DEMOGRAPHY© Volume 20, Number 2 May 1983

IMPROVING THE ACCURACY OF MIGRATION AGE DETAIL IN MULTIPLE-AREA POPULATION FORECASTS

Esther C. Schroeder
Computer Science and Applied Mathematics, Lawrence Berkeley Laboratory, Berkeley, California 94720

Donald B. Pittenger
The Demographics Laboratory, 2065 Lakemoor Drive S.W., Olympia, Washington 98502

Abstract—Population projections are often required for many geographical areas, and must be prepared with maximal computer and minimal analytical effort. At the same time, realistic age detail forecasts require a flexible means of treating age-specific net migration. This report presents a migration projection technique compatible with these constraints. A simplified version of Pittenger's model is used, where future migration patterns are automatically assigned from characteristics of historical patterns. A comparative test of age pattern accuracy for 1970-1980 indicates that this technique is superior to the commonly used plus-minus adjustment to historical rates.

INTRODUCTION

There is no universally superior method of treating migration in population projections. This assertion is made from an *applied* rather than *ideal* standpoint. In reality, demographers must deal with such constraining factors as time, cost, computer storage, and data manipulation capabilities.

This paper addresses the situation where age-specific net migration rates must be determined for many geographic areas with a minimum expenditure of analytical effort. The requirement of high production using a small staff is common in the public, private, and academic sectors, and its constraining elements are obvious. Added to this is the desire to treat age-specific migration flexibly; flexible migration rates should yield more accurate age detail than would relatively fixed rates based on historical migration patterns. However, flexibility is usually accompanied by model complexity, whereas high production and few personnel would seem to dictate simplicity.

Existing Techniques for Treating Migration

The problem indicated above included a requirement for race-sex-age detail in population projection output. This means that we are dealing with techniques used under the "cohort-component" methodological framework. This methodology refers to the case where populations are broken down into race-sex-age groups and moved through time by multiplication by race-sex-age-specific rates of fertility, mortality, and migration. For details, consult Irwin (1977), Pittenger (1976) or Shryock et al. (1973).

Until recently, most demographers or other technicians preparing sub-national projections have relied on the assumption that future race-sex-age-specific net migration rate patterns will be similar to historical patterns for the population in question (Pittenger, 1976, Chapter 8). Irwin's (1977) Census Bureau manual for local planners provides an example of a technique for modifying historical decade migration rates using a "plus-minus adjustment" so that known postcensal trends may be accomodated. The weaknesses of this technique are well known, however (Shryock et al., pp 705–706). The need for better means of treating age-specific net migration has led to improved models in recent years (Rogers, 1975; U.S. Bureau of the Census, 1979), but these models require data not available to most forecasters.

Pittenger (1978) proposed a technique for flexibly modeling age-specific net migration rate patterns. It is based on a simple typology of underlying directional (in and out) age-specific migration rate patterns. Although intermediate cases are possible, directional flow patterns tend to have either early or late age timing of the modal rates. On a five-year model, "early" might be age group 20–24 and "late" could be age group 25–29 or 30–34. One class of net patterns ("younger") occurs when in-migration timing is early and out-migration timing is late. Another case ("older") has a late in-migration mode and an early out-migration mode. Shapes of net rate patterns within each type will vary depending upon the magnitude of the rates in each direction. For example, high age-specific in-migration rates combined with low out-migration rates yield net patterns that are in-migratory. This concept is illustrated in the lower left-hand corner of Figure 1.

Pittenger has also observed that the directional flow timing patterns for areas as small as counties normally do not change over time, even though overall net migration can fluctuate between highly out and highly in. This means that, once timing patterns have been established for a given population, realistic age-specific net migration rate patterns can be modeled simply by increasing or decreasing rate magnitudes for directional flows.

Methodological Strategy

A major problem in adapting Pittenger's model entailed simplifying the computation algorithm. This algorithm is complicated and requires a great deal of analytical effort in assigning parameter values and calibration testing. When projecting a few populations, this creates no special difficulty. But it presents a significant problem if the projections are to be mass-produced, which is the concern of this paper.

Another problem had to do with the assignment of migration pattern types to individual populations. An automated assignment procedure had to be developed to replace an essentially judgmental task.

