Nonparametric Hazard Estimation with Time Varying Discrete Covariates

Leung, Siu Fai and Wing Hung Wong

Working Paper No. 181
March 1989

Nonparametric Hazard Estimation with Time Varying Discrete Covariates

Siu Fai Leung and Wing Hung Wong

Rochester Center for Economic Research
Working Paper No. 181

NONPARAMETRIC HAZARD ESTIMATION WITH TIME VARYING DISCRETE COVARIATES

by

Siu Fai Leung* and Wing Hung Wong**

Working Paper No. 181

First Draft: July 1987
Current Version: March 1989

The traditional one-sample nonparametric estimator of the cumulative hazard function is extended to allow for time varying discrete covariates. A proof for the consistency and asymptotic normality of the estimator, as well as a derivation of the asymptotic standard error, is provided. Numerical results show that the estimator performs reasonably well.

---

* Department of Economics, University of Rochester.
**Department of Statistics, University of Chicago.

We thank the referees for some useful comments. Leung's research was supported in part by the Bradley Foundation and the Hewlett Foundation. Wong's research was supported in part by National Science Foundation Grant No. DMS-8601732.

Introduction

In this paper, we extend the one-sample nonparametric estimator of the cumulative hazard function proposed by Nelson (1972) to allow for time varying covariates. Time varying covariates often appear in econometric duration analyses. For example, the monthly wage rate is one of the time varying covariates frequently found in economic duration data. Transfer payments which vary within an unemployment spell are another example. In the literature on parametric estimation of the hazard function, Heckman and Singer (1984) show that ad hoc treatment of time varying covariates can yield wildly misleading estimates. They estimate a job turnover model and find that conventional methods such as replacing time trended variables by their within spell average values or by the beginning of spell values produce very different estimates. The alternative treatment they propose is to allow the covariates to vary freely in their computation algorithm. The importance of controlling for time varying covariates in parametric estimation of the hazard function is also emphasized in Kiefer's (1988) recent survey. On the other hand, little work has been done on nonparametric estimation with time varying covariates.

In this paper, we deal with time varying covariates which satisfy two conditions. First, they take discrete values and the number of the possible discrete values is small compared to the sample size. One example is categorical variables (or qualitative variables) such as several categories of wage rates or earnings. Thus, time varying continuous covariates such as age are not treated in this paper, unless it is judged appropriate to discretize them into a few groups.

In the literature on parametric estimation of the hazard function, it is frequently implicitly assumed that the time varying covariates have a Markovian effect on the hazard function.[1] A simple example may be useful to illustrate this point. Suppose a worker has a hazard rate of unemployment $\lambda(t;w(t))$ at time t, where w(t) is the wage rate at time t. Assume there are only two wage rates, w_1 and w_2. Suppose worker A has wage rate w_1 for $t \in [0,T)$ and wage rate w_2 at time T, then his hazard rates are correspondingly $\lambda(t;w_1)$ for $t \in [0,T)$ and $\lambda(T;w_2)$ at time T. Suppose an otherwise identical worker B has wage rate w_2 for $t \in [0,T]$. It follows that both workers have the same hazard rate of unemployment $\lambda(T;w_2)$ at time T even though they have different wage paths before time T. The Markovian property lies in the presumption that worker A immediately assumes the new hazard rate after the wage rate has changed, independent of his previous wage path. His hazard rate of unemployment at time T equals that of a person who may have been exposed to the risk of unemployment for a long period, say [0,T], with wage w_2. This is an important assumption and needs to be checked in actual applications. Here, we also assume that this Markovian property holds and this is the second condition we imposed on the time varying covariates discussed in this paper.

The paper is divided into three sections. Section 1 begins with a brief review of the Nelson estimator, which is then extended to situations when there are time varying discrete covariates. Section 2 shows that the estimator we propose is consistent and asymptotically normal. The proof itself is of independent interest. In the one-sample case, Breslow and Crowley (1974) and Meier (1975) first prove that the Nelson estimator of the cumulative hazard function, or equivalently the Kaplan-Meier estimator of the survival function (Kaplan and Meier 1958), is consistent and asymptotically normal. However, their proofs seem relatively involved, especially on the derivation of the covariance of the estimators (Breslow and Crowley 1974). Here we provide a simple and concise proof which makes use of the martingale central limit theorem.[2] One advantage of our proof is that it can be easily extended to handle time varying discrete covariates. Section 3 presents some numerical results which show that the estimators we propose perform reasonably well.

Section 1 : The Estimator

Consider the standard one-sample case without time varying covariates. Let $T \geq 0$ be the time the event of interest occurred, which may be a death, a birth or an unemployment, depending on the situation under study. Throughout this paper, we use the term 'event' and leave it unspecified. Accordingly, the term 'survive' means that the 'event' has not yet occurred, and the phrase 'number at risk' denotes the number of persons who survive. Assume a random censoring mechanism and let $C \geq 0$ be the censoring time. The random variables T and C are assumed to be independent. Consider a sample of n persons. Let $T_1, T_2, \ldots, T_n$ be independent and identically distributed with distribution function F(t), and $C_1, C_2, \ldots, C_n$ be independent and identically distributed with distribution function G(t). We observe the data $(Y_1,\delta_1), (Y_2,\delta_2), \ldots, (Y_n,\delta_n)$, where $Y_i=\min\{T_i,C_i\}$, $\delta_i=1$ if $T_i \leq C_i$ (uncensored), and $\delta_i=0$ if $T_i > C_i$ (censored). Suppose the Y_i's are ordered, with $Y_1 \leq Y_2 \leq \ldots \leq Y_n$. At time t, let us divide the interval [0,t) into K(t) subintervals $[0,t_2)$, $[t_2,t_3)$, ..., $[t_{K(t)},t)$. Let $\Delta_k=[t_k,t_{k+1})$, where $t_1=0$ and $t_{K(t)+1}=t$. Let $\lambda(t)$ be the hazard function associated with F(t), then the conditional probability that the event occurred in Δ_k is given by

q_k = P(the event occurred in Δ_k | the person survived to the beginning of Δ_k)

$$= \int_{\Delta_k} \lambda(y)dy$$

$$\simeq \lambda(t_k)\Delta_k$$

Let d_k=number of events occurred in Δ_k and r_k=number at risk at the beginning of Δ_k, then q_k can be estimated by $\hat{q}_k=d_k/r_k$. Notice that r_k=number at risk at $t_k = \Sigma_{m=1}^{n} I\{m: Y_m \geq t_k\}$, where I{A} is the indicator function of the set A, and it is assumed that $r_k > 0$ for k=1,2,...,K(t). The cumulative hazard function, which is sometimes called the integrated hazard, $\Lambda(t)=\int_0^t \lambda(y)dy= \sum_{k=1}^{K(t)} q_k$ is then estimated by $\hat{\Lambda}(t)= \sum_{k=1}^{K(t)} \hat{q}_k$. Let there be n persons at time 0 and suppose the K(t) is chosen so large that we can ignore the possibility of more than one event occuring in any interval. Set $\hat{q}_k$=0 if no event occurred in Δ_k, and $\hat{q}_k=1/(n-i+1)$ if an event corresponding to Y_i occurred in Δ_k, then we obtain the Nelson estimator (Nelson 1972):

$$\hat{\Lambda}(t) = \sum_{Y_i \leq t} \frac{\delta_i}{n-i+1} \tag{1}$$

Notice that the term (n-i+1) in (1) can also be written as $r(Y_i)$, where $r(s) = \Sigma_{m=1}^{n} I\{m: Y_m \geq s\}$. The Kaplan-Meier estimator (or the product limit estimator), which is an estimator of the survivor function S(t)=1-F(t), is given by $\hat{S}(t) = \Pi_{Y_i \leq t} \{1-[\delta_i/(n-i+1)]\}$. Although the Kaplan-Meier estimator may be more popular in the econometrics and statistics literature [see, e.g. Miller (1981 p.46-48)], we choose to study the Nelson estimator because, as we shall see in Section 2, the estimator contains a martingale structure which plays an important role in deriving the asymptotic properties of the estimator and in generalizing to the case with time varying discrete covariates.

