Regional Demographic Modeling

Frans Willekens

Any modeling effort consists of a number of steps: (1) definition of the system to be modeled, (2) construction of the model, (3) estimation of the model parameters, and (4) validation of the model and tests of internal consistency. Regional demographic modeling is no exception. It is different, however, because of the impact of new and alternative ways of defining the regional demographic system on recent modeling efforts. The system being modeled has changed more than the modeling approach.

Two observations have led to the current state of regional demographic modeling. First, population changes in a particular subnational unit are heavily influenced by what happens in other regions. Modeling the interdependence between regions has been a major challenge. The lack of success for several years was due not so much to inadequate modeling as to the definition of an inadequate system. The interdependence could only be fully represented when regions were viewed as parts of a larger, multiregional system. Second, population is not homogeneous with regard to demographic behavior. Changes in the population size, the number of births or migrations, may be due to variations in the population composition rather than to variations in demographic behavior. There is, therefore, a need to decompose the population into demographic categories.

The regional population structure to be modeled has a geographic or spatial dimension and a demographic dimension. In the geographic dimension the population may be divided into discrete spatial units (a system of regions): rural/urban, states, counties, functional urban regions, and so on. In the extreme case, space is considered as a continuum and the growth and dispersion of population in this continuum is investigated. In the demographic dimension, the total population may be decomposed by age, sex, color, or any other demographic feature. It is not surprising that demographers have emphasized the study and projection of regional populations decomposed into demographic categories, whereas geographers decomposed populations along the geographic dimension. Multiregional demography combines both perspectives and proposes an integrated methodology for the analysis and projection of multiregional population systems disaggregated by age, and sex (Rogers 1975). In another approach, not age but duration of residence in the region constitutes the basis for demographic categories (Ginsberg 1971).
This chapter consists of three major sections. The first presents a short review of approaches to modeling regional demographic change, focusing on the multiregional approach. The second section treats the multiregional demographic growth model and shows how the model parameters are derived from a multiregional life table. The third section deals with a major problem in multiregional demographic modeling; namely, the lack of adequate data, in particular with regard to migration. New opportunities of better analytical techniques can only be fully explored if data are abundantly available or if accurate estimation procedures can be developed.

Models of Regional and Multiregional Demographic Change

An exposition of projection techniques is impossible without reference to the population structure to be projected (projection system). Demographic and geographic models differ not so much in the mathematical features of the model as in the projection system, the dynamics of which the model tries to describe. The basic difference between projection techniques is how migration is incorporated into the projection model. While the uniregional perspective analyzes one region at a time, where migration is a net exchange with other regions of the same system, in the multiregional perspective the interdependence between the regions is represented by directional migration flows.

The Multiregional Perspective

Regions do not grow or decline in a vacuum. They interact with other regions of the same country and with the rest of the world. In the multiregional perspective, this interaction is fully represented (Rogers and Willekens 1976). The projection system consists of a set of regions. The difference between the multiregional perspective and the open uniregional perspective is the way the interdependence is represented.

First, there is a direct interaction through the people going from one region to another, the interaction being represented through gross or directional migration flows. Second, there is an indirect interaction. Changes in the demographic behavior (fertility decline, say) in one region affect other regions indirectly through migration. Because of the greater specificity of the migration component and the ability to handle regional differential fertility and mortality levels, the assumptions going into the projection can be spelled out in greater detail.

Two types of multiregional projection models may be distinguished. Both are modifications of the Markov chain model and treat the demographic and geographic categories of the projection system as states of the state space
(elements of the state vector). They take the form of a system of linear simultaneous equations, easily expressed in matrix format:

\[
\begin{bmatrix}
K_1(t+1) \\
K_2(t+1) \\
K_3(t+1) \\
\vdots \\
K_p(t+1)
\end{bmatrix} = \begin{bmatrix}
g_{11} & g_{12} & g_{13} & \cdots & g_{1p} \\
g_{21} & g_{22} & g_{23} & \cdots & g_{2p} \\
g_{31} & g_{32} & g_{33} & \cdots & g_{3p} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
g_{p1} & g_{p2} & g_{p3} & \cdots & g_{pp}
\end{bmatrix} \begin{bmatrix}
K_1(t) \\
K_2(t) \\
K_3(t) \\
\vdots \\
K_p(t)
\end{bmatrix}
\]

