7

EVENT-HISTORY ANALYSIS FOR LEFT-TRUNCATED DATA

*Guang Guo**

A subject is left-truncated when it comes under observation after having been exposed to the risk of an event for some time. Left-truncated subjects tend to have lower risks at shorter durations than those in a normal sample because high-risk subjects tend to experience the event and drop out before reaching the point at which observation begins. When start times are unknown, left truncation is practically intractable unless the hazard rate is constant or all left-truncated subjects are discarded. When start times are known, left-truncated data can be handled by the conditional likelihood approach. The critical information on the start time of a left-truncated subject can be frequently obtained from human subjects. The concentration of covariate information in the observation period in a longitudinal social survey is well-suited for the conditional approach. We show that a piece-wise exponential hazard model and a discrete-time model based on the conditional likelihood approach can be readily estimated by extant packages like SAS. We also describe an alternative conditional partial likelihood approach. An empirical example of marital dissolution in the United States is provided.

*University of North Carolina at Chapel Hill

I am indebted to Sanders Korenman for involving me in his research and for many helpful comments. The work was supported by the Department of Sociology and the Office of Population Research at Princeton University. Additional support was provided by the Carolina Population Center at the University of North Carolina at Chapel Hill. The paper has benefited from comments by David Guilkey, Lorenzo Moreno, Thomas Mroz, and Germán Rodríguez. I also thank the editor, Peter Marsden, and the anonymous reviewers for many helpful suggestions.

1. INTRODUCTION

Social scientists' event-history data are almost never complete. Frequently, an event history is right-censored, meaning that the length of the survival time is observed to be greater than a certain value, but the precise length is unknown. For instance, in a study of divorce rates, a subject is right-censored if the subject's marriage remains intact at the end of an observation period. The subject's marriage may dissolve in the future, but we do not know when. Statistical procedures for right-censored data have been developed and used routinely (Cox and Oakes 1984; Tuma and Hannan 1984; Allison 1984; Yamaguchi 1991).

Sometimes, event-history data are left-truncated, meaning that a subject has been exposed to the risk of an event for a while when it comes under observation; the length of exposure prior to observation may or may not be known. In contrast to the clear-cut statistical solutions for right-censored event-history data, there is considerable confusion surrounding the solutions for the data that are left-truncated. Even the definitions of left censoring and left truncation in the literature can be quite puzzling. This paper aims at providing a practical guidance for coping with social science event-history data that are left-truncated, especially when the length of exposure prior to observation is known. Only the case of single events is treated, although much of the discussion should be applicable to the case of repeated events, in which only the first spell is likely to be left-truncated. Marital dissolution data from the Panel Study of Income Dynamics (PSID) are used as an example throughout the paper (Survey Research Center 1989).

1.1. *Characterization of Incomplete Data*

Figure 1 illustrates various complete and incomplete event histories. The horizontal axis indicates the calendar time denoted by τ. The observation period is usually of finite length, with the beginning and ending marked by τ_0 and τ_1. For convenience and without loss of generality, the start time of the observation window is assumed to be zero. A lowercase *e* for *event* at the right end of a line indicates the exact time at which an event occurs is observed. A lowercase *s* for *start time* at the left end of a line indicates that the start time is

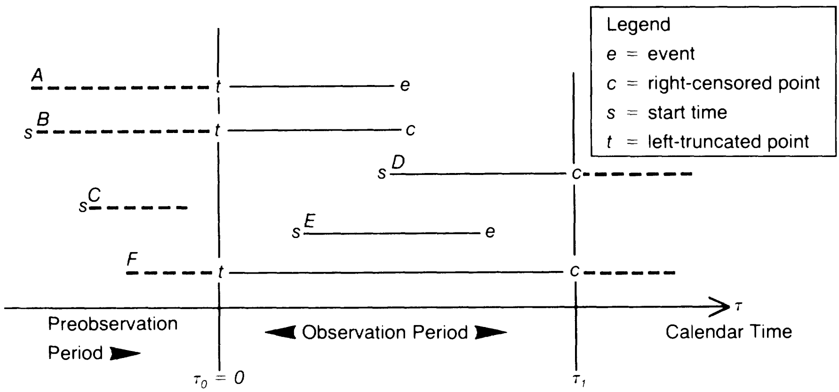

FIGURE 1. Complete and incomplete event-history observations.

known. A lowercase c for *censored* indicates that the subject is right censored at that point. Right censoring may arise from the termination of the observation period (subject D) or the loss to follow-up (subject B). The right-censored subject D can be handled routinely by conventional methods. A lowercase t for *truncated* at τ_0 signifies that the subjects A and B have been exposed to the risk for some time. Subject E is the only complete event history described in Figure 1. Subject F is both left-truncated and right-censored.

Both subjects A and B are left-truncated, but the start time is known only for the latter. An example of subject A arises in the use of longitudinal social survey data such as those collected in the PSID. The PSID began collecting information from a nationally representative sample of households in the United States in 1968 (τ_0), and the respondents were interviewed each year thereafter. For illustrative purposes, we consider 1987 as the end of the observation window (τ_1). If we are interested in the event of first marital dissolution for women, we need to know the start time of a marriage as well as the time of dissolution or censoring. Left truncation arises when a woman married some time before 1968 and entered the observation period in 1968. She might or might not have divorced after 1968. For many years before the 1985 wave of the survey that collected information on retrospective marital history, the exact length of exposure to divorce before 1968 could not be identified.

1.2. *Definitions of Left Censoring and Left Truncation*

While we consider subjects that enter the observation period after having been exposed to the risk for a while as left-truncated, some social scientists consider them as left-censored or partially left-censored (Tuma and Hannan 1984 and Yamaguchi 1991). Their definition is intuitive in that it defines the lack of information on the left and right symmetrically as left and right censoring. Econometricians usually treat this situation as a part of the problems caused by "sampling of duration data" (Ridder 1984) or "the initial conditions" (Heckman and Singer 1986) in event-history analysis.

Our definition of left truncation is drawn from the work of Cox and Oakes (1984, pp. 177–78), Miller (1981, p. 7), Trussell and Guinnane (1991), Keiding (1986), Hald (1952), and Turnbull (1974). They distinguish two kinds of data: truncated and censored. Truncation arises when sampling from an incomplete population. Left truncated event-history data are incomplete because they do not include those subjects that have not survived long enough to be observed.[1] As Keiding (1986, p. 357) explained: "On the other hand censoring occurs when we are able to sample the complete population but the individual values of observations below (or above) a given value are not specified." A left-censored subject is known to have experienced the event, but the exact failure time is unknown. In Figure 1, subject C is left-censored. The absence of e at the right end of the time line for subject C in Figure 1 signifies the lack of knowledge about the exact time at which the event occurs. Note the symmetry in the definitions of left and right censoring here. In both left and right censoring, the exact length of time until the event is unknown, but we know that the event occurs before the start of the observation window in the case of left censoring and after the end of the observation window in the case of right censoring.

