# ADVANCES IN MULTIREGIONAL DEMOGRAPHY

Andrei Rogers (Editor)

RR-81-6
May 1981

**INTERNATIONAL INSTITUTE FOR APPLIED SYSTEMS ANALYSIS
Laxenburg, Austria**

**International Standard Book Number 3-7045-0004-6**

---

*Research Reports*, which record research conducted at IIASA, are independently reviewed before publication. However, the views and opinions they express are not necessarily those of the Institute or the National Member Organizations that support it.

---

# FOREWORD

Demography is concerned with the evolution of human populations, their age and sex structure, and the way in which the components of change, births and deaths, alter this structure over time. Accordingly mathematical demographers have focused their attention on population *stocks* and on population *events*. The need to include several regional populations and the *flows* that interconnect them to form a national multiregional population *system* has led to the development of *multiregional* mathematical demography, which is concerned with the evolution of human populations over space as well as time.

The papers in this volume deal with problems concerning data and measurement, methods of constructing life tables, population projections, analyses of migration patterns and age profiles, aggregation procedures, and the evolutionary dynamics of populations experiencing changing rates of natural increase and migration. The authors are all members of an international group of scholars studying national problems of human settlement at the International Institute for Applied Systems Analysis (IIASA).

The first paper examines an important measurement problem in migration analysis: the transformation of data collected over one unit of time into information covering a different period of time. Data on migration often appear in the form of a response to the question: where did you live $n$ years ago? In Canada and the USA, for example, $n$ is usually taken to be five. Yet the data on births and deaths are reported annually. Thus it is necessary to reconcile one-year with five-year data. *Pavel Kitsul* of the Soviet Union and *Dimiter Philipov* of Bulgaria tackle this problem in their contribution to this volume. They outline an elegant mathematical procedure using matrix theory. The method is illustrated with data for a three-region disaggregation of the population of Great Britain.

Migration data and mortality data for a multiregional population system may be combined to produce estimates of the probabilities of population redistribution and survival. The demographer's normal method of assessing such probabilities is the life table. *Jacques Ledent* of France considers two alternative methods of constructing multiregional life tables, and demonstrates that a computational procedure based on probabilities specific to an individual's region of birth yields more accurate allocations of life expectancies than the more conventional Markov-based solution.

*Dimiter Philipov* of Bulgaria and *Andrei Rogers* of the USA, in work related to that of Ledent, have developed a procedure that generates multiregional population projections disaggregated by region of birth. They outline two classes of projections: *native-independent* projections, in which identical probabilities of transition are assigned to all residents of a region, and *native-dependent* projections, in which these probabilities are further disaggregated by region of birth. The results once again emphasize the importance of including region-of-birth-specific information in demographic analysis.

As part of its work on patterns of migration and settlement in individual nations, IIASA has introduced new techniques for inferring age-specific migration flows from aggregated data. *Frans Willekens* of Belgium, *András Pór* of Hungary, and *Richard Raquillet* of France report on their collaborative work dealing with this topic. They outline a general estimation procedure that incorporates both maximum-likelihood and minimum chi-square estimates. Data for Austria and Sweden are used to illustrate the methodology.

A common demographic approach in mortality studies is the decomposition of mortality rates by cause of death. *Andrei Rogers* of the USA and *Luis Castro* of Mexico use an analogous method to analyze migration rates. They show that different age profiles are associated with different causes of migration. Using data for Czechoslovakia, they demonstrate the ways in which the *levels* and *age profiles* of different cause-specific migration schedules contribute to the aggregate age patterns of migration which change over time and space.

The theory of stable population dynamics has been developed quite thoroughly in the demographic literature, but it is virtually all based on the assumption that fertility, mortality, and migration rates remain unchanged. The case of changing rates has received relatively little attention; not much is known about the influence of variable rates on the age composition and regional distribution of populations. *Young Kim* of the Republic of Korea considers how multiregional zero-growth populations evolve over time when experiencing variations in birth, death, and migration rates. Her paper identifies ways in which the age structure in each region is influenced by the pattern of recent rates and how the effect of the initial composition decreases over time until it is finally lost. Data for India and the Soviet Union illustrate some of the key concepts.

The seventh and last paper in this collection develops a formalism for determining the relationships between a linear Markovian population model and the corresponding aggregated model. *Robert Gibberd* of Australia first shows that an aggregated population model is generally non-Markovian. He then suggests several Markovian approximations, including two which provide upper and lower bounds for the aggregated population distribution. Australian migration data are used to illustrate the results.

It is hoped that the publication of this collection of papers will stimulate further contributions in the field of multiregional demographic analysis.

*Andrei Rogers*
Chairman
Human Settlements
and Services Area

## ACKNOWLEDGMENTS

The editor is grateful to Philip Rees and to Robert Gibberd for a careful review of the entire manuscript and for a number of valuable suggestions. Nathan Keyfitz and Kao-Lee Liaw also kindly read various papers in this collection and contributed comments that were useful during revision. Susie Riley assembled the papers and retyped parts of the resulting manuscript, on which Helen Gasking imposed a consistency through her untiring editorial efforts. To all, sincere thanks.

# CONTENTS

# THE ONE-YEAR/FIVE-YEAR MIGRATION PROBLEM*

*Pavel Kitsul and Dimiter Philipov*

## 1 THE PROBLEM

The analysis of mobility is often restricted by the unavailability of data. Frequently models use cross-sectional data to approximate longitudinal patterns. Problems also arise because the cross-sectional data may refer to different periods of time.

In the case of migration, registration statistics in many countries can be used to produce origin–destination tables of migration flows over a period of one year. Censuses usually also provide such flow data, but over a five- or ten-year period. Statisticians are thus faced with two sets of data, which give different information that may be difficult to reconcile (Rees 1979b). Is one set of data more accurate than the other, or do they reveal different patterns of migration?

This paper investigates the problem of reconciling demographic data collected over different periods of time. The migration example discussed above can be incorporated at an early stage in the construction of the simplest multiregional model: the multiregional life table.

Consider a multiregional population, disaggregated by age, and for which the necessary data on regional populations, births, deaths, and interregional migrations are readily available. Assume that the width of the age group is five years and that the periods of observation can be either one year or five years. Then the multiregional life-table probabilities of migrating can be computed according to eqn. (1) (Rogers and Ledent 1976, Willekens and Rogers 1978)

$$\mathbf{P}_5(x) = [\mathbf{I} + \tfrac{5}{2}\mathbf{M}_1(x)]^{-1}[\mathbf{I} - \tfrac{5}{2}\mathbf{M}_1(x)] \tag{1}$$

where $\mathbf{P}_5(x)$ is the matrix of probabilities $p_{ij}^{(5)}(x)$ that a person at exact age $x$ in region $i$ will be living in region $j$ five years later; $\mathbf{I}$ is the identity matrix; $\mathbf{M}_1(x)$ is the matrix:

$$\mathbf{M}_1(x) = \begin{bmatrix} \left[M_{1\delta}(x) + \sum_{j\neq 1} M_{1j}(x)\right] & -M_{21}(x) & \cdots & -M_{n1}(x) \\ -M_{12}(x) & \left[M_{2\delta}(x) + \sum_{j\neq 2} M_{2j}(x)\right] & \cdots & -M_{n2}(x) \\ \vdots & \vdots & & \vdots \\ -M_{1n}(x) & -M_{2n}(x) & \cdots & \left[M_{n\delta}(x) + \sum_{j\neq n} M_{nj}(x)\right] \end{bmatrix}$$

where $M_{ij}(x)$ are the one-year observed gross migration rates for people aged $x$ to

* Based on WP-80-81.

$x+4$ moving from region $i$ to region $j$, and $M_{i\delta}(x)$ is the annual death rate in region $i$ for individuals aged $x$ to $x+4$. The matrices $\mathbf{M}_1(x)$ and $\mathbf{P}_5(x)$ are of dimension $n \times n$, where $n$ is the number of regions.

A factor of five is introduced into eqn. (1) to reconcile the one-year observed data with the five-year probabilities. It is assumed that the migrations are uniformly distributed over the five-year period (Ledent 1978).

When the observed data refer to a five-year period, the above assumption is not necessary. In this case, eqn. (2) can be used

$$\mathbf{P}_5(x) = [\mathbf{I} + \tfrac{1}{2}\mathbf{M}_5(x)]^{-1}[\mathbf{I} - \tfrac{1}{2}\mathbf{M}_5(x)] \tag{2}$$

where $\mathbf{M}_5(x)$ is a matrix constructed analogously to $\mathbf{M}_1(x)$ from five-year observed gross migration and death rates.

If the assumption that the migrations are uniformly distributed over the period studied is correct, eqns. (1) and (2) should give approximately equal results. In such a case eqn. (1) would be a good approximation to eqn. (2). Results computed for people of exact age 15 migrating within the subsequent five years to another region in Great Britain (East Anglia, South East England, and the rest of Britain) are given in Table 1 both for a one-year period of observation (1970) and for a five-year period (1966–1971). The corresponding results for other ages are given in Appendix C. Two sets of data have been used for the estimations in this paper. The data from the first set refer to the five-year period from 1966 to 1971, and were taken from Rees (1978). The second set refer to the single year 1970, and can be found in Rees (1979a). In the second case the data were originally disaggregated for ten regions but were reaggregated to the three-region system considered here.

TABLE 1 Probabilities of a person at exact age 15 in one of three regions of Great Britain (East Anglia, South East England, or the rest of Britain) living in the same or another region five years later. Calculated using one-year observations and five-year observations and eqns. (1) and (2) respectively.[a]

| Region of origin | Probability of living in region: East Anglia | South East | Rest | Probability of death |
|---|---|---|---|---|
| *One-year observations (1970) and eqn. (1)* | | | | |
| East Anglia | 0.838896 | 0.084048 | 0.073464 | 0.003591 |
| South East | 0.010098 | 0.917494 | 0.069230 | 0.003178 |
| Rest of Britain | 0.005401 | 0.047277 | 0.944153 | 0.003169 |
| *Five-year observations (1966–1971) and eqn. (2)* | | | | |
| East Anglia | 0.898068 | 0.053417 | 0.044920 | 0.003595 |
| South East | 0.007041 | 0.948826 | 0.040965 | 0.003168 |
| Rest of Britain | 0.003073 | 0.030466 | 0.963210 | 0.003251 |

[a] Taken from Rees (1978, 1979a).

It is clear that the probability of leaving the region of origin is substantially higher when calculated using one-year observed data than when calculated using five-year data. Therefore, eqn. (1) must overestimate the probability of migration and underestimate the probability of living in the same region five years later. The probabilities of remaining in the region of origin are represented by the elements of the main diagonal of each table. This is also true for other ages, as shown by the data in Appendix C.

The two sets of probabilities are shown to be significantly different by comparing the corresponding distribution of expectations of life given in Table 2 (see also Appendix D).

TABLE 2 Distribution of expectations of life at exact age 15 in three regions of Great Britain. Calculated using one-year observations and five-year observations and eqns. (1) and (2) respectively.

| | Number of years spent in region | | | |
|---|---|---|---|---|
| Region of origin | East Anglia | South East | Rest | Total |
| *One-year observations (1970) and eqn. (1)* | | | | |
| East Anglia | 18.46 | 17.76 | 23.16 | 59.38 |
| South East | 2.82 | 34.36 | 22.14 | 59.32 |
| Rest of Britain | 1.62 | 11.48 | 45.70 | 58.80 |
| *Five-year observations (1966–1971) and eqn. (2)* | | | | |
| East Anglia | 28.78 | 13.86 | 17.01 | 59.65 |
| South East | 2.48 | 40.97 | 16.01 | 59.46 |
| Rest of Britain | 1.29 | 8.22 | 49.25 | 58.76 |

The distribution of the expectation of life for an individual born in the first region, East Anglia, is markedly different in the two cases. Although not so large, the differences in the distribution of life expectancy for natives of the other two regions are also significant. The same holds true for other ages (Appendix D).

Now compare the probabilities for dying, as shown in Table 1. They are obviously so close that the probability of death calculated using one-year data and eqn. (1) is a good approximation to that calculated using five-year data and eqn. (2). However, eqn. (1) still contains the assumption that invalidated the corresponding approximation in the case of migration; namely, that the observed deaths/migrations are uniformly distributed over the five-year period.

One and the same assumption gives different results: in the case of deaths it is valid, but in the case of migrations it is unjustified. The reason for this difference is that migration may be repeated, unlike death. Migrants are usually identified by comparing their places of residence at the beginning and at the end of the period of interest. Therefore, multiple moves within this period are not counted.

An example is presented in Figure 1. Let an individual reside in region 1 at time 0. He will be a resident of the same region at the end of the first year, but at the end of the second year he will be a resident of region 2. At the end of the third and fourth years he will reside in region 3. Figure 1(a) assumes that he remains in region 3 until at least the end of the fifth year while Figure 1(b) assumes that he moves back to region 1 during this year.

In a one-year data collection system this individual would be registered as a migrant either twice (Figure 1a) or three times (Figure 1b). But if data are

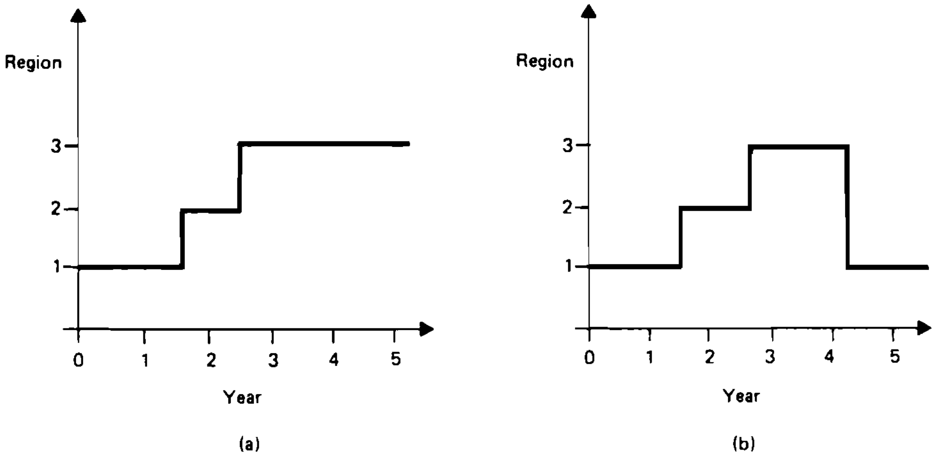



FIGURE 1 Migration of an individual among three regions over a period of five years.

collected over a five-year period, the same individual would register only one move in the case of Figure 1(a) and no move in the case of Figure 1(b).

In the above example, an additional move (from region 3 to region 1 in Figure 1(b)) was registered correctly by yearly observations but resulted in the measurement of one move *less* in the case of five-year observations. This is one type of move responsible for the inaccurate results produced by using a multiplicative factor of five. A detailed description of the ideas outlined above may be found in Rees (1977).

These ideas suggest that an individual's migratory behavior may be represented as a stochastic process. If each move is independent from every other move, and if the probability of a move does not depend on time, the process can be described as a homogeneous Markovian process.

The Markovian assumption gives rise to a new kind of estimating procedure, represented by eqn. (3)

$$\mathbf{P}_5(x) = \{[\mathbf{I} + \tfrac{1}{2}\mathbf{M}_1(x)]^{-1}[\mathbf{I} - \tfrac{1}{2}\mathbf{M}_1(x)]\}^5 \tag{3}$$

which is based on the equality

$$\mathbf{P}_5(x) = [\mathbf{P}_1(x)]^5 \tag{3a}$$

for any Markovian process. Here $\mathbf{P}_1(x)$ is the matrix of probabilities $p^1_{ij}(x)$ that an individual at exact age $x$ in region $i$ will be living in region $j$ one year later. Thus defined, this probability has little demographic meaning, because of the inconsistency between the width of the age group (5 years) and time-period of interest (1 year), but its formal definition is correct. If the Markovian assumption proves to be valid, then $[\mathbf{P}_1(x)]^5$ is already demographically meaningful.*

* However, if the matrices $\mathbf{P}_1(x)$ for $x = 1, 2, 3, \ldots$ are available, $\mathbf{P}_5(x)$ should be approximated by the matrix

$$\mathbf{P}_1(x+4) \cdot \mathbf{P}_1(x+3) \cdot \mathbf{P}_1(x+2) \cdot \mathbf{P}_1(x+1) \cdot \mathbf{P}_1(x).$$

The probabilities of people of exact age 15 living in the same or another region of Great Britain five years later have been calculated using eqn. (3) and are displayed in Table 3. These results are very similar to those obtained using eqn. (1). Hence the Markovian assumption has not introduced any significant improvement. This is also true for other ages (Appendix B).

TABLE 3 Probabilities of a person at exact age 15 in one of three regions of Great Britain living in the same or another region five years later. Calculated using the Markovian assumption.

| Region of origin | Probability of living in region: East Anglia | South East | Rest | Probability of death |
|---|---|---|---|---|
| East Anglia | 0.839297 | 0.083767 | 0.073345 | 0.003591 |
| South East | 0.010063 | 0.917623 | 0.069137 | 0.003178 |
| Rest of Britain | 0.005394 | 0.047212 | 0.944226 | 0.003169 |

Rogers (1965) and Rees (1977) have suggested that the Markovian assumption should be used in analyses of interregional migration. Rees has applied the approach to two sets of data for Great Britain: data from a questionnaire referring to the migration of heads of households, and census data on interregional migration. In the first case, the results obtained were satisfactory but in the second analysis, which included ten regions of Great Britain, the calculated rates differed significantly from the observed rates. After a detailed examination of the problem the author concluded that "... a more complex [than the Markovian] process is involved when an interregional framework is employed" (Rees 1977, p. 262).

The Markovian assumption is theoretically better than the assumption of a uniform distribution of migrations over time, because it allows return migration to be considered (see Figure 1b). It can therefore be thought of as dividing the population into two different groups. Ideas of this kind have been explored by Blumen *et al.* (1955), who gave rise to what is known today as the "mover–stayer" model. This model was later elaborated by Goodman (1961), Spilerman (1972), Boudon (1975), Bartholomew (1973), and others.

The mover–stayer model is based on the assumption that a certain part of the population has a zero probability of migration (*stayers*), while the rest of the population has a non-zero probability of migration (*movers*). Thus all the migrations are made by the "movers". The formal description of the model is

$$\mathbf{P}_5(x) = \alpha\boldsymbol{\pi}_5(x) + (1-\alpha)\mathbf{I} \tag{4}$$

where $0 < \alpha < 1$, $\mathbf{P}_5(x)$ and $\boldsymbol{\pi}_5(x)$ are matrices representing probabilities of migration within the next 5 years for people at exact age $x$, and $\mathbf{I}$ is the identity matrix. $\mathbf{P}_5(x)$ and $\boldsymbol{\pi}_5(x)$ are defined similarly but are different in magnitude.

The Markovian assumption is now applied to the matrix $\boldsymbol{\pi}(x)$, instead of the matrix $\mathbf{P}(x)$. Therefore, if only $\pi_1(x)$ were available, a possible approximation of eqn. (2) would be

$$\mathbf{P}_5(x) = \alpha[\boldsymbol{\pi}_1(x)]^5 + (1-\alpha)\mathbf{I} \tag{5}$$

Note that for $\alpha = 1$, eqn. (5) reduces to eqn. (3a). Note also that, according to this presentation of the mover–stayer model, $\alpha$ does not depend on the region of origin or destination.

Instead of elaborating on the last equation we shall proceed further by considering a possible extension, the high- and low-intensity movers model.

## 2 THE HIGH- AND LOW-INTENSITY MOVERS MODEL

The mover–stayer model was based on the existence of two homogeneous groups of individuals—movers and stayers. In the demographic literature, however, migrants themselves are often divided into two groups with respect to the "parity" or number of moves. One group comprises those migrants who move only once during the period of observation, and the other is composed of individuals who migrate more often. The latter are sometimes referred to as "chronic" migrants. Long and Hansen (1975) report that the rates of return migration to the southern states of the USA are much higher than those for first-time moves to the same destination. However, at the same time, the return migrants constitute only a small part of the total number of migrants (10–20%).

Spilerman (1972) has tried to extend the mover–stayer model by developing the suggestion made by Blumen *et al.* (1955) that a continuous range of intensities of migration should be considered. He proposes a solution to the problem on this basis. However, this model is Markovian and cannot be used in the present case. Boudon (1975) suggests that two homogeneous populations should be considered, both with probabilities of migration greater than zero. He focuses basically on inter-generational occupation tables. The solution of the resulting model is based on the maximum likelihood principle, which causes substantial computational difficulties when dealing with large numbers of equations and unknowns.

In this paper we shall assume, like Boudon, that the population consists of two groups with different intensities of migration, but we propose a different method of solution (matrix diagonalization). It is believed that this will bring the model closer to the demographic idea of migration propensities, and will provide more theoretical insight into methods of dealing with such problems as return migrants or chronic migrants.

Let $p_{ij}^1(x)$ be the probability that an individual at exact age $x$ in region $i$ will live in region $j$ one year later. Let $\sum_{j=1}^{n} p_{ij}^1(x) = 1$, where $n$ is the number of regions. This equation does not take into account the effect of mortality. This assumption is made for convenience, since the matrix of the $p_{ij}^1(x)$ will then be stochastic and its properties will be easier to describe and understand.

Note that the probabilities $p_{ij}^1(x)$ described here are linked with the estimated probabilities $\hat{p}_{ij}^1(x)$ from

$$\hat{\mathbf{P}}_1(x) = [\mathbf{I} + \tfrac{1}{2}\mathbf{M}_1(x)]^{-1}[\mathbf{I} - \tfrac{1}{2}\mathbf{M}_1(x)]$$

where $\hat{\mathbf{P}}_1(x) = [\hat{p}_{ij}^1(x)]$, by the equality

$$p_{ij}^1(x) = \frac{\hat{p}_{ij}^1(x)}{1 - \hat{p}_{i\delta}^1(x)}$$

where $\hat{p}_{i\delta}^1(x)$ is the estimated probability that a person at exact age $x$ in region $i$ will be dead one year later. Bearing in mind that $\sum_{j=1}^{n} \hat{p}_{ij}^{1}(x) + \hat{p}_{i\delta}^{1}(x) = 1$, it must be true that $\sum_{j=1}^{n} p_{ij}^{1}(x) = 1$.

The formal description of the extension of the mover–stayer model considered here is based on the equality

$$p_{ij}^{1}(x) = \alpha_{ij}(x)\pi_{ij}(x) + [1 - \alpha_{ij}(x)]\rho_{ij}(x) \tag{6}$$

where $\pi_{ij}$ and $\rho_{ij}$ are probabilities with meanings analogous to that of $p_{ij}^{1}$, and $\alpha_{ij}(x)$ is a real parameter, $0 < \alpha < 1$. The equality shows that the probability $p_{ij}^{1}(x)$, which refers to the *total* population of region $i$ at exact age $x$, is the weighted sum of two probabilities $\pi_{ij}(x)$ and $\rho_{ij}(x)$, which refer to subgroups of this regional population, with weights $\alpha_{ij}(x)$ and $[1 - \alpha_{ij}(x)]$ respectively. The model defined by the above probabilities is called the *high- and low-intensity movers model*, to differentiate it from the extension developed by Spilerman (1972).

In order to make use of this model to estimate $p_{ij}(x)$, it is necessary to know the values of $\alpha_{ij}(x)$, $\pi_{ij}(x)$, and $\rho_{ij}(x)$. Unfortunately, these data are unavailable. A number of further assumptions must therefore be made in order to find a convincing method of solving for $\alpha$, $\pi$, and $\rho$.

We shall first assume that the parameter $\alpha_{ij}(x)$ does not depend on the regions $i$ and $j$, i.e., that the two groups with different probabilities of migration are not separated on a regional basis. This means that factors other than the region of residence (for instance, social status and economic occupation) affect the probabilities of migration and the number of return migrants and chronic migrants. The validity of this and other assumptions is discussed later in the paper.

The matrix equivalent of eqn. (6) is

$$\mathbf{P}_1(x) = \alpha(x)\boldsymbol{\pi}_1(x) + [1 - \alpha(x)]\boldsymbol{\rho}_1(x) \tag{7}$$

where $\alpha(x)$ is a scalar depending on the age $x$. Note that $\alpha(x)$ and the elements of the two matrices $\boldsymbol{\pi}_1(x)$ and $\boldsymbol{\rho}_1(x)$ are all non-negative.

We shall further assume that the stochastic processes defined by the stochastic matrices $\boldsymbol{\pi}_1(x)$ and $\boldsymbol{\rho}_1(x)$ are Markovian. Thus we assume that these matrices satisfy the Kolmogoroff–Chapman equations (Chiang 1968, Karlin 1969). If so, the overall process, defined by $\mathbf{P}(x)$, is a mixture of two Markovian processes.

The mixture of two Markovian processes is generally not itself a Markovian process. Since $\alpha(x) = 1$ reduces the process to the Markovian process defined by $\mathbf{P}_1(x)$, the high- and low-intensity model is a non-Markovian extension of the Markovian model.

Equation (7) was based on a one-year period of observation. More generally, if the period of observation is $\tau$ years, the process can be represented by eqn. (8).

$$\mathbf{P}_\tau(x) = \alpha(x)\boldsymbol{\pi}_\tau(x) + [1 - \alpha(x)]\boldsymbol{\rho}_\tau(x) \tag{8}$$

The Markovian assumption for $\boldsymbol{\pi}_\tau$ and $\boldsymbol{\rho}_\tau$ gives the following relationships between processes involving different values of $\tau$ ($\tau = 1$ and $\tau = 5$, say):

$$\begin{aligned} \boldsymbol{\pi}_5(x) &= [\boldsymbol{\pi}_1(x)]^5 \\ \boldsymbol{\rho}_5(x) &= [\boldsymbol{\rho}_1(x)]^5 \end{aligned} \tag{9}$$

With a knowledge of eqns. (9), eqn. (8) can be used to form the system

$$\begin{aligned} \mathbf{P}_1(x) &= \alpha(x)\boldsymbol{\pi}_1(x) + [1 - \alpha(x)]\boldsymbol{\rho}_1(x) \\ \mathbf{P}_5(x) &= \alpha(x)[\boldsymbol{\pi}_1(x)]^5 + [1 - \alpha(x)][\boldsymbol{\rho}_1(x)]^5 \end{aligned} \tag{10}$$

If this system can be solved with respect to the unknowns $\alpha(x)$ and the elements of $\boldsymbol{\pi}_1(x)$ and $\boldsymbol{\rho}_1(x)$, the one-year/five-year migration problem can be attacked using the newly formulated model. Hence we proceed to solve system (10) with respect to $\alpha(x)$, $\pi_{ij}(x)$, and $\rho_{ij}(x)$ for each $i,j = 1, 2, \ldots, n$. There are $2n^2+1$ unknowns in system (10), where $n$ is the number of regions, and $2n^2+2n$ equations (the $2n^2$ comes from the dimension of the matrices, and the $2n$ from the restrictions $\sum_{j=1}^{n} \pi_{ij} = 1$ and $\sum_{j=1}^{n} \rho_{ij} = 1$).

In finding the solution of system (10) we are faced with a problem caused by the large number of non-linear equations. This non-linear system is also overdetermined. For instance, for $n = 3$ there are 19 unknowns and 24 equations. The two problems will be considered together.

Consider the system of Kolmogoroff differential equations (Chiang 1968)

$$\frac{\mathrm{d}}{\mathrm{d}\tau}\mathbf{P}(\tau) = \mathbf{P}(\tau)\boldsymbol{\mu}$$

with the initial condition

$$\mathbf{P}(0) = \mathbf{I}$$

The elements $\mu_{ij}$ of the matrix $\boldsymbol{\mu}$ represent the "intensity" or "force" of migration from region $i$ to region $j$. The elements satisfy the conditions

$$\sum_{j=1}^{n} \mu_{ij} = 0$$

$$\mu_{ij} \geqslant 0 \quad \text{for } i \neq j$$

$$\mu_{ii} \leqslant 0$$

Some important properties of $\boldsymbol{\mu}$ are given by Chiang (1968).

The formal solution of the system of Kolmogoroff differential equations is

$$\mathbf{P}(\tau) = \mathrm{e}^{\boldsymbol{\mu}\tau} \tag{11}$$

The definition of $\mathrm{e}^{\boldsymbol{\mu}\tau}$ as a matrix function is given by Gantmacher (1959, Chapter V).

The matrices $\boldsymbol{\pi}_1(x)$ and $\boldsymbol{\rho}_1(x)$ represent Markovian processes and (for $\tau = 1$) they can therefore be written

$$\boldsymbol{\pi}_1(x) = \mathrm{e}^{\boldsymbol{\mu}_\pi(x)}$$

$$\boldsymbol{\rho}_1(x) = \mathrm{e}^{\boldsymbol{\mu}_\rho(x)}$$

Then system (10), with $\alpha(x)$ set equal to $\alpha$, may be transformed to

$$\begin{aligned} \mathbf{P}_1(x) &= \alpha\, \mathrm{e}^{\boldsymbol{\mu}_\pi(x)} + (1-\alpha)\, \mathrm{e}^{\boldsymbol{\mu}_\rho(x)} \\ \mathbf{P}_5(x) &= \alpha\, \mathrm{e}^{5\boldsymbol{\mu}_\pi(x)} + (1-\alpha)\, \mathrm{e}^{5\boldsymbol{\mu}_\rho(x)} \end{aligned} \tag{12}$$

Note that on the right-hand side of the equation the probabilities of migration have been replaced by the corresponding intensities of migration.

Next we introduce the assumptions

$$\begin{aligned} \boldsymbol{\mu}_\rho(x) &= k(x)\boldsymbol{\mu}_\pi(x) \qquad && 0 < k(x) < 1 \\ k(x) &= k && \text{for all } x \end{aligned} \tag{13}$$

This means that the difference in the propensity to migrate for individuals from the two groups (weighted by the parameter $\alpha$) is independent of the regions $i$ and $j$. Introducing assumptions (13) into system (12) and denoting $\boldsymbol{\mu}_\pi$ by $\boldsymbol{\mu}$ we obtain

$$\begin{aligned}\mathbf{P}_1(x) &= \alpha\, e^{\boldsymbol{\mu}(x)} + (1-\alpha)\, e^{k\boldsymbol{\mu}(x)} \\ \mathbf{P}_5(x) &= \alpha\, e^{5\boldsymbol{\mu}(x)} + (1-\alpha)\, e^{5k\boldsymbol{\mu}(x)}\end{aligned} \tag{14}$$

Introducing the assumptions (13) means that the number of unknowns is reduced from $2n^2+1$ in eqns. (10) to $n^2+2$ in eqns. (14), the number of equations being reduced to $2n^2+n$, as the restrictions $\sum_{j=1}^n \pi_{ij} = 1$ and $\sum_{j=1}^n \rho_{ij} = 1$ are replaced by $\sum_{j=1}^n \mu_{ij} = 0$.

For $n = 3$, there will be 11 unknowns and 21 equations. For $n > 3$, the number of equations will increasingly exceed the number of unknowns. Therefore, for $n \geqslant 3$, the solution must be found indirectly. We shall use the method of matrix diagonalization to decrease the dimension of the problem and the degree of its overdetermination.

Assume that there are $n$ eigenvalues of $\mathbf{P}_1$ and that they are all different.* (This assumption is usual in the social sciences and adequately reflects real-world situations.) Then the transformation $\mathbf{T}_1$ which diagonalizes $\mathbf{P}_1$ is defined by the $n$ different right eigenvectors. Analogously let $\mathbf{P}_5$ be diagonalized by $\mathbf{T}_5$. By $\mathbf{T}^{-1}$ we denote the inverse of the matrix $\mathbf{T}$. Hence, $\mathbf{T}_1^{-1}$ and $\mathbf{T}_5^{-1}$ are constructed by the left eigenvectors of $\mathbf{P}_1$ and $\mathbf{P}_5$, respectively. For more details about diagonalization see, for instance, Bellman (1960), Chiang (1968), or Gantmacher (1959).

Let $\mathbf{T}_1^{-1}\mathbf{P}_1\mathbf{T}_1 = \text{diag}\,(\mathbf{P}_1) = \boldsymbol{\Lambda}_1$, where $\boldsymbol{\Lambda}_1$ is a diagonal matrix of the eigenvalues of $\mathbf{P}_1$. Correspondingly, let diag $(\mathbf{P}_5) = \boldsymbol{\Lambda}_5$. Introducing the diagonalization into eqns. (14) gives

$$\begin{aligned}\boldsymbol{\Lambda}_1 &= \mathbf{T}_1^{-1}[\alpha\, e^{\boldsymbol{\mu}} + (1-\alpha)\, e^{k\boldsymbol{\mu}}]\mathbf{T}_1 \\ \boldsymbol{\Lambda}_5 &= \mathbf{T}_5^{-1}[\alpha\, e^{5\boldsymbol{\mu}} + (1-\alpha)\, e^{5k\boldsymbol{\mu}}]\mathbf{T}_5\end{aligned}$$

and hence

$$\begin{aligned}\boldsymbol{\Lambda}_1 &= \alpha\mathbf{T}_1^{-1}\, e^{\boldsymbol{\mu}}\mathbf{T}_1 + (1-\alpha)\mathbf{T}_1^{-1}\, e^{k\boldsymbol{\mu}}\mathbf{T}_1 \\ \boldsymbol{\Lambda}_5 &= \alpha\mathbf{T}_5^{-1}\, e^{5\boldsymbol{\mu}}\mathbf{T}_5 + (1-\alpha)\mathbf{T}_5^{-1}\, e^{5k\boldsymbol{\mu}}\mathbf{T}_5\end{aligned} \tag{14a}$$

It will be necessary to use a certain class of matrices, which are defined as follows:

> The matrices $\mathbf{A}$ and $\mathbf{B}$ are *related*** if they can be diagonalized by the same transformation $\mathbf{T}$.

It is easy to show that if the matrices $\mathbf{A}$ and $\mathbf{B}$ are related, then the matric $\mathbf{C}$, where

$$\mathbf{C} = a\,\mathrm{f}(\mathbf{A}) + b\,\mathrm{g}(\mathbf{B})$$

---

* To simplify the notation, age groups will no longer be denoted.

** The authors would like to thank A. Seifelnasr, who indicated that the word "similar" which was used here originally was inappropriate because this term is used in the literature to define another class of matrices.

and f( · ) and g( · ) are scalar functions and $a$ and $b$ are real numbers, is also related to $\mathbf{A}$ and $\mathbf{B}$ (Gantmacher 1959, Chapter V). In particular, if $\mathbf{v}$ is the diagonalized matrix diag ($\boldsymbol{\mu}$), then

$$\text{diag}\,(e^{\mu}) = e^{v}$$

Consider now the system (14a). Since the left-hand side of each equation is a diagonal matrix, the same will be true of the matrix sum on the right-hand side. But the matrices $\boldsymbol{\mu}$ and $k\boldsymbol{\mu}$ are related and therefore, from the equation above, the matrices $e^{\mu}$ and $e^{k\mu}$ are also related. Hence they are diagonalized by the same transformation $\mathbf{U}$. Then $\mathbf{U}$ diagonalizes a linear combination of $e^{\mu}$ and $e^{k\mu}$, and hence diagonalizes $\mathbf{P}_1$ as well. Then $\mathbf{P}_1$ and $e^{\mu}$, or $e^{k\mu}$, are also related. If so, the transformation $\mathbf{T}_1$ diagonalizes $e^{\mu}$ and $e^{k\mu}$.

Analogously, $\mathbf{T}_5$ diagonalizes the related matrices $\mathbf{P}_5$, $e^{5\mu}$, and $e^{5k\mu}$. Then eqns. (14a) can be represented as

$$\begin{aligned} \Lambda_1 &= \alpha\, e^{v} + (1-\alpha)\, e^{kv} \\ \Lambda_5 &= \alpha\, e^{5v} + (1-\alpha)\, e^{5kv} \end{aligned} \tag{15}$$

bearing in mind the similarity between $\boldsymbol{\mu}$, $k\boldsymbol{\mu}$, $5\boldsymbol{\mu}$, and $5k\boldsymbol{\mu}$, and applying successively the property of matrix functions cited above.

Note that $e^{\mu}$ and $e^{5\mu}$ (or $e^{k\mu}$ and $e^{5k\mu}$) are related, so that the transformations $\mathbf{T}_1$ and $\mathbf{T}_5$ should diagonalize them both. This implies that the matrices $\mathbf{P}_1$ and $\mathbf{P}_5$ should also be related, and be diagonalized by either $\mathbf{T}_1$ or $\mathbf{T}_5$. However, since transformations are unique, $\mathbf{T}_1$ and $\mathbf{T}_5$ should be equal. This condition is too rigid to be met in practice, but we can relax it a little by assuming that $\mathbf{T}_1$ and $\mathbf{T}_5$ are empirically close enough to meet the theoretical requirements, i.e., that when applied to the diagonal matrices $\Lambda_1$ and $\Lambda_5$, they yield the initial matrices $\mathbf{P}_1$ and $\mathbf{P}_5$, as shown in eqns. (16)

$$\begin{aligned} \hat{\mathbf{P}}_1 &= \mathbf{T}_5 \Lambda_1 \mathbf{T}_5^{-1} \approx \mathbf{P}_1 \\ \hat{\mathbf{P}}_5 &= \mathbf{T}_1 \Lambda_5 \mathbf{T}_1^{-1} \approx \mathbf{P}_5 \end{aligned} \tag{16}$$

If the expressions (16) do not hold, the whole theory developed up to now is not valid. This would mean that the Markovian assumptions or some of the assumptions made for the matrices $\boldsymbol{\pi}$ and $\boldsymbol{\rho}$ are not justified. The accuracy of eqns. (16) therefore provides a measure of the validity of the model considered here.* The numerical expressions for $\hat{\mathbf{P}}$ and $\mathbf{P}$ are compared in the next section; at the moment it is sufficient to state that $\hat{\mathbf{P}}$ and $\mathbf{P}$ are close enough to suggest that the model is valid. The observed and estimated results are given in full in Appendix B.

Let $\lambda_i(\mathbf{P}_1)$ be the $i$th eigenvalue of $\mathbf{P}_1$ and $\lambda_i(\mathbf{P}_5)$ the $i$th eigenvalue of $\mathbf{P}_5$. Let $\nu_i$ be the $i$th eigenvalue of $\boldsymbol{\mu}$. Then the system of matrix equations (15) can be presented as the non-linear system of equations

$$\left.\begin{aligned} \lambda_i(\mathbf{P}_1) &= \alpha\, e^{\nu_i} + (1-\alpha)\, e^{k\nu_i} \\ \lambda_i(\mathbf{P}_5) &= \alpha\, e^{5\nu_i} + (1-\alpha)\, e^{5k\nu_i} \end{aligned}\right\} \quad i = 1, 2, \ldots, n \tag{17}$$

The matrices $\mathbf{P}_1$ and $\mathbf{P}_5$ are stochastic. Therefore their largest eigenvalue is equal to unity and the corresponding eigenvalue of $\boldsymbol{\mu}$ is equal to zero. Hence, two of

---

* Some theoretical aspects of this approximation are considered in Appendix A.

the equations from system (17) must be excluded. The number of the equations will then decrease to $2n-2$. At the same time, the number of unknowns is $n+1$ (since for some $i$, $\nu_i = 0$), which is a substantial decrease when compared with the $n^2+2$ resulting from eqns. (14).

Let $n = 3$ (the case $n = 2$ is better handled by eqns. 14). There are 4 equations and 4 unknowns; therefore the system is well defined.

Let $n > 3$. There are then more equations than unknowns. Therefore, if the system is consistent, we can use the same method of solution as for $n = 3$.

The solution for the case $n = 3$ is considered below.

In order to simplify the notation, let $z_i = e^{\nu_i}$. Let $\lambda_1(\mathbf{P}_1)$ and $\lambda_1(\mathbf{P}_5)$ be equal to unity, hence $\nu_1 = 0$. Then eqns. (17) can be rewritten as

$$\left.\begin{aligned} \lambda_i(\mathbf{P}_1) &= \alpha z_i + (1-\alpha) z_i^k \\ \lambda_i(\mathbf{P}_5) &= \alpha z_i^5 + (1-\alpha) z_i^{5k} \end{aligned}\right\} \quad i = 2, 3 \tag{18}$$

Let $k$ be held fixed. System (18) can be rearranged as in eqns. (19)

$$\left.\begin{aligned} \alpha &= [\lambda_i(\mathbf{P}_1) - z_i^k]/(z_i - z_i^k) \\ \alpha &= [\lambda_i(\mathbf{P}_5) - z_i^{5k}]/(z_i^5 - z_i^{5k}) \end{aligned}\right\} \quad i = 2, 3 \tag{19}$$

and hence

$$[\lambda_i(\mathbf{P}_1) - z_i^k]/(z_i - z_i^k) = [\lambda_i(\mathbf{P}_5) - z_i^{5k}]/(z_i^5 - z_i^{5k}) \qquad i = 2, 3$$

Note that the above equations are well defined, since the exclusion of the eigenvalue $\nu_1 = 0$ ensures that all the denominators are non-zero.

This leaves us with three unknowns: $k$, $z_2$, and $z_3$. An additional restriction is provided by the assumption that $\alpha$ does not depend on the regions. Therefore the solutions for $z_2$ and $z_3$ must be such that eqns. (19) yield the same value for $\alpha$. The last condition is used to construct an algorithm for solving eqns. (18).

*Step 1.* Fix an arbitrary value for $k$ such that $0 < k < 1$.

*Step 2.* Form the function

$$f(z_i) = [\lambda_i(\mathbf{P}_1) - z_i^k](z_i^5 - z_i^{5k}) - [\lambda_i(\mathbf{P}_5) - z_i^{5k}](z_i - z_i^k)$$

for the given value of $k$.

*Step 3.* Find the roots of $f(z_i) = 0$ using the Newton–Raphson approximation

$$z_{n+1} = z_n - \frac{f(z_n)}{f'(z_n)}$$

starting with $z_0 = 0.01$. Recall that $z_i$ is bounded in the interval (0, 1) because $z_i = \exp(\nu_i)$ and $\nu_i < 0$.

*Step 4.* With the values of $z_i$, estimate $\alpha$ from eqns. (19). Let $z_i$ provide an estimated value of $\alpha$ denoted by $\alpha_i$.

*Step 5.* If $\alpha_2 \neq \alpha_3$, go back to Step 1. If $\alpha_2 = \alpha_3$ (up to a predefined tolerance level), the solution has been found.

The small initial value for $z_0$ is assumed in order to exclude the trivial root $z_i = 1$, which gives $1 = z_i = e^{\nu_i}$, i.e., $\nu_i = 0$.

The solution of eqns. (18) and the values of $\alpha$, $k$, $\nu_2$, and $\nu_3$ can be used to construct the matrices $\boldsymbol{\pi}$ and $\boldsymbol{\rho}$. Thus the initial system (10) can be constructed numerically.

It is possible to find an approximate solution by minimizing a function $F$ of four variables:

$$F(z_2, z_3, \alpha, k) = \sum_{i=2}^{3} \{[\lambda_i(\mathbf{P}_i) - \alpha z_i - (1-\alpha) z_i^k]^2 + [\lambda_i(\mathbf{P}_5) - \alpha z_i^5 - (1-\alpha) z_i^{5k}]^2\}$$

This method of solution was found to give the same results as the one described above, and is to be preferred if library nonlinear-optimization routines are available.

## 3 NUMERICAL VERIFICATION

Consider the two matrices $\mathbf{P}_1$ and $\mathbf{P}_5$ for the group aged 15–19 years in the three regions of Great Britain considered in the first section. Let the effect of mortality be eliminated, so that the two matrices are stochastic, that is, with row elements summing to unity. Their numerical expressions are then

$$\mathbf{P}_1 = \begin{bmatrix} 0.96614 & 0.01829 & 0.01556 \\ 0.00220 & 0.98320 & 0.01460 \\ 0.00114 & 0.00997 & 0.98889 \end{bmatrix}$$

$$\mathbf{P}_5 = \begin{bmatrix} 0.90131 & 0.05361 & 0.04508 \\ 0.00706 & 0.95184 & 0.04109 \\ 0.00308 & 0.03056 & 0.96635 \end{bmatrix}$$

The eigenvalues are: $\lambda_1(\mathbf{P}_1) = 1$; $\lambda_2(\mathbf{P}_1) = 0.96405$; $\lambda_3(\mathbf{P}_1) = 0.97419$; $\lambda_1(\mathbf{P}_5) = 1$; $\lambda_2(\mathbf{P}_5) = 0.89477$; $\lambda_3(\mathbf{P}_5) = 0.92473$.

The eigenvalues of each matrix are different, and therefore the eigenvectors are also different. The eigenvectors define the diagonalization transformations.

The system (15) now becomes

$$\begin{bmatrix} 1.0 & 0 & 0 \\ 0 & 0.96405 & 0 \\ 0 & 0 & 0.97419 \end{bmatrix} = \alpha \exp \begin{bmatrix} 0 & 0 & 0 \\ 0 & \nu_2 & 0 \\ 0 & 0 & \nu_3 \end{bmatrix} + (1-\alpha) \exp \begin{bmatrix} 0 & 0 & 0 \\ 0 & k\nu_2 & 0 \\ 0 & 0 & k\nu_3 \end{bmatrix}$$

$$\begin{bmatrix} 1.0 & 0 & 0 \\ 0 & 0.89477 & 0 \\ 0 & 0 & 0.92473 \end{bmatrix} = \alpha \exp \begin{bmatrix} 0 & 0 & 0 \\ 0 & 5\nu_2 & 0 \\ 0 & 0 & 5\nu_3 \end{bmatrix} + (1-\alpha) \exp \begin{bmatrix} 0 & 0 & 0 \\ 0 & 5k\nu_2 & 0 \\ 0 & 0 & 5k\nu_3 \end{bmatrix}$$

The equivalent of system (18), after removing the two trivial equalities, is

$$\begin{aligned}
0.96405 &= \alpha\, e^{\nu_2} + (1-\alpha)\, e^{k\nu_2} \\
0.97419 &= \alpha\, e^{\nu_3} + (1-\alpha)\, e^{k\nu_3} \\
0.89477 &= \alpha\, e^{5\nu_2} + (1-\alpha)\, e^{5k\nu_2} \\
0.92473 &= \alpha\, e^{5\nu_3} + (1-\alpha)\, e^{5k\nu_3}
\end{aligned} \tag{18a}$$

We now search for a solution for $\alpha$, $k$, $\nu_2$, and $\nu_3$. Replacing $e^{\nu_2}$ by $z_2$ and $e^{\nu_3}$ by $z_3$, system (18a) yields

$$\begin{aligned}
\hat{\alpha}_2 &= \frac{0.96405 - z_2^k}{z_2 - z_2^k} \qquad & \hat{\alpha}_3 &= \frac{0.97419 - z_3^k}{z_3 - z_3^k} \\
\hat{\alpha}_2 &= \frac{0.89477 - z_2^{5k}}{z_2^5 - z_2^{5k}} \qquad & \hat{\alpha}_3 &= \frac{0.92473 - z_3^{5k}}{z_3^5 - z_3^{5k}}
\end{aligned} \tag{19a}$$

The algorithm at the end of the previous section was then applied. The unique value $k = 0.01$ was found to give $\hat{\alpha}_2 = \hat{\alpha}_3 = \alpha$. For this $k$, $\alpha = 0.0233$, $\nu_2 = -1.6848$, and $\nu_3 = -1.0051$ ($\nu_i = \ln z_i$).

The values of $\alpha$ and $k$ imply that 2.3% of the group aged 15–19 have a "force" of migration one hundred times greater than that of the remaining 97.7% in this age group. Note that this large difference in the intensities or "force" of migration does not imply the same difference in the probabilities of migration! The probabilities of migration may be deduced from system (10) once $\boldsymbol{\pi}_1$ and $\boldsymbol{\rho}_1$ have been calculated using the relations

$$\begin{aligned}
\boldsymbol{\pi}_1 &= e^{\boldsymbol{\mu}} \\
\boldsymbol{\rho}_1 &= e^{k\boldsymbol{\mu}}
\end{aligned} \tag{20}$$

Note that if $\mathbf{P}_1$ is diagonalized by the transformation $\mathbf{T}_1$, $\boldsymbol{\pi}_1$ and $\boldsymbol{\rho}_1$ are diagonalized by the same transformation ($\boldsymbol{\pi}_1$ and $\boldsymbol{\rho}_1$ are *related*). Then eqn. (20) yields

$$\begin{aligned}
\operatorname{diag}(\boldsymbol{\pi}_1) &= \mathbf{T}_1^{-1}\boldsymbol{\pi}_1\mathbf{T}_1 = \mathbf{T}_1^{-1}\, e^{\boldsymbol{\mu}}\,\mathbf{T}_1 = e^{\mathbf{v}} \\
\operatorname{diag}(\boldsymbol{\rho}_1) &= \mathbf{T}_1^{-1}\boldsymbol{\rho}_1\mathbf{T}_1 = e^{k\mathbf{v}}
\end{aligned}$$

Then

$$\operatorname{diag}(\boldsymbol{\pi}_1) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & e^{\nu_2} & 0 \\ 0 & 0 & e^{\nu_3} \end{bmatrix}$$

$$\operatorname{diag}(\boldsymbol{\rho}_1) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & e^{k\nu_2} & 0 \\ 0 & 0 & e^{k\nu_3} \end{bmatrix}$$

$\boldsymbol{\pi}_1$ and $\boldsymbol{\rho}_1$ can be found from the last two expressions by applying the reverse transformations

$$\begin{aligned}
\boldsymbol{\pi}_1 &= \boldsymbol{T}_1 \operatorname{diag}(\boldsymbol{\pi}_1)\mathbf{T}_1^{-1} \\
\boldsymbol{\rho}_1 &= \mathbf{T}_1 \operatorname{diag}(\boldsymbol{\rho}_1)\mathbf{T}_1^{-1}
\end{aligned}$$

The estimated values for $\boldsymbol{\pi}_1$ and $\boldsymbol{\rho}_1$ are

$$\boldsymbol{\pi}_1 = \begin{bmatrix} 0.23138 & 0.38506 & 0.38360 \\ 0.04575 & 0.58615 & 0.36809 \\ 0.02863 & 0.25083 & 0.72054 \end{bmatrix}$$

$$\boldsymbol{\rho}_1 = \begin{bmatrix} 0.98373 & 0.00952 & 0.00675 \\ 0.00116 & 0.99271 & 0.00614 \\ 0.00048 & 0.00421 & 0.99531 \end{bmatrix}$$

While $\boldsymbol{\rho}_1$ has a similar structure to $\mathbf{P}_1$, this is not true of $\boldsymbol{\pi}_1$. The elements on the main diagonal of $\boldsymbol{\pi}_1$ reflect the probabilities that the high-intensity movers will remain in the same region for one year. The values are much lower than the typical values for an average population. Note that these probabilities are very dependent on the size of the regional population; this explains why the comparatively small region of East Anglia is connected with high out-migration probabilities.

The following expressions may be derived for $(\boldsymbol{\pi}_1)^5$ and $(\boldsymbol{\rho}_1)^5$:

$$(\boldsymbol{\pi}_1)^5 = \mathbf{T}_1 \operatorname{diag}[(\boldsymbol{\pi}_1)^5]\mathbf{T}_1^{-1}$$

$$(\boldsymbol{\rho}_1)^5 = \mathbf{T}_1 \operatorname{diag}[(\boldsymbol{\rho}_1)^5]\mathbf{T}_1^{-1}$$

where

$$\operatorname{diag}[(\boldsymbol{\pi}_1)^5] = \begin{bmatrix} 1 & 0 & 0 \\ 0 & e^{5\nu_2} & 0 \\ 0 & 0 & e^{5\nu_3} \end{bmatrix}$$

$$\operatorname{diag}[(\boldsymbol{\rho}_1)^5] = \begin{bmatrix} 1 & 0 & 0 \\ 0 & e^{5k\nu_2} & 0 \\ 0 & 0 & e^{5k\nu_3} \end{bmatrix}$$

Using the expression

$$\mathbf{P}_5(x) = \alpha[\boldsymbol{\pi}_1(x)]^5 + (1-\alpha)[\boldsymbol{\rho}_1(x)]^5$$

where $x = 15$, we obtain the final numerical estimate of $\mathbf{P}_5(15)$

$$\mathbf{P}_5(15) = \begin{bmatrix} 0.90092 & 0.05367 & 0.04541 \\ 0.00645 & 0.95098 & 0.04257 \\ 0.00332 & 0.02909 & 0.96760 \end{bmatrix}$$

$$= 0.0233 \begin{bmatrix} 0.04468 & 0.38844 & 0.56692 \\ 0.04450 & 0.38908 & 0.56640 \\ 0.04397 & 0.38425 & 0.57179 \end{bmatrix}$$

$$+ 0.9767 \begin{bmatrix} 0.92142 & 0.04566 & 0.03292 \\ 0.00554 & 0.96443 & 0.03003 \\ 0.00234 & 0.02058 & 0.97707 \end{bmatrix}$$

Using the expression

$$\mathbf{P}_1(x) = \alpha\boldsymbol{\pi}_1(x) + (1-\alpha)\boldsymbol{\rho}_1(x)$$

where $x = 15$, we obtain the final numerical estimate of $\mathbf{P}_1(15)$

$$\mathbf{P}_1(15) = \begin{bmatrix} 0.96614 & 0.01830 & 0.01556 \\ 0.00220 & 0.98320 & 0.01460 \\ 0.00114 & 0.00997 & 0.98889 \end{bmatrix}$$

$$= 0.0233 \begin{bmatrix} 0.23138 & 0.38506 & 0.38360 \\ 0.04575 & 0.58615 & 0.36809 \\ 0.02863 & 0.25083 & 0.72054 \end{bmatrix}$$

$$+ 0.9767 \begin{bmatrix} 0.98373 & 0.00952 & 0.00675 \\ 0.00116 & 0.99271 & 0.00614 \\ 0.00048 & 0.00421 & 0.99531 \end{bmatrix}$$

Note that the estimated matrix $\mathbf{P}_5(15)$ is very close to the observed matrix $\mathbf{P}_5(15)$ given at the beginning of Section 3, while the estimated and observed matrices $\mathbf{P}_1(15)$ are exactly the same.

The matrix $[\boldsymbol{\pi}_1(15)]^5$ in the numerical expression for $\mathbf{P}_5(15)$ above is of particular interest because each column contains three numbers which are approximately equal. This is a consequence of the fact that $\boldsymbol{\pi}_1$ refers to the group with an intensity of migration approximately one hundred times greater than that of the other group. Since $[\boldsymbol{\pi}_1]^\tau = e^{\mu\tau}$ and $[\boldsymbol{\rho}_1]^\tau = e^{k\mu\tau}$, both processes tend to the same asymptote, but the first approaches it much more quickly. This is illustrated in Figure 2, where $[\mathbf{a}]_{ij}$ denotes an element from the $i$th row and $j$th column of a matrix $\mathbf{a}$.

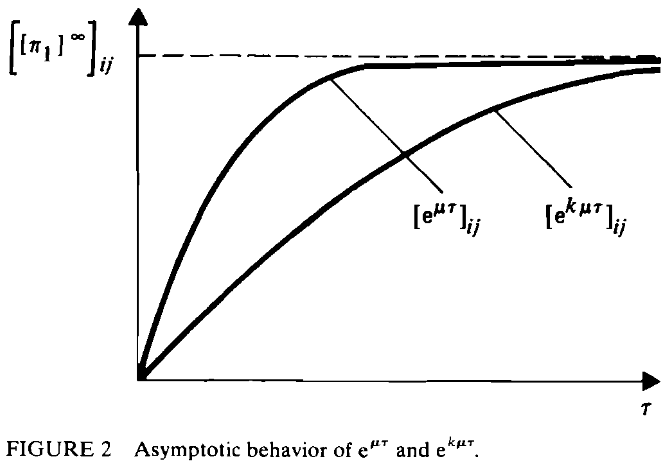



FIGURE 2 Asymptotic behavior of $e^{\mu\tau}$ and $e^{k\mu\tau}$.

$[\boldsymbol{\pi}_1]^5$ is seen to be very close to the asymptotic distribution described by $[\boldsymbol{\pi}_1]^\infty$. But $[\boldsymbol{\pi}_1]^\infty$ defines the stable state of the high-intensity movers, and therefore, even if this part of the population is not stable in the initial period of time, it should reach spatial stability over a period of 5–10 years. Since real demographic processes are quite homogeneous over such a short period of time, it is reasonable to suppose that the spatial distribution of the high-intensity movers is approximately stable at the initial point of time.

Because the matrices $\mathbf{P}_1$, $\boldsymbol{\pi}_1$ and $\boldsymbol{\rho}_1$ are related, $\mathbf{P}_1^\infty = \boldsymbol{\pi}_1^\infty = \boldsymbol{\rho}_1^\infty$. This proves that the process described by the high- and low-intensity movers model retains the important demographic properties of stabilization and ergodicity, although the model is not Markovian.

Only one age group (15–19 years) has been considered up to now. We therefore decided to repeat the procedure for the other fourteen age groups, solving system (18) with respect to $\alpha$, $k$, $\nu_2$, and $\nu_3$ using the algorithm described earlier in this paper. The method of solution failed twice, for the age groups 50–54 and 70–74, although the solutions obtained for ages greater than 50 were generally not satisfactory. The results are shown in Table 4.

It is believed that this procedure gives bad results for the older population primarily because of the method of solution. When trying to solve system (18) for ages greater than 50, it was observed that $\alpha$ and $k$ tended to zero. However, as $k \to 0$ the high- and low-intensity movers model tends to the mover–stayer model, and $\alpha \to 0$ reduces it still further to a Markovian process. It is therefore possible that the more sophisticated estimation procedures employed in the high- and low-intensity movers model are more inaccurate than those used in the simpler models when the migration movements are very low. This could explain to some extent the differences between the solutions for the age groups 45–49 and under and 55–59 and over.

Consideration of the values of $\alpha$, $k$, $\nu_2$, and $\nu_3$ for the first ten age groups in Table 4 leads to the following conclusions:

1. The values of $k$ are quite similar, the mean being 0.01202.
2. The values of $\alpha$ generate a curve which resembles a migration curve. [Different migration schedules for Great Britain are given in Rees (1979b).]
3. The absolute values of each of the $\nu_i$ also generate a curve resembling a migration curve, although the resemblance is not as close as for the curve generated by $\alpha$.

These features can be used in the implementation of the model, which is discussed in the next section.

TABLE 4 Values of $\alpha$, $k$, $\nu_2$, and $\nu_3$ for different age groups.

| Age group | $\alpha$ | $k$ | $\nu_2$ | $\nu_3$ |
|---|---|---|---|---|
| 0–4 | 0.03156 | 0.01279 | −1.35706 | −0.62757 |
| 5–9 | 0.02014 | 0.01280 | −1.11350 | −0.59498 |
| 10–14 | 0.01519 | 0.01311 | −0.92648 | −0.54148 |
| 15–19 | 0.02338 | 0.01045 | −1.67540 | −1.04744 |
| 20–24 | 0.04147 | 0.00787 | −2.72179 | −1.62251 |
| 25–29 | 0.04166 | 0.01436 | −1.35777 | −0.78607 |
| 30–34 | 0.02259 | 0.01248 | −1.33867 | −0.82405 |
| 35–39 | 0.02244 | 0.01286 | −0.91052 | −0.53732 |
| 40–44 | 0.01020 | 0.01000 | −1.02088 | −0.66838 |
| 45–49 | 0.01601 | 0.01350 | −0.51931 | −0.34236 |
| 50–54[a] | — | — | — | — |
| 55–59 | 0.00288 | 0.001830 | −2.05199 | −3.57241 |
| 60–64 | 0.00336 | 0.002410 | −1.53867 | −3.01742 |
| 65–69 | 0.00972 | 0.007235 | −0.44394 | −0.78785 |
| 70–74[a] | — | — | — | — |

[a] Solution not found.

## 4 IMPLEMENTATION OF THE MODEL

The previous two sections described the mathematical and numerical aspects of the high- and low-intensity movers model. The numerical results justify the assumptions made, and therefore verify the model itself. However, the numerical results were derived from two sets of data—one-year and five-year observations—both disaggregated by age.

In the general case, we must assume that only one set of data is available, and then use it to obtain approximations for the other set. Since one-year data are usually available in most countries, we will assume these to be given. Before considering the numerical results any further, however, the theoretical background must be developed.

In Section 2 it is shown that starting with the matrix equation

$$\mathbf{P}_1(x) = \alpha(x)\boldsymbol{\pi}_1(x) + [1-\alpha(x)]\boldsymbol{\rho}_1(x)$$

it is possible to construct the system of scalar equations

$$\begin{aligned} \lambda_2(\mathbf{P}_1) &= \alpha\ e^{\nu_2} + (1-\alpha)\ e^{k\nu_2} \\ \lambda_3(\mathbf{P}_1) &= \alpha\ e^{\nu_3} + (1-\alpha)\ e^{k\nu_3} \end{aligned} \tag{21}$$

omitting the dependence on age $x$ for clarity. System (21) contains two equations and four unknowns, $\alpha$, $k$, $\nu_2$, and $\nu_3$; two of the unknowns must therefore be specified exogenously. This is in fact the basis for the implementation of the model.

Recalling the conclusions drawn from Table 4 at the end of the previous section, it seems reasonable to search for values of $\alpha$ and $k$ which might be applicable to the total population aggregated by age ($\alpha_{tot}$ and $k_{tot}$, respectively). Then two approaches are possible: keep these values constant for all ages, or disaggregate them in accordance with the results from Table 4 [i.e., $k_{tot}$ may be kept constant, and $\alpha_{tot}$ may be used to generate a set $\alpha(x)$ for all $x$, such that the $\alpha(x)$ form a curve similar to that of the observed migration rates, and the arithmetic mean of $\alpha(x)$ is equal to $\alpha_{tot}$].

In either case, it is only necessary to obtain values for $\alpha_{tot}$ and $k_{tot}$. The derivation of these values will be discussed later in this section, but for the moment let us suppose they are available. In this case, $k_{tot}$ and $\alpha_{tot}$, or $\alpha(x)$, can be used to solve system (21) for $\nu_2(x)$ and $\nu_3(x)$. System (22) can then be solved with respect to the unknowns $\lambda_2(\mathbf{P}_5)$ and $\lambda_3(\mathbf{P}_5)$

$$\begin{aligned} \lambda_2(\mathbf{P}_5) &= \alpha\ e^{5\nu_2} + (1-\alpha)\ e^{5k\nu_2} \\ \lambda_3(\mathbf{P}_5) &= \alpha\ e^{5\nu_3} + (1-\alpha)\ e^{5k\nu_3} \end{aligned} \tag{22}$$

where the dependence on age $x$ is again omitted. The diagonalized matrix $\boldsymbol{\Lambda}_5 = \text{diag}(\mathbf{P}_5)$ therefore becomes available since it is already known that $\lambda_1(\mathbf{P}_5) = 1$. In order to find $\mathbf{P}_5$ it is necessary to know its diagonalizing transformation. But the discussion here suggests that $\mathbf{P}_5$ is a function of $\mathbf{P}_1$, i.e., $\mathbf{P}_5 = f(\mathbf{P}_1)$, where the function $f(\cdot)$ may be deduced from system (10). Therefore, $\mathbf{T}_1$ must diagonalize $\mathbf{P}_5$ and hence

$$\mathbf{P}_5 = \mathbf{T}_1\boldsymbol{\Lambda}_5\mathbf{T}_1^{-1} \tag{23}$$

Note that eqn. (23) implies $\mathbf{T}_1 = \mathbf{T}_5$. This equality was discussed on page 10, and it was concluded that it should be approximately true (Appendix B). This then implies that eqn. (23) is also an approximation. According to the structure of the model, this approximation should yield better results than those discussed in the first section.

It is still unclear how values for $\alpha$ and $k$ may be obtained, even for the total population. One possible method is to look at sociological studies: $\alpha$ can be deduced from information on which section of the population migrates more frequently, and $k$ can be estimated from discussions of the difference in migration frequency between the two groups. (It should be borne in mind that $k$ indicates differences in the intensity, and not the probability, of migration.)

However, there is another, more preferable, way of deriving $\alpha$ and $k$. Many countries hold censuses or enquiries every five or ten years, and these yield data on interregional migration flows aggregated by age (the migration-flow matrix). Since the mid-period multiregional population data are usually available, it is possible to estimate a matrix of origin–destination migration rates for the total population, aggregated by age. Let this matrix be $\mathbf{M}_5(\text{tot})$. The numerical form of $\mathbf{M}_5(\text{tot})$ for Great Britain was estimated to be

$$\mathbf{M}_5(\text{tot}) = \begin{bmatrix} 0.92659 & 0.03618 & 0.03724 \\ 0.00694 & 0.95628 & 0.03678 \\ 0.00267 & 0.01750 & 0.97982 \end{bmatrix} \tag{24a}$$

The corresponding matrix for a one-year period is

$$\mathbf{M}_1(\text{tot}) = \begin{bmatrix} 0.97494 & 0.01290 & 0.01217 \\ 0.00214 & 0.98606 & 0.01180 \\ 0.00075 & 0.00581 & 0.99344 \end{bmatrix} \tag{24b}$$

Note that these matrices have the same structure as those given at the beginning of Section 3. Their eigenvalues are: $\lambda_1(\mathbf{M}_5) = 1$; $\lambda_2(\mathbf{M}_5) = 0.91973$; $\lambda_3(\mathbf{M}_5) = 0.94296$; $\lambda_1(\mathbf{M}_1) = 1$; $\lambda_2(\mathbf{M}_1) = 0.97286$; $\lambda_3(\mathbf{M}_1) = 0.98159$. Applying the procedures described in Section 3, the unknown parameters are found to have the values

$$\begin{aligned} \alpha_{\text{tot}} &= 0.02198, & k_{\text{tot}} &= 0.01049, \\ \nu_2(\text{tot}) &= -1.1735 & \nu_3(\text{tot}) &= -0.7092 \end{aligned} \tag{25}$$

These values will be used to derive the age-specific migration-rate matrices, $\mathbf{M}_5(x)$. This can be done in two different ways. First, the parameters $\alpha$ and $k$ are kept constant at the values given in eqns. (25) for all $x$. Consider the case when $x = 15$. New values for $\nu_2$ and $\nu_3$ may be estimated from system (21). In a similar way, values for $\lambda_2\,[\mathbf{M}_5(15)]$ and $\lambda_3\,[\mathbf{M}_5(15)]$ (0.89003 and 0.92254, respectively) are calculated using system (22). The diagonalized matrix $\Lambda_5(15) = \text{diag}\,[\mathbf{M}_5(15)]$ then becomes available, since $\lambda_1[\mathbf{M}_5(15)] = 1$. Finally, the transformation $\mathbf{T}_1(15)$, which diagonalizes $\mathbf{M}_1(15)$, may be used to obtain $\mathbf{M}_5(15)$

$$\mathbf{M}_5(15) = \mathbf{T}_1(15)\Lambda_5(15)\mathbf{T}_1^{-1}(15) = \begin{bmatrix} 0.89647 & 0.05640 & 0.04714 \\ 0.00679 & 0.94911 & 0.04411 \\ 0.00343 & 0.03015 & 0.96642 \end{bmatrix} \tag{26a}$$

The second way of deriving the matrices $\mathbf{M}_5(x)$ for each $x$ is to keep $k$ constant at $k_{\text{tot}}$ once again, but to use $\alpha_{\text{tot}}$ and the observed migration schedules to yield values

$\alpha(x)$ for each $x$. Suppose that the migration schedule is given by the age-specific rates $m_1(x)$, which can be estimated at the national level. Let $n$ be the number of age groups. Then, from the expressions for the means

$$\sum_x [m_1(x)]/n = m_1, \qquad \sum_x [\alpha(x)]/n = \alpha_{tot}$$

we obtain

$$\alpha(x) = \alpha_{tot} m_1(x)/m_1$$

For $x = 15$, $\alpha(15)$ was estimated to be 0.03404. This value of $\alpha$ and $k_{tot}$ from eqns. (25) were used to derive the matrix

$$\mathbf{M}_5(15) = \begin{bmatrix} 0.91063 & 0.04545 & 0.04294 \\ 0.00555 & 0.95367 & 0.04078 \\ 0.00317 & 0.02783 & 0.96900 \end{bmatrix} \qquad (26b)$$

Each of the matrices in eqns. (26a) or (26b) can be rearranged as on p. 1, and then substituted into eqn. (2), which yields the desired matrix $\mathbf{P}_5(15)$. The results obtained are given in Table 5.

TABLE 5 Approximate probabilities of a person at exact age 15 in one of three regions of Great Britain living in the same or another region five years later. Calculated using eqns. (26a) and (26b).

| | Probability of living in region | | | |
|---|---|---|---|---|
| Region of origin | East Anglia | South East | Rest | Probability of death |
| *Calculated using eqn. (26a)* | | | | |
| East Anglia | 0.898531 | 0.052791 | 0.045082 | 0.003595 |
| South East | 0.006347 | 0.948149 | 0.042336 | 0.003168 |
| Rest of Britain | 0.003291 | 0.028926 | 0.964532 | 0.003251 |
| *Calculated using eqn. (26b)* | | | | |
| East Anglia | 0.911296 | 0.043880 | 0.041226 | 0.003598 |
| South East | 0.005237 | 0.952348 | 0.039248 | 0.003167 |
| Rest of Britain | 0.003050 | 0.026778 | 0.966927 | 0.003251 |

Both methods yield estimated probabilities very close to the probabilities calculated using eqn. (2) and shown in Table 1, and produce much better results than eqn. (1) (also shown in Table 1). It is worth noting that $\alpha_{tot}$ gives better results than $\alpha(x)$ even though the numerical values of the $\alpha(x)$ are substantially different. This shows that the high- and low-intensity movers model is relatively insensitive to the values of its parameters.

Table 6 gives the expectations of life at age 15 estimated using eqns. (26a) and (26b) as described above.

Again, in both cases, the results are very close to the values calculated using eqn. (2) given in Table 2, and $\alpha_{tot}$ yields better results than $\alpha(x)$.

These numerical results have been calculated using data for age 15, but the general conclusions are also valid for all other ages. For convenience to the reader,

TABLE 6 Distribution of expectations of life at exact age 15 in three regions of Great Britain. Calculated using eqns. (26a) and (26b).

| | Number of years spent in region | | | |
|---|---|---|---|---|
| Region of origin | East Anglia | South East | Rest | Total |
| *Calculated using eqn. (26a)* | | | | |
| East Anglia | 27.69 | 14.69 | 17.25 | 59.63 |
| South East | 2.33 | 40.65 | 16.47 | 59.45 |
| Rest of Britain | 1.17 | 8.52 | 49.07 | 58.76 |
| *Calculated using eqn. (26b)* | | | | |
| East Anglia | 30.40 | 13.03 | 16.25 | 59.68 |
| South East | 2.11 | 41.91 | 15.45 | 59.47 |
| Rest of Britain | 1.11 | 7.86 | 49.77 | 58.74 |

the complete set of expectations of life is given in Appendix D, together with the levels of migration. The latter are the regional distributions of life expectancy at age 0, and represent a measure of the accuracy of the approximations made in the various methods (see the introductory remarks to Appendix D).

We conclude that the model suggested here provides a reasonable approximation to the problem considered. A number of assumptions were made in order to find a solution, but it has been shown that these assumptions are justified. The assumption that certain variables, $\alpha$ and $k$, are independent of the regions of origin or destination may be used to show that differences in the population arising from the interpretation of $\alpha$ and $k$ do not depend on regional factors.

The fact that the transformations $\mathbf{T}_1$ and $\mathbf{T}_5$ are approximately equal may be interpreted as a preserved ranking in the attraction of the regions for migrants. That is, the magnitude of the migration flows between various regions may be different in different periods of time, but their relative proportions will remain the same.

Finally, the fact that $\alpha$ and $k$ are almost independent of the age groups was unexpected, but it has its demographic or social interpretation: the differences between the age-specific migration curves of "chronic" migrants and those of "all" migrants are insignificant when considering the one-year/five-year migration problem.

## ACKNOWLEDGMENTS

The authors would like to thank Professor Andrei Rogers, Dr. Ahmed Siefelnasr, and other participants of an HSS Area Seminar where this paper was presented and where helpful discussions took place. They are also grateful to Philip Rees for providing the data for the three-region analysis of Great Britain, and to both Philip Rees and Robert Gibberd for a careful review of the manuscript.

## REFERENCES

Bartholomew, D. (1973) Stochastic Models for Social Processes. Chichester: Wiley.
Bellman, R. (1960) Introduction to Matrix Analysis. New York, Toronto, London: McGraw-Hill.

Blumen, I., F. Kogan, and P. McCarthy (1955) The Industrial Mobility of Labor as a Probability Process. Cornell Studies of Industry and Labor Relations, 4. Ithaca, N.Y.
Boudon, R. (1975) A model for the analysis of mobility tables. In Quantitative Sociology, edited by H. Blalock. New York, San Francisco, London: Academic Press.
Chiang, C. (1968) Introduction to Stochastic Processes in Biostatistics. New York: Wiley.
Gantmacher, F. R. (1959) The Theory of Matrices. New York: Chelsea Publishing Company.
Goodman, L. (1961) Statistical methods for the mover–stayer model. Journal of the American Statistical Association 56:841–868.
Karlin, S. (1969) A First Course in Stochastic Processes. New York and London: Academic Press.
Ledent, J. (1978) Some Methodological and Empirical Considerations in the Construction of Increment–Decrement Life Tables. RM-78-25. Laxenburg, Austria: International Institute for Applied Systems Analysis.
Long, L., and K. Hansen (1975) Trends in return migration to the south. Demography 12(4):601–614.
Rees, P. H. (1977) The measurement of migration, from census data and other sources. Environment and Planning A 9:247–272.
Rees, P. H. (1978) Problems of Multiregional Analysis: Data Collection and Demographic Accounting. Working Paper 221. Leeds: University of Leeds, School of Geography.
Rees, P. H. (1979a) Migration and Settlement: 1. United Kingdom. RR-79-3. Laxenburg, Austria: International Institute for Applied Systems Analysis.
Rees, P. H. (1979b) Migration and Settlement in the United Kingdom. Unpublished paper. Leeds: University of Leeds, School of Geography.
Rogers, A. (1965) A Markovian policy of interregional migration. Papers, Regional Science Association 17:205–224. From the Philadelphia Meeting, November 1965.
Rogers, A. (1975) Introduction to Multiregional Mathematical Demography. New York: Wiley.
Rogers, A., and J. Ledent (1976) Increment–decrement life tables: a comment. Demography 13(2):287–290.
Singer, B., and S. Spilerman (1976) The representation of social processes by Markov models. American Journal of Sociology 82:1–54.
Spilerman, S. (1972) Extensions of the mover–stayer model. American Journal of Sociology 78(3).
Suleimanova, K. R. (1949) Stochastic matrices with real characteristic values. Dokl. Akad. SSSR 66:343–345.
Willekens, F., and A. Rogers (1978) Spatial Population Analysis: Methods and Computer Programs. RR-78-18. Laxenburg, Austria: International Institute for Applied Systems Analysis.

## APPENDIX A

In the text it was shown that the empirical transition matrices $\mathbf{P}_1$ and $\mathbf{P}_5$ can be diagonalized by approximately equal matrices $\mathbf{T}_1$ and $\mathbf{T}_5$, such that

$$\mathbf{P}_1 \approx \mathbf{T}_5^{-1}(\mathbf{T}_1\mathbf{P}_1\mathbf{T}_1^{-1})\mathbf{T}_5 \tag{A1}$$

$$\mathbf{P}_5 \approx \mathbf{T}_1^{-1}(\mathbf{T}_5\mathbf{P}_5\mathbf{T}_5^{-1})\mathbf{T}_1 \tag{A2}$$

This empirical fact led to the conclusion that the $n(n-1)$-dimensional problem of estimating the five-year transition matrix from the one-year matrix (or vice versa) can be reduced to the $(n-1)$-dimensional problem of estimating the eigenvalues $\lambda_i(\mathbf{P}_5)$ [or $\lambda_i(\mathbf{P}_1)$], $i = 2, 3, \ldots, n$; $\lambda_1 = 1$. Further, we will consider only the case when all the $\lambda_i$ are real and positive. For simplicity let $n = 3$. This case is presented graphically in Figure A1.

If the matrices $\mathbf{P}_1$ and $\mathbf{P}_5$ are known, it is then necessary to describe the empirical points $[1, \lambda_2(1), \lambda_2(5)]$ and $[1, \lambda_3(1), \lambda_3(5)]$ as functions of time.

In this paper we suggested making use of the approximating function

$$\lambda_i(\tau) = \alpha\, e^{\tau\nu_i} + (1-\alpha)\, e^{\tau k\nu_i} \tag{A3}$$

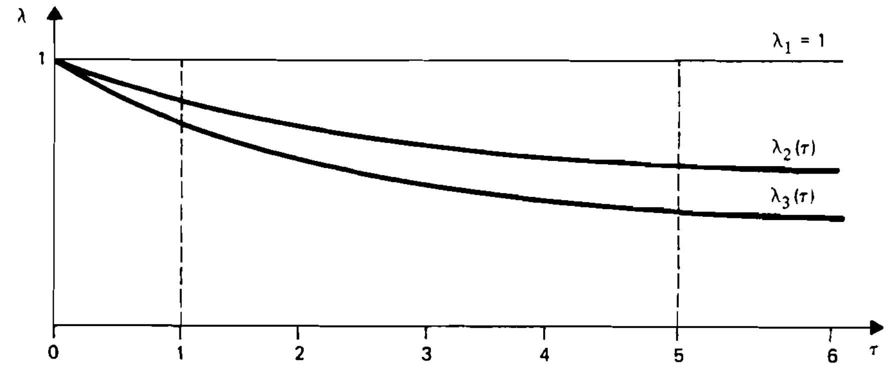



FIGURE A1 Dependence of eigenvalues λ of a transition matrix on time τ.

where $\alpha$ and $k$ are known (or can be found from aggregate data, in which case they will also be approximated).

Decreasing the dimension of the problem from $n(n-1)$ to $(n-1)$ inevitably presents additional theoretical difficulties. In this case, the problems are

1. Is it always possible to solve eqn. (A3) for $\tau = 1$ if $\alpha$ and $k$ are given?
2. Are 1, $e^{\nu_2}$, and $e^{\nu_3}$ eigenvalues of any stochastic matrix?
3. Are 1, $e^{\nu_2}$, and $e^{\nu_3}$ eigenvalues of any continuous-time Markovian transition matrix?
4. Are 1, $e^{\nu_2}$, and $e^{\nu_3}$ eigenvalues of any stochastic matrix which can be diagonalized by a given transformation $\mathbf{T}_1$?

The answers to these questions are given below.

1. *Equation (A3) has a unique non-negative solution.* It is easy to see that the function

$$f(\nu) = \alpha\, e^{\nu} + (1-\alpha)\, e^{k\nu}$$

is monotonically decreasing, $f(0) = 1$, $\lim_{\nu \to -\infty} f(\nu) = 0$, and, hence, for $0 < \lambda \leq 1$, the equation $f(\nu) = \lambda$ has a unique non-negative solution.

2. *Theorem.* (Suleimanova 1949). The set of $n+1$ real numbers $\{1, \lambda_2, \lambda_3, \ldots, \lambda_n\}$, where $|\lambda_i| < 1$ for $i = 2, 3, \ldots, n$ is a set of eigenvalues of a positive stochastic matrix provided that the sum of the modulus of the negative numbers of the set is less than unity.
3. The problem of representing some stochastic matrix as a continuous-time Markovian transition matrix (embedding problem) can be avoided by considering an integer $1/k$ and discrete time. The necessary conditions for such embedding can be found in Singer and Spilerman (1976).
4. If the transformation $\mathbf{T}_1$ of the matrix $\mathbf{P}_1$ is such that $\lambda_1$ is equal to 1, then it is easy to show that the matrix

$$\boldsymbol{\pi} = \mathbf{T}_1^{-1} \begin{pmatrix} 1 & & & & \\ & e^{\nu_2} & & & \\ & & e^{\nu_3} & & \\ & & & \ddots & \\ & & & & e^{\nu_n} \end{pmatrix} \mathbf{T}_1 \qquad \text{(A4)}$$

has the property $\sum_{i=1}^{n} \pi_{ij} = 1$. This is so because the eigenvector corresponding to $\lambda_1$ has components always equal to $(1, 1, 1, \ldots, 1)$. It is necessary only to check whether $\boldsymbol{\pi} > \mathbf{0}$.

Empirical results show that in our case $\boldsymbol{\pi}$ is always positive and hence stochastic. In the general case it is necessary to prove that the transformation

$$\rho^{-1}(\mathbf{P})$$

where

$$\rho(x) = \alpha x + (1-\alpha)x^k$$

leaves the matrix $\boldsymbol{\pi}$ positive, where

$$\boldsymbol{\pi} = \rho^{-1}(\mathbf{P})$$

and this problem is still unresolved.