In order to highlight the main ideas, we make the following simplifying assumptions. The whole approach and the proofs can easily be extended to more complicated situations. Consider a situation in which there is only one time varying discrete covariate with J possible values X_1, X_2, ..., X_J, where J is small relative to the sample size. At time 0, for each $j=1,2,\ldots,n$, let $N_j(0)$ be the number of persons whose covariates take on the value X_j, therefore, $n=N_1(0)+N_2(0)+\ldots+N_J(0)$ is the total number of people at time 0. For a person with covariate value X_j at time t, let $F_j(t)$ be the distribution of the time the event occurred and $G_j(t)$ be the distribution of the censoring time. A person can switch among the X_j's before the event or the censoring occurred. Assume that the switching time for a person with covariate value X_j is drawn from a distribution $H_j(t)$ which is independent of the distributions $F_j(t)$ and $G_j(t)$. Given these assumptions, we would observe the data $(Y_i,\delta_i,\{X_i^*(t),t \le Y_i\})$ for each person $i=1,2,\ldots,n$, where $X_i^*(t)$ is the ith person's covariate value at time t. Thus, the difference with the case without time varying covariates is that, apart from observing the information (Y_i,δ_i), we also observe the covariate history $\{X_i^*(t)\}$ before time Y_i.

Under this setup, our problem is to estimate the cumulative hazard function $\Lambda_j(t)$ for each covariate value X_j, $j=1,2,\ldots,J$. In other words, there are J cumulative hazard functions to be estimated, one for each covariate value. One may interpret $\Lambda_j(t)$ as the cumulative hazard function of a person who begins with covariate value X_j at time 0 and never changes the covariate value after time 0. The estimation method is similar to the previous case. For each person i, let $\delta_{ij} = 1$ if Y_i is the time the event occurred (i.e. uncensored) with covariate value $X_i^*(Y_i) = X_j$, and $\delta_{ij} = 0$ otherwise. At time t, divide $[0,t)$ into K(t) subintervals $\Delta_1, \Delta_2, \ldots, \Delta_{K(t)}$, where $\Delta_k=[t_k,t_{k+1})$, $t_0=0$, $t_{K(t)+1}=t$. Let

q_{kj} = P(the event occurred in Δ_k | the person survived to the beginning of Δ_k with covariate value X_j)

Then we estimate q_{kj} by $\hat{q}_{kj} = d_{kj}/r_{kj}$, where d_{kj}=number of events occurred in Δ_k with covariate value X_j and r_{kj}=number at risk at the beginning of Δ_k with covariate value X_j. Notice that r_{kj}=number at risk at t_k with covariate value $X_j = \Sigma_{m=1}^{n} I\{m\colon Y_m \geq t_k \text{ and } X_m^*(t_k) = X_j\}$. The estimator for the cumulative hazard function conditional on the the jth covariate value becomes

$$\hat{\Lambda}_j(t) = \sum_{k=1}^{K(t)} \hat{q}_{kj} = \sum_{Y_i \leq t} \frac{\delta_{ij}}{r_j(Y_i)} \tag{2}$$

where we have defined $\hat{q}_{kj}=0$ if no event occurred in Δ_k and $\hat{q}_{kj}=1/r_j(Y_i)$ if an event which corresponds to Y_i with $X_i^*(Y_i)=X_j$ occurred in Δ_k, and $r_j(s)$ is the number of people with $X_i^*(s)=X_j$ and are at risk at time s. Clearly, $r_j(s) = \Sigma_{m=1}^{n} I\{m\colon Y_m \geq s \text{ and } X_m^*(s) = X_j\}$, which is identical to r_{kj} except that the time t_k is replaced by time s.[3]

Comparing (1) and (2), one immediately notices that the only difference between the two estimators lies in the denominator, the number of people at risk. When there are time varying discrete covariates, the number of people at risk at time s with covariate value X_j, $r_j(s)$, comes from J sources: people who begin with covariate value X_j at time 0 and have not switched to any other covariate value before time s and survive to time s, and people who have covariate value X_k ($k \neq j$) before time s and have switched to the jth covariate value before time s and survive to time s with covariate value X_j. In the case without time varying covariates, the number of people at risk is decreasing over time because of the occurrence of the events and censoring over time. It follows that the denominator in (1) is decreasing over time. However, with time varying discrete covariates, the number of people at risk for each covariate value is not necessarily decreasing over time because some people may switch from one covariate value to another so that the number at risk for a certain covariate value at a certain point in time may be larger than some time before. Nevertheless, the estimator in (2) is still consistent and asymptotically normal.

Section 2 : Asymptotic Results

In this section we first present a simple proof of the classical result (Breslow and Crowley 1974, Meier 1975) that without time varying covariates, the estimator in (1) or equivalently the Kaplan-Meier estimator, is consistent and asymptotically normal. We then show with some modifications that our estimator with time varying discrete covariates is also consistent and asymptotically normal.

Without time varying covariates, let n be the initial sample size at time 0, then

$$\hat{\Lambda}(t)-\Lambda(t) = \sum_{k=1}^{K_n} (\hat{q}_k-q_k) = \sum_{k=1}^{K_n} W_{nk}$$

where $W_{nk}=\hat{q}_k-q_k$ and $K_n \equiv K(t)$ (the subscript n on K emphasizes the fact that the number of intervals depends on the sample size). Let $\mathcal{F}_{n,k}$ ($k \geq 0$) be the σ-field generated by the events occuring in the intervals Δ_1, Δ_2, ...,

Δ_k, with $F_{n,0}=\{n\}$. For each person who survives to the beginning of Δ_k, there is a probability q_k that the event will occur in Δ_k and a probability $1-q_k$ that the event will not occur in Δ_k. Therefore, given the number at risk at the beginning of Δ_k (i.e. r_k), the number of events d_k occurred in Δ_k follows a binomial distribution $\mathrm{Bin}(r_k,q_k)$. It follows that

$$E(W_{nk}|F_{n,k-1}) = E(\hat{q}_k - q_k|F_{n,k-1}) = E(\frac{d_k}{r_k} - q_k|F_{n,k-1}) = 0 \qquad (3)$$

because given $F_{n,k-1}$, the number at risk r_k is known so that the conditional expectation equals $(1/r_k)E(d_k - r_k q_k|F_{n,k-1})=0$. Similarly,

$$\mathrm{Var}(W_{nk}|F_{n,k-1}) = q_k(1-q_k)/r_k \qquad (4)$$

Since $F_{n,0} \subseteq F_{n,1} \subseteq F_{n,2} \subseteq \ldots \subseteq F_{n,k}$, the sequence of random variables $\sum_{k=1}^{m} W_{nk}$ $(m=1,2,\ldots)$ forms a martingale relative to the increasing σ-fields $F_{n,k}$ $(k=0,1,2,\ldots)$. We want to show that $n^{\frac{1}{2}} \sum_{k=1}^{K_n} W_{nk} \xrightarrow{D} N(0,\sigma^2)$ for some $\sigma^2 > 0$. This would follow from the martingale central limit theorem (see for example Hall and Heyde 1980), provided the following two conditions are satisfied:

(a) $$\sum_{k=1}^{K_n} \mathrm{Var}(n^{\frac{1}{2}} W_{nk}|F_{n,k-1}) \xrightarrow{P} \sigma^2 \qquad (\sigma^2 > 0)$$

(b) $$\sum_{k=1}^{K_n} E(\,|n^{\frac{1}{2}} W_{nk}|^z\,|F_{n,k-1}) \xrightarrow{P} 0 \qquad \text{for some } z > 2$$

