Estimating Multistate Transition Hazards From Last-Move Data

Carl P. SCHMERTMANN

Following United Nations recommendations, many countries collect or publish internal migration data in last-move form, despite continuing uncertainty among researchers about how to estimate transition rates and probabilities from such information. "Last-move" data are a form of retrospective event history in which the only available information for each observational unit are the state at the time of a survey (ω), the last previous state (ψ), and the time at which the $\psi \rightarrow \omega$ transition occurred. The statistical literature has addressed special cases, but there is still no general method for estimating transition hazards from last-move data. In this article I propose such a method, analyze its performance in a Monte Carlo simulation study, and apply it to migration data from Brazil's 1980 census.

KEY WORDS: Backward recurrence times; Censored data; Migration; Multistate demography; Open intervals; Survival analysis.

1. INTRODUCTION

1.1 Internal Migration Data

Information on migration within the United States comes primarily from the decennial census, which records each individual's place of residence on the census date and place of residence 5 years earlier. Cross-tabulating responses to these two questions provides a partial picture of internal migration flows for the second half of each decade.

This type of data, which summarizes possibly complex event histories by noting individuals' states at the start and end of a fixed time period, may be called "N-year-ago" data. Virtually all statistical methods for migration analysis and migration accounting systems require N-year-ago data as input (Courgeau 1988; Rees 1985; Rogers 1975).

Despite their prevalence, N-year-ago migration data have well-known deficiencies. Most notably, they leave gaps in the data record if surveys are taken more than N years apart, they record at most one transition per respondent over the period of interest, and they may obscure the actual origins and destinations of individual moves. (As a combined example of the latter two problems, note that an N-year-ago data set would report a single "A to C" transition for an individual who made both an A $\rightarrow$ B move and a B $\rightarrow$ C move over N years.)

Partly in response to these deficiencies, many countries use an alternative method of reporting internal migration data, the "last-move" method. Last-move data report the origins, destinations, and times of the last move made by individuals in the population (for example, respondent A last moved 1 year ago, from California to Oregon; respondent B last moved 8 years ago, from Idaho to Montana; and so on). They do not record individuals' locations at some common, fixed point in the past. Some countries for which migration data come from national population registries also publish migration information in last-move form.

In a comprehensive survey on national migration statistics (now rather old, but unfortunately the last of its kind), the United Nations (UN) found that 71 of 121 reporting countries collected last-move data; for 19 of these countries, last-move data were the only statistics available on internal migration (United Nations 1978, Annex III). The same UN survey reported that 40 national censuses included last-move questions, with 13 (including those of India, Indonesia, Bangladesh, and Brazil) asking last-move questions exclusively. UN recommendations for the 1980s round of national censuses (United Nations 1980) approved last-move questions as an acceptable alternative to N-year-ago questions. In a contemporary example, the Demographic and Health Surveys (DHS), a set of more than 80 international population surveys, currently among the most widely-used data sets in demography, collect and report migration data exclusively in last-move form.

In practice, last-move data have proven extremely awkward. Unlike the N-year-ago questions, they do not map neatly into the discrete-time transition probability matrices of multistate demographic accounting systems (Rogers 1975). Because they record only a subset of the state transitions and periods of exposure that occur over a period of interest, they cannot be used for calculating event/exposure ratios in continuous-time hazard models.

Estimating transition hazards and probabilities from last-move data is more than a good statistical puzzle—it is also a vexing, unsolved problem for demographers who study migration. Researchers remain uncertain about how to use last-move data properly.

1.2 Statistical Background

Last-move data are not unique to migration. Many surveys, censuses, and demographic registration systems collect or publish data on the most recent event experienced by a respondent. For example, health surveys in poor countries frequently include questions on women's current parity and time since last birth, and there is much literature on how to construct meaningful fertility indices from such data (see, e.g., Feeney and Ross 1984; Sheps, Menken, Ridley, and Lingner 1970; Srinivasan 1968). Marketing questionnaires may ask when the respondent last purchased a specified

Carl P. Schmertmann is Associate Professor, Department of Economics and Center for the Study of Population, Florida State University, Tallahassee, FL 32306 (E-mail: *schmertmann@fsu.edu*). The author thanks Griffith Feeney for several valuable insights about the issues addressed in this article, and Andrei Rogers and Phil Rees for their helpful comments.

item or brand (Wheat and Morrison 1994). An epidemiological survey may sample those with a particular disease and ask how long they have been infected (Keiding 1991).

From the standpoint of survival models, last-event data are somewhat unusual. Each observed spell begins with an event and ends when interrupted by the survey. By construction, no events occur during the sampled spells. Despite this peculiar censoring process, several researchers (Allison 1985; Baydar and White 1988; Sorensen 1977) have demonstrated that standard approaches for estimating hazard rates need only slight modification to work with last-event data in many types of survival models. Allison (1985) and Hamerle (1991) have provided excellent overviews of relevant statistical issues.

The problem of estimating hazard rates from last-event data is not completely solved, however. In particular, there currently is no general technique for estimating rates from last-event data in multistate survival models, which are the primary mathematical tool for demographers studying migration.

Some migration researchers (e.g., Courgeau 1988, p. 283) have concluded that the statistical difficulties posed by last-move migration data are insuperable, or that in any case they outweigh potential advantages of such data. My aim in this article is to demonstrate that last-move data can in fact be used to consistently estimate transition hazards in a multistate Markov model. Instantaneous rate estimates may then be converted into discrete-time transition probabilities for use in multistate accounting systems. I propose a fairly simple approach based on modeling the "visibility" of transitions and backprojecting the survey population to estimate exposure, study the new estimator's properties using Monte Carlo simulation, and construct an example from 1980 Brazilian census data as an illustration.

2. NOTATION

Let us first establish a system of notation for last-move data. A survey is taken at time T. The period of interest consists of the T years prior to the survey, from time 0 to T. At each moment during $[0, T]$, each of the N individuals surveyed was in exactly one of R possible states. The survey collects information on the *last* change of state, if any, made by each individual during $[0, T]$. The survey records

- ψ, the origin state of the last move prior to the survey,
- ω, the destination state of the last move (this is also the individual's state at time T),
  and
- u, the elapsed time between the last move and the survey date T.

The elapsed time u is the backward recurrence time discussed by Allison (1985). Individuals with no moves during $[0, T]$ are left-censored; by convention, they are assigned $u = T$ and $\psi = \omega$. (A separate dummy variable to indicate censoring is unnecessary; any observation with $\psi = \omega$ must be censored.) Like the migration data discussed earlier, this survey does *not* record an individual's state at time $t = 0$, denoted by α. This initial state α is known for individuals who made no moves (if $\psi = \omega$, then $\alpha = \omega$), but not for others. If an individual made only one move during $[0, T]$, then $\alpha = \psi$; however, because the number of moves is not recorded, the researcher observes α with certainty only for the censored observations.

For each pair of states (i, j), call V_{ij} the total number of observations for which $\psi = i$ and $\omega = j$. V stands for "visible" moves. Of course, many "invisible" moves may have occurred during $[0, T]$ that are not recorded in the survey, because they were not last moves. For $i \neq j$, let $V_{ij}(u)du$ denote the number of visible $i \to j$ moves with elapsed times in the interval $[u, u + du]$, let $H_{ij}(u)$ denote the probability distribution function of backward recurrence times among those visible $i \to j$ moves, and let $h_{ij}(u)$ denote the corresponding probability density functions. $\mathbf{\Lambda}$ denotes the entire set of last-move data $\{\psi_k, \omega_k, u_k\}_{k=1\ldots N}$.

