The Annals of Statistics
1990, Vol. 18, No. 1, 354–372

ESTIMATING REGRESSION PARAMETERS USING LINEAR RANK TESTS FOR CENSORED DATA[1]

BY ANASTASIOS A. TSIATIS

Harvard School of Public Health and Dana-Farber Cancer Institute

A class of estimates for regression parameters in a linear model with right censored data is proposed. These estimates are derived by using linear rank tests for right censored data as estimating equations. They are shown to be consistent and asymptotically normal with covariance matrix for which estimates are proposed.

Efficient estimates within this class are derived together with conditions when they are fully efficient.

0. Introduction. The problem of estimating regression coefficients using ranks has received much attention in the literature for uncensored data and the results are summarized very nicely in Hettmansperger (1984). There has also been some work with censored data for the two-sample problem. Specifically, Louis (1981) considered estimation of a scale change between two distributions with the use of a log rank test, whereas Wei and Gail (1983) generalize to other linear rank tests as well.

We shall consider the general problem of multiple linear regression and show how linear rank tests can be used as estimating equations. The resulting linear rank estimates will be shown to be asymptotically normal with variances that can be estimated consistently. We shall also look at the efficiency of these estimates under various situations.

1. Model and notation. As in most right censored data problems we shall assume that there are two underlying random variables T and C, corresponding to time to failure and time to censoring, of which the minimum is observed. Ultimately, we wish to make inference on the relationship between the time to failure T and other concomitant variables, say, $\mathbf{Z} = (Z_1, Z_2, \dots, Z_K)$.

We shall consider the linear model

$$h(T_i) = \boldsymbol{\beta}'\mathbf{Z}_i + e_i, \qquad i = 1, \dots, N,$$

where, for the ith individual with covariates $\mathbf{Z}_i = (Z_{i1}, \dots, Z_{iK})$, a known monotone transformation h of the survival time is linearly related to the covariates plus error e_i, where the e_i, $i = 1, \dots, N$, are assumed to be iid with common distribution function F.

Since survival times are positive, it is convenient to consider transformations to the entire real line. The log transformation is often used and the resulting

Received December 1986; revised March 1989.

[1]This work was supported by grants CA-23415 and AI-24643, awarded by the National Institutes of Health, DHHS.

AMS 1980 *subject classifications*. Primary 62G05; secondary 62J05.

Key words and phrases. Censored data, linear rank tests, counting processes, martingales, asymptotic normality, efficiency.

model is often referred to as an accelerated failure time model. From here on we shall work on the transformed time scale for which the linear model applies. The primary aim of this paper is to find semiparametric estimates of β when the transformation h is known but the error distribution is unspecified. This problem has also been studied by Ritov (1989), for censored data, where he derives information bounds in the semiparametric sense of Begun, Hall, Huang and Wellner (1983).

Another class of semiparametric models which have also received much attention recently are those where the error distribution is specified but the transformation function h is not. For example, if the error distribution follows the extreme value distributions, then such a semiparametric model corresponds to the popular proportional hazards model proposed by Cox (1972). This general class of semiparametric models has also been studied by Bickel (1986), Doksum (1987) and Cuzick (1988).

The data that are observed will consist of N iid random vectors

$$(X_i, \Delta_i, \mathbf{Z}_i), \qquad i = 1, \dots, N,$$

where $X_i = \min(T_i, C_i)$ and

$$\Delta_i = \text{failure indicator} = \begin{cases} 1, & \text{if } T_i \le C_i, \\ 0, & \text{if } T_i > C_i. \end{cases}$$

The underlying model we wish to consider is

$$T_i = \beta' \mathbf{Z}_i + e_i, \tag{1.1}$$

where e_i are iid with distribution function $F(x)$ and corresponding hazard function $\lambda(x) = -d \log S(x)/dx$, where $S(x)$ denotes the survival function $1 - F(x)$.

The covariates $\mathbf{Z}_i$, $i = 1, \dots, N$, are assumed to be fixed and bounded. The censoring times C_i, $i = 1, \dots, N$, are independent random variables whose distribution may depend on the covariates. We shall denote by $H_i(x)$ the probability that C_i exceeds x. In order to avoid any nonidentifiability problems, we shall also assume that (T_i, C_i) are statistically independent.

2. Using linear rank tests as estimating equations. Linear rank tests for testing no association to covariates for right censored data have been derived by Prentice (1978) using a score test for the marginal likelihood of generalized ranks. Also, Gill (1980) derived a class of rank tests that occur naturally from a counting process point of view. These tests have been shown to be equivalent or, at least, asymptotically equivalent; see, for example, Mehrotra, Michalek and Mihalko (1982), Cuzick (1985) and Andersen, Borgan, Gill and Keiding (1982).

For our purposes, it will be convenient to consider the counting process approach. For simplicity, we shall consider the case of a single covariate Z and, in Chapter 5, indicate how to generalize to multiple covariates. Hence, if we want to test the null hypothesis, $\beta = 0$, in our regression model (1.1), then a class of linear rank tests can be written as

$$S_N(W_N) = \sum_{i=1}^{N} \int W_N(u)\, dN_i(u)\{Z_i - \bar{Z}(u)\},$$

where if we denote by $I(A)$ the indicator function for the event A, then $N_i(u)$ is the counting process for the ith individual, defined by

$$N_i(u) = I(X_i \leq u,\ \Delta_i = 1)$$

and

$$\bar{Z}(u) = \sum_{j=1}^{N} Z_j Y_j(u) \Big/ \sum_{j=1}^{N} Y_j(u),$$

where $Y_j(u) = I(X_j \geq u)$. The stochastic function $W_N(u)$ is referred to as a weight function and its role will be discussed later.

This statistic has a nice interpretation as a weighted sum over the death times of the observed covariate at the death time minus the average of the covariates still at risk at that point in time. We can easily generalize to test the hypothesis $\beta = \beta_0$ by replacing the times X_i with the residuals $X_i - \beta_0 Z_i$. The resulting linear rank test would be

$$S_N(W_N, \beta_0) = \sum_{i=1}^{N} \int W_N(u, \beta_0)\, dN_i(u + \beta_0 Z_i)\{Z_i - \bar{Z}(u, \beta_0)\},$$

where $\bar{Z}(u, \beta_0) = \sum Z_j Y_j(u + \beta_0 Z_j) / \sum Y_j(u + \beta_0 Z_j)$.

