

ORIGINAL ARTICLE

Open Access

Recognizing duration effects in multistate population models



Robert Schoen 

Correspondence: rschoen309@att.net

Department of Sociology,
Pennsylvania State University,
University Park, USA

Abstract

The risk of many demographic events varies by both current state and duration in that state. However, the use of such semi-Markov models has been substantially constrained by data limitations. Here, a new specification of the semi-Markov transition probability matrix in terms of the underlying rates is provided, and a general procedure is developed to estimate semi-Markov probabilities and rates from adjacent population data.

Multistate models recognizing marriage and divorce by duration in state are constructed for United States Females, 1995. The results show that recognizing duration in the married and divorced states adds significantly to the model's analytical value. Extending the constant- α method to semi-Markov models, 2000–2005 U.S. population data and 1995 cross-product ratios are employed to estimate 2000–2005 duration-dependent transfer probabilities and rates.

The present analyses provide new relationships between probabilities and rates in semi-Markov models. Extending the constant cross-product ratio estimation approach opens new sources of data and expands the range of data susceptible to state-duration analyses.

Keywords: Multistate models, Semi-Markov, Remarriage by duration since divorce, Divorce by duration of marriage, Semi-Markov transition probability matrix, Cross-product ratios

Introduction

Multistate models typically follow persons as they move from state to state over age and/or time, using the Markov assumption that the risk of movement depends only on a person's current status. However, demographers have long known that, in many situations, the duration or length of time a person has been in their current state, can substantially affect the risk of an interstate movement.

Models that recognize both current state and duration in that state are known as semi-Markov models, and a substantial body of statistical, actuarial, and demographic literature has explored them. Life insurance actuaries have long used "Select and Ultimate" life tables, which are based on mortality by age, sex, and years since the policy was purchased. In the USA, an ongoing Society of Actuaries mortality investigation used a 15-year select period (Jordan Jr., 1975, p. 24–28). Actuarial experience showed

that individuals at short policy durations had higher death rates than similar persons at long durations. Schoen (1977) showed that the probability of divorce varied with both age and duration of marriage, with duration having a greater effect than age.

The statistical properties of semi-Markov models were explored in depth beginning in the 1950s (e.g., Smith 1955), and that work has since been extended (e.g., Cinlar, 1969; Feller 1968; Ginsberg, 1971; Grabski, 2016; and Hoem, 1972). With regard to multistate models, the pioneering work of Wolf (1988) was the first to examine life table construction with duration dependence. Other noteworthy contributions/applications were made by Cai et al., 2006; Cook and Lawless (2018); Hennessey (1980); Keilman and Gill (1986); Lynch and Brown (2010); and Rajulton (1985). The Appendix section provides additional references and citations to implementing software.

Here, there are two objectives. First, to relate multistate semi-Markov rates to semi-Markov probabilities and explore the implications of those relationships. Second, to extend the Constant- α approach to estimate semi-Markov probabilities and rates from data on adjacent populations in the context of a never married/married/divorced multistate model that recognizes duration of marriage and duration since divorce.

Specifying the state-duration transition probability matrix

To develop the semi-Markov model, let there be S states, where state S_i has j durations. Each of the N state-duration categories is termed a cell. For simplicity, assume no mortality or other source of attrition, and age/time intervals of n years. To be concrete, and with little loss of generality, we emphasize the 2-state model where each state has 3 specified durations: 0, 1, and 2, the first two of n years, and the last duration category open-ended.

The logical structure of an age-duration model means that the possible flows between cells are restricted. Persons at each duration in each state can only move (that is move between states, not advance to a higher duration) to duration category zero in one of the other $S-1$ states. Accordingly, in the 2-state, 3-duration, model of Eq. (1), there are 2 (origin states) \times 3 (durations) \times 1 (destination state) = 6 possible transition rates.

Markov models are generally based on the underlying forces of transition or the corresponding rates of transfer from one state to another. In the semi-Markov context, however, persons also advance to a higher duration in the same state, a status change that is not directly captured by any occurrence/exposure transfer rate. Semi-Markov probabilities reflect both interstate transfer and intrastate advancement. Specifically, with $N = 6$, the 6×6 transition probability matrix of a 2-state, 3-duration model has the form

$$A = \begin{bmatrix} \pi_{10,10} & \pi_{11,10} & \pi_{12,10} & \pi_{20,10} & \pi_{21,10} & \pi_{22,10} \\ \pi_{10,11} & 0 & 0 & 0 & 0 & 0 \\ 0 & \pi_{11,12} & \pi_{12,12} & 0 & 0 & 0 \\ \pi_{10,20} & \pi_{11,20} & \pi_{12,20} & \pi_{20,20} & \pi_{21,20} & \pi_{22,20} \\ 0 & 0 & 0 & \pi_{20,21} & 0 & 0 \\ 0 & 0 & 0 & 0 & \pi_{21,22} & \pi_{22,22} \end{bmatrix} \quad (1)$$

where $\pi_{hi,jk}$ is the probability that a person in state h at duration i at the start of an interval is in state j at duration k at the end of the interval. The rows of A represent destination states, while the columns of A represent origin states. In the model we are

considering, where there is no attrition, all columns sum to 1. Of the 36 cells of matrix A , only 18 have non-zero probabilities. There are 3 non-zero probabilities in each column, as each person can, at the end of the interval, either advance to the next higher duration (or stay at the highest duration) in the initial state or be at duration zero in *any* state. Multiple moves (i.e., changes of state) in an interval are possible, and a person can be at duration zero in the initial state by moving to another state and then returning. All persons who move during an interval are at duration zero at the end of the interval.

In much life table construction, the transition probability matrix is found from a matrix of underlying rates. Because a conventional rate matrix does not capture advancement, going from transfer rates to semi-Markov probabilities requires a cell by cell rather than a matrix-to-matrix approach. Nonetheless, given a set of occurrence/exposure transfer rates, every probability in A can be expressed in terms of those rates using established multistate calculation procedures.

