A MATRIX MODEL OF POPULATION REDISTRIBUTION

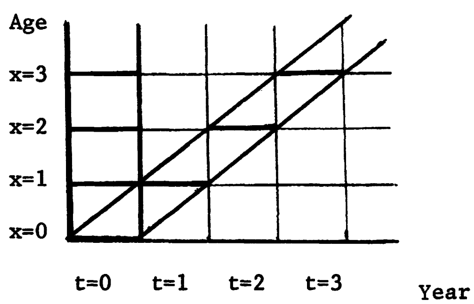
W. L. Li
Ohio State University*

This paper employs matrix theories in constructing a probabilistic model of interregional population redistribution. The analysis can be viewed as an extension of the previous contribution made by Leslie (1945), Keyfitz (1964), Tarver (1965), Roger (1968), and many others. In this paper, regional population is treated as an open system and the interregional flows of population are considered. By drawing on the merits and eliminating the defects of various previous models, this paper can present, it is hoped, a more realistic projection of regional distribution.

Population Projection as a Cohort Approach

Demographic projections are typically made by calculating the future consequences of past conditions. In this sense, a population projection is merely one way of regarding the fertility, mortality, and migration of a given period in the past. Its usefulness does not necessarily depend on whether it constitutes a good prediction, for it can be interesting to analyze a hypothetical trend based on the assumption that given rates would continue into the future.

To illustrate a general demographic analysis, a Lexis diagram is perhaps the most easily comprehensible device (see Pressat, 1961). In the diagram time is along the horizontal axis, and age along the vertical. A group of individuals that starts at x = 0 at the moment of birth in time t, and is followed up to a terminating point corresponding to both their age and time of death, is generally referred as a cohort.



As the analysis of actual cohort requires many years of observation, a very useful construct in demographic analysis is the synthetic cohort. This is formed by taking the cross-sectional experience of one age category of individuals and applying it as a constant to an actual cohort. The experience could be mortality, fertility, migration, school enrollments, or any other social events that can be treasured numerically. If mortality is analyzed the construct is a life table; if fertility, mortality, and age distribution, it could be a stable population model. In other words, a Lexis diagram illustrates a fundamental mode of demographic analysis that is the consequence of a particular social process if an actual cohort followed exactly the same series of probabilities that various cross-sectional age groups at a given date experienced. It is based on the postulates that population projections should be regarded as a legitimate exercise, and that demographers ought to seek "more accurate" population projection by basing them on more realistic assumption.

The Leslie-Keyfitz Model

In the Leslie-Keyfitz model, the procedure of projecting a population is separable from its original age distribution, following the principle sometimes referred as the Markov process. The projection operator is a matrix whose elements reflect the range in mortality and fertility of the given population in a given year. The matrix can be repeatedly multiplied by an age distribution, and it can carry the numbers at the several ages into successive time periods. In other words, to denote the age distribution after n periods from the starting point, one raises the matrix to the nth power and then multiplies it by the initial age distribution.

As an illustration, let us denote:

K_x^t = the number of females alive in age group x at time t.

S_x = the survival ratio; that is, the probability that a female aged x at time t will be live to age (x + 1) at time (t + 1).

F_x = the fertility rate of a surviving female from aged x to (x + 1).

For simplicity, let us illustrate the principal with a population of only three age groups, which can be denoted as x = 1, 2, and 3. Then, the Leslie-Keyfitz model can be expressed as follows:

$$\begin{bmatrix} S_o \cdot F_1 & S_o \cdot F_2 & S_o \cdot F_3 \\ S_1 & 0 & 0 \\ 0 & S_2 & 0 \end{bmatrix} \times \begin{bmatrix} K_1^t \\ K_2^t \\ K_3^t \end{bmatrix} = \begin{bmatrix} K_1^{t+1} \\ K_2^{t+1} \\ K_3^{t+1} \end{bmatrix} \qquad (1)$$

Or it can simply expressed as:

$$S \cdot K_t = K_{t+1} \qquad (2)$$

* This research has been supported in part by the National Sciences Foundation under contract NSF - GS2630.

The projection matrix S is obtained from the cross-sectional data of a population at a given time. It is treated as constant in order to project the population to its future size.

The empirical implementation of this matrix, as suggested by Keyfitz (1968: 30-31), is rather simple for those diagonal elements, S_1 and S_2. For instance,

$$S_x = L_{x+1}/L_x \qquad (3)$$

where L_x is adopted from the life-table notation of the survivorship function. However, it is relatively more complicated to implement the elements in the first row of the matrix. S_o denotes the survival function of those newly born during the first years. Thus,

$$S_o = L_o/l_o \qquad (4)$$

and F_x can be implemented by:

$$F_x = 1/2\left[f_x + (L_{x+1}/L_x)\cdot f_{x+1}\right] \qquad (5)$$

where f_x is the age specific birth rate at age x.

