Andrei Rogers
Lisa Jordan

Estimating Migration Flows from Birthplace-Specific Population Stocks of Infants

When adequate data on migration are unavailable, demographers infer such data indirectly, usually by turning to residual methods of estimating net *migration. This paper sets out and illustrates an inferential method that uses population totals in the first age group of birthplace-specific counts of residents in each region of a multiregional system to indirectly infer the entire age schedule of* directional *age-specific migration flows. Specifically, it uses an estimate of infant migration that is afforded by a count of infants enumerated in a region other than their region of birth to infer all other age-specific migration flows. Since infants migrate with their parents, the migration propensities of both are correlated, and the general stability of the age profiles of migration schedules then allows the association to be extended to all other age groups.*

1. INTRODUCTION

The next national census in the United States will be the first since the 1930 enumeration to lack a question on internal migration. The U.S. Census Bureau plans to drop its long-form questionnaire in 2010 and to replace it with a continuous monthly survey called the American Community Survey (ACS), which is based on its successful Current Population Survey (CPS). This change, a response to the pressure from Congress to reduce the cost of the decadal census, will provide more up-to-date information, but the absence of the larger sample provided by the census count will complicate the measurement and analysis of internal migration flows for use in small area population projections, for example. The ACS will provide more timely data, but the samples will be smaller than have been provided by the census, and the strategy of averaging accumulated samples over time will mix changing migration patterns. Finally, the migration question will refer to a one-year time interval instead of the

This research is being supported by a grant from the National Science Foundation (SES-02040808). The authors wish to acknowledge the help of James Raymer in the earlier stages of this research.

Andrei Rogers is Professor of Geography and Director of the Population Program at the Institute of Behavioral Science, University of Colorado (andrei.rogers@colorado.edu). *Lisa Jordan is a research assistant and graduate student at the Population Program, University of Colorado* (lisa.anselmi@colorado.edu).

Geographical Analysis, Vol. 36, No. 1 (January 2004) The Ohio State University
Submitted: January 1, 2003. Revised version accepted: May 7, 2003.

five-year interval used in the decennial censuses of 1960 through 2000, complicating historical comparisons and multiregional projections based on five-year age groups. Consequently, students of territorial mobility increasingly will find it necessary to complement or augment possibly inadequate data collected on migration with data obtained by means of "indirect estimation."

When adequate migration data for a particular study region are unavailable, social scientists generally estimate such data *indirectly* by combining related information (perhaps drawn from several sources, time periods, and geographical areas) with models that produce estimates of migration on the basis of information that may be only indirectly related to its value. When confronted with the absence of *any* data on migration streams, for example, analysts have turned to residual methods of estimating *net* migration (e.g., Bogue, Hinze, and White 1982). Such methods attribute differences in population totals between two dates to natural increase and to net migration. By subtracting the known or estimated contribution of the former, they obtain the latter as a residual. The estimates produced by such techniques can be subject to considerable error in instances of differential rates of net undercounts between pairs of censuses. And even in instances of similar undercount levels, residually estimated net migration totals accumulate many other possible errors (such as age misreporting) made in the estimation process.

Residual methods of estimating net migration continue to be used today, generally coupled with minor refinements in technique or data. One such refinement, birthplace-specific net migration, has been used when birthplace-specific population stock data are available (Eldridge and Kim, 1968). For example, this method was applied by Miller (1994) to analyze historical data (the censuses of 1900 and 1910) made available by the Integrated Public Use Microdata Series (IPUMS) project based at the University of Minnesota.

This paper sets out a method that uses the population totals in the first age group of birthplace-specific population data to indirectly infer the entire age schedule of *directional* origin-destination-specific migration flows, and it applies this method to data provided by the same historical censuses studied by Miller (1994). The focus is on the internal migration of the U.S.-born population. The method will be useful to at least three user communities: (1) historical demographers and geographers seeking to identify changing mobility patterns hidden in the recently available historical population censuses that lack a migration question; (2) migration analysts studying mobility patterns in data poor less-developed countries; and (3) population researchers faced with the prospective loss of migration data previously contained in the to-be-eliminated "long form" questionnaire of past U.S. decennial censuses.

We *test* our particular method of indirect estimation on the post-1950 U.S. decennial censuses that *did* ask a migration question, and then demonstrate and evaluate the methods of indirectly estimating migration using three historical IPUMS data sets drawn from the U.S. censuses of 1900, 1910, and 1920—censuses that *did not* ask a migration question. In addition, we include estimates of interregional migration for the 1935 to 1940 and 1945 to 1950 periods, intervals for which the 1940 and 1950 censuses did provide data on migration, but only for a small sample in the 1940 census and only for a one-year interval in the 1950 census. Data for the 1930 census are not yet available in a PUMS version.

2. THE INDIRECT ESTIMATION OF MIGRATION PROPENSITIES

Although indirect estimation techniques have been applied fruitfully in studies of mortality and fertility, they have not been developed as systematically and formally for the analysis of migration. For example, the United Nations manual on the subject is very explicit in its non-coverage of migration: "A further limitation of the *Manual* is that it deals mainly with the estimation of fertility and mortality in developing countries. There are other demographic processes affecting the populations of these countries (migration for example) which are not treated here" (United Nations 1983, p. 1).

The first set of "model" mortality schedules published by the United Nations summarized the age-specific death rates of 158 life tables of national populations by using " . . . regression equations which linked the probability of death in each five-year age interval with the corresponding probability in the previous age interval. . . . Thus model schedules could be calculated by assigning alternative probabilities of infant death from very high to very low, and associating with each . . . the schedule of death probabilities in successive groups calculated from the corresponding regressions" (Coale and Trussell 1996, p. 475). The set of life tables so developed would be appropriate for describing the mortality schedule of a particular population so long as the age patterns of death rates were similar in different populations at roughly the same level of mortality, and so long as the 158 life tables were based on reasonably accurate data. An analogous approach for describing the migration schedule of a particular population seems to be an obvious starting point for the estimation of territorial mobility.

The justifications listed for the persistence of regularities in the age profiles of mortality have their counterparts in the search for comparable regularities in the age profiles of migration. Young adults in their early twenties usually exhibit the highest migration rates, and teenagers, in their senior year in high school, the lowest. The migration rates of children necessarily mirror the rates of their parents; thus the migration rates of infants exceed those of adolescents. Retirees, migrating for non-job-related reasons, tend to move to regions with warmer climates and to locations with relatively high levels of social services and cultural amenities, often creating "retirement peaks" in those flows that originate in the Snowbelt states and end in the Sunbelt states.