The P's in (a) and (b) denote convergence in probability, and (b) is the conditional Lyapounov's condition. To verify (a), (4) implies that

$$\sum_{k=1}^{K_n} \mathrm{Var}(n^{\frac{1}{2}} W_{nk}|F_{n,k-1}) \approx \sum_{k=1}^{K_n} \frac{q_k}{r_k} n \approx \sum_{k=1}^{K_n} \frac{\lambda(t_k)\Delta_k}{r_k/n}$$

where we have assumed that the q_k's are small so that $1-q_k \approx 1$, and

$q_k \simeq \lambda(t_k)\Delta_k$. By the Glivenko-Cantelli theorem, r_k/n converges uniformly with probability 1 to $E(r_k/n)$, which is $(1-F(t_k))(1-G(t_k))$, the probability of surviving and being uncensored at t_k. Let $S_T(t_k) \equiv (1-F(t_k))(1-G(t_k))$. Assume that for any finite t_k, $S_T(t_k)$ is bounded below by some positive number, then $1/(r_k/n)$ converges uniformly with probability 1 to $1/S_T(t_k)$. It follows that for any positive ϵ, there exists a positive number M such that for n > M,

$$\left| \sum_{k=1}^{K_n} \frac{\lambda(t_k)\Delta_k}{r_k/n} - \sum_{k=1}^{K_n} \frac{\lambda(t_k)\Delta_k}{S_T(t_k)} \right| < \epsilon \tag{5}$$

with probability 1. Assume that $\lambda(t)$ and $S_T(t)$ are continuous functions, then the Riemann integral $\int_0^t [\lambda(x)/S_T(x)]dx$ exists. Thus, given $\epsilon > 0$, there exists a positive number M′ such that for n > M′,

$$\left| \sum_{k=1}^{K_n} \frac{\lambda(t_k)}{S_T(t_k)} \Delta_k - \int_0^t \frac{\lambda(x)}{S_T(x)}dx \right| < \epsilon \tag{6}$$

since $\Delta_k \to 0$ as $n \to \infty$. Therefore, combining (5) and (6),

$$\left| \sum_{k=1}^{K_n} \frac{\lambda(t_k)\Delta_k}{r_k/n} - \int_0^t \frac{\lambda(x)}{S_T(x)}dx \right| < 2\epsilon$$

with probability 1, for n > max{M,M′}. It follows that

$$\sum_{k=1}^{K_n} \mathrm{Var}(n^{1/2}W_{nk}|\mathcal{F}_{n,k-1}) \xrightarrow{P} \sigma^2(0,t) \tag{7}$$

where $\sigma^2(0,t)=\int_0^t [\lambda(x)/S_T(x)]dx$.

To verify (b), we show that $\sum_{k=1}^{K_n} E(|n^{1/2}W_{nk}|^4 |\mathcal{F}_{n,k-1}) \xrightarrow{P} 0$.

$$E(|n^{1/2}W_{nk}|^4 |\mathcal{F}_{n,k-1})=n^2E(|\frac{d_k}{r_k} - q_k|^4 |\mathcal{F}_{n,k-1})$$

$$=(n^2/r_k^4) E(|d_k-r_kq_k|^4 |\mathcal{F}_{n,k-1})$$

$$=(n^2/r_k^4)\{3r_k^2q_k^2(1-q_k)^2 + nq_k(1-q_k)[1-6q_k(1-q_k)]\} \tag{8}$$

$$\simeq \frac{3[\lambda(t_k)\Delta_k]^2}{[r_k/n]^2} + \frac{\lambda(t_k)\Delta_k}{n[r_k/n]^3} \tag{9}$$

where (8) comes from the fourth central moment of the binomial distribution $Bin(r_k, q_k)$ and (9) is obtained by assuming that $1-6q_k \simeq 1$ and $q_k \simeq \lambda(t_k)\Delta_k$. It follows that

$$\sum_{k=1}^{K_n} E(|n^{1/2}W_{nk}|^4 \mid \mathcal{F}_{n,k-1}) \le (\max_k \Delta_k) \sum_{k=1}^{K_n} \frac{3[\lambda(t_k)]^2\Delta_k}{[r_k/n]^2} + \left(\frac{1}{n}\right) \sum_{k=1}^{K_n} \frac{\lambda(t_k)\Delta_k}{[r_k/n]^3}$$

Using similar arguments as before, it can easily be shown that

$$\sum_{k=1}^{K_n} \frac{3[\lambda(t_k)]^2\Delta_k}{[r_k/n]^2} \xrightarrow{P} \int_0^t \frac{3[\lambda(x)]^2}{[S_T(x)]^2}dx$$

$$\sum_{k=1}^{K_n} \frac{\lambda(t_k)\Delta_k}{[r_k/n]^3} \xrightarrow{P} \int_0^t \frac{\lambda(x)}{[S_T(x)]^3}dx$$

Assume that both integrals are bounded, then

$$\sum_{k=1}^{K_n} E(|n^{1/2}W_{nk}|^4 \mid \mathcal{F}_{n,k-1}) \xrightarrow{P} 0.$$

since $\max_k \Delta_k \to 0$ and $1/n \to 0$ as $n \to \infty$. Hence, from the martingale central limit theorem, we conclude that the Nelson estimator is consistent and asymptotically normal, i.e.

$$n^{1/2}(\hat{\Lambda}(t)-\Lambda(t)) \xrightarrow{D} N(0,\sigma^2(0,t))$$

The martingale structure also allows us to calculate the asymptotic covariance $AVAR(\hat{\Lambda}(s),\hat{\Lambda}(t))$ $(s > t)$ easily. From the definition of $\hat{\Lambda}(t)$, we can write

$\hat{\Lambda}(s) = \sum_{k=1}^{K_n} W_{nk} + \sum_{k=K_n+1}^{J_n} W_{nk} = \hat{\Lambda}(t) + \hat{\varphi}(t,s)$, where $\hat{\varphi}(t,s) \equiv \sum_{k=K_n+1}^{J_n} W_{nk}$, K_n and J_n are the numbers of intervals in $[0,t)$ and $[0,s)$ respectively. Consider the random vector $Q_n \equiv (\hat{\Lambda}(t), \hat{\varphi}(t,s))'$, for any vector $e \equiv (e_1, e_2) \in R^2$,

$$e \cdot Q_n = e_1\hat{\Lambda}(t) + e_2\hat{\varphi}(t,s) = \sum_{k=1}^{K_n} e_1 W_{nk} + \sum_{k=K_n+1}^{J_n} e_2 W_{nk} = \sum_{k=1}^{J_n} W'_{nk}\ , \text{ where}$$

$W'_{nk}=e_1W_{nk}$ if $k \le K_n$ and $W'_{nk}=e_2W_{nk}$ if $k > K_n$. Clearly, the sequence of random variables $\{ \sum_{k=1}^{J_n} W'_{nk}\}$ still forms a martingale relative to the σ-fields $\{F_{n,k-1}\}$. Using the fact that $E[W_{nk}W_{nj}|F_{n,k-1}]=E[W_{nk}E(W_{nj}|F_{n,j-1})|F_{n,k-1}]=0$ for $j > k$, we can check that (assume $J_n/K_n \to t/s$ as $n \to \infty$), as $n \to \infty$,

$$\sum_{k=1}^{J_n} \mathrm{Var}(n^{1/2}W'_{nk}|F_{n,k-1}) \approx \sum_{k=1}^{K_n} \frac{e_1^2 q_k}{r_k/n} + \sum_{k=K_n+1}^{J_n} \frac{e_2^2 q_k}{r_k/n}$$

$$\xrightarrow{P} e_1^2 \int_0^t \frac{\lambda(x)}{S_T(x)}dx + e_2^2 \int_t^s \frac{\lambda(x)}{S_T(x)}dx$$

Similarly, we can easily check that

$$\sum_{k=1}^{J_n} E(|n^{1/2}W'_{nk}|^4 \mid F_{n,k-1}) \xrightarrow{P} 0$$

Therefore, $n^{1/2}e \cdot Q_n \xrightarrow{D} N(0,\theta^2)$, where $\theta^2 = e_1^2\sigma^2(0,t) + e_2^2\sigma^2(t,s)$,

