.

Figure 6 shows some of the different assumptions that can be made in constructing accounts of population change. The first part of Figure 6 (6.A) shows the standard arrangement of an unconstrained accounts based model. Initial populations, births, and deaths totals by region, and surviving migrants and infant migrants are input to the model and the remaining terms are estimated. One output of the model is the set of column totals for the survival half of the matrix which are the end of period regional populations. Data may well be available for these items and a comparison can be made that usually reveals discrepancies between model estimate and data.

The accounts matrix initially estimated can be adjusted using the well known balancing factor or RAS method so that the column totals of the survival half of the accounts matrix add up exactly to the end of period populations. The accounts based model is then said to be constrained (Figure 2.B). Let us simplify the accounts notation to K^{mn} as the typical element (persons moving from origin lifestate/region to destination lifestate/region), and let us call the row constraints R^m and the column constraints C^n, irrespective of whether the totals figures refer to populations, births, deaths, emigrants or immigrants. Then the adjusted accounts matrix estimates are derived as follows:

$$\hat{K}^{mn} = A^m B^n K^{mn} \qquad (53)$$

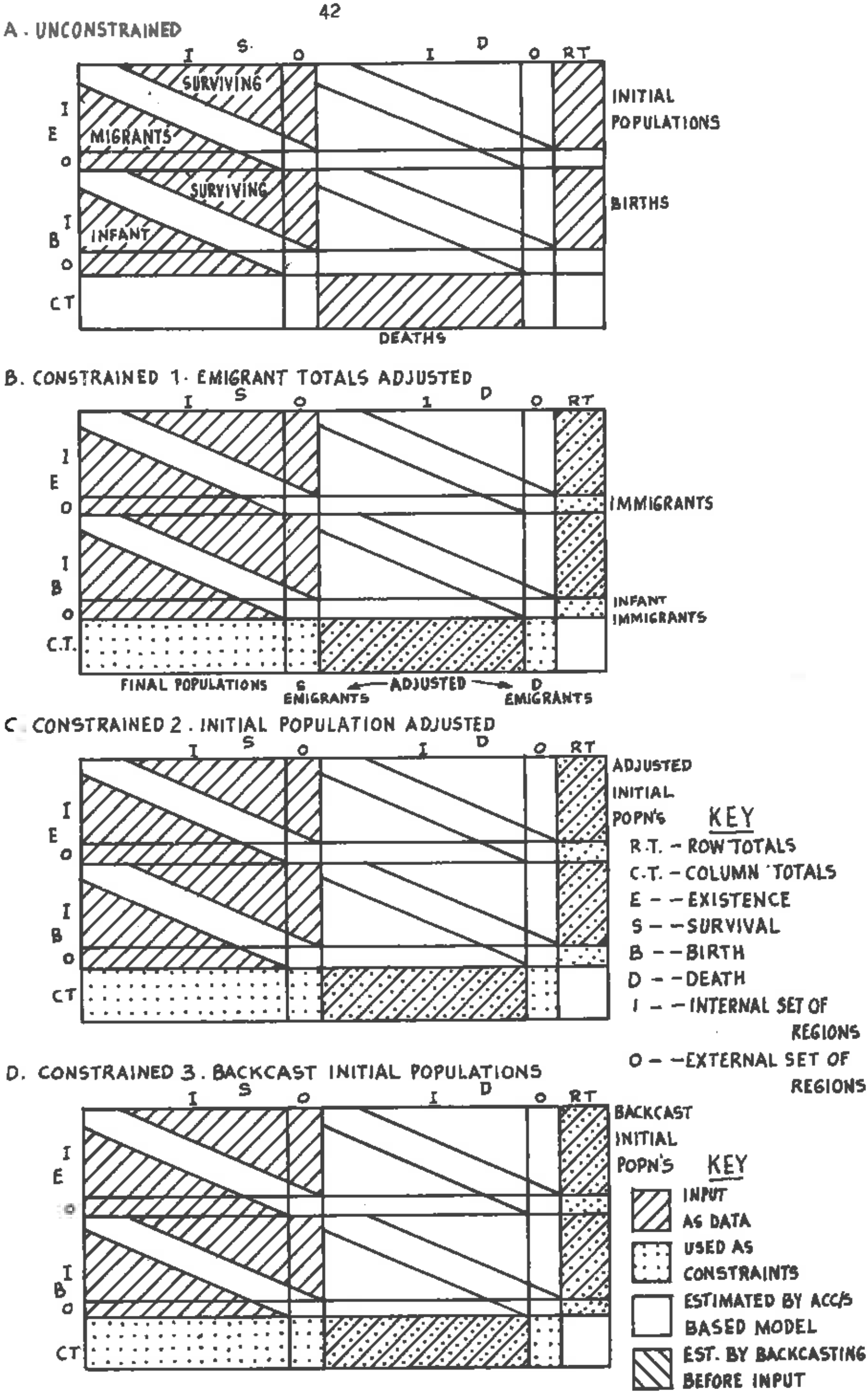

Figure 6. Types of accounts used in British region study

subject to

$$\sum_n \hat{K}^{mn} = R^m \tag{54}$$

and

$$\sum_m \hat{K}^{mn} = C^n \tag{55}$$

Substituting the right hand side of Equation (53) in the left hand sides of Equations (54) and (55), and rearranging, we obtain

$$A^m = R^m / \sum_n B^n K^{mn} \tag{56}$$

$$B^n = C^n / \sum_m A^m K^{mn} \tag{57}$$

These equations must be solved iteratively in the usual fashion (Wilson, 1974; Willekens, 1977).

One condition must be satisfied in order for a solution to this set of equations to exist and that is that the sum of the row constraints must equal the sum of the column constraints:

$$\sum_m R^m = \sum_n C^n \tag{58}$$

This is rarely the case when the relevant data, some of them estimated, have been assembled initially. There is usually a difference, d,

$$(\sum_m R^m) + d = \sum_n C^n \tag{59}$$

