CHAPTER VII

MIGRATION: A COMPUTER SIMULATION MODEL

RICHARD J. WALLAT

AND

THEODORE E. DOWNING

In an attempt to explore the potential of simulations for estimating and projecting migration, the Bureau of Ethnic Research has developed a computer model to mimic inter-state and international migration of Mexicans. This modelling exercise was aimed at providing a working theory of migration -- that could be experimentally manipulated and offer insights into the migration process. The simulation was designed as a general model capable of rapid modification so its usefulness would extend to the analysis of migration in other countries as well.

Structure of the Model

An Analogy

What is a simulation? How does it work? To explain, we offer an analogy drawn from a first-cousin of simulation, parlor games. Let's play a game called "Mexican migration." To begin, cover your living room floor with shoe boxes, each representing a state of Mexico. Partition the boxes into several parts, each representing different sectors of the state's economy, one for agriculture, another for manufacturing, yet another for construction, and so on. Next, you will need some marbles, one color to represent females, another for males. Write an age on each marble, corresponding to the number of individuals of each age in the country. These are called age and sex "attributes" of the population. Fill the partitions in each shoe box with as many marbles as there are males and females of each age in that economic sector. The total number of marbles in all the boxes, therefore, replicates the age-sex distribution of the entire population of the country at the start of the game, for example, 1960. Having laid out the game,set an alarm clock to buzz at fixed intervals, say every 20 minutes, and play as if the period between buzzes is equivalent to a year in real time.

The action begins by spinning a wheel to randomly select a state. Once selected, the wheel is spun again to select certain attributes of the state's population (for example, a male in the transportation sector between the ages of 20 and 24). Next, we move the marble between different boxes according to rules we have established concerning the likelihood that a person with those attributes will migrate. These rules are based on detailed observations (or hypotheses) about the past behavior of similar people under similar conditions. Obviously a large rule book is needed since the possible combinations of characteristics of people and economic sectors are large and the rules governing which kinds of people are moved can be complex.

We play the game over and over again; spinning the wheel, selecting a marble, consulting the rule book, and making the appropriate moves. Eventually, the alarm clock rings marking the end of a year's play. After repeating the game for 10 "years" the population has been redistributed, represented by the fact that many marbles are now in different boxes. We compare the final distribution with the original distribution to determine the correspondence of the simulated distribution to the correct resolution of the game. i.e. where the marbles should be after 10 years according to the known distribution in 1970. If the marbles turned up in the places other than where they should be in 1970, we rewrite some of the rules, and repeat the game. Each rule change is used to align the outcome to what is expected, making the game a little more realistic by discovering more accurate rules.

Just as playing the game of Monopoly tells us something about real estate, so also playing this game should tell us something about migration. By changing the rules and observing the outcome of many games, we could determine the sensitivity of the game to changes in particular rules. In this sense, the game would be a test of a theory (the rule book) as an explanation of migration. Repeated plays of the game would also indicate what rules lead to nonsensical outcomes. For instance, it might be discovered that a slight alteration of a rule quickly leads to the complete migration of all marbles to one box. This would indicate that the rule needs to be rewritten or counterbalanced by other rules which mitigate such unwarranted consequences.

It would take a long time, great patience, and a messy living room to play this hypothetical game. Fortunately, there is an alternative. A computer can be programmed to play the game, and migrate millions of imaginary people between imaginary states in minutes. Essentially this is what we have accomplished in this phase of the Bureau's Mexican Migration Project; we developed a computer program which closely resembles the geographical and population structure of federal states in Mexico and formulated rules which seem to govern migration between these states.

Theoretical Structure

For the purposes of this model, migration may be viewed as the movement of people across some arbitrarily defined boundary. The area which the migrant leaves will be called the origin; the area where the migrant arrives, the destination. Naturally, an area may be, simultaneously, a destination to some migrants, a point of origin to others. Like most migration theorists, we feel that migration is a consequence of many circumstances at the area of origin and destination, including varying opportunities and obligations facing different types of people. Above all, migration is selective; the same circumstances at an origin or destination will affect different types of people differently. Thus, the first task of our model is to take into account characteristics of the place of origin, destination, and the migrants themselves.

A theoretical structure underlies any game or model. This structure is formed by data requirements, assumptions, simplifications of reality, causal chains, correlations, concepts, and a host of other characteristics. The best way to see the theoretical structure of this model is to go through its sequence of operations, step by step.

To begin, we create a map of Mexico in the computer by digitizing a matrix to represent state boundaries. Since we are interested in inter-state migration in Mexico and Mexico to United States migration, this map consists of Mexico's 32 federal entities and the four United States border states. To form the initial conditions of the simulation, we select and store in the computer's memory thousands of values of attributes for Mexico's population and characteristics of states. Attributes and characteristics were selected which previous migration studies indicated would influence the volume of migration between states and the selection of migrants (Table 1). Next, we set the model's clock at 1960, the initial year for calculating the first set of migrations.

Table 7.1: Inputs to Mexican Migration Simulation

I. For Each State

1. Age and sex distribution in rural and urban areas 1960, 1970
2. Education distribution by sex 1960, 1970
3. Income distribution by sex 1960, 1970
4. Economic sector distribution by age 1960, 1970
5. Birth rates by year 1960, 1970
6. Death rates by age by sex 1970
7. Rural-urban death rates (mean of 1960 to 1970 period)
8. Net lifetime migration into and out of the state 1950-60

II. For Nation

Male to female birth ratio (mean of 1960 to 1970 period)

The operation of the model begins with the selection of one of Mexico's 32 federal entities as an area of origin (Figure 7.1). The selection is random, with a state's likelihood of selection being weighted according to its proportion of all of Mexico's population. For example, Jalisco had approximately 2,443,000 inhabitants in 1960, roughly seven percent of Mexico's 34 million people. Therefore, in the selection process, Jalisco has seven chances out of 100 of being chosen. This sampling procedure guarantees that all Mexicans have an equal opportunity to become migrants, although this does not mean that all states will have equal proportions of their populations actually migrating, as will soon become apparent.

Once a state is selected, we define a particular group of people as potential migrants. To do this, we calculate the proportions of people in a state's population with different attributes of sex, age, income, education, rural or urban residence, and who are members of the various sectors of the labor force. We then randomly select a specific value for each attribute; one for sex, age, income, and so on. Once again, this selection is proportionate to the frequency of each attribute in the state's population. For example, Jalisco's population has approximately 16 percent of its people in the agricultural sector, therefore, in the simulation agriculturalists had approximately 16 chances out of 100 of being considered potential migrants. The selection of attributes continues until a potential migrant group, with a sex, age, income level, and economic sector membership is defined.

Once a state's potential migrant group is selected, we turn to the rule book. The rule book contains a set of graphs for each attribute, indicating likelihood that a group defined by a particular value of the attribute will migrate. Checking the graphs, attribute by attribute, we arrive at a string of probabilities, one for each attribute. To this list we add yet another probability based on how the economic characteristics of the area of origin affect its propensity to expell out-migrants. Combining these probabilities, we establish the migration potential of the group, a number representing the probability that this group in this particular state will migrate.