Using what have now become standard results for such stochastic integrals of counting processes, we can write $S_N(W_N, \beta_0)$ as

$$\sum_{i=1}^{N} \int W_N(u, \beta_0)\, dM_i(u + \beta_0 Z_i)\{Z_i - \bar{Z}(u, \beta_0)\},$$

where $M_i(u + \beta_0 Z_i) = N_i(u + \beta_0 Z_i) - \int_{-\infty}^{u} \lambda(x) Y_i(x + \beta_0 Z_i)\, dx$ is a martingale process with respect to the filtration $F_N(u, \beta_0)$ generated by all the survival and censoring information of the residuals up to error term time u. That is,

$$F_N(u, \beta_0) = \sigma[I\{(X_i - \beta_0 Z_i) \leq u\}, \Delta_i I\{(X_i - \beta_0 Z_i) \leq u\}, Z_i;\ i = 1, \dots, N].$$

Also, if the weight function $W_N(u, \beta_0)$ is predictable with respect to $F_N(u, \beta_0)$, then the process

$$S_N(W_N, \beta_0, u) = \sum_{i=1}^{N} \int_{-\infty}^{u} W_N(x, \beta_0)\, dM_i(x + \beta_0 Z_i)\{Z_i - \bar{Z}(x, \beta_0)\}$$

is also an $F_N(u, \beta_0)$ martingale. Using results from Andersen and Gill (1982) and Andersen, Borgan, Gill and Keiding (1982), it can also be shown that under appropriate conditions

$$N^{-1/2} S_N(W_N, \beta_0) \rightarrow_D N(0, \sigma^2),$$

where $\rightarrow_D$ indicates convergence in distribution. The asymptotic variance σ^2

can be estimated consistently by

$$\sum_{i=1}^{N} \int_{-\infty}^{\infty} W_N^2(u, \beta_0)\hat{V}(u, \beta_0)\, dN_i(u + \beta_0 Z_i)/N,$$

where $\hat{V}(u, \beta_0)$ is the empirical variance of the Z's that are at risk at error term time u, i.e.,

$$\hat{V}(u, \beta_0) = \sum_{j=1}^{N} \left\{Z_j - \bar{Z}(u, \beta_0)\right\}^2 Y_j(u + \beta_0 Z_j) \Big/ \sum_{j=1}^{N} Y_j(u + \beta_0 Z_j).$$

When $\beta = \beta_0$ the statistic $S_N(W_N, \beta)$ is centered at zero. This suggests that the linear rank test can be used as an estimating equation. That is, we may wish to consider an estimate $\hat{\beta}$ obtained by solving the equation $S_N(W_N, \hat{\beta}) = 0$. However, the tests we are considering are functions of the ranks of the residuals; hence, the statistic $S_N(W_N, \beta)$ is a step function of β. Therefore, more precisely, we define $\hat{\beta}$ as any value β for which $S_N(\beta)$ changes sign. That is, $sgn(S_N(\hat{\beta}+)) = -sgn(S_N(\hat{\beta}-))$.

We ultimately want to establish the asymptotic normality of this linear rank estimate. Since $S_N(W_N, \beta)$ is a step function, the usual way of expanding $S_N(W, \beta)$ about $S_N(W_N, \beta_0)$ in a Taylor series expansion will not work. We will instead show that the statistic $S_N(W_N, \beta)$ is asymptotically linear in a neighborhood of the true value β_0 as did Jurečková (1969, 1971) for the uncensored linear rank tests. This will then enable us to establish the large sample properties of $\hat{\beta}$, as will be shown in Section 3.

3. Asymptotic linearity of $S_N(\beta)$. In order to allow for ease of presentation, we shall consider $W_N(u) = 1$ (log rank test). However, the results will be generalized to arbitrary weight functions in Section 4. Also, without loss of generality let $\beta_0 = 0$. As we noted previously, the problem can always be transformed to this by considering residuals $X_i - \beta_0 Z_i$.

The stochastic integral that will be computed from here on will be truncated at the value T^*, which satisfies the condition that for some $\xi > 0$,

$$(3.1) \qquad P(X_i \geq T^* + \xi) \geq \psi > 0 \quad \text{for all } i.$$

The truncation is similar to that used by Andersen and Gill (1982). Other conditions that will be assumed in order to prove the results in this paper are as follows.

Conditions.

(A) The density of the error term in model (1.1), $f(x) = dF(x)/dx$, exists and is bounded by K_1 for all $x \leq T^* + \xi$.

(B) The density of the censoring random variables C_i is also bounded. That is, $h_i(x) = -dH_i(x)/dx \leq K_2$ for all i and $x \leq T^* + \xi$.

NOTE. (A) and (B) together imply that the density of $X_i = \min(T_i, C_i)$ is bounded. This follows because the density of X_i is equal to $f(x)H_i(x) + h_i(x)S(x)$, which is less than $K_1 + K_2$ for all i and for $x \leq T^* + \xi$.

(C) There exists a function $\theta(u)$ such that

$$|\lambda(u+\varepsilon) - \lambda(u) - \varepsilon\lambda'(u)| \leq \varepsilon^2\theta(u)$$

for $u \leq T^*$ and $|\varepsilon| \leq \xi$;

$$\int_{-\infty}^{T^*} |\theta(u)|\, du < \infty,$$

where $\lambda(u)$ is the hazard function of the error in model (1.1) and $\lambda'(u) = d\lambda(u)/du$.

(D) The covariates are bounded. Without loss of generality we shall assume that $|Z_i| \leq 1$ for all i.

Although the covariates Z_i, $i = 1, \dots, N$, are fixed constants, we shall assume that they satisfy the following stability conditions, which are again similar to those given by Andersen and Gill (1982).

(E) There exists a continuous function $\mu(u, \beta)$ for values of β in a neighborhood of $\beta = 0$, B such that

$$\sup_{\beta \in B,\, u \leq T^* + \xi} \{|\bar{Z}(u,\beta) - \mu(u,\beta)|\} \to_P 0,$$

where $\bar{Z}(u,\beta) = \sum_{j=1}^N Z_j Y_j(u + \beta Z_j) / \sum_{j=1}^N Y_j(u + \beta Z_j)$.

(F) There exists a continuous function $A(u, \beta)$ such that

$$\sup_{\beta \in B,\, u \leq T^* + \xi} \left[\left| N^{-1} \sum_{i=1}^N \{Z_i - \bar{Z}(u,\beta)\}^2 Y_i(u + \beta Z_i) - A(u,\beta) \right| \right] \to_P 0.$$

The asymptotically linear approximation of $S_N(\beta)$ near $\beta = 0$ is motivated by the following relationship:

$$S_N(\beta) = \sum_{i=1}^N \int_{-\infty}^{T^*} dN_i(u + \beta Z_i)\{Z_i - \bar{Z}(u,\beta)\}$$

$$= \sum_{i=1}^N \int_{-\infty}^{T^*} \{dN_i(u + \beta Z_i) - \lambda(u)\, du\, Y_i(u + \beta Z_i)\}\{Z_i - \bar{Z}(u,\beta)\}$$

$$= \sum_{i=1}^N \int_{-\infty}^{T^*} dM_i(u + \beta Z_i)\{Z_i - \bar{Z}(u,\beta)\} \tag{3.2}$$

$$+ \sum_{i=1}^N \int_{-\infty}^{T^*} \{\lambda(u + \beta Z_i) - \lambda(u)\}\, du\, Y_i(u + \beta Z_i)\{Z_i - \bar{Z}(u,\beta)\}. \tag{3.3}$$

The integral (3.2) can be written as

$$(3.4)\qquad \sum_{i=1}^{N}\int_{-\infty}^{T^*+\beta Z_i} dM_i(u)\{Z_i - \bar{Z}(u-\beta Z_i,\beta)\}$$

$$\approx \sum_{i=1}^{N}\int_{-\infty}^{T^*} dM_i(u)\{Z_i - \bar{Z}(u)\} = S_N(0),$$

whereas (3.3) can be approximated by

$$\sum_{i=1}^{N}\beta\int_{-\infty}^{T^*} Z_i Y_i(u+\beta Z_i)\{Z_i - \bar{Z}(u,\beta)\}\lambda'(u)\,du$$

$$= \beta\int_{-\infty}^{T^*}\Big\{\sum(Z_i - \bar{Z}(u,\beta))^2 Y_i(u+\beta Z_i)\Big\}\lambda'(u)\,du.$$