[1]It should be emphasized that, unlike censoring, truncation is a characteristic of a sample, rather than of an individual subject. Suppose that we observe a sample of subjects A, B, D, E, and F in the observation period (Figure 1). The *sample* is left-truncated because it has left out the high-risk subject C. In this paper, however, we sometimes refer to *subjects* such as A, B, and F as left-truncated when the sample is already known to be left-truncated for convenience. Fixing a left-truncated sample usually amounts to a special treatment of the standard likelihood function of the left-truncated subjects A, B, and F while leaving the likelihood function of the subjects such as D and E unchanged.

Turnbull (1974) provides an example of left censoring that Leiderman et al. (1973) came across in a study of African infant precocity. The 65 children in the sample were born between July 1 and December 31, 1969. Starting in January 1970, each child was given a test every month to see if he or she had learned to perform a particular task. The start time is birth and the event is accomplishing the task. Left censoring occurred when some children were found to be able to perform the task at the very first test (i.e., they had experienced the event before the observation period). Sample selection is absent since these left-censored children remained in the sample after they had experienced the event.

1.3. *Problems of Left Truncation*

The characteristic problem of left truncation is sample selection, whether or not the start time is known. The left-truncated cases sampled at the beginning of the observation period tend to overrepresent low-risk cases among any given cohort. For instance, the couples sampled in 1968 in the PSID can be viewed as survivors of their marriage cohorts. The original size of a marriage cohort must be bigger than what is observed in 1968 though the exact size is unknown, because many in the original cohort did not stay in the first marriage long enough to be sampled in 1968.

The problem of sample selection can be illustrated more formally. Define the length of time from τ_0 to the start of exposure as T_r, with r standing for *retrospective,* and define the length of time from τ_0 to the time the subject is last observed as T_p, with p standing for *prospective* (Figure 2). Define T, which is the sum of T_r and T_p, as the survival time of the population in whose parameters we are interested. Let the density of the subjects that start t_r periods before τ_0 be expressed as $h(-t_r)$. The probability of these subjects surviving t_r periods must be $S_T(t_r)$, the survivor function of the population. Then the proportion of those that start t_r periods before τ_0 and have made it to τ_0 is $h(-t_r)S_T(t_r)$. Integrating over all cohorts, we obtain the probability of those that start before τ_0 and have survived to τ_0

$$P_0(\tau_0 \in T) = \int_0 h(-t_r)\, S_T\,(t_r)\, dt_r, \tag{1}$$

where $\tau_0 \in T$ means that the duration time t includes the calendar time $\tau_0 = 0$ as an element. Dividing the surviving cohort by the surviving population, we obtain

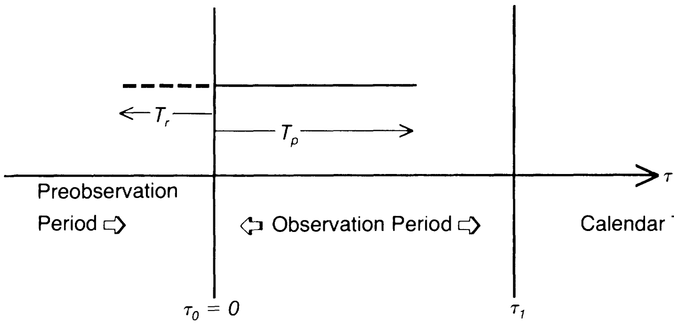

FIGURE 2. T_r and T_p.

$$f(t_r) = \frac{h(-t_r)\, S_T\,(t_r)}{P_0\,(\tau_0 \in T)}\,, \tag{2}$$

which is the density of those cases that start some time before τ_0 and have made it at least to τ_0. Assuming a constant entry rate $h(-t_r) = h$ over time, this density (2) would correspond to Cox's (1962) equilibrium density of the backward recurrence-time distribution. Due to sample selection, $f_{T_r}\,(t_r)$ (the density of left-truncated subjects) is quite different from $f_T\,(t)$ (the density of the population) and cannot be used as if it were $f_T(t)$ for the estimation of the parameters of the population.

2. SOME SOLUTIONS TO LEFT TRUNCATION WITH UNKNOWN START TIMES

The previous work on the "initial conditions problems" (Heckman and Singer 1986) and the distribution of event-history data (Ridder 1984) provides an important framework within which various solutions to left truncation can be better understood. In this paper, we give only a brief review of the results. For more details, see also Hamerle's (1991) exposition. Ridder shows that the density of the joint distribution of T_r and T_p, given covariates x and the condition that T starts at the calendar time $-t_r$ and has survived beyond τ_0, is

$$g(t_r,t_p|x,\tau_0 \in T) = \frac{h(-t_r|x) f(t_r + t_p|x)}{\int_0 h(-t_r|x)\, S_T(t_r|x)\, dt_r}\,. \tag{3}$$

Integrating (3) over t_r, the unobserved part of T, we obtain the density of a left-truncated case (observation A in Figure 1)

$$f(t_p|x,\tau_0 \in T) = \frac{\int\limits_0 h(-t_r|x)\, f(t_r + t_p|x)\, dt_r}{\int\limits_0 h(-t_r|x) S_T(t_r|x) dt_r}. \tag{4}$$

Expression (4), however, is intractable. The term $h(-t_r|x)$, the proportion of the population that starts being exposed to the risk at time $-t_r$, is almost never known.

An easy solution to left truncation is available, however, if one is willing to assume that the survival time T has an exponential distribution with the density $f_T(t) = \lambda e^{-\lambda t}$. Under this assumption, expression (4) for left-truncated observations simplifies to

$$f(t_p|\tau_0 \in T) = \lambda e^{-\lambda t_p}. \tag{5}$$

Expression (5) is in itself the density of an exponential distribution, as if the exposure of the observation started at τ_0, rather than the calendar time $-t_r$. The left-truncated observation simply *forgets* its experience before τ_0. This result comes directly from the property of *memorylessness* of the exponential distribution. Unfortunately, the assumption of a constant hazard $\lambda(t) = \lambda$ is often unrealistic. Erroneously assuming an exponential distribution may lead to severe bias in parameter estimates (Heckman and Singer 1986).

Another simple solution is to discard all left-truncated cases from analysis as Allison suggests (1984). Ridder's (1984) distribution theory shows that the resulting sample is not biased. The loss of information, however, can be huge. For example, in the first marital dissolution data of the PSID we have described, 49 percent of white women and 39 percent of black women, whose first marriages began prior to 1968, would have to be discarded if we are to follow this strategy. Moreover, some of the lost information is unique. Suppose that we are interested in the differential impact of the "no fault" divorce laws on divorce rates of the women who have been married for 20 years or less and the women who have been married for more than 20 years. In the PSID, we have observed the women for only 20 years (1968–1987). Then the use of left-truncated cases is the only way to answer questions concerning divorce rates at durations longer than 20 years.