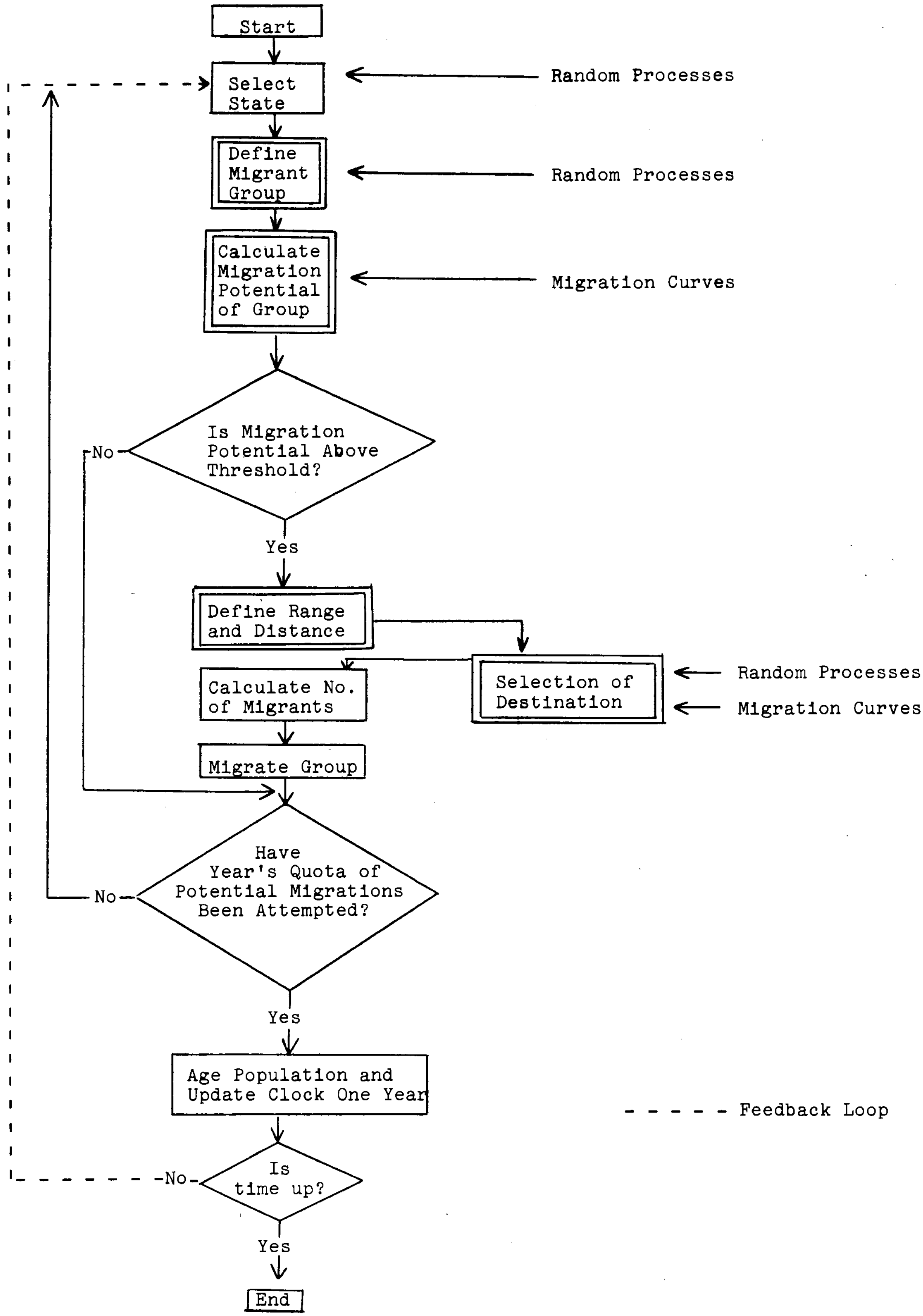
Next, we compare this group's migration potential with an arbitrarily chosen value called a "threshold." If the group's migration potential falls above the threshold, the decision is made to migrate them and we move on to another section of the rule book which determines a destination. If the group's migration potential falls below the threshold, then the group does not migrate and we check to see if any arbitrary quota for attempted migrations in one year has been exceeded. If it has not, then we begin once more the process of selecting a state, defining a migrant group and determining whether it will actually migrate (note feedback loop in Figure 7.1).

However, assuming that the migration potential for the group selected is high enough that they are chosen as migrants, the simulation must now select a destination. Once again, the characteristics of states enter into the operation of the model. Potential areas of origin are evaluated according to theoretical expectations of being attractive to this particular migrant group. After the destination is selected, the size of the group is defined and the migrants are moved to their new destination. Following this migration, the model determines whether the quota for attempted migrations has been exceeded. Assuming that it has, then the entire national population is aged and the clock is moved ahead by one year. Automatically the out-migration population and their attributes are subtracted from the population at the place of origin and added to the state of destination.

The process starts again by selecting a new state and once again defining a new group of potential migrants. The cycle continues, over and over, until a pre-established limit of years to be simulated is reached. At this point, the model automatically generates adequate statistics for evaluating the distribution of the population and its attributes in the ending year (1970) and all intervening years since 1960. These statistics include all flows of migrants into and out of states classified by their sector, income, age, and sex. It is important to remember that unlike the parlor game, all the events we have just described occur in seconds. In essence, time is compressed in the model, allowing for multiple operations of the model in a single computer run.

The "rule book" dictates the expected behavior and characteristics of migrants. It is divided into rules which determine what segment of a state's population will

Figure 7.1: Main Program



migrate and rules for determining a migrant group's destination. To see how these rules are formed and used in the simulation, we shall back track over the simulation once more, this time referring to how specific decisions are made.

Selection of Migrants

In this model, a group of potential migrants is defined by five attributes: (1) age, (2) sex, (3) income, (4) sector of employment, and (5) economic characteristics of the state of origin. Combining these factors determines whether a group of potential migrants actually moves.

Each of the first four attributes have certain dimensions or values. For sex, there are an established number of females and males; for age, a given number of people in different age categories. Likewise, there are so many people living in urban and rural areas, and so on. Possible combinations of the dimension of these attributes is large, but finite. Some combinations are quite rare, for example, a rural female construction worker making 100,000 or more pesos a year. Rather than determine the migration potential of every possible combination, we define the specific attributes of a potential migrant group which approximates the group's relative frequency in a population. We draw a random sample for each attribute, weighting the selected value in proportionto its distribution in the state's population. Thus, if agriculturalists represent 85 percent of a state's population, approximately 85 of every 100 persons selected will be agriculturalists. The selection process continues for each attribute until we have defined a potential migrant group.

We know that migration is selective; the probability that a person with a particular set of attributes will migrate varies with the value of that particular factor. For instance, it is more likely that individuals in their early twenties will migrate than individuals in their late sixties.

Phrased another way, we may say that the migration probability of a 20 year old is higher than that of a 60 year old. We represent these probabilities by means of a "migration curve" of that attribute. Ideally, we should establish migration curves for all attributes. For the attribute of age, this would mean plotting the probability of migration for people of different age categories. Then we should do the same for sex and residence. After doing this, we would be faced with the problem of resolving the interactive effects of these characteristics. For example, given that a 25-29 year old person hasa migration probability of 0.005 and that males migrate more frequently than females (0.59) and that urban people have a greater probability of migration than rural people (0.54), then we wish to calculate the probability that an urban male, between 25-29 years old, would move. Since we wish to arrive at a single migration probability we are faced with problems of not only the relative importance of each attribute, but also the relative interaction of these variables, i.e. determining the degree of interdependence of the attribute's probabilities.

The literature on Mexican migration does not always provide easily accessible information for answering these questions. The existing theories and information on Mexican migration are based on different research methodologies, geographical units of analysis, and degrees of theoretical sophistication. Some findings are derived from analyses of national census data (Cabrera 1967, 1970, 1972), others are from sample surveys of regions (Browning and Feindt 1967, 1969: Balan, Browning and Jelin 1973) or select communities (Ball 1967, 1971) and yet others from studies of migrants from a particular ethnic group or at a specific place of origin (Butterworth 1962, 1963, 1965, 1972). How did we decide which studies provided adequate rules for the simulation? We first checked sources for <u>demonstrated relations</u> between an attribute (such as age) and its effect on a group's potential for migration. "Demonstrated relation" means that a national study of the phenomena had been conducted on this characteristic. For example, Gustavo Cabrera (1972) using census materials, empirically measured the importance of age as a selectivity factor in Mexican migration. With slight modification his study fit our requirements of a demonstrated relationship and was preferred over more regional studies.

