The Annals of Statistics
1994, Vol. 22, No. 2, 732–746

MONOTONE ESTIMATING EQUATIONS FOR CENSORED DATA

BY MENDEL FYGENSON AND YA'ACOV RITOV

Rutgers University and Hebrew University of Jerusalem

The monotone class rank-test-based estimating equations for regression models with right censored data is considered. We introduce an estimator which is a solution of a monotone estimating equation that is an extension of the Gehan test. The estimator is easy to derive, $\sqrt{n}$-consistent and asymptotically normal under minimal conditions. All monotone estimating equations are characterized, and a simulation study, which shows that our suggested procedure performs well, is included.

1. Introduction. We consider the problem of estimating the regression coefficients in the accelerated failure time model. Let T and C be real random variables, and let Z be an R^d random vector of covariates distributed such that $\varepsilon = T - \beta_0^T Z$ is independent of (Z, C). Let $\Delta = I(T \geq C)$ and $Y = \min(T, C)$, where $I(\cdot)$ is the indicator function. We observe a random sample $X_1, X_2, \dots, X_n$, from $X = (Z, Y, \Delta)$ and wish to estimate β_0. The distributions of ε and (Z, C) are considered to be unknown and arbitrary.

This model is used in survival analysis where usually T is the log of the lifetime and C is the log of the censoring time. The majority of the proposed estimators for the accelerated failure time model were obtained by modifying the least squares method to accommodate right censoring. Most of these methods depend on specific assumptions concerning the distribution of X and can be used if a particular submodel of the general case is assumed [cf. Miller (1976), Koul, Susarla and Van Ryzin (1981), Wei and Gail (1983), Powell (1984), Leurgans (1987) and Fygenson and Zhou (1994)].

For estimating the regression coefficient in the case of uncensored data, many advocate the use of methods like the M-estimators or rank procedures to overcome some of the robustness limitations of the least squares method.

Ritov (1990) generalized the method of Buckley and James (1979) and introduced a family of M-estimators for the censored regression model. The estimators were defined as a generalized solution of the following estimating equations:

$$\sum_{i=1}^{n}(Z_i - \overline{Z})\left\{\Delta_i s(Y_i - \beta^T Z_i) - (1-\Delta)\frac{\int_{Y_i - \beta^T Z_i}^{\infty} s(t)\, d\widehat{F}^{\beta}(t)}{1 - \widehat{F}^{\beta}(Y_i - \beta^T Z_i)}\right\} = 0, \tag{1.1}$$

where $s(\cdot)$ is some score function $\overline{Z} = n^{-1}\sum_{i=1}^{n} Z_i$ and $\widehat{F}^{\beta}(\cdot)$ is the Kaplan–Meier estimator based on the residuals $Y_1 - \beta^T Z_1, \dots, Y_n - \beta^T Z_n$. Another approach

Received January 1991; revised May 1993.
AMS 1991 *subject classifications.* Primary 62G05; secondary 62J05, 62E20, 62E25.
Key words and phrases. Right censored data, rank tests, Kendall's tau, information bound, monotonicity, Kaplan–Meier estimator.

was suggested by Tsiatis (1990). He constructed a class of estimators using linear rank tests for censored data. His estimators were defined as a generalized solution of the following estimating equations:

$$S_n(\beta) = \sum_{i=1}^{n} w_i \Delta_i (Z_i - \overline{Z}_i) = 0, \tag{1.2}$$

where w_i is some nonnegative weight function of the observations and where

$$\overline{Z}_i = \frac{\sum_{j=1}^{n} Z_j I\left(v_j^\beta \geq v_i^\beta\right)}{\sum_{j=1}^{n} I\left(v_j^\beta \geq v_i^\beta\right)}$$

is the mean of the covariate of the "risk group" at $v_i^\beta = Y_i - \beta^T Z_i$.

Both Tsiatis (1990) and Ritov (1990) proved that their estimating equations have, under regularity conditions, a solution which is $\sqrt{n}$-consistent and asymptotically normal. However, neither was able to prove that their estimating equations have a reasonable global behavior. In particular, they were not able to show that the equations have a unique solution, nor were they able to devise an algorithm which was known to converge. In addition, both papers lack any information on the actual behavior of the suggested estimators and neither addresses the issues of robustness.

These omissions are directly related to the fact that equations (1.1) and (1.2) are not monotone in β. The lack of monotonicity also necessitates strong assumptions and lengthy complex proofs of the estimators' properties. Most important, the nonmonotonicity limits (at best) the use of the proposed estimation methods in practice. (See Section 3.)

The main results of this paper are as follows. A monotone estimating equation is proposed. This equation is based on Gehan's (1965) modification of the Wilcoxon test for right censored data. Due to the monotonicity of the estimating equation, the set of its generalized solutions is convex, and it is relatively easy to locate an estimator and to establish its properties. This estimator is $\sqrt{n}$-consistent and asymptotically normal under minimal conditions which are much weaker than the ones imposed by Tsiatis (1990). This estimator appears to be the first reasonable estimator that is generally applicable to our model and can also be used as a starting point for estimators that are more difficult to compute. Moreover, it is straightforward to construct a confidence interval for β using this estimator. It turns out that the proposed estimating equation is a member of Tsiatis' family, although it was not investigated by him in particular. Finally, the subfamily of (1.2) for which the resulting estimating equations are monotone in β is characterized.