We have mentioned that one major obstacle to the direct use of (4) as the likelihood function for left-truncated cases is the unavailability of the entry rate $h(-t_r)$. The easiest solution is to make a time homogeneity assumption: $h(-t_r) = h$. Under this assumption, which allows the cancellation of the entry rate, the density for left-truncated cases (4) becomes

$$f(t_p|x,\tau_0 \in T) = \frac{S(t_p|x)}{\int\limits_0 S_T(t_r|x)dr}. \tag{6}$$

The denominator of (6) yields $E(T|x)$, the expected value of the distribution. With further parametric assumptions on the distribution of T, (6) can be used to estimate the parameters related to the population distribution T.

The problem of unobserved entry rates is merely circumvented rather than confronted in (6). Under certain circumstances, external data sources (external to data analyzed) may be used to estimate entry rates empirically. See Tuma and Hannan (1984, pp. 131–32) for a similar argument. In the case of women's marital dissolution in the PSID, the aggregate data from the census may be used to establish the probabilities of marriage during the years before 1968. The difficulty of this strategy is the need to calculate the probabilities of marriage during those years for every possible combination of covariates x. Chances are that no external data source can match the large number of combinations of covariates, which is typical in social scientists' work today. In general, the density for left-truncated subjects (4) remains intractable.

The unknown entry rate is not the only problem for (4). The usage of time-varying covariates is a major feature of event-history analysis. Time-varying covariates offer an opportunity to examine the relation between the risk of an event and the *changing* conditions under which the event occurs. For a left-truncated subject, however, the values of time-varying covariates before τ_0 are almost always unobserved. When density (4) is used, we are forced to abandon basically all time-varying covariates even though these covariates are observed during the observation period.

In summary, when the start time is not known, left truncation remains a very difficult problem unless we are willing to assume a constant hazard rate or to delete all left-truncated subjects. The problem becomes much easier, however, when the start time is known. In the rest of this paper, we focus on a conditional likelihood approach to left truncation when start times are known. The conditional approach does not have the drawbacks of the approaches we have described in this section. We will show that many flexible hazard models based on the approach can be easily estimated with commercial statistical packages that are widely available. We believe that the conditional approach deserves special attention from social scientists because of the distinctive characteristics of social scientists' event-history data that we will describe in the next section.

3. SOCIAL SCIENTISTS' EVENT-HISTORY DATA

3.1. *Longitudinal Surveys*

Event-history analysis was originally developed in the biomedical sciences and engineering. Typical data in these fields consist of the survival times of cancer patients or the failure times of equipment components. In social sciences, the typical data are the waiting times of social and economic events in the life of human subjects. In the present context, at least two characteristics of social scientists' event-history data should be taken into consideration when statistical methods are selected. First, when start times fall before the observation period, social scientists' human subjects are much more likely to be aware of the start of the exposure to a life event (date of marriage in the PSID example) than cancer patients, for example, to be aware of the beginning date of their cancer. Second, social scientists' event-history data are often collected by a longitudinal survey that follows a respondent for many years. Before discussing these characteristics in detail, we give a brief description of several major longitudinal social surveys.

The past 25 years have seen the start and growth of several nationally representative longitudinal surveys of households and individuals. Two of the most important ones were initiated in the 1960s: the National Longitudinal Surveys of Labor Market Experience (NLS) and the Panel Study of Income Dynamics (PSID) (Center for Human Resources Research 1991). The NLS has collected five separate longitudinal data bases: men aged 45 to 59 in 1966, young men aged 14 to 24 in 1966, women aged 30 to 40 in 1967, young women aged 14 to 24 in 1968, and youth aged 14 to 21 in 1979. Each of the first four data bases contains approximately 5,000 individuals and the last one contains more than 10,000 young men and young women. The PSID data set contained about 6,000 households and 15,000 individuals at the beginning of the survey in 1968. With a few exceptions, annual interviews were conducted. Most of the surveys have continued to the present. More recent years saw the start of the U.S. Survey of Income and Program Participation (SIPP) and the German Socio-Economic Panel (SOEP) (U.S. Department of Commerce 1987; Deutsches Institut 1991). Both have been conducted much the same way as the NLS and the PSID. The results are a wealth of information on the respondent's history of marriage, fertility, labor force participation, education, personal income, health, and so on.

Taken together, these and other longitudinal surveys have been and will probably remain primary data sources for social scientists' event-history analysis.

3.2. *Acquiring Start Times*

The conditional approach requires that the start time of a left-truncated case be known—that is, that we have data like subject *B* rather than *A* in Figure 1. It is often difficult to ascertain the outset of a cancer, for instance, in an analysis of cancer patients. The situation changes radically, however, when human individuals and their important life events are the subjects of event-history analysis.

The date of an important life event can be recalled even years after it occurred. Moreover, the dates of the past events should be recalled relatively accurately in social surveys because social scientists tend to be interested in only well-defined, important social and economic occurrences. Freedman et al. (1988) report a remarkable correspondence between the data acquired in 1980 about current life events and the data about these events recalled retrospectively in 1985 via the life-history-calendar technique. It has often been suggested that the social scientists' task is more complicated because their subjects are people. But in this case, human subjects work to our advantage: People have memories.

When the PSID started in 1968, the start times of most first marriages were not collected. This was the case until 1985 when retrospective marital histories were taken, which naturally include the start date of the first marriage. Subject A in Figure 1 thus becomes subject B, a left-truncated case with known start time. In fact, this information could have been acquired in any of the 17 yearly surveys conducted between 1968 and 1985. We have checked the reliability of the retrospective information by comparing the date of the first marriage reported retrospectively in 1985 against the date reported in the yearly surveys among those who married after 1968. The two sets of reports are quite consistent. In the more recent SIPP and SOEP, retrospective questions on dates of past events were administered much earlier during the observation period than they were in the PSID.

It should be recognized that a substantial amount of event-history data is from archival sources from which acquiring start times is more difficult. Unlike the observations in an ongoing survey, the archival subjects usually cannot be reached again for more information. Nevertheless, researchers working with left-truncated archival data should be aware that start times are sometimes available.

One such example arises from the official data on the life insurance industry in New York (Lehrman 1989). Most of the data come from the *Annual Report of the Superintendent of the Insurance Department*. These annual reports include abstracts of the annual financial statements of all assessment companies and fraternal societies actively selling life insurance in New York state from 1881 onward. The event of interest here is the failure of these insurance businesses. The problematic or left-truncated subjects are those that started operating before 1881, the beginning of the observation period. These businesses tend to be more successful. The less successful ones tend to fail before 1881 and thus are excluded from the official reports. Analyzing the left-truncated businesses as if they were regular subjects would lead to an underestimation of failure rates. Fortunately, the annual reports contain the date of creation, or start time, for each business, which can be used in the conditional likelihood approach to address the problem of sample selection.

