The Annals of Statistics
1990, Vol. 18, No. 1, 303–328

ESTIMATION IN A LINEAR REGRESSION MODEL WITH CENSORED DATA[1]

BY Y. RITOV

The Hebrew University of Jerusalem

We consider the semiparametric linear regression model with censored data and with unknown error distribution. We describe estimation equations of the Buckley–James type that admit $\sqrt{n}$-consistent and asymptotically normal solutions. The derived estimator is efficient at a particular error distribution. We show the equivalence between this type of estimator and an estimator based on a linear rank test suggested by Tsiatis. This equivalence is an extension of a basic equivalence between Doob type martingales and counting process martingales shown by Ritov and Wellner. An extension to an estimator that is efficient everywhere is discussed.

1. Introduction. Let $(Y, Z, C) \in R \times R^m \times R$ and suppose we observe $X = (Y \wedge C, Z, \Delta)$, where $\Delta = 1_{\{Y \le C\}}$. Let μ be a product of the Lebesgue measure on the real line with some measure of R^{m+1}. We assume that the joint density of (Y, Z, C) relative to μ is given by $f_0(Y - \beta_0^T Z)h(Z, C)$. Let F_0 be the distribution function with density f_0. That is, $\varepsilon \equiv Y - \beta_0^T Z$ is distributed according to F_0 and is independent of Z and C. The latter have an arbitrary joint distribution.

This model can arise in a regular regression situation if the measuring device fails to give a true measurement above a given level. In survival analysis this model may be used to describe the log of the failure time in the "accelerated time model," see Kalbfleisch and Prentice (1981) and Lawless (1982). There it is assumed that T, the survival time, follows a conditional scale model

$$\tilde{f}_0(t|z) = e^{-\beta_0^T z}\tilde{f}_0(e^{-\beta_0^T z}t). \tag{1.1}$$

If $Y = \log(T)$ and $f_0(y) = e^y \tilde{f}_0(e^y)$, then (1.1) is identical to the linear regression model.

In this paper we address the problem of estimating β_0 when we are given a random sample $X_1, \dots, X_n$ from X with f_0 and h unknown. Miller (1976), Buckley and James (1979) and Koul, Susarla and Van Rayzin (1981) suggested estimators based on different modifications of the ordinary least squares method. Hopkins (1984) suggested an adaptive estimator of β and made some simulation experiments to compare his estimator with other estimators. Recently, Tsiatis (1990), building on some earlier work of Louis (1981) and Wei and Gail (1983), suggested a family of estimators based on linear rank tests for the slope.

Received September 1987; revised February 1989.

[1]Research supported in part by the Office of Naval Research, contract N00014-80-C-0136. Part of the research was done while the author was at the University of California at Berkeley and at the University of Pennsylvania.

AMS 1980 *subject classifications*. 62G20, 62G05.

Key words and phrases. Buckley–James estimator, counting process, Kaplan–Meier estimator, martingales.

We suggest an estimator that is a close relative of the Buckley–James estimator. It is a modification of the M-estimator for regression, except that the unobserved Y values are replaced by their (estimated) conditional expectations. Somewhat surprisingly, the estimator resembles quite closely the Tsiatis estimator, although the motivation for the score function of the Buckley–James estimator and the weights of the ranks in the Tsiatis construction are different.

The theoretical background for this resemblance was given in Ritov and Wellner (1988). In this paper the authors investigate a link between counting process martingales and Doob martingales (conditional expectation martingales) that arise via censoring. The same type of link is the base for the equivalence between the two seemingly unrelated estimators.

Ritov (1984, 1986) gave a detailed construction of a $\sqrt{n}$-consistent estimator under the minimal condition of nonzero information and described the construction of efficient estimators. We do not pursue these topics here. Some comments on them are given in Section 6.

The paper is organized as follows. In Section 2 the Buckley–James type and Tsiatis estimators are described. In Section 3 we investigate the relations between the estimators and exhibit martingale representations of them. Then, in Section 4, their asymptotic equivalence is proved. The asymptotic properties of the Buckley–James estimators are investigated further in Section 5. Section 6 is devoted to some additional comments on the estimators.