Hence, by (3.4) and condition (F) we would expect $S_N(\beta)$ to be asymptotically close to

$$\tilde{S}_N(\beta) = S_N(0) + N\beta g(0),$$

where $g(0) = \int_{-\infty}^{T^*} A(u,0)\lambda'(u)\,du$ in a neighborhood of $\beta = 0$. We shall make these ideas precise shortly, but basically if we can show that $S_N(\beta)$ and $\tilde{S}_N(\beta)$ are asymptotically equivalent, then the estimates $\hat{\beta}$, β^* (where $\hat{\beta}$ is the value of β where $S_N(\beta)$ changes sign, and where $\tilde{S}_N(\beta^*) = 0$) will also be asymptotically equivalent. We note that $N^{1/2}\beta^* = N^{-1/2}S_N(0)/g(0)$, which clearly converges in distribution to a normal with mean 0 and variance $\sigma^2(0)/g^2(0)$, where $\sigma^2(0) = \int_{-\infty}^{T^*} A(u,0)\lambda(u)\,du$ is the asymptotic variance of the log rank test. Hence, if we can show that $N^{1/2}(\hat{\beta} - \beta^*) \to_P 0$, then this would imply that $N^{1/2}(\hat{\beta})$ converges to the same distribution as $N^{1/2}(\beta^*)$. Arguing as in Jurečková (1969, 1971), it would suffice to show that

$$(3.5)\qquad \sup_{|\beta|\le CN^{-1/2}} N^{-1/2}|S_N(\beta) - \tilde{S}_N(\beta)| \to_P 0$$

for any $C > 0$.

The derivation of (3.5) will be in two steps. In Theorem 3.1 we shall show pointwise convergence. That is, for any fixed d,

$$N^{-1/2}\{S_N(N^{-1/2}d) - \tilde{S}_N(N^{-1/2}d)\} \to_P 0.$$

This can then be used to show uniform convergence at a fixed finite number of points that form a mesh from $-C$ to $+C$. That is, if we form a mesh $d_0,\dots,d_m$ from $-C$ to C, then we can show that

$$\max_{i\le m} N^{-1/2}|S_N(N^{-1/2}d_i) - \tilde{S}_N(N^{-1/2}d_i)| \to_P 0.$$

Finally, in order to complete the proof of uniform convergence, we must show that $N^{-1/2}S_N(\beta)$ as a function of β cannot fluctuate too greatly within any interval in the mesh. More precisely, if we take the mesh size equal to δ then we may show that for any $\varepsilon > 0$, there exists $\delta > 0$ such that

$$(3.6)\quad \lim_{N\to\infty} P\left\{ \sup_{dN^{-1/2} \le \beta^* \le (d+\delta)N^{-1/2}} N^{-1/2}|S_N(\beta^*) - S_N(dN^{-1/2})| \ge \varepsilon \right\} = 0$$

for any $|d| \le C$. This will be done in Theorem 3.2.

THEOREM 3.1. *For any fixed d*

$$N^{-1/2}\{S_N(N^{-1/2}d) - \tilde{S}_n(N^{-1/2}d)\} \to_P 0.$$

PROOF. Theorem 3.1 will be proved by a series of lemmas. As we have already noted, $S_N(N^{-1/2}d)$ can be written as a sum of (3.2) and (3.3). We shall establish that $N^{-1/2}\{(3.2) - S_N(0)\} \to_P 0$ by the use of Lemmas (3.1) and (3.2). Lemma (3.3) will then be used to show that $N^{-1/2}\{(3.3) - N^{1/2}dg(0)\} \to_P 0$. This will complete the proof of Theorem 3.1 since $\tilde{S}_N(N^{-1/2}d) = S_N(0) + N^{1/2}dg(0)$. □

LEMMA 3.1. *Let β_N denote a sequence of constants converging to* 0. *Then*

$$(3.7)\quad N^{-1/2}\left[\sum_{i=1}^{N} \int_{-\infty}^{T^*} dM_i(u + \beta_N Z_i)\{Z_i - \bar{Z}(u, \beta_N)\} - \int_{-\infty}^{T^*} dM_i(u + \beta_N Z_i)\{Z_i - \mu(u, \beta_N)\} \right]$$

converges in probability to 0, *where $\mu(u, \beta)$ is defined in condition* (E).

PROOF. We note that (3.7) is equal to $R(T^*)$, where

$$R(u) = N^{-1/2} \sum_{i=1}^{N} \int_{-\infty}^{u} dM_i(x + \beta_N Z_i)\{\bar{Z}(x, \beta_N) - \mu(x, \beta_n)\}$$

is an $F_N(u, \beta_N)$ martingale. Therefore, by using the version of Lenglart's inequality [Lenglart (1977)] given by Andersen and Gill [(1982); Appendix I, I.2] we get that

$$P\{|R(T^*)| > \varepsilon\} \le \delta/\varepsilon^2 + P\left[N^{-1} \sum_{i=1}^{N} \int_{-\infty}^{T^*} \{\bar{Z}(u, \beta_N) - \mu(u, \beta_N)\}^2 \times \lambda(u + \beta_N Z_i) Y_i(u + \beta_N Z_i)\, du > \delta \right].$$

By condition (E), we can find a value $N(\varepsilon, K)$ such that, for any $N > N(\varepsilon, K)$, $P[\sup_{u \le T^*}\{|\bar{Z}(u, \beta_N) - \mu(u, \beta_N)|\} > K] < \varepsilon$.

Hence, with probability exceeding $1 - \varepsilon$, the integral above cannot exceed $K^2\Lambda(T^* + \xi)$ whenever $N > N(\varepsilon, K)$, where $\Lambda(u)$ denotes the cumulative hazard function

$$\Lambda(u) = \int_{-\infty}^{u} \lambda(x)\,dx = -\log\{S(u)\}.$$

NOTE. This is true as long as N is sufficiently large that β_N is less than ξ. This, of course, creates no problem as β_N goes to 0 and we can choose the larger of $N(\varepsilon, K)$ or $N(\xi)$ such that $\beta_{N(\xi)} < \xi$. By assumption, $\Lambda(T^* + \xi)$ is finite, since $S(T^* + \xi)$ is bounded away from 0. Hence if we choose $K \le \{\delta/\Lambda(T^* + \xi)\}^{1/2}$, then

$$P\left[N^{-1}\sum_{i=1}^{N}\int_{-\infty}^{T^*}\{\bar{Z}(u,\beta_N) - \mu(u,\beta_N)\}^2\lambda(u+\beta_N Z_i)Y_i(u+\beta_N Z_i)\,du > \delta\right] < \varepsilon$$

$$\text{for } N > N(\varepsilon, K).$$

The proof is complete by choosing $\delta = \varepsilon^3$. □

LEMMA 3.2.

$$N^{-1/2}\left\{\sum\int_{-\infty}^{T^*} dM_i(x+\beta_N Z_i)\{Z_i - \bar{Z}(x,\beta_N)\} - S_N(0)\right\} \to_P 0. \tag{3.8}$$

PROOF. Expression (3.8) can be written as

$$N^{-1/2}\left[\sum\int_{-\infty}^{T^*} dM_i(x+\beta_N Z_i)\{Z_i - \bar{Z}(x,\beta_N)\} - \sum\int_{-\infty}^{T^*} dM_i(x+\beta_N Z_i)\{Z_i - \mu(x,\beta_N)\}\right]$$

$$+ N^{-1/2}\sum\left[\int_{-\infty}^{T^*} dM_i(x+\beta_N Z_i)\{Z_i - \mu(x,\beta_N)\} - \int_{-\infty}^{T^*} dM_i(x)\{Z_i - \mu(x,0)\}\right] \tag{3.9}$$

$$+ N^{-1/2}\left[\sum\int_{-\infty}^{T^*} dM_i(x)\{Z_i - \mu(x,0)\} - S_N(0)\right].$$

By Lemma 3.1, the first and third terms in the summand converge in probability to 0. Hence, we need only to show that the second term, (3.9), converges to 0 in probability.