3.3. *Asymetry of the Preobservation and Observation Periods*

Previous work on left truncation tends to view the preobservation and observation periods as more or less symmetrical with respect to the amount of information they contain. In the development of equation (6), a left-truncated case is considered to consist of the information (t_r) from the preobservation period and the information (t_p) from the observation period. Then, the joint distribution of T_r and T_p is sought for the construction of the likelihood function. The unobservability of T_r requires integration of the joint distribution under parametric assumptions. The inclusion of t_r in this manner is preferred because it increases the statistical efficiency of the estimation.

In longitudinal social surveys, however, the distribution of information between the preobservation and observation periods is heavily skewed to the latter. During the observation period, the respondents are interviewed year after year, even three times a year in the SIPP. Not surprisingly, an enormous amount of information would accumulate on the respondent's changing social, economic, and personal circumstances during the observation period. By 1988, the PSID had accumulated nearly 15,000 variables, almost all of which are concerned with the observation period. Although dates of important life events before the observation period can be retrieved quite accurately through retrospective questions, the possibility of obtaining information such as detailed income and employment history from the distant past is extremely slim.

One's objectives may be limited to estimating the parameters in the distribution of the survival time and its relation to covariates, such as gender, that do not change in a respondent's lifetime. In that case, the preobservation and observation periods may well be viewed as symmetrical. But social scientists are usually interested in much more complex questions, the answers to which demand detailed information over time. In some recent studies on the relationship between individual divorce rates over time and women's economic status, for instance, an examination of the detailed income history of women is essential (e.g., Korenman, Guo, and Geronimus 1991; Dechter 1992).

4. A CONDITIONAL LIKELIHOOD APPROACH

4.1. *Conditional Likelihood*

The examination of longitudinal social surveys reveals that most complications occur before the observation period, and most information is contained in the observation period. In this section, we describe an approach of conditional likelihood that avoids the complications in the preobservation period and maximizes the use of the available information in the observation period.

Recall the two problems of left truncation: unknown start time and sample selection. Availability of start times removes the first, but the second remains. Treating a left-truncated subject with a known start time as if it were a standard subject still leads to an underestimation of hazard rates at shorter durations (to be illustrated by an example in Section 6). The conditional likelihood approach addresses the problem of sample selection by conditioning the density of a left-truncated case on the case's having survived to t_r, the amount of time the case spends in the preobservation period. The idea is to use only those pieces of exposure that have not been selected.

The idea has been around for many years. Schoen (1975) studied increment-decrement life tables that a subject can enter at different durations, so that the selection bias can be eliminated. The increments mean entrants here. Thompson (1977) suggests that observations be grouped into intervals when ties are present and that the likelihood function be constructed for each interval conditionally on having made it to the beginning of the interval. Lancaster (1979) and Elandt-Johnson and Johnson (1980, p. 349) also describe the conditional likelihood.

The conditional density of a left-truncated case is defined as

$$f_T[t|T>t_r,x(t)] = \frac{\lambda[t|x(t)]\ S_T\ [t|\ \{x(u)\}_{u=0}^{u=t}]}{S_T[t_r|\{x(u)\}_{u=0}^{u=t_r}]}, \tag{7}$$

$$= \lambda[t|x(t)]\ exp\ [-\int_{t_r}^{t}\lambda(u)\ e^{x(u)'\beta}du], \tag{8}$$

where $\{x(u)\}_{u=0}^{u=t_r}$ and $\{x(u)\}_{u=0}^{u=t}$ stand for the time-varying covariates from 0 to t_r and from 0 to t, respectively. Note that u in (7) is used in exactly the same way as that in the integral in (8) to represent the points along t. The second term in (8) can be interpreted as the survivor function of T given having survived from $t=0$ to $t=t_r$. The values of time-varying covariates need to be observed only in the observation period from t_r to t. This implies that all information required in the conditional likelihood approach is supplied from the observation period except the start time of t. Note that model (8) assumes that the covariates act multiplicatively on the hazard function.

Except for t_r, the information required by the conditional approach amounts to the collection of all the solid lines (both exposure and accompanying covariates) within the observation period in Figure 1. The interrupted lines that extend to the left of τ_0 or to the right of τ_1 are excluded. These required data correspond almost exactly to the data normally collected during an observation period by a longitudinal social survey. In this sense, the conditional approach is probably the most effective way of using the collected information. In contrast, discarding left-truncated cases equals wasting the solid parts of subjects A, B, and F in Figure 1. The data collected in the observation period in a social survey bear a resemblance to those for the construction of a period life table. The estimates from analyzing these data via the conditional approach can be readily interpreted in terms of a calendar period. On the other hand, the same data with left-truncated subjects excluded resemble those for the construction of a cohort life table. An analysis of these data is relevant only to the cohort that starts in the observation period.

The conditional likelihood (8) can be alternatively derived by conditioning $f(t_p)$ (4), the density for left-truncated cases with unknown start times, on t_r. The resulting conditional density (8) is much simpler than (4) and does not depend on the entry rate $h(-t_r)$. The conditional likelihood approach has been compared with approach (6), and criticized for being less efficient than (6), since the information contained in t_r is used only for eliminating the complications caused by left truncation and not for estimating parameters β (Hamerle 1991). There seems to be more, however, to the comparison of the two approaches than efficiency. Whereas the more efficient approach (6) depends on the convenient assumption of time homogeneity, the less efficient conditional likelihood does not. Besides, approach (6) is more efficient only in the use of information on the distribution of time. When time-varying covariates are available, the conditional likelihood approach is much more "efficient" than approach (6), which is usually compelled to give up the use of any time-varying covariates collected in the observation period because these covariates are almost never observed before the observation period.

4.2. *Continuous-Time Piece-Wise Exponential Models*

A key issue for most data analysts is the accessibility of the estimation methods. The maximum likelihood estimates from this conditional likelihood function are, in general, different from those yielded by the methods for conventional event-history data. For instance, when T has a Weibull distribution with density $f(t) = \lambda p(\lambda t)^{p-1} exp[-(\lambda t)^p]$, the cumulative hazard functions $\int_0^t \lambda(u;x)\, du$ and $\int_{t_r}^t \lambda(u;x)\, du$ yield different expressions of $(\lambda t)^p e^{x'\beta}$ and $(\lambda)^p[t^p - (t_r)^p]e^{x'\beta}$ respectively. This means that software developed for standard event-history analysis cannot be used for the type of data under consideration unless the more general form of the cumulative hazard function [$\int_{t_r}^t \lambda(u;x)\, du$, which includes $\int_0^t \lambda(u;x)\, du$ as a special case] is a built-in feature in the software.

