2. POPULATION PROJECTION: DUTCH AND ENGLISH MULTIREGIONAL METHODS

Philip Rees and Frans Willekens

2.1 Introduction

In both the Netherlands and the United Kingdom national government departments responsible for the production of population projections for both the nation and subnational units have recently sought to improve those projections through the development and implementation of new multiregional models. These are models which project the population of many regions simultaneously and which include the migration flows between them in the analysis. Researchers and consultants outside of central government were asked to prepare these models (Gordijn et al. 1983, Willekens and Drewe 1981; 1984, Martin and Voorhees Associates and Bates 1981). In this chapter these efforts are described and compared in order to assess what has been achieved to date and what improvements could be effected. The earlier sections of the chapter contain an assessment of the needs of users of subnational population projections and a review of the conceptual bases of multiregional models.

2.2 Forecasting requirements

Here we review the demands of users of population projections rather than the wishes of population modellers. The latter need to assess the requirements of the 'market' for their product carefully, and to invest rather more attention than hitherto in the presentation of model outputs. Because of the large number of variables used in multiregional population models, this issue must be the concern of the model designer/systems analyst, and not just a task relegated to a later stage in the development of a projection system.

2.2.1 *Population numbers, age disaggregation and time intervals*

Users of population projections, be they planners or academics, most usually require knowledge of population numbers for short intervals at varying dates in the future. The populations are broken down into detailed age groups, not for their own sake, but in order to yield good estimates of target client or service or market populations. Examples are the school populations in various grades, the numbers of potential students qualified to enter higher education, the labour force, or the retired population. The age classification of the population needed to estimate these target populations must be as fine as possible, preferably by single years of age. Full details of all single year age populations are rarely required but flexible aggregations of these ages are needed. For example, an assessment of the viability of alternative plans for school reorganization might require projections for the 5-8, 9-13, 14-18 age groups in one scheme, and the 5-10, 11-16, 17-18 age groups in another.

Very often, population numbers themselves are insufficient to define the target population. Further disaggregation may be needed. In the school planning case, an estimate is required

of the proportions in any age group entering independent and state supported schools, and within the state sector the proportions or numbers likely to enter denominational or secular schools.

2.2.2 Detailed inputs and summary outputs

Most population projection models can be designed to yield not only fine detail on the population stocks at regular intervals but also fine detail on the demographic components that link successive population stocks. Users do not normally require, for example, detailed information on mortality or migration rates used in the projections by single years of age, but they probably wish to know what those rates mean in terms of summary measures such as the crude death rate, life expectancy, the crude migration rate or gross migration rate. To generalize this point, we can say that multiregional population projection models operate with variables at a level of disaggregation much finer than that required by most users and that any implementation of such models should aim to provide relevant summaries of the information generated.

To exemplify these points, we examine the practice adopted by the United Nations (UN) in their latest, very useful set of population estimates and projections as assessed in 1982 (UN 1985). In this volume two pages of results are devoted to each national projection. The second page shows the estimates and projections of the populations for each country at 5 year intervals from 1950 to 2025 for both sexes and for 5 year age groups to 80+ years in a one page format (compressing two pages of computer printout). The first page of each pair summarizes the population stock information by broad age groups and provides a set of indicators of the age and sex structure of the population in the first of three sub-tables. The second sub-table provides information on component rates in the 5 year time intervals and summarizes the fertility and mortality levels using nine indicators. The final sub-table gives a selection of the indicators for high fertility and low fertility variants. The same set of statistics is provided by the UN for country aggregations such as continental, sub-continental and more/less developed areas, so that comparisons across and between all scales are a simple matter. All the statistics are available to users on magnetic tape for further analyses.

A key feature of the UN projections is the emphasis on presenting statistics over time: time points or intervals are used as the column dimension in all tables. Time series analysis of the results is intended and made easy by such a presentation (e.g. Keyfitz and Vaupel 1986). The UN projections provide guidelines against which the presentation of subnational projection results can be assessed. What is less satisfactory about the UN projections, however, is the lack of detail provided about the models and programs used to generate the results. UN(1985 p.3) suggests that 5 year time and age intervals were used throughout to generate projections shown on the second page, but no details are provided on the methods used to produce the annual population figures and other age group numbers contained on the first page.

2.2.3 A specific or a general model?

Population projection models and their associated software may be specific to the particular

system being studied or they may be used more generally with a variety of systems of interest. Academic researchers incline to the latter kind of model because they wish to study many different systems; government researchers prefer the former, both because the model and associated software are easier to design and because their attention is focussed on the current set of subnational units.