Assuming that $\beta_N > 0$, (3.9) can be written as

$$\text{A}\qquad N^{-1/2}\left[\sum_{i=1}^{N}\int_{-\infty}^{T^*} dM_i(x)\{\mu(x-\beta_N Z_i,\beta_N) - \mu(x,0)\}\right]$$

$$\text{B}\qquad +N^{-1/2}\left[\sum_{\{i:\,Z_i\ge 0\}}\int_{T^*}^{T^*+\beta_N Z_i} dM_i(x)\{Z_i - \mu(x,0)\}\right]$$

$$\text{C}\qquad -N^{-1/2}\left[\sum_{\{i:\,Z_i< 0\}}\int_{T^*+\beta_N Z_i}^{T^*} dM_i(x)\{Z_i - \mu(x-\beta_N Z_i,\beta_N)\}\right].$$

Terms A, B, and C integrated up to time u are all $F_N(u,0)$ martingales. Hence, the variances of these terms are:

$$\operatorname{Var}(A) = N^{-1}\sum_{i=1}^{N}\int_{-\infty}^{T^*}\{\mu(x-\beta_N Z_i,\beta_N)-\mu(x,0)\}^2\lambda(x)P(X_i\ge x)\,dx,$$

$$\operatorname{Var}(B) = N^{-1}\sum_{\{i:\,Z_i\ge 0\}}\int_{T^*}^{T^*+\beta_N Z_i}\{Z_i-\mu(x,0)\}^2\lambda(x)P(X_i\ge x)\,dx,$$

$$\operatorname{Var}(C) = N^{-1}\sum_{\{i:\,Z_i< 0\}}\int_{T^*+\beta_N Z_i}^{T^*}\{Z_i-\mu(x-\beta_N Z_i,\beta_N)\}^2\lambda(x)P(X_i\ge x)\,dx.$$

Since the $|Z|$'s are bounded by 1, by assumption (D), this implies that

$$\operatorname{Var}(A) \le 4N^{-1}\sum_{i=1}^{N}\int_{-\infty}^{T^*}\lambda(x)S(x)H_i(x)\,dx \le 4\int_{-\infty}^{T^*}\lambda(x)S(x)\,dx \le 4.$$

Therefore, by the continuity of $\mu(u,\beta)$ and the dominated convergence theorem, Var(A) converges to 0 and hence A converges in probability to 0.

The variance of B is less than

$$4N^{-1}\sum_{\{i:\,Z_i\ge 0\}}\int_{T^*}^{T^*+\beta_N Z_i}\lambda(x)S(x)H_i(x)\,dx \le 4\int_{T^*}^{T^*+\beta_N}f(x)\,dx,$$

which by assumption (A) is less than $4\beta_N K_1$. Hence Var(B) converges to 0 and thus B converges in probability to 0. Similarly, we can show that C converges in probability to 0, which completes the proof. □

LEMMA 3.3. *The integral* (3.3) *satisfies the following property*:

$$N^{-1}\left[\sum_{i=1}^{N}\int_{-\infty}^{T^*}\{\lambda(u+\beta_N Z_i)-\lambda(u)\}Y_i(u+\beta_N Z_i)\{Z_i-\bar{Z}(u,\beta_N)\}\right] = \beta_N\{g(0)+o_p(1)\}, \tag{3.10}$$

where $g(0) = \int_{-\infty}^{T^*}A(u,0)\lambda'(u)\,du$, *and* $o_p(1)$ *is a term which converges in probability to* 0.

PROOF. Expression (3.10) can be written as the sum

$$\beta_N N^{-1}\sum_{i=1}^{N}\int_{-\infty}^{T^*}Z_iY_i(u+\beta_N Z_i)\{Z_i-\bar{Z}(u,\beta_N)\}\lambda'(u)\,du \tag{3.11}$$

$$+N^{-1}\sum_{i=1}^{N}\int_{-\infty}^{T^*}[\{\lambda(u+\beta_N Z_i)-\lambda(u)\}-\beta_N Z_i\lambda'(u)] \times Y_i(u+\beta_N Z_i)\{Z_i-\bar{Z}(u,\beta_N)\}\,du. \tag{3.12}$$

We note that (3.11) is equal to

$$\beta_N \int_{-\infty}^{T^*} \left[N^{-1} \sum_{i=1}^{N} \{Z_i - \bar{Z}(u, \beta_N)\}^2 Y_i(u + \beta_N Z_i) \right] \lambda'(u)\, du. \tag{3.13}$$

Using assumption (F), we can show that the integrand in (3.13) converges in probability to $g(0)$.

By assumptions (C) and (D), we see that for N sufficiently large and $u \le T^*$

$$|\{\lambda(u + \beta_N Z_i) - \lambda(u)\} - \beta_N Z_i \lambda'(u)| \le \beta_N^2 \theta(u);$$

hence (3.12) is bounded by

$$2\beta_N^2 \int_{-\infty}^{T^*} |\theta(u)| \left\{ N^{-1} \sum Y_i(u + \beta_N Z_i) \right\} du. \tag{3.14}$$

The integrand in (3.14) is positive with mean equal to

$$\int_{-\infty}^{T^*} |\theta(u)| \left\{ N^{-1} \sum_{i=1}^{N} S(u + \beta_N Z_i) H_i(u + B_N Z_i) \right\} du \le \int_{-\infty}^{T^*} |\theta(u)|\, du,$$

which is bounded by assumption (C). Therefore, the integrand in (3.14) is bounded in probability. Thus, (3.13) and (3.14) establish (3.10), and the proof of Lemma 3.3 is complete. The proof of Theorem 3.1 is completed by letting $\beta_N = N^{-1/2} d$. □

In order to complete the proof of uniform convergence of $N^{-1/2}\tilde{S}_N(\beta)$ to $N^{-1/2}S_N(\beta)$ as described in (3.5), we must establish relationship (3.6). This will be accomplished by putting a probabilistic bound on the maximum change that can occur for the statistics $S_N(\beta^*)$ as β^* varies from $dN^{-1/2}$ to $(d + \delta)N^{-1/2}$. We note that the statistic $S_N(\beta^*)$ is a function of the ranks of the residuals $X_i - \beta^* Z_i$. Hence, as β^* varies from $dN^{-1/2}$ to $(d + \delta)N^{-1/2}$, a change in the statistic $S_N(\beta^*)$ occurs whenever any of the ranks of $X_i - \beta^* Z_i$, $i = 1, \dots, N$, change.