Two Interregional Models

The Leslie-Keyfitz model is dynamic but essentially aspatial, based on the assumption that population is a closed system with no migration. The model has to be modified to be useful in studies of interregional distribution of a population.

A Markov-chain model similar to Equation (2) was used by Tarver and Gurley to analyze the 1955-60 interregional mobility in order to make long-range projections of the populations of the nine census divisions of the United States. In their model the projection matrix is implemented by the transition probabilities calculated from data on the interdivisional migration streams, Let us call it a M matrix, where the element m_{ij} denotes the probability of moving from region i to region j. Thus, in a n-region model, the matrix is:

$$M = \begin{bmatrix} m_{11} & m_{12} & \cdots & m_{1n} \\ m_{21} & m_{22} & \cdots & m_{2n} \\ m_{n1} & m_{n2} & \cdots & m_{nn} \end{bmatrix} \qquad (6)$$

Then, the population in various regions is projected by raising the power of this matrix and multiplied by the regional populations in 1960.

Despite their success in analyzing a process of theoretical interest, Tarver and Gurley (1965: 139) expressed their dissatisfaction with the stringent assumptions they had to make. They concluded:

> The Markov process is a rather elegant model to specify the future population distribution of various geographic areas which will result, providing the migration patterns observed in a recent interval continue indefinitely. It is concise, for it includes only one of the three components (births, deaths, and migrants) affecting future population numbers. One limitation is that the model excludes children born and deaths after the observed period, since it applies to a constant total population. A modification in the model to allow for varying natality and mortality processes in each area would make the technique much more realistic, for natural population increase tends to be highest in areas having the highest out-migration tendencies.

A model proposed by Andrei Rogers (1968) can be viewed as an attempt to remedy these limitations of Tarver and Gurley's model. It is relatively a more realistic approach in that birth and death are incorporated with migration. The three components are integrated into what he calls a "multi-regional matrix growth operator." The procedure of constructing the model is basically as follows:

1). The Leslie-Keyfitz approach is extended by adding an in-migration component to their model. Thus,

$$K_{t+1} = S\cdot K_{t+1} + N_t \qquad (7)$$

where N_t is a vertor denoting the number of migrants into each region.

2). The in-migration component is obtained by:

$$N_t = M\cdot K_t \qquad (8)$$

where M is the migration transition matrix, as defined by Tarver and Gurley.

3). Substituting Equation (8) in Equation (7), an interregional matrix growth operator G is obtained:

$$K_{t+1} = S\cdot K_t + M\cdot K_t = G\cdot K_t \qquad (9)$$

The operator G is a supermatrix with the number of regions as its dimension, and within each region there is a submatrix with the number of age groups as its dimension. For instance, in a model of two regions (e.g., California and rest of the United States) and of three age groups within each region, the supermatrix can be expressed as:

$$\begin{array}{cc} & \begin{array}{cc} \text{Region 1} & \text{Region 2} \end{array} \\ \begin{array}{c} \text{Region 1} \\ \\ \\ \text{Region 2} \end{array} & \left[\begin{array}{ccc|ccc} S_o\cdot F_1 & S_o\cdot F_2 & S_o\cdot F_3 & 0 & 0 & 0 \\ S_1 & 0 & 0 & m_1 & 0 & 0 \\ 0 & S_2 & 0 & 0 & m_2 & 0 \\ \hline 0 & 0 & 0 & S_oF_1 & S_oF_2 & S_oF_3 \\ m_1 & 0 & 0 & S_1 & 0 & 0 \\ 0 & m_2 & 0 & 0 & S_2 & 0 \end{array}\right] \end{array} \qquad (10)$$

$$X \begin{bmatrix} K_1^t \\ K_2^t \\ K_3^t \\ \hline K_1^t \\ K_2^t \\ K_3^t \end{bmatrix} = \begin{bmatrix} K_1^{t+1} \\ K_2^{t+1} \\ K_3^{t+1} \\ \hline K_1^{t+1} \\ K_2^{t+1} \\ K_3^{t+1} \end{bmatrix} \qquad \text{(10 cont.)}$$

In structure Roger has made a mechanical synthesis of the Leslie-Keyfitz model, on the one hand, and Tarver and Gurley's model, on the other. The regional populations in each age group are estimated by using the original regional populations as distribution vector, post-multiplying it to the fertility-mortality matrix S and migration-transition matrix M, respectively, and summing the results of the two multiplications. An obvious weakness of this calculation is that it does not allow for the interaction between the numbers of survivors and of migrants. The population of any region, it is true, comprises two components, survivors and in-migrants, but survivors also migrate during the period, and migrants follow a certain mortality schedule. Moreover, it would seem that Rogers's model takes into consideration only the input of migration, which is relatively acceptable in a two-region model. But in a multi-regional model Rogers's formulation is rather cumbersome. It is theoretically more appropriate to consider simultaneously the in-migration to and the out-migration from each region.