or

\[
\{K(t+1)\} = G\{K(t)\}
\]

where \(K(t)\) is the state vector at time \(t\) with elements or subvectors \(K_i(t)\), and \(G\) is the transition matrix or growth operator with elements \(g_{ij}\). The simultaneity of the equation system, incorporated in the matrix multiplication, enables one to determine the indirect influences of one region upon another. It also makes possible the study of the dynamics of demographic change. Rogers (1966:177) points out: "Expressing the process of population growth as a matrix 'operator,' separates it from the population to which it is applied and thereby allows a clearer focus on the intrinsic characteristics of a particular growth structure, its impact on another population, and its long-run distributional consequences."

Common to both types of models is their reliance on the Markovian assumption: the probability of surviving and migrating during a given time interval depends only on the region of residence (and age) at the beginning of the interval and is independent of the previous life history. They differ, however, in the composition of the state vector and in the way the transition matrix is calculated.

In the first type of models of multiregional demographic change, which may be labeled aggregate models, the state space consists of the regions of the projection system. There is no disaggregation by age. These models, which have been studied predominantly in geography and regional science, are used to project the total population by region. Conventional Markov models are used to study population redistribution and do not consider natural increase. Population redistribution is represented as an ergodic Markov chain, in which it is possible to move from an arbitrary state to any other state in a finite number of steps but in which it is impossible to leave the system. Applications of ergodic Markov chains in migration research are given by Rogers (1966), McKinnon and Skarke (1977), and Salkin, Lianos, and Paris (1975). Rogers (1968) extended the Markov model to include all the important components of demographic change: migration and regional levels of fertility and mortality.
The second type of models are age-specific. The strong age effect of migration has long been recognized, and it led Rogers (1966, 1975) to consider a projection system which is both region- and age-specific. The modeling approach combined the efforts of geographers focusing on the directional migration flows in a multiregional system, with the efforts of demographers focusing on age as the most important variable. From the geographer's perspective, multiregional demography is an extension of conventional Markov models of interregional mobility along the age dimension. From the demographer's perspective, it generalizes the conventional demography of fertility and mortality by adding a spatial dimension. For a review of recent research in multiregional demography see Rogers (1978).

**The Multiregional Demographic Growth Model**

In multiregional demography, all regions are investigated simultaneously. The processes of change are represented by sets of simultaneous equations, which are easily expressed in matrix notation and studied using matrix algebra. Here, we review the multiregional demographic growth model, developed by Rogers (1973, 1975). It is a generalization of the cohort-survival model of Leslie (1945). The parameters of the model are not directly estimated but calculated from observed or estimated data by the multiregional life table. The procedure is described in the second part of this section.

**The Multiregional Cohort-Survival Model**

Denote the age composition and regional distribution of the population at time \( t \) by the vector \( \{K^t(x)\} \). Assuming age groups of equal length of five years, the vector can be partitioned as follows:

\[
\begin{bmatrix}
K^t(0) \\
K^t(5) \\
\vdots \\
K^t(z)
\end{bmatrix}
\]

and \( \{K^t(x)\} = K^t(x) \) (15.2)

where \( K^t(x) \) denotes the number of people in region \( i \) at time \( t \), who are \( x \) to \( +h \) years of age, with the age interval, five years say, \( K^t(x) \) is the regional distribution of the population in age group \( x \) to \( x+4 \).
The multiregional demographic growth model describes the change of the population vector \( \mathbf{K}^0 \), that is, the population by age and region. The change is recorded for a unit time interval, which is equal to the age interval, usually five years. To describe the model in detail, we distinguish between the projection of the population already alive in the base year (dynamics of a cohort) and the projection of births and the subsequent children in the 0–4 year age group. A two-region system is considered for convenience.