The article is constructed as follows. In Section 2, we introduce the monotone equation, derive the properties of its estimators and construct confidence intervals for β_0. In Section 3, we investigate the existence of other monotone estimating equations that belong to Tsiatis' family. A simulation study for the estimator proposed in Section 2 is presented in Section 4.

2. A monotone estimating equation. For the uncensored case, that is, $C_i = \infty$ for $i = 1, 2, \dots$, it is, proposed by Sen (1968) and others to consider the estimating equation

$$W_n(\beta) = n^{-3/2} \sum_{i=1}^{n} \sum_{j=1}^{n} (Z_i - Z_j) I\left(v_j^\beta > v_i^\beta\right), \tag{2.1}$$

where $v_i^\beta = Y_i - \beta^T Z_i$.

Since $EW_n(\beta_0) = 0$, one may try to estimate the slope by β, which makes $W_n(\beta)$ as close to zero as possible. This is an extension of the test for β_0 based on the Kendall correlations between the residuals and the covariates. Since $W_n(\beta)$ is not continuous in β, the equation $W_n(\beta) = 0$ may fail to have a solution. We call β a generalized solution of $W_n(\beta) = 0$ if slight perturbations of any of its components change the sign of W_n.

It is important to note that $W_n(\beta)$ is a monotone nondecreasing field. [Note that a function $W(\beta)$: $R^d \to R^d$ is called a monotone nondecreasing field if, for any $\beta, \xi \in R^d$, $\xi^T W(\beta + x\xi)$ is a monotone nondecreasing function of the real variable x]. It follows that all the generalized solutions of $W_n(\beta) = 0$ belongs to a convex set whose diameter is $O(n^{-1})$.

The question is how to generalize this method to the censored case. There are at least two possible ways. The simplest way is to include only those pairs for which we know that $T_i - \beta^T Z_i < T_j - \beta^T Z_j$, that is, to consider

$$W_n(\beta) = n^{-3/2} \sum_{i=1}^{n} \sum_{j=1}^{n} (Z_i - Z_j) \Delta_i I\left(Y_j - \beta^T Z_j > Y_i - \beta^T Z_i\right). \tag{2.2}$$

This estimating equation is monotone under right censoring, and therefore the set of its generalized solutions is convex and it is relatively easy to locate a well-behaved solution. Another way is to compare all possible pairs and to weight each comparison according to the number of censored residuals in the pair. As an example we consider Efron's (1967) modification to the Wilcoxon test in which the resulting estimating equation is

$$\begin{aligned} W_n^*(\beta) = \sum_{i,j=1}^{n} (Z_j - Z_i) \Bigg[& \Delta_i \Delta_j + \left(1 - \frac{2\hat{\bar{F}}_i^\beta}{\hat{\bar{F}}_j^\beta}\right) \Delta_i (1 - \Delta_j) \\ & + \left(1 - \frac{\hat{\bar{F}}_i^\beta}{\hat{\bar{F}}_j^\beta}\right)(1 - \Delta_i)(1 - \Delta_j) + \Delta_j (1 - \Delta_i) \Bigg] I\left(v_i^\beta > v_j^\beta\right), \end{aligned} \tag{2.3}$$

where $\hat{\bar{F}}_i^\beta$ is the Kaplan–Meier estimator of the survival function at v_i^β, based on all the residuals. This equation is not monotone and therefore will not be analyzed further here. However, it can be shown that this equation is also a member of Tsiatis' class.

To investigate the properties of (2.2), we need the following notation. Denote the probability density function and the distribution function of ε by f and F, respectively. A bar above a distribution function denotes a survival function; thus $\overline{F}(\cdot) = 1 - F(\cdot)$. Let $G(\cdot)$ be the marginal distribution function of the centered censoring variable $C - \beta_0^T Z$, and let $G(\cdot \mid Z)$ be its conditional distribution given the covariate. Let $H(\cdot)$ denote the distribution of the residuals without regard to whether they are censored, that is, $\overline{H}(t) = \overline{G}(t)\overline{F}(t)$. We denote by H_u the substochastic distribution of the uncensored residuals, that is, $dH_u(t) = \overline{G}(t)f(t)\,dt$. Let

$$R = \int \operatorname{var}(Z \mid C - \beta_0^T Z \geq u)\overline{H}(u)\left(\frac{f'(u)}{f(u)} + \frac{f(u)}{\overline{F}(u)}\right)\overline{G}(u)f(u)\,du$$

and

$$J = \int \operatorname{var}(Z \mid C - \beta_0^T Z \geq u)\overline{H}^2(u)\overline{G}(u)f(u)\,du,$$

where

$$\operatorname{var}(Z \mid C - \beta_0^T Z \geq u) = \frac{E\left[\{Z - D(u)\}\{Z - D(u)\}^T I(C - \beta_0^T Z \geq u)\right]}{\overline{G}(u)}$$

and

$$D(u) = E(Z \mid C - \beta_0^T \geq u).$$

We obtain the following main results.

THEOREM 2.1. *Suppose that $E(\|Z\|^2) < \infty$, that the distribution of ε has finite Fisher information and that the distribution of Z given $\Delta = 1$ is not concentrated on a proper hyperplane of R^d.*

(i) *Let*

$$B_n = \left\{ \max_{\substack{\Delta_i = 1\\ 1 \leq i \leq n}} \beta^T Z_i > \min_{1 \leq i \leq n} \beta^T Z_i,\ \forall\, \beta \in R^d \right\}.$$

Then $P(B_n) \to 1$ and $W_n(\beta) = 0$ has a generalized solution on B_n.