To do so, we follow the procedures in Schoen (Schoen, 1988, Chap.4; Schoen, 2006, Chap.1). It is convenient to start with the (2,1) cell of A , where $\pi_{10,11}$ is the probability that a person initially in state 1 at duration 0 (i.e., years 0 through 4) will be in state 1 at duration 1 at the end of the interval. That probability of advancement to duration category 1 is simply the probability of never leaving state 1, or

$$\pi_{10,11} = [1 - (n/2) m_{10,20}] / [1 + (n/2) m_{10,20}] \quad (2)$$

where $m_{hi,j0}$ is the rate of movement from state h , duration i to state j , duration 0, and the implicit survivorship function is assumed to be linear (see also Preston et al., 2001, Chap. 3). This approach serves to specify all of the six π 's in rows 2, 3, 5, and 6 of A . For present purposes, we assume linear survivorship as that assumption is generally reasonable and yields explicit algebraic solutions. Alternative solution procedures, such as those discussed in Schoen (1988), Chap. 4), are possible but would make little substantive difference here. Should a large rate be encountered that makes the linear assumption problematic when 5-year intervals are used, interval length can be reduced to 1 year.

Continuing down the first column of A , $\pi_{10,20}$ is the probability that a person initially in state 1 at duration 0 ends the interval in state 2 at duration 0. To find that semi-Markov probability, we need a multistate calculation. Consider a 2-state Markov model that does not recognize duration but that has transfer rates from the initial state-duration equal to those prevailing in the semi-Markov model. The other transfer rates in this 2-state Markov model are considered equal to those at duration 0 in the semi-Markov model. Then, $\pi_{10,20}$ is the same as p_{12} , the Markov model probability of starting in state 1 and ending in state 2. Under the linear assumption, in a 2-state model where m_{ij} is the occurrence/exposure rate of transfer from state i to state j , the matrix of rates is of the form

$$M = \begin{bmatrix} -m_{12} & m_{21} \\ m_{12} & -m_{21} \end{bmatrix} \quad (3)$$

and the transition probability matrix is

$$\mathbf{PP} = \begin{bmatrix} p_{11} & p_{21} \\ p_{12} & p_{22} \end{bmatrix} \quad (4)$$

where p_{ij} is the probability that a person initially in state i will be in state j at the end of the interval. As there is no attrition, each column of \mathbf{M} sums to zero, with the diagonal element equal to minus the sum of the off-diagonal rates. Each column of \mathbf{PP} sums to one.

Under the linear assumption (Schoen, 2006, Section 1.6), Markov transition probability matrix \mathbf{PP} can be found from rate matrix \mathbf{M} by

$$\mathbf{PP} = [\mathbf{I} - (n/2) \mathbf{M}]^{-1} [\mathbf{I} + (n/2) \mathbf{M}] \quad (5)$$

where \mathbf{I} is the $N \times N$ identity matrix (which has ones on the main diagonal and zeros elsewhere). Thus $\pi_{10,20}$ follows from p_{12} using rates from duration zero to duration zero. The algebraic expression is

$$\pi_{10,20} = 2n m_{10,20} / [2 + n(m_{10,20} + m_{20,10})] \quad (6)$$

The values of $\pi_{11,20}$ and $\pi_{12,20}$ can also be obtained from the p_{12} element, but by using rates from state 1 at durations 1 and 2, respectively. All first transfer rates take persons to the other state at duration zero, while subsequent rates take those persons from duration zero to duration zero.

To find $\pi_{10,10}$, the remaining probability in the first column, we need only use the fact that the sum of each column is one, hence

$$\pi_{10,10} = 1 - \pi_{10,11} - \pi_{10,20} \quad (7)$$

Algebraically, the result is

$$\pi_{10,10} = (2n^2 m_{10,20} m_{20,10}) / [(2 + n m_{10,20})(2 + n m_{10,20} + n m_{20,10})] \quad (8)$$

The same approach yields all of the remaining probabilities in \mathbf{A} in terms of the set of underlying rates. By straightforward extension, it takes any set of transition rates and provides the elements of the associated semi-Markov transition probability matrix. Table 1 provides the complete algebraic solution for the probability matrix of a 2-state, 3-duration model in terms of the underlying transition rates.

While Table 1 gives the 18 probabilities in \mathbf{A} in terms of the underlying 6 rates, we can also take those 18 equations and solve for the 6 rates and for 12 probabilities in terms of the other 6 probabilities. In the case of probability matrix \mathbf{A} of Eq. (1), Table 2 gives explicit solutions for the 6 rates and for 12 probabilities in terms of the other 6 π s. The implications of doing so are substantial and give rise to what may be termed a *rate principle*: the number of independent probabilities in a semi-Markov model is given by the number of independent rates underlying that model. Matrix \mathbf{A} , like its underlying rate matrix, has only 6 independent elements. It follows that an arbitrary probability matrix of the form of \mathbf{A} is likely to have *no underlying set* of semi-Markovian rates because its elements would not be constrained by the relationships in Table 2. In other words, that probability matrix is not “embedded” in a Markovian process (cf. Singer & Spilerman, 1976). The problem of finding rates from such probabilities is addressed in a later section.

Table 1 Elements of the 2-state, 3-duration transition probability matrix

The 2-state, 3-duration transition probability matrix of Eq(1) is

$$A = \begin{bmatrix} \pi_{10,10} & \pi_{11,10} & \pi_{12,10} & \pi_{20,10} & \pi_{21,10} & \pi_{22,10} \\ \pi_{10,11} & 0 & 0 & 0 & 0 & 0 \\ 0 & \pi_{11,12} & \pi_{12,12} & 0 & 0 & 0 \\ \pi_{10,20} & \pi_{11,20} & \pi_{12,20} & \pi_{20,20} & \pi_{21,20} & \pi_{22,20} \\ 0 & 0 & 0 & \pi_{20,21} & 0 & 0 \\ 0 & 0 & 0 & 0 & \pi_{21,22} & \pi_{22,22} \end{bmatrix} \quad (A.1)$$