**Dynamics of a Cohort.** In a closed multiregional system, the members of a given cohort may migrate to other regions of the country, or die, that is, move to the absorbing state of death. The movement of an individual between the regions and the absorbing state may be illustrated by a Lexis diagram (Rogers 1973, 1975). Figure 15–1 is a Lexis diagram for a system of two regions. There are six classes of life lines, denoted by A, B, C, D, E, and F, respectively. Life lines B and E refer to individuals in region 1 who die or emigrate during the unit age (and time) interval. Life line C represents an individual who outmigrates from region 1 to region 2 and returns before the end of the time interval. Finally, D refers to an individual in region 1 who outmigrates to 2, survives the unit age interval, and stays in 2 until the beginning of the next interval.

![Two-Region Lexis Diagram](image_url)

Reprinted from A. Rogers (1973, page 5) with permission of the author and Pion Ltd.

**Figure 15–1. Two-Region Lexis Diagram**
In the diagram, we see that at time \( t \), there are five people in region 1, three of them in age group \((x - 5, x)\) and two in age group \((x, x + 5)\). Of the people aged \( x - 5 \) to \( x \) in region 1 at time \( t \), only two, namely, A and C, survive to be in region 1 at time \( t + 1 \), and \( x \) to \( x + 5 \) years of age. Two, namely, B and E, die. Only one, namely D, survives to be in region 2 at time \( t + 1 \) and \( x + 5 \) to \( x + 9 \) years of age. The people aged \( x \) to \( x + 4 \) at time \( t \) can survive, migrate within the country, or die in the unit interval.\(^1\)

Denote by \( s_i^{(j)}(x) \) the proportion of the people in region 1 and \( x \) to \( x + 4 \) years old at time \( t \), who survive to be \( x + 5 \) to \( x + 9 \) years old five years later at time \( t + 1 \) and are then in region 2. Equivalently, \( s_i^{(j)}(x) \) denotes the proportion of people \( x \) to \( x + 4 \) years old who remain in region 1. Ignoring immigration, the number of people of age \( x + 5 \) to \( x + 9 \) in region 1 at time \( t + 1 \) is given by

\[
K_1^{(t+1)}(x+5) = s_{i1}^{(j)}(x) K_i^{(0)}(x) + s_{i2}^{(j)}(x) K_2^{(0)}(x) \tag{15.3}
\]

Note that \( s_i^{(j)}(x) \) includes in principle the persons who left region 1 but returned in the same time interval. For projection purposes the complete migration history of an individual is not important, but the places of residence at the beginning and at the end of the projection interval are. Equation 15.3 written for region 2 yields

\[
K_2^{(t+1)}(x+5) = s_{i1}^{(j)}(x) K_1^{(0)}(x) + s_{i2}^{(j)}(x) K_2^{(0)}(x) \tag{15.4}
\]

Expressions 15.3 and 15.4 may be combined in the matrix operation:

\[
\begin{bmatrix}
K_1^{(t+1)}(x+5) \\
K_2^{(t+1)}(x+5)
\end{bmatrix} =
\begin{bmatrix}
s_{i1}^{(j)}(x) & s_{i2}^{(j)}(x) \\
s_{i1}^{(j)}(x) & s_{i2}^{(j)}(x)
\end{bmatrix}
\begin{bmatrix}
K_i^{(0)}(x) \\
K_1^{(0)}(x)
\end{bmatrix}
\]

\(
[K_i^{(t+1)}(x+5)] = S_i^{(0)}(x) [K_i^{(0)}(x)] \quad 0 \leq x < z
\)

The matrix of survivorship proportions \( S_i^{(0)}(x) \) may be derived directly from observed age-specific mortality and migration rates. In general, however, it is derived from the multiregional life table. The computation procedure is discussed later.

**Births.** The children of 0–4 years at time \( t+1 \) are born during the unit projection interval. Let \( F_i^{(0)}(x) \) be the annual birthrate of people aged \( x \) to \( x + 4 \) in region \( i \). It is assumed that children, born in the unit time interval \((t, t + 1)\), are born in the region of residence of the parents at time \( t \). The annual number of births in region 1 to people aged \( x \) to \( x + 4 \) is

\[
B_i^{(i)}(x) = F_i^{(0)}(x) K_i^{(0)}(x) \tag{15.5}
\]
The multiregional distribution of births is
\[
[B^t(x)] = F^t(x) \{K^t(x)\}
\]
where
\[
[B^t(x)] = \begin{bmatrix} B_1^t(x) \\ B_2^t(x) \end{bmatrix}
\]
and
\[
F^t(x) = \begin{bmatrix} F_1^t(x) & 0 \\ 0 & F_2^t(x) \end{bmatrix}
\]