This issue is far more serious if a multiregional modelling strategy is adopted. A single region cohort survival model with net migration rates or flows does not require alteration when used with more regions. The same model is simply used more times. However, in multiregional models, an expansion from N regions to $2N$ may require redesign of the software (because interaction variable dimensions increase in proportion to N squared) or redesign of the model (because with very many regions in relation to the size of the population, many of the cells in the migration matrix may contain small and unreliable numbers). The model used will vary with the size of the system being modelled. Although system specific models are easier to design and program, even in governmental use there will be problems occasioned by the frequent re-specification of subnational areas prompted by politicians either because the settlement system has changed or because electoral advantage may be thereby gained. It may be worthwhile, therefore, investing in a more general model for long term use designed to withstand severe 'spatial shock'.

2.2.4 Issues concerned with the spatial units adopted

The strategies for modelling the populations of a large number of regions have been reviewed in Rogers (1976) and Rees (1979a; 1981b) and may be summarized as follows. (1) Fully multiregional. All regions in the system are modelled simultaneously in a multiregional model in full age and sex detail. Problems of small numbers or unreliable data are ignored.

(2) Partitioned. The system of regions may be partitioned into sub-systems, within which interregional migrations are modelled explicitly while migration between regions in different subsystems is modelled at the subsystem level.

(3) Aggregated. Some of the variables in the system are aggregated or pooled across the spatial or the age dimension of the model. For example, in the subnational projections for England (OPCS 1984a) migration between 108 subnational units is modelled for three aggregate age groups (see section 4 for more details). Another popular choice is to divide the migration process into a process of choosing to move and a process of selecting a destination (Frey 1984).

(4) Biregional. In this strategy the system is collapsed into N sets of 2 regions consisting of the area of interest and the rest of the nation (although flows to and from the rest of the world should also be incorporated in the model).

All the strategies listed above can be termed 'multiregional' because they retain at least some of the interactions and interdependency between regions. Solutions (1) and (4) lead to general and portable models. Solutions (2) and (3) result in models specific to the system being studied. Each solution will yield a different set of projections, and the subnational projections will not necessarily add to the figures obtained in projecting the national population. A decision has then to be made about whether to adjust the subnational

populations or to accept their sum as the national projection.

2.2.5 Summary of the user's requirements

To summarize, users of population projections for subnational units require:

- (1) future populations classified by single years of age and sex;
- (2) summary measures of the inputs to the projections for each interval;
- (3) summary measures of the components of change in each future period;
- (4) time oriented presentation of the projection results;
- (5) standard presentation of the projection results at all scales;
- (6) a model (or its software) general enough to cope with a redefinition of the spatial units of projection; and
- (7) an effective strategy for projecting the population of a large number of units, either across the country or within a particular region.

2.3 Multiregional population modelling: the bare necessities

2.3.1 The case for multiregional population modelling

A strong case has been made for two decades now that the only proper way to project regional populations is to model explicitly the migration flows between them along with the mortality, fertility and external migration components involved (see Rogers 1985a; 1985b, Willekens 1985b, and Rees 1986a for recent surveys). Use of net migration rates or flows can lead in the long run to absurd results. By explicitly modelling interregional migration flows using interregional transition rates, the influence of one region's population dynamics on that of another is captured. The same case can be made for explicitly modelling international migration in a similar way but data difficulties usually mean that a net flow or rate term is used. However, external migration must be included in a regional population projection in order to close the population account (Rees 1986b).

2.3.2 Migration concepts and projection models

For nearly a decade now a clear link has been recognized between the form in which migration activity is recorded and the form which the multiregional population projection model should take (Rees 1977, Ledent 1978; 1979, Ledent and Rees 1980; 1986, Rees 1985). There are three principal sources of migration data: censuses, population registers and surveys. Survey data on migration are not as yet used much for projection purposes and they are not considered further here.

The national census frequently contains a retrospective question on the place of residence at some date in the past. The date can be fixed, for example, when the place of residence n years ago is recorded, or variable, for example, when the place of birth or place of residence at a given age is recorded. Alternatively, the census may contain a question on the previous

residence, without any information on the time of migration. Population registers typically contain information on each change of usual address.

Cross-classifications of regions of residence at two points in time are referred to as 'transition' data and migration data based on changes of residence are denoted as 'movement' data. Transition data are used to compute transition probabilities directly which are entered in a multiregional cohort survival model as rates of migration and survival or survivorship rates. Movement data are used to compute occurrence-exposure rates from which probabilities suitable for entry into multiregional cohort survival models can be estimated.

Fertility and mortality data are treated similarly in the two approaches. It is possible, however, to measure survivorship rates in the transition approach directly without recourse to mortality data. Usually, surviving stayers are estimated as a residual by subtracting internal out-migrants, emigrants and non-survivors from the initial population. The same calculations are carried out in the movement approach but the intra-region components are merely accounting residuals.

2.3.3 The role of population accounts

These computations are best organized in a population accounting framework. Population accounts are tables which display in a consistent fashion all the inputs from and outputs to a regional population. It is not appropriate here to give a full description (see Rees and Wilson 1977, and Rees 1981a for details of transition accounts and Rees 1984 for details of movement accounts), nor is it necessary to construct accounts before computing the rates that enter multiregional population models. But population accounts do provide yardsticks of the consistency with which population models have been put together, and later we use these yardsticks in the assessment of the projection models being reviewed.