3. A SIMPLE MODEL

Assume that during $[0, T]$ the column vector of state-specific populations $\mathbf{N}(t) = [N_1(t) \ldots N_R(t)]'$ evolves according to a continuous-time, first-order Markov process with a constant $R \times R$ hazard matrix $\boldsymbol{\mu}$. For a given initial population $\mathbf{N}(0)$,

$$E[\mathbf{N}(t)] = \mathbf{P}(t)\mathbf{N}(0) \tag{1}$$

and

$$\mathbf{P}(0) = \mathbf{I}; \qquad \frac{d}{dt}[\mathbf{P}(t)] = \boldsymbol{\mu}\mathbf{P}(t), \tag{2}$$

where the element in the ith *column* and jth *row* of $\mathbf{P}(t)$, $P_{ij}(t)$, represents the probability that an individual in state i at time 0 will be in state j at time t. The corresponding element of the hazard matrix $\boldsymbol{\mu}$ represents the instantaneous hazard of an $i \to j$ movement. Following the standard in multistate demography (Rogers 1975), this notation reverses the usual (row, column) order of matrix subscripts to preserve the intuitive interpretation of μ_{ij} as the $i \to j$ transition hazard. Diagonal elements of $\boldsymbol{\mu}$ are defined so that each column sums to 0, $\mu_{jj} = -\sum_{k \neq j} \mu_{jk}$.

It is important to understand what the Markov model implies for the densities $V_{ij}(u)$. Moves from $i \to j$ are recorded in the survey only if the individual makes no subsequent transitions. Thus the expected density of visible $i \to j$ moves at any particular value of u is

$$V_{ij}(u) = \mu_{ij} N_i(T - u) S_j(u), \qquad i \neq j, \tag{3}$$

where $N_i(T - u)$ represents the number of individuals at risk of an $i \to j$ transition u periods before the survey and $S_j(u) = \exp(-u \cdot \mu_{jj}) = \exp(-u \sum_{k \neq j} \mu_{jk})$ represents the probability of surviving in state j for u periods without a move. The expected total number of individuals with visible $i \to j$ moves in the survey will be

$$V_{ij} = \mu_{ij} \int_0^T N_i(T - u) S_j(u)\, du, \qquad i \neq j. \tag{4}$$

The expected value of V_{ij} thus depends positively on the hazard rate for $i \to j$ moves, negatively on the hazards of moves from j to anywhere else, and positively on the

(changing and unobserved) population of state i during the $[0, T]$ period.

4. ESTIMATION PROBLEM

The problem is to estimate the transition hazards $\boldsymbol{\mu}$ from the last-move data $\boldsymbol{\Lambda}$. Standard methods based on tabulations of α and ω (see, e.g., Courgeau 1988; Rees 1986; Rogers 1975; Singer and Spilerman 1976) will not work, because they require knowledge of the initial state α for each observation. The same problem applies to maximum likelihood approaches that use the conditional likelihood functions $L(\psi, \omega, u|\alpha)$. Event/exposure ratios will not work because, curiously, at the individual level the data contain only events with no corresponding exposure ($\psi \to \omega$ moves, with no information about time spent in ψ prior to the move) and exposure with no corresponding events (the time spent during $(T-u, T]$, during which there must, be construction, be no moves away from ω).

Stated plainly, it is difficult to estimate transition rates from last-move data, because events and relevant periods of exposure may have been rendered "invisible" by moves later in the period. This problem is evident in (3) and (4), where the populations at risk, $N_i(T-u)$, are unknown. In general, $N_i(T-u)$ depends on every element of $\boldsymbol{\mu}$, via (1), and on the unknown initial population $\mathbf{N}(0)$. The researcher can observe only the end of period values $N_j(T) = \Sigma_i(V_{ij})$; any earlier populations must estimated in some manner.

One way around the problem of unknown exposure is to assume that the period length T is very large relative to the hazard rates. For example, Allison (1985, pp. 320–321) demonstrated that for a special case ($R = 2, T \to \infty$), one can consistently estimate $\boldsymbol{\mu}$ from $\boldsymbol{\Lambda}$. Specifically, he showed that μ_{12} can be estimated as the inverse of the sample mean of u for those who last made a $2 \to 1$ move, and μ_{21} can be estimated as the inverse of the sample mean of u for those who last made a $1 \to 2$ move.

Allison's results highlight the manner in which backward recurrence time distributions $h_{ij}(u)$ change as $T \to \infty$. As T increases, the population $\mathbf{N}(T)$ converges to a constant vector $\mathbf{N}^*$ that is proportional to the eigenvector associated with the zero eigenvalue of $\boldsymbol{\mu}$ (simplifying the distribution of individuals across states), and the proportion of the population experiencing at least one event in $[0, T]$ rises to unity (simplifying the distribution of spell lengths u). As $T \to \infty$, (3) simplifies to

$$V_{ij}(u) = [\mu_{ij} N_i^*] S_j(u) = [\mu_{ij} N_i^*] \exp(u \cdot \mu_{jj}). \qquad (5)$$

Because the terms in square brackets do not vary with u, (4) similarly converges to

$$V_{ij} = -\mu_{ij} N_i^* / \mu_{jj}, \qquad i \neq j \qquad (6)$$

and, among individuals with $i \to j$ last moves, the distribution of u simplifies to the exponential,

$$h_{ij}(u) = -\mu_{jj} \exp(u \cdot \mu_{jj}), \qquad u \geq 0, \qquad (7)$$

with mean

$$E_{ij}(u) = -\mu_{jj}^{-1} \quad \forall i. \qquad (8)$$

Thus, regardless of the number of states R, as $T \to \infty$, one can use simple method-of-moments estimators for the diagonal elements of the hazard matrix. Specifically,

$$\hat{\mu}_{jj} = -\bar{u}_j^{-1}, \qquad (9)$$

where $\bar{u}_j$ is the sample mean of u for all individuals in state j at time T. In Allison's example ($R = 2$), the diagonals identify $\boldsymbol{\mu}$ completely, but for $R > 2$, additional information is needed.

5. ESTIMATION STRATEGIES

In this section I discuss three alternative methods for estimating hazards from last-move data and briefly compare their performance in a Monte Carlo simulation. The first method is a straightforward generalization of Allison's (1985) large-T estimator, the second method is a simple ad hoc procedure suggested in the demographic literature on migration, and the third method—based on modeling the visibility of moves and backprojecting from the end-of-period population to estimate exposure—is new.

5.1 Asymptotic Estimator (A)

The asymptotic approach suggested by Allison (1985) in his analysis of backward recurrence times can be generalized using some of the results from the previous section. Specifically, by combining (6) and (9), one can derive estimators for the off-diagonal elements of the hazard matrix that are consistent in the limit as $T \to \infty$:

$$\hat{\mu}_{ij} = V_{ij} / [\bar{u}_j \cdot N_i(T)], \qquad i \neq j. \qquad (10)$$

This A estimator may work well in two rather different situations; when T is very large relative to the rates in $\boldsymbol{\mu}$ or (more surprisingly) when T is very small.

For large T, the asymptotic assumptions will be approximately satisfied. Individuals are likely to have made many transitions over $[0, T]$, the interstate distribution will be nearly unchanging, and the A estimator will converge to the correct parameter value.

