METHODS OF MIGRATION ANALYSIS

by

ANITA GAYLE SCHROEDER

B.S., Baker University, 1966

A MASTER'S REPORT

submitted in partial fulfillment of the

requirements for the degree

MASTER OF SCIENCE

Department of Statistics
and Computer Science

KANSAS STATE UNIVERSITY
Manhattan, Kansas

1968

Approved by:

H. C. Fryer

Major Professor

TABLE OF CONTENTS

I. Introduction

A. Definitions

Barclay (1958) defines demography as the "numerical portrayal of a human population." According to Thomlinson (1960), demography consists principally of three concerns: the measurement (especially numerical), the causes, and the consequences of the actions of an aggregate of persons. For the sociologist, measurement and measurement analysis are only the initial steps toward understanding a human process; but for the statistician, analysis is an end in itself. This paper will focus on the methods used in measuring and predicting migration, one of the more important demographic variables. The paper is intended primarily as a survey of the various methods of migration analysis which are currently available to the population researcher.

Several concepts prevalent in the demographic vocabulary should be clarified. A migrant may be defined as a person who has changed the location of his residence, severing all previous community ties. Since it is difficult to collect data on "community ties", a more practical definition of a migrant, given by Bogue (1957) is "a person who crosses a boundary [usually political] in changing residence." This paper will be concerned primarily with internal migration, or migration which does not involve international boundaries.

In general, internal migration may be studied from two different aspects - that of migration streams and that of migration differentials. Migration stream analysis is concerned with the volume and direction of place-to-place movements, while migration differential analysis deals with the differences in the characteristics of migrants and nonmigrants and the differences among migrant subgroups. According to Rogers (1966), the analysis of migration streams is "concerned primarily with the effect that variations in environmental conditions at origins and destinations have on volumes of flow, the study of differentials is concerned with the traits of migrants in various age-sex-income-race classifications."

Other factors given by Bogue (1957) which must be determined in measuring migration are the boundaries of residence areas, the time intervals to be considered, and the classification of origins and destinations of the migrants. The comparability of any migration data depends on the constancy of the boundaries involved, the equality of the time intervals, and the use of similar classifications of origins and destinations. Also the study of migration with respect to a given area or community must differentiate among in-migration, out-migration, and net migration.

B. Standard migration rates

In reference to volume of migration, demographers usually prefer the use of relative numbers instead of enumerated numbers, since relative numbers give more valid comparisons and projections. In general, the numerator of a demographic rate must be the total number of events classified as having occurred, and the denominator must be the total population exposed to occurrence or to the risk of the event. The following ratios cited by Bogue (1957) have been used to compute crude migration rates:

$$\text{crude in-migration rate} = \frac{\text{migrants received}}{\text{population of receiving area}}$$

$$\text{crude out-migration rate} = \frac{\text{migrants lost}}{\text{population of sending area}}$$

$$\text{crude net migration rate} = \frac{\text{in-migrants minus out-migrants}}{\text{population of given area}} \quad .$$

If attention is to be focused on the stream of movement between two areas instead of on a particular area, both the origin and destination areas are used in computing the migration rate. For example, the following rates, given by Bogue (1957), have been used:

$\frac{M \cdot k}{O \cdot D}$ estimates the number of migrants per unit of population at origin per unit of population at destination, multiplied by a constant,

$\frac{M}{\sqrt{O \cdot D}}$ estimates the number of migrants divided by the geometric mean of the origin and destination populations, and

$\frac{M}{\frac{1}{2}(O+D)}$ estimates the number of migrants divided by the arithmetic mean of the origin and destination populations,

where

M = number of migrants (in one direction only)

O = population at origin

D = population at destination

k = constant.

These rates are not valid if either the sending or receiving population is zero, or approaches zero.

C. Inadequecies of data and analysis

Migration studies have always been hampered by the lack of accurate data and by the absence of qualified statistical information and analysis.

It is difficult not only to collect and record migration data, but also to define exactly what constitutes a migrant. (Political boundaries, boundary changes, a mover's purpose, time periods involved, and the occurrence of multiple moves must all be considered in defining a "migrant".) Once data are obtained, simple migration is usually given in terms of the number of migrants from A to B in a given period of time, qualified to some unknown extent by the time periods involved, the number of boundary lines, and the distances moved. Thus while calculation of the crude standard migration rates given above is quite straightforward, interpretation is at best subjective, and often incoherent. It is clear that more sophisticated statistical techniques are needed.

The difficulties and deficiencies existent in migration data and analysis are numerous and not all of them are mentioned above. The preceding summary is intended only as a brief introduction to the concepts and terms used in the field of migration study.

II. Migration Models

A. E. G. Ravenstein

Ravenstein (1885) made one of the first attempts to determine some of the quantitative rules which govern migration. He listed the following "laws of migration", where absorbing and dispersing centers are receiving and sending areas respectively:

1. The great body of migrants will move only a short distance, causing a kind of universal displacement which produces "currents of migration" in the direction of large industrial and commercial centers.
2. A growing town will "recruit" the inhabitants of the surrounding neighborhoods, whose places are then filled by migrants from more remote districts. The process continues step-by-step to the most remote areas of the country.
3. The process of dispersion is the inverse of that of absorption and exhibits similar features.
4. A main current of migration will produce a compensating counter current.
5. Long-distance migrants will usually go to the large commercial and industrial areas.
6. In general, residents of towns are less apt to migrate than rural residents.
7. Females are more likely to migrate than males.

Some of these laws are currently valid, and some are not, but more importantly, they represent a functional scheme in which several independent variables are related to the dependent variable, migration.

B. Models for migration stream analysis

1. Samuel Stouffer - "intervening opportunities" hypothesis

Stouffer (1940) was one of several later migration analysts who not only formulated a functional scheme for migration, but also defined a relationship between the dependent and independent variables. He proposed that "the number of persons going a given distance is directly proportional to the number of opportunities at that distance and inversely proportional to the number of intervening opportunities."

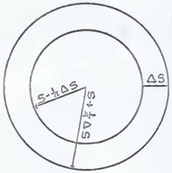

Figure 1. Circular model of intervening opportunities.

Symbolically, if Δy is the number of persons moving from the origin to the circular band with width Δs, if x is the cumulated number of opportunities between the origin and distance s, and if Δx is the number of opportunities within the circular band of width Δs, then

(1) $$\frac{\Delta y}{\Delta s} = \frac{a}{x}\frac{\Delta x}{\Delta s} \quad .$$

Note that here we are considering $\frac{\Delta x}{x}$, or the ratio of opportunities in the circular band (receiving area) to the intervening opportunities. Rather than a direct and invariant relation between mobility and distance, Equation (1) formulates a direct relation between mobility and opportunities. Mobility and distance may be related if desired in any auxiliary manner which expresses intervening opportunities as a function of distance.

