# THE INDIRECT ESTIMATION OF MIGRATION: METHODS FOR DEALING WITH INACCURATE AND MISSING DATA

Andrei Rogers
Population Program
Institute of Behavioral Science
University of Colorado
Boulder, Colorado

December 2009

<u>Acknowledgments</u>: Several generations of students have aided me in carrying out this work over the past three decades; among them were Frans Willekens, Jacques Ledent, Luis Castro, and most recently, Jani Little, James Raymer and Bryan Jones. My thanks to all.

# THE INDIRECT ESTIMATION OF MIGRATION: METHODS FOR DEALING WITH INACCURATE AND MISSING DATA

Andrei Rogers

## ABSTRACT

This paper especially focuses on formal methods for describing, smoothing, imposing, and indirectly inferring migration flows in the absence of adequate, accurate migration data. The term "indirect estimation" is used in demography to describe inferential techniques that produce estimates of a certain variable on the basis of data that may only be indirectly related to its value. We seek a formal model-based approach for using such data to estimate directional (i.e., origin-destination-specific) migration flows. The research reported here should be useful to several three user communities: (1) population researchers faced with the prospective loss of the detailed migration data formerly contained in the to-be-eliminated "long form" questionnaire of past U.S. decennial censuses and its replacement by a significantly smaller continuous monthly sampling survey called the American Community Survey (ACS), (2) historical demographers seeking to identify changing mobility patterns hidden in the increasingly available historical population censuses that lack a migration question, and (3) migration analysts studying migration patterns in data poor less-developed countries, and (4) although the focus is on internal migration, it is believed that the methods can be modified to apply also to international migration. But this is not attempted here.

## 1. INTRODUCTION

In countries with well-developed data reporting systems, demographic estimation is based on data collected by censuses and vital registration systems. Demographic estimation in countries with inadequate or inaccurate data reporting systems often must rely on methods that are "indirect." For instance, the use of the proportion of children dead among those ever borne by women aged 20-24 years to estimate the probability of dying before age 2 is an example of indirect estimation. Such estimation techniques usually rely on parameterized "model" schedules—collections of age-specific rates that are based on patterns observed in various populations other than the one being studied—and select one of them on the basis of some data describing the observed population. The justification for such an approach is that age profiles of observed schedules of rates vary within predetermined limits for most human populations. Rates for one age group are highly correlated with those of other age groups, and expressions of such interrelationships are the basis of model schedule construction.

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 UN 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).

Unlike fertility and mortality, which involve single populations, migration links two populations: the population of the origin region and that of the destination region. This greatly complicates its estimation by indirect methods. What this means in practical terms is that a focus on *age patterns* (as in the case of fertility and mortality) is not enough—one also must focus on *spatial patterns*. The imposition of observed regularities in both the age and spatial patterns of interregional migration to "discipline" inadequate data on internal or international territorial mobility holds great promise as a means for developing detailed age- and destination-specific migration flow data from inaccurate, partial, and even non-existent information on this most fundamental process underlying population redistribution.

## 2. MODELS FOR DESCRIBING MIGRATION DATA

The estimation of migration from aggregate and incomplete data generally has been carried out with a focus on net migration, which is approximated by the population change that cannot be attributed to natural increase. Given data on population sizes at two points in time, and estimates of birth and death rates for the interval defined by these two points, net migration may be approximated by the difference between the observed population at the second point in time and the hypothetical projected population that would have resulted if only natural increase were added to the initial population. Such methods are reviewed in, for example, United Nations (1967) and Bogue et al. (1982).

Methods of inferring gross (directional) migration streams have a much more limited history, (Rogers, 1968; Rogers, 1975; Rogers and Willekens, 1986). In the early years, methods of indirect estimation were geared to particular missing data problems. Consequently, the methods had an ad-hoc character (as do many methods of indirect estimation in demography). More recently, however, the indirect estimation of migration has relied on the use of models and on the theory of statistical inference to approximate the necessary parameters from available data. Some models describe age patterns of migration, while others describe spatial interaction patterns. Both categories of models are considered in an issue of Mathematical Population Studies (Vol. 7, Nov. 3, 1999).

### 2.1 Representing Age Patterns of Migration : Model Schedules

Recognizing that most human populations experience rates of age-specific fertility and mortality that exhibit remarkably persistent regularities, demographers have found it possible to summarize and codify such regularities by means of mathematical expressions called parameterized model schedules. Although the development of model fertility and mortality schedules has received considerable attention in demographic studies, the use of model migration schedules until recently has played a more limited role.

In several studies of regularities in age patterns of migration, we have (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), and 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).

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 several components. For example, as

1. A single negative exponential curve of the pre-labor force ages;
2. A left-skewed unimodal curve of the labor force ages;
3. An almost bell-shaped curve of the post-labor force ages;
4. A constant curve $a_0$, the inclusion of which improves the fit of the mathematical expression to the observed schedule.

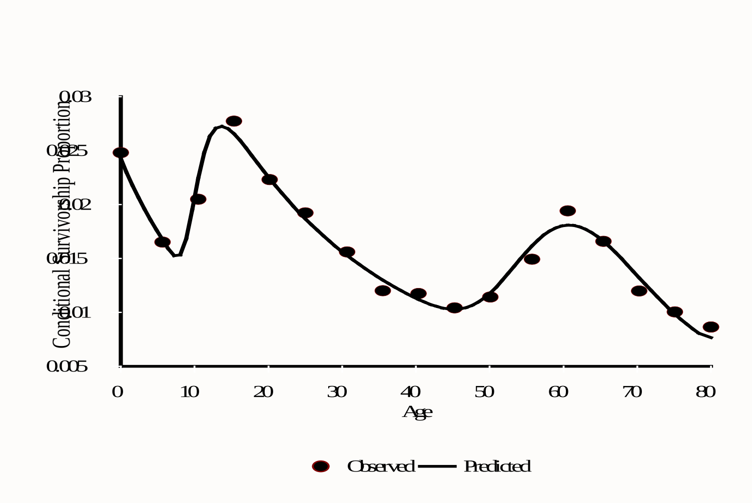


***Figure 1. Observed and Predicted (11 Parameter Model Migration Schedule) U.S. Foreign-Born Internal Migration from the Northeast to the South: 1955-1960***