Counterintuitively, the A estimator should also work well for small T values. Few moves will be rendered invisible during a short period; thus the numerator in the A estimator approximately equals the total number of $i \to j$ events. State-specific populations $N_i(t)$ will be nearly constant over a short period $[0, T]$, and $\bar{u}_j \approx T$ (because it is a weighted average of $u = T$ for the many individuals who remain in state j, and $u_k \in [0, T]$ for the few who make a last move to j). Thus the denominator in (10) will approximately equal total exposure to $i \to j$ moves, because $\int_0^T N_i(t)\, dt \approx T N_i(T) \approx \bar{u}_j N_i(T)$. Consequently, when T is small, the A estimator will (somewhat coincidentally) approximate the true events/exposure ratio for $i \to j$ transitions.

5.2 Naive Estimator (N)

A far less elegant, but possibly less dangerous, approach is to ignore the last-move nature of the data entirely. If individuals made at most one transition during $[0, T]$, then *last* moves and *all* moves are synonymous, $\alpha = \psi$ for every observation, and all histories are completely observed. These

may be reasonable assumptions when T is low (relative to rates $\boldsymbol{\mu}$) and multiple moves over $[0, T]$ are rare. Under these assumptions, the maximum likelihood estimator (MLE) is

$$\hat{\mu}_{ij} = V_{ij}/X_i, \qquad i \neq j, \tag{11}$$

where (using I_{ijk} as a dummy indicator for an $i \to j$ last move by observation k)

$$X_i = \sum_{k=1}^{N} \sum_{j=1}^{R} [I_{ijk}(T - u_k) + I_{jik}(u_k)] \tag{12}$$

measures estimated total exposure in state i.

This estimator should work well in low-rate cases with relatively small T. Many migration estimation problems in demography satisfy these conditions, and demographers have suggested this as an approximate estimator for last move data (Courgeau 1988, p. 165).

5.3 Backprojection Estimator (B)

Both of the preceding methods require strong a priori assumptions about the size of T relative to the hazard rates $\boldsymbol{\mu}$. It clearly would be be desirable to develop a procedure for estimating transition rates from last-move data that does not depend on such simplifying (and possibly wrong) assumptions.

Equation (4) can be reinterpreted for this purpose. For a given hazard matrix, the missing populations at risk, $N_i(T-u)$, can be estimated by backprojecting the end-of-period population from time T to time $T-u$. Estimated midperiod populations may then be used to calculate an estimate of the expected total number of visible $i \to j$ moves:

$$\hat{V}_{ij}(\boldsymbol{\mu}) = \mu_{ij} \int_0^T \hat{N}_i[T - u|\boldsymbol{\mu}, \mathbf{N}(T)] S_j(u|\boldsymbol{\mu})\, du, \qquad i \neq j. \tag{13}$$

The right side of (13) must be calculated numerically. Accurate backprojection can be accomplished by repeatedly applying the approximation rule (Keyfitz 1985, p. 356),

$$\mathbf{N}(t - \Delta) \approx \left[\mathbf{I} + \frac{\Delta}{2}\boldsymbol{\mu}\right]^{-1} \left[\mathbf{I} - \frac{\Delta}{2}\boldsymbol{\mu}\right] \mathbf{N}(t), \tag{14}$$

with a small time increment Δ. For a given hazard matrix $\boldsymbol{\mu}$, one can thus calculate the right side of (13) by first backprojecting the end-of-period population $\mathbf{N}(T)$ to times $T - \Delta, T - 2\Delta, \ldots, 0$, and then using a simple rectangular approximation to the integral. I used this procedure in all calculations that follow.

With this numerical procedure, it is possible to rapidly calculate the expected number of visible moves V_{ij} for any given hazard matrix. A method-of-moments estimator for the $R(R-1)$ off-diagonal elements may then be defined as the solution $\boldsymbol{\mu}$ to the nonlinear equations

$$\hat{V}_{ij}(\boldsymbol{\mu}) = V_{ij}, \qquad i \neq j, \tag{15}$$

solved by Gauss–Newton or some other appropriate technique. The Appendix contains some additional details about this procedure.

Unlike the A and N estimators, the B estimator does not rely on strong assumptions about T and $\boldsymbol{\mu}$. Consequently, it should work well under a wide variety of rates and time periods. I investigate this question next.

5.4 Monte Carlo Comparison

To compare the estimators, I generated sets of pseudorandom transition histories from a Markov model. In each of the Monte Carlo datasets, a population of $N = 5,000$ was distributed uniformly across three states at the start of a period of interest, and subsequently evolved according to a Markov process with transition rates

$$\boldsymbol{\mu} = \begin{bmatrix} -.040 & .003 & .010 \\ .010 & -.011 & .015 \\ .030 & .008 & -.025 \end{bmatrix}. \tag{16}$$

For each of several period lengths ($T = 5, 10, 50, 100, 150$), I used the following procedure:

- Generate 200 independent sets of 5,000 individual histories over $[0, T]$.
- Aggregate the complete histories into 200 last-move datasets $\{\psi_k, \omega_k, u_k\}_{k=1\ldots5,000}$.
- For each last-move dataset, estimate the six off-diagonal elements of $\boldsymbol{\mu}$ using the A, N, and B methods.
- Construct summary measures (mean estimate, mean percent error, mean absolute error) of the estimators by averaging over the 200 estimates.

The T values span a wide range. At the low end, $T = 5$, most individuals ($\sim 90\%$) will make no transitions, and most observations will be truncated. This situation favors the N estimator, which assumes small numbers of transitions over $[0, T]$. In contrast, when $T = 100$ or 150 with these rates, the population will be very close to its equilibrium interstate distribution, better matching the assumptions of the A estimator.

The chosen hazard rates are arbitrary, but it is possible to extrapolate from these experiments by recognizing that an increase in all of the hazards, together with a equiproportional decrease in T, would leave the process essentially unchanged. For example, the results reported for $T = 50$ are virtually identical to the results one would get if rates were all five times higher ($\mu_{12} = .05, \mu_{13} = .15$, etc.) and T were equal to 10 rather than 50. Thus in addition to the literal interpretation, one can interpret results for $T = 5, 10, \ldots, 150$ as representing experiments in which the time period remains constant at $T = 5$ and the hazard rates increase thirtyfold.

Figure 1 presents a graphical summary of Monte Carlo results for the three estimation procedures. Figure 1 (a), (c), and (e), reports bias (i.e., mean estimation error across the 200 simulated data sets); Figure 1, (b), (d), and (f), report mean absolute errors. Each plot contains results for all six transition rates, and all calculations in the figure are expressed as percentages of the true rates. For example, the true value of μ_{12} is 10 per 1,000. Over the 200 Monte Carlo datasets with $T = 50$, the A estimator for μ_{12} ranged from a minimum of 13.00 to a maximum of 17.83 per 1,000. The