To conveniently express the elements of **A** in terms of the underlying occurrence / exposure rates of interstate transfer, separate matrices, **A₁₃** and **A₄₆**, are shown below for the first 3 and the second 3 columns of **A**. We then have

$$A_{13} = \begin{bmatrix} 2 n^2 m_{10,20} m_{20,10} / (D_0 E_0) & 2 n^2 m_{11,20} m_{20,10} / (D_1 E_1) & 2 n^2 m_{12,20} m_{20,10} / (D_2 E_2) \\ (2-n m_{10,20}) / (D_0) & 0 & 0 \\ 0 & (2-n m_{11,20}) / (D_1) & (2-n m_{12,20}) / (D_2) \\ 2 n m_{10,20} / (E_0) & 2 n m_{11,20} / (E_1) & 2 n m_{12,20} / (E_2) \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad (A.2a)$$

where for $j = 0, 1,$ and $2, D_j = 2 + n m_{1j,20}$ and $E_j = 2 + n m_{1j,20} + n m_{20,10}$; and

$$A_{46} = \begin{bmatrix} 2 n m_{20,10} / (G_0) & 2 n m_{21,10} / (G_1) & 2 n m_{22,10} / (G_2) \\ 0 & 0 & 0 \\ 2 n^2 m_{10,20} m_{20,10} / (F_0 G_0) & 2 n^2 m_{10,20} m_{21,10} / (F_1 G_1) & 2 n^2 m_{10,20} m_{22,10} / (F_2 G_2) \\ (2-n m_{20,10}) / (F_0) & 0 & 0 \\ 0 & (2-n m_{21,10}) / (F_1) & (2-n m_{22,10}) / (F_2) \end{bmatrix} \quad (A.2b)$$

where for $j = 0, 1,$ and $2, F_j = 2 + n m_{2j,10}$ and $G_j = 2 + n m_{2j,10} + n m_{10,20}$.

Table 2 Expressions for 6 rates and 12 probabilities in a 2-state, 3-duration semi-Markov model in terms of the other 6 probabilities.

In a model like that of Eq. (1), expressing the 6 underlying rates and 12 probabilities in terms of the probabilities $\pi_{m0v0}, \pi_{m1m2}, \pi_{m2v0}, \pi_{v0m0}, \pi_{v1m0},$ and π_{v2v2} , yields

$$\begin{aligned} m_{m0v0} &= 2 \pi_{m0v0} / [n (2 - \pi_{m0v0} - \pi_{v0m0})] \\ m_{m1v0} &= 2 (1 - \pi_{m1m2}) / [n (1 + \pi_{m1m2})] \\ m_{m2v0} &= 2 \pi_{m2v0} (2 - \pi_{m0v0}) / [n (2 - \pi_{m0v0} - \pi_{v0m0}) (2 - \pi_{m2v0})] \\ m_{v0m0} &= 2 \pi_{v0m0} / [n (2 - \pi_{m0v0} - \pi_{v0m0})] \\ m_{v1m0} &= 2 \pi_{v1m0} (2 - \pi_{v0m0}) / [n (2 - \pi_{m0v0} - \pi_{v0m0}) (2 - \pi_{v1m0})] \\ m_{v2m0} &= 2 (1 - \pi_{v2v2}) / [n (1 + \pi_{v2v2})] \\ \pi_{m0m0} &= \pi_{m0v0} \pi_{v0m0} / (2 - \pi_{v0m0}) \\ \pi_{m0m1} &= (2 - 2 \pi_{m0v0} - \pi_{v0m0}) / (2 - \pi_{v0m0}) \\ \pi_{m1m0} &= \pi_{v0m0} (1 - [\pi_{m1m2}]^2) / [(4 - 2 \pi_{m0v0} - \pi_{v0m0} + \pi_{v0m0} \pi_{m1m2})] \\ \pi_{m1v0} &= 2(2 - 2 \pi_{m1m2} - \pi_{m0v0} - \pi_{v0m0} + \pi_{v0m0} \pi_{m1m2} + \pi_{m0v0} \pi_{m1m2}) / \\ & [(4 - 2 \pi_{m0v0} - \pi_{v0m0} + \pi_{v0m0} \pi_{m1m2})] \\ \pi_{m2m0} &= \pi_{v0m0} \pi_{m2v0} (2 - \pi_{m2v0}) / [(4 - 2 \pi_{m0v0} - 2 \pi_{v0m0} + \pi_{m2v0} \pi_{v0m0})] \\ \pi_{m2m2} &= (4 - 4 \pi_{m2v0} - 2 \pi_{m0v0} - 2 \pi_{v0m0} + 2 \pi_{m2v0} \pi_{m0v0} + \pi_{m2v0} \pi_{v0m0}) / \\ & [(4 - 2 \pi_{m0v0} - 2 \pi_{v0m0} + \pi_{v0m0} \pi_{m2v0})] \\ \pi_{v0v0} &= \pi_{m0v0} \pi_{v0m0} / (2 - \pi_{m0v0}) \\ \pi_{v0v1} &= (2 - 2 \pi_{v0m0} - \pi_{m0v0}) / (2 - \pi_{m0v0}) \\ \pi_{v1v0} &= \pi_{v1m0} \pi_{m0v0} (2 - \pi_{v1m0}) / [(4 - 2 \pi_{m0v0} - 2 \pi_{v0m0} + \pi_{m0v0} \pi_{v1m0})] \\ \pi_{v1v2} &= (4 - 4 \pi_{v1m0} - 2 \pi_{m0v0} - 2 \pi_{v0m0} + 2 \pi_{v1m0} \pi_{v0m0} + \pi_{v1m0} \pi_{m0v0}) / \\ & [(4 - 2 \pi_{m0v0} - 2 \pi_{v0m0} + \pi_{v1m0} \pi_{m0v0})] \\ \pi_{v2m0} &= 2 (2 - 2 \pi_{v2v2} - \pi_{m0v0} - \pi_{v0m0} + \pi_{v2v2} \pi_{v0m0} + \pi_{v2v2} \pi_{m0v0}) / \\ & [(4 - 2 \pi_{v0m0} - \pi_{m0v0} + \pi_{v2v2} \pi_{m0v0})] \\ \pi_{v2v0} &= \pi_{m0v0} (1 - [\pi_{v2v2}]^2) / [(4 - 2 \pi_{v0m0} - \pi_{m0v0} + \pi_{m0v0} \pi_{v2v2})] \end{aligned}$$

The situation where duration effects are implicit

Duration effects are not explicitly recognized in Markovian analyses. Nonetheless, the states of a Markov model do have an implicit duration composition. The objective of this section is to determine that distribution in multistate models. Implicit durations in fertility models are examined in Schoen (2019).