Therefore, we can bound the maximum change of the statistic by computing the number of pairs of ranks that will be interchanged times the maximum change of the statistic for each such interchange.

We shall first show how to bound the maximum change of the statistic at an interchange. Note that whenever an interchange in ranks occurs as β^* increases from $dN^{-1/2}$ to $(d + \delta)N^{-1/2}$, it must occur between neighboring order statistics of the residuals $X_i - \beta^* Z_i$, $i = 1, \dots, N$. Let's order the residuals $(X_i - \beta^* Z_i)$, $i = 1, \dots, N$, and denote the corresponding covariate and failure indicator of the ith ordered residuals as $Z_{(i)}(\beta^*)$ and $\Delta_{(i)}(\beta^*)$. The statistic $S_N(\beta^*)$ can then be written as

$$\sum \Delta_{(i)}(\beta^*) \left\{ Z_{(i)}(\beta^*) - \bar{Z}_{(i)}(\beta^*) \right\}, \quad \text{where } \bar{Z}_{(i)}(\beta^*) = \sum_{f=i}^{N} Z_{(f)}(\beta^*)/(N - i + 1).$$

If the next change in ranks occurs between the j and $(j + 1)$st ordered residuals, then the new statistic $S_N(\beta^{*+})$ would be equal to

$$\sum_{i=1}^{j-1} \Delta_{(i)}(\beta^*)\{Z_{(i)}(\beta^*) - \bar{Z}_{(i)}(\beta^*)\} + \Delta_{(j+1)}(\beta^*)\{Z_{(j+1)}(\beta^*) - \bar{Z}_{(j)}(\beta^*)\}$$
$$+ \Delta_{(j)}(\beta^*)\big[Z_{(j)}(\beta^*) - \{\bar{Z}_{(j+2)}(\beta^*)(N-j-1) + Z_{(j)}(\beta^*)\}/\{N-j\}\big]$$
$$+ \sum_{i=j+2}^{N} \Delta_{(i)}(\beta^*)\{Z_{(i)}(\beta^*) - \bar{Z}_{(i)}(\beta^*)\}.$$

Hence, the difference in the statistic before and after the interchange in ranks is equal to

$$\text{(3.15)} \quad \{\Delta_{(j+1)}(\beta^*) - \Delta_{(j)}(\beta^*)\}\{\bar{Z}_{(j+2)}(\beta^*)(N-j-1)/(N-j) - \bar{Z}_{(j)}(\beta^*)\}$$
$$+ \Delta_{(j+1)}(\beta^*)Z_{(j+1)}(\beta^*)/(N-j) - \Delta_{(j)}(\beta^*)Z_{(j)}(\beta^*)/(N-j).$$

If both $\Delta_{(j)}$ and $\Delta_{(j+1)}$ are equal to 1, then (3.15) is equal to $(Z_{(j+1)} - Z_{(j)})/(N-j)$, whereas, if $\Delta_{(j)} = 1$ and $\Delta_{(j+1)} = 0$, then (3.15) is equal to $[Z_{(j+1)} - \{\bar{Z}_{(j+2)}(N-j-1) + Z_{(j)}\}/(N-j+1)]$. Finally, if $\Delta_{(j)} = 0$ and $\Delta_{(j+1)} = 1$, then (3.15) is equal to $(\bar{Z}_{(j+1)} - Z_{(j)})/(N-j+1)$. For any of these cases, since the Z's are bounded by 1 in absolute value, then the change in the statistic is bounded by $2/(N-j)$, where $(N-j)$ is the number of residuals at risk at the point where the interchange takes place.

REMARK. If there is no censoring, then the change in the statistic must be positive. Therefore, for the uncensored problem, the statistic $S_N(\beta^*)$ is monotone. This fact can be used to establish uniform convergence of $N^{-1/2}\tilde{S}_N(\beta)$ to $N^{-1/2}S_N(\beta)$ rather easily, as is done in Jurečková (1969, 1971). However, when censoring is introduced, or for other weight functions $W_N(u)$, the statistic $S_N(W_N, \beta)$ is not necessarily monotone in β. This necessitates a more complex proof for uniform convergence.

Recall that the value T^* was chosen so that $P(X_i \geq T^* + \xi) \geq \psi > 0$ for all i. Hence, with arbitrarily large probability we can find N large enough so that $N^{-1}\sum_{i=1}^{N} I(X_i \geq T^* + \beta_N Z_i) > \psi/2$. Since the statistic is computed for values of residuals which are less than $T^* + \beta_N Z_i$, this means that the number of residuals at risk will exceed $N\psi/2$ whenever an interchange in ranks takes place. Hence, with arbitrarily large probability the change in the statistic is bounded by $(4/\psi)N^{-1}$.

The next problem is to find a bound on the number of interchanges that will occur as β^* varies from $dN^{-1/2}$ to $(d+\delta)N^{-1/2}$. An interchange between two pairs of values (i, j) will occur for the value β_{ij} such that $X_i - \beta_{ij}Z_i = X_j - \beta_{ij}Z_j$, or $\beta_{ij} = (X_j - X_i)/(Z_j - Z_i)$. Therefore, an interchange (of pairs) occurring for values of β^* between $dN^{-1/2}$ and $(d+\delta)N^{-1/2}$ implies that $dN^{-1/2} \leq \beta_{ij} \leq (d+\delta)N^{-1/2}$ or

$$\text{(3.16)} \qquad dN^{-1/2}(Z_j - Z_i) \leq (X_j - X_i) \leq (d+\delta)N^{-1/2}(Z_j - Z_i).$$

The total number of such interchanges is therefore equal to

$$M = \sum_{i=1}^{N} \sum_{j \neq i} I(A_{ij}),$$

where A_{ij} denotes the event (3.16).

We are now in position to establish (3.6), which we state as a theorem.

THEOREM 3.2. *For any* $\varepsilon > 0$, *there exists* $\delta > 0$ *such that*

$$\lim_{N \to \infty} P\left\{ \sup_{dN^{-1/2} \le \beta^* \le (d+\delta)N^{-1/2}} N^{-1/2} |S_N(\beta^*) - S_N(dN^{-1/2})| \ge \varepsilon \right\} = 0$$

for any $|d| < C$.

PROOF. We have already noted that the maximum change of the statistic after each change is bounded by $(4/\psi)N^{-1}$ with arbitrarily large probability for sufficiently large N. Also, the number of such interchanges that can occur as β^* varies from $dN^{-1/2}$ to $(d+\delta)N^{-1/2}$ is equal to M. This implies it suffices to show that

$$\lim_{N \to \infty} P\{N^{-3/2} M \ge \varepsilon\} = 0 \tag{3.17}$$

for an appropriately chosen $\delta > 0$.