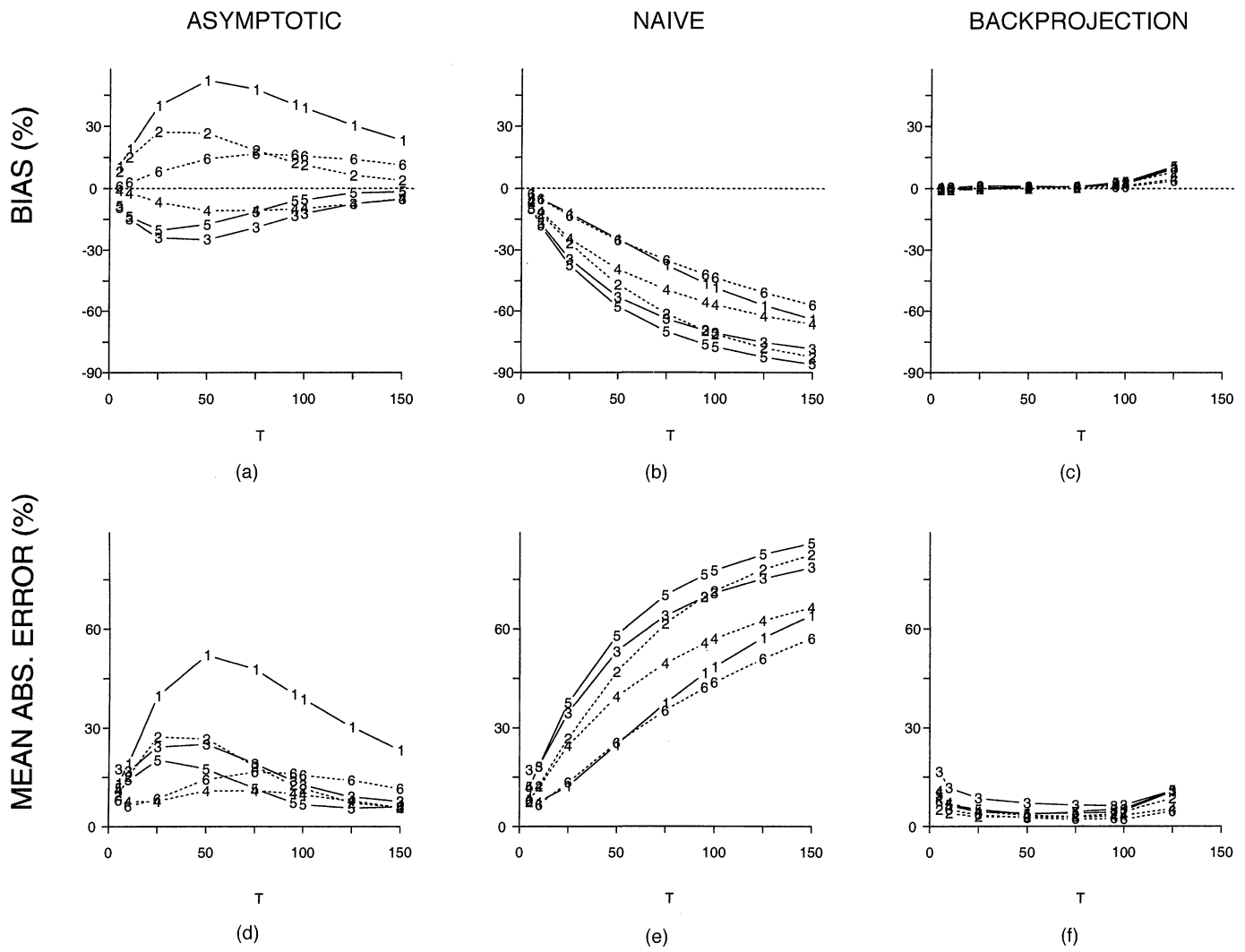

Figure 1. Monte Carlo Results for Three Alternative Estimators (Asymptotic, Naive, and Backprojection) of the Six Transition Hazards in a Three-State Markov Model. Lines in each panel correspond to the six transition rates, as follows: 1 = μ_{12}, 2 = μ_{13}, 3 = μ_{21}, 4 = μ_{23}, 5 = μ_{31}, 6 = μ_{32}. For each period length T, estimates were made from 200 independent sets of N = 5,000 individual transition histories, and the results compared to the true rates used to generate the data. (a), (c), and (e): the Monte Carlo estimates of bias (as percentages of the underlying true rates); (b), (d), and (f): the estimated mean absolute error (also as percentages). Backprojection estimators became numerically unstable at high values of T, and often failed to converge for T = 150. Backprojection estimators have very low bias relative to the other estimators, causing almost complete overlap of the six lines in 1c.

mean of the 200 estimates was 15.21, making the estimated bias in this case equal to $+52.1\%$, which is the number reported for flow 1, $T = 50$ in Figure 1(a). In this particular case, all errors were positive, so the mean absolute error was also equal to $+52.1\%$, as reported for flow 1, $T = 50$ in 1b.

Several notable findings emerge from the Monte Carlo experiments for estimating transition rates from last-move data:

- The B estimator is superior to both of the other procedures. Over a wide range of period lengths T, it has both the lowest bias and the lowest mean absolute error among the three estimators. The only exception is for $T > 125$, when the B estimator generally did not converge.
- For experiments with $T > 125$, the B estimator was numerically unstable and almost always failed to converge. As a Markov process gets sufficiently close to equilibrium (unchanging state distributions and exponentially distributed times since last move), the B method cannot distinguish between scalar multiples of $\boldsymbol{\mu}$, because all would imply the same set of V_{ij}'s. This caused the convergence problems. As $T \to \infty$, one must use information on the timing of moves, rather than simply the volumes of movement, to scale the columns of $\boldsymbol{\mu}$.
- A estimators work reasonably well when the switching process is just beginning, or when it is close to equilibrium. But it was inferior to backprojection (i.e., its bias and mean absolute error were larger) in almost all cases. In these experiments the A estimator was the preferred method only for the very longest T value, $T = 150$, the cases in which backprojection failed.
- The N estimator ignores the possibility of "invisible" moves during $[0, T]$, and consequently is biased down-

ward for all period lengths. This bias grows large as T increases. For the T values that demographers are most likely to deal with ($T \leq 10$ with rates like these), the negative bias is only moderate, and in these small samples (5,000 observations) the N estimator has mean absolute errors only slightly larger than the B estimator. Thus, although the B estimator is better, the (far simpler) N estimator is in fact reasonably good for short period lengths.

6. APPLICATION TO BRAZILIAN CENSUS DATA

6.1 Brazilian Internal Migration Data

Brazil is among the countries that collect internal migration data primarily in last-move form. Brazil does not have a system of population registration, making its decennial population censuses by far the most important source of information on internal migration (Martine 1990). Brazilian censuses prior to 1991 included only questions about previous place of residence and time of last move, with no questions about place of residence at some fixed, common date. (The 1991 census asked both types of questions.) Because researchers lack standardized methods for using last-move data to estimate migration rates, much of the data on the internal redistribution of Brazil's population during the post-World War II period remains underutilized.

The 1970s were an especially interesting period. Large-scale changes in land use and ownership altered the economies of several agricultural frontier regions. Previously these areas had grown rapidly through net in-migration, but in the 1970s they reversed course and became large net population exporters. The primary example of this migration reversal occurred in the southern state of Paraná, which went from a net in-migration of $+577,000$ during the 1960s to -1.3 million during the 1970s (Martine 1990, p. 38). Many of those leaving Paraná migrated to neighboring São Paulo, Brazil's most industrialized state. Many others moved to the western Amazon region, where the impact of migrant settlers on the local ecology has become an international concern.

Detailed study of internal migration in the 1970s requires using Brazil's 1980 Demographic Census (IBGE 1980). Having established the utility of the backprojection method, I now apply it to last-move data tabulated from the 3% and 25% public use samples of that census. The 3% public use sample omitted migration information but was used to tabulate age-specific total populations. A special tabulation of the 25% sample containing only individuals who reported a change of *município* over 1970–1980 was used for tabulating the interstate movers (V_{ij} for $i \neq j$). I thank Cláudio Machado for making the special 25% tabulation available.