Martingale techniques are used extensively in the proofs, showing the importance of this device for censored data models, even for those models that are not hazard function oriented: Some four different martingales and three families of σ-fields are used in this paper.

Some parts of the proofs, those that are either short or were found by us to be interesting, are given in Sections 3 to 5. The remainder of the proofs are given in Section 7.

2. Two classes of estimators of the slope. Let $X = (Y \wedge C, Z, 1_{\{Y \le C\}})$ and let X_i, $i = 1, 2, \dots, n$, be a random sample from X. Assume first that $f_0(\cdot)$ is known. Then the maximum likelihood estimator $\hat{\beta}$ of the slope is a solution of

$$0 = \sum_{i=1}^{n} Z_i \left\{ \Delta_i \frac{f_0'}{f_0}\left(Y_i - \hat{\beta}^T Z_i\right) - (1 - \Delta_i) \frac{f_0}{1 - F_0}\left(C_i - \hat{\beta}^T Z_i\right) \right\}.$$

More generally, we can replace $-f_0'/f_0$ by any reasonable score function $s(\cdot)$ and get an M-estimator that is a solution of the following set of equations:

$$0 = \sum_{i=1}^{n} Z_i \left[\Delta_i s\left(Y_i - \hat{\beta}^T Z_i\right) + (1 - \Delta_i) \frac{\int_{C_i - \hat{\beta}^T Z_i}^{\infty} s(t)\, dF_0(t)}{1 - F_0\left(C_i - \hat{\beta}^T Z_i\right)} \right].$$

Normally, f_0 is known only up to a shift, or equivalently, there is an intercept term in the regression equation. In that case the above estimating equations do not have mean zero when we plug in the correct slope with misspecified interception. The simple remedy is centering the Z_i's:

$$(2.1)\quad 0 = \sum_{i=1}^{n} (Z_i - \bar{Z})\left[\Delta_i s(Y_i - \hat{\beta}^T Z_i) + (1 - \Delta_i)\frac{\int_{C_i - \hat{\beta}^T Z_i}^{\infty} s(t)\, dF_0(t)}{1 - F_0(C_i - \hat{\beta}^T Z_i)}\right],$$

where $\bar{Z} = n^{-1}\sum_{i=1}^{n} Z_i$.

If $s(t) \equiv t$, then (2.1) defines a natural generalization of the least squares estimators to the censored model. A standard approximation of (2.1) is given by the one-step Newton–Raphson approximation: Begin with some auxiliary estimator, $\tilde{\beta}$, which is $\sqrt{n}$-consistent, and construct

$$(2.2)\quad \begin{aligned}\hat{\beta} = \tilde{\beta} - \hat{K}_n^{-1} n^{-1} \sum_{i=1}^{n} (Z_i - \bar{Z})\Bigg[&\Delta_i s(Y_i - \tilde{\beta}^T Z_i) \\ &+ (1 - \Delta_i)\frac{\int_{C_i - \tilde{\beta}^T Z_i}^{\infty} s(t)\, dF_0(t)}{1 - F_0(C_i - \tilde{\beta}^T Z_i)}\Bigg],\end{aligned}$$

where $\hat{K}_n^{-1}$ is some estimate of the derivative (with respect to β) of the r.h.s. of (2.1).

It is well known that, without censoring and under some mild conditions, the estimator defined by either (2.1) or (2.2) is well behaved, whether the error distribution is known or not, that is, $\sqrt{n}(\hat{\beta} - \beta)$ converges in law to a normal distribution with mean 0. The situation is changed drastically when we permit some censored observations. Neither (2.1) nor (2.2) defines an asymptotically unbiased estimator.