Here, let all rates from every given state to another given state be the same at all durations. In the context of our 2-state, 3-duration model, there are only 2 distinct rates, m_{12} and m_{21} . Using the approach of the preceding section, we can write all of the elements of the semi-Markov transition probability matrix in terms of those two rates. The dominant right eigenvector of that probability matrix provides the long-term (stable population) state-duration composition (Schoen, 2006). That eigenvector, \mathbf{u} , can readily be found from \mathbf{A} using mathematical software such as Maple or Mathematica.

Here, the 6×1 state-duration composition vector \mathbf{u} reflects the relative number in each state-duration, beginning with state 1 at durations 0, 1, and 2, and following with state 2 at durations 0, 1, and 2. That vector can be written as

$$\mathbf{u} = \begin{bmatrix} 1 \\ \pi_{10,11} \\ \pi_{10,11}^2 (2 - \pi_{20,10}) / (2\pi_{10,20}) \\ (2 - \pi_{20,10}) / (2 - \pi_{10,20}) \\ \pi_{20,21} (2 - \pi_{20,10}) / (2 - \pi_{10,20}) \\ \pi_{20,21}^2 (2 - \pi_{20,10}) / (2\pi_{20,10}) \end{bmatrix} \quad (9)$$

where the number in state 1 at duration 0 is scaled to one. The relative size of the 2 states is given by the ratio $(2 - \pi_{20,10}) / (2 - \pi_{10,20})$. The larger $\pi_{20,10}$ is to $\pi_{10,20}$, the smaller the proportion in state 2. Within each state, the proportion decreases with duration, by a factor of $\pi_{10,11}$ in state 1 and $\pi_{20,21}$ in state 2. The highest, open-ended duration has an additional factor representing all of the fractions at (unrecognized) higher 5-year durations.

Estimating transition probabilities from adjacent populations under constant- α

There are a number of situations where population figures by state and duration at both the beginning and end of an age/time interval are known, but there is no information on the transitions during that interval. The constant- α approach, presented in Schoen (2020), can be extended to the semi-Markov case and allow interstate probabilities to be estimated. This section describes how to do so.

The constant- α approach is based on the assumption that the cross-product ratios (α 's) of the multistate transition probability matrix are fixed. Cross-product ratios are analogous to odds ratios, can be formed from any rectangular set of 4 non-zero matrix elements, and equal the product of the upper left and lower right elements divided by the product of the lower left and upper right elements. For example, in \mathbf{A} , we can define

$$\alpha_{1142} = \pi_{10,10} \pi_{11,20} / (\pi_{10,20} \pi_{11,10}) \quad (10)$$

which is one of the 7 distinct α 's in A . The subscripts "1142" represent the upper-left (1,1) and lower right (4,2) elements of the ratio. A distinct cross-product ratio includes at least one cell that is not included in any other cross-product ratio.

If the transition probability matrix is viewed as a contingency table, the constant α 's can be interpreted as the fixed interaction effects of a saturated log linear model. Preserving α 's can provide maximum likelihood estimates that maximize entropy, as they find the pattern of interstate flows that can arise in the greatest number of ways. In multistate Markov models, Schoen (2020) described how to estimate transition probabilities from a variety of data sources and found that the approach provided good estimates of movements between poverty states in the USA.

Here, we seek to implement the constant- α approach in the semi-Markov context where data are available on adjacent populations. Let x_{jk} represent the start of interval population in state j at duration k , and let y_{jk} represent the end of interval population in state j at duration k . Then

$$y = P x \tag{11}$$

where P , which has the form of A , is the transition probability matrix and vectors x and y contain the x_{jk} and y_{jk} population values, respectively.

In the no-mortality semi-Markov case, let us rewrite Eq. (11), using a base transition probability matrix, B , whose elements imply the set of cross-product ratios that are being held constant. Matrix B should be chosen with care and needs to reflect a population with the same state-duration structure and the same interstate movements as the population whose probabilities are to be estimated.

To satisfy the projection relationship, matrix B is pre-multiplied by a diagonal matrix R of row factors and post-multiplied by a diagonal matrix, C , of column factors. The i -th diagonal element of R is r_i , with $r_1 = 1$, and the j -th diagonal element of C is c_j . Hence, we can write

$$y = R B C x \tag{12}$$

where the desired transition probability matrix, P , is given by

$$P = R B C \tag{13}$$

and

$$z = \begin{bmatrix} z_{10} \\ z_{11} \\ z_{12} \\ z_{20} \\ z_{21} \\ z_{22} \end{bmatrix} \tag{14}$$

where z can be either x or y . By the definition of α , matrix P has the same cross-product ratios as matrix B . However, the elements of P generally do *not* satisfy the constraints of Table 2 even when the elements of B do.

With N state-durations, Eq. (12) has $(2N - 1)$ unknowns, the N diagonal elements of C and $(N - 1)$ diagonal elements of R . Those $(2N - 1)$ unknowns can be found from the $(N - 1)$ independent scalar projection equations contained in Eq. (12) and the N equations that require that the N columns of P sum to 1. An iterative solution can be found, but here we proceed by solving the $(2N - 1)$ equations. That approach has the advantage of finding all of the possible solutions. There can be more than one valid (i.e., real and non-negative) demographic solution, while there may be *no* valid solutions at all. The latter can arise if the cross-product ratios are incompatible with the given populations, the most obvious case being when a large ending population at one duration arises solely from a small initial population at the previous duration.

When the probabilities and rates are known, they can be used in life tables or other demographic models. For example, the life course of a cohort can be traced by a multi-state life table, and all of the life table summary measures calculated. We now turn to applying the approaches presented here, first to use rates to calculate a state-duration life table, and second to estimate interstate transfer probabilities and rates using the constant- α method.

Calculating a state-duration model from duration-specific rates

Here, we calculate a semi-Markov model by starting with a Markovian multistate model and extending it through the introduction of duration-specific rates. Marital status models are particularly appropriate for such extensions, as both divorce and remarriage after divorce are known to vary by duration in state.

We begin with the age-state-specific rates used in the construction of the marital status life table for United States Females, 1995 (cf. Schoen & Standish, 2001). To simplify matters, the semi-Markov calculations proceed from age 15 to age 50, ignoring mortality. That yields a 3-state model with states never married (s), married (m), and divorced (v).