In this and the next sections, we will show that piece-wise exponential hazard models based on the conditional approach can be estimated by any package estimating a Poisson regression and that the conditional discrete-time models can be estimated by any package estimating a logistic regression. The important point is that these results are not package-specific. A data analyst has to know only how to estimate a Poisson model or a logistic model, which can be estimated by almost any general commercial statistical package.

Because of its flexible functional form, the piece-wise exponential model is one of the most widely used hazard models in the social sciences. The basic idea is simple. We divide the maximum duration into a number of intervals. The baseline hazard within each interval is assumed to be constant, but the baseline hazards are allowed to vary across intervals. The baseline hazards are a vector of parameters that determine the shape of hazard rate over time.

We first consider data that are not left-truncated. We construct J duration intervals with $J+1$ cut points $0 = \tau_1 < \tau_2 < \ldots < \tau_{J+1} = \infty$ and assume that the baseline hazard $\lambda(t)$ takes the value λ_j when t is in interval j–that is, $\tau_j < t < \tau_{j+1}$. Define t_j as the exposure time spent in the j-th interval by a subject. Define δ_j as one if the subject experiences the event in the interval and as zero otherwise.

Holford (1980) shows that the log-likelihood of the above model is equivalent to the kernel of the log-likelihood that treats the event indicators $\delta_j s$ as if they were independent Poisson random variables with means $\mu_j = t_j \lambda_j e^{x'\beta}$. Note that the means depend on a proportional hazards assumption. We point out that this result holds for left-truncated data with known start time as well. We write the conditional log-likelihood of a piece-wise exponential hazard model for subject i as

$$logL_i = \sum_{j=j_i}^{m_j} \delta_{ij}(log\lambda_j + x_i'\beta) - \sum_{j=j_i}^{m_j} \lambda_j t_{ij} e^{x_i'\beta}, \tag{9}$$

where interval m_i is the interval in which subject i is last observed with $m_i = J$ as a special case when subject i is last observed in interval J and interval j_i is the interval in which subject i is first observed with j_i taking the value *1* as a special case when the first interval starts within the observation period. Adding to and subtracting from (9) the quantity $\sum_{j=j_i}^{m_j} \delta_{ij} logt_{ij}$ yield

$$\log L_i = \sum_{j=j_i}^{m_j} [\delta_{ij} \log(\lambda_j t_{ij} e^{x_i'\beta}) - \lambda_j t_{ij} e^{x_i'\beta}] - \sum_{j=j_i}^{m_j} \delta_{ij} \log t_{ij}. \tag{10}$$

The first summation term can, again, be considered equivalent to the kernel of the log-likelihood of independent Poisson random variables with the log of exposure log t_{ij} and linear predictors $\alpha_j + x_i'\beta$, where $\alpha_j = \log \lambda_j$. This differs from the case Holford (1980) considered in that the summation now can start from any interval. But this difference is of no consequence for using Poisson regression routines in standard statistical software. The routines treat each interval as if it were an independent observation without regard to whether its previous interval is observed or not. The second summation term is a constant and therefore may be ignored in the maximization of the likelihood. An empirical example is provided in Section 6 to show how estimation based on (10) can be accomplished in practice. When the explanatory variables are constant over time, a piece-wise exponential hazard model can also be estimated with a program for exponential accelerated life models.[2]

4.3. *Discrete-time Models*

Discrete-time models have become one of the most popular class of hazard models for social scientists since Allison's (1982) work, because of their connection in conceptualization with the familiar statistical models for binary data, their uncomplicated way of incorporating time-varying covariates, and their ease of estimation via standard software developed for binary data. In this section, we point out that

[2]Pointed out by an anonymous reviewer.

the whole class of discrete-time models based on the conditional likelihood can be estimated by the same standard software for binary data.

The discrete-time models arise when the random variable T takes values at distinct $t_1 < t_2 < \ldots < t_J$ with the corresponding probability function

$$f(t) = Pr(T=t) = \lambda_J \prod_{j=1}^{J-1} (1-\lambda_j), \tag{11}$$

and survivor function

$$S(t) = Pr(T > t) = \prod_{j=1}^{J} (1-\lambda_j), \tag{12}$$

where the discrete-time hazard rate $\lambda_j = Pr(T=t_j | T \geq t_j)$ is the conditional probability of experiencing the event at t_j.

To develop the conditional likelihood, we again use t_r (Figure 2) to represent the length of exposure before the observation period with $t_r \leq \tau_0 < t_{r+1}$, where t_r and t_{r+1} are two adjacent distinct survival times. Using δ_i again as the event indicator, conditioning the likelihood on having survived to t_r, taking logs, and multiplying over the sample, we obtain the conditional log-likelihood for the discrete-time models

$$\log L = \sum_{i=1}^{n} \sum_{j=1}^{J} \delta_{ij} \log \left(\frac{\lambda_{ij}}{1-\lambda_{ij}} \right) + \sum_{i=1}^{n} \sum_{j=r+1}^{J} \log (1-\lambda_{ij}). \tag{13}$$

Since the hazard rate $\lambda_j = 1 - S(t_{j+1})/S(t_j)$ can be treated as a probability, (13) can be considered the log-likelihood of binomial data except that the second summation in the second term sums from $r+1$, rather than 1. Whenever $\log(1-\lambda_{ij})$ in the second term is conditioned away (when $j \leq r$), however, its corresponding part in the first term of (13) is always zero, since by definition only the last of the series of trials may be a success in a discrete-time model. Thus the conditional log-likelihood (13) remains a binomial log-likelihood.

The hazard rate λ_{ij} for observation i at time j may, in turn, be linked to the observation's covariates through one of the four link functions: the logistic, probit, complementary log-log, and log-log functions (McCullagh and Nelder 1989; Kalbfleisch and Prentice

1980, ch. 2). The logistic link function is by far the most popular choice for discrete-time models partly because the software for logistic regressions is widely available. The empirical example in Section 6 will demonstrate how a model based on (13) can be estimated.

5. CONDITIONAL PARTIAL LIKELIHOOD

The methods discussed so far all need parametric assumptions on the functional form of the baseline hazards. The information on which such assumptions are made is often scarce. In standard event-history analysis, Cox's (1972) partial likelihood hazard model has been used extensively to avoid making such assumptions. In this section, we describe an extension of Cox's partial likelihood, so that it can be used for left-truncated event-history data with known start times. See Helsen (1990) for another discussion of this approach.