If Equation (1) is written in terms of differentials, and if we assume that x is a continuous function, say f(s), then

$$\frac{dy}{ds} = \frac{a}{f(s)} \cdot \frac{d\ f(s)}{d(s)} .$$

Integrating, we have

$$(2) \qquad y = a \cdot \log f(s) + c .$$

If y is the cumulated number of movers between the origin and a circle of radius s, and f(s) is the cumulated number of opportunities within the circle, then Equation (2) states that the total number of persons who migrate to any point within the circle is directly proportional to the logarithm of the number of opportunities within the circle.

2. G. K. Zipf - the P_1P_2/D hypothesis

Zipf (1946) hypothesized that "the inter-community movement of goods (by value) and of persons between any two communities, P_1 and P_2, that are separated by an easiest transportation distance, D, will be directly proportionate to the product, P_1P_2, and inversely proportionate to the distance D." Zipf based his theory on the following principles. Assuming an equal average income and an equal proportion of gainfully employed, a given community, P, will contribute to the total production, C, of the system an amount proportional in value to P/C.[1] Community P will also receive from the system an amount that is proportional to P/C (during a certain time interval). Then if community P_1 has a share of the total flow of goods equal to P_1/C, and community P_2 during the same time interval has a share equal to P_2/C, then the value interchange between P_1 and P_2 would be proportional to

$$\frac{P_1}{C} \cdot \frac{P_2}{C} = \frac{P_1 \cdot P_2}{C^2} \quad .$$

Then using Lemma 1, Zipf obtained the final result.

Lemma 1. The number and sizes and locations of communities in a given social economy represent equilibria in the minimizing of work in transporting raw materials through industrial processes to consumers.

Thus a value interchange, or interchange of goods, between P_1 and P_2 should be inversely proportional to D, the easiest intervening transportation distance. Finally, the interchange, in value, for any two communities P_1 and P_2 will be directly proportional to P_1P_2/D. In equation form,

$$Y = \frac{P_1P_2}{D} \text{ , where}$$

Y = value of goods between any P_1 and P_2 .

---

[1] This theory applies only if all members of the population get an approximately equal share of the national income, i.e., the average real income per person is about the same in any community regardless of its size, and also if an approximately equal percentage of persons in each community are gainfully employed.

Zipf substantiated his hypothesis with the use of highway, railway, and airway travel data. As a result of his findings, he disagreed with Stouffer's general theory relating mobility and distance.

3. T. R. Anderson - comparison of Zipf and Stouffer

According to Anderson (1955), the hypotheses of both Zipf and Stouffer made use of the basic model,

$$m = a \cdot \frac{X}{Y}$$

where m is the number of migrants in a given stream, a is a constant, and X and Y are independent variables. For Stouffer,

$$m = a \cdot \frac{\text{"opportunities" or total in-migrants}}{\text{distance (as a function of intervening opportunities)}} .$$

In Zipf's model, assuming uniform income and unemployment,

$$m = a \cdot \frac{\text{population size of area}}{\text{distance (in terms of easiest intervening transportation distance)}} .$$

After testing the hypotheses of Zipf and Stouffer, Anderson concluded that Zipf's formula, especially, had several sources of error. He noted that the P_1P_2/D hypothesis did not take into account a migrant's inclination to stay in a given state or to move to areas of high employment. Anderson included the following results:

1) Distance as a function of intervening opportunities is no more accurate than distance in terms of highway mileage.

2) In Zipf's hypothesis, population size should be corrected for unemployment, more closely corresponding to the concept of "opportunities" in Stouffer's hypothesis.

3) The accuracy of Zipf's formula is reduced, as already noted, by the influence of state boundaries.

4) Zipf's formula is more accurate if both population size and distance are raised to powers other than one -- size to a constant power, and distance to a variable power.

Anderson suggested that the powers to which the basic variables X and Y are raised be treated as variables from one application to another, rather than being held constant. This idea agrees with the finding that Zipf's model is more accurate if the distance, Y, is raised to a variable power.

4. S. C. Dodd - the interactance hypothesis

Dodd (1950) adopted the basic P_1P_2/D hypothesis of Zipf, adding the dimensions of time, per capita activity, and a constant of interacting. Defining the index of interacting, I_o, as the <u>observed</u> number of interacts between each of the $\binom{n}{2}$ pairs in a set of n groups, and the index of interactance, I_e, as the <u>expected</u> interacting, Dodd then hypothesized that the correlation index, r_{eo}, between the observed interacting and the expected interactance would approach unity, where

$$I_e = \frac{k\ I_A\ P_A\ I_B\ P_B\ T}{L^1} \quad , \tag{3}$$

or I_e represents the $\binom{n}{2}$ calculated products of the seven observed factors:

T = total time of interacting,

P_A, P_B = populations of any two groups, A and B,

L^l = distance between the two groups A and B, where the exponent l (small L) weights the base factor,

I_A, I_B = the specific indices of level, or the per capita activity, characterizing each group or subset of groups in a unit period, and

k = a constant for each type of interacting in a given culture and period.

According to Dodd, the interactance hypothesis can be roughly stated as follows: "Groups of people interact more as they become faster, nearer, larger, and leveled up in activity." The hypothesis thus states the factors determining the quantity of group interaction regardless of the form of the activity. The converse of the hypothesis states that people will interact proportionately less as their groups (a) have fewer actions per period, (b) are further apart, (c) have smaller populations, and (d) have a greater diversity in average activity.

The interactance hypothesis includes the P_1P_2/D hypothesis as a special case. If the factors of time (T), the specific indices of level (I_A, I_B), and the constant of interacting (k) are either controlled or neglected (and hence are equal to unity), then Equation (3) reduces to

$$I_e = \frac{P_A P_B}{L^l} .$$

A further explanation of the factors I_A and I_B may be of interest. These indices of specific level are weighing factors which equate the heterogeneity of the groups A and B. The index of level for each group may be the average acting of a given kind in a unit period. For example, the average number of per capita phone calls handled by the exchange of a given city might represent the city's "telephonic" level for some period. In Equation (3), the average level is I_A (= $(\sum I)_A$ / P), where P is the number of persons in a group who are interacting in some way. When multiplied by P, the average level becomes the total number of acts, $(\sum I)_A$, for some period. Then,

$$(\sum I)_A = I_A \; P_A = \text{the "activity" of group A,}$$

for a unit period. It follows that the interactance of two groups in a unit period and a unit distance apart (i.e., T = 1, L = 1) would be in proportion to the product of their activities, or

$$(4) \qquad I_e = k \; I_A \; P_A \; I_B \; P_B = k \; (\sum I)_A \; (\sum I)_B \; .$$

Thus the index of level of activity, I, for each group may be found as the per capita activity of that group.

The subfactors which determine an index of specific level of activity may be such influences as age, income, occupation, etc. For Dodd, I_A and I_B represented those differential group characteristics which increased the correlation index, r_{eo}. In the future, other variables might be included as indices of specific level depending on whether or not their inclusion increases the index r_{eo}.

Support of Dodd's hypothesis has come from the inspection of scattergrams of interacting vs interactance, and from related studies of types of interacting. It has not yet been completely verified.

A direct application of the interactance hypothesis may be made to migration between communities. In this case, I_A and I_B might represent average moves per person between communities A and B in a given time period. The other factors in Equation (3) would be defined as they were previously.