$\sigma^2(0,t) = \int_0^t [\lambda(x)/S_T(x)]dx$ and $\sigma^2(t,s) = \int_t^s [\lambda(x)/S_T(x)]dx$. Hence, by Cramer-Wold (1936),

$$n^{1/2}\begin{pmatrix} \hat{\Lambda}(t) \\ \hat{\varphi}(t,s) \end{pmatrix} \xrightarrow{D} N\left(0, \begin{bmatrix} \sigma^2(0,t) & 0 \\ 0 & \sigma^2(t,s) \end{bmatrix}\right)$$

It follows easily that $(\hat{\Lambda}(t),\hat{\Lambda}(s))$ is also asymptotically jointly normal with $\mathrm{AVAR}(\hat{\Lambda}(t),\hat{\Lambda}(s)) = \mathrm{AVAR}(\hat{\Lambda}(t)) = \sigma^2(0,t)/n$. Hence, in general, for any t and s,

$$\text{AVAR}(\hat{\Lambda}(t),\hat{\Lambda}(s)) = \text{AVAR}(\hat{\Lambda}(j)), \quad j=\min\{t,s\} \tag{10}$$

We can also derive the Greenwood formula (Greenwood 1926) as follows. From (7), the asymptotic variance is given by

$$\text{AVAR}(\hat{\Lambda}(t)) = \frac{\sigma^2(0,t)}{n} = \int_0^t \frac{\lambda(x)}{nS_T(x)}dx \simeq \sum_{k=1}^{K_n} \frac{q_k}{nS_T(t_k)} \simeq \sum_{k=1}^{K_n} \frac{q_k}{r(t_k)}$$

Hence, an estimator for the asymptotic variance is

$$\hat{\text{AVAR}}(\hat{\Lambda}(t)) = \sum_{k=1}^{K_n} \hat{q}_k/r(t_k) = \sum_{Y_i \le t} \left(\frac{\delta_i}{n-i+1}\right)\left(\frac{1}{n-i+1}\right) \tag{11}$$

The last equality in (11) is obtained by substituting the terms for $\hat{q}_k$ and r_k respectively. By the δ-method, the corresponding estimator for the survival function S(t) is given by

$$\hat{\text{AVAR}}(\hat{S}(t))=(\hat{S}(t))^2 \sum_{Y_i \le t} \frac{\delta_i}{(n-i+1)^2} \simeq (\hat{S}(t))^2 \sum_{Y_i \le t} \frac{\delta_i}{(n-i+1)(n-i)} \tag{12}$$

The last expression in (12) is just the Greenwood formula for the asymptotic variance of the Kaplan-Meier estimator. [See, e.g., Miller (1981, p.65).]

Our proof not only provides a concise alternative to the traditional proofs, but it also offers an easy extension to the case with time varying discrete covariates. Consider the setup described in Section 1 in which there is one time varying covariate with J possible values. For each $j=1,2,\ldots,J$, the sequence of random variables $\sum_{k=1}^{m} W_{nkj} = \sum_{k=1}^{m} (\hat{q}_{kj}-q_{kj})$ $(m=1,2,\ldots)$ still forms a martingale relative to the increasing σ-fields $F_{n,k}$ $(k=0,1,2,\ldots)$. Therefore, if we can verify conditions (a) and (b), then the martingale central limit theorem implies that $\hat{\Lambda}_j(t)$ is a consistent estimator of $\Lambda_j(t)$ and is asymptotically normally distributed with some variance $\sigma_j{}^2 > 0$. For the asymptotic results to hold, we need to assume that

$N_j(0)=\omega_j n$, where $\omega_1+\omega_2+...+\omega_J=1$, and the weights ω_j's are fixed as $n \to \infty$. Apart from these minor modifications, the methods we have employed to verify conditions (a) and (b) for the case without time varying covariates are also applicable here. However, there is one important difference. As we have mentioned in Section 1, the expression for the number at risk with covariate value X_j (r_{kj}) is different from r_k. More specifically, $r_{kj} = \Sigma_{m=1}^{n} I\{m: Y_m \geq t_k \text{ and } X_m^*(t_k) = X_j\}$ and $r_k = \Sigma_{m=1}^{n} I\{m: Y_m \geq t_k\}$. Since we have repeatedly made use of the uniform convergence of r_k/n in the verifications of conditions of (a) and (b) above, we need to show that r_{kj}/n is also uniformly convergent. If this is accomplished, then the verifications of conditions (a) and (b) performed above will immediately carry over to the case with time varying covariates. The Glivenko-Cantelli theorem, which we have used to prove the uniform convergence of r_k/n, cannot be simply applied to prove the uniform convergence of r_{kj}/n because the sets $\{m: Y_m \geq t_k \text{ and } X_m^*(t_k) = X_j\}$ and $\{m: Y_m \geq t_k\}$ are different.

Since there are J covariate values and each person can switch among the covariate values, let $Z_i(t)$ be the number of switches the ith person made in the interval $[0,t)$, $i=1,2,...,n$. The total number of switches in $[0,t)$ becomes $Z(t) = \Sigma_{i=1}^{n} Z_i(t)$. In the appendix, we show that, if for some M, $0 < M < \infty$, $P\{Z(t) \leq nM\}$ converges in probability to 1 as $n \to \infty$, then r_k/n will converge uniformly with probability 1 to $E(r_{kj}/n)$. It follows that conditions (a) and (b) are satisfied, as long as r_{kj}/n and $E(r_{kj}/n)$ are bounded below by some positive numbers.

The condition that for some finite positive number M, $P\{Z(t) \leq nM\}$ converges in probability to 1 as $n \to \infty$, is a weak one. It is clearly weaker than the requirement that $Z(t)/n$ obeys the weak law of large numbers.

In general, the expression for $E(r_{kj}/n)$ will be very complicated since one has to enumerate an infinite number of possible switches. Although it may not be possible to write down an explicit expression for $E(r_{kj}/n)$, the computation and the asymptotic properties of the estimators are not affected since these results do not require $E(r_{kj}/n)$ to be known explicitly. Although the asymptotic variance of $\hat{\Lambda}_j(t)$ is a function of $E(r_{kj}/n)$ and will therefore also be a complicated expression, it can be consistently estimated. Specifically, the arguments that lead to equation (11) above are also valid here, therefore $AVAR(\hat{\Lambda}_j(t))$ can be estimated by equation (11), with $\hat{q}_k$, $r(t_k)$, δ_i, n-i+1 replaced by $\hat{q}_{kj}$, $r_j(t_k)$, δ_{ij}, $r_j(Y_i)$ respectively. On the other hand, if there is some prior information, derived from economic theory or from the data itself, that allows one to make some restrictions on the numbers and the directions of switches, then it may be possible to express $E(r_{kj}/n)$ explicitly. Below we will provide a simple example to illustrate how the expression $E(r_{kj}/n)$ can be calculated. The simulation experiment in Section 3 is also based on this example. The simulation experiment demonstrates that it is not necessary to know $E(r_{kj}/n)$ explicitly in order to perform the estimation. We need to know $E(r_{kj}/n)$ explicitly only when we have to calculate $E(r_{kj}/n)$ to obtain the true (theoretical) values in order to compare with the estimates. Hence, the explicit expression of $E(r_{kj}/n)$ is used only when we evaluate the performance of the estimators.

Consider a situation in which there is only one time varying discrete covariate with two possible values X_1 and X_2. Let there be $N_1(0)$ persons with covariate value X_1 at time 0 and $N_2(0)$ persons with covariate value X_2 at time 0. A person can switch between X_1 and X_2 before the event or the censoring occurred. For simplicity, assume each person has at most one switch. These assumptions imply that the condition $Z(t) \leq nM$ is trivially satisfied with M=1.

