CHAPTER 22

Event History Models for Life Course Analysis

LAWRENCE L. WU

The questions posed by life course researchers often differ in fundamental ways from those posed by sociologists, developmental psychologists, or economists (Elder, 1998; Mayer & Tuma, 1990). For example, life course researchers often focus analytic attention on transitions marking adolescence or early adulthood and the roles and statuses accompanying such transitions (Hogan & Astone, 1986; Modell, Furstenberg, & Hershberg, 1976; Shanahan, 2000). Prototypical questions along these lines include whether certain social groups experience a more rapid transition to adulthood or whether the timing of such transitions (or the duration spent in selected life course statuses) has changed for successive cohorts (Winsborough, 1980). As Mayer and Tuma (1990) note, work in this vein often implicitly conceives of social structure as arising out of individual experiences of varying duration, as opposed to alternative perspectives that see social structure in terms of collectivities of persons with particular fixed attributes (Blau, 1977), as generated from relational networks (and resulting "structural holes") among individual actors (White, Boorman, & Breiger, 1976) or from the aggregate behavior of rational actors (Becker, 1991).

Consider, for example, the linkage between early and later life course events, roles, and attitudes often posited by life course researchers (Elder, 1999).* Positing such linkages often generates a quite wide range of questions, which in turn carries important implications for the types of covariates and statistical models appropriate for life course research.

*This focus is not unique to the life course; indeed, precursors can be found in much work in the status attainment literature (see e.g., Blau & Duncan, 1967; Sewell & Hauser, 1975).

LAWRENCE L. WU • Department of Sociology, University of Wisconsin, Madison, Wisconsin 53706

Handbook of the Life Course, edited by Jeylan T. Mortimer and Michael J. Shanahan. Kluwer Academic/Plenum Publishers, New York, 2003.

Focusing on current and future trajectories of behavior has often led life course researchers to think more deeply about the various temporal and dynamic processes underlying past events, experiences, and the larger social forces shaping individual biographies and trajectories. One consequence of such a focus is that life course researchers have come to reject, by and large, the notion that the life course can be understood simply as a process of unilinear aging (Mayer & Tuma, 1990; Settersten & Mayer, 1997). Instead, life course researchers have increasingly emphasized the analytical importance of multiple dimensions of time, for example, age, duration in various statuses, and historical dimensions of time as measured by period or cohort.*

A related issue concerns interrelationships in the life course, two aspects of which are often stressed. One concerns the assertion, testable in principle, that domains such as work, marriage, childbearing, and emotional development, which are typically analyzed in isolation, in fact cannot be adequately understood without considering these domains as a unified whole. Another is the assertion, again testable in principle, concerning "linked" lives—that, for example, the events, behaviors, and outcomes experienced by one individual in a couple profoundly influences the course of events, behaviors, and outcomes experienced by a spouse.

In this chapter, I review methods relevant to life course research when event history data—that is, data on one or more discrete outcomes followed through time—are available. Several excellent monographs and textbooks (Allison, 1985; Blossfeld, Hamerle, & Mayer, 1989; Cox & Oakes, 1984; Fleming & Harrington, 1991; Hougaard, 2000; Lancaster, 1990; Tuma & Hannan, 1984; Yamaguchi, 1991) provide extensive coverage of relevant statistical models and issues. In this chapter, I provide a condensed summary of such issues, but also depart from these accounts by devoting attention to how existing models speak to (or, in some cases, do not speak to) the types of questions often posed by life course researchers.

CONCEPTUALIZING TRANSITIONS BETWEEN DISCRETE LIFE COURSE STATUSES

Event history analysis is well-suited to an analysis of life course transitions.† Indeed, the very concept of a transition is central both to research on the life course and to the conceptual and statistical modeling of event histories. For concreteness, consider Figure 22-1, which sketches two ways a researcher might conceptualize a particular life course transition for women—that of the transition to single motherhood. Panel A of Figure 22-1 presents a simple and highly stylized conceptualization of this process in which women can occupy two statuses of interest—that of being a single mother and that of not being a single mother. All women thus begin life at birth in the status labeled "1" (not a single mother), with some women subsequently transiting to the status labeled "2." Those who have become single mothers can exit this status as well, transiting back to the status labeled "1" and so forth.

*Another important line of life course research has emphasized the long historical view. The most influential work of this sort has emphasized the collective life trajectories of birth cohorts (Elder, 1999; Mayer, 1988). Other work in this vein has traced, for example, the increasing diversity of family and work experiences (Bumpass, 1990; Spain & Bianchi, 1996) and institutional influences on the increasing historical differentiation of individuals (Meyer, 1986). For an overview of statistical issues that arise in this line of research, see Glenn (this volume).

†For an alternative view, see Abbott and Tsay (2000) and the resulting commentary (Abbott, 2000; Levine, 2000; Wu, 2000).

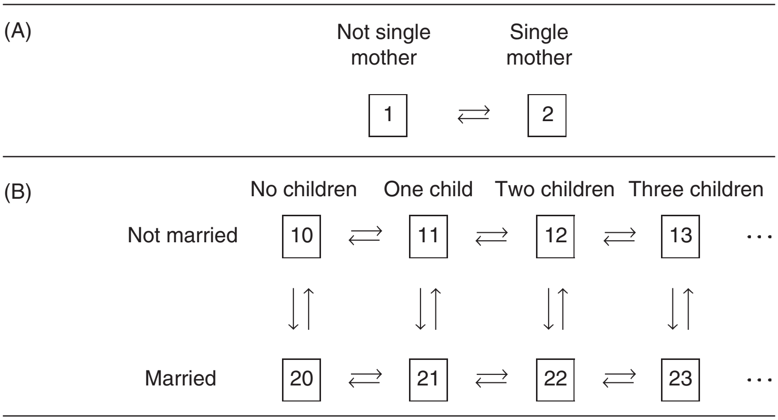

FIGURE 22-1. Two alternative "state spaces" for the transition into and out of single motherhood.

Panel B presents an alternative conceptualization of this process that cross-classifies a woman's marital status and number of children. Women begin life unmarried and with no children (status labeled "10"), and may then transit to subsequent statuses. (Child mortality is depicted in this panel by the leftward pointing arrows.) Note that there are now several possible transitions to single motherhood—the transition 10 → 11 consisting of a non-marital first birth and the transitions 21 → 11, 22 → 12, and 23 → 13 representing changes to not married statuses of married women.

Because panel B is an elaboration of panel A, the corresponding models are nested within one another, with the model in panel A equivalent to that in panel B when a variety of behavioral assumptions are imposed on the model in panel B. Equivalently, panel A can be said to "pool" across the various transitions in panel B, where, as in more standard contexts, such "pooling" assumptions can be tested empirically in ways that are formally equivalent to tests of pooling across race and ethnicity. Note that one could add further conceptual distinctions to the model in panel B—one could, for example, distinguish between the childbearing of women who are never married, divorced, or widowed, or between the childbearing of women in cohabiting and marital unions compared to the childbearing of single non-cohabitating women. Still, the point to be emphasized is the importance of considering alternative conceptualizations of the transitions of *theoretical* interest, even when not all such transitions are observed or available in the data at hand.

Another important notion is the pool of individuals at *risk* of a particular transition. In panel A, the risk set is straightforward—women at risk of a transition to single motherhood consist of those who are not single mothers, while those at risk of the other transition consist of single mothers—while in panel B, different risk sets of women are distinguished. Under either conceptualization of the process, however, it is clearly important to restrict the sample analyzed to those at risk of a particular transition, with departures from this rule of thumb undertaken only when the researcher has clear substantive or theoretical grounds to do so.

SINGLE TRANSITION FOR A HOMOGENEOUS POPULATION

I now formalize ideas starting with the simple case of a single transition for a homogeneous population, for example, age-specific mortality in a population that is assumed to be behaviorally identical.* An important issue is that the event in question (mortality in this example) will often vary substantially with time (age in this example), with this pattern of time variation often exhibiting substantial nonlinearities. For example, age variation in mortality typically follows a so-called "bathtub" pattern, in which mortality is high at young and old ages, but low during the adolescent and adult years. Accounting for such patterns of time variation is of critical importance; in particular, estimation of other quantities of interest will, in general, be biased, sometimes substantially, if the underlying pattern of time variation is not accounted for adequately.

In modeling life course transitions, the quantity of fundamental interest is the so-called *hazard rate*. Equation (1) gives three equivalent definitions of this quantity:

$$r(t) = \lim_{\Delta t \downarrow 0} \frac{\Pr(T \leq t + \Delta t | T \geq t)}{\Delta t} = \frac{f(t)}{1 - F(t)} = \frac{f(t)}{S(t)} \qquad (1)$$

where t denotes time (e.g., age), T denotes the random variable for the time of the event (age at death), and $f(t)$, $F(t)$, and $S(t) = 1 - F(t) = \Pr(T \geq t)$ denote the *density, cumulative distribution*, and so-called *survivor probability*, respectively.

The hazard rate provides the dependent variable of interest when modeling a single transition and is also a key building block for more complicated problems like that depicted in panels A and B of Figure 22-1. It is, however, not an immediately intuitive quantity. Some insight can be gained by considering its component parts. For example, note that the units for $f(t)$ and $S(t)$ are percent per unit time and percent, respectively; thus, the unit for $r(t)$ is "per unit time" as in "the age-specific first birth rate per month." Note also that because $f(t)$ and $S(t)$ are non-negative quantities, $r(t)$ is also a non-negative quantity that can assume any value between 0 and ∞. Finally note that, as is the case for a logistic regression model or for the probability p governing a coin flip, the outcome—the probability p, the log odds $\log(p/1 - p)$, or the hazard rate, $r(t)$—is not directly observed in the way a continuous outcome y_i is observed in a static linear regression model.

The quantity $\Pr(T \leq t + \Delta t | T \geq t)$ in Equation (1) gives the probability of having the event between time t and $t + \Delta t$, conditional on the event of interest not yet having occurred; hence, conditional on the population at risk, this quantity provides the probability that the event will occur between "now," as indexed by t, and some time in the future, as indexed by $t + \Delta t$. For events occuring in continuous time, it is desirable to define $\Pr(T \leq t + \Delta t | T \geq t)$ over all possible positive t; this is done via the limit in Equation (1), with the limit restricted to positive values of t so as to restrict intervals to those in the future. Combining these two parts—the limit and the conditional probability—yields the hazard rate, which is typically

*Within an event history context, homogeneity does not imply that individuals in such a population will experience the event of interest at the same time. As a rough analogy, consider a hypothetical population of coins in which the probability of heads is 0.5, where flipping 100 coins sampled from such a population will not yield 100 identical outcomes. A closer analogy is to radioactive material, in which individual atoms, even when chemically identical, will decay at different times. In the latter case, the *distribution* of event times can be shown to be exponential even though the event time for any given atom cannot be predicted.

interpreted as the "risk" of an event, where risk refers to the "instantaneous" conditional probability that the event of interest occurs at time t.