(ii) *For any fixed β, $W_n(\beta) - \sqrt{n}A(\beta)$ is asymptotically normal with mean 0, where $A(\cdot)$ is nonstochastic such that $A(\beta_0) = 0$ and $A(\beta)$: $R^d \to R^d$ is monotone. The derivative $\dot{A}$ of $A(\cdot)$ at $\beta_0 = 0$ is invertible. In particular, $W_n(\beta_0)$ is asymptotically normal with mean zero and finite variance. Moreover, for any $M < \infty$ and $v > 0$ there are $C_1, C_2 < \infty$ such that*

$$P\left[\sup_{\|t\| < M} \left|W_n(\beta_0 + t/\sqrt{n}) - \dot{A}(\beta_0)t\right| > C_1\right] < v$$

and

$$P\left[\sup_{\|t\| < M} \left|W_n(\beta_0 + t) - \sqrt{n}A(\beta_0 + t)\right| > C_2\right] < v.$$

(iii) *Let $\widehat{\beta}$ be any generalized solution of $W_n(\beta) = 0$. Then $\sqrt{n}(\widehat{\beta} - \beta_0)$ is asymptotically normal with mean* 0 *and variance-covariance matrix $R^{-1}JR^{-1}$.*

PROOF. Since the support of ΔZ spans R^d, we obtain that $P(B_n) \to 1$. Now the function $W_n(\cdot)$ is the gradient of the function

$$Q_n(\beta) = n^{-3/2} \sum_{i=1}^{n} \sum_{j=1}^{n} \beta^T(Z_i - Z_j) I\{\beta^T(Z_i - Z_j) > Y_i - Y_j\}.$$

Since $W_n(\cdot)$ is monotone, $Q_n(\cdot)$ is convex. On B_n

$$\lim_{\|\beta\| \to \infty} Q_n(\beta) = \infty.$$

However, $Q_n(\beta_0)$ is finite and $Q_n(\cdot)$ is convex, and, hence, $Q_n(\cdot)$ has a minimum at some point $\widehat{\beta}$. Of course, $\widehat{\beta}$ is a generalized solution of $W_n(\beta) = 0$. The first claim follows.

For any fixed β, $W_n(\beta)$ is a simple U-statistic and, hence, its asymptotic normality and its expectation are immediate. Assume, without any loss of generality, that $\beta_0 = 0$. Write now $W_n(\beta)$ as a U-statistic with a symmetric kernel:

$$W_n(\beta) = n^{-3/2} \sum_{i=2}^{n} \sum_{j=1}^{i-1} (Z_i - Z_j)\Big\{\Delta_i I\big(v_j^\beta > v_i^\beta\big) - \Delta_j I\big(v_i^\beta > v_j^\beta\big)\Big\}. \tag{2.4}$$

A direct calculation yields that the expectation of $W_n(\beta)$ is $\sqrt{n}A(\beta) + o(n^{-1/2})$, where

$$\begin{aligned} A(\beta) = \tfrac{1}{2}E\Big[&(Z_1 - Z_2)(Z_1 - Z_2)^T \int \overline{G}\big(t - \beta^T Z_1 \mid Z_1\big)\overline{G}\big(t - \beta^T Z_2 \mid Z_2\big) \\ &\times \Big\{\overline{F}\big(t - \beta^T Z_2\big) f\big(t - \beta^T Z_1\big) - \overline{F}\big(t - \beta^T Z_1\big) f\big(t - \beta^T Z_2\big)\Big\}\, dt\Big]. \end{aligned}$$

If $\beta \to 0$, then, since f has finite Fisher information and $E(Z_i^2) < \infty$,

$$\begin{aligned} A(\beta) &= \frac{1}{2}E\Big[(Z_1 - Z_2)(Z_1 - Z_2)^T \\ &\qquad \times \int \overline{G}\big(t - \beta^T Z_1 \mid Z_1\big)\overline{G}\big(t - \beta^T Z_2 \mid Z_2\big) \\ &\qquad\qquad \times \Big\{\overline{F}(t)\frac{f'(t)}{f(t)} - f(t)\Big\} f(t)\, dt\Big]\beta + o(\beta) \\ &= \frac{1}{2}E\Big[(Z_1 - Z_2)(Z_1 - Z_2)^T \\ &\qquad \times \int \overline{G}(t \mid Z_1)\overline{G}(t \mid Z_2)\Big\{\overline{F}(t)\frac{f'(t)}{f(t)} - f(t)\Big\} f(t)\, dt\Big]\beta + o(\beta). \end{aligned}$$

The linearity of A near β_0 follows. Moreover, the matrix derivative of A given by

$$\dot{A}\beta = \frac{1}{2}E\left[(Z_1 - Z_2)(Z_1 - Z_2)^T \int \overline{G}(t \mid Z_1)\overline{G}(t \mid Z_2)\left\{\overline{F}(t)\frac{f'(t)}{f(t)} - f(t)\right\}f(t)\,dt\right]$$

$$= \frac{1}{2}E\left[(Z_1 - Z_2)(Z_1 - Z_2)^T \int \overline{F}(t)f(t)\,dK(t \mid Z_1, Z_2)\right],$$

where $K(t \mid Z_1, Z_2) = -\overline{G}(t \mid Z_1)\overline{G}(t \mid Z_2)$ is an increasing function. From the assumption that the distribution of Z given $\Delta = 1$ is not concentrated on a proper hyperplane in R^d, it follows that $\dot{A}$ is positive definite.