5. Ralph Thomlinson - mathematical model

Thomlinson (1960) proposed a migration model which would enable demographers to compare migration rates by controlling for the following seven spatial factors: (1) size of area of origin, (2) size of area of destination, (3) shape of area of origin, (4) shape of area of destination, (5) distribution of population within area of origin, (6) distribution of population within area of destination, and (7) distance moved. The model was designed to estimate the number of migrants attributable to the seven spatial factors. If this model is satisfactory, demographers will be able to isolate those social factors in which they are interested by controlling for the spacial factors.

Thomlinson constructed his model in the following manner. Given a circle of radius r_i with center in area A, and with a segment of the perimeter of the circle falling in another area (B), then the probability that a person starting at the center of the circle and moving a distance r_i will be an inter-area migrant is given by

$$P = \frac{\text{length of arc in area B}}{\text{circumference of circle}} \quad .$$

Averaging this probability for all possible points in A would give the probability that a man starting from any point in A and moving r_i miles would go to area B, assuming A had a uniform density of population. Since this latter assumption is rarely met, a weighted average, obtained by multiplying the probability at each point times the population at that point, must be used.

In setting up his model, Thomlinson found that two short cuts were necessary to make the study practicable. First, he restricted the number of center points to as small a number as would give an acceptable amount of detail for area A. Second, he used a set of distance bands instead of computing every possible (theoretical) distance.

The relative "desirability" of a given area might be indicated by the ratio of actual to expected migration, using in, out, or net migration. For a symbolic representation of this design, consider Fig. 2, where A is a source area of irregular shape and density D, B is a terminal area of irregular shape, the distance of migration is r_i from the point of origin (h,k), and s_B is the arc of intersection of the circle with area B.

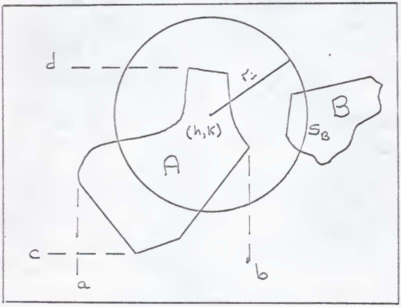

Fig. 2. Schematic representation of the model, Thomlinson (1960).

Then,

$$(5) \qquad P\,[\text{migrant settles in B} \mid \text{move of } r_i \text{ miles from } (h,k)]$$

$$= \frac{s_B\ D_{h,k}}{2\Pi r_i} \quad .$$

The probability that a migrant traveling r_i miles from a point in A settles in B is given by

$$(6) \qquad P(A,\ B;\ r_i) = \frac{\int_c^d \int_a^b s_{B_{xy}}\ D_{xy}\ dx\ dy}{2\Pi r_i \int_c^d \int_a^b D_{xy}\ ds\ dy} \quad .$$

Then if $M(A,\ B;\ r_i)$ is the total number of migrants from A to B, and $M(A;\ r_i)$ is the total number of migrants going r_i miles from A,

$$M(A,\ B:\ r_i) = P(A,\ B;\ r_i) \cdot M(A;\ r_i) \quad .$$

The total pattern of migration between all sub-areas of A and all sub-areas of B may be expressed by the following set of equations (for simplification, assume that there are three sub-areas $(a_1,\ a_2,\ a_3)$ of A, two sub-areas $(b_1,\ b_2)$ of B, and four distances $(r_1,\ r_2,\ r_3,\ r_4)$):

$$(7) \quad M(a_1,\ a_1) = P(a_1,a_1;r_1) \cdot M(a_1;r_1) + P(a_1,a_1;r_2) \cdot M(a_1;r_2)$$
$$+ P(a_1,a_1;r_3) \cdot M(a_1;r_3) + P(a_1,a_1;r_4) \cdot M(a_1;r_4)$$

$$M(a_1,\ a_2) = P(a_1,a_2;r_1) \cdot M(a_1;r_1) + P(a_1,a_2;r_2) \cdot M(a_1;r_2)$$
$$+ P(a_1,a_2;r_3) \cdot M(a_1;r_3) + P(a_1,a_2;r_4) \cdot M(a_1;r_4)$$

$$\vdots$$

$$M(b_2,\ b_2) = P(b_2,b_2;r_1) \cdot M(b_2;r_1) + P(b_2,b_2;r_2) \cdot M(b_2;r_2)$$
$$+ P(b_2,b_2;r_3) \cdot M(b_2;r_3) + P(b_2,b_2;r_4) \cdot M(b_2;r_4) \ .$$

There are 25 such equations in all. Migration between areas is represented as follows:

$$M(A,A) = \sum_{i}\sum_{j}\sum_{k} M(a_j,a_k;r_i)$$

$$M(A,B) = \sum_{i}\sum_{j}\sum_{k} M(a_j,b_k;r_i)$$

$$M(B,A) = \sum_{i}\sum_{j}\sum_{k} M(b_j,a_k;r_i)$$

$$M(B,B) = \sum_{i}\sum_{j}\sum_{k} M(b_j,b_k;r_i)$$

where $i = 1,2,3,4$; $j = 1,2,3$; $k = 1,2$.

Each of these equations may also be determined as the sum of several of the sub-equations of (7). For example,

$$M(B,B) = M(b_1,b_1) + M(b_1,b_2) + M(b_2,b_1) + M(b_2,b_2) .$$

In order to compute the expected frequencies of migration (for comparison with the observed frequencies), the following steps are necessary:

1) Determine the centers of the sub-areas, and find 2a below.
2) Multiply together
    a) proportion of circumference of circle falling in sub-area of B
    b) reciprocal of distance to base sub-area (area A)
    c) population of sub-area of origin (area A)
    d) population of sub-area of destination (area B).
3) Add together all products (one for each distance for every sub-area of A).
4) Add the sums of products for each base sub-area.

5) Divide this grand sum into the total known number of migrants between areas A and B, to get k.

6) Multiply k times each product in 2 above to get the expected number of persons who move between each base sub-area and each other sub-area.

The comparison of the computed and observed frequencies is then made through a series of indices. For each area, the following indices are computed:

$$I_o = \text{index of out migration} = \frac{\text{observed out migrants}}{\text{computed out migrants}} \cdot 100$$

$$I_I = \text{index of in migration} = \frac{\text{observed in migrants}}{\text{computed in migrants}} \cdot 100$$

$$I_N = \text{index of net migration} = \frac{\text{obs. net migr - comp. net migr}}{\text{computed gross migrants}} \cdot 100$$

$$I_L = \text{index of local migration} = \frac{\text{observed migrants within area}}{\text{computed migrants within area}} \cdot 100$$

$$I_G = \text{index of gross migration} = \frac{\text{observed gross migrants}}{\text{computed gross migrants}} \cdot 100$$