An Alternative Model

It appears to be preferrable to adopt a multiplicative interregional model, rather than the additive model that Rogers suggests. The idea has been discussed by Feeney (1970) and independently by Li (1970). Following the Leslie-Keyfitz formulation, an interregional model can be constructed that is structurally based on the cohort-survival analysis, as depicted by a Lexis diagram--in other words, on the postulate that whenever a cohort reaches a certain age, it will assume that age's rates of fertility, mortality, and interregional migration. Consequently, the population at any given time can be estimated by summing the various cohorts at that time.

In distinction from Rogers's formulation, our supermatrix is arranged according to the number of age groups, and any element in the supermatrix is a submatrix whose dimensionality is determined by the number of regions. The elements in the first row of the supermatrix designate the probabilities that in each region any surviving non-migrant, in-migrant, and out-migrant will bear a live child. According to the multiplicative model, a migrant woman tends to follow the fertility-mortality schedule of the region of her origin, an assumption that can be substantiated by demographic evidence from a number of sources.

Thus, within a given age cohort x, the submatrix of two-region model can be mathematically expressed as follows:

$$R_x = \begin{bmatrix} S_oF_x^1 & 0 \\ 0 & S_oF_x^2 \end{bmatrix} \begin{bmatrix} m_{11}^x & m_{12}^x \\ m_{21}^x & m_{22}^x \end{bmatrix} = \begin{bmatrix} m_{11}^x S_oF_x^1 & m_{12}^x S_oF_x^1 \\ m_{21}^x S_oF_x^2 & m_{22}^x S_oF_x^2 \end{bmatrix} \qquad (11)$$

where m_{ij} is an element of the migration transition matrix; $S_oF_x$is the number of surviving births per female, as before. Also note that the sub- or super-script x denotes age, and the subscripts 1 and 2 denote the regions.

The diagonal elements of the supermatrix are constituted by the joint probabilities of migration transition and survivorship. Thus, for a given age cohort x, the two-region submatrix will be:

$$Q_x = \begin{bmatrix} S_x^1 & 0 \\ 0 & S_x^2 \end{bmatrix} \times \begin{bmatrix} m_{11}^x & m_{12}^x \\ m_{21}^x & m_{22}^x \end{bmatrix} = \begin{bmatrix} m_{11}^x S_x^1 & m_{12}^x S_x^1 \\ m_{21}^x S_x^2 & m_{22}^x S_x^2 \end{bmatrix} \qquad (12)$$

Incorporating these two components into a cohort surviving supermatrix, we obtain a model with which we can project over a given future period the regional populations by age, sex, and color. The model can be expressed as:

$$K_{t+1} = G \cdot K_t \qquad (13)$$

where G is a supermatrix that includes such submatrices as:

$$G = \begin{bmatrix} R_1 & R_2 & R_3 \\ Q_1 & 0 & 0 \\ 0 & Q_2 & 0 \end{bmatrix} \qquad (14)$$

and K is a supervector with three elements (e.g., age groups) within each subvector of two elements (e.g., regions).

An Application

The model presented in Equation (13) can be tested using the U. S. data for 1960. Census and vital statistics publications give the following information, necessary for the implementation:

1). interdivisional migration transition matrix by age, sex, and color, computed from 1960 census data.

2). age-specific birth rates by color for each division, available from U. S. National Center of Health Statistics (1968).

3). survival ratios by age, sex, and color, obtained from the life table of each division.

4). population by age, sex, and color for each division, from the 1960 census.

These data represent the necessary inputs for the model. Table 1 illustrates for white females aged 20-24 the results of implementing the matrix of R_x in Equation (11). A 9 x 9 matrix is derived because there are 9 census divisions in the United States.

An application of the matrix R_x (such as shown in Table 1) to the female population yeilds estimates of total births, rather than only female births, since age-specific total birth rates are used, (rather than age-specific female birth rates). Consequently, to obtain separately the estimated female births and male births, the figures have to be adjusted with a constant sex ratio at birth. The estimated number of male births can then be substituted in the first row of the supermatrix for males.