We extend the 1995 life table by adding 5-year duration categories 0 and 1, and open-ended duration category 2, to both the married and divorced states. Data on second marriages by duration of first divorce and age at divorce are available for 1995 from Bramlett and Mosher (Bramlett & Mosher, 2001, Table 7) and provide the basis for allocating age-specific remarriage rates (m_{vm}) to the three duration categories. Age-duration-specific divorce rates (m_{mv}) for first marriages in California, 1969, are provided in Schoen (Schoen, 1975, Table 2). While somewhat old, they appear to be the most suitable values available. The relative sizes of those published duration-specific rates were then weighted by the initial state composition at each age interval in the extended life table. The weighted differential values, by duration, were multiplicatively adjusted to reproduce the all-durations rate in the 1995 life table. Those adjusted duration-specific rates were the inputs used to calculate the extended multistate life table.

The construction of the extended life table proceeded age by age, beginning with 100,000 persons in the never married state at exact age 15. The state-duration composition of the extended table at the end of each age interval is generated from the initial state-duration composition survived, per Eq. (11), by a 7×7 state-duration transition probability matrix. That transition matrix is the 6×6 matrix of Eq. (1), with a top row and left-most column added to reflect the never married state. The expressions for the

marriage and divorce cells of the matrix are shown in Table 1. There is no re-entry to the never married state, and the probabilities that a never married person ends the interval never married (π_{ss}), married at duration 0 ($\pi_{s,m0}$), and divorced at duration 0 ($\pi_{s,v0}$) are

$$\begin{aligned}\pi_{ss} &= (2 - nm_{s,m0}) / (2 + nm_{s,m0}) \\ \pi_{s,m0} &= 2nm_{s,m0}(2 + nm_{v0,m0}) / ((2 + nm_{s,m0})(2 + nm_{m0,v0} + nm_{v0,m0})) \\ \pi_{s,v0} &= 2n^2m_{s,m0}m_{m0,v0} / ((2 + nm_{s,m0})(2 + nm_{m0,v0} + nm_{v0,m0}))\end{aligned}\quad (15)$$

The linear assumption is used throughout.

Persons moving between states always begin the next interval at duration 0. The extended life table terminates at exact age 50, after which mortality is more salient and there are fewer marital status transitions. The source 1995 rates and the extended life table functions are given in Table 3.

Selected extended marital status life table measures are presented in Table 4. Panel A shows that over the 15 to 50 age interval, the ratio of divorces to all marriages was 0.438 in the state-duration life table and 0.403 in the 1995 life table. The ratio of remarriages to divorces was 0.586 in the extended life table and 0.655 in the 1995 no-durations table. Thus, there is more divorce and less remarriage in the extended life table. At the same time, the extended life table has a longer average duration of marriage and a shorter average duration of divorce.

Those results may seem inconsistent at first, but the figures in Table 4, panel B and the first panel of Table 3 offer an explanation. Divorces are rather evenly distributed over the three duration categories, but remarriages are heavily (71%) concentrated at duration 0. Divorce rates decline gradually over age, while remarriage rates drop sharply after age 35. Thus, the 3-duration extended life table has faster and earlier remarriage, which shortens the average duration of a divorce and lengthens the average duration of a marriage. Recognizing duration does make a difference.

Estimating probabilities from adjacent populations using constant- α

The approach here uses the cross-product ratios from the 1995 extended life table of the previous section to estimate duration-specific probabilities from marital status life table populations for United States Females, 2000–2005, at ages 30 to 35. The input values are the 1995 table state-duration population distributions at ages 30 and 35, and the 7×7 array of 1995 probabilities, which have the form of Eq. (1) augmented by a first row and left-most column to reflect the never married (s) state. The 2000–2005 life table populations are based on Schoen (2016). Following the procedure described after the presentation of Eqs. (11)–(13), the $(2N - 1) = 13$ equations were solved for the row and column adjustment factors to the 1995 base probabilities. There were multiple solutions, but only one was demographically appropriate (i.e., with all rates between 0 and 1; though rates can exceed one, such a rate would be unrealistic here). All of the adjustment factors were fairly close to 1, varying only from 0.70 to 1.61. The 2000–2005 estimated matrix of probabilities, \mathbf{P} , for ages 30 to 35 was then calculated using Eq. (13). The result is

Table 3 Values from the source and extended marital status life tables for United States Females, 1995

Item	Age						
	15–20	20–25	25–30	30–35	35–39	40–45	45–50
<i>I. Rates from the source 1995 marital status life table</i>							
1. M_{sm}	.0200	.0923	.1268	.0861	.0531	.0324	.0166
2. M_{mv}	.0528	.0404	.0388	.0316	.0258	.0216	.0152
3. M_{vm}	.0629	.2263	.1824	.1521	.1017	.0720	.0586
<i>II. Values from the extended life table</i>							
<i>A. Persons (ℓ) by state at the end of the age interval</i>							
1. ℓ_s	90476	56341	29219	18868	14446	12281	11302
2. ℓ_{m0}	8549	33261	29822	16307	10170	7280	5667
3. ℓ_{m1}	0	6980	28339	26045	14228	8872	6590
4. ℓ_{m2}	0	0	4853	27754	47383	55564	59879
5. ℓ_{v0}	975	3147	6487	7532	7211	6395	4609
6. ℓ_{v1}	0	270	1154	2954	4294	4563	4270
7. ℓ_{v2}	0	0	127	540	2268	5045	7684
<i>B. Person-years lived (L) by state during the age interval</i>							
1. L_s	476190	367043	213900	120217	83284	66816	58957
2. L_{m0}	21372	121975	157707	115323	66194	43626	32367
3. L_{m1}	0	0	100431	135961	100684	57751	38654
4. L_{m2}	0	0	0	81516	187841	257366	288607
5. L_{v0}	2438	10982	24085	35045	36856	34016	27510
6. L_{v1}	0	0	3878	10270	18120	22143	22084
7. L_{v2}	0	0	0	1668	7020	18282	31821
<i>C. Interstate transfers (d) during the age interval</i>							
1. $d_{s,m0}$	9524	34135	27122	10351	4422	2165	979
2. $d_{m0,v0}$	1128	4928	5040	3118	1803	1190	645
3. $d_{m1,v0}$	0	0	7222	4512	2549	1465	716
4. $d_{m2,v0}$	0	0	0	4150	4773	4954	4056
5. $d_{v0,m0}$	153	2485	4465	5245	4036	3060	2192
6. $d_{v1,m0}$	0	0	561	1722	1600	1245	1100
7. $d_{v2,m0}$	0	0	0	201	481	822	1268
8. All divorces	1128	4928	12262	11780	9125	7609	5417
10 All remarriages	153	2485	5026	7168	6117	5127	4560