For $1 \le i < j \le N$, we write $W_{ij} = I(A_{ij}) + I(A_{ji})$,

$$V_i = \sum_{j \neq i} \left\{ E(W_{ij}|X_i) - E(W_{ij}) \right\},$$

$$V_{ij} = W_{ij} - E(W_{ij}|X_i) - E(W_{ij}|X_j) + E(W_{ij}).$$

Then

$$M - E(M) = \sum_{i=1}^{N} V_i + \sum\sum_{i<j} V_{ij}$$

is Hoeffding's expansion for $(M - E(M))$ as a sum of pairwise uncorrelated random variables [cf. van Zwet (1984)]. Denote by $f_i^*(u)$ the density of the random variable X_i. That is $f_i^*(u) = f(u)H_i(u) + h_i(u)S(u)$. We note that

$$P(A_{ij}|X_i) = \int_{X_i + dN^{-1/2}(Z_j - Z_i)}^{X_i + (d+\delta)N^{-1/2}(Z_j - Z_i)} f_j^*(u)\, du,$$

which, by conditions A, B, and D, can be shown to be less than $2(K_1 + K_2)\delta N^{-1/2}$. Similarly, we can show that $P(A_{ji}|X_i) \le 2(K_1 + K_2)\delta N^{-1/2}$, hence

$$|E(W_{ij}|X_i)| \le 4(K_1 + K_2)\delta N^{-1/2}.$$

Therefore,

$$E(M) \leq 2(K_1 + K_2)\delta N^{3/2},$$

$$\sigma^2(M) = \sum_{i=1}^{N} \sigma^2(V_i) + \sum_{i<j} \sum \sigma^2(V_{ij}) = O(N^2).$$

Application of Chebyshev's inequality proves (3.17) with $\delta = \varepsilon/\{3(K_1 + K_2)\}$ and establishes Theorem 3.2 □

Theorems 3.1 and 3.2 suffice to establish uniform convergence of $N^{-1/2}S_N(\beta)$ to the linear function $N^{-1/2}\tilde{S}_N(\beta)$ in an $O(N^{-1/2})$ neighborhood of the true value β_0. This implies that there exists a sequence of solutions $\hat{\beta}_N$, i.e., $\text{sgn}(S_N(\hat{\beta}+)) = -\text{sgn}(S_N(\hat{\beta}-))$ such that $N^{1/2}(\hat{\beta}_N - \beta_0)$ converges to a normal distribution with mean 0 and variance $\sigma^2(0)/g^2(0)$. Since $S_N(\beta)$ is not necessarily monotone in β, this means that there could be multiple solutions. However, since we have proven (3.5), or that $S_N(\beta)$ is asymptotically linear, then any sequence of solutions in an $N^{-1/2}$ neighborhood of β_0 would also have the property that

$$N^{1/2}(\hat{\beta}_N - \beta_0) \to_D N(0, \sigma^2(0)/g^2(0)).$$

The results above only apply in the neighborhood of the true value β_0. We have not yet been able to prove whether other solutions exist outside this neighborhood, although this never occurred in any of the numerical examples we considered.

In the next section, we will generalize the results to linear rank tests with arbitrary weight functions and discuss some of the efficiency properties of the resulting estimates. Since the method of proof is very similar to that used for the log rank test, the details will be omitted.

4. Arbitrary weight functions. Consider the linear rank test for $\beta_0 = 0$ as

$$S_N(W_N) = \sum_{i=1}^{N} \int_{-\infty}^{T^*} W_N(u)\, dN_i(u)\{Z_i - \bar{Z}(u)\},$$

where $W_N(u)$ is an $F_N(u)$ measurable, left continuous nonnegative function of the observations (X_i, Δ_i, Z_i), $i = 1, \ldots, N$. We shall also assume that $W_N(u)$ converges uniformly in probability to the deterministic function $W^*(u)$. Examples of such tests include: the generalized Wilcoxon test proposed by Gehan (1965), where $W_N(u) = \sum Y_j(u)/N$; the generalized Wilcoxon test of Prentice (1978) and Peto and Peto (1972), where $W_N(u) = KM(u^-)$; the left continuous version of the Kaplan–Meier estimate; and the class of tests proposed by Harrington and Fleming (1982), where $W_N(u) = \{KM(u^-)\}^{\rho}$.

If we denote by $W_N(u, \beta)$ the same function applied to the residuals $(X_i - \beta Z_i, \Delta_i, Z_i)$, $i = 1, \ldots, N$, then we shall consider estimate $\hat{\beta}(W_N)$ which is a solution to the estimating equation

$$S_N(W_N, \beta) = \sum_{i=1}^{N} \int_{-\infty}^{T^*} W_N(u, \beta)\, dN_i(u + \beta Z_i)\{Z_i - \bar{Z}(u, \beta)\} = 0.$$

Using methodology similar to that in Section 3, we can show that

$$N^{1/2}\{\hat{\beta}(W_N) - \beta_0\} \to_D N(0, \sigma^2(W^*)/g^2(W^*)),$$

where

$$\sigma^2(W^*) = \int_{-\infty}^{T^*} \{W^*(u, \beta_0)\}^2 A(u, \beta_0)\lambda(u)\, du,$$

$$g(W^*) = \int_{-\infty}^{T^*} W^*(u, \beta_0)\}A(u, \beta_0)\lambda'(u)\, du$$

and $A(u, \beta_0)$ is defined by regularity condition (F) in Section 3.

Remark. A simple application of the Cauchy–Schwartz inequality can be used to show that the asymptotic variance of the estimate $\hat{\beta}(W_N)$ will be minimized by choosing $W^*(u)$ [the limit of $W_N(u)$] to be proportional to $d\{\log \lambda(u)\}/du$ or $\lambda'(u)/\lambda(u)$. This result is similar to that of finding the optimal weight function for the hypothesis testing problem as shown by Schoenfeld (1981) and Gill (1980). It is shown by Ritov (1989) that the estimate $\hat{\beta}(W_N)$, assuming we can take T^* equal to infinity, is also the most efficient estimate of β among all semiparametric estimates when the error distribution indeed has hazard function $\lambda(u)$. Therefore, it seems likely that a fully efficient semiparametric estimate of β can be obtained by using a weight function which adaptively estimates $\lambda'(u)/\lambda(u)$. This, however, has not been considered in this paper.

We wish to note that the most efficient linear rank estimate $\hat{\beta}(W_N)$ is generally not fully efficient. That is, if we assume the linear model $T_i = \alpha + \beta Z_i + e_i$, where the e_i are iid with known distribution function $F(x)$ and corresponding hazard function $\lambda(x)$, then we can use standard likelihood methods to show that the maximum likelihood estimate $\hat{\beta}_M$ is fully efficient with asymptotic variance equal to the inverse of the expected information. In general, the asymptotic variance of $\hat{\beta}_M$ is smaller than the asymptotic variance of the most efficient linear rank estimate, although usually by not very much. In the special case, however, when the censoring is such that the distribution of $C_i - \beta_0 Z_i$ is independent of Z_i, then the most efficient linear rank estimate is fully efficient. This in practice would not be expected to occur except possibly in the case when $\beta_0 = 0$. The details of these calculations are straightforward but rather tedious and, therefore, have been omitted.