The number of births during the five-year period starting at \(t\) to people aged \(x\) to \(x+4\) is
\[
[B^{t+1}(x)] = \int_0^h F^t(x+t) \{K^t(x+t)\}\ dt
\]

The integral equation may be approximated by the linear interpolation:
\[
[B^{t+1}(x)] = \frac{5}{2} \left[ F^t(x) \{K^t(x)\} + F^{t+1}(x+5) \{K^{t+1}(x+5)\} \right]
\]
\[
= \frac{5}{2} \left[ F^t(x) + F^{t+1}(x+5) S^t(x) \right] \{K^t(x)\}
\]

Not all the children born during the interval will still be alive at the end of the interval, that is, at \(t+1\). Define the matrix \(\hat{\beta}^t\):
\[
\hat{\beta}^t = \begin{bmatrix} \hat{\beta}_{ij}^t \\ \hat{\beta}_{2j}^t \end{bmatrix}
\]

where an element \(\hat{\beta}_{ij}^t\) is the proportion of babies born in region \(i\) during time interval \((t, t+1)\), who survive and are in region \(j\) at the end of the time interval. This matrix takes into account the migration of children in the first age group.

Writing
\[
[B^t(x)] = \frac{5}{2} \hat{\beta}^t \left[ F^t(x) + F^{t+1}(x+5) S^t(x) \right]
\]

the population in the first age group at time \(t+1\) is
\[
[K^{t+1}(0)] = \sum_x B^t(x) \{K^t(x)\}
\]

(15.7)
The summation is over all the fertile age groups. If \( \bar{\alpha} \) and \( \bar{\beta} \) are respectively the lowest and the highest age group of the reproductive period, the summation is from \( \bar{\alpha} \) to \( \bar{\beta} \).

**The Complete Growth Model.** The two equation systems (15.4 and 15.7) describe the growth of a multiregional population. Both systems may be combined into a single matrix expression of an extremely simply form:

\[
\{ \bar{X}(t+1) \} = Q(t) \{ \bar{X}(t) \}
\]  

(15.8)

where

\[
\{ \bar{X}(t) \} = 
\begin{bmatrix}
\{ X^{0}(0) \} \\
\{ X^{0}(5) \} \\
\vdots \\
\{ X^{0}(x) \} \\
\vdots \\
\{ X^{0}(z) \}
\end{bmatrix}
\]

and

\[
Q(t) = 
\begin{bmatrix}
0 & 0 & \cdots & \cdots & \cdots & \cdots & 0 \\
S^{0}(0) & 0 & \cdots & \cdots & \cdots & \cdots & 0 \\
0 & S^{0}(5) & \cdots & \cdots & \cdots & \cdots & 0 \\
\vdots & \vdots & \ddots & \cdots & \cdots & \cdots & \vdots \\
0 & \cdots & \cdots & S^{0}(z-5) & 0 \\
\end{bmatrix}
\]  

(15.9)

with \( z \) being the last age group. The matrix \( Q(t) \) is called the generalized Leslie matrix (Feeney 1970:36; Rogers 1975:123).

If the growth matrix is constant in time, the population growth model may be written as:

\[
\{ X(t) \} = G \{ X(0) \}
\]  

(15.10)

with \( \{ X(0) \} \) the base year population.

**The Multiregional Life Table**

The multiregional life table is a table expressing the mortality and migration history of a hypothetical number of people born at the same time and in the same region (birth cohort), as they age. The multiregional life table was developed by Rogers (1975, chapter 2) and is a fundamental concept of multiregional demography. It is a descriptive device; it does not explain but
describes in different ways the mortality and migration pattern of a population. It is static since the life table is constructed from a set of mortality and migration schedules of a single time period. The life table is finally a model; it applies the observed age-specific rates of mortality and migration to a birth cohort (100,000 births, say) and assumes that the cohort follows a life history prescribed by the vital rates of a given period.