## APPENDIX B

This Appendix contains data which verify the assumption that the transition matrices $\mathbf{P}_1(x)$ and $\mathbf{P}_5(x)$ can be diagonalized by the same transformation matrix $\mathbf{T}(x)$ for each age group. The following matrices are compared:

$\mathbf{P}_5$ (five-year observed migration probabilities)

$\hat{\mathbf{P}}_5 = \mathbf{T}_1 \mathbf{\Lambda}_5 \mathbf{T}_1^{-1}$ (five-year estimated migration probabilities)

$\mathbf{P}_1$ (one-year observed migration probabilities)

$\hat{\mathbf{P}}_1 = \mathbf{T}_5 \mathbf{\Lambda}_1 \mathbf{T}_5^{-1}$ (one-year estimated migration probabilities)

Migration probabilities calculated using the "Markovian" approximation

$$\mathbf{P}_5 = (\mathbf{P}_1)^5$$

$$\mathbf{P}_1 = (\mathbf{P}_5)^{1/5}$$

are also given (fifth degree and fifth root).

age group 1 (0–4 years)

| five-year obs. | | | five-year est. | | | fifth degree | | |
|---|---|---|---|---|---|---|---|---|
| 0.88975 | 0.05502 | 0.05523 | 0.88912 | 0.05138 | 0.05950 | 0.82383 | 0.08312 | 0.09305 |
| 0.00893 | 0.94485 | 0.04623 | 0.00916 | 0.94525 | 0.04559 | 0.01490 | 0.91527 | 0.06983 |
| 0.00447 | 0.01899 | 0.97654 | 0.00314 | 0.02010 | 0.97677 | 0.00484 | 0.03087 | 0.96430 |

| one-year obs. | | | one-year est. | | | fifth root | | |
|---|---|---|---|---|---|---|---|---|
| 0.96181 | 0.01835 | 0.01984 | 0.96203 | 0.01975 | 0.01823 | 0.97671 | 0.01091 | 0.01238 |
| 0.00331 | 0.98213 | 0.01456 | 0.00318 | 0.98200 | 0.01481 | 0.00195 | 0.98868 | 0.00937 |
| 0.00102 | 0.00646 | 0.99253 | 0.00149 | 0.00606 | 0.99244 | 0.00065 | 0.00414 | 0.99522 |

age group 2 (5–9 years)

| five-year obs. | | | five-year est. | | | fifth degree | | |
|---|---|---|---|---|---|---|---|---|
| 0.91464 | 0.03916 | 0.04620 | 0.91401 | 0.04079 | 0.04520 | 0.87660 | 0.05936 | 0.06404 |
| 0.00720 | 0.95475 | 0.03805 | 0.00639 | 0.95703 | 0.03658 | 0.00932 | 0.93954 | 0.05114 |
| 0.00259 | 0.01564 | 0.98178 | 0.00277 | 0.01710 | 0.98013 | 0.00391 | 0.02394 | 0.97216 |

| one-year obs. | | | one-year est. | | | fifth root | | |
|---|---|---|---|---|---|---|---|---|
| 0.97393 | 0.01269 | 0.01338 | 0.97413 | 0.01222 | 0.01365 | 0.98215 | 0.00854 | 0.00932 |
| 0.00200 | 0.98745 | 0.01056 | 0.00226 | 0.98676 | 0.01098 | 0.00134 | 0.99118 | 0.00748 |
| 0.00081 | 0.00495 | 0.99424 | 0.00075 | 0.00453 | 0.99472 | 0.00057 | 0.00350 | 0.99593 |

age group 3 (10–14 years)

| five-year obs. | | | five-year est. | | | fifth degree | | |
|---|---|---|---|---|---|---|---|---|
| 0.92914 | 0.03458 | 0.03628 | 0.92893 | 0.03424 | 0.03683 | 0.90449 | 0.04674 | 0.04877 |
| 0.00570 | 0.96346 | 0.03084 | 0.00540 | 0.96311 | 0.03149 | 0.00740 | 0.95134 | 0.04127 |
| 0.00227 | 0.01593 | 0.98181 | 0.00233 | 0.01530 | 0.98237 | 0.00307 | 0.02007 | 0.97686 |

| one-year obs. | | | one-year est. | | | fifth root | | |
|---|---|---|---|---|---|---|---|---|
| 0.98008 | 0.00984 | 0.01008 | 0.98014 | 0.00993 | 0.00993 | 0.98534 | 0.00711 | 0.00755 |
| 0.00156 | 0.98997 | 0.00847 | 0.00164 | 0.99006 | 0.00830 | 0.00112 | 0.99245 | 0.00642 |
| 0.00063 | 0.00412 | 0.99525 | 0.00061 | 0.00429 | 0.99509 | 0.00048 | 0.00312 | 0.99640 |

age group 4 (15–19 years)

| five-year obs. | | | five-year est. | | | fifth degree | | |
|---|---|---|---|---|---|---|---|---|
| 0.89593 | 0.05704 | 0.04704 | 0.89551 | 0.05717 | 0.04733 | 0.83971 | 0.08572 | 0.07457 |
| 0.00755 | 0.94978 | 0.04267 | 0.00688 | 0.94890 | 0.04421 | 0.01030 | 0.91949 | 0.07021 |
| 0.00318 | 0.03178 | 0.96504 | 0.00344 | 0.03023 | 0.96634 | 0.00548 | 0.04795 | 0.94657 |

| one-year obs. | | | one-year est. | | | fifth root | | |
|---|---|---|---|---|---|---|---|---|
| 0.96553 | 0.01869 | 0.01578 | 0.96567 | 0.01866 | 0.01566 | 0.97812 | 0.01208 | 0.00981 |
| 0.00225 | 0.98296 | 0.01479 | 0.00246 | 0.98326 | 0.01427 | 0.00146 | 0.98941 | 0.00913 |
| 0.00115 | 0.01010 | 0.98874 | 0.00107 | 0.01062 | 0.98831 | 0.00071 | 0.00625 | 0.99305 |

age group 5 (20–24 years)

| five-year obs. | | | five-year est. | | | fifth degree | | |
|---|---|---|---|---|---|---|---|---|
| 0.86049 | 0.07308 | 0.06643 | 0.86276 | 0.07901 | 0.05823 | 0.74953 | 0.13111 | 0.11936 |
| 0.01060 | 0.92992 | 0.05948 | 0.01089 | 0.92545 | 0.06366 | 0.01763 | 0.85576 | 0.12662 |
| 0.00410 | 0.04145 | 0.95445 | 0.00345 | 0.03990 | 0.95665 | 0.00768 | 0.07882 | 0.91350 |

| one-year obs. | | | one-year est. | | | fifth root | | |
|---|---|---|---|---|---|---|---|---|
| 0.94359 | 0.03039 | 0.02601 | 0.94287 | 0.02865 | 0.02849 | 0.97080 | 0.01709 | 0.01211 |
| 0.00412 | 0.96806 | 0.02782 | 0.00405 | 0.96979 | 0.02617 | 0.00236 | 0.98432 | 0.01332 |
| 0.00164 | 0.01735 | 0.98101 | 0.00188 | 0.01810 | 0.98002 | 0.00071 | 0.00836 | 0.99093 |

age group 6 (25–29 years)

| five-year obs. | | | five-year est. | | | fifth degree | | |
|---|---|---|---|---|---|---|---|---|
| 0.87004 | 0.06207 | 0.06790 | 0.86996 | 0.06733 | 0.06271 | 0.78704 | 0.11137 | 0.10159 |
| 0.01085 | 0.92755 | 0.06160 | 0.00967 | 0.92711 | 0.06322 | 0.01602 | 0.88154 | 0.10244 |
| 0.00384 | 0.03156 | 0.96460 | 0.00404 | 0.03085 | 0.96512 | 0.00654 | 0.05000 | 0.94346 |

| one-year obs. | | | one-year est. | | | fifth root | | |
|---|---|---|---|---|---|---|---|---|
| 0.95297 | 0.02530 | 0.02173 | 0.95298 | 0.02318 | 0.02384 | 0.97244 | 0.01451 | 0.01306 |
| 0.00365 | 0.97442 | 0.02193 | 0.00411 | 0.97457 | 0.02132 | 0.00209 | 0.98474 | 0.01317 |
| 0.00139 | 0.01071 | 0.98790 | 0.00130 | 0.01098 | 0.98772 | 0.00084 | 0.00643 | 0.99273 |

age group 7 (30–34 years)

| five-year obs. | | | five-year est. | | | fifth degree | | |
|---|---|---|---|---|---|---|---|---|
| 0.90142 | 0.04568 | 0.05290 | 0.90127 | 0.04661 | 0.05212 | 0.85469 | 0.06812 | 0.07719 |
| 0.00842 | 0.94319 | 0.04839 | 0.00825 | 0.94423 | 0.04751 | 0.01204 | 0.91742 | 0.07054 |
| 0.00293 | 0.02177 | 0.97530 | 0.00295 | 0.02263 | 0.97441 | 0.00440 | 0.03357 | 0.96203 |

| one-year obs. | | | one-year est. | | | fifth root | | |
|---|---|---|---|---|---|---|---|---|
| 0.96897 | 0.01478 | 0.01625 | 0.96903 | 0.01449 | 0.01648 | 0.97938 | 0.00984 | 0.01079 |
| 0.00263 | 0.98262 | 0.01476 | 0.00268 | 0.98230 | 0.01503 | 0.00175 | 0.98846 | 0.00979 |
| 0.00091 | 0.00703 | 0.99205 | 0.00090 | 0.00677 | 0.99233 | 0.00061 | 0.00467 | 0.99472 |

age group 8 (35–39 years)

| five-year obs. | | | five-year est. | | | fifth degree | | |
|---|---|---|---|---|---|---|---|---|
| 0.92492 | 0.03457 | 0.04051 | 0.92414 | 0.03547 | 0.04039 | 0.88834 | 0.05373 | 0.05793 |
| 0.00659 | 0.95708 | 0.03634 | 0.00582 | 0.95937 | 0.03481 | 0.00890 | 0.94221 | 0.04890 |
| 0.00223 | 0.01636 | 0.98141 | 0.00246 | 0.01763 | 0.97990 | 0.00347 | 0.02485 | 0.97168 |

| one-year obs. | | | one-year est. | | | fifth root | | |
|---|---|---|---|---|---|---|---|---|
| 0.97654 | 0.01141 | 0.01205 | 0.97681 | 0.01118 | 0.01200 | 0.98432 | 0.00738 | 0.00830 |
| 0.00190 | 0.98801 | 0.01009 | 0.00217 | 0.98730 | 0.01054 | 0.00121 | 0.99167 | 0.00712 |
| 0.00072 | 0.00513 | 0.99415 | 0.00063 | 0.00478 | 0.99459 | 0.00050 | 0.00361 | 0.99589 |

age group 9 (40–44 years)

| five-year obs. | | | five-year est. | | | fifth degree | | |
|---|---|---|---|---|---|---|---|---|
| 0.94373 | 0.02775 | 0.02853 | 0.94308 | 0.02762 | 0.02930 | 0.92505 | 0.03651 | 0.03844 |
| 0.00541 | 0.96619 | 0.02839 | 0.00504 | 0.96742 | 0.02753 | 0.00669 | 0.95723 | 0.03608 |
| 0.00224 | 0.01291 | 0.98485 | 0.00189 | 0.01383 | 0.98428 | 0.00249 | 0.01813 | 0.97939 |

| one-year obs. | | | one-year est. | | | fiftn root | | |
|---|---|---|---|---|---|---|---|---|
| 0.98451 | 0.00761 | 0.00788 | 0.98469 | 0.00766 | 0.00765 | 0.98833 | 0.00569 | 0.00597 |
| 0.00140 | 0.99122 | 0.00739 | 0.00149 | 0.99089 | 0.00762 | 0.00104 | 0.99335 | 0.00560 |
| 0.00051 | 0.00371 | 0.99578 | 0.00060 | 0.00347 | 0.99594 | 0.00038 | 0.00282 | 0.99680 |

age group 10 (45–49 years)

| five-year obs. | | | five-year est. | | | fifth degree | | |
|---|---|---|---|---|---|---|---|---|
| 0.95535 | 0.02340 | 0.02125 | 0.95530 | 0.02250 | 0.02222 | 0.94190 | 0.03100 | 0.02710 |
| 0.00482 | 0.97118 | 0.02401 | 0.00444 | 0.96985 | 0.02572 | 0.00614 | 0.96158 | 0.03228 |
| 0.00238 | 0.01038 | 0.98724 | 0.00138 | 0.01000 | 0.98862 | 0.00166 | 0.01257 | 0.98576 |

| one-year obs. | | | one-year est. | | | fifth root | | |
|---|---|---|---|---|---|---|---|---|
| 0.98808 | 0.00642 | 0.00550 | 0.98806 | 0.00666 | 0.00528 | 0.99089 | 0.00462 | 0.00450 |
| 0.00127 | 0.99215 | 0.00658 | 0.00136 | 0.99251 | 0.00613 | 0.00091 | 0.99387 | 0.00522 |
| 0.00034 | 0.00256 | 0.99710 | 0.00060 | 0.00264 | 0.99675 | 0.00028 | 0.00203 | 0.99769 |

age group 11 (50–54 years)

| five-year obs. | | | five-year est. | | | fifth degree | | |
|---|---|---|---|---|---|---|---|---|
| 0.96875 | 0.01678 | 0.01448 | 0.96408 | 0.01508 | 0.02084 | 0.93340 | 0.03557 | 0.03103 |
| 0.00441 | 0.97358 | 0.02201 | 0.00242 | 0.97688 | 0.02070 | 0.00586 | 0.96359 | 0.03055 |
| 0.00227 | 0.00797 | 0.98976 | 0.00110 | 0.00777 | 0.99113 | 0.00153 | 0.01158 | 0.98689 |

| one-year obs. | | | one-year est. | | | fifth root | | |
|---|---|---|---|---|---|---|---|---|
| 0.98629 | 0.00739 | 0.00632 | 0.98900 | 0.00864 | 0.00236 | 0.99271 | 0.00308 | 0.00422 |
| 0.00122 | 0.99256 | 0.00622 | 0.00207 | 0.99038 | 0.00755 | 0.00049 | 0.99532 | 0.00419 |
| 0.00031 | 0.00236 | 0.99733 | 0.00060 | 0.00258 | 0.99682 | 0.00022 | 0.00157 | 0.99821 |

age group 12 (55–59 years)

| five-year obs. | | | five-year est. | | | fifth degree | | |
|---|---|---|---|---|---|---|---|---|
| 0.97825 | 0.01158 | 0.01016 | 0.97524 | 0.01531 | 0.00945 | 0.96624 | 0.01810 | 0.01565 |
| 0.00516 | 0.97224 | 0.02260 | 0.00491 | 0.97452 | 0.02058 | 0.00574 | 0.96667 | 0.02759 |
| 0.00153 | 0.00628 | 0.99220 | 0.00066 | 0.00640 | 0.99294 | 0.00119 | 0.00854 | 0.99027 |

| one-year obs. | | | one-year est. | | | fifth root | | |
|---|---|---|---|---|---|---|---|---|
| 0.99314 | 0.00371 | 0.00314 | 0.99386 | 0.00286 | 0.00328 | 0.99499 | 0.00312 | 0.00189 |
| 0.00118 | 0.99321 | 0.00561 | 0.00129 | 0.99268 | 0.00604 | 0.00100 | 0.99484 | 0.00417 |
| 0.00024 | 0.00174 | 0.99803 | 0.00048 | 0.00168 | 0.99784 | 0.00013 | 0.00130 | 0.99857 |

age group 13 (60–64 years)

| five-year obs. | | | five-year est. | | | fifth degree | | |
|---|---|---|---|---|---|---|---|---|
| 0.97819 | 0.00968 | 0.01213 | 0.97438 | 0.01610 | 0.00952 | 0.96503 | 0.01935 | 0.01562 |
| 0.00570 | 0.96726 | 0.02704 | 0.00664 | 0.97065 | 0.02272 | 0.00793 | 0.96163 | 0.03044 |
| 0.00116 | 0.00601 | 0.99283 | 0.00080 | 0.00596 | 0.99324 | 0.00138 | 0.00796 | 0.99066 |

| one-year obs. | | | one-year est. | | | fifth root | | |
|---|---|---|---|---|---|---|---|---|
| 0.99289 | 0.00398 | 0.00313 | 0.99379 | 0.00245 | 0.00376 | 0.99481 | 0.00328 | 0.00190 |
| 0.00163 | 0.99217 | 0.00620 | 0.00143 | 0.99135 | 0.00722 | 0.00136 | 0.99404 | 0.00461 |
| 0.00027 | 0.00162 | 0.99810 | 0.00037 | 0.00160 | 0.99803 | 0.00016 | 0.00121 | 0.99863 |

age group 14 (65–69 years)

| five-year obs. | | | five-year est. | | | fifth degree | | |
|---|---|---|---|---|---|---|---|---|
| 0.97596 | 0.00906 | 0.01498 | 0.97453 | 0.01192 | 0.01355 | 0.96548 | 0.01767 | 0.01684 |
| 0.00514 | 0.96885 | 0.02601 | 0.00472 | 0.96985 | 0.02543 | 0.00702 | 0.95750 | 0.03549 |
| 0.00160 | 0.00599 | 0.99240 | 0.00114 | 0.00601 | 0.99285 | 0.00140 | 0.00839 | 0.99020 |

| one-year obs. | | | one-year est. | | | fifth root | | |
|---|---|---|---|---|---|---|---|---|
| 0.99299 | 0.00364 | 0.00338 | 0.99343 | 0.00272 | 0.00385 | 0.99485 | 0.00243 | 0.00273 |
| 0.00144 | 0.99132 | 0.00724 | 0.00155 | 0.99100 | 0.00744 | 0.00096 | 0.99388 | 0.00516 |
| 0.00028 | 0.00171 | 0.99801 | 0.00040 | 0.00172 | 0.99788 | 0.00023 | 0.00122 | 0.99855 |

age group 15 (70–74 years)

| five-year obs. | | | five-year est. | | | fifth degree | | |
|---|---|---|---|---|---|---|---|---|
| 0.97468 | 0.00927 | 0.01605 | 0.97160 | 0.01105 | 0.01735 | 0.95386 | 0.02401 | 0.02213 |
| 0.00461 | 0.97831 | 0.01708 | 0.00199 | 0.98045 | 0.01756 | 0.00440 | 0.97284 | 0.02276 |
| 0.00172 | 0.00571 | 0.99257 | 0.00093 | 0.00556 | 0.99351 | 0.00114 | 0.00725 | 0.99161 |

| one-year obs. | | | one-year est. | | | fifth root | | |
|---|---|---|---|---|---|---|---|---|
| 0.99059 | 0.00494 | 0.00448 | 0.99212 | 0.00390 | 0.00399 | 0.99425 | 0.00225 | 0.00350 |
| 0.00091 | 0.99449 | 0.00461 | 0.00214 | 0.99322 | 0.00464 | 0.00040 | 0.99605 | 0.00354 |
| 0.00023 | 0.00147 | 0.99830 | 0.00031 | 0.00166 | 0.99803 | 0.00019 | 0.00112 | 0.99869 |

## APPENDIX C

This Appendix presents the probabilities of a person at exact age $x$ in one of three regions of Great Britain (East Anglia, South East England, and the rest of Britain) living in the same or another region five years later.

The probabilities are calculated using four different methods:

1. Calculated using eqn. (2). These estimates are accepted in this paper as the correct ones.
2. Estimated using the parameters $\alpha(x)$ and $k_{tot}$, $\alpha$ being disaggregated by age such that the schedule is the observed migration schedule for Britain, and the area under the curve is equal to $\alpha_{tot}$.
3. Estimated using the parameters $\alpha_{tot}$ and $k_{tot}$, where $\alpha_{tot}$ is aggregated by age.
4. Estimated using eqn. (1).

The results given by the second and third methods are approximately equal. Both are much closer to the correct values obtained by the first method than are the results obtained using the last method.

Probabilities of death and migration within the subsequent five years for a person aged $x$ resident in East Anglia.

Calculated using eqn. (2)

| age | death | migration from e.anglia to e.anglia | s.east | r.brit |
|---|---|---|---|---|
| 0 | 0.019077 | 0.878803 | 0.050263 | 0.051856 |
| 5 | 0.001796 | 0.916665 | 0.037008 | 0.044531 |
| 10 | 0.001555 | 0.930246 | 0.033027 | 0.035172 |
| 15 | 0.003595 | 0.898068 | 0.053417 | 0.044920 |
| 20 | 0.004013 | 0.866501 | 0.067041 | 0.062445 |
| 25 | 0.003356 | 0.875430 | 0.057047 | 0.064167 |
| 30 | 0.004319 | 0.902390 | 0.042690 | 0.050601 |
| 35 | 0.006141 | 0.922117 | 0.032730 | 0.039012 |
| 40 | 0.010595 | 0.935395 | 0.026427 | 0.027583 |
| 45 | 0.018549 | 0.938719 | 0.022245 | 0.020486 |
| 50 | 0.028144 | 0.942074 | 0.015878 | 0.013903 |
| 55 | 0.048522 | 0.931112 | 0.010774 | 0.009592 |
| 60 | 0.079041 | 0.901212 | 0.008709 | 0.011038 |
| 65 | 0.129096 | 0.850309 | 0.007746 | 0.012849 |
| 70 | 1.000000 | 0.000000 | 0.000000 | 0.000000 |

Estimated using parameters $k_{tot}$ and $\alpha(x)$

| age | death | migration from e.anglia to e.anglia | s.east | r.brit |
|---|---|---|---|---|
| 0 | 0.019073 | 0.883325 | 0.045025 | 0.052577 |
| 5 | 0.001796 | 0.922943 | 0.035134 | 0.040127 |
| 10 | 0.001555 | 0.936659 | 0.029015 | 0.032771 |
| 15 | 0.003598 | 0.911296 | 0.043880 | 0.041226 |
| 20 | 0.004018 | 0.882393 | 0.056432 | 0.057157 |
| 25 | 0.003353 | 0.890947 | 0.053094 | 0.052606 |
| 30 | 0.004316 | 0.915103 | 0.036296 | 0.044285 |
| 35 | 0.006138 | 0.925940 | 0.031534 | 0.036388 |
| 40 | 0.010592 | 0.939213 | 0.023136 | 0.027059 |
| 45 | 0.018549 | 0.940257 | 0.020454 | 0.020740 |
| 50 | 0.028197 | 0.925367 | 0.023514 | 0.022921 |
| 55 | 0.048548 | 0.926070 | 0.012639 | 0.012743 |
| 60 | 0.079067 | 0.895656 | 0.013026 | 0.012251 |
| 65 | 0.129099 | 0.847444 | 0.010987 | 0.012469 |
| 70 | 1.000000 | 0.000000 | 0.000000 | 0.000000 |

Estimated using parameters $k_{tot}$ and $\alpha_{tot}$

| age | death | migration from e.anglia to e.anglia | s.east | r.brit |
|---|---|---|---|---|
| 0 | 0.019085 | 0.873151 | 0.051594 | 0.056170 |
| 5 | 0.001796 | 0.924027 | 0.034317 | 0.039859 |
| 10 | 0.001555 | 0.939917 | 0.026843 | 0.031685 |
| 15 | 0.003595 | 0.898531 | 0.052791 | 0.045082 |
| 20 | 0.004000 | 0.826082 | 0.094781 | 0.075137 |
| 25 | 0.003356 | 0.858027 | 0.076874 | 0.061743 |
| 30 | 0.004317 | 0.907994 | 0.040963 | 0.046726 |
| 35 | 0.006138 | 0.926855 | 0.030814 | 0.036193 |
| 40 | 0.010590 | 0.941873 | 0.021251 | 0.026286 |
| 45 | 0.018546 | 0.942748 | 0.018184 | 0.020521 |
| 50 | 0.028188 | 0.929261 | 0.020337 | 0.022214 |
| 55 | 0.048547 | 0.926934 | 0.011493 | 0.013026 |
| 60 | 0.079066 | 0.896416 | 0.012007 | 0.012511 |
| 65 | 0.129101 | 0.847997 | 0.010167 | 0.012736 |
| 70 | 1.000000 | 0.000000 | 0.000000 | 0.000000 |

Calculated using eqn. (1)

| age | death | migration from e.anglia to e.anglia | s.east | r.brit |
|---|---|---|---|---|
| 0 | 0.018940 | 0.810806 | 0.080174 | 0.090080 |
| 5 | 0.001843 | 0.876374 | 0.058454 | 0.063329 |
| 10 | 0.001657 | 0.903852 | 0.046153 | 0.048338 |
| 15 | 0.003591 | 0.838896 | 0.084048 | 0.073464 |
| 20 | 0.004028 | 0.751251 | 0.127714 | 0.117006 |
| 25 | 0.002720 | 0.788667 | 0.108686 | 0.099927 |
| 30 | 0.003434 | 0.853717 | 0.066789 | 0.076059 |
| 35 | 0.006096 | 0.884107 | 0.052742 | 0.057055 |
| 40 | 0.011283 | 0.915184 | 0.035784 | 0.037749 |
| 45 | 0.018095 | 0.925257 | 0.030197 | 0.026451 |
| 50 | 0.029244 | 0.906652 | 0.034231 | 0.029872 |
| 55 | 0.047840 | 0.920240 | 0.017128 | 0.014792 |
| 60 | 0.077441 | 0.890556 | 0.017734 | 0.014269 |
| 65 | 0.132126 | 0.838018 | 0.015340 | 0.014516 |
| 70 | 1.000000 | 0.000000 | 0.000000 | 0.000000 |

## APPENDIX C (*continued*)

Probabilities of death and migration within the subsequent five years for a person aged $x$ resident in South East England.

Calculated using eqn. (2)

| age | death | migration from e.anglia | s.east | s.east to r.brit |
|---|---|---|---|---|
| 0 | 0.020863 | 0.008171 | 0.927202 | 0.043764 |
| 5 | 0.001684 | 0.006789 | 0.954557 | 0.036970 |
| 10 | 0.001458 | 0.005433 | 0.963036 | 0.030073 |
| 15 | 0.003168 | 0.007041 | 0.948826 | 0.040965 |
| 20 | 0.003254 | 0.009658 | 0.930742 | 0.056346 |
| 25 | 0.003467 | 0.009916 | 0.928048 | 0.058569 |
| 30 | 0.004590 | 0.007838 | 0.941089 | 0.046483 |
| 35 | 0.006757 | 0.006217 | 0.951901 | 0.035125 |
| 40 | 0.011733 | 0.005153 | 0.955671 | 0.027443 |
| 45 | 0.019975 | 0.004582 | 0.952380 | 0.023062 |
| 50 | 0.031934 | 0.004181 | 0.942966 | 0.020920 |
| 55 | 0.053023 | 0.004801 | 0.921191 | 0.020985 |
| 60 | 0.085046 | 0.005125 | 0.885635 | 0.024193 |
| 65 | 0.133732 | 0.004395 | 0.839813 | 0.022060 |
| 70 | 1.000000 | 0.000000 | 0.000000 | 0.000000 |

Estimated using parameters $k_{tot}$ and $\alpha$ $(x)$

| age | death | migration from e.anglia | s.east | s.east to r.brit |
|---|---|---|---|---|
| 0 | 0.020860 | 0.008018 | 0.930649 | 0.040473 |
| 5 | 0.001684 | 0.005486 | 0.959888 | 0.032942 |
| 10 | 0.001458 | 0.004559 | 0.965485 | 0.028498 |
| 15 | 0.003167 | 0.005237 | 0.952348 | 0.039248 |
| 20 | 0.003254 | 0.007474 | 0.929418 | 0.059853 |
| 25 | 0.003467 | 0.007611 | 0.935941 | 0.052982 |
| 30 | 0.004590 | 0.006329 | 0.948042 | 0.041040 |
| 35 | 0.006756 | 0.005160 | 0.956572 | 0.031513 |
| 40 | 0.011733 | 0.004162 | 0.958370 | 0.025735 |
| 45 | 0.019977 | 0.004030 | 0.952138 | 0.023855 |
| 50 | 0.031938 | 0.003855 | 0.941627 | 0.022581 |
| 55 | 0.053023 | 0.003991 | 0.922484 | 0.020503 |
| 60 | 0.085028 | 0.005328 | 0.887998 | 0.021647 |
| 65 | 0.133744 | 0.004353 | 0.838682 | 0.023222 |
| 70 | 1.000000 | 0.000000 | 0.000000 | 0.000000 |

Estimated using parameters $k_{tot}$ and $\alpha_{tot}$

| age | death | migration from e.anglia | s.east | s.east to r.brit |
|---|---|---|---|---|
| 0 | 0.020860 | 0.009287 | 0.928479 | 0.041374 |
| 5 | 0.001684 | 0.005357 | 0.959990 | 0.032969 |
| 10 | 0.001458 | 0.004196 | 0.966404 | 0.027942 |
| 15 | 0.003168 | 0.006347 | 0.948149 | 0.042336 |
| 20 | 0.003263 | 0.012957 | 0.902510 | 0.081270 |
| 25 | 0.003468 | 0.011095 | 0.923058 | 0.062379 |
| 30 | 0.004590 | 0.007229 | 0.945432 | 0.042749 |
| 35 | 0.006756 | 0.005030 | 0.956674 | 0.031540 |
| 40 | 0.011732 | 0.003780 | 0.959332 | 0.025156 |
| 45 | 0.019976 | 0.003563 | 0.953380 | 0.023081 |
| 50 | 0.031938 | 0.003311 | 0.942816 | 0.021935 |
| 55 | 0.053020 | 0.003606 | 0.923694 | 0.019680 |
| 60 | 0.085024 | 0.004903 | 0.889201 | 0.020873 |
| 65 | 0.133736 | 0.004016 | 0.839950 | 0.022298 |
| 70 | 1.000000 | 0.000000 | 0.000000 | 0.000000 |

Calculated using eqn. (1)

| age | death | migration from e.anglia | s.east | s.east to r.brit |
|---|---|---|---|---|
| 0 | 0.018586 | 0.014359 | 0.899220 | 0.067835 |
| 5 | 0.001533 | 0.009175 | 0.938596 | 0.050695 |
| 10 | 0.001381 | 0.007299 | 0.950357 | 0.040962 |
| 15 | 0.003178 | 0.010098 | 0.917494 | 0.069230 |
| 20 | 0.003520 | 0.017154 | 0.855275 | 0.124051 |
| 25 | 0.003489 | 0.015630 | 0.880170 | 0.100711 |
| 30 | 0.004472 | 0.011791 | 0.914194 | 0.069542 |
| 35 | 0.006149 | 0.008730 | 0.936894 | 0.048227 |
| 40 | 0.011434 | 0.006551 | 0.946564 | 0.035451 |
| 45 | 0.019270 | 0.005977 | 0.943300 | 0.031453 |
| 50 | 0.031272 | 0.005636 | 0.933706 | 0.029386 |
| 55 | 0.052188 | 0.005428 | 0.916422 | 0.025961 |
| 60 | 0.083825 | 0.007268 | 0.881265 | 0.027643 |
| 65 | 0.133965 | 0.006090 | 0.829459 | 0.030486 |
| 70 | 1.000000 | 0.000000 | 0.000000 | 0.000000 |

Probabilities of death and migration within the subsequent five years for a person aged $x$ resident in the "rest of Britain".

Calculated using eqn. (2)

| age | death | migration from r.brit lo | | |
|---|---|---|---|---|
| | | e.anglia | s.east | r.brit |
| 0 | 0.022136 | 0.004181 | 0.017998 | 0.955684 |
| 5 | 0.001871 | 0.002506 | 0.015177 | 0.980446 |
| 10 | 0.001633 | 0.002209 | 0.015518 | 0.980641 |
| 15 | 0.003251 | 0.003073 | 0.030466 | 0.963210 |
| 20 | 0.003800 | 0.003934 | 0.039181 | 0.953085 |
| 25 | 0.003777 | 0.003682 | 0.029948 | 0.962593 |
| 30 | 0.004903 | 0.002833 | 0.020878 | 0.971386 |
| 35 | 0.007725 | 0.002162 | 0.015796 | 0.974317 |
| 40 | 0.013437 | 0.002167 | 0.012477 | 0.971920 |
| 45 | 0.023507 | 0.002290 | 0.009976 | 0.964227 |
| 50 | 0.036279 | 0.002165 | 0.007582 | 0.953974 |
| 55 | 0.061317 | 0.001438 | 0.005829 | 0.931416 |
| 60 | 0.097875 | 0.001059 | 0.005380 | 0.895686 |
| 65 | 0.152017 | 0.001377 | 0.005082 | 0.841524 |
| 70 | 1.000000 | 0.000000 | 0.000000 | 0.000000 |

Estimated using parameters $k_{tot}$ and $\alpha(x)$

| age | death | migration from r.brit to | | |
|---|---|---|---|---|
| | | e.anglia | s.easl | r.bril |
| 0 | 0.022139 | 0.002782 | 0.017830 | 0.957248 |
| 5 | 0.001871 | 0.002470 | 0.015382 | 0.980277 |
| 10 | 0.001633 | 0.002095 | 0.013823 | 0.982449 |
| 15 | 0.003251 | 0.003050 | 0.026778 | 0.966921 |
| 20 | 0.003801 | 0.003803 | 0.037141 | 0.955255 |
| 25 | 0.003778 | 0.003401 | 0.025834 | 0.966988 |
| 30 | 0.004904 | 0.002594 | 0.019444 | 0.973058 |
| 35 | 0.007725 | 0.002228 | 0.015956 | 0.974091 |
| 40 | 0.013437 | 0.001815 | 0.012860 | 0.971888 |
| 45 | 0.023511 | 0.001297 | 0.009262 | 0.965931 |
| 50 | 0.036282 | 0.001155 | 0.008531 | 0.954033 |
| 55 | 0.061318 | 0.000999 | 0.006328 | 0.931356 |
| 60 | 0.097873 | 0.001103 | 0.005652 | 0.895372 |
| 65 | 0.152017 | 0.001051 | 0.005482 | 0.841449 |
| 70 | 1.000000 | 0.000000 | 0.000000 | 0.000000 |

Estimated using parameters $k_{tot}$ and $\alpha_{tot}$

| age | death | migration from r.brit to | | |
|---|---|---|---|---|
| | | e.anglia | s.east | r.brit |
| 0 | 0.022138 | 0.002882 | 0.018335 | 0.956644 |
| 5 | 0.001871 | 0.002461 | 0.015392 | 0.980277 |
| 10 | 0.001633 | 0.002038 | 0.013536 | 0.982793 |
| 15 | 0.003251 | 0.003291 | 0.028926 | 0.964532 |
| 20 | 0.003797 | 0.004592 | 0.050819 | 0.940792 |
| 25 | 0.003777 | 0.003927 | 0.030489 | 0.961807 |
| 30 | 0.004903 | 0.002671 | 0.020336 | 0.972089 |
| 35 | 0.007725 | 0.002237 | 0.015947 | 0.974091 |
| 40 | 0.013437 | 0.001795 | 0.012538 | 0.972229 |
| 45 | 0.023511 | 0.001305 | 0.008951 | 0.966232 |
| 50 | 0.036282 | 0.001136 | 0.008269 | 0.954313 |
| 55 | 0.061319 | 0.001035 | 0.006063 | 0.931584 |
| 60 | 0.097874 | 0.001138 | 0.005441 | 0.895547 |
| 65 | 0.152019 | 0.001084 | 0.005267 | 0.841631 |
| 70 | 1.000000 | 0.000000 | 0.000000 | 0.000000 |

Calculated using eqn. (1)

| age | death | migration from r.brit to | | |
|---|---|---|---|---|
| | | e.anglia | s.easl | r.brit |
| 0 | 0.021738 | 0.004690 | 0.029974 | 0.943598 |
| 5 | 0.001798 | 0.003871 | 0.023720 | 0.970611 |
| 10 | 0.001556 | 0.003043 | 0.019920 | 0.975481 |
| 15 | 0.003169 | 0.005401 | 0.047277 | 0.944153 |
| 20 | 0.003588 | 0.007540 | 0.077211 | 0.911661 |
| 25 | 0.003647 | 0.006435 | 0.049150 | 0.940768 |
| 30 | 0.004800 | 0.004348 | 0.033083 | 0.957769 |
| 35 | 0.007492 | 0.003422 | 0.024503 | 0.964583 |
| 40 | 0.013599 | 0.002444 | 0.017808 | 0.966149 |
| 45 | 0.023844 | 0.001625 | 0.012251 | 0.962281 |
| 50 | 0.038534 | 0.001478 | 0.011135 | 0.948853 |
| 55 | 0.061789 | 0.001125 | 0.008033 | 0.929053 |
| 60 | 0.098387 | 0.001260 | 0.007228 | 0.893125 |
| 65 | 0.153253 | 0.001212 | 0.007210 | 0.838325 |
| 70 | 1.000000 | 0.000000 | 0.000000 | 0.000000 |

## APPENDIX D

This Appendix gives the distribution of expectations of life at exact age $x$ in three regions of Great Britain. The data are calculated in four different ways, as in Appendix C. The expectations of life provide a better empirical verification of the discussion in the text than the probabilities of migration and death given previously.

The Appendix also includes the regional distribution of life expectancies at age 0, as a proportion of the total life expectancy. This is called the migration level.

Migration levels (regional distribution of life expectancy at exact age 0 as a proportion of total life expectancy).

Calculated using eqn. (2)

| | e.anglia | s.east | r.brit |
|---|---|---|---|
| e.anglia | 0.560325 | 0.035449 | 0.018489 |
| s.east | 0.198130 | 0.737912 | 0.116309 |
| r.brit | 0.241544 | 0.226639 | 0.865202 |
| total | 1.000000 | 1.000000 | 1.000000 |

Estimated using parameters $k_{tot}$ and $\alpha(x)$

| | e.anglia | s.east | r.brit |
|---|---|---|---|
| e.anglia | 0.583179 | 0.030217 | 0.015845 |
| s.east | 0.185816 | 0.751615 | 0.111445 |
| r.brit | 0.231004 | 0.218168 | 0.872710 |
| total | 1.000000 | 1.000000 | 1.000000 |

Estimated using parameters $k_{tot}$ and $\alpha_{tot}$

| | e.anglia | s.east | r.brit |
|---|---|---|---|
| e.anglia | 0.545893 | 0.033476 | 0.016616 |
| s.east | 0.209131 | 0.734537 | 0.120364 |
| r.brit | 0.244976 | 0.231987 | 0.863019 |
| total | 1.000000 | 1.000000 | 1.000000 |

Calculated using eqn. (1)

| | e.anglia | s.east | r.brit |
|---|---|---|---|
| e.anglia | 0.408922 | 0.041217 | 0.023187 |
| s.east | 0.257889 | 0.643877 | 0.163357 |
| r.brit | 0.333190 | 0.314906 | 0.813456 |
| total | 1.000000 | 1.000000 | 1.000000 |

Distribution of life expectancy at exact age $x$ for people born in East Anglia.

Calculated using eqn. (2)

| age | total | e.anglia | s.east | r.brit |
|---|---|---|---|---|
| 0 | 73.06435 | 40.93980 | 14.47626 | 17.64828 |
| 5 | 69.43669 | 36.94765 | 14.62970 | 17.85934 |
| 10 | 64.55702 | 32.71232 | 14.32012 | 17.52458 |
| 15 | 59.65361 | 28.78300 | 13.86263 | 17.00798 |
| 20 | 54.85471 | 25.22608 | 13.26733 | 16.36129 |
| 25 | 50.05791 | 22.07278 | 12.46714 | 15.51799 |
| 30 | 45.22318 | 19.29715 | 11.48181 | 14.44422 |
| 35 | 40.41728 | 16.82754 | 10.39531 | 13.19442 |
| 40 | 35.67406 | 14.58276 | 9.25866 | 11.83264 |
| 45 | 31.06776 | 12.52962 | 8.11201 | 10.42613 |
| 50 | 26.66482 | 10.65217 | 6.98460 | 9.02805 |
| 55 | 22.45778 | 8.91441 | 5.88743 | 7.65594 |
| 60 | 18.59584 | 7.34651 | 4.87481 | 6.37453 |
| 65 | 15.12872 | 5.95201 | 3.97142 | 5.20529 |
| 70 | 12.15285 | 4.76494 | 3.20632 | 4.18159 |

Estimated using parameters $k_{tot}$ and $\alpha(x)$

| age | total | e.anglia | s.east | r.brit |
|---|---|---|---|---|
| 0 | 73.09180 | 42.62563 | 13.58164 | 16.88453 |
| 5 | 69.46439 | 38.65459 | 13.73097 | 17.07884 |
| 10 | 64.58482 | 34.38632 | 13.44905 | 16.74944 |
| 15 | 59.68147 | 30.39919 | 13.02908 | 16.25321 |
| 20 | 54.88312 | 26.75449 | 12.49320 | 15.63542 |
| 25 | 50.08735 | 23.47217 | 11.78339 | 14.83179 |
| 30 | 45.25232 | 20.53234 | 10.90085 | 13.81913 |
| 35 | 40.44584 | 17.87718 | 9.91503 | 12.65363 |
| 40 | 35.70110 | 15.44244 | 8.87327 | 11.38538 |
| 45 | 31.09241 | 13.20630 | 7.81328 | 10.07283 |
| 50 | 26.68634 | 11.15388 | 6.76988 | 8.76258 |
| 55 | 22.47504 | 9.26905 | 5.74510 | 7.46089 |
| 60 | 18.60916 | 7.58867 | 4.78855 | 6.23193 |
| 65 | 15.13769 | 6.09829 | 3.92913 | 5.11027 |
| 70 | 12.15793 | 4.82587 | 3.19806 | 4.13401 |

Estimated using parameters $k_{tot}$ and $\alpha_{tot}$

| age | total | e.anglia | s.east | r.brit |
|---|---|---|---|---|
| 0 | 73.03786 | 39.87083 | 15.27450 | 17.89253 |
| 5 | 69.41029 | 35.87259 | 15.44019 | 18.09751 |
| 10 | 64.53058 | 31.64669 | 15.13105 | 17.75284 |
| 15 | 59.62715 | 27.69015 | 14.69195 | 17.24505 |
| 20 | 54.82837 | 24.08595 | 14.12933 | 16.61309 |
| 25 | 50.03131 | 20.95878 | 13.31690 | 15.75563 |
| 30 | 45.19673 | 18.30097 | 12.25106 | 14.64470 |
| 35 | 40.39109 | 15.95318 | 11.06757 | 13.37034 |
| 40 | 35.64829 | 13.80612 | 9.84316 | 11.99900 |
| 45 | 31.04238 | 11.82975 | 8.61960 | 10.59303 |
| 50 | 26.63955 | 10.01059 | 7.43094 | 9.19802 |
| 55 | 22.43355 | 8.33453 | 6.27926 | 7.81976 |
| 60 | 18.57315 | 6.83635 | 5.21392 | 6.52289 |
| 65 | 15.10771 | 5.50704 | 4.26024 | 5.34044 |
| 70 | 12.13194 | 4.37217 | 3.44912 | 4.31064 |

Calculated using eqn. (1)

| age | total | e.anglia | s.east | r.brit |
|---|---|---|---|---|
| 0 | 72.80315 | 29.77078 | 18.77510 | 24.25727 |
| 5 | 69.16043 | 25.73113 | 18.93327 | 24.49603 |
| 10 | 64.28154 | 21.89120 | 18.44443 | 23.94591 |
| 15 | 59.38100 | 18.45856 | 17.76056 | 23.16188 |
| 20 | 54.57787 | 15.47465 | 16.90228 | 22.20094 |
| 25 | 49.77751 | 13.07009 | 15.76754 | 20.93987 |
| 30 | 44.92973 | 11.18631 | 14.39406 | 19.34936 |
| 35 | 40.11012 | 9.62420 | 12.92924 | 17.55667 |
| 40 | 35.36221 | 8.25938 | 11.44214 | 15.66070 |
| 45 | 30.77094 | 7.04332 | 9.97619 | 13.75143 |
| 50 | 26.37483 | 5.93988 | 8.56031 | 11.87464 |
| 55 | 22.21351 | 4.94600 | 7.21045 | 10.05706 |
| 60 | 18.37084 | 4.06105 | 5.96118 | 8.34861 |
| 65 | 14.91985 | 3.27413 | 4.85012 | 6.79560 |
| 70 | 11.98053 | 2.60649 | 3.92115 | 5.45289 |

## APPENDIX D *(continued)*

Distribution of life expectancy at exact age $x$ for people born in South East England.

Calculated using eqn. (2)

| age | total | e.anglia | s.east | r.brit |
|---|---|---|---|---|
| 0 | 72.76063 | 2.57926 | 53.69096 | 16.49041 |
| 5 | 69.25768 | 2.61335 | 49.91430 | 16.73003 |
| 10 | 64.37093 | 2.56135 | 45.36145 | 16.44813 |
| 15 | 59.46222 | 2.48366 | 40.96908 | 16.00947 |
| 20 | 54.64418 | 2.38796 | 36.80819 | 15.44803 |
| 25 | 49.81932 | 2.26641 | 32.85591 | 14.69700 |
| 30 | 44.98650 | 2.11804 | 29.13494 | 13.73352 |
| 35 | 40.18504 | 1.95048 | 25.63597 | 12.59859 |
| 40 | 35.44996 | 1.77230 | 22.32324 | 11.35442 |
| 45 | 30.85555 | 1.59060 | 19.20091 | 10.06404 |
| 50 | 26.46224 | 1.40968 | 16.27994 | 8.77262 |
| 55 | 22.28231 | 1.23027 | 13.55335 | 7.49869 |
| 60 | 18.44316 | 1.05817 | 11.09296 | 6.29203 |
| 65 | 14.99926 | 0.89583 | 8.93211 | 5.17132 |
| 70 | 12.02812 | 0.75029 | 7.11128 | 4.16656 |

Estimated using parameters $k_{tot}$ and $\alpha(x)$

| age | total | e.anglia | s.east | r.brit |
|---|---|---|---|---|
| 0 | 72.76698 | 2.19881 | 54.69272 | 15.87545 |
| 5 | 69.26402 | 2.22518 | 50.92850 | 16.11033 |
| 10 | 64.37720 | 2.17621 | 46.34756 | 15.85343 |
| 15 | 59.46841 | 2.10607 | 41.91064 | 15.45170 |
| 20 | 54.65029 | 2.02167 | 37.69820 | 14.93042 |
| 25 | 49.82510 | 1.91699 | 33.69706 | 14.21106 |
| 30 | 44.99222 | 1.79054 | 29.92350 | 13.27818 |
| 35 | 40.19072 | 1.64767 | 26.35215 | 12.19090 |
| 40 | 35.45534 | 1.49471 | 22.95306 | 11.00757 |
| 45 | 30.86038 | 1.33819 | 19.73841 | 9.78378 |
| 50 | 26.46599 | 1.18213 | 16.72888 | 8.55498 |
| 55 | 22.28550 | 1.02862 | 13.92323 | 7.33364 |
| 60 | 18.44527 | 0.88313 | 11.39010 | 6.17204 |
| 65 | 14.99966 | 0.74464 | 9.15718 | 5.09784 |
| 70 | 12.02564 | 0.61683 | 7.26773 | 4.14108 |

Estimated using parameters $k_{tot}$ and $\alpha_{tot}$

| age | total | e.anglia | s.east | r.brit |
|---|---|---|---|---|
| 0 | 72.74550 | 2.43525 | 53.43427 | 16.87598 |
| 5 | 69.24203 | 2.46342 | 49.64874 | 17.12988 |
| 10 | 64.35522 | 2.40892 | 45.07616 | 16.87014 |
| 15 | 59.44641 | 2.33465 | 40.64504 | 16.46672 |
| 20 | 54.62826 | 2.24561 | 36.44219 | 15.94046 |
| 25 | 49.80346 | 2.12610 | 32.50816 | 15.16919 |
| 30 | 44.97081 | 1.97334 | 28.86644 | 14.13102 |
| 35 | 40.16946 | 1.80167 | 25.44020 | 12.92759 |
| 40 | 35.43467 | 1.62196 | 22.17944 | 11.63327 |
| 45 | 30.84056 | 1.44174 | 19.09189 | 10.30692 |
| 50 | 26.44775 | 1.26500 | 16.19757 | 8.98518 |
| 55 | 22.26844 | 1.09366 | 13.49517 | 7.67961 |
| 60 | 18.43030 | 0.93320 | 11.05284 | 6.44426 |
| 65 | 14.98720 | 0.78243 | 8.89777 | 5.30700 |
| 70 | 12.01601 | 0.64498 | 7.07244 | 4.29859 |

Calculated using eqn. (1)

| age | total | e.anglia | s.east | r.brit |
|---|---|---|---|---|
| 0 | 72.79819 | 3.00049 | 46.87308 | 22.92462 |
| 5 | 69.12948 | 3.02074 | 42.92278 | 23.18596 |
| 10 | 64.23332 | 2.93499 | 38.53592 | 22.76242 |
| 15 | 59.32033 | 2.81985 | 34.36416 | 22.13633 |
| 20 | 54.50202 | 2.68534 | 30.48161 | 21.33507 |
| 25 | 49.68726 | 2.52106 | 26.96223 | 20.20396 |
| 30 | 44.85318 | 2.32707 | 23.80667 | 18.71943 |
| 35 | 40.04630 | 2.11663 | 20.90770 | 17.02197 |
| 40 | 35.29754 | 1.89919 | 18.18192 | 15.21644 |
| 45 | 30.70520 | 1.68388 | 15.62465 | 13.39666 |
| 50 | 26.31229 | 1.47354 | 13.23545 | 11.60331 |
| 55 | 22.15569 | 1.27213 | 11.02200 | 9.86156 |
| 60 | 18.32352 | 1.08362 | 9.01872 | 8.22118 |
| 65 | 14.88562 | 0.90651 | 7.25817 | 6.72094 |
| 70 | 11.94291 | 0.74655 | 5.79005 | 5.40631 |

Distribution of life expectancy at exact age $x$ for people born in the "rest of Britain".

Calculated using eqn. (2)

| age | total | e.anglia | s.east | r.brit |
|---|---|---|---|---|
| 0 | 71.96190 | 1.33051 | 8.36979 | 62.26160 |
| 5 | 68.53434 | 1.34994 | 8.51325 | 58.67115 |
| 10 | 63.65786 | 1.32549 | 8.40152 | 53.93085 |
| 15 | 58.75750 | 1.29045 | 8.21698 | 49.25008 |
| 20 | 53.94091 | 1.24672 | 7.94257 | 44.75162 |
| 25 | 49.13517 | 1.19011 | 7.52361 | 40.42144 |
| 30 | 44.31017 | 1.11931 | 6.97281 | 36.21805 |
| 35 | 39.51413 | 1.03806 | 6.34748 | 32.12859 |
| 40 | 34.79619 | 0.95081 | 5.68466 | 28.16071 |
| 45 | 30.22595 | 0.86057 | 5.00967 | 24.35571 |
| 50 | 25.87470 | 0.76876 | 4.34338 | 20.76256 |
| 55 | 21.73273 | 0.67466 | 3.69290 | 17.36516 |
| 60 | 17.95303 | 0.58378 | 3.09287 | 14.27637 |
| 65 | 14.58047 | 0.49951 | 2.55763 | 11.52333 |
| 70 | 11.68486 | 0.42570 | 2.10235 | 9.15681 |

Estimated using parameters $k_{tot}$ and $\alpha(x)$

| age | total | e.anglia | s.east | r.brit |
|---|---|---|---|---|
| 0 | 71.94492 | 1.13994 | 8.01793 | 62.78704 |
| 5 | 68.51715 | 1.15864 | 8.15388 | 59.20464 |
| 10 | 63.64065 | 1.14079 | 8.04166 | 54.45820 |
| 15 | 58.74026 | 1.11224 | 7.86004 | 49.76799 |
| 20 | 53.92361 | 1.07444 | 7.60031 | 45.24886 |
| 25 | 49.11797 | 1.02334 | 7.20830 | 40.88632 |
| 30 | 44.29305 | 0.95847 | 6.69441 | 36.64017 |
| 35 | 39.49707 | 0.88367 | 6.11145 | 32.50195 |
| 40 | 34.77940 | 0.80264 | 5.48742 | 28.48933 |
| 45 | 30.20950 | 0.71847 | 4.84514 | 24.64590 |
| 50 | 25.85888 | 0.63433 | 4.20931 | 21.01524 |
| 55 | 21.71770 | 0.55159 | 3.58729 | 17.57882 |
| 60 | 17.93922 | 0.47417 | 3.00984 | 14.45521 |
| 65 | 14.56795 | 0.40186 | 2.49125 | 11.67485 |
| 70 | 11.67320 | 0.33647 | 2.04715 | 9.28958 |

Estimated using parameters $k_{tot}$ and $\alpha_{tot}$

| age | total | e.anglia | s.east | r.brit |
|---|---|---|---|---|
| 0 | 71.96233 | 1.19575 | 8.66171 | 62.10488 |
| 5 | 68.53492 | 1.21545 | 8.81093 | 58.50854 |
| 10 | 63.65845 | 1.19724 | 8.69739 | 53.76382 |
| 15 | 58.75809 | 1.16848 | 8.51504 | 49.07458 |
| 20 | 53.94149 | 1.13023 | 8.25159 | 44.55966 |
| 25 | 49.13568 | 1.07687 | 7.82424 | 40.23458 |
| 30 | 44.31060 | 1.00737 | 7.24297 | 36.06027 |
| 35 | 39.51449 | 0.92694 | 6.58950 | 31.99805 |
| 40 | 34.79636 | 0.84035 | 5.89883 | 28.05717 |
| 45 | 30.22581 | 0.75099 | 5.19515 | 24.27967 |
| 50 | 25.87410 | 0.66206 | 4.50344 | 20.70860 |
| 55 | 21.73193 | 0.57490 | 3.83035 | 17.32668 |
| 60 | 17.95193 | 0.49349 | 3.20769 | 14.25076 |
| 65 | 14.57885 | 0.41759 | 2.64964 | 11.51162 |
| 70 | 11.68218 | 0.34908 | 2.17223 | 9.16087 |

Calculated using eqn. (1)

| age | total | e.anglia | s.east | r.brit |
|---|---|---|---|---|
| 0 | 72.04422 | 1.67049 | 11.76890 | 58.60482 |
| 5 | 68.58958 | 1.69563 | 11.95382 | 54.94012 |
| 10 | 63.70807 | 1.66608 | 11.76848 | 50.27349 |
| 15 | 58.80295 | 1.62132 | 11.48491 | 45.69672 |
| 20 | 53.98223 | 1.56295 | 11.07999 | 41.33929 |
| 25 | 49.16759 | 1.48281 | 10.44365 | 37.24113 |
| 30 | 44.33633 | 1.38078 | 9.60925 | 33.34629 |
| 35 | 39.53420 | 1.26460 | 8.69119 | 29.57841 |
| 40 | 34.80239 | 1.14097 | 7.73416 | 25.92726 |
| 45 | 30.23069 | 1.01562 | 6.77442 | 22.44065 |
| 50 | 25.87479 | 0.89212 | 5.84091 | 19.14176 |
| 55 | 21.76488 | 0.77351 | 4.94922 | 16.04214 |
| 60 | 17.97826 | 0.66222 | 4.12198 | 13.19406 |
| 65 | 14.59352 | 0.55830 | 3.38529 | 10.64993 |
| 70 | 11.69842 | 0.46493 | 2.76441 | 8.46908 |

# CONSTRUCTING MULTIREGIONAL LIFE TABLES USING PLACE-OF-BIRTH-SPECIFIC MIGRATION DATA

*Jacques Ledent*

## 1 INTRODUCTION

The ordinary life table is a device for following a closed group of people, born at the same time, as it decreases in size until the death of its last member. The emphasis is put on the nonreversible transition from one state (being alive) to another (being dead). A straightforward extension of this model is the multiple-decrement life table which recognizes transitions to more than one final absorbing state (e.g., decrements due to various causes of death).

However, when recurrent, non-final transitions occur, the latter model does not permit one to follow persons who have moved from one state to another and to analyze their subsequent experiences. Such a problem may be handled with the help of more-complex life tables which recognize entries, or increments into states, as well as exits, or decrements from states. Because of their general nature, such life tables, known as increment–decrement life tables, are valuable in the analysis of marital status, labor-force participation, birth parity, and interregional migration; in the last case, they are often referred to as multiregional life tables (Rogers 1973).

Among such generalized life tables, a distinction is often made between uniradix increment–decrement life tables, for which the initial cohort is concentrated in a unique state, and multiradix increment–decrement life tables, for which the initial cohort is allocated to several, if not all, of the intercommunicating states.

The key feature of all increment–decrement life tables—whether uniradix or multiradix—lies in their formulation as simple Markov-chain models. As a consequence, such generalized life tables rely on stringent assumptions (population homogeneity and Markovian behavior) which are far from reflecting reality and thus often lead to faulty results (Ledent 1980a). This is especially true in the case of multiregional life tables since, as is well known, individuals with identical demographic characteristics (age, sex, and race) can exhibit quite different propensities for migration depending on past events in their lives.

In particular, consider perhaps the most interesting results that may be drawn from a multiregional life table, namely, the number of years (both total and distributed according to the regions in which they are to be spent) that an individual born in any of the regions can expect to live. These results are likely to be highly inaccurate if they are derived from a multiregional life table calculated with the traditional approach (that is, as a multiradix increment–decrement life table based on the type of migration data commonly available). This inaccuracy arises because the application, in the traditional approach, of the same age schedules of mobility to the individuals of a given region (regardless of their region of birth) ignores the generally well-established fact that migration propensities are heavily dependent on the birthplace of the individuals concerned. (For a quantitative observation of this effect in the United States, see Long and Hansen 1975; see also Ledent 1981).

Therefore, to provide more-acceptable values of the regional expectations of life at birth (both total values and regional shares), multiregional life tables should rely on interregional migration data cross-classified by place of birth. This paper demonstrates the construction of such multiregional life tables, which involves the calculation of a uniradix increment–decrement life table for each of the regional shares of the initial cohort. It also compares such an approach (hereafter called the place-of-birth-dependent approach) with the traditional approach based on commonly available migration data (the place-of-birth-independent approach). An illustration is provided by applying it to a system consisting of the four US Census Regions observed during the period 1965–1970, for females only; the necessary migration data can be readily derived from published census information (US Bureau of the Census 1973).

The rest of this paper is divided into five sections. Section 2, intended as a background section, presents a brief reminder of the theory and mathematical treatment of increment–decrement life tables. Section 3 is a discussion of the issue at hand, i.e., the influence of the population-homogeneity assumption on the calculation of such tables: the discussion is centered on the particular role of the birthplace in migration decisions. Section 4 reports on the implementation of the place-of-birth-dependent approach and Section 5 provides some perspectives on the contrast which this approach offers with respect to the usual place-of-birth-independent approach. Finally, Section 6 summarizes and presents the general conclusions of the paper. The general method used to construct the various increment–decrement life tables considered in this paper is described in the Appendix.

## 2 INCREMENT-DECREMENT LIFE TABLES: A REMINDER

Although some of the issues underlying the construction of increment–decrement life tables were considered long ago, it is only recently that thorough and systematic discussion of the methodological and empirical problems raised by such construction has appeared in the literature. Nevertheless, in less than a decade, the contributions of a number of researchers (Rogers 1973, 1975; Schoen and Nelson 1974; Rogers and Ledent 1975, 1976; Schoen 1975; Hoem and Fong 1976; Schoen and Land 1977; Ledent 1978, 1980a; Krishnamoorthy 1979) have led to the development of a formal mathematical treatment which now gives increment–decrement life tables a status comparable to that of the ordinary life table.

Perhaps the single most important factor responsible for this development was the realization that an increment–decrement life table can be regarded as a generalized life table in which elements in matrix format are substituted for the scalar elements of the ordinary life table (Rogers and Ledent 1975, 1976; Rogers 1975).

In this section we present an overview of a mathematical treatment of increment–decrement life tables that parallels the classical exposition of the ordinary life table: the correspondence between the formulas relevant to the ordinary and the increment–decrement life tables, respectively, is stressed in Table 1. Equation

TABLE 1 A tabular comparison of the theoretical exposition of ordinary and increment–decrement life tables[a].

| Ordinary life table | | Increment–decrement life table | |
|---|---|---|---|
| $\mu(y) = \lim_{dy \to 0} \frac{d(y)}{l(y)\,dy}$ | (1) | ${}^{i}\mu^{j}(y) = \lim_{dy \to 0} \frac{{}^{i}d^{j}(y)}{l^{i}(y)\,dy}$ | (1′) |
| $l(y + dy) = l(y) - d(y)$ | (2) | $l^{i}(y + dy) = l^{i}(y) - \sum_{\substack{j=1 \\ j \neq i}}^{r+1} {}^{i}d^{j}(y) + \sum_{\substack{j=1 \\ j \neq i}}^{r} {}^{j}d^{i}(y)$ | (2′) |
| $\frac{d}{dy} l(y) = -\mu(y) l(y)$ | (3) | $\frac{d}{dy} \mathbf{l}(y) = -\boldsymbol{\mu}(y) \mathbf{l}(y)$ | (3′) |
| $l(y) = \Omega(y) l(0)$ | (4) | $\mathbf{l}(y) = \boldsymbol{\Omega}(y) \mathbf{l}(0)$ | (4′) |
| $\Omega(y) = \exp\left[-\int_0^y \mu(t)\,dt\right]$ | (5) | | |
| $l_{x+n} = p_x l_x$ | (6) | $\mathbf{l}_{x+n} = \mathbf{p}_x \mathbf{l}_x$ | (6′) |
| $p_x = \Omega(x + n)/\Omega(x)$ | (7) | $\mathbf{p}_x = \boldsymbol{\Omega}(x + n) \boldsymbol{\Omega}(x)^{-1}$ | (7′) |
| $L_x = \int_0^n l(x + t)\,dt$ | (8) | $\mathbf{L}_x = \int_0^n \mathbf{l}(x + t)\,dt$ | (8′) |
| $T_x = \int_0^\infty l(x + t)\,dt$ | (9) | $\mathbf{T}_x = \int_0^\infty \mathbf{l}(x + t)\,dt$ | (9′) |
| $e_x = T_x / l_x$ | (10) | $\mathbf{e}_x = \mathbf{T}_x \mathbf{l}_x^{-1}$ | (10′) |

[a] Taken from Ledent (1980a, p. 536 and 542).

numbers (1)–(10) and (1′)–(10′) used below refer to the numbered equations in Table 1.

Suppose we have a system of $r + 1$ states ($r$ intercommunicating states plus the state of death) in which the initial cohort is allocated among $s$ states ($1 \leq s \leq r$): let $l^{i}(0)$ be the "radix" of state $i$. The principal problem here is one of estimating the state-specific curves of survivors $l^{i}(y)$ at each age $y$. Such estimation is centered around the differential equation (3′); it presents a vector notation of the $r$ scalar equations arrived at by substituting the equations (1′) defining the instantaneous mobility rates into the accounting equations (2′) showing the increments and decrements to each $l^{i}(y)$ group. [Note that eqn. (3′) is a straightforward vector extension of the basic differential equation (3) of the ordinary life table.]

Equation (3′) admits $r$ linearly independent solutions, which can be expressed as eqn. (4′), a straightforward matrix extension of the ordinary life table solution, eqn. (4). These independent solutions of eqn. (3′) are the $r$ stationary populations that are generated by an arbitrary radix in each of the $r$ states (regardless of whether some of the states are initially empty or not).

The matrix $\boldsymbol{\Omega}(y)$ is a proper transition probability matrix showing the state-specific survival probabilities at age $y$ of the members of each radix. [Note that unlike its counterpart in the ordinary life table, this matrix cannot be simply expressed in terms of the instantaneous mobility rates, but has to be determined by the infinitesimal calculus of Volterra (Schoen and Land 1977).] The number of survivors $\mathbf{l}_x$, at fixed ages 0, $n$, $2n$, . . . , may be derived by applying in succession, as shown in eqn. (6′), a set of age-specific transition probability matrices $\mathbf{p}_x$ (generalizing the age-specific survival probabilities $p_x$ of the ordinary life table).

Now, it is possible to define multistate life-table functions generalizing the usual statistics found in a life table. Equation (8′) defines the multistate life-table function $\mathbf{L}_x$ whose $(i, j)$th element represents the number of people born in state $j$ and alive in state $i$ of the life table between ages $x$ and $x+n$, or alternatively, the number of person-years lived in state $i$ between those ages by the members of the $j$th radix. From there, it is possible to define generalized $T$-statistics [eqn. (9′)] and, finally, generalized $e$-statistics [eqn. (10′)]: the $(i, j)$th element of $\mathbf{e}_x$ denotes the number of future years that an $x$-year-old individual present in state $j$ can expect to spend in state $i$.

Another generalization of interest is that of the mortality rates $m_x$ and survivorship proportions $s_x$ of the ordinary life table, because the calculation of applied increment–decrement life tables is centered around the equalization of the life-table values of the generalized $m$- or $s$-statistics with their observed counterparts. The relevant approaches are known as the movement and transition approaches, and were devised by Schoen (1975) and by Rogers (1973, 1975), respectively.

On the one hand, interstate "passage" can be observed as a move, that is, an instantaneous event similar to a death. This leads to the movement approach—consistent with the approach taken in the ordinary life table—in which the linkage with the observed population is ensured through an equalization of the life-table mortality and mobility rates with their observed counterparts. On the other hand, interstate "passage" can be observed as a change in an individual's state of presence between two points in time (regardless of the number of moves made in the meantime). This is the essence of the transition approach, in which the linkage with the observed population is ensured through an equalization of the life-table survivorship proportions with their observed counterparts.

These two alternative approaches are not competitive but complementary, in that the choice of either is dictated by the type of data at hand (for a detailed comparison, see Ledent 1980a). In fact, in most applications of increment–decrement life tables to real situations, the movement approach is the more relevant. The major exception, which requires the use of the transition approach, occurs in the field of interregional migration when data are obtained from population censuses that describe changes of residence between two points in time.

## 3 THE ISSUE ADDRESSED IN THIS PAPER

The most important feature of increment–decrement life tables is the formulation of their underlying model as a simple Markov-chain model. It follows that all the individuals of a given age present at the same time in a given state have identical propensities for moving out of that state (the population-homogeneity assumption) and that these propensities are independent of the past history of the individuals concerned (the Markovian assumption).

Clearly, in some instances, such an assumption is far from being realistic. Take, for example, the case of interregional migration in which the place of birth of the prospective migrant heavily influences his decision to move and his choice of destination. For example, in their study of migration flows to the South from the rest of the USA, Long and Hansen (1975) present convincing evidence that the probability of moving to the South is considerably higher for those born in the South than for those born elsewhere. Also, the present author (Ledent 1981) has described some more general evidence of the influence of the place of birth on migration patterns, with reference to a four-region disaggregation of the USA.

The migration data set used in this paper was obtained by reordering data taken from the volume *Lifetime and Recent Migration* published by the US Bureau of the Census (1973). The lengthy Table 11 of that volume provides estimates of the numbers of residents in each Federal state in 1970, cross-classified by place of birth and place of residence in 1965 (ten geographical units have been used: the state of residence in 1970 and the nine US Census Divisions). These estimates are provided for each sex and for each race and are subdivided into ten age groups: 0–4, 5–9, 10–14, 15–19, 20–24, 25–29, 30–39, 40–49, 50–59, 60 and over (all ages referring to 1965). The data concerning females in the ten age groups were aggregated and rearranged to show the changes of residence (cross-classified by place of birth) which were made between 1965 and 1970 in the US Census four-region system. The interregional migration streams thus obtained for the highly migratory group of women aged 20–24 in 1965 are shown in Table 2. For example, 73,703 women in that age group moved from the South region to the North Central region. Of that total, 43,047 were born in the South, and 30,656 elsewhere. Interestingly enough, most of the non-Southern-born migrants—24,847 or 81%—were born in the North Central region.

These figures do indeed show large differences in the propensity to migrate according to the place of birth. If we ignore for a moment the place of birth, then the average female 20–24-year-old Southern resident has a 0.0392 probability of moving to the North Central region over a five-year period; however, when place of birth is taken into account, we find that the probability is either lower (0.0260 for the Southern-born) or higher (for the non-Southern-born) than the average value. For the case of those born outside the South, the probability reaches 0.0449 and 0.0495, for women born in the Northeast and the West, respectively, but increases to almost 25% (0.2491) for those born in the North Central region. More generally, someone living outside his or her region of birth appears to have a high probability of returning there. (For a detailed analysis of this subject, see Ledent 1981.)

Clearly, the large mobility differentials according to place of birth, just described, sharply contradict the population-homogeneity assumption which underlies the calculation of a multiregional life table from migration data relating to the total national population. Thus, we may reasonably predict that statistics on expectations of life at birth obtained from such multiregional life tables will be inaccurate, because they are based on average mobility propensities rather than on mobility propensities specific to the regional shares of the initial cohort.

However, the availability of interregional migration data cross-classified by place of birth, such as those shown in Table 2, immediately suggests the possibility of circumventing or, more exactly, reducing the effects of the population-homogeneity assumption that underlies the calculation of a multiregional life table from aggregate

TABLE 2 Place-of-birth-specific interregional migration flows over the period 1965–1970, for females aged 20–24 in 1965[a].

| From | To Northeast | North Central | South | West |
|---|---|---|---|---|
| *Born in the Northeast* | | | | |
| Northeast | 1,110,763 | 18,637 | 36,184 | 24,299 |
| North Central | 10,491 | 33,482 | 3,597 | 3,256 |
| South | 21,675 | 4,051 | 59,628 | 4,808 |
| West | 9,562 | 2,331 | 3,798 | 46,020 |
| *Born in the North Central* | | | | |
| Northeast | 21,364 | 7,887 | 3,297 | 2,887 |
| North Central | 16,550 | 1,285,304 | 38,998 | 48,157 |
| South | 3,546 | 24,847 | 64,224 | 7,117 |
| West | 2,833 | 23,586 | 7,282 | 133,306 |
| *Born in the South* | | | | |
| Northeast | 83,292 | 3,112 | 3,847 | 2,905 |
| North Central | 2,919 | 159,435 | 28,663 | 6,217 |
| South | 23,652 | 43,047 | 1,553,583 | 23,064 |
| West | 2,218 | 4,876 | 33,119 | 108,464 |
| *Born in the West* | | | | |
| Northeast | 7,088 | 660 | 891 | 3,595 |
| North Central | 834 | 22,878 | 2,134 | 8,216 |
| South | 976 | 1,758 | 23,212 | 9,554 |
| West | 5,611 | 10,563 | 13,470 | 579,719 |
| *Place of birth not considered* | | | | |
| Northeast | 1,222,507 | 30,296 | 54,219 | 33,686 |
| North Central | 30,794 | 1,501,099 | 73,392 | 65,846 |
| South | 49,849 | 73,703 | 1,700,647 | 54,598 |
| West | 20,224 | 41,356 | 47,614 | 867,509 |

[a] Data obtained by aggregating data from the US Bureau of the Census (1973, Table 11).

(place-of-birth-independent) migration data. The main idea here is to construct separate uniradix increment–decrement life tables for each of the radices, i.e., the regional shares of the (arbitrary) initial cohort. In this way, multistate life-table statistics can be obtained which no longer relate to a single homogeneous population but to a population divided into $r$ homogeneous groups (as many as there are regions), defined by place of birth.

## 4 AN ILLUSTRATION OF THE PLACE-OF-BIRTH-DEPENDENT APPROACH

Methodologically, the implementation of the approach just suggested does not raise any problem: it simply requires the calculation of $r$ increment–decrement life tables instead of one (the fact that they are uniradix rather than multiradix increment–decrement life tables does not have any bearing on the actual calculation of the multistate life-table functions). Thus, in this section, we simply demonstrate this new approach by applying it to the set of US place-of-birth-specific migration data discussed in Section 3.

Because the migration information available here evidently concerns changes of residence between two points in time, the relevant approach here is the transition approach. The actual calculation method used (an overview is presented in the Appendix) combines the estimation of the age-specific survival probability using a method developed elsewhere by this author (Ledent 1980b) and the calculation of the number of person-years lived, $\mathbf{L}_x$, from a linear integration approach (Rogers 1973, 1975).

Note that because no mortality information cross-classified by place of birth is available we simply use the same set of age-specific mortality rates: those observed for the population of each region regardless of the region of birth (US Department of Health, Education, and Welfare, selected years). Actually, this treatment hardly constitutes a problem. In effect, although it does not yield the most precise values for the multistate statistics referring to each regional cohort, the consideration of identical mortality rates for the calculation of the four uniradix increment–decrement life tables appears to be quite acceptable: the dependence of mortality on the place of birth is probably minimal as long as the spatial units considered are broad geographical areas (this is certainly less true in the case of rural–urban systems, especially in developing countries). Therefore, comparison of the multistate life-table statistics relating to each radix offers an assessment of the influence of differential mobility according to the place of birth, with the effect of mortality differentials removed.

Let us now examine the actual results obtained for the application described above. Table 3, which sets out the transition probabilities according to their region of birth for women exactly 20 years old, confirms the general observation that the probability of moving from region $i$ to region $j$ is smaller for those born in region $i$ and much higher for those born in region $j$ than for those who were born neither in region $i$ nor in region $j$.

Table 4 shows the numbers of remaining years—disaggregated into periods specific to the regions in which they are spent—that 20-year-old residents of each region can expect to live, depending on their place of birth. For example, a resident of the South region who was born in the South is expected to survive a further 56.55 years, of which 49.29 years (about 87.2%) will be spent in the South. However, if this Southern resident had been born in another region, a much smaller part of her remaining lifetime (from 56.27 to 57.53 years according to the region of birth) would be spent in the South: 22.08 years if born in the Northeast, 20.09 years if born in the North Central region, and 16.45 years if born in the West.

Observe the regional variations in the total expectations of remaining life according to the place of birth, in spite of the fact that the mortality pattern is independent of the place of birth. For example, the total expectation of remaining life for a Southern resident is much higher (lower) if she was born in the West (Northeast) than if born in the South: this is indeed a consequence of the assumption underlying multiregional life tables that an in-migrant adopts the mortality regime of the region to which she has just moved.

TABLE 3 Place-of-birth-dependent approach: transition probabilities for females exactly 20 years old.

| From | To: Northeast | North Central | South | West | Death |
|---|---|---|---|---|---|
| *Born in the Northeast* | | | | | |
| Northeast | 0.9138 | 0.0206 | 0.0397 | 0.0228 | 0.003060 |
| North Central | 0.2114 | 0.6429 | 0.0726 | 0.0696 | 0.003389 |
| South | 0.2359 | 0.0475 | 0.6611 | 0.0517 | 0.003738 |
| West | 0.1264 | 0.0352 | 0.0538 | 0.7812 | 0.003390 |
| *Born in the North Central* | | | | | |
| Northeast | 0.5914 | 0.2639 | 0.0839 | 0.0490 | 0.003133 |
| North Central | 0.0137 | 0.9064 | 0.0357 | 0.0480 | 0.003418 |
| South | 0.0339 | 0.2438 | 0.6527 | 0.0659 | 0.003774 |
| West | 0.0173 | 0.1292 | 0.0468 | 0.8033 | 0.003409 |
| *Born in the South* | | | | | |
| Northeast | 0.7795 | 0.0372 | 0.1499 | 0.0303 | 0.003119 |
| North Central | 0.0164 | 0.7954 | 0.1496 | 0.0352 | 0.003443 |
| South | 0.0184 | 0.0332 | 0.9212 | 0.0234 | 0.003837 |
| West | 0.0167 | 0.0370 | 0.1613 | 0.7815 | 0.003435 |
| *Born in the West* | | | | | |
| Northeast | 0.5943 | 0.0583 | 0.0876 | 0.2568 | 0.003119 |
| North Central | 0.0271 | 0.6707 | 0.0673 | 0.2315 | 0.003421 |
| South | 0.0299 | 0.0498 | 0.6720 | 0.2446 | 0.003778 |
| West | 0.0104 | 0.0207 | 0.0277 | 0.9378 | 0.003406 |

## 5 COMPARISON BETWEEN THE PLACE-OF-BIRTH-DEPENDENT AND PLACE-OF-BIRTH-INDEPENDENT APPROACHES

This section tries to provide a meaningful comparison between the place-of-birth-dependent and place-of-birth-independent approaches to the construction of a multiregional life table. In principle, this requires firstly the aggregation of the separate uniradix increment–decrement life tables previously calculated and secondly the comparison of the results thus obtained with those of the multiradix increment–decrement life table based on the same set of data aggregated over all birthplaces.

There is, however, an interesting conclusion which we can derive even before aggregating the various uniradix life tables. This relates to the life expectancies at birth and their regional distributions. Instead of focusing on expectations of life at age 20, let us consider the analogous expectations of life at age zero. In this case, the only expectations of life with any meaning are those calculated for people born and resident in the same region. Each of the uniradix increment–decrement life tables calculated provides a value for the expectation of life at birth for females born in the region concerned, broken down into several numbers indicating the time to be spent in each region. The values obtained from each uniradix life table can then be grouped into a single matrix, such as the one shown in the first part of Table 5. It appears that

TABLE 4 Place-of-birth-dependent approach: totals and regional distributions of remaining life (in years) for females exactly 20 years old.

| Region of residence | Number of years spent in region | | | | |
|---|---|---|---|---|---|
| | Northeast | North Central | South | West | Total |
| *Born in the Northeast* | | | | | |
| Northeast | 47.31 | 1.66 | 4.48 | 2.67 | 56.11 |
| North Central | 23.53 | 19.25 | 7.49 | 6.15 | 56.42 |
| South | 25.26 | 3.45 | 22.08 | 5.48 | 56.27 |
| West | 18.76 | 3.20 | 6.97 | 27.85 | 56.78 |
| *Born in the North Central* | | | | | |
| Northeast | 14.86 | 25.17 | 7.62 | 9.18 | 56.82 |
| North Central | 0.99 | 46.73 | 3.80 | 5.30 | 56.83 |
| South | 2.15 | 25.76 | 20.09 | 8.78 | 56.78 |
| West | 1.29 | 17.48 | 4.79 | 33.72 | 57.27 |
| *Born in the South* | | | | | |
| Northeast | 27.88 | 3.73 | 21.47 | 3.41 | 56.49 |
| North Central | 1.58 | 30.21 | 21.10 | 3.87 | 56.76 |
| South | 1.51 | 3.17 | 49.29 | 2.58 | 56.55 |
| West | 1.66 | 4.01 | 23.04 | 28.37 | 57.08 |
| *Born in the West* | | | | | |
| Northeast | 12.59 | 3.64 | 4.99 | 36.37 | 57.60 |
| North Central | 1.50 | 17.41 | 4.38 | 34.31 | 57.60 |
| South | 1.64 | 3.39 | 16.45 | 36.05 | 57.53 |
| West | 0.66 | 1.53 | 1.99 | 53.66 | 57.83 |

an American woman has a total expectation of life greater than 74 years (from 74.20 years if born in the South to 75.85 years if born in the West), of which more than 60 years are to be spent in the region of birth (from 60.21 years if born in the North Central region to 68.51 years if born in the West).

How do these expectations of life compare with those obtained with the place-of-birth-independent approach, that is, from the multiradix increment–decrement life table based on the same data set aggregated over all birthplaces? The matrix of expectations of life at birth produced by the latter approach is shown in the second part of Table 5; it indicates a much smaller proportion of total lifetime spent in the region of birth (from 64.7% to 70.1%) than does the place-of-birth-dependent approach where it ranges between 81.0% and 90.3%.

Thus, the substitution of place-of-birth-specific migration data for the more traditional place-of-birth-independent data increases the expected numbers of years to be spent in the region of birth by about ten years (9.69 years in the case of the Northeast, 9.78 years in the case of the North Central, 9.99 years in the case of the South) except in the case of the West where the increase is almost twice this size (19.81 years). This result is consistent with the earlier observation that, once an American woman—and most particularly one born in the West region—has moved out of her region of birth, she is very likely to return. In addition, note that the use of place-of-birth-specific migration data implies increased differentials between the

TABLE 5 Totals and regional distributions of expectations of life at birth (in years): comparison between the place-of-birth-dependent and place-of-birth-independent approaches.

| Region of birth | Number of years spent in region: Northeast | North Central | South | West | Total |
|---|---|---|---|---|---|
| *Place-of-birth-dependent approach* | | | | | |
| Northeast | 61.78 | 2.53 | 6.06 | 3.87 | 74.24 |
| North Central | 1.44 | 60.21 | 5.22 | 7.43 | 74.30 |
| South | 2.49 | 5.10 | 62.63 | 3.99 | 74.20 |
| West | 1.10 | 2.71 | 3.52 | 68.51 | 75.85 |
| *Place-of-birth-independent approach* | | | | | |
| Northeast | 52.09 | 5.80 | 10.99 | 5.56 | 74.43 |
| North Central | 4.10 | 50.43 | 11.38 | 8.36 | 74.26 |
| South | 5.55 | 8.83 | 52.64 | 7.23 | 75.25 |
| West | 4.45 | 9.25 | 12.88 | 48.70 | 75.27 |

total regional expectations of life, which take on values nearing those they would have if migration were ignored.

An equivalent and perhaps more telling way of assessing the impact on the calculation of a multiregional life table of using place-of-birth-specific migration data is to look at the changes in the regional percentage shares of the expectations of life at birth caused by using such data. From the values shown in Table 5, it can be readily established that the introduction of such disaggregated data cuts the proportion of lifetime to be spent outside the region of birth by about half, except in the case of the Western-born women for whom the cut amounts to slightly more than 70%: the proportion decreases from 30.0 to 16.8% for women born in the Northeast, from 32.1 to 19.0% for women born in the North Central, from 29.1 to 15.6% for Southern-born women, and from 35.3 to 9.7% for Western-born women.

We now turn to the aggregation of the four uniradix increment–decrement life tables—calculated for each regional share of the initial cohort—into a multiregional life table directly comparable to that obtained from the approach based on commonly available data. This aggregation raises the fundamental question of how to choose the most appropriate regional distribution of the initial cohort.

We suggest here that, since the mobility and mortality patterns studied in our USA illustration are those of a given period (1965–1970), the radices or regional shares of the initial cohort ought to be in proportion to the numbers of female births observed in each region over the same period. On this basis, the initial cohort, which we can arbitrarily set equal to 100,000 persons, should be allocated as follows: 22,735 (Northeast), 27,791 (North Central), 32,245 (South), and 17,229 (West).

The aggregated transition probabilities (for females of exact age 20) which result from such a regional allocation are shown in the first part of Table 6; the second part of the table shows the corresponding transition probabilities obtained with the place-of-birth-independent approach.

The two corresponding sets of transition probabilities are very similar, with significant discrepancies arising only in the probabilities of migration out of the West. The retention probability for the West region calculated using the place-of-birth-dependent approach is 0.9106 as compared with 0.8907 from the place-of-birth-

TABLE 6 Transition probabilities for females exactly 20 years old: comparison between the place-of-birth-dependent and place-of-birth-independent approaches.

| From | To | | | | |
|---|---|---|---|---|---|
| | Northeast | North Central | South | West | Death |
| *Place-of-birth-dependent approach* | | | | | |
| Northeast | 0.8975 | 0.0262 | 0.0461 | 0.0272 | 0.003064 |
| North Central | 0.0191 | 0.8869 | 0.0448 | 0.0458 | 0.003418 |
| South | 0.0285 | 0.0444 | 0.8901 | 0.3320 | 0.003828 |
| West | 0.0156 | 0.0312 | 0.0392 | 0.9106 | 0.003408 |
| *Place-of-birth-independent approach* | | | | | |
| Northeast | 0.8959 | 0.0265 | 0.0480 | 0.0265 | 0.003065 |
| North Central | 0.0194 | 0.8839 | 0.0485 | 0.0448 | 0.003420 |
| South | 0.0289 | 0.0443 | 0.8921 | 0.0309 | 0.003829 |
| West | 0.0187 | 0.0395 | 0.0476 | 0.8907 | 0.003409 |

independent approach; this corresponds to an absolute difference of 19.9 per thousand, compared to a maximum difference of 3.0 per thousand for the other regions. Similar results may also be observed for all the other age groups.

By contrast, the expectations of life which we obtain by aggregating the four place-of-birth-specific increment–decrement life tables calculated for each radix are quite different from those derived from the place-of-birth-independent approach. For example, the aggregated expectations of life for females of exact age 20 indicate that the number of remaining years to be spent in the region of residence are as follows: 45.4 years (if resident in the Northeast), 44.3 (if resident in the North Central region), 45.8 (if resident in the South), and 49.4 (if resident in the West); these should be compared with the values of 42.0, 41.1, 43.3, and 40.6 years, respectively, obtained from the place-of-birth-independent approach (see Table 7).

TABLE 7 Totals and regional distributions of expectations of remaining life (in years) for females exactly 20 years old: comparison between the place-of-birth-dependent and place-of-birth-independent approaches.

| Region of residence | Number of years spent in region | | | | |
|---|---|---|---|---|---|
| | Northeast | North Central | South | West | Total |
| *Place-of-birth-dependent approach* | | | | | |
| Northeast | 45.39 | 2.25 | 5.26 | 3.26 | 56.16 |
| North Central | 1.60 | 44.27 | 5.01 | 5.94 | 56.83 |
| South | 2.53 | 4.27 | 45.79 | 4.00 | 56.58 |
| West | 1.43 | 3.07 | 3.81 | 49.39 | 57.70 |
| *Place-of-birth-independent approach* | | | | | |
| Northeast | 41.95 | 3.44 | 7.26 | 3.61 | 56.26 |
| North Central | 2.57 | 41.05 | 7.52 | 5.66 | 56.80 |
| South | 3.46 | 5.38 | 43.32 | 4.40 | 56.56 |
| West | 2.80 | 5.87 | 8.10 | 40.62 | 57.39 |

Comparison of the results given in Tables 6 and 7 thus indicates that the place-of-birth-dependent approach leads to aggregate multistate life-table functions which are either similar to or largely different from those obtained from the place-of-birth-independent approach, depending on whether they relate to events occurring over a single age interval or over a longer period of time.

However, we should note that the results just derived rely on an aggregation of the four uniradix increment–decrement life tables calculated for each regional share of the initial cohort using one particular system of statistical weighting: the system seems intuitively reasonable, but arguments can be made for and against it. This raises the problem of whether alternative allocations of the initial cohort among the regions would lead to quite different aggregated multiregional life tables. In view of this uncertainty, we performed an alternative aggregation of the four uniradix increment–decrement life tables, this time using identical weights (radices). The multistate life-table functions thus obtained (which are not shown here) did not appear to differ very significantly from those calculated earlier. Thus we conclude that as long as the state allocation of the initial cohort consists of radices which more or less reflect the weights of the regions with regard to some meaningful socioeconomic factor—these weights are expected to represent an allocation which does not depart too much from an allocation into $r$ equal parts—very similar estimates of the aggregate multistate life-table functions should be produced.

To summarize briefly, unlike the place-of-birth-independent approach, the more desirable place-of-birth-dependent approach leads to aggregate multistate life-table functions which depend on the regional allocation of the initial cohort. However, as long as the radices are chosen on a reasoned basis, this "radix problem" does not exert an overly large influence on the values obtained.

## 6 SUMMARY AND CONCLUSION

An important assumption common to all life-table models is the population-homogeneity assumption stemming from the Markovian formulation of the models. This assumption is in sharp contrast to the observation that, in the real world, equally aged individuals of a given status category (i.e., belonging to a given state of the system) generally exhibit quite different tendencies to move out of their current status category.

These mobility differentials can be related first to different personal characteristics (e.g., sex, race) or socioeconomic characteristics (e.g., occupation) which affect the level of mobility at a given instant. Thus, to obtain more accurate estimates of increment–decrement life tables, it is possible simply to calculate separate life tables for those groups of people which can be easily distinguished, such as men and women or whites and non-whites, etc.

Second, and more important in the case of increment–decrement life tables, mobility differentials may also depend on whether the phenomenon may be repeated or not, and on the frequency of this repetition. Unfortunately, such differentials cannot generally be attributed to an easily identifiable characteristic and it is generally not possible to calculate separate life tables for more homogeneous groups. An exception to this statement occurs in the analysis of migration, when adequate census data allow one to distinguish homogeneous groups of migrants on the basis of their place of birth. In this case an alternative multiregional life table can be constructed as a set of uniradix increment–decrement life tables corresponding to each of the regional shares of the initial cohort.

Compared with the traditional approach to the calculation of a multiregional life table, this alternative approach appears to provide not only more detail (in the case of the transition probabilities) but also more accuracy (in the case of the expectations of life at birth): this improvement is the result of considering a more realistic migration pattern, one which explicitly accounts for return migration to the birthplace (a demographic phenomenon of considerable importance, as shown in Section 3).