In several studies of regularities in age patterns of migration, the senior author and his students (e.g., Rogers and Castro 1981; Rogers and Watkins 1987; Rogers and Little 1994) discovered that the mathematical expression called the *multiexponential function* provides a remarkably good fit to a wide variety of empirical interregional migration schedules. That goodness-of-fit has led a number of demographers and geographers to adopt it in various studies of migration all over the world. The multiexponential model migration schedule has been fitted successfully, for example, to migration flows between local authorities in England (Bates and Bracken 1982; 1987); Sweden's regions (Holmberg 1984); Canada's metropolitan and nonmetropolitan areas (Liaw and Nagnur 1985); Indonesia's regions (United Nations 1992); the regions of Japan, Korea, and Thailand (Kawabe 1990); and South Africa's and Poland's national patterns (Hofmeyr 1988; Potrykowska 1988). Most recently, Statistics Canada has adopted the multiexponential model migration schedule to produce its provincial population projections (George et al. 1994), and doctoral dissertations have applied it to represent interregional migration flows in Mexico (Pimienta 1999) and in Indonesia (Muhidin 2002). Collectively these studies have shown that age profiles of migration fall into two fundamental "families": those exhibiting a retirement peak and those that do not. The multiexponential function, described below, accommodates both families of age profiles. (Yet a third "family" may be specified: one that includes an "upward slope" at the oldest ages; see Rogers and Watkins 1987.)

Figure 1 illustrates a typical observed migration schedule (the round dots) and its graduation by a multiexponential model migration schedule (the superimposed smooth outline) defined as the sum of four components:

1. A single negative exponential curve of the *pre-labor force ages,* with its descent parameter α_1;

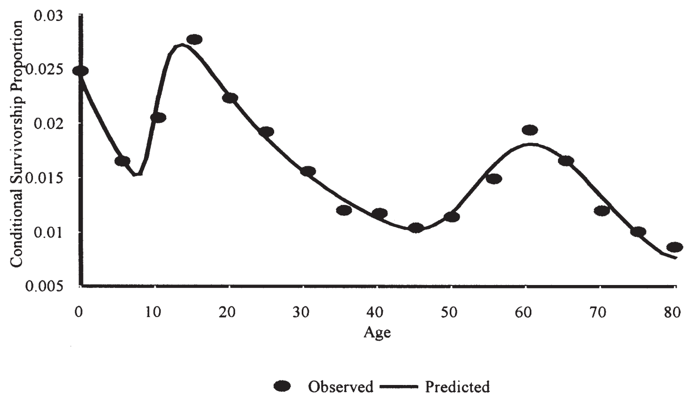

FIG.1. Observed and Predicted (11 Parameter Model Migration Schedule) U.S. Internal Migration from the Northeast to the South, 1955–1960

2. A left-skewed unimodal curve of the *labor force* ages positioned in μ_2 on the age axis and exhibiting parameters of ascent (λ_2) and descent (α_2);
3. An almost bell-shaped curve of the *post-labor force* ages positioned in μ_3 on the age axis and exhibiting parameters of ascent (λ_3) and descent (α_3);
4. A constant curve A_0, the inclusion of which improves the fit of the mathematical expression to the observed schedule, and three constants A_1, A_2, and A_3, that define the relative levels of their three associated curves.

Estimates of the parameters of the model migration schedule are obtained using a nonlinear algorithm that searches for the "best" parameter values for the parameterized model migration schedule:

$$\left.\begin{aligned} S(x) = {} & A_1 \exp(-\alpha_1 x) \\ & + A_2 \exp\{-a_2(x-\mu_2) - \exp[-\lambda_2(x-\mu_2)]\} \\ & + A_3 \exp\{-a_3(x-\mu_3) - \exp[-\lambda_3(x-\mu_3)]\} \\ & + A_0 \end{aligned}\right\} \tag{1}$$

where $S(x)$ denotes the conditional migration probability at age x. Frequently the retirement peak is absent, and the function then is defined by 7 parameters. (In yet other instances, an upward slope at the oldest ages is evident, in which case a positive exponential curve is added, and the function then is defined by 9 parameters if a retirement peak is absent and by 13 parameters if it is present.)

The observed regularities in the age patterns of countless migration schedules suggest that, as in the case of mortality patterns, information on the probabilities associated with infants may be linked with the corresponding probabilities in each of the subsequent age intervals, and therefore also of all age intervals aggregated together, by means of regression equations. Empirical patterns show one of three prototypical age profiles of migration but exhibit different levels under those profiles (i.e., related areas under the different parts of prototypical age profiles). This paper draws on such observed regularities to set out a procedure for obtaining estimates of age and origin-destination-specific migration propensities from knowledge only of the propensity exhibited by infants.

3. OBTAINING AN INITIAL ESTIMATE

A single census distribution that is birthplace-specific (as well as age- and residence-specific) offers in its very first age group an indicator of migration. Children who are, say, 0–4 years old at the time of the census and living in region *j*, having been born in region *i*, must have migrated during the immediately preceding five-year interval. Given their young age, and the fact that they were on average born $2^1/2$ years ago, it is unlikely that they experienced more than one migration. These data provide our initial estimate of spatial patterns and of migration level. Regression equations may be used to expand these child migration levels and spatial patterns into age-specific levels and patterns.

To illustrate the method, consider the data from the 1990 census presented in Figure 2 below, which shows a plot of the aggregate conditional survivorship proportion, $S_{ij}(+)$, against the corresponding first age-group-specific component of that aggregate proportion, $S_{ij}(-5)$. The former represents the fraction of persons *of all ages* who resided in region *i* at the start of the time interval and in region *j* at the end of it. The latter is the first member of the set of age-group-specific proportions $S_{ij}(x)$, that in a suitably weighted linear combination comprise the former; it represents the fraction of all births born in region *i* during the past, say, five years who survived to the census date to enter the 0–4 years age group resident in region *j* at that date. Consequently, it can be calculated by back-casting to region *i* all i-born 0–4 year olds enumerated at the time of the census, no matter where they lived, and then deriving the fraction of that number who ended up in region *j* at the time of the census count. (The $S_{ij}(-5)$ measure is defined on pages 98–99 of Rogers 1995.) Note that there are twelve observations, one for each interregional flow present in the 4 by 4 matrix of aggregate migration streams reported in the 1990 census for the U.S. regional disaggregation into the Northeast, Midwest, South, and West.

Returning to the scatter plot in Figure 2 we notice that a straight line offers a good approximation of the relationship between the infant migration level ($S_{ij}(-5)$) between regions *i* and *j*, and the corresponding level across all ages. Table 1 shows that the adjusted R^2 is 0.86 and the *t*-statistic is 8.3. Adding the corresponding data for three other censuses (1960, 1970, and 1980) lowers the adjusted R^2 to 0.77, but increases the *t*-statistic to 12.4.