which must be distributed among the row and column totals so as to satisfy Equation (58). The decision as to which row or column sums should be first adjusted will depend very much on the situation being studied. In the case of a study of population change in the republics of the Soviet Union (Thomas and Rees, 1978) the difference was apportioned to the initial populations as at 1st January, 1968, since these were regarded as the least reliable statistics among the row and column totals.

In a study of British regions it was felt that the emigrant terms (surviving and non-surviving) were the most unreliable statistics and the difference was apportioned to those terms in developing the final set of population accounts (Rees, 1978b).

$$\hat{K}^{*(*)\sigma(0)} = K^{*(*)\sigma(0)} + d \left[\frac{K^{*(*)\sigma(0)}}{K^{*(*)\sigma(0)} + K^{*(*)\delta(0)}} \right] \quad (60)$$

and

$$\hat{K}^{*(*)\delta(0)} = K^{*(*)\delta(0)} + d \left[\frac{K^{*(*)\delta(0)}}{K^{*(*)\sigma(0)} + K^{*(*)\delta(0)}} \right] \quad (61)$$

where $K^{*(*)\sigma(0)}$ is the column total of surviving emigrants and $K^{*(*)\delta(0)}$ is the column total of non-surviving emigrants, with the ^ referring to adjusted estimate.

An illustration of such calculations is given in Table 7. The constraints initially assembled differ

$$\sum_m R^m = 55{,}128{,}902$$

$$\sum_n C^n = 55{,}146{,}099$$

$$d = 17{,}197 \quad (62)$$

Surviving emigrants were adjusted downwards from 468,422 to 451,325 and non-surviving emigrants from 2,736 to 2,636 to yield a new total of column sums of 55,128,902 equal to the total of row sums. By way of aside it might be noted that the emigration totals are derived from applying the ratio of emigrants to immigrants in the International Passenger Survey (I.P.C.) to the Census derived immigrant total to avoid underestimates of the emigrant terms.

Figure 6.C illustrates an alternative set of adjustments – to the initial populations. This is the method used in the U.S.S.R. republics case (Thomas and Rees, 1978) but is less appropriate for British regions. Another possible approach to accounts construction if it is felt that the initial populations of a period are unreliable is to use the accounting equations in a backwards or "backcast" fashion (the Figure 6.D situation). Details of this method are given in Thomas and Rees (1978). There are thus a variety of ways in which the matrix of demographic change can be "forced" to reproduce observed population change in a period.