2.3.4 Age matters

In constructing any population model care must be exercised in adopting the correct age-time plan (ATP) in which the data are observed or estimated. Four age-time plans may be distinguished (Willekens 1985a): the age-cohort plan, in which events are classified by the cohort to which a person belongs and the age at the time of the event; the age-period (or just period) plan, in which the age and the calendar period at the time of the event are recorded; the period-cohort plan, in which the cohort is recorded to which a person belongs and the calendar period in which the event occurs; and the age-period-cohort plan in which the age and the calendar period at the event together with the cohort are recorded. The period-cohort plan (see Rees and Woods 1986, Figure 12.2) is the appropriate one for projection purposes, and the implementation of any regional projection model is considerably simplified if all age classified data are assembled using this ATP or are converted to it through estimation prior to model specification.

A second age matter is the existence of regularities in the profiles of migration rates by age, thoroughly researched by Rogers and Castro (1986) and by many other authors. Figure 2.1

Figure 2.1 Examples of migration rate profiles for Greater London and West Netherlands

shows two examples from the Netherlands and the United Kingdom. These schedules can be summarized as being composed of three or more of the five components shown in the inset to Figure 2.1. For the Greater London profile we would use a model with a retirement peak. For the West Netherlands, only three components would be necessary. The large number of age specific migration rates (16 for Greater London, 92 for West Netherlands) can be reduced to 9, 11 or 13 model parameters. This proves to be of considerable utility in implementing single year of age multiregional population models.

2.3.5 Summary of the modelling requirements for a subnational population projection model

Work in the field of spatial population dynamics over the past decade suggests that the following guidelines be followed in designing a subnational projection model.

- (1) It should incorporate multiregional features.
- (2) Clear cognizance must be taken of the conceptual type of migration data available in designing the projection model. It has proved too easy in the past to muddle model design in this respect.
- (3) Although it is not essential to construct population accounts prior to projection, all the components included in population accounts must be consciously incorporated in the projection, either implicitly or explicitly.

(4) It is preferable to prepare all input data for the base period and all rate or parameter changes in the projection periods using the period-cohort age-time observation plan. It has also proved very easy in the past to confuse this problem, which is one of estimation, with that of projection model design.

(5) One useful method for handling the problem of too many variables that occurs in multiregional modelling is to use model parameters rather than rate schedules to represent age-disaggregated migration components (Rogers 1986).

2.4 The projection models

We now examine what has been accomplished with the data bases available in each country. The Dutch PRIMOS and MUDEA models are reviewed together with the British model used to generate the internal migration component of the official subnational population projections for England and Wales.

2.4.1 *The PRIMOS model: projection linked to housing construction*

In the early 1980s a PROjection Information and MONitoring System (PRIMOS) was developed at the Delft Centre for Planning Studies for the Directorate General for Housing. Details of the model are given in Gordijn et al. (1983) and an application that assesses the impact of dwelling construction plans in Brouwer et al. (1984). The primary intention of the model was to enable national government to evaluate the influence of new housing construction on both housing shortages and population change in municipalities. To make such an assessment possible, the PRIMOS model includes migration between municipalities in its structure, and hence can be treated as a multiregional model.

The model consists of four parts.

(1) Births, deaths and external migration for the 774 municipalities of the Netherlands are projected and then controlled to the corresponding quantities produced in the national projections of the Central Bureau of Statistics.

(2) Two sub-models are employed for the projection of internal migration. The first computes the component of migration that is a direct consequence of the building of new dwellings. The second computes the structural component of migration, involving the bulk of transfers between existing dwellings.

(3) Potential numbers of households are estimated using population outputs of the first two parts together with headship rates.

(4) The results of the projection are monitored (see Gordijn et al. 1984).

Attention is concentrated here on parts one and two, which involve the projection of the population, disaggregated by age and sex.

Since the model is designed to project the populations of 774 municipalities, averaging 18.5 thousand inhabitants, a projection model that involves a considerable degree of aggregation of the underlying variables is a necessity. For fertility, mortality, immigration and emigration components this involves the following type of model:

age disaggregated
component for
municipality

=

age-specific
population for
municipality

x

national
age-specific rate
for the Netherlands

x

all age
balancing
factor for
municipality

x

age-specific
balancing
factor for
the Netherlands

(2.1)

where the first balancing factor is designed to ensure that the age disaggregated estimates for the municipality add up correctly to the observed all age total, and the second balancing factor ensures that the age disaggregated estimates for the municipality sum to the observed age disaggregated figures for the Netherlands as a whole. This method assumes that the national patterns of age-specific fertility rates, age-specific mortality rates and age-specific external migration rates apply uniformly across the country. Only the level of the component changes. As a first approximation this may be a reasonable way to handle the 'small number' problem, but there is no doubt that age-specific patterns will vary from municipality to municipality. For example, it is likely that age-specific fertility curves will differ among municipalities as a result of residential relocation: some municipalities will be favoured by families in the expansion phase and this will raise fertility rates for women in their twenties compared with other women. To capture this type of variation would, however, be very difficult at such a fine scale.