The asymptotic normality of $W_n(\beta)$ follows from the standard U-statistics theory [see Serfling (1980)]. In particular, for $\beta = 0$, we obtain that $W_n(0) = \frac{1}{2}n^{-1/2}\Sigma_{i=1}^n \xi_i + o(1)$, where $\xi_1, \xi_2, \dots, \xi_n$ are independent and identically distributed and

$$\xi_1 = E\Big[(Z_1 - Z_2)\big\{\Delta_1 I(Y_2 > Y_1) - \Delta_2 I(Y_1 > Y_2)\big\} \mid Z_1, Y_1, \Delta_1\Big]$$

$$= \Delta_1\left[\big\{Z_1 - D(Y_1)\overline{H}(Y_1)\big\} - \int_{-\infty}^{Y_1} \big\{Z_1 - D(t)\big\}\,dH_u(t)\right]$$

$$- (1 - \Delta_1)\int_{-\infty}^{Y} \big\{Z_1 - D(t)\big\}\,dH_u(t).$$

The uniform convergence follows from the monotonicity of $W_n(\cdot)$ and $A(\cdot)$ [see Brown (1985) and Ritov (1987)]. Having proved the first two parts, the last part is an immediate conclusion from Tsiatis (1990) and Ritov (1990): First, since $\dot{A}$ is invertible and ξ_i has a finite variance, it follows that the generalized solution for $W_n(\beta) = 0$ is $\sqrt{n}$-consistent. Moreover, $W_n(\beta)$ is equivalent to Tsiatis' estimating function with a weight function $\overline{H}(\cdot)$. It follows from Ritov [(1990), Remark 6.5] that the estimator is, therefore, equivalent to the solution of (1.1) with score function

$$s(t) = \overline{H}(t) - \frac{\int_t^\infty \overline{H}(\tau)f(\tau)\,d\tau}{\overline{F}(t)}.$$

The asymptotoic variance formula follows from Ritov [(1990), Theorem 5.1]. □

The information bound for estimating β_0 is a lower bound, derived by semiparametric methods, on the spread of the best possible estimator of β_0. It is derived in Ritov and Wellner (1988) and Bickel, Klaassen, Ritov and Wellner (1991), and it is equal to J.

COROLLARY 2.1. *Suppose that the information bound for estimating β is finite and invertible. Then the generalized solution of $W_n(\beta) = 0$ is $\sqrt{n}$-consistent.*

PROOF. The corollary follows since the conditions of the theorem are exactly the conditions for a finite and invertible information bound [see Ritov (1990), Remark 6.3]. □

Finally in this section, we establish a confidence interval for β_0. For simplicity we take β_0 to be a real parameter. The asymptotic variance of the generalized solution of an estimating equation $W_n(\beta) = 0$ is given by $\operatorname{var}\{W_n(\beta_0)\}/\dot{A}^2(\beta_0)$. Since $W_n(\beta_0)$ is a simple U-statistic [see (2.4)], its asymptotic variance is

$$V = \operatorname{var}\Big[E(Z_1 - Z_2)\big\{\Delta_1 I(Y_2 - \beta_0 Z_2 > Y_1 - \beta_0 Z_1) - \Delta_2 I(Y_1 - \beta_0 Z_1 > Y_2 - \beta_0 Z_2) \mid Z_1, Y_1, \Delta_1\big\}\Big]$$

which can be estimated by

$$\widehat{V} = \frac{1}{n}\sum_{i=1}^{n}\left[\sum_{j=2}^{n}(Z_i - Z_j)\big\{\Delta_i I(Y_j - \widehat{\beta} Z_j > Y_i - \widehat{\beta} Z_i) - \Delta_j I(Y_i - \widehat{\beta} Z_i > Y_j - \widehat{\beta} Z_j)\big\}\right]^2, \tag{2.5}$$

where $\widehat{\beta}$ is the estimator of β_0. The derivative can be estimated by

$$\widehat{A} = \frac{\{W_n(\widehat{\beta} + c_n) - W_n(\widehat{\beta} - d_n)\}}{2c_n}, \tag{2.6}$$

where $c_n + d_n \to 0$ but $\liminf \sqrt{n}(c_n + d_n) > 0$. As a result we obtain that $\sqrt{\widehat{V}}/|\widehat{A}|$ is a natural estimator of the standard deviation of the solution. One can therefore use

$$\left(\widehat{\beta} - z_{\alpha/2}\frac{\widehat{V}^{1/2}}{|\widehat{A}\,|},\ \widehat{\beta} + z_{\alpha/2}\frac{\widehat{V}^{1/2}}{|\widehat{A}\,|}\right) \tag{2.7}$$

as a $(1-\alpha)$ confidence interval for β_0.

3. The class of monotone estimating equations. Consider Tsiatis' estimating equation,

$$S_n(\beta) = \sum_{i=1}^{n}\Delta_i(Z_i - \overline{Z}_i)w_i = 0. \tag{3.1}$$

In general, $S_n(\beta)$ is not a monotone function of β, primarily due to the presence of right censoring. The nonmonotonicity may cause the estimating equation to be practically useless in some applications. To see this we consider the behavior of the function $S_n(\beta)$.