A natural extension of (2.1) to cope with unknown F_0 was suggested by Buckley and James (1979). The idea is that the unknown F_0 in (2.1) should be replaced by its generalized maximum likelihood estimator, namely the Kaplan–Meier (KM) estimator based on the residuals (assuming that the estimated slope is the true value). Let $\mathbb{F}_n^{\beta}$ be the KM estimator, assuming β is the true slope. We consider therefore an estimator that is an approximate solution of $\Psi_n(\beta; s) = 0$, where

$$(2.3)\quad \begin{aligned}\Psi_n(\beta; s) &= n^{-1/2} \sum_{i=1}^{n} (Z_i - \bar{Z}) \\ &\quad\times\left[\Delta_i s(Y_i - \beta^T Z_i) + (1 - \Delta_i)\frac{\int_{(C_i - \beta^T Z_i, \infty)} s(t)\, d\mathbb{F}_n^{\beta}(t)}{1 - \mathbb{F}_n^{\beta}(C_i - \beta^T Z_i)}\right] \\ &= n^{-1/2} \sum_{i=1}^{n} (Z_i - \bar{Z})\left[\Delta_i s(\varepsilon_i^{\beta}) + (1 - \Delta_i)\frac{\int_{(\zeta_i^{\beta}, \infty)} s(t)\, d\mathbb{F}_n^{\beta}(t)}{1 - \mathbb{F}_n^{\beta}(\zeta_i^{\beta})}\right],\end{aligned}$$

$\varepsilon_i^{\beta} \equiv Y_i - \beta^T Z_i$ and $\zeta_i^{\beta} \equiv C_i - \beta^T Z_i$. Note that the KM estimator is based on $\{(\varepsilon_i^{\beta} \wedge \zeta_i^{\beta}, \Delta_i),\ i = 1, \dots, n\}$. Hence it is itself a function of β. Moreover, due to this dependence, $\Psi_n(\cdot\,; s)$ is not monotone and it is not continuous, even when $s(\cdot)$ is. As a result, $\Psi_n(\beta; s) = 0$ may fail to have a solution.

Buckley and James (1979) do not give a proper theoretical justification for their suggestion nor a full analysis of the behavior of the estimator. A partial result concerning its consistency is given in James and Smith (1984). However, it is not clear from their analysis, for example, whether it is important to use the KM estimator in the estimate of the conditional distribution or whether any other reasonable estimator can be used. It will be shown in the following discussion that the properties of the KM estimator are essential for the analysis. Actually we believe that any other (nonequivalent) estimator will not yield a regular estimator.

Motivated by linear rank tests of the slope, Tsiatis (1990) suggested a family of estimators. Let

$$\Gamma_n(\beta; w) = n^{-1/2} \sum_{i=1}^{n} \Delta_i \left[Z_i - \frac{\sum_{j=1}^{n} Z_j 1_{\{\varepsilon_j^\beta \wedge \zeta_j^\beta \geq u\}}}{\sum_{j=1}^{n} 1_{\{\varepsilon_j^\beta \wedge \zeta_j^\beta \geq u\}}} \right] w(\varepsilon_i^\beta), \tag{2.4}$$

for some weight function w. $\Gamma_n(\beta; w)$ may serve as a test statistic for β. Consequently, a solution of $\Gamma_n(\beta; w) = o_p(1)$ can be used as an estimator of the slope.

The reader used to the counting process point of view may find it convenient to rewrite (2.4) as

$$\Gamma_n(\beta; w) = n^{-1/2} \sum_{i=1}^{n} \int_{(-\infty, \zeta_i^\beta]} \left[Z_i - \frac{\sum_{j=1}^{n} Z_j 1_{\{\varepsilon_j^\beta \wedge \zeta_j^\beta \geq u\}}}{\sum_{j=1}^{n} 1_{\{\varepsilon_j^\beta \wedge \zeta_j^\beta \geq u\}}} \right] w(u)\, dN_i^\beta(u),$$

where $N_i^\beta(u) = 1_{\{\varepsilon_i^\beta \leq u\}}$. We now add a term that is identically 0 to obtain

$$\begin{aligned}\Gamma_n(\beta; w) = n^{-1/2} \sum_{i=1}^{n} \int_{(-\infty, \zeta_i^\beta]} &\left[Z_i - \frac{\sum_{j=1}^{n} Z_j 1_{\{\varepsilon_j^\beta \wedge \zeta_j^\beta \geq u\}}}{\sum_{j=1}^{n} 1_{\{\varepsilon_j^\beta \wedge \zeta_j^\beta \geq u\}}} \right] w(u) \\ &\times \left\{ dN_i^\beta(u) - 1_{\{\varepsilon_i^\beta \geq u\}}\, d\Lambda(u) \right\},\end{aligned} \tag{2.5}$$

for *any* "hazard function" $\Lambda(\cdot)$.