5. Extensions to multiple covariates. In this section we shall generalize the results of the previous sections to the multiple regression problem. Since the concepts are similar to the single covariate problem, the proofs will be omitted and only the main results will be presented. The assumptions will be the same as in Section 2, except that the underlying survival times T_i are linearly related to K covariates. That is,

$$T_i = \boldsymbol{\beta}'\mathbf{Z}_i + e_i, \quad \text{where } \boldsymbol{\beta} = (\beta_1, \ldots, \beta_K)' \text{ and } \mathbf{Z}_i = (Z_{i1}, \ldots, Z_{iK}).$$

We define a vector of estimating equations as the K linear rank tests for each of the K covariates in the regression model. That is, $\mathbf{S}_N(W_N, \boldsymbol{\beta}) = \{S_{Nj}(W_N, \boldsymbol{\beta}), j = 1, \ldots, K\}$, where

$$S_{Nj}(W_N, \boldsymbol{\beta}) = \sum_{i=1}^{N} \int_{-\infty}^{T^*} W_N(u)\, dN_i(u + \boldsymbol{\beta}'\mathbf{Z}_i)\{Z_{ij} - \bar{Z}_j(u, \boldsymbol{\beta})\},$$

where $\bar{Z}_j(u, \boldsymbol{\beta}) = \sum_{f=1}^{N} Z_{fj} Y_f(u + \boldsymbol{\beta}'\mathbf{Z}_f)/\sum_{f=1}^{N} Y_f(u + \boldsymbol{\beta}'\mathbf{Z}_f)$, and the weight function $W_N(u)$ satisfies the assumptions given in Section 4. Since $\mathbf{S}_N(W_N, \boldsymbol{\beta})$ is a discontinuous function, the corresponding linear rank estimates $\hat{\boldsymbol{\beta}}(W_N)$ are defined as the values $\boldsymbol{\beta}$ for which slight perturbations of its component would change the sign of $\mathbf{S}_N$. The main result is that the limiting distribution of $N^{1/2}(\hat{\boldsymbol{\beta}}(W_N) - \boldsymbol{\beta}_0)$ is a multivariate normal with mean 0 and covariance matrix equal to $G^{-1}V^*G^{-1}$, where V^* denotes the $K \times K$ matrix with elements

$$V_{jf}^* = \int_{-\infty}^{T^*} \{W^*(u, \beta_0)\}^2 A_{jf}(u, \beta_0)\lambda(u)\, du, \qquad j, f = 1, \ldots, K,$$

$A_{jf}(u, \beta_0)$ is the limit of

$$\sum_{i=1}^{N} \{Z_{ij} - \bar{Z}_j(u, \boldsymbol{\beta}_0)\}\{Z_{if} - \bar{Z}_f(u, \boldsymbol{\beta}_0)\} Y_i(u + \boldsymbol{\beta}_0'\mathbf{Z}_i)/N$$

and G denotes the $K \times K$ matrix with elements

$$G_{jf} = \int_{-\infty}^{T^*} W^*(u, \beta_0)\} A_{jf}(u, \beta_0)\lambda'(u)\, du.$$

6. Estimate of the covariance matrix. In the previous sections, it was shown that the linear rank estimate $\hat{\boldsymbol{\beta}}(W_N)$ is asymptotically normal. In order, however, to be able to use this estimate for purposes of statistical inference, we must construct a consistent estimate of the asymptotic variance. We showed in Section 5 that the asymptotic variance was given by $G^{-1}V^*G^{-1}$, where the (f, j)th element of V and G are given by

$$V_{fj}^* = \int \{W^*(u)\}^2 A_{fj}(u, 0)\lambda(u)\, du, \tag{6.1}$$

and

$$G_{fj} = \int_{-\infty}^{T^*} W^*(u) A_{fj}(u, 0)\lambda'(u)\, du. \tag{6.2}$$

These computations were made under the assumption that the true value $\boldsymbol{\beta}_0$ was without loss of generality taken to be equal to $\mathbf{0}$. If indeed $\boldsymbol{\beta}_0$ were equal to zero, then well-established results for linear rank tests [Andersen and Gill (1982) and Gill (1980)] can be used to show that a consistent estimate of V_{fj}^* can be obtained by substituting the empirical estimates in the integrand of (6.1). Namely, the estimate V_{fj}^* is equal to

$$\int_{-\infty}^{T^*} \{W_N(u)\}^2 \hat{A}_{fj}(u)\, d\hat{\Lambda}(u),$$

where

$$\hat{A}_{fj}(u) = \sum_{i=1}^{N} \left\{ \left(Z_{if} - \bar{Z}_f(u)\right)\left(Z_{ij} - \bar{Z}_j(u)\right) Y_i(u) \right\} / N \tag{6.3}$$

and $\hat{\Lambda}(u)$ is taken to be the Nelson (1969) estimate of the cumulative hazard function, namely,

$$\hat{\Lambda}(u) = \int_{-\infty}^{u} \left\{ \sum_{i=1}^{N} dN_i(x) \right\} \Big/ \left\{ \sum_{i=1}^{N} Y_i(x) \right\}.$$

Hence the estimate $\hat{V}_{fj}^*$ is equal to

$$N^{-1} \sum_{i=1}^{N} \int_{-\infty}^{T^*} \{W_N(u)\}^2 \times \left[\sum_{k=1}^{N} \left\{ \left(Z_{kf} - \bar{Z}_f(u)\right)\left(Z_{kj} - \bar{Z}_j(u)\right) Y_k(u) \right\} \Big/ \sum_{k=1}^{N} Y_k(u) \right] dN_i(u).$$

This estimate is a consistent estimate of V_{fj}^* when $\boldsymbol{\beta}_0 = \mathbf{0}$. If $\boldsymbol{\beta}_0$ were not equal to 0, the problem could be transformed to that above by considering the residuals $X_i - \boldsymbol{\beta}_0'\mathbf{Z}_i$ and applying the estimate to the residuals. However, since the value $\boldsymbol{\beta}_0$ is not known, we propose computing the estimate of V_{fj}^* based on the estimated residuals $(X_i - \hat{\boldsymbol{\beta}}(W_N)'\mathbf{Z}_i)$, $i = 1, \dots, N$. The proof that the estimate $\hat{V}_{fj}^*$, using these residuals, is a consistent estimate of V_{fj}^* has not been verified at this point; however, some preliminary numerical results based on simulations seem to indicate that this estimate indeed works well.

We must also find an estimate for G_{fj}. Integrating (6.2) by parts we obtain that

$$G_{fj} = -\int_{-\infty}^{T^*} \lambda(u)\, d\left\{ W^*(u) A_{fj}(u) \right\}. \tag{6.4}$$

We propose estimating G_{fj} by substituting consistent estimates for each of the terms in the integrand of (6.4). Namely, we propose estimating G_{fj} by

$$\hat{G}_{fj} = -\int_{-\infty}^{T^*} \hat{\lambda}(u)\, d\left\{ W_n(u) \hat{A}_{fj}(u) \right\}.$$

Since $\hat{A}_{fj}(u)$ given in (6.3) is a step function changing at each value of X_i, as is the weight function $W_N(u)$ in most cases, then the estimate $\hat{G}_{fj}$ can be written as

$$\hat{G}_{fj} = \sum_{i=1}^{N} \hat{\lambda}(X^{(i)}) \left\{ W_N(X^{(i)}) \hat{A}_{fj}(X^{(i)}) - W_N(X^{(i+1)}) \hat{A}_{fj}(X^{(i+1)}) \right\}, \tag{6.5}$$

where $X^{(i)}$ denotes the ith order statistic of X_i, $i = 1, \dots, N$. In order that this estimate be consistent, we must find a consistent estimate for the hazard function $\lambda(u)$. We propose using the kernel estimate

$$\hat{\lambda}(u) = \left[\int \{dN(x)/Y(x)\} w\{(x-u)/h_N\}\right] / h_N,$$

where w denotes the kernel function, h_N denotes the window width, $N(x) = \sum N_i(x)$ and $Y(x) = \sum Y_i(x)$. Properties of this estimate, including consistency, have been studied by Ramlau-Hansen (1983) and Tanner (1983). The estimate $\hat{G}_{fj}$ must be applied to the estimated residuals $(X_i - \hat{\beta}(W_N)'\mathbf{Z}_i)$, $i = 1, \ldots, N$. This creates extra difficulty as $\hat{\beta}(W_N)$ is itself an estimate and therefore a rigorous proof of the consistency of $\hat{G}_{fj}$ has not been established, but again numerical results based on simulations seems to indicate this estimate performs well. We also noticed in the numerical calculations that changes in the window width within reasonable limits did not affect the estimate by much. However, more investigation in the choice of $\hat{\lambda}(u)$ has to be made before making any definitive recommendations.