To construct a life table, first transform observed rates into a set of probabilities. These probabilities form the basis for the other life-table statistics. Two types of probabilities may be distinguished: unconditional and conditional probabilities.

Unconditional Probabilities. Let \( \hat{\ell}_i(x) \) denote the probability that a person born in region \( i \) will be in region \( j \) at exactly age \( x \). The set of possible probabilities in a two-region system (urban-rural) is contained in the matrix \( \hat{\ell}(x) \):

\[
\hat{\ell}(x) = \begin{bmatrix} \hat{\ell}_1(x) & \hat{\ell}_2(x) \\ \hat{\ell}_3(x) & \hat{\ell}_4(x) \end{bmatrix}
\]  

For example, \( \hat{\ell}_1(x) \) denotes the probability that a person born in region 1 will be in region 2 at age \( x \). The diagonal element, \( \hat{\ell}_i(x) \) is the probability that he is born in region 1 and is there at age \( x \). Note that this does not imply a continuous stay in this region. The person may spend some time in another area before reaching age \( x \). The matrix \( \hat{\ell}(x) \) tells something about the regions of residence of a person at two points in time, that is, time at birth and at age \( x \).

Conditional Probabilities. Assuming that the probabilities of survival and of migrating at a certain age depend only on the region of residence at that age and are independent of previous residences, then \( \hat{\ell}(x) \) may be written as the product of conditional probabilities:

\[
\hat{\ell}(x) = P(x-5) P(x-10), \ldots, P(y), \ldots, P(0)
\]  

and

\[
\hat{\ell}(x) = P(x-5) \hat{\ell}(x-5)
\]

where

\[
P(y) = \begin{bmatrix} p_{11}(y) & p_{21}(y) \\ p_{12}(y) & p_{22}(y) \end{bmatrix}
\]

and an element \( p_{ij}(y) \) denotes the probability that a person of region \( i \) and \( y \) years old will survive and be in region \( j \) five years later (age interval). This is
conditional upon reaching age \( y \). Conditional probabilities may be computed from observed age-specific rates of mortality and migration (Rogers and Ledent 1976).

\[
\mathbb{P}(y) = \left[ I + \frac{5}{2} \mathbb{M}(y) \right]^{-1} \left[ I - \frac{5}{2} \mathbb{M}(y) \right]
\]  

(15.13)

where

\[
\mathbb{M}(y) = \begin{bmatrix}
M_{13}(y) + M_{14}(y) & -M_{21}(y) \\
-M_{15}(x) & M_{25}(y) + M_{14}(y)
\end{bmatrix}
\]

(15.14)

with \( M_{ij}(y) \) being age-specific death rates in region \( i \) and \( M_{ij}(y) \) age-specific rates of migration from \( i \) to \( j \).

From the knowledge of \( \xi(x) \), the number of people at exact age \( x \) may be obtained through multiplication by the base population, which is the regional radix or birth cohort (number of births in the life-table population):

\[
\xi(x) = \hat{\xi}(x) \xi(0)
\]

where \( \xi(0) \) is a diagonal matrix of regional radices.

On the basis of these probability measures a set of useful demographic statistics may be calculated. For instance, how long will a person of a specific demographic category stay in a region? How many times will he change residence before reaching age 65? How many people will migrate? How many 20-to-24-year-olds are in a given region if the number of births there is known? What proportion of an age class is expected to migrate? The answers to these and other questions may constitute columns of the multiregional life table. The various life-table statistics describe therefore different aspects of the same population.

Multiregional Demographic Analysis with Incomplete Data

Better analytical techniques can only be fully explored if data are abundantly available or if accurate estimation procedures can be developed. Recently, attention has been devoted to the design of techniques to infer detailed migration patterns from available data. This research still needs expansion. Four groups of techniques may be considered.