In the next sections we show that for any $s(\cdot)$ there is a $w(\cdot)$ that is a function of $s(\cdot)$ and $\mathbb{F}_n^\beta$ only, such that $\Gamma_n(\beta; w) = \Psi_n(\beta; s)$. Moreover, for any $s(\cdot)$ there is a $w(\cdot)$ that is a function of $s(\cdot)$ and F_0 only such that if $\beta_n = \beta_0 + O_p(n^{-1/2})$, then $\Psi_n(\beta_n; s) = \Gamma_n(\beta_n; w) + o_p(1)$.

We conclude this section with a list of the major assumptions under which $\Psi_n(\beta; s)$ will be analyzed.

ASSUMPTIONS.

(A.1) Let c_0 be such that $\Pr(Y \wedge C - \beta^T Z < c_0) < 1$ for all β in some neighborhood of β_0. Then $s(\cdot)$ belongs to a family $\mathscr{S}_0 \subseteq L_2(F_0)$, where any $s \in \mathscr{S}_0$ satisfies:

$$\lim_{\nu \to 0} \int \sup\{|s(t+\xi) - s(t)|^2 : |\xi| \leq \nu\}\, dF_0(t) = 0. \tag{i}$$

$$s(t) = s(t \wedge c_0). \tag{ii}$$

(A.2) Z has compact support.

(A.3) F_0 has finite Fisher information for location.

REMARK. Assumption (A.1)(ii) is essentially equivalent to the more standard demand in the survival analysis literature for considering only intervals that are bounded from the right.

3. Martingale representation of $\Gamma_n(\beta; w)$ and $\Psi_n(\beta; s)$ and their correspondence. Consider

$$\mathbb{Q}_s(t; \varepsilon|F) \equiv s(\varepsilon)1_{\{\varepsilon \le t\}} + \frac{\int_{(t,\infty)} s(u)\, dF(u)}{1 - F(t)} 1_{\{\varepsilon > t\}}.$$

Let $\mathscr{G} = \{\mathscr{G}_t\}$ be a family of increasing σ-fields, where $\mathscr{G}_t$ is the minimal σ-field such that $(\varepsilon \wedge t, 1_{\{\varepsilon \le t\}})$ is $\mathscr{G}_t$-measurable. If ε is a random variable distributed according to F, then $\mathbb{Q}_s(\cdot; \varepsilon|F) = E_F(s(\varepsilon)|\mathscr{G}_t)$, and hence it is a martingale with respect to $\mathscr{G}$. Note that $\mathbb{Q}_s(C - \beta^T Z; Y - \beta^T Z|\mathbb{F}_n^\beta)$ is the "building block" of $\Psi_n(\beta; s)$ with

$$\Psi_n(\beta; s) = n^{-1/2} \sum_{i=1}^n (Z_i - \bar{Z})\mathbb{Q}_s(\zeta_i^\beta; \varepsilon_i^\beta|\mathbb{F}_n^\beta). \tag{3.1}$$

On the other hand, define the counting process martingale

$$\mathbb{M}(t; \varepsilon|F) = 1_{\{\varepsilon \le t\}} - \int_{(-\infty, t]} 1_{\{\varepsilon \ge u\}} \frac{dF(u)}{1 - F(u-)}.$$

It is a martingale under the same assumptions as $\mathbb{Q}_s(t; \varepsilon|F)$. Considering (2.5), it can be seen that $\int w(u)\, d\mathbb{M}(u)$ is the "building block" of $\Gamma_n(\beta; w)$:

$$\begin{aligned} &\Gamma_n(\beta; w) \\ &= n^{-1/2} \sum_{i=1}^n \int_{(-\infty, \zeta_i^\beta]} \left[Z_i - \frac{\sum_{j=1}^n Z_j 1_{\{\varepsilon_j^\beta \wedge \zeta_j^\beta \ge u\}}}{\sum_{j=1}^n 1_{\{\varepsilon_j^\beta \wedge \zeta_j^\beta \ge u\}}} \right] w(u)\, d\mathbb{M}(u; \varepsilon_i^\beta|F), \end{aligned} \tag{3.2}$$

for any distribution function F.