The second part of the PRIMOS model is concerned with internal migration. The structural component captures the historical pattern of migration flows while the housing construction component measures the migration flows resulting from the creation of new dwellings. For the model to work as a projection model, new housing construction plans need to be used as inputs. Presumably, the historical/structural component includes migrations resulting from past plans: there is no attempt to extract these from the structural component.

The housing construction component works by translating new housing supply in a municipality into demand in other municipalities, discounting this demand by a negative exponential function of travel time squared to that municipality. The supply of dwellings in a municipality is also affected by the supply in other areas. An iterative procedure is used to bring supply and demand into balance (Figure 14.1 in Brouwer et al. 1984). The structural migration component is computed by using three aggregations of the full origin by destination by age group array for the years 1972 to 1979 as inputs to an adjusted multiproportional fitting algorithm. The three aggregations are out-migrations from municipalities by age, in-migrations to municipalities by age and municipality to municipality migrations for all ages. The standard three-face technique is modified through aggregating the inter-municipal matrix to those cells containing the largest flows and using sets of remainder regions (rest of housing market, rest of province, rest of the Netherlands). This reduces the size of the intermunicipal matrix from 599,076 (774 x 774) cells to only 5,851. The estimation problem is also reduced by shrinking the number of age groups considered to only six: 0-14, 15-19, 20-24, 25-34, 35-59 and 60 and over.

Without detailed access to the computer program delivered to the Directorate-General of

Housing, it is not possible to say much in detail about the input requirements and standard outputs of the PRIMOS model beyond the description above. For example, published descriptions of the PRIMOS model do not make it clear how the broad age groups used in the structural migration component are linked to the single year age groups specified in the fertility, mortality and external migration parts of the model. In terms of the user's requirements for a subnational projection model listed earlier in section 2, the PRIMOS model scores well on point 7 (ability to deal with a large number of spatial units), but the aggregation decisions made make it a specific rather than a general and portable model.

In a second Dutch projection model, now considered, the multiregional content is much greater, though the scale of application is much coarser (province or region).

2.4.2 MUDEA: a multiregional model for the Netherlands

Between 1982 and 1984 a multiregional model for subnational population projection was developed at the Netherlands Interuniversity Demographic Institute (NIDI), under the direction of Frans Willekens of NIDI and Paul Drewe of Delft University of Technology under contract to the Physical Planning Agency (RPD) of the Netherlands. The project was overseen by and in response to an inter-ministry Committee for Regional Population Projections (CORBEP). The model structure is fully described in Willekens and Drewe (1984) and its intended application is briefly outlined in ter Heide (1984, p.346). Full details of the computer program and applications are available from NIDI. This model was developed

- (1) for two sexes;
- (2) for a variable number of regions (up to 40; the number is limited by computer memory);
- (3) for flexible age groups (e.g. 1 year or 5 year) and a variable highest age (e.g. 80+ or 90+);
- (4) for a projection interval equal to the age interval (generally 1 or 5 years); and
- (5) for regional projections that are consistent with official national projections (optional).

The modelling strategy adopted was the fully multiregional model in which the complete array of migration variables (migration disaggregated by 2 sexes, N regions of origin, N regions of destination and NA ages) was used. The original intention (Willekens and Drewe 1981) was to base the projection model on earlier models developed at the International Institute for Applied Systems Analysis (IIASA) (Willekens and Rogers 1978; Willekens 1979; Ramchandran 1980). In the event, the MUDEA model incorporated substantial changes from the IIASA models.

- (1) The model deals with two sexes using a female dominant fertility sub-model. That is, male births are a product of the proportion of births that are male multiplied by a set of fertility rates multiplied by a set of female populations at risk (Willekens and Drewe 1984, pp.329-330).

- (2) The age-time plan for the input data was changed from ATP I (the period plan) to ATP II (period-cohort plan) because the focus of interest was projection rather than life table analysis. The original intention to report on multiregional life expectancies was dropped, although Ledent and Rees (1980; 1986) have shown that such expectancies, at birth, can be generated directly from the projection model's survivorship rates and Rees (1986a) has suggested that the whole life table could be estimated from ATP II data.

(3) The MUDEA model is clearly and unambiguously based on migration data of the movement type (as generated by the Dutch registration system), whereas the original IIASA models were ambiguous in their input requirements (because the issues involved were only resolved in the course of the associated project).

(4) The MUDEA model is clearly linked to an underlying set of population accounts of the movement variety, although the flow accounting equations are used rather than the full accounts based model.

(5) One consequence of this is that the MUDEA model fully and carefully incorporates international migration into its structure. Emigration flows are treated in the model in the same way as internal out-migration flows, namely as the product of migration rates transformed into period-cohort transition probabilities and the base population. Immigration flows are treated as exogeneous inputs (which are subjected to survival, fertility and migration rates appropriate to their sojourn after immigration).