To see that the three alternative definitions in Equation (1) are equivalent, first consider the quantity $[\Pr(T < t + \Delta t | T \geq t)]/S(t)$, putting aside momentarily consideration of the limit and where for analytical convenience I assume an absolutely continuous $f(t)$. Recall from elementary probability theory that for $\Pr(B) > 0$ one can write

$$\Pr(A|B) = \frac{\Pr(A \text{ and } B)}{\Pr(B)}$$

Let $A = T < t + \Delta t$ and $B = T \geq t$, in which case

$$A \text{ and } B = \{T < t + \Delta t\} \text{ and } \{T \geq t\} = t \leq T < t + \Delta t$$

Then

$$\begin{aligned}
\Pr(T < t + \Delta t | T \geq t) &= \Pr(A|B) \\
&= \frac{\Pr(A \text{ and } B)}{\Pr(B)} \\
&= \frac{\Pr(t \leq T < t + \Delta t)}{\Pr(T \geq t)} \\
&= \frac{\Pr(t \leq T < t + \Delta t)}{S(t)}
\end{aligned} \tag{2}$$

recalling that $S(t) = 1 - F(t) = \Pr(T \geq t)$.

Note that the limit in Equation (1) is absent from Equation (2). To reintroduce the limit, recall from elementary probability theory that for an absolutely continuous density $f(t)$, one can write

$$f(t) = \lim_{\Delta t \downarrow 0} \frac{\Pr(t \leq T < t + \Delta t)}{\Delta t}$$

Then applying the limit in Equation (1) to Equation (2) yields

$$\begin{aligned}
\lim_{\Delta t \downarrow 0} \frac{\Pr(T < t + \Delta t | T \geq t)}{\Delta t} &= \lim_{\Delta t \downarrow 0} \frac{\Pr(t \leq T < t + \Delta t)}{\Delta t S(t)} \\
&= \frac{f(t)}{S(t)}
\end{aligned}$$

which establishes the equivalence between the alternative definitions in Equation (1).

A somewhat more intuitive quantity than $r(t)$ is the survivor probability $S(t)$; in population terms, this quantity for a single transition can be thought of as giving the proportion of the population that survives to time t without having experienced the event of interest. The hazard rate and survivor function are related according to

$$S(t) = \exp\left[-\int_0^t r(s)ds\right] = \exp[-H(t)] \tag{3}$$

where the so-called *integrated hazard* is given by

$$H(t) = \int_0^t r(s)\mathrm{d}s \tag{4}$$

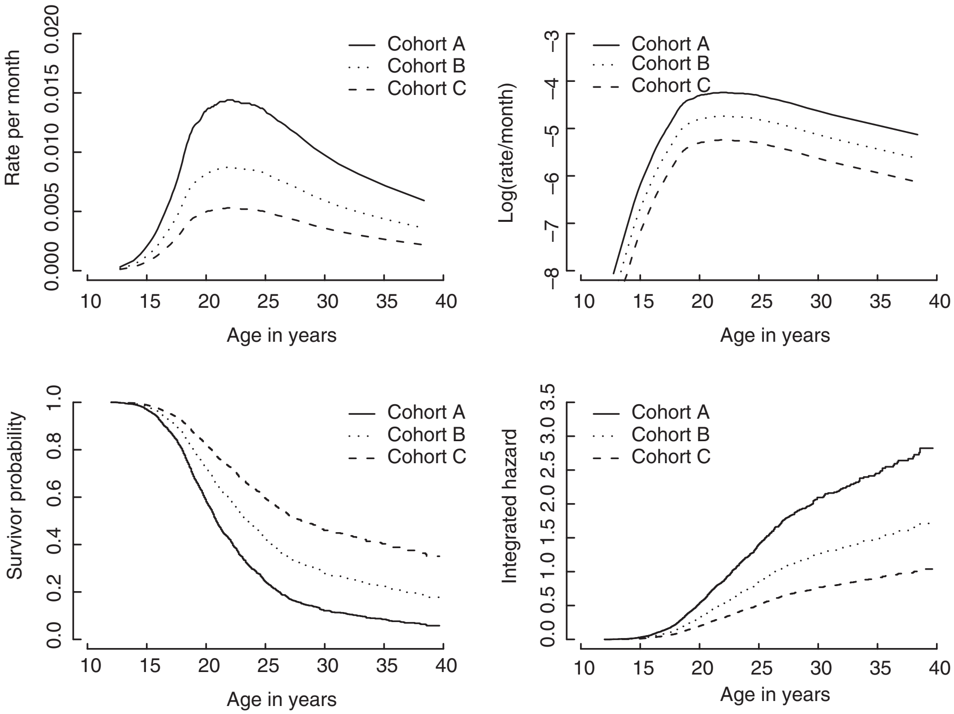

FIGURE 22-2. Hypothetical examples of a proportional decline in the Hazard rate of first marriage for three successive birth cohorts. Implications for $S(t)$, $H(t)$, $r(t)$, and log $r(t)$.

To gain further insight into the quantities $S(t)$, $H(t)$, $r(t)$, and $\log r(t)$, consider some hypothetical data on age at first marriage for three successive birth cohorts as presented in Figure 22-2. In this example, I constructed the curves for $r(t)$ from those for $\log r(t)$, with the three curves for $\log r(t)$ shifted vertically from one another by an additive constant:

$$\log r_A(t) = \log r_B(t) + c = \log r_C(t) + 2c$$

The two upper panels present graphs of $r(t)$ and $\log r(t)$ for these three cohorts, with the solid curves in both panels lying uniformly above the dotted and dashed curves and the dashed curve lying uniformly below the dotted curve. Exponentiating the above shows that the corresponding relationships for $r(t)$ take a multiplicative form, with

$$r_A(t) = r_B(t) \times \exp(c) = r_C(t) \times \exp(2c)$$

This is an example of the widely used *proportional hazard specification*, in which plots of $\log r(t)$ are parallel or in which the ratio of hazard rates is a constant that does not vary with t:

$$\frac{r_A(t)}{r_B(t)} = \exp(c), \quad \frac{r_A(t)}{r_C(t)} = \exp(2c), \quad \text{and} \quad \frac{r_B(t)}{r_C(t)} = \exp(c)$$

The quantities $\exp(c)$ and $\exp(2c)$ can be interpreted as relative risks. For example, in constructing Figure 22-2, I took $c = 0.5$; thus, for this hypothetical example, the age-specific "risk" of first marriage is $\exp(+0.5) = 1.65$ or 65% higher for the cohort *A* (solid curve), relative to that for cohort *B* (dotted curve), while the age-specific risk of first marriage for cohort *C* (dashed curve) is $\exp(-0.5) = 0.61$ or 39% lower relative to cohort *B* (dotted curve) and $\exp(-1.0) = 0.37$ or 63% lower relative to cohort *A* (solid curve).

The bottom two panels of Figure 22-2 plot the corresponding survival probabilities and integrated hazard functions. Note in particular the inverse relationship between $r(t)$ and $S(t)$, with higher rates corresponding to lower survival probabilities. Similarly, because the solid curve for $r(t)$ lies uniformly above the dotted curve for $r(t)$, it then follows that the solid curve for $S(t)$ will lie uniformly below the dotted curve for $S(t)$. In the context of age at first marriage, then, cohort *A* experiences uniformly higher age-specific rates of entry into first marriage than cohorts *B* or *C* (upper panels for $r(t)$ and $\log r(t)$); likewise, the proportions remaining never-married (lower left-hand panel) are uniformly lower in cohort *A* than in cohorts *B* or *C*. Similarly, by age 40, 5.6, 17.4, and 34.6% of individuals in cohorts *A*, *B*, and *C*, respectively, remain single and never-married. Thus, if cohorts *A*, *B*, and *C* represented successive birth cohorts of women, these results would indicate both a delay in age at first marriage and a greater propensity to forgo entry into first marriage for successive birth cohorts of women.

The four panels of Figure 22-2 illustrate how graphical plots of $r(t)$, $\log r(t)$, and $S(t)$ can convey useful information. For example, the two upper panels of Figure 22-2 reveal age-graded differences in first marriage. The quantity $S(t)$ similarly provides information on how the proportion of the population remaining single and never-married, that is surviving in the origin state, varies with age. Similar analyses can be conducted for successive birth cohorts, by race and ethnicity, or for cross-classifications by other characteristics.

Although the proportional hazard model is heavily used, many of the questions posed by life course researchers in fact imply violations of proportionality. For example, there has been considerable debate on whether the behavior of successive cohorts of women is best understood as a "retreat" from marriage (e.g., see Gilder, 1986; Popenoe, 1996) or whether observed behaviors instead reflect delayed but eventual entry from marriage (Oppenheimer, 1997). The important point is that a proportional specification carries strong assumptions concerning this question—if invoked for successive cohorts of women, it implies both delay and retreat. The difficulty is that, while retreat logically implies delay, delay need not logically imply retreat; hence, invoking proportionality in fact is equivalent to staking a strong a priori position on such an issue. Figure 22-3 presents a hypothetical example in which retreat—in the sense of increases in the proportions who never marry—need not follow from delayed marriage. In the two upper panels of Figure 22-3, the curves for $r(t)$ and $\log r(t)$ are shifted horizontally, with the peak age at entry into first marriage occurring at ages 18, 20, and 22 for cohorts *A*, *B*, and *C*, respectively. The lower left-hand panel traces the consequences by age in the proportions who remain single and never-married. The differences in these proportions are large between 15 and 30 but narrow substantially at later ages. Under this hypothetical marriage regime, then, there is no change in the proportion of women who never marry across successive cohorts, despite marked delays in entry into marriage. Thus in Figure 22-3, delay does not imply retreat; in Figure 22-3, both delay and retreat occur.