Tsiatis (1990) has shown that $S_n(\beta)$ is asymptotically linear in a neighborhood of the true value β_0 (see Figure 1). In practice, one begins with some initial value β_1 and iterates the equation $S_n(\beta) = 0$ in the hope of reaching the true β_0. Unless the starting value chosen belongs to the "right" neighborhood, one

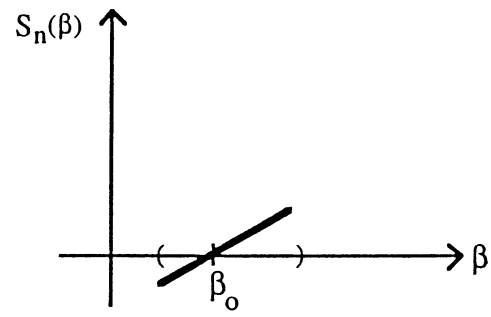

FIG. 1. *The asymptotic behavior of $S_n(\beta)$ in a neighborhood of the true value β_0.*

may not reach the true β_0 or any solution at all. This may occur if $S_n(\beta)$ behaves as in Figure 2. The situation is much worse when the vector parameter β is of high dimension since one cannot plot the function $S_n(\beta)$. To avoid these difficulties, one can use a monotone estimating equation which will insure that the iteration reaches (with probability converging to 1) a small neighborhood of the true β_0 regardless of the intial value β_1.

Although it has drawbacks, the idea of using (3.1) as an estimating equation for censored data is very important. It provides us with a method of estimation which is very flexible and introduces many estimators for which the corresponding test statistics have been widely used and investigated. In particular, letting $w_i = R_i^\beta$, where R_i^β is the risk group at v_i^β, one can derive (2.2) by simple rearrangement of the terms.

It follows from our discussion that it would be of central importance to identify all the monotone estimating equations that belong to Tsiatis' family. These equations will differ only with respect to their weight functions. Thus, equivalently, we characterize the class of nonnegative weights for which the resulting estimating equations are monotone.

Throughout what follows, we denote $v_{(1)}^\beta < \cdots < v_{(n)}^\beta$ as the ordered residuals and $Z_1^\beta, \dots, Z_n^\beta, \Delta_1^\beta, \dots \Delta_n^\beta$ and $W_1^\beta, \dots, W_n^\beta$ as their covariates, indicators and weights, respectively. With the above notation, the function $S_n(\beta)$ can be written

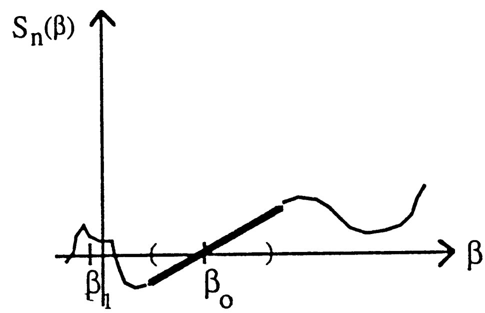

FIG. 2. *The asymptotic behavior of $S_n(\beta)$ including "bad" neighborhoods.*

as

$$S_n(\beta) = \sum_{j=1}^{n} \Delta_j^\beta W_j^\beta \left[Z_j^\beta - \sum_{i=1}^{n} \frac{Z_i^\beta}{n-j+1} I\left(v_{(j)}^\beta \geq v_{(i)}^\beta\right) \right] \tag{3.2}$$

Since these equations are linear rank statistics, we restrict ourselves to the class $\mathcal{W}$ of those weights which are predictable and functions of the residuals through their ranks, that is, $W_k^\beta = W_{nk}(\Delta_1^\beta, \dots, \Delta_{k-1}^\beta)$.

THEOREM 3.1. *The function $S_n(\beta)$: $R^d \to R^d$ is a monotone nondecreasing random field if and only if*

$$W_k^\beta = (n-k+1) b_n\left(u_k^\beta\right),$$

where $u_k^\beta = \Sigma_{1<k} \Delta_i^\beta$ and the sequence $\{b_n(i)\}_{i=1}^n$ satisfies

$$b_n(i) \leq b_n(i-1) \frac{n-i+1}{n-i-1}.$$

PROOF. It is clear that a change in $S_n(\beta)$ occurs only when two or more residuals interchange positions. Let $\delta \in R^d$ be a unit vector, and suppose that the kth- and $(k+1)$st-order residuals interchange position between $\beta- = \beta + (0-)\delta$ and $\beta+ = \beta + (0+)\delta$. When this takes place, we have

$$\Delta_k^{\beta\pm} = \Delta_{k+1}^{\beta\mp}, \quad Z_k^{\beta\pm} = Z_{k+1}^{\beta\mp} \quad \text{and} \quad \Delta_j^{\beta\pm} = \Delta_j^\beta, \quad Z_j^{\beta\pm} = Z_j^\beta \quad \text{for all } j \neq k, k+1.$$

The resulting change in the value of the function S_n is

$$\begin{aligned} D_n(\beta, \delta) = {} & \left(Z_{k+1}^{\beta+} - \overline{Z}_{k+2}^\beta\right) \left(\Delta_{k+1}^{\beta+} W_{k+1}^{\beta+} \frac{n-k-1}{n-k} - \Delta_k^{\beta+} W_k^\beta \frac{1}{n-k+1} \right. \\ & \qquad \left. - \Delta_{k+1}^{\beta+} W_k^\beta \frac{n-k}{n-k+1} \right) \\ & + \left(Z_k^{\beta+} - \overline{Z}_{k+2}^\beta\right) \left(\Delta_k^{\beta+} W_k^{\beta+} \frac{n-k}{n-k+1} - \Delta_k^{\beta+} W_{k+1}^{\beta-} \frac{n-k-1}{n-k} \right. \\ & \qquad \left. + \Delta_{k+1}^{\beta+} W_k^\beta \frac{1}{n-k+1} \right) \\ & + \sum_{j=k+2}^{n} \Delta_j^\beta \left(W_j^{\beta+} - W_j^{\beta-}\right)\left(Z_j^\beta - \overline{Z}_j^\beta\right). \end{aligned} \tag{3.3}$$