If A represents one <u>area</u> and beta all other areas, i.e., $\beta = B,C,D,\ldots,$ then

$$(8) \quad I_{O_A} = \frac{\text{Observed migrants from A to } \beta}{\sum_{ijk}\sum\sum M(a_j,\beta_k;r_i)} \cdot 100$$

$$I_{I_A} = \frac{\text{Observed migrants from } \beta \text{ to A}}{\sum_{ijk}\sum\sum M(\beta_k,a_j;r_i)} \cdot 100$$

$$I_{L_A} = \frac{\text{Observed migrants from A to A}}{\sum_{ij}\sum M(a_j,a_j;r_i)} \cdot 100$$

$$I_{N_A} = \frac{(\text{Observed migr from A to } \beta) - (\text{Obs migr from } \beta \text{ to A}) -}{\sum\sum\sum_{ijk} M(a_j, \beta_k; r_i) +}$$

$$\frac{M(a_j, \beta_k; r_i) - \sum\sum\sum_{ijk} M(\beta_k, a_j; r_i)}{+ \sum\sum\sum_{ijk} M(\beta_k, a_j; r_i)} \cdot 100$$

There appears to be little need for the calculation of I_{G_A}. Indices may also be computed for every possible move, giving

$$(9) \quad I_s = \text{index of simple migr} = \frac{\text{observed migrants}}{\text{computed migrants}} \cdot 100$$

$$I_n = \text{index of net migr} = \frac{\text{obs net migr - comp net migr}}{\text{computed gross migration}} \cdot 100$$

$$= \frac{\left[\begin{matrix}\text{(obs migr in}\\\text{dir of move)}\end{matrix} - \begin{matrix}\text{(obs migr}\\\text{opp move)}\end{matrix}\right] - \left[\begin{matrix}\text{(comp migr in}\\\text{dir of move)}\end{matrix} - \begin{matrix}\text{(comp migr}\\\text{opp move)}\end{matrix}\right]}{\text{computed gross migration}} \cdot 100$$

$$I_l = \text{index of local migration (identical with } I_L) .$$

Interpretation of the area and move indices is straightforward. An index of about 100 indicates average desirability, indices above 200 show great desirability, and indices under 50 indicate undesirable moves or areas.

Thomlinson noted that the composite index which he called F, the Force of Attraction, might be a better measure of migration than a single index. He used the following form:

$$F = I_I - I_O + 2I_N .$$

This index was intended only for use with areas.

6a. Richard Morrill - migration simulation

In his study of migration and urban settlement, Morrill (1965) presented a migration model which used simulation techniques. He stated that there were three important factors which affected the destination of migrants from a given area:

1) the distance of the (destination) area,

2) the differential attractiveness of the area, and

3) previous migration.

The relative attractiveness, called I_k, is an index of factors tending toward regular net gains or losses. (In this discussion, sources will be designated by $\underline{j}$, destinations by $\underline{k}$, both from 1 to n.) Let b_k be the total population of area k, b_{k_u} be the urban population, and $\bar{b}$ be the weighted mean population of all areas. Then,

$$I_k = \frac{b_k + b_{k_u}}{\bar{b}}, \text{ where } \bar{b} = \frac{\sum b_k + \sum b_{k_u}}{n} .$$

The probability of migration between two areas j and k is expressed by

$$(10) \qquad p_{jk} = \frac{a\ I_k}{d_{jk}^{\ b}} ,$$

where d_{jk} is the distance between j and k, and a is a constant weighing factor.

Variations in "regional attraction" are often reflected in the number of migrants, since populations with similar potential mobility will have different volumes of movement, depending on the number of opportunities in the surrounding areas. The actual migration volume may then be found by multiplying the basic sum of probabilities ($\sum_k p_{jk}$, or the sum of the probabilities of migrating from one area j to the k possible destinations) times the "ideal" number of migrants. If we let N_j represent the ideal number of migrants from j as a function of population and economic conditions, then the actual number of migrants from j is given by

$$M_j = N_j \sum_k p_{jk} \quad (11)$$

A summary of all migrations may be represented by

$$\sum_k \sum_j m_{jk}$$

which gives the total migrants from each area to every other area. The value m_{jk} is found by 1) normalizing $\sum_k p_{jk}$, i.e.,

$$k \sum_k p_{jk} = 1 = \sum_k p'_{jk} \; ;$$

2) converting

$$\sum_k p'_{jk} \rightarrow \sum_k |q_{jk}| \; ;$$

3) assigning random numbers to M_j,

$$M_j \simeq \sum r_j \; ;$$

and 4) selecting a migration path

$$r_j \simeq |q_{jk}| \rightarrow m_{jk}$$

until all M_j are assigned.

The total number of migrants to a single area k is represented by

$$\sum_j m_{jk} \; .$$

Then the net internal migration is given by

$$g_k = \sum_j m_{jk} - \sum_k m_{jk} \; (M_j)$$

where g_k is the net gain or loss.

6b. Programming of migration simulation

The migration model which Morrill (1965) designed involves a great deal of computation. The distances between every pair of areas must be calculated, and the probabilities of migrating from a given area to all other areas must be determined. Then the random numbers which were selected to represent the migrants must be matched against the accumulated probabilities, area by area, and finally the whole process must be repeated for each time period desired. Fortunately, this sequence of steps is not hard to program for a computer.

The program consists of the following basic steps:

1) Select the first area.
2) Compute the distance from this area to the second area by the formula $d_{12} = \sqrt{(x_1-x_2)^2 + (y_1-y_2)^2}$.
3) Enter this distance into the distance probability function (10) and compute p_{12}.
4) Multiply this probability times the index of attraction I_2, and store the product.
5) Repeat steps 2-4 for all other values of k from 3 to n.
6) Store these products and add them accumulatively in the order computed.
7) Multiply the total sum times the ideal number of migrants (N_1) to get the number of random digits needed.
8) Normalize the products to total 1 and express as a tally of discrete numbers.

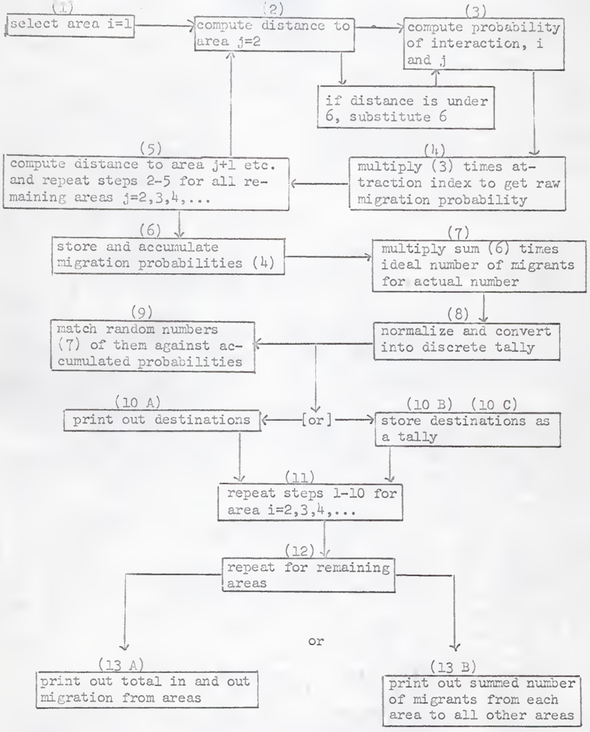

Fig. 3. Flow chart for simulation of migration program, Morrill (1965).