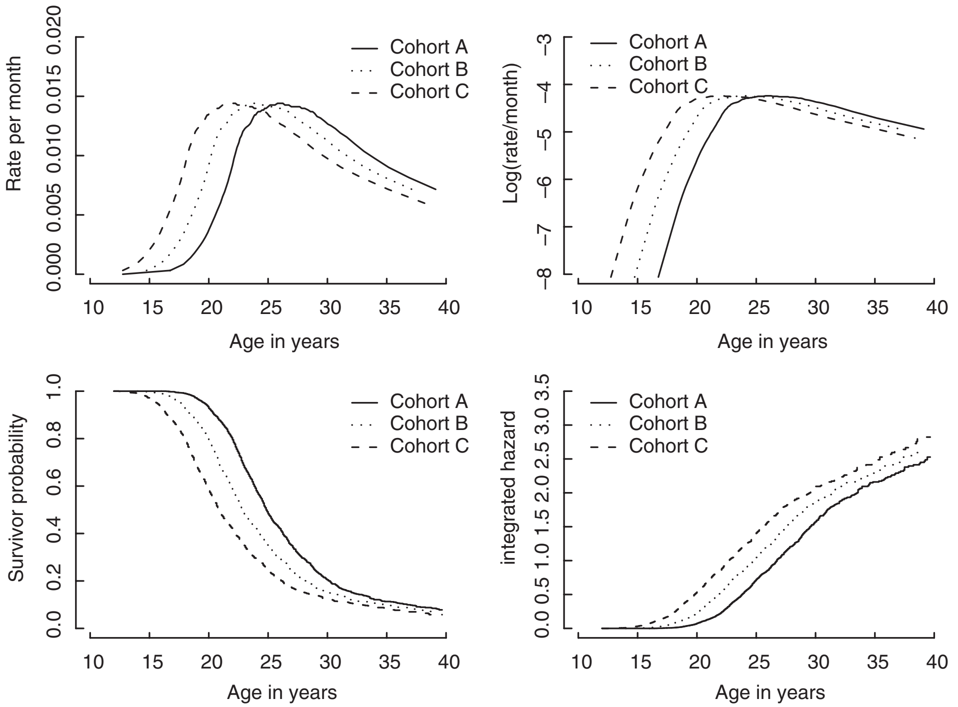

FIGURE 22-3. Hypothetical example of a "pure" delay in the hazard rate of first marriage for three successive birth cohorts. Implications for $S(t)$, $H(t)$, $r(t)$, and log $r(t)$.

NONPARAMETRIC ANALYSES OF A SINGLE TRANSITION

As Figures 22-2 and 22-3 suggest, much can be learned from exploratory analyses that permit the visual inspection of quantities such as $S(t)$, $H(t)$, $r(t)$, and log $r(t)$. Such analyses typically make heavy use of *nonparametric estimators* of quantities such as the survival probability (Kaplan & Meier, 1958), integrated hazard (Aalen, 1978; Nelson, 1972), and hazard rate (Cox & Oakes, 1984), where "nonparametric" refers to the lack of strong distributional assumptions concerning the distribution of event times, that is the shape of the hazard rate with respect to t.

As in the previous section, I begin by considering the case of a single transition in a homogeneous population. Let T_i denote the random variable for individual i's time at the event of interest. Not all individuals may experience the event by the time of last interview, in which case these individuals are then said to be *right censored*. As a result, representing data on the outcome requires a pair of variables, (t_i, δ_i), where t_i is the realization of the random variable T_i and $\delta_i = 1$ if the event is observed for individual i and 0 if the outcome for i is right-censored.

The Kaplan–Meier (1958) estimator of the survival probability, which can be shown to be the nonparametric maximum-likelihood estimator for this quantity, is given by

$$S_{\mathrm{KM}}(t) = \prod_{\mathrm{R}(s):s \le t} \left[1 - \frac{\mathrm{d}(s)}{\#R(s)} \right] \tag{5}$$

where the product is taken over all individuals at risk of the event at time t, $\mathrm{d}(t)$ denotes the number of individuals with events at time t, $\mathrm{R}(t)$ is the set of individuals at risk of the event at time t, and $\#\mathrm{R}(t)$ denotes the number of individuals at risk at time t. The Nelson–Aalen estimator

(Aalen, 1978; Nelson, 1972), which can be shown to be the nonparametric maximum-likelihood estimator for the integrated hazard, is given by

$$H_{\mathrm{NA}}(t) = \sum_{\mathrm{R}(s):s \le t} \frac{\mathrm{d}(s)}{\#\mathrm{R}(s)} \tag{6}$$

For technical statistical reasons, no sensible nonparametric maximum-likelihood estimator of the hazard rate exists. Still, one can obtain a nonparametric estimator of the hazard rate that possesses good properties using the classic demographic life table or methods for nonparametric density estimation (e.g., see Allison, 1995; Cox & Oakes, 1984; Wu, 1989).

Figure 22-4 presents nonparametric estimates of the two most intuitive quantities $S(t)$ and $r(t)$ for age at first marriage for White and Black women in the 1980 Current Population

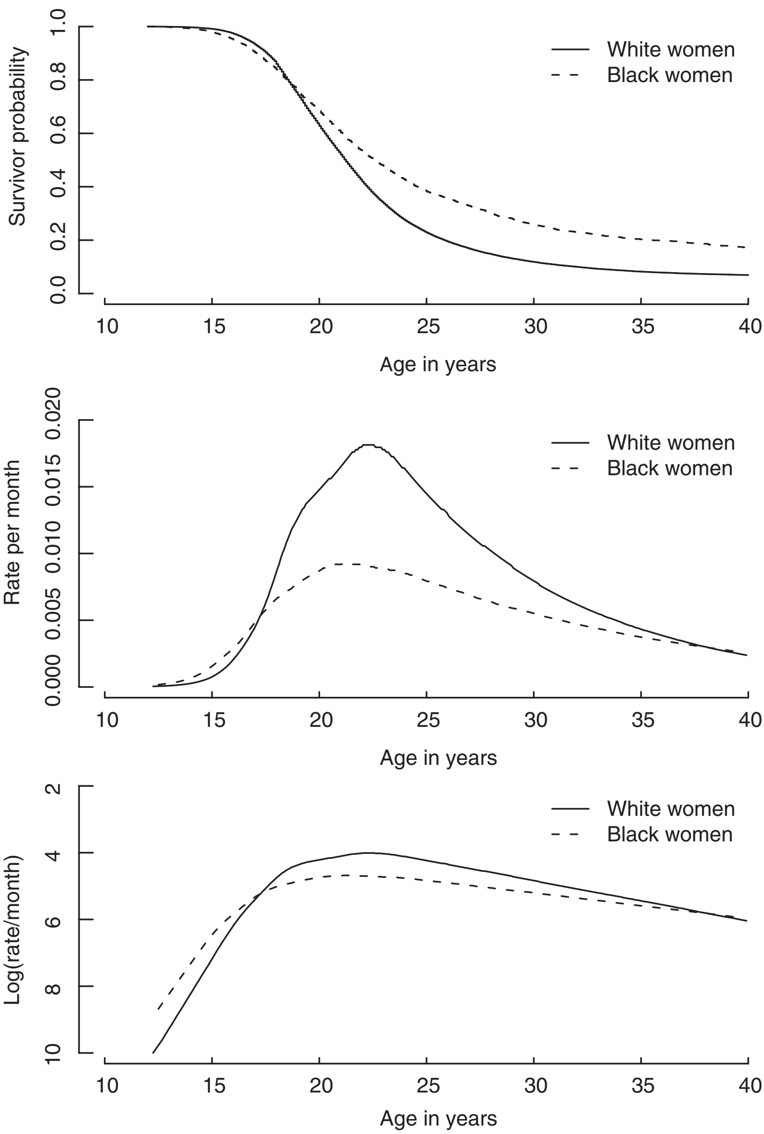

FIGURE 22-4. Non-parametric estimates of $S(t)$, $H(t)$, and $r(t)$, for White and Black women, 1980 Current Population Survey.

Survey (CPS). The upper panel, which presents Kaplan–Meier estimates of $S(t)$, shows that $S_{KM}(t)$ yields a monotonically declining step function, with values varying between 1 and 0. For these cohorts of women, the curves decline steeply between ages 20 and 40; roughly 64% of White women were never married by age 20, with about 7% never married by age 40. By contrast, about 69 and 17% of Black women were never married by ages 20 and 40, respectively. The middle panel of Figure 22-4, which presents smoothed nonparametric estimates of $r(t)$, shows that Black/White differences in the transition to first marriage are indeed most pronounced between ages 20 and 40. The lower panel, which presents smoothed nonparametric estimates of $r(t)$, shows that Black/White differences are not constant with age, but rather widen during early and middle adulthood, and then narrow at later ages. This panel makes clear that proportionality appears to be violated for White and Black patterns in age at first marriage; note also that Black/White first marriage rates do not appear to follow a pattern of "pure" delay such as that exhibited in Figure 22-3.

These examples also illustrate the prototypical steps in how one might conduct exploratory analyses of event history data. In particular, observed characteristics of individuals can be used in conjunction with the nonparametric methods outlined above, for example, by classifying individuals using observed values of a discrete covariate or by discretizing the values of a metric covariate, followed by visual inspection of graphical displays such as those in Figure 22-4, with such analyses providing, roughly speaking, an event history equivalent to exploratory methods using cross-tabulations or comparisons of group means.

PARAMETRIC MODELS FOR A SINGLE TRANSITION

To this point, our discussion has focused on a single transition for a homogeneous population; thus, we have proceeded by and large without consideration of right-hand side covariates. In this section, we consider how one might obtain estimates of the association of observed covariates with the hazard rate $r(t)$. To simplify matters, we continue to focus on a single transition.

The most common parametric specification for covariates is the so-called *proportional hazard* model given by

$$r[t|\mathbf{x}_i(t)] = q(t)\exp[\mathbf{b}\mathbf{x}_i(t)] \tag{7}$$

or

$$\log r[t|\mathbf{x}_i(t)] = \log q(t) + \mathbf{b}\mathbf{x}_i(t) \tag{8}$$

where i indexes individuals and $\mathbf{x}(t)$ denotes a vector of observed covariates, some of which may be time varying. A key assumption in Equations (7) and (8) is that time variation in $\log r(t)$ is captured by a single function, the so-called baseline hazard $\log q(t)$, with all additional heterogeneity across individuals captured by the additive effects of the covariates $\mathbf{x}_i(t)$.

The types of covariates allowed by Equations (7) and (8) span a remarkably wide range; indeed, one can include any aspect of an individual's history up to time t (Aalen, 1978; Tuma & Hannan, 1984). Thus, event history models let life course researchers examine an unusually rich set of covariates in ways that capture changes in the social and economic contexts of individuals that might influence the outcome of interest. One can, for example, investigate the effect of earlier life events, current social circumstances, cumulative experiences, and exposure to particular statuses, which allows researchers to contrast effects of past, current, and cumulative experience on current behavior (Wu & Martinson, 1993; Wu & Thomson, 2001). See also Wu (1996), who compared three alternative effects of income in an individual's family of origin—a simple measure of income at time *t*, permanent and transitory measures of income, and measures of income level and income change. One can also incorporate future predictions (e.g., projected income, marriage market opportunities, or characteristics of potential mates; e.g., see Dechter, 1992) when the predicted values of such covariates are obtained from models using current and past covariate information for a given individual. As Mayer and Tuma (1990) noted, this flexibility meshes well with a central presupposition of many life course researchers, which is that an individual's social context can vary considerably over time in ways rarely reflective of some stable social equilibrium. If so, then a central analytical task is to capture central features of this variation over time for individuals.