We need $\delta^T D_n(\beta,\delta) \geq 0$ for the monotonicity. This should hold in particular when $\Delta_k^{\beta+} + \Delta_{k+1}^{\beta+} = 1$, whatever the values of $Z_j - \overline{Z}_j$, for $j > k+1$. Hence the third term on the RHS of (3.3) must be 0. We conclude that W_j^β should be independent of any permutation of $\Delta_1, \dots, \Delta_{j-1}$, that is,

$$W_j^\beta = w_n\left(j, u_j^\beta\right), \tag{3.4}$$

where $u_j^\beta = \sum_{i<j} \Delta_i^\beta$. Next consider the first two terms on the RHS of (3.3), and suppose first that $\Delta_{k+1}^{\beta+} = 1$ while $\Delta_k^{\beta+} = 0$. We obtain that, in this case,

$$\begin{aligned} D_n(\beta,\delta) &= \left(Z_{k+1}^{\beta+} - \overline{Z}_{k+2}^{\beta}\right)\left(W_{k+1}^{\beta+}\frac{n-k-1}{n-k} - W_k^\beta \frac{n-k}{n-k+1}\right) \\ &\quad + \left(Z_k^{\beta+} - \overline{Z}_{k+2}^{\beta}\right) W_k^\beta \frac{1}{n-k+1}. \end{aligned} \tag{3.5}$$

We conclude from (3.5) that, to ensure that $\delta^T D_n(\beta,\delta) \geq 0$ whenever $\delta^T(Z_k - Z_{k+1}) \geq 0$, we need

$$W_k^\beta \frac{1}{n-k+1} = -\left(W_{k+1}^{\beta+}\frac{n-k-1}{n-k} - W_k^\beta \frac{n-k}{n-k+1}\right) \geq 0. \tag{3.6}$$

Since in this case $u_{k+1}^{\beta+} = u_k^\beta$, we obtain from (3.4) and (3.6) that

$$\begin{aligned} w_n(k+1,u) &= \frac{n-k}{n-k+1} w_n(k,u) \\ &= \frac{n-k}{n-u} w_n(u+1,u), \end{aligned}$$

that is,

$$w_n(k,u) \equiv b_n(u)(n-k+1), \tag{3.7}$$

for positive b. The case $\Delta_{k+1}^{\beta+} = 0$ and $\Delta_k^{\beta+} = 1$ is symmetrical.

Finally, we should consider the case $\Delta_{k+1}^{\beta+} = \Delta_k^{\beta+} = 1$. We obtain from (3.3) that

$$\delta^T D_n(\beta,\delta) = \delta^T\left(Z_k^{\beta+} - Z_{k+1}^{\beta+}\right)\left(b_n\left(u_k^\beta\right)(n-k+1) - b_n\left(u_k^\beta + 1\right)(n-k-1)\right).$$

The RHS will be positive whenever $\delta^T(Z_k - Z_{k+1}) \geq 0$ if and only if

$$\begin{aligned} b_n(u+1) &\leq \min_{k>u} b_n(u)\frac{n-k+1}{n-k-1} \\ &= b_n(u)\frac{n-u}{n-u-2}. \end{aligned}$$

□

Normally one would generate the weight by choosing a function b and then defining

$$b_n(i) = \frac{1}{n} b\left(\frac{i}{n}\right)$$

(for the Gehan's statistic $b = 1$).

The condition for monotonicity is then

$$(3.8)\qquad \begin{aligned} 0 \le b_n(i-1)\frac{n-i+1}{n-i-1} - b_n(i) &= \frac{1}{n} b\left(\frac{i-1}{n}\right)\frac{n-i+1}{n-i-1} - \frac{1}{n} b\left(\frac{i}{n}\right) \\ &= \frac{1}{n}\left[\frac{2}{1+i} b\left(\frac{i}{n}\right) - \left(\frac{1}{n}\right) b'\left(\frac{i}{n}\right)\right] + o\left(\frac{1}{n^2}\right). \end{aligned}$$

Taking the limit, this implies

$$(3.9)\qquad 0 \ge \frac{b'(t)}{b(t)} - \frac{2}{1-t} = \frac{d}{dt}\ln\left[(1-t)^2 b(t)\right],$$

that is, $(1-t)^2 b(t)$ should be monotone nonincreasing. It can easily be shown that (3.9) implies (3.8).

Note that, without loss of generality, under the assumption that $\beta_0 = 0$, the weights converge to

$$b\left(H_u(y)\right)\overline{F}(y)\overline{G}(y), \quad \text{where } H_u(y) = P(\Delta = 1, Y \le y).$$

4. A simulation study. In this section we consider the actual behavior of our proposed estimator and compare its efficiency to two other rank estimators: log-rank where the weight function is unity; and Peto–Prentice, where the weight function is the Kaplan–Meier estimator of the survival function of the residuals [Peto (1972) and Prentice (1987)]. Before presenting the simulation results, let us consider theoretically the advantages and disadvantages of these three estimators.