When national studies were unavailable, we relied on less exhaustive regional studies for determining an attribute's influence on migration (such as the work on Monterrey by Balan, Browning, and Jelin 1973). Alternatively, had data been available for other areas of Mexico, we might have compared this regional study to other similar studies. When we based our migration probability on a regional study, such as the Monterrey work, we assumed that the pattern of migration for Monterrey is representative of the entire nation.

If neither national nor regional studies for the attributes impact on migration potential could be found, we formed migration rules based on the best available information, i.e. relations which appear to be true according to studies conducted in other countries. For example, in some situations we had to rely on generalized propositions made by Ravenstein (1885) and Lee (1969). Finally, if we are unable to base our probabilities on either demonstrated relations, regional studies, or hypothesized relations tested in another context, then we made an educated guess based on what we intuitively felt was a valid relation. Naturally, a guess was the weakest of the four alternatives and avoided as much as possible. Some might object to this style of modelling, however, an operational model requires that all relations between variables be defined, otherwise it would not run.

Age Selectivity

We can examine these difficulties in determining who will migrate by looking at one of the more critical decisions made in this part of the simulation. Cabrera (1972) examines the relative influence of age, sex, and place of residence on migration selectivity. His data, for the period 1930-60, is based on inter-state migration, migration to urban zones (with populations over 15,000), and Mexico City. Using the method of intercensal survival indices, he discovers that age selectivity is greater among migrants to urban centers than among migrants to rural areas. He suggests an "inverse relationship between the growth in the absolute number of migrants and the importance of age selectivity." That is, the greater the volume of migration the less likely it is that migrants concentrate in certain age groups. Furthermore, Cabrera discovers that females are more likely than males to migrate to urban areas and Mexico City, with a trend toward equilibrium of the sex of migrants occurring as the migrant stream ages, a finding consistent with the argument made by Balan, Browning, and Jelin's (1973).

Cabrera's work permits us to solve the previous mentioned problem of the interdependence of attributes. Reanalyzing his data, we can estimate the probability that males and females of different ages will migrate. The manner in which his results are tabulated permits us to ignore the interaction of age and sex probabilities, for he provides data on the intersecting sets of males by age and females by age. Furthermore, Cabrera discovered that the selectivity of migrants to urban areas differs from that of all migrants. By inference, we may assume that the curves of rural to rural migrants are more closely aligned to the between-states curves than they are to the urban curves. Thus, the migration simulation model incorporated Cabrera's findings into its rule book by using them to set the migration potential curves for age, sex, and residence (Figures 7.2 through 7.5).

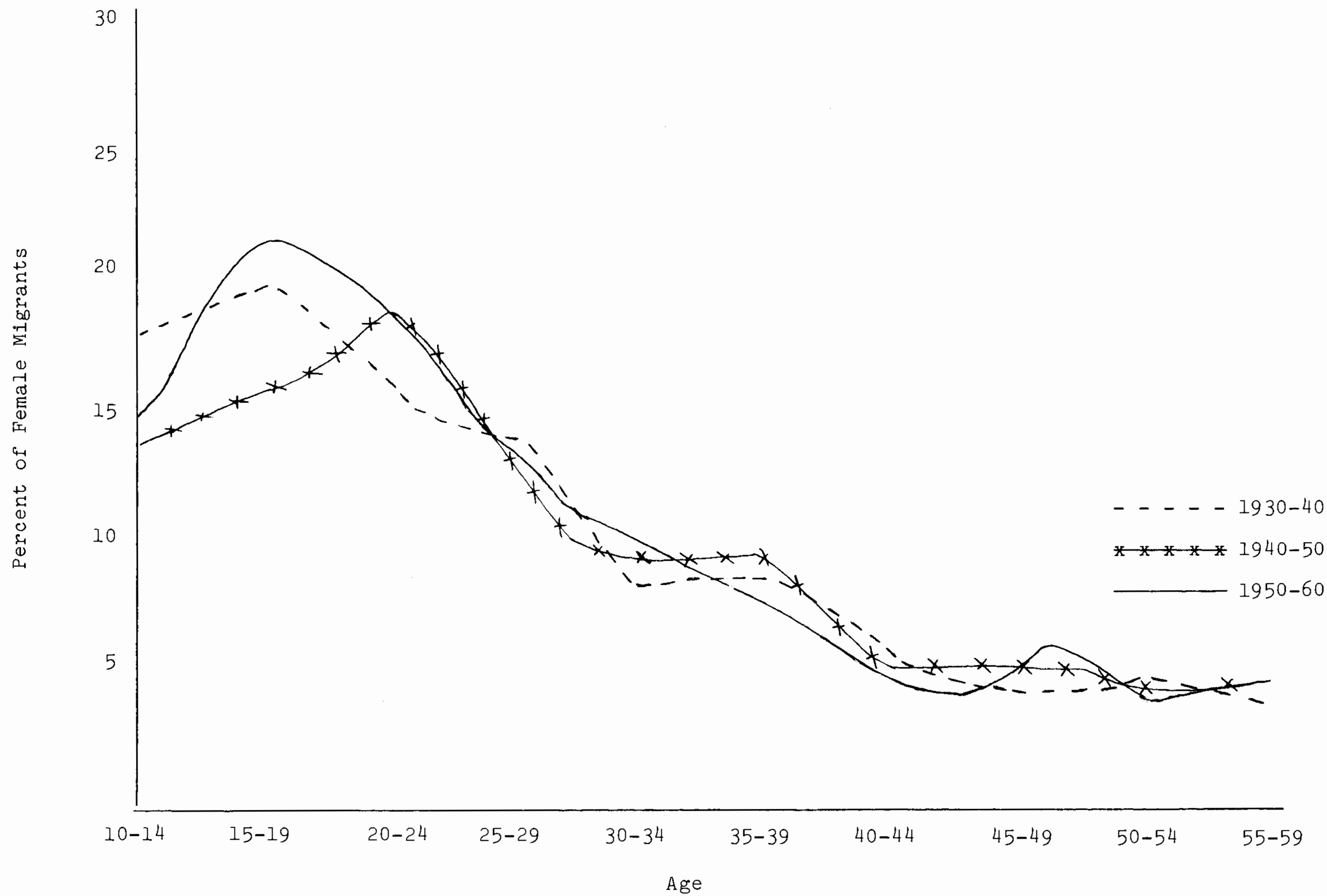
It must be stressed that certain assumptions are required before assignment of the relative importance of any attribute on a group's migration potential. For example, accepting Cabrera's information as part of the "rules" requires the following assumptions:

Assumption 1: The selectivity of migrants for age and sex does not change from 1950-70.

Cabrera's data is based on the 1950-60 period, but the model is being calibrated for the 1960-70 period. If we use the curves in Figures 7.2-7.5, then we must assume that no changes took place during the second decade (1960-70). Naturally, if additional information were to become available, then this assumption could be relaxed or modified.