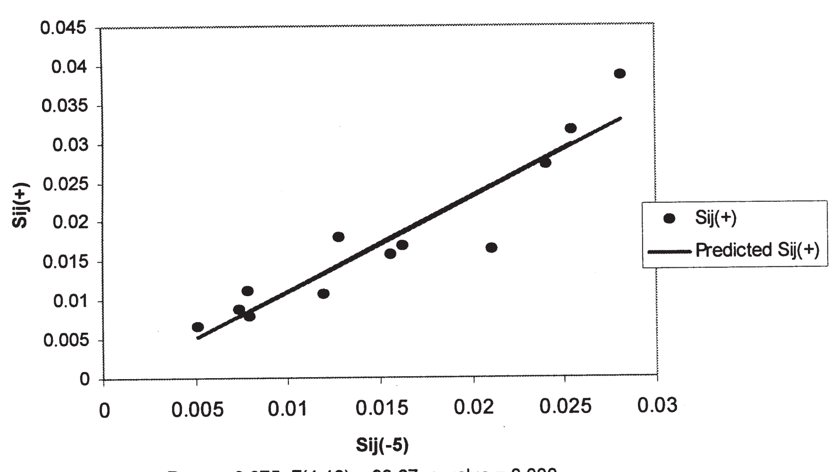

FIG.2. Aggregate Survivorships ($S_{ij}(+)$) as a Function of Infant Survivorships ($S_{ij}(-5)$) for 1990

TABLE 1
The Simple Regression Model

Year	Constant	*t*-score	Slope $S_{ij}(-5)$	*t*-score	*R*-sq	Adj *R*-sq	*N*
1960	0.002	0.615	0.904	5.929	0.779	0.756	12
1970	0.002	0.608	0.805	5.054	0.719	0.691	12
1980	−0.001	−0.322	1.147	7.193	0.838	0.822	12
1990	−0.001	−0.405	1.198	8.347	0.875	0.862	12
1960–1990	0.001	0.416	0.991	12.406	0.780	0.765	48

Somewhat more robust results can be obtained by adding a second independent variable to the regression equation. Let ${}_iK_j(+)\%$ denote the percentage of *i*-borns of all ages who are enumerated in region *j* at census time, and let

$$S_{ij}(+) = a + bS_{ij}(-5) + c_iK_j(+)\% + error\ term \tag{2}$$

Table 2 sets out the relevant statistics for this multiple linear regression model. Note that the adjusted R^2 for the 12 observations from the 1990 census increases from 0.86 to 0.92, and the adjusted R^2 for the 48 observations generated by the four decadal censuses grows from 0.77 to 0.83.

A visual examination of the data suggests that further improvements might be achieved by disaggregating the forty-eight observations into the twelve with a retirement peak (the flows from the Northeast to the Midwest and to the South, and the flows from the Midwest to the West) and the remaining thirty-six without one. Table 3 reveals that such a disaggregation does indeed improve the quality of the fit achieved by the regression model. The adjusted R^2 for the twelve schedules with a retirement peak is 0.92; that for the remaining thirty-eight schedules without one is 0.85.

Figure 3 demonstrates how these two regression equations, fitted to the forty-eight interregional migration streams enumerated by the 1960, 1970, 1980, and 1990

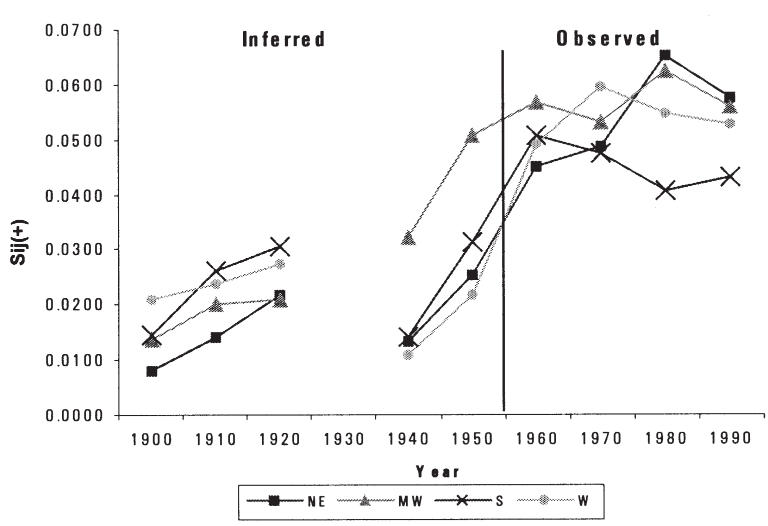

FIG. 3. Historical Regional Survivorships ($S_{ij}(+)$) from 1900–1990

TABLE 2
The Multiple Regression Model

Year	Constant	t-score	Slope $S_{ij}(-5)$	t-score	Slope ${}_iK_j(+)\%$	t-score	R-sq	Adj R-sq	N
1960	0.001	0.322	0.740	4.702	0.0007	1.980	0.846	0.812	12
1970	0.001	0.243	0.776	4.349	0.0003	0.670	0.725	0.663	12
1980	−0.003	−1.059	0.856	4.601	0.0011	2.254	0.897	0.834	12
1990	−0.002	−0.980	0.865	5.290	0.0010	2.772	0.932	0.917	12
1960–1990	−0.001	−0.87	0.778	9.474	0.0009	4.493	0.841	0.834	48

TABLE 3
Two Multiple Regression Models

Year	Constant	t-score	Slope $S_{ij}(-5)$	t-score	Slope ${}_iK_j(+)\%$	t-score	R-sq	Adj R-sq	N
Observations Without Retirement Curve:									
1960	0.002	1.497	0.774	8.619	0.00005	0.146	0.955	0.940	9
1970	0.006	2.251	0.935	8.124	−0.00146	−2.534	0.928	0.904	9
1980	0.001	0.478	0.761	4.052	0.00040	0.582	0.880	0.840	9
1990	0.001	0.542	0.796	4.093	0.00041	0.630	0.878	0.838	9
1960–1990	0.002	2.169	0.769	10.683	0.00001	0.030	0.862	0.854	36
Observations With Retirement Curve:									
1960–1990	−0.006	−1.716	1.252	11.061	0.00059	2.811	0.935	0.921	12

censuses, perform when applied to the infant migration propensities found in the IPUMS census files for 1900, 1910, 1920, 1940, and 1950. The estimated aggregate interregional migration levels that they "predict" for the first half of the twentieth century are significantly lower than those prevailing during the second half, a finding that accords with what has been reported in the scholarly literature on the population history of the United States (Haines and Steckel 2000).

Table 4 sets out the associated predicted and observed 4 by 4 matrices of migration flow numbers and the percent errors for the 1985 to 1990 time period.

Clearly, the best fits were obtained for the destination-specific out-migration flows from the Midwest, and the worst fits for the out-migration streams from the West. These fits are reflected in the respective values of the corresponding statistics in the table, showing the Mean Average Percent Error (MAPE). For all 1990 flows, the MAPE is 10.45. On average, the flow values were underestimated by 7 percent.