9) Match the random numbers against the accumulated products (8) to find the migration destinations.

10) Either print out the destinations or store them as a tally.

11) Select the second area and repeat steps 1-10.

12) Repeat for all remaining areas, 3 to n.

13) Depending upon available storage, the net total migration for an area, or the summed number of migrants from each area to each other area, may be printed out.

The flow chart for this program is illustrated in Fig. 3.

Morrill concentrated his research on the technique of simulation as a way of looking at human processes. Although his model was greatly simplified, and undoubtably left out some relevant factors, it is useful. Morrill pointed out that while the simulation technique was not to be used for its own sake, it could provide an "operational framework for explanations of spatial behavior."

7. A. T. Bharucha-Reid - stochastic models

It is possible to describe one-way migration models as finite birth processes, with the "birth" of an individual in the destination area being due to the migration of an individual from the origin area. The following models were presented by Bharucha-Reid (1960).

Consider two disjoint regions R_1 and R_2, and assume that (1) only one-way migration from R_1 to R_2 can occur, and (2) at time t=0 there are n_1 individuals in R_1 and n_2 in R_2. The models described here will determine the probability that in the time interval $[0,t)$, x individuals will migrate from R_1 to R_2.

Let $X(t)$ denote the number of migrants from R_1 who are in R_2 at time t. Note that at any time $t > 0$, there will be $n_2 + x$ individuals in R_2, where $x = 0,1,\ldots,n_1$. Next, let $p(x,t)\Delta t$ denote the probability that in the time interval $(t, t+\Delta t)$ one individual will migrate from R_1 to R_2, given that x individuals have migrated in the interval $[0,t)$. Then let $q_x(y,t)\Delta t$ denote the probability that in the interval $(t, t+\Delta t)$ exactly y individuals will migrate from R_1 to R_2, given the migration of x individuals in the time interval $[0,t)$. Assuming that the migrants are statistically independent, we get the following equations:

$$(12) \qquad q_x(1,t)\Delta t = p(x,t)(n_1-x)\Delta t + o(\Delta t)$$

$$(13) \qquad \begin{aligned} q_x(0,t)\Delta t &= 1 - q_x(1,t) + o(\Delta t) \\ &= 1 - p(x,t)(n_1-x)\Delta t + o(\Delta t) \ . \end{aligned}$$

Finally in order to get the probability that x individuals will migrate from R_1 to R_2 in the interval $[0,t)$, we use the migration probability function $p(x,t)$, and derive differential equations for $P_x'(t)$, where

$$P_x(t) = P\ [X(t)=x]\ , \quad x=0,1,\ldots,n_1\ .$$

The following models will give examples of this process.

(1) Let $P(x,t) = f(x)$ be some function of time which is subject only to the condition

$$(n_1-j)f(j) \neq (n_1-k)f(k) \qquad j,k = 0,1,\ldots,n_1\ .$$

In this case $P_x(t)$ satisfies the differential-difference equations:

$$\text{(14)} \qquad \frac{d\,P_x(t)}{dt} = \lambda_{x-1}\,P_{x-1}(t) - \lambda_x\,P_x(t), \quad x=1,2,\ldots,n_1$$

$$\frac{d\,P_0(t)}{dt} = -\,\lambda_0 P_0(t)$$

where $\lambda_x = (n_1-x)f(x)$. The solution of the above system is

$$P_x(t) = \beta_x \sum_{k=0}^{\infty} \left(\frac{1}{\alpha_{xk}}\right) e^{-\lambda_k t} \qquad x=1,2,\ldots,n_1$$

$$P_0(t) = e^{-\lambda_0 t}$$

where $\alpha_{xk} = \prod_{\substack{i=0 \\ i\neq k}}^{x} \gamma_{ik}$ for $k \leq x$,

$$\gamma_{ik} = \lambda_i - \lambda_k \quad ,$$

and $\beta_x = \prod_{i=0}^{x-1} \lambda_i$.

(2) In the second model, let

$$p(x,t) = \frac{f(t)}{n_1-x} \qquad \text{for } x < n_1$$

$$= \;0 \qquad\qquad x=n_1 \quad ,$$

where $f(t)$ is an almost arbitrary function of time. $P_x(t)$ will satisfy Equation (14) with

$$\lambda_x(t) = f(t)\;(n_1-x)^{-1} \qquad x=0,1,\ldots,n_1-1 \quad .$$

The solution for this λ-value is

$$P_x(t) = \frac{(\Lambda(t))^x}{x!} e^{-\Lambda(t)} \qquad x=0,1,\ldots,n_1-1$$

where $\Lambda(t) = \int_0^t f(\tau)\, d\tau$. If $x = n_1$, the solution is

$$P_{n_1}(t) = 1 - \sum_{x=0}^{n_1-1} P_x(t) .$$

For this particular model, $P_x(t)$ is the Poisson distribution. If two-way migration were to be considered, the migration models could be formulated as birth-and-death processes.

C. Models for migration differential analysis

1. James D. Tarver - migration prediction

Tarver (1961) proposed a migration model which would predict the movement of people during a given time period, in order to test his hypothesis that demographic, economic, and social conditions were interdependent in explaining social mobility. According to his model, "migration among the various subdivisions of states, or of the United States, is a corollary of the distinctive demographic, economic, and social structure existing therein." Tarver also presented a technique which partitioned the coefficients of multiple correlation, based on three sets of structural variates, into their independent and interactive components.

For his study, Tarver let the dependent variable (Y) be the state net migration rate, and the independent variables (X_i) be the initial demographic, economic, and social variables. Thus Y is a function of the X-variates, or

$$Y = f(X_1, X_2, \ldots, X_n) .$$

The following mathematical model was used:

$$(15)\quad Y_{hiaj} = \mu + (Z_1)_h + (Z_2)_i + (Z_3)_a + (Z_1,Z_2)_{hi} + (Z_1,Z_3)_{ha}$$

$$+ (Z_2,Z_3)_{ia} + (Z_1,Z_2,Z_3)_{hia} + \varepsilon_{hiaj}$$

where Z_1 contains the economic variables, Z_2 the social variables, Z_3 the demographic variables, and $\varepsilon \sim N(0,\sigma^2)$. The various groupings of the variables are given in Table 1. In this case regression analysis was used to find a solution.

	Z_1 - Economic Variables
X_1	Percent of employed persons working in the construction industry
	Z_2 - Social Variables
X_2	Median years of school completed by persons 25 years and over
X_3	Percent of population, 25 years and over, completing four or more years of high school
	Z_3 - Demographic Variables
X_4	Percent of population under 20 years of age
X_5	Percent of population white
	Y
Net migration rate, 1940-1950	

Table 1. Dependent and independent variables, by states, Tarver (1961).

Tarver postulated that not only was Y dependent on the X_i's, but also that Y was some linear function of the X_i's, and that interaction was present among the independent variables. Regressions were computed using a step-wise procedure.[2] Table 2 shows the results of the analysis.