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, in this illustration, as

$$\begin{aligned}\bar{S}(x) = {} & a_1 \exp(-\alpha_1 x) \\ & + a_2 \exp\{-\alpha_2(x-\mu_2) - \exp[-\lambda_2(x-\mu_2)]\} \\ & + a_3 \exp\{-\alpha_3(x-\mu_3) - \exp[-\lambda_3(x-\mu_3)]\} \\ & + a_0\end{aligned} \tag{1}$$

where $\bar{S}(x)$ 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. Finally, if both the retirement curve and an upward slope are introduced then the function is defined by 13 parameters.

The linkage between age structure and the analysis of fertility, mortality, and migration processes is central to demographic study. Mathematical representations of age patterns of rates or probabilities, called model schedules, have allowed demographers accurately to define such age-specific patterns with continuous functions described with relatively few parameters. The corresponding mathematical representation of spatial patterns calls for a somewhat more complex statistical structure. We believe that a powerful, yet conceptually simple, instrument for the study of aggregate migration spatial structures is offered by the family of generalized linear models, particularly the log-linear model. Just as model schedules are used to make comparisons across time and place, the log-linear specification can be employed as a statistical model that is especially valuable for comparing interregional migration structures across time. The parameters of the log-linear model can be used to not only gauge the relative emissiveness and attractiveness of specific regions, but also to identify the level of interaction between pairs of regions. Because the parameters of the model are interpretable and can be used to characterize migration spatial structure, the log-linear model has the potential for standardizing and enhancing demographic analysis.

## 2.2 Representing Spatial Patterns of Migration : Spatial Interaction Models

Models that describe and predict the numbers of migrations between two regions by relating them to variables describing the characteristics of the origin, the destination, and the “friction” associated with their separation are often called gravity models (Fotheringham and Williams, 1983; Sen and Smith, 1995) even though their current formulations barely resemble the model’s original expression drawn from “social physics.”

The problem of estimating gravity models or, more accurately, spatial interaction models, has been approached from different perspectives over the past four decades. First formulated as an analogy to Newton’s law of gravitation, the resulting purely mechanical approach was revised by Alan Wilson some 30 years ago in terms of entropy maximization theory (Wilson, 1970). This was followed by a behavioral micro-theoretical approach called random utility modeling (MacFadden, 1978). And 20 years ago geographers recognized that models developed in the field of discrete multivariate analysis could be applied fruitfully to express spatial interaction patterns. Foremost among these models has been the log-linear model, and a principal contributor to this recognition has been Professor Frans Willekens (Willekens, 1980, 1982, 1983).

The log-linear model is a powerful instrument for the study of complex data structures. Its use to express traditional models of spatial interaction enhances the opportunities for structural analysis. Questions that the data are expected to help answer may be expressed in terms of the parameters of the model. Furthermore, the model clarifies and simplifies the estimation of spatial interaction flows. And when particular interaction effects cannot be derived from available data, they often may be calculated using other comparable data sets (usually historical data on such interaction effects). In this way different data sets may be combined to develop best estimates of spatial interaction.

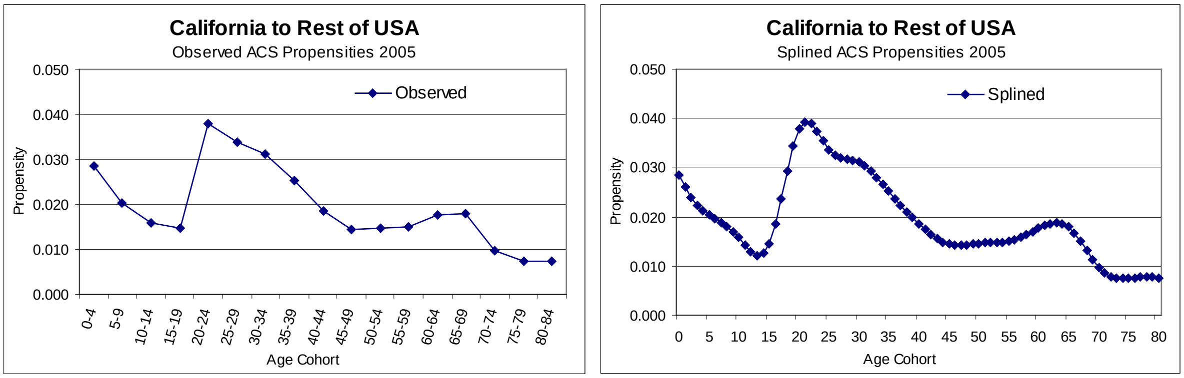
The corpus of work on generalized linear models that we use in our efforts to describe spatial patterns of migration are set out in several published articles that deal with log-linear and logistic models, that may be used to describe, smooth or impose a particular spatial structure on a set of observed migration flows (Rogers, et al., 2001; Rogers, et al., 2002 a, b; Rogers, et al., 2003; Raymer and Rogers, 2007; Raymer and Rogers, 2008).