MODEL IMPLEMENTATION

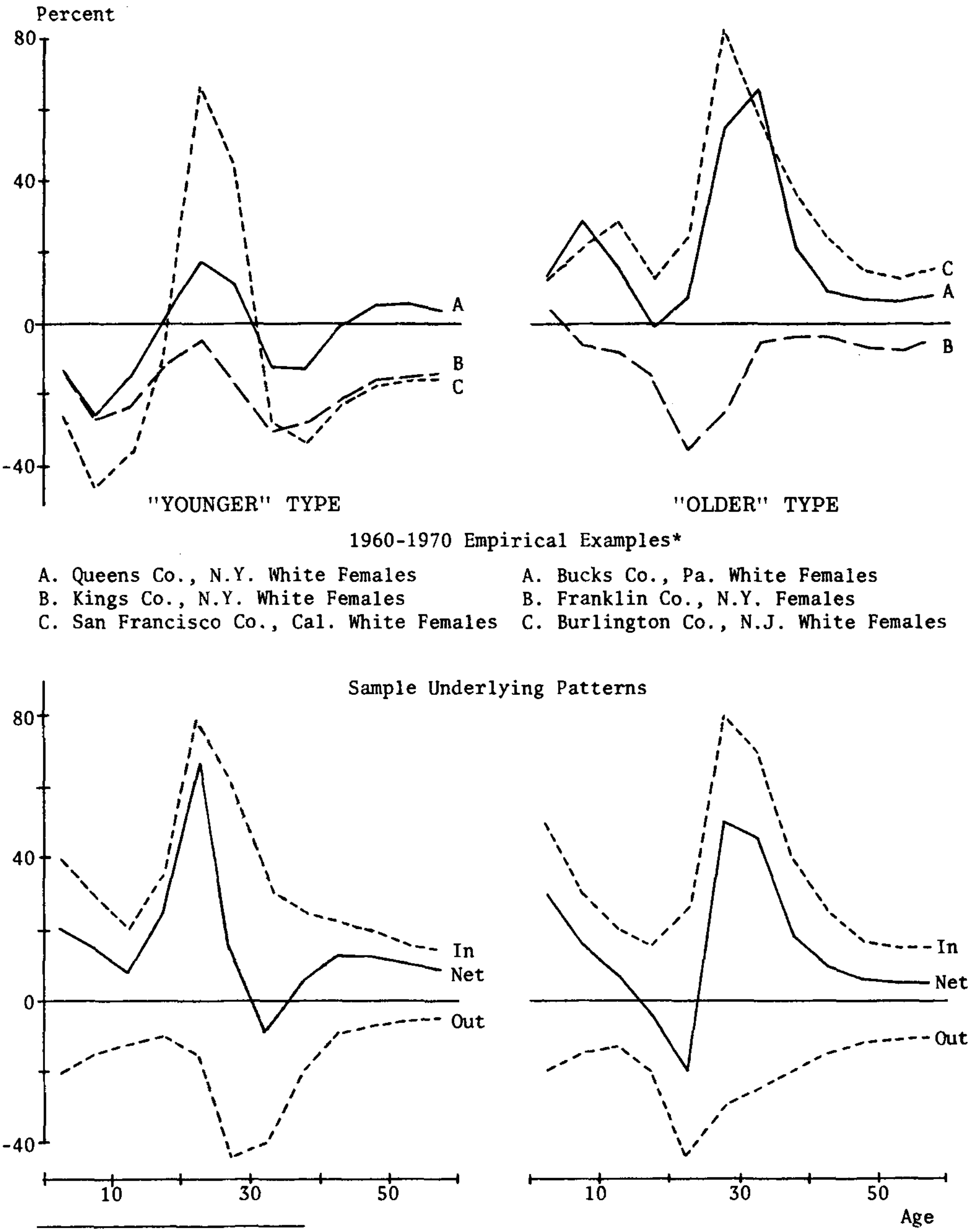
The implementation of an automated version of Pittenger's model requires essentially two steps—first, defining a set of typical direction flow patterns, and second, developing an algorithm for assigning a particular pattern to each population group.

Definition of Migration Patterns

The model contains a file of directional flow patterns that, when correctly selected and properly scaled, yield in- and out-migration rates that can be combined to mimic closely the historical age-specific net patterns. Changes in scale permit flexibility in pattern shape for forecasts. Since retirement migration can vary considerably within migration patterns, it is handled separately. Thus the migration patterns used by the model are defined only up to age 65.

The source for analyzing age-specific directional migration flows was data published for State Economic Areas (SEAs) (U. S. Bureau of the Census, 1963, 1972). SEA data are very useful for several reasons: they show migration rates over a five-year time period; they are defined for the same geographic areas for two time periods, 1955–1960 and 1965–1970; and they represent a variety of demographic conditions—central city counties, suburban counties, growing and declining areas, etc.

One finding was that the downward



* Source: Bowles, et. al., 1975.