The reference period for the last-move questions on the Brazilian census was $T = 10$ years. Brazil, like the United States, has a federal government and is subdivided into states. Respondents were asked about their state of residence at the time of the census, their last previous state of residence, and the number of years that they had resided in their current state (coded as < 1 year, 1 year, 2 years, ..., 5 years, 6–9 years, 10+ years).

I analyze 1970–1980 movements for a three-region system comprising Paraná, the state of São Paulo, and the rest of Brazil. The relatively rapid reversal of migration streams experienced by Paraná means that one must carefully consider the possibility that last moves, particularly moves away from Paraná, may not have been the only moves by individuals over the period of interest.

6.2 Backprojection Estimates

Appendix Table A.1 contains V_{ij} data for this system, with males and females tabulated separately, and each row corresponding to a 5-year age cohort (those 10–14, 15–19, ..., 70+ at the time of the 1980 census). To keep the example simple, I omit those younger than age 10 at the time of the census. Hazard estimation for these youngest individuals must be handled differently, because they were not at risk of migration for the entire 1970–1980 period.

Figure 2 displays B estimates of migration hazards for 1970–1980 in graphical form. Six transition hazards were estimated separately for each (sex, cohort) pair. The results are plotted in the figure, with straight lines joining the estimates for adjacent cohorts. Once the programs are written, the marginal computational cost of backprojection is low. A moderately powered desktop personal computer (c. 1997) calculated the (6 flows) $\times$ (13 ages) $\times$ (2 sexes) $= 156$ rate estimates in Figure 2 in less than 5 seconds.

The estimated rates look sensible, and they have features common to many migration-by-age schedules—for example, peak rates are at young adult ages, and female rates are slightly higher than male rates at young ages and lower afterwards. It is important to highlight that these are, in a sense, the first proper migration rate estimates from this census. To my knowledge, no previous estimates of Brazilian migration for this period have accounted carefully for the last-move nature of the census data.

Figure 3 compares the A and N estimation methods to backprojection for the Brazilian data. Rates are very low for this system (or, equivalently, $T = 10$ is a relatively short period), so the A and N estimators should have small biases, similar to those in the $T = 5$ or $T = 10$ Monte Carlo experiments. This is indeed the case; A estimators range from 14% below the corresponding B estimator to 19% above, whereas N estimators range from 11% below to equal. (See the figure caption for more details.) The biases in the A and N estimators are apparently only moderate, and these estimators may be acceptable for some applications. But it is clearly useful to have a more accurate method, both for producing improved estimates and for using as a baseline. In applications with higher transition rates than those here (e.g., studies of short-distance migration), careful consideration of last-move data becomes even more important, tipping the balance in favor of backprojection.

6.3 Additional Use of Timing Information

One potential advantage of last-move data over standard "Where did you live 5 years ago/1 year ago?" questions is that the last-move data contain richer information on the timing of movements. If, for example, migration rates for a

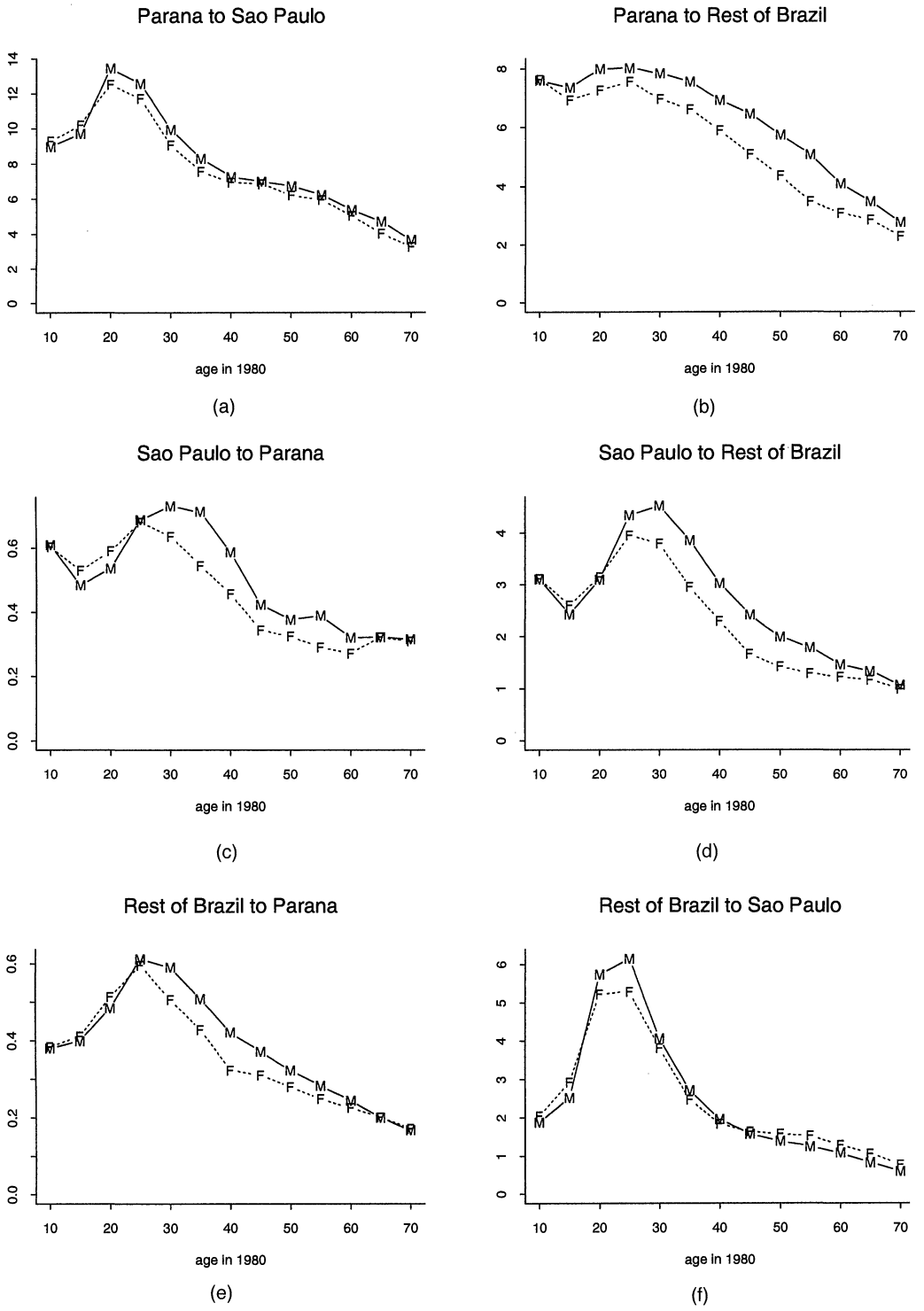

Figure 2. 1970—1980 Migration Rates Estimated From Last-Move Data in Brazil's 1980 Census, Using Backprojection. With three regions, there are six gross flows and thus six transition hazards for each of the 26 (sex, age group) cells. The six panels correspond to the six flows. Separate backprojection estimates for each (sex, age group) cell produce six rate estimates, which are then plotted as points in the corresponding panels, using F for females and M for males. Points in each panel are joined with line segments for ease of interpretation. All rates are in per 1,000 terms.

particular (origin, destination) pair had been increasing or decreasing prior to the survey, then it should be possible to observe these changes. Figure 4 illustrates this point in a very simple fashion. I estimated rates for the *second* half of the decade, 1975–1980, by reassigning individuals who moved and reported $u = 6$–9 years as truncated, nonmover observations. I then recalculated the B estimators with these new data and $T = 5$. For brevity, the actual estimates for 1975–1980 are not shown. Instead, the figure reports ratios of 1975–1980 hazard rates to 1970–1980 rates calculated earlier, disaggregated by cohort and sex. Ratios above or below unity would indicate that migration was accelerating or decelerating, respectively, over the course of the 1970s.