Several measurement issues arise when incorporating time-varying covariates into Equations (7) and (8). For example, if a time-varying covariate $x_i(t)$ is discrete and gathered as an event history, the analyst can determine the value of $x_i(t)$ at all observed times t during which person i is at risk. In other circumstances, the value of $x_i(t)$ may be difficult to determine at all possible times. Examples include covariates such as income, work hours, expenditures, depression scales, or attitudes that can vary from one moment to another but which will often be measured only sporadically. In such cases, a commonly invoked and analytically convenient assumption is that the value of such a covariate is constant between measurements.*

The popularity of proportional hazard models stems in part from practical and theoretical considerations. The asymptotic properties of these models are well understood under a variety of conditions (e.g., Cox & Oakes, 1984; Fleming & Harrington, 1991), including quite general conditions on the distribution of censoring times. Furthermore, under proportionality, covariates have linear effects on the logarithm of $r(t)$; hence, intuitions from ordinary and logistic regression carry over in a straightforward way to the proportional hazard model. Empirically, proportionality is often adequate in that estimated coefficients under proportionality are often similar to those when proportionality is relaxed; this can often hold when the observation period is short relative to the mechanisms that generate variation over time in the effect of a covariate.

Proportionality is nevertheless a strong assumption and violations can occur empirically. As noted above, Figure 22-4 provides an empirical example where proportionality appears suspect; indeed, visual displays like those in the lower panel of Figure 22-4 provide a useful exploratory way to check proportionality. Fortunately, the models in Equations (7) and (8) can be generalized to incorporate non-proportional effects of covariates; indeed, by appropriate definition of time-varying covariates, one can adopt standard software to estimate certain non-proportional models. Recall that when proportionality is violated, the effect of a covariate on the logarithm of the hazard rate will not be an additive constant, but rather will vary with t. Thus, one way to relax the proportionality assumption is to code a set of time-varying dummy variables that represent an exhaustive and mutually exclusive partition of the observation period. Interacting these dummy variables with a covariate x hypothesized to have a non-proportional effect then yields a specification that models the effect of x as a step function of time. Other approaches suggested in the literature include a log multiplicative specification (Xie, 1994), piecewise linear splines (Wu & Martinson, 1993), and a local likelihood approach that makes few assumptions about the form of time variation in covariate effects (Wu & Tuma, 1990).

*For alternative specifications relaxing this constancy assumption, see, for example, Tuma and Hannan (1984). Note also that the plausibility of such an assumption will depend not only on the number and frequency of measurements, but also on the temporal variation in the covariate relative to the frequency of measurements.

In practice, assumptions concerning the baseline hazard $q(t)$ are typically of greater concern than possible non-proportionalities in the effects of covariates; very roughly speaking, this often occurs when the baseline hazard accounts for a greater proportion of the observed variation in $r(t)$ than do the covariates $\mathbf{x}(t)$, as is often true in practice. A popular choice is a model due to Cox (1972), in which the baseline hazard $q(t)$ is allowed to be an unknown and unspecified function of time. Under this model, maximum likelihood estimation is not possible, but the method of partial likelihood (Cox, 1975) can be shown to yield consistent and asymptotically efficient estimates of the parameters $\mathbf{b}$ under quite general conditions, including mild assumptions concerning the distribution of censoring (Andersen & Gill, 1982).

While the Cox model yields estimates of the parameters $\mathbf{b}$ for the effects of covariates, it does not provide any direct estimate of the baseline hazard function $q(t)$. Knowledge of $q(t)$ is often unnecessary for a number of important analytical purposes, for example comparisons of control and treatment groups in medical trials evaluating the efficacy of a new drug or treatment in which the outcome is mortality from a specific form of cancer. However, many of the questions routinely posed by life course researchers require knowledge of $q(t)$; examples include comparisons across appropriately defined cohorts of individuals in the specific pace of childbearing, the timing of entry into marriage, or median time to divorce. Although estimates of $q(t)$ can be recovered in the Cox model, obtaining inferences about $q(t)$ under the Cox model is more difficult. As a result, it can be useful to consider parametric alternatives to the Cox model.

Various parameterizations for $q(t)$ have been proposed, including the exponential, Weibull, Gompertz, Makeham, log logistic, log Gaussian, Hernes, sickle, and Coale–McNeil models (Blossfeld et al., 1989; Tuma & Hannan, 1984; Wu, 1990). Sometimes theory provides grounds to motivate a particular choice, but more often practical considerations (e.g., software availability) underlie these choices. Unfortunately, estimated effects of covariates can sometimes vary considerably across different functional forms, complicating matters for the analyst.

One reason for this sensitivity is that the models vary markedly in their specification of time variation in the baseline hazard. For example, some models assume that the baseline hazard increases or decreases monotonically (e.g., the Weibull, Gompertz, and Makeham models), while other models assume a unimodal shape for the baseline hazard, with the rate rising and then declining (e.g., the log logistic, log Gaussian, Coale–McNeil, Hernes, and sickle models). In addition, some models yield a distribution of event times that integrates to unity, implying that all individuals will experience the event of interest if observed for a sufficiently long period of time (e.g., the exponential, Weibull, Makeham, log logistic, and log Gaussian models), while other models can, in some cases, yield a so-called defective distribution in which some individuals will never experience the event of interest, even if observed for an arbitrarily long time. Note that a defective distribution of event times is often substantively plausible; examples include marriage or sexual initiation in which some individuals, for example, those who have taken vows of celibacy or chastity, may never marry; parity-specific fertility, in which some individuals may never proceed to, say, a fifth birth; and residential moves, where some individuals may live all their lives in the residence in which they were born. Similarly, there are other instances in which a non-defective distribution of event times is desirable, with a classic example being human mortality.

As a result, many researchers use models that mimic the Cox model in the sense of providing a flexible functional form for the baseline hazard. One popular and easily implemented parametric alternative to the Cox model specifies the baseline hazard as a piecewise constant function, that is as a step function of t. More formally, consider P time intervals, $(0, \tau_1]$,

$(\tau_1, \tau_2], \ldots, (\tau_{P-2}, \tau_{P-1}], (\tau_{P-1}, \infty]$, where the τ_P are prespecified by the analyst; then let $q(t)$ be defined by a series of constants on these intervals, that is

$$q(t) = \begin{cases} \exp(\lambda_1), & t \in (0, \tau_1], \\ \exp(\lambda_2), & t \in (\tau_1, \tau_2], \\ \cdots & \\ \exp(\lambda_P), & t \in (\tau_{P-1}, \infty] \end{cases} \tag{9}$$

Equivalently,

$$\log q(t) = \begin{cases} \lambda_1, & t \in (0, \tau_1], \\ \lambda_2, & t \in (\tau_1, \tau_2], \\ \cdots & \\ \lambda_P, & t \in (\tau_{P-1}, \infty] \end{cases} \tag{10}$$

The resulting piecewise exponential model for the baseline hazard is easily implemented, for example, by defining P time-varying dummy variables corresponding to the P time intervals.*

A slight variant of the above lets log $r(t)$ vary linearly within intervals, with the linear segments splined to yield a continuous function:

$$\log q(t) = \begin{cases} \lambda_1 + \gamma_1 t, & t \in (0, \tau_1], \\ \lambda_2 + \gamma_2 t, & t \in (\tau_1, \tau_2], \\ \cdots & \\ \lambda_P + \gamma_P t, & t \in (\tau_{P-1}, \infty] \end{cases}. \tag{11}$$

subject to the $P - 1$ equality constraints

$$\begin{aligned} \lambda_1 + \gamma_1\tau_1 &= \lambda_2 + \gamma_2\tau_1 \\ \lambda_2 + \gamma_2\tau_2 &= \lambda_3 + \gamma_3\tau_2 \\ &\cdots \\ \lambda_{P-1} + \gamma_{P-1}\tau_{P-1} &= \lambda_P + \gamma_P\tau_{P-1} \end{aligned} \tag{12}$$

yielding the so-called splined piecewise Gompertz model for the baseline hazard.

Both the piecewise constant function in Equation (11) and the piecewise linear spline in Equation (12) provide very flexible specifications for log $q(t)$. Note, for example, that both can accommodate a variety of shapes for the baseline hazard, including monotonically increasing, monotonically decreasing, unimodal, or multimodal patterns of time variation. Note that, given P prespecified time intervals, the piecewise constant specification in Equation (11) uses P degrees of freedom, while the piecewise linear spline in Equation (12) uses $P + 1$ degrees of freedom.†

Table 22-1 presents estimates from three proportional hazard models for age at entry into first marriage using data on White women from the 1980 CPS. We contrast estimates from a

*Note that Equations (7) and (8) lack a time-invariant constant term—what would be the intercept in a linear regression. One can retain such a constant term by omitting one of the P intervals from estimation, in which case the estimates λ_p and $\exp(\lambda_p)$ provide contrasts with respect to the overall constant term.

†The piecewise linear spline uses $2P$ parameters but then invokes $P - 1$ equality constraints, thus requiring $2P - (P - 1) = P + 1$ free parameters.

Cox model, a piecewise constant, and piecewise linear spline model; for the latter two models, we have specified three time intervals, corresponding to ages 12 and 18, 18 and 25, and 25 and older. Note that these data provide large samples but relatively few covariates; hence, Table 22-1 reports estimates only for respondent's years of schooling completed by time of interview (discretized into 0–11, 13–15, and 16 or higher, with 12 years the omitted category) and year of birth (1930–1949 and 1950 or later, with 1929 or earlier the omitted category). Estimates agree closely across the three models, with estimates from the Cox and three-period linear spline model agreeing particularly closely. For the Cox and three-period linear spline models, White women with 11 or fewer years of schooling completed have a 28% ($\exp(0.25) = 1.28$) higher rate of first marriage than White women with 12 years of schooling, while White women with some college education have a 30% ($\exp(-0.35) = 0.70$) lower rate of first marriage than White women with 12 years of schooling.

Results from these models also show that the piecewise linear spline specification for $q(t)$ provides a substantially better fit to the data than the piecewise constant specification. Table 22-2 compares log likelihood values corresponding to four models for these data using a piecewise constant baseline with and without covariates and a piecewise linear spline baseline with and without covariates. The lower panel of Table 22-2 provides two sets of χ^2 comparisons that provide tests of adding the five covariates in Table 22-1 and for modeling log $r(t)$ using three constants versus three linear splines. Adding covariates yields a χ^2 increment in fit of around 4,300, while allowing log $r(t)$ to vary linearly yields a χ^2 increment in fit around 13,500. This is a typical result when analyzing life course data, with careful modeling of the baseline hazard often yielding much more substantial improvements in fit than the introduction of covariates. Intuitively, such large increments in fit are often observed because typical life course outcomes exhibit substantial within-individual time variation, with this time variation often greater than the variation observed across individuals. As a consequence, it is generally advisable to devote careful modeling attention to the form of time variation in $q(t)$ to ensure that conclusions about other parameters are not biased by incorrect or inappropriate assumptions about time dependence in $q(t)$.