Table 7 Constraint calculations illustrated for the 1970-71 British region accounts

Region	Initial constraints: Row Totals (1)	Initial constraints: Column Totals (2)	Adjusted constraints: Row Totals (3)	Adjusted constraints: Column Totals (4)
	Initial populations	Final populations	Initial populations	Final populations
North	3133638	3136785	3133638	3136785
Yorks. & Humb.	4852819	4865284	4852819	4865284
North West	6587191	6599893	6587191	6599893
East Midlands	3602557	3629412	3602557	3629412
West Midlands	5088926	5116516	5088926	5116516
East Anglia	1659731	1681818	1659731	1681818
South East	16961020	16988179	16961020	16988179
South West	4052831	4082495	4052831	4082495
Wales	2715987	2722405	2715987	2722405
Scotland	5212579	5216731	5212579	5216731
	Immigrants	Surviving emigrants	Immigrants	Surviving emigrants
Rest of World	387442	468422	387442	451325
	Births	Deaths	Births	Deaths
North	49831	38152	49831	38152
Yorks. & Humb.	82308	59192	82308	59192
North West	109772	83770	109772	83770
East Midlands	60227	40656	60227	40656
West Midlands	88973	54118	88973	54118
East Anglia	26229	18965	26229	18965
South East	262607	191970	262607	191970
South West	61364	50766	61364	50766
Wales	42696	34930	42696	34930
Scotland	87103	62904	87103	62904
	Infant immigrants	Non-surviving emigrants	Infant immigrants	Non-surviving emigrants
Rest of World	3071	2736	3071	2636
Totals	55128902	55146099	55128902	55128902

7. Reproduction of observed change from the base period up to the present

The question must now be posed: "which method or which set of accounts should be adopted for the base period before embarking on proper projection?" To answer this question we can exploit the gap between the end of the base period and the "then" of Figure 2.E. Projections of the multiregional system of interest can be carried out for that time gap and the results compared with the population estimates available. The projection model can be run in two modes:

(i) In the first mode all rates (birth, death, migration) are assumed fixed. In this case the projections are to some extent being tested (the Figure 2.H situation) although the model is rather naive.

(ii) In the second mode information on birth rates, death rates and external migration flows can be used in the projections, thus eliminating the influence of their misspecification on the projections for the most part. Internal migration rates are assumed to remain constant in the absence of any published alternatives.

Application of the projection model in either or both of these modes will enable us to "calibrate" the base period accounts matrix - that is, select the one that gives the best projective performance in the recent past.

This type of analysis has been carried out for British (standard regions) using the aggregate accounts based model. The accounts being calibrated are those for the year prior to the 1971 Census for which internal migrant data are available. Selected results from this analysis are given in Table 8 for two regions - the South East and Yorkshire and Humberside. All statistics in the analysis apply to post-1st April, 1974 regional boundaries.

The first row under the region name in Table 8 gives the population estimate figures for each year listed in the columns or rather the figure for 24th/25th April in each year interpolated between successive 30th June mid-year home population estimates. These figures are regarded as the best estimates of regional populations available, even though they become less reliable with time elapsed from the 1971 Census. Then the populations projected in four runs of the accounts based model are reported: run UF

Table 8 Selected results from the calibration runs, British regions, 1970-76

	Population at Census date in year indicated (in millions)						
"Model"	1970	1971	1972	1973	1974	1975	1976
	SOUTH EAST						
Estimate	16.961	16.988	17.015	17.019	16.976	16929	16.899
Run UF	16.961	17.019	17.076	17.133	17.189	17.246	17.302
Run UV	16.961	17.019	17.082	17.130	17.134	17.123	17.123
Run C1F	16.961	16.988	17.016	17.044	17.072	17.101	17.131
Run C1V	16.961	16.988	17.028	17.053	17.035	17.002	16.981
	YORKSHIRE AND HUMBERSIDE						
Estimate	4.853	4.865	4.880	4.889	4.896	4.899	4.894
Run UF	4.853	4.854	4.856	4.857	4.859	4.861	4.864
Run UV	4.853	4.854	4.854	4.846	4.837	4.820	4.805
Run C1F	4.853	4.865	4.878	4.890	4.903	4.916	4.928
Run C1V	4.853	4.865	4.872	4.870	4.868	4.857	4.848

Notes:

1. Estimate = population figure derived from interpolating between official mid-year estimate from O.P.C.S. Population estimates, Series PP1.