Given these assumptions, the expected number of people at risk at time t with covariate value X_1 is given by

$$E[r_1(t)] = N_1(0)(1-F_1(t))(1-G_1(t))(1-H_1(t)) + N_2(0)\int_0^t (1-F_2(x))(1-G_2(x))\frac{(1-F_1(t))(1-G_1(t))}{(1-F_1(x))(1-G_1(x))}dH_2(x) \qquad (13)$$

On the right-side of (13), $(1-F_1(t))(1-G_1(t))(1-H_1(t))$ is the probability that a person begins with covariate value X_1, survives to time t and has not switched to covariate value X_2 in $[0,t)$. The integral in the second term is the probability that a person begins with covariate value X_2 has switched to covariate value X_1 before time t and survives to time t with covariate value X_1. The first two terms inside the integral is the probability that neither the event nor the censoring occurs in $[0,x)$, for a person with covariate value X_2 in $[0,x)$. The fraction inside the integral in (13) comes from the Markov assumption, namely that, the conditional probability of a person surviving at time t given that he has switched to covariate value X_1 at time x is the same as the conditional probability of a person surviving at time t with covariate value X_1 given that he survives to time x with covariate value X_1. Integrating all the terms inside the integral with respect to the density of the distribution of switching time ($H_2(x)$) yields the probability that a person begins with covariate value X_2, has switched to covariate value X_1 in $[0,t)$, and survives to time t. Multiplying the probabilities by the respective initial numbers at risk ($N_1(0)$ and $N_2(0)$) and adding them up gives the expected number of people at risk at time t with covariate value X_1. When there are no time varying covariates, $N_1(0)=n$, $N_2(0)=0$, $H_1(t)=0$, $E[r_1(t)]$ becomes $n(1-F_1(t))(1-G_1(t))$. Therefore, $E[r_1(t)/n]=(1-F_1(t))(1-G_1(t))$, which is essentially the same as $E(r_k/n)$ discussed before.

The above example illustrates the principle to calculate $E(r_{kj}/n)$, and it is easy to see that the expression will become more complicated when less restrictions and assumptions are made on the numbers and the directions of switches.

Since there are J estimated cumulative hazard functions at each time t, one may be interested in deriving the asymptotic distribution of the vector $\hat{\Lambda}^*(t)=(\hat{\Lambda}_1(t),\hat{\Lambda}_2(t),\ldots,\hat{\Lambda}_J(t))'$. Let $\hat{\Lambda}_j(t) = \sum_{k=1}^{K_n} W_{nkj}$ where $W_{nkj}=\hat{q}_{kj}-q_{kj}$. For any vector $e^* \equiv (e_1,e_2,\ldots,e_J) \in R^J$, $e^*\cdot\hat{\Lambda}^*(t) = \sum_{k=1}^{K_n} W''_{nk}$, where $W''_{nk} = e_1W_{nkj}+e_2W_{nk2}+\ldots+e_JW_{nkJ}$. By the independence of the J binomial experiments,

$$\mathrm{Var}(n^{1/2}W''_{nk}|F_{n,k-1}) \simeq \sum_{j=1}^{J} e_j^2\frac{\lambda_j(t_k)\Delta_k}{r_{kj}/n}$$

therefore

$$n^{1/2}e^*\cdot\hat{\Lambda}^*(t) \xrightarrow{D} N(0,\phi^2(0,t))$$

where $\phi^2(0,t) = \Sigma_{j=1}^{J} e_j^2\sigma_j^2(0,t)$, $\sigma_j^2(0,t) = \int_0^t \{\lambda_j(x)/E[r_j(x)/n]\}dx$. Hence, by the Cramer-Wold device (Cramer and Wold 1936), $(\hat{\Lambda}_1(t),\hat{\Lambda}_2(t),\ldots,\hat{\Lambda}_J(t))$ is asymptotically jointly normally distributed, and $\hat{\Lambda}_m(t)$ and $\hat{\Lambda}_j(t)$ are asymptotically independent for any $m\neq j$. The last result looks counterintuitive because one may expect that the estimators should be dependent because people may switch from one covariate value to another so that the numbers at risk of the two groups at the same time point would be correlated. Our theory shows that the estimators are asymptotically independent because of the martingale structure and the independence of the J binomial distributions.

The extension of the estimator and the asymptotic results to the case with more than one time varying discrete covariate is straightforward because one can simply treat each of the Xj's ($j=1,2,\ldots,J$) as a vector (instead of a scalar) and modify the above results accordingly.

Section 3 : Numerical Results

In this section, we provide some numerical results which show that the estimator we propose behaves well. For simplicity, we exclude censoring. The Monte Carlo experiment we performed is as follows: First, draw a random variable t_{11} from the Weibull distribution $F_1(t)=1-\exp(-t^2)$ and regard this as the distribution of the time of occurrence of the event (death, birth or unemployment etc.) for group one, then draw a random variable t_{12} from another Weibull distribution $H_1(t)=1-\exp(-t^2)$ and regard this as the distribution of the switching time for group one. If $t_{11} \le t_{12}$, then record t_{11} as the time the event occurred. Similarly for the second group: first draw a random variable t_{21} from the exponential distribution $F_2(t)=1-\exp(-2t)$ and regard this as the distribution of the time of occurrence of the event for group two, then draw a random variable t_{22} from another exponential distribution $H_2(t)=1-\exp(-2t)$ and regard this as the distribution of the switching time for group two. If $t_{21} \le t_{22}$, then record t_{21} as the time the event occurred. If $t_{11} > t_{12}$, then a switching from group one to group two occurred at time t_{12} and we draw a random variable Y_2 from the conditional exponential distribution $1-\exp[-(t-t_{12})]$ and record Y_2 as the time the event occurred. Similarly, if $t_{21} > t_{22}$, then a switching from group two to group one occurred at time t_{22} and we draw a random variable Y_1 from the conditional Weibull distribution $1-\exp[-(t-t_{22})^2]$ and record Y_1 as the time the event occurred. In this experiment, the event occurred for everyone and there is no censoring. The procedure is repeated 1000 times so that the initial sample size for each group is 1000. From the data, we can use the estimator defined in (2) to estimate the cumulative hazard functions for group one (the Weibull distribution $F_1(t)$) and group two (the exponential distribution $F_2(t)$) respectively. For each group, we calculate the estimated cumulative hazard at each time point at which an event occurred. Figures 1 and 2 report the results in which we have plotted the difference between the estimated cumulative hazard and the true cumulative hazard against the observation. Each observation corresponds to a time point at which the event occurred. The true cumulative hazard function is t^2 for the Weibull distribution $F_1(t)$ and $2t$ for the exponential distribution $F_2(t)$. We see that the estimator performs very well since the difference is very close to zero for all the observations, except at the right ends of the curves.[4] The estimators do not perform well at the right ends because our asymptotic results are based on the assumption that q_{kj} is very small such that $1-q_{kj} \simeq 1$ and $1-6q_{kj} \simeq 1$, and this assumption has been used many times in the verification of conditions (a) and (b) in Section 2. At the right ends of these curves, the number of events occurred is large relative to the number at risk, so that q_{kj} can no longer be neglected.

Table 1 reports the mean values and standard deviations of the

estimates at three time points (0.4,0.8,1.2) when we replicate the experiment 1000 times. We see that the average values of the estimates are close to the true values and the standard deviations are relatively small. In addition, we also calculate the covariance matrix of the estimators at these three time points. The results are shown in Table 2. From (10), our theory suggests that for each group, the asymptotic covariance of the estimators at two time points s and t should equal the asymptotic variance of the estimator at min{s,t}. We see that in Table 2,
$Cov(\hat{\Lambda}_1(0.4),\hat{\Lambda}_1(0.8)) \approx Cov(\hat{\Lambda}_1(0.4),\hat{\Lambda}_1(1.2)) \approx Var(\hat{\Lambda}_1(0.4))$,
$Cov(\hat{\Lambda}_1(0.8),\hat{\Lambda}_1(1.2)) \approx Var(\hat{\Lambda}_1(0.8))$,
$Cov(\hat{\Lambda}_2(0.4),\hat{\Lambda}_2(0.8)) \approx Cov(\hat{\Lambda}_2(0.4),\hat{\Lambda}_2(0.8)) \approx Var(\hat{\Lambda}_2(0.4))$, and
$Cov(\hat{\Lambda}_2(0.8),\hat{\Lambda}_2(1.2)) \approx Var(\hat{\Lambda}_2(0.8))$. The numbers agree quite well with the theory. In addition, our theory also suggests that the estimators of the first and second groups at the same time point should be uncorrelated. We see that $Cov(\hat{\Lambda}_1(0.4),\hat{\Lambda}_2(0.4)) \approx Cov(\hat{\Lambda}_1(0.8),\hat{\Lambda}_2(0.8)) \approx Cov(\hat{\Lambda}_1(1.2),\hat{\Lambda}_2(1.2)) \approx 0$, and again the numbers agree quite well with the theory.