Notes: The model states are never married (s), married (m), and divorced (v), with 5-year duration categories 0, 1, and 2. The initial population, $\ell_s(15)$, is 100,000. All entries are rounded independently. Cross-product ratio α_{hijk} equals the product of the (h,i) and (j,k) elements of the matrix divided by the product of the (h,k) and (i,j) elements
 Source: See text and Schoen and Standish (2001); access was provided to unpublished tabulations

Table 4 State-duration life table summary measures of marriage and divorce, United States Females, 1995

A. Events over the 15 to 50 age interval			
	Extended life table	1995 life table	
1. Ratio of divorces/marriages	.438	.403	
2. Ratio of remarriages/divorces	.586	.655	
3. Average duration of a marriage	15.15	14.69	
4. Average duration of a divorce	5.86	6.33	
B. Events by duration in state, extended life table			
Duration:	0	1	2
1. Fraction of divorces, ages 15–50	.342	.315	.343
2. Fraction of remarriages, ages 15–50	.706	.203	.091

Notes: The average duration of a marriage is the number of person-years lived in the married state by persons aged 15–50 divided by the number of marriages between the ages of 15 and 50. The average duration of a divorce is the number of person-years lived in the divorced state by persons aged 15–50 divided by the number of divorces between the ages of 15 and 50
Source: See text

$$\mathbf{P} = \begin{bmatrix} .7211 & 0 & 0 & 0 & 0 & 0 & 0 \\ .2629 & .0264 & .0236 & .0345 & .4214 & .3655 & .2580 \\ 0 & .8803 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & .8921 & .8367 & 0 & 0 & 0 \\ .0160 & .0933 & .0843 & .1288 & .0256 & .0216 & .0166 \\ 0 & 0 & 0 & 0 & .5530 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & .6129 & .7254 \end{bmatrix} \tag{16}$$

with all columns summing to 1. The largest interstate movement probabilities are from the divorced states to state m_0 . Married persons have probabilities of remaining married of greater than 80%.

In sum, the calculation of an estimated transition probability matrix from a base probability matrix and adjacent populations is straightforward. However, the calculation of the interstate movement *rates* (and decrements) from the adjacent populations and matrix \mathbf{P} probabilities is more complicated and is examined next.

Calculating the non-Markovian marriage and divorce rates and decrements

Estimated transition probability matrix \mathbf{P} is non-Markovian because constraints such as those given in Table 2 generally do not hold. Finding appropriate rates consistent with the input populations and estimated probabilities is a non-trivial problem that, to the best of my knowledge, has not been carefully examined in the demographic literature.

In order to find occurrence/exposure rates satisfying Eqs. (12) and (16), more than 7 distinct rates are needed, and there is no unique solution. Here, a 2-step approach is proposed. Step 1 distinguishes between rates that describe a person’s first interstate movement and those that relate to a subsequent movement. Let M_f denote a first move rate, and M a subsequent move rate. To introduce decrements, let df_{jk} be the number of first moves from persons in state-duration j at the start of the interval who move to state k during the interval.

There are 7 first decrement rates, one from each state-duration, and every first move has to be to duration zero in the other state. These rates are related to the probability

of a first decrement, and in the linear case can be described by an expression like Eq. (2). Rewriting Eq. (2) to solve for Mf in terms of π yields

$$Mf_{jk} = (2/n) (1 - \pi_{jh}) / (1 + \pi_{jh}) \quad (17)$$

where h is the state-duration where persons initially in state-duration j would be at the end of the interval, absent a move. Eq. (17) provides all 7 first transfer rates. Again using established linear life table relationships, the 7 first decrements produced by those rates are of the form

$$df_{jk} = x_j (2 n Mf_{jk}) / (2 + n Mf_{jk}) \quad (18)$$

where x_j is the beginning of interval population in the initial state-duration.

To find the subsequent rates and decrements, it is helpful to set out the 7 state-duration model algebraically by writing 7 equations that describe all of the interstate flows. Those 7 flow equations are

$$\begin{aligned} y_s &= x_s - df_{s,m_0} \\ y_{m_1} &= x_{m_0} - df_{m_0,v_0} \\ y_{m_2} &= x_{m_1} - df_{m_1,v_0} + x_{m_2} - df_{m_2,v_0} \\ y_{v_1} &= x_{v_0} - df_{v_0,m_0} \\ y_{v_2} &= x_{v_1} - df_{v_1,m_0} + x_{v_2} - df_{v_2,m_0} \\ y_{m_0} &= df_{s,m_0} + df_{v_0,m_0} + df_{v_1,m_0} + df_{v_2,m_0} + (n/2)y_{v_0}M_{v_0,m_0} - (n/2)y_{m_0}M_{m_0,v_0} \\ y_{v_0} &= df_{m_0,v_0} + df_{m_1,v_0} + df_{m_2,v_0} - (n/2)y_{v_0}M_{v_0,m_0} + (n/2)y_{m_0}M_{m_0,v_0} \end{aligned} \quad (19)$$

The first five flow equations follow from the first decrements as defined above, that is the first movements based on the person's initial state-duration. The move of a person initially in state-duration m_0 who advances to state-duration m_1 and then moves to state-duration v_0 during the interval is included in df_{m_0,v_0} , and hence in Mf_{m_0,v_0} . Since there is no attrition, summing all of the seven flow equations confirms that the total ending population equals the total initial population. Thus, there are only six independent flow equations.

The last two flow equations are conceptually different and include subsequent moves between state-durations m_0 and v_0 . Those two equations do not include terms for x_{m_0} and x_{v_0} because those persons, absent a move, would be in state-durations m_1 and v_1 , respectively at the end of the interval. All subsequent moves from state-durations m_0 and v_0 must come from entrants during the interval, i.e., the df terms in those flow equations. Under the linear assumption, those entries are, on average, at mid-interval. It follows that $(n/2)$ times the ending (y_{m_0} or y_{v_0}) population reflects the number of person-years lived in state-duration m_0 or v_0 during the interval. Multiplying those person-years by the $M_{m_0v_0}$ or $M_{v_0m_0}$ rate of subsequent movement provides the number of subsequent moves between state-durations m_0 and v_0 .