2. Run UF = run of accounts based model, unconstr. version,fixed vital rates

3. Run UV = " " " " " " " variable " "

4. Run C1F = " " " " " constr. version 1,fixed " "

5. Run C1V = " " " " " " " variable " "

refers to the unconstrained types of accounts of Figure 6.F and the constant projection as referred to in paragraph (i) referred to above; run UV refers to the same type of accounts with variable rates projection as in paragraph (ii) above; run CF refers to the Figure 6.B type of accounts with constant rates; and run CV refers to the Figure 6.B accounts with variable rates. The results are graphed in Figure 7.

The first point made by the graphs is that the constrained accounts model performs better than its unconstrained equivalent in almost all cases, and in both regions in variable rate mode. This result has very general application in the field of population projection and justifies empirically a shift in methodology from Figure 2.B and 2.C situations to 2.D structure.

The relationship between the various curves has rather different meanings in the two regions. In the South East the constrained variable model run (CV) is the only one that follows the substantial turndown in the population of the South East. The gap between the CV run and the CF represents the error from not correctly projecting the downturn in natural increase in the region over the period less the partially counterbalancing upturn in net emigration. The gap between the CV and the E curve represents the error from not correctly projecting the downturn in net internal migration balance for the region over the 1970s. The UF-CF and UV-CV gaps represent the error involved in not adjusting the base period accounts to known constraints. In Yorkshire and Humberside, the errors involved in not correctly projecting a downturn in vital rates (gaps CF-CV, UF-UV) are compensated for by an increase in internal net immigration so that the fixed rate projections appear to do better than the variable rate ones. This result points up the need for updating internal migration data between Censuses.

Table 9 gathers together some summary statistics for both the runs graphed in Figure 7 and two other alternative versions of the constrained accounts. The first statistic given is merely the difference between the total of all the regional populations (that is, the Great Britain population) in the 1976 estimates and those in the projections. The second statistic sums the absolute value of the differences between estimated and projected populations, and the third column gives the sum of the absolute values of these differences as a percentage of the estimated population. The first two constrained accounts versions come out best with little to choose between them. They perform better than either the unconstrained or constrained backcast versions.