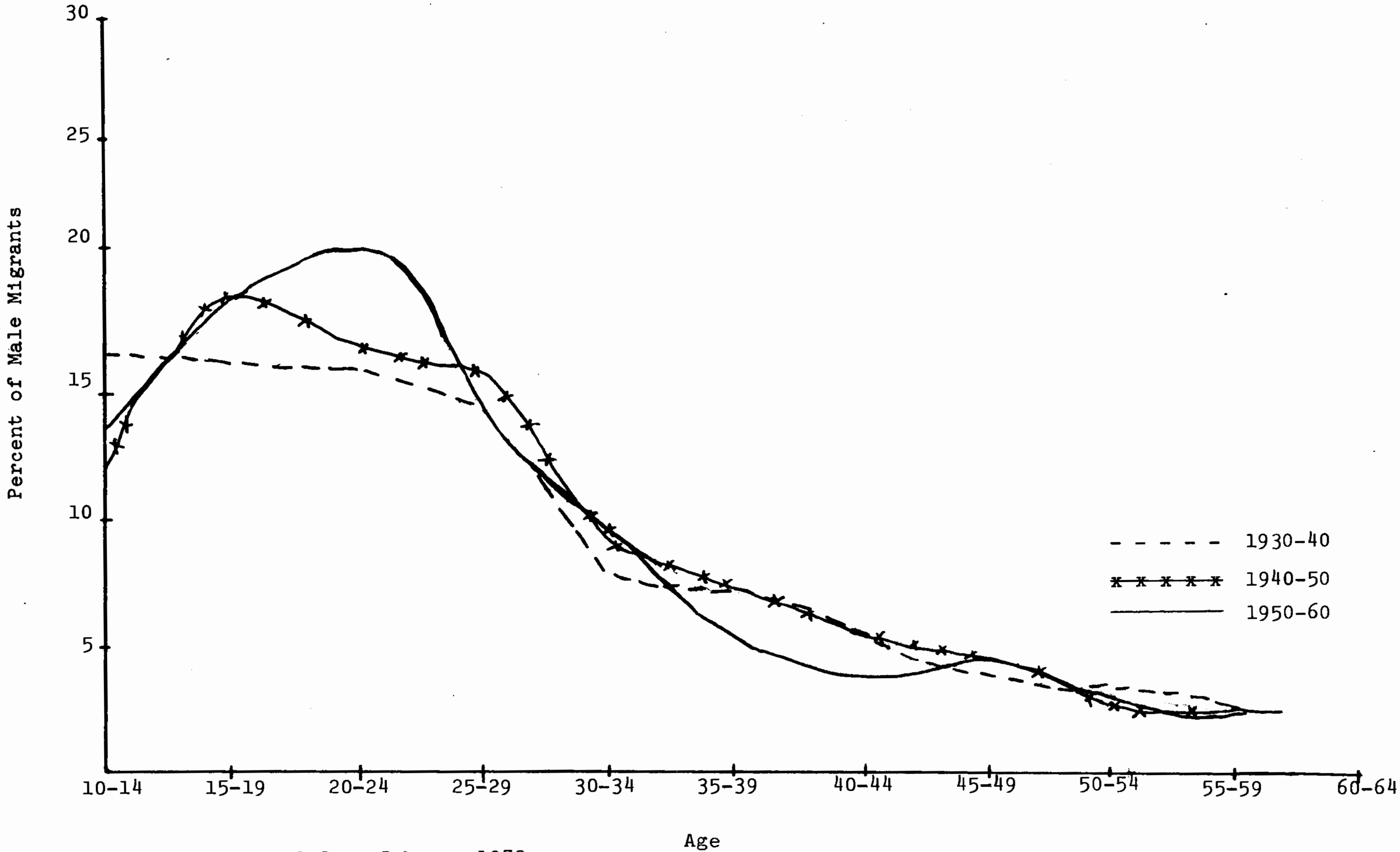
As an alternative, we might have extrapolated a trend from the 1930-60 period into the 1960-70 period. Unfortunately, inspection of the tables reveals there are no distinguishable national trends toward greater or lesser selectivity of age, sex, and place of residence. If anything, there appears to be a cyclical pattern. For females, selectivity for age was more pronounced in the 1930-40 period and the 1950-60 period than in the intermediate decade (1940-50). In contrast, male migrants to all areas show a long term trend,