In general, the first and subsequent rates for the same transition differ. Assuming $M = Mf$ produces values that do not satisfy the flow equations. Furthermore, those last two flow equations reveal a further difficulty: they only determine *net* subsequent decrements, that is the difference $[(n/2) y_{v_0} M_{v_0,m_0} - (n/2) y_{m_0} M_{m_0,v_0}]$.

To surmount that difficulty and calculate the subsequent rates and decrements, we go to Step 2. Borrowing from Schoen and Jonsson (2003), we assume that the *product* of the rates of divorce and remarriage remains constant. The heuristic argument is one

of “attractiveness”: if (re)marriage becomes more (or less) attractive, one of the rates is likely to rise and the other to fall, so their product can remain unchanged. Thus, we can write

$$M_{f_{m_0,v_0}} M_{f_{v_0,m_0}} = M_{m_0,v_0} M_{v_0,m_0} \tag{20}$$

Using Eq. (20) with one of the last two flow equations in Eq. (19) allows the calculation of the two subsequent (M) rates and decrements.

The results of the 2-step calculations for the rates and decrements values are shown in Table 5, along with the beginning and ending populations by state-duration. First move divorces occur in roughly equal numbers in the three duration groups, while first move remarriages are concentrated at duration zero.

Table 6 summarizes the seven state-duration model. At ages 30 to 35, the cohort of 100,000 women have a total of 13,471 marriages and 8726 divorces. First decrement divorces were 76% of all divorces, while first decrement remarriages were only 59% of all remarriages, a reflection of the high remarriage rates in the years immediately following a divorce.

The 2-step approach presented in this section permits the calculation of rates and decrements from estimated non-Markovian transition probability matrices, such as the one in Eq. (16). While the solution is not unique because there is insufficient information to fully identify the model’s non-Markovian aspects, a reasonable, demographically sound solution is presented. These procedures extend the constant- α approach to fully provide semi-Markov probabilities, rates, and decrements from a base probability matrix and adjacent population values.

Summary and conclusion

Semi-Markov multistate models, which recognize both current state and duration in that state, are frequently useful in demographic analyses. The risk of many vital and health events, such as marriage, divorce, and recovery from disability, can vary greatly by duration in state, and that differential risk is often worth examining.

Table 5 First (Mf) and subsequent (M) movement rates and decrements (df and d) in the 7 state-duration model, United States Females, 2000–2005, ages 30 to 35

Move	Origin population		First move		Subsequent moves	
	Initial (x)	Ending (y)	Rate (Mf)	Decrement (df)	Rate (M)	Decrement (d)
$s \rightarrow m_0$	29055	20952	.06481	8103	0	0
$m_0 \rightarrow v_0$	21794	12632	.02547	2609	.02658	839
$m_1 \rightarrow v_0$	25626	19185	.02281	2765	0	0
$m_2 \rightarrow v_0$	15387	35735	.03557	2513	0	0
$v_0 \rightarrow m_0$	6071	6840	.11515	2714	.11032	1886
$v_1 \rightarrow m_0$	1782	3357	.09599	690	0	0
$v_2 \rightarrow m_0$	285	1299	.06366	78	0	0

Note: Model states are never married (s), married 0–5 years (m_0), married 5–10 years (m_1), Married 10 or more years (m_2), divorced 0–5 years (v_0), divorced 5–10 years (v_1), and divorced 10 or more years (v_2)
 Source: See text

Table 6 A summary of rates and decrements in the 7 state-duration model, United States Females, 2000–2005, ages 30 to 35

	First decrements	Subsequent decrements	Total decrements	Overall rate
$m_0 \rightarrow v_0$	2609	839	3448	.02573
$v_0 \rightarrow m_0$	2714	1886	4600	.11311
Over all durations:				
Total number of marriages:	13,471			
First marriages	8,103			
Remarriages	5,368			
Total number of divorces:	8,726			
Ratio divorces/all marriages:	.648			
Ratio remarriages/all marriages:	.398			

Notes: Subsequent decrements are calculated by assuming that the product of duration zero divorce and remarriage rates is constant

The overall rate is all duration zero decrements divided by the appropriate total person-years of exposure. The person-years for first move divorces is $(n/2)*(x_{m0} + y_{m1}) = 102,447.5$ and for subsequent divorces is $(n/2)*y_{m0} = 31,580$, totaling 134,027.5. The person-years for first move remarriages is $(n/2)*(x_{v0} + y_{v1}) = 23,570$ and for subsequent remarriages is $(n/2)*y_{v0} = 17,100$, totaling 40,670

Source: See text

A new procedure for writing a semi-Markov transition probability matrix in terms of underlying occurrence/exposure rates of interstate transfer is presented. A rate principle is propounded, which equates the number of independent probabilities in a transition matrix to the number of independent rates in the underlying multistate model.

Standard Markov models, such as conventional multistate life tables, have an implicit duration composition that can be worth examining. Procedures for doing so, in both the long and short term, are described, and the duration structure of a 2-state, 3-duration model is provided.

Using data-derived rates of transfer by duration of marriage and divorce, a 3-state, 7-rate marital status life table is calculated for United States Females, 1995. The results indicate that recognizing duration in state not only provides finer detail, but also enhances the analytical value of the table.

The constant- α approach to estimating multistate transition rates from data on adjacent populations and known cross-product ratios is then extended to semi-Markov models, and applied to estimating duration-specific probabilities in a marital status model for United States Females, 2000–2005. The calculation of the probabilities is straightforward, and a demographically valid 2-step procedure is presented to calculate a consistent set of transfer rates and decrements.

The use of semi-Markov models in demography has been limited, not primarily for substantive reasons, but because of data limitations. The procedures described here facilitate the construction of duration-dependent models from data on both transfer rates and the composition of adjacent populations. The application of semi-Markov models to a broader range of data can give researchers greater descriptive detail and enhanced analytical power.