Figure 1.—Illustration of Age-Specific Net Migration Rate Typology Based on Underlying Directional Patterns.

slope of the curve after the peak seems to vary with the level of migration. Figure 2 shows the exponential slope of the directional migration rate at ages 40–44 for male populations in selected SEAs plotted against the percentage directional migration for the same cohort. The upper plot shows out-migration and the

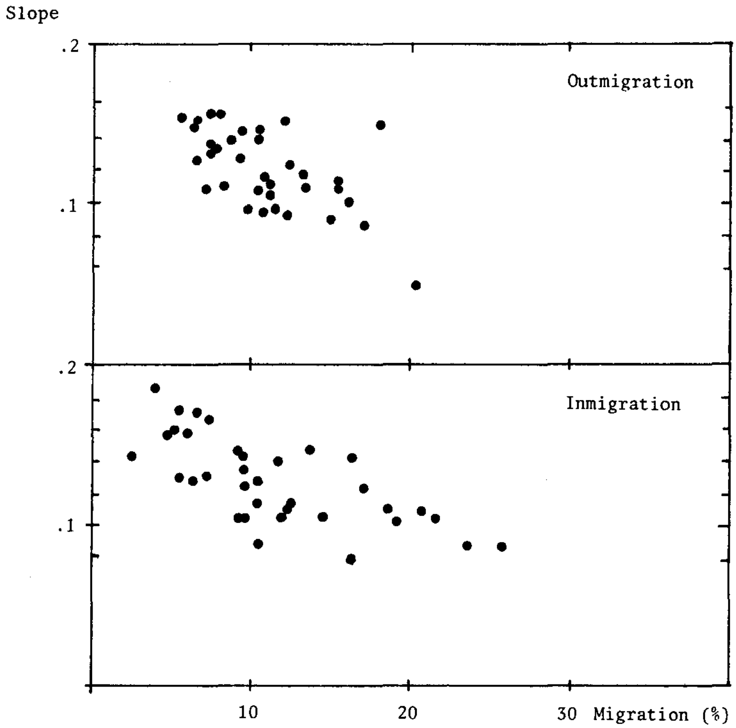


Figure 2.—Relationship of Migration Slope to Migration Level at Ages 40–44, Selected State Economic Areas, 1965–1970, Males

lower, in-migration. Both show that at ages 40–44, the slope of the directional migration rate becomes less steep with increasing migration. To preserve the relationship of slope to migration rate level, migration rate patterns were developed for three different slope values, 0.09, 0.12 and 0.15.

Model migration patterns were designed with reference to three factors—the height of the peak age-specific rate, the age at which this peak occurs, and the general slope of the curve as defined by the rate for age group 40–44. Thirty-six patterns were created for the present application. For each of three slope categories, there are three amplitudes at the peak (short, tall, and extreme), and four timings of the peak ("college," "early," "intermediate," and "late"). These are presented in Table 1.

Since most of the migration occurs in the ages 15 to 34, the migration model is most concerned with estimating the migration flows for those ages. At present, it does not seem necessary to distinguish different migration patterns within each slope value for the remaining age groups. Thus, in Table 1, the numbers printed under CS for these other age groups are to be applied across all patterns within the slope category. Further study may indicate that this procedure should be modified.

Procedure Used

Since migration flows in a particular area are apt to be considerably different for each race-sex group, the following procedure is used once for each group. Slope assignments are made on the basis of independent net migration estimates or forecasts—not on historical rates. The