4. INTRODUCING THE AGE DISTRIBUTION

Since equation (2) successfully describes the association between the aggregate proportion of migrating as a function of the corresponding proportion for infant migrants, and of the percentage of i-borns of all ages who are residing in region j at census time, it is likely that it also does so for each age group contained in the aggregate. Therefore, let

$$S_{ij}(x) = a + b(x)S_{ij}(-5) + c(x)_iK_j(+)\% + \textit{error term} \tag{3}$$

for $x = 0, 5, \ldots, 80$. Table 5 presents the resulting fit of this model to the 1960–1990 data.

TABLE 4
Predicted and Observed U.S.-Born Migration Flows for 1985–1990 and percent error

Residence in 1985	Residence in 1990* NE	MW	S	W
NE		329,847 **335,024**	1,603,751 **1,537,269**	469,284 **334,806**
MW	343,920 **317,937**		1,672,414 **1,720,916**	946,995 **886,917**
S	760,298 **800,489**	1,180,046 **1,310,702**		1,122,041 **1,005,956**
W	339,742 **304,300**	658,281 **570,839**	1,052,769 **805,489**	

*predicted values are in bold

Residence in 1985	Residence in 1990 NE	MW	S	W	MAPE
NE		1.57	4.15	28.7	11.47
MW	7.55		2.90	6.34	5.60
S	5.29	11.07		10.35	8.90
W	10.43	13.28	23.49		15.73
MAPE	7.76	8.64	10.18	15.13	

TABLE 5

Two Age-Specific Multiple Regression Models

Age	Intercept	*t*-score	Slope $S_{ij}(-5)$	*t*-score	Slope ${}_iK_j(+)\%$	*t*-score	*R*-sq	Adj *R*-sq	*N*
Observations Without Retirement Peak, 1960–1990:									
0	0.001	1.08	1.041	13.72	−0.0002	0.59	0.907	0.901	36
5	0.001	0.77	0.749	9.95	−0.00003	0.91	0.842	0.832	36
10	0.002	1.54	0.656	6.79	0.0002	0.49	0.735	0.719	36
15	0.006	2.01	1.114	6.10	0.0007	1.02	0.717	0.700	36
20	0.008	3.73	1.508	10.97	−0.0003	−0.47	0.861	0.853	36
25	0.005	3.29	1.059	10.15	−0.0001	−0.35	0.843	0.833	36
30	0.003	2.24	0.883	8.78	−0.0002	−0.61	0.793	0.78	36
35	0.002	1.22	0.657	7.10	−0.00002	−0.06	0.731	0.715	36
40	0.001	1.02	0.461	6.33	0.0008	0.29	0.699	0.68	36
45	0.001	0.58	0.389	6.58	0.000004	0.02	0.703	0.645	36
50	0.001	0.53	0.346	5.72	−0.0001	−0.42	0.618	0.595	36
55	0.0002	0.17	0.314	4.92	−0.000003	−0.01	0.568	0.542	36
60	0.0002	0.20	0.337	5.15	−0.0001	−0.41	0.565	0.539	36
65	0.001	0.79	0.269	5.78	−0.0001	−0.36	0.626	0.603	36
70	0.001	0.87	0.265	5.85	−0.0001	−0.28	0.637	0.615	36
75	0.001	1.05	0.239	4.83	0.0001	0.34	0.583	0.558	36
80	0.001	0.76	0.235	3.81	0.0002	0.84	0.515	0.486	36
Observations With Retirement Peak, 1960–1990:									
0	−0.007	−1.92	1.280	9.83	0.0006	2.71	0.920	0.902	12
5	−0.011	−3.60	1.176	11.04	0.0008	4.19	0.939	0.926	12
10	−0.001	−0.13	1.256	4.85	0.0002	0.31	0.724	0.663	12
15	0.034	3.29	1.273	3.52	−0.0009	−1.34	0.613	0.527	12
20	0.010	2.95	1.126	9.51	0.0011	4.85	0.926	0.910	12
25	−0.0002	−0.05	1.202	7.32	0.0008	2.49	0.868	0.839	12
30	−0.006	−1.48	1.162	8.85	0.0007	3.07	0.907	0.886	12
35	−0.008	−2.34	1.003	7.92	0.0008	3.53	0.892	0.869	12
40	−0.010	−3.06	0.905	7.77	0.0008	3.90	0.893	0.869	12
45	−0.012	−3.21	0.991	7.43	0.0007	2.77	0.874	0.846	12
50	−0.017	−3.19	1.272	6.64	0.0007	1.96	0.841	0.806	12
55	−0.025	−3.18	1.864	6.65	0.0007	1.29	0.835	0.799	12
60	−0.016	−1.51	1.977	5.22	−0.0001	−0.21	0.752	0.697	12
65	−0.004	−0.42	1.215	3.98	−0.0004	−0.71	0.646	0.568	12
70	−0.002	−0.30	0.716	3.50	−0.00001	3.50	0.577	0.483	12
75	−0.003	−0.52	0.665	3.39	0.0001	0.27	0.562	0.465	12
80	−0.006	−1.24	0.619	3.76	0.0005	1.75	0.655	0.578	12

For the thirty-six schedules without a retirement peak, the adjusted R^2 values range from a high of 0.90, for $x = 0$, to a low of 0.49 for $x = 80$. All of the t-scores for the first independent variable are statistically significant. For the twelve schedules with retirement peaks the range for the adjusted R^2 values extend from a high of 0.93 for $x = 20$ to a low of 0.47 for $x = 75$. And here again, the first independent variable is statistically significant, having the expected positive signs.

Figure 4 illustrates some of the age patterns that are *predicted* for 1960 to 1990 census periods by the simple model of equation (3), for the particular origin-destination pair of Northeast to South and of South to Northeast, with the corresponding *observed* age patterns set out below the predicted. However, the simple multiple regression model set out in equation (3) does not guarantee a non-negative estimate for $S_{ij}(x)$. To ensure that the estimated conditional survivorship proportions always are non-negative, and range between zero or unity, we turn next to *logistic* regression.

Instead of predicting the survivorship proportions using a linear estimation approach, the logged odds of the survivorship are predicted, then converted back into probabilities. Specifically, let

$$\ln\left(\frac{S_{ij}(x)}{1-S_{ij}(x)}\right) = a(x) + b(x)S_{ij}(-5) + c(x)_i K_j(+)\% + error \tag{4}$$

for $x = 0, 5, \ldots, 80$. Table 6 illustrates the separate regressions of equation (4) for each age category, with and without retirement peaks, using least squares estimation.