However, we must note that the improvements in the calculation of multiregional life tables thus achieved represent only a partial step toward the total removal of the population-homogeneity assumption implicit in the traditional approach: this assumption is still present within the stationary population associated with each radix.

## ACKNOWLEDGMENTS

The author is grateful to Andrei Rogers for critical comments and constructive suggestions on an earlier draft. Thanks also go to Philip Rees who provided helpful advice on the paper.

## REFERENCES

Hoem, J. M., and M. S. Fong (1976) A Markovian Chain Model of Working Life Tables. Working Paper No. 2. Copenhagen: Laboratory of Actuarial Mathematics, University of Copenhagen.

Krishnamoorthy, S. (1979) Classical approach to increment–decrement life tables: an application to the study of marital status of United States females, 1970. Mathematical Biosciences 44(1–2):139–154.

Ledent, J. (1978) Some Methodological and Empirical Considerations in the Construction of Increment–Decrement Life Tables. RM-78-25. Laxenburg, Austria: International Institute for Applied Systems Analysis.

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

Ledent, J. (1980b) An Improved Methodology for Constructing Increment–Decrement Life Tables from the Transition Perspective. WP-80-104. Laxenburg, Austria: International Institute for Applied Systems Analysis.

Ledent, J. (1981) The Influence of the Birthplace on Geographical Mobility in the US. II: Analysis of Migration Propensities. Laxenburg, Austria: International Institute for Applied Systems Analysis (forthcoming).

Long, L. H., and K. A. Hansen (1975) Trends in return migration to the South. Demography 12(4):601–614.

McNeil, D. R., T. J. Trussell, and J. C. Turner (1977) Spline interpolation of demographic data. Demography 14:245–252.

Rees, P. H., and A. G. Wilson (1977) Spatial Population Analysis. London: Edward Arnold.

Rogers, A. (1973) The multiregional life table. Journal of Mathematical Sociology 3:127–137.

Rogers, A. (1975) Introduction to Multiregional Mathematical Demography. New York, N.Y.: John Wiley.

Rogers, A., and J. Ledent (1975) Multiregional Population Projection. Discussion paper No. 128. Evanston, Illinois: The Center for Mathematical Studies in Economics and Management Science, Northwestern University. Published in revised form in Optimization Techniques Modeling and Optimization in the Service of Man, Part I, edited by Jean Cea (pages 31–58), Springer-Verlag, Berlin, 1976.

Rogers, A., and J. Ledent (1976) Increment–decrement life tables: a comment. Demography 13(2):287–290.

Schoen, R. (1975) Constructing increment–decrement life tables. Demography 12(2):313–324.

Schoen, R., and K. C. Land (1977) A General Algorithm in Estimating a Markov-Generated Increment–Decrement Life Table with Applications to Marital Status Patterns. WP-77-15. Champaign, Illinois: Program in Applied Social Statistics, Department of Sociology, University of Illinois at Urbana. (Published in revised form in the Journal of the American Statistical Association 74:761–776 (1979).)

Schoen, R., and V. E. Nelson (1974) Marriage, divorce and mortality: a life table analysis. Demography 11:267–290.

US Bureau of the Census (1973) Census of Population: 1970: Lifetime and Recent Migration. Final Report PC(2)-2D. Washington, D.C.: US Government Printing Office.

US Department of Health, Education and Welfare (selected years) Vital Statistics of the United States. Washington, D.C.: US Government Printing Office.

## APPENDIX A: A BRIEF DESCRIPTION OF THE METHOD OF CALCULATION

The general calculation procedure used to construct both the uniradix and the multiradix increment–decrement life tables considered in this paper consists of two main steps:

1. First, estimating a set of transition probabilities $\bar{\mathbf{p}}_x$ conditional on survival, from the observed transition proportions conditional on survival (obtained from the matrices shown in Table 3 by dividing each element by the sum of the elements in the same row).
2. Second, transforming the $\bar{\mathbf{p}}_x$ into the required set of transition probabilities $\mathbf{p}_x$ by introducing mortality information.

More specifically, this procedure means that $\mathbf{p}_x$ is derived from

$$\mathbf{p}_x = \bar{\mathbf{p}}_x \mathbf{p}_x^{\sigma} \tag{A1}$$

where $\bar{\mathbf{p}}_x$ is a matrix of transition probabilities conditional on survival, evaluated in terms of the observed transition proportions conditional on survival ($\bar{\mathbf{S}}_x$), and $\mathbf{p}_x^{\sigma}$ is a diagonal matrix of survival probabilities.

As a first approximation, $\bar{\mathbf{p}}_x$ can be estimated using the averaging formula proposed by Rees and Wilson (1977)

$$\bar{\mathbf{p}}_x = \tfrac{1}{2}(\bar{\mathbf{S}}_{x-n} + \bar{\mathbf{S}}_x) \tag{A2}$$

However, a better estimation can be performed by interpolating between the conditional transition proportions in a less crude fashion. Ledent (1980a) suggests that for each pair of states $i$ and $j$ ($j \neq i$), one could interpolate between the conditional transition proportions ${}^i\bar{S}_x^j$ by using cubic-spline functions, which are increasingly coming into use in the field of demography (McNeil *et al.* 1977). Since we are dealing here with a five-year time interval (1965–1970), the ordinate—for age $y$—of the continuous curve thus obtained represents the probability that an individual present at age $y$ in region $i$ will be present in region $j$ five years later. In this way, it is possible, for each region $i$, to estimate at evenly spaced ages, $0, 5, 10, 15, \ldots$, migration probabilities ${}^{i}\bar{p}^{j}_{x}$ $(j = 1, 2, \ldots, r;\ j \neq i)$, from which retention probabilities ${}^{i}\bar{p}^{i}_{x} = 1 - \sum_{j=1, j\neq i}^{r} {}^{i}\bar{p}^{j}_{x}$ immediately follow.

The estimation of the set of survival probabilities $\mathbf{p}^{\sigma}_{x}$, assuming the availability of conventional mortality-rate data for each region, is not so straightforward, because the two-step estimation procedure suggested by eqn. (A1) requires the mortality pattern to be a characteristic of the place of residence at the exact age $x (x = 0, n, 2n, \ldots)$ immediately below the age at which death occurs (rather than a characteristic of the place of death between ages $x$ and $x + n$, as for conventional mortality rates).

The methodology used for such an estimation relies on an iterative procedure based on the estimation of mortality rates dependent on the place of residence at age $x$ from the conventionally observed mortality rates (for details of the iterative procedure, see Ledent 1980b). Then, the required set of survival probabilities is obtained by analogy with the survival probability $p_x$ of an ordinary life table from the relation

$$\mathbf{p}^{\sigma}_{x} = \left(\mathbf{I} + \frac{n}{2}\hat{\mathbf{M}}^{\delta}_{x}\right)^{-1}\left(\mathbf{I} - \frac{n}{2}\hat{\mathbf{M}}^{\delta}_{x}\right) \tag{A3}$$

where $\hat{\mathbf{M}}^{\delta}_{x}$ is a diagonal matrix containing the mortality rates previously estimated.

Finally, combining the estimates of $\bar{\mathbf{p}}_x$ and $\mathbf{p}^{\sigma}_{x}$ as shown in eqn. (A1) leads to the required estimates of $\mathbf{p}_x$ and then, by using eqn. (6′), to the estimates of $\mathbf{l}_x$. The numbers of person-years lived, $\mathbf{L}_x$, are then simply obtained using the usual linear-integration method (Rogers 1973, 1975) by assuming

$$\mathbf{L}_x = \frac{n}{2}\left(\mathbf{l}_x + \mathbf{l}_{x+n}\right) \tag{A4}$$

The [illegible] statistics are then calculated from

$$\mathbf{T}_x = \sum_{k=0}^{z-x/n} \mathbf{L}_{x+kn} \tag{A5}$$

where [illegible]denotes the last age group, and the expectations of life follow by use of eqn. (1

# MULTISTATE POPULATION PROJECTIONS

*Dimiter Philipov and Andrei Rogers*

## 1 INTRODUCTION

Much of mathematical demography is concerned with the measurement and projection of changes of state, or *status*, experienced by individuals during their lifetime, e.g., changes in marital status, in employment status, in educational status, and in residential location. The study of such transitions from state to state and the evolution of the associated status-specific populations is the focus of a growing body of methodological techniques and applications sometimes referred to as *multistate demography* (Rogers 1980).

Recent work in multistate mathematical demography has identified a unifying matrix-based generalization of classical techniques, which illuminates the common features of many of the well-known methods for dealing with transfers between multiple states of existence. For example, it is now understood that multiple decrement life tables, marital status life tables, tables of working life, tables of educational life, and multiregional life tables are all members of a general class of increment–decrement life tables known as *multistate life tables.* It has also become evident that projections of populations disaggregated by status can be carried out using a common methodology—*multistate projection.*

Although traditional single-state methods are more parsimonious in their data requirements and provide reasonably adequate results for many purposes, they cannot deal with interstate transitions differentiated by origins and destinations and must, therefore, account for changes in stocks by reference to *net* totals, e.g., net migration. In a recent paper we have shown that such an approach may introduce biases and inconsistencies into a projection and that multistate models have a decisive advantage over single-state models as a consequence of their ability to produce disaggregated projections that trace the evolution of subcategories of a population over time and space (Rogers and Philipov 1979). This feature of multistate projection methods is developed in this paper, in the particular context of *multiregional* demography.

## 2 STATIONARY AND STABLE POPULATION DISTRIBUTIONS

To make our argument less abstract, imagine a single-sex population (females) disaggregated into five-year age groups, and for simplicity consider its spatial distribution to extend over only two regions, North and South. For a numerical illustration let us draw on 1965–1970 data for the United States previously examined in Rogers and Castro (1976) and, more recently, in Ledent (1980). These data are set out in the Appendixes and will be used throughout this paper.* Note that three Census Regions, Northeast, North Central, and West, have been aggregated together to form a single region: the "rest of the United States" or, more simply, the North.

In 1968, the female population of the USA stood at 102.3 million, with 32.5 million in the South and 69.8 million in the North (Appendix A). Conventional single-region life table calculations give a Southern-born baby girl a life expectancy of 74.11 years, just four months less than the corresponding life expectancy of a baby girl born in the North. The gross reproduction rates in the two regions are 1.18 and 1.16, respectively.

Consider next the results of a multiregional (two-state) analysis (Rogers 1975). First, computing a biregional life table (Appendix B) we observe that about 27% of a Southern-born baby girl's life expectancy can be expected to be lived in the North. Projecting the biregional population 30 years forward on the assumption of constant rates gives a 1998 national total of 138.6 million, with 33.0% residing in the South (Appendix B). Continuing this projection to stability yields an ultimate share for the South of 34.5% and an intrinsic rate of growth of 4.361 per thousand (Appendix B).

The expectation of life at birth in a conventional single-state life table with a unit radix may be interpreted as the stationary population that underlies the life table calculations. This is also true for multistate life tables; hence, we may conclude that in the stationary biregional population set out in Appendix B about 72.6% of the total Southern-born population resides in the South as *natives*, whereas 84.8% of the Northern-born population lives in the North. This leaves the remaining 15.2% to live in the South as *aliens* (i.e., individuals living in a place different from their place of birth).

Multiplying the stationary population in each age group by $\exp[-r(x+2.5)]$, where $r$ is the intrinsic rate of growth and $x$ is the starting age of the age group, gives the relative age distribution of the place-of-birth-specific stable population resident in each region. Since $r$ is relatively small in our USA illustration ($r = 0.004361$), the stable share of natives and aliens in each region differs only slightly from the stationary (life-table) share, with the percentages of natives given above (72.6% and 84.8%) shifting to 72.3% and 86.4%, respectively. Multiplying each of these by the stable shares of the national population in each region (i.e., 34.5% and 65.5%, respectively) gives the stable shares of the national population in each of the four place-of-residence-by-place-of-birth (PRPB) subcategories, as shown in the bottom line of Table 1.

## 3 NATIVE-INDEPENDENT MULTISTATE PRPB POPULATION PROJECTIONS

Several recent studies of migration have emphasized the importance of analyzing the flow patterns of *return* migrants, pointing to the not-surprising empirical fact that the migration rates of people returning to their region of birth are significantly higher than average (Ledent 1980, Lee 1974, Long and Hansen 1975, Miller 1977).

* Appendix A contains the Rogers and Castro data; Appendix C presents the Ledent data.

TABLE 1 PRPB distribution at stability of national and regional female population of the USA ($r = 0.004361$).

| | Resident in South | | Resident in North | |
|---|---|---|---|---|
| | Born in South | Born in North | Born in North | Born in South |
| Percentage of regional population | 72.3 | 27.7 | 86.4 | 13.6 |
| Percentage of national population | 34.5 | | 65.5 | |
| Percentage distribution of national population | 24.9 | 9.6 | 56.6 | 8.9 |

In the next section we follow this advice and introduce higher transition probabilities for return migrants in the multistate projection model. We shall call the outputs of such models *native-dependent projections*. In this section, however, we treat first the simpler case of *native-independent projections*, i.e., projections carried out with models assuming that all of the individuals in a regional population experience identical age-specific probabilities of moving, dying, and bearing offspring.*

## 3.1 Fertility

In projecting a multistate population forward over time, we shall at times refer to people by where they live and at other times by where they were born. This poses no difficulties when we are dealing with survivors of a current population; it simply becomes a matter of keeping track of individuals born in each region. It is the births of new individuals that need to be examined, because the babies may be born in the region of residence of their parents at the start or at the end of the unit interval of time, and they themselves may migrate during the same interval into yet another region.

In the conventional multistate projection model, some of the babies born in a given region during a unit time interval $(t, t+1)$ may be living in another region at the end of that interval. Consequently, at time $t+1$ these babies can be distinguished both by their place of residence, $j$, and by their place of birth, $i$. Moreover, they may also be classified by the region of residence, say $k$, of their parent at the *start* of the time interval, because each regional population of parents is a potential contributor of babies to each PRPB-specific category of babies. For example, in our two-region illustration based on USA data, we distinguish four categories of babies for parents initially resident in each region. Figure 1 shows the four categories corresponding to parents initially resident in the South; there are of course four equivalent categories for babies born to parents initially resident in the North.

* Because of the unavailability of the necessary fertility and mortality data, we are unable to introduce native-dependency in birth and death rates.

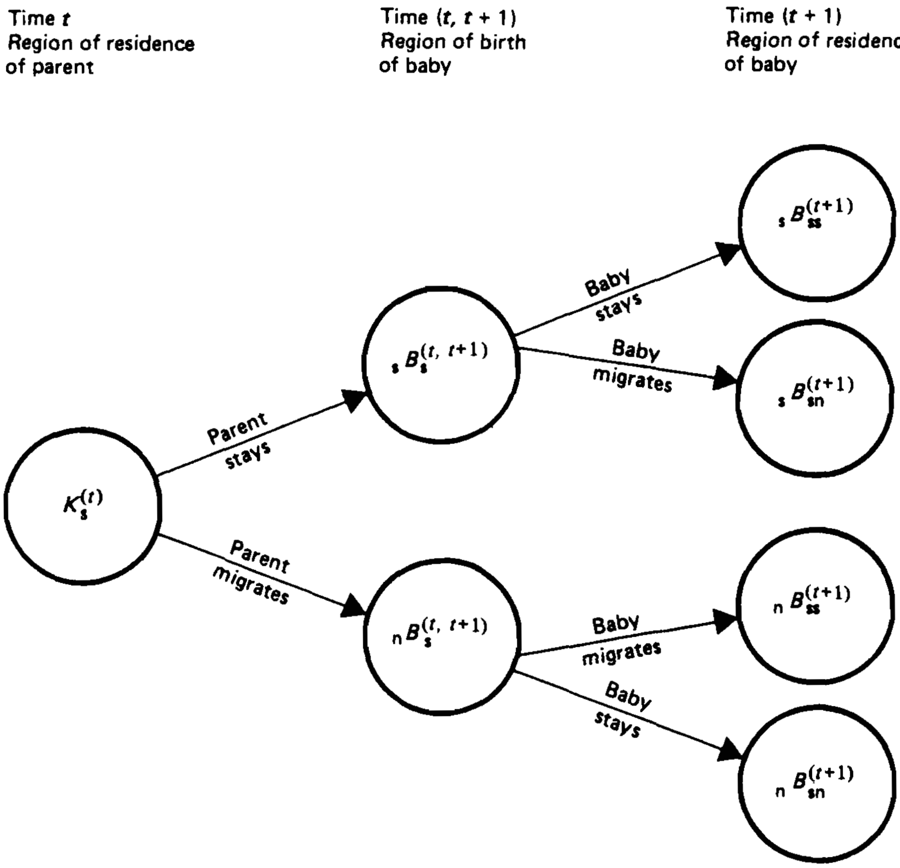



FIGURE 1 The four categories of babies born to parents resident in the South at time $t$.

Let

$$b^i_{kj}(x)=\frac{{}_iB^{(t+1)}_{kj}(x)}{K^{(t)}_k(x)} \tag{1}$$

denote the average number of babies born during the five-year time interval $(t, t+1)$ in region $i$ and alive in region $j$ at time $t+1$, for every individual between the ages of $x$ and $x+4$ living in region $k$ at time $t$. Summing over all birthplaces $i$ gives the conventional multiregional birth rate (Rogers 1975, p. 121)

$$b_{kj}(x)=\frac{1}{2}\left[\frac{{}_{k0}L_j(0)}{l_k(0)}F_k(x)+\sum_{h=1}^{m}s_{kh}(x)\frac{{}_{h0}L_j(0)}{l_h(0)}F_h(x+5)\right] \tag{2}$$

where

$F_h(x)$ is the annual birth rate of people aged $x$ to $x+4$ residing in region $h$

${}_{h0}L_j(0)$ is the total number of person-years lived between ages 0 and 5 in region $j$, per person born in region $h$

$s_{kh}(x)$ is the proportion of people living in region $k$ and aged $x$ to $x+4$ that survive to be in region $h$ and aged $x+5$ to $x+9$, five years later

$l_h(0)$ is the radix of region $h$ (set equal to unity in our calculations)

$m$ is the total number of regions

Since, by definition

$$b_{kj}(x) = \sum_{i=1}^{m} b_{kj}^{i}(x) \tag{3}$$

it is easy to develop computational formulas for $b_{kj}^{i}(x)$ by taking the appropriate components from eqn. (2). For our two-region (South–North) example, this gives four equations of the form

$$b_{kj}^{k}(x) = \frac{1}{2} \frac{{}_{k0}L_j(0)}{l_k(0)} [F_k(x) + s_{kk}(x) F_k(x+5)] \tag{4}$$

for $(k, j) = (\text{n, n}), (\text{s, s}), (\text{n, s}), (\text{s, n})$

for parents in two regions who do not migrate between time $t$ and the birth of the infant $(i = k)$, but whose child may or may not migrate before $t+1 (j = k \text{ or } j \neq k)$; and four equations

$$b_{kj}^{i}(x) = \frac{1}{2} \frac{{}_{i0}L_j(0)}{l_i(0)} [s_{ki}(x) F_i(x+5)] \tag{5}$$

for $(i, j, k) = (\text{n, s, s}), (\text{n, n, s}), (\text{s, n, n}), (\text{s, s, n})$

corresponding to parents who do migrate between time $t$ and the birth of the infant $(i \neq k)$, but whose child may or may not migrate before $t+1 (j = i \text{ or } j \neq i)$. This implies that a child may migrate without its parents between the ages of 0 and 5.

## 3.2 Projection

The age-specific birth rates, by region of birth of child, may be incorporated into the standard multiregional projection model (Rogers 1975, Chap. 5) transforming that model into a *multistate* projection model, where the states of interest are places of birth. This transformation makes it possible to generate projections that keep track of the regions of birth, i.e., that produce PRPB projections.

Appendix B describes the matrix model. Note that the Markovian assumption is still retained. All individuals in a region, recent in-migrants as well as established residents, aliens as well as natives, are assumed to experience identical probabilities of transition. This assumption is relaxed in Section 4.

Appendix B sets out the multistate growth matrix for our two-region (South–North), two-state (natives and aliens) example. Appendix B also presents the stable distribution across states that ultimately arises if this projection matrix is applied to *any* observed population. The stable distribution depends only on the elements of the growth matrix and not on the initial (base-year) population distribution. (Since it is also of some interest to use the matrix to generate projections, a 30-year projection based on the 1968 population is included in Appendix B.)

The stable growth results in Appendix B may be compared with the results of the conventional projection presented earlier in the same Appendix. Note that the intrinsic rate of growth remains the same ($r = 0.004361$), as does the spatial distribution of the national population ($sha_s = 34.46\%$ and $sha_n = 65.54\%$). The national and regional age compositions remain unchanged, with the mean age in the South being 37.94 years and that in the North 36.65. In short, the two projections to stability give identical results, as they indeed must. The multistate projection, however, includes additional information: it disaggregates regional populations by place of birth. It reveals, for example, that, at stability, the mean age of the alien population in the South will be about 15.3 years older than that of the native population and some 2.5 years older than the North's alien population. All of these stable growth results, however, could be obtained *without* the multistate growth matrix. We have shown earlier (Table 1) that a simple weighting of the stationary multiregional life table population gives identical results. The usefulness of the growth matrix, therefore, lies in generating projections such as that presented at the end of Appendix B.

## 4 NATIVE-DEPENDENT MULTISTATE PRPB POPULATION PROJECTIONS

### 4.1 Data

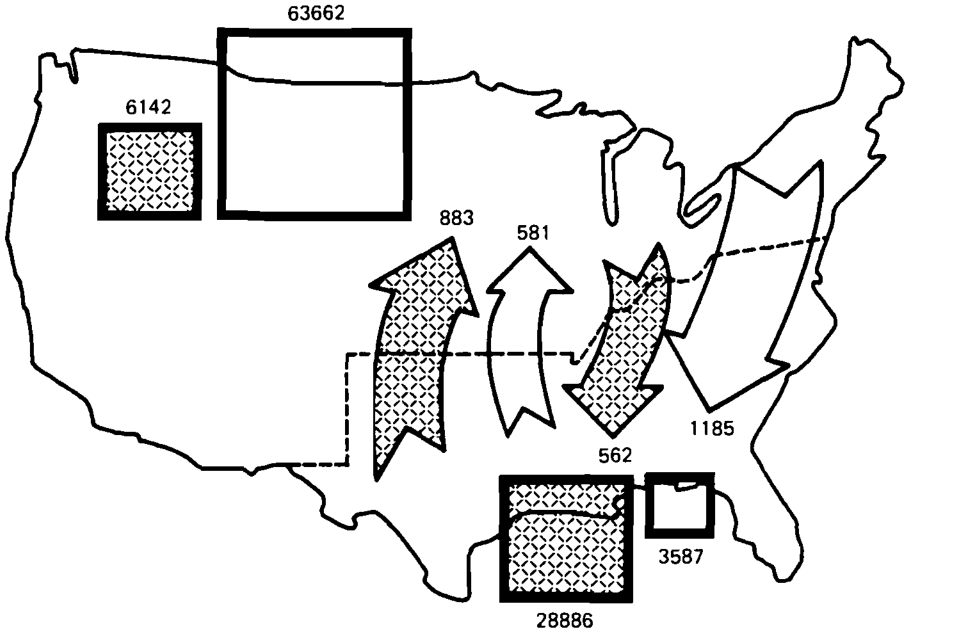
It is widely recognized that the migration rates of return migrants are significantly greater than the average rates of migration to the same destination (Ledent 1980, Long and Hansen 1975, Miller 1977). Migration data published by the 1970 US Census provide empirical support for this observation (US Bureau of the Census 1973). Appendix C contains the relevant figures for our two-region example.

The first table in the Appendix presents data on the Southern-born population residing in the South in 1968. It shows that the crude rate of migration of Southern-born females to the North was 6.12 per thousand. The next table sets out the corresponding data for the Southern-born population living in the North; this group has a crude migration rate to the South (i.e., return migration) of 18.30 per thousand, roughly three times as large. Nevertheless, because the Southern-born population resident in the South is much larger than that resident in the North, the corresponding *net* migration of Southern-borns into the South is negative.

The data on the Northern-born population living in the South show that the crude rate of return migration to the North is 32.39 per thousand, about ten times the rate of Northern-borns migrating to the South (3.72 per thousand, according to Appendix C). Once again, the net migration of natives into their region of birth is negative.

Appendix C also indicates that the native–alien composition of the flows in the two directions differs. The five-year flow from the South to the North consists of 883.4 thousand Southern-borns and 580.9 thousand Northern-borns, a 1.5 : 1 ratio. The flow from the North to the South, on the other hand, consists of 1.2 million Northern-borns and 562.1 thousand returning Southern-borns, a 2.1 : 1 ratio. The principal reason for this difference is the 2:1 ratio of the two populations under consideration. The North, with about two-thirds of the national population, sends roughly 1.2 times more migrants to the South than it receives in return (Figure 2).

Although native-dependent migration data are available for the USA, there is apparently no comparable data on fertility and mortality. Thus, in the next sections we retain the Markovian assumption for birth and death rates, assuming that




Population stocks, 1968 (in thousands)
Born in North
Born in South

Migration flows, 1965–1970 (in thousands)
Born in North
Born in South

Migration rates per thousand

| | N→S | S→N |
|---|---|---|
| Born in North | 3.72 | 32.39 |
| Born in South | 18.30 | 6.12 |

FIGURE 2 Population flows between two regions (North and South) of the USA, disaggregated by region of birth. Note that the migration flows cover a five-year period (1965–1970) but the migration rates are crude (annual) values for 1968.

everyone residing in a given region is exposed to the same fertility and mortality regimes. Consequently, our development of a native-dependent multistate projection model will treat only migration as being native-dependent. The necessary extensions to include fertility and mortality should be straightforward, but for the USA, at least, it is not likely that such an extension would produce significantly different results. There are situations, however, where it could make a great deal of difference, e.g., in projecting urban and rural populations.

## 4.2 Life Table

The computation of a PRPB native-dependent life table is a straightforward exercise (Ledent 1980). One simply calculates a separate table for each cohort, applying to it the appropriate PRPB probabilities. No conceptual innovations are required; indeed, a standard multiregional life table program (Willekens and Rogers 1978) may be used. A program of this type and the data in Appendix C were used to produce the native-dependent life table summarized in Appendix D.

Appendix D shows that the probabilities of return migration are significantly larger than those of non-return migration. For example, the probability that a Southern-born 20-year-old female living in the South will be in the North five years later is 0.0551. For the corresponding Northern-born females living in the South this probability is 0.2749; i.e., return migration is five times more probable than non-return migration. Roughly the same ratio is exhibited by the probabilities of aliens and natives migrating to the South (0.0263 compared with 0.1300).

Applying these probabilities to Southern-born and Northern-born cohorts in a multistate life table results in the expectations of life given in Appendix D. As an example, Table 2 presents the expected distribution of remaining lifetime for the two cohorts at age 20. The table illustrates the striking effect that place of birth has on the locations where the individual is expected to spend the rest of her life. A Southern-born female living in the North at age 20 is likely to spend over half of her remaining expected lifetime of 56.59 years in her region of birth, while a Northern-born female of the same age and place of residence is only likely to spend five years of her remaining lifetime in the South, i.e., six times less than the native Southerner.

TABLE 2 Distribution of life expectancy at exact age 20 by place of birth, place of residence at age 20, and place of future residence.

| | Born in South | | Born in North | |
|---|---|---|---|---|
| | Resident in South (age 20) | Resident in North (age 20) | Resident in South (age 20) | Resident in North (age 20) |
| Future years spent in South | 46.41 | 32.43 | 13.06 | 5.08 |
| Future years spent in North | 10.10 | 24.16 | 43.52 | 51.55 |
| Total remaining life expectancy | 56.51 | 56.59 | 56.58 | 56.63 |

## 4.3 Fertility

The introduction of native-dependent migration behavior into the calculation of the fertility elements of the multistate growth matrix is straightforward and uses the native-dependent probabilities and survivorship proportions defined in the native-dependent life table. The formulas for $b^i_{kj}(x)$ become

$$b^i_{kj} = \sum_{h=1}^{m} {}_h b^i_{kj}(x) \qquad (6)$$

where the rates now include a subscript on the left-hand side to denote the place of birth of the parent and hence the place-of-birth-specific probabilities used to calculate expected births.

The required computation procedure can be more readily understood if eqns. (4) and (5) are first re-expressed in the alternative form (Willekens and Rogers 1978, p. 59)

$$b^k_{kj}(x) = \tfrac{5}{4}p_{kj}(0)[F_k(x) + s_{kk}(x)F_k(x+5)] \qquad j \neq k \tag{7}$$

$$= \tfrac{5}{4}[1 + p_{kk}(0)][F_k(x) + s_{kk}(x)F_k(x+5)] \qquad j = k \tag{8}$$

$$b^i_{kj}(x) = \tfrac{5}{4}p_{ij}(0)[s_{ki}(x)F_i(x+5)] \qquad j \neq i \tag{9}$$

$$= \tfrac{5}{4}[1 + p_{ii}(0)][s_{ki}(x)F_i(x+5)] \qquad j = i \tag{10}$$

since

$$\frac{{}_{k0}L_j(0)}{l_k(0)} = \tfrac{5}{2}p_{kj}(0) \qquad k \neq j \tag{11}$$

and

$$\frac{{}_{k0}L_k(0)}{l_k(0)} = \tfrac{5}{2}[1 + p_{kk}(0)] \qquad k = j \tag{12}$$

when the linear integration formula is used to calculate person-years on a unit radix.

Equations (7–10) may be made native-dependent by replacing $p_{kj}(0)$ by ${}_hp_{kj}(0)$ and $s_{ki}(x)$ by ${}_hs_{ki}(x)$. The native-dependent probabilities and survivorship proportions may be obtained from the multistate life table (see, for example, Appendix D). In our two-region numerical example, the birth rates with $h$ equal to the baby's place of birth may be found as a residual

$${}_hb^h_{kj}(x) = {}_hb_{kj}(x) - {}_hb^i_{kj}(x) \tag{13}$$

## 4.4 Projection

The various native-dependent birth rates and survivorship proportions are collected to form the matrices $\mathbf{B}(x)$ and $\mathbf{S}(x)$ defined in eqn. (B6) of Appendix B and organized in the structure of the growth matrix defined in eqn. (B5), and illustrated in Appendix D. This yields a native-dependent multistate projection model that distinguishes among transition probabilities and regional populations according to place of birth. Such a model produces projections somewhat different to those of the native-independent counterpart discussed in Section 3 of this paper. Table 3 compares selected outputs; more detailed results from the native-dependent model may be found in Appendix D.

Table 3 identifies two very important characteristics of native-dependent and native-independent projections. First, aggregate totals and growth rates are the same in both projections if the Markovian assumption is retained for fertility and mortality rates. For example, both methods predict that the total female population of the USA will be 138.6 million in 1998 and will ultimately attain a stable rate of growth of 0.00436. Second, the percentage share of *natives* in each regional population is

TABLE 3 Native-dependent and native-independent PRPB projections of the 1968 female population of the USA: to 1998 and to stability.

| Year | Resident in South | | Resident in North | | Total population[a] |
|---|---|---|---|---|---|
| | Born in South | Born in North | Born in North | Born in South | |
| (*a*) *Native-dependent projections*[b] | | | | | |
| 1968 | 28,885,548 | 3,586,779 | 63,662,232 | 6,142,451 | 102,277,016 |
| (% of total) | (28.2) | (3.5) | (62.2) | (6.0) | (100) |
| 1998 | 38,495,044 | 6,289,250 | 86,446,904 | 7,378,696 | 138,609,888 |
| (% of total) | (27.8) | (4.5) | (62.4) | (5.3) | (100) |
| (% of stable population) | (26.9) | (5.0) | (63.3) | (4.7) | (100) |
| (*b*) *Native-independent projections*[c] | | | | | |
| 1968 | 28,885,548 | 3,586,779 | 63,662,232 | 6,142,451 | 102,277,016 |
| (% of total) | (28.2) | (3.5) | (62.2) | (6.0) | (100) |
| 1998 | 34,966,964 | 10,832,081 | 81,580,392 | 11,213,492 | 138,592,928 |
| (% of total) | (25.2) | (7.8) | (58.9) | (8.1) | (100) |
| (% of stable population) | (24.9) | (9.6) | (56.6) | (8.9) | (100) |

[a] Totals may not equal the sums of the columns due to independent rounding.
[b] From Appendix D ($r = 0.004360$).
[c] From Appendix B ($r = 0.004361$).

consistently underestimated in the native-independent projections because they do not take into account the higher probabilities of return migration. This suggests that disaggregations by place of birth may not lead to significant improvements in the accuracy with which national population *growth* is projected; however, they are important in analyzing projected *redistributions* of national populations.

Note that in the native-dependent projection the South's share of the national population consistently hovers at the level of 32%, whereas in the native-independent projection it increases slightly over time to an ultimate share of just over 34%. A comparison of the mean ages of natives and aliens as given in Appendixes B and D suggests that the native-dependent projection generates a slightly *older* native population and a *younger* alien population in each region.

## 5 EXTENSIONS

The fundamental concepts discussed in this paper have been illustrated with a four-state projection model in which two of the states referred to regions of residence and the other two to regions of birth. This disaggregation produced PRPB population projections, i.e., projections of regional populations disaggregated into *natives* and *aliens*. The extension of this projection methodology to a larger number of states is relatively straightforward. For example, we may further disaggregate natives into *stayers**, who have never left the region of birth, and *returners*, who have left the

* Stayers can only be approximated by assuming that individuals present in a region both at the beginning and end of a unit interval of time did not leave the region during this period.

region of birth and come back again. Similarly, aliens may be disaggregated into *recent aliens*, aliens who have arrived during the most recent time interval, and *established aliens*, who arrived previously. Thus we have the disaggregation

$$\begin{aligned} \text{residents} &= \text{natives} + \text{aliens} \\ &= \text{stayers} + \text{returners} + \text{established aliens} + \text{recent aliens} \end{aligned}$$

The projected native-independent stable population presented in Appendix B is disaggregated in this way in Appendix E. An analogous result could be obtained for the native-dependent stable population in Appendix D.

Table 4 contains selected results from Appendix E. It is interesting to note the surprisingly low shares of the native and alien populations accounted for by returners and recent aliens, respectively, and to observe the large variations in the mean ages of the various status-specific populations.

TABLE 4 Characteristics of four different resident categories in the stable population[a], calculated using the native-independent projection ($r = 0.004361$).

| Type of resident | South | | | North | | |
|---|---|---|---|---|---|---|
| | Stable regional population | % of regional total | Mean age | Stable regional population | % of regional total | Mean age |
| Natives | | | | | | |
| Stayers | 30,996,010 | 69.3 | 32.71 | 71,193,688 | 83.7 | 34.48 |
| Returners | 1,377,703 | 3.1 | 56.29 | 2,320,481 | 2.7 | 53.83 |
| Aliens | | | | | | |
| Established aliens | 10,674,229 | 23.8 | 51.82 | 10,225,156 | 12.0 | 49.35 |
| Recent aliens | 1,700,556 | 3.8 | 31.42 | 1,360,090 | 1.6 | 25.42 |
| All residents | 44,748,500 | 100 | 37.94 | 85,099,416 | 100 | 36.65 |

[a] The stable population given here is proportional to that listed in Appendix B, and can be scaled to the same totals.

## 6 CONCLUSION

Multistate population projections disaggregate conventional population projections into a number of state-specific subcategories, such as region of residence, region of birth, and duration of residence in the current location. The disaggregated projections should produce more accurate results if interstate transition probabilities are dependent on the categories chosen. This appears to be particularly true for projections of the distribution of an aggregate population across categories of several types. In our numerical example it was necessary to assume native-independent fertility and mortality rates, and so the aggregate growth rate of the population, not surprisingly, was unaffected by the disaggregation. However, more interesting results are likely to be obtained, for example, by using urban–rural fertility data in a projection for a typical developing country.

## ACKNOWLEDGMENTS

The authors are grateful to Tatiana Zaslavskaya for useful suggestions and criticisms. Jacques Ledent kindly provided the native-dependent data and commented on an earlier draft. Thanks also go to Denise Baker-Promper for a careful job of typing.

## REFERENCES

Ledent, J. (1980) Constructing Multiregional Life Tables Using Place-of-Birth-Specific Migration Data. WP-80-96. Laxenburg, Austria: International Institute for Applied Systems Analysis.

Lee, A. S. (1974) Return migration in the United States. International Migration Review 8:283–300.

Long, L. H., and K. A. Hansen (1975) Trends in return migration to the South. Demography 12:601–614.

Miller, A. R. (1977) Interstate migrants in the United States: Some social–economic differences by type of move. Demography 14:1–17.

Rogers, A. (1975) Introduction to Multiregional Mathematical Demography. New York: Wiley.

Rogers, A., Ed. (1980) Essays in Multistate Demography. RR-80-10. Laxenburg, Austria: International Institute for Applied Systems Analysis (reprinted from a special issue of Environment and Planning A 12(5) May 1980).

Rogers, A., and L. J. Castro (1976) Model Multiregional Life Tables and Stable Populations. RR-76-9. Laxenburg, Austria: International Institute for Applied Systems Analysis.

Rogers, A., and D. Philipov (1979) Multiregional Methods for Subnational Population Projections. WP-79-40. Laxenburg, Austria: International Institute for Applied Systems Analysis (forthcoming in Sistemi Urbani).

US Bureau of the Census (1973) Census of Population: 1970: Lifetime and Recent Migration. Final Report PC(2)-2D, Washington D.C.: US Government Printing Office.

Willekens, F., and A. Rogers (1978) Spatial Population Analysis: Methods and Computer Programs. RR-78-18. Laxenburg, Austria: International Institute for Applied Systems Analysis.

# APPENDIX A

Native-independent input data, disaggregated by region of residence, for the female population of the USA in 1968*.

Data for South

| age | population number | population - % - | births number | births - % - | deaths number | deaths - % - | arrivals number | arrivals - % - | departures number | departures - % - | observed rates ( x 1000 ) birth | death | inmig | outmig | net mig |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 3334898. | 10.27 | 0. | 0.00 | 17782. | 8.24 | 209403. | 11.99 | 199715. | 13.64 | 0.000 | 5.332 | 12.558 | 11.977 | 0.581 |
| 5 | 3542782. | 10.91 | 0. | 0.00 | 1504. | 0.70 | 172787. | 9.89 | 153848. | 10.51 | 0.000 | 0.425 | 9.754 | 8.685 | 1.069 |
| 10 | 3339162. | 10.28 | 2626. | 0.49 | 1161. | 0.54 | 162400. | 9.30 | 148174. | 10.12 | 0.786 | 0.348 | 9.727 | 8.875 | 0.852 |
| 15 | 2891785. | 8.91 | 114390. | 21.48 | 1853. | 0.86 | 272908. | 15.62 | 272329. | 18.60 | 39.557 | 0.641 | 18.875 | 18.835 | 0.040 |
| 20 | 2311419. | 7.12 | 183919. | 34.54 | 1786. | 0.83 | 215741. | 12.35 | 219342. | 14.98 | 79.570 | 0.773 | 18.667 | 18.979 | -0.312 |
| 25 | 1990180. | 6.13 | 121960. | 22.90 | 1888. | 0.87 | 141976. | 8.13 | 125653. | 8.58 | 61.281 | 0.949 | 14.268 | 12.627 | 1.640 |
| 30 | 2037422. | 6.27 | 67009. | 12.58 | 2846. | 1.32 | 107328. | 6.14 | 89143. | 6.09 | 32.889 | 1.397 | 10.536 | 8.751 | 1.785 |
| 35 | 1985913. | 6.12 | 32854. | 6.17 | 4479. | 2.08 | 85610. | 4.90 | 67244. | 4.59 | 16.544 | 2.255 | 8.622 | 6.772 | 1.850 |
| 40 | 1986324. | 6.12 | 9189. | 1.73 | 6585. | 3.05 | 72598. | 4.16 | 51118. | 3.49 | 4.626 | 3.315 | 7.310 | 5.147 | 2.163 |
| 45 | 1978231. | 6.09 | 584. | 0.11 | 9402. | 4.36 | 57991. | 3.32 | 34586. | 2.36 | 0.295 | 4.753 | 5.863 | 3.497 | 2.366 |
| 50 | 1734004. | 5.34 | 0. | 0.00 | 11382. | 5.27 | 54761. | 3.13 | 25819. | 1.76 | 0.000 | 6.564 | 6.316 | 2.978 | 3.338 |
| 55 | 1607191. | 4.95 | 0. | 0.00 | 15289. | 7.08 | 68248. | 3.91 | 20932. | 1.43 | 0.000 | 9.513 | 8.493 | 2.605 | 5.888 |
| 60 | 1086849. | 3.35 | 0. | 0.00 | 14476. | 6.71 | 55626. | 3.18 | 17635. | 1.20 | 0.000 | 13.319 | 10.236 | 3.245 | 6.991 |
| 65 | 932570. | 2.87 | 0. | 0.00 | 19581. | 9.07 | 33834. | 1.94 | 14781. | 1.01 | 0.000 | 20.997 | 7.256 | 3.170 | 4.086 |
| 70 | 709600. | 2.19 | 0. | 0.00 | 22675. | 10.51 | 19852. | 1.14 | 13320. | 0.91 | 0.000 | 31.955 | 5.595 | 3.754 | 1.841 |
| 75 | 495358. | 1.53 | 0. | 0.00 | 26323. | 12.20 | 11912. | 0.68 | 7993. | 0.55 | 0.000 | 53.139 | 4.809 | 3.227 | 1.582 |
| 80 | 299348. | 0.92 | 0. | 0.00 | 25558. | 11.84 | 2979. | 0.17 | 1998. | 0.14 | 0.000 | 85.379 | 1.990 | 1.335 | 0.655 |
| 85 | 209286. | 0.64 | 0. | 0.00 | 31275. | 14.49 | 994. | 0.06 | 667. | 0.05 | 0.000 | 149.437 | 0.950 | 0.637 | 0.312 |
| tot | 32472326. | 100.00 | 532531. | 100.00 | 215845. | 100.00 | 1746948. | 100.00 | 1464297. | 100.00 | | | | | |
| gross | | | | | | | | | | | 1.178 | 1.952 | 0.809 | 0.625 | |
| crude(x1000) | | | | | | | | | | | 16.400 | 6.647 | 10.760 | 9.019 | 1.741 |
| m.age | | 30.80 | | 25.08 | | 62.15 | | 25.86 | | 22.10 | 25.80 | 77.51 | 34.60 | 29.13 | |
| e(0) | | | | | | | | | | | | 74.11 | | | |

Data for North

| age | population number | population - % - | births number | births - % - | deaths number | deaths - % - | arrivals number | arrivals - % - | departures number | departures - % - | observed rates ( x 1000 ) birth | death | inmig | outmig | net mig |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 7452446. | 10.68 | 0. | 0.00 | 36704. | 8.08 | 199715. | 13.64 | 209403. | 11.99 | 0.000 | 4.925 | 5.360 | 5.620 | -0.260 |
| 5 | 7848825. | 11.24 | 0. | 0.00 | 2855. | 0.63 | 153848. | 10.51 | 172787. | 9.89 | 0.000 | 0.364 | 3.920 | 4.403 | -0.483 |
| 10 | 7183312. | 10.29 | 2296. | 0.21 | 2015. | 0.44 | 148174. | 10.12 | 162400. | 9.30 | 0.320 | 0.281 | 4.126 | 4.522 | -0.396 |
| 15 | 6089654. | 8.72 | 165521. | 15.25 | 3426. | 0.75 | 272329. | 18.60 | 272908. | 15.62 | 27.181 | 0.563 | 8.944 | 8.963 | -0.019 |
| 20 | 4912590. | 7.04 | 370626. | 34.16 | 3229. | 0.71 | 219342. | 14.98 | 215741. | 12.35 | 75.444 | 0.657 | 8.930 | 8.783 | 0.147 |
| 25 | 4167663. | 5.97 | 280926. | 25.89 | 3245. | 0.71 | 125653. | 8.58 | 141976. | 8.13 | 67.406 | 0.779 | 6.030 | 6.813 | -0.783 |
| 30 | 4380822. | 6.28 | 162729. | 15.00 | 4860. | 1.07 | 89143. | 6.09 | 107328. | 6.14 | 37.146 | 1.109 | 4.070 | 4.900 | -0.830 |
| 35 | 4279997. | 6.13 | 79516. | 7.33 | 7808. | 1.72 | 67244. | 4.59 | 85610. | 4.90 | 18.579 | 1.824 | 3.142 | 4.000 | -0.858 |
| 40 | 4459036. | 6.39 | 22108. | 2.04 | 12563. | 2.77 | 51118. | 3.49 | 72598. | 4.16 | 4.958 | 2.817 | 2.293 | 3.256 | -0.963 |
| 45 | 4572721. | 6.55 | 1358. | 0.13 | 18994. | 4.18 | 34586. | 2.36 | 57991. | 3.32 | 0.297 | 4.154 | 1.513 | 2.536 | -1.024 |
| 50 | 3706203. | 5.31 | 0. | 0.00 | 22826. | 5.03 | 25819. | 1.76 | 54761. | 3.13 | 0.000 | 6.159 | 1.393 | 2.955 | -1.562 |
| 55 | 3319385. | 4.76 | 0. | 0.00 | 30054. | 6.62 | 20932. | 1.43 | 68248. | 3.91 | 0.000 | 9.054 | 1.261 | 4.112 | -2.851 |
| 60 | 2069925. | 2.97 | 0. | 0.00 | 27637. | 6.08 | 17635. | 1.20 | 55626. | 3.18 | 0.000 | 13.352 | 1.704 | 5.375 | -3.671 |
| 65 | 1731326. | 2.48 | 0. | 0.00 | 36548. | 8.05 | 14781. | 1.01 | 33834. | 1.94 | 0.000 | 21.110 | 1.707 | 3.908 | -2.201 |
| 70 | 1438705. | 2.06 | 0. | 0.00 | 49052. | 10.80 | 13320. | 0.91 | 19852. | 1.14 | 0.000 | 34.095 | 1.852 | 2.760 | -0.908 |
| 75 | 1066579. | 1.53 | 0. | 0.00 | 60513. | 13.32 | 7993. | 0.55 | 11912. | 0.68 | 0.000 | 56.736 | 1.499 | 2.234 | -0.735 |
| 80 | 668357. | 0.96 | 0. | 0.00 | 59863. | 13.18 | 1998. | 0.14 | 2979. | 0.17 | 0.000 | 89.567 | 0.598 | 0.891 | -0.294 |
| 85 | 457131. | 0.65 | 0. | 0.00 | 71996. | 15.85 | 667. | 0.05 | 994. | 0.06 | 0.000 | 157.495 | 0.292 | 0.435 | -0.143 |
| tot | 69804680. | 100.00 | 1085080. | 100.00 | 454188. | 100.00 | 1464297. | 100.00 | 1746948. | 100.00 | | | | | |
| gross | | | | | | | | | | | 1.157 | 2.025 | 0.293 | 0.382 | |
| crude(x1000) | | | | | | | | | | | 15.545 | 6.507 | 4.195 | 5.005 | -0.810 |
| m.age | | 30.40 | | 26.05 | | 63.52 | | 22.10 | | 25.86 | 26.63 | 78.09 | 29.50 | 35.17 | |
| e(0) | | | | | | | | | | | | 74.45 | | | |

* Note that the numbers of arrivals and departures cover a five-year period (1965–1970); all other numbers and rates refer to a single year (1968).

## APPENDIX A *(continued)*

Native-independent input data—continued*.
Data for the whole USA

| age | population number | population - % - | births number | births - % - | deaths number | deaths - % - | arrivals number | arrivals - % - | departures number | departures - % - | observed rates (x 1000) birth | death | inmig | outmig | net mig |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 10787344. | 10.55 | 0. | 0.00 | 54486. | 8.13 | 409118. | 12.74 | 409118. | 12.74 | 0.000 | 5.051 | 7.585 | 7.585 | 0.000 |
| 5 | 11391607. | 11.14 | 0. | 0.00 | 4359. | 0.65 | 326635. | 10.17 | 326635. | 10.17 | 0.000 | 0.383 | 5.735 | 5.735 | 0.000 |
| 10 | 10522474. | 10.29 | 4922. | 0.30 | 3176. | 0.47 | 310574. | 9.67 | 310574. | 9.67 | 0.468 | 0.302 | 5.903 | 5.903 | 0.000 |
| 15 | 8981439. | 8.78 | 279911. | 17.30 | 5279. | 0.79 | 545237. | 16.98 | 545237. | 16.98 | 31.165 | 0.588 | 12.141 | 12.141 | 0.000 |
| 20 | 7224009. | 7.06 | 554545. | 34.28 | 5015. | 0.75 | 435083. | 13.55 | 435083. | 13.55 | 76.764 | 0.694 | 12.045 | 12.045 | 0.000 |
| 25 | 6157843. | 6.02 | 402886. | 24.91 | 5133. | 0.77 | 267629. | 8.33 | 267629. | 8.33 | 65.426 | 0.834 | 8.692 | 8.692 | 0.000 |
| 30 | 6418244. | 6.28 | 229738. | 14.20 | 7706. | 1.15 | 196471. | 6.12 | 196471. | 6.12 | 35.795 | 1.201 | 6.122 | 6.122 | 0.000 |
| 35 | 6265910. | 6.13 | 112370. | 6.95 | 12287. | 1.83 | 152854. | 4.76 | 152854. | 4.76 | 17.934 | 1.961 | 4.879 | 4.879 | 0.000 |
| 40 | 6445360. | 6.30 | 31297. | 1.93 | 19148. | 2.86 | 123716. | 3.85 | 123716. | 3.85 | 4.856 | 2.971 | 3.839 | 3.839 | 0.000 |
| 45 | 6550952. | 6.41 | 1942. | 0.12 | 28396. | 4.24 | 92577. | 2.88 | 92577. | 2.88 | 0.296 | 4.335 | 2.826 | 2.826 | 0.000 |
| 50 | 5440207. | 5.32 | 0. | 0.00 | 34208. | 5.11 | 80580. | 2.51 | 80580. | 2.51 | 0.000 | 6.288 | 2.962 | 2.962 | 0.000 |
| 55 | 4926576. | 4.82 | 0. | 0.00 | 45343. | 6.77 | 89180. | 2.78 | 89180. | 2.78 | 0.000 | 9.204 | 3.620 | 3.620 | 0.000 |
| 60 | 3156774. | 3.09 | 0. | 0.00 | 42113. | 6.29 | 73261. | 2.28 | 73261. | 2.28 | 0.000 | 13.341 | 4.642 | 4.642 | 0.000 |
| 65 | 2663896. | 2.60 | 0. | 0.00 | 56129. | 8.38 | 48615. | 1.51 | 48615. | 1.51 | 0.000 | 21.070 | 3.650 | 3.650 | 0.000 |
| 70 | 2148305. | 2.10 | 0. | 0.00 | 71727. | 10.70 | 33172. | 1.03 | 33172. | 1.03 | 0.000 | 33.388 | 3.088 | 3.088 | 0.000 |
| 75 | 1561937. | 1.53 | 0. | 0.00 | 86836. | 12.96 | 19905. | 0.62 | 19905. | 0.62 | 0.000 | 55.595 | 2.549 | 2.549 | 0.000 |
| 80 | 967705. | 0.95 | 0. | 0.00 | 85421. | 12.75 | 4977. | 0.15 | 4977. | 0.15 | 0.000 | 88.272 | 1.029 | 1.029 | 0.000 |
| 85 | 666417. | 0.65 | 0. | 0.00 | 103271. | 15.41 | 1661. | 0.05 | 1661. | 0.05 | 0.000 | 154.965 | 0.498 | 0.498 | 0.000 |
| tot | 102277000. | 100.00 | 1617611. | 100.00 | 670033. | 100.00 | 3211245. | 100.00 | 3211245. | 100.00 | | | | | |
| gross | | | | | | | | | | | 1.164 | 2.002 | 0.459 | 0.459 | |
| crude(x1000) | | | | | | | | | | | 15.816 | 6.551 | 6.280 | 6.280 | 0.000 |
| m.age | | 30.53 | | 25.73 | | 63.08 | | 24.15 | | 24.15 | 26.36 | 77.91 | 32.54 | 32.54 | |
| e(0) | | | | | | | | | | | | 74.34 | | | |

* Note that the numbers of arrivals and departures cover a five-year period (1965–1970); all other numbers and rates refer to a single year (1968).

## APPENDIX B: NATIVE-INDEPENDENT MULTISTATE ANALYSIS

Notation used in Appendix B:

$1$ = region South
$2$ = region North
$q(x, i)$ = probability of dying in region $i$ between ages $x$ and $x+4$
$p(x, i, j)$ = probability of living in region $i$ at age $x+4$ having lived in region $j$ at age $x$
$l(x, i, j)$ = number of people living in region $i$ at exact age $x$ who were born in region $j$
$ll(x, i, j)$ = number of years lived in region $i$ between ages $x$ and $x+4$ by an individual born in region $j$
$m(x, i, j)$ = age-specific rates of migration from region $j$ to region $i$ for the life-table population
$md(x, i)$ = age-specific rates of dying in region $i$ for the life-table population
$s(x, i, j)$ = proportion of people aged $x$ to $x+4$ in region $j$ surviving to be in region $i$ aged $x+5$ to $x+9$ five years later
$e(x, i, j)$ = expectation of life at exact age $x$ of an individual born in region $j$ and living in region $i$

The computer output representing the growth matrix is labeled in the following manner. The block of data headed "region $i \to j$" refers to individuals born in region $i$ but resident in region $j$ at time $t$. Each block is divided into two main sections, headed "first row" and "survivorship proportions". The matrix headed "first row" provides information on the number of offspring born to individuals in the "region $i \to j$" category. The columns $k \to l$ describe the number of children born in region $k$ between time $t$ and $t+5$, and surviving in region $l$ at time $t+5$, for each individual aged $x$ in category "region $i \to j$". The matrix headed "survivorship proportions" gives the proportion of individuals born in region $i$ and resident in region $j$ at time $t$ who are living in region $m$ at time $t+5$. The columns are labeled $i \to m$; residence in region $j$ at time $t$ is implicit in the block heading. Note that two of the four columns in the blocks of survivorship proportions must be zero because only one region of birth is considered in each block.

## APPENDIX B *(continued)*

Native-independent biregional and multistate analysis.

Biregional (North–South) life table for 1968.

| age | q(x,1) | p(x,1,1) | p(x,2,1) | l(x,1,1) | l(x,2,1) | ll(x,1,1) | ll(x,2,1) | m(x,2,1) | md(x,1) | s(x,1,1) | s(x,2,1) | e(x,1,1) | e(x,2,1) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 0.026254 | 0.917796 | 0.055950 | 100000. | 0. | 4.79449 | 0.13987 | 0.011977 | 0.005332 | 0.935958 | 0.049674 | 53.81 | 20.34 |
| 5 | 0.002114 | 0.955917 | 0.041968 | 91780. | 5595. | 4.49081 | 0.37282 | 0.008685 | 0.000425 | 0.955671 | 0.042408 | 50.34 | 20.74 |
| 10 | 0.001730 | 0.955400 | 0.042870 | 87853. | 9318. | 4.29977 | 0.55462 | 0.008875 | 0.000348 | 0.932966 | 0.064580 | 45.82 | 20.40 |
| 15 | 0.003182 | 0.909020 | 0.087799 | 84138. | 12867. | 4.02897 | 0.81370 | 0.018835 | 0.000641 | 0.908391 | 0.088108 | 41.47 | 19.86 |
| 20 | 0.003830 | 0.907739 | 0.088430 | 77021. | 19681. | 3.69352 | 1.13255 | 0.018979 | 0.000773 | 0.920653 | 0.075086 | 37.43 | 19.09 |
| 25 | 0.004707 | 0.935336 | 0.059957 | 70720. | 25621. | 3.44241 | 1.36386 | 0.012627 | 0.000949 | 0.942773 | 0.051420 | 33.74 | 17.98 |
| 30 | 0.006930 | 0.951020 | 0.042050 | 66976. | 28933. | 3.28382 | 1.49600 | 0.008751 | 0.001397 | 0.953487 | 0.037473 | 30.30 | 16.64 |
| 35 | 0.011179 | 0.956178 | 0.032643 | 64377. | 30907. | 3.16321 | 1.57597 | 0.006772 | 0.002255 | 0.957411 | 0.028812 | 27.05 | 15.18 |
| 40 | 0.016409 | 0.958764 | 0.024827 | 62152. | 32132. | 3.05613 | 1.62128 | 0.005147 | 0.003315 | 0.959195 | 0.020903 | 23.98 | 13.67 |
| 45 | 0.023460 | 0.959691 | 0.016849 | 60093. | 32720. | 2.95411 | 1.63448 | 0.003497 | 0.004753 | 0.956670 | 0.015511 | 21.07 | 12.14 |
| 50 | 0.032276 | 0.953504 | 0.014220 | 58071. | 32659. | 2.84757 | 1.61731 | 0.002978 | 0.006564 | 0.947575 | 0.013182 | 18.30 | 10.62 |
| 55 | 0.046446 | 0.941315 | 0.012239 | 55832. | 32033. | 2.72515 | 1.56780 | 0.002605 | 0.009513 | 0.931320 | 0.013439 | 15.65 | 9.12 |
| 60 | 0.064451 | 0.920663 | 0.014885 | 53174. | 30679. | 2.57216 | 1.48528 | 0.003245 | 0.013319 | 0.904106 | 0.014378 | 13.15 | 7.69 |
| 65 | 0.099752 | 0.886180 | 0.014068 | 49712. | 28732. | 2.35660 | 1.36962 | 0.003170 | 0.020997 | 0.862672 | 0.014670 | 10.78 | 6.32 |
| 70 | 0.148032 | 0.836188 | 0.015780 | 44552. | 26053. | 2.05271 | 1.21038 | 0.003754 | 0.031955 | 0.798310 | 0.013793 | 8.64 | 5.09 |
| 75 | 0.234636 | 0.753037 | 0.012326 | 37556. | 22363. | 1.65070 | 0.98608 | 0.003227 | 0.053139 | 0.706043 | 0.008502 | 6.76 | 3.97 |
| 80 | 0.351842 | 0.643685 | 0.004474 | 28472. | 17080. | 1.17125 | 0.69970 | 0.001335 | 0.085379 | 1.043796 | 0.008439 | 5.26 | 3.06 |
| 85 | 1.000000 | 0.000000 | 0.000000 | 18378. | 10907. | 1.22661 | 0.69560 | 0.000637 | 0.149437 | 0.000000 | 0.000000 | 4.19 | 2.38 |

| age | q(x,2) | p(x,2,2) | p(x,1,2) | l(x,2,2) | l(x,1,2) | ll(x,2,2) | ll(x,1,2) | m(x,1,2) | md(x,2) | s(x,2,2) | s(x,1,2) | e(x,2,2) | e(x,1,2) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 0.024352 | 0.949396 | 0.026252 | 100000. | 0. | 4.87349 | 0.06563 | 0.005620 | 0.004925 | 0.962689 | 0.024097 | 63.16 | 11.29 |
| 5 | 0.001820 | 0.976904 | 0.021276 | 94940. | 2625. | 4.69492 | 0.17886 | 0.004403 | 0.000364 | 0.976834 | 0.021552 | 59.74 | 11.50 |
| 10 | 0.001405 | 0.976753 | 0.021842 | 92857. | 4529. | 4.59374 | 0.27212 | 0.004522 | 0.000281 | 0.966459 | 0.031430 | 55.03 | 11.34 |
| 15 | 0.002817 | 0.955401 | 0.041781 | 90893. | 6355. | 4.45724 | 0.39826 | 0.008963 | 0.000563 | 0.955595 | 0.041349 | 50.39 | 11.07 |
| 20 | 0.003293 | 0.955783 | 0.040924 | 87397. | 9575. | 4.29440 | 0.54608 | 0.008783 | 0.000657 | 0.959360 | 0.037042 | 45.93 | 10.69 |
| 25 | 0.003899 | 0.963750 | 0.032351 | 84379. | 12268. | 4.16088 | 0.66182 | 0.006813 | 0.000779 | 0.967108 | 0.028165 | 41.64 | 10.17 |
| 30 | 0.005548 | 0.970906 | 0.023546 | 82056. | 14205. | 4.05805 | 0.74114 | 0.004900 | 0.001109 | 0.971198 | 0.021478 | 37.49 | 9.52 |
| 35 | 0.009101 | 0.971616 | 0.019283 | 80266. | 15441. | 3.96894 | 0.79382 | 0.004000 | 0.001824 | 0.970921 | 0.017534 | 33.47 | 8.80 |
| 40 | 0.014008 | 0.970285 | 0.015707 | 78492. | 16312. | 3.87640 | 0.82961 | 0.003256 | 0.002817 | 0.968736 | 0.013995 | 29.60 | 8.05 |
| 45 | 0.020573 | 0.967205 | 0.012222 | 76564. | 16872. | 3.77255 | 0.85000 | 0.002536 | 0.004154 | 0.961465 | 0.013131 | 25.88 | 7.28 |
| 50 | 0.030341 | 0.955548 | 0.014110 | 74338. | 17128. | 3.64036 | 0.86271 | 0.002955 | 0.006159 | 0.946179 | 0.016611 | 22.32 | 6.50 |
| 55 | 0.044290 | 0.936389 | 0.019321 | 71277. | 17380. | 3.45581 | 0.87795 | 0.004112 | 0.009054 | 0.924003 | 0.021793 | 18.92 | 5.74 |
| 60 | 0.064600 | 0.910747 | 0.024653 | 66955. | 17738. | 3.20497 | 0.89297 | 0.005375 | 0.013352 | 0.897232 | 0.020939 | 15.72 | 4.97 |
| 65 | 0.100253 | 0.882401 | 0.017345 | 61243. | 17981. | 2.88844 | 0.87445 | 0.003908 | 0.021110 | 0.858492 | 0.014408 | 12.76 | 4.18 |
| 70 | 0.157026 | 0.831374 | 0.011600 | 54294. | 16997. | 2.49253 | 0.79598 | 0.002760 | 0.034095 | 0.791290 | 0.009920 | 10.13 | 3.42 |
| 75 | 0.248372 | 0.743096 | 0.008532 | 45407. | 14842. | 1.98330 | 0.66016 | 0.002234 | 0.056736 | 0.695346 | 0.005863 | 7.85 | 2.73 |
| 80 | 0.365879 | 0.631134 | 0.002987 | 33925. | 11564. | 1.38469 | 0.47773 | 0.000891 | 0.089567 | 0.980017 | 0.005810 | 6.04 | 2.16 |
| 85 | 1.000000 | 0.000000 | 0.000000 | 21463. | 7545. | 1.36105 | 0.50670 | 0.000435 | 0.157495 | 0.000000 | 0.000000 | 4.69 | 1.75 |

Biregional population projection to 1998.

**population**

| age | total | south | north |
|---|---|---|---|
| 0 | 11191921. | 3665005. | 7526916. |
| 5 | 11330953. | 3653294. | 7677659. |
| 10 | 11523384. | 3719317. | 7804067. |
| 15 | 10917141. | 3559833. | 7357308. |
| 20 | 9664246. | 3166043. | 6498203. |
| 25 | 8365251. | 2744050. | 5621202. |
| 30 | 10471771. | 3335232. | 7136539. |
| 35 | 11122236. | 3573490. | 7548746. |
| 40 | 10164837. | 3315739. | 6849099. |
| 45 | 8537646. | 2822215. | 5715432. |
| 50 | 6708680. | 2231955. | 4476725. |
| 55 | 5522833. | 1882367. | 3640466. |
| 60 | 5470576. | 1885838. | 3584738. |
| 65 | 4943318. | 1740152. | 3203166. |
| 70 | 4501880. | 1570753. | 2931126. |
| 75 | 3750906. | 1300461. | 2450444. |
| 80 | 2258454. | 819373. | 1439081. |
| 85 | 2146892. | 813921. | 1332971. |
| total | 138592928. | 45799032. | 92793888. |

**percentage distribution**

| age | total | south | north |
|---|---|---|---|
| 0 | 8.0754 | 8.0024 | 8.1114 |
| 5 | 8.1757 | 7.9768 | 8.2739 |
| 10 | 8.3146 | 8.1210 | 8.4101 |
| 15 | 7.8771 | 7.7727 | 7.9287 |
| 20 | 6.9731 | 6.9129 | 7.0028 |
| 25 | 6.0358 | 5.9915 | 6.0577 |
| 30 | 7.5558 | 7.2823 | 7.6907 |
| 35 | 8.0251 | 7.8025 | 8.1350 |
| 40 | 7.3343 | 7.2398 | 7.3810 |
| 45 | 6.1602 | 6.1622 | 6.1593 |
| 50 | 4.8406 | 4.8734 | 4.8244 |
| 55 | 3.9849 | 4.1101 | 3.9232 |
| 60 | 3.9472 | 4.1176 | 3.8631 |
| 65 | 3.5668 | 3.7995 | 3.4519 |
| 70 | 3.2483 | 3.4297 | 3.1587 |
| 75 | 2.7064 | 2.8395 | 2.6407 |
| 80 | 1.6296 | 1.7891 | 1.5508 |
| 85 | 1.5491 | 1.7772 | 1.4365 |
| total | 100.0000 | 100.0000 | 100.0000 |
| m.ag | 34.8294 | 35.4515 | 34.5224 |
| sha | 100.0000 | 33.0457 | 66.9543 |
| lam | 1.040688 | 1.045199 | 1.038476 |
| r | 0.007976 | 0.008841 | 0.007551 |

Stable equivalent population.

**stable equivalent to original population**

| age | total | south | north |
|---|---|---|---|
| 0 | 10085119. | 3412733. | 6672386. |
| 5 | 9733339. | 3282591. | 6450748. |
| 10 | 9507037. | 3205440. | 6301597. |
| 15 | 9281257. | 3119844. | 6161413. |
| 20 | 9051945. | 3022177. | 6029768. |
| 25 | 8822863. | 2940894. | 5881969. |
| 30 | 8588644. | 2874887. | 5713757. |
| 35 | 8337015. | 2802114. | 5534902. |
| 40 | 8056895. | 2719865. | 5337030. |
| 45 | 7739962. | 2625686. | 5114277. |
| 50 | 7374423. | 2523439. | 4850984. |
| 55 | 6941856. | 2418413. | 4523444. |
| 60 | 6421507. | 2300187. | 4121320. |
| 65 | 5769569. | 2119189. | 3650379. |
| 70 | 4936833. | 1840190. | 3096642. |
| 75 | 3889733. | 1467411. | 2422322. |
| 80 | 2687830. | 1027604. | 1660227. |
| 85 | 2659351. | 1058911. | 1600440. |
| total | 129885200. | 44761572. | 85123616. |

**percentage distribution**

| age | total | south | north |
|---|---|---|---|
| 0 | 7.7646 | 7.6242 | 7.8385 |
| 5 | 7.4938 | 7.3335 | 7.5781 |
| 10 | 7.3196 | 7.1611 | 7.4029 |
| 15 | 7.1457 | 6.9699 | 7.2382 |
| 20 | 6.9692 | 6.7517 | 7.0835 |
| 25 | 6.7928 | 6.5701 | 6.9099 |
| 30 | 6.6125 | 6.4227 | 6.7123 |
| 35 | 6.4188 | 6.2601 | 6.5022 |
| 40 | 6.2031 | 6.0763 | 6.2697 |
| 45 | 5.9591 | 5.8659 | 6.0081 |
| 50 | 5.6776 | 5.6375 | 5.6988 |
| 55 | 5.3446 | 5.4029 | 5.3140 |
| 60 | 4.9440 | 5.1388 | 4.8416 |
| 65 | 4.4421 | 4.7344 | 4.2883 |
| 70 | 3.8009 | 4.1111 | 3.6378 |
| 75 | 2.9947 | 3.2783 | 2.8457 |
| 80 | 2.0694 | 2.2957 | 1.9504 |
| 85 | 2.0475 | 2.3657 | 1.8801 |
| total | 100.0000 | 100.0000 | 100.0000 |
| m.ag | 37.0983 | 37.9440 | 36.6536 |
| sha | 100.0000 | 34.4624 | 65.5376 |
| lam | 1.022046 | 1.022046 | 1.022046 |
| r | 0.004361 | 0.004361 | 0.004361 |

Biregional population projection: stable equivalent components and intrinsic rates.

| | births | | deaths | | outmigration | | inmigration | |
|---|---|---|---|---|---|---|---|---|
| | number | rate | number | rate | number | rate | number | rate |
| south | 700743. | 0.015659 | 554107. | 0.012383 | 359335. | 0.008030 | 407843. | 0.009114 |
| north | 1363633. | 0.016024 | 943958. | 0.011092 | 407843. | 0.004793 | 359335. | 0.004223 |
| total | 2064376. | 0.015898 | 1498065. | 0.011537 | 767178. | 0.005908 | 767178. | 0.005908 |

stable growth rate 0.004361

normalizing factor 74.0755

**APPENDIX B** *(continued)*

**Biregional Model for Native-independent Multistate Projection**

Expressing each set of four age-specific birth rates defined in eqns. (4) and (5) in the form of a matrix, and ignoring the place of birth of the parent (the subscript on the left-hand side) gives

$$ {}_{\mathrm{s}}\mathbf{B}^{\mathrm{s}}(x) = {}_{\mathrm{n}}\mathbf{B}^{\mathrm{s}}(x) = {}_{\cdot}\mathbf{B}^{\mathrm{s}}(x) = \begin{bmatrix} b^{\mathrm{s}}_{\mathrm{ss}}(x) & b^{\mathrm{s}}_{\mathrm{ns}}(x) \\ b^{\mathrm{s}}_{\mathrm{sn}}(x) & b^{\mathrm{s}}_{\mathrm{nn}}(x) \end{bmatrix} \tag{B1} $$

$$ {}_{\mathrm{s}}\mathbf{B}^{\mathrm{n}}(x) = {}_{\mathrm{n}}\mathbf{B}^{\mathrm{n}}(x) = {}_{\cdot}\mathbf{B}^{\mathrm{n}}(x) = \begin{bmatrix} b^{\mathrm{n}}_{\mathrm{ss}}(x) & b^{\mathrm{n}}_{\mathrm{ns}}(x) \\ b^{\mathrm{n}}_{\mathrm{sn}}(x) & b^{\mathrm{n}}_{\mathrm{nn}}(x) \end{bmatrix} \tag{B2} $$

Setting out the corresponding survivorship proportions* as the matrix

$$ {}_{\mathrm{s}}\mathbf{S}(x) = {}_{\mathrm{n}}\mathbf{S}(x) = {}_{\cdot}\mathbf{S}(x) = \begin{bmatrix} s_{\mathrm{ss}}(x) & s_{\mathrm{ns}}(x) \\ s_{\mathrm{sn}}(x) & s_{\mathrm{nn}}(x) \end{bmatrix} \tag{B3} $$

with the place-of-birth dependence suppressed once again, we obtain the usual population growth process defined by

$$ \{\mathbf{K}^{(t+1)}\} = \mathbf{G}\{\mathbf{K}^{(t)}\} \tag{B4} $$

where

$$ \mathbf{G} = \begin{bmatrix} \mathbf{B}(0) & \mathbf{B}(5) & \cdots \\ \mathbf{S}(0) & \mathbf{0} & \cdots \\ \mathbf{0} & \mathbf{S}(5) & \\ \vdots & & \ddots \end{bmatrix} \tag{B5} $$

$$ \mathbf{B}(x) = \begin{bmatrix} {}_{\cdot}\mathbf{B}^{\mathrm{s}}(x) & {}_{\cdot}\mathbf{B}^{\mathrm{s}}(x) \\ {}_{\cdot}\mathbf{B}^{\mathrm{n}}(x) & {}_{\cdot}\mathbf{B}^{\mathrm{n}}(x) \end{bmatrix} \qquad \mathbf{S}(x) = \begin{bmatrix} {}_{\cdot}\mathbf{S}(x) & \mathbf{0} \\ \mathbf{0} & {}_{\cdot}\mathbf{S}(x) \end{bmatrix} $$

$$ \{\mathbf{K}^{(t)}\} = \begin{bmatrix} \{\mathbf{K}^{(t)}(0)\} \\ \{\mathbf{K}^{(t)}(5)\} \\ \vdots \end{bmatrix} \qquad \{\mathbf{K}^{(t)}(x)\} = \begin{bmatrix} {}_{\mathrm{s}}K^{(t)}_{\mathrm{s}}(x) \\ {}_{\mathrm{s}}K^{(t)}_{\mathrm{n}}(x) \\ {}_{\mathrm{n}}K^{(t)}_{\mathrm{s}}(x) \\ {}_{\mathrm{n}}K^{(t)}_{\mathrm{n}}(x) \end{bmatrix} $$

The extension to the native-dependent case is straightforward. The place-of-birth subscript on the left-hand side is no longer suppressed and

$$ \mathbf{B}(x) = \begin{bmatrix} {}_{\mathrm{s}}\mathbf{B}^{\mathrm{s}}(x) & {}_{\mathrm{n}}\mathbf{B}^{\mathrm{s}}(x) \\ {}_{\mathrm{s}}\mathbf{B}^{\mathrm{n}}(x) & {}_{\mathrm{n}}\mathbf{B}^{\mathrm{n}}(x) \end{bmatrix} \qquad \mathbf{S}(x) = \begin{bmatrix} {}_{\mathrm{s}}\mathbf{S}(x) & \mathbf{0} \\ \mathbf{0} & {}_{\mathrm{n}}\mathbf{S}(x) \end{bmatrix} \tag{B6} $$

---

* Survivorship proportions are defined in the normal way (Rogers 1975, p. 79) as

$$ \mathbf{S}(x) = \mathbf{L}(x+5)\mathbf{L}^{-1}(x) $$

Native-independent multistate projection: growth matrix (1968). (See p. 65 for explanation of labeling.)

region s -> s

| age | first row s -> s | s -> n | n -> s | n -> n |
|---|---|---|---|---|
| 0 | 0.000000 | 0.000000 | 0.000000 | 0.000000 |
| 5 | 0.001802 | 0.000053 | 0.000000 | 0.000033 |
| 10 | 0.090356 | 0.002636 | 0.000058 | 0.004277 |
| 15 | 0.268101 | 0.007822 | 0.000218 | 0.016198 |
| 20 | 0.325997 | 0.009511 | 0.000166 | 0.012333 |
| 25 | 0.221237 | 0.006454 | 0.000063 | 0.004654 |
| 30 | 0.116658 | 0.003403 | 0.000023 | 0.001696 |
| 35 | 0.050277 | 0.001467 | 0.000005 | 0.000348 |
| 40 | 0.011769 | 0.000343 | 0.000000 | 0.000015 |
| 45 | 0.000708 | 0.000021 | 0.000000 | 0.000000 |
| 50 | 0.000000 | 0.000000 | 0.000000 | 0.000000 |
| 55 | 0.000000 | 0.000000 | 0.000000 | 0.000000 |
| 60 | 0.000000 | 0.000000 | 0.000000 | 0.000000 |
| 65 | 0.000000 | 0.000000 | 0.000000 | 0.000000 |
| 70 | 0.000000 | 0.000000 | 0.000000 | 0.000000 |
| 75 | 0.000000 | 0.000000 | 0.000000 | 0.000000 |
| 80 | 0.000000 | 0.000000 | 0.000000 | 0.000000 |

| age | survivorship proportions s -> s | s -> n | n -> s | n -> n |
|---|---|---|---|---|
| 0 | 0.935958 | 0.049674 | 0.000000 | 0.000000 |
| 5 | 0.955671 | 0.042408 | 0.000000 | 0.000000 |
| 10 | 0.932966 | 0.064580 | 0.000000 | 0.000000 |
| 15 | 0.908390 | 0.088108 | 0.000000 | 0.000000 |
| 20 | 0.920653 | 0.075086 | 0.000000 | 0.000000 |
| 25 | 0.942773 | 0.051420 | 0.000000 | 0.000000 |
| 30 | 0.953487 | 0.037474 | 0.000000 | 0.000000 |
| 35 | 0.957411 | 0.028812 | 0.000000 | 0.000000 |
| 40 | 0.959195 | 0.020903 | 0.000000 | 0.000000 |
| 45 | 0.956670 | 0.015511 | 0.000000 | 0.000000 |
| 50 | 0.947575 | 0.013182 | 0.000000 | 0.000000 |
| 55 | 0.931320 | 0.013439 | 0.000000 | 0.000000 |
| 60 | 0.904106 | 0.014379 | 0.000000 | 0.000000 |
| 65 | 0.862672 | 0.014670 | 0.000000 | 0.000000 |
| 70 | 0.798310 | 0.013793 | 0.000000 | 0.000000 |
| 75 | 0.706043 | 0.008502 | 0.000000 | 0.000000 |
| 80 | 1.043797 | 0.008439 | 0.000000 | 0.000000 |

region s -> n

| age | first row s -> s | s -> n | n -> s | n -> n |
|---|---|---|---|---|
| 0 | 0.000000 | 0.000000 | 0.000000 | 0.000000 |
| 5 | 0.000041 | 0.000001 | 0.000010 | 0.000761 |
| 10 | 0.002980 | 0.000087 | 0.000872 | 0.064790 |
| 15 | 0.007887 | 0.000230 | 0.003258 | 0.241907 |
| 20 | 0.005442 | 0.000159 | 0.004598 | 0.341414 |
| 25 | 0.002221 | 0.000065 | 0.003391 | 0.251789 |
| 30 | 0.000852 | 0.000025 | 0.001811 | 0.134482 |
| 35 | 0.000194 | 0.000006 | 0.000768 | 0.057001 |
| 40 | 0.000010 | 0.000000 | 0.000172 | 0.012782 |
| 45 | 0.000000 | 0.000000 | 0.000010 | 0.000724 |
| 50 | 0.000000 | 0.000000 | 0.000000 | 0.000000 |
| 55 | 0.000000 | 0.000000 | 0.000000 | 0.000000 |
| 60 | 0.000000 | 0.000000 | 0.000000 | 0.000000 |
| 65 | 0.000000 | 0.000000 | 0.000000 | 0.000000 |
| 70 | 0.000000 | 0.000000 | 0.000000 | 0.000000 |
| 75 | 0.000000 | 0.000000 | 0.000000 | 0.000000 |
| 80 | 0.000000 | 0.000000 | 0.000000 | 0.000000 |

| age | survivorship proportions s -> s | s -> n | n -> s | n -> n |
|---|---|---|---|---|
| 0 | 0.024097 | 0.962689 | 0.000000 | 0.000000 |
| 5 | 0.021552 | 0.976834 | 0.000000 | 0.000000 |
| 10 | 0.031430 | 0.966459 | 0.000000 | 0.000000 |
| 15 | 0.041349 | 0.955595 | 0.000000 | 0.000000 |
| 20 | 0.037042 | 0.959360 | 0.000000 | 0.000000 |
| 25 | 0.028165 | 0.967107 | 0.000000 | 0.000000 |
| 30 | 0.021478 | 0.971198 | 0.000000 | 0.000000 |
| 35 | 0.017534 | 0.970921 | 0.000000 | 0.000000 |
| 40 | 0.013995 | 0.968736 | 0.000000 | 0.000000 |
| 45 | 0.013131 | 0.961465 | 0.000000 | 0.000000 |
| 50 | 0.016611 | 0.946179 | 0.000000 | 0.000000 |
| 55 | 0.021793 | 0.924003 | 0.000000 | 0.000000 |
| 60 | 0.020939 | 0.897232 | 0.000000 | 0.000000 |
| 65 | 0.014408 | 0.858492 | 0.000000 | 0.000000 |
| 70 | 0.009920 | 0.791290 | 0.000000 | 0.000000 |
| 75 | 0.005863 | 0.695346 | 0.000000 | 0.000000 |
| 80 | 0.005810 | 0.980017 | 0.000000 | 0.000000 |

## APPENDIX B *(continued)*

Native-independent multistate projection: growth matrix (1968)—continued.