We also calculate the estimates for the asymptotic variance. When there are time varying discrete covariates, the formulas for the estimators of the asymptotic variances have been explained at the end of Section 2. The results are shown in Figures 3 and 4 in which we have plotted the difference between the estimated asymptotic variance and the theoretical asymptotic variance against the observation. The theoretical asymptotic variances are given by $\sigma_j^2(0,t)/n = (1/n) \int_0^t \{\lambda_j(x)/E[r_j(x)/n]\}dx$ for j=1,2. These expressions involve double integrals which have no closed form solutions (see the note under Table 3), hence numerical integration is used to obtain the theoretical values. Figures 3 and 4 show that the difference is close to zero except at the right ends of the curves. Again, the estimators do not perform well at the right ends of the curves because q_{kj} is not negligible at these regions. We also replicate the experiment 1000 times and calculate the estimates of the asymptotic variance at three time points (0.4,0.8,1.2) and the average values of the estimates are shown in Table 3. We see that the average values are very close to the theoretical values.

Footnotes

1. See, e.g., Flinn and Heckman (1983). The assumption is even more pervasive in theoretical models of economic duration analyses, see, e.g., the labor turnover model in Jovanovic (1979) and the examples in Heckman and Singer (1984).

2. The martingale central limit theorem, under the framework of point process and stochastic integral, has been applied to survival analysis by Aalen (1978), but he has not explicitly verified the needed regularity conditions in the survival setting. It is our hope that the present simple and self-contained approach may provide a more intuitive and accessible alternative.

3. Throughout this paper, $r_k \equiv r(t_k)$ and $r_{kj} \equiv r_j(t_k)$. For convenience, we will use the simpler symbols r_k and r_{kj} wherever necessary.

4. We have also performed the experiment using other distributions such as a loglogistic distribution, and a Weibull distribution with decreasing hazard rate, and the plots are very similar to Figures 1 and 2.

Table 1
Means and Standard Deviations
of the estimates of the cumulative hazard
(1000 replications)

Group 1: Weibull Distribution $F_1(t)=1-\exp(-t^2)$, Cumulative Hazard=t^2
Sample Size=1000

Time (t)	True Value (t^2)	Average of Estimates	Standard Deviation
0.4	0.16	0.1604	0.0119
0.8	0.64	0.6408	0.0281
1.2	1.44	1.4416	0.0612

Group 2: Exponential Distribution $F_2(t)=1-\exp(-2t)$, Cumulative Hazard=2t
Sample Size=1000

Time (t)	True Value (2t)	Average of Estimates	Standard Deviation
0.4	0.8	0.7997	0.0399
0.8	1.6	1.5996	0.0671
1.2	2.4	2.4023	0.0897

Table 2
Covariance Matrix of
the estimates of the cumulative hazard
(1000 replications)

	$\hat{\Lambda}_1(0.4)$	$\hat{\Lambda}_1(0.8)$	$\hat{\Lambda}_1(1.2)$	$\hat{\Lambda}_2(0.4)$	$\hat{\Lambda}_2(0.8)$	$\hat{\Lambda}_2(1.2)$
$\hat{\Lambda}_1(0.4)$	0.000142					
$\hat{\Lambda}_1(0.8)$	0.000149	0.000787				
$\hat{\Lambda}_1(1.2)$	0.000172	0.000903	0.003750			
$\hat{\Lambda}_2(0.4)$	-0.000009	0.000009	-0.000022	0.001592		
$\hat{\Lambda}_2(0.8)$	-0.000019	0.000062	0.000033	0.001553	0.004497	
$\hat{\Lambda}_2(1.2)$	-0.000032	0.000066	0.000086	0.001410	0.004299	0.008045

Table 3
Means and Standard Deviations
of the estimates of the asymptotic
variance of the cumulative hazard
(1000 replications)

Group 1: Weibull Distribution
Sample Size=1000

Time (t)	Theoretical Value	Average of Estimates	Standard Deviation
0.4	0.0001394	0.0001398	0.0000112
0.8	0.0007683	0.0007707	0.0000474
1.2	0.0033777	0.0033991	0.0003024

Group 2: Exponential Distribution
Sample Size=1000

Time (t)	Theoretical Value	Average of Estimates	Standard Deviation
0.4	0.0016755	0.0016761	0.0001272
0.8	0.0047211	0.0047278	0.0003604
1.2	0.0086205	0.0086552	0.0006521

Note: The formulas for the theoretical values of the asymptotic variances for Group 1 and Group 2 are respectively given by:

$$\frac{1}{n}\int_0^t \frac{2x}{\exp(-2x^2)+2\exp(-x^2)\int_0^x \exp(-4y+y^2)\,dy}\,dx$$

and

$$\frac{1}{n}\int_0^t \frac{2}{\exp(-4x)+2\exp(-2x)\int_0^x y\exp(2y-2y^2)\,dy}\,dx$$

where n=1000.

Appendix

In this appendix, we prove that if for some M, $0 < M < \infty$, $P\{Z(t) \le nM\}$ converges in probability to 1 as $n \to \infty$, then r_k/n will converge uniformly with probability 1 to $E(r_{kj}/n)$. Our proof is partly based on the proof of the Glivenko-Cantelli theorem in Pollard (1984 p.13-16). There are five steps in Pollard's proof. It is easy to see that the first two steps (the two symmetrizations) are also valid in our model. To save space, we will not reproduce these steps here. Using our notations, Pollard's equation (11) becomes

$$P\{\sup_t |(r_j(t)/n) - E(r_j(t)/n)| > \epsilon\}$$

$$\le 4P\{\sup_t |n^{-1}\Sigma_{m=1}^{n} \pi_m I\{m: Y_m \ge t \text{ and } X_m^*(t) = X_j\}| > \epsilon/4\} \quad \text{for } n \ge 8\epsilon^{-2},$$

where $\pi_1, \pi_2, \ldots, \pi_n$ are independent sign random variables with $P\{\pi_i=+1\}=P\{\pi_i=-1\}=1/2$. Let $A_{mj}(t)=\{m: Y_m \ge t \text{ and } X_m^*(t) = X_j\}$. Equation (12) of Pollard is not valid here because of the presence of the random variables $X_m^*(t)$ and the possibility of an infinite number of switches such that the number of intervals will not be a finite number. Therefore, we consider the conditional probability $P\{\sup_t |n^{-1}\Sigma_{m=1}^{n} \pi_m I\{A_{mj}(t)\}| > \epsilon/4 \mid Y,X,Z(t) \le nM\}$, where $Y=(Y_1, Y_2, \ldots, Y_n)$ and $X=\{X_m^*(s)\}$, $s\in[0,t)$. That is, given the realizations of the observed times Y, the realizations of the covariate values X, and the condition that the number of switches is bounded above by nM for some finite positive number M, then $\sup_t |n^{-1}\Sigma_{m=1}^{n} \pi_m I\{A_{mj}(t)\}|$ is reduced to finding the maximum of $|n^{-1}\Sigma_{m=1}^{n} \pi_m I\{A_{mj}(t)\}|$ over a given set of intervals $[0,t_\gamma)$, $\gamma=1,2,\ldots,\Gamma$, where $\Gamma \le n+nM=n(1+M)$. The number of intervals Γ is less than or equal to n+nM because there are n observed times