Appendix

Most of the analytical work on semi-Markov models has been done by statisticians, with some significant applied work by actuaries. Jordan (Jordan Jr., 1975, p. 24–28)

provides a brief, non-technical introduction from an actuarial perspective. Hoem (1972) and Cook and Lawless (2018) are more statistical, but provide good introductory treatments. More advanced treatments can be found in Cai et al. (2006) and Barbu et al. (2017).

The computer programs in this paper were written using Maple software, and other mathematical packages, such as Mathematica, can also be used. The computer package R has the most developed semi-Markov software. Willekens and Putter (2014) give an excellent discussion of multistate software in general, with some useful information for semi-Markov modeling. Some specific semi-Markov packages in R are examined in Alvares et al. (2018) and in Krol and Saint-Pierre (2015).

Acknowledgements

Valuable comments from Lowell Hargens are acknowledged with thanks.

Author's contributions

The author read and approved the final manuscript.

Author information

Robert Schoen, PhD is a Distinguished Senior Scholar, Department of Sociology, Pennsylvania State University (USA).

Funding

The author declares that he received no funding support.

Availability of data and materials

The author declares that unpublished figures from marital status life tables for the United States, 1995, and the United States 2000–2005, are available from the author. All other data used are from the published sources cited.

Declarations

Consent for publication

The single author declares that he is solely responsible for the content and contributions of the paper.

Competing interests

The author declares that he has no competing interests.

Received: 29 March 2021 Accepted: 5 June 2021

Published online: 06 November 2021

References

- Alvares, D., S. Haneuse, C. Lee and K.H. Lee. 2018. SemiCompRisks: An R package for independent and cluster-correlated analyses of semi-competing risks data. (<https://arxiv.org/abs/1801.03567>)
- Barbu, V. S., Karagrigoriou, A., & Makrides, A. (2017). Semi-Markov modeling for multi-state systems. *Methodology and Computing in Applied Probability*, 19(4), 1011–1028. <https://doi.org/10.1007/s11009-016-9510-y>.
- Bramlett, M. D., & Mosher, W. D. (2001). *First marriage dissolution, divorce, and remarriage: United States. Advance data from vital and health statistics, no. 323*. Hyattsville: National Center for Health Statistics.
- Cai, L., Schenker, N., & Lubitz, J. (2006). Analysis of functional status transitions by using a semi-Markov process model in the presence of left-censored spells. *Journal of the Royal Statistical Society, Series C*, 55, 447–491.
- Cinlar, E. (1969). Markov renewal theory. *Advances in Applied Probability*, 1, 123–187.
- Cook, R. J., & Lawless, J. F. (2018). Multistate models for the analysis of life history data. In *Monographs on Statistics and Applied Probability 158*. Boca Raton: Chapman and Hall.
- Feller, W. (1968). *An introduction to probability theory and its applications. Vol 1*, (3d ed.,). New York and London: Wiley.
- Ginsberg, R. B. (1971). Semi-Markov processes and mobility. *The Journal of Mathematical Sociology*, 1, 233–262.
- Grabski, F. (2016). Concept of semi-Markov process. *Scientific Journal of Polish Naval Academy*, 57, 25–36.
- Hennessey, J. C. (1980). An age-dependent, absorbing semi-Markov model of work histories of the Disabled. *Mathematical Biosciences*, 51, 283–304.
- Hoem, J. M. (1972). Inhomogeneous semi-Markov processes, select actuarial tables, and duration-Dependence in demography. In T. N. E. Greville (Ed.), *Population Dynamics*, (pp. 251–296). New York: Academic Press.
- Jordan Jr., C. W. (1975). *Life contingencies*, (2nd ed.,). Chicago: Society of Actuaries.
- Keilman, N., & Gill, R. (1986). *On the estimation of multidimensional demographic models with population registration data. Working paper No. 68*. Voorburg: Netherland Interuniversity Demographic Institute.
- Krol, A. and P. Saint-Pierre. 2015. Semi-Markov: an R package for parametric estimation in multi-state semi-Markov models. (/article_zx/0000.53003)
- Lynch, S. M., & Brown, J. S. (2010). Obtaining multistate life table distributions for highly refined subpopulations from cross-sectional data: a Bayesian extension of Sullivan's method. *Demography*, 47, 1053–1077.
- Preston, S. H., Heuveline, P., & Guillot, M. (2001). *Demography: measuring and modeling demographic processes*. Malden: Blackwell.

- Rajulton, F. (1985). Heterogeneous marital behavior in Belgium, 1970 and 1977: an application of the semi-Markov model to period data. *Mathematical Biosciences*, 73, 197–225.
- Schoen, R. (1975). California divorce rates by age at first marriage and duration of first marriage. *Journal of Marriage and the Family*, 37, 548–555.
- Schoen, R. (1977). On choosing an indexing variable in demographic analysis. *Social Science Research*, 6, 246–256.
- Schoen, R. (1988). *Modeling multigroup populations*. New York: Plenum.
- Schoen, R. (2006). *Dynamic population models*. Dordrecht: Springer.
- Schoen, R. (2016). The continuing retreat from marriage: figures from marital status life tables for United States Females, 2000-2005 and 2005-2010. In R. Schoen (Ed.), *Dynamic Demographic Analysis*, (pp. 203–215). Dordrecht: Springer.
- Schoen, R. (2019). On the implications of age-specific fertility for sibships and birth spacing. In R. Schoen (Ed.), *Analytical family demography*, (pp. 201–214). Dordrecht: Springer.
- Schoen, R. (2020). Dynamic multistate models with constant cross-product ratios: Applications To poverty status. *Demography*, 57, 779–797.
- Schoen, R., & Jonsson, S. H. (2003). Estimating multistate transition rates from population distributions. *Demographic Research*, 9(29 August), 1–24.
- Schoen, R., & Standish, N. (2001). The retrenchment of marriage: results from marital status life tables for the United States, 1995. *Population and Development Review*, 27, 553–563.
- Singer, B., & Spilerman, S. (1976). The representation of social processes by Markov models. *American Journal of Sociology*, 82, 1–54.
- Smith, W.L. (1955) Regenerative stochastic processes. *Proceedings of the Royal Society of London, Series A*, 232, 6-31.
- Willekens, F., & Putter, H. (2014). Software for multistate analysis. *Demographic Research* 31, (14), 381–420.
- Wolf, D.A. 1988. The multistate life table with duration-dependence. *Mathematical Population Studies* 1: 217-245.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)