Table 1.—Model Directional Migration Rates by Age, Slope, and Pattern Type

Age	Pattern Type											
	CS	CT	CX	ES	ET	EX	IS	IT	IX	LS	LT	LX
						Slope 0.09						
0-4	.1490											
5-9	.2450											
10-14	.1980											
15-19	.3800	.493	.620	.235	.235	.235	.220	.220	.220	.180	.180	.180
20-24	.4000	.400	.450	.400	.500	.620	.330	.475	.580	.280	.280	.280
25-29	.3260	.326	.326	.326	.385	.385	.330	.420	.520	.326	.450	.550
30-34	.2650	.265	.265	.265	.265	.265	.265	.290	.290	.265	.330	.330
35-39	.2150											
40-44	.1750											
45-49	.1420											
50-54	.1160											
55-59	.0940											
60-64	.0764											
						Slope 0.12						
0-4	.1250											
5-9	.1900											
10-14	.1140											
15-19	.3500	.498	.620	.175	.175	.175	.160	.160	.160	.125	.125	.125
20-24	.3770	.377	.425	.377	.450	.550	.290	.420	.520	.240	.240	.240
25-29	.2860	.286	.286	.286	.330	.330	.290	.370	.470	.286	.390	.480
30-34	.2170	.217	.217	.217	.217	.217	.217	.235	.235	.217	.255	.255
35-39	.1650											
40-44	.1250											
45-49	.0948											
50-54	.0719											
55-59	.0546											
60-64	.0414											
						Slope 0.15						
0-4	.0715											
5-9	.1040											
10-14	.0722											
15-19	.2200	.337	.425	.090	.090	.090	.080	.080	.080	.060	.060	.060
20-24	.2390	.239	.280	.239	.300	.540	.180	.275	.350	.120	.120	.120
25-29	.1690	.169	.169	.169	.200	.200	.180	.240	.310	.169	.250	.325
30-34	.1200	.120	.120	.120	.120	.120	.120	.130	.130	.120	.150	.150
35-39	.0848											
40-44	.0600											
45-49	.0425											
50-54	.0301											
55-59	.0213											
60-64	.0151											

NOTE: C = College, E = Early, I = Intermediate, L = Late
S = Short, T = Tall, X = Extreme

pattern selection is based on historical migration rates, however.

Slope estimation. The model depends upon an exogenous estimate of total population to determine the total net migration for each race-sex group for the initial projection or estimation interval. The volume of net migration is calculated as a residual after comparing this independent estimate of population with the initial population survived over one time period. The ratio of this total net migration to the survived population is the total net migration rate.

Figure 3 shows a plot of net migration rates for age group 40–44 vs. the total net migration rate for the race-sex group. It appears from this plot that the net migra-

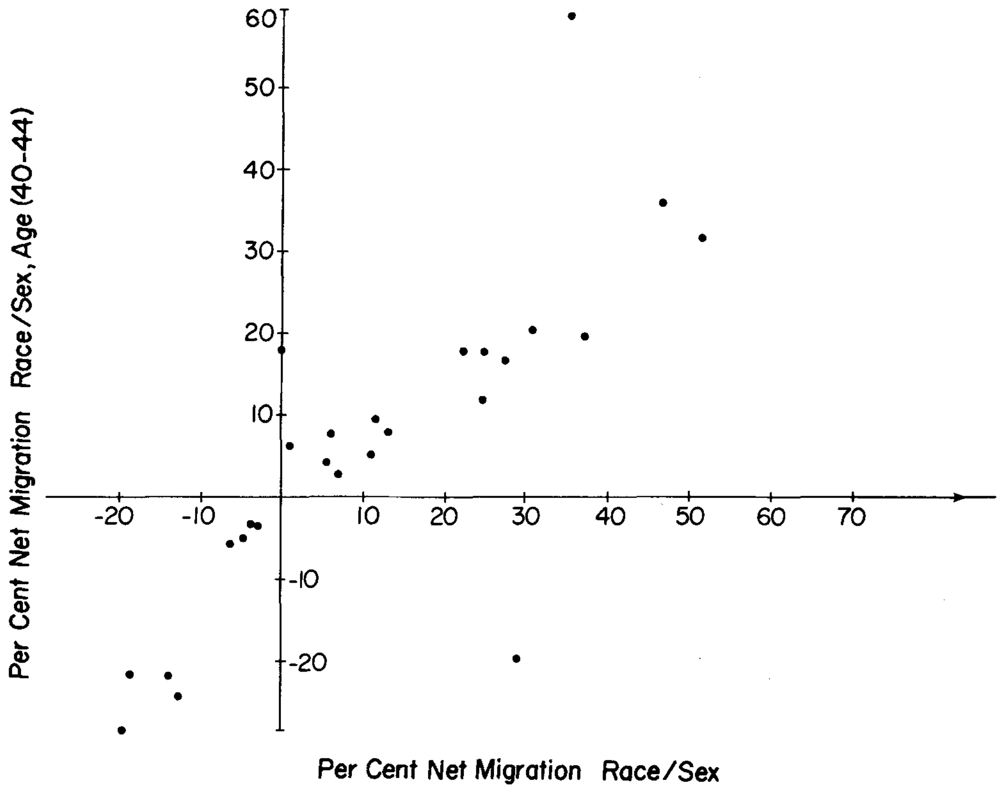


Figure 3.—Relationship of Percent Migration for a Race, Sex, Age Cohort (40–44) to Percent Migration for Total Race, Sex Group

tion rate at the age group 40–44 can be approximated by that of the total race-sex group.