**region n -> s**

| age | first row s -> s | s -> n | n -> s | n -> n |
|---|---|---|---|---|
| 0 | 0.000000 | 0.000000 | 0.000000 | 0.000000 |
| 5 | 0.001802 | 0.000053 | 0.000000 | 0.000033 |
| 10 | 0.090356 | 0.002636 | 0.000058 | 0.004277 |
| 15 | 0.268101 | 0.007822 | 0.000218 | 0.016198 |
| 20 | 0.325997 | 0.009511 | 0.000166 | 0.012333 |
| 25 | 0.221237 | 0.006454 | 0.000063 | 0.004654 |
| 30 | 0.116658 | 0.003403 | 0.000023 | 0.001696 |
| 35 | 0.050277 | 0.001467 | 0.000005 | 0.000348 |
| 40 | 0.011769 | 0.000343 | 0.000000 | 0.000015 |
| 45 | 0.000708 | 0.000021 | -0.000000 | 0.000000 |
| 50 | 0.000000 | 0.000000 | 0.000000 | 0.000000 |
| 55 | 0.000000 | 0.000000 | 0.000000 | 0.000000 |
| 60 | 0.000000 | 0.000000 | 0.000000 | 0.000000 |
| 65 | 0.000000 | 0.000000 | 0.000000 | 0.000000 |
| 70 | 0.000000 | 0.000000 | 0.000000 | 0.000000 |
| 75 | 0.000000 | 0.000000 | 0.000000 | 0.000000 |
| 80 | 0.000000 | 0.000000 | 0.000000 | 0.000000 |

| age | survivorship proportions s -> s | s -> n | n -> s | n -> n |
|---|---|---|---|---|
| 0 | 0.000000 | 0.000000 | 0.935958 | 0.049674 |
| 5 | 0.000000 | 0.000000 | 0.955671 | 0.042408 |
| 10 | 0.000000 | 0.000000 | 0.932966 | 0.064580 |
| 15 | 0.000000 | 0.000000 | 0.908390 | 0.088108 |
| 20 | 0.000000 | 0.000000 | 0.920653 | 0.075086 |
| 25 | 0.000000 | 0.000000 | 0.942773 | 0.051420 |
| 30 | 0.000000 | 0.000000 | 0.953487 | 0.037474 |
| 35 | 0.000000 | 0.000000 | 0.957411 | 0.028812 |
| 40 | 0.000000 | 0.000000 | 0.959195 | 0.020903 |
| 45 | 0.000000 | 0.000000 | 0.956670 | 0.015511 |
| 50 | 0.000000 | 0.000000 | 0.947575 | 0.013182 |
| 55 | 0.000000 | 0.000000 | 0.931320 | 0.013439 |
| 60 | 0.000000 | 0.000000 | 0.904106 | 0.014379 |
| 65 | 0.000000 | 0.000000 | 0.862672 | 0.014670 |
| 70 | 0.000000 | 0.000000 | 0.798310 | 0.013793 |
| 75 | 0.000000 | 0.000000 | 0.706043 | 0.008502 |
| 80 | 0.000000 | 0.000000 | 1.043797 | 0.008439 |

**region n -> n**

| age | first row s -> s | s -> n | n -> s | n -> n |
|---|---|---|---|---|
| 0 | 0.000000 | 0.000000 | 0.000000 | 0.000000 |
| 5 | 0.000041 | 0.000001 | 0.000010 | 0.000761 |
| 10 | 0.002980 | 0.000087 | 0.000872 | 0.064790 |
| 15 | 0.007887 | 0.000230 | 0.003258 | 0.241907 |
| 20 | 0.005442 | 0.000159 | 0.004598 | 0.341414 |
| 25 | 0.002221 | 0.000065 | 0.003391 | 0.251789 |
| 30 | 0.000852 | 0.000025 | 0.001811 | 0.134482 |
| 35 | 0.000194 | 0.000006 | 0.000768 | 0.057001 |
| 40 | 0.000010 | 0.000000 | 0.000172 | 0.012782 |
| 45 | 0.000000 | 0.000000 | 0.000010 | 0.000724 |
| 50 | 0.000000 | 0.000000 | 0.000000 | 0.000000 |
| 55 | 0.000000 | 0.000000 | 0.000000 | 0.000000 |
| 60 | 0.000000 | 0.000000 | 0.000000 | 0.000000 |
| 65 | 0.000000 | 0.000000 | 0.000000 | 0.000000 |
| 70 | 0.000000 | 0.000000 | 0.000000 | 0.000000 |
| 75 | 0.000000 | 0.000000 | 0.000000 | 0.000000 |
| 80 | 0.000000 | 0.000000 | 0.000000 | 0.000000 |

| age | survivorship proportions s -> s | s -> n | n -> s | n -> n |
|---|---|---|---|---|
| 0 | 0.000000 | 0.000000 | 0.024097 | 0.962689 |
| 5 | 0.000000 | 0.000000 | 0.021552 | 0.976834 |
| 10 | 0.000000 | 0.000000 | 0.031430 | 0.966459 |
| 15 | 0.000000 | 0.000000 | 0.041349 | 0.955595 |
| 20 | 0.000000 | 0.000000 | 0.037042 | 0.959360 |
| 25 | 0.000000 | 0.000000 | 0.028165 | 0.967107 |
| 30 | 0.000000 | 0.000000 | 0.021478 | 0.971198 |
| 35 | 0.000000 | 0.000000 | 0.017534 | 0.970921 |
| 40 | 0.000000 | 0.000000 | 0.013995 | 0.968736 |
| 45 | 0.000000 | 0.000000 | 0.013131 | 0.961465 |
| 50 | 0.000000 | 0.000000 | 0.016611 | 0.946179 |
| 55 | 0.000000 | 0.000000 | 0.021793 | 0.924003 |
| 60 | 0.000000 | 0.000000 | 0.020939 | 0.897232 |
| 65 | 0.000000 | 0.000000 | 0.014408 | 0.858492 |
| 70 | 0.000000 | 0.000000 | 0.009920 | 0.791290 |
| 75 | 0.000000 | 0.000000 | 0.005863 | 0.695346 |
| 80 | 0.000000 | 0.000000 | 0.005810 | 0.980017 |

Native-independent multistate projection: stable equivalent population*.

stable equivalent to original population

| age | total | s -> s | s -> n | n -> s | n -> n | south | north |
|---|---|---|---|---|---|---|---|
| 0 | 10085108. | 3324184. | 96980. | 88548. | 6575398. | 3412731. | 6672377. |
| 5 | 9733327. | 3046470. | 252910. | 236119. | 6197828. | 3282589. | 6450738. |
| 10 | 9507027. | 2853957. | 368129. | 351481. | 5933459. | 3205438. | 6301589. |
| 15 | 9281247. | 2616531. | 528440. | 503311. | 5632966. | 3119842. | 6161405. |
| 20 | 9051934. | 2346942. | 719647. | 675233. | 5310112. | 3022175. | 6029760. |
| 25 | 8822855. | 2140195. | 847929. | 800698. | 5034033. | 2940893. | 5881962. |
| 30 | 8588633. | 1997563. | 910026. | 877321. | 4803722. | 2874884. | 5713748. |
| 35 | 8337005. | 1882691. | 937992. | 919421. | 4596901. | 2802112. | 5534893. |
| 40 | 8056885. | 1779722. | 944147. | 940141. | 4392876. | 2719863. | 5337022. |
| 45 | 7739954. | 1683207. | 931299. | 942478. | 4182971. | 2625685. | 5114270. |
| 50 | 7374416. | 1587504. | 901642. | 935933. | 3949336. | 2523437. | 4850978. |
| 55 | 6941849. | 1486486. | 855188. | 931926. | 3668249. | 2418412. | 4523437. |
| 60 | 6421499. | 1372767. | 792697. | 927419. | 3328617. | 2300186. | 4121313. |
| 65 | 5769562. | 1230596. | 715204. | 888592. | 2935169. | 2119188. | 3650373. |
| 70 | 4936827. | 1048784. | 618416. | 791406. | 2478221. | 1840190. | 3096638. |
| 75 | 3889730. | 825197. | 492945. | 642214. | 1929374. | 1467411. | 2422319. |
| 80 | 2687827. | 572886. | 342238. | 454718. | 1317986. | 1027603. | 1660224. |
| 85 | 2659347. | 587023. | 332895. | 471887. | 1267543. | 1058910. | 1600437. |
| total | 129885040. | 32382706. | 11588724. | 12378844. | 73534760. | 44761552. | 85123488. |

percentage distribution

| age | total | s -> s | s -> n | n -> s | n -> n | south | north |
|---|---|---|---|---|---|---|---|
| 0 | 7.7646 | 10.2653 | 0.8368 | 0.7153 | 8.9419 | 7.6242 | 7.8385 |
| 5 | 7.4938 | 9.4077 | 2.1824 | 1.9074 | 8.4284 | 7.3335 | 7.5781 |
| 10 | 7.3196 | 8.8132 | 3.1766 | 2.8394 | 8.0689 | 7.1611 | 7.4029 |
| 15 | 7.1457 | 8.0800 | 4.5599 | 4.0659 | 7.6603 | 6.9699 | 7.2382 |
| 20 | 6.9692 | 7.2475 | 6.2099 | 5.4547 | 7.2212 | 6.7517 | 7.0835 |
| 25 | 6.7928 | 6.6091 | 7.3168 | 6.4683 | 6.8458 | 6.5701 | 6.9099 |
| 30 | 6.6125 | 6.1686 | 7.8527 | 7.0873 | 6.5326 | 6.4227 | 6.7123 |
| 35 | 6.4188 | 5.8139 | 8.0940 | 7.4274 | 6.2513 | 6.2601 | 6.5022 |
| 40 | 6.2031 | 5.4959 | 8.1471 | 7.5947 | 5.9739 | 6.0763 | 6.2697 |
| 45 | 5.9591 | 5.1979 | 8.0363 | 7.6136 | 5.6884 | 5.8659 | 6.0081 |
| 50 | 5.6776 | 4.9023 | 7.7803 | 7.5607 | 5.3707 | 5.6375 | 5.6988 |
| 55 | 5.3446 | 4.5904 | 7.3795 | 7.5284 | 4.9885 | 5.4029 | 5.3140 |
| 60 | 4.9440 | 4.2392 | 6.8402 | 7.4920 | 4.5266 | 5.1388 | 4.8416 |
| 65 | 4.4421 | 3.8002 | 6.1716 | 7.1783 | 3.9915 | 4.7344 | 4.2883 |
| 70 | 3.8009 | 3.2387 | 5.3364 | 6.3932 | 3.3701 | 4.1111 | 3.6378 |
| 75 | 2.9947 | 2.5483 | 4.2537 | 5.1880 | 2.6238 | 3.2783 | 2.8457 |
| 80 | 2.0694 | 1.7691 | 2.9532 | 3.6733 | 1.7923 | 2.2957 | 1.9504 |
| 85 | 2.0475 | 1.8128 | 2.8726 | 3.8120 | 1.7237 | 2.3657 | 1.8801 |
| total | 100.0000 | 100.0000 | 100.0000 | 100.0000 | 100.0000 | 100.0000 | 100.0000 |
| m.ag | 37.0983 | 33.7122 | 46.5403 | 49.0142 | 35.0955 | 37.9440 | 36.6536 |
| sha | 200.0000 | 24.9318 | 8.9223 | 9.5306 | 56.6153 | 34.4624 | 65.5376 |
| lam | 1.022046 | 1.022046 | 1.022045 | 1.022045 | 1.022046 | 1.022045 | 1.022046 |
| r | 0.004361 | 0.004361 | 0.004361 | 0.004361 | 0.004361 | 0.004361 | 0.004361 |

*The first letter in the heading of each column refers to region of birth, the second to region of residence.

## APPENDIX B (*continued*)

Native-independent multistate projection: 1998*.

population

| age | total | s -> s | s -> n | n -> s | n -> n | south | north |
|---|---|---|---|---|---|---|---|
| 0 | 11191920. | 3565044. | 104007. | 99961. | 7422908. | 3665005. | 7526915. |
| 5 | 11330953. | 3371462. | 279890. | 281833. | 7397768. | 3653295. | 7677658. |
| 10 | 11523383. | 3282105. | 423356. | 437212. | 7380711. | 3719317. | 7804066. |
| 15 | 10917140. | 2955789. | 596957. | 604044. | 6760351. | 3559833. | 7357308. |
| 20 | 9664246. | 2434663. | 746544. | 731380. | 5751659. | 3166043. | 6498204. |
| 25 | 8365251. | 1974378. | 782234. | 769672. | 4838967. | 2744050. | 5621201. |
| 30 | 10471771. | 2155525. | 1113533. | 1179707. | 6023006. | 3335233. | 7136539. |
| 35 | 11122237. | 2303537. | 1213501. | 1269953. | 6335245. | 3573490. | 7548747. |
| 40 | 10164838. | 2163192. | 1154347. | 1152548. | 5694752. | 3315739. | 6849039. |
| 45 | 8537646. | 1865530. | 1009215. | 956685. | 4706216. | 2822215. | 5715431. |
| 50 | 6708679. | 1547194. | 835889. | 684761. | 3640835. | 2231955. | 4476725. |
| 55 | 5522834. | 1375220. | 672039. | 507147. | 2968426. | 1882368. | 3640466. |
| 60 | 5470576. | 1382930. | 601518. | 502908. | 2983219. | 1885838. | 3584737. |
| 65 | 4943318. | 1276115. | 512907. | 464037. | 2690258. | 1740152. | 3203166. |
| 70 | 4501880. | 1139681. | 429441. | 431072. | 2501685: | 1570753. | 2931127. |
| 75 | 3750906. | 943571. | 354143. | 356891. | 2096301. | 1300461. | 2450444. |
| 80 | 2258454. | 610544. | 200112. | 208828. | 1238969. | 819373. | 1439081. |
| 85 | 2146892. | 620479. | 183859. | 193442. | 1149112. | 813921. | 1332971. |
| total | 138592928. | 34966964. | 11213492. | 10832081. | 81580392. | 45799036. | 92793888. |

percentage distribution

| age | total | s -> s | s -> n | n -> s | n -> n | south | north |
|---|---|---|---|---|---|---|---|
| 0 | 8.0754 | 10.1955 | 0.9275 | 0.9228 | 9.0989 | 8.0024 | 8.1114 |
| 5 | 8.1757 | 9.6418 | 2.4960 | 2.6018 | 9.0681 | 7.9768 | 8.2739 |
| 10 | 8.3146 | 9.3863 | 3.7754 | 4.0363 | 9.0472 | 8.1210 | 8.4101 |
| 15 | 7.8771 | 8.4531 | 5.3236 | 5.5764 | 8.2867 | 7.7727 | 7.9287 |
| 20 | 6.9731 | 6.9628 | 6.6576 | 6.7520 | 7.0503 | 6.9129 | 7.0028 |
| 25 | 6.0358 | 5.6464 | 6.9758 | 7.1055 | 5.9315 | 5.9915 | 6.0577 |
| 30 | 7.5558 | 6.1645 | 9.9303 | 10.8909 | 7.3829 | 7.2823 | 7.6907 |
| 35 | 8.0251 | 6.5878 | 10.8218 | 11.7240 | 7.7656 | 7.8025 | 8.1350 |
| 40 | 7.3343 | 6.1864 | 10.2943 | 10.6401 | 6.9805 | 7.2398 | 7.3810 |
| 45 | 6.1602 | 5.3351 | 9.0000 | 8.8320 | 5.7688 | 6.1622 | 6.1593 |
| 50 | 4.8406 | 4.4247 | 7.4543 | 6.3216 | 4.4629 | 4.8734 | 4.8244 |
| 55 | 3.9849 | 3.9329 | 5.9931 | 4.6819 | 3.6387 | 4.1101 | 3.9232 |
| 60 | 3.9472 | 3.9550 | 5.3642 | 4.6428 | 3.6568 | 4.1176 | 3.8631 |
| 65 | 3.5668 | 3.6495 | 4.5740 | 4.2839 | 3.2977 | 3.7995 | 3.4519 |
| 70 | 3.2483 | 3.2593 | 3.8297 | 3.9796 | 3.0665 | 3.4297 | 3.1587 |
| 75 | 2.7064 | 2.6985 | 3.1582 | 3.2948 | 2.5696 | 2.8395 | 2.6407 |
| 80 | 1.6296 | 1.7461 | 1.7846 | 1.9279 | 1.5187 | 1.7891 | 1.5508 |
| 85 | 1.5491 | 1.7745 | 1.6396 | 1.7858 | 1.4086 | 1.7772 | 1.4365 |
| total | 100.0000 | 100.0000 | 100.0000 | 100.0000 | 100.0000 | 100.0000 | 100.0000 |
| m.ag | 34.8294 | 33.3544 | 42.8966 | 42.2211 | 33.3714 | 35.4515 | 34.5224 |
| sha | 200.0000 | 25.2300 | 8.0910 | 7.8158 | 58.8633 | 33.0457 | 66.9543 |
| lam | 1.040688 | 1.026520 | 1.070213 | 1.110426 | 1.034260 | 1.045199 | 1.038476 |
| r | 0.007976 | 0.005235 | 0.013572 | 0.020949 | 0.006737 | 0.008841 | 0.007551 |

*The first letter in the heading of each column refers to region of birth, the second to region of residence.

# APPENDIX C

Native-dependent input data, disaggregated by region of birth and region of residence, for the female population of the USA in 1968*.

Region of birth, South; Region of residence, South.

| age | population number | population - % - | births number | births - % - | deaths number | deaths - % - | arrivals number | arrivals - % - | departures number | departures - % - | observed rates ( x 1000 ) birth | death | inmig | outmig | net mig |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 3111200. | 10.77 | 0. | 0.00 | 16589. | 8.84 | 55363. | 9.85 | 138291. | 15.65 | 0.000 | 5.332 | 3.559 | 8.890 | -5.331 |
| 5 | 3252861. | 11.26 | 0. | 0.00 | 1381. | 0.74 | 58046. | 10.33 | 90082. | 10.20 | 0.000 | 0.425 | 3.569 | 5.539 | -1.970 |
| 10 | 3041703. | 10.53 | 2392. | 0.51 | 1057. | 0.56 | 46802. | 8.33 | 97055. | 10.99 | 0.786 | 0.348 | 3.077 | 6.382 | -3.304 |
| 15 | 2549596. | 8.83 | 100854. | 21.57 | 1634. | 0.87 | 59746. | 10.63 | 194217. | 21.99 | 39.557 | 0.641 | 4.687 | 15.235 | -10.548 |
| 20 | 2014624. | 6.97 | 160303. | 34.28 | 1557. | 0.83 | 80736. | 14.36 | 122898. | 13.91 | 79.570 | 0.773 | 8.015 | 12.201 | -4.186 |
| 25 | 1755229. | 6.08 | 107562. | 23.00 | 1665. | 0.89 | 60874. | 10.83 | 67339. | 7.62 | 61.281 | 0.949 | 6.936 | 7.673 | -0.737 |
| 30 | 1794786. | 6.21 | 59029. | 12.62 | 2507. | 1.34 | 45896. | 8.16 | 45655. | 5.17 | 32.889 | 1.397 | 5.114 | 5.088 | 0.027 |
| 35 | 1749412. | 6.06 | 28942. | 6.19 | 3945. | 2.10 | 35970. | 6.40 | 34458. | 3.90 | 16.544 | 2.255 | 4.112 | 3.939 | 0.173 |
| 40 | 1739408. | 6.02 | 8046. | 1.72 | 5766. | 3.07 | 28055. | 4.99 | 25869. | 2.93 | 4.626 | 3.315 | 3.226 | 2.974 | 0.251 |
| 45 | 1732323. | 6.00 | 511. | 0.11 | 8234. | 4.39 | 21837. | 3.88 | 17446. | 1.97 | 0.295 | 4.753 | 2.521 | 2.014 | 0.507 |
| 50 | 1529112. | 5.29 | 0. | 0.00 | 10037. | 5.35 | 16854. | 3.00 | 13216. | 1.50 | 0.000 | 6.564 | 2.204 | 1.729 | 0.476 |
| 55 | 1417283. | 4.91 | 0. | 0.00 | 13483. | 7.18 | 19924. | 3.54 | 10694. | 1.21 | 0.000 | 9.513 | 2.812 | 1.509 | 1.302 |
| 60 | 931085. | 3.22 | 0. | 0.00 | 12401. | 6.61 | 13856. | 2.46 | 8263. | 0.94 | 0.000 | 13.319 | 2.976 | 1.775 | 1.201 |
| 65 | 798916. | 2.77 | 0. | 0.00 | 16775. | 8.94 | 8632. | 1.54 | 6842. | 0.77 | 0.000 | 20.997 | 2.161 | 1.713 | 0.448 |
| 70 | 607902. | 2.10 | 0. | 0.00 | 19425. | 10.35 | 5292. | 0.94 | 6140. | 0.70 | 0.000 | 31.954 | 1.741 | 2.020 | -0.279 |
| 75 | 424364. | 1.47 | 0. | 0.00 | 22550. | 12.01 | 3175. | 0.56 | 3684. | 0.42 | 0.000 | 53.138 | 1.496 | 1.736 | -0.240 |
| 80 | 256447. | 0.89 | 0. | 0.00 | 21895. | 11.67 | 795. | 0.14 | 921. | 0.10 | 0.000 | 85.378 | 0.620 | 0.718 | -0.098 |
| 85 | 179292. | 0.62 | 0. | 0.00 | 26793. | 14.27 | 266. | 0.05 | 307. | 0.03 | 0.000 | 149.438 | 0.297 | 0.342 | -0.046 |
| tot | 28885548. | 100.00 | 467639. | 100.00 | 187694. | 100.00 | 562119. | 100.00 | 883377. | 100.00 | | | | | |
| gross | | | | | | | | | | | 1.178 | 1.952 | 0.296 | 0.407 | |
| crude(x1000) | | | | | | | | | | | 16.189 | 6.498 | 3.892 | 6.116 | -2.224 |
| m.age | | 30.34 | | 25.08 | | 61.57 | | 26.62 | | 20.46 | 25.80 | 77.51 | 34.39 | 27.06 | |
| e(0) | | | | | | | | | | | | 74.11 | | | |

Region of birth, South; Region of residence, North.

| age | population number | population - % - | births number | births - % - | deaths number | deaths - % - | arrivals number | arrivals - % - | departures number | departures - % - | observed rates ( x 1000 ) birth | death | inmig | outmig | net mig |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 263013. | 4.28 | 0. | 0.00 | 1295. | 3.12 | 138291. | 15.65 | 55363. | 9.85 | 0.000 | 4.924 | 105.159 | 42.099 | 63.060 |
| 5 | 356268. | 5.80 | 0. | 0.00 | 130. | 0.31 | 90082. | 10.20 | 58046. | 10.33 | 0.000 | 0.365 | 50.570 | 32.586 | 17.984 |
| 10 | 401326. | 6.53 | 128. | 0.10 | 113. | 0.27 | 97055. | 10.99 | 46802. | 8.33 | 0.319 | 0.282 | 48.367 | 23.324 | 25.043 |
| 15 | 484489. | 7.89 | 13169. | 10.56 | 273. | 0.66 | 194217. | 21.99 | 59746. | 10.63 | 27.181 | 0.563 | 80.174 | 24.664 | 55.510 |
| 20 | 561604. | 9.14 | 42370. | 33.98 | 369. | 0.89 | 122898. | 13.91 | 80736. | 14.36 | 75.445 | 0.657 | 43.767 | 28.752 | 15.015 |
| 25 | 537998. | 8.76 | 36264. | 29.09 | 419. | 1.01 | 67339. | 7.62 | 60874. | 10.83 | 67.405 | 0.779 | 25.033 | 22.630 | 2.403 |
| 30 | 545053. | 8.87 | 20246. | 16.24 | 605. | 1.46 | 45655. | 5.17 | 45896. | 8.16 | 37.145 | 1.110 | 16.753 | 16.841 | -0.088 |
| 35 | 528164. | 8.60 | 9813. | 7.87 | 963. | 2.32 | 34458. | 3.90 | 35970. | 6.40 | 18.579 | 1.823 | 13.048 | 13.621 | -0.573 |
| 40 | 510761. | 8.32 | 2532. | 2.03 | 1439. | 3.47 | 25869. | 2.93 | 28055. | 4.99 | 4.957 | 2.817 | 10.130 | 10.986 | -0.856 |
| 45 | 523046. | 8.52 | 155. | 0.12 | 2172. | 5.24 | 17446. | 1.97 | 21837. | 3.88 | 0.296 | 4.153 | 6.671 | 8.350 | -1.679 |
| 50 | 396699. | 6.46 | 0. | 0.00 | 2443. | 5.89 | 13216. | 1.50 | 16854. | 3.00 | 0.000 | 6.158 | 6.663 | 8.497 | -1.834 |
| 55 | 353663. | 5.76 | 0. | 0.00 | 3202. | 7.72 | 10694. | 1.21 | 19924. | 3.54 | 0.000 | 9.054 | 6.048 | 11.267 | -5.220 |
| 60 | 189579. | 3.09 | 0. | 0.00 | 2531. | 6.10 | 8263. | 0.94 | 13856. | 2.46 | 0.000 | 13.351 | 8.717 | 14.618 | -5.900 |
| 65 | 158295. | 2.58 | 0. | 0.00 | 3342. | 8.06 | 6842. | 0.77 | 8632. | 1.54 | 0.000 | 21.112 | 8.645 | 10.906 | -2.262 |
| 70 | 131129. | 2.13 | 0. | 0.00 | 4471. | 10.78 | 6140. | 0.70 | 5292. | 0.94 | 0.000 | 34.096 | 9.365 | 8.071 | 1.293 |
| 75 | 97578. | 1.59 | 0. | 0.00 | 5536. | 13.35 | 3684. | 0.42 | 3175. | 0.56 | 0.000 | 56.734 | 7.551 | 6.508 | 1.043 |
| 80 | 61485. | 1.00 | 0. | 0.00 | 5507. | 13.28 | 921. | 0.10 | 795. | 0.14 | 0.000 | 89.567 | 2.996 | 2.586 | 0.410 |
| 85 | 42301. | 0.69 | 0. | 0.00 | 6662. | 16.06 | 307. | 0.03 | 266. | 0.05 | 0.000 | 157.490 | 1.452 | 1.258 | 0.194 |
| tot | 6142451. | 100.00 | 124677. | 100.00 | 41472. | 100.00 | 883377. | 100.00 | 562119. | 100.00 | | | | | |
| gross | | | | | | | | | | | 1.157 | 2.025 | 2.256 | 1.438 | |
| crude(x1000) | | | | | | | | | | | 20.298 | 6.752 | 28.763 | 18.303 | 10.460 |
| m.age | | 35.47 | | 26.66 | | 65.99 | | 20.46 | | 26.62 | 26.63 | 78.09 | 21.21 | 29.28 | |
| e(0) | | | | | | | | | | | | 74.45 | | | |

* Note that the numbers of arrivals and departures cover a five-year period (1965–1970); all other numbers and rates refer to a single year (1968).

## APPENDIX C (*continued*)

Native-dependent input data—continued*.

Region of birth, North; Region of residence, South.

| age | population number | population - % - | births number | births - % - | deaths number | deaths - % - | arrivals number | arrivals - % - | departures number | departures - % - | observed rates ( x 1000 ) birth | death | inmig | outmig | net mig |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 223698. | 6.24 | 0. | 0.00 | 1193. | 4.24 | 154040. | 13.00 | 61424. | 10.57 | 0.000 | 5.333 | 137.721 | 54.917 | 82.804 |
| 5 | 289921. | 8.08 | 0. | 0.00 | 123. | 0.44 | 114741. | 9.68 | 63766. | 10.98 | 0.000 | 0.424 | 79.153 | 43.989 | 35.165 |
| 10 | 297459. | 8.29 | 234. | 0.36 | 103. | 0.37 | 115598. | 9.76 | 51119. | 8.80 | 0.787 | 0.346 | 77.724 | 34.370 | 43.353 |
| 15 | 342189. | 9.54 | 13536. | 20.86 | 219. | 0.78 | 213162. | 17.99 | 78112. | 13.45 | 39.557 | 0.640 | 124.587 | 45.654 | 78.933 |
| 20 | 296795. | 8.27 | 23616. | 36.39 | 229. | 0.81 | 135005. | 11.39 | 96444. | 16.60 | 79.570 | 0.772 | 90.975 | 64.990 | 25.985 |
| 25 | 234951. | 6.55 | 14398. | 22.19 | 223. | 0.79 | 81102. | 6.85 | 58314. | 10.04 | 61.281 | 0.949 | 69.037 | 49.639 | 19.398 |
| 30 | 242636. | 6.76 | 7980. | 12.30 | 339. | 1.20 | 61432. | 5.18 | 43488. | 7.49 | 32.889 | 1.397 | 50.637 | 35.846 | 14.791 |
| 35 | 236501. | 6.59 | 3913. | 6.03 | 533. | 1.89 | 49640. | 4.19 | 32786. | 5.64 | 16.545 | 2.254 | 41.979 | 27.726 | 14.253 |
| 40 | 246916. | 6.88 | 1142. | 1.76 | 819. | 2.91 | 44543. | 3.76 | 25249. | 4.35 | 4.625 | 3.317 | 36.079 | 20.451 | 15.628 |
| 45 | 245908. | 6.86 | 72. | 0.11 | 1169. | 4.15 | 36154. | 3.05 | 17140. | 2.95 | 0.293 | 4.754 | 29.404 | 13.940 | 15.464 |
| 50 | 204892. | 5.71 | 0. | 0.00 | 1345. | 4.78 | 37907. | 3.20 | 12603. | 2.17 | 0.000 | 6.564 | 37.002 | 12.302 | 24.700 |
| 55 | 189908. | 5.29 | 0. | 0.00 | 1807. | 6.42 | 48324. | 4.08 | 10238. | 1.76 | 0.000 | 9.515 | 50.892 | 10.782 | 40.110 |
| 60 | 155764. | 4.34 | 0. | 0.00 | 2075. | 7.37 | 41770. | 3.53 | 9372. | 1.61 | 0.000 | 13.321 | 53.632 | 12.034 | 41.599 |
| 65 | 133654. | 3.73 | 0. | 0.00 | 2806. | 9.97 | 25202. | 2.13 | 7939. | 1.37 | 0.000 | 20.995 | 37.712 | 11.880 | 25.832 |
| 70 | 101698. | 2.84 | 0. | 0.00 | 3250. | 11.54 | 14560. | 1.23 | 7180. | 1.24 | 0.000 | 31.957 | 28.634 | 14.120 | 14.514 |
| 75 | 70994. | 1.98 | 0. | 0.00 | 3773. | 13.40 | 8737. | 0.74 | 4309. | 0.74 | 0.000 | 53.145 | 24.613 | 12.139 | 12.474 |
| 80 | 42901. | 1.20 | 0. | 0.00 | 3663. | 13.01 | 2184. | 0.18 | 1077. | 0.19 | 0.000 | 85.383 | 10.182 | 5.021 | 5.161 |
| 85 | 29994. | 0.84 | 0. | 0.00 | 4482. | 15.92 | 728. | 0.06 | 360. | 0.06 | 0.000 | 149.430 | 4.854 | 2.400 | 2.454 |
| tot | 3586779. | 100.00 | 64891. | 100.00 | 28151. | 100.00 | 1184829. | 100.00 | 580920. | 100.00 | | | | | |
| gross | | | | | | | | | | | 1.178 | 1.952 | 4.924 | 2.361 | |
| crude(x1000) | | | | | | | | | | | 18.092 | 7.849 | 66.066 | 32.392 | 33.674 |
| m.age | | 34.57 | | 25.04 | | 65.99 | | 25.50 | | 24.60 | 25.80 | 77.51 | 30.67 | 28.94 | |
| e(0) | | | | | | | | | | | | 74.11 | | | |

Region of birth, North; Region of residence, North.

| age | population number | population - % - | births number | births - % - | deaths number | deaths - % - | arrivals number | arrivals - % - | departures number | departures - % - | observed rates ( x 1000 ) birth | death | inmig | outmig | net mig |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 7189433. | 11.29 | 0. | 0.00 | 35409. | 8.58 | 61424. | 10.57 | 154040. | 13.00 | 0.000 | 4.925 | •1.709 | 4.285 | -2.576 |
| 5 | 7492557. | 11.77 | 0. | 0.00 | 2726. | 0.66 | 63766. | 10.98 | 114741. | 9.68 | 0.000 | 0.364 | 1.702 | 3.063 | -1.361 |
| 10 | 6781986. | 10.65 | 2168. | 0.23 | 1903. | 0.46 | 51119. | 8.80 | 115598. | 9.76 | 0.320 | 0.281 | 1.507 | 3.409 | -1.901 |
| 15 | 5605165. | 8.80 | 152352. | 15.86 | 3153. | 0.76 | 78112. | 13.45 | 213162. | 17.99 | 27.181 | 0.563 | 2.787 | 7.606 | -4.819 |
| 20 | 4350986. | 6.83 | 328256. | 34.18 | 2860. | 0.69 | 96444. | 16.60 | 135005. | 11.39 | 75.444 | 0.657 | 4.433 | 6.206 | -1.773 |
| 25 | 3629665. | 5.70 | 244662. | 25.47 | 2826. | 0.68 | 58314. | 10.04 | 81102. | 6.85 | 67.406 | 0.779 | 3.213 | 4.469 | -1.256 |
| 30 | 3835769. | 6.03 | 142483. | 14.84 | 4255. | 1.03 | 43488. | 7.49 | 61432. | 5.18 | 37.146 | 1.109 | 2.267 | 3.203 | -0.936 |
| 35 | 3751833. | 5.89 | 69704. | 7.26 | 6844. | 1.66 | 32786. | 5.64 | 49640. | 4.19 | 18.579 | 1.824 | 1.748 | 2.646 | -0.898 |
| 40 | 3948275. | 6.20 | 19575. | 2.04 | 11124. | 2.70 | 25249. | 4.35 | 44543. | 3.76 | 4.958 | 2.817 | 1.279 | 2.256 | -0.977 |
| 45 | 4049675. | 6.36 | 1203. | 0.13 | 16821. | 4.08 | 17140. | 2.95 | 36154. | 3.05 | 0.297 | 4.154 | 0.846 | 1.786 | -0.939 |
| 50 | 3309504. | 5.20 | 0. | 0.00 | 20383. | 4.94 | 12603. | 2.17 | 37907. | 3.20 | 0.000 | 6.159 | 0.762 | 2.291 | -1.529 |
| 55 | 2965722. | 4.66 | 0. | 0.00 | 26852. | 6.51 | 10238. | 1.76 | 48324. | 4.08 | 0.000 | 9.054 | 0.690 | 3.259 | -2.568 |
| 60 | 1880346. | 2.95 | 0. | 0.00 | 25106. | 6.08 | 9372. | 1.61 | 41770. | 3.53 | 0.000 | 13.352 | 0.997 | 4.443 | -3.446 |
| 65 | 1573031. | 2.47 | 0. | 0.00 | 33206. | 8.05 | 7939. | 1.37 | 25202. | 2.13 | 0.000 | 21.110 | 1.009 | 3.204 | -2.195 |
| 70 | 1307576. | 2.05 | 0. | 0.00 | 44582. | 10.80 | 7180. | 1.24 | 14560. | 1.23 | 0.000 | 34.095 | 1.098 | 2.227 | -1.129 |
| 75 | 969001. | 1.52 | 0. | 0.00 | 54977. | 13.32 | 4309. | 0.74 | 8737. | 0.74 | 0.000 | 56.736 | 0.889 | 1.803 | -0.914 |
| 80 | 606872. | 0.95 | 0. | 0.00 | 54356. | 13.17 | 1077. | 0.19 | 2184. | 0.18 | 0.000 | 89.567 | 0.355 | 0.720 | -0.365 |
| 85 | 414830. | 0.65 | 0. | 0.00 | 65333. | 15.83 | 360. | 0.06 | 728. | 0.06 | 0.000 | 157.493 | 0.174 | 0.351 | -0.177 |
| tot | 63662232. | 100.00 | 960403. | 100.00 | 412716. | 100.00 | 580920. | 100.00 | 1184829. | 100.00 | | | | | |
| gross | | | | | | | | | | | 1.157 | 2.025 | 0.137 | 0.286 | |
| crude(x1000) | | | | | | | | | | | 15.086 | 6.483 | 1.825 | 3.722 | -1.897 |
| m.age | | 29.92 | | 25.97 | | 63.28 | | 24.60 | | 25.50 | 26.63 | 78.09 | 33.12 | 35.80 | |
| e(0) | | | | | | | | | | | | 74.45 | | | |

* Note that the numbers of arrivals and departures cover a five-year period (1965–1970); all other numbers and rates refer to a single year (1968).

# APPENDIX D

Native-dependent multistate life tables: probabilities of death and migration (1968).

region s -> s

| age | death | migration from s -> s to s -> s | s -> n | n -> s | n -> n |
|---|---|---|---|---|---|
| 0 | 0.026271 | 0.935243 | 0.038486 | 0.000000 | 0.000000 |
| 5 | 0.002117 | 0.972648 | 0.025236 | 0.000000 | 0.000000 |
| 10 | 0.001733 | 0.968610 | 0.029657 | 0.000000 | 0.000000 |
| 15 | 0.003186 | 0.927746 | 0.069068 | 0.000000 | 0.000000 |
| 20 | 0.003841 | 0.941010 | 0.055149 | 0.000000 | 0.000000 |
| 25 | 0.004717 | 0.959769 | 0.035515 | 0.000000 | 0.000000 |
| 30 | 0.006943 | 0.969089 | 0.023969 | 0.000000 | 0.000000 |
| 35 | 0.011195 | 0.970124 | 0.018681 | 0.000000 | 0.000000 |
| 40 | 0.016421 | 0.969423 | 0.014156 | 0.000000 | 0.000000 |
| 45 | 0.023472 | 0.966923 | 0.009604 | 0.000000 | 0.000000 |
| 50 | 0.032285 | 0.959546 | 0.008169 | 0.000000 | 0.000000 |
| 55 | 0.046454 | 0.946558 | 0.006989 | 0.000000 | 0.000000 |
| 60 | 0.064449 | 0.927557 | 0.007994 | 0.000000 | 0.000000 |
| 65 | 0.099752 | 0.892744 | 0.007504 | 0.000000 | 0.000000 |
| 70 | 0.148000 | 0.843577 | 0.008422 | 0.000000 | 0.000000 |
| 75 | 0.234596 | 0.758812 | 0.006592 | 0.000000 | 0.000000 |
| 80 | 0.351822 | 0.645776 | 0.002402 | 0.000000 | 0.000000 |
| 85 | 1.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 |

region n -> s

| age | death | migration from n -> s to s -> s | s -> n | n -> s | n -> n |
|---|---|---|---|---|---|
| 0 | 0.026080 | 0.000000 | 0.000000 | 0.740360 | 0.233560 |
| 5 | 0.002089 | 0.000000 | 0.000000 | 0.801482 | 0.196429 |
| 10 | 0.001704 | 0.000000 | 0.000000 | 0.841508 | 0.156787 |
| 15 | 0.003156 | 0.000000 | 0.000000 | 0.795963 | 0.200881 |
| 20 | 0.003772 | 0.000000 | 0.000000 | 0.721279 | 0.274949 |
| 25 | 0.004642 | 0.000000 | 0.000000 | 0.777616 | 0.217742 |
| 30 | 0.006845 | 0.000000 | 0.000000 | 0.830833 | 0.162321 |
| 35 | 0.011069 | 0.000000 | 0.000000 | 0.861339 | 0.127592 |
| 40 | 0.016330 | 0.000000 | 0.000000 | 0.888331 | 0.095339 |
| 45 | 0.023373 | 0.000000 | 0.000000 | 0.911002 | 0.065625 |
| 50 | 0.032212 | 0.000000 | 0.000000 | 0.910253 | 0.057535 |
| 55 | 0.046390 | 0.000000 | 0.000000 | 0.903825 | 0.049785 |
| 60 | 0.064464 | 0.000000 | 0.000000 | 0.881348 | 0.054188 |
| 65 | 0.099752 | 0.000000 | 0.000000 | 0.848494 | 0.051754 |
| 70 | 0.148212 | 0.000000 | 0.000000 | 0.793739 | 0.058049 |
| 75 | 0.234869 | 0.000000 | 0.000000 | 0.719607 | 0.045524 |
| 80 | 0.351960 | 0.000000 | 0.000000 | 0.631334 | 0.016706 |
| 85 | 1.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 |

region s -> n

| age | death | migration from s -> n to s -> s | s -> n | n -> s | n -> n |
|---|---|---|---|---|---|
| 0 | 0.024503 | 0.182255 | 0.793243 | 0.000000 | 0.000000 |
| 5 | 0.001832 | 0.148470 | 0.849698 | 0.000000 | 0.000000 |
| 10 | 0.001413 | 0.108392 | 0.890195 | 0.000000 | 0.000000 |
| 15 | 0.002825 | 0.111811 | 0.885363 | 0.000000 | 0.000000 |
| 20 | 0.003317 | 0.129964 | 0.866718 | 0.000000 | 0.000000 |
| 25 | 0.003931 | 0.104743 | 0.891326 | 0.000000 | 0.000000 |
| 30 | 0.005591 | 0.079343 | 0.915066 | 0.000000 | 0.000000 |
| 35 | 0.009154 | 0.064592 | 0.926254 | 0.000000 | 0.000000 |
| 40 | 0.014053 | 0.052284 | 0.933663 | 0.000000 | 0.000000 |
| 45 | 0.020618 | 0.039815 | 0.939568 | 0.000000 | 0.000000 |
| 50 | 0.030365 | 0.040155 | 0.929480 | 0.000000 | 0.000000 |
| 55 | 0.044326 | 0.052180 | 0.903494 | 0.000000 | 0.000000 |
| 60 | 0.064592 | 0.065837 | 0.869570 | 0.000000 | 0.000000 |
| 65 | 0.100257 | 0.047782 | 0.851961 | 0.000000 | 0.000000 |
| 70 | 0.156924 | 0.033653 | 0.809422 | 0.000000 | 0.000000 |
| 75 | 0.248239 | 0.024708 | 0.727053 | 0.000000 | 0.000000 |
| 80 | 0.365827 | 0.008648 | 0.625525 | 0.000000 | 0.000000 |
| 85 | 1.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 |

region n -> n

| age | death | migration from n -> n to s -> s | s -> n | n -> s | n -> n |
|---|---|---|---|---|---|
| 0 | 0.024345 | 0.000000 | 0.000000 | 0.018225 | 0.957431 |
| 5 | 0.001820 | 0.000000 | 0.000000 | 0.013677 | 0.984504 |
| 10 | 0.001405 | 0.000000 | 0.000000 | 0.015551 | 0.983045 |
| 15 | 0.002816 | 0.000000 | 0.000000 | 0.033466 | 0.963718 |
| 20 | 0.003289 | 0.000000 | 0.000000 | 0.026254 | 0.970457 |
| 25 | 0.003894 | 0.000000 | 0.000000 | 0.019603 | 0.976504 |
| 30 | 0.005542 | 0.000000 | 0.000000 | 0.014505 | 0.979954 |
| 35 | 0.009093 | 0.000000 | 0.000000 | 0.012177 | 0.978730 |
| 40 | 0.014002 | 0.000000 | 0.000000 | 0.010518 | 0.975480 |
| 45 | 0.020567 | 0.000000 | 0.000000 | 0.008406 | 0.971027 |
| 50 | 0.030338 | 0.000000 | 0.000000 | 0.010714 | 0.958948 |
| 55 | 0.044285 | 0.000000 | 0.000000 | 0.015047 | 0.940667 |
| 60 | 0.064601 | 0.000000 | 0.000000 | 0.020006 | 0.915393 |
| 65 | 0.100253 | 0.000000 | 0.000000 | 0.013959 | 0.885788 |
| 70 | 0.157038 | 0.000000 | 0.000000 | 0.009155 | 0.833807 |
| 75 | 0.248387 | 0.000000 | 0.000000 | 0.006763 | 0.744851 |
| 80 | 0.365884 | 0.000000 | 0.000000 | 0.002395 | 0.631721 |
| 85 | 1.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 |

## APPENDIX D *(continued)*

Native-dependent multistate life tables: expectations of life (1968).

initial region of cohort s -> s

| age | total | s -> s | s -> n | n -> s | n -> n |
|---|---|---|---|---|---|
| 0 | 74.14005 | 63.18850 | 10.95154 | 0.00000 | 0.00000 |
| 5 | 71.07285 | 59.92465 | 11.14820 | 0.00000 | 0.00000 |
| 10 | 66.21751 | 55.28968 | 10.92784 | 0.00000 | 0.00000 |
| 15 | 61.32692 | 50.72458 | 10.60235 | 0.00000 | 0.00000 |
| 20 | 56.51323 | 46.41336 | 10.09986 | 0.00000 | 0.00000 |
| 25 | 51.71765 | 42.32948 | 9.38817 | 0.00000 | 0.00000 |
| 30 | 46.94445 | 38.37098 | 8.57347 | 0.00000 | 0.00000 |
| 35 | 42.24437 | 34.51814 | 7.72622 | 0.00000 | 0.00000 |
| 40 | 37.67913 | 30.79958 | 6.87955 | 0.00000 | 0.00000 |
| 45 | 33.25034 | 27.20755 | 6.04279 | 0.00000 | 0.00000 |
| 50 | 28.97207 | 23.74648 | 5.22559 | 0.00000 | 0.00000 |
| 55 | 24.84492 | 20.41054 | 4.43438 | 0.00000 | 0.00000 |
| 60 | 20.92360 | 17.23384 | 3.68977 | 0.00000 | 0.00000 |
| 65 | 17.19334 | 14.19163 | 3.00172 | 0.00000 | 0.00000 |
| 70 | 13.82308 | 11.42644 | 2.39663 | 0.00000 | 0.00000 |
| 75 | 10.81459 | 8.95461 | 1.85998 | 0.00000 | 0.00000 |
| 80 | 8.39719 | 6.97299 | 1.42419 | 0.00000 | 0.00000 |
| 85 | 6.63259 | 5.53472 | 1.09787 | 0.00000 | 0.00000 |

initial region of cohort n -> s

| age | total | s -> s | s -> n | n -> s | n -> n |
|---|---|---|---|---|---|
| 0 | 74.24090 | 0.00000 | 0.00000 | 25.41264 | 48.82825 |
| 5 | 71.16196 | 0.00000 | 0.00000 | 21.62573 | 49.53624 |
| 10 | 66.30126 | 0.00000 | 0.00000 | 18.23078 | 48.07048 |
| 15 | 61.40276 | 0.00000 | 0.00000 | 15.41447 | 45.98829 |
| 20 | 56.57965 | 0.00000 | 0.00000 | 13.06390 | 43.51575 |
| 25 | 51.76950 | 0.00000 | 0.00000 | 11.19910 | 40.57040 |
| 30 | 46.97432 | 0.00000 | 0.00000 | 9.74482 | 37.22949 |
| 35 | 42.23798 | 0.00000 | 0.00000 | 8.53261 | 33.70537 |
| 40 | 37.62153 | 0.00000 | 0.00000 | 7.48109 | 30.14044 |
| 45 | 33.13837 | 0.00000 | 0.00000 | 6.54067 | 26.59770 |
| 50 | 28.79987 | 0.00000 | 0.00000 | 5.67831 | 23.12156 |
| 55 | 24.63294 | 0.00000 | 0.00000 | 4.87541 | 19.75753 |
| 60 | 20.66832 | 0.00000 | 0.00000 | 4.11519 | 16.55312 |
| 65 | 16.92252 | 0.00000 | 0.00000 | 3.37896 | 13.54356 |
| 70 | 13.52774 | 0.00000 | 0.00000 | 2.69510 | 10.83264 |
| 75 | 10.55460 | 0.00000 | 0.00000 | 2.10008 | 8.45453 |
| 80 | 8.17843 | 0.00000 | 0.00000 | 1.63872 | 6.53971 |
| 85 | 6.41651 | 0.00000 | 0.00000 | 1.31095 | 5.10556 |

initial region of cohort s -> n

| age | total | s -> s | s -> n | n -> s | n -> n |
|---|---|---|---|---|---|
| 0 | 74.38078 | 36.62362 | 37.75716 | 0.00000 | 0.00000 |
| 5 | 71.18633 | 37.07647 | 34.10985 | 0.00000 | 0.00000 |
| 10 | 66.31605 | 35.92097 | 30.39507 | 0.00000 | 0.00000 |
| 15 | 61.41255 | 34.29254 | 27.12001 | 0.00000 | 0.00000 |
| 20 | 56.58736 | 32.43096 | 24.15640 | 0.00000 | 0.00000 |
| 25 | 51.77921 | 30.33603 | 21.44319 | 0.00000 | 0.00000 |
| 30 | 46.99195 | 28.02473 | 18.96721 | 0.00000 | 0.00000 |
| 35 | 42.27297 | 25.59182 | 16.68114 | 0.00000 | 0.00000 |
| 40 | 37.68459 | 23.12437 | 14.56021 | 0.00000 | 0.00000 |
| 45 | 33.23359 | 20.65763 | 12.57596 | 0.00000 | 0.00000 |
| 50 | 28.93270 | 18.22321 | 10.70948 | 0.00000 | 0.00000 |
| 55 | 24.79180 | 15.83815 | 8.95365 | 0.00000 | 0.00000 |
| 60 | 20.85652 | 13.52005 | 7.33647 | 0.00000 | 0.00000 |
| 65 | 17.12225 | 11.24400 | 5.87825 | 0.00000 | 0.00000 |
| 70 | 13.74583 | 9.12206 | 4.62376 | 0.00000 | 0.00000 |
| 75 | 10.74843 | 7.20269 | 3.54574 | 0.00000 | 0.00000 |
| 80 | 8.34301 | 5.64732 | 2.69569 | 0.00000 | 0.00000 |
| 85 | 6.58009 | 4.50800 | 2.07209 | 0.00000 | 0.00000 |

initial region of cohort n -> n

| age | total | s -> s | s -> n | n -> s | n -> n |
|---|---|---|---|---|---|
| 0 | 74.44974 | 0.00000 | 0.00000 | 5.49861 | 68.95113 |
| 5 | 71.24503 | 0.00000 | 0.00000 | 5.58911 | 65.65591 |
| 10 | 66.37069 | 0.00000 | 0.00000 | 5.48144 | 60.88926 |
| 15 | 61.46108 | 0.00000 | 0.00000 | 5.32023 | 56.14085 |
| 20 | 56.62837 | 0.00000 | 0.00000 | 5.07865 | 51.54972 |
| 25 | 51.80863 | 0.00000 | 0.00000 | 4.75990 | 47.04873 |
| 30 | 47.00400 | 0.00000 | 0.00000 | 4.41807 | 42.58593 |
| 35 | 42.25630 | 0.00000 | 0.00000 | 4.07144 | 38.18485 |
| 40 | 37.62709 | 0.00000 | 0.00000 | 3.72945 | 33.89764 |
| 45 | 33.13234 | 0.00000 | 0.00000 | 3.39240 | 29.73995 |
| 50 | 28.78265 | 0.00000 | 0.00000 | 3.06050 | 25.72215 |
| 55 | 24.60922 | 0.00000 | 0.00000 | 2.73158 | 21.87764 |
| 60 | 20.63810 | 0.00000 | 0.00000 | 2.39260 | 18.24550 |
| 65 | 16.89049 | 0.00000 | 0.00000 | 2.02643 | 14.86406 |
| 70 | 13.49295 | 0.00000 | 0.00000 | 1.65340 | 11.83956 |
| 75 | 10.52485 | 0.00000 | 0.00000 | 1.31133 | 9.21352 |
| 80 | 8.15384 | 0.00000 | 0.00000 | 1.03641 | 7.11743 |
| 85 | 6.39218 | 0.00000 | 0.00000 | 0.83589 | 5.55630 |

Native-dependent multistate projection: growth matrix (1968). (See p. 65 for explanation of labeling.)

region s -> s

| age | first row s -> s | s -> n | n -> s | n -> n |
|---|---|---|---|---|
| 0 | 0.000000 | 0.000000 | 0.000000 | 0.000000 |
| 5 | 0.001846 | 0.000037 | 0.000002 | 0.000020 |
| 10 | 0.092669 | 0.001843 | 0.000303 | 0.002985 |
| 15 | 0.275558 | 0.005480 | 0.001066 | 0.010490 |
| 20 | 0.333269 | 0.006628 | 0.000707 | 0.006954 |
| 25 | 0.224945 | 0.004473 | 0.000255 | 0.002506 |
| 30 | 0.118359 | 0.002354 | 0.000091 | 0.000893 |
| 35 | 0.050870 | 0.001012 | 0.000019 | 0.000183 |
| 40 | 0.011881 | 0.000236 | 0.000001 | 0.000008 |
| 45 | 0.000714 | 0.000014 | 0.000000 | 0.000000 |
| 50 | 0.000000 | 0.000000 | 0.000000 | 0.000000 |
| 55 | 0.000000 | 0.000000 | 0.000000 | 0.000000 |
| 60 | 0.000000 | 0.000000 | 0.000000 | 0.000000 |
| 65 | 0.000000 | 0.000000 | 0.000000 | 0.000000 |
| 70 | 0.000000 | 0.000000 | 0.000000 | 0.000000 |
| 75 | 0.000000 | 0.000000 | 0.000000 | 0.000000 |
| 80 | 0.000000 | 0.000000 | 0.000000 | 0.000000 |

| age | survivorship proportions s -> s | s -> n | n -> s | n -> n |
|---|---|---|---|---|
| 0 | 0.952905 | 0.032724 | 0.000000 | 0.000000 |
| 5 | 0.970381 | 0.027694 | 0.000000 | 0.000000 |
| 10 | 0.948549 | 0.048993 | 0.000000 | 0.000000 |
| 15 | 0.934461 | 0.062030 | 0.000000 | 0.000000 |
| 20 | 0.949700 | 0.046026 | 0.000000 | 0.000000 |
| 25 | 0.964085 | 0.030094 | 0.000000 | 0.000000 |
| 30 | 0.969504 | 0.021439 | 0.000000 | 0.000000 |
| 35 | 0.969718 | 0.016490 | 0.000000 | 0.000000 |
| 40 | 0.968146 | 0.011938 | 0.000000 | 0.000000 |
| 45 | 0.963298 | 0.008873 | 0.000000 | 0.000000 |
| 50 | 0.953213 | 0.007535 | 0.000000 | 0.000000 |
| 55 | 0.937345 | 0.007414 | 0.000000 | 0.000000 |
| 60 | 0.910767 | 0.007719 | 0.000000 | 0.000000 |
| 65 | 0.869526 | 0.007846 | 0.000000 | 0.000000 |
| 70 | 0.804772 | 0.007377 | 0.000000 | 0.000000 |
| 75 | 0.710013 | 0.004564 | 0.000000 | 0.000000 |
| 80 | 1.047918 | 0.004523 | 0.000000 | 0.000000 |

region s -> n

| age | first row s -> s | s -> n | n -> s | n -> n |
|---|---|---|---|---|
| 0 | 0.000000 | 0.000000 | 0.000000 | 0.000000 |
| 5 | 0.000247 | 0.000005 | 0.000063 | 0.000621 |
| 10 | 0.010412 | 0.000207 | 0.005578 | 0.054884 |
| 15 | 0.023237 | 0.000462 | 0.021252 | 0.209105 |
| 20 | 0.017634 | 0.000351 | 0.030661 | 0.301684 |
| 25 | 0.007403 | 0.000147 | 0.022991 | 0.226210 |
| 30 | 0.002894 | 0.000058 | 0.012358 | 0.121590 |
| 35 | 0.000656 | 0.000013 | 0.005283 | 0.051982 |
| 40 | 0.000033 | 0.000001 | 0.001193 | 0.011739 |
| 45 | 0.000000 | 0.000000 | 0.000068 | 0.000664 |
| 50 | 0.000000 | 0.000000 | 0.000000 | 0.000000 |
| 55 | 0.000000 | 0.000000 | 0.000000 | 0.000000 |
| 60 | 0.000000 | 0.000000 | 0.000000 | 0.000000 |
| 65 | 0.000000 | 0.000000 | 0.000000 | 0.000000 |
| 70 | 0.000000 | 0.000000 | 0.000000 | 0.000000 |
| 75 | 0.000000 | 0.000000 | 0.000000 | 0.000000 |
| 80 | 0.000000 | 0.000000 | 0.000000 | 0.000000 |

| age | survivorship proportions s -> s | s -> n | n -> s | n -> n |
|---|---|---|---|---|
| 0 | 0.169317 | 0.817454 | 0.000000 | 0.000000 |
| 5 | 0.129917 | 0.868458 | 0.000000 | 0.000000 |
| 10 | 0.108810 | 0.889071 | 0.000000 | 0.000000 |
| 15 | 0.120724 | 0.876200 | 0.000000 | 0.000000 |
| 20 | 0.118955 | 0.877412 | 0.000000 | 0.000000 |
| 25 | 0.093050 | 0.902175 | 0.000000 | 0.000000 |
| 30 | 0.072320 | 0.920298 | 0.000000 | 0.000000 |
| 35 | 0.058664 | 0.929738 | 0.000000 | 0.000000 |
| 40 | 0.046230 | 0.936451 | 0.000000 | 0.000000 |
| 45 | 0.039902 | 0.934666 | 0.000000 | 0.000000 |
| 50 | 0.045809 | 0.916951 | 0.000000 | 0.000000 |
| 55 | 0.058394 | 0.887408 | 0.000000 | 0.000000 |
| 60 | 0.056805 | 0.861372 | 0.000000 | 0.000000 |
| 65 | 0.040613 | 0.832407 | 0.000000 | 0.000000 |
| 70 | 0.028797 | 0.772561 | 0.000000 | 0.000000 |
| 75 | 0.017028 | 0.684281 | 0.000000 | 0.000000 |
| 80 | 0.016774 | 0.969646 | 0.000000 | 0.000000 |

## APPENDIX D (*continued*)

Native-dependent multistate projection: growth matrix (1968)—continued.

region n -> s

| age | first row s -> s | s -> n | n -> s | n -> n |
|---|---|---|---|---|
| 0 | 0.000000 | 0.000000 | 0.000000 | 0.000000 |
| 5 | 0.001402 | 0.000188 | 0.000001 | 0.000140 |
| 10 | 0.072404 | 0.009717 | 0.000109 | 0.011711 |
| 15 | 0.218049 | 0.029263 | 0.000402 | 0.043198 |
| 20 | 0.272323 | 0.036546 | 0.000386 | 0.041486 |
| 25 | 0.190591 | 0.025578 | 0.000164 | 0.017610 |
| 30 | 0.101947 | 0.013681 | 0.000062 | 0.006660 |
| 35 | 0.044793 | 0.006011 | 0.000013 | 0.001365 |
| 40 | 0.010651 | 0.001429 | 0.000001 | 0.000059 |
| 45 | 0.000646 | 0.000087 | 0.000000 | 0.000000 |
| 50 | 0.000000 | 0.000000 | 0.000000 | 0.000000 |
| 55 | 0.000000 | 0.000000 | 0.000000 | 0.000000 |
| 60 | 0.000000 | 0.000000 | 0.000000 | 0.000000 |
| 65 | 0.000000 | 0.000000 | 0.000000 | 0.000000 |
| 70 | 0.000000 | 0.000000 | 0.000000 | 0.000000 |
| 75 | 0.000000 | 0.000000 | 0.000000 | 0.000000 |
| 80 | 0.000000 | 0.000000 | 0.000000 | 0.000000 |

| age | survivorship proportions s -> s | s -> n | n -> s | n -> n |
|---|---|---|---|---|
| 0 | 0.000000 | 0.000000 | 0.766006 | 0.219649 |
| 5 | 0.000000 | 0.000000 | 0.819376 | 0.178729 |
| 10 | 0.000000 | 0.000000 | 0.821482 | 0.176091 |
| 15 | 0.000000 | 0.000000 | 0.762531 | 0.234013 |
| 20 | 0.000000 | 0.000000 | 0.744276 | 0.251539 |
| 25 | 0.000000 | 0.000000 | 0.800543 | 0.193751 |
| 30 | 0.000000 | 0.000000 | 0.844562 | 0.146519 |
| 35 | 0.000000 | 0.000000 | 0.873766 | 0.112561 |
| 40 | 0.000000 | 0.000000 | 0.898939 | 0.081251 |
| 45 | 0.000000 | 0.000000 | 0.910685 | 0.061558 |
| 50 | 0.000000 | 0.000000 | 0.907257 | 0.053562 |
| 55 | 0.000000 | 0.000000 | 0.893223 | 0.051534 |
| 60 | 0.000000 | 0.000000 | 0.865871 | 0.052603 |
| 65 | 0.000000 | 0.000000 | 0.823295 | 0.053871 |
| 70 | 0.000000 | 0.000000 | 0.760900 | 0.050938 |
| 75 | 0.000000 | 0.000000 | 0.682606 | 0.031749 |
| 80 | 0.000000 | 0.000000 | 1.019620 | 0.031410 |

region n -> n

| age | first row s -> s | s -> n | n -> s | n -> n |
|---|---|---|---|---|
| 0 | 0.000000 | 0.000000 | 0.000000 | 0.000000 |
| 5 | 0.000025 | 0.000003 | 0.000007 | 0.000769 |
| 10 | 0.002085 | 0.000280 | 0.000610 | 0.065536 |
| 15 | 0.005059 | 0.000679 | 0.002282 | 0.245143 |
| 20 | 0.003123 | 0.000419 | 0.003213 | 0.345069 |
| 25 | 0.001244 | 0.000167 | 0.002363 | 0.253808 |
| 30 | 0.000485 | 0.000065 | 0.001261 | 0.135400 |
| 35 | 0.000115 | 0.000015 | 0.000534 | 0.057309 |
| 40 | 0.000006 | 0.000001 | 0.000120 | 0.012838 |
| 45 | 0.000000 | 0.000000 | 0.000007 | 0.000727 |
| 50 | 0.000000 | 0.000000 | 0.000000 | 0.000000 |
| 55 | 0.000000 | 0.000000 | 0.000000 | 0.000000 |
| 60 | 0.000000 | 0.000000 | 0.000000 | 0.000000 |
| 65 | 0.000000 | 0.000000 | 0.000000 | 0.000000 |
| 70 | 0.000000 | 0.000000 | 0.000000 | 0.000000 |
| 75 | 0.000000 | 0.000000 | 0.000000 | 0.000000 |
| 80 | 0.000000 | 0.000000 | 0.000000 | 0.000000 |

| age | survivorship proportions s -> s | s -> n | n -> s | n -> n |
|---|---|---|---|---|
| 0 | 0.000000 | 0.000000 | 0.016331 | 0.970457 |
| 5 | 0.000000 | 0.000000 | 0.014759 | 0.983629 |
| 10 | 0.000000 | 0.000000 | 0.024232 | 0.973658 |
| 15 | 0.000000 | 0.000000 | 0.029224 | 0.967723 |
| 20 | 0.000000 | 0.000000 | 0.023422 | 0.972985 |
| 25 | 0.000000 | 0.000000 | 0.017384 | 0.977897 |
| 30 | 0.000000 | 0.000000 | 0.013476 | 0.979210 |
| 35 | 0.000000 | 0.000000 | 0.011446 | 0.977017 |
| 40 | 0.000000 | 0.000000 | 0.009539 | 0.973198 |
| 45 | 0.000000 | 0.000000 | 0.009541 | 0.965059 |
| 50 | 0.000000 | 0.000000 | 0.012816 | 0.949979 |
| 55 | 0.000000 | 0.000000 | 0.017359 | 0.928437 |
| 60 | 0.000000 | 0.000000 | 0.016935 | 0.901235 |
| 65 | 0.000000 | 0.000000 | 0.011484 | 0.861402 |
| 70 | 0.000000 | 0.000000 | 0.007861 | 0.793333 |
| 75 | 0.000000 | 0.000000 | 0.004699 | 0.696499 |
| 80 | 0.000000 | 0.000000 | 0.004638 | 0.981125 |

Native-dependent multistate projection: 1998*.

population

| age | total | s -> s | s -> n | n -> s | n -> n | south | north |
|---|---|---|---|---|---|---|---|
| 0 | 11196813. | 3459609. | 123621. | 121584. | 7491999. | 3581193. | 7615620. |
| 5 | 11336667. | 3352658. | 217972. | 224193. | 7541845. | 3576850. | 7759817. |
| 10 | 11526699. | 3349397. | 288225. | 301517. | 7587560. | 3650914. | 7875785. |
| 15 | 10918060. | 3104168. | 405434. | 407530. | 7000927. | 3511699. | 7406361. |
| 20 | 9664265. | 2668703. | 493979. | 455106. | 6046477. | 3123810. | 6540456. |
| 25 | 8365127. | 2267579. | 484798. | 417376. | 5195374. | 2684954. | 5680172. |
| 30 | 10472025. | 2613974. | 654084. | 564914. | 6639054. | 3178888. | 7293137. |
| 35 | 11122760. | 2796649. | 718683. | 598967. | 7008460. | 3395617. | 7727144. |
| 40 | 10165524. | 2624486. | 690665. | 545279. | 6305094. | 3169765. | 6995759. |
| 45 | 8538368. | 2256039. | 615889. | 462802. | 5203638. | 2718841. | 5819527. |
| 50 | 6708972. | 1839828. | 540732. | 357586. | 3970826. | 2197415. | 4511558. |
| 55 | 5522848. | 1581665. | 463757. | 299359. | 3178067. | 1881023. | 3641824. |
| 60 | 5470592. | 1552931. | 430268. | 331373. | 3156020. | 1884304. | 3586288. |
| 65 | 4943313. | 1413890. | 374535. | 327847. | 2827041. | 1741737. | 3201576. |
| 70 | 4501860. | 1250901. | 318541. | 313722. | 2618696. | 1564623. | 2937237. |
| 75 | 3750827. | 1033038. | 266129. | 261182. | 2190478. | 1294220. | 2456607. |
| 80 | 2258369. | 661776. | 150538. | 154037. | 1292018. | 815813. | 1442556. |
| 85 | 2146821. | 667762. | 140846. | 144875. | 1193338. | 812637. | 1334184. |
| total | 138609888. | 38495044. | 7378696. | 6289250. | 86446904. | 44784300. | 93825616. |

percentage distribution

| age | total | s -> s | s -> n | n -> s | n -> n | south | north |
|---|---|---|---|---|---|---|---|
| 0 | 8.0779 | 8.9872 | 1.6754 | 1.9332 | 8.6666 | 7.9965 | 8.1168 |
| 5 | 8.1788 | 8.7093 | 2.9541 | 3.5647 | 8.7243 | 7.9868 | 8.2705 |
| 10 | 8.3159 | 8.7009 | 3.9062 | 4.7942 | 8.7771 | 8.1522 | 8.3941 |
| 15 | 7.8768 | 8.0638 | 5.4947 | 6.4798 | 8.0985 | 7.8414 | 7.8938 |
| 20 | 6.9723 | 6.9326 | 6.6947 | 7.2363 | 6.9944 | 6.9752 | 6.9709 |
| 25 | 6.0350 | 5.8906 | 6.5702 | 6.6363 | 6.0099 | 5.9953 | 6.0540 |
| 30 | 7.5550 | 6.7904 | 8.8645 | 8.9822 | 7.6799 | 7.0982 | 7.7731 |
| 35 | 8.0245 | 7.2650 | 9.7400 | 9.5237 | 8.1072 | 7.5822 | 8.2356 |
| 40 | 7.3339 | 6.8177 | 9.3603 | 8.6700 | 7.2936 | 7.0778 | 7.4561 |
| 45 | 6.1600 | 5.8606 | 8.3469 | 7.3586 | 6.0195 | 6.0710 | 6.2025 |
| 50 | 4.8402 | 4.7794 | 7.3283 | 5.6857 | 4.5934 | 4.9067 | 4.8084 |
| 55 | 3.9845 | 4.1087 | 6.2851 | 4.7598 | 3.6763 | 4.2002 | 3.8815 |
| 60 | 3.9468 | 4.0341 | 5.8312 | 5.2689 | 3.6508 | 4.2075 | 3.8223 |
| 65 | 3.5663 | 3.6729 | 5.0759 | 5.2128 | 3.2703 | 3.8892 | 3.4123 |
| 70 | 3.2479 | 3.2495 | 4.3170 | 4.9882 | 3.0293 | 3.4937 | 3.1305 |
| 75 | 2.7060 | 2.6836 | 3.6067 | 4.1528 | 2.5339 | 2.8899 | 2.6183 |
| 80 | 1.6293 | 1.7191 | 2.0402 | 2.4492 | 1.4946 | 1.8216 | 1.5375 |
| 85 | 1.5488 | 1.7347 | 1.9088 | 2.3035 | 1.3804 | 1.8146 | 1.4220 |
| total | 100.0000 | 100.0000 | 100.0000 | 100.0000 | 100.0000 | 100.0000 | 100.0000 |
| m.ag | 34.8266 | 34.4130 | 43.3132 | 42.5775 | 33.7225 | 35.5595 | 34.4768 |
| sha | 200.0000 | 27.7722 | 5.3234 | 4.5374 | 62.3671 | 32.3096 | 67.6904 |
| lam | 1.040731 | 1.036608 | 1.026943 | 1.063249 | 1.042165 | 1.040268 | 1.040952 |
| r | 0.007985 | 0.007191 | 0.005317 | 0.012266 | 0.008260 | 0.007896 | 0.008027 |

*The first letter in the heading of each column refers to region of birth, the second to region of residence.

## APPENDIX D (*continued*)

Native-dependent multistate projection: stable equivalent population*.

**stable equivalent to original population**

| age | total | s -> s | s -> n | n -> s | n -> n | south | north |
|---|---|---|---|---|---|---|---|
| 0 | 10093513. | 3099809. | 110375. | 107086. | 6776243. | 3206895. | 6886618. |
| 5 | 9741735. | 2908414. | 187531. | 188533. | 6457257. | 3096947. | 6644789. |
| 10 | 9515354. | 2785250. | 238160. | 244395. | 6247549. | 3029645. | 6485708. |
| 15 | 9289493. | 2610332. | 340689. | 344562. | 5993911. | 2954893. | 6334599. |
| 20 | 9060114. | 2426895. | 450501. | 428461. | 5754257. | 2855356. | 6204758. |
| 25 | 8831009. | 2307557. | 496042. | 443888. | 5583522. | 2751445. | 6079564. |
| 30 | 8596857. | 2221869. | 505812. | 442661. | 5426515. | 2664530. | 5932327. |
| 35 | 8345431. | 2143451. | 502067. | 437342. | 5262571. | 2580793. | 5764639. |
| 40 | 8065594. | 2062540. | 491309. | 432833. | 5078910. | 2495373. | 5570220. |
| 45 | 7748977. | 1976005. | 474258. | 428104. | 4870610. | 2404109. | 5344867. |
| 50 | 7383601. | 1880951. | 450870. | 426929. | 4624851. | 2307880. | 5075721. |
| 55 | 6950989. | 1774493. | 418377. | 436977. | 4321141. | 2211470. | 4739519. |
| 60 | 6430207. | 1651348. | 376137. | 455294. | 3947427. | 2106642. | 4323564. |
| 65 | 5777358. | 1492467. | 329480. | 451132. | 3504278. | 1943599. | 3833758. |
| 70 | 4942713. | 1282848. | 279805. | 402781. | 2977279. | 1685629. | 3257084. |
| 75 | 3892667. | 1018021. | 220764. | 322768. | 2331114. | 1340789. | 2551879. |
| 80 | 2688178. | 710900. | 152353. | 226291. | 1598634. | 937191. | 1750987. |
| 85 | 2653692. | 731401. | 147689. | 233010. | 1541592. | 964411. | 1689281. |
| total | 130007496. | 35084556. | 6172220. | 6453048. | 82297664. | 41537600. | 88469880. |

**percentage distribution**

| age | total | s -> s | s -> n | n -> s | n -> n | south | north |
|---|---|---|---|---|---|---|---|
| 0 | 7.7638 | 8.8353 | 1.7883 | 1.6595 | 8.2338 | 7.7205 | 7.7841 |
| 5 | 7.4932 | 8.2897 | 3.0383 | 2.9216 | 7.8462 | 7.4558 | 7.5108 |
| 10 | 7.3191 | 7.9387 | 3.8586 | 3.7873 | 7.5914 | 7.2937 | 7.3310 |
| 15 | 7.1454 | 7.4401 | 5.5197 | 5.3395 | 7.2832 | 7.1138 | 7.1602 |
| 20 | 6.9689 | 6.9173 | 7.2988 | 6.6397 | 6.9920 | 6.8741 | 7.0134 |
| 25 | 6.7927 | 6.5771 | 8.0367 | 6.8787 | 6.7845 | 6.6240 | 6.8719 |
| 30 | 6.6126 | 6.3329 | 8.1950 | 6.8597 | 6.5938 | 6.4147 | 6.7055 |
| 35 | 6.4192 | 6.1094 | 8.1343 | 6.7773 | 6.3946 | 6.2131 | 6.5159 |
| 40 | 6.2039 | 5.8788 | 7.9600 | 6.7074 | 6.1714 | 6.0075 | 6.2962 |
| 45 | 5.9604 | 5.6321 | 7.6837 | 6.6341 | 5.9183 | 5.7878 | 6.0415 |
| 50 | 5.6794 | 5.3612 | 7.3048 | 6.6159 | 5.6197 | 5.5561 | 5.7372 |
| 55 | 5.3466 | 5.0578 | 6.7784 | 6.7716 | 5.2506 | 5.3240 | 5.3572 |
| 60 | 4.9460 | 4.7068 | 6.0940 | 7.0555 | 4.7965 | 5.0717 | 4.8870 |
| 65 | 4.4439 | 4.2539 | 5.3381 | 6.9910 | 4.2581 | 4.6791 | 4.3334 |
| 70 | 3.8019 | 3.6564 | 4.5333 | 6.2417 | 3.6177 | 4.0581 | 3.6816 |
| 75 | 2.9942 | 2.9016 | 3.5767 | 5.0018 | 2.8325 | 3.2279 | 2.8845 |
| 80 | 2.0677 | 2.0262 | 2.4684 | 3.5067 | 1.9425 | 2.2562 | 1.9792 |
| 85 | 2.0412 | 2.0847 | 2.3928 | 3.6109 | 1.8732 | 2.3218 | 1.9094 |
| total | 100.0000 | 100.0000 | 100.0000 | 100.0000 | 100.0000 | 100.0000 | 100.0000 |
| m.ag | 37.0971 | 35.9564 | 43.6530 | 46.7766 | 36.3328 | 37.6374 | 36.8435 |
| sha | 200.0000 | 26.9866 | 4.7476 | 4.9636 | 63.3022 | 31.9502 | 68.0498 |
| lam | 1.022039 | 1.022039 | 1.022039 | 1.022039 | 1.022039 | 1.022039 | 1.022039 |
| r | 0.004360 | 0.004360 | 0.004360 | 0.004360 | 0.004360 | 0.004360 | 0.004360 |

*The first letter in the heading of each column refers to region of birth, the second to region of residence.

## APPENDIX E

Native-independent multistate projection to stability ($r = 0.004361$). Regional population disaggregated into stayers (stay), returners (return), established aliens (est. al), and recent aliens (rec. al).

**region of residence - south**

| age | total | stay | return | est.al | rec.al |
|---|---|---|---|---|---|
| 0 | 3411697. | 3323213. | 0. | 0. | 88484. |
| 5 | 3281607. | 3043111. | 2523. | 80968. | 155004. |
| 10 | 3204495. | 2845275. | 7861. | 221016. | 130343. |
| 15 | 3118879. | 2597105. | 18674. | 320859. | 182241. |
| 20 | 3021253. | 2308070. | 38226. | 447397. | 227560. |
| 25 | 2940080. | 2078979. | 60639. | 608277. | 192184. |
| 30 | 2874060. | 1917582. | 79436. | 738677. | 138365. |
| 35 | 2801331. | 1788790. | 93394. | 818593. | 100554. |
| 40 | 2719116. | 1675649. | 103586. | 861193. | 78689. |
| 45 | 2624978. | 1572615. | 110141. | 882495. | 59728. |
| 50 | 2522720. | 1471915. | 115131. | 882074. | 53600. |
| 55 | 2417691. | 1364559. | 121509. | 867496. | 64127. |
| 60 | 2299483. | 1243404. | 128976. | 849239. | 77863. |
| 65 | 2118588. | 1099914. | 130362. | 820264. | 68048. |
| 70 | 1839632. | 928322. | 120188. | 749780. | 41343. |
| 75 | 1466969. | 725095. | 99888. | 617987. | 23999. |
| 80 | 1027285. | 500875. | 71839. | 443624. | 10946. |
| 85 | 1058634. | 511537. | 75331. | 464289. | 7477. |
| total | 44748500. | 30996010. | 1377703. | 10674229. | 1700556. |

**percentage distribution**

| age | total | stay | return | est.al | rec.al |
|---|---|---|---|---|---|
| 0 | 7.6242 | 10.7214 | 0.0000 | 0.0000 | 5.2033 |
| 5 | 7.3334 | 9.8178 | 0.1832 | 0.7585 | 9.1149 |
| 10 | 7.1611 | 9.1795 | 0.5706 | 2.0706 | 7.6647 |
| 15 | 6.9698 | 8.3788 | 1.3554 | 3.0059 | 10.7165 |
| 20 | 6.7516 | 7.4463 | 2.7746 | 4.1914 | 13.3815 |
| 25 | 6.5702 | 6.7072 | 4.4014 | 5.6986 | 11.3013 |
| 30 | 6.4227 | 6.1865 | 5.7658 | 6.9202 | 8.1365 |
| 35 | 6.2602 | 5.7710 | 6.7790 | 7.6689 | 5.9130 |
| 40 | 6.0764 | 5.4060 | 7.5187 | 8.0680 | 4.6272 |
| 45 | 5.8661 | 5.0736 | 7.9945 | 8.2675 | 3.5123 |
| 50 | 5.6376 | 4.7487 | 8.3567 | 8.2636 | 3.1519 |
| 55 | 5.4028 | 4.4024 | 8.8197 | 8.1270 | 3.7709 |
| 60 | 5.1387 | 4.0115 | 9.3617 | 7.9560 | 4.5787 |
| 65 | 4.7344 | 3.5486 | 9.4622 | 7.6845 | 4.0015 |
| 70 | 4.1110 | 2.9950 | 8.7238 | 7.0242 | 2.4312 |
| 75 | 3.2783 | 2.3393 | 7.2504 | 5.7895 | 1.4112 |
| 80 | 2.2957 | 1.6159 | 5.2144 | 4.1560 | 0.6437 |
| 85 | 2.3657 | 1.6503 | 5.4678 | 4.3496 | 0.4397 |
| m.age | 37.94 | 32.71 | 56.29 | 51.82 | 31.42 |
| share | 100.00 | 69.27 | 3.08 | 23.85 | 3.80 |

## APPENDIX E (*continued*)

Native-independent multistate projection to stability ($r = 0.004361$)—continued.

region of residence - north
***************************

| age | total | stay | return | est.al | rec.al |
|---|---|---|---|---|---|
| 0 | 6670372. | 6573470. | 0. | 0. | 96902. |
| 5 | 6448941. | 6191142. | 4969. | 91555. | 161276. |
| 10 | 6299802. | 5917069. | 14720. | 241664. | 126350. |
| 15 | 6159685. | 5594918. | 36468. | 348240. | 180058. |
| 20 | 6028093. | 5230638. | 78035. | 493941. | 225480. |
| 25 | 5880246. | 4909523. | 123055. | 675491. | 172177. |
| 30 | 5712154. | 4645590. | 156783. | 802176. | 107604. |
| 35 | 5533321. | 4414482. | 181139. | 864623. | 73077. |
| 40 | 5335551. | 4193576. | 198126. | 890811. | 53038. |
| 45 | 5112815. | 3974697. | 207108. | 894626. | 36384. |
| 50 | 4849622. | 3738852. | 209373. | 875878. | 25520. |
| 55 | 4522238. | 3461032. | 206244. | 834620. | 20342. |
| 60 | 4120114. | 3129016. | 198649. | 772965. | 19484. |
| 65 | 3649363. | 2746801. | 187554. | 695805. | 19202. |
| 70 | 3095755. | 2306997. | 170543. | 600641. | 17575. |
| 75 | 2421656. | 1785927. | 142949. | 478726. | 14055. |
| 80 | 1659712. | 1214971. | 102588. | 335291. | 6861. |
| 85 | 1599979. | 1164990. | 102178. | 328103. | 4707. |
| total | 85099416. | 71193688. | 2320481. | 10225156. | 1360090. |

percentage distribution
***********************

| age | total | stay | return | est.al | rec.al |
|---|---|---|---|---|---|
| 0 | 7.8383 | 9.2332 | 0.0000 | 0.0000 | 7.1246 |
| 5 | 7.5781 | 8.6962 | 0.2141 | 0.8954 | 11.8578 |
| 10 | 7.4029 | 8.3112 | 0.6343 | 2.3634 | 9.2898 |
| 15 | 7.2382 | 7.8587 | 1.5716 | 3.4057 | 13.2387 |
| 20 | 7.0836 | 7.3471 | 3.3629 | 4.8306 | 16.5783 |
| 25 | 6.9099 | 6.8960 | 5.3030 | 6.6062 | 12.6592 |
| 30 | 6.7123 | 6.5253 | 6.7565 | 7.8451 | 7.9116 |
| 35 | 6.5022 | 6.2007 | 7.8061 | 8.4558 | 5.3730 |
| 40 | 6.2698 | 5.8904 | 8.5382 | 8.7120 | 3.8996 |
| 45 | 6.0080 | 5.5829 | 8.9252 | 8.7493 | 2.6751 |
| 50 | 5.6988 | 5.2517 | 9.0228 | 8.5659 | 1.8764 |
| 55 | 5.3141 | 4.8614 | 8.8880 | 8.1624 | 1.4956 |
| 60 | 4.8415 | 4.3951 | 8.5607 | 7.5594 | 1.4325 |
| 65 | 4.2884 | 3.8582 | 8.0826 | 6.8048 | 1.4118 |
| 70 | 3.6378 | 3.2405 | 7.3495 | 5.8741 | 1.2922 |
| 75 | 2.8457 | 2.5085 | 6.1603 | 4.6818 | 1.0334 |
| 80 | 1.9503 | 1.7066 | 4.4210 | 3.2791 | 0.5045 |
| 85 | 1.8801 | 1.6364 | 4.4033 | 3.2088 | 0.3461 |
| m.age | 36.65 | 34.48 | 53.83 | 49.35 | 25.42 |
| share | 100.00 | 83.66 | 2.73 | 12.02 | 1.60 |

# ENTROPY, MULTIPROPORTIONAL, AND QUADRATIC TECHNIQUES FOR INFERRING PATTERNS OF MIGRATION FROM AGGREGATE DATA

*Frans Willekens, András Pór, and Richard Raquillet*

## 1 INTRODUCTION

The lack of adequate regional statistics is frequently given as a reason for not endorsing the development of sophisticated regional or multiregional models. The data problem is particularly severe in the case of studies of interregional migration disaggregated by category of migrant.

Improved modeling of migration and of multiregional demographic phenomena in general can frequently only be carried out if additional data requirements are met. However, in cases where the collection of the necessary data is either impossible or too expensive, estimation procedures may sometimes be used as substitutes.

This paper presents a particular class of estimation methods which have great potential in migration analysis. They may be used whenever detailed migration flows (e.g., flows classified into various migrant categories) are required, but only aggregate data on migration patterns are initially available. For instance, how can we disaggregate a total origin–destination migration-flow matrix into age-specific flows if the only information available is a national estimate of the age composition of migrants? How would this estimate differ from one based on information about the age structure of migrants arriving in and departing from each region, or on an initial crude estimate of the age-specific flow matrices? Does additional information improve our estimates significantly? (This is of particular relevance in deciding on the extent of data collection required for a given migration analysis.) Questions like these may be answered by applying the techniques proposed in this paper. In addition, these methods are useful for updating migration tables, because this updating consists of nothing more than estimating migration flows under new conditions (constraints) when estimates under existing conditions are available.

The applicability of the methodology is not limited to migration studies alone, but may also be extended to input–output analysis, transportation analysis (e.g., trip distribution, freight flows), regional economics (e.g., journey-to-work tables), and other fields: in short, to the analysis of various kinds of interaction tables or contingency tables. In fact these techniques, in their simplified (two-dimensional) forms, have received considerable attention in the above-mentioned fields. Quite recently it was realized that the mathematical and statistical characteristics of the basic problem addressed in this paper have also been investigated in contingency-table analysis (the analysis of qualitative, cross-classified data) (see, for example, Bishop *et al.* 1975). A number of results from this research may be used to study the problem of estimating detailed migration flows from aggregate data and, in particular, to interpret the accuracy of the estimates. The relationship between the estimation methods presented in this paper and the techniques of contingency-table analysis has been discussed by Willekens (1980).

After this introduction, the paper consists of four main sections (Sections 2–5) and two appendixes. In Section 2, the general problem is dealt with in mathematical terms, and some special cases, which are of particular practical interest, are formulated. Section 3 presents the solutions to the basic problem and its variants. In Section 3.1 it is shown that, in terms of mathematical programming, the basic or primal problem is equivalent to another nonlinear problem (dual problem) that may be solved using standard methods. The solution algorithms, which are implemented on the computer, follow in Section 3.2. Section 4 presents some numerical illustrations, namely, the estimated values of interregional migration flows in Austria, disaggregated by age; the estimates are compared with the observed values to investigate the validity of the estimation procedure. A similar validity analysis using Swedish data is also discussed. Conclusions and some suggestions for further research are presented in Section 5. Proofs of five theorems used in the paper, and multiproportional solutions for three special cases of the entropy problem are given in Appendixes A and B, respectively. This paper is a shortened and modified version of an earlier report (Willekens *et al.* 1979), which contains additional applications and a description and listing of the computer program MULTENTROPY that was used to obtain the numerical results presented in this paper.

The entropy method discussed here (the 3F variant) was used in the IIASA Comparative Migration and Settlement Study to obtain estimates of interregional age-specific migration flows for countries with incomplete migration data (Bulgaria and the Netherlands). The applications to Bulgaria and the Netherlands have been described by Philipov (1978) and by Drewe and Willekens (1980), respectively. The method was also used by Tan (1980) to estimate missing migration data for Belgium.

## 2 PROBLEM FORMULATION

The estimation procedures proposed in this paper for inferring detailed migration patterns from aggregate data have one common feature: the aggregate data appear as sums (hereafter referred to as "marginal sums") of different groups of elements of two- or $n$-dimensional arrays, the elements of which are unknown and must be estimated. (In the two-dimensional case these marginal sums usually represent row or column sums.) Two main groups of problems may be distinguished: (i) the entropy problem, and (ii) the quadratic adjustment problem.

(i) *The entropy problem.* The problem here is to produce a "maximally unbiased" estimate of the elements of an array under the given marginal conditions. In the application of entropy models, two model types may be distinguished: (a) entropy maximizing models, and (b) information-divergence (I-divergence) minimizing models.

(a) In the entropy maximizing problem one tries to determine the "most probable" elements of an array, under the given marginal conditions. No initial array is known *a priori* and the values of the elements of the array are seen as equally likely, apart from the marginal constraints specified. The entropy maximizing method was introduced in regional science by Wilson (1967, 1970). The two-dimensional case has received considerable attention in the literature and has been elaborated in several ways to recover interregional flow matrices (of people or of commodities) from various forms of aggregate data (Chilton and Poet 1973, Evans and Kirby 1974, Nijkamp 1975, Willekens 1977).

(b) In the I-divergence minimizing problem one tries to estimate an "*a posteriori*" array which is as "close" as possible to an "*a priori*" array, and which satisfies some given constraints (row and column sums).
The distance function used here to measure the "closeness" of the arrays is the I-divergence or Kullback–Leibler information number (Kullback 1959), also called information for discrimination, *information gain*, or entropy of an *a posteriori* distribution relative to an *a priori* distribution (Renyi 1970). If the known *a priori* array is uniformly distributed, i.e., all elements of the array are equal, then the I-divergence measure is equivalent to the negentropy (entropy with a negative sign). I-divergence with a negative sign was also defined as a measure of the average conditional entropy (Nijkamp and Paelinck 1974a, Theil 1967).
A procedure for minimizing I-divergence in spatial interaction models has been developed by Batty and March (1976). In the two-dimensional case,* the biproportional adjustment method, better known as the RAS method, is the technique most extensively used to solve this type of problem. It was developed independently by Leontief and Stone to update input–output tables and has been studied in detail by Bacharach (1970). Hewings (1977), Thumann (1978), and Hewings and Janson (1980) have provided recent evaluations of RAS and related techniques with special emphasis on their capabilities for exchanging, updating, and predicting regional input–output coefficients. The RAS method is equivalent to the Furness method used in traffic models (e.g., Evans 1970, Evans and Kirby 1974). In the analysis of contingency tables the method is known as iterative proportional fitting and is attributed to Bartlett in a paper written as early as 1935.

(ii) *The quadratic adjustment problem.* This problem applies to situations where initial estimates of the entries of an array are available but where the estimates do not conform with predefined (measured) row and column totals. The reason may be that different sources have been used to calculate the estimates and the totals. The sum-constrained adjustment problem is then to find the array which is as close as possible to the initial array and also satisfies the predefined totals.

The distance function used here to measure the "closeness" of the arrays is a quadratic-type function. Examples can easily be found to show that, unlike the I-divergence minimizing problem, the fact that the elements of the initial array are positive does not ensure that the unique solution of the quadratic adjustment problem will also be positive. To overcome this weakness of the quadratic adjustment method, we suggest an adjustment technique which is not quadratic but which is very closely related to the quadratic adjustment method.

* A two-dimensional array is a matrix.