As with the previous age-specific model of equation (3), a moderate range of goodness of fit values (adjusted R^2) across age categories is obtained. The adjusted R^2 values are highest for the younger age groups: 0.90 for $x = 0$ in the model *without* retirement peaks, and 0.93 for $x = 5$ in the model *with* retirement peaks. Conversely, the adjusted R^2 values are lowest in the oldest age categories: 0.50 for $x = 80$, in the model without retirement peaks, and 0.54 for $x = 75$ in the model with retirement peaks. Once again, as in the estimation results for equation (3), the coefficient for regional distribution of migrants, ${}_iK_j(+)\%$ is not significant in the model without retirement peaks. However, it is sometimes a significant predictor of the logged odds of $S_{ij}(x)$ in the model with retirement peaks.

The coefficients are described in table 6 in terms logged-odds of $S_{ij}(x)$, making interpretation somewhat cumbersome. Figures 5, 6, and 7 show the results of the predicted survivorships, after the logged odds are transformed back to probabilities. Figure 5 examines the relationships between predicted and observed estimates of the $S_{ij}(x)$ for best and worst fits. The predicted migration schedule from the Midwest to the Northeast for 1960 is the best-fitting with a mean average percent error of 9 percent. The worst-fitting migration schedule is from the Northeast to West for 1980, which had a mean average percent error of 50 percent.

Figure 6 illustrates the age profiles of predicted $S_{ij}(x)$ for four different flows from 1960 to 1990; notice that the Northeast to South flow includes a retirement peak, which produces pronounced rises in the flows at the older ages.

Finally, using the regression results in table 7, figure 7 sets out some of the estimated $S_{ij}(x)$ for 1900 to 1950 (with 1930 missing because of a lack of data). These are our "best estimates" of the historical migration flows in the United States for those time periods.

On average, the age patterns follow the standard regularities observed in empirical schedules, but the flows tend to be underestimated by 9 percent. The age category that is best predicted by the model is that of 20–24 year olds, with a mean average percent error (MAPE) of 11 percent. The worst predicted category is that of the 80–84 year olds with a MAPE of 34 percent. This result is intuitively plausible, since

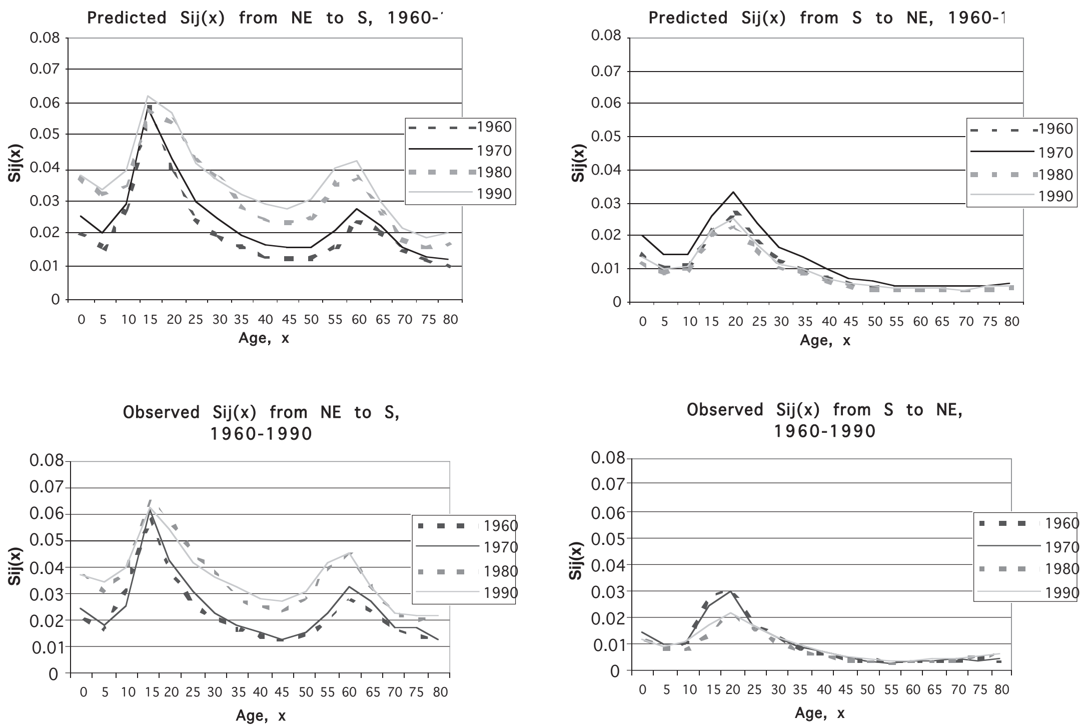

FIG. 4A. Predicted and Observed Migration Schedules from the Northeast to South and South to Northeast