Figure 7.2 : Net Migration of Inter-State Females of Different Ages



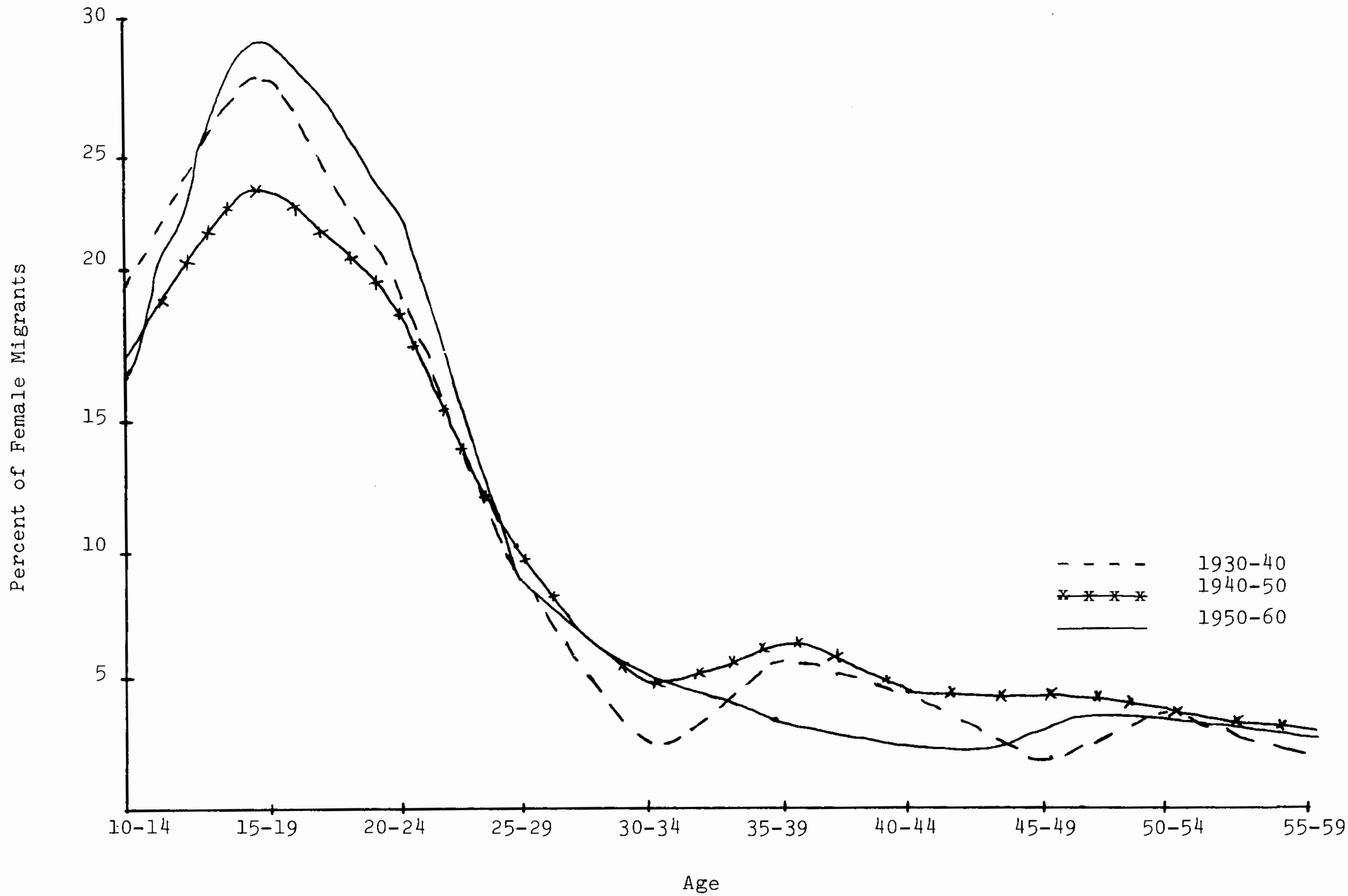
Source: Adapted from Cabrera 1972

Figure 7.3: Net Migration of Inter-State Males of Different Ages



Source: Adapted from Cabrera 1972

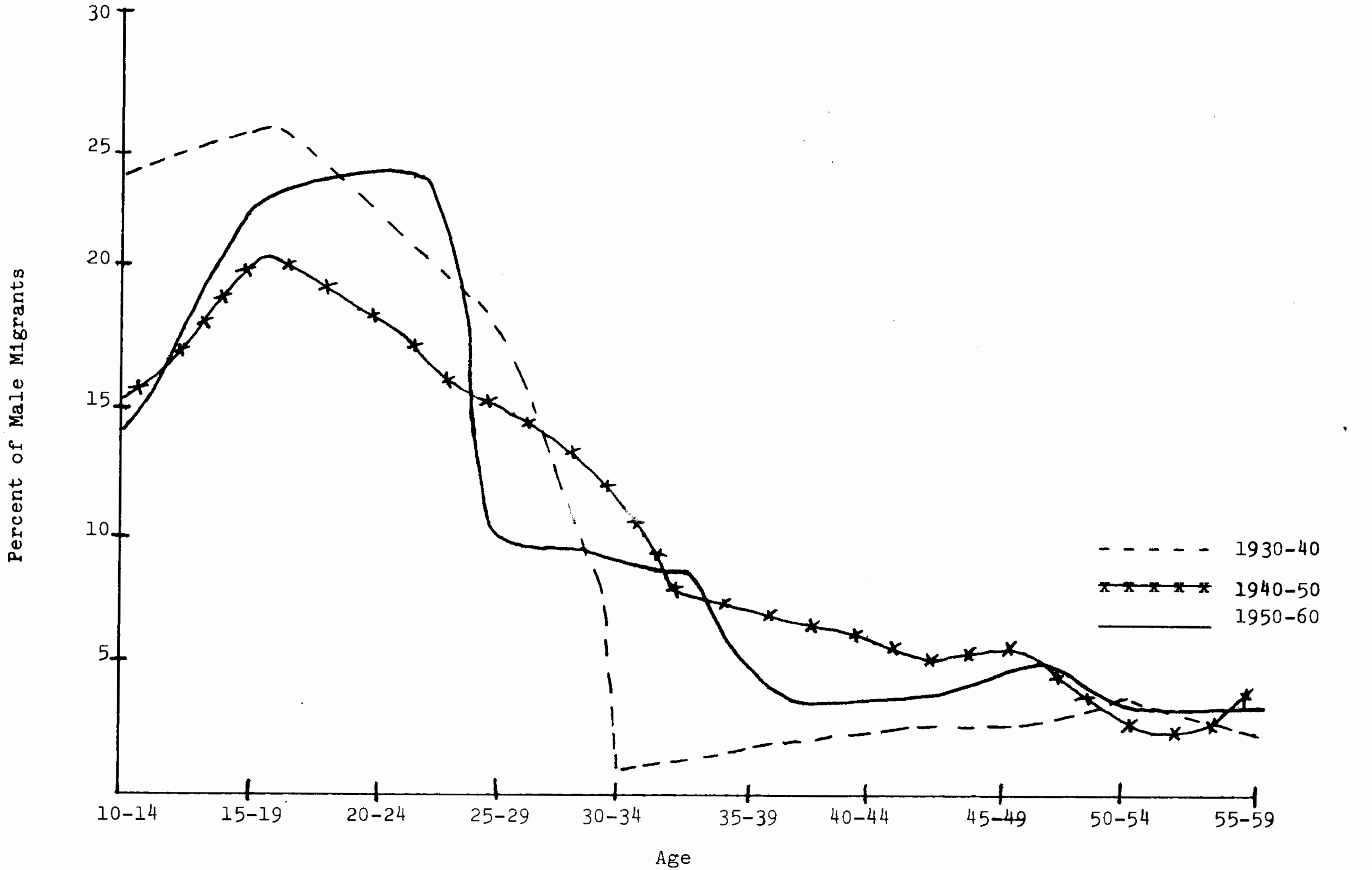
Figure 7.4: Net Migration of Inter-state Females of Different Ages in Urban Zones



Source: Adapted from Cabrera 1972

Note: Urban means 15,000+ people

Figure 7.5: Net Migration of Males of Different Ages in Urban Zones



Source: Adapted from Cabrera 1972

Note: Urban means 15,000+ people

with the age selectivity increasing in favor of 10-24 year olds from 1930-60 (Figure 7.3). However, male migrants to urban areas appear to be following a different trend, decreasing their selectivity during the intermediate decade (1940-50). The lack of data points, however, prevents extrapolation of these shifts into a trend.

Assumption 2: The selectivity of non-urban migrants resembles the inter-state migration pattern more than it does the urban migrant pattern.

Cabrera did not present data on rural migrants which corresponds to Figures 7.4 and 7.5. Figures 7.2 and 7.3 include urban and rural migrants, they more closely approximate the movement of rural migrants than any other data we have. Comparing these four tables, it appears that migration to urban areas is more selective for age than migration to rural areas, a conclusion holding for males and females. We also find from studies by Tobah and Cosio (1970) that migration to urban areas is more selective for age than migration to rural areas.

Assumption 3: The selectivity of migrants for age and sex is the same for all migrant streams.

In reality, this assumption is probably false since it appears that the selectivity of migrant streams changes as the migrant stream gets larger (in absolute numbers) or older. However, until more quantitative studies which specify this relationship are available, we decided to use these four curves for all migrant streams between states.

In the case of income, education, and labor force sector membership, it was immediately apparent that there were no useful empirical studies comparable to Cabrera's work by which we could define the impact of these variables on migration potential. Most studies have been too specialized to apply to nation-wide data, and as a consequence, assumptions were made about how these characteristics were to be treated. Those unfamiliar with formal modelling might be puzzled as to the need for so many, often unrealistic assumptions. In response, we wish to point out that assumptions are a necessity in any modelling. In verbal models they are made unconsciously or left unstated. Formal model building forces the theorist to be aware and make explicit the assumptions. Furthermore, any assumption can be modified or "relaxed" to test their relative importance to the results of the model. Unlike verbal models, a simulation model's design permits a rather rapid relaxation or modification of many of its assumptions, a process which would permit an experimental test of the theory's sensitivity to the modified assumptions.

To return to the decision as to who will migrate, it may be noted that all the attributes considered thus far are essentially demographic, common to people but not to the region itself. However, we know that out-migration is also related to economic conditions in the state of origin (Rogers 1965, Bogue 1969, also see Sayers and Weaver, Chapter II in this report). To mimic the economic situation, we need a measure of the economic well-being of the major labor sectors in each state. Unfortunately, Cauthorn and Hubbard's analyses of the economic sectors coincided with the building of this model, preventing the inclusion of detailed information on economic sectors. As a result, the current version of the model employed a less-than-optimal data. The project's economist estimated the relative health of the economic sectors in each state. This estimate establishes a migration probability for all state economic sectors. We assume that the poorer the economic condition of a state's non-agricultural sector, the higher the probability of out-migration. The sectors selected for consideration were agriculture, oil, gas, mining, manufacturing, construction, electrical, commerce, transportation, service, and a category called "insufficient information." The specific impact of each of these sectors on migration is discussed by Cauthorn and Hubbard in Chapter V of this report.