Table 22-1. Comparison of Covariate Effects for Cox, Three-period Exponential, and Three-period Splined Gompertz Models: Age at First Marriage, White Women, 1980 Current Population Survey

	Cox	Three-period exponential	Three-period Gompertz
Years of schooling completed			
0–11	0.25***	0.18***	0.25***
	(0.01)	(0.01)	(0.01)
13–15	−0.35***	−0.35***	−0.35***
	(0.02)	(0.02)	(0.02)
16+	−0.71***	−0.68***	−0.70***
	(0.02)	(0.02)	(0.02)
Year of birth			
1930–1949	0.48***	0.54***	0.48***
	(0.01)	(0.01)	(0.01)
1950+	0.16***	0.16***	0.17***
	(0.01)	(0.01)	(0.01)

Note: Standard errors in parentheses. See text for additional details.
*$p < 0.01$, **$p < 0.001$, ***$p < 0.0001$ (two-tailed test).

TABLE 22-2. Selected Comparisons of Model Fit for the Three-period Exponential and Three-period Splined Gompertz Models: Age at First Marriage, White Women, 1980 Current Population Survey

Model	Model for baseline hazard	Covariates?	log £
Panel A: Model description and statistics			
1	Three-period constant	No	−214,452.8
2	Three-period constant	Yes	−212,293.0
3	Three-period linear spline	No	−207,678.2
4	Three-period linear spline	Yes	−205,490.1
Comparison	Test for	df	χ^2
Panel B: Model comparisons			
1 versus 2	Adding covariates	5	4,319.6
3 versus 4	Adding covariates	5	4,376.2
1 versus 3	Three-period constant versus linear spline	1	13,549.2
2 versus 4	Three-period constant versus linear spline	1	13,605.8

The proportional models in Equations (7)–(12) can be easily extended to accommodate multiple dimensions of time. Consider, for example, the transition between a first and second child as depicted in panel B of Figure 22-1. Empirically, the second birth rate varies less with the age of a woman than with the duration since first birth, yet age variation in the second birth typically cannot be ignored. Because of examples like this, analysts often have sound reasons to extend Equations (7)–(12) to multiple dimensions of time. Let t denote age and u the duration since a first birth; then a straightforward extension of the proportional hazard model in Equation (7) is

$$r[t, u|\mathbf{x}_i(t)] = q_1(t)q_2(u)\exp[\mathbf{b}\mathbf{x}_i(t)] \tag{13}$$

or, equivalently,

$$\log r[t, u|\mathbf{x}_i(t)] = \log q_1(t) + \log q_2(u) + \mathbf{b}\mathbf{x}_i(t) \tag{14}$$

Note that Equations (13) and (14) yield an age- and duration-specific model of second births under the assumption that the second birth rate is separable into two components, $q_1(t)$ and $q_2(u)$ (Lillard, 1993; Wu & Martinson, 1993). As for a single time dimension, one can use numerous parameterizations for $q_1(t)$ and $q_2(u)$, including those in Equations (10) or (11).

It has long been recognized that linear regression models can be unidentified when controlling for multiple dimensions of time, with the classic example being simultaneous linear terms for age, period, and cohort (Glenn, this volume). Although t and u will co-vary in strong ways for each individual in the sample, identification in Equations (13) and (14) is often possible because $q_1(t)$ and $q_2(u)$ are typically highly non-linear functions of t and u, with these non-linearities helping to identify model parameters. For example, Wu and Martinson (1993) presented models that control duration, age, period, and cohort; however, identification will in general become increasingly problematic as the number of time dimensions or parameters used to model the baseline functions increases. Given these issues, one sensible procedure is to identify, theoretically or empirically, those time dimensions which induce the greatest variation in the underlying hazard rate. One can then invest greater modeling effort for the "primary" time dimensions, for example using the specification in Equation (11) and less effort for "secondary" time dimensions, for example by using the specifications in Equations (10) or (11) coupled with relatively widely-spaced intervals.

A wide class of diagnostics for the above models can be obtained using so-called martingale residuals (e.g., see Fleming & Harrington, 1991, Chapter 4). Graphical displays of such residuals can be used to assess the influence of particular observations and to check assumptions concerning proportionality and the functional form of covariate effects (i.e., are effects linear in x or in log x) in ways analogous to diagnostic residual displays in linear regression.

MULTIPLE ORIGIN AND DESTINATION STATES

Thus far, we have formalized issues for a single transition, but the models discussed above generalize in straightforward ways to more complicated processes such as those depicted in panel B of Figure 22-1. To simplify details, let us return to a single homogeneous population but generalize the above to multiple origin and destination states. Let j and k index the origin and destination states, respectively; then let the transition rate $r_{jk}(t)$ be defined as

$$r_{jk}(t) = \lim_{\Delta t \downarrow 0} \frac{\Pr(T_{jk} \leq t + \Delta t | T_{jk} \geq t)}{\Delta t} = \frac{f_{jk}(t)}{1 - F_{jk}(t)} = \frac{f_{jk}(t)}{S_{jk}(t)} \qquad (15)$$

Thus, Equation (15) generalizes Equation (1) by representing each transition $j \rightarrow k$ by a unique transition rate $r_{jk}(t)$.

The generalization of the survivor probability to multiple origins and destinations involves some subtle but important shifts. Returning momentarily to the case of a single event, recall that the survivor probability has two equivalent interpretations: (1) as the probability of not yet having experienced the event of interest and (2) as the probability of remaining in the origin state. When multiple destination states exist, (1) and (2) will, in general, differ; in addition, the interpretation of (1) is complicated by an identifiability issue when so-called competing risks are present.

To make issues concrete, consider panel A of Figure 22-1, in which there is only one transition out of each origin status. In this case, matters reduce to the case for a single transition, conditional on status at origin. By contrast, in panel B of Figure 22-1, each origin state has multiple destination states, for example from the origin state 01, there are two possible transitions, 01 $\rightarrow$ 11 and 01 $\rightarrow$ 20, while from the origin state 21, there are three possible transitions, 21 $\rightarrow$ 11, 21 $\rightarrow$ 20, and 21 $\rightarrow$ 22. When individuals in an origin state are subject to multiple destination states, they are said to be subject to competing risks. When competing risks are present, it can be shown that the interpretation given in (1) is unaffected but that in (2) must be modified in ways detailed below.

To formalize these ideas, suppose that individuals observed in origin state j are subject to multiple destination states, indexed by $k = 1, \ldots, K_j$, where $K_j > 1$. Let $r_{jk}(t)$ denote the K_j transition rates corresponding to each of these transitions; then the probability of surviving to time t in state j, is given by

$$S_j(t) = \exp\left[- \int_0^t \sum_{k=1}^{K_j} r_{jk}(s) \mathrm{d}s \right] = \exp\left[- \sum_{k=1}^{K_j} \int_0^t r_{jk}(s) \mathrm{d}s \right] = \prod_{k=1}^{K_j} \exp\left[- \int_0^t r_{jk}(s) \mathrm{d}s \right] \qquad (16)$$

Thus, the second interpretation of $S_j(t)$—the probability of surviving in origin state j—is identical to that for a single transition, except that $S_j(t)$ conditions on origin state j—that is, it refers to the survival to time t of those individuals who have not exited the origin state j by time t.

Now consider the probability in (1)—that of not having experienced the event of interest. When competing risks are present, there is not a single event but rather multiple events that must be considered when accounting for exits from an origin state j. Consider a classic example in which age-specific mortality in humans is classified by cause of death—for example, deaths due to (1) cardiovascular disease, (2) cancer, (3) homicides and other acts of violence, (4) accidents, and (5) a residual category for all other causes of death. Distinguishing between multiple types of events gives rise to additional complications; in particular, the interpretation of $S_{jk}(t)$ and $r_{jk}(t)$ will in general differ from more familiar quantities such as the proportion in the population observed to experience the event k conditional on origin state j.

To see this, consider two birth cohorts followed until death and suppose that cohort members are identical in all respects except that deaths due to cardiovascular disease have been eliminated in cohort A but not in cohort B. It then follows that, in cohort A, if one cause of death is eliminated, the proportion in A experiencing other causes of death will necessarily increase. Let p_k^A and p_k^B denote the proportion of deaths of type k that occur in cohorts A and B; then if cohorts A and B are identical save for $p_{A1} = 0$, it will nevertheless follow that $p_k^A \geq p_k^B$ for all remaining causes of death. The researcher only possessing estimates of p_k^A and p_k^B will observe that $p_{Ak} \geq p_{Bk}$ for $k \neq 1$ and, hence, might be tempted to conclude that mortality in A and B differ fundamentally, with mortality from causes other than cardiovascular disease systematically higher in cohort A than in B. Yet by construction, the two cohorts have identical mortality risks save for the elimination of deaths from cardiovascular disease in cohort A. This apparent paradox would be avoided were comparisons based on the quantities $r_k(t)$ or $S_k(t)$. Thus, sufficiently large samples would reveal that $r_1^A(t) = 0$ while $r_1^B(t) > 0$—mortality from cardiovascular disease is eliminated in cohort A but not B—but that mortality from other causes is otherwise identical, that is, that $r_k^A(t) = r_k^B(t)$ for $k \neq 1$.

Turning this example on its head makes it clear that, under competing risks, the $S_{jk}(t)$ cannot be interpreted as if they provided the proportions of those in origin state j who experience the event k, a statement that holds even when censoring is absent. Rather, the $S_{jk}(t)$ should be interpreted as giving the proportion in an origin state who would have experienced the kth transition were all other competing transitions to be eliminated. Note, moreover, that the plausibility of this interpretation rests heavily on this independence assumption.* Violations of this assumption would include situations in which, say, those who are observed to die of one chronic condition—for example, cardiovascular disease—differ systematically in other ways that lead them to have higher (or lower) risk of another chronic condition. A difficulty is that the independence assumption under competing risks has been shown to be non-identifiable in the sense that one cannot obtain formal tests of this assumption (Tsiatis, 1975).†

*When covariates are included in the model, this assumption becomes one of conditional independence, that is, the competing risks are assumed to be independent conditional on the covariates $\mathbf{x}(t)$. This assumption is similar to the so-called "irrelevance of alternatives" in a multinomial logistic regression. It is possible to state conditions that are slightly weaker than full independence for the competing risk model; however, such conditions carry little practical import (see Cox & Oakes, 1984, for details).

†Consider the promotion of assistant professors in an academic department. If some junior faculty depart in anticipation of non-promotion, then simply distinguishing between two sorts of events (promotion vs. departure) will not correct the upward bias in estimates of tenure rates. Allison (1995) suggest a simple procedure to assess the sensitivity of estimates to such a possibility. Consider two situations, one in which all departing junior faculty would have in fact been fired at a time ϵ after they are observed to have departed and another in which all departing junior faculty would have been promoted had they remained. Estimates under these two behavioral extremes can be used to construct Manski-type bounds for the usual naive estimate (Manski 1995).