## 2.1 Entropy Maximization

The techniques described in this section address problems in which no input matrix is given *a priori.* Wilson (1967, 1970) used an index to produce the matrix entries which are *most probable.* This index is called the entropy of the matrix.

### *2.1.1 The Entropy Concept*

Suppose that we are given the total number of arrivals $I_j$ and departures $O_i$ disaggregated by region in a two-region system, and that the problem is to estimate the complete origin–destination migration-flow matrix **M** with elements $m_{ij}$. For example, to estimate

$$\mathbf{M} = \begin{bmatrix} m_{11} & m_{21} \\ m_{12} & m_{22} \end{bmatrix}$$

where

$$I_1 = m_{11} + m_{21} = m_{.1} = 3 \qquad O_1 = m_{11} + m_{12} = m_{1.} = 4$$

$$I_2 = m_{12} + m_{22} = m_{.2} = 3 \qquad O_2 = m_{21} + m_{22} = m_{2.} = 2$$

$$m_{11} + m_{12} + m_{21} + m_{22} = m_{..} = 6$$

In contrast to the biproportional adjustment method, no initial estimates of the matrix elements are available. Therefore, our information is limited to the row and column totals of the matrix to be estimated. For given row and column totals of a matrix, there may be a large number of arrangements of entries that satisfy the marginal conditions. For example, if for a $2 \times 2$ migration matrix the row sums are three and three, and the column sums are four and two, then there are three possible arrangements of the entries

$$\mathbf{M}_a = \begin{bmatrix} 3 & 0 \\ 1 & 2 \end{bmatrix} \qquad \mathbf{M}_b = \begin{bmatrix} 2 & 1 \\ 2 & 1 \end{bmatrix} \qquad \mathbf{M}_c = \begin{bmatrix} 1 & 2 \\ 3 & 0 \end{bmatrix}$$

Each arrangement of the entries of **M** is called a *macrostate* of the system.

The true migration flow is represented by one of the three macrostates, $\mathbf{M}_a$, $\mathbf{M}_b$, or $\mathbf{M}_c$. Given the limited information we have about the migration behavior, we do not know which macrostate is the true one. Therefore, we must make a guess. It is here that the entropy method is useful. It selects the macrostate which has the highest probability of occurring. A certain macrostate may be generated by various so-called *microstates.* A microstate is an assignment of *individual* migrants to the origin–destination table. In other words, a microstate is a description of the location of every individual in the system, whereas a macrostate gives the number of people in each cell of the table. Consider, for example, the matrix $\mathbf{M}_a$, and denote the individual migrants by $m1$, $m2$, $m3$, $m4$, $m5$, and $m6$. According to $\mathbf{M}_a$, three people migrate from region 1 to region 1, i.e., move within the region. We can select these three from the six migrants in 20 different ways. The number of possible combinations of three people from six can easily be computed using the well-known combinatorial formula

$${}^3C_6 = \frac{6!}{3!(6-3)!}$$

Once we have chosen three people to constitute $m_{11}$, we must select one person out of the remaining three to constitute $m_{12}$. There are only three possible ways of doing this. Finally, the two remaining individuals constitute $m_{22}$, since $m_{21}=0$. Therefore, the total number of ways of selecting three out of six, one out of the remaining three, and two out of the remaining two, is

$$\frac{6!}{3!(6-3)!}\,\frac{3!}{1!(3-1)!}\,\frac{2!}{2!}=60$$

Each of the 60 ways constitutes a separate microstate, or assignment of individuals. In general, the number of ways in which we can select a *particular* macrostate from the total number of migrants $m_{..}$ is given by the combinatorial formula

$$W=\frac{m_{..}!}{m_{11}!\,m_{12}!\,m_{21}!\,m_{22}!} \tag{1}$$

$$W=\frac{m_{..}!}{\prod\limits_{i,j} m_{ij}!}$$

Using eqn. (1), we obtain $W=60$ for $\mathbf{M}_a$, $W=180$ for $\mathbf{M}_b$, and $W=60$ for $\mathbf{M}_c$. The value $W$ is the number of microstates which give rise to a particular macrostate, and is called the *entropy* of the macrostate.

The macrostate $\mathbf{M}_b$ with the highest entropy value is then chosen as the best estimate of the true migration flow. The use of this selection criterion relies on two critical assumptions:

1. The probability that a macrostate represents the true migration-flow matrix is proportional to the number of microstates of the system which give rise to this macrostate (entropy) and which satisfy the marginal conditions.
2. Each microstate is equally probable.

The formal solution to the two-dimensional entropy maximization problem is derived in the next section.

### 2.1.2 *Solving the Entropy Problem*

The first assumption given above may be stated in a slightly different way: the true arrangement of a system is one which maximizes the entropy, i.e., one in which the elements can be organized in as many ways as possible (maximum "disorder"). This is the second law of thermodynamics. The analogy between the behavior of social and physical systems is not accidental. Several authors (e.g., Isard 1960) have attempted to describe social phenomena by laws borrowed from physics. This approach, known as *social physics*, was developed in the early regional-science literature.

The use of the entropy concept in the social sciences may be illustrated by information theory (Jaynes 1957), by Bayes's theorem for conditional probabilities (Hyman 1969), and by the use of maximum-likelihood estimators (Evans 1971, Batty and MacKie 1972).* In information theory, entropy represents expected

* For a comparison, see Wilson (1970, pp. 1–10) and Nijkamp (1977, pp. 18–20).

information. It indicates the degree of uncertainty about the occurrence of events in information systems. Hence, a high entropy value (low uncertainty) is associated with events which are likely to occur. In the maximum-likelihood approach, entropy maximization is equivalent to maximizing the likelihood of a macrostate.

The problem of estimating the most probable migration-flow matrix to satisfy given row and column totals may now be formulated as follows: find the macrostate with maximum entropy $W$, subject to the marginal conditions. The solution is given by

$$\max W = \frac{m_{..}!}{\prod_{i,j} m_{ij}!} \tag{2}$$

subject to

$$\sum_j m_{ij} = m_{i.} = O_i \qquad \text{for all } i \tag{3}$$

$$\sum_i m_{ij} = m_{.j} = I_j \qquad \text{for all } j \tag{4}$$

Since the maximum of eqn. (2) coincides with the maximum of any monotonic function of $W$, we may replace $W$ by the Naperian logarithm of $W$ ($\ln W$) in the objective function

$$\begin{aligned} \ln W &= \ln m_{..}! - \ln \prod_{i,j} m_{ij}! \\ &= \ln m_{..}! - \sum_i \sum_j \ln m_{ij}! \end{aligned} \tag{5}$$

To make differentiation of the complex function (5) easier, we replace $\ln m_{ij}!$ by Stirling's approximation

$$\ln m_{ij}! = m_{ij} \ln m_{ij} - m_{ij}$$

Since $\ln m_{..}!$ is a constant, we may write the objective function as

$$\max \ln \hat{W} = -\sum_i \sum_j (m_{ij} \ln m_{ij} - m_{ij}) \tag{6}$$

or equivalently

$$\min \ln \hat{W} = \sum_i \sum_j m_{ij} \ln m_{ij} - m_{ij} \tag{7}$$

## 2.2 I-Divergence Minimization (Biproportional Adjustment)

The biproportional adjustment method, independently developed by Leontief (1941) and Stone (1963), and equivalent to the iterative proportional fitting procedure of Bartlett (1935), uses the I-divergence measure as a distance function

$$I(\mathbf{M} \| \mathbf{M}^0) = \sum_i \sum_j m_{ij} \ln (m_{ij}/m_{ij}^0) \tag{8}$$

which is defined for $m_{ij}^0 \neq 0$. This technique is better known as the RAS method (Stone 1963).

The basic features of the biproportional adjustment problem are described in this section. Quite often our information is not limited to the row and column sums of a two-dimensional migration matrix; we may also know the migration patterns of specific categories of the population (e.g., sexes, age groups). This information may then be used to adjust the original estimates. A more general version of the biproportional adjustment process, the multiproportional adjustment process, allows more *a priori* information to be taken into consideration; this is treated in Section 2.5.

Suppose we are given the total number of migrants arriving in each region, and departing from each region ($I_j$ and $O_i$, respectively) in a two-region system, and that, as before, the problem is to estimate the complete origin–destination migration-flow matrix $\mathbf{M}$ with elements $m_{ij}$. For example, to estimate

$$\mathbf{M} = \begin{bmatrix} m_{11} & m_{21} \\ m_{12} & m_{22} \end{bmatrix}$$

where

$$I_1 = m_{.1} \qquad O_1 = m_{1.}$$

$$I_2 = m_{.2} \qquad O_2 = m_{2.}$$

Suppose we are also given initial estimates of the elements $m_{ij}$; these estimates are the elements $m^0_{ij}$ of the matrix $\mathbf{M}^0$

$$\mathbf{M}^0 = \begin{bmatrix} m^0_{11} & m^0_{21} \\ m^0_{12} & m^0_{22} \end{bmatrix}$$

The initial estimates may be derived from migration tables of previous years, from experts' opinions, or from other sources.* However, the column sums $m^0_{i.}$ and the row sums $m^0_{.j}$ of $\mathbf{M}^0$ are not necessarily equal to the predefined number of departures $O_i = m_{i.}$, and arrivals $I_j = m_{.j}$. We therefore have to adjust the elements of $\mathbf{M}^0$ so that they add up to the required totals. The sum-constrained biproportional adjustment problem may be formulated as follows:

Find the $p \times q$ matrix $\mathbf{M}$ such that the I-divergence measure is minimized (in migration tables, $p = q$):

$$\min I(\mathbf{M} \| \mathbf{M}^0) \tag{9}$$

subject to

$$m_{i.} = \sum_j m_{ij} = O_i \qquad (i = 1, 2, \ldots, q) \tag{10}$$

$$m_{.j} = \sum_i m_{ij} = I_j \qquad (j = 1, 2, \ldots, p) \tag{11}$$

Note that since $\sum_i O_i = \sum_j I_j$, the $p + q$ row and column totals are not independent. The number of independent constraints is $p + q - 1$. Constraints (10) and (11) are found in all existing methods for constrained adjustment of a two-dimensional array, i.e., of a matrix.

---

* This methodology has been particularly successful for updating input–output tables.

## 2.3 A Comparison of Entropy Maximization and I-Divergence Minimization

This section compares the entropy maximizing and I-divergence minimizing techniques. Both techniques have the same set of constraints,* and any differences may therefore be explained by differences in the objective functions.

(i) Note that the entropy objective function to be maximized

$$\hat{W} = -\sum_i \sum_j (m_{ij} \ln m_{ij} - m_{ij})$$

is also given by

$$\hat{W} = -\sum_i \sum_j (m_{ij} \ln m_{ij}) + m_{..}$$

Since $m_{..}$ is a constant, the maximization of $\hat{W}$ gives the same result as the maximization of the simpler function $\hat{\hat{W}} = -\sum_i \sum_j m_{ij} \ln m_{ij}$. The objective function $\hat{\hat{W}} = -\sum_i \sum_j m_{ij} \ln m_{ij}$ is in fact more widely used. If $m_{ij}$ is interpreted as a probability by scaling $(\sum_i \sum_j m_{ij} = 1)$, then the objective function defines a quantity called *statistical entropy* or the entropy of the probability distribution. This quantity (multiplied by a constant) has been defined by Shannon as a measure of the uncertainty contained in a probability distribution (Shannon and Weaver 1949).

(ii) The biproportional process differs from the entropy maximizing process in that it contains a term $m_{ij}^0$ representing an initial estimate of the elements. Again, if $m_{ij}$ and $m_{ij}^0$ can be interpreted as probabilities, the quantity**

$$W = \sum_i \sum_j m_{ij} \ln (m_{ij}/m_{ij}^0)$$

represents the *information divergence.*

(iii) It appears that all these objective functions belong to the same family and merely express the concept of entropy in different ways; the same basic idea is common to biproportional "minimum deviation" and to statistical entropy. In information theory, the biproportional objective function is interpreted as a measure of the "surprise" produced by new values when compared with old values. Therefore, minimizing the "surprise" or the deviation from our initial information and finding the most probable flows are the basic aims of both the entropy and the biproportional approaches. Both of these techniques measure the deviation with the same function

$$\sum_i \sum_j m_{ij} \ln (m_{ij}/m_{ij}^0) \tag{12}$$

when *a priori* information is available, and use the function

$$\sum_i \sum_j m_{ij} \ln m_{ij} \tag{13}$$

* If a cost constraint is considered in an entropy problem, then the reciprocal of the cost function plays the role of an *a priori* array $\mathbf{M}^0$ (Willekens 1980).

** When some *a priori* information exists and is expressed in terms of *a priori* probabilities $(p_i^0, i = 1, 2, \ldots, n)$, the expected value or information content of a message changing *a priori* information into *a posteriori* information is measured by $-\sum_i p_i \ln (p_i/p_i^0)$ (Jaynes 1957, Theil 1967).

when no such information exists. Therefore, entropy maximizing problems can generally be formulated as I-divergence minimizing problems

$$\min d[\mathbf{M}, \mathbf{M}^0] = \sum_i \sum_j m_{ij} \ln (m_{ij}/m_{ij}^0) \tag{14}$$

where $m_{ij}^0$ can be uniformly set equal to 1.*

## 2.4 Quadratic Adjustment

We may formulate the sum-constrained quadratic adjustment problem in the following way.

Find a matrix **M** such that a quadratic distance measure $d[\mathbf{M}, \mathbf{M}^0]$ is minimized, subject to

$$\sum_j m_{ij} = O_i \qquad \text{for all } i$$

$$\sum_i m_{ij} = I_j \qquad \text{for all } j$$

The techniques may be distinguished by the distance measures $d[\mathbf{M}, \mathbf{M}^0]$ used.

(i) *Least-squares adjustment.* The most obvious distance measure is the euclidean norm, defined as

$$d[\mathbf{M}, \mathbf{M}^0] = \tfrac{1}{2} \sum_i \sum_j (m_{ij} - m_{ij}^0)^2 \tag{15}$$

(ii) *Friedlander adjustment.* Friedlander (1961) used a distance norm of the $\chi^2$ type

$$d[\mathbf{M}, \mathbf{M}^0] = \tfrac{1}{2} \sum_i \sum_j (m_{ij} - m_{ij}^0)^2 / m_{ij}^0 \tag{16}$$

to adjust contingency tables. A related method has been developed by Hortensius (1970). The distance norm is the weighted deviation

$$d[\mathbf{M}, \mathbf{M}^0] = \sum_i \sum_j (m_{ij} - m_{ij}^0)^2 / s_{ij} \tag{17}$$

where $1/s_{ij}$ is the weight corresponding to the difference $(m_{ij} - m_{ij}^0)$. The weight used is a measure of the uncertainty. The underlying idea is that the more accurate $m_{ij}^0$ is, the less adjustment is needed. Hence, for accurate estimates of $m_{ij}^0$, the weight $1/s_{ij}$ should be small. If the elements of $m_{ij}^0$ are estimated from a sample, then both the estimate or mean $m_{ij}^0$ and the whole frequency distribution are known. An appropriate value for $s_{ij}$ is therefore the variance of the distribution (assuming normal distribution). This measure of uncertainty is used by Hortensius.

---

* In general, the I-divergence or multiproportional problem reduces to an entropy problem whenever the elements of the initial array are multiproportional (with the uniform distribution as a special case). The results are therefore independent of the initial array. Friedmann (1978) recently showed that the RAS solution is the same for different initial biproportional matrices, although he did not connect this with the entropy problem. MacGill (1977) provides a further discussion on the formal equivalence between entropy and biproportional adjustment methods.

(iii) *Modified Friedlander adjustment.* An important weakness of the least-squares and Friedlander approaches is that a strictly positive matrix $\mathbf{M}^0$ does not necessarily yield a strictly positive matrix $\mathbf{M}$ of estimates. Some elements of $\mathbf{M}$ may be negative. To overcome this failure, we suggest the following measure, which is no longer quadratic:

$$d[\mathbf{M}, \mathbf{M}^0] = \tfrac{1}{2} \sum_i \sum_j (m_{ij} - m_{ij}^0)^2 / m_{ij}$$

## 2.5 Generalization of Flow-Estimation Problems to *N*-Dimensional Arrays

The formulation of the problem can easily be extended to more than two dimensions. Suppose that we are given the migration-flow matrix of the total population, and that we are interested in the migration patterns of subsets of the population, e.g., sexes, age groups, nationalities, professional categories. Suppose that we also know the number of arrivals and departures made by each subset in each region. The migration from $i$ to $j$ by category $k$ is denoted by $m_{ijk}$. The total migration from $i$ to $j$ is $m_{ij.} = c_{ij}$, the number of departures from $i$ by category $k$ is $m_{i.k} = b_{ik}$, and the number of arrivals in $j$ by category $k$ is $m_{.jk} = a_{jk}$. Therefore, the following marginal constraints must be met:

$$\sum_{i=1}^{n} m_{ijk} = a_{jk} \qquad (j = 1, 2, \ldots, m;\ k = 1, 2, \ldots, l)$$

$$\sum_{j=1}^{m} m_{ijk} = b_{ik} \qquad (i = 1, 2, \ldots, n;\ k = 1, 2, \ldots, l)$$

$$\sum_{k=1}^{l} m_{ijk} = c_{ij} \qquad (i = 1, 2, \ldots, n;\ j = 1, 2, \ldots, m)$$

The information available is, however, not always presented in this way. For example, in an extreme case, we may not know the flow matrix of the total population but only the total number of arrivals and departures disaggregated by region; we may know the composition of the migrant categories only at the national level. In this case the constraints would take on a different form:

$$\sum_{i=1}^{n} \sum_{j=1}^{m} m_{ijk} = u_k \qquad (k = 1, 2, \ldots, l)$$

$$\sum_{i=1}^{n} \sum_{k=1}^{l} m_{ijk} = v_j \qquad (j = 1, 2, \ldots, m)$$

$$\sum_{j=1}^{m} \sum_{k=1}^{l} m_{ijk} = w_i \qquad (i = 1, 2, \ldots, n)$$

where $u_k = m_{..k}$ is the total number of migrants in category $k$, $v_j = m_{.j.}$ is the total number of arrivals in region $j$, and $w_i = m_{i..}$ is the total number of departures from region $i$.

Various combinations of bivariate and univariate marginal sums are possible: a known total-flow matrix with the composition of migrant categories at the national level, numbers of arrivals and departures disaggregated by type of migrant only, etc.

Before proceeding to specific problems, we may formulate our problem in general terms. The mathematical formulation of the *basic adjustment problem* is as follows:

$$\min \sum_i \sum_j \sum_k d[m_{ijk}, m^0_{ijk}] \tag{18}$$

subject to

$$\sum_i m_{ijk} = a_{jk} \qquad (\forall j \in J, k \in K) \tag{19}$$

$$\sum_j m_{ijk} = b_{ik} \qquad (\forall i \in I, k \in K) \tag{20}$$

$$\sum_k m_{ijk} = c_{ij} \qquad (\forall j \in J, i \in I) \tag{21}$$

$$\sum_i \sum_j m_{ijk} = u_k \qquad (\forall k \in K) \tag{22}$$

$$\sum_i \sum_k m_{ijk} = v_j \qquad (\forall j \in J) \tag{23}$$

$$\sum_j \sum_k m_{ijk} = w_i \qquad (\forall i \in I) \tag{24}$$

$$\sum_k u_k = \sum_j v_j = \sum_i w_i = ST \tag{25}$$

$$m_{ijk} \geq 0 \qquad (\forall i \in I, j \in J, k \in K) \tag{26}$$

where $I = \{1, 2, \ldots, n\}$, $J = \{1, 2, \ldots, m\}$, and $K = \{1, 2, \ldots, l\}$ are the index sets; $a_{jk}$, $b_{ik}$, and $c_{ij}$ are bivariate marginal sums; $u_k$, $v_j$, and $w_i$ are univariate marginal sums; and $ST$ is the total sum.

The elements to be estimated, $m_{ijk}$, may be arranged in a three-dimensional array, $\mathbf{M} = [m_{ijk}]$; the initial estimates constitute the array $\mathbf{M}^0 = [m^0_{ijk}]$. Both arrays contain only non-negative elements. Some of the elements $m_{ijk}$ may be known exactly. For instance, if intraregional migration is not considered, then the diagonal elements $m_{iik} = m^0_{iik} = 0$. If other migration flows are known *a priori* (i.e., are fixed to the initial estimate $m^0_{ijk}$), we have to consider the cell constraints

$$m_{ijk} = m^0_{ijk} \qquad (i, j, k) \notin \Gamma \tag{27}$$

where the set $\Gamma$ is defined by setting $\Gamma = \{(i, j, k) | \text{migration from } i \text{ to } j \text{ by category } k \text{ is possible and not fixed}\}$.

The right-hand sides of the constraints are matrices $\mathbf{A} = [a_{jk}]$, $\mathbf{B} = [b_{ik}]$, and $\mathbf{C} = [c_{ij}]$, and vectors $\boldsymbol{U} = [u_k]$, $\boldsymbol{V} = [v_j]$, and $\boldsymbol{W} = [w_i]$. We shall call the constraints (19)–(21) *face constraints*, because they present given values for the three faces of the "cube" or three-dimensional array $\mathbf{M}$. Analogously, the univariate constraints (22)–(24) will be labeled *edge constraints* because they allocate values to the edges of the "cube" or three-dimensional array $\mathbf{M}$. If the face constraints are given, the edge constraints are redundant since the elements on the edges are the sums of the elements on the faces. We shall therefore refer to the basic problem as the "three prescribed faces" problem, or, more concisely, the "three-face (3F)" problem.

Three special cases are of particular interest:

1. The two-face (2F) problem—minimize (18) subject to (19), (21), (26), and (27);
2. The one-face, one-edge (1FE) problem—minimize (18) subject to (21), (22), (26), and (27);
3. The three-edge (3E) problem—minimize (18) subject to (22)–(26), and (27).

To find the best estimates of the elements $m_{ijk}$, given the constraints and given an array of initial estimates $m^0_{ijk}$, we generalize the different distance measures as follows:

(i) *Three-dimensional I-divergence measure*

$$d[\mathbf{M}, \mathbf{M}^0] = \sum_{(i,j,k)\in\Gamma} m_{ijk} \ln (m_{ijk}/m^0_{ijk}) \tag{28}$$

Note that the entropy function is equivalent to the I-divergence measure when the original array $m^0_{ijk}$ over the index set $\Gamma$ consists of ones. The problem of minimizing the I-divergence measure subject to various marginal-sum constraints is the *multi-proportional adjustment problem.*

(ii) *Three-dimensional $\chi^2$ measure*

$$d[\mathbf{M}, \mathbf{M}^0] = \tfrac{1}{2} \sum_{(i,j,k)\in\Gamma} (m_{ijk} - m^0_{ijk})^2/m^0_{ijk} \tag{29}$$

The problem of minimizing the $\chi^2$ measure subject to various marginal-sum constraints is the *multidimensional Friedlander adjustment problem.*

(iii) *Modified three-dimensional $\chi^2$ measure*

$$d[\mathbf{M}, \mathbf{M}^0] = \tfrac{1}{2} \sum_{(i,j,k)\in\Gamma} (m_{ijk} - m^0_{ijk})^2/m_{ijk} \tag{30}$$

It is understood that $(0 - m^0_{ijk})^2/0 = 0$ if $m^0_{ijk} = 0$, and $(0 - m^0_{ijk})^2/0 = +\infty$ if $m^0_{ijk} > 0$. The problem of minimizing the modified $\chi^2$ measure subject to various marginal-sum constraints is the *modified multidimensional Freidlander adjustment problem.*

In the next section, the basic 3F problem and its variants, considered as nonlinear mathematical programming problems (primal problems), are converted into equivalent dual problems by the duality correspondence, in order to derive the solution algorithms more easily. In establishing the duality results and the solution algorithms for our basic adjustment problem, we rely on results obtained by Rockafellar (1970) in the field of "perturbation" functions and separable programming. We assume that there exists a strictly positive feasible solution [$m_{ijk} > 0$ for all $(i, j, k) \in \Gamma$] which satisfies the constraints (19)–(26).

We prove the existence of a strictly positive feasible solution for the case in which all migration flows are possible in Appendix A. In this case we also estimate the number of people remaining in the same region; the same proof applies however when migration flows from one region to the same region are impossible. For the case in which no strictly positive feasible solution exists, we can establish an asymptotic duality result and prove the convergence of an iterative solution procedure for the I-divergence minimizing problem.

## 3 SOLUTION OF THE MULTIPROPORTIONAL AND THE MODIFIED FRIEDLANDER (QUADRATIC) ADJUSTMENT PROBLEMS

In order to solve the multiproportional and the modified multidimensional Friedlander adjustment problems, we first derive the corresponding dual problems, for whose solution simple algorithms have been developed. A solution algorithm for the primal multiproportional adjustment problem is also given.

### 3.1 Duality Results

Before deriving the dual problems, we formulate a number of theorems that demonstrate the existence and uniqueness of the optimal solutions. Proofs of these theorems are given in Appendix A.

*Theorem 1*

If the solution set defined by constraints (19)–(27) contains a feasible solution $\mathbf{M} = [m_{ijk}]$, such that

$$m_{ijk} > 0 \qquad (i, j, k) \in \Gamma$$

where $\Gamma$ is the index set for which $m_{ijk}$ is not fixed, then both the multiproportional and the modified multidimensional Friedlander adjustment problems have unique optimal solutions. Further, the optimal solutions are strictly positive for all indices $(i, j, k) \in \Gamma$.

In order to establish the duality results, we introduce some new notation:

$$I(j, k) = \{i \in I \mid (i, j, k) \in \Gamma\} \qquad (\forall j \in J, k \in K)$$

$$J(i, k) = \{j \in J \mid (i, j, k) \in \Gamma\} \qquad (\forall i \in I, k \in K)$$

$$K(i, j) = \{k \in K \mid (i, j, k) \in \Gamma\} \qquad (\forall i \in I, j \in J)$$

Further,

$$\mathbf{\Lambda} = (\lambda_{ij})_{i,j=1}^{n,m}$$

$$\mathbf{N} = (\nu_{ik})_{i,k=1}^{n,l}$$

$$\mathbf{H} = (\zeta_{jk})_{j,k=1}^{m,l}$$

are real-valued matrices denoting the Lagrangian multipliers for the face constraints.

(i) *Unconstrained dual multiproportional adjustment problem*

Minimize

$$L_1(\mathbf{\Lambda}, \mathbf{N}, \mathbf{H}) = \sum_{(i,j,k)\in\Gamma} m_{ijk}^0 \exp\left[-(1 + \lambda_{ij} + \nu_{ik} + \zeta_{jk})\right] + \sum_{i=1}^{n} \sum_{j=1}^{m} \bar{c}_{ij}\lambda_{ij} + \sum_{i=1}^{n} \sum_{k=1}^{l} \bar{b}_{ik}\nu_{ik} + \sum_{j=1}^{m} \sum_{k=1}^{l} \bar{a}_{jk}\zeta_{jk} \tag{31}$$

subject to

$$\lambda_{11} = 1$$

$$\lambda_{ij}, \nu_{ik}, \zeta_{jk} \in R \qquad (\forall i \in I, j \in J, k \in K)$$

where

$$\bar{c}_{ij} = c_{ij} - \sum_{k \notin K(i,j)} m^{0}_{ijk}$$

$$\bar{b}_{ik} = b_{ik} - \sum_{j \notin J(i,k)} m^{0}_{ijk}$$

$$\bar{a}_{jk} = a_{jk} - \sum_{i \notin I(j,k)} m^{0}_{ijk}$$

*Theorem 2*

Let the multiproportional adjustment problem have an optimal solution $\hat{\mathbf{M}} = (\hat{m}_{ijk})$ such that

$$\hat{m}_{ijk} > 0 \qquad (i, j, k) \in \Gamma$$

If $(\bar{\mathbf{\Lambda}}, \bar{\mathbf{N}}, \bar{\mathbf{H}})$ is an optimal solution of the dual problem (31), then

$$\hat{m}_{ijk} = m^{0}_{ijk} \exp\left[-(1 + \bar{\lambda}_{ij} + \bar{\nu}_{ik} + \bar{\zeta}_{jk})\right]$$

(ii) *Dual modified multidimensional Friedlander adjustment problem*
Minimize

$$L_2(\mathbf{\Lambda}, \mathbf{N}, \mathbf{H}) = -2 \sum_{(i,j,k) \in \Gamma} m^{0}_{ijk}(1 + \lambda_{ij} + \nu_{ik} + \zeta_{jk})^{1/2} + \sum_{i=1}^{n} \sum_{j=1}^{m} \bar{c}_{ij}\lambda_{ij} + \sum_{i=1}^{n} \sum_{k=1}^{l} \bar{b}_{ik}\nu_{ik} + \sum_{j=1}^{m} \sum_{k=1}^{l} \bar{a}_{jk}\zeta_{jk} + 2|\Gamma| \qquad (32)$$

subject to

$$\lambda_{11} = 1$$

$$\lambda_{ij}, \nu_{ik}, \zeta_{jk} \in R \qquad (\forall i \in I, j \in J, k \in K)$$

$$\lambda_{ij} + \nu_{ik} + \zeta_{jk} > -1$$

*Theorem 3*

Let the modified multidimensional Friedlander adjustment problem have an optimal solution $\hat{\mathbf{M}} = (\hat{m}_{ijk})$ such that

$$\hat{m}_{ijk} > 0 \qquad (i, j, k) \in \Gamma$$

If $(\bar{\mathbf{\Lambda}}, \bar{\mathbf{N}}, \bar{\mathbf{H}})$ is an optimal solution of the dual problem (32) then

$$\hat{m}_{ijk} = \frac{m^{0}_{ijk}}{(1 + \bar{\lambda}_{ij} + \bar{\nu}_{ik} + \bar{\zeta}_{jk})^{1/2}}$$

## 3.2 Solution Algorithms

The algorithms presented here solve the multiproportional and the modified multidimensional Friedlander adjustment problems for the basic three-face (3F) case. The solutions for the special cases (3E), (1FE), and (2F) may be obtained in a similar way.*

### *3.2.1 Multiproportional Adjustment Problem*

*Algorithm for solving the unconstrained dual problem (31)*

*Step 0*

Set

$S(\text{step}) = 0$

$\lambda_{ij}^{(0)} = 1$

$\nu_{ik}^{(0)} = 1 \qquad (\forall i \in I, j \in J, k \in K)$

$\zeta_{jk}^{(0)} = 1$

*Step 1*

Set

$\lambda_{11}^{(S+1)} = \lambda_{11}^{(S)} = 1$

$$\lambda_{ij}^{(S+1)} = \ln \sum_{k \in K(i,j)} m_{ijk}^{0} \exp[-(1 + \nu_{ik}^{(S)} + \zeta_{jk}^{(S)})] - \ln c_{ij}$$

for all $i \in I, j \in J$, except for $i = j = 1$.

*Step 2*

Set

$$\nu_{ik}^{(S+1)} = \ln \sum_{j \in J(i,k)} m_{ijk}^{0} \exp[-(1 + \lambda_{ij}^{(S+1)} + \zeta_{jk}^{(S)})] - \ln b_{ik}$$

*Step 3*

Set

$$\zeta_{jk}^{(S+1)} = \ln \sum_{i \in I(j,k)} m_{ijk}^{0} \exp[-(1 + \lambda_{ij}^{(S+1)} + \nu_{ik}^{(S+1)})] - \ln a_{jk}$$

If

$|\lambda_{ij}^{(S+1)} - \lambda_{ij}^{(S)}| \leq \varepsilon$

$|\nu_{ik}^{(S+1)} - \nu_{ik}^{(S)}| \leq \varepsilon \qquad (\forall i \in I, j \in J, k \in K)$

$|\zeta_{jk}^{(S+1)} - \zeta_{jk}^{(S)}| \leq \varepsilon$

is fulfilled then STOP, otherwise $S \leftarrow S + 1$, and go to Step 1.

---

* If the elements $m_{ijk}^{0}$ are uniformly distributed, then the special cases of the multiproportional adjustment problem become entropy problems with closed-form solutions (Appendix B).

*Theorem 4*

Let $\mathbf{M} = (m_{ijk})$ be a feasible solution of the basic problem such that

$$m_{ijk} > 0 \qquad (i, j, k) \in \Gamma$$

Then the above algorithm converges to an optimal solution of the unconstrained dual problem (31).

Using the result of Theorem 2, we can easily derive the direct primal algorithm.

*Algorithm for solving the primal multiproportional adjustment problem*

*Step 0*

Set

$$m_{ijk}^{(0)} = m_{ijk}^{0} \qquad (\forall i \in I, j \in J, k \in K)$$
$$S = 0$$

In the entropy problem, the *a priori* distribution $m_{ijk}^{0}$ is not known and may therefore be set uniformly equal to unity, i.e., $m_{ijk}^{0} = 1$, $\forall i, j, k$.

*Step 1*

Set

$$m_{ijk}^{(3S+1)} = m_{ijk}^{(3S)} c_{ij} \Big/ \sum_{k=1}^{l} m_{ijk}^{(3S)}$$

for all $i \in I$, $j \in J$

*Step 2*

Set

$$m_{ijk}^{(3S+2)} = m_{ijk}^{(3S+1)} a_{jk} \Big/ \sum_{i=1}^{n} m_{ijk}^{(3S+1)}$$

*Step 3*

Set

$$m_{ijk}^{(3S+3)} = m_{ijk}^{(3S+2)} b_{ik} \Big/ \sum_{j=1}^{m} m_{ijk}^{(3S+2)}$$

If

$$\left| \frac{m_{ijk}^{(3S+3)}}{m_{ijk}^{(3S+2)}} - 1 \right| \leq \varepsilon \qquad (i, j, k) \in \Gamma$$

is fulfilled then STOP, otherwise $S \leftarrow S + 1$, and go to Step 1.

### *3.2.2 Modified Multidimensional Friedlander Adjustment Problem*

*Algorithm for solving the dual problem (32)*

*Step 0*

Set

$S = 0$

$\lambda_{ij}^{(0)} = 0$

$\nu_{ik}^{(0)} = 0 \qquad (\forall i \in I, j \in J, k \in K)$

$\zeta_{jk}^{(0)} = 0$

*Step 1*

Let $\lambda_{ij}^{(S+1)}$ be the solution of

$$\sum_{k \in K(i,j)} \frac{m_{ijk}^{0}}{[1 + \lambda_{ij}^{(S+1)} + \nu_{ik}^{(S)} + \zeta_{jk}^{(S)}]^{1/2}} = c_{ij}$$

for all $i \in I, j \in J$, except for $i = j = 1$.
(The left-hand side of this equation is a strictly monotonically decreasing function of the variable $\lambda_{ij}^{(S+1)}$ in the range $(0, +\infty)$. Therefore, there is a unique solution $\lambda_{ij}^{(S+1)}$. A possible method of solution is the Newton method.)

*Step 2*

Let $\nu_{ik}^{(S+1)}$ be the solution of

$$\sum_{j \in J(i,k)} \frac{m_{ijk}^{0}}{[1 + \lambda_{ij}^{(S+1)} + \nu_{ik}^{(S+1)} + \zeta_{jk}^{(S)}]^{1/2}} = b_{ik}$$

for $i \in I$, $k \in K$, and where $\lambda_{ij}^{(S+1)}$, $\zeta_{jk}^{(S)}$ are considered to be given.

*Step 3*

Let $\zeta_{jk}^{(S+1)}$ be the solution of

$$\sum_{i \in I(j,k)} \frac{m_{ijk}^{0}}{[1 + \lambda_{ij}^{(S+1)} + \nu_{ik}^{(S+1)} + \zeta_{jk}^{(S+1)}]^{1/2}} = a_{jk}$$

If

$|\lambda_{ij}^{(S+1)} - \lambda_{ij}^{(S)}| \leq \varepsilon$

$|\nu_{ik}^{(S+1)} - \nu_{ik}^{(S)}| \leq \varepsilon \qquad (\forall i \in I, j \in J, k \in K)$

$|\zeta_{jk}^{(S+1)} - \zeta_{jk}^{(S)}| \leq \varepsilon$

is fulfilled then STOP, otherwise $S \leftarrow S + 1$, and go to Step 1.

*Theorem 5*

Let $\mathbf{M} = (m_{ijk})$ be a feasible solution of the basic problem (3F) such that

$m_{ijk} > 0 \qquad (i, j, k) \in \Gamma$

Then the above algorithm for solving the dual program of the modified Friedlander adjustment problem converges to an optimal solution.

Because of the nature of the modified Friedlander adjustment problem, we could not find a direct primal algorithm. However, using the results of Theorem 3 we can compute the solution of the primal problem $\hat{m}_{ijk}$ for $(i, j, k) \in \Gamma$ from the dual solution using the formula

$$\hat{m}_{ijk} = \frac{m^0_{ijk}}{(1 + \bar{\lambda}_{ij} + \bar{\nu}_{ik} + \bar{\zeta}_{jk})^{1/2}}$$

## 4 VALIDITY ANALYSIS AND NUMERICAL ILLUSTRATIONS (USING AUSTRIAN AND SWEDISH DATA)

The techniques developed in the previous sections may be used to deduce detailed migration patterns from aggregate data. But how accurate are the estimates, and which method gives the best results under a given set of conditions? Although research has not yet proceeded far enough to give definite answers to these questions, this section attempts to answer them numerically.

The accuracy of the estimation procedures is examined by applying them to multiregional systems for which complete data sets exist. Using only the marginal conditions, the detailed migration flows are estimated and the estimates are then compared with the observed data. Austria and Sweden are among the few countries that make detailed migration data available. These data may be aggregated in various ways to yield different sets of marginal totals, which may in turn be used to simulate different levels of data availability. Estimates of the detailed flows are calculated using the entropy maximizing method and the quadratic adjustment method.

The "validity" test of the multiproportional and modified Friedlander adjustment methods is basically a comparison between the estimated and observed migration flows and a judgment on the "closeness" of both sets of values. The quality of an estimation procedure is determined by the accuracy with which it can replicate observed data. A key problem in validity analysis is the definition of a composite index that measures the "closeness" of two arrays. In this paper, two such indices are used: the $\chi^2$, and the absolute percentage error. Note that the $\chi^2$ is the distance measure used in the modified Friedlander method.

The $\chi^2$ statistic measures the relative squared deviation between the estimates and the observed values

$$\chi^2 = \sum_{i,j,k} (m_{ijk} - m^0_{ijk})^2 / m_{ijk} \qquad m_{ijk} \neq 0 \tag{33}$$

The absolute percentage error ($APE$) measures the deviation in absolute terms

$$APE = \sum_{i,j,k} d_{ijk}$$

where

$$d_{ijk} = |m_{ijk} - m^0_{ijk}| / m^0_{ijk} \qquad m^0_{ijk} \neq 0 \tag{34}$$

The average absolute percentage error ($\overline{APE}$), also known as the relative mean deviation or the mean prediction error, is given by

$$\overline{APE} = \sum_{i,j,k} |m_{ijk} - m^0_{ijk}| \Big/ \sum_{i,j,k} m^0_{ijk} \qquad m^0_{ijk} \neq 0$$

Both measures give similar results, but the $\chi^2$ statistic attaches a greater penalty to large deviations.

An important observation of the validity study was that the error is not uniformly distributed among the elements of the array. The relative error, expressed by both the $\chi^2$ and the *APE*, is greatest among the small elements (i.e., minor flows). The reason is the small denominator in (33) and (34). This observation is consistent with results obtained in input–output analysis (e.g., Hinojosa 1978). In addition, we found that a small number of elements of the array are responsible for most of the error; most elements are in low-error categories.

Because of these observations, and to get a better idea about the composition of the overall squared or absolute percentage deviation, the error analysis is carried out for subgroups of migration flows. The subgroups are formed on the basis of age (three broad age categories are considered: 0–14, 15–64, 65+), and volume of migration flow (eleven size classes are distinguished: 0–199, 200–399, 400–599, . . . , 1800–1999, 2000+). Note that the size classes are in terms of *observed* flows and not in terms of the estimates. The error analysis is also carried out for twelve error categories ranging from less than 2% to greater than 100%.

## 4.1 Entropy Maximization

The Austrian migration data are taken from the 1971 census and compare the place of residence in 1966 with that in 1971. The data were kindly provided by Dr. M. Sauberer, of the Austrian Institute for Regional Planning. Various aggregations of the migration data were made, and the techniques presented in the previous sections were then used to reproduce the original flow matrix.

The multidimensional entropy maximization method is a special variant of the multiproportional adjustment method: the initial estimates of the elements of the array are set equal to any scalar value, for example, unity: $m^0_{ijk} = 1$, for all $i, j, k$, except if $i = j$.* In other words, it is assumed that initial estimates are only available for the diagonal elements.

### *4.1.1 The Three-face (3F) Problem*

Suppose the problem is to deduce origin–destination migration flows disaggregated by age for Austria (four regions) from the available information on the flow matrix of the total population and on the age composition of the arriving and departing migrants for each region. The data are given in Table 1, and present the three faces of a cube (array), the content (elements) of which must be estimated. The estimates can only be obtained by iteration. The computer program solves the 3F problem using the direct primal algorithm (see the second algorithm in Section 3.2.1).

The results of the (3F) estimation procedure are shown in Table 2, together with the observed migration-flow data. The algorithm took only five iterations to converge (tolerance level $10^{-4}$).

---

* Intraregional migration is assumed to be nonexistent (structural zero). The contingency tables obtained therefore belong to the category of the so-called "incomplete contingency tables" and the analysis is one of "quasi-independence" (Bishop *et al.* 1975, p. 177).

TABLE 1 Internal migration in Austria,[a] 1966–1971.

*(a) Migration flow matrix of total population*

| To | From: East | South | North | West | Total |
|---|---|---|---|---|---|
| East | 0 | 12564 | 10587 | 3091 | 26242 |
| South | 7460 | 0 | 4532 | 3543 | 15535 |
| North | 11471 | 7715 | 0 | 3629 | 22815 |
| West | 3272 | 7494 | 4158 | 0 | 14924 |
| Total | 22203 | 27773 | 19277 | 10263 | 79516 |

*(b) Departures ("from") and arrivals ("to") by region and age*

| | Region: East | | South | | North | | West | | Total | |
|---|---|---|---|---|---|---|---|---|---|---|
| Age | To | From | To | From | To | From | To | From | To | From |
| 0 | 1896 | 1783 | 1428 | 1909 | 1733 | 1445 | 985 | 905 | 6042 | 6042 |
| 5 | 1086 | 930 | 734 | 1115 | 981 | 917 | 577 | 416 | 3378 | 3378 |
| 10 | 2264 | 1597 | 1052 | 3662 | 2118 | 1568 | 1961 | 568 | 7395 | 7395 |
| 15 | 7424 | 4172 | 3212 | 8323 | 4801 | 5297 | 4595 | 2240 | 20032 | 20032 |
| 20 | 4582 | 4227 | 2806 | 4625 | 4391 | 3250 | 2558 | 2235 | 14337 | 14337 |
| 25 | 2612 | 2807 | 1883 | 2625 | 2757 | 1885 | 1438 | 1373 | 8690 | 8690 |
| 30 | 1122 | 1123 | 835 | 1062 | 1155 | 924 | 603 | 606 | 3715 | 3715 |
| 35 | 906 | 915 | 622 | 903 | 929 | 750 | 519 | 408 | 2976 | 2976 |
| 40 | 950 | 871 | 540 | 807 | 824 | 688 | 413 | 361 | 2727 | 2727 |
| 45 | 625 | 579 | 374 | 517 | 515 | 447 | 237 | 208 | 1751 | 1751 |
| 50 | 601 | 618 | 441 | 514 | 541 | 478 | 268 | 241 | 1851 | 1851 |
| 55 | 675 | 689 | 458 | 521 | 552 | 489 | 251 | 237 | 1936 | 1936 |
| 60 | 576 | 700 | 434 | 455 | 568 | 439 | 201 | 185 | 1779 | 1779 |
| 65 | 431 | 543 | 331 | 340 | 433 | 328 | 147 | 131 | 1342 | 1342 |
| 70 | 274 | 353 | 213 | 217 | 280 | 209 | 94 | 82 | 861 | 861 |
| 75 | 145 | 194 | 114 | 118 | 156 | 109 | 52 | 46 | 467 | 467 |
| 80 | 49 | 68 | 37 | 39 | 53 | 35 | 17 | 14 | 156 | 156 |
| 85 | 24 | 34 | 21 | 21 | 28 | 19 | 8 | 7 | 81 | 81 |
| Total | 26242 | 22203 | 15535 | 27773 | 22815 | 19277 | 14924 | 10263 | 79516 | 79516 |

[a] The nine Austrian federal states (Bundesländer) have been aggregated here into four regions, as follows: East = Wien, Niederösterreich, Burgenland; South = Steiermark, Kärnten; North = Salzburg, Oberösterreich; West = Tirol, Vorarlberg.

The estimates are generally very close to the observed values. The average absolute percentage error is 4.27%, a very low figure compared with that obtained using existing biproportional or two-dimensional entropy methods (Nijkamp and Paelinck 1974b, Hinojosa 1978). About half of the number of migration flows (number of cells in the migration table) and almost two-thirds of the migration volume are estimated with less than 4% error (Table 3b). About 69% of the total absolute percentage error is due to minor migration flows (less than 200 migrants) representing only 11% of the flow volume (Table 3a). A similar pattern is obtained if the $\chi^2$ statistic is used. The error distribution is, however, more explicit. The minor flows account for 34% of the total $\chi^2$ value.

Similar observations were obtained in applying the estimation procedure to Swedish data. For a system composed of eight regions and eighteen age groups, the average absolute percentage error is 6.32%. About 28% of the cells in the migration table and about half of the migration volume are estimated with an error of less than 4%. The lower level of accuracy as compared with the Austrian estimates is a consequence of the larger share of small flows in the Swedish table. Whereas 52% of the migration flows for Austria contain less than 200 migrants, this figure is 79% for Sweden. About 94% of the total absolute percentage error is due to these minor flows, which account for 72% of the total $\chi^2$ value.

The contribution of minor flows to the overall error is further illustrated by the cross-classification of error categories and flow-size classes (Table 3c). Minor flows are concentrated in the larger error categories.

The importance of small flow values in the overall error raises an additional problem, namely, rounding. Since the most probable estimates are rounded to the nearest integer to represent the number of migrants, error due to rounding may be substantial in the case of minor flows. The error measures in this study do not take into account the effects of rounding. The value of $m_{ijk}$ used in the calculation of the error statistics is the original estimate before rounding.

The effect of age on the error distribution was also investigated. The results are not shown, since no significant difference between the contribution of each age category to the overall error could be observed. This may be a consequence of the uniform age pattern of the migrants.

### *4.1.2 The Three-edge (3E) Problem*

In the (3E) problem, it is assumed that the only known information is for the edges of the box, i.e., the total number of arrivals and departures disaggregated by region, and the migrant age structure at the national level. The data are shown in the row and column sums of Table 1(a) and in the last column of Table 1(b).

The entropy or most probable estimates of the migration flows by age are calculated directly using eqn. (B1) from Appendix B. The (3E) entropy method yields estimates with an average absolute percentage error of 31%. The error is not concentrated in the minor flows but is evenly distributed among the classes of flow size. The high inaccuracy of the estimates may be traced back to the lack of data. Estimating migration flows from a uniform distribution as an initial guess ($m^0_{ijk} = 1$), and with only three edges given as new marginals, yields a migration pattern in which the three variables, region of origin, region of destination, and age, are independent when taken all together or in pairs.

### *4.1.3 The One-face, One-edge (1FE) Problem*

In the (1FE) problem we estimate the values of $m_{ijk}$ if the total flow matrix $c_{ij}$ and the age structure of the migrants at the national level $u_k$ are given; the estimates are obtained from eqn. (B2) and are shown in Table 4. This may occur when the available *a priori* information consists of the total population and a model migration schedule (see the paper by Rogers and Castro in this volume). By introducing information on the total flow matrix, the average absolute percentage error drops by half, from 31 to 16%. The contribution of minor flows to the overall error increases

TABLE 2 Estimated (3F) and observed migration flows by age for four regions of Austria, 1966–1971.

| | | | Migration from east to | | | | | | | Migration from south to | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Age | Total | | East | South | North | West | Age | Total | | East | South | North | West |
| 0 | 1783 | Est. | 0 | 674 | 874 | 234 | 0 | 1909 | Est. | 882 | 0 | 556 | 471 |
| | | Obs. | 0 | 670 | 877 | 236 | | | Obs. | 853 | 0 | 575 | 481 |
| 5 | 930 | Est. | 0 | 328 | 482 | 120 | 5 | 1115 | Est. | 493 | 0 | 348 | 274 |
| | | Obs. | 0 | 328 | 468 | 134 | | | Obs. | 530 | 0 | 329 | 256 |
| 10 | 1597 | Est. | 0 | 483 | 821 | 293 | 10 | 3662 | Est. | 1371 | 0 | 1075 | 1216 |
| | | Obs. | 0 | 537 | 828 | 232 | | | Obs. | 1342 | 0 | 1044 | 1276 |
| 15 | 4172 | Est. | 0 | 1351 | 2029 | 793 | 15 | 8323 | Est. | 3800 | 0 | 2022 | 2501 |
| | | Obs. | 0 | 1280 | 2192 | 700 | | | Obs. | 3760 | 0 | 1950 | 2613 |
| 20 | 4227 | Est. | 0 | 1336 | 2246 | 645 | 20 | 4625 | Est. | 2097 | 0 | 1324 | 1203 |
| | | Obs. | 0 | 1289 | 2231 | 707 | | | Obs. | 2081 | 0 | 1381 | 1163 |
| 25 | 2807 | Est. | 0 | 939 | 1477 | 392 | 25 | 2625 | Est. | 1187 | 0 | 781 | 656 |
| | | Obs. | 0 | 910 | 1464 | 433 | | | Obs. | 1225 | 0 | 800 | 600 |
| 30 | 1123 | Est. | 0 | 378 | 596 | 149 | 30 | 1062 | Est. | 465 | 0 | 333 | 263 |
| | | Obs. | 0 | 368 | 588 | 167 | | | Obs. | 464 | 0 | 346 | 252 |
| 35 | 915 | Est. | 0 | 295 | 493 | 127 | 35 | 903 | Est. | 392 | 0 | 281 | 230 |
| | | Obs. | 0 | 293 | 480 | 142 | | | Obs. | 408 | 0 | 276 | 219 |
| 40 | 871 | Est. | 0 | 280 | 474 | 117 | 40 | 807 | Est. | 409 | 0 | 223 | 174 |
| | | Obs. | 0 | 312 | 438 | 121 | | | Obs. | 411 | 0 | 240 | 156 |
| 45 | 579 | Est. | 0 | 204 | 306 | 69 | 45 | 517 | Est. | 277 | 0 | 140 | 100 |
| | | Obs. | 0 | 222 | 289 | 68 | | | Obs. | 269 | 0 | 149 | 99 |
| 50 | 618 | Est. | 0 | 229 | 314 | 75 | 50 | 514 | Est. | 256 | 0 | 147 | 111 |
| | | Obs. | 0 | 241 | 312 | 65 | | | Obs. | 263 | 0 | 134 | 117 |
| 55 | 689 | Est. | 0 | 261 | 346 | 81 | 55 | 521 | Est. | 289 | 0 | 133 | 99 |
| | | Obs. | 0 | 280 | 331 | 78 | | | Obs. | 296 | 0 | 132 | 93 |
| 60 | 700 | Est. | 0 | 259 | 373 | 67 | 60 | 455 | Est. | 246 | 0 | 133 | 76 |
| | | Obs. | 0 | 269 | 357 | 74 | | | Obs. | 253 | 0 | 137 | 65 |
| 65 | 543 | Est. | 0 | 203 | 290 | 51 | 65 | 340 | Est. | 185 | 0 | 100 | 55 |
| | | Obs. | 0 | 210 | 280 | 53 | | | Obs. | 190 | 0 | 102 | 48 |
| 70 | 353 | Est. | 0 | 131 | 189 | 33 | 70 | 217 | Est. | 118 | 0 | 64 | 35 |
| | | Obs. | 0 | 138 | 182 | 33 | | | Obs. | 121 | 0 | 65 | 31 |
| 75 | 194 | Est. | 0 | 71 | 105 | 18 | 75 | 118 | Est. | 63 | 0 | 35 | 19 |
| | | Obs. | 0 | 74 | 101 | 19 | | | Obs. | 65 | 0 | 36 | 17 |
| 80 | 68 | Est. | 0 | 24 | 37 | 6 | 80 | 39 | Est. | 22 | 0 | 11 | 6 |
| | | Obs. | 0 | 26 | 35 | 7 | | | Obs. | 22 | 0 | 12 | 5 |
| 85 | 34 | Est. | 0 | 13 | 18 | 3 | 85 | 21 | Est. | 11 | 0 | 7 | 3 |
| | | Obs. | 0 | 13 | 18 | 3 | | | Obs. | 11 | 0 | 7 | 3 |
| Total | 22203 | | 0 | 7460 | 11471 | 3272 | Total | 27773 | | 12564 | 0 | 7715 | 7494 |

TABLE 2 *Continued.*

| | | | Migration from north to | | | | | | | Migration from west to | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Age | Total | | East | South | North | West | Age | Total | | East | South | North | West |
| 0 | 1445 | Est. | 764 | 402 | 0 | 280 | 0 | 905 | Est. | 250 | 352 | 303 | 0 |
| | | Obs. | 814 | 363 | 0 | 268 | | | Obs. | 229 | 395 | 281 | 0 |
| 5 | 917 | Est. | 482 | 251 | 0 | 184 | 5 | 416 | Est. | 111 | 155 | 150 | 0 |
| | | Obs. | 448 | 282 | 0 | 187 | | | Obs. | 108 | 124 | 184 | 0 |
| 10 | 1568 | Est. | 745 | 371 | 0 | 452 | 10 | 568 | Est. | 148 | 197 | 222 | 0 |
| | | Obs. | 771 | 344 | 0 | 453 | | | Obs. | 151 | 171 | 246 | 0 |
| 15 | 5297 | Est. | 2888 | 1107 | 0 | 1302 | 15 | 2240 | Est. | 736 | 754 | 750 | 0 |
| | | Obs. | 2892 | 1123 | 0 | 1282 | | | Obs. | 772 | 809 | 659 | 0 |
| 20 | 3250 | Est. | 1806 | 734 | 0 | 710 | 20 | 2235 | Est. | 678 | 736 | 821 | 0 |
| | | Obs. | 1861 | 701 | 0 | 688 | | | Obs. | 640 | 816 | 779 | 0 |
| 25 | 1885 | Est. | 1029 | 466 | 0 | 390 | 25 | 1373 | Est. | 395 | 479 | 499 | 0 |
| | | Obs. | 998 | 482 | 0 | 405 | | | Obs. | 389 | 491 | 493 | 0 |
| 30 | 924 | Est. | 492 | 241 | 0 | 191 | 30 | 606 | Est. | 165 | 216 | 226 | 0 |
| | | Obs. | 485 | 255 | 0 | 184 | | | Obs. | 173 | 212 | 221 | 0 |
| 35 | 750 | Est. | 402 | 187 | 0 | 162 | 35 | 408 | Est. | 113 | 140 | 155 | 0 |
| | | Obs. | 379 | 213 | 0 | 158 | | | Obs. | 119 | 116 | 173 | 0 |
| 40 | 688 | Est. | 419 | 146 | 0 | 122 | 40 | 361 | Est. | 121 | 113 | 127 | 0 |
| | | Obs. | 411 | 141 | 0 | 136 | | | Obs. | 128 | 87 | 146 | 0 |
| 45 | 447 | Est. | 277 | 101 | 0 | 68 | 45 | 208 | Est. | 71 | 69 | 69 | 0 |
| | | Obs. | 275 | 102 | 0 | 70 | | | Obs. | 81 | 50 | 77 | 0 |
| 50 | 478 | Est. | 273 | 124 | 0 | 81 | 50 | 241 | Est. | 73 | 88 | 80 | 0 |
| | | Obs. | 273 | 119 | 0 | 86 | | | Obs. | 65 | 81 | 95 | 0 |
| 55 | 489 | Est. | 304 | 114 | 0 | 71 | 55 | 237 | Est. | 82 | 82 | 72 | 0 |
| | | Obs. | 295 | 114 | 0 | 80 | | | Obs. | 84 | 64 | 89 | 0 |
| 60 | 439 | Est. | 271 | 110 | 0 | 58 | 60 | 185 | Est. | 59 | 64 | 61 | 0 |
| | | Obs. | 263 | 114 | 0 | 62 | | | Obs. | 60 | 51 | 74 | 0 |
| 65 | 328 | Est. | 204 | 83 | 0 | 41 | 65 | 131 | Est. | 42 | 46 | 43 | 0 |
| | | Obs. | 198 | 84 | 0 | 46 | | | Obs. | 43 | 37 | 51 | 0 |
| 70 | 209 | Est. | 130 | 53 | 0 | 26 | 70 | 82 | Est. | 26 | 29 | 27 | 0 |
| | | Obs. | 125 | 54 | 0 | 30 | | | Obs. | 28 | 21 | 33 | 0 |
| 75 | 109 | Est. | 67 | 27 | 0 | 14 | 75 | 46 | Est. | 15 | 16 | 16 | 0 |
| | | Obs. | 66 | 27 | 0 | 16 | | | Obs. | 14 | 13 | 19 | 0 |
| 80 | 35 | Est. | 22 | 8 | 0 | 4 | 80 | 14 | Est. | 5 | 5 | 5 | 0 |
| | | Obs. | 22 | 8 | 0 | 5 | | | Obs. | 5 | 3 | 6 | 0 |
| 85 | 19 | Est. | 11 | 6 | 0 | 2 | 85 | 7 | Est. | 2 | 3 | 2 | 0 |
| | | Obs. | 11 | 6 | 0 | 2 | | | Obs. | 2 | 2 | 3 | 0 |
| Total | 19277 | | 10587 | 4532 | 0 | 4158 | Total | 10263 | | 3091 | 3543 | 3629 | 0 |

TABLE 3 Error analysis of (3F) migration estimates for Austria.

*(a) Analysis by size of flow*

| Size of flow | Number of flows | | Number of migrants | | Cumulative absolute % error | | $\chi^2$ | |
|---|---|---|---|---|---|---|---|---|
| | Total | % | Total | % | Value | % | Value | % |
| 0–200 | 112 | 51.85 | 8452 | 10.63 | 1043 | 68.56 | 0.9121e 02 | 33.70 |
| 200–400 | 45 | 20.83 | 12742 | 16.02 | 241 | 15.81 | 0.5711e 02 | 21.10 |
| 400–600 | 20 | 9.26 | 9481 | 11.92 | 74 | 4.87 | 0.2255e 02 | 8.33 |
| 600–800 | 11 | 5.09 | 7687 | 9.67 | 73 | 4.81 | 0.4191e 02 | 15.49 |
| 800–1000 | 9 | 4.17 | 7705 | 9.69 | 36 | 2.36 | 0.1924e 02 | 7.11 |
| 1000–1200 | 3 | 1.39 | 3330 | 4.19 | 8 | 0.52 | 0.2466e 01 | 0.91 |
| 1200–1400 | 7 | 3.24 | 9075 | 11.41 | 25 | 1.63 | 0.1295e 02 | 4.78 |
| 1400–1600 | 1 | 0.46 | 1464 | 1.84 | 1 | 0.06 | 0.1074e 00 | 0.04 |
| 1600–1800 | 0 | 0.00 | 0 | 0.00 | 0 | 0.00 | 0.0000e 00 | 0.00 |
| 1800–2000 | 2 | 0.93 | 3811 | 4.79 | 7 | 0.43 | 0.4201e 01 | 1.55 |
| 2000+ | 6 | 2.78 | 15769 | 19.83 | 14 | 0.95 | 0.1887e 02 | 6.97 |
| Total | 216 | 100.00 | 79516 | 100.00 | 1522 | 100.00 | 0.2706e 03 | 100.00 |

*(b) Analysis by error category*

| Error category | Percentage error[a] | Number of flows | | Number of migrants | | Average flow |
|---|---|---|---|---|---|---|
| | | Total | % | Total | % | |
| 1 | 0–2 | 46 | 21.30 | 24037 | 30.23 | 522.543 |
| 2 | 2–4 | 57 | 26.39 | 24756 | 31.13 | 434.316 |
| 3 | 4–6 | 31 | 14.35 | 13604 | 17.11 | 438.839 |
| 4 | 6–8 | 18 | 8.33 | 6463 | 8.13 | 359.056 |
| 5 | 8–10 | 12 | 5.56 | 4026 | 5.06 | 335.500 |
| 6 | 10–15 | 28 | 12.96 | 5021 | 6.31 | 179.321 |
| 7 | 15–20 | 10 | 4.63 | 798 | 1.00 | 79.800 |
| 8 | 20–30 | 10 | 4.63 | 650 | 0.82 | 65.000 |
| 9 | 30–40 | 3 | 1.39 | 158 | 0.20 | 52.667 |
| 10 | 40–60 | 1 | 0.46 | 3 | 0.00 | 3.000 |
| 11 | 60–100 | 0 | 0.00 | 0 | 0.00 | 0.000 |
| 12 | 100+ | 0 | 0.00 | 0 | 0.00 | 0.000 |
| Total | | 216 | 100.00 | 79516 | 100.00 | 368.130 |

*(c) Analysis by size of flow and error category*

| Size of flow | Error category 0–2 | 2–4 | 4–6 | 6–8 | 8–10 | 10–15 | 15–20 | 20–30 | 30–40 | 40–60 | 60–100 | 100+ | Total |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0–200 | 21 | 23 | 13 | 8 | 4 | 20 | 10 | 9 | 3 | 1 | 0 | 0 | 112 |
| 200–400 | 9 | 12 | 10 | 5 | 3 | 5 | 0 | 1 | 0 | 0 | 0 | 0 | 45 |
| 400–600 | 6 | 9 | 0 | 2 | 2 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 20 |
| 600–800 | 1 | 2 | 4 | 0 | 2 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 11 |
| 800–1000 | 2 | 4 | 0 | 2 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 9 |
| 1000–1200 | 1 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 3 |
| 1200–1400 | 1 | 3 | 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 7 |
| 1400–1600 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| 1600–1800 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 1800–2000 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 |
| 2000+ | 4 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 6 |
| Total | 46 | 57 | 31 | 18 | 12 | 28 | 10 | 10 | 3 | 1 | 0 | 0 | 216 |

[a] Average absolute percentage error (relative mean deviation) = 4.27.

in importance, in particular if the squared deviation is used as a measurement of error (Table 5).

### *4.1.4 The Two-face (2F) Problem*

Assume that the following two faces are known: the total flow matrix ($c_{ij}$) and the age structure of the arriving migrants disaggregated by region ($a_{jk}$). The estimates may be obtained directly by applying eqn. (B3). The additional information on the age composition of arrivals reduces the average absolute percentage deviation only slightly, from 16 to 12% ($\chi^2$ drops from 3662 to 2006). However, the error distribution changes. Minor flows get a greater share of the total absolute percentage error (72% of total absolute percentage deviation and 28% of total $\chi^2$ value).

Results similar to those reported for Austria were obtained for Sweden, where the number of regions is twice as large. The average absolute percentage errors in the (3F), (3E), (1FE), and (2F) problems are, respectively, 6.32, 34.58, 15.26, and 11.90%. The share of the minor flows in the total error was always much higher: in the Austrian case the minor flows accounted for 51–72% of the total percentage absolute deviation, while in the Swedish case they amounted to 90–94%. This may be explained in part by the proportion of the total number and volume of flows constituted by minor flows.

The cases of data availability considered here lead to a firm conclusion: expanding the data set not only reduces the estimation error, but also increases the implicit weight attached to minor flows. This is implicit in the error statistics used: as the deviations between estimates and observations decline, the effects of small denominators become more apparent. The error analysis carried out here may also be used to investigate the marginal value of information on migration. Not every subset of the migration data has the same impact on the quality of the estimates. Hence, in further research we should consider the question: what kind of information on the migration pattern is required in order to obtain estimates of the detailed flow with an acceptable minimum level of accuracy?

## 4.2 Quadratic (Modified Friedlander) Method

As before, we assume that no *a priori* information on the elements $m^0_{ijk}$ is available. The elements are set equal to unity except for the diagonal; hence the objective function of the modified Friedlander problem becomes

$$\min d[m_{ijk}, m^0_{ijk}] = \tfrac{1}{2} \sum_{i,j,k} (m_{ijk} - 1)^2 / m_{ijk}$$

The modified Friedlander method is applied to the 3F problem only. The computer program solves the problem using the dual algorithm (see Section 3.2.2). The estimated values of the migration flows $m_{ijk}$ are shown in Table 6. The algorithm required 60 iterations to converge (tolerance level $10^{-4}$). The convergence is therefore much slower than in the entropy algorithm. In addition to slow convergence, rounding errors may cause problems (all variables are single precision). The value of the $\chi^2$ statistic is equal to 1614 (Table 7), and lies between the values obtained for the (2F) and (3F) problems. The $\chi^2$ value is given here for illustrative

TABLE 4 Observed and estimated (1FE) migration flows by age for four regions of Austria, 1966–1971.

| Age | Total | | Migration from east to | | | | Age | Total | | Migration from south to | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | East | South | North | West | | | | East | South | North | West |
| 0 | 1687 | Est. | 0 | 567 | 872 | 249 | 0 | 2110 | Est. | 955 | 0 | 586 | 569 |
| | | Obs. | 0 | 670 | 877 | 236 | | | Obs. | 853 | 0 | 575 | 481 |
| 5 | 943 | Est. | 0 | 317 | 487 | 139 | 5 | 1180 | Est. | 534 | 0 | 328 | 318 |
| | | Obs. | 0 | 328 | 468 | 134 | | | Obs. | 530 | 0 | 329 | 256 |
| 10 | 2065 | Est. | 0 | 694 | 1067 | 304 | 10 | 2583 | Est. | 1168 | 0 | 717 | 697 |
| | | Obs. | 0 | 537 | 828 | 232 | | | Obs. | 1342 | 0 | 1044 | 1276 |
| 15 | 5593 | Est. | 0 | 1879 | 2890 | 824 | 15 | 6997 | Est. | 3165 | 0 | 1944 | 1888 |
| | | Obs. | 0 | 1280 | 2192 | 700 | | | Obs. | 3760 | 0 | 1950 | 2613 |
| 20 | 4003 | Est. | 0 | 1345 | 2068 | 590 | 20 | 5008 | Est. | 2265 | 0 | 1391 | 1351 |
| | | Obs. | 0 | 1289 | 2231 | 707 | | | Obs. | 2081 | 0 | 1381 | 1163 |
| 25 | 2426 | Est. | 0 | 815 | 1254 | 358 | 25 | 3035 | Est. | 1373 | 0 | 843 | 819 |
| | | Obs. | 0 | 910 | 1464 | 433 | | | Obs. | 1225 | 0 | 800 | 600 |
| 30 | 1037 | Est. | 0 | 349 | 536 | 153 | 30 | 1298 | Est. | 587 | 0 | 360 | 350 |
| | | Obs. | 0 | 368 | 588 | 167 | | | Obs. | 464 | 0 | 346 | 252 |
| 35 | 831 | Est. | 0 | 279 | 429 | 122 | 35 | 1039 | Est. | 470 | 0 | 289 | 280 |
| | | Obs. | 0 | 293 | 480 | 142 | | | Obs. | 408 | 0 | 276 | 219 |
| 40 | 761 | Est. | 0 | 256 | 393 | 112 | 40 | 952 | Est. | 431 | 0 | 265 | 257 |
| | | Obs. | 0 | 312 | 438 | 121 | | | Obs. | 411 | 0 | 240 | 156 |
| 45 | 489 | Est. | 0 | 164 | 253 | 72 | 45 | 612 | Est. | 277 | 0 | 170 | 165 |
| | | Obs. | 0 | 222 | 289 | 68 | | | Obs. | 269 | 0 | 149 | 99 |
| 50 | 517 | Est. | 0 | 174 | 267 | 76 | 50 | 647 | Est. | 292 | 0 | 180 | 174 |
| | | Obs. | 0 | 241 | 312 | 65 | | | Obs. | 263 | 0 | 134 | 117 |
| 55 | 541 | Est. | 0 | 182 | 279 | 80 | 55 | 676 | Est. | 306 | 0 | 188 | 182 |
| | | Obs. | 0 | 280 | 331 | 78 | | | Obs. | 296 | 0 | 132 | 93 |
| 60 | 497 | Est. | 0 | 167 | 257 | 73 | 60 | 621 | Est. | 281 | 0 | 173 | 168 |
| | | Obs. | 0 | 269 | 357 | 74 | | | Obs. | 253 | 0 | 137 | 65 |
| 65 | 375 | Est. | 0 | 126 | 194 | 55 | 65 | 469 | Est. | 212 | 0 | 130 | 126 |
| | | Obs. | 0 | 210 | 280 | 53 | | | Obs. | 190 | 0 | 102 | 48 |
| 70 | 240 | Est. | 0 | 81 | 124 | 35 | 70 | 301 | Est. | 136 | 0 | 84 | 81 |
| | | Obs. | 0 | 138 | 182 | 33 | | | Obs. | 121 | 0 | 65 | 31 |
| 75 | 130 | Est. | 0 | 44 | 67 | 19 | 75 | 163 | Est. | 74 | 0 | 45 | 44 |
| | | Obs. | 0 | 74 | 101 | 19 | | | Obs. | 65 | 0 | 36 | 17 |
| 80 | 44 | Est. | 0 | 15 | 23 | 6 | 80 | 54 | Est. | 25 | 0 | 15 | 15 |
| | | Obs. | 0 | 26 | 35 | 7 | | | Obs. | 22 | 0 | 12 | 5 |
| 85 | 23 | Est. | 0 | 8 | 12 | 3 | 85 | 28 | Est. | 13 | 0 | 8 | 8 |
| | | Obs. | 0 | 13 | 18 | 3 | | | Obs. | 11 | 0 | 7 | 3 |
| Total | 22203 | | 0 | 7460 | 11471 | 3272 | Total | 27773 | | 12564 | 0 | 7715 | 7494 |

TABLE 4 *Continued.*

| Age | Total | | Migration from north to East | South | North | West | Age | Total | | Migration from west to East | South | North | West |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 1465 | Est. | 804 | 344 | 0 | 316 | 0 | 780 | Est. | 235 | 269 | 276 | 0 |
| | | Obs. | 814 | 363 | 0 | 268 | | | Obs. | 229 | 395 | 281 | 0 |
| 5 | 819 | Est. | 450 | 193 | 0 | 177 | 5 | 436 | Est. | 131 | 151 | 154 | 0 |
| | | Obs. | 448 | 282 | 0 | 187 | | | Obs. | 108 | 124 | 184 | 0 |
| 10 | 1793 | Est. | 985 | 421 | 0 | 387 | 10 | 954 | Est. | 287 | 329 | 337 | 0 |
| | | Obs. | 771 | 344 | 0 | 453 | | | Obs. | 151 | 171 | 246 | 0 |
| 15 | 4856 | Est. | 2667 | 1142 | 0 | 1048 | 15 | 2585 | Est. | 779 | 893 | 914 | 0 |
| | | Obs. | 2892 | 1123 | 0 | 1282 | | | Obs. | 772 | 809 | 659 | 0 |
| 20 | 3476 | Est. | 1909 | 817 | 0 | 750 | 20 | 1850 | Est. | 557 | 639 | 654 | 0 |
| | | Obs. | 1861 | 701 | 0 | 688 | | | Obs. | 640 | 816 | 779 | 0 |
| 25 | 2107 | Est. | 1157 | 495 | 0 | 454 | 25 | 1122 | Est. | 338 | 387 | 397 | 0 |
| | | Obs. | 998 | 482 | 0 | 405 | | | Obs. | 389 | 491 | 493 | 0 |
| 30 | 901 | Est. | 495 | 212 | 0 | 194 | 30 | 479 | Est. | 144 | 166 | 170 | 0 |
| | | Obs. | 485 | 255 | 0 | 184 | | | Obs. | 173 | 212 | 221 | 0 |
| 35 | 721 | Est. | 396 | 170 | 0 | 156 | 35 | 384 | Est. | 116 | 133 | 136 | 0 |
| | | Obs. | 379 | 213 | 0 | 158 | | | Obs. | 119 | 116 | 173 | 0 |
| 40 | 661 | Est. | 363 | 155 | 0 | 143 | 40 | 352 | Est. | 106 | 122 | 124 | 0 |
| | | Obs. | 411 | 141 | 0 | 136 | | | Obs. | 128 | 87 | 146 | 0 |
| 45 | 424 | Est. | 233 | 100 | 0 | 92 | 45 | 226 | Est. | 68 | 78 | 80 | 0 |
| | | Obs. | 275 | 102 | 0 | 70 | | | Obs. | 81 | 50 | 77 | 0 |
| 50 | 449 | Est. | 246 | 105 | 0 | 97 | 50 | 239 | Est. | 72 | 82 | 84 | 0 |
| | | Obs. | 273 | 119 | 0 | 86 | | | Obs. | 65 | 81 | 95 | 0 |
| 55 | 469 | Est. | 258 | 110 | 0 | 101 | 55 | 250 | Est. | 75 | 86 | 88 | 0 |
| | | Obs. | 295 | 114 | 0 | 80 | | | Obs. | 84 | 64 | 89 | 0 |
| 60 | 431 | Est. | 237 | 101 | 0 | 93 | 60 | 230 | Est. | 69 | 79 | 81 | 0 |
| | | Obs. | 263 | 114 | 0 | 62 | | | Obs. | 60 | 51 | 74 | 0 |
| 65 | 325 | Est. | 179 | 76 | 0 | 70 | 65 | 173 | Est. | 52 | 60 | 61 | 0 |
| | | Obs. | 198 | 84 | 0 | 46 | | | Obs. | 43 | 37 | 51 | 0 |
| 70 | 209 | Est. | 115 | 49 | 0 | 45 | 70 | 111 | Est. | 33 | 38 | 39 | 0 |
| | | Obs. | 125 | 54 | 0 | 30 | | | Obs. | 28 | 21 | 33 | 0 |
| 75 | 113 | Est. | 62 | 27 | 0 | 24 | 75 | 60 | Est. | 18 | 21 | 21 | 0 |
| | | Obs. | 66 | 27 | 0 | 16 | | | Obs. | 14 | 13 | 19 | 0 |
| 80 | 38 | Est. | 21 | 9 | 0 | 8 | 80 | 20 | Est. | 6 | 7 | 7 | 0 |
| | | Obs. | 22 | 8 | 0 | 5 | | | Obs. | 5 | 3 | 6 | 0 |
| 85 | 20 | Est. | 11 | 5 | 0 | 4 | 85 | 10 | Est. | 3 | 4 | 4 | 0 |
| | | Obs. | 11 | 6 | 0 | 2 | | | Obs. | 2 | 2 | 3 | 0 |
| Total | 19277 | | 10587 | 4532 | 0 | 4158 | Total | 10263 | | 3091 | 3543 | 3629 | 0 |

TABLE 5 Error analysis of (1FE) migration estimates for Austria.

*(a) Analysis by size of flow*

| Size of flow | Number of flows | | Number of migrants | | Cumulative absolute % error | | $\chi^2$ | |
|---|---|---|---|---|---|---|---|---|
| | Total | % | Total | % | Value | % | Value | % |
| 0–200 | 112 | 51.85 | 8452 | 10.63 | 3809 | 70.07 | 0.7952e 03 | 21.72 |
| 200–400 | 45 | 20.83 | 12742 | 16.02 | 774 | 14.24 | 0.6374e 03 | 17.41 |
| 400–600 | 20 | 9.26 | 9481 | 11.92 | 232 | 4.27 | 0.1920e 03 | 5.24 |
| 600–800 | 11 | 5.09 | 7687 | 9.67 | 208 | 3.83 | 0.2945e 03 | 8.04 |
| 800–1000 | 9 | 4.17 | 7705 | 9.69 | 106 | 1.96 | 0.1565e 03 | 4.27 |
| 1000–1200 | 3 | 1.39 | 3330 | 4.19 | 49 | 0.90 | 0.1751e 03 | 4.78 |
| 1200–1400 | 7 | 3.24 | 9075 | 11.41 | 141 | 2.59 | 0.7689e 03 | 21.00 |
| 1400–1600 | 1 | 0.46 | 1464 | 1.84 | 14 | 0.26 | 0.3530e 02 | 0.96 |
| 1600–1800 | 0 | 0.00 | 0 | 0.00 | 0 | 0.00 | 0.0000e 00 | 0.00 |
| 1800–2000 | 2 | 0.93 | 3811 | 4.79 | 3 | 0.05 | 0.1222e 01 | 0.03 |
| 2000+ | 6 | 2.78 | 15769 | 19.83 | 99 | 1.83 | 0.6055e 03 | 16.54 |
| Total | 216 | 100.00 | 79516 | 100.00 | 5437 | 100.00 | 0.3662e 04 | 100.00 |

*(b) Analysis by error category*

| Error category | Percentage error[a] | Number of flows | | Number of migrants | | Average flow |
|---|---|---|---|---|---|---|
| | | Total | % | Total | % | |
| 1 | 0–2 | 19 | 8.80 | 10024 | 12.61 | 527.579 |
| 2 | 2–4 | 12 | 5.56 | 4089 | 5.14 | 340.750 |
| 3 | 4–6 | 18 | 8.33 | 5945 | 7.48 | 330.278 |
| 4 | 6–8 | 4 | 1.85 | 5277 | 6.64 | 1319.250 |
| 5 | 8–10 | 12 | 5.56 | 4602 | 5.79 | 383.500 |
| 6 | 10–15 | 38 | 17.59 | 13163 | 16.55 | 346.395 |
| 7 | 15–20 | 28 | 12.96 | 14785 | 18.59 | 528.036 |
| 8 | 20–30 | 31 | 14.35 | 9764 | 12.28 | 314.968 |
| 9 | 30–40 | 20 | 9.26 | 7422 | 9.33 | 371.100 |
| 10 | 40–60 | 16 | 7.41 | 3523 | 4.43 | 220.188 |
| 11 | 60–100 | 10 | 4.63 | 748 | 0.94 | 74.800 |
| 12 | 100+ | 8 | 3.70 | 174 | 0.22 | 21.750 |
| Total | | 216 | 100.00 | 79516 | 100.00 | 368.130 |

*(c) Analysis by size of flow and error category*

| Size of flow | Error category | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 0–2 | 2–4 | 4–6 | 6–8 | 8–10 | 10–15 | 15–20 | 20–30 | 30–40 | 40–60 | 60–100 | 100+ | Total |
| 0–200 | 7 | 6 | 7 | 2 | 7 | 19 | 10 | 15 | 8 | 13 | 10 | 8 | 112 |
| 200–400 | 2 | 4 | 7 | 0 | 2 | 7 | 5 | 9 | 8 | 1 | 0 | 0 | 45 |
| 400–600 | 4 | 1 | 2 | 0 | 1 | 5 | 4 | 3 | 0 | 0 | 0 | 0 | 20 |
| 600–800 | 1 | 0 | 0 | 0 | 1 | 1 | 5 | 1 | 2 | 0 | 0 | 0 | 11 |
| 800–1000 | 2 | 0 | 1 | 0 | 0 | 3 | 1 | 2 | 0 | 0 | 0 | 0 | 9 |
| 1000–1200 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 3 |
| 1200–1400 | 1 | 0 | 1 | 0 | 0 | 2 | 1 | 0 | 0 | 2 | 0 | 0 | 7 |
| 1400–1600 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| 1600–1800 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 1800–2000 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 |
| 2000+ | 0 | 0 | 0 | 2 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 6 |
| Total | 19 | 12 | 18 | 4 | 12 | 38 | 28 | 31 | 20 | 16 | 10 | 8 | 216 |

[a] Average absolute percentage error (relative mean deviation) = 16.24.

purposes only. It is not an appropriate goodness-of-fit test for the modified Friedlander method, since it is not independent of the objective function used. As a consequence, an objective comparison of the quality of the entropy and the quadratic methods is not straightforward. Further research on the development of appropriate test statistics is needed.

## 5 CONCLUSIONS

In migration analysis it frequently occurs that data on migration flows disaggregated by regions of origin and destination do not exist for subgroups of the population (such as age groups, income classes, educational levels, etc.) and that one has to infer these detailed flows from some more-aggregate information. This paper presents and generalizes two classes of estimation methods with great potential for solving this type of problem.

The classes and individual techniques may formally be presented as mathematical optimization problems with nonlinear objective functions and linear constraints. The various techniques differ in the objective function used.

The first class, namely, the bi- and multiproportional adjustment methods, combines the entropy maximization and the I-divergence minimizing problems. The two formulations have much in common, although they were developed independently. This paper presents a generalized formulation integrating both estimation techniques. The entropy method may be considered a special case of the more general I-divergence method. In the entropy problem, no initial values of the elements to be estimated are available and they are therefore uniformly set equal to unity. A single-solution algorithm is developed for both problems. Which of the techniques should be used will depend on the data available. The entropy method is more suited to estimating detailed migration flows on the basis of aggregate information only; the I-divergence method lends itself to updating migration-flow tables.

In the quadratic adjustment problems, weighted squared deviations between estimates and initial guesses are minimized. The latter may be derived from outdated migration tables or may represent *a priori* information on the migration pattern. A disadvantage of the quadratic adjustment methods, among them the Friedlander method, is that the estimates are not always of the appropriate sign. To avoid such anomalies as negative gross out-migration flows, a modified Friedlander method is presented.

The validity of the various estimation procedures has been demonstrated using migration data referring to Austria and Sweden. The numerical illustration shows that the estimates obtained by the multidimensional entropy method are very close to the observed data in the (3F) case. The estimates in other cases were not so good, but it was assumed that fewer data were available. However, one remarkable observation was that the (2F) case yielded estimates that did not deviate much from those obtained in the (3F) case, even though the latter required much more initial data. The reason is the high age specificity of migration. There seems to be a threshold level of information about the age composition of migrants that is sufficient to yield good estimates of detailed flows: after this threshold is reached, having more *a priori* information on the age structure does not add significantly to the quality of the estimates. This may lead to some interesting further research.