In these articles we outline our formal model-based approach to defining the spatial structure of a particular observed (or hypothetical) migration flow pattern. We define migration spatial structure to be a particular description of a matrix of interregional migration flows—one that provides an analyst with the means to: (1) reconstruct that matrix of flows, (2) identify the implied relative “push” at each origin and “pull” at each destination (called “emissiveness” and “attractiveness” respectively), and (3) express the origin-destination-specific levels of spatial interaction implied by that matrix. Spatial interaction is here taken to reflect the degree of deviation exhibited by that matrix when compared to the corresponding matrix generated under the assumption of no spatial interaction, i.e., a situation in which origin-destination-specific migration flows, rates, or probabilities do not depend on regions of origin; the larger the deviation the stronger the degree of spatial interaction.

## 3. SMOOTHING MIGRATION DATA

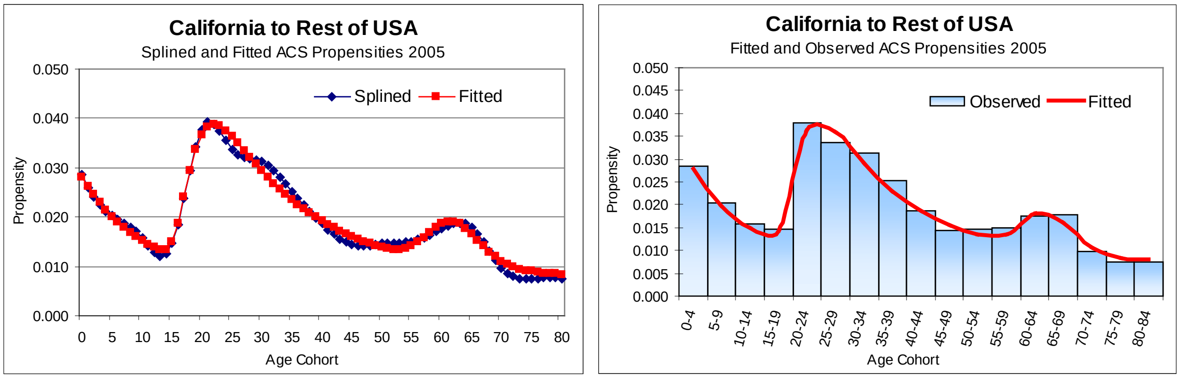
Smoothing an observed migration schedule is normally carried out with a model migration schedule, such as the one set out in Figure 1. The 5-year age groups used by the observed data now give rise to a smooth curve from which an observation can be obtained for any specific age simply by inserting the desired age, for the x variable in Equation 1 or its variants.

A more detailed illustration is offered by Figure 2. Here the observed data comes from American Community Survey and a 1-year age group (and time interval) is used. The jagged curve may be turned into a smooth profile by the following sequence of steps. First, aggregate the data into 5-year age groups, then fit a cubic sphere to the data and fit a model migration schedule to that data. A comparison of the observed data with the fitted model migration schedule shows a satisfactory goodness-of-fit.



A. Observed Data

B. Splined Data



C. Model Schedule Fit

D. Observed vs. Fitted Data

***Figure 2. Smoothing the Data: The Cubic Spline and Model Schedule Fits***

But this isn't always the case. California being the largest state in population size, received the largest sample in the American Community Survey. Wyoming, being the least populated state, received the smallest, and its fitted curve, consequently, gives an unsatisfactory result. Clearly, smoothing alone does not work; "repairing" the data is a possible option. We next explore the effectiveness of "imposing" a migration schedule obtained from a collection of larger states on the defective data for Wyoming.

## 4. IMPOSING MIGRATION DATA

After fitting model migration schedules to observed ACS outmigration patterns for all 50 states plus the District of Columbia, it became evident that the fitted schedules for the larger state populations exhibited satisfactory results. The parameters of the 26 most populated states (with over 4 million residents) were averaged and then applied, imposed that is, on the observed data for the remaining 25 schedules. The resulting revised age profiles were then scaled to give the levels of outmigration exhibited by the originally observed (irregular) migration schedules. Figure 3 shows the resulting fitted curve for Wyoming. To obtain greater accuracy, the average parameters of the largest 26 states, were divided into 3 families of model schedules: those with retirement peaks, those without but with high $a_2$ to $a_1$ ratios, and finally, those with low ratios. Further details are set out in Rogers, Jones and Ma (2008).