$(Y_1, Y_2, \ldots, Y_n)$ and at most nM switching times (since there are at most nM switches). This leads to the inequality

$$P\{\sup_t |n^{-1}\Sigma_{m=1}^{n} \pi_m I\{A_{mj}(t)\}| > \epsilon/4 \mid Y,X,Z(t) \leq nM\}$$

$$\leq \Sigma_{\gamma=1}^{\Gamma} P\{|n^{-1}\Sigma_{m=1}^{n} \pi_m I\{A_{mj}(t_\gamma)\}| > \epsilon/4 \mid Y,X,Z(t) \leq nM\}$$

$$\leq n(1+M)\max_\gamma P\{|n^{-1}\Sigma_{m=1}^{n} \pi_m I\{A_{mj}(t_\gamma)\}| > \epsilon/4 \mid Y,X,Z(t) \leq nM\}$$

$$\leq 2n(1+M)\exp(-n\epsilon^2/32)$$

The last inequality is an application of Hoeffding's inequality as described in Pollard (p.16), since the indicator function $I\{A_{mj}(t_\gamma)\} \leq 1$. Taking expectations over Y and X, we finally get

$$P\{\sup_t |(r_j(t)/n) - E(r_j(t)/n)| > \epsilon \mid Z(t) \leq nM\}$$

$$\leq 8n(1+M)\exp(-n\epsilon^2/32)$$

For any events B, Q, and Q^c (the complement of Q),

$P(B) = P(B|Q)P(Q) + P(B|Q^c)P(Q^c) \leq P(B|Q) + P(Q^c)$. Hence,

$$P\{\sup_t |(r_j(t)/n) - E(r_j(t)/n)| > \epsilon\}$$

$$\leq P\{\sup_t |(r_j(t)/n) - E(r_j(t)/n)| > \epsilon \mid Z(t) \leq nM\} + P\{Z(t) > nM\}$$

$$\leq 8n(1+M)\exp(-n\epsilon^2/32) + P\{Z(t) > nM\}$$

$$\longrightarrow 0, \text{ as } n \to \infty.$$

Since $\sup_t |(r_j(t)/n) - E(r_j(t)/n)|$ converges in probability to zero, it follows from the arguments in Pollard (p.21-22) that $r_j(t)/n$ converges to $E(r_j(t)/n)$ uniformly with probability 1.

References

Aalen ,O., 1978, Nonparametric Inference for a Family of Counting Processes, Annals of Statistics 6, 701-726.

Breslow, N. and J. Crowley, 1974, A Large Sample Study of the Life Table and Product Limit Estimates Under Random Censorship, Annals of Statistics 2, 437-453.

Cramer, H. and H. Wold, 1936, Some Theorems on Distribution Functions, Journal of London Mathematical Society 11, 290-295.

Flinn, C. and J. Heckman, 1983, Are Unemployment and Out of the Labor Force Behaviorally Distinct Labor Force States ? Journal of Labor Economics 1, 28-42.

Greenwood, M., 1926, The Natural Duration of Cancer, Reports on Public Health and Medical Subjects 33 (H. M. Stationery Office, London).

Hall, P. and C. Heyde, 1980, Martingale Limit Theory and Its Applications (Academic Press, New York).

Heckman, J. and B. Singer, 1984, Econometric Duration Analysis, Journal of Econometrics 24, 63-122.

Jovanovic , B., 1979, Job Matching and the Theory of Turnover, Journal of Political Economy 87, 972-990.

Kaplan, E. and P. Meier, 1958, Nonparametric Estimation from Incomplete Observations, Journal of the American Statistical Association 53, 457-481.

Kiefer, N., 1988, Economic Duration Data and Hazard Functions, Journal of Economic Literature 26, 646-679.

Meier, P., 1975, Estimation of a Distribution Function from Incomplete Observations, in: J. Gani, ed., Perspectives in Probability and Statistics. Papers in Honour of M.S. Bartlett (Academic Press, New York) 67-82.

Miller, R., 1981, Survival Analysis (Wiley, New York).

Nelson, W., 1972, Theory and Applications of Hazard Plotting for Censored Failure Data, Technometrics 14, 945-966.

Pollard, D., 1984, Convergence of Stochastic Processes (Springer-Verlag, New York).

Figure 1 : Group One (Weibull Distribution)

Difference=Estimated Cumulative Hazard
− True Cumulative Hazard

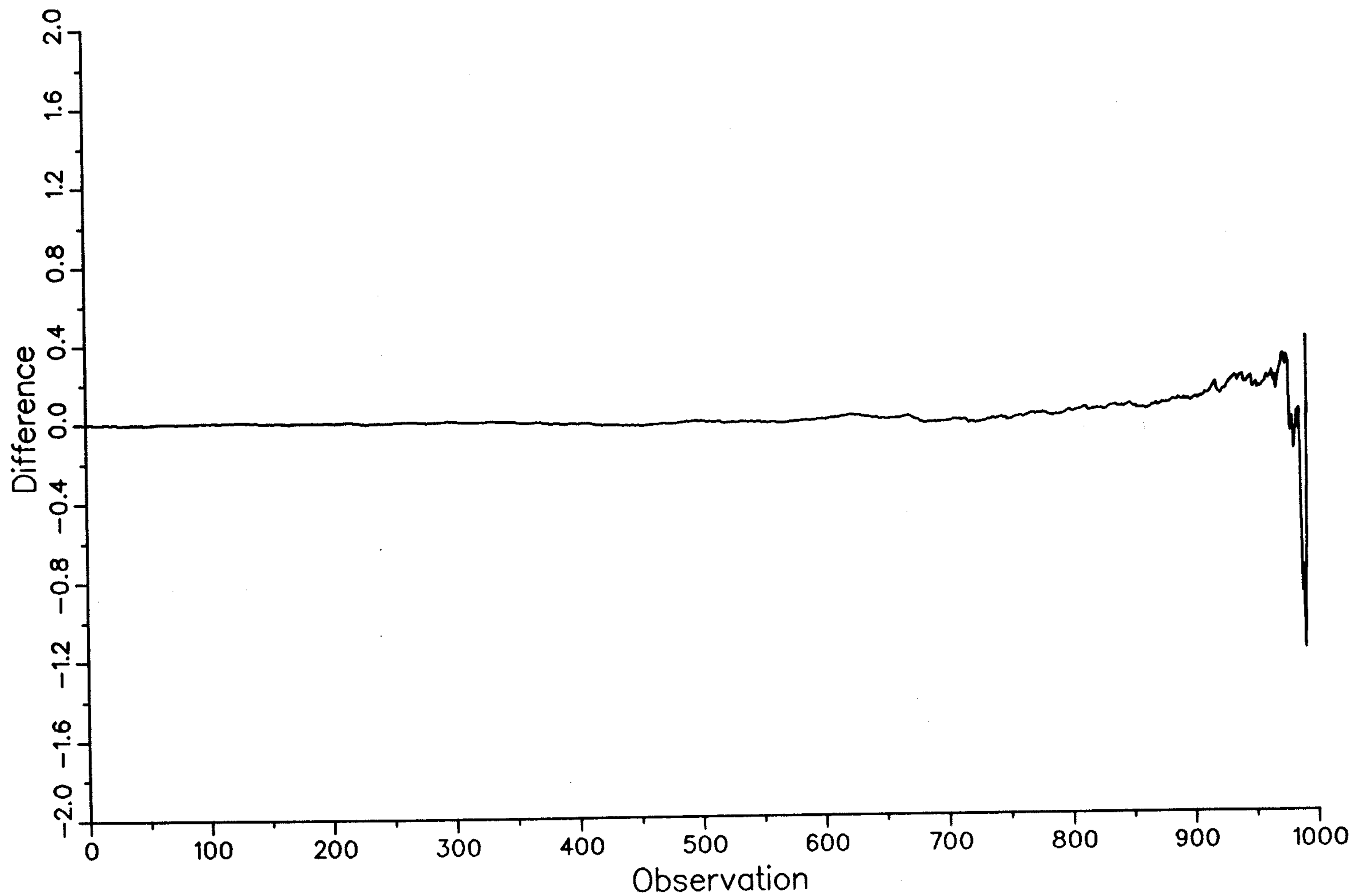

Figure 2 : Group Two (Exponential Distribution)