It is well known that the most efficient rank estimator must be a solution of an estimating equation in which the weight function is proportional to the derivative of the log-hazard rate of the distribution of the residual survival times. Since the weight function should not depend on the censoring distribution, it was pointed out to us by one of the referees that our proposed estimator cannot be efficient and the Peto–Prentice estimator may be more efficient. On the other hand, our estimating equation (2.2) provides a unique estimator, while the log-rank and the Peto–Prentice estimating equations may provide more than one estimator for the same data set (see our discussion at the beginning of Section 3).

To examine these issues in practice, we conducted a Monte Carlo experiment. The results reported in Tables 1 and 2 are based on 1000 replications of each experiment. In all the experiments, the covariate Z had a $U(0,1)$ distribution. In the experiments reported in Table 1 the error term had a $N(0,1)$ while the

TABLE 1

Normal error: all entries are for $\beta = 1.5$ and sample size 100. The columns are the censored percentages, the bias, the standard deviation of the Gehan estimator, the probability of β not included in the confidence interval, the average length of the confidence interval and the standard deviation of the Peto–Prentice and log-rank estimators. All estimates are based on 1000 repetitions

CP	$E(\widehat{\beta}_G - \beta)$	$\sigma(\widehat{\beta}_G)$	α	$(E\|I\|)$	$\sigma(\widehat{\beta}_{PP})$	$\sigma(\widehat{\beta}_{LR})$
0.0	−0.0158	0.37	0.111	0.5930	0.36	0.40
18.3	0.0025	0.39	0.1115	0.6317	0.40	0.44
30.2	−0.0224	0.40	0.108	0.6577	0.40	0.44

conditional distribution given Z of the censoring variable was $N(c - 1.2Z, 4)$. The parameter c was used to control the censoring level. Table 2 was constructed with the accelerated time model in mind. The error term had a log–Weibull distribution with parameters 1 and 0.25. The censoring variables were arbitrarily chosen to be equal to $c - 0.8\beta_0 Z + 3\log(\varepsilon)$ where ε is a standard exponential random variable and the constant c was chosen to give a desired probability of censoring.

For each sample the slope was estimated as a generalized solution of (2.2). This was done as follows. The search for the solution began by locating an integer β_1 such that $W_n(\beta_1) \leq 0 \leq W_n(\beta_1 + 1)$, from which we calculated $\widehat{A} = W_n(\beta_1 + 1) - W_n(\beta_1)$. We then used (2.5)–(2.7) with $\alpha = 10\%$ to construct a confidence interval for the slope. The actual probability coverage and the average length of the confidence set are reported in the tables.

The results in Tables 1 and 2 are representative of all the simulations we conducted. The following is a summary of our findings. When the distribution of the error term is normal, all three estimators achieve their large-sample properties in samples of moderate size. In particular, our estimator has a small relative bias and its efficiency is compatible with the other two rank estimators. However, when the error term has a log–Weibull distribution, all three estimators are much less efficient. This is noticeable in the increase of their standard deviations to a point where the null hypothesis H_0: $\beta = 0$ cannot be rejected even when $\beta = 1.5$. The relative bias of our estimator is still small but its standard deviation is somewhat (not significantly) larger than that of the two other rank estimators.

Finally, we consider the question of how often the Peto–Prentice and the log-rank estimating equations do not provide a unique solution. We plotted the estimating equation of all three rank procedures for the data that generated Tables 1 and 2. While only equation (2.2) was monotone, the other estimating equations did provide only one solution. We tried other simulations and found an example (see Figure 3) in which the log-rank method provided more than one estimator; however, the Peto–Prentice method continued to provide only one solution.

We conclude that when the slope is a scalar our estimator is not significantly less efficient than the other two rank methods and that the other two methods do

TABLE 2

Log-Weibull error: The columns are the true slope, the sample size, the censored percentages (CP), the bias, the standard deviation of the Gehan estimator, the probability of β not included in the confidence interval, and the standard deviation of the Peto–Prentice and log-rank estimators. All estimates are based on 1000 *repetitions*

β	n	CP	$E(\hat{\beta}_G - \beta)$	$\sigma(\hat{\beta}_G)$	α	$(E\|I\|)$	$\sigma(\hat{\beta}_{PP})$	$\sigma(\hat{\beta}_{LR})$
1.50	100	0.1	0.0985	1.60	0.114	2.6558	1.49	1.45
		18.1	−0.0552	1.73	0.112	2.9174	1.59	1.55
		49.3	0.1918	2.31	0.101	3.7089	2.11	2.07
	200	0.1	0.0507	1.20	0.126	1.8724	1.11	1.05
		18.3	−0.0228	1.27	0.107	2.0591	1.17	1.13
		49.1	0.1236	1.59	0.105	2.5845	1.45	1.41
15.0	100	2.2	0.0441	1.66	0.117	2.6822	1.52	1.43
		35.8	0.0179	2.53	0.112	4.1920	2.33	2.27
		53.4	0.4439	3.70	0.128	5.9180	3.33	3.24
	200	2.2	0.0559	1.17	0.103	1.8972	1.08	1.03
		36.0	0.0117	1.86	0.106	2.9660	1.69	1.63
		53.3	0.1764	2.58	0.128	4.1530	2.33	2.27