TABLE 6 Observed and estimated (modified Friedlander) migration flows by age for four regions of Austria, 1966–1971.

| | | | Migration from east to | | | | | | | Migration from south to | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Age | Total | | East | South | North | West | Age | Total | | East | South | North | West |
| 0 | 1783 | Est. | 0 | 681 | 799 | 304 | 0 | 1909 | Est. | 921 | 0 | 628 | 360 |
| | | Obs. | 0 | 670 | 877 | 236 | | | Obs. | 853 | 0 | 575 | 481 |
| 5 | 930 | Est. | 0 | 316 | 428 | 186 | 5 | 1115 | Est. | 503 | 0 | 413 | 199 |
| | | Obs. | 0 | 328 | 468 | 134 | | | Obs. | 530 | 0 | 329 | 256 |
| 10 | 1597 | Est. | 0 | 466 | 729 | 402 | 10 | 3662 | Est. | 1426 | 0 | 1194 | 1042 |
| | | Obs. | 0 | 537 | 828 | 232 | | | Obs. | 1342 | 0 | 1044 | 1276 |
| 15 | 4172 | Est. | 0 | 1522 | 2156 | 494 | 15 | 8323 | Est. | 3388 | 0 | 1815 | 3120 |
| | | Obs. | 0 | 1280 | 2192 | 700 | | | Obs. | 3760 | 0 | 1950 | 2613 |
| 20 | 4227 | Est. | 0 | 1321 | 2411 | 494 | 20 | 4625 | Est. | 2197 | 0 | 1143 | 1285 |
| | | Obs. | 0 | 1289 | 2231 | 707 | | | Obs. | 2081 | 0 | 1381 | 1163 |
| 25 | 2807 | Est. | 0 | 915 | 1486 | 407 | 25 | 2625 | Est. | 1268 | 0 | 790 | 567 |
| | | Obs. | 0 | 910 | 1464 | 433 | | | Obs. | 1225 | 0 | 800 | 600 |
| 30 | 1123 | Est. | 0 | 362 | 560 | 201 | 30 | 1062 | Est. | 470 | 0 | 388 | 204 |
| | | Obs. | 0 | 368 | 588 | 167 | | | Obs. | 464 | 0 | 346 | 252 |
| 35 | 915 | Est. | 0 | 273 | 465 | 176 | 35 | 903 | Est. | 404 | 0 | 325 | 174 |
| | | Obs. | 0 | 293 | 480 | 142 | | | Obs. | 408 | 0 | 276 | 219 |
| 40 | 871 | Est. | 0 | 252 | 471 | 148 | 40 | 807 | Est. | 439 | 0 | 234 | 134 |
| | | Obs. | 0 | 312 | 438 | 121 | | | Obs. | 411 | 0 | 240 | 156 |
| 45 | 579 | Est. | 0 | 190 | 304 | 85 | 45 | 517 | Est. | 298 | 0 | 143 | 76 |
| | | Obs. | 0 | 222 | 289 | 68 | | | Obs. | 269 | 0 | 149 | 99 |
| 50 | 618 | Est. | 0 | 223 | 300 | 95 | 50 | 514 | Est. | 265 | 0 | 162 | 87 |
| | | Obs. | 0 | 241 | 312 | 65 | | | Obs. | 263 | 0 | 134 | 117 |
| 55 | 689 | Est. | 0 | 257 | 339 | 92 | 55 | 521 | Est. | 305 | 0 | 137 | 80 |
| | | Obs. | 0 | 280 | 331 | 78 | | | Obs. | 296 | 0 | 132 | 93 |
| 60 | 700 | Est. | 0 | 253 | 375 | 72 | 60 | 455 | Est. | 258 | 0 | 132 | 65 |
| | | Obs. | 0 | 269 | 357 | 74 | | | Obs. | 253 | 0 | 137 | 65 |
| 65 | 543 | Est. | 0 | 198 | 292 | 53 | 65 | 340 | Est. | 195 | 0 | 98 | 47 |
| | | Obs. | 0 | 210 | 280 | 53 | | | Obs. | 190 | 0 | 102 | 48 |
| 70 | 353 | Est. | 0 | 128 | 191 | 34 | 70 | 217 | Est. | 124 | 0 | 62 | 30 |
| | | Obs. | 0 | 138 | 182 | 33 | | | Obs. | 121 | 0 | 65 | 31 |
| 75 | 194 | Est. | 0 | 68 | 107 | 19 | 75 | 118 | Est. | 67 | 0 | 34 | 17 |
| | | Obs. | 0 | 74 | 101 | 19 | | | Obs. | 65 | 0 | 36 | 17 |
| 80 | 68 | Est. | 0 | 23 | 38 | 6 | 80 | 39 | Est. | 24 | 0 | 10 | 5 |
| | | Obs. | 0 | 26 | 35 | 7 | | | Obs. | 22 | 0 | 12 | 5 |
| 85 | 34 | Est. | 0 | 13 | 19 | 3 | 85 | 21 | Est. | 11 | 0 | 7 | 3 |
| | | Obs. | 0 | 13 | 18 | 3 | | | Obs. | 11 | 0 | 7 | 3 |
| Total | 22203 | | 0 | 7460 | 11471 | 3272 | Total | 27773 | | 12564 | 0 | 7715 | 7494 |

TABLE 6 *Continued.*

| Age | Total | | Migration from north to East | South | North | West | Age | Total | | Migration from west to East | South | North | West |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 1445 | Est. | 678 | 445 | 0 | 322 | 0 | 905 | Est. | 297 | 302 | 306 | 0 |
| | | Obs. | 814 | 363 | 0 | 268 | | | Obs. | 229 | 395 | 281 | 0 |
| 5 | 917 | Est. | 443 | 283 | 0 | 192 | 5 | 416 | Est. | 140 | 135 | 141 | 0 |
| | | Obs. | 448 | 282 | 0 | 187 | | | Obs. | 108 | 124 | 184 | 0 |
| 10 | 1568 | Est. | 650 | 401 | 0 | 517 | 10 | 563 | Est. | 188 | 186 | 194 | 0 |
| | | Obs. | 771 | 344 | 0 | 453 | | | Obs. | 151 | 171 | 246 | 0 |
| 15 | 5297 | Est. | 3481 | 835 | 0 | 981 | 15 | 2240 | Est. | 555 | 855 | 830 | 0 |
| | | Obs. | 2892 | 1123 | 0 | 1282 | | | Obs. | 772 | 809 | 659 | 0 |
| 20 | 3250 | Est. | 1796 | 676 | 0 | 779 | 20 | 2235 | Est. | 589 | 809 | 837 | 0 |
| | | Obs. | 1861 | 701 | 0 | 688 | | | Obs. | 640 | 816 | 779 | 0 |
| 25 | 1885 | Est. | 911 | 510 | 0 | 464 | 25 | 1373 | Est. | 433 | 459 | 482 | 0 |
| | | Obs. | 998 | 482 | 0 | 405 | | | Obs. | 389 | 491 | 493 | 0 |
| 30 | 924 | Est. | 444 | 282 | 0 | 198 | 30 | 606 | Est. | 208 | 191 | 207 | 0 |
| | | Obs. | 485 | 255 | 0 | 184 | | | Obs. | 173 | 212 | 221 | 0 |
| 35 | 750 | Est. | 362 | 220 | 0 | 168 | 35 | 408 | Est. | 141 | 129 | 139 | 0 |
| | | Obs. | 379 | 213 | 0 | 158 | | | Obs. | 119 | 116 | 173 | 0 |
| 40 | 688 | Est. | 381 | 176 | 0 | 131 | 40 | 361 | Est. | 130 | 111 | 120 | 0 |
| | | Obs. | 411 | 141 | 0 | 136 | | | Obs. | 128 | 87 | 146 | 0 |
| 45 | 447 | Est. | 253 | 119 | 0 | 75 | 45 | 208 | Est. | 74 | 66 | 68 | 0 |
| | | Obs. | 275 | 102 | 0 | 70 | | | Obs. | 81 | 50 | 77 | 0 |
| 50 | 478 | Est. | 251 | 142 | 0 | 86 | 50 | 241 | Est. | 85 | 77 | 79 | 0 |
| | | Obs. | 273 | 119 | 0 | 86 | | | Obs. | 65 | 81 | 95 | 0 |
| 55 | 489 | Est. | 284 | 126 | 0 | 79 | 55 | 237 | Est. | 87 | 75 | 76 | 0 |
| | | Obs. | 295 | 114 | 0 | 80 | | | Obs. | 84 | 64 | 89 | 0 |
| 60 | 439 | Est. | 253 | 122 | 0 | 64 | 60 | 185 | Est. | 65 | 59 | 60 | 0 |
| | | Obs. | 263 | 114 | 0 | 62 | | | Obs. | 60 | 51 | 74 | 0 |
| 65 | 328 | Est. | 190 | 91 | 0 | 47 | 65 | 131 | Est. | 46 | 42 | 43 | 0 |
| | | Obs. | 198 | 84 | 0 | 46 | | | Obs. | 43 | 37 | 51 | 0 |
| 70 | 209 | Est. | 121 | 58 | 0 | 30 | 70 | 82 | Est. | 29 | 26 | 27 | 0 |
| | | Obs. | 125 | 54 | 0 | 30 | | | Obs. | 28 | 21 | 33 | 0 |
| 75 | 109 | Est. | 61 | 31 | 0 | 17 | 75 | 46 | Est. | 16 | 15 | 15 | 0 |
| | | Obs. | 66 | 27 | 0 | 16 | | | Obs. | 14 | 13 | 19 | 0 |
| 80 | 35 | Est. | 20 | 9 | 0 | 5 | 80 | 14 | Est. | 5 | 4 | 5 | 0 |
| | | Obs. | 22 | 8 | 0 | 5 | | | Obs. | 5 | 3 | 6 | 0 |
| 85 | 19 | Est. | 10 | 6 | 0 | 3 | 85 | 7 | Est. | 2 | 2 | 2 | 0 |
| | | Obs. | 11 | 6 | 0 | 2 | | | Obs. | 2 | 2 | 3 | 0 |
| Total | 19277 | | 10587 | 4532 | 0 | 4158 | Total | 10263 | | 3091 | 3543 | 3629 | 0 |

TABLE 7 Error analysis of the modified Friedlander migration estimates for Austria.

*(a) Analysis by size of flow*

| Size of flow | Number of flows | | Number of migrants | | Cumulative absolute % error | | $\chi^2$ | |
|---|---|---|---|---|---|---|---|---|
| | Total | % | Total | % | Value | % | Value | % |
| 0–200 | 112 | 51.85 | 8452 | 10.63 | 1355 | 54.16 | 0.1903e 03 | 11.79 |
| 200–400 | 45 | 20.83 | 12742 | 16.02 | 545 | 21.79 | 0.2963e 03 | 18.35 |
| 400–600 | 20 | 9.26 | 9481 | 11.92 | 153 | 6.11 | 0.9538e 02 | 5.91 |
| 600–800 | 11 | 5.09 | 7687 | 9.67 | 168 | 6.73 | 0.3418e 03 | 21.17 |
| 800–1000 | 9 | 4.17 | 7705 | 9.69 | 63 | 2.50 | 0.6438e 02 | 3.99 |
| 1000–1200 | 3 | 1.39 | 3330 | 4.19 | 51 | 2.02 | 0.1301e 03 | 8.06 |
| 1200–1400 | 7 | 3.24 | 9075 | 11.41 | 90 | 3.61 | 0.2404e 03 | 14.89 |
| 1400–1600 | 1 | 0.46 | 1464 | 1.84 | 1 | 0.06 | 0.3149e 00 | 0.02 |
| 1600–1800 | 0 | 0.00 | 0 | 0.00 | 0 | 0.00 | 0.0000e 00 | 0.00 |
| 1800–2000 | 2 | 0.93 | 3811 | 4.79 | 10 | 0.42 | 0.1237e 02 | 0.77 |
| 2000+ | 6 | 2.78 | 15769 | 19.83 | 65 | 2.60 | 0.2431e 03 | 15.06 |
| Total | 216 | 100.00 | 79516 | 100.00 | 2502 | 100.00 | 0.1614e 04 | 100.00 |

*(b) Analysis by error category*

| Error category | Percentage error[a] | Number of flows | | Number of migrants | | Average flow |
|---|---|---|---|---|---|---|
| | | Total | % | Total | % | |
| 1 | 0–2 | 22 | 10.19 | 9606 | 12.08 | 436.636 |
| 2 | 2–4 | 37 | 17.13 | 10261 | 12.90 | 277.324 |
| 3 | 4–6 | 18 | 8.33 | 7370 | 9.27 | 409.444 |
| 4 | 6–8 | 30 | 13.89 | 10016 | 12.60 | 333.867 |
| 5 | 8–10 | 17 | 7.87 | 10995 | 13.83 | 646.765 |
| 6 | 10–15 | 25 | 11.57 | 7359 | 9.25 | 294.360 |
| 7 | 15–20 | 26 | 12.04 | 10460 | 13.15 | 402.308 |
| 8 | 20–30 | 33 | 15.28 | 12085 | 15.20 | 366.212 |
| 9 | 30–40 | 5 | 2.31 | 1064 | 1.34 | 212.800 |
| 10 | 40–60 | 2 | 0.93 | 68 | 0.09 | 34.000 |
| 11 | 60–100 | 1 | 0.46 | 232 | 0.29 | 232.000 |
| 12 | 100+ | 0 | 0.00 | 0 | 0.00 | 0.000 |
| Total | | 216 | 100.00 | 79516 | 100.00 | 368.130 |

*(c) Analysis by size of flow and error category*

| Size of flow | Error category 0–2 | 2–4 | 4–6 | 6–8 | 8–10 | 10–15 | 15–20 | 20–30 | 30–40 | 40–60 | 60–100 | 100+ | Total |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0–200 | 10 | 22 | 7 | 15 | 6 | 13 | 16 | 17 | 4 | 2 | 0 | 0 | 112 |
| 200–400 | 3 | 9 | 5 | 5 | 4 | 5 | 4 | 9 | 0 | 0 | 1 | 0 | 45 |
| 400–600 | 3 | 2 | 3 | 5 | 3 | 3 | 0 | 1 | 0 | 0 | 0 | 0 | 20 |
| 600–800 | 1 | 1 | 1 | 2 | 0 | 1 | 1 | 3 | 1 | 0 | 0 | 0 | 11 |
| 800–1000 | 3 | 0 | 1 | 1 | 2 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 9 |
| 1000–1200 | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 1 | 0 | 0 | 0 | 0 | 3 |
| 1200–1400 | 0 | 2 | 0 | 1 | 0 | 0 | 3 | 1 | 0 | 0 | 0 | 0 | 7 |
| 1400–1600 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| 1600–1800 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 1800–2000 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 |
| 2000+ | 1 | 0 | 1 | 0 | 2 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 6 |
| Total | 22 | 37 | 18 | 30 | 17 | 25 | 26 | 33 | 5 | 2 | 1 | 0 | 216 |

[a] Average absolute percentage error (relative mean deviation) = 11.03.

A number of topics that need more study have been mentioned in this paper. First, there is the question of how much we need to know in advance to generate acceptable or good estimates of migration flows for subgroups of the population. The answer depends on the homogeneity of the population structure with regard to variables underlying the classification scheme. For instance, the age curve of migrants is very homogeneous across populations. Limited information on the age composition may therefore give good estimates of the flows disaggregated by age.

A related research topic is the development of ways to improve the initial guesses $m_{ijk}^0$. One strategy is to derive the *a priori* distribution $m_{ijk}^0$ from a behavioral migration model in which moves are explained on the basis of push and pull factors, intervening factors, and personal factors or migrant characteristics. This approach was applied by Nijkamp (1976) to estimate commuting flows. Another possible strategy may be to introduce *a priori* expert opinions on detailed migration patterns. Yet another approach is to identify a structure in the interaction flows to be estimated, to represent this structure by a specific mathematical expression, and to introduce it as additional *a priori* information to improve the quality of the estimates. In the entropy methods presented in this paper, it is assumed that the estimates $m_{ijk}$ are independent. However, some kind of interdependence may be introduced without over-complicating the estimation procedure. Current research on model migration schedules is of particular relevance in this regard (Castro and Rogers 1979). The increased attention devoted to the analysis of contingency tables or cross-classified data may also be very useful (Bishop *et al.* 1975, Goodman 1978). In fact, the limited data available require implicit or explicit assumptions to be made about the interdependence of the cross-classified variables. In the 3F problem, pairwise interactions between variables are introduced in the estimation procedure through the marginal constraints. These interaction patterns are absent in the 3E problem, and only partially present in the 1FE problem. The potential of contingency-table analysis for improved estimation of complex migration tables has been explored by Willekens (1980).

A third topic for further research is the extension and application of the techniques presented in this paper for updating migration tables. In countries such as the United States, the United Kingdom, Canada, Japan, and France, censuses are the main sources of migration statistics. Changes in migration patterns during the period between censuses are therefore difficult to assess at a detailed level. The multiproportional adjustment method may be relevant for updating census migration tables using marginal constraints derived from aggregate data for the years between censuses.

Finally, we need to know objectively how well the techniques perform. It is necessary, therefore, to develop one or a set of appropriate statistics to measure in objective terms the "goodness-of-fit" of the methods presented. The $\chi^2$ statistic is used most frequently. Related statistics are also used by Nijkamp and Paelinck (1974a), Hinojosa (1978), and others. However, the use of these test statistics poses some severe problems; for example their values are affected heavily by smaller flows. (See also Forslund and Schoettner 1979.)

The research reported in this paper has led to some potentially very useful estimation methods for migration analysis. It has also led to the formulation of a number of challenging research topics to improve our knowledge of migration patterns. The results presented here and the research priorities identified are not limited, however, to the study of migration. They are applicable to any kind of $n$-dimensional interaction tables, whether they are spatial (e.g., trip distribution, commuting), sectoral (input–output studies), or of some other type. It is not the nature of the data that is important, but their cross-classification; the methods may be applied to all cross-classified data or $n$-dimensional contingency tables regardless of their subject.

## REFERENCES

Bacharach, M. (1970) Biproportional Matrices and Input–Output Analysis. London: Cambridge University Press.

Bartlett, M. S. (1935) Contingency table interactions. Journal of the Royal Statistical Society Series A(2):248–252.

Batty, M., and S. MacKie (1972) The calibration of gravity, entropy and related models of spatial interaction. Environment and Planning A 4:205–233.

Batty, M., and L. March (1976) The method of residues in urban modelling. Environment and Planning A 8:198–215.

Bishop, Y. M., S. E. Fienberg, and P. W. Holland (1975) Discrete Multivariate Analysis. Cambridge, Mass.: M.I.T. Press.

Castro, L., and A. Rogers (1979) Migration Age Patterns: I. Measurement and Analysis. WP-79-16. Laxenburg, Austria: International Institute for Applied Systems Analysis.

Chilton, R., and R. Poet (1973) An entropy maximizing approach to the recovery of detailed migration patterns from aggregate census data. Environment and Planning A 5:135–146.

Drewe, P., and F. Willekens (1980) Maximum Likelihood Estimates of Age-Specific Migration Flows in the Netherlands. Delft Progress Report 5:92–111.

Evans, A. W. (1970) Some properties of trip distribution methods. Transportation Research 4:19–36.

Evans, A. W. (1971) The calibration of trip distribution models with exponential or similar cost functions. Transportation Research 5:15–38.

Evans, S. P., and H. R. Kirby (1974) A three-dimensional Furness procedure for calibrating gravity models. Transportation Research 8:105–122.

Forslund, P., and J. Schoettner (1979) Estimation of Migration Flows: a Validation of Entropy Solutions. WP-79-59. Laxenburg, Austria: International Institute for Applied Systems Analysis.

Friedlander, D. (1961) A technique for estimating a contingency table, given the marginal totals and some supplementary data. Journal of the Royal Statistical Society, Series A:412–420.

Friedmann, R. (1978) Updating input coefficients. Zeitschrift für die Gesamte Staatswissenschaften 134:144–165 (in German).

Goodman, L. (1978) The Analysis of Categorical Data. Cambridge, Mass.: ABT Books, ABT Associates.

Hewings, G. J. D. (1977) Evaluating the possibilities for exchanging regional input–output coefficients. Environment and Planning A 9:927–944.

Hewings, G. J. D., and B. N. Janson (1980) Exchanging regional input–output coefficients: a reply and further comments. Environment and Planning A 12:843–854.

Hinojosa, R. C. (1978) A performance test of the biproportional adjustment of input–output coefficients. Environment and Planning A 10:1047–1052.

Hortensius (1970) Estimation of the Elements of a Table (unpublished manuscript in Dutch). The Hague: Central Planning Agency.

Hyman, G. M. (1969) The calibration of trip distribution models. Environment and Planning A 1:105–112.

Isard, W. (1960) Methods of Regional Analysis. Cambridge, Mass.: M.I.T. Press.

Jaynes, E. T. (1957) Information theory and statistical mechanics. Physical Review 106:620–630.

Kullback, S. (1959) Information Theory and Statistics. New York: Wiley.

Leontief, W. (1941) The Structure of the American Economy 1919–1939. New York: Oxford University Press.

MacGill, S. M. (1977) Theoretical properties of biproportional matrix adjustments. Environment and Planning A 9:687–701.

Nijkamp, P. (1975) Reflections on gravity and entropy models. Regional Science and Urban Economics 5:203–225.
Nijkamp, P. (1976) Spatial Mobility and Settlement Patterns: An Application of Behavioral Entropy. RM-76-45. Laxenburg, Austria: International Institute for Applied Systems Analysis.
Nijkamp, P. (1977) Gravity and Entropy Models: The State of the Art. Research Memorandum No. 1977-2. Amsterdam: Vrije Universiteit, Economische Fakulteit.
Nijkamp, P., and J. Paelinck (1974a) A dual interpretation and generalization of entropy maximizing models in regional science. Papers of the Regional Science Association 33:13–31.
Nijkamp, P., and J. Paelinck (1974b) Some methods for updating input–output tables, in Foundations of Empirical Economic Research. Rotterdam: Netherlands Economic Institute (mimeographed).
Philipov, D. (1978) Migration and settlement in Bulgaria. Environment and Planning A 10:593–617.
Renyi, A. (1970) Probability Theory. Amsterdam: North-Holland Publishing Company.
Rockafellar, R. T. (1970) Convex Analysis. Princeton, N.J.: Princeton University Press.
Shannon, C. E., and W. Weaver (1949) The Mathematical Theory of Communication. Urbana, Ill.: University of Illinois Press.
Stone, R., Ed. (1963) Input–Output Relations, 1954–1966. London: Chapman and Hall.
Tan, E. (1980) On the Estimation of Migration Flows by Migrant Categories. MA Thesis. Inter-university Program in Demography, Brussels.
Theil, H. (1967) Economics and Information Theory. Amsterdam: North-Holland Publishing Company.
Thumann, R. G. (1978) A comment on "evaluating the possibilities for exchanging input–output coefficients." Environment and Planning A 10:321–325.
Willekens, F. (1977) The Recovery of Detailed Migration Patterns from Aggregate Data: An Entropy Maximizing Approach. RM-77-58. Laxenburg, Austria: International Institute for Applied Systems Analysis.
Willekens, F. (1980) Entropy, Multiproportional Adjustment and Analysis of Contingency Tables. Forthcoming in Sistemi Urbani, Vol. 2, No. 2–3.
Willekens, F., and A. Rogers (1978) Spatial Population Analysis: Methods and Computer Programs. RR-78-18. Laxenburg, Austria: International Institute for Applied Systems Analysis.
Willekens, F., A. Pór, and R. Raquillet (1979) Entropy, Multiproportional, and Quadratic Techniques for Inferring Detailed Migration Patterns from Aggregate Data. WP-79-88. Laxenburg, Austria: International Institute for Applied Systems Analysis.
Wilson, A. (1967) A statistical theory of spatial distribution models. Transportation Research 1:253–269.
Wilson, A. (1970) Entropy in Urban and Regional Modeling. London: Pion Limited.

## APPENDIX A: PROOFS OF THEOREMS 1–5

### Proof of Theorem 1

Let $S$ denote the set of feasible solutions to constraints (19)–(27). $S$ is a nonempty, bounded, closed convex set in $R_+^\gamma$ ($\gamma = |\Gamma|$; the cardinality of set $\Gamma$).

Since the objective functions of both the multiproportional and the modified multidimensional Friedlander adjustment problems are strictly convex over $R_+^\gamma$, we have a unique optimal solution $s^* \in S$ in both cases. $s^*$ is a vector containing the elements of the set $\{m_{ijk} | (i, j, k) \in \Gamma\}$.

We now show that if one element $s_i^* = 0$ for some $i$, then a contradiction ensues. By assumption, we know that there exists $\hat{s} \in S$ such that $\hat{s}_i > 0$ for $0 \leqslant i \leqslant \gamma$.

By definition, the value of the modified three-dimensional $\chi^2$ measure ($xe$:30) is $+\infty$ at point $s^*$ and finite at $\hat{s}$, which obviously contradicts the fact that $s^*$ is an optimal value for the modified multidimensional Friedlander adjustment problem.

For the multiproportional adjustment problem, and since $S$ is a convex set, we have

$$\lambda\hat{s}+(1-\lambda)s^* \in S \qquad (\forall 0 \leq \lambda \leq 1)$$

$$\lambda\hat{s}+(1-\lambda)s^* > 0 \qquad (\forall 0 < \lambda \leq 1)$$

Let

$$f(\lambda)=\sum_{i=1}^{\gamma}(\lambda\hat{s}_i+(1-\lambda)s_i^*)\ln\left[\frac{\lambda\hat{s}_i+(1-\lambda)s_i^*}{s_i^0}\right] \qquad 0<\lambda<1$$

where the vector $s^0$ represents the values of the set $\{m_{ijk}^0|(i,j,k)\in\Gamma\}$ in the same way as the vector $s \in S$ represents the values of $\{m_{ijk}|(i,j,k)\in\Gamma\}$.

Consider now the derivative of $f(\lambda)$ for $0<\lambda<1$. We have

$$\frac{\partial f(\lambda)}{\partial\lambda}=\sum_{s_i^*=0}\hat{s}_i\ln\left(\frac{\lambda\hat{s}_i}{s_i^0}\right)+\sum_{s_i^*\neq 0}(\hat{s}_i-s_i^*)\ln\left[\frac{\lambda\hat{s}_i+(1-\lambda)s_i^*}{s_i^0}\right]+\sum_{i=1}^{\gamma}(\hat{s}_i-s_i^*)$$

Since

$$\sum_{s_i^*\neq 0}(\hat{s}_i-s_i^*)\ln\left[\frac{\lambda\hat{s}_i+(1-\lambda)s_i^*}{s_i^0}\right]$$

is bounded over the interval [0, 1] and

$$\lim_{\lambda\to 0}\sum_{s_i^*=0}\hat{s}_i\ln\left(\frac{\lambda\hat{s}_i}{s_i^0}\right)=-\infty$$

there exist such $0<\lambda_0<1$ that

$$\frac{\partial f(\lambda)}{\partial\lambda}<0 \qquad (\forall 0<\lambda<\lambda_0)$$

which obviously contradicts the fact that $s_i^*$ is an optimal solution.

## Proof of Theorems 2 and 3

Before starting the proof of Theorems 2 and 3, we first summarize some features of convex analysis, as presented by Rockafellar (1970).

Let f be a closed, proper convex function on $R^n$ (i.e., f is convex and lower semicontinuous and never assumes the value $-\infty$, although $+\infty$ is allowed). The effective domain of f is the convex set dom f $=\{x: f(x)<+\infty\}$. The conjugate function of $f(x)$ is denoted by $f^*(x^*)$ and is defined as

$$f^*(x^*)=\sup\{\langle x,x^*\rangle-f(x)|x\in \text{ri}(\text{dom } f)\}$$

where ri denotes the relative interior of the set dom f. The proof of the duality results will rely on the decomposition principle suggested by Rockafellar (1970).

First, we shall consider the problem of minimizing

$$f_1(x_1)+\cdots+f_n(x_n) \qquad x=(x_1,x_2,\ldots,x_n)\in R^n \tag{A1}$$

subject to

$$\mathbf{A}\boldsymbol{x} = \boldsymbol{b}$$

and

$$\boldsymbol{x} \geqslant 0$$

where $\mathrm{f}_i(x_i)$ is a proper convex function and $\boldsymbol{b}$ is a given vector in $R^n$.

*Theorem A1*

If the infimum in the above problem is finite and there exists $\boldsymbol{x} \in R^n$ such that

$$\mathbf{A}\boldsymbol{x} = \boldsymbol{b}$$

and

$$\boldsymbol{x} > 0$$

(i.e., all the components of $\boldsymbol{x}$ are strictly positive), then by minimizing the convex function

$$w(\boldsymbol{\lambda}) = \mathrm{f}_1^*(-\langle \boldsymbol{a}^{(n)}, \boldsymbol{\lambda}\rangle) + \cdots + \mathrm{f}_n(-\langle \boldsymbol{a}^{(n)}, \boldsymbol{\lambda}\rangle) + \langle \boldsymbol{b}, \boldsymbol{\lambda}\rangle \qquad \boldsymbol{\lambda} \in R^n$$

(where $\mathrm{f}_i^*$ is conjugate to $\mathrm{f}_i$ and $\boldsymbol{a}^{(i)}$ represents the $i$th column of matrix $\mathbf{A}$), and using any member of the minimum set $\boldsymbol{\lambda}^* \in R^n$ of it, the following subproblems define an optimal solution to the primal problem

$$\min_{x_i} \mathrm{f}_i(x_i) + x_i\langle \boldsymbol{a}^{(i)}, \boldsymbol{\lambda}^*\rangle \qquad (i = 1, 2, \ldots, n)$$

For the proof of Theorem A1, see Section 28 of Rockafellar (1970).

Both the I-divergence measure eqn. (28) and the modified three-dimensional $\chi^2$ measure can be separated into single-valued convex functions of the variables $m_{ijk}$.

Since the constraints (19)–(27) are linear equations of variables $m_{ijk}$, we can decompose both the multiproportional and the modified multidimensional Friedlander adjustment problems as discussed above.

The conjugate of the functions

$$\mathrm{f}(x) = \begin{cases} x \ln(x/a) & \text{if } x > 0 \\ 0 & \text{if } x = 0 \\ +\infty & \text{if } x < 0 \end{cases} \qquad a > 0 \qquad x \in \boldsymbol{R}_+$$

and

$$\mathrm{g}(x) = \begin{cases} (x-a)^2/x & \text{if } x > 0 \\ +\infty & \text{if } x \leq 0 \end{cases}$$

can be obtained in the form

$$\mathrm{f}^*(x^*) = a\,\mathrm{e}^{x^*-1} \qquad \forall x^* \in \boldsymbol{R}$$

$$\mathrm{g}^*(x^*) = \begin{cases} -2a\sqrt{1-x^*}+2 & \text{if } x^* < 1 \\ +\infty & \text{if } x^* \geqslant 1 \end{cases} \qquad x \in \boldsymbol{R}_+$$

It is now easy to see that the functions $W$ for the multiproportional and for the modified multidimensional Friedlander problems will take the forms $L_1(\mathbf{\Lambda}, \mathbf{N}, \mathbf{H})$ and $L_2(\mathbf{\Lambda}, \mathbf{N}, \mathbf{H})$, respectively.

It is easy to prove that one of the dual variables can be fixed at any constant value in both dual problems without changing the optimal value of the objective function. Theorems 2 and 3 are direct corollaries of Theorem A1.

## Proof of Theorems 4 and 5

Before starting the proof of Theorems 4 and 5, we must summarize a number of points from convex analysis.

The recession function of f is denoted by $f0^+$ and defined as

$$(f0^+)(\boldsymbol{x}) = \lim_{\alpha \downarrow 0} \alpha f(\alpha^{-1}\boldsymbol{x}) \quad \boldsymbol{x} \in \text{dom } f$$

If, for a nonzero vector, $\boldsymbol{x} \in \text{dom } f$, $(f0^+)(\boldsymbol{x}) \leq 0$, then vector $\boldsymbol{x}$ is known as the "direction in which f recedes" or the direction of recession of f.

In the following analysis, our attention will be focused on the properties of the parameterized nest of level sets

$$\text{lev}_\alpha f = \{\boldsymbol{x} | f(\boldsymbol{x}) \leq \alpha\} \qquad \alpha \in R$$

belonging to a given proper convex function f.

Let inf f denote the infimum of $f(\boldsymbol{x})$ as $\boldsymbol{x}$ ranges over $R^n$. For $\alpha = \inf f$, $\text{lev}_\alpha f$ consists of the points $\boldsymbol{x}$ where the infimum of f is attained. We call this level set the minimum set of f.

We shall need the following properties of the level sets:

*Lemma 1*

Let f be a closed proper convex function which has no direction of recession. The infimum of f is then finite and attained. Moreover, all the level sets $\text{lev}_\alpha f (\alpha \geq \inf f)$ are nonempty, closed, bounded convex sets.

For the proof of Lemma 1, see Theorem 27.1 of Rockafellar (1970).

To prove the special algorithms given in Section 3 it is necessary to formulate a general procedure for a class of differentiable, closed proper convex functions which have no direction of recession and converge to the minimum set.

Let $f(\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_m)$ be a closed proper convex function on $R^n$, where $\boldsymbol{x}_i \in R^{n_i}$ for all $1 \leq i \leq m$ and $\sum n_i = n$.

For a given $\boldsymbol{x}^0 \in \text{dom } f$, let us define the sequence of points $\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)}, \ldots$ recursively, by letting $\boldsymbol{x}^{(k)}$ be one of the solutions of the minimizing problem

$$\min_{\boldsymbol{x}_l \in R^{n_l}} f(\boldsymbol{x}_1^{(s)}, \ldots, \boldsymbol{x}_{l-1}^{(s)}, \boldsymbol{x}_l, \boldsymbol{x}_{l+1}^{(s-1)}, \ldots, \boldsymbol{x}_m^{(s-1)}) \tag{A2}$$

where

$$k = ms + l \qquad 1 \leq l \leq m$$

*Lemma 2*

Let f be a closed, proper convex function on $R^n$ as defined in the above algorithm. Further, let $\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)}, \ldots,$ be the sequence generated by the algorithm.

If f is continuously differentiable at every $x \in \text{dom f}$ and has no direction of recession, then

$$\lim_{k\to\infty} f(x^{(k)}) = \inf f$$

and $x^{(1)}, x^{(2)}, \ldots$ is a bounded sequence and all its cluster points belong to the minimum set of f.

*Proof*

By definition of the sequence $x^{(1)}, x^{(2)}, \ldots$ we have

$$+\infty > f(x^{(0)}) \geq f(x^{(1)}) \geq, \ldots, \geq f(x^{(k-1)}) \geq f(x^{(k)}) \geq \alpha$$

Hence

$$\lim_{k\to\infty} f(x^{(k)}) = \alpha \geq \inf f \qquad \text{(A3)}$$

From Lemma 1, it follows that the sequence $x^{(1)}, x^{(2)}, \ldots$ is bounded and $\inf f < +\infty$.

Let $x_c$ denote a cluster point of the sequence $x^{(1)}, x^{(2)}, \ldots$. From (A3) we obtain

$$x_c \in \{x \mid f(x) = \alpha\} \qquad \text{(A4)}$$

We shall assume that $\alpha > \inf f$ and demonstrate that this implies a contradiction. Since function f is differentiable over dom f, we have from eqn. (A2)

$$\left.\frac{\partial f(x_1, \ldots, x_l, \ldots, x_m)}{\partial x_{l_i}}\right|_{x=x^{(k)}} = 0 \qquad (\forall 1 \leq i \leq n_l)$$

for all $l$, $i$, and $k$, where $k = ms + l$ and $x_l = (x_{l1}, x_{l2}, \ldots, x_{ln_l}) \in R^{n_l}$. Hence

$$\left.\frac{\partial f(x)}{\partial x_i}\right|_{x=x_c} = 0 \qquad (\forall 1 \leq i \leq n) \qquad \text{(A5)}$$

which contradicts the fact that $\alpha > \inf f$, because a necessary and sufficient condition for a given point $x$ to belong to the minimum set is given by eqn. (A5).

*Lemma 3*

Let $L_1(\mathbf{\Lambda}, \mathbf{N}, \mathbf{H})$ and $L_2(\mathbf{\Lambda}, \mathbf{N}, \mathbf{H})$ be the functions defined in eqns. (31) and (32).

If there exists a feasible solution to the constraints (19)–(27) such that

$$m_{ijk} > 0 \qquad (i, j, k) \in \Gamma$$

then $L_1(\mathbf{\Lambda}, \mathbf{N}, \mathbf{H})$ and $L_2(\mathbf{\Lambda}, \mathbf{N}, \mathbf{H})$ are closed proper convex functions and have no direction of recession.

*Proof*

From the simple form of functions $L_1$ and $L_2$, it is almost a trivial matter to see that the function is a closed proper function over all values of variables $\mathbf{\Lambda}, \mathbf{N}, \mathbf{H}$.

By definition, the recession function of $L_1$ is given as

$$(L_1 0^+)(\mathbf{\Lambda}, \mathbf{N}, \mathbf{H}) = \lim_{\alpha\downarrow 0} \alpha L_1(\alpha^{-1}\mathbf{\Lambda}, \alpha^{-1}\mathbf{N}, \alpha^{-1}\mathbf{H})$$

Recall that $\lambda_{11}$ is not treated as a variable but has a constant value of 1.

If there exist $(i, j, k) \in \Gamma$ such that

$$\lambda_{ij} + \nu_{ik} + \zeta_{jk} < 0 \qquad \text{if } i \neq 1 \text{ or } j \neq 1$$

$$\nu_{1k} + \zeta_{1k} < 0 \qquad \text{if } i = j = 1$$

then

$$(L_1 0^+)(\mathbf{\Lambda}, \mathbf{N}, \mathbf{H}) = +\infty$$

In the case when $\nu_{1k} + \zeta_{1k} \geqslant 0$ for all $k \in K$, and $\lambda_{ij} + \nu_{ik} + \zeta_{jk} \geqslant 0$ for all $(i, j, k) \in \Gamma$, we have

$$(L_1 0^+)(\mathbf{\Lambda}, \mathbf{N}, \mathbf{H}) = \sum_{i=1}^{n} \sum_{j=1}^{m} c_{ij}\lambda_{ij} + \sum_{i=1}^{n} \sum_{k=1}^{l} b_{ik}\nu_{ik} + \sum_{j=1}^{m} \sum_{k=1}^{l} a_{jk}\zeta_{jk}$$

Since there exists a feasible solution to the basic problem such that

$$m_{ijk} > 0 \qquad (i, j, k) \in \Gamma$$

$$m_{ijk} = \overset{0}{m}_{ijk} \qquad (i, j, k) \notin \Gamma$$

we have

$$\sum_{i=1}^{n} \sum_{j=1}^{m} c_{ij}\lambda_{ij} + \sum_{i=1}^{n} \sum_{k=1}^{l} b_{ik}\nu_{ik} + \sum_{j=1}^{m} \sum_{k=1}^{l} a_{jk}\zeta_{jk} = \sum_{(i,j,k)\in\Gamma} m_{ijk}(\lambda_{ij} + \nu_{ik} + \zeta_{jk})$$

Since

$$\lambda_{11} + \nu_{1k} + \zeta_{1k} > 0 \qquad k \in K$$

and since there exists $k \in K$ such that

$$(1, 1, k) \in \Gamma$$

we have

$$(L_1 0^+)(\mathbf{\Lambda}, \mathbf{N}, \mathbf{H}) > 0$$

The proof for function $L_2(\mathbf{\Lambda}, \mathbf{N}, \mathbf{H})$ can be derived in the same way as for function $L_1(\mathbf{\Lambda}, \mathbf{N}, \mathbf{H})$. It is easy to see that the solution algorithms for our dual problems (31) and (32) are special cases of our general procedure.

The proof of Theorems 4 and 5 obviously follows from Lemmas 1, 2, and 3.

## APPENDIX B: MULTIPROPORTIONAL SOLUTIONS FOR THE SPECIAL CASES (3E), (1FE), AND (2F)

Note that in the two-dimensional entropy problem excluding any cost factor, the most probable values of the elements of interaction tables can be obtained analytically and are given by the simple expression

$$m_{ij} = \frac{O_i I_j}{\sum_i I_i} = \frac{m_{i.} m_{.j}}{m_{..}}$$

The element $m_{i.}$ denotes the total number of departures out of $i$, and $m_{.j}/m_{..}$ is the proportion of all migrants that go to $j$. This proportion is independent of the region of origin. The most probable number of migrants from $i$ to $j$ is simply the geometric average of the numbers of departures from $i$ and arrivals at $j$. In other words, the best estimates are obtained by multiplying the number of out-migrants by a fixed in-migration profile (multiplicative spatial interaction model).

In the multidimensional case, there are also some analytical solutions to the I-divergence (entropy maximization) problem. A necessary condition for these solutions to exist, however, is that the initial distribution is uniform and that the constraints correspond to (3E), (1FE), or (2F).

### The Three-edge (3E) Problem

In the (3E) problem it is necessary to minimize (18) subject to (22)–(26). Let $i$ denote the region of origin, $j$ the region of destination, and $k$ the age group. The solution to the (3E) problem is given by

$$m_{ijk} = u_k v_j w_i / (ST)^2$$

or

$$m_{ijk} = m_{.j.} m_{..k} m_{i..} / (m_{...})^2 \qquad \text{(B1)}$$

where

$$m_{i..} = \sum_{j,k} m_{ijk}$$

$$m_{.j.} = \sum_{i,k} m_{ijk}$$

$$m_{..k} = \sum_{i,j} m_{ijk}$$

Note that the solution assumes that all three classifications are mutually independent (e.g., number of arrivals is independent of number of departures).

### The One-face, One-edge (1FE) Problem

In the (1FE) problem it is necessary to minimize (18) subject to (21), (22), and (26). Suppose the given face consists of the flow matrix of the total population $c_{ij}$ and the edge represents the age structure of the migrants at the national level. The best estimates of the flow matrices disaggregated by age are given by

$$m_{ijk} = c_{ij} u_k / ST$$

or

$$m_{ijk} = m_{ij.} m_{..k} / m_{...} \qquad \text{(B2)}$$

where

$$m_{ij.} = \sum_{k} m_{ijk}$$

The ratio $m_{..k}/m_{...}$ is the national age composition of the migrants. It is applied to all values of the flow matrix $m_{ij.}$ and hence the age structure is uniform for all flows.

The solution (B2) is therefore a model describing the migration flow under the assumption of a uniform age profile for migrants, and may be referred to as the age-profile model. If the given face consists of the age composition of the departing migrants ($m_{i.k}$), and the edge is the number of in-migrants by region ($m_{.j.}$), the model is an in-migration profile model. Conversely, if the age structure of arriving migrants ($m_{.jk}$) is known, together with the total number of out-migrants by region ($m_{i..}$), we have an out-migration profile model. Which of the models yields the best results depends on the homogeneity of the categories considered with respect to the profile adopted. Since the real age profile of migrants is quite uniform, the age-profile model may well yield better estimates.

### The Two-face (2F) Problem

In the (2F) problem it is necessary to minimize (18) subject to (19), (21), and (26). The solution is for given values of $c_{ij}$ and $a_{jk}$.

$$m_{ijk} = c_{ij}a_{jk}/v_j$$

or

$$m_{ijk} = m_{ij.}m_{.jk}/m_{.j.} \tag{B3}$$

The ratios $m_{.jk}/m_{.j.}$ are conditional probabilities. The solution therefore implies the assumption of conditional independence between the $j$ (destination) and $i$ (origin) classifications for every $k$ (migrant category, e.g., age group). In eqn. (B3) the age structure of migrants is destination-specific. The ratio $m_{.jk}/m_{.j.}$ represents the age curve of in-migrants, and is independent of the region of origin.

# AGE PATTERNS OF MIGRATION: CAUSE-SPECIFIC PROFILES*

*Andrei Rogers and Luis J. Castro*

## 1 INTRODUCTION

Studies have shown that the age pattern of deaths varies systematically with the level of mortality. For example, as the expectation of life at birth increases, the largest absolute declines in mortality generally occur at ages below 5 and above 65. This is a consequence of the dramatic reduction in the contribution to overall deaths made by infectious diseases, which have a U-shaped age profile of mortality. Are there analogous systematic variations in age patterns of migration? Does the age pattern of migration vary with the level of migration? For example, if divorce is a reason for migration, and if the level of migration and the number of divorces per capita both increase with economic development, should one then expect a particular shift in the age profile of aggregate migration?

Preston (1976, p. 109) pointed out that:

> ... roughly half of the variance in age-curves of mortality at a particular mortality level can be accounted for by variance in relative importance of 6 or 7 cause of death categories among populations at that level ... it suggests that causes of death have substantial value in accounting for disparities in age patterns in mortality.

Can the same be said of disparities in age patterns of migration?

This paper seeks to illuminate the role played by various reasons for migration in the observed variations of age-specific migration rates. The focus is on the *levels* and *age profiles* of different reason-specific migration schedules and on their contribution to age curves of aggregate migration and their changes over time and space. Because we follow the "mortality" analogy in disaggregating migration schedules, we refer to reasons as causes. In future work, however, we shall consider the "fertility" analogy in analyzing such schedules, adopting a disaggregation by status instead of by cause.

## 2 MIGRATION DISAGGREGATED BY CAUSE

Why people move is a question that needs to be considered with respect to (1) those characteristics of potential migrants that condition receptivity to migration and

* Based on WP-79-65.

(2) those environmental factors that stimulate migration from one community to another. Nevertheless, some insight into motivations for migration may be obtained simply by asking people why they moved. This approach has been adopted, for example, in nationwide surveys conducted by the U.S. Bureau of the Census (Long and Hansen 1979) and in national migration registers maintained in such countries as Czechoslovakia (Kühnl 1978).

Studies of reported causes for migration within a given country are subject to a number of serious limitations. First, usually only the "main" cause is tabulated and examined, yet multiple interdependent causes underlie migration behavior. Second, the number of alternative causes listed in migration questionnaires are typically broad aggregations of a much wider range of causes and therefore may inadequately reflect the true importance of motivations connected with migration. Finally, problems arise when the causes are not separately classified for the initiators of migration (e.g., household heads) and for their dependents (e.g., children). In short, reported causes of migration are often mutually interdependent, usually insufficient in number, and generally are not linked directly to the true decision maker. However, analogous limitations also appear in studies of mortality disaggregated by cause, without presenting insuperable obstacles. As noted by Preston (1976, p. 2):

> Causes are undoubtedly recorded with considerable inaccuracy and interpopulation incomparability, and these problems have discouraged the exploitation of cause-of-death statistics. But demographic data are never perfectly accurate, and the choice is between neglecting them altogether and producing qualified statements about the tendencies they suggest.

Part A of Table 1 gives the percentage of household heads moving for each of five causes in the USA and in Hungary. These data confirm that it is a great oversimplification to explain migration solely in terms of economic motivations, i.e., employment. Although approximately half of the migrating household heads cited employment as the main reason for moving, education, marriage, housing, and other

TABLE 1 Migration data disaggregated by cause: USA[a], Hungary[b], and Czechoslovakia[c], various dates.

| | | Percentage of migrants citing the cause | | | | |
|---|---|---|---|---|---|---|
| Region | Date | Employment | Education | Marriage | Housing | Other |
| *A. Household heads only* | | | | | | |
| USA | 1974–1976 | 56.6 | 5.4 | 1.6 | 8.1 | 28.3 |
| Hungary | 1958 | 49.7 | 2.5 | 15.4 | 12.0 | 20.4 |
| Hungary | 1968 | 43.8 | 1.7 | 21.5 | 14.1 | 18.9 |
| *B. All migrants* | | | | | | |
| USA | 1974–1976 | 59.8 | 3.9 | 1.4 | 8.0 | 26.9 |
| Czechoslovakia | 1973 | 28.1 | 1.0 | 17.0 | 41.8 | 12.1 |
| Czech Republic | 1973 | 20.9 | 0.7 | 17.1 | 47.8 | 13.5 |

[a] USA data are taken from Long and Hansen (1979) and refer to inter-state migration.
[b] Hungarian data are taken from Compton (1971) and refer to all inter-community migration.
[c] Czechoslovakian data are taken from Kühnl (1978) and refer to all inter-community migration; the Czech Republic and the Slovak Republic together comprise the nation of Czechoslovakia.

reasons together provided the motivation for the other half to migrate. Moreover, the data indicate that, in Hungary, employment as a cause of migration has been declining in relative importance over time.

Part B of Table 1 presents comparable data for all migrants, including the household head. Only 36% of all migrants were found to be household heads in the USA survey; in Hungary the corresponding proportion ranged from 55% in 1958 to 63% in 1968. The data for Czechoslovakia do not distinguish between household heads and their accompanying dependents.

Housing reasons accounted for over 40% of all migration between communities (communes) in Czechoslovakia in 1973; this total is about five times as high as the figure for the USA. Data for the USA, however, refer to *inter-state* migration, and one would expect housing reasons to decline in importance relative to employment reasons when considering migrations over such relatively greater distances.

Less than 30% of migration within Czechoslovakia was caused by changes in employment. This relatively low share of the total is somewhat surprising and apparently reflects a leveling of regional economic differences (Kühnl 1978, p. 4):

> . . . for the major part, it is the outcome of specific Czechoslovak conditions: In recent years, development in the whole Czechoslovak Socialist Republic, but particularly in the Czech Socialist Republic, was characterized by a relatively balanced territorial development of productive forces, accompanied by a levelling-out of regional differences in economic and income structure and in the training requirements of available jobs.

Table 2 shows that the cause-specific pattern of migration varies with distance migrated. For example, the two principal causes of migration in the Czech Republic,

TABLE 2 Migration data[a] disaggregated by cause and by distance migrated: Czech Republic, 1966–1973.

| Distance migrated | Year | Percentage of migrants citing the cause[b] | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| Between regions | 1967 | 30.6 | 6.3 | 0.9 | 5.2 | 13.5 | 2.0 | 36.0 | 4.3 | 1.2 |
| | 1969 | 28.4 | 5.6 | 1.5 | 5.0 | 15.3 | 3.0 | 37.0 | 4.1 | 0.1 |
| | 1971 | 26.4 | 6.0 | 1.3 | 5.2 | 13.6 | 2.9 | 39.1 | 5.4 | 0.1 |
| | 1973 | 24.4 | 5.7 | 1.4 | 5.6 | 15.5 | 3.4 | 38.3 | 5.7 | 0.0 |
| Between districts within regions | 1967 | 23.3 | 8.1 | 0.6 | 4.4 | 13.0 | 2.1 | 43.6 | 4.0 | 0.9 |
| | 1969 | 21.2 | 7.4 | 0.6 | 4.0 | 16.0 | 3.1 | 43.9 | 3.8 | 0.0 |
| | 1971 | 19.1 | 7.1 | 0.6 | 4.2 | 15.2 | 3.0 | 45.2 | 5.5 | 0.1 |
| | 1973 | 16.8 | 6.6 | 0.6 | 4.4 | 17.5 | 3.3 | 45.4 | 5.4 | 0.0 |
| Between communities within districts | 1967 | 12.2 | 7.9 | 0.2 | 4.5 | 13.1 | 1.7 | 55.8 | 3.9 | 0.7 |
| | 1969 | 9.5 | 6.8 | 0.2 | 4.8 | 18.4 | 2.7 | 54.3 | 3.7 | 0.1 |
| | 1971 | 8.7 | 6.1 | 0.2 | 3.8 | 16.6 | 2.6 | 56.5 | 5.3 | 0.2 |
| | 1973 | 7.6 | 5.9 | 0.2 | 4.5 | 17.9 | 2.8 | 55.4 | 5.7 | 0.0 |

[a] Taken from Kühnl (1978, p. 7).
[b] Causes for migration are abbreviated as follows:
1, change of employment; 2, moving closer to place of work; 3, education; 4, health; 5, marriage; 6, divorce; 7, housing; 8, other; 9, unknown.

employment and housing, exhibit opposite relationships with the distance migrated.* They mirror the relationship shown in the USA and Czechoslovakian data given earlier in Part B of Table 1, which reflect the commonly observed tendency that short-distance migration is primarily motivated by housing reasons and long-distance migration by employment reasons. For, as Hoover (1971, p. 169) pointed out in a study of USA data:

> ... those men who only moved within the same county were predominantly influenced by housing considerations. Since all of a county is generally regarded as being in a single labor market or commuting range, job changes are related only to a minor extent with intracounty moves: most such movers are not changing jobs.
>
> For those who moved to a different county ... the picture is quite different, with employment changes (including entry to or exit from military service) emerging as the major reasons for migrating. This reflects the fact that an intercounty migration generally involves shifting to a different labor market beyond the commuting range for the former job.

Causes of migration are related to a person's age and sex. For example, migration motivated by health reasons is a phenomenon characteristic of old persons, whereas education-related migration is predominantly associated with young people. Wives tend to be younger than their husbands; therefore the age profile of female migration peaks at an earlier age than the corresponding profile for males. Thus, in order to understand better why people move, it is important to disaggregate cause-specific migration data by age and by sex.

Table 3 illustrates typical migration-by-cause data disaggregated by age and sex, and it may be seen that most moves occur before the age of 40. Only health- and housing-related migrations occur in significant amounts at later ages. Finally, the concentration of migration within certain age groups is most pronounced for migration associated with marriage (60–80% between the ages of 20 and 30) and education (94–97% between the ages of 16 and 30).

## 3 DESCRIPTION: CAUSE-SPECIFIC AGE PROFILES

If the age pattern of migration is influenced by its cause-specific structure then it should be possible to attribute differences in age patterns of migration in two or more populations, at least partially, to differences in their cause-specific structures. Unfortunately, detailed age-specific migration data that are disaggregated by cause are exceedingly scarce, and we have been able to find only one source for this study: the Czechoslovakian migration register.**

* There are twelve administrative regions in Czechoslovakia: eight in the Czech Republic and four in the Slovak Republic. The nation is composed of about 100 districts and approximately 10,000 communities (communes).

** Identification of causes of migration has been a part of the regular internal migration register of Czechoslovakia since 1966. The data are based on responses given by migrants at the time that they notify local authorities of their change of address. Dependents are not distinguished from household heads in these data.

TABLE 3 Age and sex differentials in cause-specific patterns of migration: Hungary, 1958 and Czechoslovakia, 1973[a].

| Causes of migration | Country[b] | Percentage citing a given cause in age group | | | | | |
|---|---|---|---|---|---|---|---|
| | | 16–19 | 20–29 | 30–39 | 40–49 | 50–59 | 60+ |
| *A. Males* | | | | | | | |
| Change of | Hu | 9.8 | 41.7 | 26.8 | 11.8 | 7.1 | 2.8 |
| employment | Cz | 7.9 | 48.2 | 23.1 | 15.1 | 4.7 | 1.0 |
| Closer to | Hu | 5.7 | 47.0 | 28.2 | 10.4 | 6.3 | 2.4 |
| place of work | Cz | 7.0 | 58.3 | 18.2 | 11.1 | 4.2 | 1.2 |
| Education | Hu | 68.8 | 28.0 | 2.4 | 0.5 | 0.3 | — |
| | Cz | 40.9 | 53.6 | 4.1 | 1.0 | 0.2 | 0.2 |
| Health | Hu | 2.5 | 15.0 | 17.2 | 13.0 | 18.6 | 33.7 |
| | Cz | 5.3 | 13.2 | 9.1 | 8.4 | 9.1 | 54.9 |
| Marriage | Hu | 3.9 | 73.0 | 14.0 | 4.6 | 2.8 | 1.7 |
| | Cz | 3.8 | 81.8 | 8.9 | 3.2 | 1.4 | 0.9 |
| Housing | Hu | 6.0 | 27.1 | 23.4 | 12.7 | 12.4 | 18.4 |
| | Cz | 7.1 | 53.0 | 17.5 | 8.5 | 4.4 | 9.5 |
| *B. Females* | | | | | | | |
| Change of | Hu | 20.8 | 44.8 | 18.7 | 8.8 | 4.8 | 2.1 |
| employment | Cz | 12.5 | 48.8 | 22.0 | 12.2 | 3.2 | 1.3 |
| Closer to | Hu | 15.0 | 52.9 | 19.4 | 7.2 | 4.0 | 1.5 |
| place of work | Cz | 12.9 | 61.2 | 14.8 | 7.4 | 2.8 | 0.9 |
| Education | Hu | 70.1 | 26.8 | 2.5 | 0.3 | 0.2 | 0.1 |
| | Cz | 62.8 | 34.6 | 1.2 | 0.8 | 0.6 | 0.0 |
| Health | Hu | 4.2 | 15.5 | 15.2 | 9.4 | 17.6 | 38.2 |
| | Cz | 2.9 | 11.9 | 5.3 | 4.5 | 8.1 | 67.3 |
| Marriage | Hu | 24.8 | 59.5 | 9.6 | 3.3 | 2.0 | 0.8 |
| | Cz | 22.3 | 69.1 | 4.7 | 2.3 | 1.0 | 0.6 |
| Housing | Hu | 6.6 | 27.8 | 18.6 | 11.0 | 12.0 | 24.0 |
| | Cz | 10.8 | 52.6 | 12.3 | 7.1 | 5.6 | 11.6 |

[a] Taken from Compton (1971, pp. 90, 91) and Czechoslovakian Federal Statistical Office (1974). The migration of those less than 16 years of age is assumed here to be dependent migration.
[b] Countries are abbreviated as Hu, Hungary and Cz, Czechoslovakia.

Figure 1 displays histograms and their associated cubic-spline interpolations (McNeil *et al.* 1977) for age-specific male and female migration rates in Czechoslovakia. Figure 2 presents the age-specific cause-of-migration structures that underlie these rates. For ease of visual comparison all age profiles have been scaled so that the area under the curve (the Gross Migraproduction Rate, or *GMR*) is unity.

The age profiles reveal that the causes of migration have quite different age patterns. Of the eight causes illustrated, the age profile of *housing* reasons is most similar to that of the aggregate migration schedule, exhibiting roughly the same four peaks: during infancy, during the early years of labor-force participation, at retirement, and in the oldest age group. Migrations due to *marriage* and *education*, on the other hand, are concentrated between the ages of 10 and 30 and are essentially unimodal in age profile. Migrations caused by *divorce*, *change of employment*, and *moving closer to the place of work* have profiles that are bimodal, with local peaks during infancy and during the early years of labor-force participation. Finally, *health*

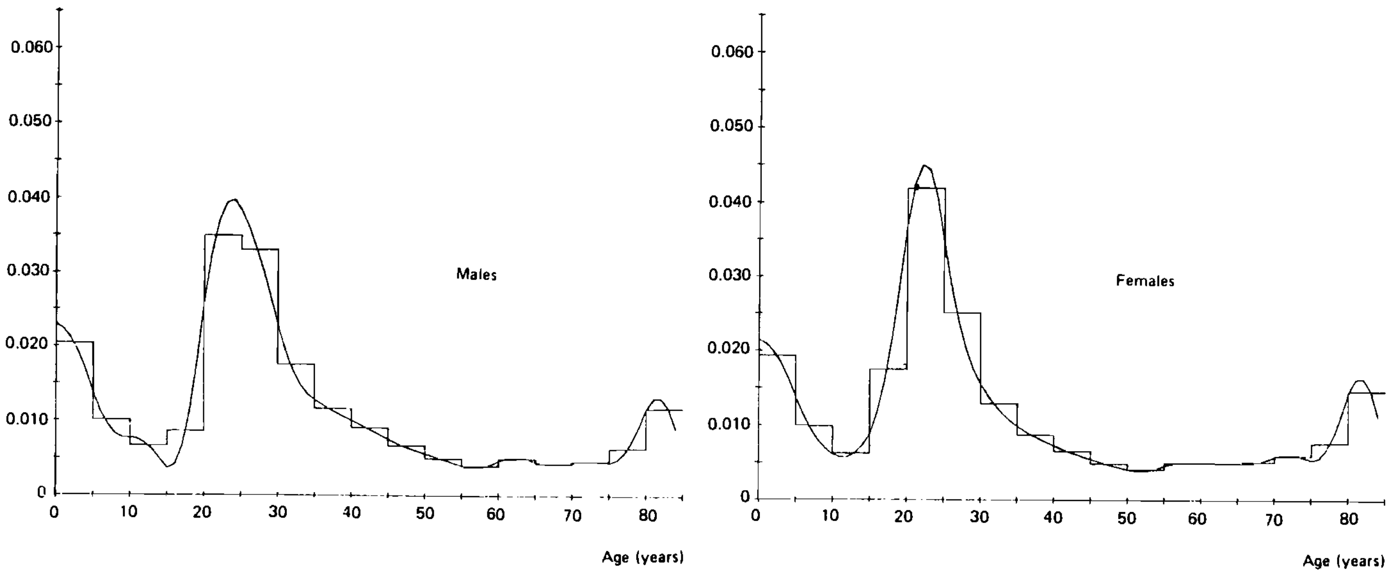



FIGURE 1 Observed annual migration rates for all causes combined: Czechoslovakia, males and females, 1973.

is apparently an important cause of migration only for the elderly. (The residual category "all other reasons" is aggregated with divorce in Figure 2 in order to give it a profile that is more amenable for our subsequent analysis.)

The different cause-specific age patterns may be interpreted within a life-cycle framework in which individuals pass through different states of existence. Starting with birth and then entry into the educational system at the elementary level, the "passage" may also include entry into military service or university, marriage, multiple entries into and withdrawals from the labor force, perhaps divorce and remarriage, retirement, death of spouse, and moves to enter sanatoria or to rejoin relatives.

Associated with this individual life-cycle perspective is a family life cycle which begins with marriage, passes on to procreation and child rearing (possibly interrupted by divorce or death), continues with child "launching", retirement, and ultimately ends with the death of both spouses. Young (1975, p. 61) described this family life cycle as follows:

> The main stages through which a family passes during its lifetime are the initial childless stage, usually lasting one or two years, the childbearing stage, defined by the interval between the first and last birth, usually extending over approximately eight years, and the intermediate stage occurring between the birth of the last child and the first child leaving home, extending for about 16 years and representing the only period during which all members of the family are living in the household. This is followed by what is often referred to as the "launching" stage, occurring between the events of the first and last child leaving home, usually of comparable length to the childbearing stage. Beyond the launching stage, when the parents are again alone in the household is the post-parental stage ending when one of the spouses dies. Then follows a period of widowhood until the death of the other parent. Obviously, the sequence and timing of the stages of the life cycle differ from this for families prematurely broken by death or divorce of the parents.

Figure 2 presented cause-specific age profiles scaled to a *GMR* of unity for ease of comparison. The *relative levels* of the cause-specific contributions are illustrated in Figure 3, which also includes comparable data for an earlier year, 1970, to permit an examination of changes over time. The aggregate migration profile, however, is still scaled such that the area under the curve is unity.

Figure 3 shows that both the *levels* and the *age profiles* of the 1970 and 1973 cause-specific structures of migration are roughly similar. Housing reasons account for approximately 40% of the total gross migraproduction rate, economic reasons (change of employment, moving closer to place of work) for an additional 25–30%, marriage for 12–15%, health for 8–11%, and all other reasons for the remaining 5–15%. The major difference in age profiles occurs in the post-retirement ages, with the 1973 schedules exhibiting a much more pronounced old-age peak.

Figure 4 compares the levels and age profiles of male *intra*-republic moves with those of male *inter*-republic moves in an effort to identify the possible effects of distance on the cause-specific structure of migration. Here the findings are not so clear cut. Although intra-republic migration levels disaggregated by cause seem to be similar in both the Czech and the Slovak republics, this is not true for the corresponding inter-republic migration data. Housing reasons, for example, contribute

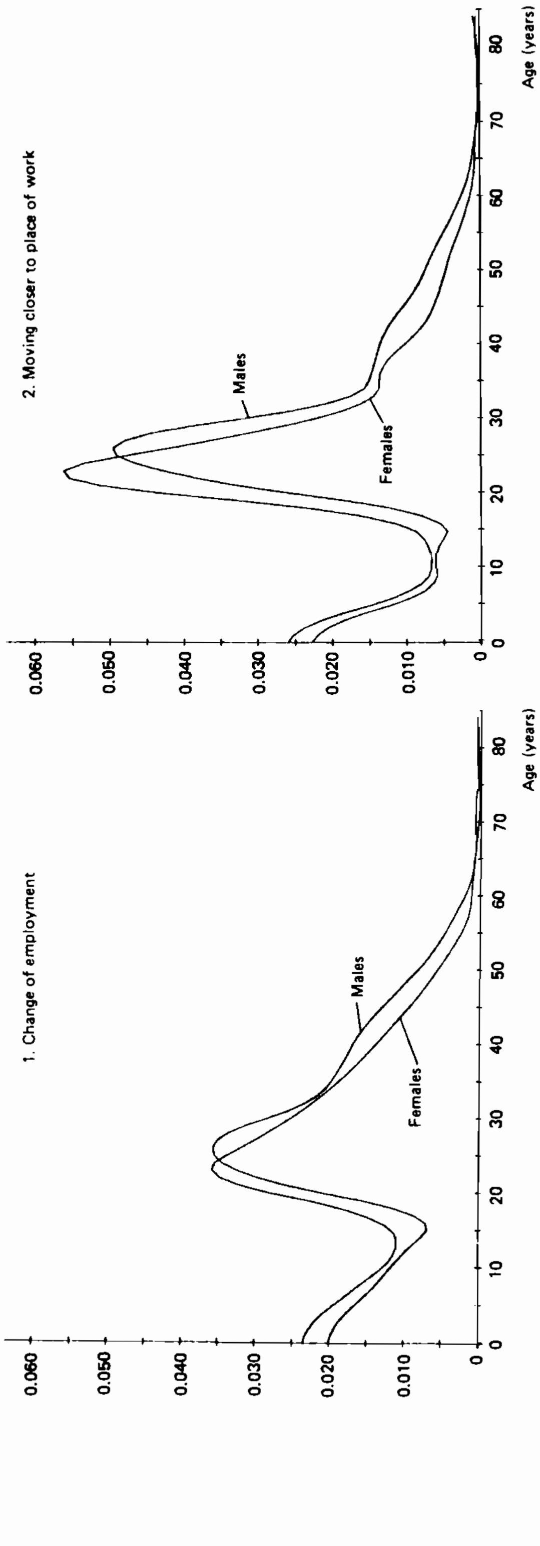



FIGURE 2 Observed cause-specific migration rates: Czechoslovakia, males and females, 1973.

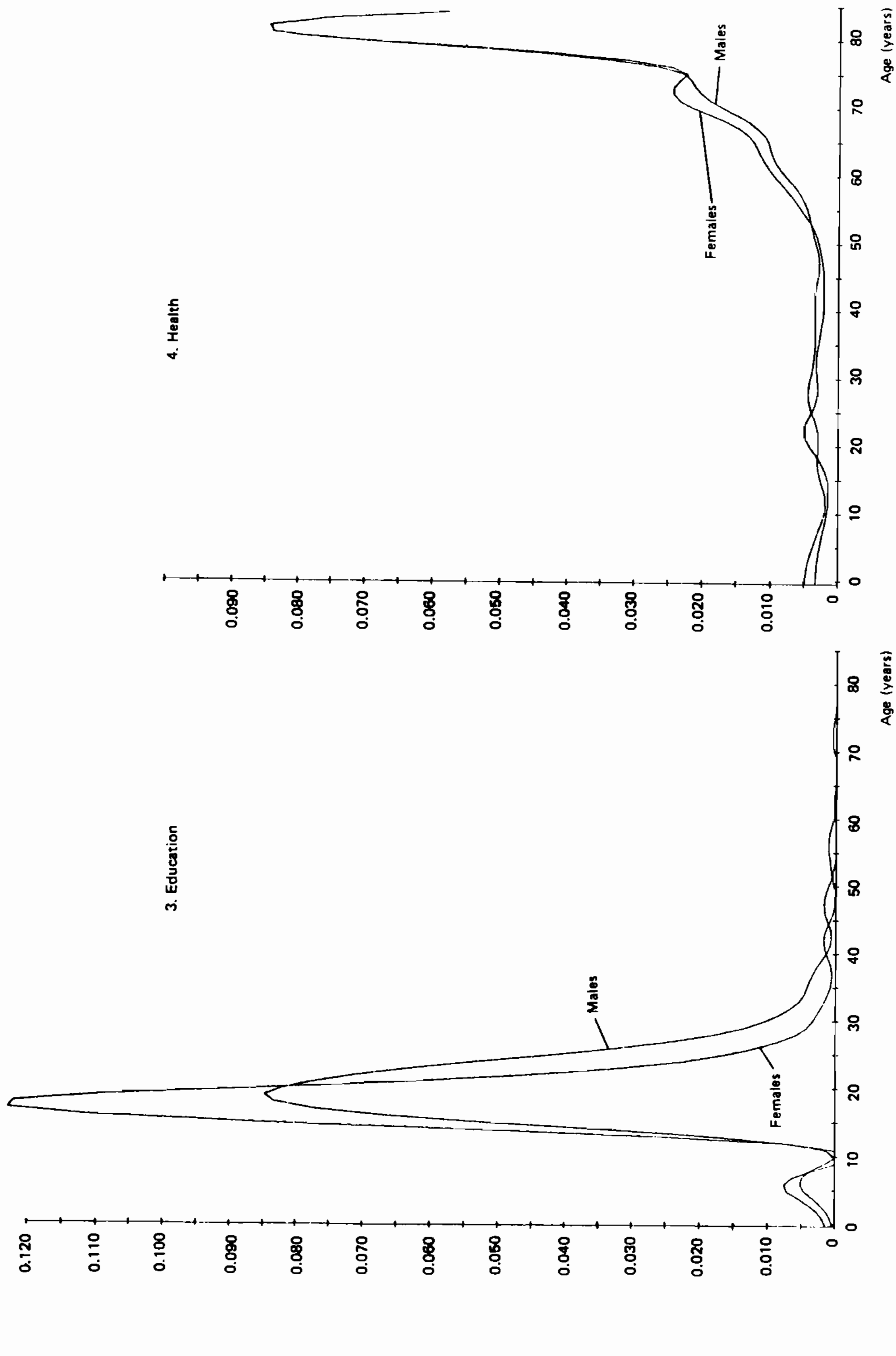



FIGURE 2 *(continued)*

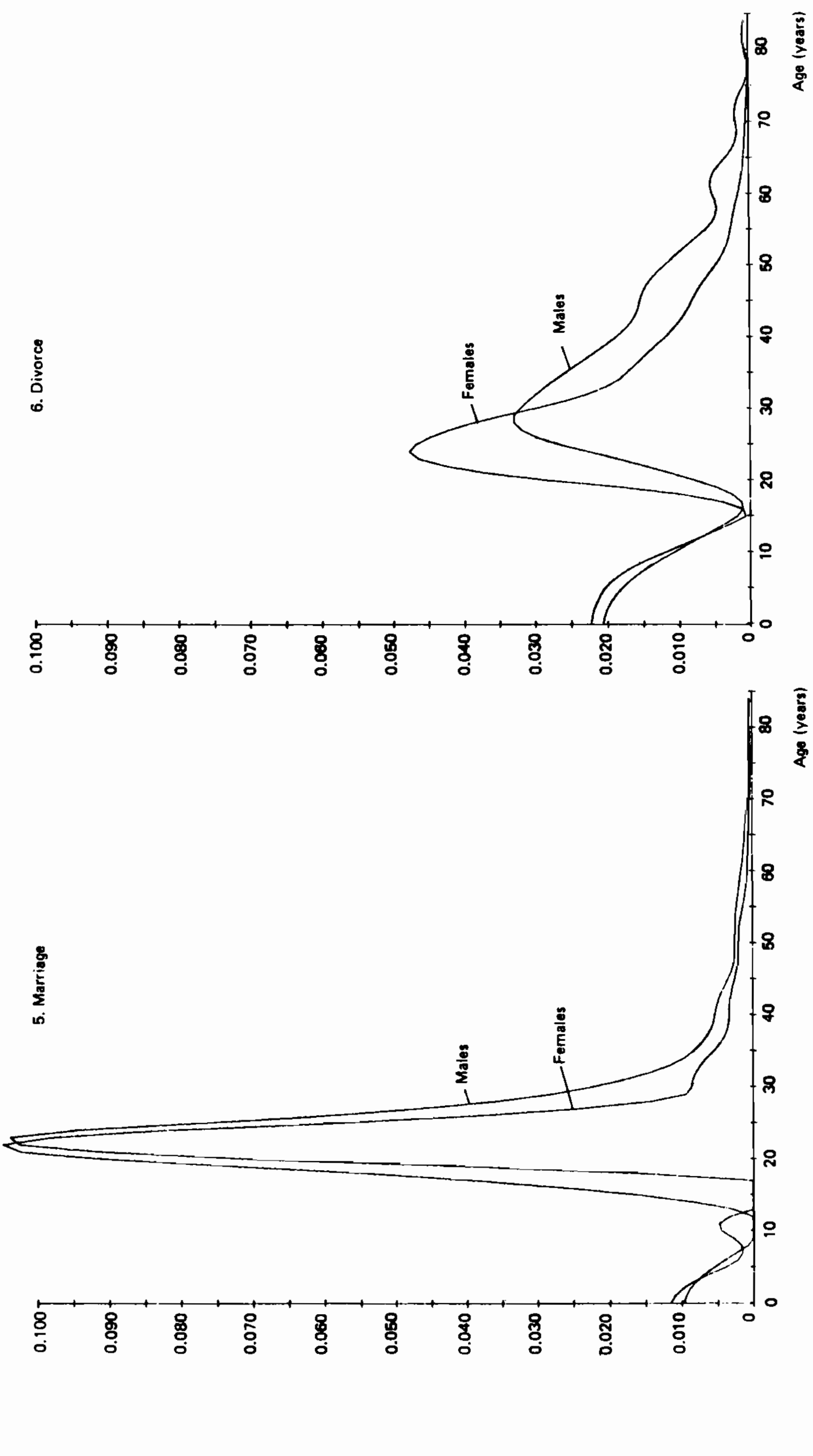



FIGURE 2 *(continued)*

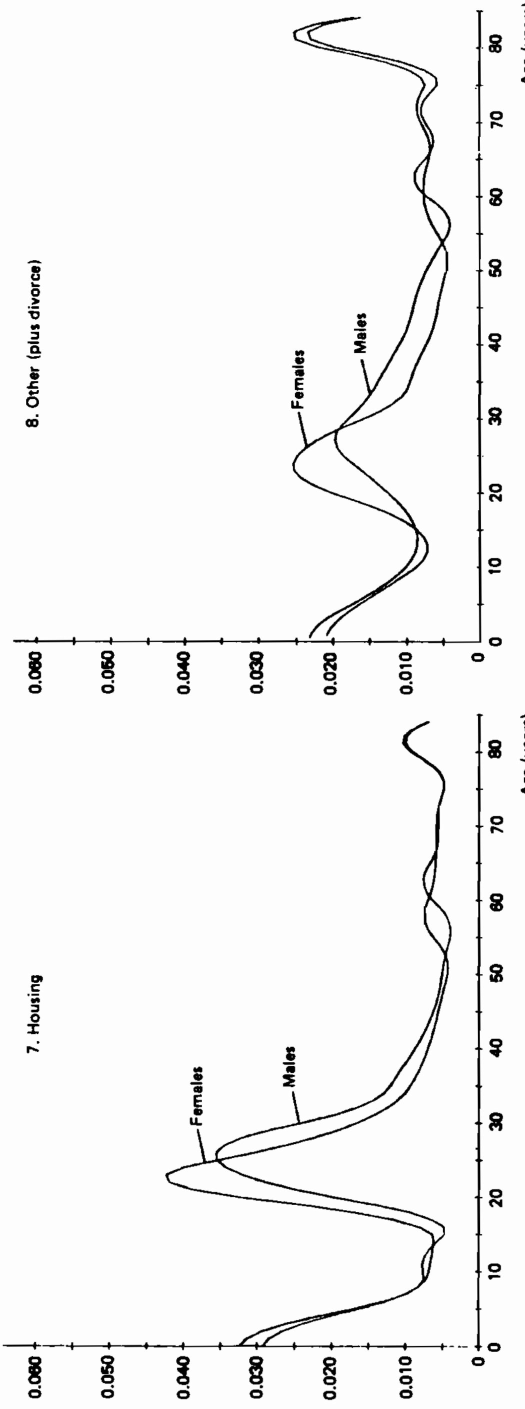



FIGURE 2 *(continued)*

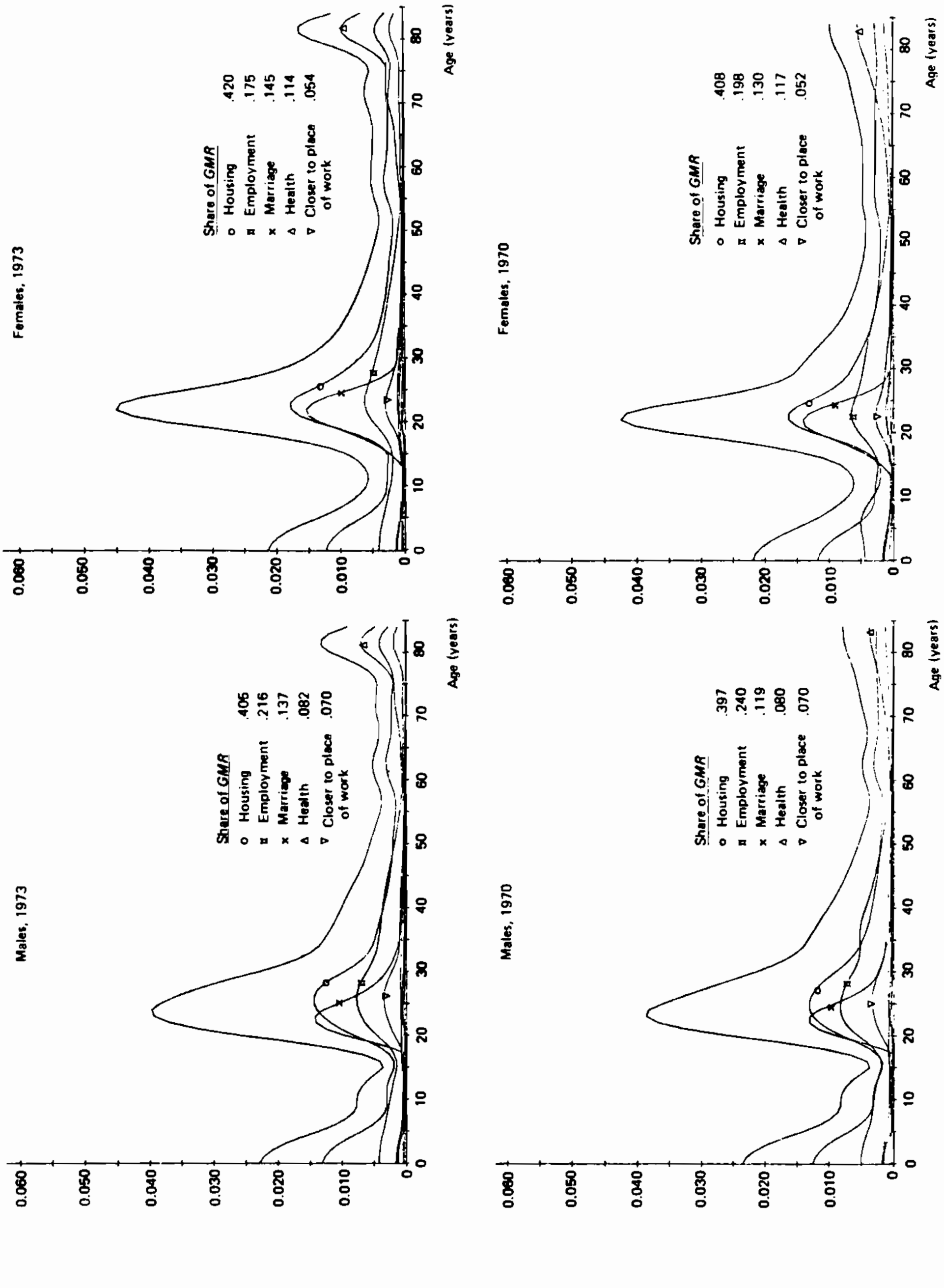



FIGURE 3 Observed cause-specific migration rates: Czechoslovakia, males and females, 1973 and 1970.

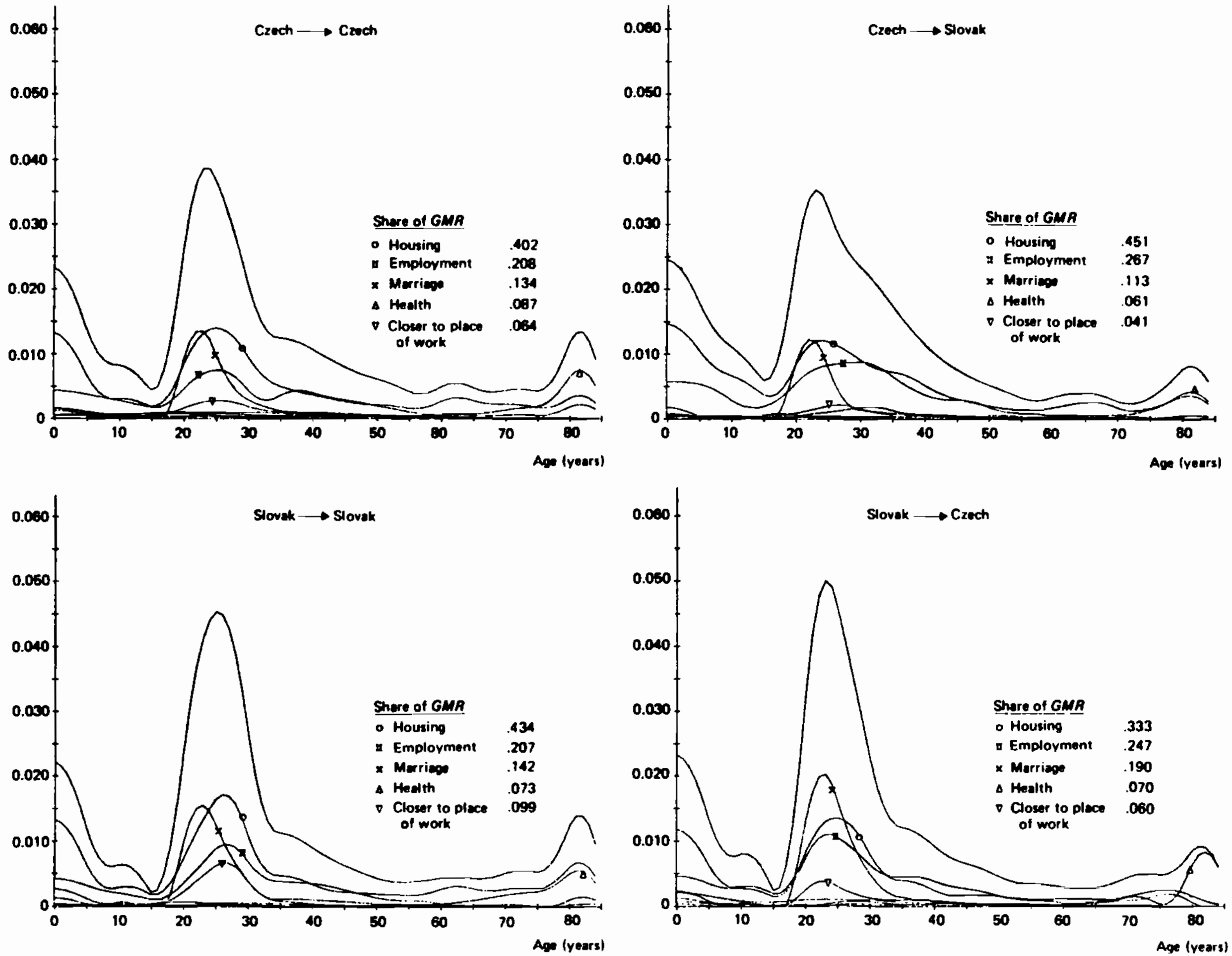



FIGURE 4 Observed cause-specific migration rates: intra- and inter-republic migrations in Czechoslovakia, males only, 1973.

45% of the total *GMR* in the flow from the Czech to the Slovak republic, but account for only 33% of the flow in the reverse direction. The age profiles also seem to differ, but a more accurate assessment requires that the profiles be scaled to a unit *GMR*, as in Figure 5.

Figure 5 shows the (scaled) age profiles of the following three causes of migration:

- change of employment
- moving closer to place of work
- housing

Differences in age profiles that are associated with differences in the distances migrated are hard to identify. Indeed, more significant differences seem to be associated with the republic of origin than with the category (intra- or inter-republic) of flow. For example, the age profiles of migration originating in Slovakia are more *labor dominant*, i.e., a higher proportion of the total *GMR* is associated with the young labor-force ages. Such labor-dominant curves are characterized by a relatively narrow, high peak between the ages of 20 and 30.

We conclude, therefore, that although the migration levels associated with various causes vary with distance migrated, the cause-specific age profiles may in fact be quite similar. However, only when more data of the sort used above become available will it be possible to clarify this matter: at present, our conclusions are merely conjectural.

## 4 ANALYSIS: MODEL SCHEDULES

Figure 6 illustrates both the *observed* age-specific migration schedules (histograms, scaled to unit *GMR*) for Czechoslovakian males and females in 1973 and their representation by a *model* schedule (the superimposed smooth curves) defined as the sum of four components:

- A single negative exponential curve for the *pre-labor-force* ages, with a rate of descent $\alpha_1$.
- A unimodal curve for the *labor-force* ages, with rates of ascent and descent $\lambda_2$ and $\alpha_2$, respectively.
- Another unimodal curve for the *post-labor-force* ages, with rates of ascent and descent $\lambda_3$ and $\alpha_3$, respectively.
- A constant curve $c$, the inclusion of which improves the quality of fit provided by the mathematical expression of the schedule.

The decomposition described above suggests the following simple sum of four curves (Rogers *et al.* 1978)*:

$$\begin{aligned} M(x) = {} & a_1 e^{-\alpha_1 x} \\ & + a_2 e^{-\alpha_2(x-\mu_2) - e^{-\lambda_2 (x-\mu_2)}} \\ & + a_3 e^{-\alpha_3(x-\mu_3) - e^{-\lambda_3(x-\mu_3)}} \\ & + c \end{aligned} \qquad (x = 0, 1, 2, \ldots, z) \qquad (1)$$

* Both the labor-force and the post-labor-force components in eqn. (1) are described by the "double-exponential" curve formulated by Coale and McNeil (1972) for their studies of nuptiality and fertility.