It is now necessary to go from the percentage net migration at age 40–44 to directional migration for this age group. Figure 4 shows plots of the in-migration rate at age 40–44 vs. net migration at age 40–44. The upper graph is for males and the lower is for females. Although there is some scatter in the plots, it does appear that it is possible to use the net migration rate for the race-sex group to estimate the in-migration rate for age cohort 40–44. The out-migration rate is then obtained using the identity Out = In-Net.

As was mentioned earlier, the slope of the directional migration rate at ages 40–44 tends to become less steep with increasing migration. This information was used in defining the directional migration rates. By using the linear relationships implied by Figure 2, it is possible to estimate a slope for each in- and out-migration rate.

Chosing the directional migration patterns. Once the slope has been determined for a race-sex group in a particular area, a decision still must be made as to which of the 12 patterns within the slope grouping best describe the character of this area. The applicable pattern is identified by examining the historical intercensal net migration rates for the age groups 15–19 through 35–39. Since these are the age groups where the majority of the migration occurs and where changes in inflection of net rate patterns are usually found, the differences between migration patterns are most evident here. Let the net migration ratio be defined as

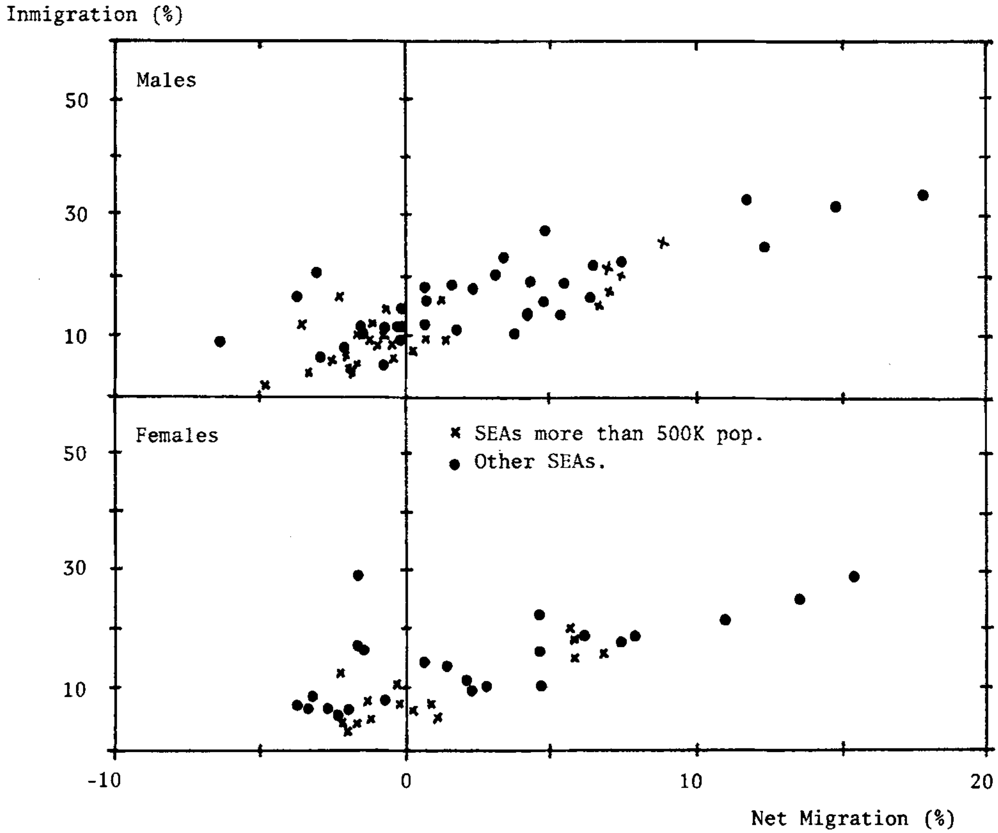


Figure 4.—Relationship of Immigration to Net Migration at Ages 40–44, Selected State Economic Areas, 1965–1970

one plus the net migration rate. Then we can calculate the net migration ratio for age cohorts 15–19 through 35–39, i.e., for age groups with indices 4 through 8.

As an example, Table 2 shows data for white males in West Virginia for the decade 1960–1970 (Bowles et al., 1975, Part 3, page 64). Of the five age groups of interest, the eighth group, ages 35–39, has the highest ratio and the fifth group, ages 20–24, has the lowest ratio. By convention, denote this rank pattern as 85, i.e., the index of the highest ratio is first and the index of the lowest ratio is second. Furthermore, let the amplitude, A, be defined as the high ratio minus the low ratio, or, in this example,

$$A = .917 - .607 = .310.$$

Table 3 was established to assign in- and out-migration patterns according to the rank pattern and the amplitude A. The pattern assignments were defined after studying plots of past migration rates for many areas. The rank pattern locates the position of the peak for both the in- and out-migration flows—whether college (C), early (E), intermediate (I), or late (L). The amplitude (A) is used to estimate the height of the peak—short (S), tall (T), or extreme (E).