PARAMETRIC MODELS FOR MULTIPLE TRANSITIONS

Construction of parametric models for the $r_{jk}(t)$ in Equation (15) is straightforward, with the underlying issues similar to those for modeling a single transition. For example, a proportional hazard specification incorporating covariates will be given by

$$r_{jk}[t|\mathbf{x}_{jk}(t)] = q_{jk}(t)\exp[\mathbf{b}_{jk}\mathbf{x}_{jk}(t)] \tag{17}$$

where $q_{jk}(t)$ denotes the baseline for the transition from state j to k and $\mathbf{x}_{jk}(t)$ denotes a vector of (possibly time-varying) covariates. The jk subscript on $\mathbf{x}$ emphasizes that one can specify different covariates across transitions and that the effects themselves will differ across transitions. In practice, however, researchers often employ the same set of covariates across the multiple transitions to aid the substantive interpretation of the coefficients $\mathbf{b}_{jk}$.

One can estimate Equation (17) using a suitable generalization of the Cox model to multiple origin and destination states. Under this specification, the $q_{jk}(t)$ are assumed to be arbitrary unspecified functions of t that vary in arbitrary but unspecified ways across transitions. If the Cox model is not used, then Equation (17) will require parameterizing the $q_{jk}(t)$. Guidance for these parameterizations can be obtained from suitable generalizations to the nonparametric methods discussed above. The resulting patterns can then be used to select a particular parameterization of the $q_{jk}(t)$, with the underlying issues essentially identical to those for a single transition. Multiple dimensions of time can also be handled in ways similar to those discussed for a single transition.

Estimation of Equation (17) will, in general, yield a very large number of parameter estimates—there will, in general, be coefficients for each covariate and transition pair. For substantive and interpretive parsimony, one may wish to determine if the effect of a covariate is similar across selected transitions—for example, for the transitions depicted in panel B of Figure 22-1, if income effects are similar for third and higher-order marital births. Such hypotheses can be evaluated using log likelihood ratio or BIC (Bayesian Information Criterion) tests (Raftery, 1995; Schwarz, 1978) under equality constraints on the appropriate parameters across transitions. Suitable extensions of this same idea can be used to determine if one can "pool" across transitions, for example, if, in Figure 22-1, model fit is not substantially degraded under the more parsimonious model in panel A relative to the more complex model in panel B.

UNOBSERVED HETEROGENEITY

As noted at the outset of this chapter, life course theorists often assert that one must approach the life course holistically—that domains such as work, marriage, childbearing, and emotional development, typically analyzed in isolation, cannot be understood adequately without considering these domains as one unified whole. The models considered to this point help sharpen this idea. Take, for example, panel B of Figure 22-1, which depicts transitions in two domains—marriage and childbearing. One can show that the application of the nonparametric estimator in Equation (5) to these transitions will reproduce the distribution of individuals, both across the multiple statuses and over time. Extensions of these methods to incorporate covariates—the Cox model or the models in Equations (9)–(12)—will likewise also reproduce the across-state and across-time distribution of individuals when proportionality holds. This implies, then, that the importance of this assertion lies not at the level of aggregate distributions, but rather at the level at which individual behaviors are modeled.

One approach formalizing this idea lets events in one life course domain affect transition rates in another domain. The models considered thus far accomodate this type of dependence by letting the researcher condition on any aspect of an individual's past history—including the individual's trajectory or past history in another life domain—when modeling $r_{ijk}(t)$. Thus, measures constructed from a person's trajectory of work and labor force experience up to time t can be used as right-hand side covariates in modeling risks at time t of transitions in the realms of fertility, marriage, divorce, retirement, and so forth. Another possibility, often raised by economists, is that an individual's decisions about work, fertility, and marriage reflect an attempt to maximize utility across these joint spheres. If all relevant decision inputs are observed, then one can condition on suitable measures of these inputs, in which case no new issues arise. However, when some relevant inputs are unobserved, a number of subtle issues arise in ways that are substantially more troublesome in a hazard regression context than in a static linear regression.*

Recall that a standard result for linear regression is that, if an unobservable z is uncorrelated with a covariate x of interest, then the OLS (ordinary least squares) estimate of the effect of x will be unbiased although not optimally efficient. No analogous result holds in the hazard regression context; in particular, even if an unobservable z is initially uncorrelated with a covariate x, in general, as time passes, x and z will not remain uncorrelated. This gives rise to a number of difficulties and subtleties. One difficulty, first noted by Sheps and Menken (1973), is that unobserved heterogeneity can play havoc with attempts to make inferences about time or age dependence. A classic example is to consider a population which, when observed initially, is comprised of equal numbers of individuals from two groups, A and B. Suppose that in both groups, mortality is governed by a constant (exponential) hazard rate, with $r_A > r_B$ and suppose further that the analyst is ignorant of the existence of the two subgroups. Sheps and Menken (1973) observed that, in this situation, the analyst will observe a monotonically declining rate with age, despite the fact that, by assumption, mortality does not vary with age in either group. This is an example of unobserved "frailty": because members in A are frailer than those in B, they will die more quickly. As a consequence, the composition of the sample population will shift over time, moving from equal numbers of individuals from A and B to a population more heavily weighted towards those from B. Thus, what the analyst observes (technically, a distribution consisting of a mixture of the underlying r_A and r_B) will be mortality that is initially close to r_A but which will over time decline to r_B, as the sample composition of the population is increasingly selected against individuals from group A. See, for example, Trussell and Richards (1985) and Vaupel and Yashin (1985), with the latter providing informative examples showing how unobserved heterogeneity may affect conclusions about time variation in $r(t)$.

A second issue concerns potential biases in estimated covariate effects, an issue emphasized in a series of influential papers by Heckman and Singer (1982, 1984, 1985). In reanalyses of unemployment data of Kiefer and Neumann (1981) employing a Weibull specification for duration dependence, they found that covariate estimates fluctuated markedly depending on whether the unobservable z was assumed to be normal, log normal, or gamma distributed.† Their results, coupled with a series of Monte Carlo studies, led them to advocate an alternative approach to modeling unobserved heterogeneity using a discrete mixing distribution with

*A fixed-effects strategy, often used to analyze panel data on metric outcomes (e.g., see Halaby, this volume), was shown by Chamberlain (1985) to yield inconsistent estimates in an event history context; hence, researchers have concentrated attention on random effects models.

†Heckman and Singer (1982) motivated their use of a Weibull distribution on search theoretical grounds. In subsequent work (Heckman & Singer, 1985), they proposed an alternative Box–Cox-type parameterization for time dependence that has as special cases the exponential, Weibull, and Gompertz models.

finite points of support. One difficulty with this approach is that, while it proceeds nonparametrically with respect to the distribution of the unobservables, it nevertheless requires strong parametric assumptions concerning the distribution of event times (Trussell & Richards, 1985). Yet another possibility would be to proceed with reasonably flexible specifications for both the distribution of the unobservables (e.g., using a discrete mixing distribution) and of the event times (e.g., using the piecewise constant or spline specifications in Equations (10) or (11)). For theoretical results concerning such an approach, see Elbers and Ridder (1982) and Heckman and Singer (1984).

Another approach to handling unobserved heterogeneity is due to Lillard and colleagues (e.g., see Lillard, 1993; Lillard, Brien, & Waite, 1995; Lillard & Waite, 1993; Upchurch, Lillard, & Panis, 2001). It has long been recognized that when multiple transitions are observed for a sample of individuals, this added information might permit identification of additional distributional aspects of the unobservables. Lillard and colleagues built upon this insight to estimate models akin to simultaneous equation models for metric outcomes in which unobservables may be correlated across transitions. Consider extending the proportional specification in Equation (17) for multiple origin and destination states by adding an error term e_{jk}, where e_{jk} is assumed to capture unobserved heterogeneity specific to the transition from origin state j to destination state k:

$$\log r_{jk}[t|\mathbf{x}_{jk}(t)] = \log q_{jk}(t) + \mathbf{b}_{jk}\mathbf{x}_{jk}(t) + e_{jk} \tag{18}$$

Let $m = 1, \ldots, M$ index the full set of transitions $j \rightarrow k$ and let the e_{jk} be assumed normally distributed with mean 0 and covariance matrix

$$\operatorname{cov}(e_m, e_p) = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1M} \\ \sigma_{21} & \sigma_2^2 & \cdots & \sigma_{2M} \\ \vdots & \vdots & \vdots & \vdots \\ \sigma_{M1} & \sigma_{M2} & \cdots & \sigma_M^2 \end{pmatrix} \tag{19}$$

with the unobservables for two transitions, m and m', correlated according to

$$\rho_{mm'} = \frac{\sigma_{mm'}}{\sigma_m \sigma_{m'}} \tag{20}$$

Thus, the models in Equations (18)–(20) link different life course domains using observed *and* unobserved attributes of individuals and by allowing the effects of such unobservables to be correlated across life domains. Lillard, Brien, and Waite (1995) used such a modeling strategy to address whether the observed higher rate of divorce among those who have cohabited prior to marriage is an artifact of unobservables that differentially select couples into cohabitation. Thus, an unusual strength of these models is that they let researchers address endogenous selection. Nevertheless, it can be difficult to achieve identification of the many parameters in the variance–covariance matrix in Equation (19), with identification achieved in most empirical work to date by imposing exclusion restrictions or by exploiting data on repeated events. Alternatively, one might achieve identification structurally via instrumental variables (e.g., see Duncan, 1975), although finding adequate instruments can be difficult in practice.

It is nevertheless important to emphasize that all models for unobserved heterogeneity proposed to date assume that what is unobserved does not vary with time and can be proxied by a one-dimensional correction term. Often, these assumptions are plausible; this is especially true for economic models of behavior, where it is commonly assumed that behavior is determined by a parsimonious set of influences, including those not observed by the researcher. Moreover, in many economic models of behavior, individuals are assumed to act rationally, optimizing over the life span; under these assumptions, characteristics such as permanent income and long-term horizons play an important role. By contrast, a recurrent theme among many life course researchers, particularly those drawn from sociology and demography, concerns the relatively fluid nature of the social circumstances of individuals. If so, then extant models for unobserved heterogeneity address only one aspect of is unobserved—those unobserved characteristics of individuals that are fixed and unchanging—but not other aspects of what may be unobserved—unobserved characteristics of individuals that may be more fluid in nature.

COUPLED PROCESSES

To this point, the models I have considered focus on a single individual's life course transitions, but as noted earlier, life course theorists have often posited linkages across the life courses of multiple individuals, for example, that the events, behaviors, and outcomes for one member of a married couple might have profound effects on the events, behaviors, and outcomes for the person's spouse or partner. If such events, behaviors, or outcomes for both members of the couple are observed, if such covariates for a spouse or partner are exogenous to the outcome of the other, and if, conditional on these covariates, the processes for members of the couple can be assumed independent, then the problem reduces to the usual one of modeling effects of covariates on a transition of interest, with the set of covariates now expanded to include observed characteristics of a spouse or partner.