Regression Number	Function	R^2	Component
1	$Y = f(Z_1)$	.6948	$r_{y.1}^2$
2	$Y = f(Z_2)$	.4089	$r_{y.2}^2$
3	$Y = f(Z_3)$	.1964	$r_{y.3}^2$
4	$Y = f(Z_1, Z_2)$	.8588	$R_{y.12}^2$
5	$Y = f(Z_1, Z_3)$	.7204	$R_{y.13}^2$
6	$Y = f(Z_2, Z_3)$	.5431	$R_{y.23}^2$
7	$y = f(Z_1, Z_2, Z_3)$	.9779	$R_{y.123}^2$

Table 2. Functional relations and R^2 for final equation (white population), Tarver (1961).

The results show, for example, that the two social variables in Z_2, both of which are indices of educational attainment, explain about 41% of the variation in Y. The most significant correlation coefficient is that

[2]This technique enters the independent variable having the highest simple correlation with the dependent variable in the regression analysis first; thereafter, it selects, successively, the next independent variable having the highest explanatory power for the residuals. It excludes the variables which are linear combinations of others, as well as those which insignificantly augment the coefficients of multiple correlation.

for Equation 7, which contains all three types of index variables. Thus the demographic, economic, and social variables, considered together, explain over 97% of the net migration rate variations among the white population.

The analysis demonstrates that migration is a composite of inter-related demographic, economic, and social factors. Economic factors are clearly more significant than demographic and social factors in explaining net migration.

3. E. S. Lee - census survival and state-of-birth data

Applied studies in demography are often limited to the type of data available for analysis. It is for this reason that most studies of migration in the United States are based upon the U.S. Decennial Census, virtually the only complete source of U.S. population data. The following two methods of analysis, given by Lee (1957), have more practical value then they have theoretical interest, and they are included only for completeness.

Census survival ratios may be computed in two forms, forward or reverse. A forward census survival ratio is the ratio of the number of persons in an age-sex group of a closed population at a given census to the number of persons ten years younger at the preceding census. For example,

$$r = \frac{\text{native white males aged 20-24, U.S., 1940}}{\text{native white males aged 10-14, U.S., 1930}} .$$

This ratio, when multiplied by the enumerated native white males aged 10-14 in each state in 1930, gives an estimate of the number of persons aged 10-14 who, under the assumptions of this method would have lived to be enumerated in 1940. The difference between this expected population and the enumerated population for each state is defined as <u>net migration</u>. In symbols,

$$p' - rp = M_{for}$$

where M_{for} = (forward) estimate of net migration for the state

p' = state's enumerated native white male population aged 20-24 in 1940,

p = state's enumerated native white male population aged 10-14 in 1930, and

r = forward census survival ratio.

The reverse census survival ratio is actually the inverse of the forward ratio. Symbolically,

$$(p')\,\frac{1}{r} - p = M_{rev}$$

where M_{rev} is the (reverse) estimate of net migration for the state.

Several assumptions are inherent in census survival estimates of migration. First it is necessary that the census enumerations be consistent from one census to the next. Although this assumption is rarely if ever met, demographers usually ignore the discrepancy. (As the data is manipulated and ratios computed, enumeration errors tend to cancel each other, at least in part.) It is also assumed that the census survival ratio is the same for each state as for the nation. Since this assumption can be tested directly, it is possible to adjust for discrepancies in the ratios.

For a state-of-birth analysis of migration, the following data may be obtained directly from census tabulations:

1) the number of persons born in each state and living there at the time of the census (the "nonmigrants"),
2) the number of persons living in each state and born in any other state, and
3) the number of persons born in each state and living in any other state.

Using these values, the following derived measures can be obtained:

1) the net gain and loss in the interchange between two states,
2) the in-migrants into a state,
3) the out-migrants from a state, and
4) the "birth-residence index", found by subtracting (3) from (2) directly above.

State-of-birth indicators are obviously incomplete and vague, for they fail to specify exact time periods and they omit the effects of certain types of mortality as well as the occurrence of return migrations. In spite of these defects, state-of-birth methods are invaluable for practical studies, since they make use of the most readily available and the most complete data source, the U.S. Census. Lee, et al. (1957), Tarver (1962), Zachariah (1962), and Siegel and Hamilton (1952) give more complete investigations of census-related migration models.

4. Andrei Rogers - Markovian model

Parzen (1962) defines a stochastic process as a random phenomenon that arises through a process which is developing in time in a manner controlled by probabilistic laws. If we accept this definition, then a

Markov process is defined as follows. Let $\{X(t), t=0,1,\ldots\}$ be a discrete parameter stochastic process[3] such that for any set of n time points $t_1 < t_2 < \ldots < t_n$ in the index set of the process, the conditional distribution of $X(t_n)$, for given values of $X(t_1),\ldots,X(t_{n-1})$, depends only on $X(t_{n-1})$. Symbolically,

$$\begin{aligned} & P[X(t_n) \leq x_n \mid X(t_1) = x_1,\ldots,X(t_{n-1}) = x_{n-1}] \\ (16) \qquad & \\ & \qquad = P[X(t_n) \leq x_n \mid X(t_{n-1}) = x_{n-1}] \ . \end{aligned}$$

Then $\{X(t), t \geq 0\}$ is said to be a Markov process.

Markov chain theory has provided a useful tool for analyzing mobility in all forms - social, industrial, and geographic. While the concepts involved in Markovian analysis have had only a limited success in migration stream analysis, they have proved to be very useful in differential analysis. Rogers (1966) investigated migration differentials in California, and the following section is based on his work.

Suppose we have a geographical system of m regions with a population consisting of n cohorts. Then for some cohort r, assume each member of the cohort behaves independently of the other members and according to an m x m transition matrix P_r. Each element of P_r may be estimated by the following formula

$$(17) \qquad {}_rP_{ij} = \frac{{}_rk_{ij}}{\sum_{j=1}^{m} {}_rk_{ij}} \qquad r=1,\ldots,n;\ i,j=1,\ldots,m$$

[3] Here only discrete parameter processes are used, and the possibility of the occurrence of continuous parameter stochastic processes will be ignored.

where k_{ij} = number of persons moving from region i to region j in an unspecified time period. The following concepts are basic to the construction of a Markov model for migration: the cohort transition matrix, the mean first passage times, and the equilibrium vector (related to the property of a long-run or stationary distribution).

The mobility of a migrant class is illustrated in a cohort-specific transition matrix. If for example,

$$(18) \qquad P_r = \begin{array}{cc} & \begin{array}{cc} A & B \end{array} \\ \begin{array}{c} A \\ B \end{array} & \begin{bmatrix} 1/2 & 1/2 \\ 1/4 & 3/4 \end{bmatrix} \end{array} ,$$

then there is a 50% chance that an individual in region A will move to region B in a unit time period. There is a probability of only 1/4 however that an individual in region B will move to region A in the same time period. (Here it is assumed that A and B exhaust all possible moves.) Thus P_r is actually the matrix of probabilities given in Equation 19, where m and n are a unit time period apart and hence will be omitted in the future.

$$(19) \quad P_r = \begin{bmatrix} p_{A,A}(n,m) & p_{A,B}(n,m) \\ p_{B,A}(n,m) & p_{B,B}(n,m) \end{bmatrix} = \begin{bmatrix} p_{AA} & p_{AB} \\ p_{BA} & p_{BB} \end{bmatrix} ,$$

where $p_{A,B}(n,m) = P(X_m=B|X_n=A)$,

and $\sum_j p_{i,j}(n,m) = 1 \qquad$ for all i .