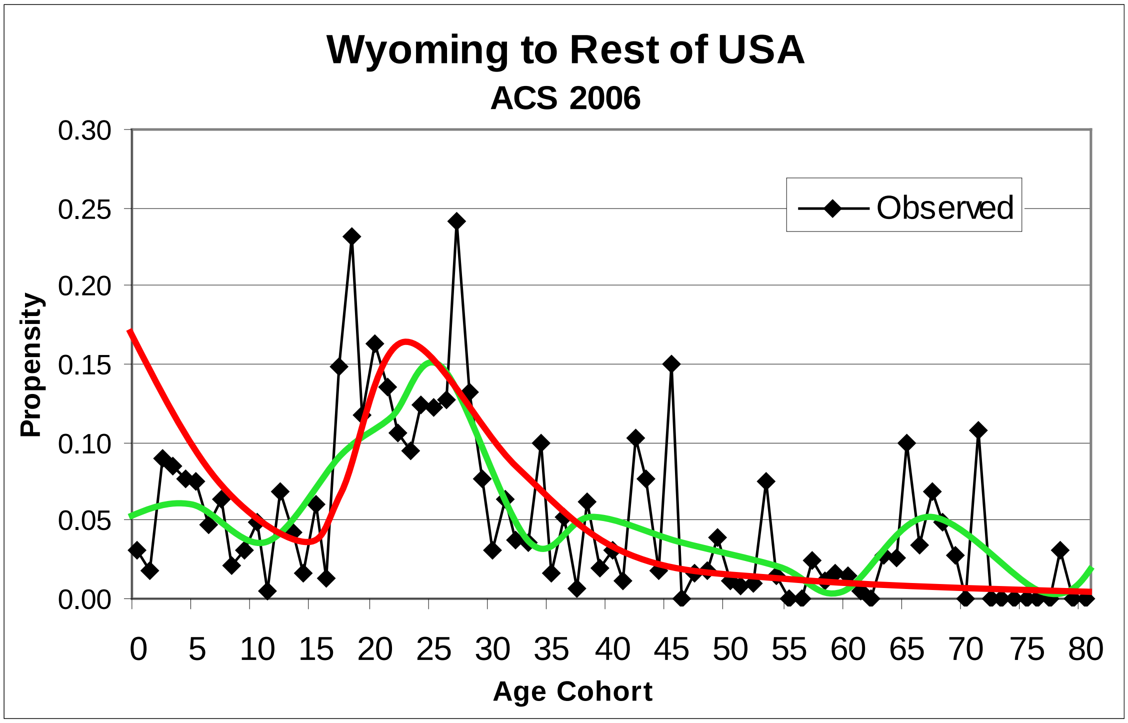


***Figure 3. Repairing the ACS Data: Wyoming 2006***

An example of imposing a spatial structure is presented in Tables 1 and 2. The focus is the 1980-1985 CPS (Current Population Survey) estimate of the four region matrix of migration flows in Table 1. How does it compare to the corresponding much larger sample census enumerations in the 1980 and 1990 censuses? Do the numbers exhibit a spatial structure that roughly matches the two defined by the censuses? Apparently not. What repairing should one carry out?

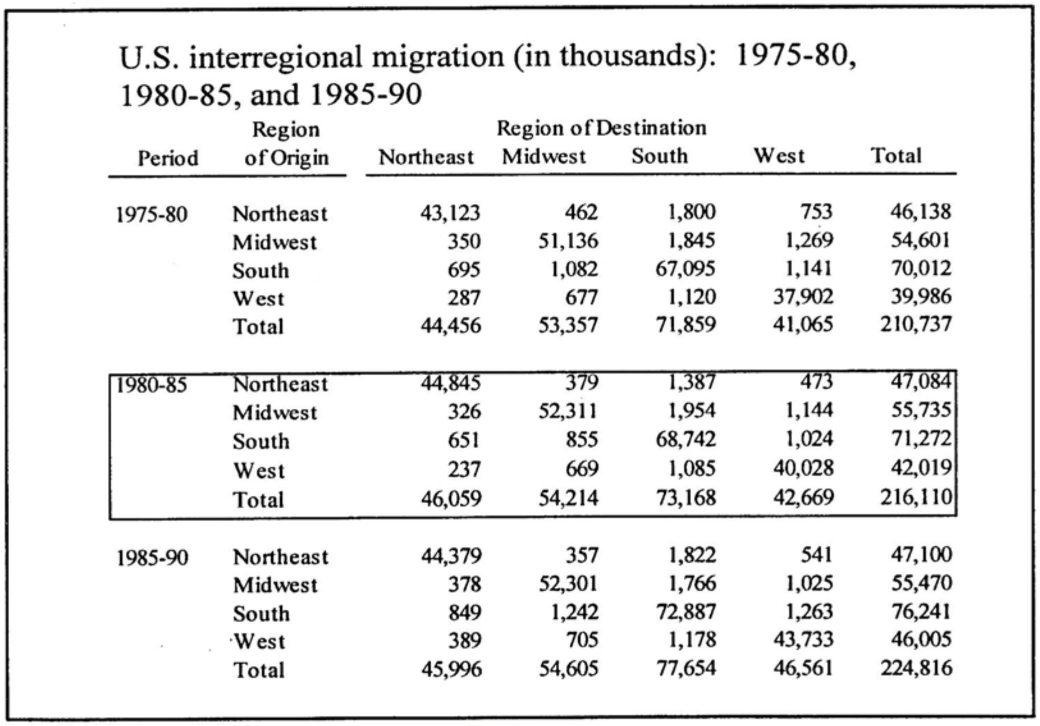
The data in Table 1 represent numbers of persons by region of residence at time of census or survey and region of residence five years prior to that census or survey. The regions in the analysis are the Northeast, Midwest, South, and West regions, as defined by the Census Bureau. The 1975-1980 and 1985-1990 migration data are based on a much larger sample size (about 1.5 million to 2.0 million persons- i.e., 5% of the U.S. decennial census enumerations) compared to the 1980-1985 migration data (with a sample size of about 50,000 households). Hence, the accuracy of the latter understandably will be viewed with some question.

The log-linear model that describes the data set out in Table 1 perfectly is the *saturated* log-linear model. If the data are incomplete, auxiliary information may be used to predict migration flows. The migration-flow table for the current period may be predicated on the basis of, for example, information regarding the *aggregate* total number of persons living in regions *i* and *j* at the beginning and end of the time interval, and the historical data on the number of origin-destination-specific migration flows.

To illustrate the method, we predict the 1980-1985 CPS migration-flow matrix based on the marginal totals of that data and the spatial structure of the 1975-1980 migration-flow matrix.

Table 2 sets out the resulting predicted migration-flow table and ratios of the predicted-to-observed migration-flow tables for the period between 1980 and 1985. Note that the interaction parameters of the saturated log-linear model are equal to those set out in Table 1 for the period between 1975 and 1980. For example, the number of migrants from the Northeast to the South between 1980 and 1985 predicted by the model is 1,614.