Thus, the MUDEA model meets the first four modelling requirements set out in section 2.3.5. There was a clear intention in the original specification of the project (Willekens and Drewe 1981) that the fifth requirement, that the user be able to 'control' the projection by inputting a limited set of key parameters rather than a large body of age specific rates, be met, but this has yet to be accomplished. There is also the ambition (Willekens and Drewe 1981, ter Heide 1984) that the MUDEA model be linked with explanatory models of migration being developed under other RPD research contracts in a second phase. The goal has been partly achieved (Willekens 1986, Drewe 1986, Willekens and Drewe 1986).

How general is MUDEA (recalling the discussion in sections 2.2.4 and 2.2.4)? It is general in the sense that it can be used with any number of regions (up to 40) and any number of ages, and can be employed in other countries where the same input information is available or can be estimated. In principle, the number of regions could be expanded given sufficient computer memory for the resulting arrays, but the flow and rate estimates would rapidly become unreliable because of the small number problem. The MUDEA model could not be directly used at the local level in the Netherlands (as ter Heide 1984 recognizes). Another modelling strategy such as partitioning, aggregation, parameterization or biregional modelling would be required. Inputs required for the MUDEA model are:

- (1) the start of base year and end of base year populations by sex and age (e.g. 0,1,2, ..., 89,90+) for each region;
- (2) births in a year by age (of mother) by region;
- (3) births in a year by sex of child;
- (4) deaths in a year by age by region and sex;
- (5) internal migrations by age by origin, sex and destination;
- (6) emigrations by age by region and sex; and
- (7) immigrations by age by region and sex.

All age classifications for flows are of the period-cohort type. The migration data are all published by the Central Bureau of Statistics on this basis, but data on childbearing and mortality are published according to age at time of the event (period data). They were converted to period cohort form before being entered in the MUDEA model.

Table 2.1 summarizes the considerable printed output that the MUDEA model produces for each period of analysis. Virtually all the model variables are printed out in full age detail (in most tables the row classification is by age). The outputs of the model are clearly voluminous. In a two region test run there were 169 pages of output for one period. The MUDEA program, however, contains parameters to control table printing. The presentation of the results of the multiregional model compares rather unfavourably with that of the UN. Time-oriented tables are essential for easy and fruitful use of projection outputs. Tables should be standard for each spatial unit across all spatial scales. The user should be able to suppress difficult to assimilate matrix style multiregional output (really needed only for error diagnosis) and concentrate on the essential results of the multiregional model.

Table 2.1 Summary of the output tables produced by MUDEA

TABLE VARIABLES		FORM	MACRO-CLASS	ROW-CLASS	COLUMN-CLASS
1.	Population (initial)	Numbers, percentages	Sex	Age(1)	Region
2.	Population (initial)	Numbers, percentages	Sex	Age(5)	Region
3.	Person/years, births,deaths interregional migration, external migration	Numbers, percentages	Sex by region	Age(1)	Components (as under variables)
4.	Person years, births, deaths interregional migration, external migration	Numbers, percentages	Sex by region	Age (5)	Components (as under variables)
5.	Components as in 4 (except person years)	Age-specific rates per 1000	Sex by region	Age(1)	Components (as under variables)
6.	Components as in 5	Age-specific rates per 1000	Sex by region	Age(5)	Components (as under variables)
7.	Components (summarized)	Numbers, percentages years, rates per 1000 mid-year population	-	Sex by region	Components
8.	Birth rates	Age-specific rates	Sex by origin	Age(1)	Destination region
9.	Survivorship rates	Age-specific rates (decimal proportional)	Sex by origin region	Age(1)	Destination region

Table 2.1 (Continued) Summary of the output tables produced by MUDEA

TABLE VARIABLES		FORM	MACRO-CLASS	ROW-CLASS	COLUMN-CLASS
10.	Internal migration	Numbers	Sex by origin region	Age (1)	Destination region
11.	Internal migration	Numbers	Sex by origin region	Age (5)	Destination region
12.	Internal migration	Percentages	Sex by origin region	Age (1)	Destination region
13.	Internal migration	Percentages	Sex by origin region	Age (5)	Destination region
14.	Internal migration	Rate per 1000 (out-migration)	Sex by origin region	Age (1)	Destination region
15.	Internal migration	Rate per 1000 (out-migration)	Sex by origin region	Age (5)	Destination region
16.	Internal migration (summarized)	Numbers, percentages, rates	Sex	Form by origin region	Destination region
17.	Population (final)	Numbers, percentages	Sex	Age (1)	Residence region
18.	Population (final)	Numbers, percentages	Sex	Age (5)	Residence region