Retirement Migration

Retirement migration is handled separately for two important reasons:

(1) The migration patterns for the retirement age population can vary consid-

Table 2.—White Males in West Virginia, 1960–1970

Age Index	Age Group	Migration Rate	Migration Ratio	Rank
1	0–4	-.033		
2	5–9	-.099		
3	10–14	-.088		
4	15–19	-.140	.860	
5	20–24	-.393	.607	Low
6	25–29	-.346	.654	
7	30–34	-.096	.904	
8	35–39	-.083	.917	High

erably within migration patterns exhibited by the population less than 65. The factors that cause the retirement age population to in-migrate or out-migrate are often independent of those affecting the population less than 65.

(2) Fairly good estimates of retirement age migration can be obtained by using Medicare data on the population over 65, if the model is to be used for making intercensal or postcensal age estimates rather than for forecasting.

Migration of the population over 65. Overall retirement age net migration is forecast exogenously and distributed by age. From postcensal estimates, this migration might be treated as in the following example. By comparing the 1970 population that survived to 1975 with the 1975 Medicare based estimates (U. S. Bureau of the Census, 1980), estimates of net migration can be obtained for those cohorts over 65. For each cohort, the net migration rate is calculated as the net migration divided by the 1970 population survived to 1975. These net migration rates are used for the age groups 65–69, 70–74, and 75+.

Retirement-related migration. Retirement related migration does not necessarily begin at age 65. Some people, for reasons of health or finance, retire well before they reach the age of 65. Since wives are apt to be younger than their husbands, there appears to be considerable retirement-related migration for females less than age 65.

Figure 5 illustrates this for areas that are well known for their in- or out-migration of the retirement age population. The upper plots show net migration rates in Arizona and Florida for males and females. The lower plots show net migration rates in New York and Illinois for males and females. In all cases, the bulge due to retirement migration starts well before the age group 65–69. Thus, in areas experiencing large retirement migration, the migration rates of the age groups just below age 65 should be modified to account for this. Furthermore, it should be noted that the change in migration rates due to retirement for females preceeds that for males.

An area is considered to have "retirement" migration for a given race-sex group if the net migration rates for all age groups 65 and over of that race-sex have the same sign. Since this retirement migration also is having some impact on the age groups just under age 65, the migration rates of these age groups must be modified accordingly.

The following modification is made for

Table 3.—Migration Pattern Assignments

Rank Pattern	Assignment
45, 46, 47, 48[a]	CS - ET if A < .30 CT - ET if .30 < A < .65 CX - IX if .65 < A
56, 57, 58	ES - LS if A < .20 ET - LT if .20 < A < .60 EX - LT if .60 < A
64, 65, 67, 68, 54, 78	IT - LS if A < .35 IX - LS if A > .35
74, 75, 76, 84	LS - ES if A < .20 LT - ET if .20 < A < .60 LX - ET if .60 < A
85, 86, 87	LS - IT if A < .30 LS - IX if A > .30

NOTE: A = amplitude

[a]If, for rank patterns 45, 46 or 47, the ratio of the net migration ratio for age group 4 to that for age group 8 is < 1.125, then the assignment should be that for rank pattern 86. This is to distinguish "true" college patterns from patterns more symmetrical in their outflow of young adults.