***Table1. U.S. Interregional Migration (in Thousands): 1975-80, 1980-85, and 1985-90***



U.S. interregional migration (in thousands): 1975-80, 1980-85, and 1985-90

| Period | Region of Origin | Region of Destination<br>Northeast | Midwest | South | West | Total |
|---|---|---|---|---|---|---|
| 1975-80 | Northeast | 43,123 | 462 | 1,800 | 753 | 46,138 |
| | Midwest | 350 | 51,136 | 1,845 | 1,269 | 54,601 |
| | South | 695 | 1,082 | 67,095 | 1,141 | 70,012 |
| | West | 287 | 677 | 1,120 | 37,902 | 39,986 |
| | Total | 44,456 | 53,357 | 71,859 | 41,065 | 210,737 |
| 1980-85 | Northeast | 44,845 | 379 | 1,387 | 473 | 47,084 |
| | Midwest | 326 | 52,311 | 1,954 | 1,144 | 55,735 |
| | South | 651 | 855 | 68,742 | 1,024 | 71,272 |
| | West | 237 | 669 | 1,085 | 40,028 | 42,019 |
| | Total | 46,059 | 54,214 | 73,168 | 42,669 | 216,110 |
| 1985-90 | Northeast | 44,379 | 357 | 1,822 | 541 | 47,100 |
| | Midwest | 378 | 52,301 | 1,766 | 1,025 | 55,470 |
| | South | 849 | 1,242 | 72,887 | 1,263 | 76,241 |
| | West | 389 | 705 | 1,178 | 43,733 | 46,005 |
| | Total | 45,996 | 54,605 | 77,654 | 46,561 | 224,816 |

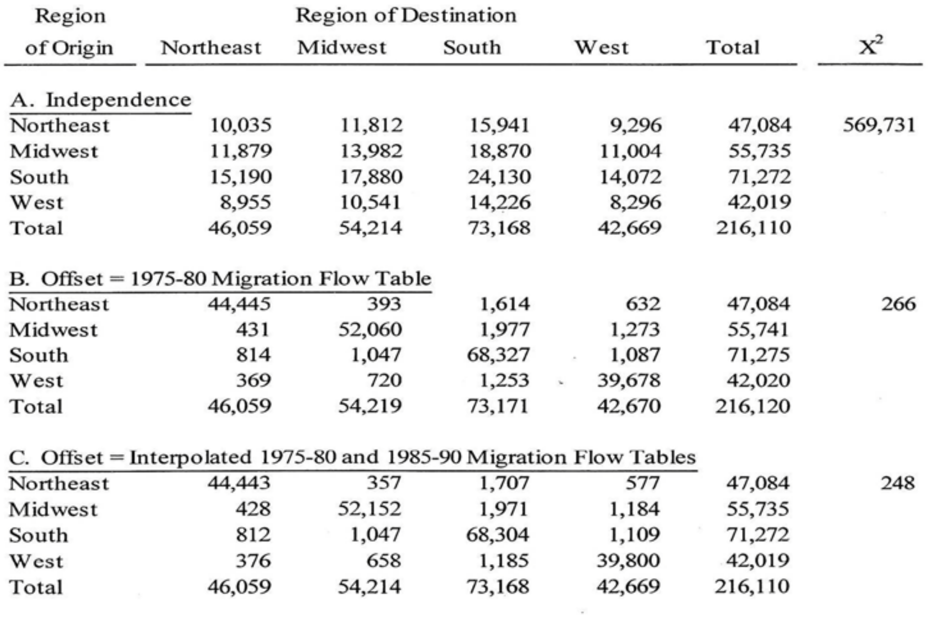
The main effects log-linear model with offset that was used to impose the spatial structure of the 1975-1980 flows reported by the 1980 census, as well as the imposed structure of the average of the two census enumerations of the flows, is set out below:

$$\hat{m}_{ijx} = m^*_{ijx}(T)(O_i)(D_j)(A_x) \tag{2}$$

where $m^*_{ijx}$ is a historical table of age-specific interregional migration. Essentially what is achieved here is an imposition of the spatial interaction pattern of the census data scaled to sum of the marginal column and row totals of the observed CPS data. The same result can be obtained with the well-known Iterative Proportional Fitting (IPF) method (Willekens, 1980). The difference is that the log-linear model is a statistical model and therefore offers richer possibilities for analysis. It not only predicts flow numbers, but it also offers estimates of parameter values with standard errors; IPF methods only provide the predicted flows.

Table 2 presents the "repaired" flow data. The goodness-of-fit of each of the three alternatives is measured by the $X^2$ reported in the final column of the table of numbers. Unsurprisingly, an assumption of independence (i.e., no spatial interaction) gives a very poor result. The 1975-1980 spatial interaction structure gives a reasonable fit, one that is improved when that spatial structure is averaged with the one following a decade later.

***Table 2. Estimated U.S. Migration Flows for the 1980-85 Period***



| Region |  | Region of Destination |  |  |  |  |
|---|---|---|---|---|---|---|
| of Origin | Northeast | Midwest | South | West | Total | $X^2$ |
| A. Independence |  |  |  |  |  |  |
| Northeast | 10,035 | 11,812 | 15,941 | 9,296 | 47,084 | 569,731 |
| Midwest | 11,879 | 13,982 | 18,870 | 11,004 | 55,735 |  |
| South | 15,190 | 17,880 | 24,130 | 14,072 | 71,272 |  |
| West | 8,955 | 10,541 | 14,226 | 8,296 | 42,019 |  |
| Total | 46,059 | 54,214 | 73,168 | 42,669 | 216,110 |  |
| B. Offset = 1975-80 Migration Flow Table |  |  |  |  |  |  |
| Northeast | 44,445 | 393 | 1,614 | 632 | 47,084 | 266 |
| Midwest | 431 | 52,060 | 1,977 | 1,273 | 55,741 |  |
| South | 814 | 1,047 | 68,327 | 1,087 | 71,275 |  |
| West | 369 | 720 | 1,253 | 39,678 | 42,020 |  |
| Total | 46,059 | 54,219 | 73,171 | 42,670 | 216,120 |  |
| C. Offset = Interpolated 1975-80 and 1985-90 Migration Flow Tables |  |  |  |  |  |  |
| Northeast | 44,443 | 357 | 1,707 | 577 | 47,084 | 248 |
| Midwest | 428 | 52,152 | 1,971 | 1,184 | 55,735 |  |
| South | 812 | 1,047 | 68,304 | 1,109 | 71,272 |  |
| West | 376 | 658 | 1,185 | 39,800 | 42,019 |  |
| Total | 46,059 | 54,214 | 73,168 | 42,669 | 216,110 |  |

Finally, Figures 4 and 5 show the originally observed and the corresponding repaired age patterns of migration. Note that in addition to extending the age range past 60, the repair also gives rise to more "reasonable" looking age profiles. Further details appear in Rogers, Willekens and Raymer (2003).