To summarize, we used five personal attributes and one statewide characteristic to calculate the migration potential of randomly selected groups. They were:

1. Age
2. Sex
3. Rural/Urban Residence

} Combined into one set of curves, see Figure 7.2-7.5

4. Education
5. Income
6. Labor Sector Condition

Fortunately, we had Cabrera's analysis which enables us to combine the first three of the five personal attributes into a single probability.

Weighting Probabilities

Thusfar, we have assumed that a group's potential for migration is a function of its age, sex, residence in rural or urban areas, education, income, and the relative economic condition of the labor sector to which the group belongs. We have not, however, specified the relative importance of the list of factors, with respect to each other, in determining if a group will actually migrate. The best empirical research we have concerning the relative importance of these factors is Tarver's (1961) study of intercensal net migration in the United States between 1940-50. He performs a linear multiple-regression analysis of white and non-white migration in which the independent variables were represented by clusters of social, economic, and demographic variables. This analysis revealed that information on all three clusters of independent variables, including their interdependence, was necessary to derive an accurate estimate of internal migration. It is possible, of course, to adopt Tarver's relative weightings in the Mexican migration model, but it must be assumed that the factors underlying migration in Mexico are operating in a similar fashion to those in the United States. Tarver's own analysis makes us hesitant to make this assumption. To attain the same level of accuracy for estimating non-white migration that he had for white migration, he had to introduce nine more independent variables into his clusters (14 vs. 5). If we assume that the category "non-white" is indicative of a cultural difference in migration patterns, then his analysis is telling us that the relative influence of the same variables may vary for different cultural groups. Since, we know that Mexico is radically different, culturally and institutionally from the United States, it is likely that Tarver's weightings are inapplicable to Mexico.

What, then, do we do? Unlike verbal theories, computer models cannot encompass ambiguities: the operation of a simulation requires an explicit specification of the relative importance of factors assumed to influence migration. Lacking any empirical research on this topic, we opted for an experimental determination of the weights. Initially, we decided to arbitrarily set the relative influence of the independent probabilities of migration at 35 percent for the cluster of age-sex and residence, 35 percent for labor force sector membership, 15 percent income, and 15 percent education. The relative weightings of these factors are adjustable, meaning that various combinations could be tried on different runs of the model to determine the effects of different weightings.

The combination of the four probabilities yields a single value which we call a "group's migration potential." The specific value of this probability is based on the heterogeneity of the population within a state and on the economic conditions of the potential migrant's labor sector. It depends on (1) the particular attributes of people selected during the random sampling of the state's population, (2) the economic condition of the area of origin and (3) the weightings assigned to the algorithm for calculating a group's migration potential. The operators of the model have no deterministic control over who is selected to be considered for migration. However, they can control the theoretical assumptions underlying how a particular attribute effects a group's pattern of migration. For example, if a subsequent study shows the precise relation between income and migration, then this information can be immediately added to the model to improve its estimation and projection of migration. This ability of the model to assimilate future studies is one of the strengths of simulation modelling.

The Threshold

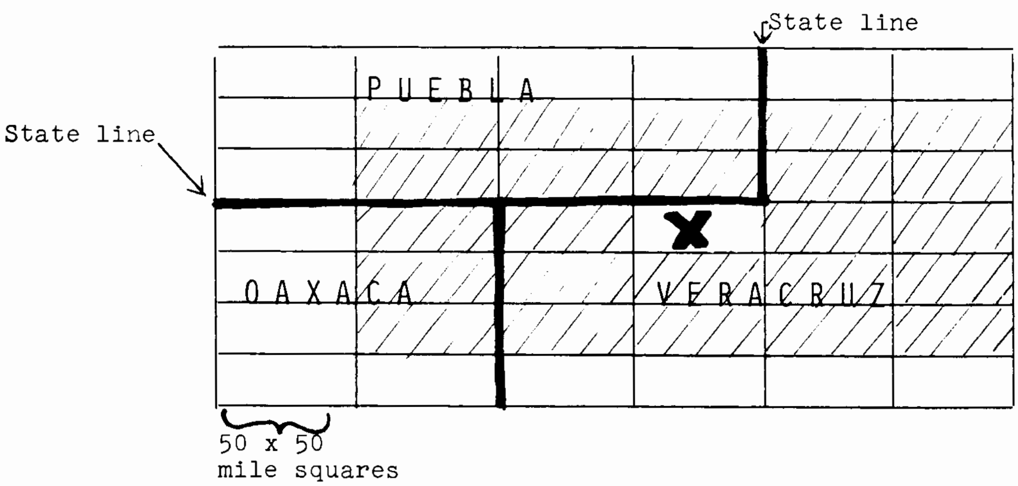
The newly calculated "migration potential" of a group is compared to an arbitrary value called the "threshold." The threshold value forms a very convenient valve, controling the national gross rate of migration. Indirectly, the threshold also effects the composition of the migrant group. Set very high, only individuals with attributes that give them a high probability of migration will migrate. Set at zero, any combination of attributes, no matter how unlikely, will result in the group migrating. When the model is used to estimate intercensal migration, the threshold value permits us to align the total intercensal net migration of the model to the volume of migration which actually occurred in the population. In future experiments with the model we hope that the threshold value can be used to control the influence of national economic trends, such as GNP or the rate of inflation, on gross migration within the country.

Selection of a Destination

Deciding whether a group will leave an area is only half the task of modelling migration. A destination for the group must also be chosen. We divided the task of selecting a specific destination into two problems. First, we developed an algorithm to calculate the farthest range that a particular group may migrate. Then, we evaluated all potential destinations within this range to determine exactly where the migrants move.

To perform this task, we programmed into the computer a more refined picture of Mexico's geography. A grid of 50 by mile squares is superimposed over a map of Mexico. For example, the intersection of the states of Oaxaca, Puebla and Veracruz might be represented as in Figure 7.6.

Figure 7.6: Representation of Distance in Model

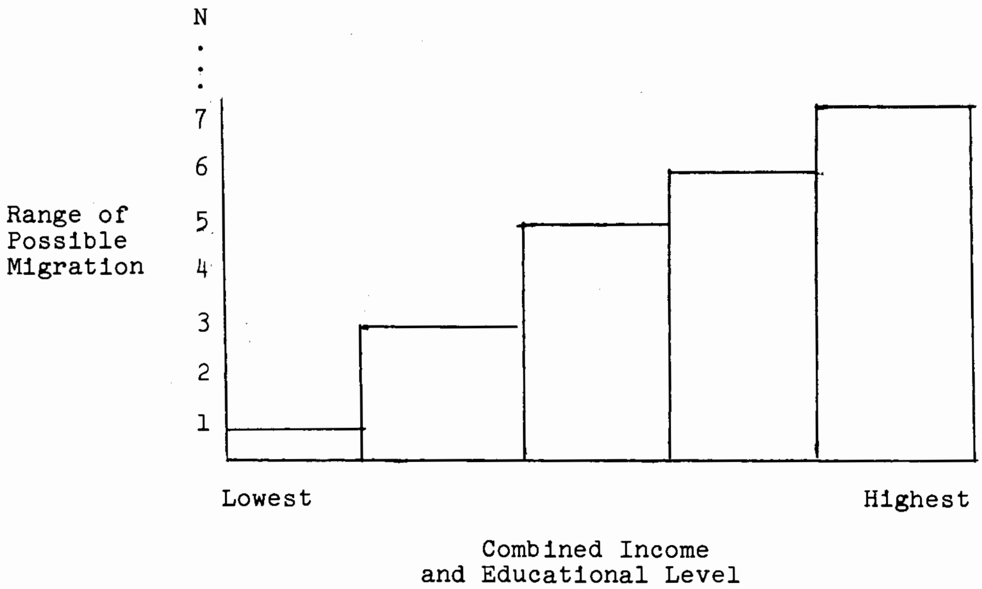


Thus, the larger the state in absolute area, the more squares it contains. The exact location of the migrant's origin is determined by assigning an urban classification to all squares with over 10,000 population. All others are classified as rural. If the migrant group selected is of an urban origin, then we make a list of all urban areas within the migrant's state of origin and randomly select one of them as the urban area of origin. For example, if Veracruz was selected as origin for an urban migrant group, we would list all the areas in this state having populations over 10,000 and assign an equal probability of being selected in a random draw. We then pick one, which we indicate by X on Figure 7.6.