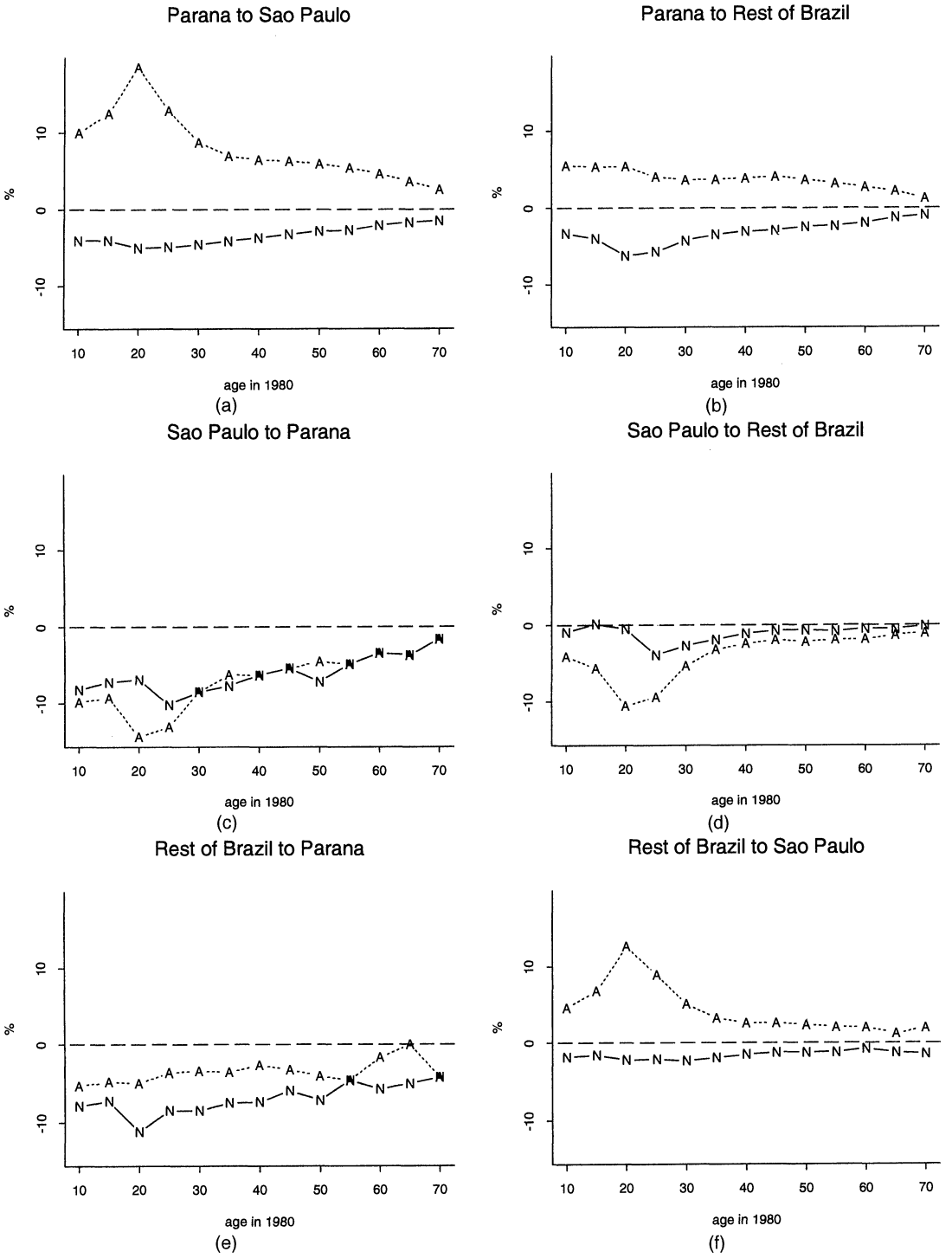

Figure 3. Asymptotic (A) and Naive (N) Estimates of Brazilian 1970—1980 Migration Flows, Relative to the Corresponding Backprojection (B) Estimates. For each age group and flow, the plot displays values of 100 · (estimate/B − 1) for both A and N estimates. Monte Carlo results suggest that B estimates are likely to be close to the true hazards, thus plotted values are likely to be close to the percentage errors in A and N estimates. Data in the figure are for males only; results for females are similar.

The top pair of lines in Figure 4 report ratios for movements from Paraná to the rest of Brazil. The 1975–1980 rates are much higher than the rates for the full decade, indicating (sensibly, given the history of Brazilian migration) a sharp increase in movements from Paraná to destinations outside of São Paulo during the late 1970s. The bottom pair of lines, for movements into Paraná from locations other than São Paulo, indicates that this flow was also accelerating, but more slowly. Of course, differences between 1970–1980 and 1975–1980 rates are due to age effects, as well as period effects. Cohort members were all 5 years older in 1975–1980 than in 1970–1975, and this would cause the ratios in Figure 4 to differ from unity even in the absence of intraperiod trends. However, for all six flows (four of which are omitted from the plot), rate changes go in the same direction for every cohort, providing strong evidence for period effects.

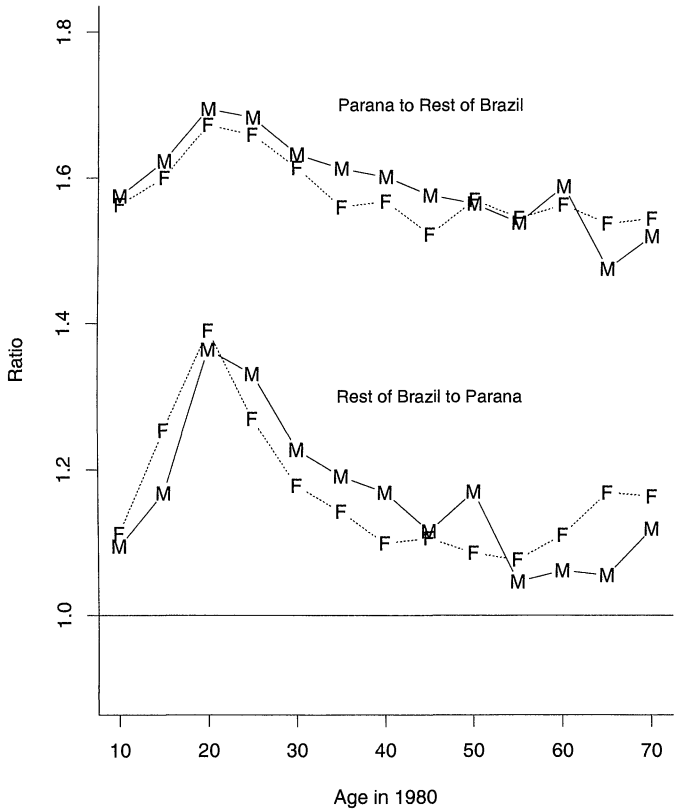

Figure 4. Ratio of 1975—1980 Rates to 1970—1980 Rates, for Migration to and From Paraná. Values above 1 imply an acceleration of migration over the decade, with higher values corresponding to more rapid acceleration. Points Marked M and F represent males and females.