7. Concluding remarks. In this paper we have proposed a class of linear rank estimates $\hat{\beta}(W_N)$ for estimating regression parameter β in a linear model with right censored data. We have shown that there exist estimators which are solutions to estimating equations derived from linear rank tests for censored data that are asymptotically normal with covariance matrix for which an estimate is proposed. We have also shown that with the appropriate weight function $W_N(u)$, these estimates are close to being fully efficient.

The results presented here are primarily theoretical. Before proposing such estimates for general use, a more thorough applied investigation must be conducted. We already noted that the proposed estimate of the covariance matrix uses kernel methods for estimating the hazard function. Recommendations for reasonable kernel functions and, more importantly, the window width still need to be considered.

Efficient numerical methods for computing these estimates have to also be examined. For some preliminary numerical work, we considered using a Newton–Raphson type algorithm for finding solutions to the equations $\mathbf{S}_N(W_N, \hat{\beta}) = \mathbf{0}$. This was motivated by the fact that the estimating equation is approximately linear. That is,

$$\mathbf{S}_N(W_N, \beta) = \mathbf{S}_N(W_N, \beta_0) + G^{K \times K}(\beta_0)(\beta - \beta_0).$$

This suggests that we update our estimates by

$$\beta^{(i+1)} = \beta^{(i)} - \hat{G}^{-1}(\beta^{(i)})\mathbf{S}_N(W_N, \beta^{(i)}),$$

where $\beta^{(i)}$ is the ith iteration and $\hat{G}$ is the estimate given by (6.5), and continue until the values converge. We found that such a procedure converged quickly to a range near the solution, but did not necessarily give unique answers.

The estimator proposed here is a competitor to the Buckley–James estimate [Buckley and James (1979)] for linear regression with right censored data. A numerical comparison of the relative performance of these two estimates will be very useful.

There also remain other theoretical questions to be considered. No doubt many of the stringent regularity conditions made in this paper can be substantially weakened.

Acknowledgments. I would like to thank Myrto Lefkopoulou for helpful discussions during the preparation of this manuscript. I would also like to thank the Associate Editor and the referees for their many helpful suggestions.

I owe a very special thanks to Professor van Zwet for his careful attention to this paper and who proposed the use of Hoeffding's decomposition theorem, which substantially simplified the proof of Theorem 3.2.

REFERENCES

ANDERSEN, P. K. and GILL, R. D. (1982). Cox's regression model for counting processes: A large sample study. *Ann. Statist.* **10** 1100–1120.

ANDERSEN, P. K., BORGAN, O., GILL, R. and KEIDING, N. (1982). Linear nonparametric tests for comparison of counting processes, with applications to censored survival data. *Internat. Statist. Rev.* **50** 219–258.

BEGUN, J. M., HALL, W. J., HUANG, W. and WELLNER, J. A. (1983). Information and asymptotic efficiency in parametric–nonparametric models. *Ann. Statist.* **11** 432–452.

BICKEL, P. J. (1986). Efficient testing in a class of transformation models. In *Papers on Semiparametric Models at the ISI Centenary Session* (R. D. Gill and M. N. Voors, eds.) 63–81. Report MS-R8614, Centre for Mathematics and Computer Science, Amsterdam.

BUCKLEY, J. and JAMES I. (1979). Linear regression with censored data. *Biometrika* **66** 429–436.

COX, D. R. (1972). Regression models and life tables (with discussion). *J. Roy. Statist. Soc. Ser. B* **34** 187–220.

CUZICK, J. (1985). Asymptotic properties of censored linear rank tests. *Ann. Statist.* **13** 133–141.

CUZICK, J. (1988). Rank regression. *Ann. Statist.* **16** 1369–1389.

DOKSUM, K. A. (1987). An extension of partial likelihood methods for proportional hazards models to general transformation models. *Ann. Statist.* **15** 325–345.

GEHAN, E. (1965). A generalized Wilcoxon test for comparing arbitrarily singly-censored samples. *Biometrika* **52** 203–223.

GILL, R. D. (1980). *Censoring and Stochastic Integrals*. Mathematical Center Tract 124. Mathematische Centrum, Amsterdam.

HARRINGTON, D. P. and FLEMING, T. R. (1982). A class of rank test procedures for censored survival data. *Biometrika* **69** 553–566.

HETTMANSPERGER, T. P. (1984). *Statistical Inference Based on Ranks*. Wiley, New York.

JUREČKOVÁ, J. (1969). Asymptotic linearity of a rank statistic in regression parameter. *Ann. Math. Statist.* **40** 1889–1900.

JUREČKOVÁ, J. (1971). Nonparametric estimate of regression coefficients. *Ann. Math. Statist.* **42** 1328–1338.

LENGLART, E. (1977). Relation de domination entre deux processus. *Ann. Inst. H. Poincaré* **13** 171–179.

LOUIS, T. A. (1981). Nonparametric analysis of an accelerated failure time model. *Biometrika* **68** 381–390.

MEHROTRA, K. G., MICHALEK, J. E. and MIHALKO, D. (1982). A relationship between two forms of linear rank procedures for censored data. *Biometrika* **69** 674–676.

NELSON, W. (1969). Hazard plotting for incomplete failure data. *J. Qual. Technol.* **1** 27–52.

PETO, R. and PETO, J. (1972). Asymptotically efficient rank invariant test procedures. *J. Roy. Statist. Soc. Ser. A* **135** 185–206.

PRENTICE, R. L. (1978). Linear rank tests with right censored data. *Biometrika* **65** 167–179.

RAMLAU-HANSEN, H. (1983). Smoothing counting process intensities by means of kernel functions. *Ann. Statist.* **11** 453–466.

RITOV, Y. (1989). Estimation in a linear regression model with censored data. Unpublished manuscript.

SCHOENFELD, D. (1981). The asymptotic properties of nonparametric tests for comparing survival distributions. *Biometrika* **68** 316–319.

TANNER, M. A. (1983). A note on the variable kernel estimator of the hazard function from randomly censored data. *Ann. Statist.* **11** 994–997.

WEI, L. J. and GAIL, M. H. (1983). Nonparametric estimation for a scale-change with censored observations. *J. Amer. Statist. Assoc.* **78** 382–388.

VAN ZWET, W. R. (1984). A Berry–Esseen bound for symmetric statistics. *Z. Wahrsch. Verw. Gebiete* **66** 425–440.

DIVISION OF BIOSTATISTICS
DANA-FARBER CANCER INSTITUTE
44 BINNEY STREET
BOSTON, MASSACHUSETTS 02115