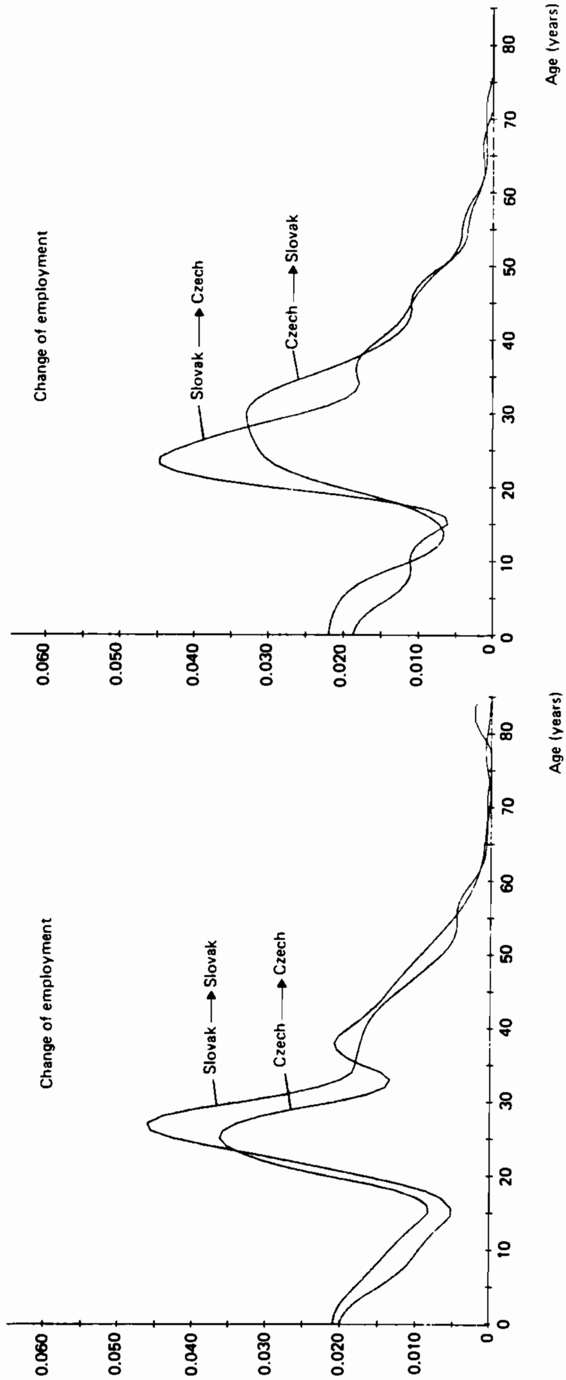
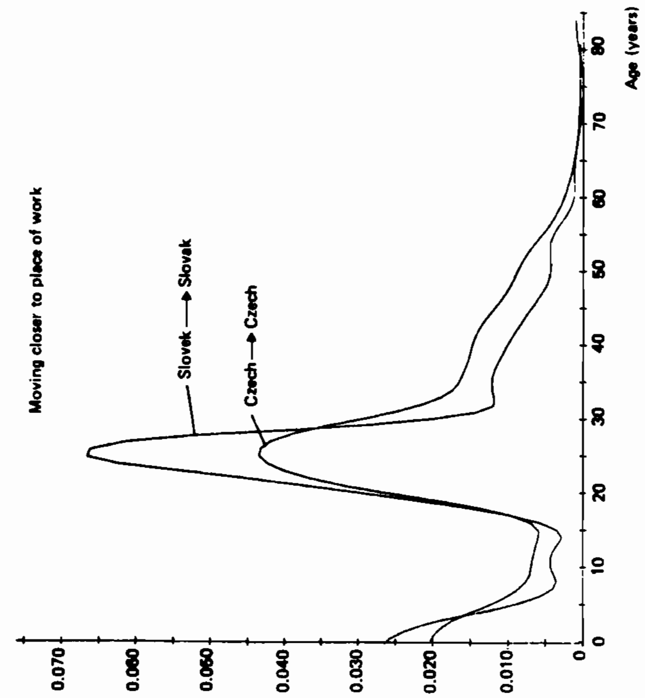



FIGURE 5 Observed cause-specific migration rates: intra- and inter-republic migrations in Czechoslovakia, males only, 1973, specific causes.



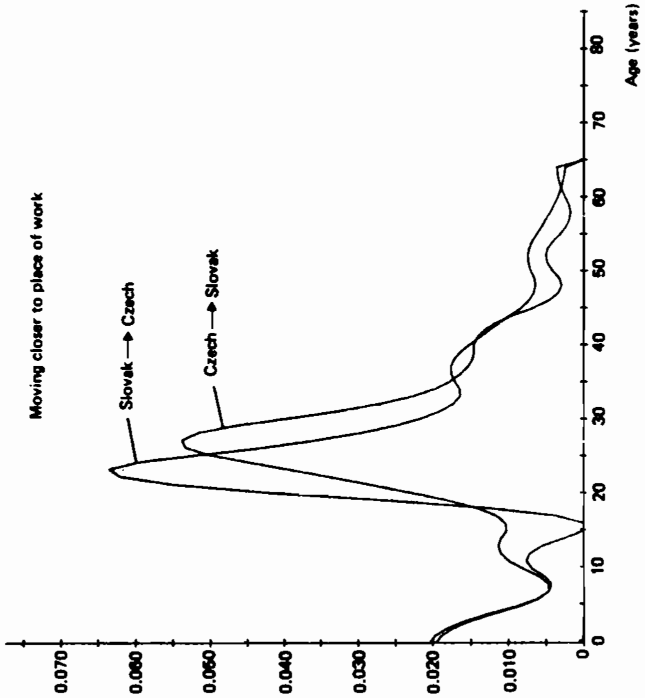




FIGURE 5 *(continued)*

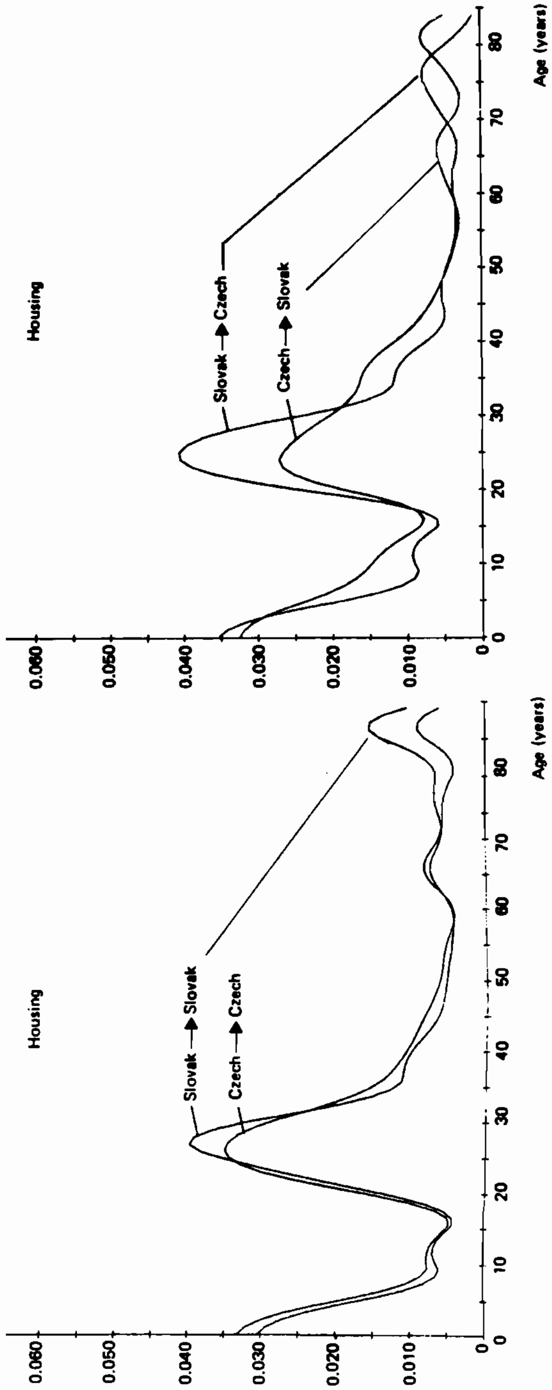



FIGURE 5 *(continued)*

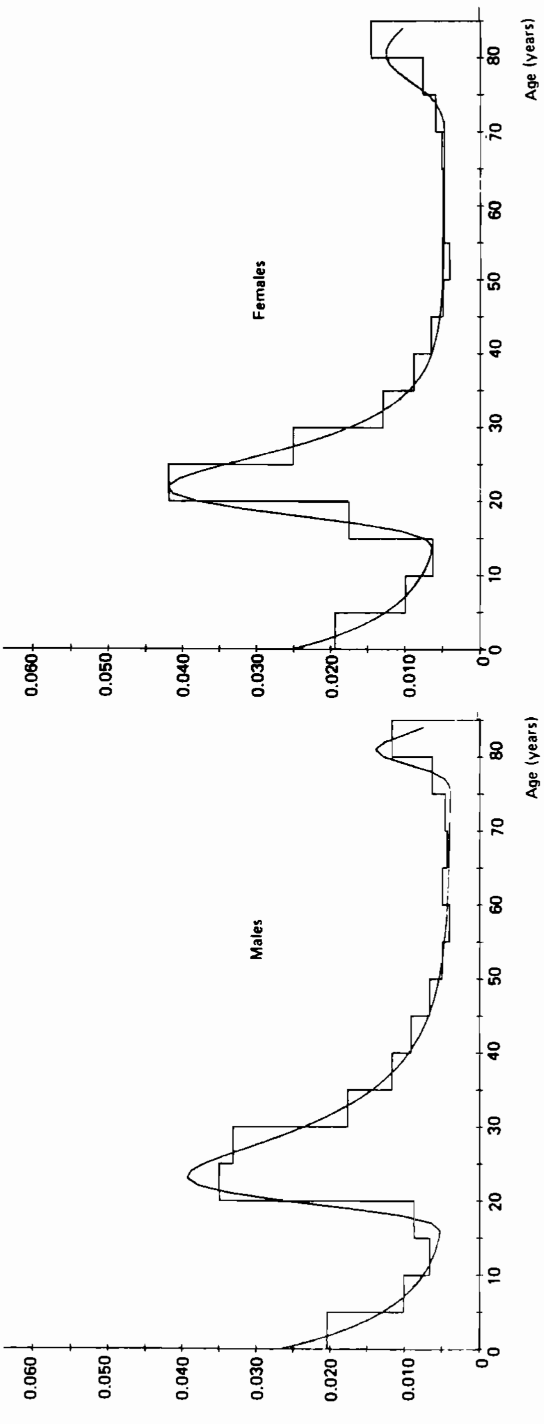



FIGURE 6 Model schedules of observed migration rates for all causes combined: Czechoslovakia, males and females, 1973.

The "full" model schedule in eqn. (1) has eleven parameters: $a_1$, $\alpha_1$, $a_2$, $\mu_2$, $\alpha_2$, $\lambda_2$, $a_3$, $\mu_3$, $\alpha_3$, $\lambda_3$, and $c$. The *profile* of the full model schedule is defined by seven of the eleven parameters: $\alpha_1$, $\mu_2$, $\alpha_2$, $\lambda_2$, $\mu_3$, $\alpha_3$, and $\lambda_3$. Its *level* is determined by the remaining four parameters: $a_1$, $a_2$, $a_3$, and $c$. A change in the value of the gross migraproduction rate of a particular model schedule alters proportionally the values of the level but does not affect the profile. Finally, migration schedules without a retirement or old-age peak may be represented by a "reduced" model with seven parameters, because in such instances the third component of eqn. (1) is omitted.

The model schedule defined in eqn. (1) may be used to fit all of the cause-specific profiles illustrated in Figures 2 and 5. The two profiles concerned with change of employment and moving closer to place of work and the profiles of migration associated with marriage and with divorce may be described by the reduced, seven-parameter model. Education-motivated migration profiles follow the model schedule with both the first and the third components omitted (i.e., $a_1 = a_3 = 0$). The age pattern of health-related migration can be described by the model schedule with both the first and the second components omitted (i.e., $a_1 = a_2 = 0$). Finally, migration caused by housing reasons and by the remaining "all other causes" (including divorce) takes on the profile of the full, eleven-parameter model, as does the aggregate schedule. Figures 7 and 8 display the results of such fits, and Appendixes A and B set out some of the numerical estimates of the various parameters and variables that define these model migration schedules. More detailed numerical outputs, together with their interpretation and methods of derivation, are described in Rogers and Castro (1979) and Castro and Rogers (1979), respectively.

The model-schedule profiles displayed in Figures 6, 7, and 8 repeat the patterns exhibited by the cubic-spline interpolations of the same data in Figures 1, 2, and 5, respectively. The principal advantage of the model-schedule representation is that the mathematical description is particularly well-suited for analytical studies of the properties of each cause-specific curve and of the impacts on the aggregate migration-age pattern of changes in each curve's relative importance.

The Czechoslovakian age patterns of internal migration disaggregated by cause may be characterized in a number of different ways, as described by Rogers and Castro (1979). We begin by observing that, among the seven causes examined, only health, housing, and the category "other reasons" exhibit a retirement (in fact a post-retirement) peak, with the one for housing occurring more than a dozen years earlier than the peak representing migration for health reasons. We also note that the *low point* and the *high peak*, $x_l$ and $x_h$, respectively, occur earlier in female profiles than in male profiles, in all cases for which these measures have been calculated. This is undoubtedly a reflection of the differences in ages at marriage.

The age profiles of reduced-form model migration schedules (i.e., those with no retirement peak and only seven parameters) are determined by the four parameters $\alpha_1$, $\mu_2$, $\alpha_2$, $\lambda_2$, and the ratio $\delta_{12} = a_1/a_2$. The parameters $\alpha_1$ and $\alpha_2$ define the rates of decrease with age of the migration rates of children and their parents, respectively; $\mu_2$ locates the labor-force curve on the age axis, while $\lambda_2$ defines the rate of increase of the migration rates of those young adults who leave the family home primarily for marital, educational, or economic reasons. Finally, the ratio $\delta_{12}$ relates the height of the pre-labor-force curve to that of the labor-force component.

The model schedules illustrated in Figures 6–8 exhibit a wide range of values for these parameters of interest. The curve with the highest peak, that for marriage,

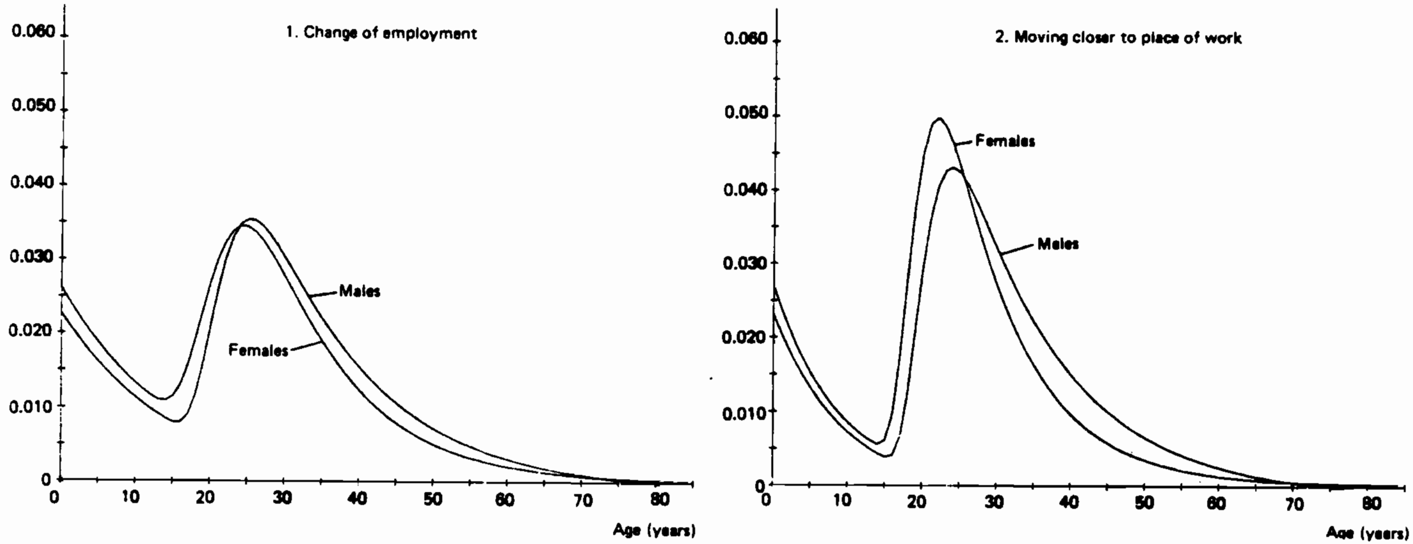



FIGURE 7 Model schedules of observed cause-specific migration rates: Czechoslovakia, males and females, 1973.

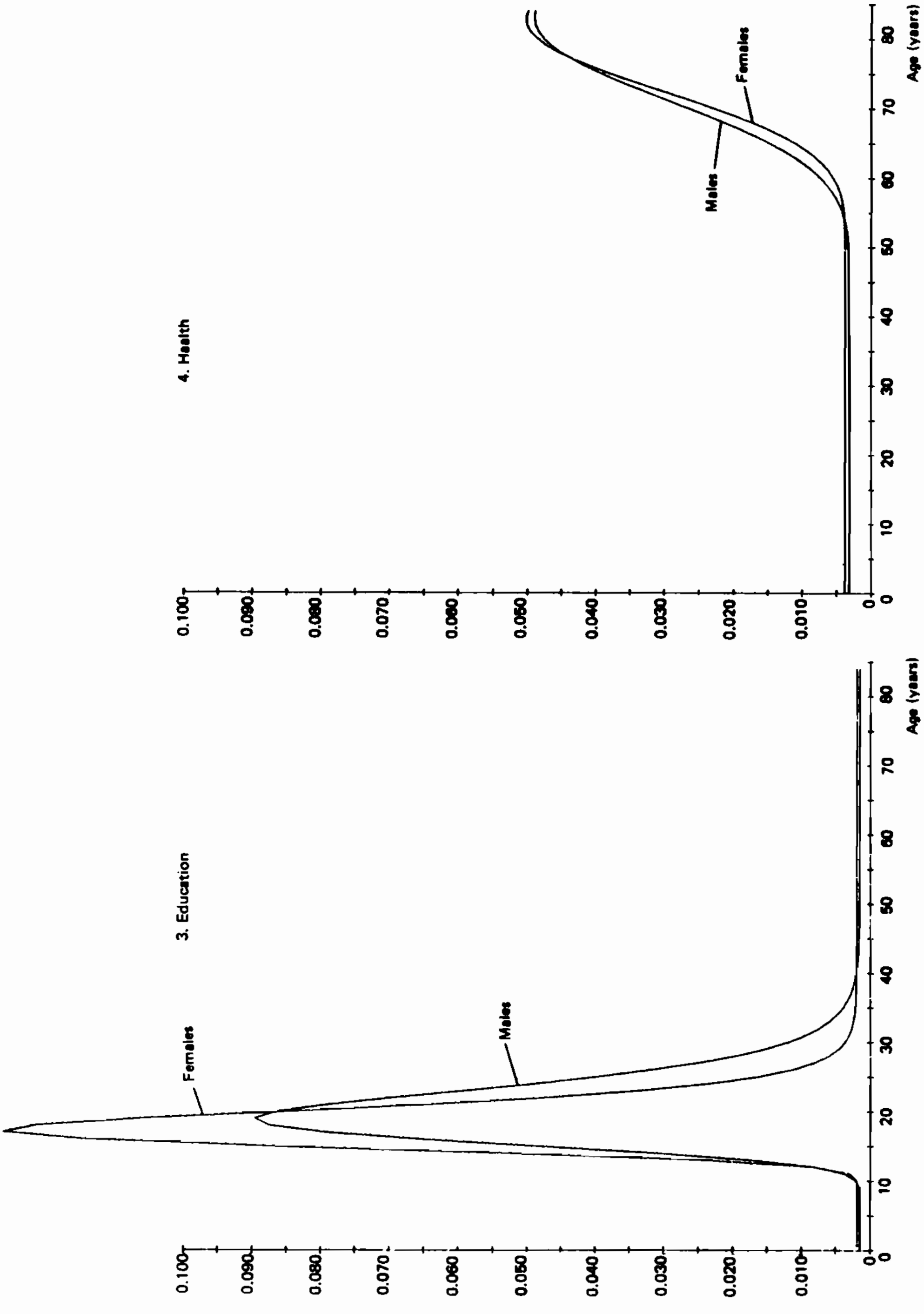



FIGURE 7 (*continued*)

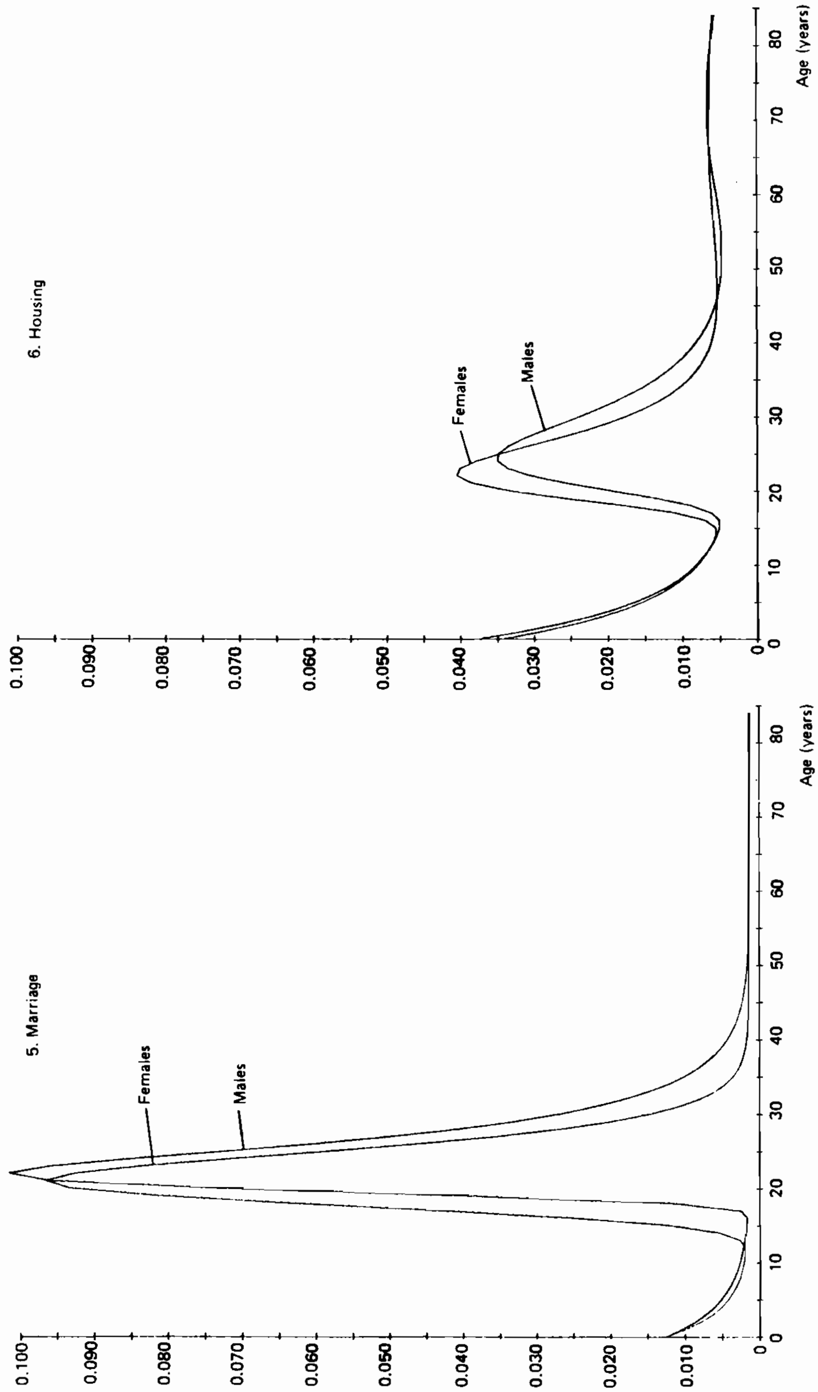



FIGURE 7 *(continued)*

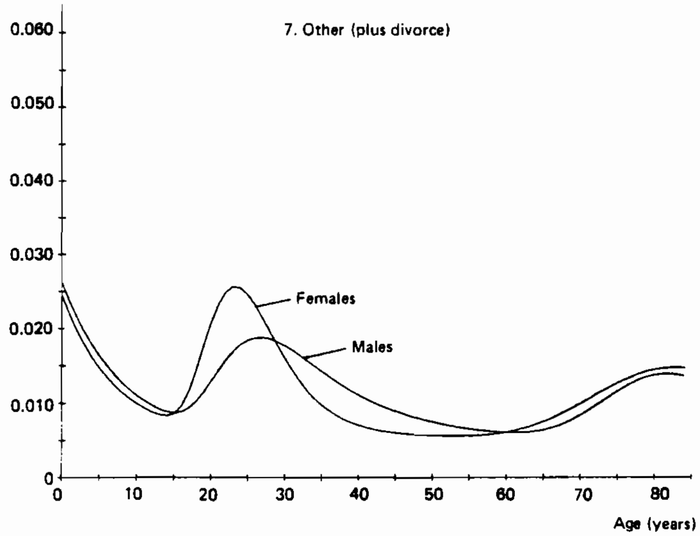



FIGURE 7 (*continued*)

may be characterized by its relatively low value for $\delta_{12}$, a value that is about one-fifth as large as those of the other profiles. This profile, with a relatively low $\delta_{12}$ value, may be termed *labor dominant*. In contrast, age patterns with much higher values of $\delta_{12}$ tend to show relatively higher migration rates for children, and may thus be said to be *child dependent*.

*Labor dominance* reflects the dominance of the migration levels of those in the working ages relative to those of children and pensioners. *Labor asymmetry* refers to the shape of the central bell-shaped curve and is measured by the ratio $\sigma_2 = \lambda_2/\alpha_2$. The numerical values for $\sigma_2$ given in the Appendixes and in Rogers and Castro (1979) indicate that the national profile for moving closer to place of work exhibits the most asymmetrical pattern of all the causes considered.

## 5 SYNTHESIS: SENSITIVITY EXPERIMENTS

The two preceding sections have been devoted to a description and an analysis of cause-specific age profiles of migration. We now turn to an examination of cause-deleted age profiles, focusing in particular on the impact that the deletion of a particular cause has on the remaining age pattern of aggregate migration.

Figure 9 illustrates four aggregate model-schedule age profiles*; a total of seven such profiles have been numerically defined in Appendixes A and B, and in Rogers and Castro (1979). In each case, the aggregate age profile with the contribution from a single cause deleted is compared with the profile in which the share of that single cause is increased to five times its observed level. The resulting contrasts clearly identify the contribution of each cause of migration to the aggregate age profile.

* The area under each curve is once again scaled to a *GMR* of unity.

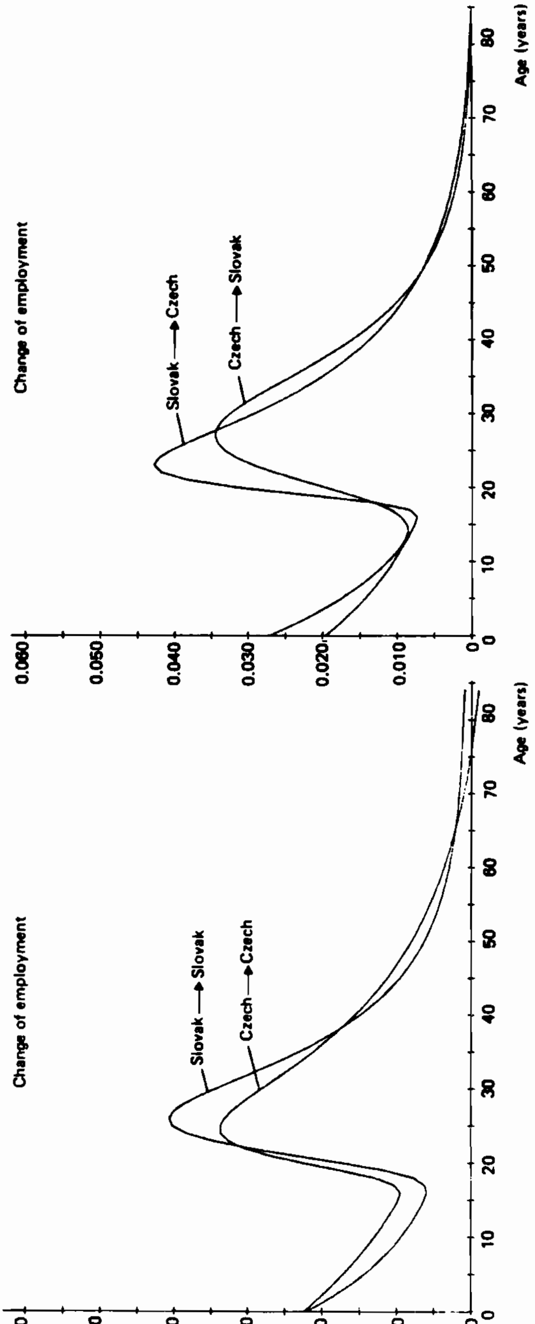



FIGURE 8 Model schedules of observed cause-specific migration rates: intra- and inter-republic migrations in Czechoslovakia, males only, 1973, specific causes.

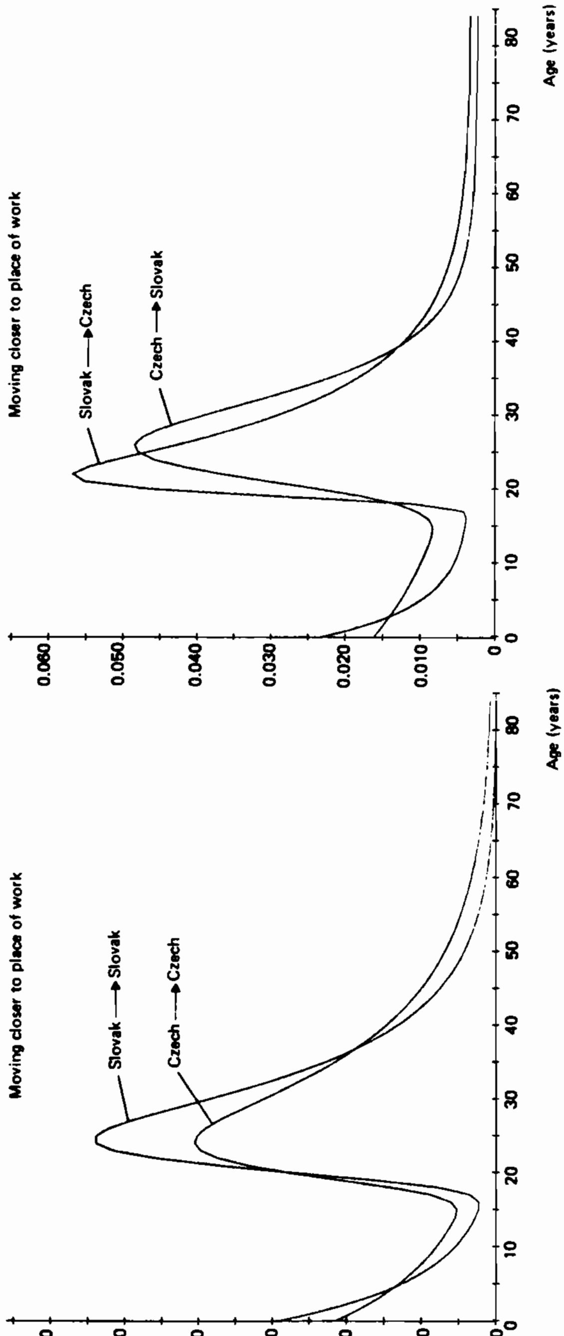



FIGURE 8 (*continued*)

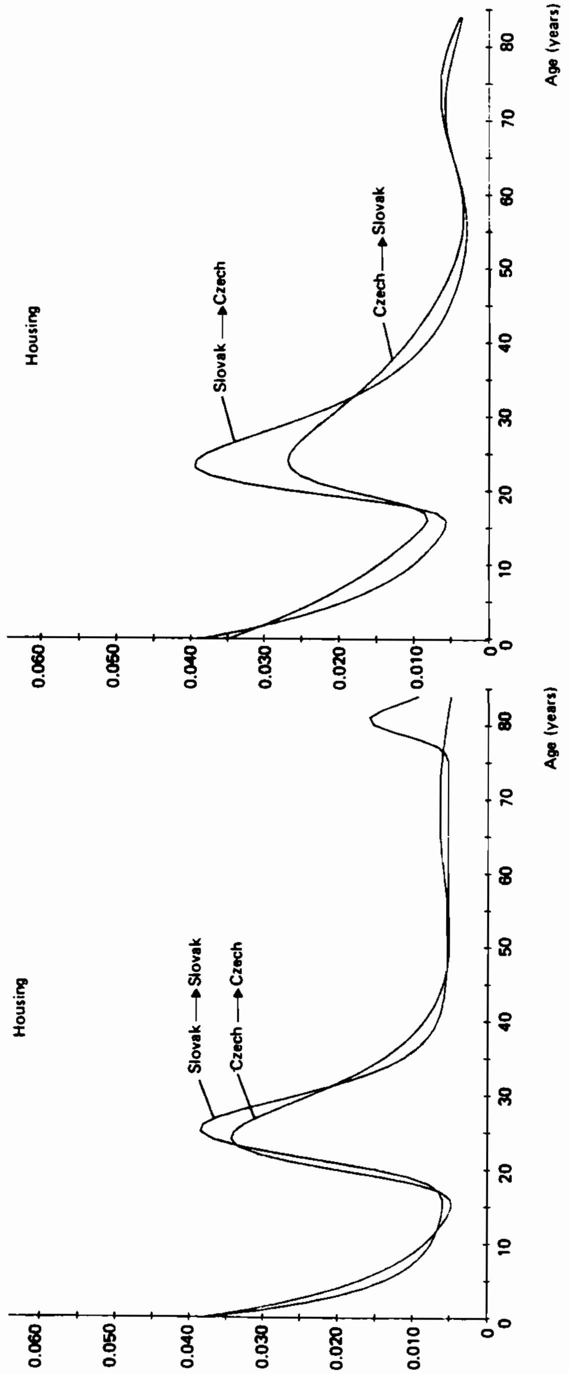



FIGURE 8 *(continued)*

Deleting change of employment as a cause of migration from the aggregate migration curve results in an increase in the mean age of migration because of the artificially increased relative importance of migration in the post-labor-force age groups. Increasing the relative importance of this cause of migration five-fold lowers the rates of descent of the migration rates of young adults, $\alpha_2$, and of their children, $\alpha_1$.

Health becomes an important reason for migrating only in the post-retirement age groups. Thus, deleting this cause removes the old-age peak; increasing the importance of this contribution five-fold simply increases the height of the peak. The slope parameters $\alpha_1$, $\alpha_2$, and $\lambda_2$ and the location parameter $\mu_2$ remain unchanged.

A five-fold increase in the relative importance of marriage as a cause for migration dramatically increases the maximum value of the young adult peak. The rate of ascent $\lambda_2$ increases, but the rates of descent $\alpha_1$ and $\alpha_2$ are only slightly affected. The curve then becomes a member of the labor-dominant family, as analyzed by Castro and Rogers (1979). Deleting the contribution of marriage leaves the aggregate curve virtually unchanged.

Finally, the desire for different housing influences the aggregate migration profile at very early ages and at retirement ages. Deleting this contribution reduces the rate of descent of pre-labor-force migration, $\alpha_1$, and increases the importance of old-age migration. The mean age of migration remains virtually unchanged.

Figure 9 shows that changes in the relative importance of different causes of migration produce predictable changes in the age profile of aggregate migration. Employment influences the rates of descent $\alpha_1$ and $\alpha_2$. Health reasons affect only the post-retirement profile, influencing $\lambda_3$ and the position of the post-retirement peak, $x_r$. Marriage dramatically affects the rate of ascent $\lambda_2$, but leaves $\alpha_1$ and $\alpha_2$ relatively unchanged. Housing, on the other hand, influences $\alpha_1$ and $\alpha_2$, but particularly $\alpha_1$, and changes the pattern of old-age migration.

## 6 CONCLUSION

This paper was motivated by the conjecture that regularities in the age patterns of migration of different national populations are likely to be stronger and more evident in cause-specific schedules than in aggregate schedules. The implications of this for model migration schedules should be the same as for model mortality schedules, in the context of which Preston (1976, p. 118) observed (italics added for emphasis):

> Model mortality patterns are typically required for demographic estimation only if death registration is incomplete. But if the degree of incompleteness is largely invariant with respect to cause of death, then the *cause-structure* of mortality can be reliably estimated. In such a case the *age pattern* of mortality should be largely recoverable without reference to any external models. The *level* of mortality can then be estimated through conventional stable population or census survival techniques.

Substituting the word "migration" for "mortality" and "death" in the above quotation yields the observation that the age pattern (i.e., profile) of migration may be estimated by weighting each cause-specific profile by the migration structure (i.e.,

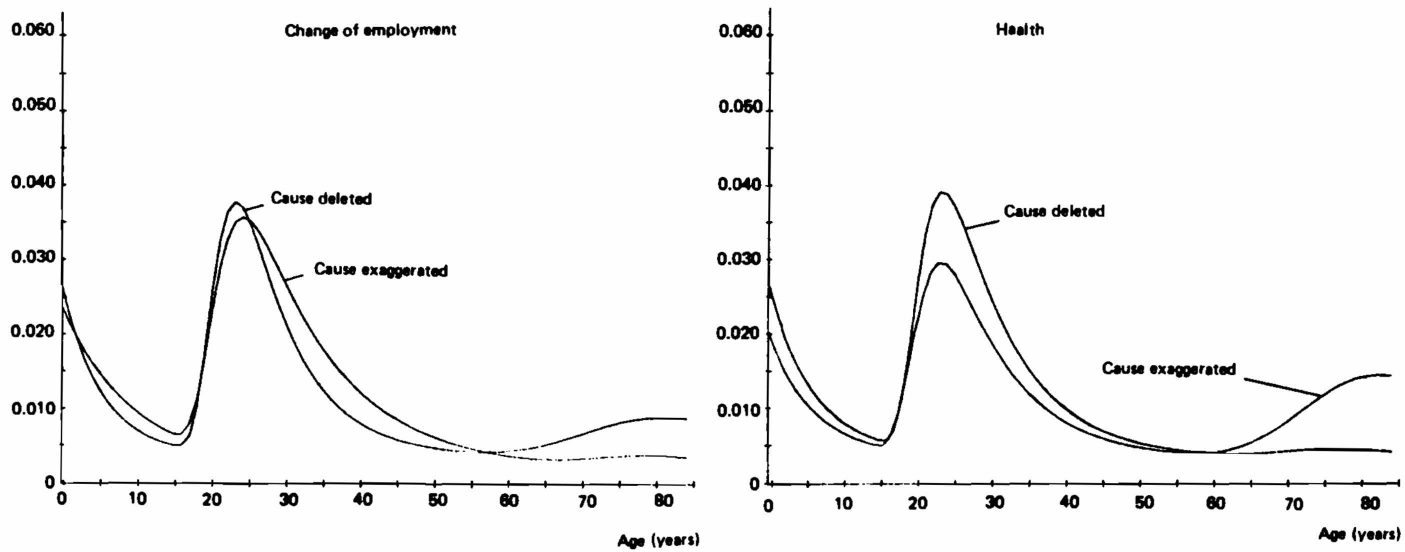



FIGURE 9 Model schedules of cause-deleted and cause-exaggerated aggregate migration rates.

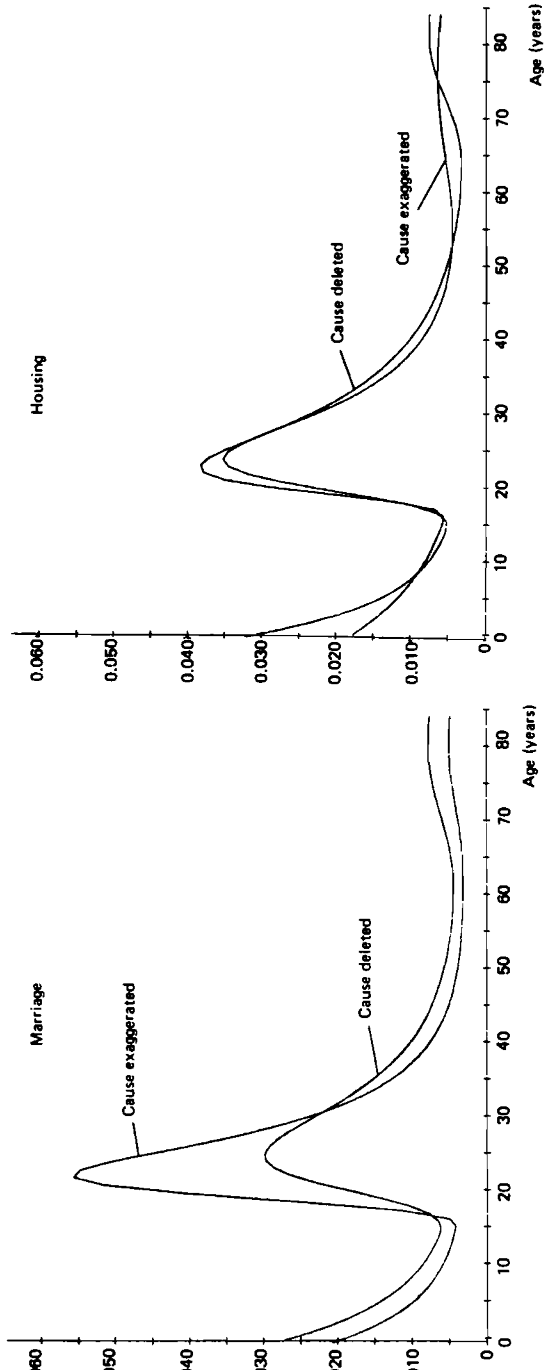



FIGURE 9 *(continued)*

the fraction of the total *GMR* that is attributable to each of $k$ causes)

$$\bar{M}(x) = \sum_{i=1}^{k} \overline{GMR}^{(i)} \bar{M}^{(i)}(x) \qquad (x = 0, 1, 2, \ldots, z) \tag{2}$$

where

$$\sum_{x}^{z} \bar{M}(x) = 1$$

$$\sum_{i}^{k} \overline{GMR}^{(i)} = 1$$

$$\sum_{i}^{k} \bar{M}^{(i)}(x) = 1$$

Equation (2) may also be expressed in the form of migration proportions

$$N(x) = \sum_{i=1}^{k} N^{(i)} N^{(i)}(x) \tag{3}$$

where $N^{(i)}(x)$ denotes the proportion of migrants at age $x$ among those citing cause $i$, and $N^{(i)}$ is the proportion of all migrants who cite cause $i$.

The estimation problem can also be "turned on its head" as Preston (1976, p. 116) observes:

> Just as the cause-structure of mortality can be used to predict the age pattern, the age pattern implies a special cause of death structure.

By analogy, if we apply these comments to migration this suggests that given, for example, the Hungarian aggregate age pattern of migration and the Czechoslovakian cause-specific age patterns of migration (or migration profiles) one could use eqns. (2) or (3) to develop estimates of the implied Hungarian cause-of-migration structure and compare it with the structure described earlier in this paper.

## ACKNOWLEDGMENTS

The data used in this paper were collected as part of a comparative migration and settlement study conducted at the International Institute for Applied Systems Analysis (IIASA). The authors gratefully acknowledge the generous provision of detailed Czechoslovakian data by Karel Kühnl of the Department of Geography at Charles University in Prague. We are also grateful for the opportunity to discuss the subject of this paper with Nathan Keyfitz, who called our attention to Preston's "inverse" use of the mortality cause–structure relationship.

## REFERENCES

Castro, L. J., and A. Rogers (1979) Migration Age Patterns: I. Measurement and Analysis. WP-79-16. Laxenburg, Austria: International Institute for Applied Systems Analysis.

Coale, A. J., and D. R. McNeil (1972) The distribution by age of the frequency of first marriage in a female cohort. Journal of the American Statistical Association 67:743–749.

Compton, P. A. (1971) Some Aspects of the Internal Migration of Population in Hungary Since 1957. Budapest: Central Statistical Office, Demographic Research Institute.

Czechoslovakian Federal Statistical Office (1974) Demographic Trends in Czechoslovakia. Prague.

Hoover, E. M. (1971) An Introduction to Regional Economics. New York: A. A. Knopf.

Kühnl, K. (1978) Selected aspects of migration motivation in the Czech Socialist Republic. Acta Universitatis Carolinae, Geographica 13(1):3–11.

Long, L. H., and K. A. Hansen (1979) Reasons for Interstate Migration: Jobs, Retirement, Climate, and Other Influences. Current Population Reports, Series P-23, No. 81. Washington, D.C.: Bureau of the Census, U.S. Department of Commerce.

McNeil, D. R., T. J. Trussell, and J. C. Turner (1977) Spline interpolation of demographic data. Demography 14:245–252.

Preston, S. (1976) Mortality Patterns in National Populations. New York: Academic Press.

Rogers, A., and L. J. Castro (1979) Migration Age Patterns: II. Cause-specific Profiles. WP-79-65. Laxenburg, Austria: International Institute for Applied Systems Analysis.

Rogers, A., R. Raquillet, and L. J. Castro (1978) Model migration schedules and their applications. Environment and Planning A 10(5):475–502.

Young, C. M. (1975) Factors associated with the timing and duration of the leaving-home stage of the family life cycle. Population Studies 29(1):61–73.

## APPENDIXES

The parameters $a_1$, $a_2$, $a_3$, $\alpha_1$, $\alpha_2$, $\alpha_3$, etc., used in the following tables are defined in the text. The columns 1–8 represent causes of migration, as follows:

1 Change of employment
2 Moving closer to place of work
3 Education
4 Health
5 Marriage
6 Housing
7 Other causes
8 All causes

## APPENDIX A

Numerical estimates of various parameters and variables that define the model migration schedule: males, 1973.

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| gmr (obs) | 0.22832 | 0.07414 | 0.00973 | 0.08687 | 0.14475 | 0.42813 | 0.08501 | 1.05694 |
| gmr (mms) | 0.99706 | 0.99759 | 1.05827 | 1.36446 | 1.05302 | 1.04005 | 1.10948 | 1.02932 |
| mae%m | 8.28649 | 16.13420 | 12.62849 | 46.06667 | 13.52592 | 12.37815 | 16.78544 | 8.57164 |
| a1 | 0.02461 | 0.02570 | 0.00000 | 0.00000 | 0.01258 | 0.03812 | 0.02265 | 0.02489 |
| alfa1 | 0.06397 | 0.10403 | 0.00000 | 0.00000 | 0.21349 | 0.19165 | 0.12484 | 0.18486 |
| a2 | 0.05519 | 0.06871 | 0.21037 | 0.00000 | 0.21069 | 0.07477 | 0.03021 | 0.06803 |
| mu2 | 21.35782 | 20.22688 | 21.50009 | 0.00000 | 20.65884 | 22.66046 | 23.78325 | 20.88616 |
| alfa2 | 0.06853 | 0.07191 | 0.31682 | 0.00000 | 0.21405 | 0.15098 | 0.09431 | 0.12812 |
| lambda2 | 0.32095 | 0.41243 | 0.20419 | 0.00000 | 0.57113 | 0.29236 | 0.19419 | 0.42862 |
| a3 | 0.00000 | 0.00000 | 0.00000 | 0.00038 | 0.00000 | 0.00031 | 0.00024 | 0.00005 |
| mu3 | 0.00000 | 0.00000 | 0.00000 | 129.25330 | 0.00000 | 116.40599 | 116.25908 | 90.44624 |
| alfa3 | 0.00000 | 0.00000 | 0.00000 | 0.23520 | 0.00000 | 0.16319 | 0.26248 | 1.28572 |
| lambda3 | 0.00000 | 0.00000 | 0.00000 | 0.03902 | 0.00000 | 0.03521 | 0.04909 | 0.20373 |
| c | -0.00131 | -0.00143 | 0.00143 | 0.00366 | 0.00133 | 0.00311 | 0.00490 | 0.00394 |
| mean age | 26.92261 | 26.69720 | 24.64404 | 72.80155 | 26.81527 | 34.46154 | 42.99207 | 35.15611 |
| %( 0-14) | 21.92932 | 17.50977 | 7.34002 | 4.10135 | 7.14720 | 22.40590 | 20.56418 | 17.88818 |
| %(15-64) | 77.76012 | 83.16303 | 88.41094 | 15.26308 | 89.27063 | 60.47754 | 50.00044 | 65.75255 |
| %(65+ ) | 0.31055 | -0.67281 | 4.24904 | 80.63557 | 3.58218 | 17.11657 | 29.43537 | 16.35927 |
| deltalc | -18.72077 | -17.98383 | 0.00000 | 0.00000 | 9.49213 | 12.24543 | 4.62645 | 6.31150 |
| delta12 | 0.44587 | 0.37401 | 0.00000 | 0.00000 | 0.05972 | 0.50985 | 0.74968 | 0.36582 |
| delta32 | 0.00000 | 0.00000 | 0.00000 | 0.00000 | 0.00000 | 0.00419 | 0.00780 | 0.00073 |
| beta12 | 0.93339 | 1.44665 | 0.00000 | 0.00000 | 0.99737 | 1.26936 | 1.32373 | 1.44285 |
| sigma2 | 4.68306 | 5.73504 | 0.64452 | 0.00000 | 2.66813 | 1.93644 | 2.05913 | 3.34538 |
| sigma3 | 0.00000 | 0.00000 | 0.00000 | 0.16588 | 0.00000 | 0.21577 | 0.18703 | 0.15845 |
| x low | 16.07025 | 15.85025 | 5.34001 | 0.00000 | 16.55026 | 15.99025 | 15.57024 | 16.23026 |
| x high | 25.77048 | 24.31044 | 19.36033 | 0.00000 | 22.39040 | 24.89046 | 27.13051 | 23.68043 |
| x ret. | 0.00000 | 0.00000 | 0.00000 | 83.20175 | 0.00000 | 72.57948 | 82.01150 | 81.41137 |
| x shift | 9.70022 | 8.46019 | 14.02032 | 0.00000 | 5.84013 | 8.90020 | 11.56026 | 7.45017 |
| a | 34.33701 | 37.80360 | 0.00000 | 0.00000 | 33.88362 | 29.07706 | 29.08875 | 32.32034 |
| b | 0.02759 | 0.03944 | 0.08813 | 0.04653 | 0.09999 | 0.03007 | 0.01012 | 0.03416 |

Numerical estimates of various parameters and variables that define the model migration schedule: females, 1973.

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| gmr (obs) | 0.18916 | 0.05818 | 0.00679 | 0.12306 | 0.15853 | 0.45577 | 0.09250 | 1.08399 |
| gmr (mms) | 0.99914 | 1.00747 | 1.09491 | 1.39222 | 1.05011 | 1.04442 | 1.12564 | 1.05941 |
| mae%m | 7.07582 | 12.64702 | 14.05034 | 47.88653 | 20.66070 | 11.48562 | 18.16703 | 12.67905 |
| a1 | 0.02719 | 0.02848 | 0.00000 | 0.00000 | 0.01301 | 0.03388 | 0.02054 | 0.02230 |
| alfa1 | 0.06730 | 0.11599 | 0.00000 | 0.00000 | 0.27833 | 0.19080 | 0.14362 | 0.18986 |
| a2 | 0.06362 | 0.08654 | 0.28366 | 0.00000 | 0.17396 | 0.09271 | 0.05328 | 0.09538 |
| mu2 | 21.37095 | 19.17327 | 19.43375 | 0.00000 | 24.90277 | 21.39939 | 24.12472 | 21.03097 |
| alfa2 | 0.09251 | 0.10477 | 0.46279 | 0.00000 | 0.40363 | 0.20259 | 0.21418 | 0.20711 |
| lambda2 | 0.23451 | 0.41851 | 0.28093 | 0.00000 | 0.20020 | 0.31701 | 0.20494 | 0.30105 |
| a3 | 0.00000 | 0.00000 | 0.00000 | 0.00061 | 0.00000 | 0.00030 | 0.00028 | 0.00012 |
| mu3 | 0.00000 | 0.00000 | 0.00000 | 134.39360 | 0.00000 | 137.68983 | 127.29836 | 97.25676 |
| alfa3 | 0.00000 | 0.00000 | 0.00000 | 0.19788 | 0.00000 | 0.10295 | 0.19910 | 0.60703 |
| lambda3 | 0.00000 | 0.00000 | 0.00000 | 0.03436 | 0.00000 | 0.02262 | 0.03773 | 0.10699 |
| c | -0.00007 | 0.00022 | 0.00189 | 0.00305 | 0.00142 | 0.00342 | 0.00533 | 0.00463 |
| mean age | 25.03554 | 24.46350 | 23.72281 | 73.78278 | 24.92222 | 34.53135 | 44.09703 | 36.56535 |
| %( 0-14) | 25.58745 | 20.30622 | 10.04876 | 3.35996 | 6.76956 | 20.82524 | 18.45527 | 17.11870 |
| %(15-64) | 73.26627 | 78.60558 | 84.57401 | 14.60548 | 89.36448 | 61.69809 | 48.62447 | 62.20044 |
| %(65+ ) | 1.14629 | 1.08820 | 5.37724 | 82.03456 | 3.86596 | 17.47667 | 32.92026 | 20.68086 |
| deltalc | -366.79382 | 132.11542 | 0.00000 | 0.00000 | 9.13307 | 9.90647 | 3.85487 | 4.81541 |
| delta12 | 0.42738 | 0.32904 | 0.00000 | 0.00000 | 0.07476 | 0.36545 | 0.38556 | 0.23385 |
| delta32 | 0.00000 | 0.00000 | 0.00000 | 0.00000 | 0.00000 | 0.00328 | 0.00529 | 0.00126 |
| beta12 | 0.72750 | 1.10717 | 0.00000 | 0.00000 | 0.68956 | 0.94179 | 0.67055 | 0.91672 |
| sigma2 | 2.53505 | 3.99467 | 0.60704 | 0.00000 | 0.49599 | 1.56478 | 0.95686 | 1.45353 |
| sigma3 | 0.00000 | 0.00000 | 0.00000 | 0.17365 | 0.00000 | 0.21969 | 0.18952 | 0.17625 |
| x low | 14.26021 | 14.75022 | 7.70006 | 0.00000 | 12.01016 | 15.00023 | 14.44022 | 14.13021 |
| x high | 24.84045 | 22.38040 | 17.67029 | 0.00000 | 21.41038 | 22.79041 | 23.81043 | 22.26040 |
| x ret. | 0.00000 | 0.00000 | 0.00000 | 83.43180 | 0.00000 | 70.53905 | 83.22176 | 81.04129 |
| x shift | 10.58024 | 7.63017 | 9.97023 | 0.00000 | 9.40022 | 7.79018 | 9.37021 | 8.13019 |
| a | 29.73326 | 30.90177 | 0.00000 | 0.00000 | 28.99369 | 26.51705 | 25.22472 | 27.32467 |
| b | 0.02371 | 0.04437 | 0.12430 | 0.04594 | 0.09476 | 0.03514 | 0.01741 | 0.03573 |

## APPENDIX B

Numerical estimates of various parameters and variables that define the model migration schedule: males, 1970.

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| gmr (obs) | 0.29635 | 0.08633 | 0.01197 | 0.09902 | 0.14644 | 0.48970 | 0.10358 | 1.23339 |
| gmr (mms) | 1.01490 | 1.00524 | 1.05739 | 1.26116 | 1.06026 | 1.02457 | 1.08862 | 1.03365 |
| mae%m | 9.11786 | 18.20402 | 13.44149 | 7.42840 | 13.83925 | 6.26912 | 6.84090 | 5.31996 |
| a1 | 0.02481 | 0.02661 | 0.00000 | 0.00000 | 0.01596 | 0.03760 | 0.03054 | 0.02554 |
| alfa1 | 0.06229 | 0.09435 | 0.00000 | 0.00000 | 0.26587 | 0.16101 | 0.11857 | 0.16790 |
| a2 | 0.05466 | 0.05322 | 0.23743 | 0.00000 | 0.23200 | 0.07236 | 0.02462 | 0.06010 |
| mu2 | 21.41127 | 19.24976 | 19.25745 | 0.00000 | 20.70897 | 22.34701 | 21.45374 | 20.44023 |
| alfa2 | 0.06993 | 0.05193 | 0.26361 | 0.00000 | 0.23683 | 0.14367 | 0.06815 | 0.11497 |
| lambda2 | 0.29810 | 0.50858 | 0.25130 | 0.00000 | 0.57681 | 0.29268 | 0.24080 | 0.48173 |
| a3 | 0.00000 | 0.00000 | 0.00000 | 0.00700 | 0.00000 | 0.00113 | 0.00032 | 0.00099 |
| mu3 | 0.00000 | 0.00000 | 0.00000 | 144.75052 | 0.00000 | 140.82730 | 149.12640 | 120.73910 |
| alfa3 | 0.00000 | 0.00000 | 0.00000 | 0.12361 | 0.00000 | 0.08521 | 0.15639 | 0.16812 |
| lambda3 | 0.00000 | 0.00000 | 0.00000 | 0.02880 | 0.00000 | 0.02133 | 0.02890 | 0.04242 |
| c | -0.00101 | -0.00250 | 0.00151 | 0.00312 | 0.00146 | 0.00129 | 0.00234 | 0.00345 |
| mean age | 27.13773 | 27.90430 | 24.65069 | 75.00883 | 26.61332 | 35.16151 | 44.05198 | 36.21602 |
| %( 0-14) | 22.44630 | 17.61143 | 7.14535 | 3.63024 | 7.50273 | 22.52970 | 22.99731 | 18.41443 |
| %(15-64) | 76.53215 | 82.10146 | 88.75771 | 14.63406 | 88.57130 | 59.43447 | 44.00539 | 63.24842 |
| %(65+ ) | 1.02155 | 0.28712 | 4.09695 | 81.73570 | 3.92598 | 18.03583 | 32.99730 | 18.33715 |
| deltalc | -24.67950 | -10.64596 | 0.00000 | 0.00000 | 10.95189 | 29.11929 | 13.06424 | 7.41177 |
| delta12 | 0.45389 | 0.50000 | 0.00000 | 0.00000 | 0.06880 | 0.51960 | 1.24073 | 0.42503 |
| delta32 | 0.00000 | 0.00000 | 0.00000 | 0.00000 | 0.00000 | 0.01568 | 0.01300 | 0.01646 |
| beta12 | 0.89072 | 1.81687 | 0.00000 | 0.00000 | 1.12262 | 1.12067 | 1.73971 | 1.46036 |
| sigma2 | 4.26284 | 9.79323 | 0.95333 | 0.00000 | 2.43556 | 2.03720 | 3.53323 | 4.19007 |
| sigma3 | 0.00000 | 0.00000 | 0.00000 | 0.23295 | 0.00000 | 0.25031 | 0.18478 | 0.25233 |
| x low | 15.78025 | 15.81025 | 6.37003 | 0.00000 | 16.53026 | 15.93025 | 15.63024 | 16.40026 |
| x high | 25.83048 | 23.49042 | 19.08032 | 0.00000 | 22.26040 | 24.72045 | 25.99048 | 23.38042 |
| x ret. | 0.00000 | 0.00000 | 0.00000 | 94.15409 | 0.00000 | 75.33007 | 90.43330 | 88.20282 |
| x shift | 10.05023 | 7.68018 | 12.71029 | 0.00000 | 5.73013 | 8.79020 | 10.36024 | 6.98016 |
| a | 33.67702 | 39.66355 | 0.00000 | 0.00000 | 33.03363 | 32.36700 | 26.43486 | 31.65367 |
| b | 0.02600 | 0.03475 | 0.08745 | 0.04977 | 0.10661 | 0.02890 | 0.00891 | 0.03242 |

Numerical estimates of various parameters and variables that define the model migration schedule: females, 1970.

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| gmr (obs) | 0.25620 | 0.06765 | 0.00713 | 0.15162 | 0.16827 | 0.52903 | 0.11641 | 1.29630 |
| gmr (mms) | 1.00283 | 1.02095 | 1.10634 | 1.26649 | 1.05566 | 1.03134 | 1.08222 | 1.04361 |
| mae%m | 10.68861 | 16.82710 | 21.40944 | 10.51390 | 21.67021 | 6.72738 | 5.22975 | 7.05139 |
| a1 | 0.03050 | 0.03101 | 0.00000 | 0.00000 | 0.01747 | 0.03401 | 0.02742 | 0.02269 |
| alfa1 | 0.06115 | 0.09488 | 0.00000 | 0.00000 | 0.31199 | 0.17930 | 0.15458 | 0.16517 |
| a2 | 0.05225 | 0.05554 | 0.28905 | 0.00000 | 0.17416 | 0.08398 | 0.04628 | 0.08209 |
| mu2 | 21.05717 | 18.17378 | 18.36008 | 0.00000 | 24.94529 | 21.12695 | 21.00349 | 20.36508 |
| alfa2 | 0.08771 | 0.06215 | 0.41998 | 0.00000 | 0.41376 | 0.20297 | 0.17692 | 0.18910 |
| lambda2 | 0.26106 | 0.65986 | 0.31031 | 0.00000 | 0.20119 | 0.34715 | 0.33487 | 0.35138 |
| a3 | 0.00000 | 0.00000 | 0.00000 | 0.00113 | 0.00000 | 0.00397 | 0.00179 | 0.00278 |
| mu3 | 0.00000 | 0.00000 | 0.00000 | 158.81168 | 0.00000 | 191.60085 | 167.78363 | 111.57628 |
| alfa3 | 0.00000 | 0.00000 | 0.00000 | 0.13826 | 0.00000 | 0.03228 | 0.09642 | 0.16549 |
| lambda3 | 0.00000 | 0.00000 | 0.00000 | 0.02553 | 0.00000 | 0.01086 | 0.02210 | 0.05004 |
| c | -0.00019 | -0.00165 | 0.00298 | 0.00267 | 0.00125 | 0.00178 | 0.00447 | 0.00449 |
| mean age | 24.15170 | 24.70141 | 25.91103 | 75.21085 | 24.28372 | 35.83890 | 44.50188 | 37.45120 |
| %( 0-14) | 29.50291 | 21.96146 | 11.71107 | 3.25016 | 7.36534 | 20.98272 | 20.85575 | 18.41419 |
| %(15-64) | 69.33871 | 78.60107 | 80.39476 | 14.49330 | 89.27168 | 59.54916 | 44.74255 | 59.26760 |
| %(65+ ) | 1.15838 | -0.56254 | 7.89417 | 82.25653 | 3.36298 | 19.46813 | 34.40170 | 22.31821 |
| delta1c | -156.61819 | -18.73866 | 0.00000 | 0.00000 | 13.93153 | 19.10191 | 6.13840 | 5.04880 |
| delta12 | 0.58370 | 0.55832 | 0.00000 | 0.00000 | 0.10033 | 0.40496 | 0.59235 | 0.27645 |
| delta32 | 0.00000 | 0.00000 | 0.00000 | 0.00000 | 0.00000 | 0.04729 | 0.03867 | 0.03388 |
| beta12 | 0.69719 | 1.52680 | 0.00000 | 0.00000 | 0.75404 | 0.88339 | 0.87374 | 0.87344 |
| sigma2 | 2.97637 | 10.61806 | 0.73887 | 0.00000 | 0.48626 | 1.71036 | 1.89281 | 1.85813 |
| sigma3 | 0.00000 | 0.00000 | 0.00000 | 0.18469 | 0.00000 | 0.33657 | 0.22921 | 0.30239 |
| x low | 14.92023 | 15.49024 | 7.86007 | 0.00000 | 12.06016 | 15.30024 | 15.45024 | 14.65022 |
| x high | 24.60045 | 21.55038 | 17.39028 | 0.00000 | 21.37037 | 22.67040 | 22.81041 | 22.10039 |
| x ret. | 0.00000 | 0.00000 | 0.00000 | 92.66377 | 0.00000 | 91.22346 | 101.12558 | 87.68271 |
| x shift | 9.68022 | 6.06014 | 9.53022 | 0.00000 | 9.31021 | 7.37017 | 7.36017 | 7.45017 |
| a | 26.90043 | 32.57695 | 0.00000 | 0.00000 | 28.56203 | 32.77700 | 22.64613 | 26.22182 |
| b | 0.01966 | 0.03712 | 0.11248 | 0.04722 | 0.09768 | 0.03288 | 0.01756 | 0.03276 |

# MULTIREGIONAL ZERO-GROWTH POPULATIONS WITH CHANGING RATES

*Young J. Kim*

## 1 INTRODUCTION

In recent years, fertility in most developed countries has fallen toward or has already nearly reached replacement levels, with perturbations from time to time resulting from social and economic conditions. We are interested in describing mathematically the dynamics of such populations. To do this we begin with the result of the *weak ergodic theorem*, which states that the age structure of a population subject to an arbitrary sequence of fertility and mortality schedules over time eventually loses its dependence on the initial age distribution and comes to be a function only of its relatively recent history of fertility and mortality rates (e.g., Lopez 1961). Nothing is said in this theorem, however, about how age structure is determined by recent vital rates or how the effect of an initial age structure is lost. This has led us to examine the dynamics of populations with arbitrarily changing vital rates but with a Net Reproduction Rate (*NRR*) of unity (Kim and Sykes 1978).

As levels and changes in levels of fertility and mortality diminish, in-migration and out-migration play an increasingly important role in determining the dynamics of regional populations. Rogers (1975) has developed a model of multiregional population dynamics in which migration schedules as well as mortality and fertility schedules play an important role. He has thereby extended stable population theory to include multiregional populations. By analogy to stable-population theory, multiregional stable-population theory states that if regional age-specific schedules of fertility, mortality, and migration are fixed for a long time, the population evolves into a multiregional stable population with fixed regional shares and regional age compositions.

Stable theory for populations with fixed vital rates cannot be extended in a predictable way to populations with rates that are arbitrarily changing over time. We can, however, obtain specific formulas for various attributes of such populations and see how weak ergodicity works explicitly for populations that are almost stationary but with otherwise arbitrary rates. To do this we restrict the number of age groups to two for a closed population and the number of regions to two for a multiregional population without age structure. The results obtained are *qualitatively* true for populations with a greater number of age groups and regions. We also follow the usual restriction of a one-sex model when age structure is considered.

In Section 2 we review the dynamics of closed populations with changing rates but with an *NRR* equal to unity. In Section 3 we formulate general biregional population dynamics with changing vital rates, and obtain specific formulas for the dynamics of special zero-growth biregional populations in Section 4. In Section 5, examples are given and discussed. We discuss the dynamics of populations with age groups *and* regions in Section 6, and conclude the paper (Section 7) with a discussion of several interpretations of some old and some new concepts.

## 2 DYNAMICS OF CLOSED ZERO-GROWTH POPULATIONS

In this section we summarize the results of Kim and Sykes (1978). However, the vectors and matrices describing population dynamics are given here in more conventional forms. This change makes the generalization to multiregional dynamics easier, allowing one to retain the representations usually used by demographers.

We consider a closed population with two age groups expressed in vector form

$$\boldsymbol{X}_t = (X_{t1} X_{t2})' \qquad (t = 0, 1, 2, \ldots) \tag{1}$$

and introduce the $2 \times 2$ population projection matrix (ppm)

$$\mathbf{A}_t = \begin{bmatrix} b_{t1} & b_{t2} \\ s_t & 0 \end{bmatrix} \qquad (t = 1, 2, \ldots) \tag{2}$$

(See Kim and Sykes 1978, for a more detailed explanation of the notation used.) Since the dynamics of a population with age structure $\boldsymbol{X}_t$ at time $t$ are given by

$$\boldsymbol{X}_{t+1} = \mathbf{A}_{t+1} \boldsymbol{X}_t \qquad (t = 0, 1, 2, \ldots) \tag{3}$$

it follows that the age distribution at time $t$ is

$$\boldsymbol{X}_t = \mathbf{A}_t \mathbf{A}_{t-1} \cdots \mathbf{A}_2 \mathbf{A}_1 \boldsymbol{X}_0 = \mathbf{M}_t \boldsymbol{X}_0 \tag{4}$$

where we have written the backward product of $t$ ppms as $\mathbf{M}_t$ to avoid writing a long string of matrices in the rest of the analysis.

We first consider the dynamics of populations when the ppm is row-stochastic, i.e.,

$$\mathbf{A}_t = \begin{bmatrix} 1 - b_t & b_t \\ 1 & 0 \end{bmatrix} \qquad (t = 1, 2, \ldots) \tag{5}$$

in which fertility is split arbitrarily between the two age groups and mortality is set to zero. Note that the period net reproduction rate $NRR_t = 1$ for all $t$, but that the cohort (generation) $NRR_t = 1 - b_t + b_{t+1} \neq 1$. By directly multiplying the matrices, it can be shown that

$$\mathbf{M}_t = \begin{bmatrix} G_t & 1 - G_t \\ G_t - \gamma_t & 1 - G_t + \gamma_t \end{bmatrix} \qquad (t = 1, 2, \ldots) \tag{6}$$

where

$$\begin{aligned} G_t &= 1 - b_1 + b_1 b_2 - \cdots + (-1)^t b_1 \cdots b_t \\ \gamma_t &= (-1)^t b_1 \cdots b_t \end{aligned} \tag{7}$$

If the sequence of birth rates $\{b_t\}$ is bounded below 1 (a sufficient, but not a necessary condition) then the two limits

$$\begin{aligned} \lim_{t \to \infty} \gamma_t &= 0 \\ \lim_{t \to \infty} G_t &= G \end{aligned} \tag{8}$$

exist. Hence the product matrix $\mathbf{M}_t$ converges to a constant matrix of rank 1, i.e.

$$\lim_{t\to\infty} \mathbf{M}_t = \begin{bmatrix} G & 1-G \\ G & 1-G \end{bmatrix} \tag{9}$$

where $G$ satisfies

$$1-b_1 < G < 1-b_1+b_1b_2 \tag{10}$$

and thus is completely determined by the first two values of the sequence $\{b_t\}$. From eqns. (4) and (9), the age distribution at time $t$, for large $t$, is given by

$$\boldsymbol{X}_t = [GX_{01}+(1-G)X_{02}]\boldsymbol{I} \tag{11}$$

where $\boldsymbol{I}$ is the unit vector. The eventual population size is a weighted average of the initial population in the two age groups, with the weight determined by early vital rates; the relative age distribution is uniform. Thus, even with a changing fertility pattern over time, strong ergodicity holds and all of the usual measures describing the population eventually become constant.

We next consider the dynamics of populations with column-stochastic ppms, i.e., $\mathbf{A}_t$ now has the form

$$\mathbf{A}_t = \begin{bmatrix} 1-s_t & 1 \\ s_t & 0 \end{bmatrix} \qquad (t=1,2,\ldots) \tag{12}$$

In this case, fertility in the second age group is constant, while that in the first age group varies with mortality. For this ppm, both period and cohort *NRR* are unity. By directly multiplying the ppms, it can be shown that

$$\mathbf{M}_t = \begin{bmatrix} H_t & H_t-\eta_t \\ 1-H_t & 1-H_t+\eta_t \end{bmatrix} \qquad (t=1,2,\ldots) \tag{13}$$

where

$$\begin{aligned} H_t &= 1-s_t+s_ts_{t-1}-\cdots+(-1)^t s_t\cdots s_1 \\ \eta_t &= (-1)^t s_t\cdots s_1 \end{aligned} \tag{14}$$

Although the form of $H_t$ is superficially similar to the expression for $G_t$ given in eqn. (7), it differs crucially in that $H_t$ is mainly determined by the later elements of the sequence $\{s_t\}$. Because of this, the sequence $\{H_t\}$ has no limit as $t$ increases, although

$$1-s_t < H_t < 1-s_t+s_ts_{t-1} \tag{15}$$

The value of $H_t$ can be calculated to any required degree of accuracy by using more terms. The sequence of product matrices $\{\mathbf{M}_t\}$ satisfies

$$\lim_{t\to\infty} \left\{ \mathbf{M}_t - \begin{bmatrix} H_t & H_t \\ 1-H_t & 1-H_t \end{bmatrix} \right\} = 0 \tag{16}$$

Thus, from eqns. (4) and (16), the population at time $t$, for large $t$, is given by

$$\boldsymbol{X}_t = (X_{01}+X_{02}) \begin{bmatrix} H_t \\ 1-H_t \end{bmatrix} \tag{17}$$

i.e., although the total population size is fixed at all times, the number of births and the age distribution constantly change over time and are determined by the most recent values of the vital rates. (See Kim and Sykes 1978 for more detailed discussion and generalization.)

## 3 BIREGIONAL POPULATION DYNAMICS

We now consider populations without age structure located in two regions. A major formal difference between this population and the "closed" population with two age groups is that now all four transitions are possible, whereas for the closed population the contribution from the second age group to itself was zero. When this is translated into a transition matrix (also a ppm) all four cells have non-zero entries for the two-region (biregional) dynamics.

We can formulate the dynamics of regional populations in two ways. First, we may take a period approach by using the two-region accounting relationship (Rogers 1968):

$$\begin{bmatrix} P_1(t) \\ P_2(t) \end{bmatrix} = \begin{bmatrix} 1+b_1(t)-d_1(t)-o_1(t) & o_2(t) \\ o_1(t) & 1+b_2(t)-d_2(t)-o_2(t) \end{bmatrix} \begin{bmatrix} P_1(t-1) \\ P_2(t-1) \end{bmatrix}$$

$$(t=1,2,\ldots) \qquad (18)$$

where $P_i(t)$, $(i=1,2)$, is the population size in region $i$ at time $t$, and $b_i(t)$, $d_i(t)$, and $o_i(t)$, $(i=1,2)$, are the crude rates of birth, death, and out-migration, respectively, for region $i$ at time $t$. If the rates given are the usual single-year rates, the time unit of eqn. (18) is also one year. If we denote the matrix in eqn. (18) as $\mathbf{C}(t)$, i.e.

$$\mathbf{C}(t) = \begin{bmatrix} 1+b_1(t)-d_1(t)-o_1(t) & o_2(t) \\ o_1(t) & 1+b_2(t)-d_2(t)-o_2(t) \end{bmatrix} \qquad (19)$$

then the dynamics of regional populations at time $t$ are given by

$$\boldsymbol{P}(t) = \mathbf{C}(t)\mathbf{C}(t-1)\cdots\mathbf{C}(2)\mathbf{C}(1)\boldsymbol{P}(0) \qquad (20)$$

The dynamics of a two-region population can also be described using the generation method. The birth sequence in two regions satisfies the expression

$$\begin{bmatrix} B_1(t) \\ B_2(t) \end{bmatrix} = \begin{bmatrix} R_{11}(t) & R_{21}(t) \\ R_{12}(t) & R_{22}(t) \end{bmatrix} \begin{bmatrix} B_1(t-1) \\ B_2(t-1) \end{bmatrix} \qquad (21)$$

where $B_i(t)$, $(i=1,2)$, is the number of births in region $i$ at time $t$, and $R_{ij}(t)$, $(i,j=1,2)$, is the Spatial Net Reproduction Rate (*SNRR*) in region $j$ of women born in region $i$ at time $t-1$. Here, the time unit is the length of a generation. $R_{ij}(t)$ is given by

$$R_{ij}(t) = \int_\alpha^\beta p_{ij}^t(x) m_j^t(x)\, \mathrm{d}x \qquad (i,j=1,2) \qquad (22)$$

where $p_{ij}^t(x)$ is the probability of surviving to age $x$ in region $j$ for those born in region $i$ at time $t$, and $m_j^t(x)$ is the age-specific fertility rate for age $x$ in region $j$ at time $t$. Note that although Rogers and Willekens (1976a, eqn. 4.2; 1976b, eqn. 3) described expressions similar to eqn. (21), they restricted themselves to the limiting stationary birth sequence of populations with zero growth rates. Also note that $R_{ij}(t)$ in eqns.

(21) and (22) is ${}_iR_j(0)$ at time $t$ in the notation of Rogers and Willekens. From eqn. (21), the dynamics of births at time $t$ are given by

$$\boldsymbol{B}(t) = \mathbf{R}(t)\mathbf{R}(t-1)\cdots\mathbf{R}(2)\mathbf{R}(1)\boldsymbol{B}(0) \tag{23}$$

and are thus determined by the product of *SNRR* matrices.

We may consider the dynamics of population size in a similar way to the dynamics of births. The population at time $t$ in terms of births at time $t$ may be expressed as

$$\begin{bmatrix} Y_1(t) \\ Y_2(t) \end{bmatrix} = \begin{bmatrix} a_{11}(t) & a_{21}(t) \\ a_{12}(t) & a_{22}(t) \end{bmatrix} \begin{bmatrix} B_1(t) \\ B_2(t) \end{bmatrix} \tag{24}$$

where $Y_i(t)$ is the population in region $i$ at time $t$, and $a_{ij}(t)$ represents the "discounted" number of years lived in region $j$, on the average, by individuals born in region $i$ at time $t$. The $a_{ij}(t)$ may be formally expressed as

$$a_{ij}(t) = \int_0^w \exp\,[-r_j(t)x]p^t_{ij}(x)\,\mathrm{d}x$$

whereas the usual multiregional expectation of life at birth at time $t$ is

$$e_{ij}(t) = \int_0^w p^t_{ij}(x)\,\mathrm{d}x$$

$p^t_{ij}(x)$ has already been defined in eqn. (22), and $r_j(t)$ is the annual growth rate of the population of region $j$ at time $t$. We may note that the $a_{ij}$ correspond to an actuarial function, $a_0$, which is the value at birth of a *life annuity*.

By combining eqns. (21) and (24), we obtain the dynamics of populations

$$\boldsymbol{Y}(t) = [\mathbf{A}(t)\mathbf{R}(t)\mathbf{A}^{-1}(t-1)]\boldsymbol{Y}(t-1) = \mathbf{N}(t)\boldsymbol{Y}(t-1) \tag{25}$$

$\mathbf{N}(t)$ in eqn. (25) is defined as

$$\mathbf{N}(t) = \mathbf{A}(t)\mathbf{R}(t)\mathbf{A}^{-1}(t-1) \tag{26}$$

and may be called the Spatial Net Reproduction Rate for a Population (*SNRRP*). Its element $N_{ij}(t)$ has the same interpretation as the more widely used $R_{ij}(t)$, except that $N_{ij}(t)$ refers to populations and $R_{ij}(t)$ to births. From eqn. (25), it follows that the regional population distribution at time $t$ is given by

$$\boldsymbol{Y}(t) = \mathbf{N}(t)\mathbf{N}(t-1)\cdots\mathbf{N}(2)\mathbf{N}(1)\boldsymbol{Y}(0) \tag{27}$$

The three formulations of biregional population dynamics given by eqns. (20), (23), and (27) are formally identical in that the dynamics are completely determined by the product of $2\times2$ ppms. In addition, the three ppms $\mathbf{C}$, $\mathbf{R}$, and $\mathbf{N}$ are related in the following manner. First, the *SNRR* matrix $\mathbf{R}$ and *SNRRP* matrix $\mathbf{N}$ are related by eqn. (26). Moreover, the product matrices have the relationship

$$\mathbf{N}(t)\mathbf{N}(t-1)\cdots\mathbf{N}(2)\mathbf{N}(1) = \mathbf{A}(t)[\mathbf{R}(t)\mathbf{R}(t-1)\cdots\mathbf{R}(2)\mathbf{R}(1)]\mathbf{A}^{-1}(0)$$

We notice that when the rates are constant over time, $\mathbf{N}$ is a similarity transform of $\mathbf{R}$, and the product matrix of $\mathbf{N}$s is a similarity transform of the product matrix of $\mathbf{R}$s. Second, since $\mathbf{C}(t)$ projects the population forward one year, while $\mathbf{N}(t)$ projects the population forward the length of a generation, they satisfy the relationship

$$\mathbf{C}(t+T-1)\cdots\mathbf{C}(t+1)\mathbf{C}(t) = \mathbf{N}(t) \tag{28}$$

where $T$ is the length of a generation. When the rates are assumed to be constant during a generation, eqn. (28) reduces to

$$[\mathbf{C}(t)]^T = \mathbf{N}(t) \tag{29}$$

To give a better idea of the magnitude of the elements of **C** and **N**, 1970 data for India and the USSR are used in the following illustration. The single-year ppms, **C**, are given by Rogers (1980) as

$$\mathbf{C}(\text{India}) = \begin{bmatrix} 1.010 & 0.007 \\ 0.010 & 1.015 \end{bmatrix}$$
$$\mathbf{C}(\text{USSR}) = \begin{bmatrix} 0.998 & 0.035 \\ 0.011 & 0.975 \end{bmatrix} \tag{30}$$

from which we obtain the ppms, **N**,

$$\mathbf{N}(\text{India}) = \begin{bmatrix} 1.373 & 0.290 \\ 0.415 & 1.581 \end{bmatrix}$$
$$\mathbf{N}(\text{USSR}) = \begin{bmatrix} 1.068 & 0.741 \\ 0.233 & 0.581 \end{bmatrix} \tag{31}$$

where it is assumed that $T$ is 28 years, and that the birth, death, and migration rates are constant over time. The ppm **R** may be obtained by using the relationship

$$\mathbf{R} = \mathbf{A}^{-1}\mathbf{N}\mathbf{A}$$

However, the *SNRR* matrix calculated in this way will not agree with values calculated from age-specific rates of fertility, mortality, and migration. This discrepancy is due to a bias introduced by assuming that crude rates of birth, death, and migration are constant over time.

## 4 DYNAMICS OF BIREGIONAL ZERO-GROWTH POPULATIONS

Before exploring the dynamics of regional populations with changing vital rates, let us briefly summarize the dynamics of stable stationary populations (see Rogers and Willekens 1976b). When the rates are constant over time, the limiting distribution is determined by the maximum eigenvalue and the corresponding eigenvectors of the ppm. A stationary multiregional population results if the maximum eigenvalue is unity, i.e., in the case of a $2 \times 2$ matrix

$$\mathbf{N} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

if the elements satisfy the relationship

$$(1-a)(1-d) = bc \tag{32}$$

A special case of eqn. (32) is a row-stochastic matrix, which may be written as

$$\mathbf{N} = \begin{bmatrix} 1-b & b \\ c & 1-c \end{bmatrix} \tag{33}$$

An example of this form is given by Rogers and Willekens (1976b, p. 8). The limiting stationary distribution becomes

$$\boldsymbol{Y} = \left(\frac{bY_1(0) + cY_2(0)}{b + c}\right)\boldsymbol{I} \tag{34}$$

which states that the two regions will contain equal numbers of people, and that the ultimate population will be determined by a weighted average of the initial populations in the two regions.

Another special case of eqn. (32) is a column-stochastic matrix (Rogers and Willekens 1976b, p. 7)

$$\mathbf{N} = \begin{bmatrix} 1-c & b \\ c & 1-b \end{bmatrix} \tag{35}$$

When the matrix is of this form, the limiting stationary distribution is given by

$$\boldsymbol{Y} = [Y_1(0) + Y_2(0)]\begin{bmatrix} b/(b+c) \\ c/(b+c) \end{bmatrix} \tag{36}$$

so that the ratio of regional shares will be $b/c$.

Now we consider the dynamics of populations in two regions with changing rates specified by any of the eqns. (20), (23), and (27). Before going into details, we first summarize and interpret useful theorems given by Chatterjee and Seneta (1977) on backward products of row-stochastic matrices. For backward products

$$\mathbf{M}_t = \mathbf{A}_t\mathbf{A}_{t-1} \cdots \mathbf{A}_2\mathbf{A}_1$$

of row-stochastic matrices $\{\mathbf{A}_t\}$, weak and strong ergodicity are equivalent, i.e.,

$$\lim_{t\to\infty} \mathbf{M}_t = \boldsymbol{IP}' \tag{37}$$

where $\boldsymbol{P}'$ is a probability vector that depends on the elements of the first two matrices in the series $\mathbf{A}_t$ (Theorem 1). A sufficient condition for ergodicity is

$$\lim_{t\to\infty} \sum_{k=1}^{t} \varepsilon_k = \infty \tag{38}$$

where $\varepsilon_k$ is the minimum element of the matrix $\mathbf{N}_t$ (corollary of Theorem 4).

Although the limit theorem given in eqn. (37) is known, it does not tell us anything about how the sequence converges, or what the elements of the limiting matrix are. To see how the sequence converges, we proceed as in the previous section.

When the matrices representing changing rates are of the form

$$\mathbf{A}_t = \begin{bmatrix} 1-a_t & a_t \\ c_t & 1-c_t \end{bmatrix} \qquad (t = 1, 2, \ldots) \tag{39}$$

the row elements sum to one. This condition is equivalent to

$$d_i(t) + o_i(t) = b_i(t) + o_j(t) \qquad (i, j = 1, 2) \tag{40}$$

for the period model of eqn. (18), and to

$$\boldsymbol{R}_{ii}(t) + \boldsymbol{R}_{ji}(t) = 1 \quad \text{or} \quad \boldsymbol{N}_{ii}(t) + \boldsymbol{N}_{ji}(t) = 1 \qquad (i, j = 1, 2) \tag{41}$$

for the generation models. By directly multiplying the matrices, we obtain

$$\mathbf{M}_1 = \mathbf{A}_1 = \begin{bmatrix} 1-a_1 & a_1 \\ c_1 & 1-c_1 \end{bmatrix}$$

$$\mathbf{M}_2 = \mathbf{A}_2\mathbf{A}_1 = \begin{bmatrix} 1-\{a_1+a_2[1-(a_1+c_1)]\} & a_1+a_2[1-(a_1+c_1)] \\ c_1+c_2[1-(a_1+c_1)] & 1-\{c_1+c_2[1-(a_1+c_1)]\} \end{bmatrix}$$

and so on.