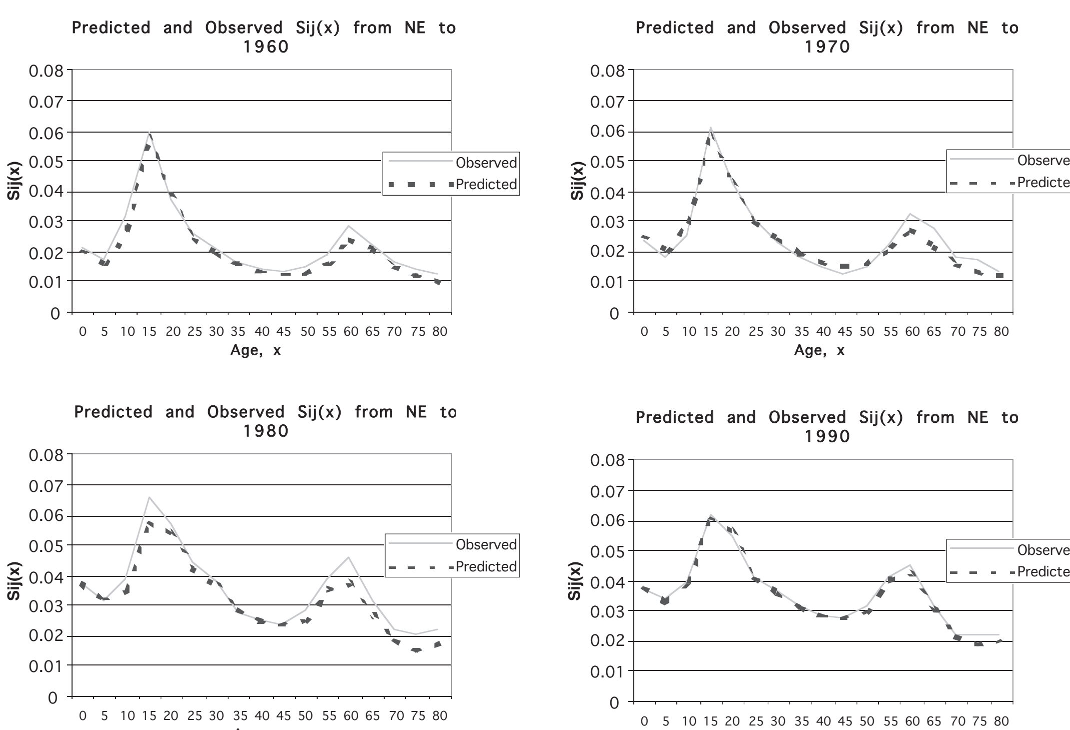

FIG. 4B. Predicted and Observed Migration Schedules from the Northeast to South

TABLE 6

Two Age-Specific Logistic Regression Models

Age	Intercept	t-score	Slope $S_{ij}(-5)$	t-score	Slope ${}_iK_j(+)\%$	t-score	R-sq	Adj R-sq	N
Observations Without Retirement Peak, using a logistic function, 1960–1990:									
0	−5.190	−76.52	55.655	12.33	0.029	1.647	0.908	0.903	36
5	−5.489	−67.911	52.282	9.720	0.039	1.859	0.869	0.861	36
10	−5.286	−49.521	43.571	6.134	0.044	1.572	0.743	0.727	36
15	−4.582	−39.526	39.305	5.094	0.057	1.881	0.697	0.679	36
20	−4.377	−58.376	47.100	9.439	0.019	0.990	0.848	0.839	36
25	−4.726	−66.816	44.533	9.460	0.025	1.331	0.855	0.846	36
30	−5.038	−61.615	46.953	8.628	0.022	1.017	0.826	0.816	36
35	−5.415	−57.053	47.060	7.451	0.033	1.348	0.794	0.782	36
40	−5.748	−56.515	45.201	6.678	0.043	1.627	0.771	0.757	36
45	−6.113	−53.636	49.811	6.568	0.043	1.429	0.759	0.744	36
50	−6.363	−41.667	54.391	5.352	0.031	0.785	0.655	0.634	36
55	−6.596	−34.683	54.477	4.304	0.058	1.161	0.591	0.567	36
60	−6.668	−33.298	58.627	4.399	0.052	0.991	0.588	0.563	36
65	−6.643	−38.462	54.363	4.730	0.047	1.048	0.621	0.598	36
70	−6.572	−38.898	54.717	4.866	0.035	0.783	0.616	0.592	36
75	−6.434	−34.681	50.422	4.084	0.035	0.732	0.536	0.508	36
80	−6.429	−33.504	43.915	3.439	0.075	1.490	0.532	0.504	36
Observations With Retirement Peak, using a logistic function, 1960–1990:									
0	−4.890	−33.848	49.428	9.719	0.021	2.188	0.917	0.898	12
5	−5.273	−39.619	53.083	11.328	0.033	3.772	0.940	0.927	12
10	−4.627	−16.048	48.713	4.800	−0.003	−0.150	0.719	0.657	12
15	−3.302	−15.371	27.583	3.647	−0.021	−1.472	0.664	0.552	12
20	−3.878	−46.681	26.458	9.047	0.024	4.384	0.918	0.899	12
25	−4.428	−29.499	37.759	7.145	0.022	2.283	0.862	0.831	12
30	−4.824	−33.094	44.275	8.628	0.026	2.694	0.900	0.878	12
35	−5.222	−30.405	47.191	7.806	0.035	3.102	0.886	0.861	12
40	−5.554	−30.698	50.058	7.859	0.042	3.562	0.892	0.867	12
45	−5.837	−25.846	62.399	7.849	0.031	2.101	0.880	0.853	12
50	−5.992	−19.205	75.340	6.86	0.022	1.065	0.842	0.807	12
55	−5.912	−16.399	89.677	7.066	0.004	0.149	0.847	0.813	12
60	−5.319	−13.711	86.504	6.335	−0.034	−1.327	0.824	0.785	12
65	−5.205	−12.733	71.987	5.002	−0.041	−1.549	0.754	0.699	12
70	−5.39	−13.762	54.471	3.951	−0.013	−0.52	0.639	0.559	12
75	−5.554	−13.974	53.930	3.854	−0.005	−0.182	0.624	0.540	12
80	−5.707	−17.435	45.639	3.960	0.033	1.564	0.667	0.593	12

[a]Note that the coefficients are the logged-odds of $S_{ij}(-5)$ or ${}_iK_j(+)\%$

Best fit: Worst fit:

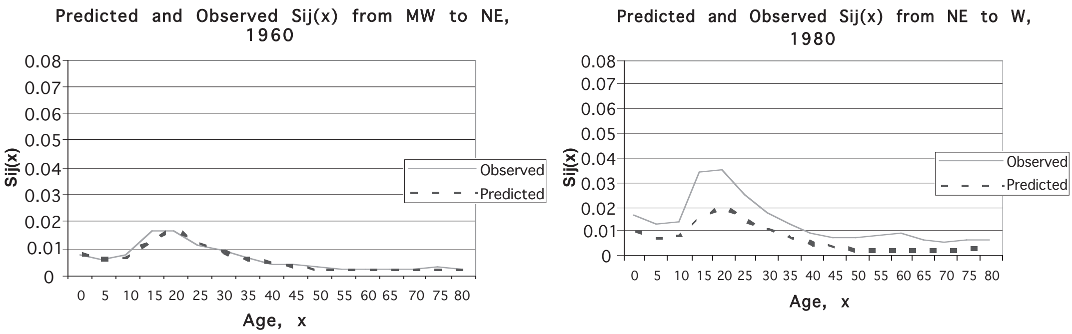

FIG. 5. Predicted versus Observed Migration Schedules: Best Fit and Worst Fit

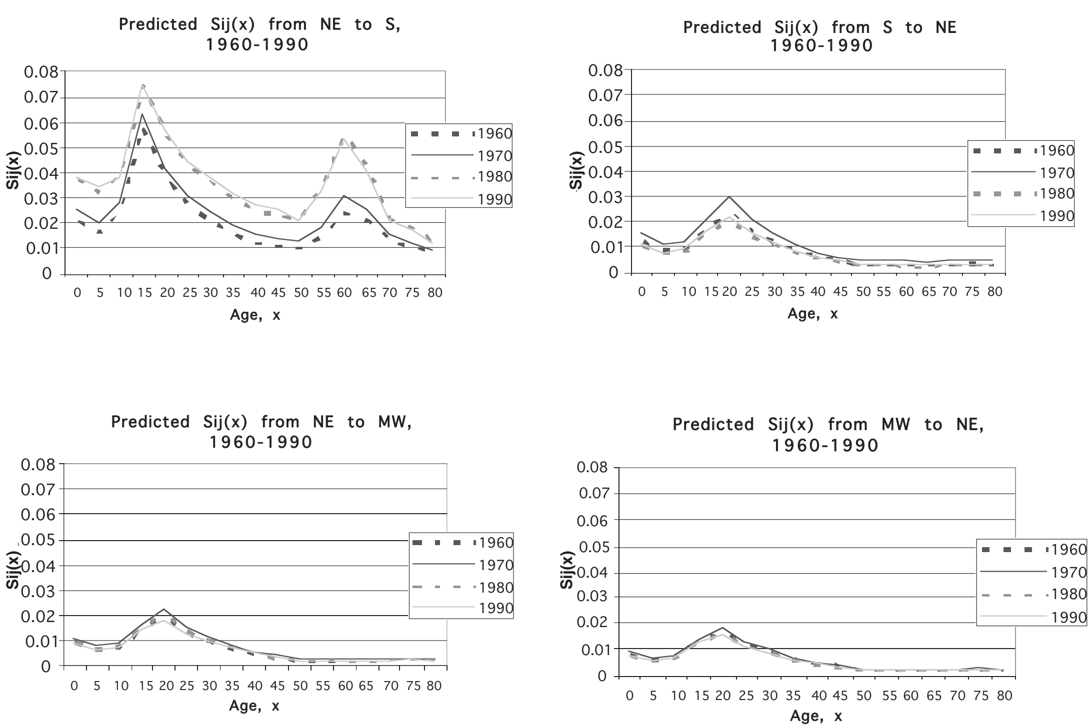

FIG. 6. Predicted 1960–1990 Migration Schedules

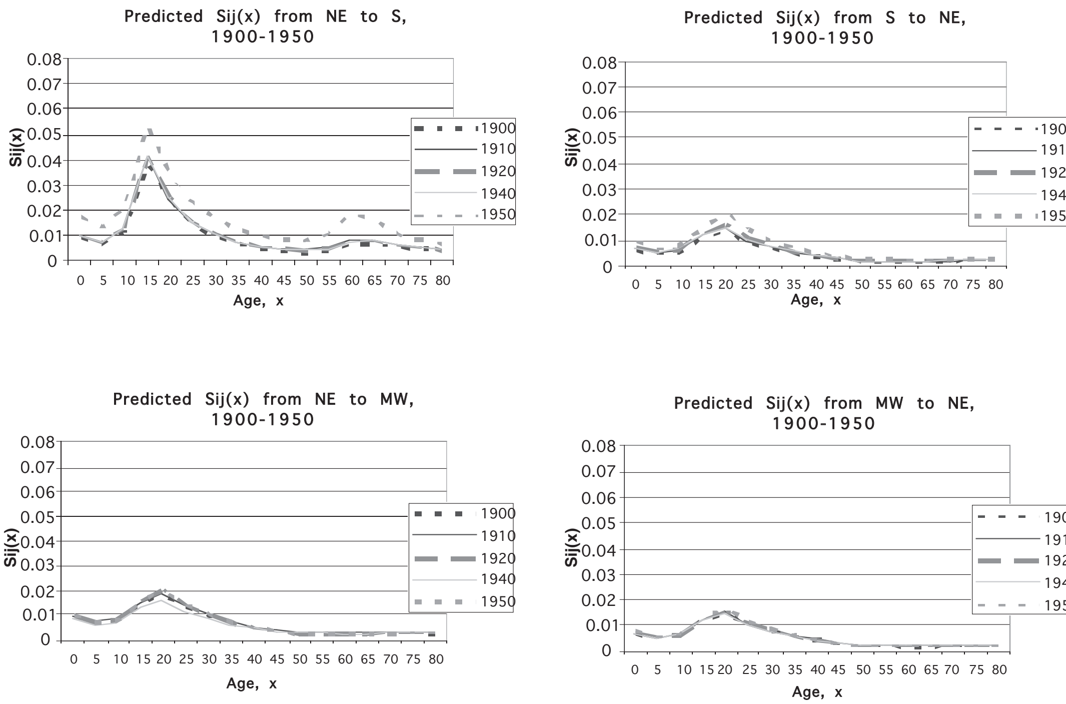

FIG. 7. Predicted 1900–1950 Migration Schedules

infant migration should better predict the migration flow of parents instead of grandparents. The best predicted migration schedules over all the years are from Northeast to South (MAPE = 12 %); the worst predictions for the migration schedules are from the Northeast to the West (MAPE = 39%). Across the four census years, the best predicted schedules were for 1955 to 1960 (MAPE = 19%), and the worst were for 1985 to 1990 (MAPE = 25%).

6. CONCLUSION

The indirect estimation of migration flows (numbers) and propensities (probabilities, or proportions) is made possible by the persistent regularities that are observed in the demographic data. Capturing those regularities with models, therefore, is an important component of any estimation strategy. In the case of migration, two patterns are central to such a strategy: the first must deal with the age structure, the second with the spatial structure. Model migration schedules of the multiexponential variety have been shown to be valuable in accomplishing the former task (Rogers and Castro 1981; Rogers and Raymer 1999); log-linear and logistic models have been successfully used to accomplish the latter task (Sweeney and Konty 2002; Rogers et al. 2002; Rogers, Willekens, and Raymer 2003). It is important to recognize, however, that both patterns are *scale-dependent*, and that the models, therefore, need to be reestimated when applied to changing spatial resolution levels of the underlying data.

In this paper, we have outlined a method for obtaining an initial estimate of migration age and spatial patterns using just one propensity to infer all of the rest: namely, the estimate of infant migration that is afforded by the number of infants enumerated in a survey or census, who at the time of the survey reside in an area different from their place of birth. Since infants migrate with their parents, the two migration propensities have a natural association, one that may be captured by means of a regression equation. And the stability of the age profile of migration schedules then allows the association to be extended to all other age groups.

It could be said that inferring an entire schedule of age-specific migration proportions from just a single infant migration propensity is like reconstructing an entire dinosaur from just a single hip bone. But that is the way of indirect estimation methods in demography. For example, a single infant mortality rate often has been used to estimate an entire life table, because scores of studies have demonstrated that strong regularities in age profiles are exhibited by such tables (United Nations 1955). That is what makes life insurance a workable proposition. We believe that analogous regularities in age profiles, and spatial structures, are exhibited by migration data, and our results are surprisingly successful and merit testing in other settings. A multinational comparative study is being launched by Rogers and an international team of collaborators to carry out such a test (Rogers 2002). In the meantime, we offer a few conclusions that might guide such efforts.