Notes:
(i) Percentages - age-specific variable as % of all age total
(ii) Table 8: the column class in the MUDEA output is incorrectly labelled as 'birth region': it is the region of destination of the infant
(iii) Table 10 includes migration between localities within regions which do not figure in the MUDEA model as such
(iv) Age (1) - single years of age Age (5) - 5 year age groups

2.4.3 The OPCS/DOE subnational model for England

Between 1979 and 1981 a model was developed for the projection of the internal migration component of the official subnational population projections for England by Martin and Voorhees Associates and John Bates Services under contract to the Department of the Environment (DOE) of the United Kingdom in close liaison with the Regional Demography Unit of OPCS. Ian Bracken of the University of Wales Institute of Science and Technology

(Cardiff) acted as consultant to the project. The model is fully described in Martin and Voorhees Associates and John Bates Services (1981), in associated programmer's guide and user's manual, and in two journal papers, Bates and Bracken (1982) and Bracken and Bates (1983). Results of using this new model are reported in the fifth of the series of subnational population projections for England (OPCS 1984). A new set of projections incorporating data from the 1981 Census as well as NHSCR data from the post-1981 period have since been prepared.

The model is concerned only with internal migration and is designed to supply a pre-existing single region cohort survival model with net migration values by sex and single years of age, values which can be modified in consultation with local authorities. One ironic benefit of using net migration in the projection model is that it doesn't matter whether the net figures are derived from transition or movement data as long as all transition types have been accounted for (Rees 1985).

Since the project brief required that single year of age populations for males, females and persons be projected for 108 local authority areas in England, a fully multiregional strategy was rejected from the outset because of the small number problem. Rather, an approach was developed which used trended gross migration rates, model generated rates of out-migration and in-migration (a 'parameterized' strategy) and interregion allocation rates for three broad age groups only.

It is useful in exposing the structure of the model in full to present it in general equation form (which the authors do not) as this will reveal a number of interesting features. All the variables in the model are disaggregated by sex, but, since the two populations are treated in the same way in the migration model, no notation for the sex classification is introduced in the description below. The model comprises four stages:

(1) Estimation of the total number of out-migrations

The total number of out-migrations from an area i by age a for a future period t is projected by

$$M^i_a(t) = (gmr^i(t) om^I_a(s) P^i_a(t)) , \quad i \in I \tag{2.2}$$

where

$gmr^i(t)$ = the gross migraproduction rate (gmr) of out-migration from area i in period t , which is a linear extrapolation of the gmr of the area for a standard period in the past.

$om^I_a(s)$ = the proportion of the out-migration gmr accounted for by age group a . It is derived from model migration rates standardized to a gmr of 1 for area cluster I for standard period s ; and

$P^i_a(t)$ = population of region i at age a in period t .

Migration data from the 1971 Census was used to calibrate model migration rate schedules for the 108 areas for males and females. A cluster analysis was then carried out to combine together similar profiles into 12 groups. New model schedules were then defined using the pooled data for the 12 groups.

(2) Assignment of out-migrations to destinations

The total of out-migrants from an area is distributed between available destinations by multiplication by an allocation proportion:

$$M_{a}^{ij}(t) = M_{A}^{i.}(t) k_{A}^{ij}(s) \quad (2.3)$$

where

$M_{a}^{ij}(t)$ = the number of moves from area i to area j in period t in broad age group A ;

$M_{A}^{i.}(t) = \sum_{a \in A} M_{a}^{i.}(t)$; and

$k_{A}^{ij}(s)$ = the proportion of origin i moves in broad age group A in standard period s that have destinations in area j .

In OPCS (1984) NHSCR data for mid-1977 to mid-1982, supplemented with 1971 census data where the spatial disaggregation of the NHSCR data was insufficient, were used. The broad age groups were

- (i) ages 0-16 and 29-59 (family ages)
- (ii) ages 17-28 (labour force ages)
- (iii) ages 60 and over (retirement ages).

The broad age groups were determined by the similarity of their origin-destination patterns.

(3) Estimation of the total number of in-migrations

The inter-area migrations by broad age group are summed and then disaggregated to single years of age:

$$M_{a}^j(t) = \sum_A \sum_j M_{A}^{ij}(t) im_{a}^J(s) \quad (2.4)$$

where

$M_{a}^j(t)$ = total number of in-migrations to area j in period t in age group a ; and

$im_{a}^J(s)$ = the proportion of the in-migration GMR accounted for by age group a . The proportions are derived from model migration rates standardized to a GMR of 1 for area cluster J for standard period s .

(4) Generation of the net migration assumptions

Net migration estimates for each area j by single year of age are then derived as a residual:

$$N_a^j(t) = M_a^j(t) - M_a^j(t) \quad (2.5)$$

and fed into the single region cohort survival model. Methods were also developed to adjust the migration matrices resulting at stage (2) of the model to exogeneous in-migration totals.

The OPCS/DOE model represents an elegant and thoroughly tested solution to the problem of projecting the internal flows of a large number of subnational areal units, carried out by the consultants in a remarkably short period of 18 months. However, a number of further comments can be made based on this algebraic reformulation of the model.