The data in Figure 4 illustrate only one rather crude use of the timing information available with last-move data. More sophisticated approaches might model and fit time trends for migration rates (Doeve 1987). The interesting point is that standard period migration data, although easier for demographers to use and interpret, do not allow analysis of potentially important intraperiod trends. In contrast, last-move data provide finer detail on the timing of moves and make such analysis possible.

7. DISCUSSION

Properly estimating transition rates and probabilities from last-move data has been an ongoing problem for demographers studying human migration. Last-move data are widely collected and published, but there has been no theoretically sound approach to estimating rates from event histories in multistate models with this peculiar kind of censoring. The statistical literature has covered special cases of the problem, such as repeatable events in one-state models (Allison 1985; Baydar and White 1988; Sorensen 1977), asymptotic results for multistate processes that have reached equilibrium (Allison 1985) or multistate models with irreversible transitions and known initial state distributions (Keiding 1991). The general estimation problem has remained unsolved, however.

In this article I have proposed a general estimation method that yields consistent rate estimates even in the most general case (multiple states, no equilibrium, unknown initial distributions). The backprojection estimator works well over a very wide range of transition rates. Monte Carlo experiments demonstrate that the backprojection estimator is very accurate, even in small datasets. The experiments also demonstrate that backprojection is superior to two other available estimators, both of which are based on strong assumptions about the Markov process generating the data. The backprojection method fails only in extreme cases of very long time periods or very high transition rates—specifically, it fails to converge when the Markov process gets very close to its equilibrium state during the period under study. This problem is unlikely to cause any difficulties in real-world estimation problems.

An application of backprojection to internal migration data from the 1980 Brazilian census shows that the method produces reasonable estimates for interstate migration rates. The results also illustrate, albeit crudely, how one can use last-move data to learn about the timing of migration flows over the reference period.

The backprojection method developed in this article is appropriate for first-order Markov survival models with constant hazard rates. These models are the workhorses of applied migration research, but they are obviously limited. Future research might consider estimators for last-move data generated by processes exhibiting duration dependence in transition hazards, or by processes in which the hazards vary over calendar time. It might also be fruitful to consider estimation approaches other than method of moments. I hope that this article provides a starting point for such explorations.

APPENDIX: EXAMPLE OF RATE ESTIMATION FROM BRAZILIAN CENSUS DATA

This Appendix briefly discusses estimation for one specific case: 1970–1980 rates for males aged 20–24 years in 1980, for migration flows between Paraná (state 1), São Paulo (state 2), and rest of Brazil (state 3). The last-move data in Table A.1 indicate that in this age category there were 337,009 males who resided in Paraná over the entire 1970–1980 period ($V_{11} = 337,009$), 5,988 who last moved from São Paulo to Paraná ($V_{21} = 5,988$), and so forth. Arranging the visible moves into a matrix with elements indexed in (column, row) order, we have

$$\mathbf{V} = \begin{bmatrix} 337{,}009 & 5{,}988 & 17{,}701 \\ 51{,}314 & 1{,}073{,}885 & 229{,}494 \\ 30{,}077 & 37{,}140 & 3{,}910{,}328 \end{bmatrix},$$

and the observed population at the end of the period is equal to

$$\mathbf{N}(1980) = \begin{bmatrix} 337{,}009 + 5{,}988 + 17{,}701 \\ 51{,}314 + 1{,}073{,}885 + 229{,}494 \\ 30{,}077 + 37{,}140 + 3{,}910{,}328 \end{bmatrix} = \begin{bmatrix} 360{,}698 \\ 1{,}354{,}693 \\ 3{,}977{,}545 \end{bmatrix}.$$

The question to answer is "Given these end-of-period populations, for what 3×3 matrix of migration hazards μ is the observed $\mathbf{V}$ for 1970–1980 equal to the expected value $\hat{\mathbf{V}}(\mu)$?" The columns of any valid μ must each sum to 0, and off-diagonal elements must be nonnegative. Thus the search for μ is a search through $\mathbb{R}^6_+$. The rows of $\mathbf{V}$ must sum to $\mathbf{N}(1980)$, so we may take as independent "targets" the six off-diagonal elements of $\mathbf{V}$. This yields six nonlinear equations in six parameters,

$$\begin{aligned} V_{12} &= \hat{V}_{12}(\mu_{12}, \mu_{13}, \mu_{21}, \mu_{23}, \mu_{31}, \mu_{32}), \\ V_{13} &= \hat{V}_{13}(\mu_{12}, \mu_{13}, \mu_{21}, \mu_{23}, \mu_{31}, \mu_{32}), \\ V_{21} &= \hat{V}_{21}(\mu_{12}, \mu_{13}, \mu_{21}, \mu_{23}, \mu_{31}, \mu_{32}), \\ V_{23} &= \hat{V}_{23}(\mu_{12}, \mu_{13}, \mu_{21}, \mu_{23}, \mu_{31}, \mu_{32}), \\ V_{31} &= \hat{V}_{31}(\mu_{12}, \mu_{13}, \mu_{21}, \mu_{23}, \mu_{31}, \mu_{32}), \end{aligned}$$

and

$$V_{32} = \hat{V}_{32}(\mu_{12}, \mu_{13}, \mu_{21}, \mu_{23}, \mu_{31}, \mu_{32}),$$

which can be solved by Gauss–Newton. The only unusual feature of the estimation procedure is the calculation of expected counts of visible moves on the right sides of the equations. These calculations must be performed repeatedly during the search through μ space. Calculations are done in two steps, using (13) and (14). Specifically, for any given trial value $\tilde{\boldsymbol{\mu}}$, we first backproject the regional populations from 1980 to 1970 in increments of, say, $\Delta = .1$ years using (14):

$$\hat{\mathbf{N}}(1979.9) \approx [\mathbf{I} + .05\tilde{\boldsymbol{\mu}}]^{-1}[\mathbf{I} - .05\tilde{\boldsymbol{\mu}}]\mathbf{N}(1980),$$

$$\hat{\mathbf{N}}(1979.8) \approx [\mathbf{I} + .05\tilde{\boldsymbol{\mu}}]^{-1}[\mathbf{I} - .05\tilde{\boldsymbol{\mu}}]\hat{\mathbf{N}}(1979.9),$$

and so forth. After backprojecting, we plug the estimated state-specific populations into (13) to calculate the expected number of visible moves for each interstate flow. The integral is calculated numerically, with a simple rectangular approximation. For example, with $\Delta = .1$, the calculation for visible moves from Paraná to São Paulo is

$$\hat{V}_{12}(\tilde{\boldsymbol{\mu}}) = \tilde{\boldsymbol{\mu}}_{12} \sum_{u \in \mathbf{U}} \hat{N}_1[1980 - u|\tilde{\boldsymbol{\mu}}] \exp[u \cdot (-\tilde{\mu}_{21} - \tilde{\mu}_{23})]\Delta,$$

where $\mathbf{U} = \{.1, .2, \ldots, 10\}$.

I solved the system of equations using a standard Gauss–Newton algorithm, with numerical derivatives (see, e.g., Amemiya 1985, chap. 4). I wrote the program in Pascal and ran it on a desktop computer. Estimated rates for the aged 20–24 years male data are

$\mu_{12} = .013466$[Paraná to São Paulo],
$\mu_{13} = .007998$[Paraná to Rest of Brazil],
$\mu_{21} = .000537$[São Paulo to Paraná],
$\mu_{23} = .003096$[São Paulo to Rest of Brazil],
$\mu_{31} = .000484$[Rest of Brazil to Paraná],
$\mu_{32} = .005745$[Rest of Brazil to São Paulo].