What is usually deemed implausible is the assumption of conditional independence; said another way, we often suspect that the set of observed covariates do not exhaust the set of what we would wish to observe theoretically and, in particular, that certain key characteristics for a couple are unobserved. A classic example concerns the problem of modeling the mortality of married couples, where the assumption of conditional independence is suspect if researchers do not observe aspects of diet or health-related behaviors that might affect both members of the couple, cumulative but unobserved health insults that are reflective of a couple's physical or social environment, or mate selection on unobserved characteristics that would tend to make couples more similar on health outcomes than two randomly chosen individuals in the population.

The specifications in Equations (18)–(20) can, in principle, be adapted to cover some, but not all, of these cases. Suppose, for example, that some unobserved characteristic of couples generates a positive correlation between their mortality experience relative to two individuals drawn at random from the same population. Concretely, let T_1 and T_2 denote the mortality of husbands and wives, respectively; then a model analogous to Equations (18)–(20) for the mortality for couple i can be written as

$$\begin{aligned} \log r_{i1}[t|\mathbf{x}_{i1}(t)] &= \log q_1(t) + \mathbf{b_1}\mathbf{x}_{i1}(t) + e_{i1} \\ \log r_{i2}[t|\mathbf{x}_{i2}(t)] &= \log q_2(t) + \mathbf{b_2}\mathbf{x}_{i2}(t) + e_{i2} \end{aligned} \tag{21}$$

with $\mathbf{x}_{i1}(t)$, $\mathbf{x}_{i2}(t)$, e_{i1}, and e_{i2} denoting observed covariates and unobserved components for the husband and wife in couple i and where

$$\rho = \frac{\sigma_{12}}{\sigma_1^2\sigma_2^2}$$

gives the correlation between members of a couple in age-specific mortality risks. As before, this model incorporates strong behavioral assumptions, in particular that the unobservables e_{i1} and e_{i2} do not vary over time; note, in particular, that such an assumption would not cover the case in which mortality is affected by cumulative health insults as shared by a couple. In addition, identification of the model parameters in Equation (21) can be difficult when the data available to the researcher do not contain instruments that would, for example, plausibly affect the mortality of husbands, but not wives. An alternative approach is to model the *joint* distribution, $f(t_1,t_2)$, for the two event times, which yields so-called bivariate survivor models. In practice, this approach also has proven difficult to implement, in part because researchers often have little guidance for specifying the parametric form of the resulting two-dimensional baseline hazard function $q(t_1,t_2)$. Mare and Palloni (1988), Mare (1994), and Poetter (2000) provide empirical examples and comprehensive discussions of these and other issues.*

CONCLUSION

Good methods often help sharpen theory—the translation of theoretical ideas stated verbally into testable propositions linked to data very often provides greater theoretical insight, for example, by revealing conceptual ambiguities or gaps in a theoretical formulation. In this chapter, I have reviewed some examples that typify how the interplay between life course theory and event history methods might yield such analytical insights. Examples include consideration of the populations deemed to be at "risk" of particular life course transitions; careful specification of the states and transitions between states characterizing a problem, not excluding those transitions of theoretical importance even if they may be difficult to observe empirically; how one might operationalize notions of age grading in the timing of various adult transitions, including extensions to other temporal dimensions—not just age—that might plausibly govern such transitions; how one might distinguish, conceptually and theoretically, notions such as marital "retreat" versus "delay" for successive birth cohorts; and what might be meant by an assertion about linkages across domains in the life course for a given individual and similarly what might be meant by an assertion that couple processes are linked.

Another theme running throughout this chapter concerns the implicit trade-offs between nonparametric, semi-parametric, and parametric methods. In my own work, I have found it useful to begin with exploratory analyses via nonparametric techniques. Because nonparametric methods make few parametric assumptions, they can often provide important indications for how one might formulate more parametric models, especially when choosing among different models for the baseline hazard. I have also found it useful to use, where possible, flexible parametric models, for example, those utilizing piecewise linear splines in the place of a simple linear (or other parametric) specification. Such techniques can be used to model nonlinear effects of covariates or non-proportional effects of covariates; piecewise linear splines also yield quite flexible models for the baseline hazard. Note also that such models can be used to obtain tests of standard assumptions, for example, assumptions concerning linearity or proportionality in the effect of a covariate.

The ability to relax such assumptions has led to increased interest in nonparametric and semi-parametric methods in the statistical and econometric literatures (e.g., see Härdle, 1990; Hastie & Tibshirani, 1990, 1993). Thus, when sufficiently large sample sizes are available to the researcher, it is often possible to devise methods that rely more heavily on information

*For a brief discussion of software to estimate these and other models, see the Appendix.

contained within the available data, for example, using this information to guide the choice of an appropriate functional form (e.g., as opposed to linearity or proportionality). Conversely, important insights follow from those models in which it is not possible to relax maintained assumptions even when arbitrarily large samples are available. In such circumstances, parameter estimates are obtained both from the observed data *and* from assumptions that cannot, even in principle, be tested (Manski, 1995). It is worth noting that, to date, while the classic work of Heckman and Singer (1982, 1985) on models for unobserved heterogeneity provides a semi-parametric framework for uncorrelated unobservables, researchers have not yet devised nonparametric or semi-parametric alternatives to the multiple transition models with correlated unobservables such as those in Equations (18)–(20).

The intersection of these issues—of how formal methods may clarify theoretical ideas and how nonparametric techniques can shed light on model identification—provides insight into a central assertion in the life course: that one cannot understand seemingly disparate life course domains in isolation or that the life courses of spouses or partners exert mutual influences on one another. For example, consider life course transitions defined by cross-classifying an individual's social statuses across two or more life course domains, as in Figure 22-1, which depicts the transitions of women through a cross-classification of fertility and marital statuses. It can be shown that standard nonparametric methods will reproduce the observed distribution across persons and across time through these multiple transitions and statuses. This result implies that the notion of linked life course domains must lie at some deeper level than the distribution of individuals across statuses over time.

What might constitute a "deeper" notion of linked lives? One possibility is that the researcher directly observes all relevant aspects of that which is hypothesized to drive such linkages across life course domains. If so, one can proceed in the usual way by incorporating these data as standard covariates in a hazard regression. However, researchers usually worry that one or more key factors driving such linkages are in fact unobserved. If so, then such unobservables can induce correlations across the life course transitions observed for individuals. The models outlined in Equations (18)–(20), which were developed to address precisely this problem, are thus of great intrinsic substantive interest to life course, since they would in principle allow researchers to address linkages and to obtain point estimates of the correlation between transitions for a sample of individuals or for the transitions observed for members of a couple paired. But, as noted above, it has been difficult to date to devise obvious nonparametric or semi-parametric alternatives to such models; hence, it remains unclear the degree to which parameter estimates from such models are identified solely or in part from such untestable model assumptions. Such statistical difficulties may, in turn, be reflective not just of technical issues but may be revealing of gaps in theoretical accounts, for example, ambiguities or incompleteness by life course theorists in specifying the range of theoretical and behavioral mechanisms that might generate behavioral linkages across different parts of the life course or across individuals in a couple.

APPENDIX: SOFTWARE

It is important to emphasize that, because software for event history models continues to evolve rapidly, any survey of available software will become dated rapidly. This being said, several readily available software packages (e.g., SAS, SPSS, Stata, S, S-Plus, and R) have modules that permit estimation of most basic models, including nonparametric estimation of the survivor, integrated hazard, and (sometimes) hazard functions, and estimation of the Cox model. Many of these packages also have provisions that permit the user to define time-varying covariates; in such cases, these models can be used to obtain estimates of the piecewise constant baseline model in Equations (9) and (10) by the appropriate coding of time-varying dummy variables representing the appropriate time intervals.

Some of these packages also permit the estimation of models with multiple origin and destination states. When this is not possible, one can obtain estimates for standard models (i.e., those that assume conditional independence in the transitions to the multiple destination states) using a two-step procedure: first, by separating the problem into each origin state and, second, for each origin state, estimating parameters for a given destination state *k* by treating as censored those transitions to the other destination states.* Allison (1995) provides a useful and comprehensive survey of hazard estimation using SAS.

There are also several packages that have been developed for estimation of less standardized event history models. An incomplete list of such packages includes aML (Lillard & Panis, 2000), CTM (Yi, Honoré, & Walker, 1987), RATE (Tuma, 1979), and TDA (Blossfeld et al., 1989), and a variety of supplemental libraries developed for S, S-Plus, and R (see e.g., Loader, 1999). Most (but not all) of these specialized packages also permit estimation of basic models, but they otherwise differ considerably in the model coverage; hence, it is difficult to identify any one package as providing superior coverage relative to another. For example, aML and RATE permit estimation of the piecewise linear spline for the baseline hazard function; provisions for piecewise linear splines in these packages also permit estimation of nonlinear and non-proportional covariate effects. TDA provides an extensive array of more parametric baseline hazard functions such as the Gompertz, Weibull, log logistic, log Gaussian, and sickle models. Local likelihood models (Wu & Tuma, 1990) are most easily estimated using supplemental libraries available in S, S-Plus, and R (Loader, 2000). CTM and aML have the most comprehensive routines for estimating models for unobserved heterogeneity, with CTM permitting estimation of the models discussed by Heckman and Singer (1985) and aML the models discussed by Lillard (1993).

ACKNOWLEDGMENT: Research funding from the National Institute of Child Health and Human Development (HD 29550) and research facilities provided under HD 05876 to the Center for Demography and Ecology, are gratefully acknowledged.

REFERENCES

Aalen, O. O. (1978). Nonparametric inference for a family of counting processes. *Annals of Statistics, 6*(4), 701–726.

Abbott, A. (2000). Reply to Levine and Wu. *Sociological Methods and Research, 29*(1), 65–76.

Abbott, A., & Tsay, A. (2000). Sequence analysis and optimal matching methods in sociology: review and prospect. *Sociological Methods and Research, 29*(1), 3–33.

Allison, P. D. (1985). *Event history analysis: regression for longitudinal event data*. Beverly Hills, CA: Sage.

Allison, P. D. (1995). *Survival analysis using the SAS® system: A practical guide*. Cary, NC: SAS Institute.

Andersen, P. K., & Gill, R. D. (1982). Cox's regression model for counting processes: a large sample study. *Annals of Statistics, 10*(4), 1100–1120.

Becker, G. S. (1991). *A treatise on the family* (Enlarged ed.) Cambridge, MA: Harvard University Press.