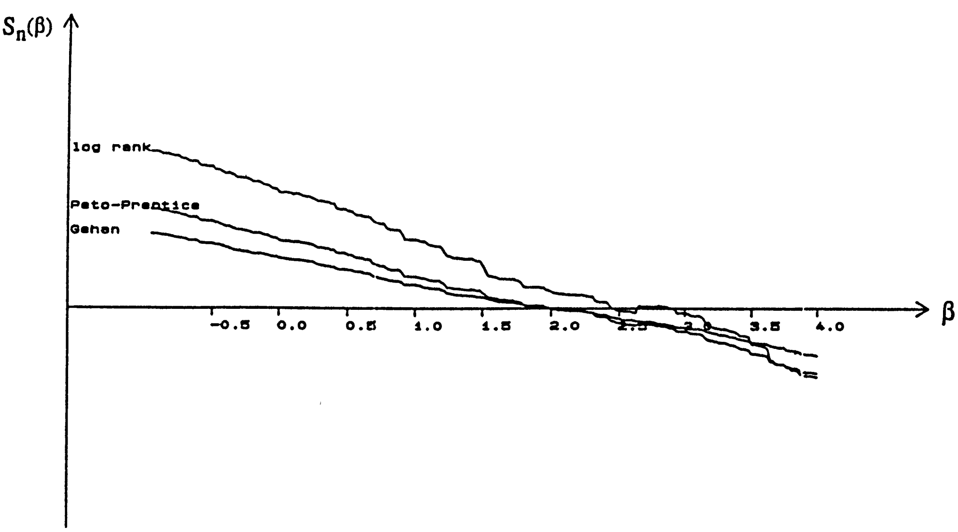

FIG. 3. *Plots of the estimating equations: sample size* 100; *uncensored observations* 58; *independent variable* $U(0,1)$; *slope* 1.5; *error term log of a Weibull*(1,4.0); *censoring* $0.0 + 0.8(1-z) + e$, *where e, is the log of Weibull*(1,3.0).

not, in general, provide more than one solution. However, in practice, especially when one estimates the parameters in a multiple regression, our estimators should be derived first since they are easier and faster to obtain and can be used as a starting point for other rank estimators. We stress that in higher dimensions the guaranteed uniqueness of our method is of ample importance as it is not known how often other rank methods provide a unique solution.

Acknowledgment. The first author thanks Butch Tsiatis for insightful discussions.

REFERENCES

ANDERSEN, P. K. (1982). Testing goodness of fit of Cox's regression and life model. *Biometrics* **38** 67–77.

BICKEL, P. J., KLAASSEN, C. A. J., RITOV, Y. and WELLNER, J. A. (1991). *Efficient and Adaptive Estimation for Semiparametric Models*. Johns Hopkins Univ. Press.

BROWN, B. M. (1985). Multiparameter linearization theorems. *J. Roy Statist. Soc. Ser. B* **47** 325–331.

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

EFRON, B. (1967). The two sample problem with censored data. *Proc. Fifth Berkeley Symp. Math. Statist. Probab.* **4** 831–853. Univ. California Press.

FYGENSON, M. and ZHOU, M. (1994). On using stratification in the analysis of linear regression models with right censoring. *Ann. Statist.* **22** 747–762.

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

HARRINGTON, D. P. and FLEMING, T. R. (1982). A class of rank test procedures for censored survival data. *Biometrika* **69** 553–556.

KALBFLEISH, J. D. and PRENTICE, R. L. (1973). Marginal likelihoods based on Cox's regression and life model. *Biometrika* **60** 267–278.

KOUL, H., SUSARLA, V. and VAN RYZIN, J. (1981). Regression analysis with randomly right censored data. *Ann. Statist.* **9** 1276–1288.

LEURGANS, S. (1987). Linear models, random censoring and synthetic data. *Biometrika* **74** 301–309.

MILLER, R. G. (1976). Least squares regression with censored data. *Biometrika* **63** 449–464.

PETO, R. (1972). Rank tests of maximal power against Lehmann-type alternatives. *Biometrika* **59** 472–475.

POWELL, L. L. (1984). Least absolute deviation estimation for censored regression model. *J. Econometrics* **25** 303–325.

PRENTICE, R. L. (1978). Linear rank tests with right censored data. *Biometrika* **65** 167–180.

RITOV, Y. (1987). Tightness of monotone random fields. *J. Roy. Statist. Soc. Ser. B* **49** 331–333.

RITOV, Y. (1990). Estimation in a linear regression model with censored data. *Ann. Statist.* **18** 303–328.

RITOV, Y. and WELLNER, J. A. (1988). Censoring, martingales and the Cox model. *Contemp. Math.* **80** 191–220.

SCHOENFELD, D. (1980). Chi-squared goodness of fit tests for the proportional hazards regression model. *Biometrika* **67** 145–153.

SEN, P. K. (1968). Estimates of the regression coefficient based on Kendall's tau. *J. Amer. Statist. Assoc.* **63** 1379–1389.

SERFLING, R. J. (1980). *Approximation Theorems of Mathematical Statistics.* Wiley, New York.

STORER, B. E. and CROWLEY, L. (1985). A diagnostic for Cox regression and general conditional likelihood. *J. Amer. Statist. Assoc.* **80** 139–147.

Therneau, T. M., Grambsch, P. M. and Fleming, T. R. (1990). Martingale-based residuals for survival models. *Biometrika* **77** 147–160.

Tsiatis, A. (1990). Estimating regression parameters using linear rank tests for censored data. *Ann. Statist.* **18** 354–372.

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

Department of Statistics
Busch Campus
Rutgers University
New Brunswick, New Jersey 08903

Department of Statistics
Faculty of Social Sciences
Hebrew University
Jerusalem 91905
Israel