First, there is the question of spatial scale. Indirect estimation of the twelve migration flows associated with a four-region disaggregation of the United States is of limited interest to most people. To be practically useful, it is necessary to increase the resolution level. A first step in that direction is moving toward a larger number of regions, for example, the eight observed destination-specific migration schedules associated with each of the nine U.S. Census Divisions, according to the 1990 census. Our preliminary efforts in this direction indicate that the methodology continues to produce satisfactory estimates, with the same regression models when they are fitted to data reported for the changed resolution level. For example, deleting the three of the seventy-two schedules that exhibit a retirement peak leaves a sample size of sixty-nine schedules, for which the quality of fit afforded by even our simplest univariate age-specific (non-logistic) regression model is impressive with R^2 values that range from a high of 0.99 (for $x = 25$) to a low of 0.77 (for $x = 55$) with all but four exceeding 0.90. All of the slope coefficients associated with the infant migration proportion have the expected positive sign and are statistically significant, but most of those associated with ${}_iK_j(+)\%$ have a negative sign, and almost half are not statistically significant. Nevertheless, even the simplest model predicts reasonable migration schedules, and this suggests that further disaggregations are feasible.

Second, having obtained initial estimates of all migration schedules, one could seek improvements in the quality of the results by fitting model migration schedules to the initially estimated age profiles and logistic models to the initially estimated spatial structures. The latter would involve judgments by the analyst, or a panel of experts, regarding the reasonableness of deviations from past and expected future trends in parameter values or odds ratios. In an earlier article we illustrate such a procedure by adjusting the spatial structure of an observed interregional migration pattern for 1980 to 1985, as exhibited by the (small) sample Current Population Survey (CPS) data, to better reflect the expected trends exhibited by the bordering (large) sample census data for 1975 to 1980 and 1985 to 1990 (Rogers, Willekens, and Raymer 2003).

Third, there is the added option of "cleaning up" the data by fitting model migration schedules to the input data *before* carrying out the regression analysis. Underreporting of the infant population can be a problem in some regions, and such model schedules would "improve" the data. And there also is the possibility of introducing additional independent variables, such as regional dummy variables, into the regression analysis.

Finally, there is the unresolved issue of how to deal with the internal migration of the foreign-born population. Their infant population stocks reflect international, not internal, migration flow totals and so cannot be used to predict the interregional migration patterns inside the United States. Perhaps such patterns could be approximated by linking of the internal migration patterns of the foreign-born to the net migration induced changes in the population stocks of foreign-born adolescents, i.e., by estimating the same regression equations, but using net migration-based estimates of the migration propensities of adolescents instead of infants. We are currently experimenting with this approach.

LITERATURE CITED

Bates, J., and I. Bracken. (1987). "Migration Age Profiles for Local Authority Areas in England, 1971–1981." *Environment and Planning A* 19, 521–35.

———. (1982). "Estimation of Migration Profiles in England and Wales." *Environment and Planning A* 14, 889–900.

Bogue, D. J., K. Hinze, and M. White. (1982). *Techniques of Estimating Net Migration*. Chicago: Community and Family Study Center, University of Chicago.

Coale, A. J., and J. Trussell. (1996). "The Development and Use of Demographic Models." *Population Studies* 50(3), 469–84.

Eldridge, H. T., and Y. Kim. (l968). "The Estimation of Intercensal Migration from Birth-Residence Statistics." *Analytical and Technical Report No. 7*. Philadelphia: Population Studies Center, University of Pennsylvania.

George, M. V. et al. (1994). *Population Projections for Canada, Provinces, and Territories: 1993–2016*. Ottawa: Statistics Canada.

Haines, M. R., and R. H. Steckel, eds. (2000). *A Population History of America*. Cambridge: Cambridge University Press.

Hofmeyr, B. E. (1988). "Application of a Mathematical Model to South African Migration Data, 1975–1980." *Southern African Journal of Demography* 2(1), 24–28.

Holmberg, I. (1984). "Model Migration Schedules: the Case of Sweden." In *Proceedings of the Scandinavian Demographic Symposium*, vol. 6. Stockholm: Scandinavian Demographic Society.

Kawabe, H. (1990). *Migration Rates by Age Group and Migration Patterns: Application of Rogers' Migration Schedule Model to Japan, the Republic of Korea, and Thailand*. Tokyo: Institute of Developing Economies.

Liaw, K. L., and D. N. Nagnur. (1985). "Characterization of Metropolitan and Nonmetropolitan Outmigration Schedules of the Canadian Population System, 1971–1976." *Canadian Studies of Population* 12(1), 81–102.

Miller, A. R. (1994). "Estimating Interregional Migration in the United States from Sample Data." *Historical Methods* 27(1), 5–23.

Muhidin, S. (2002). *The Population of Indonesia*. Amsterdam: Rosenberg Publishers.

Pimienta, R. (1999). "Analisis Multirregional de los Patrones des Edad y Sexo de la Migraciion Interna el Caso de Mexico." Ph.D. diss., Center for the Study of Demography and Urban Development, El Colegio de Mexico, Mexico, D.F.

Potrykowska, A. (1988). "Age Patterns and Model Migration Schedules in Poland." *Geographia Polonica* 54, 63–80.

Rogers, A. (2002). "The Indirect Estimation of Migration: A Proposal for a Multinational Study." Working Paper 02-04, Population Program, Institute of Behavioral Science, University of Colorado, Boulder, Colorado.

Rogers, A. (1995). *Multiregional Demography: Principles, Methods, and Extensions*. London: John Wiley.

Rogers, A., and L. Castro. (1981). *Model Migration Schedules*, Research Report. Laxenburg, Austria: International Institute for Applied Systems Analysis.

Rogers, A., and J. Little. (1994). "Parameterizing Age Patterns of Demographic Rates with the Multiexponential Model Schedule." *Mathematical Population Studies* 4(3), 175–94.

Rogers, A., and J. Raymer. (1999). "Fitting Observed Demographic Rates with the Multiexponential Model Schedule: An Assessment of Two Estimation Programs." *Review of Urban and Regional Development Studies* 11(1), 1–10.

Rogers, A., and J. Watkins. (1987). "General Versus Elderly Interstate Migration and Population Redistribution in the United States." *Research on Aging* 9(4), 483–529.

Rogers, A., F. Willekens, J. Little, and J. Raymer. (2002). "Describing Migration Spatial Structure." *Papers in Regional Science* 81, 29–48.

Rogers, A., F. W. Willekens, and J. Raymer. (2003). Imposing Age and Spatial Structures on Inadequate Migration-Flow Data Sets. *The Professional Geographer* 55(1), 56–69.

Sweeney, S. H., and K. J. Konty. (2002). "Population Forecasting with Nonstationary Multiregional Growth Matrices." *Geographical Analysis* 34(4), 289–312.

United Nations. (1992). *Preparing Migration Data for Subnational Population Projections*. New York: Department of International Economic and Social Affairs, United Nations.

United Nations. (1983). *Manual X: Indirect Techniques for Demographic Estimation*. New York: Department of International Economic and Social Affairs, United Nations.

United Nations. (1955). "Age and Sex Patterns of Mortality: Model Life Tables of Under-Developed Countries." *Population Studies No. 22.* New York: United Nations.