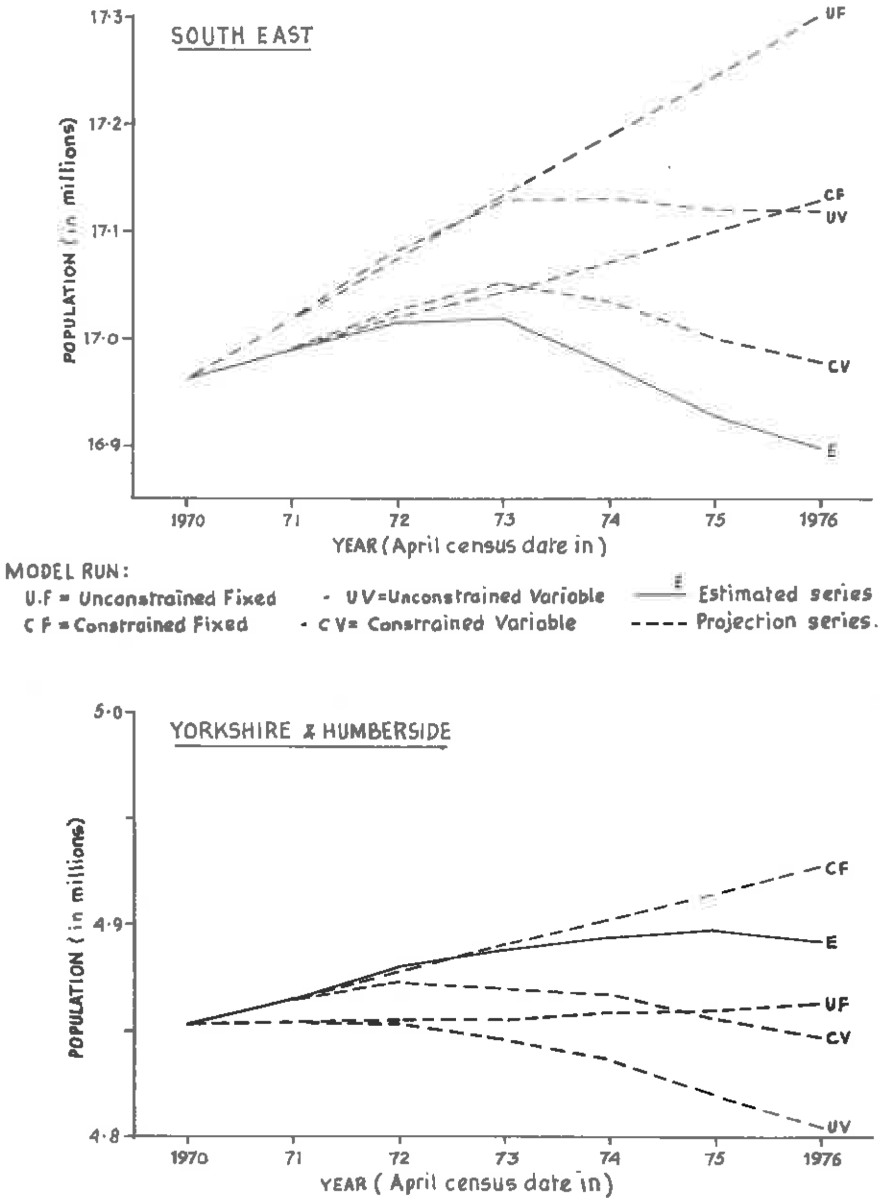

Figure 7. Calibration run results for the South-East and Yorkshire and Humberside, 1970-1976.

Table 9 Calibration statistics for British regions, 1970-76, for 1976 population comparison

Model Run	Difference in total	Sum of absolute differences	Sum of absolute % differences
UF	398870	683248	9.53
C1F	504742	612589	9.81
C2F	418722	554132	9.17
C3F	415790	625963	8.58
UV	-168241	392967	11.39
C1V	-150591	384358	8.20
C2V	-150582	387776	8.15
C3V	-150772	538976	10.13

Notes:

1.

Run	Model	Version	Status of Birth and death rates	Status of external migrants	Type in Figure 6
UF	abm	unconstrained	constant	constant	A
C1F	"	constrained 1	"	"	B
C2F	"	" 2	"	"	C
C3F	"	" 3	"	"	D
UV	"	unconstrained	variable	variable	A
C1V	"	constrained 1	"	"	B
C2V	"	" 2	"	"	C
C3V	"	" 3	"	"	D

2. abm = accounts based model, aggregate version.

3. Difference in total = Population of GB from model run − Population of GB from estimate

Sum of absolute differences

$$= \sum_{i=1}^{10} \left| P^i(\text{model run}) - P^i(\text{estimate}) \right|$$

Sum of absolute % differences

$$= \sum_{i=1}^{10} \left| 100((P^i(\text{model run})-P^i(\text{estimate}))/P^i(\text{estimate})) \right|$$

The final choice made was of the first version of the constrained accounts in which the emigrant terms are adjusted. Using the internal migration rates from these accounts new accounts were prepared using the same constraint techniques for 1971-72, 1972-73, 1973-74, 1974-75 and 1975-76 (Rees, 1978b). The migrant terms in these updated accounts do differ in pattern from those of 1970-71, and reflect in part the unpublished migration information built into the official population estimate series. Projection proper with these updated accounts would start with 1975-76 as a base period.

8. Concluding remarks

In this paper only a small portion of the territory mapped in Figure 1 has been explored. Others are busily working in many of the fields not touched here, further along the modelling sequence (Rogers, 1978; Liaw, 1977; and many others). What this paper has aimed to show is that it is worthwhile tarrying a little on the journey into the future of our regional populations in towns called "data", "accounts", "projection model form" and "calibration".

References

Batty, M. (1976) Urban modelling: algorithms, calibrations, predictions Cambridge University Press, Cambridge.