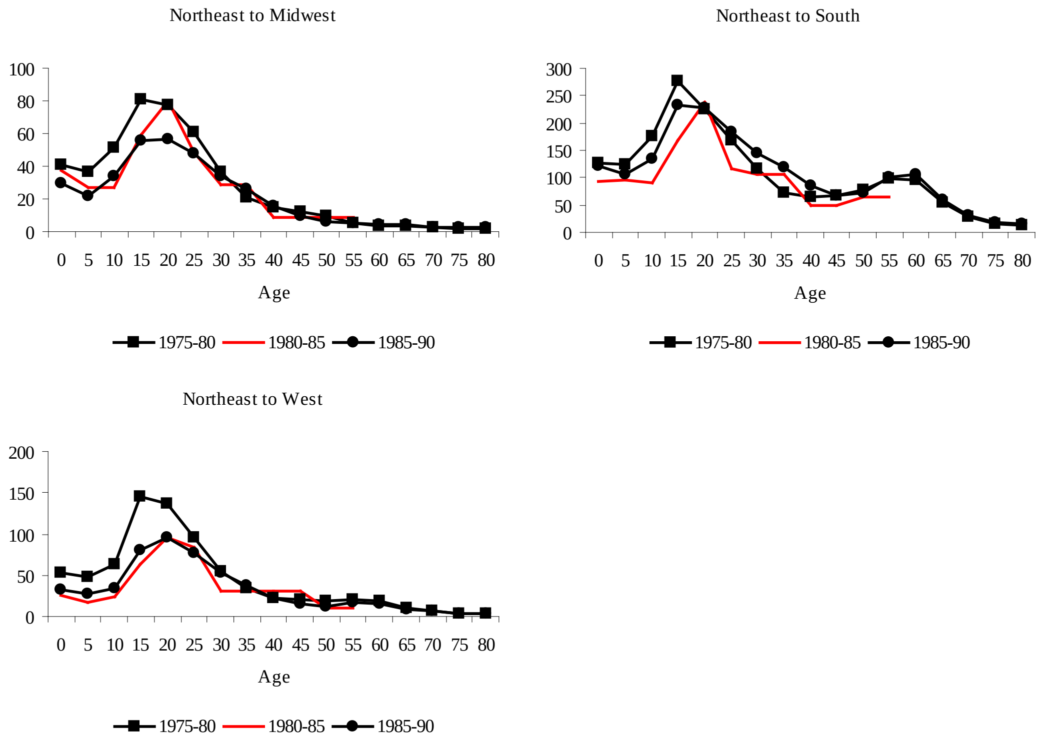


***Figure 4. Observed Age Patterns of Migration from the Northeast: U.S. Census (1980 and 1990) and Current Population Survey (1985) Data***

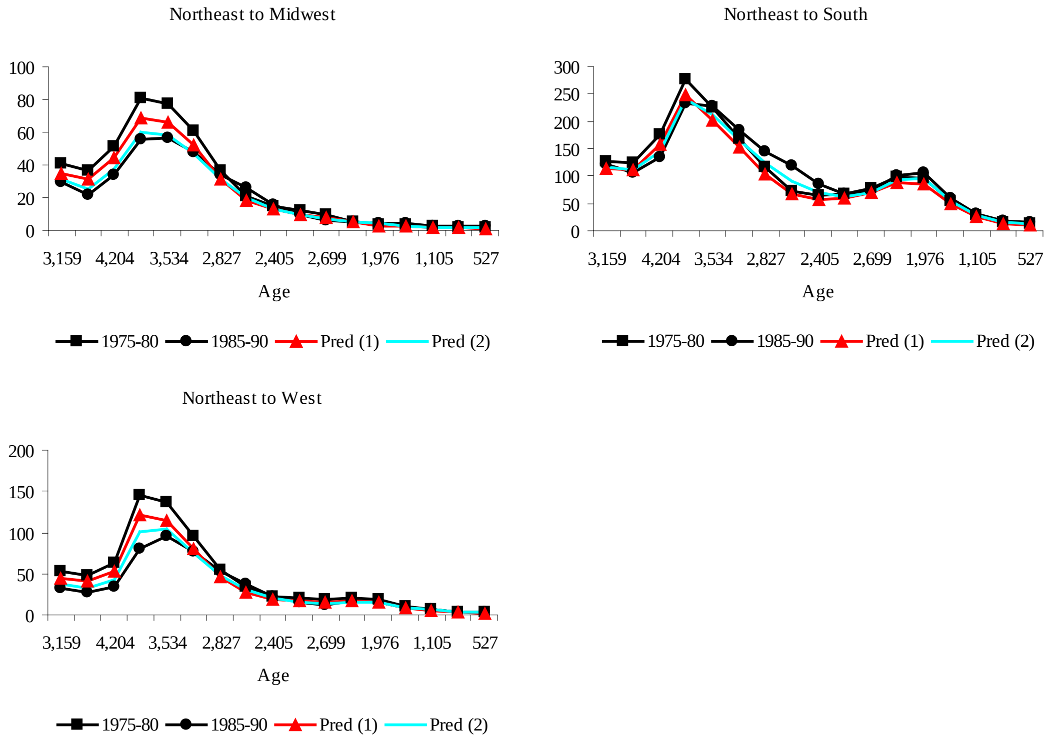


***Figure 5. Age Patterns of Migration from the Northeast: U.S. Census (1980 and 1990) and Repaired Current Population Survey (1985) Data***

## 5. INFERRING MISSING MIGRATION DATA USING INFANT MIGRATION PROPENSITIES

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).