Then, for example, $p_{B,A}(n,m) = P(X_m=A|X_n=B)$ is the probability that an individual is residing in area A at time t=m given he had been a resident of area B at t=n, a unit time period before.

The diagonal elements of transition matrices give an indication of the non-mobility or "stability" of the populations in each area. If the probability given by the diagonal element for some area Q is low, then the relative mobility of area Q is high. Also if more than two regions were to be considered, then the relative "attraction" of different destinations for a given source area could be compared.

It is sometimes of interest to consider the length of time required for an individual to move from state i to state j for the first time, or the first passage time. The mean of the distribution of this random variable is called the <u>mean first passage time</u>.

Returning to Equation 19, consider the probability that an individual currently in region A will move for the first time to region B in n time periods. Call this probability $g_{AB}^{(n)}$. Beginning with n equal to 1,

$$g_{AB}^{(1)} = p_{AB}$$

$$g_{AB}^{(2)} = p_{AA} \cdot p_{AB}$$

$$= p_{AA} \cdot g_{AB}^{(1)} \quad .$$

Thus the probability of going from A to B for the first time in one time period is by definition p_{AB}. The probability of going from A to B for the first time in two steps is equal to the probability of remaining in A for the first time period multiplied by the probability of moving from A to B in the next unit period. Generalizing, we have

$$(20) \qquad g_{AB}^{(n)} = p_{AA} \cdot g_{AB}^{(n-1)}$$

$$= (p_{AA})^{n-1} \, p_{AB} \; ,$$

which gives the first passage time distribution. Since $p_{AA} = 1-p_{AB}$,

$$g_{AB}^{(n)} = p_{AB}\,(1-p_{AB})^{n-1} \ .$$

This function is the geometric distribution and hence the mean first passage time is

$$m_{AB} = \frac{1}{p_{AB}} \ .$$

The mean first passage times are described in matrix form by

$$M = \begin{bmatrix} m_{AA} & m_{AB} \\ m_{BA} & m_{BB} \end{bmatrix} \ .$$

In general, the elements of the matrix M may be found by the following formula due to Parzen (1962),

$$(21) \qquad m_{i,j} = 1 + \sum_{k \neq j} p_{i,k}\, m_{k,j} \ , \quad i \neq j \ .$$

Other methods for computing the mean first passage time matrix are given by Kemeny and Snell (1960) and Parzen (1962).

Mean first passage times provide a measure of contiguity based on interchange probabilities rather than distance. Thus they might be viewed as indices of "aspatial interregional distance". This measure of proximity is defined as "migrant distance".

From the transition matrix P we may derive information about the Markov process illustrated above. For example, suppose we want to find the probability that an individual currently residing in area A will be in area B two unit time periods later. This event can occur only if

1) the individual stays in area A for the first time period and migrates to B during the second period, or

2) the individual first migrates from A to B at time t=1 and then remains in B during the next time period.

Thus for the matrix given in Equation 18,

$$p_{AB}^{(2)} = p_{AA}\, p_{AB} + p_{AB}\, p_{BB}$$

$$= (1/2)(1/2) + (1/2)(3/4)$$

$$= 5/8 \ .$$

In the same manner,

$$p_{AA}^{(2)} = 3/8 \ ,$$

$$p_{BA}^{(2)} = 5/16 \ ,$$

$$p_{BB}^{(2)} = 11/16 \ .$$

These values may then be put in the matrix

$$(22) \qquad P_r^{(2)} = \begin{array}{c} \\ A \\ B \end{array} \begin{array}{c} \begin{array}{cc} A & \quad B \end{array} \\ \begin{bmatrix} 3/8 & 5/8 \\ 5/16 & 11/16 \end{bmatrix} \end{array} \ ,$$

which describes movement between two periods of time. This result may be generalized to n time periods. Hence the transition matrix P_r completely determines the migration process. It is thus possible to use this data to compare present movement patterns, to project them into the future, and to determine the distributional consequences of a particular movement structure.

The nth order transition probabilities are easily derived by matrix multiplication. In particular the nth order transition probabilities of a transition matrix P may be found by multiplying P by itself, n number of times, or

$$P^{(n)} = P^n \quad .$$

As a demonstration of this property, note that

$$P_r^{\,2} = \begin{bmatrix} 3/8 & 5/8 \\ 5/16 & 11/16 \end{bmatrix} = P_r^{\,(2)} \quad .$$

The final concept to be considered is the <u>equilibrium probability vector</u>. It can be shown that as n increases, $p_{i,j}^{(n)}$, the probability of migrating from i to j in n time periods, approaches a limit Π_j, or

$$(23) \qquad \lim_{n\to\infty} p_{i,j}^{(n)} = \Pi_j \ ,$$

where Π_j is independent of i. A Markov chain with this property is said to possess a long-run distribution, and hence is stationary, or in "equilibrium". The values Π_j form a vector. This equilibrium vector gives an index of the long-term implications of current migratory behavior, providing the migrant doesn't die and the transition probabilities remain constant.

The Markovian model described by Rogers is based on strict and not necessarily valid assumptions concerning human behavior. (Transition probabilities are known to vary over time as well as space, and an individual will ultimately die). For this reason Markovian analysis must be limited in its usefulness as a long-term forecaster of interregional <u>flows</u> of migration. Nonetheless, when used for analyzing <u>differential</u> behavior during observed time periods, Markovian analysis provides insights not readily obtainable by other means.

III. Conclusion

This paper was intended as a survey of the different theories of migration prediction, measurement, and analysis. The models may be classified into two basic groups - those dealing with migration streams and those concerned with migration differentials. Examples of both types were presented, explained, and evaluated in the preceding paper.

It is evident that the migration models available for use vary in their sociological or statistical points of view, in their structural bases, and in their practicability and applicability to real life situations. The researcher should pick the model which fits his data and his hypothesis. The models of Stouffer (1940), Zipf (1946), and Dodd (1950), while similar, will yield varying results. Each also involves certain data which may not always be available. Stouffer's theory of intervening opportunities, for example, could be used when data are available on such "opportunities" as job or housing vacancies in a given area. The researcher might use Zipf's P_1P_2/D hypothesis when data on human movement from commercial travel sources or hotel registers are obtainable. Finally the use of Dodd's interactance hypothesis would involve some intuitive understanding of the factors contributing to a level of activity. In this instance, the researcher would need not only a complete data source, but also a thorough acquaintance with the economic and social aspects of the areas being studied.