Obviously, the model presented at Equation (13) can be carried out as an iterative process, for each division's population by age, sex, and color can be projected for every 5 years. Table 2 shows how the model can contribute to a presentation of the population redistribution among the various regions. For instance, with respect to this particular age group, 20-24, the North East Central, Mountain, and Pacific divisions are growing relatively and the other divisions are declining or, with fluctuations, remaining more or less constant.

A Suggestion

Our model does not take international migration into account. Since the United States has a relatively stable inflow of international migrants, we should add this variable to the analysis of interregional population redistribution.

Assume that immigrants follow the same pattern of fertility, mortality, and interregional flow as natives. In other words, the interregional growth operator G can be applied to both the immigrants and natives. Second, assume that the regional distribution of the immigrants is known for every age-sex-color group, and constant for every time period (which may not be too far from the truth in any country with a restricted immigration policy.)

Thus, the immigrant distribution can be represented in a supervector structurally similar to K in Equation (13). Le us denote it as I. Then the distribution of total regional population, combining the natives and immigrants, for the nth period from the starting point t will be:

$$K_{t+n} = G^{n} \cdot K_t + \sum_{j=0}^{n} G^{j} \cdot I_t \qquad (15)$$

The empirical implementation of this model is currently in progress. It is obvious that the results can yield a more realistic projection of regional population, especially in those countries where a great amount of immigration occurs.

References

Feeney, Griffith M.
1970 "Stable Age by Region Distributions," Demography 7: 341-348.

Keyfitz, Nathan
1964 "The Population Projection as a Matrix Operator," Demography 1: 57-73.

Table 1:
Interregional Growth Submatrix
White Females, 20-24

	1	2	3	4	5	6	7	8	9
1	0.195	0.009	0.004	0.001	0.007	0.001	0.002	0.001	0.006
2	0.004	0.192	0.006	0.001	0.008	0.001	0.001	0.001	0.004
3	0.001	0.003	0.224	0.005	0.005	0.002	0.002	0.002	0.007
4	0.001	0.001	0.011	0.204	0.003	0.001	0.006	0.009	0.014
5	0.002	0.005	0.009	0.002	0.184	0.004	0.003	0.002	0.005
6	0.001	0.002	0.021	0.002	0.016	0.175	0.008	0.002	0.005
7	0.001	0.001	0.005	0.006	0.005	0.003	0.214	0.006	0.010
8	0.002	0.003	0.010	0.013	0.007	0.002	0.016	0.129	0.059
9	0.002	0.003	0.008	0.007	0.006	0.002	0.008	0.015	0.185

Sources: Vital Statistics Rates in the United States, 1940-1960, Table 20, pp. 142-145, Life Tables for the Geographic Divisions of the United States: 1959-61, Tables 1 to 36. Lifetime and Recent Migration, Table 6.

Table 2:
Population Distribution by Census Division,
White Females, Age 20-24

Census Division	1965	1970	1975	1980
1	6.1	6.1	6.3	6.3
2	17.9	18.6	19.4	19.3
3	20.8	21.8	23.5	22.5
4	9.3	9.4	10.1	9.6
5	12.8	12.6	12.1	11.2
6	15.1	7.0	6.8	6.5
7	9.2	14.0	10.0	9.4
8	2.4	2.7	3.1	4.0
9	6.3	7.8	8.6	11.0
Total	100.0	100.0	100.0	100.0
Population	5,936,038	6,996,386	7,238,838	7,014,382

Sources: 1960 Census of Population: U. S. Summary, Part I, Table 59.

1968 Introduction to the Mathematics of Population, Reading, Mass.: Addison-Wesley Publishing Co.

Leslie, P. H.
1945 "On the Use of Matrices in Certain Population Mathematics," Biometrika, 33: 183-212.

Li, Wen L.
1970 "Matrix Analysis of Migration Streams," Paper presented at the Population Association of America Meetings, April, 1970.

Pressat, Roland
1961 L'analyse démographique, Paris: Presses Universitaires de France.

Rogers, Andrei
1968 Matrix Analysis of Interregional Population Growth and Distribution. Berkeley, Calif.: University of California Press.

Tarver, James D., and William R. Gurley
1965 "A Stochastic Analysis of Geographic Mobility and Population Projections of the Census Divisions in the United States," Demography 32: 134-139

U. S. Bureau of the Census
1963 1960 Census of Population: United States Summary. Washington, D.C.: U. S. Government Printing Office.

1963 1960 Census of Population: Lifetime and Recent Migration. Washington, D.C.: U. S. Government Printing Office.

U. S. National Center for Health Statistics
1965 Life Tables for the Geographic Divisions of the United States: 1959-61, Public Health Service, Publication No. 1252. Washington, D.C.

1968 Vital Statistics Rates in the United States, 1940-1960. Washington, D.C.: U. S. Government Printing Office.