Consider a situation in which there are J observed events among a total of K survival times and in which events occur at distinct ordered $\tau_1 < \tau_2 < \ldots < \tau_J$. The extension to Cox's partial likelihood is based on the same idea as that behind the conditional likelihood, hence the term conditional partial likelihood. We start from the full conditional likelihood (7). Following Cox's original development, we factorize the full likelihood into two products. Retaining only the product of the likelihood that contains the information on the order, rather than the exact time, of events, and canceling baseline hazards, we obtain the conditional partial likelihood function

$$L_P = \prod_{j \in J} \frac{e^{x_j' \beta}}{\sum_{k:\, t_k \geq \tau_j,\, t_r \leq \tau_j} e^{x_k' \beta}}. \tag{14}$$

This differs from the standard partial likelihood in that the summation space in the denominator is restricted by one additional condition, $t_r \leq \tau_j$, where t_r is again the length of time that has elapsed before the observation period. This restriction ensures that all the left-truncated cases are excluded from the risk set at τ_j if they have not entered the observation period at τ_j, even though the total length of the survival time of this left-truncated case is longer than τ_j.

Figure 3 illustrates the conditional partial likelihood. Unlike that in Figure 1, the axis indicates duration time. All event histories are left-aligned. The peculiar feature of the conditional partial likeli-

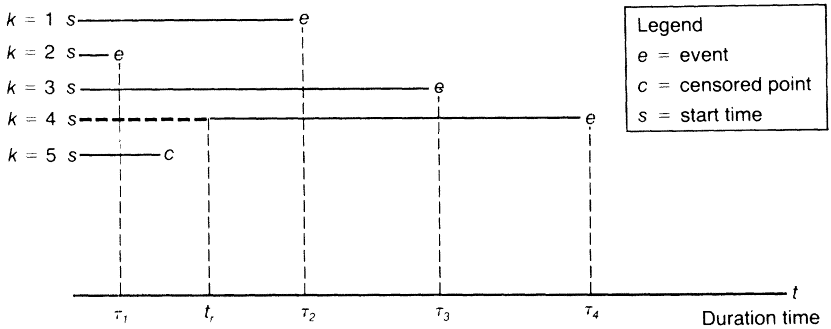

FIGURE 3. The conditional partial likelihood.

hood is illustrated by case $k=4$. Case 4 comes into observation after having spent t_r periods of time. As in (7), the length of t_r is used to determine the position of the observed part of case 4 along t. The important point is that, at τ_1, the risk set is $\mathcal{R}_1 = \{1, 2, 3, 5\}$ rather than $\mathcal{R}_1 = \{1, 2, 3, 4, 5\}$. Case 4 has yet to be under observation to be included in the risk set. The conditional partial likelihood can then be treated as an ordinary likelihood exactly as Cox's (1972) partial likelihood is. Helsen (1990) shows that a program for a partial likelihood model such as BMDP2L can be manipulated to estimate a conditional partial likelihood model.

6. AN EMPIRICAL EXAMPLE OF MARITAL DISSOLUTION IN THE U.S.

In this section we provide an empirical example of marital dissolution in the United States, using the PSID data that we have referred to frequently. This example tries to achieve two objectives. First, the example demonstrates how standard statistical packages may be used to estimate a conditional likelihood piece-wise exponential model and a conditional likelihood discrete-time model. A conditional partial likelihood model will also be estimated. Second, the example shows the consequences of ignoring left truncation when the start time is known.

The event of interest is first divorce over the period between 1970 and 1987. We select those women who were aged 65 or younger in 1970, who had married at least once by 1985, and who reported positive durations married. This sample includes 3701 first marriages. We rely on the 1985 retrospective questions to construct marital histories. Marital duration is calculated from the start of the first marriage to the date of the first separation. About 28 percent of the sample are blacks.

We first describe the procedures of estimating a conventional piece-wise exponential hazard model (model 1b in Tables 1 and 2) via a Poisson regression. To prepare data for a Poisson regression, we

TABLE 1
Estimates of the Baseline Hazard Rates or Odds Ratios of First Divorce in the U.S. from a Piece-wise Exponential Model (1a and 1b) and a Discrete-time Model (2a and 2b).[a]

	A Conditional Likelihood Approach			
	(1a)		(2a)	
	$exp(\beta)$[b]	$\|t\|$[c]	$exp(\beta)$	$\|t\|$
Baseline hazards or odds ratios (years)				
0–3	0.0035[d]	88.4	0.0423[f]	48.5
4–8	0.0025	3.6[e]	0.0301	3.6
9–15	0.0016	7.5	0.0190	7.7
16–40	0.0006	16.1	0.0075	16.3
	Standard Models Ignoring Left Truncation			
	(1b)		(2b)	
Baseline hazards or odds ratios (years)				
0–3	0.0019	98.0	0.0235	58.3
4–8	0.0014	3.4	0.0170	3.5
9–15	0.0008	7.7	0.0105	7.9
16–40	0.0004	14.1	0.0052	14.3

[a]Models (1a) and (2a) are based on a conditional likelihood approach. Models (1b) and (2b) show the consequences of ignoring left truncation. N = 3701.

[b]Relative risks.

[c]Absolute values of t statistics.

[d]Monthly rate of divorce.

[e]The t-statistics for the second through the fourth intervals test whether the baseline hazard rates in these intervals are significantly different from the rate for the first interval. The same can be said for the discrete-time model.

[f]Yearly odds ratio of divorce.

TABLE 2
Estimates of Cohort and Race Effects on First Divorce Rates in the U.S. from a Piece-wise Exponential Model (1a and 1b), a Discrete-time Model (2a and 2b), and a Partial Likelihood Model (3a and 3b).[a]

	A Conditional Likelihood Approach					
	(1a)		(2a)		(3a)	
	$exp(\beta)$[b]	$\|t\|$[c]	$exp(\beta)$	$\|t\|$	$exp(\beta)$	$\|t\|$
Baseline hazards or odds ratios (years)						
0–3	0.016[d]	22.8	0.187[f]	9.2	na[g]	na
4–8	0.013	2.3[e]	0.149	2.4	na	na
9–15	0.010	4.2	0.113	4.6	na	na
16–40	0.008	4.3	0.088	4.8	na	na
Age in 1970	0.961	7.5	0.960	7.4	0.977	3.8
Black	1.000		1.000		1.000	
White	0.665	5.6	0.663	5.6	0.666	5.6
	Standard Models Ignoring Left Truncation					
	(1b)		(2b)		(3b)	
Baseline hazards or odds ratios (years)						
0–3	0.064	18.3	0.723	2.1	na	na
4–8	0.059	0.8	0.663	1.0	na	na
9–15	0.056	1.2	0.602	1.7	na	na
16–40	0.088	2.4	0.907	1.7	na	na
Age in 1970	0.913	21.9	0.915	21.4	0.913	20.9
Black	1.000		1.000		1.000	
White	0.661	5.7	0.659	5.7	0.660	5.7

[a]Models (1a), (2a), and (3a) are based on a conditional likelihood approach. Models (1b), (2b), and (3b) show the consequences of ignoring left truncation. N = 3701.