Difference=Estimated Cumulative Hazard
− True Cumulative Hazard

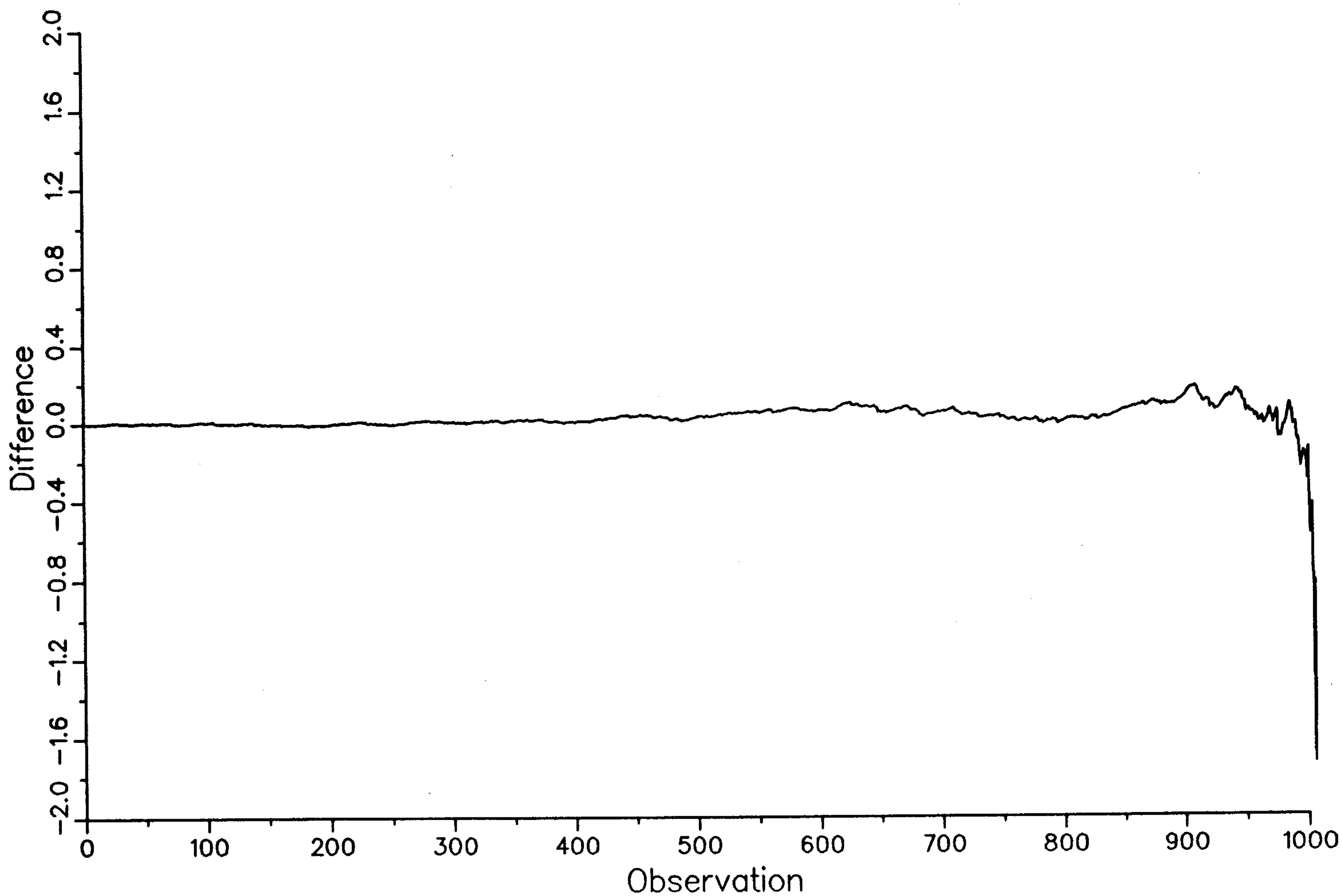

Figure 3 : Group One (Weibull Distribution)

Difference=Estimated Variance−Theoretical Variance

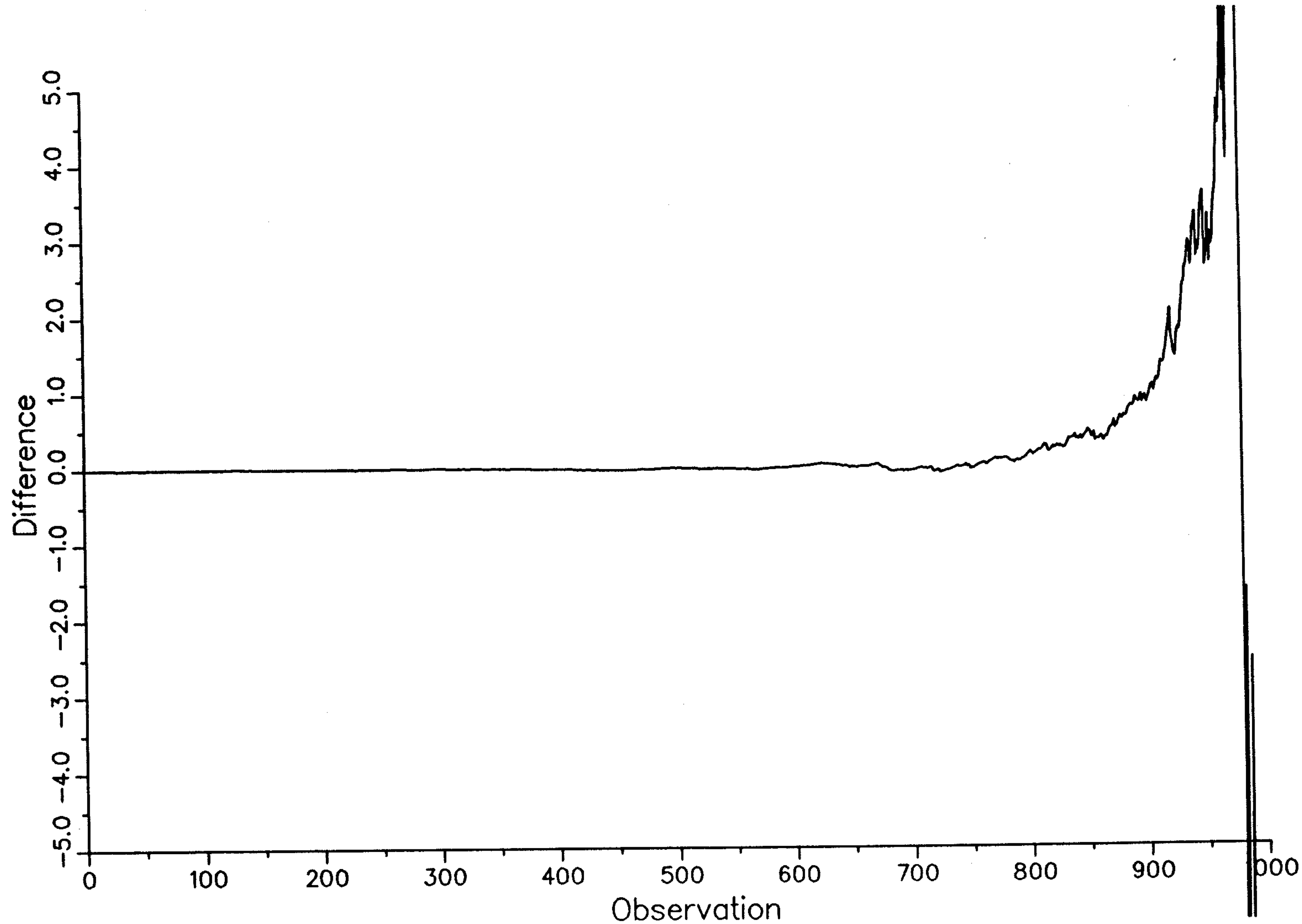

Figure 4 : Group Two (Exponential Distribution)

Difference=Estimated Variance–Theoretical Variance