The purpose of the research and the consequent information desired should be considered when a model is being selected. The intervening opportunities theory is useful, for example, in analyzing the observed differentials in the distances moved by persons from different occupational groups. (It may be noted that in this case Stouffer's theory could also be classified under migration differential analysis.) The theories of both Zipf and Dodd may also apply to other forms of movement than human migration. Zipf's hypothesis, for instance, may be applied to the movement of commercial goods between two areas, while Dodd's hypothesis may be used to study any form of interaction between areas. It is also important to remember that the hypotheses of Zipf and Stouffer conflict to some extent, and corrections for error may be necessary.

The models involving probability, specifically those of Morrill (1965), Bharucha-Reid (1960), Thomlinson (1960), and Rogers (1966), may be of interest to the researcher who is attempting to predict future migration, basing his prediction on data from past migrations. Although some of these models involve complex mathematics and unrealistic assumptions, they do offer valid indications of possible migration trends in both migration stream and migration differential analysis.

Other distinctions are also evident. The mathematical model given by Thomlinson (1960), the Markovian analysis and the census survival analysis, for example, may all be used directly on census data. The model given by Morrill was completely simulated. Also those models based on strict mathematical or statistical reasoning, such as the stochastic models presented by Bharucha-Reid and the Markovian model, must be differentiated from those of Stouffer and Dodd, which are more sociological. It might also be noted that the models of Stouffer, Zipf, Dodd, and especially Ravenstein (1885) represent some of the earliest efforts regarding migration analysis, while most of the others are more recent.

Migration methodology could be improved from the sociological standpoint as well as from the statistical. Demographers need to take a more definitive approach to the study of migration, i.e., they need to formulate more concise terminology and set up a relevant, coherent framework in which migration could be studied. Better definitions and greater clarity in the presentation of information concerning migration are definitely needed. Demographers will continue however to study migration from its many different aspects, and thus migration will probably remain a complex field, but it need not remain an incoherent one.

From the other direction, there is a need for statistical techniques and models which give good approximations with incomplete data. Due to inconsistencies in real life, there is usually a gap between theoretical and practical application. The solution could be more accurate data, or it might be better "approximation" techniques, and the latter seems to be the most relevant approach.

Probability models seem to hold the most promise for future developments in predicting migration. It may be that human as well as other forms of migration will occur according to a probabilistic design, much as natural phenomena follow mathematical laws. In this case, probability theory could serve as the mathematical tool necessary for a complete understanding of migration.

Other approaches to the study of migration might use the concepts involved in vector analysis, the theory of games, and directional derivatives. For instance, migration might be conceptualized as a two-person game between the resident of an area (who must make a decision whether or not to move) and his environment. Game theory might be applicable in this situation.

The approach to demography has become more closely associated with statistics in recent years. According to Landry (1945),

> "...la démographie comme histoire n'existerait pour ainsi dire pas sans la statistique. Quant à la démographie comme théorie, elle n'a pas dans la statistique non seul instrument; mais la statistique est pour elle, en diverses manières, un instrument précieux. C'est au développement et aux progrès de la statistique qu'elle est redevable, pour une grande partie, de son propre développement et de ses progrès. Les liens qui unissent aujourd'hui la théorie démographique et la statistique ne pouront que se reserrer davantage encore dans l'avenir."

LISTING OF FIGURES

LISTING OF TABLES

BIBLIOGRAPHY

Anderson, Theodore R. (1955). "Intermetropolitan migration: a comparison of the hypotheses of Zipf and Stouffer." American Sociological Review, 20:287-91.

Barclay, George W. (1958). Techniques of Population Analysis. New York, John Wiley and Sons.

Bharucha-Reid, A. T. (1960). Elements of the Theory of Markov Processes and Their Applications. New York, McGraw Hill.

Bogue, Donald J., Shyrock, Jr., Henry S., and Hoermann, Siegfried A. (1957). Subregional Migration in the United States, 1935-1940. Volume I: "Streams of migration between subregions." Scripps Foundation Studies in Population Distribution No. 5.

Dodd, Stuart C. (1950). The interactance hypothesis: a gravity model fitting physical masses and human groups." American Sociological Review, 15:245-56.

Kemeny, John G., and Snell, J. Laurie. (1960). Finite Markov Chains. Princeton, N. J., Van Nostrand.

Landry, Adolphe. (1945). Traite de Demographie, Payot.

Lee, Everett S., Miller, Ann Ratner, Brainerd, Carol P., and Easterlin, Richard A. (1957). Population Redistribution and Economic Growth, United States, 1780-1950. Volume I: "Methodological considerations and reference tables." American Philosophical Society.

Morrill, Richard. (1965). Migration and the Spread and Growth of Urban Settlement. Royal University of Lund, Sweden, Dept. of Geography. (Lund Studies in Geography. Series B: Human Geography, no 26.)

Parzen, Emanuel. (1962). Stochastic Processes. San Francisco, Holden-Day.

Ravenstein, E. G. (1885). "The laws of migration." Journal of the Royal Statistical Society, 48:167-235.

Rogers, Andrei. (1966). "A Markovian analysis of migration differentials." Proceedings of the Social Statistics Section, American Statistical Association, 542-66.

Siegel, Jacob S., and Hamilton, C. Horace. (1952). "Some considerations in the use of the survival rate method of estimating net migrations" Journal of the American Statistical Association, 47:475-500.

Stouffer, Samuel A. (1940). "Intervening opportunities: a theory relating mobility and distance." American Sociological Review, 5:845-67.

Tarver, James D. (1961). "Predicting migration." Social Forces, 39: 207-213.

Thomlinson, Ralph. (1960). A Mathematical Model for Migration: A Methodological Study to Improve the Quantitative Analysis. Ph.D. dissertation, Columbia University.

Zachariah, K. D. (1962). "A note on the census survival ratio method of estimating net migration." Journal of the American Statistical Association, 57:175-83.

Zipf, George K. (1946). "The P_1P_2/D hypothesis: on the intercity movement of persons." American Sociological Review, 11:677-86.

METHODS OF MIGRATION ANALYSIS

by

ANITA GAYLE SCHROEDER

B.S., Baker University, 1966

AN ABSTRACT OF A MASTER'S REPORT

submitted in partial fulfillment of the

requirements for the degree

MASTER OF SCIENCE

Department of Statistics
and Computer Science

KANSAS STATE UNIVERSITY
Manhattan, Kansas

1968

Migration may be analyzed from the basis of migration streams or that of migration differentials. Migration stream analysis is concerned with the volume and direction of place-to-place movements, while migration differential analysis deals with the differences among migrant subgroups. Most methods of migration analysis may be placed in one of these two categories.

This paper deals with some of the more characteristic attempts at migration prediction and measurement. Migration stream analysis is represented by the following models: Stouffer's theory of intervening opportunities, Zipf's P_1P_2/D hypothesis, Dodd's interactance hypothesis, Thomlinson's mathematical model, and the stochastic models given by Bharucha-Reid. Morrill's model for migration simulation is also included. The methods cited as examples of migration stream analysis include Roger's Markovian model, the census-related methods given by Lee, and Tarver's model for migration prediction.

This paper was intended primarily as a survey of the methods available for migration analysis. The method chosen for use in any one study will depend on the purpose of the research and on the available data.