[b]Relative risks.

[c]Absolute values of t statistics.

[d]Monthly rate of divorce.

[e]The t-statistics for the second through the fourth intervals test whether the baseline hazard rates in these intervals are significantly different from the rate for the first interval. The same can be said for the discrete-time model.

[f]Yearly odds ratio of divorce.

[g]Not applicable.

break the observed duration time into pieces of exposure according to the specified intervals: 0–3, 4–8, 9–15, and 16–40 years. The resulting pieces of exposure are treated as independent cases of a Poisson regression. For a woman who divorces at the end of the nineteenth year of marriage, we would have four pieces of exposure with lengths of 3, 5, 7, and 4 years. The last piece of four years is the amount of time the woman spends in marriage during the fourth interval of 16 to 40 years. A set of dummy variables needs to be created to index the four pieces of exposure and capture interval effects. Corresponding to each of the four pieces of exposure is a divorce indicator and covariates. Divorce indicators are usually coded as 1 if the woman divorces during the exposure or 0 if the marriage remains intact. Only the last indicator can be 1. For this woman, the divorce indicators are 0, 0, 0, and 1 for the four pieces of exposure, respectively.

In this example, we consider only two variables: age in 1970 and race. Age in 1970 is expected to capture the cohort effect of rising divorce rate over the past 20 years. The values of the covariates such as race and age in 1970 are the same for the four pieces of exposure unless the covariates are time-dependent. A Poisson regression procedure treats the divorce indicator as the dependent variable and pieces of exposure as a fixed intercept to be included as part of the linear predictor. The resulting parameter estimates can be interpreted as those of a piece-wise exponential hazard model. The set of dummy variables gives the baseline hazards.

When a case is left-truncated, only slightly more data work is needed for estimating a piece-wise exponential model using the conditional approach (model 1a in Tables 1 and 2). Suppose that the woman still divorced at the end of the nineteenth year, but did not enter the observation period until the end of the fifth year of marriage. We would then have three pieces of exposure with lengths of 3, 7, and 4 years for intervals 2, 3, and 4, respectively. Discarded is the exposure of three years in the first interval and the first two years of exposure in the second interval. The remaining three pieces of exposure with their corresponding divorce indicators and covariates can then be treated as independent cases. The rest remains the same. Model (1a) in Tables 1 and 2 are estimated using a Poisson regression procedure in SAS (SAS Institute 1990, p. 224).

The data preparation for a standard discrete-time model

(model 2b in Tables 1 and 2) is very similar to that for a piece-wise exponential model. We use year as the unit of discrete time. Each time unit is viewed as an independent Bernoulli trial. A woman who divorces at the end of the nineteenth year of marriage is considered having experienced 19 trials. Corresponding to these 19 trials are 19 trial success indicators, which are defined as 1 if a divorce occurs and 0 otherwise. Only the indicator for the last trial can possibly be one. Accompanying each trial is one or more covariates. Time-varying covariates can be easily incorporated by changing the value of a covariate at each time unit. Dummy variables can be used to capture duration effects. The same specifications of duration effects as those in model (1a) in Tables 1 and 2 amount to assigning the first three trials to group 1, the next five trials to group 2, the next seven trials to group 3, and the last 25 trials to group 4. Alternatively, duration effects can be captured by a linear or quadratic function of time since the beginning of marriage (Allison 1982).

For a left-truncated subject first observed at τ_0 at the end of the fifth year of marriage, the only additional thing that a conditional discrete-time model (model 2a in Tables 1 and 2) requires is to drop the first five trials from the analysis. The results are obtained using the logistic procedures in SAS (SAS Institute 1990, p. 175). The default logit link function is used in this example. SAS (p. 190) and GLIM (Payne 1986) also offer the option of the complementary log-log link function.

Models (3a) and (3b) in Table 2 present estimates from a conditional partial likelihood hazard model. We estimate the model by maximizing the conditional partial likelihood function (14), using a general purpose numerical optimization package (Quant and Goldfeld 1989).

Table 1 compares the baseline risks of marital dissolution estimated by the conditional approach with those estimated by standard event-history analysis. Models (1a)—piece-wise exponential—and (2a)—discrete-time—are based on the conditional approach. The risk of divorce is assumed to be constant within each of the four marital durations: 0–3 years, 4–8 years, 9–15 years, and 16–40 years, but the levels of the risk across the four durations are allowed to vary. Marital durations longer than 40 years are censored at the end of the fortieth year. Both models (1a) and (2a) show a monotonically declining risk of divorce over marital duration. In contrast, models

(1b) and (2b) are estimated by standard event-history analysis that makes use of the information on start time but ignores sample selection due to left truncation. As we predicted in Section 1.3, ignoring sample selection has led to an underestimation of the risks of marital dissolution. The risks of divorce estimated by models (1b) and (2b) are much lower than those estimated by models (1a) and (2a).

Table 2 examines the impact of ignoring sample selection on the effects of covariates. According to the conditional approach, shown by models (1a) and (2a), the rate of divorce is negatively correlated with duration, implying that the longer the woman has been married, the less likely she experiences a divorce. Age in 1970 is negatively related to risk of divorce, confirming our expectation that an older cohort tends to experience lower divorce rates. White Americans experienced divorce rates about 34 percent lower than Black Americans from 1970 to 1987. The three sets of results from models (1a), (2a), and (3a) are very similar except that the age effect estimated by the conditional partial likelihood hazard model (3a) is somewhat smaller than those estimated by models (1a) and (2a).

The three standard hazard models (1b, 2b, and 3b) that ignore left truncation also yield very similar results among themselves (Table 2). These results, however, are very different from those estimated by the three conditional hazard models (1a, 2a, and 3a). In contrast to the monotonically declining baseline risks of divorce in models (1a) and (2a), the baseline risks of divorce in the first three intervals (0–3, 4–8, and 9–15) in models (1b) and (2b) are not significantly different from one another and the baseline risk in the last interval (16–40) is the highest of all. Moreover, the age effects estimated in models (1b), (2b), and (3b) are much larger (0.91 vs. 0.96). than those estimated in models (1a), (2a), and (3a). Considering that age is used as a continuous variable, these differences are quite large.

The problems with models (1b), (2b), and (3b) are directly related to sample selection. The sample selection occurs when some high-risk women divorced and dropped out of the sample before 1968. As a result, women with left-truncated marriage histories (those who married before 1968 and whose marriage remained intact beyond 1968) tend to have lower rates of divorce. The problem is that the left-truncated women also tend to be older, resulting in an exaggeration of age effect. Apparently, the age variable in models

(1b), (2b), and (3b) has picked up some of the effect of sample selection.