From this area, we decide on a likely range for migration. Previous migration research had established that the probability of migration decreases with distance (Zipf 1946, Stouffer 1940, 1960). Moreover, the importance of distance seems to vary with characteristics of the migrant group. Highly educated migrants are more likely to make long distance moves than less educated migrants, richer people move longer distances than poorer people, younger people (15-30 years old) move longer distances than other age groups, and so on. To combine all the possible combinations of characteristics (income, age, sex, etc.) and assign these combinations an accurate distance relation is impossible. Not only are there too many combinations but also we did not have the necessary empirical data.

Lacking empirical evidence, we assume that the range of migration increases for migrant group's in proportion to the income and educational status. We expressed this assumption as a discrete mathematical function (Figure 7.7).

Figure 7.7: Assignment of the Range of Migration



Using the income and educational level of the migrant group, we decide on a distance to be searched as the area of potential destinations for the migrants. For initial runs of the model, we assume that education and income exert an equal influence on the range of migration. Thus we calculate the average of the values of these two variables, using this number as the radius of our range. It should also be noted that any characteristic or combination can be used to produce a likely range. The number selected for the range indicates the farthest potential migration and may be as far as the distance from one end of the country to the other or as close as the adjacent square. We demark an area as large as the number selected. To continue with our example in Figure 7.6, if the number is 2, we choose the cross-hatched area.

This decision means that any migrant from area X must move within the cross-hatched zone. It also means it is possible that migrants will not move to another state, i.e. an interstate move will not occur. The occurance of inter-state migration depends on (1) the migrant's area of origin (2) the range of possible destinations for the migrants. Allowing for such intra-state moves permits us to simulate, in principle, intra-state rural to urban migration, although the current model does not include this capability.

Within this range, we evaluate the attractiveness of all potential destinations (small squares). The evaluation is based on three characteristics. First, we determine the economic health of the labor sector to which the migrants belonged, assuming that migrants will be more likely to move to states with better economic conditions. Second, we increase likelihood of migration to areas where previous migration has occurred. That is, a previous flow of migrants between two states increases the probability of future moves between the migrants' origin and destination. Finally, we give preferential treatment to urban over rural migration, since the overwhelming weight of migration research points to the predominance of a marked rural to urban and urban to urban movement of migrants.

Once again we were faced with the problem of determining the relative importance of several factors known to influence migration, but for which we lack empirical evidence. That is, we do not know the relative importance or weighting of these three characteristics: economic pull, previous migration listing, and urban attraction. And once more, we employ the same ad hoc method mentioned previously for weighting the attributes of potential out-migrants. We assign arbitrary weights, with the expectation of changing them when we calibrate the model to the 1960-70 migration data.

After the most likely area of origin for the migrant group is selected, the size of the migrant group is calculated based on the expected national gross migration. The actual migration is devoid of emotion, consisting of nothing more than migrants being deleted from their areas of origin by subtracting their characteristics from the state of origin's totals. Likewise, these totals are added to the state of destination's population. A check is made to determine if the year's quota of attempted and completed migrations has been reached. If it has, then a massive up-dating procedure takes place. The age-sex distributions for each state are adjusted to correspond to all the migrations that occur within the year, the distributions for all characteristics are recalculated to include changes due to migration, and a time index and any time varying functions are also updated. Then the model is prepared for another yearly cycle.

Calibration

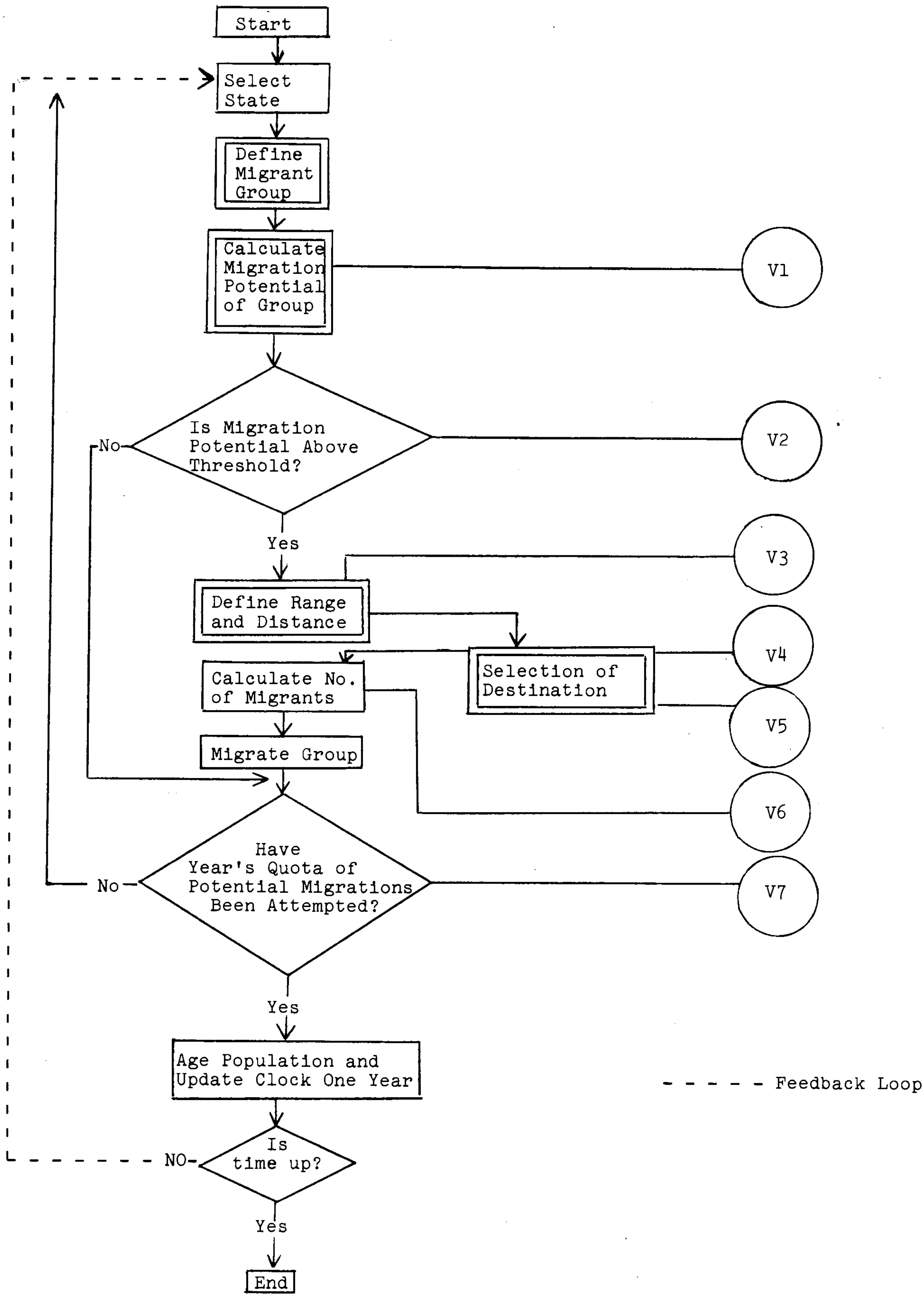
The lack of theoretical formulations specifying the relations of many of the variables used in this model and the absence of empirical data on some of the model's parameters forced us to make several arbitrary decisions during its construction. We had no choice but to set up some weighting for combining of probabilities if the model was to give us any results. In the calibration phase of the model's construction, we are adjusting and modifying these weightings so as to align the output of the model with empirically observed rates of interstate migration.