Baxter, R. and Williams, I. (1978) Population forecasting and uncertainty at the national and local scale, Progress in Planning, 9, Part 1, pp. 1-72, Pergamon Press, Oxford.

Bonsall, P.W., Champernowne, A.F., Mason, A.C. and Wilson, A.G. (1977) Transport modelling: sensitivity analysis and policy testing, Progress in Planning, 7, Part 1, pp. 153-237, Pergamon Press, Oxford.

Jenkins, J.C. and Rees, P.H. (1977) Computer programs and aggregate accounts-based and associated forecasting models of population, Working Paper 205, School of Geography, University of Leeds.

Ledent, J. (1978a) 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. (1978b) Temporal and spatial aspects in the conception, estimation and use of migration rates, Paper presented at the I.I.A.S.A. Conference on "Analysis of Multiregional Population Systems: Techniques and Applications", 19th-22nd September, 1978, International Institute for Applied Systems Analysis, Laxenburg, Austria.

Leslie, P.H. (1945) On the use of matrices in certain population mathematics, Biometrika, 23, pp. 183-212.

Liaw, Kao-Lee (1977) Discrete-time sex- and age-disaggregated spatial population model: theoretical analysis and Canadian case study, Research Report to the Canada Council, McMaster University, Hamilton, Ontario.

Liaw, Kao-Lee (1978) Dynamic properties of the 1966-1971 Canadian spatial population system, Environment and Planning A, 10(4), pp. 389-398.

Lotka, A.J. (1907) Studies on the mode of growth of material aggregates, American Journal of Science, XXIV, pp. 199-216.

McEvedy, C. and Jones, R. (1978) Atlas of world population history, Penguin Books, Harmondsworth, Middlesex.

Office of Population Censuses and Surveys (1977) Population projections, 1975-2015, Series PP2, No. 7, H.M.S.O., London.

Office of Population Censuses and Surveys (1978) Population projections: Area, 1975-1991, England Series PP3, No. 2, H.M.S.O., London.

Pittenger, D.B. (1976) Projecting state and local populations, Ballinger, Cambridge, Massachusetts.

Rees, P.H. (1976) Modelling the regional system: the population component, Working Paper 148, School of Geography, University of Leeds. Revue de Geographie de Lyons, forthcoming.

Rees, P.H. (1978a) Problems of multiregional population analysis: data collection and demographic accounting, Working Paper 229, School of Geography, University of Leeds.

Rees, P.H. (1978b) The spatial analysis of demographic growth in British regions, Final Report to the Social Science Research Council on Grant HR/2914, forthcoming, School of Geography, University of Leeds.

Rees, P.H. (1978c) Migration and settlement in the United Kingdom: dynamics and policy, Working Paper WP-78-36, International Institute for Applied Systems Analysis, Laxenburg, Austria.

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

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

Rogers, A. (1975) Introduction to multi-regional mathematical demography, Wiley, New York.

Rogers, A. (1976) Shrinking large scale population projection models, Environment and Planning A, 8, pp. 515-541.

Rogers, A. (1978) The formal demography of migration and redistribution: measurement and dynamics, Research Memorandum, RM-78-15, International Institute for Applied Systems Analysis, Laxenburg, Austria.

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

Stillwell, J.C.H. (1978) Interzonal migration: some historical tests of spatial interaction models, Environment and Planning A, forthcoming.

Stone, R. (1971) Demographic accounting and model-building, O.E.C.D., Paris.

Thomas, D. and Rees, P.H. (1978) The analysis of population change in the republics of the Soviet Union using multiregional accounting and projection methods, Working Paper, School of Geography, University of Leeds, forthcoming.

Whelpton, P.K. (1928) Population of the United States, 1925 to 1975, American Journal of Sociology, 34(2), pp. 253-270. Cited in Pittenger (1976).

Whelpton, P.K. (1936) An empirical method of calculating future population, Journal of the American Statistical Association, XXXI, pp. 457-473. Cited in Keyfitz, N. (1968) Introduction to mathematics of population, Addison-Wesley, Reading, Massachusetts.

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

Wilson, A.G. (1974) Urban and regional models in geography and planning, Wiley, London.