7. SUMMARY AND CONCLUSIONS

A subject is left-truncated when it comes under observation after having been exposed to the risk of an event for some time; the amount of time may or may not be known. The characteristic problem of left truncation is sample selection. Left truncated subjects tend to have lower risks at shorter durations than do subjects in a normal sample.

When the length of exposure before a left-truncated subject comes under observation is unknown, left truncation is practically intractable except for the special cases in which the hazard rate is constant or all left-truncated subjects are discarded. When the length of exposure or start time is known, left-truncated data can be handled by the conditional likelihood approach. This approach conditions the likelihood function on the subject's having survived to the start of the observation period. We believe that the conditional approach deserves special attention from social scientists not only because it is free from the complications caused by left truncation but also because of the characteristics of social scientists' event-history data. The critical information on the start time of a left-truncated subject can be frequently obtained from human subjects in social surveys even years after the event occurs. The concentration of covariate information in the observation period in a longitudinal social survey is also well-suited for the conditional approach.

An important practical issue is how to estimate a model based on the conditional approach since the software designed for conventional event-history analysis can no longer be used in general. We show that a piece-wise exponential hazard model can still be estimated readily using any software that estimates a Poisson regression, and that a discrete-time model can still be calculated easily via any package designed for binary data.

Additionally, we describe a conditional partial likelihood approach that combines Cox's partial likelihood and the conditional likelihood. The advantage is that no parametric assumption on the functional form of the baseline hazard is needed.

REFERENCES

Allison, Paul D. 1982. "Discrete-Time Methods for the Analysis of Event Histories." In *Sociological Methodology 1982,* edited by S. Leinhardt, 61–98. San Francisco: Jossey-Bass.

———. 1984. *Event History Analysis, Regression for Longitudinal Event Data.* Beverly Hills, Calif.: Sage.

Center for Human Resources Research. 1991. *NLS Handbook 1991.* Columbus: Center for Human Resources Research, Ohio State University.

Cox, David R. 1962. *Renewal Theory.* London: Methuen.

———. 1972. "Regression Models and Life-tables (with Discussion)." *Journal of the Royal Statistical Society B* 34:187–220.

Cox, David R., and D. Oakes. 1984. *Analysis of Survival Data.* New York: Chapman and Hall.

Dechter, Aimée R. 1992. "The Effect of Women's Economic Independence on Marital Dissolution." Paper presented at the 1992 annual meeting of the Population Association of America, Denver.

Deutsches Institut für Wirtschaftsforschung. 1991. *Das Sozio-ökonomische Panel (SOEP): Benutzerhandbuch.* Berlin.

Elandt-Johnson, Regina C., and Norman L. Johnson. 1980. *Survival Models and Data Analysis.* New York: Wiley.

Freedman, Deborah, Arland Thornton, Donald Camburn, Duane Alwin, and Linda Young-DeMarco, 1988. "*The Life History Calendar: A Technique for Collecting Retrospective Data.*" In *Sociological Methodology 1988,* edited by Clifford Clogg, 37–68. Washington, DC: American Sociological Association.

Hald, A. 1952. *Statistical Theory with Engineering Applications.* New York: Wiley.

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

Heckman, James, and B. Singer. 1986. "Econometric Analysis of Longitudinal Data." In *Handbook of Econometrics,* vol. 3, edited by Z. Griliches and M.D. Intriligator, 1689–1763. Amsterdam: North-Holland.

Helsen, Kristiaan. 1990. "New Developments in Duration Time Modeling with Applications to Marketing." Ph.D. diss., University of Pennsylvania.

Holford, T. R. 1980. "The Analysis of Rates and Survivorship Using Log-linear Models." *Biometrics* 36:299–305.

Kalbfleisch, J. D., and R. L. Prentice. 1980. *The Statistical Analysis of Failure Time Data.* New York: Wiley.

Keiding, Niels. 1986. "Nonparametric Estimation Under Truncation." In *Encyclopedia of Statistical Sciences,* vol. 9, edited by Samuel Kotz and Norman L. Johnson, 357–59. New York: Wiley.

Korenman, Sanders, Guang Guo, and Arline T. Geronimus. 1991. "Relative Wage and Marital Separation." Paper prepared for the PSID Event History Conference, Stanford University, June.

Lancaster, T. 1979. "Econometric Methods for the Duration of Unemployment." *Econometrica* 47:939–56.

Lehrman, William G. 1989. "Organizational Form and Failure in the Life Insurance Industry." Ph.D. diss., Princeton University.

Leiderman, P. H., D. Babu, J. Kagia, H.C. Kraemer, and G.F. Leiderman. 1973. "African Infant Precocity and Some Social Influences during the First Year." *Nature* 242:247–49.

McCullagh, P., and J. A. Nelder. 1989. *Generalized Linear Models.* London: Chapman and Hall.

Miller, Rupert G., Jr. 1981. *Survival Analysis.* New York: Wiley.

Payne, C. D., ed. 1986. *The GLIM System, Release 3.77.* Oxford: Numerical Algorithms Group.

Quant, R. E., and S. M. Goldfeld. 1989. *GQOPT/I.* (Version 5.00). Princeton: Department of Economics, Princeton University.

Ridder, G. 1984. "The Distribution of Single Spell Duration Data." In *Studies in Labor Market Dynamics,* edited by G. R. Neumann and N. C. Westergard-Nielsen, 45–73. New York: Springer.

SAS Institute. 1990. *SAS Technical Report P-200, SAS/STAT Software: CALIS and LOGISTIC Procedures, Release 6.04.* Cary, N.C.

Schoen, Robert. 1975. "Constructing Increment-Decrement Life Tables." *Demography* 12:313–24.

Survey Research Center. 1989. *A Panel Study of Income Dynamics: Procedures and Tape Codes.* Ann Arbor: Institute for Social Research, University of Michigan.

Thompson, W. A., Jr. 1977. "On the Treatment of Grouped Observations in Life Studies." *Biometrics* 33:463–70.

Trussell, James, and Timothy Guinnane. 1991. "Techniques of Event History Analysis." Paper prepared for an IUSSP seminar on Old and New Methods in Historical Demography, Palma de Mallorca, Spain, June.

Tuma, Nancy B., and Michael T. Hannan. 1984. *Social Dynamics, Models and Methods.* New York: Academic Press.

Turnbull, B. W. 1974. "Nonparametric Estimation of a Survivorship Function with Doubly Censored Data." *Journal of the American Statistical Association* 69:169–73.

U.S. Department of Commerce and Bureau of the Census. 1987. *Survey of Income and Program Participation: Users' Guide.* Washington, D.C.

Yamaguchi, Kazuo. 1991. *Event History Analysis.* Newbury Park, Calif.: Sage Publications.