Calibration is accomplished by building in "control values" at several points in the model's construction. Changing the setting on these control valves instantly alters critical values in the simulation. During calibration, we are adjusting these values to align the model's behavior to a pattern of internal Mexican migration between 1960 and 1970.

Table 7.2 lists the main control valves and Figure 7.8 shows their approximate location in the flow diagram of the model. For example, valve V1 changes the relative weights of age/sex/residence, education, income,and labor sector conditions used in calculating the migration potential of out-migrants. Phrase in theoretical terms, shifting of these weights not only permits us to align the model, but also permits the experimental testing of the relative importance of four factors which enter into a theoretical explanation of internal migration. In a similar manner, valve V4 permits us to alter the relative attractiveness of urban over rural areas, again permitting the experimentation and hypotheses testing. The next phase of the model's development, currently in progress, consists of a careful alteration of the settings of these valves to make the model mimic Mexican migration.

Our first step is to align the vital statistics of the model (births and deaths) to that of the period from 1960-70. The control values on the model area are set so that no migration could occur, i.e. we allow the population of each state to change only in response to births and deaths. The results, as could be expected, only roughly correspond to the state populations observed in 1970 (Table 7.3). We are encouraged, however, by two indications that the simulation is moving in the right direction. First, the model comes within one percent of the national population of Mexico (Table 7.3). A more detailed breakdown

Figure 7.8: Location of Control Valves in the Main Program



of our results by age categories shows that the model tends to underestimate the younger age categories and overestimate the older age categories. Second, the state-by-state population estimates show wider discrepancies between the model and the expected, 1970, distribution. This is expected, since we are not permitting any interstate migration. However, when the direction and magnitude of these discrepancies are compared to the direction and magnitude of internal migration which actually occurs during this period, the model again is making errors in the right direction. That is, if we correct the model's estimates by the amount of net migration known, through the Mexican census, to have taken place between 1960 and 1970, then the magnitude of the model's error is greatly reduced (compare the means and standard deviations on Table 7.4).

Table 7.2: Control Valves Used in Calibrating the Model

V1	Relative weights of four migration potentials
V2	Threshold
V3	Range of migrants group - weighting income vs. education
V4	Attractiveness of urban vs. rural areas
V5	Selection of Destination Weightings: importance of previous migration history vs. economic conditions of states.
V6	National gross rate of internal migration
V7	Number of attempted migrations

Table 7.3: Estimation of Age Distribution

Age	Model	Census	Percent Difference
0 - 4	8,386,114	8,167,510	+2.68
5 - 9	7,085,891	7,722,996	-8.25
10 - 14	6,034,860	6,396,174	-5.65
15 - 19	5,136,109	5,054,391	+1.62
20 - 24	4,304,507	4,032,341	+5.38
25 - 29	3,567,955	3,260,418	+9.43
30 - 34	2,949,383	2,596,263	+10.83
35 - 39	2,454,573	2,511,647	-2.27
40 - 44	2,050,530	1,933,340	+6.06
45 - 64	4,988,566	4,758,773	+4.83
65 +	1,912,908	1,791,385	+6.78
Total	48,871,396	48,225,238	+1.34

Table 7.4: Estimation of State Population Change, Based on Model's Orientation Allowing no Inter-State Migration

	a Model's Estimate	b Census 1970	Percent Difference Between a and b	Percent Difference* Including Observed Net Migration
Aguascalientes	356,601	338,142	+5.46%	-1.74%
Baja California	707,754	870,421	-17.69	-8.83
Baja California T	117,671	128,019	-8.08	-0.76
Campeche	238,856	247,114	-3.34	-3.28
Coahuila	1,288,396	1,114,956	+15.56	-1.79
Colima	232,905	241,153	-3.42	-2.82
Chiapas	1,642,792	1,569,053	+4.70	+2.14
Chihuahua	1,690,259	1,612,525	+4.82	-2.10
Federal District	6,167,584	6,874,165	-10.28	-8.18
Durango	1,143,551	939,208	+21.76	+4.23
Guanajuato	2,437,091	2,270,370	+7.34	+0.33
Guerrero	1,822,939	1,597,360	+14.12	+5.38
Hidalgo	1,452,897	1,193,845	+21.70	+7.02
Jalisco	3,414,065	3,296,586	+3.56	-1.19
México	2,596,424	3,833,185	-32.26	-2.98
Michoacán	2,758,895	2,324,226	+18.70	+2.98
Morelos	555,016	616,119	-9.92	-2.35
Nayarit	587,206	544,031	+7.94	+2.54
Nuevo León	1,505,357	1,694,689	-11.17	-6.30
Oaxaca	2,306,078	2,015,424	+14.42	+0.72
Puebla	2,734,215	2,508,226	+9.01	+0.98
Querétaro	542,550	485,523	+11.75	+4.38
Quintana Roo T	70,650	88,150	-19.85	-0.85
San Luis Potosí	1,572,270	1,281,996	+22.64	+5.76
Sinaloa	1,298,177	1,266,528	+2.50	+1.74
Sonora	1,128,227	1,098,720	+2.69	-1.70
Tabasco	758,891	768,327	-1.23	-1.17
Tamaulipas	1,408,530	1,456,858	-3.32	-2.22
Tlaxcala	513,329	420,638	+22.04	+3.65
Veracruz	3,662,541	3,815,422	-4.01	-0.33
Yucatán	857,823	758,355	+13.12	+0.95
Zacatecas	1,301,856	951,462	+36.83	+10.81
			$\overline{m}$ = 4.25	$\overline{m}$ = 0.16
			$\overline{sd}$ = 14.51	$\overline{sd}$ = 4.19

*Difference between a and b, if observed migration is added to model's estimate

Current Status

As of March 1, 1976, a complete, running simulation was operational on the University of Arizona CDC-6400/DEC 10. The model is on-line and interactive, permitting rapid adjustments to many of its parameters, including all the control values, the migration curves, and probability distributions. It is written in Fortran F.4, a common language found at most computer centers.

Calibration had just begun when the National Science Foundation funds for this project were exhausted. The authors felt this is a great tragedy since, to their knowledge, this is the only generalized, flexible model of internal migration capable of allowing operational tests of migration theories on a national level. The University of Arizona has provided a small fund to continue exploration of the model, but the amount is inadequate to begin the next phase. If funds may be obtained, we plan to continue development of this model. Our tentative strategy is to:

1. More closely align the model to mimic the rates of change in vital statistics. This will require computation of mortality tables for each state.

2. Set the threshold value so that the gross national migration of the model matches that of Mexico.

3. Continue the analysis of economic sectors, recalculating the changes in the health of state economic sectors for the calibration period.

4. Complete the calibration of inter-state migration by adjusting the control values, constantly comparing the models' output with the known distribution of the population in 1970, and the intercensal rates of migration.

5. Using economic and vital statistics for the 1970-76 period, project internal migration up to 1976.

6. Develop alternative economic and vital rate projections for the 1976 to 1980 period.

7. Using these projections, project internal migration for 1980. This will be a census year against which we can test the validity of this method.

8. Refine the model to make projections for economic regions in Mexico, rather than states.

9. Test alternative hypotheses concerning the importance of different variables used in the model on estimating and projecting internal migration.

To return to the analogy, we have laid out the game board, written the rules, and moved the first set of migrants. The next step is to play the game and learn about migration.