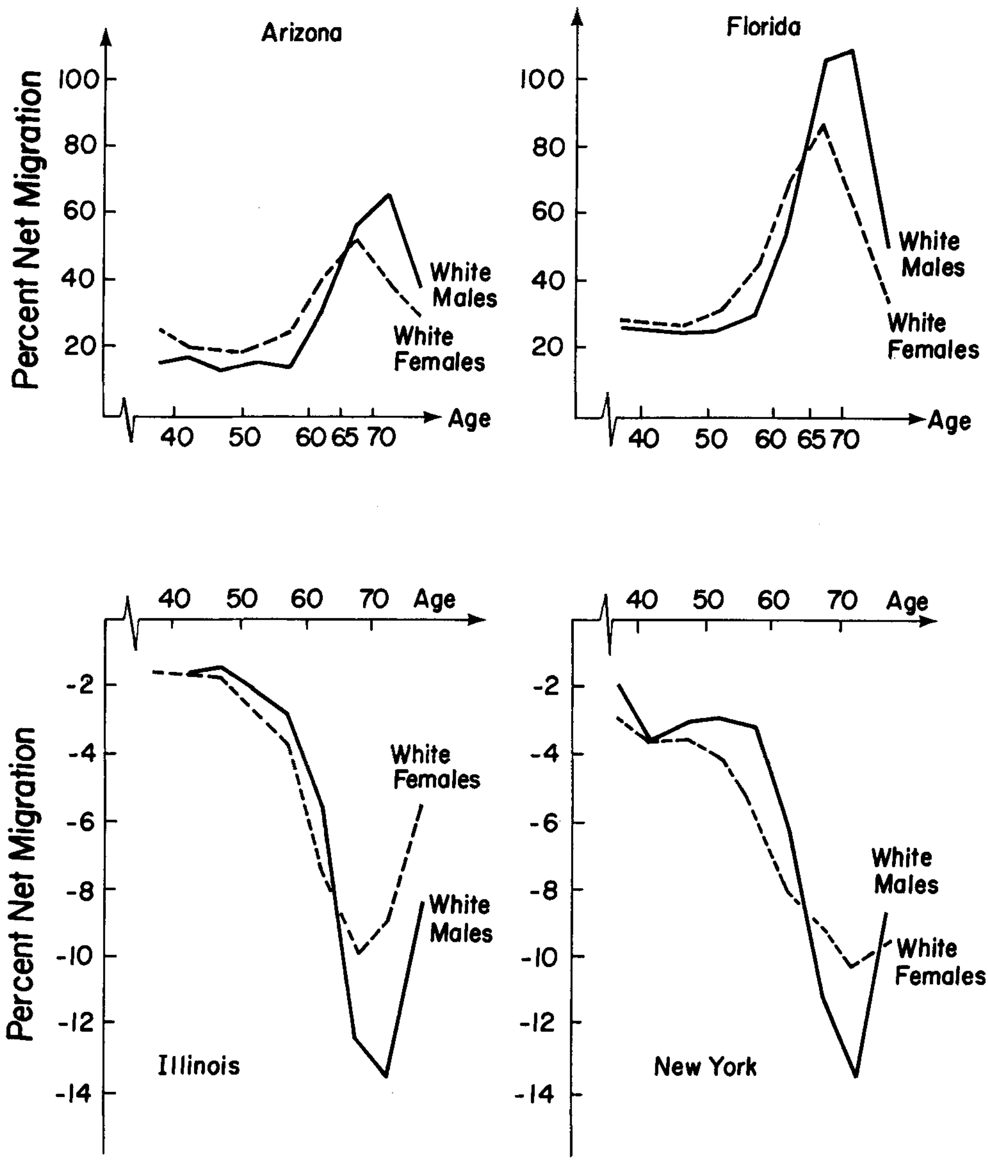


Figure 5.—1960–1970 Net Migration Rates for Selected Age Cohorts

those areas that are experiencing retirement migration, whether in or out. The net migration of the population 65 and over for each race-sex group is calculated by summing over the age groups 65–69, 70–74, and 75+. Using Table 4, a retirement-related migration is calculated for each age group 45–49 through 60–64 by multiplying the total retirement migration by the percentage corresponding to that age group. This retirement-related migration is converted to a rate by dividing by the corresponding survived population. These adjustments are added to the estimated net migration rates defined by the model. These percentages were estimated from 1965–1970 Census data for selected states.

Table 4.—Retirement Related Migration As a Percentage of Migration of the Population Over 65

Age	Males	Females
45-49	0.0	2.5
50-54	2.5	7.0
55-59	7.5	17.0
60-64	22.5	32.5

Adjustments for Special Populations

As is common with most cohort component population projection models, the special populations are handled separately (Schroeder, 1980). Since the population projections in the project that funded this research are only concerned with the civilian population, once the military population has been subtracted out of the base population, they are left out for the rest of the projection process. The college population is also subtracted out of the base population and is then added back in after the projection process to obtain the population at the end of the period.

Both net migration rates and directional migration rates based on census data are usually calculated for the total resident population. Census data limitations make it difficult to delete the migration of college students or the military. Thus, the 1960–1970 patterns from which the model is calibrated in this example also include the military and student migration. Other researchers may be able to correct for this if data on college and military populations in both censuses are available in convenient form.