We have carried out exploratory efforts to adopt a similar perspective for estimating migration probabilities from data on infant 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 5-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. Regression equations and model migration schedules can be used to expand these child-migration levels and spatial patterns into age-specific levels and patterns. ( Rogers and Jones, 2008)

To illustrate the method, consider the data presented in Figure 6 below, which shows a plot of the aggregate conditional survivorship proportion, $\bar{S}_{ij}(+)$, against the corresponding first age-group component of that aggregate proportion, $\bar{S}_{ij}(-5)$. 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 $\bar{S}_{ij}(x)$, 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, 5 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 $\bar{S}_{ij}(-5)$ measure is defined on pages 98-99 of Rogers, 1995.)

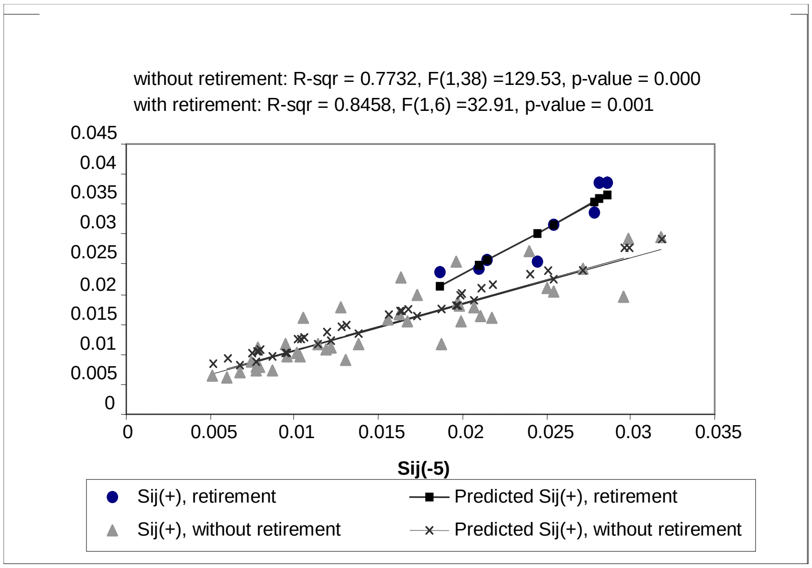


***Figure 6. Aggregate Conditional Survivorships as a Function of Infant Survivorships for 1960-1990: Observations with and without Retirement Peaks***

Returning to the scatter plot in Figure 6, notice that a straight line offers a good approximation of the relationship between the infant migration level ( $\bar{S}_{ij}(-5)$ ) between regions *i* and *j* and the corresponding level across all ages. Note that separate regression equations needed to be estimated for migration schedules with and without a retirement peak; similar strong linear associations also obtain for each age group. The latter observation is reflected by the illustration in Figure 7 of the typical age profiles that are produced by the regression equations, in this instance, of the predicted migration streams from the West to the Midwest.

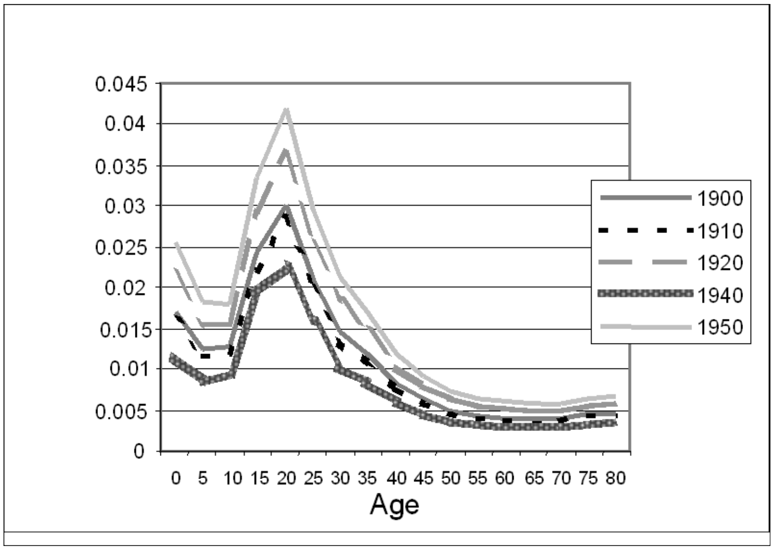


***Figure 7. Predicted from W to MW, 1900-1950***

Inferring a schedule of age-specific migration propensities 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 all indirect estimation methods in demography. Just as there are no photographs or drawings of unobserved dinosaurs, so too there are no data on unobserved migration flows.

## 6. CONCLUSION

The indirect estimation of the levels and age patterns of fertility and mortality has a long history in demography. A dominant strategy there has been to combine empirical regularities with other information to fill in the missing data. The indirect estimation of migration is of a more recent date, in part because the problem is more complicated. The *age* patterns of migrants depend on the direction of migration. To be acceptable, therefore, a method must somehow integrate the age structure with the spatial structure.