Firstly, it is unfortunate that the migration model is decoupled from the population projection model. This means that the indicators of the components of change offered in the projection model output are very limited. Secondly, it would be very easy to carry out this integration and make the combined migration and projection model fully multiregional by merging equations (2.3) and (2.4):

$$M_a^{ij}(t) = gmr^i(t) om_a^I(s) k_{A}^{ij}(s) P_a^i(t), \quad i \in I, a \in A \quad (2.6)$$

or in other words the migration rate would be modelled as

$$m_a^{ij}(t) = gmr^i(t) om_a^I(s) k_{A}^{ij}(s) \quad (2.7)$$

where s refers to the standard or base period (which may differ between variables). Stage (3) of the model is then unnecessary. Constraints on the number of interregional migrations in each period-cohort can easily be introduced using multiproportional fitting techniques, generalizations of the Furness method referred to by Martin and Voorhees Associates and John Bates Services (1981). When we examine the terms on the right hand side of equation (2.6), it can be seen that the only statistic that is projected forward in time is the gross migraproduction rate. There is no reason why the age factor and the allocation proportion could not also be projected forward given reliable time series, so a more general model would substitute period label t for the standard or base period label s in equations (2.5) and (2.7).

A third point to make about the OPCS/DOE model is that, in exposition, it suffers from a lack of clarity as to age-time plan of observation. The subscript a should refer to a period-cohort ATP throughout. However, the NHSCR data are published using the period ATP and Census migration data are collected using the period-cohort ATP. It is also unclear what populations at risk are being used in rate computation: initial, mid-point, average or final populations. A related fourth point refers to the migration concept being used. Both Census and NHSCR data are used (and have to be used) in the model but it is never made clear which conceptual type of migration (movements or transitions) is being estimated and projected. In the exposition above it is assumed that the model works with the movement concept. OPCS

(1984, p.viii) states that 'the level of migration activity' derives from NHSCR data, whereas in the project report (Martin & Voorhees Associates and John Bates Services 1981) the gross migraproduction rates reported are from census data. This conceptual ambiguity arises from the aim of the model being to generate age disaggregated net migration vectors for areas.

Fifthly, we need to consider the roles of aggregation and sub-model representation in the overall model. The interarea allocation terms, the k 's, involve aggregation of actual migration data into three broad age groups. The alternative of using gravity models of such interarea migration was rejected after investigation because the predicted matrices were felt to be insufficiently accurate. The out-migration factors (om's), involve both aggregation over area cluster and model representation of rates. A great deal of detailed work is involved in the calibration of model migration schedules and their classification though it needs only to be carried out in the British context every ten years. In other situations the use of pooled, observed data to generate standard rates might provide a short cut since the key features of the out-migration rate schedules that vary from area to area (Bracken and Bates 1983, p.353), namely presence/absence of a retirement peak and/or slope, the age at which the labour force migration peak occurs and the extent to which migration activity is concentrated around the peak could be determined by simpler methods.

How general is the OPCS/DOE model (recalling again the discussion of sections 2.4 and 2.5)? It is clearly a very specific model in terms of its implementation and exposition, designed for one set of clients (DOE/OPCS) with particular requirements in terms of area populations to be projected. It is also specific in the sense that the aggregation decisions are peculiar to the system being studied. It is doubtful whether the underlying software would be portable, and the mixture of census and register sources of migration data characteristic of the United Kingdom seems to be repeated only in Japan (Nanjo et al. 1982). However, the methods employed to solve the problem of handling many regions simultaneously are of general utility and could be applied in other contexts.

Table 2.2 summarizes the readily available outputs from the subnational projections. Output tables 1 to 4 appear in OPCS (1984); tables 5 and 6 are the standard detailed projection tables provided on request. A good feature of these outputs is their time orientation, although the published tables are woefully inadequate in the number of future time points for which data are provided - only two! For any serious use of these projections the user needs to obtain the detailed single years of age and time outputs from the Regional Demography Unit. Fuller results of the subnational population projections are now available from the National Online Manpower Information System (NOMIS) at the Department of Geography of the University of Durham (O'Brien 1987). Output tables 3 and 4 provide some idea of the components in number form but not in rate form. Summary indicators such as gross migraproduction rates or total fertility rates or life expectancies are not provided.

As with the MUDEA model, the outputs of the OPCS/DOE model are not as well organized and accessible as the projections of the UN.