Correcting for military population. The observed migration rates on which the pattern selection depends, in this illustration, are the net migration rates from 1960 to 1970. In that period, there were relatively few females in the military. In an area with a considerable military population, male migration rates would be more affected by the presence of the military than would female migration rates. It was felt that the civilian male migration rates could be better approximated by the female migration rates of the same race, rather than by the observed male migration rate of that race. Thus, in these areas, the observed male migration rates are replaced by the observed female migration rates.

Student population. In those areas with a sizable student population, the college students are subtracted from the 1970 civilian population. The 1970 civilian noninstitutional population is then projected to 1975. Just before forcing the individual race-age-sex cohorts to sum to an independent population control total, the student population is added back in. Due to the lack of more current nationwide data comparable to the 1970 Census data on students, it is assumed that the student population in a given area does not vary after 1970, i.e., the student population is held fixed. Users of the model may choose to incorporate data based on an alternative assumption.

By handling the student population separately rather than with the cohort component procedure, a net migration is implicitly assumed. To avoid a double counting of student migrants (by the student model and by the migration model), the net migration estimated by the migration model is adjusted by subtracting out the net migration implied by the student model.

Calibration

This section covers the steps involved in calibrating the patterns and applying the various adjustments to yield the desired net migration total.

First, the in- and out-migration patterns are each scaled so that the rates for age group 40–44 are the percentage in- and out-migration, respectively, that were estimated as described above. A trial net migration rate vector is formed by subtracting the scaled out-migration pattern from the scaled in-migration pattern and adding the various adjustments for retirement and the special populations. This trial net migration rate vector is applied to the survived population plus births in that period, to obtain an estimated net migration. The sum over this estimated net migration is compared with the desired net migration to obtain an error term. The scaled in-migration rates are then multiplied by another scalar to correct for this error. (In-migration is assumed to be more volatile than out-migration and therefore is the vector that is altered.) Each final net migration rate is the rescaled in-migration rate minus the scaled out-migration rate plus the various adjustments for retirement and the special populations.

Migration patterns are essentially held constant from one forecast interval to the next. Small details such as slope class may be permitted to change when the exogenously determined overall net migration values differ considerably from historical data.

VALIDATION

A thorough testing of this model can not be done until the complete 1980 Census is available. When the authors are able to access Census data from Summary Tape File 4 (STF4), many comparisons of this model can be run. Population data from STF1 and STF2 represent the entire resident population, whereas the model being tested projects just the civilian population. Nonetheless, some comparisons have been run comparing the migration model and the plus-minus technique with the early 1980 Census data.

State Level Comparisons

Some measure of its performance can be obtained by comparing the 1980 state population projections with the available 1980 figures on state population by age (U. S. Bureau of the Census, 1981). Our population projection model was run twice for each state—once using the migration model just described and once using the plus-minus technique. The plus-minus technique was applied to the observed 1960–1970 age-specific net migration rates (Bowles et al., 1975), after these rates had been divided by two to obtain half decade rates; this procedure is not uncommon.

Figure 6 shows the results of comparing both of our 1980 state level population estimates with the early results from the 1980 Census. The first two columns indicate the mean absolute percentage error over all age groups. The two remaining columns present the percentage error for the age group having the greatest relative difference between the projection and the census result.

The mean absolute percentage error of the migration submodel is less than that of the plus-minus technique in all but two states. Furthermore, it is usually less by almost a factor of two. Looking at the maximum error, again the migration model outperforms the plus-minus technique and usually by a factor of two. For several states, Illinois, Louisiana, Montana, New Hampshire, New Jersey and Texas, the percentage error of the worst fitted age group using the migration model is *less than* the mean absolute percentage error using the plus-minus technique!

In calculating the 1980 population figures, the projections were controlled to the 1980 Census totals for each state in order to isolate the effects of the migration techniques with respect to age detail. As was mentioned above, military populations were deleted from the data for 1970 and were not included in the 1980 projections in either technique. The 1980 Census figures do include military personnel which could not be subtracted out. However, it was felt that the errors due to this incomparability were quite small in most states.

Substate Comparisons

With the release of Summary Tape File1 (STF1) of the 1980 Census, some comparisons of the model can be made at substate levels as well. The procedure followed was the same at that used at the

Estimates of the Population of Counties, by Age and Sex: July 1, 1975. Washington, D.C.: U. S. Government Printing Office.
———. 1981. Supplementary Reports, 1980 Census of Population. Age, Sex, Race and Spanish Origin of the Population by Regions, Divisions, and States: 1980. PC80-S1-1. Washington, D.C.: U.S. Government Printing Office.
———. 1982. U. S. Census of Population and Housing, 1980. Summary Tape File 1, 2, and 4.