It can be shown that the upper right-hand corner element of $\mathbf{M}_t$ is given by

$$G_t = a_1 + a_2[1-(a_1+c_1)] + a_3[1-(a_1+c_1)][1-(a_2+c_2)] + \cdots + a_t[1-(a_1+c_1)] \cdots [1-(a_{t-1}+c_{t-1})] \tag{42}$$

for all $t$. Since $0 < |1-(a_t+c_t)| < 1$, for all $t$, the sequence $\{G_t\}$ converges to some value $G$, i.e.

$$\lim_{t\to\infty} G_t = G \tag{43}$$

exists and hence

$$\lim_{t\to\infty} \mathbf{M}_t = \begin{bmatrix} 1-G & G \\ 1-G & G \end{bmatrix}$$

The value of $G$ can be calculated explicitly to any required degree of accuracy from eqn. (42), by using only the early rates. The speed of convergence depends on the values of $a_t$ and $c_t$; more specifically, we see from eqn. (42) that the smaller the value of $|1-(a_t+c_t)|$, $(t=1,2,\ldots)$, the faster is the convergence. Consequently the regional population is given by

$$\lim_{t\to\infty} \begin{bmatrix} y_1(t) \\ y_2(t) \end{bmatrix} = [(1-G)y_1(0) + Gy_2(0)] \begin{bmatrix} 1 \\ 1 \end{bmatrix} \tag{44}$$

Equation (44) should be interpreted with $y_i(t)$, $a_t$, and $c_t$ replaced by $P_i(t)$, $o_2(t)$, and $o_1(t)$ in the period model, and by $Y_i(t)$, $N_{21}(t)$, and $N_{12}(t)$, or $B_i(t)$, $R_{21}(t)$, and $R_{12}(t)$ in the generation models. Thus, despite constantly changing rates over time, the regional population will eventually have a constant (stationary) distribution: the population will have equal regional shares and the size of the population will be a weighted average of the initial population distribution, where the weights (spatial reproductive values) are given by $(1-G)$ and $G$.

We next consider the dynamics of populations when the ppm is column-stochastic, i.e.

$$\mathbf{A}_t = \begin{bmatrix} 1-a_t & c_t \\ a_t & 1-c_t \end{bmatrix} \qquad (t=1,2,\ldots) \tag{45}$$

This condition is equivalent to

$$b_i(t) = d_i(t) \qquad (i=1,2) \tag{46}$$

with arbitrary $o_i(t)$ for the period model, and to

$$R_{ii}(t) + R_{ij}(t) = 1 \quad \text{or} \quad N_{ii}(t) + N_{ij}(t) = 1 \qquad (i,j=1,2) \tag{47}$$

for the generation models. Letting $H_t$ denote the upper right-hand element of the product matrix $\mathbf{M}_t$, and multiplying directly, we obtain

$$H_1 = c_1$$

$$H_2 = c_2 + c_1[1-(a_2+c_2)]$$

and, in general

$$H_t = c_t + c_{t-1}[1-(a_t+c_t)] + c_{t-2}[1-(a_t+c_t)][1-(a_{t-1}+c_{t-1})] + \cdots + c_1[1-(a_t+c_t)]\cdots[1-(a_2+c_2)] \quad (48)$$

for all $t$.

Since $0 < |1-(a_t+c_t)| < 1$, for all $t$, the value of $H_t$ depends only on the most recent rates. Although $H_t$ has no limit, the product matrix $\mathbf{M}_t$ satisfies

$$\lim_{t\to\infty}\left\{\mathbf{M}_t - \begin{bmatrix} H_t & H_t \\ 1-H_t & 1-H_t \end{bmatrix}\right\} = 0 \quad (49)$$

(In fact, it is a transpose of a non-homogeneous Markov chain.) For large $t$, the population in two regions is given by

$$\begin{bmatrix} y_1(t) \\ y_2(t) \end{bmatrix} = [y_1(0)+y_2(0)]\begin{bmatrix} H_t \\ 1-H_t \end{bmatrix} \quad (50)$$

The size of the total population is constant, but regional shares change over time. This occurs in the period model when the period rate of natural increase in each region is zero but interregional migration occurs at arbitrary rates. In the generation model, this situation arises when the *SNRRP* in the region of origin is unity, i.e., one child is born per person on the average in such a way that $N_{ii}$ is associated with the region of origin and the rest $(1-N_{ii})$ with the other region.

## 5 EXAMPLES

Table 1 shows an example of the dynamics of regional populations with row-stochastic ppms. The second and third columns give off-diagonal elements $a_t$ and $c_t$ of the ppm. Columns 4 and 5 give the populations $Y_1(t)$, $Y_2(t)$ in regions 1 and 2, column 6 gives the total combined population, $Y(t)$, and column 7 gives the percentage of the total population residing in region 1, % $Y_1(t)$; the data correspond to an initial population such that there are 100 persons in region 1 and zero in region 2. Columns 8–11 correspond to columns 4–7 for an initial population of zero in region 1 and 100 persons in region 2. Since any arbitrary initial distribution may be expressed as a linear combination of these two populations, it is possible to compute the dynamics of any population evolving under these given rates.

The elements $a_t$ and $c_t$ may be thought of either as $o_2(t)$ and $o_1(t)$ in the period model, or as $R_{21}(t)$ and $R_{12}(t)$, or $N_{21}(t)$ and $N_{12}(t)$ in the generation models. However, the values used in this example are more compatible with the generation models, as suggested by the actual data for India and the USSR presented in eqns. (30) and (31). Even with values that reflect period out-migration rates, our results hold but the time needed for convergence is too long (about 28 years are required for

TABLE 1 Biregional population dynamics with row-stochastic ppms.

| | | | $\boldsymbol{Y}(0)=(100\ \ 0)'$ | | | | $\boldsymbol{Y}(0)=(0\ \ 100)'$ | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| $t$ | $a_t$ | $c_t$ | $Y_1(t)$ | $Y_2(t)$ | $Y(t)$ | $\%\,Y_1(t)$ | $Y_1(t)$ | $Y_2(t)$ | $Y(t)$ | $\%\,Y_1(t)$ |
| 0 | | | 100 | 0 | 100 | 100 | 0 | 100 | 100 | 0 |
| 1 | 0.10 | 0.05 | 90.0 | 5.0 | 95.0 | 94.7 | 10.0 | 95.0 | 105.0 | 9.5 |
| 2 | 0.15 | 0.10 | 77.3 | 13.5 | 90.8 | 85.1 | 22.8 | 86.5 | 109.3 | 20.8 |
| 3 | 0.20 | 0.15 | 64.5 | 23.1 | 87.6 | 73.7 | 35.5 | 76.9 | 112.4 | 31.6 |
| 4 | 0.25 | 0.20 | 54.1 | 31.4 | 85.5 | 63.3 | 45.9 | 68.7 | 114.6 | 40.0 |
| 5 | 0.30 | 0.25 | 47.3 | 37.1 | 84.4 | 56.1 | 52.7 | 63.0 | 115.7 | 45.6 |
| 6 | 0.35 | 0.30 | 43.7 | 40.1 | 83.8 | 52.1 | 56.3 | 59.9 | 116.2 | 48.5 |
| 7 | 0.40 | 0.35 | 42.3 | 41.4 | 83.7 | 50.5 | 57.7 | 58.6 | 116.3 | 49.6 |
| 8 | 0.45 | 0.40 | 41.9 | 41.7 | 83.6 | 50.1 | 58.1 | 58.3 | 116.4 | 49.9 |
| 9 | 0.50 | 0.45 | 41.8 | 41.8 | 83.6 | 50.0 | 58.2 | 58.2 | 116.4 | 50.0 |
| 10 | 0.45 | 0.40 | 41.8 | 41.8 | 83.6 | 50.0 | 58.2 | 58.2 | 116.4 | 50.0 |
| 11 | 0.40 | 0.35 | 41.8 | 41.8 | 83.6 | 50.0 | 58.2 | 58.2 | 116.4 | 50.0 |
| 12 | 0.35 | 0.30 | 41.8 | 41.8 | 83.6 | 50.0 | 58.2 | 58.2 | 116.4 | 50.0 |

convergence in this example). Values for $a_t$ and $c_t$ were chosen so that $a_t > c_t$ at all times. This implies that for the period model

$$o_2(t) > o_1(t) \qquad \text{(for all } t\text{)}$$

and for the generation model

$$\begin{aligned} N_{11}(t)+N_{12}(t) &= N_{1.}(t) > 1 \\ N_{21}(t)+N_{22}(t) &= N_{2.}(t) > 1 \end{aligned} \qquad \text{(for all } t\text{)} \tag{51}$$

(In the cohort model, we may consider either the ppm **R** for births, or the ppm **N** for populations. Here we chose the ppm **N** for our explanation.)

It is obvious that, in spite of the condition imposed, populations stabilize with an equal and constant number of people in each region, regardless of their initial distributions. However, how a population achieves this and what the eventual population size is depends on the initial distribution. When the initial population is entirely in region 1, the population in region 1 decreases, while the population in region 2 increases over time. The total population decreases before it stabilizes. On the other hand, when the initial population is entirely in region 2, the opposite occurs. The size of the final population is a combined effect of the initial distribution and the spatial reproductive values. A mathematical expression of this result is given in eqn. (44). It is interesting to note that this strong ergodicity results when eqn. (51) holds. We also note that

$$N_{.1}(t) = N_{.2}(t) = 1 \qquad \text{(for all } t\text{)} \tag{52}$$

i.e., the *SNRRP* at the destination is conserved in this case.

Table 2 shows an example of the dynamics of regional populations with column-stochastic ppms. We consider two populations with the same initial distributions and the same set of rates as in Table 1. Since the total population size $Y(t) = 100$ is fixed at *all* times, the sizes of population in each region give their respective percentages, and hence $Y(t)$ and $\%\,Y_1(t)$ are omitted from the table. We

TABLE 2 Biregional population dynamics with column-stochastic ppms.

| | | | $Y(0) = (100\ 0)'$ | | $Y(0) = (0\ 100)'$ | |
|---|---|---|---|---|---|---|
| $t$ | $a_t$ | $c_t$ | $Y_1(t)$ | $Y_2(t)$ | $Y_1(t)$ | $Y_2(t)$ |
| 0 | | | 100 | 0 | 0 | 100 |
| 1 | 0.10 | 0.05 | 95.0 | 5.0 | 10.0 | 90.0 |
| 2 | 0.15 | 0.10 | 86.3 | 13.7 | 22.5 | 77.5 |
| 3 | 0.20 | 0.15 | 76.1 | 23.9 | 34.6 | 65.4 |
| 4 | 0.25 | 0.20 | 66.8 | 33.2 | 44.0 | 56.0 |
| 5 | 0.30 | 0.25 | 60.1 | 39.9 | 49.8 | 50.2 |
| 6 | 0.35 | 0.30 | 56.0 | 44.0 | 52.4 | 47.6 |
| 7 | 0.40 | 0.35 | 54.0 | 46.0 | 53.1 | 46.9 |
| 8 | 0.45 | 0.40 | 53.1 | 46.9 | 53.0 | 47.0 |
| 9 | 0.50 | 0.45 | 52.7 | 47.3 | 52.6 | 47.4 |
| 10 | 0.45 | 0.40 | 52.9 | 47.1 | 52.9 | 47.1 |
| 11 | 0.40 | 0.35 | 53.2 | 46.8 | 53.2 | 46.8 |
| 12 | 0.35 | 0.30 | 53.6 | 46.4 | 53.6 | 46.4 |
| 13 | 0.30 | 0.25 | 54.1 | 45.9 | 54.1 | 45.9 |

notice that regional populations reflect the effect of their initial distributions at first, although this effect is gradually lost and they eventually become identical. However, the populations do not stabilize over time but keep changing constantly. This result is contained in eqn. (50). We notice that this occurs when

$$N_{1.}(t) = N_{2.}(t) = 1 \tag{53}$$

is satisfied.

## 6 DYNAMICS OF ZERO-GROWTH MULTIREGIONAL POPULATIONS

The dynamics of multiregional populations, taking age structure into account, may be written (Rogers 1975, pp. 122, 123) as

$$\boldsymbol{K}^{(t)} = \mathbf{G}(t)\boldsymbol{K}^{(t-1)} \tag{54}$$

or

$$\bar{\boldsymbol{K}}^{(t)} = \mathbf{H}(t)\bar{\boldsymbol{K}}^{(t-1)} \tag{55}$$

where the multiregional ppms now depend on time $t$, in contrast to the fixed multiregional ppms of Rogers. (See Rogers 1975 for a description of the elements of the vectors and matrices.) When we consider populations with two age groups and two regions, eqns. (54) and (55) become, respectively,

$$\begin{bmatrix} K_1^t(1) \\ K_2^t(1) \\ \hline K_1^t(2) \\ K_2^t(2) \end{bmatrix} = \left[\begin{array}{cc|cc} b_{11}^t(1) & b_{21}^t(1) & b_{11}^t(2) & b_{21}^t(2) \\ b_{12}^t(1) & b_{22}^t(1) & b_{12}^t(2) & b_{22}^t(2) \\ \hline s_{11}^t & s_{21}^t & 0 & 0 \\ s_{12}^t & s_{22}^t & 0 & 0 \end{array}\right] \begin{bmatrix} K_1^{t-1}(1) \\ K_2^{t-1}(1) \\ \hline K_1^{t-1}(2) \\ K_2^{t-1}(2) \end{bmatrix} \tag{56}$$

$$\begin{bmatrix} K_1^t(1) \\ K_1^t(2) \\ \hdashline K_2^t(1) \\ K_2^t(2) \end{bmatrix} = \left[\begin{array}{cc:cc} b_{11}^t(1) & b_{11}^t(2) & b_{21}^t(1) & b_{21}^t(2) \\ s_{11}^t & 0 & s_{21}^t & 0 \\ \hdashline b_{12}^t(1) & b_{12}^t(2) & b_{22}^t(1) & b_{22}^t(2) \\ s_{12}^t & 0 & s_{22}^t & 0 \end{array}\right] \begin{bmatrix} K_1^{t-1}(1) \\ K_1^{t-1}(2) \\ \hdashline K_2^{t-1}(1) \\ K_2^{t-1}(2) \end{bmatrix} \tag{57}$$

where $K_i^t(j)$ denotes the number of persons in region $i$ in age group $j$ at time $t$; $b_{ij}^t(k)$ is the number of persons in the first age group in region $j$ at time $t$, per person in region $i$ in age group $k$ at time $t-1$; and $s_{ij}^t$ is the proportion surviving in the second age group in region $j$ at time $t$, per person in the first age group in region $i$ at time $t-1$. Since eqn. (54) seems to be simpler to manipulate, we shall use it in the rest of our analysis. The dynamics of multiregional population at time $t$ are given by

$$\begin{aligned} \boldsymbol{K}^{(t)} &= \mathbf{G}(t)\mathbf{G}(t-1)\cdots\mathbf{G}(2)\mathbf{G}(1)\boldsymbol{K}^{(0)} \\ &= \mathbf{M}(t)\boldsymbol{K}^{(0)} \end{aligned} \tag{58}$$

and hence the dynamics are completely determined by the backward product $\mathbf{M}(t)$ of multiregional ppms $\{\mathbf{G}(t)\}$.

As a special case of zero-growth dynamics, we first consider the case of row-stochastic multiregional ppms. Since the multiregional ppm $\mathbf{G}(t)$ is regular, from the theorems of Chatterjee and Seneta (1977) given by eqn. (37), we conclude that

$$\lim_{t\to\infty} \mathbf{M}(t) = \boldsymbol{I}\boldsymbol{p}'$$

holds, so that

$$\lim_{t\to\infty} \boldsymbol{K}^{(t)} = (\boldsymbol{p}, \boldsymbol{K}^{(0)})\boldsymbol{I} \tag{59}$$

where $(\boldsymbol{p}, \boldsymbol{K}^{(0)})$ represents the inner product of the vectors $\boldsymbol{p}$ and $\boldsymbol{K}^{(0)}$. Notice that the elements of the vector $\boldsymbol{p}$ represent spatial reproductive values. Thus when fertility, mortality, and migration rates change over time with the constraints

$$\begin{cases} \sum_{j=1}^{m} \sum_{k=1}^{n} b_{ji}^t(k) = 1 & (i = 1, 2, \ldots, m) \\ \sum_{j=1}^{m} s_{ji}^t(k) = 1 & (i = 1, 2, \ldots, m;\ k = 1, 2, \ldots, n-1) \end{cases} \tag{60}$$

for populations with $n$ age groups and $m$ regions, the population evolves into a multiregional stationary population with constant age distribution and regional shares.

When the multiregional ppm $\mathbf{G}(t)$ is column-stochastic, for all $t$, the sequence of product matrices $\{\mathbf{M}(t)\}$ is a transpose of a non-homogeneous Markov chain, so that

$$\lim_{t\to\infty} [\mathbf{M}(t) - \boldsymbol{h}_t\boldsymbol{I}'] = 0 \tag{61}$$

holds and hence, for large $t$, the multiregional population becomes

$$\boldsymbol{K}^{(t)} = (\boldsymbol{I}, \boldsymbol{K}^{(0)})\boldsymbol{h}_t \tag{62}$$

where the inner product $(\boldsymbol{I}, \boldsymbol{K}^{(0)})$ gives the initial total population, and the vector $\boldsymbol{h}_t$

satisfies $(\boldsymbol{I}, \boldsymbol{h}_t) = 1$. Here the vector $\boldsymbol{h}_t$ is determined by recent rates. The result of eqn. (62) holds, for all $t$, when

$$\sum_{j=1}^{m} [b_{ij}^{t}(k) + s_{ij}^{t}(k)] = 1 \qquad (i = 1, 2, \ldots, m; k = 1, 2, \ldots, n-1) \tag{63}$$

for populations with $n$ age groups and $m$ regions.

## 7 DISCUSSION

We have seen that the dynamics of populations (both single-region and multiregional) represented by a sequence of row-stochastic ppms lead to strong ergodicity, while the dynamics of populations represented by a sequence of column-stochastic ppms result in only weak ergodicity. Specific expressions originally obtained for populations with two age groups in a single region were extended to populations without age structure in two regions. From these expressions we can see explicitly how strong and weak ergodicity work. The dynamics of populations with $n$ age groups and $m$ regions are qualitatively the same as the examples discussed above, although we cannot give explicit formulas in such general cases.

Populations spread over two regions without age structure were modeled in two different ways: a period formulation, which involves the crude rates of birth, death, and migration over each period; and a generation formulation. For the generation model, we formulated the dynamics of births and populations separately, because when the rates change over time an unchanging birth sequence does not generate an unchanging population, and vice versa. We defined the *SNRRP* for the dynamics of populations, in addition to the *SNRR* for the dynamics of births: when we discuss whether the system is stationary we must make it clear whether we are referring to births or populations.

Demographic interpretations of row- and column-stochastic ppms merit some further discussion. For populations with two age groups in a single region, if mortality is fixed, and if a lifetime fertility of unity is split arbitrarily into two age groups (only the period *NRR* is equal to unity), all important features of the population eventually become constant. When fertility in the first age group and the survivorship proportions adjust themselves to a sum of unity, with fertility in the second age group being unity (both period and cohort *NRR* = 1), then the birth sequence and the age structure of the population change over time.

In the case of biregional populations, strong ergodicity results when the *SNRRP* at the region of destination is unity. This may be denoted as "location *SNRRP* = 1" or the *location replacement alternative*. When the *SNRRP* at the region of origin is unity, weak ergodicity results. This may be defined as "cohort *SNRRP* = 1," or the *cohort replacement alternative* (Rogers and Willekens 1976b, p. 6). As noted earlier, the equivalent conditions for the *SNRR* generate the corresponding dynamics of births. It is important to note that, when we consider age distribution, vital rates in most developed countries have ppms which are almost row-stochastic, but when we consider regional distribution, the ppms are more nearly column-stochastic (e.g., USSR data). The implication of this fact is that, under the conditions described above, changes in age distribution will probably be small, but changes in regional distribution will persist in the future.

It may be instructive to demonstrate how strong ergodicity is produced by row-stochastic ppms. The row-stochastic **N** implies

$$Y_i(t+1) = Y_i(t) + N_{ji}(t)[Y_j(t) - Y_i(t)] \qquad (i, j = 1, 2)$$

so that, when $Y_j(t) > Y_i(t)$

$$Y_i(t+1) > Y_i(t)$$

$$Y_j(t+1) < Y_j(t)$$

This process continues until regional populations satisfy $Y_i(t) = Y_j(t)$, and from this point on everything stays constant, as seen in Table 1.

Next we illustrate how weak ergodicity results from column-stochastic ppms. The column-stochastic ppm **N** yields

$$Y_1(t+1) = Y_1(t) + [N_{21}(t)Y_2(t) - N_{12}(t)Y_1(t)]$$

$$Y_2(t+1) = Y_2(t) - [N_{21}(t)Y_2(t) - N_{12}(t)Y_1(t)]$$

We immediately see that the total population $Y_1(t) + Y_2(t)$ is fixed at all times, and that the regional shares $Y_1(t)$ and $Y_2(t)$ change constantly according to the current migration rates and regional distribution. This was illustrated in Table 2. Only when

$$\frac{N_{21}(t)}{N_{12}(t)} = \frac{Y_1(t)}{Y_2(t)}$$

do we get

$$Y_i(t+1) = Y_i(t) \qquad (i = 1, 2)$$

Finally, the main advantage of having specific formulas for the $2 \times 2$ ppms, in addition to the limit theorems, is that we can see the particular workings of strong and weak ergodicity, and can identify the specific features required for the stable stationary case. For example, in eqns. (42) and (48), if $a_t = a$ and $c_t = c$ for all $t$, $G_t$ for large $t$ reduces to

$$G = a\{1 + [1 - (a+c)] + [1 - (a+c)]^2 + \cdots\}$$
$$= \frac{a}{a+c}$$

and $H_t$, for large $t$, reduces to

$$H = c\{1 + [1 - (a+c)] + [1 - (a+c)]^2 + \cdots\}$$
$$= \frac{c}{a+c}$$

which thus gives the results of eqns. (34) and (36) which are obtained from eigenvectors of the constant ppm.

## ACKNOWLEDGMENTS

The author would like to thank Professor Andrei Rogers for helpful discussions and the hospitality extended to her at the International Institute for Applied Systems Analysis (IIASA). She also wishes to thank Dr. Ahmed Seifelnasr for his help.

## REFERENCES

Chatterjee, S., and E. Seneta (1977) Towards consensus: some convergence theorems on repeated averaging. Journal of Applied Probability 14:89–97.

Kim, Y. J., and Z. M. Sykes (1978) Dynamics of some special populations with *NRR* = 1. Demography 15:559–569.

Lopez, A. (1961) Problems in Stable Population Theory. Princeton, N.J.: Office of Population Research, Princeton University.

Rogers, A. (1968) Matrix Analysis of Interregional Population Growth and Distribution. Berkeley, Cal.: University of California Press.

Rogers, A. (1975) Introduction to Multiregional Mathematical Demography. New York, N.Y.: Wiley.

Rogers, A. (1980) Sources of Urban Population Growth and Urbanization, 1950–2000: A Demographic Accounting. Forthcoming.

Rogers, A., and F. Willekens (1976a) Spatial population dynamics. Papers of the Regional Science Association 36:3–34.

Rogers, A., and F. Willekens (1976b) The Spatial Reproductive Value and the Spatial Momentum of Zero Population Growth. RM-76-81. Laxenburg, Austria: International Institute for Applied Systems Analysis. Also published in Environment and Planning A 10(1978):503–518.

# AGGREGATION OF POPULATION PROJECTION MODELS

*Robert Gibberd*

## 1 INTRODUCTION

The issue of aggregation arises frequently in the study of dynamic systems. This is particularly true in mathematical demography, where the analyst is concerned with measuring and projecting flows from one state, such as region of residence, or marital or employment status, to another. When modeling any demographic system it is necessary to consciously or unconsciously aggregate over some of the above variables in order to obtain a manageable model. The aggregation is usually performed with respect to the variables which are of no immediate interest to the particular problem being studied. For example, when performing projections by region, age, and sex, the variables marital status, employment, and education level are frequently ignored, despite the fact that migration rates are related to these variables. Similarly, projections for household formation may neglect region of residence or occupation. Whether ignoring these extra variables produces an important bias in the results is not known, although it is recognized that erroneous results are obtained if the variables age or sex are aggregated. Now that multistate demographic models are being used to provide projections at very fine levels of detail, there is a growing need to understand the effect of aggregation on the results of these models.

Demographic projection models are often written as linear, discrete, time-invariant systems of the form

$$\boldsymbol{K}(t+1) = \mathbf{G}\boldsymbol{K}(t) \tag{1}$$

where both $\mathbf{G}$, an $n \times n$ matrix with non-negative elements, and the initial $n$-dimensional population vector $\boldsymbol{K}(0)$ are assumed to be known. Equation (1) exhibits the Markovian property (or assumption), which states that the population vector $\boldsymbol{K}(t+1)$ is determined entirely in terms of the vector $\boldsymbol{K}(t)$. This property greatly simplifies the use of the model and the empirical calculation of the elements of the transition matrix $\mathbf{G}$.

From the $n$-dimensional population vector $\boldsymbol{K}(t)$ it is possible to form an $m$-dimensional vector $\hat{\boldsymbol{K}}(t)$ by partitioning the $n$ original elements into a new set of $m$ elements. A property of aggregated population vectors is that if a linear, time-invariant equation is used to model $\hat{\boldsymbol{K}}(t)$, as given in eqn. (2)

$$\hat{\boldsymbol{K}}(t+1) = \hat{\mathbf{G}}\hat{\boldsymbol{K}}(t) \tag{2}$$

then the vector $\hat{\boldsymbol{K}}(t)$ will generally be inconsistent with $\boldsymbol{K}(t)$. The special case where $\hat{\boldsymbol{K}}(t)$ does obey eqn. (2) for all time intervals and initial population vectors, and $\hat{\boldsymbol{K}}(t)$ is consistent with $\boldsymbol{K}(t)$ has been denoted "perfect aggregation" (Rogers 1976).

The concept of aggregation of multistate systems has been used in input–output analysis in economics (Hatanaka 1952; McManus 1956; Ara 1959; Simon and Ando 1961; Ando and Fisher 1963; Fisher 1969), in control theory for dynamic systems (Aoki 1968; Luenberger 1978), in Markov-chain theory (Burke and Rosenblatt 1958; Kemeny and Snell 1960; Lewis 1979), and in demographic analysis (Keyfitz 1972; Rogers 1969, 1971, 1975, 1976; Rogers and Philipov 1979).

The aggregation problem in demographic analysis was first noted by Rogers (1969) and was studied with regard to matrix cohort-survival models of interregional population growth. The fact that conditions for perfect aggregation were not likely to be met in practice was emphasized, and Rogers suggested that aggregation procedures which gave the "best possible" accuracy rather than "perfect" accuracy should be considered. In his most recent article on the subject, Rogers (1976) showed the need to develop improved methods for reducing large-scale population-projection models by considering a combination of aggregation and decomposition techniques.

One of the simplest examples of aggregation error was illustrated by Keyfitz (1972), who considered two independent populations with different growth rates. Keyfitz showed that if both populations were combined and an average growth rate were applied to the total population, the sum of the separate projections was always greater than the projection for the combined population. For more general results and a good summary of the aggregation problem see the article by Rogers (1976).

This paper develops a formalism for determining the general relationship between an $n$-dimensional population vector and the corresponding $m$-dimensional aggregated vector. The following main results will be presented. Firstly, that an aggregated population vector generally obeys a non-Markovian equation of the form

$$\hat{\boldsymbol{K}}(t+1)=\hat{\mathbf{G}}_0\hat{\boldsymbol{K}}(t)+\hat{\mathbf{G}}_1\hat{\boldsymbol{K}}(t-1)+\hat{\mathbf{G}}_2\hat{\boldsymbol{K}}(t-2)+\cdots \tag{3}$$

Secondly, that after a large number of time intervals, the above non-Markovian equation can be replaced by an almost equivalent Markovian equation, with the aggregated transition matrix having the same maximum eigenvalue and corresponding eigenvector as $\mathbf{G}$. Thirdly, Luenberger's use of upper and lower aggregated transition matrices is employed to estimate the error caused by aggregation. The paper concludes with some comments on the "best" aggregation procedure.

## 2 TYPES OF AGGREGATION

If the population vector $\boldsymbol{K}(t)$ is assumed to obey eqn. (1), there are three possible ways to aggregate this linear model

(a) aggregate with respect to time only
(b) aggregate the elements of the vector $\boldsymbol{K}(t)$
(c) aggregate the vector elements and time simultaneously

When eqn. (1) is aggregated with respect to time, the demographer is usually concerned with going from, say, a one-year interval to a five-year interval. This can be achieved by writing eqn. (1) as

$$\boldsymbol{K}(t+5)=\mathbf{G}^5\boldsymbol{K}(t) \qquad (t=0,5,10,\ldots) \tag{4}$$

The dimension of the population vector is unaffected and the model clearly retains the Markovian property, with the transition matrix now being taken as $\mathbf{G}^5$. This result can be generalized to any time interval and illustrates that aggregation with respect to time still preserves the linear, time-invariant, Markovian properties of the model.

The second kind of aggregation occurs when a vector $\boldsymbol{K}(t)$ of dimension $n$ is aggregated to an $m$-dimensional vector $\hat{\boldsymbol{K}}(t)$ $(m < n)$. This can be represented formally by the equation

$$\hat{\boldsymbol{K}}(t) = \mathbf{C}\boldsymbol{K}(t) \tag{5}$$

where $\mathbf{C}$ is an $m \times n$ dimensional matrix, and generally has only one non-zero element in each column. Rogers (1975) has called $\mathbf{C}$ the consolidation matrix.

Consider, as an example, a three-region population vector where

$$\boldsymbol{K}(t) = \begin{pmatrix} k_1(t) \\ k_2(t) \\ k_3(t) \end{pmatrix}$$

If the populations for regions 1 and 2 are to be combined, then

$$\mathbf{C} = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

and

$$\hat{\boldsymbol{K}}(t) = \mathbf{C}\boldsymbol{K}(t) = \begin{pmatrix} k_1(t) + k_2(t) \\ k_3(t) \end{pmatrix}$$

In the general case, we will write the vector $\boldsymbol{K}(t)$ as a partitioned vector

$$\boldsymbol{K}(t) = \begin{bmatrix} \boldsymbol{K}_1(t) \\ \boldsymbol{K}_2(t) \\ \vdots \\ \boldsymbol{K}_m(t) \end{bmatrix}$$

where the $\boldsymbol{K}_i(t)$ are vectors whose elements are going to be summed to form the $i$th element of the aggregated vector $\hat{\boldsymbol{K}}(t)$.

The consolidation matrix $\mathbf{C}$ will be partitioned as

$$\mathbf{C} = \begin{bmatrix} \boldsymbol{c}_1 & 0 & \cdots & 0 \\ 0 & \boldsymbol{c}_2 & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & & \boldsymbol{c}_m \end{bmatrix}$$

where $\boldsymbol{c}_i$ is a row vector of ones. Similarly, the transition matrix $\mathbf{G}$ can be partitioned as

$$\mathbf{G} = \begin{bmatrix} \mathbf{G}_{11} & \mathbf{G}_{12} & \cdots & \mathbf{G}_{1m} \\ \mathbf{G}_{21} & \mathbf{G}_{22} & \cdots & \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{G}_{m1} & & \cdots & \mathbf{G}_{mm} \end{bmatrix}$$

In the example considered above

$$\boldsymbol{K}_1(t) = \begin{pmatrix} k_1(t) \\ k_2(t) \end{pmatrix}$$

$$\boldsymbol{K}_2(t) = k_3(t)$$

$$\boldsymbol{c}_1 = (1 \quad 1)$$

$$\boldsymbol{c}_2 = 1$$

$$\mathbf{G}_{11} = \begin{pmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{pmatrix}$$

$$\mathbf{G}_{12} = \begin{pmatrix} g_{13} \\ g_{23} \end{pmatrix}$$

$$\mathbf{G}_{21} = (g_{31} \quad g_{32})$$

$$\mathbf{G}_{22} = g_{33}$$

This notation will be used extensively in the remaining sections.

In demography it is often desirable to aggregate the population vector in conjunction with a change in the unit time interval. A common example is the Leslie projection model using one-year cohorts, where the Leslie matrix **G** consists of the transition rates from one year to the next. A natural simplification is to aggregate adjacent state variables forming five-year cohort classes while also extending the basic time period from one year to five years, so that the time step is equal to the cohort time width. Mathematically, this can be seen as equivalent to performing a time aggregation and then aggregating the population vector. That is, it is necessary to aggregate eqn. (4) with respect to the elements of $\boldsymbol{K}(t)$ only. Hence, in principle, aggregating with respect to time and vector elements together is no different to aggregating with respect to vector elements alone.

## 3 DERIVATION OF NON-MARKOVIAN EQUATIONS FOR AN AGGREGATED SYSTEM

In this section, an equation for $\hat{\boldsymbol{K}}(t+1)$ is derived assuming that $\boldsymbol{K}(t+1)$ obeys eqn. (1). Equations (1) and (5) give the equation

$$\hat{\boldsymbol{K}}(t+1) = \mathbf{C}\boldsymbol{K}(t+1) = \mathbf{C}\mathbf{G}\boldsymbol{K}(t) \tag{6}$$

To express the right-hand side of eqn. (6) in terms of aggregated vectors, we construct a projection operator, **P**, defined by

$$\mathbf{P} = \mathbf{C}^{+}\mathbf{C} \tag{7}$$

where

$$\mathbf{C}^{+} = \mathbf{W}'(\mathbf{C}\mathbf{W}')^{-1} \tag{8}$$

$\mathbf{W}'$ is an $n \times m$ matrix which can be partitioned as follows

$$\mathbf{W}' = \begin{bmatrix} \boldsymbol{W}_1 & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{W}_2 & \cdots & \\ \vdots & & \ddots & \\ \boldsymbol{0} & \cdots & & \boldsymbol{W}_m \end{bmatrix} \tag{9}$$

and the $\boldsymbol{W}_i$ are column vectors. A particular example of $\mathbf{W}'$ is the transpose of $\mathbf{C}$, denoted by $\mathbf{C}'$. The matrix $\mathbf{CW}'$ has dimensions $m \times m$ and a sufficient condition for its inverse to exist is that $\boldsymbol{W}_i$ $(i = 1, 2, \ldots, m)$ contains no negative elements and is not a null vector.

We define the operator $\mathbf{Q}$ by

$$\mathbf{Q} = \mathbf{I} - \mathbf{P} \tag{10}$$

where $\mathbf{I}$ is the identity matrix and eqns. (7)–(10) can be used to show that $\mathbf{P}$ and $\mathbf{Q}$ are idempotent operators satisfying the conditions

$$\mathbf{P}^2 = \mathbf{P} \qquad \mathbf{Q}^2 = \mathbf{Q} \qquad \mathbf{PQ} = \mathbf{QP} = \mathbf{0} \tag{11}$$

Using these operators, eqns. (1) and (6) can be written as

$$\mathbf{C}\boldsymbol{K}(t+1) = \mathbf{CG}\boldsymbol{K}(t) = \mathbf{CGP}\boldsymbol{K}(t) + \mathbf{CGQ}\boldsymbol{K}(t) \tag{12}$$

$$\mathbf{Q}\boldsymbol{K}(t+1) = \mathbf{QG}\boldsymbol{K}(t) = \mathbf{QGP}\boldsymbol{K}(t) + \mathbf{QGQ}\boldsymbol{K}(t) \tag{13}$$

Continued substitution for $\mathbf{Q}\boldsymbol{K}(t)$ in eqn. (12), using eqn. (13), gives

$$\begin{aligned} \mathbf{C}\boldsymbol{K}(t+1) = {} & \mathbf{CGP}\boldsymbol{K}(t) + \mathbf{CGQ} \cdot \mathbf{QGP}\boldsymbol{K}(t-1) \\ & + \mathbf{CGQGQGP}\boldsymbol{K}(t-2) + \cdots \\ & + \mathbf{CGQ}(\mathbf{QGQ})^{t-1}\mathbf{QGP}\boldsymbol{K}(0) \\ & + \mathbf{CGQ}(\mathbf{QGQ})^t\boldsymbol{K}(0) \end{aligned} \tag{14}$$

This equation can be written as

$$\begin{aligned} \hat{\boldsymbol{K}}(t+1) = {} & \hat{\mathbf{G}}_0\hat{\boldsymbol{K}}(t) + \hat{\mathbf{G}}_1\hat{\boldsymbol{K}}(t-1) + \hat{\mathbf{G}}_2\hat{\boldsymbol{K}}(t-2) + \cdots \\ & + \hat{\mathbf{G}}_t\hat{\boldsymbol{K}}(0) + \mathbf{G}_{t+1}\boldsymbol{K}(0) \end{aligned} \tag{15}$$

where

$$\hat{\mathbf{G}}_0 = \mathbf{CGC}^+$$

$$\hat{\mathbf{G}}_k = \mathbf{CGQ}(\mathbf{QGQ})^{k-1}\mathbf{QGC}^+ \qquad (k = 1, 2, \ldots, t)$$

$$\mathbf{G}_{t+1} = \mathbf{CGQ}(\mathbf{QGQ})^t$$

Equation (15) is clearly a non-Markovian equation, satisfied by the aggregated population vector $\hat{\boldsymbol{K}}(t+1)$. A special case occurs when $\mathbf{CGQ}$ is a null matrix. In this case, eqn. (15) reduces to an exact Markovian equation, namely

$$\hat{\boldsymbol{K}}(t+1) = \mathbf{CGC}^+\hat{\boldsymbol{K}}(t) = \hat{\mathbf{G}}_0\hat{\boldsymbol{K}}(t) \tag{16}$$

Population models which satisfy eqn. (16) are of the special type studied by Burke and Rosenblatt (1958), Kemeny and Snell (1960), and Rogers (1968, 1975). Kemeny and Snell used the term "lumpable" to describe a Markovian chain with the property that an aggregated process is also a Markovian chain. Rogers recognized a similar situation in the modeling of demographic systems, and used the term "perfect aggregation." It is possible to show that the conditions for "lumpability" and "perfect aggregation" are equivalent to the condition $\mathbf{CGQ} = \mathbf{0}$, and this condition will be discussed in more detail in Section 4.

For most systems, $\mathbf{CGQ} \neq \mathbf{0}$, and hence if a Markovian assumption is valid for a certain population vector $\boldsymbol{K}(t)$, then a coarser or aggregated classification using $\hat{\boldsymbol{K}}(t)$ will no longer be Markovian; extra terms, called "memory" terms, will need to be included in the equation. This is a disturbing result since almost all demographic accounting models assume that the Markovian assumption is valid. By neglecting the "memory" terms, an error or bias in the results will occur. However, because of the complicated nature of the expressions given in eqn. (15), it would be unrealistic to expect future models to use the non-Markovian equation. What is required is to choose $\mathbf{W}'$ such that the "memory" terms are minimized, and eqn. (15) can be approximated by

$$\hat{\boldsymbol{K}}(t+1) \simeq \hat{\mathbf{G}}_0 \hat{\boldsymbol{K}}(t) \tag{17}$$

In the next three sections five alternative formulations of $\hat{\mathbf{G}}_0$ are considered; in Section 7 of the paper, these alternatives are applied to a numerical example.

## 4 ASYMPTOTIC APPROXIMATION TO THE NON-MARKOVIAN EQUATION

The stable or equilibrium solution of eqn. (1) is often of interest, since the long-term behavior of $\boldsymbol{K}(t)$ magnifies any trends or spatial changes that are occurring in the system. In this section it will be shown that the long-term behavior of an aggregated system can be closely approximated by a Markovian equation with a transition matrix that has the same maximum eigenvalue as the original transition matrix. The derivation of this result is given in detail by Lewis (1979), and is only briefly derived here.

For the case when the eigenvalues of $\mathbf{G}$ are distinct, the population vector $\boldsymbol{K}(t)$ $(t = 0, 1, 2, \ldots, m)$ can be written as

$$\boldsymbol{K}(t) = \lambda_1^t \left[ \boldsymbol{u}_1 a_1 + \boldsymbol{u}_2 a_2 \left(\frac{\lambda_2}{\lambda_1}\right)^t + \cdots + \boldsymbol{u}_n a_n \left(\frac{\lambda_n}{\lambda_1}\right)^t \right] \tag{18}$$

where $\boldsymbol{u}_i$ is the right eigenvector of $\mathbf{G}$ corresponding to eigenvalue $\lambda_i$, and $a_i$ is a constant $(i = 1, 2, \ldots, n)$. It is assumed that $\lambda_1$ is the eigenvalue with maximum absolute value, and hence, for $i = 2, 3, \ldots, n$

$$\lim_{t \to \infty} \left(\frac{\lambda_i}{\lambda_1}\right)^t = 0$$

Consequently, for large $t$, a first approximation to eqn. (18) is

$$\boldsymbol{K}(t) \simeq \lambda_1^t \boldsymbol{u}_1 a_1$$

If we rewrite eqn. (14) as

$$\mathbf{C}\boldsymbol{K}(t+1)=\mathbf{CGP}\boldsymbol{K}(t)+\mathbf{CGQ}\sum_{i=0}^{t-1}(\mathbf{QGQ})^{i}\mathbf{QGP}\boldsymbol{K}(t-1-i)$$
$$+\mathbf{CGQ}(\mathbf{QGQ})^{t}\boldsymbol{K}(0) \tag{19}$$

Then, for large $t$

$$\sum_{i=0}^{t-1}(\mathbf{QGQ})^{i}\mathbf{QGP}\boldsymbol{K}(t-1-i)\simeq\sum_{i=0}^{\infty}(\mathbf{QGQ})^{i}\mathbf{QGP}\lambda_1^{t-1-i}\boldsymbol{u}_1 a_1$$
$$\simeq\lambda_1^{-1}\sum_{i=0}^{\infty}(\mathbf{QGQ}/\lambda_1)^{i}\mathbf{QGP}\boldsymbol{K}(t) \tag{20}$$

Provided the series $\sum_{i=0}^{\infty}(\mathbf{QGQ}/\lambda_1)^i$ converges, it can be replaced by the expression $[\mathbf{Q}-(\mathbf{QGQ}/\lambda_1)]^{-1}$.

Equation (20) becomes, for large $t$

$$\sum_{i=0}^{t-1}(\mathbf{QGQ})^{i}\mathbf{QGP}\boldsymbol{K}(t-1-i)\simeq\lambda_1^{-1}[\mathbf{Q}-(\mathbf{QGQ}/\lambda_1)]^{-1}\mathbf{QGP}\boldsymbol{K}(t)$$
$$=(\lambda_1\mathbf{Q}-\mathbf{QGQ})^{-1}\mathbf{QGP}\boldsymbol{K}(t)$$

Thus, eqn. (19) can be approximated as

$$\mathbf{C}\boldsymbol{K}(t+1)\simeq\mathbf{CGP}\boldsymbol{K}(t)+\mathbf{CGQ}(\lambda_1\mathbf{Q}-\mathbf{QGQ})^{-1}\mathbf{QGP}\boldsymbol{K}(t)$$
$$+\mathbf{CGQ}(\mathbf{QGQ})^{t}\boldsymbol{K}(0)$$

or

$$\hat{\boldsymbol{K}}(t+1)\approx\hat{\mathbf{G}}_{\mathrm{as}}\hat{\boldsymbol{K}}(t)$$

where

$$\hat{\mathbf{G}}_{\mathrm{as}}=\mathbf{CGC}^{+}+\mathbf{CGQ}(\lambda_1\mathbf{Q}-\mathbf{QGQ})^{-1}\mathbf{QGC}^{+} \tag{21}$$

and the term $\mathbf{CGQ}(\mathbf{QGQ})^{t}\boldsymbol{K}(0)$ has been neglected since it becomes relatively small as $t$ becomes large.

The maximum eigenvalue of $\mathbf{G}$, $\lambda_1$, and the corresponding eigenvector $\boldsymbol{u}_1$ satisfy the relation

$$\mathbf{G}\boldsymbol{u}_1=\lambda_1\boldsymbol{u}_1$$

It can be shown by substitution that, if $\hat{\mathbf{G}}_{\mathrm{as}}$ is defined by eqn. (21), then

$$\hat{\mathbf{G}}_{\mathrm{as}}\hat{\boldsymbol{u}}_1=\lambda_1\hat{\boldsymbol{u}}_1$$

Thus, $\hat{\mathbf{G}}_{\mathrm{as}}$ has the corresponding maximum eigenvalue and eigenvector of $\mathbf{G}$, guaranteeing that eqns. (1) and (21) are consistent for large $t$.

The asymptotic properties of an aggregated system can be illustrated by considering a simple example, the $2\times 2$ Leslie matrix in which the two age groups are aggregated into one age group

$$\begin{pmatrix}k_1(t+1)\\k_2(t+1)\end{pmatrix}=\begin{pmatrix}b_1 & b_2\\s_1 & 0\end{pmatrix}\begin{pmatrix}k_1(t)\\k_2(t)\end{pmatrix} \tag{22}$$

The aggregated population vector is $\hat{\boldsymbol{K}}(t) = k_1(t) + k_2(t)$ and hence

$$\mathbf{C} = (1 \quad 1)$$

We then choose $\mathbf{W}'$ such that $(\mathbf{CW}')$ is the identity matrix, that is

$$\mathbf{W}' = \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \mathbf{C}^+$$

Matrix algebra then yields the following results

$$\mathbf{P} = \mathbf{W}'(\mathbf{CW}')^{-1}\mathbf{C} = \mathbf{W}'\mathbf{C} = \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}$$

$$\mathbf{Q} = \begin{pmatrix} 0 & -1 \\ 0 & 1 \end{pmatrix}$$

$$\mathbf{CGC}^+ = b_1 + s_1$$

$$\mathbf{CGQ} = (0 \quad b_2 - b_1 - s_1)$$

$$\mathbf{CGQ}(\mathbf{QGQ})^k\mathbf{QGC}^+ = -(b_2 - b_1 - s_1)(-s_1)^{k+1}$$

Substituting into eqn. (15) gives the following equation for the total population $\hat{\boldsymbol{K}}(t+1)$

$$\begin{aligned}\hat{\boldsymbol{K}}(t+1) &= (b_1 + s_1)\hat{\boldsymbol{K}}(t) \\ &\quad + (b_2 - b_1 - s_1)[s_1\hat{\boldsymbol{K}}(t-1) - s_1^2\hat{\boldsymbol{K}}(t-2) + s_1^3\hat{\boldsymbol{K}}(t-3) - \cdots] \\ &\quad + (-s)^t(b_2 - b_1 - s_1)k_2(0)\end{aligned} \tag{23}$$

It can be seen that, if $b_2 = b_1 + s_1$, then eqn. (23) becomes Markovian. That is

$$\hat{\boldsymbol{K}}(t+1) = (b_1 + s_1)\hat{\boldsymbol{K}}(t) = b_2\hat{\boldsymbol{K}}(t)$$

This corresponds to the case $\mathbf{CGQ} = 0$ or "perfect aggregation" as discussed earlier.

In the case when $b_1 + s_1 \neq b_2$, we assume that the population has a constant growth rate $r$ for large $t$. Then

$$\hat{\boldsymbol{K}}(t) \simeq r^n\hat{\boldsymbol{K}}(t-n) \tag{24}$$

and substituting in eqn. (23) gives

$$\begin{aligned}\hat{\boldsymbol{K}}(t+1) &\simeq (b_1 + s_1)\hat{\boldsymbol{K}}(t) \\ &\quad + \frac{s_1(b_2 - b_1 - s_1)}{r}\left[1 - \left(\frac{s_1}{r}\right) + \left(\frac{s_1}{r}\right)^2 - \cdots\right]\hat{\boldsymbol{K}}(t) \\ &\simeq \left[b_1 + s_1 + \frac{s_1(b_2 - b_1 - s_1)}{r + s_1}\right]\hat{\boldsymbol{K}}(t)\end{aligned} \tag{25}$$

However, for eqns. (24) and (25) to be consistent, we require

$$b_1 + s_1 + \frac{s_1(b_2 - b_1 - s_1)}{r + s_1} = r$$

That is

$$r = \frac{b_1 \pm \sqrt{(b_1^2 + 4s_1b_2)}}{2}$$

However, these values of $r$ are the same as the eigenvalues of the original Leslie matrix, and hence the aggregated eqn. (23) becomes approximately Markovian for large $t$ and can be approximated by

$$\hat{\boldsymbol{K}}(t+1)=\left[\frac{b_1+\sqrt{(b_1^2+4s_1b_2)}}{2}\right]\hat{\boldsymbol{K}}(t)$$

This is a particular case of the general equation derived in eqn. (17).

The matrix $\hat{\mathbf{G}}_{\mathrm{as}}$ depends on the choice of the matrix $\mathbf{W}'$, and two possible alternative choices for the above example are

$$\begin{pmatrix}0\\1\end{pmatrix} \quad \text{or} \quad \begin{pmatrix}\frac{1}{2}\\\frac{1}{2}\end{pmatrix}$$

Although the non-Markovian equation corresponding to eqn. (23) will now be different, the property that $\hat{\mathbf{G}}_{\mathrm{as}}$ has the maximum eigenvalue and eigenvector of $\mathbf{G}$ is invariant to the choice of $\mathbf{W}'$.

The implications for demographic modeling when studying long-term behavior are that the transition matrix $\hat{\mathbf{G}}_{\mathrm{as}}$ should be used. However, as discussed in Section 8, this may not be computationally feasible for very large systems.

## 5 UPPER AND LOWER AGGREGATION MATRICES

Luenberger (1978) introduced the idea of developing two aggregated systems: one serving as an upper bound and the other as a lower bound. To construct the upper and lower aggregated systems, it is necessary to determine in detail what "perfect aggregation" requires. If the two population vectors $\boldsymbol{K}(t+1)$ and $\hat{\boldsymbol{K}}(t+1)$ are consistent, then the equations

$$\mathbf{C}\boldsymbol{K}(t+1)=\mathbf{C}\mathbf{G}\boldsymbol{K}(t)$$

$$\mathbf{C}\boldsymbol{K}(t+1)=\hat{\boldsymbol{K}}(t+1)=\hat{\mathbf{G}}\hat{\boldsymbol{K}}(t)=\hat{\mathbf{G}}\mathbf{C}\boldsymbol{K}(t)$$

should be consistent (Rogers 1975). For this to be true for all vectors $\boldsymbol{K}(t)$, we require

$$\mathbf{C}\mathbf{G}=\hat{\mathbf{G}}\mathbf{C} \tag{26}$$

Writing eqn. (26) in terms of the submatrices $\mathbf{G}_{ij}$ and vectors $\boldsymbol{c}_{ii}$ gives

$$\boldsymbol{c}_{ii}\mathbf{G}_{ij}=\hat{\mathbf{G}}_{ij}\boldsymbol{c}_{jj} \qquad (i,j=1,2,\ldots,m) \tag{27}$$

Since the $\boldsymbol{c}_{ii}$ are row vectors with each element being equal to one, eqn. (27) is only true when the column sums of each submatrix $\mathbf{G}_{ij}$ are equal to the element $\hat{\mathbf{G}}_{ij}$. If this is true for all $i, j=1,2,\ldots,m$ then the system can be aggregated without producing any error or bias. Alternatively, if the column sums of any $\mathbf{G}_{ij}$ are not equal, then eqns. (26) and (27) cannot be satisfied. To construct the upper aggregated transition matrix $\hat{\mathbf{G}}_{\mathrm{u}}$, we define the elements of $[\hat{\mathbf{G}}_{\mathrm{u}}]_{ij}$ as equal to the maximum column sum of $\mathbf{G}_{ij}$ $(i,j=1,2,\ldots,m)$. Then it can be seen that

$$\hat{\mathbf{G}}_{\mathrm{u}}\mathbf{C}\geqslant\mathbf{C}\mathbf{G}$$

and hence, that [for non-negative $\boldsymbol{K}(t)$]

$$\hat{\boldsymbol{K}}_{\mathrm{u}}(t+1)=\hat{\mathbf{G}}_{\mathrm{u}}\mathbf{C}\boldsymbol{K}(t)\geqslant\mathbf{C}\mathbf{G}\boldsymbol{K}(t)=\mathbf{C}\boldsymbol{K}(t+1)=\hat{\boldsymbol{K}}(t+1)$$

where the inequality refers to the components of the matrix equations above.

Similarly, a lower aggregated transition matrix $\hat{\mathbf{G}}_l$ is defined as having elements $[\hat{\mathbf{G}}_l]_{ij}$ equal to the minimum column sum of $\mathbf{G}_{ij}$ $(i, j = 1, 2, \ldots, m)$. Thus

$$\hat{\mathbf{G}}_l\mathbf{C} \leq \mathbf{CG}$$

and

$$\hat{\boldsymbol{K}}_l(t+1) = \hat{\mathbf{G}}_l\mathbf{C}\boldsymbol{K}(t) \leq \mathbf{CG}\boldsymbol{K}(t) = \hat{\boldsymbol{K}}(t+1)$$

The transition matrices $\hat{\mathbf{G}}_u$ and $\hat{\mathbf{G}}_l$ can be easily constructed from the matrix $\mathbf{G}$, and when applied $t$ times to the initial population vector $\hat{\boldsymbol{K}}(0)$ will yield upper and lower bounds for the vector $\hat{\boldsymbol{K}}(t)$. A numerical example is given in Section 7.

## 6 ALTERNATIVE AGGREGATION MATRICES

In Section 5 it was shown how upper and lower aggregated systems could be obtained. We now consider the aggregated systems whose transition matrix $\hat{\mathbf{G}}_a$ is obtained by defining $[\hat{\mathbf{G}}_a]_{ij}$ as the unweighted or weighted mean of the column sums of $\mathbf{G}_{ij}$. Since the mean must lie between the maximum and minimum values, it can be shown that

$$\hat{\mathbf{G}}_l \leq \hat{\mathbf{G}}_a \leq \hat{\mathbf{G}}_u$$

and hence that the population vectors obtained from $\hat{\mathbf{G}}_a$ lie between the upper and lower vectors obtained from $\hat{\mathbf{G}}_u$ and $\hat{\mathbf{G}}_l$.

To derive the matrices $\hat{\mathbf{G}}_a$, recall that we need to find $\hat{\mathbf{G}}$ such that

$$\mathbf{CG} = \hat{\mathbf{G}}\mathbf{C} \tag{26}$$

This is an $m \times n$ matrix equation with $m^2$ unknown parameters $\hat{\mathbf{G}}_{ij}$ $(i, j = 1, 2, \ldots, m)$. As there are more equations than unknowns, there is, in general, no exact solution. However, a least-squares solution can be obtained from the generalized inverse of $\mathbf{C}$, defined as $\mathbf{C}^+$ in eqn. (8). The matrix $\mathbf{W}'$ provides the weights for the least-squares solution. Multiplying eqn. (26) by $\mathbf{C}^+$ gives

$$\hat{\mathbf{G}}_a = \mathbf{CGC}^+ = \mathbf{CGW}'(\mathbf{CW}')^{-1} \tag{28}$$

The unweighted least-squares solution is obtained by defining $\mathbf{W}'$ as the transpose of $\mathbf{C}$, and hence

$$\hat{\mathbf{G}}_{l.s.} = \mathbf{CGC}'(\mathbf{CC}')^{-1}$$

It can be shown that the $i, j$ element of $\hat{\mathbf{G}}_{l.s.}$ is the arithmetic mean of the column sums of $\mathbf{G}_{ij}$, and hence represents the least-squares solution to eqn. (26).

Since the elements of $\mathbf{G}_{ij}$ are rates (such as migration, mortality, and fertility rates), it would appear more appropriate to use a weighted average, where the weights are proportional to some population distribution. There are two obvious choices: namely, that the weights should be proportional either to the initial population vector $\boldsymbol{K}(0)$ or to the stable population vector $\boldsymbol{u}_1$. If the vectors $\boldsymbol{W}_i$ are set equal to the corresponding elements of $\boldsymbol{u}_1$, then it can be shown that

$$\mathbf{W}'(\mathbf{CW}')^{-1}\mathbf{C}\boldsymbol{u}_1 = \boldsymbol{u}_1$$

and hence it follows that $\hat{\mathbf{G}}_{eq} = \mathbf{CGC}^+$ has an eigenvalue $\lambda_1$ and eigenvector $\mathbf{C}\boldsymbol{u}_1$. This is the same property that the matrix $\hat{\mathbf{G}}_{as}$ was shown to have in the earlier section on asymptotic approximation, and hence both matrices will describe the asymptotic behavior of the aggregated system correctly.

The most pragmatic weighting is to construct an aggregated matrix using the initial population vector $\boldsymbol{K}(0)$. Defining the weighting matrix $\mathbf{W}'$ in terms of $\boldsymbol{K}(0)$ gives the following expression for the components of the aggregated matrix $\hat{\mathbf{G}}_0$

$$(\hat{\mathbf{G}}_0)_{ij} = \sum_l{}' g_{il}k_l(0) \Big/ \sum_l{}' k_l(0) \tag{29}$$

where the summation over $l$ is restricted to those elements that are summed to form the $j$th component of the aggregated system. With such a choice for $\mathbf{W}$ it can be shown that

$$\mathbf{W}'(\mathbf{C}\mathbf{W}')^{-1}\mathbf{C}\boldsymbol{K}(0) = \boldsymbol{K}(0)$$

and hence that

$$\hat{\boldsymbol{K}}(1) = \hat{\mathbf{G}}_0\hat{\boldsymbol{K}}(0) = \mathbf{C}\mathbf{G}\mathbf{W}'(\mathbf{C}\mathbf{W}')^{-1}\mathbf{C}\boldsymbol{K}(0) = \mathbf{C}\mathbf{G}\boldsymbol{K}(0) = \mathbf{C}\boldsymbol{K}(1)$$

Thus, the matrix $\hat{\mathbf{G}}_0$ gives consistent answers for the vectors $\hat{\boldsymbol{K}}(1)$ and $\boldsymbol{K}(1)$ and it can be assumed to be reasonably consistent for small values of $t$. Three other aspects to note are as follows. Firstly, in this case, $\mathbf{W}'(\mathbf{C}\mathbf{W}')^{-1}\mathbf{C}$ is equivalent to the deconsolidation operator $\mathbf{D}(t)$ for $t = 0$ used by Rogers (1975). Secondly, the matrix $\hat{\mathbf{G}}_0$ as defined above is equivalent to the observed transition matrix for the aggregated system in the time period $t = 0$ to $t = 1$. And finally, $\hat{\mathbf{G}}_0$ will be a better approximation, the closer $\boldsymbol{K}(0)$ is to the equilibrium vector $\boldsymbol{u}_1$.

## 7 NUMERICAL EXAMPLE

The results derived in Sections 3–6 will now be illustrated in an example using interstate migration data for Australia during the period 1966–1971. The data which form the basis of the migration model are given in Table 1 in the form of a migration matrix of total movers and stayers. The population vector is eight dimensional, representing the population in the eight states and territories. Note that the total population remains constant, since births, deaths, and migration into and out of Australia have been ignored. For this reason, the vector $\boldsymbol{K}(t)$ will be normalized to give the proportion of people in each state.

The transition matrix $\mathbf{G}$ is obtained from the migration matrix in Table 1 by dividing each element of a column by the sum of the elements in that column, as shown in Table 2.

To illustrate the properties of aggregated models, the population vector $\boldsymbol{K}(t)$ will be reduced to a 3-dimensional vector by combining regions 3–8 into a single region. Also, it is assumed that the 8-dimensional population vector satisfies eqn. (1).

The expression for the consolidation matrix $\mathbf{C}$ is given by

$$\mathbf{C} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix}$$

TABLE 1 Internal migration between Australian states and territories[a], 1966–1971[b].

| Destination in 1971 | Origin in 1966 N.S.W. | Vic. | Qld. | S.A. | W.A. | Tas. | A.C.T. | N.T. | Total |
|---|---|---|---|---|---|---|---|---|---|
| N.S.W. | 3,493,601 | 44,759 | 38,621 | 13,131 | 8,179 | 4,862 | 8,227 | 2,374 | 3,613,754 |
| Vic. | 37,813 | 2,672,388 | 15,949 | 17,139 | 7,649 | 8,764 | 2,453 | 1,448 | 2,763,603 |
| Qld. | 46,033 | 27,124 | 1,341,316 | 6,021 | 3,971 | 2,998 | 1,653 | 2,555 | 1,431,671 |
| S.A. | 12,158 | 13,937 | 4,148 | 897,411 | 3,370 | 1,544 | 755 | 2,041 | 935,364 |
| W.A. | 14,104 | 15,496 | 6,285 | 9,248 | 696,630 | 2,199 | 812 | 1,540 | 746,314 |
| Tas. | 3,848 | 6,399 | 1,641 | 1,553 | 814 | 302,638 | 196 | 146 | 317,235 |
| A.C.T. | 22,211 | 7,205 | 3,684 | 2,622 | 1,261 | 721 | 68,557 | 539 | 106,800 |
| N.T. | 4,148 | 3,143 | 4,640 | 5,106 | 1,876 | 319 | 323 | 33,043 | 52,598 |
| Total | 3,633,916 | 2,790,451 | 1,416,284 | 952,231 | 723,750 | 324,045 | 82,976 | 43,686 | 9,967,339 |

[a] Abbreviations used are as follows: N.S.W., New South Wales; Vic., Victoria; Qld., Queensland; S.A., South Australia; W.A., Western Australia; Tas., Tasmania; A.C.T., Australian Capital Territory; N.T., Northern Territories.
[b] Taken from Rowland (1976), p. 25, Table 3.5.

TABLE 2 Transition matrix[a] **G** for an 8-state Markov chain[b].

| | N.S.W. | Vic. | Qld. | S.A. | W.A. | Tas. | A.C.T. | N.T. |
|---|---|---|---|---|---|---|---|---|
| N.S.W. | 0.961387 | 0.016040 | 0.027269 | 0.013790 | 0.011301 | 0.015004 | 0.099149 | 0.054342 |
| Vic. | 0.010406 | 0.957690 | 0.011261 | 0.017999 | 0.010569 | 0.027046 | 0.029563 | 0.033146 |
| Qld. | 0.012668 | 0.009720 | 0.947067 | 0.006323 | 0.005487 | 0.009252 | 0.019921 | 0.058486 |
| S.A. | 0.003346 | 0.004995 | 0.002929 | 0.942430 | 0.004656 | 0.004765 | 0.009099 | 0.046720 |
| W.A. | 0.003881 | 0.005553 | 0.004438 | 0.009712 | 0.962528 | 0.006786 | 0.009786 | 0.035252 |
| Tas. | 0.001059 | 0.002293 | 0.001159 | 0.001631 | 0.001125 | 0.933938 | 0.002362 | 0.003342 |
| A.C.T. | 0.006112 | 0.002582 | 0.002601 | 0.002754 | 0.001742 | 0.002225 | 0.826227 | 0.012338 |
| N.T. | 0.001141 | 0.001126 | 0.003276 | 0.005362 | 0.002592 | 0.000984 | 0.003893 | 0.756375 |

[a] Transition probabilities have been rounded to six decimal places and consequently column sums may not always equal unity.
[b] Taken from Lewis (1979).

In this example, the operators **P** and **Q** are given by

$$
\mathbf{P}=\begin{bmatrix}
1 & 0 & 0 & \cdot & \cdot & \cdot & \cdot & \cdot \\
0 & 1 & 0 & \cdot & \cdot & \cdot & \cdot & \cdot \\
0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 \\
0 & 0 & 0 & 0 & \cdot & \cdot & \cdot & \cdot \\
0 & 0 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
\cdot & \cdot & & & & & & \\
\cdot & \cdot & & & & & & \\
\cdot & \cdot & & & & & & 0
\end{bmatrix}
$$

$$
\mathbf{Q}=\begin{bmatrix}
0 & 0 & 0 & \cdot & \cdot & \cdot & \cdot & \cdot \\
0 & 0 & 0 & \cdot & \cdot & \cdot & \cdot & \cdot \\
0 & 0 & 0 & -1 & -1 & -1 & -1 & -1 \\
\cdot & \cdot & 0 & 1 & 0 & 0 & 0 & 0 \\
\cdot & \cdot & \cdot & & 1 & 0 & 0 & 0 \\
\cdot & \cdot & \cdot & & & 1 & 0 & 0 \\
\cdot & \cdot & \cdot & & & & 1 & 0 \\
0 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 1
\end{bmatrix}
$$

The first result to note is that the column sums of the submatrices $\mathbf{G}_{13}$, $\mathbf{G}_{23}$, and $\mathbf{G}_{33}$ are not equal and that $\mathbf{CGQ} \neq \mathbf{0}$. Also, since the model is closed (in other words, the matrix **G** is stochastic), the maximum eigenvalue of **G** is 1. Thus the approximate aggregated matrix $\hat{\mathbf{G}}_{\text{as}}$ derived earlier, which is valid for long time intervals, can be calculated, and it is given in Table 3.

To obtain the upper and lower aggregated transition matrices, we note that the column sums of $\mathbf{G}_{13}$, $\mathbf{G}_{23}$, and $\mathbf{G}_{33}$ range from 0.011301, 0.010569, and 0.871288 to 0.099149, 0.033146, and 0.978131, respectively. Hence, the matrices $\hat{\mathbf{G}}_{\text{l}}$ and $\hat{\mathbf{G}}_{\text{u}}$ are easily obtained and these are also shown in Table 3.

The remaining three aggregated matrices are obtained using different weights for $\mathbf{W}'$. The unweighted least-squares approximation is given by $\mathbf{W}' = \mathbf{C}'$. For a

TABLE 3 Alternative aggregation matrix elements derived from the data.

| | | | | | |
|---|---|---|---|---|---|
| $\hat{\mathbf{G}}_{\text{as}}$ | | | $\hat{\mathbf{G}}_{\text{u}}$ | | |
| 0.9611 | 0.0131 | 0.0256 | 0.9614 | 0.0160 | 0.0991 |
| 0.0118 | 0.9591 | 0.0125 | 0.0104 | 0.9577 | 0.0331 |
| 0.0270 | 0.0278 | 0.9620 | 0.0282 | 0.0263 | 0.9781 |
| $\hat{\mathbf{G}}_{\text{l}}$ | | | $\hat{\mathbf{G}}_{\text{l.s.}}$ | | |
| 0.9614 | 0.0160 | 0.0113 | 0.9614 | 0.0160 | 0.0368 |
| 0.0104 | 0.9577 | 0.0106 | 0.0104 | 0.9577 | 0.0216 |
| 0.0282 | 0.0263 | 0.8713 | 0.0282 | 0.0263 | 0.9416 |
| $\hat{\mathbf{G}}_{\text{eq}}$ | | | $\hat{\mathbf{G}}_{0}$ | | |
| 0.9614 | 0.0160 | 0.0238 | 0.9614 | 0.0160 | 0.0213 |
| 0.0104 | 0.9577 | 0.0144 | 0.0104 | 0.9577 | 0.0151 |
| 0.0282 | 0.0263 | 0.9618 | 0.0282 | 0.0263 | 0.9636 |

weighted least-squares approximation, where the weights are proportional to the initial population vector $\boldsymbol{K}(0)$, in this case the population given by the 1966 population, the aggregated matrix can be derived by either summing the appropriate elements in Table 1, or using the expression given in eqn. (28). To obtain the $\hat{\mathbf{G}}_{eq}$ matrix it is necessary to calculate the eigenvector corresponding to the maximum eigenvalue of $\mathbf{G}$. This was computed to be (0.3526, 0.2291, 0.1692, 0.0712, 0.1263, 0.0217, 0.0216, 0.0083), where each element represents the proportion of people in each state. Using this vector to define $\mathbf{W}'$, the matrix $\hat{\mathbf{G}}_{eq}$ was calculated as shown in Table 3.

The aggregated matrices were then used to project the population share for each region. The results for the years 1981, 1991, and 2001, and the stable or equilibrium vectors are given in Table 4 and Figure 1.

The results obtained by using $\hat{\mathbf{G}}_{as}$ and $\hat{\mathbf{G}}_{eq}$ agree closely with the "exact" result, (obtained from the $8\times8$ transition matrix), and both have the correct equilibrium value. The results obtained using $\hat{\mathbf{G}}_0$ have errors ranging up to only 0.84% for the year 2001, but have errors of 5% for the equilibrium eigenvector. The bounds

TABLE 4 Alternative projections giving proportion of total population in N.S.W., Victoria, and the remainder of Australia.

| Matrix used | Year 1981 | 1991 | 2001 | Equilibrium |
|---|---|---|---|---|
| Exact $\mathbf{G}$ | 0.3595 | 0.3574 | 0.3560 | 0.3526 |
| | 0.2724 | 0.2682 | 0.2644 | 0.2291 |
| | 0.3681 | 0.3744 | 0.3796 | 0.4183 |
| $\hat{\mathbf{G}}_{as}$ | 0.3607 | 0.3587 | 0.3570 | 0.3526 |
| | 0.2722 | 0.2657 | 0.2637 | 0.2291 |
| | 0.3671 | 0.3756 | 0.3793 | 0.4183 |
| $\hat{\mathbf{G}}_{u}$ | 0.4149 | 0.4675 | 0.5210 | |
| | 0.2857 | 0.2958 | 0.3076 | |
| | 0.3800 | 0.4024 | 0.4274 | |
| $\hat{\mathbf{G}}_{l}$ | 0.3515 | 0.3397 | 0.3278 | |
| | 0.2688 | 0.2597 | 0.2503 | |
| | 0.3059 | 0.2637 | 0.2306 | |
| $\hat{\mathbf{G}}_{l.s.}$ | 0.3697 | 0.3758 | 0.3811 | 0.4150 |
| | 0.2769 | 0.2764 | 0.2759 | 0.2653 |
| | 0.3534 | 0.3478 | 0.3431 | 0.3197 |
| $\hat{\mathbf{G}}_{eq}$ | 0.3607 | 0.3591 | 0.3578 | 0.3526 |
| | 0.2719 | 0.2671 | 0.2629 | 0.2291 |
| | 0.3674 | 0.3738 | 0.3794 | 0.4183 |
| $\hat{\mathbf{G}}_0$ | 0.3589 | 0.3558 | 0.3530 | 0.3347 |
| | 0.2724 | 0.2681 | 0.2642 | 0.2355 |
| | 0.3687 | 0.3762 | 0.3828 | 0.4298 |

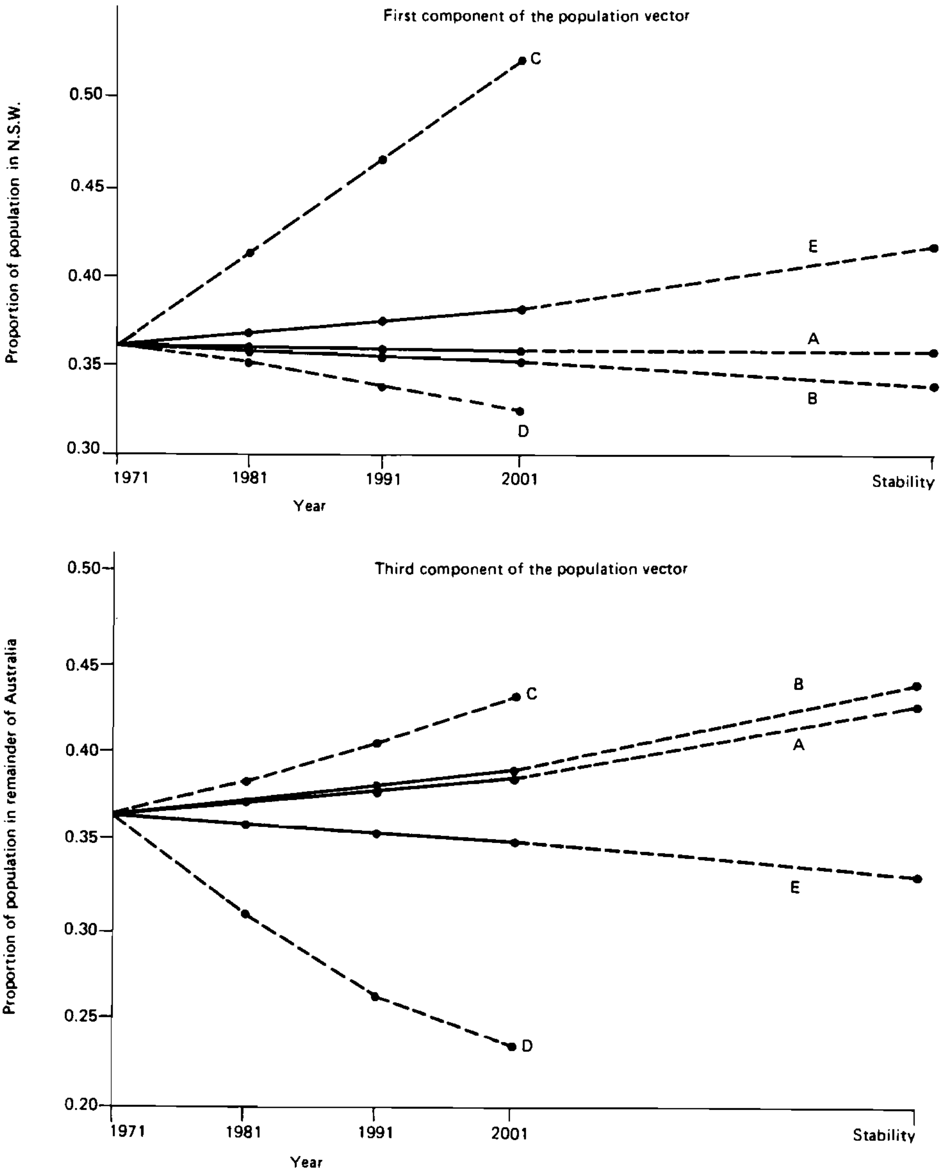



FIGURE 1 Values obtained for two components of the population vector: A, exact result; B, approximate result using $\hat{\mathbf{G}}_0$; C, upper aggregation; D, lower aggregation; E, unweighted least-squares using $\hat{\mathbf{G}}_{l.s.}$. The "remainder of Australia" includes all regions except N.S.W. and Victoria.

provided by the upper and lower aggregation matrices range from 22% to over 50% for the year 2001. Furthermore, they do not provide any reasonable bounds on the equilibrium vector. The unweighted least-squares aggregated matrix $\hat{\mathbf{G}}_{l.s.}$ did not give very good results.

The results above will not be duplicated for all systems, but it can be assumed that, if the initial population vector is not too far from the equilibrium vector, $\hat{\mathbf{G}}_0$ will be a good approximation. One method of determining whether there will be any bias is to compare the equilibrium vector of $\hat{\mathbf{G}}_0$ to $\hat{\boldsymbol{K}}(0)$. If they are different then error will be present due to aggregation.

Better results can be obtained by using $\hat{\mathbf{G}}_{as}$ and $\hat{\mathbf{G}}_{eq}$ but both matrices require disaggregated data for their construction. If these are not available $\hat{\mathbf{G}}_0$ is the only matrix that can be used.

## 8 IMPLICATIONS FOR POPULATION MODELING

Earlier work on the aggregation problem has focused on the conditions necessary for "perfect aggregation." However, as noted by Rogers (1976), these conditions are rarely obtained in demographic models. This paper has developed a method for constructing the non-Markovian equation which an aggregated system satisfies when the conditions for "perfect aggregation" do not hold. Further, various Markovian approximations have been derived, and Luenberger's (1978) upper and lower aggregated models have been used.

In qualitative terms, the implications of these results for population modeling are as follows. (The quantitative implications are still an area for further research.) Firstly, if disaggregated data are not available or have not been collected, then the only aggregated transition matrix which can be constructed is the observed transition matrix which is equivalent to the matrix $\hat{\mathbf{G}}_0$ [see eqn. (28)]. Where the data required to construct the upper or lower aggregated systems are not available, attempts could be made to estimate the missing data. Use of the matrix $\hat{\mathbf{G}}_0$, the linear Markovian model [eqn. (2)], should produce qualitatively good results for the initial vectors $\hat{\boldsymbol{K}}(t)$ (e.g., $t = 1, 2, 3, 4$). However, for larger values of $t$ there will be increasing sources of error and the magnitude of these errors will depend on how far the vector $\hat{\boldsymbol{K}}(0)$ is from the stable vector $\hat{\boldsymbol{u}}_1$. Since the vector $\boldsymbol{u}_1$ for the disaggregated model cannot be obtained in this case, it will be necessary to check how far $\hat{\boldsymbol{K}}(0)$ deviates from the stable vector of $\hat{\mathbf{G}}_0$. The qualitative differences between these two vectors probably give an idea of the size of the error caused by aggregation, but attempts to quantify the error are still needed. Rees (1980), in carrying out multiregional projections for the United Kingdom, found that the initial population vector was reasonably close to the stable distribution. However, Liaw (1980) found that for regional models of Canada the difference was much larger and hence that aggregation could lead to more serious errors.

Secondly, in the situation where the disaggregated data are available, the matrices $\hat{\mathbf{G}}_{as}$ and $\hat{\mathbf{G}}_{eq}$ can, in principle, be constructed. However, the calculation of these matrices involves knowing either the maximum eigenvalue or the corresponding eigenvector of $\mathbf{G}$. For large systems, this may be almost as difficult a calculation as the computation involved for the original disaggregated system [eqn. (1)]. However, the calculation of $\hat{\mathbf{G}}_u$ and $\hat{\mathbf{G}}_l$ would be relatively straightforward and hence these could be used to obtain upper and lower bounds for the population vector. How useful these bounds would be depends on the differences between the column sums of the $\mathbf{G}_{ij}$ submatrices. In the numerical example given in Section 7 the bounds were not very useful, but for larger systems a better result may be obtained, and this should be investigated.

In general, the linear, time-invariant Markovian models used in population projections are only an approximation, and the observed transition rates will be in error if the system is sufficiently far from the stable distribution. However, this is not the only source of error facing the demographer, and in many cases it may not be the major source of error.

## ACKNOWLEDGMENTS

The author is grateful to Andrei Rogers for providing hospitality at IIASA and for the helpful suggestions from him, Helen Lewis, and Joel Cohen.

## REFERENCES

Ando, A., and F. M. Fisher (1963) Near-decomposability, partition and aggregation, and the relevance of stability discussions. International Economic Review 4:53–67.

Aoki, M. (1968) Control of large scale dynamic systems by aggregation. IEEE Transactions on Automatic Control AC-13:246–253.

Ara, K. (1959) The aggregation problem in input–output analysis. Econometrica 27:257–262.

Burke, C. J., and M. Rosenblatt (1958) A Markovian function of a Markov chain. Annals of Mathematical Statistics 29:1112–1122.

Fisher, W. D. (1969) Clustering and Aggregation in Economics. Baltimore, Md.: Johns Hopkins University Press.

Hatanaka, M. (1952) Note on consolidation within a Leontief system. Econometrica 20:301–303.

Kemeny, J. G., and J. L. Snell (1960) Finite Markov Chain. Princeton, N.J.: Van Nostrand.

Keyfitz, N. (1972) On future population. Journal of the American Statistical Association 67:347–363.

Liaw, K.-L. (1980) Multistate dynamics: the convergence of an age-by-region population system. Environment and Planning A 12:589–613.

Lewis, H. M. (1979) Aggregation of States in Markov Chains. Ph.D. Thesis. Newcastle, N.S.W.: University of Newcastle.

Luenberger, D. G. (1978) Aggregation and Dual Aggregation of Positive Dynamic Systems. Preprints, Congress of the International Federation of Automatic Control (IFAC), Helsinki, June.

McManus, M. (1956) On Hatanaka's note on consolidation. Econometrica 24:482–487.

Rees, P. H. (1980) Multistate demographic accounts: measurement and estimation procedures. Environment and Planning A 12:499–531.

Rogers, A. (1968) Matrix Analysis of Interregional Population Growth and Distribution. Berkeley, Cal.: University of California Press.

Rogers, A. (1969) Note on perfect aggregation in the matrix cohort survival model of interregional population growth. Journal of Regional Science 9:417–424.

Rogers, A. (1971) Matrix Methods in Urban and Regional Analysis. San Francisco, Cal.: Holden-Day.

Rogers, A. (1975) Introduction to Multiregional Mathematical Demography. New York: Wiley.

Rogers, A. (1976) Shrinking large-scale population projection models by aggregation and decomposition. Environment and Planning A 8:515–541.

Rogers, A., and D. Philipov (1979) Multiregional Methods for Subnational Population Projections. WP-79-40. Laxenburg, Austria: International Institute for Applied Systems Analysis.

Rowland, D. T. (1976) Patterns of Internal Migration in Australia 1966–1971. In Population Redistribution: Policy Research edited by R. J. Pryor. Proceedings of a seminar on internal migration: research for policy-making in Australia, July. Canberra: Department of Demography, Australian National University.

Simon, H. A., and A. Ando (1961) Aggregation of variables in dynamic systems. Econometrica 29:111–138.

# THE AUTHORS

*Luis J. Castro* came to IIASA from the Universidad Nacional Autonoma de México to work on a comparative study of migration and settlement in the 17 nations associated with the Institute and on a case study of urbanization and development in Mexico. His current research is focused on age patterns of migration and their underlying cause-specific components.

*Robert Gibberd* is a Senior Lecturer in the Department of Mathematics at the University of Newcastle, Australia. His current interests include the application of mathematical models to demographic phenomena. He visited IIASA in 1980 to work on optimal aggregation in population projection models.

*Young J. Kim*, originally from the Republic of Korea, is currently an Assistant Professor in the Department of Population Dynamics at Johns Hopkins University, Maryland, USA, where she is engaged in a study of the dynamics of populations with changing vital rates. She visited IIASA in 1980 to apply this research in the field of multiregional demography.

*Pavel Kitsul*, from the Institute of Control Sciences of the USSR Academy of Sciences, Moscow, USSR, is primarily interested in the theory and application of stochastic processes and identification and control in complex systems. He has worked on estimating morbidity rates for unstable and non-stationary populations and on other related demographic problems.

*Jacques Ledent* studied engineering at the Ecole Nationale des Ponts et Chaussées, Paris, France, and at Northwestern University, Illinois, USA. In recent years he has worked extensively on problems of multiregional demography, concentrating in particular on the construction of multistate life tables. He is also involved with the comparative migration and settlement study at IIASA.

*Dimiter Philipov* came to IIASA from the Scientific Institute of Statistics at Sofia, Bulgaria, to work on the comparative migration and settlement study. His interests include mathematical statistics, the mathematics of population growth, and demoeconomics.

*András Pór* obtained his degree from the Mathematical Faculty of the Eötvos Lorand University in Budapest, Hungary, and in recent years has been working on the application of probability theory and methods of nonlinear optimization to problems of systems analysis. While at IIASA he has, among other things, been involved in the use of entropy maximization and multiproportional adjustment techniques to infer detailed migration flows from aggregate data.

*Richard Raquillet* is a graduate of the Ecole Nationale des Ponts et Chaussées, Paris, France, and of the Technological Institute at Northwestern University, Illinois, USA. He joined the Human Settlements and Services Area at IIASA for eight months in 1977 to work on the comparative analysis of migration and settlement and on related topics in multiregional demography.

*Andrei Rogers* has been Chairman of the Human Settlements and Services Area at IIASA since October 1976. He came to IIASA from Northwestern University, Illinois, USA. His current research focuses on migration patterns and the evolution of human settlement systems in both developed and developing countries.

*Frans Willekens* received a doctorate in Urban Systems Engineering at Northwestern University, Illinois, USA, after completing his undergraduate work in agricultural engineering, economics, and sociology at the University of Leuven, Belgium. He coordinated the comparative migration and settlement study at IIASA and has worked extensively in the field of multiregional demography.

## RELATED IIASA PUBLICATIONS

| | |
|---|---|
| Migration and Settlement: Selected Essays<br>(Reprinted from a Special Issue of Environment and Planning A)<br>Andrei Rogers (Editor) | RR-78-6 |
| Migration and Settlement: Measurement and Analysis<br>Andrei Rogers and Frans Willekens | RR-78-13 |
| Spatial Population Analysis: Methods and Computer Programs<br>Frans Willekens and Andrei Rogers | RR-78-18 |
| Migration Patterns and Population Redistribution<br>(Reprinted from Regional Science and Urban Economics)<br>Andrei Rogers | RR-80-7 |
| Essays in Multistate Mathematical Demography<br>(Reprinted from a Special Issue of Environment and Planning A)<br>Andrei Rogers (Editor) | RR-80-10 |
| Multidimensionality in Population Analysis<br>(Reprinted from Sociological Methodology 1980)<br>Nathan Keyfitz | RR-80-33 |