Inferring Period Migration Data from Lifetime Migration Statistics. In developing nations a migrant is defined as a person living in a region other than his region of birth. This lifetime migration measure is not suited for multiregional population analysis, which is based on a set of transition
probabilities during unit time intervals. Period migrations may be derived from lifetime data by applying life-table or stable-population theory. A review of some procedures is given by the United Nations (1970). These techniques, which differ in data required and assumptions made, yield net migration flows. Gross flows may be obtained with the place-of-birth-by-place-of-residence method developed by Rogers (1975:172–185), which relies on multiregional life-table theory. If for two consecutive censuses tabulations are available with the population by age, region of birth, and region of residence, survivorship proportions may be computed applying a formula similar to the one used in life-table construction:

\[ S(x) = K^{12}(x + h) \left[ K^{11}(x) \right]^{-1} \]

where \( K^{ij}(x) \) is a matrix of the population aged \( x \) to \( x + h \) by region of birth and residence at the \( i \)th census, and \( h \) is the census interval. A disadvantage of this method is that it may yield survivorship proportions that are negative or greater than one, if the observed population deviates much from the stationary population.

Model Migration Schedules. The regularity exhibited by empirical schedules of age-specific migration rates initiated research on ways to describe these schedules by a limited number of parameters. Rogers, Raquillet, and Castro (1978) consider schedules of gross migration rates; Pittenger (1976:187–197; 1978) presents a typology and parameterization of age curves of net migration rates. The approach is similar to that used by demographers to develop model life tables and model fertility schedules. Once "standard" migration schedules have been obtained, they can be used in situations where no information is available on the age structure of migrants.

Marginal Adjustment Methods. Frequently, data are published for aggregate migrant categories only. Marginal adjustment methods, of which the entropy maximization technique and the biproportional and multiproportional adjustment methods are illustrations, are designed to infer statistics for subgroups of migrants when only aggregate information is available (Wilson 1970; Chilton and Poet 1973). Subgroups may be defined on the basis of age, sex, nationality, income and so on. Willekens (1977) applies the entropy method to differentiate internal migration patterns by nationality and by age. Recent work focuses on methodological improvements (Willekens, Pör, and Raquillet 1979).

Demographic Accounting. A spatial demographic account provides a framework for presenting the available data on regional fertility, mortality, and migration in a consistent and correct way, and for estimating the missing items in the accounts matrix (Rees and Wilson 1977). The missing elements consist
of minor migration flows in a given time interval. The estimation involves an
iterative procedure in which the minor flows are derived from the known major
flows and population at the beginning and end of the time interval considered.
The emphasis is on a consistent system that correctly reproduces the
population changes in the base period. Demographic accounting may there-
fore be used for calibration purposes and hence in the construction of
population projection models (Rees 1978).

Conclusion

Recent activity in regional demographic modeling is aimed mainly at a better
definition of the population system to be modeled. Demographic detail was
introduced to benefit from the regularities in demographic behavior; the spatial
system considered became multiregional in order to represent more accurately
the migration component and the interdependence between regions. Multi-
regional demography combines the demographer's interest in the age dimen-
sion and the geographer's interest in spatial interaction.

The model of multiregional demographic change presented here is linear
and is embedded in Markov theory. It is a generalization of the conventional
demographic growth model to a system of regions. The steps involved in
constructing the model and the assumptions made are reviewed. Several of the
assumptions are related to the way the multiregional life table, which contains
the model parameters, is constructed. The life table transforms demographic
data observed during a given period into cohort measures, such as the
parameters of the growth model.

Multiregional demographic analysis is handicapped by the lack of
statistical data. Until more and better data become available, estimation
procedures to infer missing data may be applied.

The first phase of the new field of multiregional demography, which was
heavily demographic analysis-oriented, is coming to an end. The methodology
is fairly well developed; a package of computer programs exists, and initial
results of research on techniques to overcome data limitations are promising.
Applications are becoming available at an increasing rate. For instance, the
International Institute for Applied Systems Analysis (IIASA) adopted the
multiregional perspective in a quantitative assessment of recent patterns of
migration and population distribution in its National Member Organizations
countries. In the second phase, the perspective will be widened and socio-
economic factors will be introduced to explain the values taken by the model
parameters. Behavioral aspects will gain in importance. The development of a
demoeconomic or economic-demographic model of multiregional change will
remain a challenge for years to come.
Notes

1. The projection interval is assumed to be the same as the age interval, that is, five years. The superscript $t$ refers to the time period, and not to the exact year.