The age-structure of a population is a fundamental concept in demography, one that normally is depicted as an age pyramid. The age structure of migration has also become a fundamental concept, one that is describable with a model migration schedule. However, the spatial structure of a multiregional system of directional migration flows is a notion that lacks a widely accepted definition. We adopt the definition put forward by Rogers, et al. (2002a), which draws on a log-linear specification of a spatial interaction model. Such a formulation allows one to capture different features of a particular spatial structure, with one set of parameters representing the effects of sizes of origin populations, another set representing the corresponding effects of the sizes of destination populations, and still another set representing the strengths of the linkages between the two populations. Using these models of structure, we develop methods for describing, smoothing, imposing, and inferring migration flows in situations where the observed data are irregular, inadequate or totally unavailable.

In summary, the following observations need to be made. First, the model migration schedules and loglinear multiplicative component models offer a flexible and powerful framework for describing and analyzing observed migration flows. Second, these models offer the analyst a useful vehicle for smoothing observed data, as well as for imposing structures borrowed from other data onto incomplete or irregular migration data sets. Finally, when no observed migration data are available, these models suggest techniques of “indirect” estimation that in most instances yield reasonable values for age-specific migration propensities.

Future work on the general topic of indirect estimation of migration flows should explore extensions of the models we have used, for example, by introducing covariates into the analysis. Alternatively, other formulations of the estimation methods should be tested further. For example, we have examined methods that decompose net migration schedules into in- and outmigration schedules (Rogers and Liu, 2005), methods that use historical regularities exhibited by ratios of secondary to primary migration levels (Rogers and Raymer, 2005), and methods that predict the age compositions of outmigrants from data on the age compositions of origin populations generating these outmigrants (Little and Rogers, 2007). Finally, the age patterns of international migration generally mirror those of internal migration. Thus it is likely that several of the methods of indirect estimation outlined here and in the above referenced publications may be applicable for estimating those flows as well.

## REFERENCES

Bates, J. and I. Bracken (1987). "Migration Age Profiles for Local Authority Areas in England, 1971-1981," *Environment and Planning A*, 19:521-535.

Bates, J. and I. Bracken (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):469-484.

Fotheringham, A.S. and P.A. Williams (1983). "Further Discussion on the Poisson Interaction Model," *Geographical Analysis*, 15(4):343-347.

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

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.

Little, J. and A. Rogers (2007). "What Can the Age Composition of a Population Tell Us About the Age Composition of Its Out-Migrants?" *Population, Space and Place*, 13(1):23-39.

MacFadden, D. (1978). "Modeling the Choice of Residential Location," in *Spatial Interaction Theory and Planning Models*, A. Karlqvist, L. Lundqvist, F. Snickars, and J. Weibull, eds. (Pages 75-96).

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," unpublished Ph.D. dissertation, 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.

Raymer, J. and A. Rogers (2007). "Using Age and Spatial Flow Structures in the Indirect Estimation of Migration Streams," *Demography*, 44(2):199-223.

Raymer, J. and A. Rogers (2008). Applying model migration schedules to represent age-specific migration flows. In J. Raymer and F. Willekens (Eds.), *International migration in Europe: Data, models and* estimates (pp175-192) Chicchester: Wiley

Rogers, A., ed. (1999). *The Indirect Estimation of Migration*, special issue of *Mathematical Population Studies*, 7(3):181-309.

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

Rogers A. (1975). *Introduction to Multiregional Mathematical Demography*. New York: John Wiley.

Rogers, A. (1968). *Matrix Analysis of Interregional Population Growth and Distribution*. Berkeley: University of California Press.

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

Rogers, A., B. Jones and Wanran Ma (2008). "Repairing the Migration Data Reported by the American Community Survey," *Working Paper: POP2008-1*, Population Program, Institute of Behavioral Science, University of Colorado, Boulder, Colorado.

Rogers, A. and B. Jones (2008). Inferring directional migration propensities from the migration propensities of infants in the United States. *Mathematical Population Studies, 15(3),* 182-211.

Rogers, A. and J. Little (1994). "Parameterizing Age Patterns of Demographic Rates with the Multiexponential Model Schedule," *Mathematical Population Studies*, 4(3):175-194.

Rogers, A. and J. Liu (2005). "Estimating Directional Migration Flows from Age-Specific Net Migration Data," *Review of Urban and Regional Development Studies*, 17(3):177-196.

Rogers, A. and B. von Rabenau (1971). "Estimation of Interregional Migration Streams from Place-of-Birth-by-Residence Data," *Demography*, 8(2):185-194.

Rogers, A. and J. Raymer (2005). "Origin Dependence, Secondary Migration, and the Indirect Estimation of Migration Flows from Population Stocks," *Journal of Population Research*, 22(1):1-19.

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, and J. Raymer (2001). "Modeling Interregional Migration Flows: Continuity and Change," *Mathematical Population Studies*, 9:231-263.

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

Rogers, A., F. Willekens, and J. Raymer (2002b). "Capturing the Age and Spatial Structures of Migration," *Environment and Planning A*, 34:341-359.

Rogers, A., F. Willekens, and J. Raymer (2003). "Imposing Age and Spatial Structures on Inadequate Migration Flow Data Sets," *Professional Geographer*, 55(1):57-70

Sen, A. and T. Smith (1995). *Gravity Models of Spatial Interaction Behavior*. Berlin: Springer-Verlag.

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 (1967). *Manual IV: Methods of Estimating Basic Demographic Measures from Incomplete Data*. New York: Department of Economics and Social Affairs, United Nations.

Willekens, F. (1983). "Log-linear Modelling of Spatial Interaction," *Papers of the Regional Science Association*, 52:187-205.

Willekens, F. (1982). "Multidimensional Population Analysis with Incomplete Data," in *Multidimensional Mathematical Demography*, K. Land and A. Rogers, eds. (pp. 43-111). New York: Academic Press.

Willekens, F. (1980). “Entropy, Multiproportional Adjustment and the Analysis of Contingency Tables,” *Systemi Urbani*, 2:171-201.

Wilson, A.G. (1970). *Entropy in Urban and Regional Modeling*. London: Pion.