Table 2.2 Summary of the output tables produced by the OPCS/DOE model

TABLE VARIABLES		MACRO-CLASS	ROW-CLASS	COLUMN-CLASS
1	Populations (mid-year projected)	England regions	Age (5), ages (summary)	Time by sex
2	Populations (mid-year projected)	Areas (see note iii)	Time by sex	Ages (summary)
3	Natural comp. (births, deaths, natural change)	-	Areas (see note iii)	Components (as variables), two summary time periods
4	Projected in-out and net migration	-	Areas (as in 3)	Migration components (as variables), two summary time periods
5	Populations (mid-year projected)	Areas (as in 3)	Age (1)	Time (years) by sex
6	Populations (mid-year projected)	Areas (as in 3)	Age (5)	Time (years) by sex

Notes:
(i) Tables 1 to 4 appear in OPCS (1984a) and report on population projections from 1981 to 2001
(ii) Tables 5 and 6 are examples of unpublished tabulations obtained from the Regional Demography Unit (Dept. ARV), Office of Population Censuses and Surveys, St. Catherine's House, 10 Kingsway, London WC2B 6JP. Populations to 2011 are provided
(iii) Areas: Table 2: Inner London boroughs, Outer London boroughs, Metropolitan Counties and Districts, Non-Metropolitan Counties; Table 3: England, standard regions, and Table 2 areas

2.5 Conclusions

The purpose of this review has been to look critically at new applications of multiregional methods in order to learn from the experience. Here an attempt is made to summarize what has been learnt.

The theoretical base and mathematical specification of the Dutch MUDEA model are very sound. The projection model is carefully matched with the movement type of migration data used in a period-cohort age-time plan. The English OPCS/DOE model is not as well specified as a set of equations, and there is a lack of clarity concerning both the migration concept aimed at in the model and the age-time plans employed. However, it has the merit of making fruitful use of both census and register derived migration data, using the former to overcome deficiencies in the latter. The Dutch PRIMOS model is specified fairly fully as

a set of equations and solution procedures. However, there is a great deal missing from the published descriptions, which prevents a full evaluation of the methods employed. Occasionally, there appears to be a little confusion as to migration concept with the same variable being referred to as both 'migrations' and 'migrants'.

One essential difference between the two Dutch models is that PRIMOS in relating migration dwelling construction, contains an explanatory element, whereas MUDEA has no mechanism for explanation and produces propositions which are trend-based. Plans to develop an explanatory capability in the latter are currently in hand. The PRIMOS model incorporates migration flows between municipalities into the projection process but with a great deal of aggregation across time, space and age, so perhaps it should be labelled a model with multiregional elements rather than a multiregional model. The Dutch MUDEA model, on the other hand, is a multiregional model that handles all migration flows in age-sex-origin-destination specificity with a considerable degree of flexibility. So, for example, projections of the 11 provinces or of 15 functional regions or 41 urbanization regions could be carried out. The English OPCS/DOE model involves modelling of pooled profiles of out- and in-migration at a detailed age scale and an age pooled method of assigning out-migration to destinations. It is designed to prepare the migration projections for 108 areas and is capable of handling the other subnational areas of the United Kingdom outside England as well either separately or together with those of England.

Whereas the Dutch MUDEA model is easily portable to other demographic systems for which the same kinds of data are available, it is difficult to envisage either the English model or the Dutch PRIMOS model being directly used because of some system specific features. In the English model these are the need to arrive at a classification of in- and out-migration profiles and a pooling of the age specific interarea migration matrices into broad age ranges with common interaction features. The classes and age ranges will vary from situation to situation. Nevertheless the ideas behind the English model could well be applied elsewhere because it provides a solution to the problem of using multiregional models with a large number of regions. The PRIMOS model also uses broad age ranges to incorporate inter-area migration but in a non-portable way, because it is dealing with a system that is almost an order of magnitude more disaggregated than that tackled by the OPCS/DOE model. It provides some useful methodological suggestions on how to tackle such a system.

In terms of user requirements it is possible to comment only on the MUDEA and OPCS/DOE models. They both provide the essential projections by single year of age and sex of the populations of subnational units, controlled by sets of time-varying assumptions as to levels of fertility, mortality and migration, and thus represent considerable advances on previous models and software. The organization and presentation of the results of both models, however, fall short of the ideal. The MUDEA model provides voluminous outputs on a single year of age basis for both inputs and results, and initially it was difficult for users to find the essential statistics amid the haystack of paper. Output tables were provided a year at a time, whereas most users find tabulations of statistics across time points or intervals much easier to contemplate and from which trends can be grasped more easily. However, improvements to the output are in hand. A time orientation is employed in the presentation of the results of the English area projections, but the published volume provides very little information at the level of detail on the key statistics input to each time interval. The outputs of both Dutch and

English models contrast unfavourably with the projection statistics published by the UN for countries of the world. Such a volume and associated tapes, with appropriate modifications, for the subnational areas of the Netherlands and the United Kingdom would be a very valuable planning and research resource.

These efforts in the application of multiregional methods to subnational population projection represent considerable pioneering achievements. There is ample room, however, for improvement in the 'packaging' of models and outputs. Developments in the UK (e.g. the NOMIS system - see O'Brien 1987) suggest that perhaps the most efficient way to provide projection outputs is via a database accessible from a communications network. To this might be added facilities for users to generate their own projections using a centrally maintained software version of the model.