*This follows because maximum likelihood estimation for the expression in Equation (15) can be shown to be separable by origin; intuitively, this follows because the origin state *j* determines the set of individuals at risk of the destination states *k*. Note that such an estimation strategy cannot be used for models that weaken the conditional independence assumption, for example, for those models in which terms for unobserved heterogeneity affect the transitions to the multiple destination states.

Blau, P. M. (1977). *Inequality and heterogeneity: A primitive theory of social structure*. New York: Free Press.

Blau, P. M., & Duncan, O. D. (1967). *The American occupational structure*. New York: Wiley.

Blossfeld, H.-P., Hamerle, A., & Mayer, K. U. (1989). *Event history analysis*. Hillsdale, NJ: Erlbaum.

Bumpass, L. L. (1990). What's happening to the family: interaction between demographic and institutional change. *Demography, 27*(4), 483–498.

Chamberlain, G. (1985). Heterogeneity, omitted variable bias, and duration dependence. In J. J. Heckman & B. Singer (Eds.), *Longitudinal analysis of labor market data* (pp. 3–38). Cambridge: Cambridge University Press.

Cox, D. R. (1972). Regression models and life tables [with discussion]. *Journal of the Royal Statistical Society, B34*(2), 187–220.

Cox, D. R. (1975). Partial likelihood. *Biometrika, 62*(2), 269–276.

Cox, D. R., & Oakes, D. (1984). *Analysis of survival data*. London: Chapman & Hall.

Dechter, A. R. (1992). The effect of women's economic independence on union dissolution. (Working Paper No. 92–28) Center for Demography and Ecology, University of Wisconsin-Madison, Madison, WI.

Duncan, O. D. (1975). *Introduction to structural equation models*. New York: Academic Press.

Elbers, C., & Ridder, G. (1982). True and spurious duration dependence: The identifiability of the proportional hazard model. *Review of Economic Studies, 49*(3), 403–410.

Elder, G. H., Jr. (1998). The life course as developmental theory. *Child Development, 69*(1), 1–12.

Elder, G. H., Jr. (1999). *Children of the Great Depression: Social change in life experience* (25th anniversary ed.). Boulder, CO: Westview.

Fleming, T. R., & Harrington, D. P. (1991). *Counting processes and survival analysis*. New York: Wiley.

Gilder, G. F. (1986). *Men and marriage*. Gretna, LA: Pelican.

Härdle, W. (1990). *Applied nonparametric regression*. Cambridge: Cambridge University Press.

Hastie, T., & Tibshirani, R. (1993). Varying-coefficient models [with discussion]. *Journal of the Royal Statistical Society, B55*(4), 757–796.

Hastie, T., & Tibshirani, R. (1990). *Generalized additive models*. London: Chapman & Hall.

Heckman, J. J., & Singer, B. (1982). The identification problem in econometric models for duration data. In W. Hildenbrand (Ed.), *Advances in econometrics* (pp. 39–77). Cambridge: Cambridge University Press.

Heckman, J. J., & Singer, B. (1984). A method for minimizing the impact of distributional assumptions in econometric models for duration data. *Econometrica, 52*(2), 271–320.

Heckman, J. J., & Singer, B. (1985). Social science duration analysis. In J. J. Heckman & B. Singer (Eds.), *Longitudinal analysis of labor market data* (pp. 39–110). Cambridge: Cambridge University Press.

Hogan, D. P., & Astone, N. M. (1986). The transition to adulthood. *Annual Review of Sociology, 12*, 109–130.

Hougaard, P. (2000). *Analysis of multivariate survival data*. New York: Springer-Verlag.

Kaplan, E. L., & Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association, 53*(282), 437–481.

Kiefer, N. M., & Neumann, G. R. (1981). Individual effects in a nonlinear model: Explicit treatment of heterogeneity in the empirical job-search model. *Econometrica, 49*(4), 965–979.

Lancaster, T. (1990). *The econometric analysis of transition data*. Cambridge: Cambridge University Press.

Levine, J. (2000). But what have you done for us lately? Commentary on Abbott and Tsay. *Sociological Methods and Research, 29*(1), 34–40.

Lillard, L. (1993). Simultaneous equations for hazards: Marriage duration and fertility timing. *Journal of Econometrics, 56*(1/2), 189–217.

Lillard, L. A., Brien, M. J., & Waite, L. J. (1995). Premarital cohabitation and subsequent marital dissolution: A matter of self-selection? *Demography, 32*(3), 437–457.

Lillard, L. A., & Waite, L. J. (1993). A joint model of childbearing and marital disruption. *Demography, 30*(4), 653–681.

Loader, Clive. (1999). *Local Regression and Likelihood*. New York: Springer.

Manski, C. F. (1995). *Identification problems in the social sciences*. Cambridge, MA: Harvard University Press.

Mare, R. D. (1994). Discrete-time bivariate hazards with unobserved heterogeneity: An indirect contingency table approach. *Sociological Methodology, 24*, 341–383.

Mare, R. D., & Palloni, A. (1988). Couple models for socioeconomic effects on mortality among older persons (CDE Working Paper No. 88–07). Center for Demography and Ecology, University of Wisconsin-Madison.

Mayer, K. U. (1988). German survivors of World War II: The impact on the life course of the collective experience of birth cohorts. In M. W. Riley (Ed.), *Social structures and human lives* (Vol. 1, pp. 229–246). Newbury Park, CA: Sage.

Mayer, K. U., & Tuma, N. B. (1990). Life course research and event history analysis: An overview. In K. U. Mayer & N. B. Tuma (Eds.), *Event history analysis in life course studies* (pp. 1–20). Madison, WI: University of Wisconsin Press.

Meyer, J. W. (1986). The self and the life course: Institutionalization and its effects. In A. B. Sørensen, F. E. Weinert, & L. R. Sherrod (Eds.), *Human development and the life course: Multidisciplinary perspectives* (pp. 199–216). Hillsdale, NJ: Erlbaum Associates.

Modell, J., Furstenberg, F., & Hershberg, T. (1976). Social change and transitions to adulthood in historical perspective. *Journal of Family History, 1*(1), 7–32.

Nelson, W. (1972). Theory and applications of hazard plotting for censored failure data. *Technometrics, 14*(4), 945–966.

Oppenheimer, V. K. (1997). Women's employment and the gain to marriage: the specialization and trading model. *Annual Review of Sociology, 23*, 481–453.

Poetter, U. (2000). A multivariate Buckley–James estimator. In T. Kollo, E.-M. Tiit, & M. Srivastava (Eds.), *Multivariate statistics: New trends in probability and statistics* (Vol. 5). Utrecht: VSP.

Popenoe, D. (1996). *Life without father*. New York: Free Press.

Raftery, A. E. (1995). Bayesian model selection in social research. In P. V. Marsden (Ed.), *Sociological methodology 1995* (pp. 111–165). Cambridge, MA: Blackwell.

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics, 6*(2), 461–464.

Settersten, R. A., Jr., & Mayer, K. U. (1997). The measurement of age, age structuring, and the life course. *Annual Review of Sociology, 23*, 233–261.

Sewell, W. H., & Hauser, R. M. (1975). *Education, occupation, and earnings: Education in the early career*. New York: Academic Press.

Shanahan, M. J. (2000). Pathways to adulthood in changing societies: Variability and mechanisms in life course perspective. *Annual Review of Sociology, 26,* 667–692.

Sheps, M. C., & Menken, J. A. (1973). *Mathematical models of conception and birth*. Chicago: University of Chicago Press.

Spain, D., & Bianchi, S. M. (1996). *Balancing act: Motherhood, marriage, and employment among American women*. New York: Russell Sage.

Trussell, J., & Richards, T. (1985). Correcting for unmeasured heterogeneity in Hazard models using the Heckman–Singer procedure. In N. B. Tuma (Ed.), *Sociological methodology 1985* (pp. 242–276). San Francisco: Jossey-Bass.

Tsiatis, A. (1975). A nonidentifiability aspect of the problem of competing risks. *Proceedings of the National Academy of Sciences, 72*(1), 20–22.

Tuma, N. B. (1979). *Invoking RATE* (2nd ed.). Menlo Park: SRI International.

Tuma, N. B., & Hannan, M. T. (1984). *Social dynamics: Models and methods*. Orlando, FL: Academic Press.

Upchurch, D. M., Lillard, L. A., & Panis, C. W. A. (2001). The impact of nonmarital childbearing on subsequent marital formation and dissolution. In L. L. Wu & B. Wolfe (Eds.), *Out of wedlock: Causes and consequences of nonmarital fertility* (pp. 344–380). New York: Russell Sage Foundation.

Vaupel, J. W., & Yashin, A. I. (1985). Heterogeneity's ruses: Some surprising effects of selection on population dynamics. *American Statistician, 39*(3), 176–185.

White, H. C., Boorman, S. A., & Breiger, R. L. (1976). Social structure from multiple networks. I. Blockmodels of roles and positions. *American Journal of Sociology, 81*(4), 730–780.

Winsborough, H. H. (1980). A demographic approach to the life cycle. In K. W. Back (Ed.), *Life course: Integrative theories and exemplary populations* (pp. 65–75). Boulder, CO: Westview.

Wu, L. L. (1989). Issues in smoothing empirical hazards. In C. C. Clogg (Ed.), *Sociological methodology 1989* (pp. 127–159). Washington, DC: American Sociological Association.

Wu, L. L. (1990). Simple graphical goodness-of-fit tests for Hazard rate models. In K. U. Mayer & N. B. Tuma (Eds.), *Event History Analyses in Life Course Research* (pp. 184–199). Madison, WI: University of Wisconsin Press.

Wu, L. L. (1996). Effects of family instability, income, and income instability on the risk of a premarital birth. *American Sociological Review, 61*(3), 386–406.

Wu, L. L. (2000). Some comments on 'Sequence Analysis and Optimal Matching Methods in Sociology: Review and Prospect'. *Sociological Methods and Research, 29*(1), 41–64.

Wu, L. L., & Martinson, B. C. (1993). Family structure and the risk of a premarital birth. *American Sociological Review, 58*(2), 210–232.

Wu, L. L., & Thomson, E. (2001). Racial differences in family change and early sexual initiation. *Journal of Marriage and the Family, 63*(3), 682–696.

Wu, L. L., & Tuma, N. B. (1990). Local Hazard models. In C. C. Clogg (Ed.), *Sociological methodology 1990* (pp. 141–180). Washington, DC: American Sociological Association.

Xie, Y. (1994). Log-multiplicative models for discrete-time discrete-covariate event-history data. In P. V. Marsden (Ed.), *Sociological methodology 1994* (pp. 301–340). Washington, DC: American Sociological Association.

Yamaguchi, K. (1991). *Event history analysis*. Newbury Park, CA: Sage.

Yi, K.-M., Honoré, B. & Walker, J. (1987). *CTM: A program for the estimation and testing of continuous time multi-state multi-spell models.* Chicago: NORC/ERC and University of Chicago.