Each of these rate estimates appears as an "M" point for 20–24-year-olds on the corresponding plot in Figure 2. Similar calculations yield six estimated rates for each of the other (sex, age) combinations. All results are then presented graphically in Figure 2.

Table A1. 1980 Brazilian Census, Last-Move Data 1970–1980

		To Paraná from			To São Paulo from			To rest of Brazil from		
Sex	Age in 1980	Paraná	São Paulo	Rest of Brazil	Paraná	São Paulo	Rest of Brazil	Paraná	São Paulo	Rest of Brazil
M	10–14	*476,965*	6,905	18,979	46,895	*1,132,182*	99,809	39,967	37,730	*5,325,589*
M	15–19	*429,886*	5,617	18,110	46,015	*1,156,263*	122,478	34,898	30,110	*4,829,669*
M	20–24	*337,009*	5,988	17,701	51,314	*1,073,885*	229,494	30,077	37,140	*3,910,328*
M	25–29	*269,695*	6,781	17,885	38,266	*942,381*	194,357	24,248	45,642	*3,093,982*
M	30–34	*226,063*	6,040	14,299	24,995	*812,679*	104,742	19,749	39,863	*2,556,728*
M	35–39	*189,321*	4,801	10,220	17,310	*672,851*	57,710	15,833	27,610	*2,120,879*
M	40–44	*179,184*	3,560	7,837	14,175	*613,078*	38,284	13,562	19,439	*1,959,746*
M	45–49	*143,852*	2,110	5,558	10,888	*507,644*	24,738	10,095	12,757	*1,572,124*
M	50–54	*121,494*	1,695	4,314	8,835	*458,537*	19,453	7,536	9,449	*1,398,944*
M	55–59	*92,341*	1,305	2,969	6,178	*341,449*	13,533	5,036	6,322	*1,094,983*
M	60–64	*70,443*	801	2,003	4,025	*252,458*	9,176	3,066	3,808	*850,091*
M	65–69	*54,976*	634	1,389	2,729	*198,967*	5,875	2,021	2,726	*716,134*
M	70+	*64,259*	782	1,510	2,457	*250,931*	5,547	1,867	2,738	*927,468*
F	10–14	*465,416*	6,723	18,958	47,387	*1,108,707*	107,614	39,223	37,221	*5,260,418*
F	15–19	*437,686*	6,198	19,353	49,222	*1,152,049*	147,551	33,418	32,487	*5,005,372*
F	20–24	*343,155*	6,533	19,513	48,341	*1,061,281*	214,856	27,750	37,171	*4,042,830*
F	25–29	*277,651*	6,668	18,633	36,520	*941,941*	177,765	23,456	41,242	*3,308,653*
F	30–34	*227,194*	5,174	12,912	22,705	*796,873*	103,134	17,437	32,610	*2,680,709*
F	35–39	*189,309*	3,694	9,102	15,627	*678,702*	55,179	13,671	21,135	*2,230,456*
F	40–44	*163,655*	2,734	6,246	12,278	*604,075*	37,264	10,463	14,486	*2,024,042*
F	45–49	*135,436*	1,759	4,844	10,014	*518,141*	26,795	7,415	8,925	*1,634,751*
F	50–54	*114,318*	1,490	3,910	7,607	*465,494*	22,931	5,337	6,832	*1,453,772*
F	55–59	*86,694*	1,034	2,667	5,477	*358,132*	16,993	3,199	4,783	*1,115,049*
F	60–64	*63,336*	747	1,928	3,382	*276,650*	11,386	2,061	3,472	*884,611*
F	65–69	*49,520*	710	1,487	2,097	*221,095*	8,005	1,486	2,651	*760,102*
F	70+	*62,555*	973	1,848	2,135	*315,858*	8,495	1,503	3,191	*1,102,661*

NOTE: Figures in italics are for nonmovers (left-censored observations).

[Received July 1997. Revised September 1998.]

REFERENCES

Allison, P. D. (1985), "Survival Analysis of Backward Recurrence Times," *Journal of the American Statistical Association*, 80, 315–322.

Amemiya, T. (1985), *Advanced Econometrics*, Cambridge, MA: Harvard University Press.

Baydar, N., and White, M. (1988), "A Method for Analyzing Backward Recurrence Time Data on Residential Mobility," in *Sociological Methodology 1988*, ed. C. C. Clogg, San Francisco, CA: Jossey-Bass.

Courgeau, D. (1988), *Méthodes de Mesure de la Mobilité Spatiale: Migrations Internes, Mobilité Temporaire, Navettes* (Methods for Measuring Spatial Mobility: Internal Migration, Temporary Mobility, Commuting), Paris: Éditions de L'Institut National D'Études Démographiques.

Doeve, W. L. J. (1987), "How Do We Measure Migration? The Preferred Migration Questions for the Global 1990 Round of Population Censuses," paper presented at the meeting of the International Union for the Scientific Study of Population, Tianjin, China, October 1987.

Feeney, G., and Ross, J. A. (1984), "Analyzing Open Birth Interval Distributions," *Population Studies*, 38, 473–478.

Hamerle, A. (1991), "On the Treatment of Interrupted Spells and Initial Conditions in Event History Analysis," *Sociological Methods and Research*, 19, 388–414.

Instituto Brasileiro de Geografia e Estatística (IBGE) (1980), *Censo Demográfico/80-Amostra de 3%* (1980 Demographic Census—3% Sample).

Keiding, N. (1991), "Age-Specific Incidence and Prevalence: A Statistical Perspective," *Journal of the Royal Statistical Society*, Ser. A, 154, 371–412.

Keyfitz, N. (1985), *Applied Mathematical Demography*, New York: Springer-Verlag.

Martine, G. (1990), "Brazil," in *International Handbook of Internal Migration*, eds. C. B. Nam, W. J. Serow, and D. F. Sly, New York: Greenwood Press.

Rees, P. H. (1985), "Does It Really Matter Which Migration Data You Use in a Population Model?," in *Contemporary Studies of Migration: Proceedings of the Second British-Dutch Symposium on Population Geography*, eds. P. E. White and B. Van der Knaap, Norwich, U.K.: Geo Books.

——— (1986), "Developments in the Modeling of Spatial Populations," in *Population Structures and Models*, eds. R. Woods and P. H. Rees, London: Allen and Unwin.

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

Sheps, M. C., Menken, J. A., Ridley, J. C., and Lingner, J. W. (1970), "Truncation Effect in Closed and Open Birth Interval Data," *Journal of the American Statistical Association*, 65, 678–693.

Singer, B., and Spilerman, S. (1976), "The Representation of Social Processes by Markov Models," *American Journal of Sociology*, 82, 1–54.

Sorensen, A. B. (1977), "Estimating Rates From Retrospective Questions," in *Sociological Methodology 1977*, ed. D. R. Heise, San Francisco: Jossey-Bass.

Srinivasan, K. (1968), "A Set of Analytical Models for the Study of Open Birth Intervals," *Demography*, 5, 34–44.

United Nations (1978), *Statistics of Internal Migration: A Technical Report*, U.N. Statistical Papers, Ser. F, No. 23.

United Nations (1980), *Principles and Recommendations for Population and Housing Censuses*, U.N. Statistical Papers, Ser. M, No. 67.

Wheat, R. D., and Morrison, D. G. (1994), "Regularity, Recency, and Rates," *European Journal of Operational Research*, 76, 283–289.