We next explore the connections between $\mathbb{Q}_s(\cdot; \varepsilon|F)$ and $\mathbb{M}(\cdot; \varepsilon|F)$. For any distribution function F on the real line we define the transformations W_F and S_F by

$$W_F s(t) \equiv s(t) - \frac{\int_{(t,\infty)} s(u)\, dF(u)}{1 - F(t)}, \qquad s \in L_2(F)$$

(with $0/0 = 0$) and

$$S_F w(t) \equiv w(t) - \int_{(-\infty, t]} w(u)[1 - F(u-)]^{-1} dF(u), \qquad w \in L_2(F).$$

These transformations were investigated by Ritov and Wellner (1988) for F continuous (in this paper W_F was called R and S_F was called L). See Efron and Johnstone (1990) for an independent discussion. If F is continuous, then W_F: $L_2(F) \to L_2(F)$ is the adjoint of S_F: $L_2(F) \to L_2(F)$ and both have norm 1. In the noncontinuous case the situation is not so simple. We will be satisfied with the following result.

Lemma 3.1. *Let $S = \{s\colon s \in L_2(F)$ and $s(t) \equiv s(t \wedge c_0)$ for $t \in (-\infty, \infty)\}$ and $W \equiv \{w\colon L_2(F)$ and $w(t) = 0$ for $t \geq c_0\}$, where $F(c_0) < 1$. Then*

(i) $W_F\colon S \to W$ and $S_F\colon W \to S$ are bounded linear operators. If F is continuous, then their norm is 1.
(ii) $W_F S_F w = w$ and $S_F W_F s = s - \int s(u)\, dF(u)$.

Thus $W_F = S_F^{-1}$ on $\{s\colon s \in L_2(F)$ and $\int s\, dF = 0\}$.

Proof. That $S_F\colon L_2(F) \to L_2(F)$ is a bounded linear operator with norm not greater than 1 follows from the first proof of Ritov and Wellner (1988), Proposition 2.1(i). It is easy to verify that for $s \in S$ the norm of the second term in the definition of $W_F s$ is at most $(1 - F(c_0))^{-1/2}\|s\|$. Part (ii) follows now by a simple application of Fubini's theorem. □

The martingales $\mathbb{Q}_s$ and $\mathbb{M}$ are connected via Proposition 3.1.

Proposition 3.1. *Suppose that $s \in L_2(F)$ and $\int s(t)\, dF(t) = 0$. Then the martingale $\mathbb{Q}_s(\cdot; \varepsilon|F)$ is related to the counting process martingale $\mathbb{M}(\cdot; \varepsilon|F)$ by*

$$\mathbb{Q}_S(t; \varepsilon|F) = \int_{(-\infty, t]} W_F s(u)\, d\mathbb{M}(u; \varepsilon|F), \qquad t \in (-\infty, \infty).$$

Consequently, for any $L_2(F)$ functions $s(\cdot)$ and $D(\cdot)$ with $\int s\, dF = 0$,

$$\int_{(-\infty, t]} D(u)\, d\mathbb{Q}_s(u; \varepsilon|F) = \int_{(-\infty, t]} D(u) w(u)\, d\mathbb{M}(u; \varepsilon|F),$$

where $w = W_F s$ (or $s = S_F w$).

Note that although the proposition was expressed in probability terms, it is actually just an equality between two real variables, ε and t, a function $s(\cdot)$ and a distribution function F. If ε is a random variable distributed according to F, we get an identity between two martingales. Proposition 3.1 will be used to establish the connection between the two classes of estimators. The proposition was proved in Ritov and Wellner (1988) for a continuous distribution function F. We give here its proof for completeness.

Proof. First note that

$$\begin{aligned}
\int_{(-\infty, \xi]} W_F s(u) \frac{dF(u)}{1 - F(u-)} &= \int_{(-\infty, \xi]} s(u) \frac{dF(u)}{1 - F(u-)} \\
&\quad - \int_{(-\infty, \xi]} \frac{\int_{(\tau, \infty)} s(u)\, dF(u)}{1 - F(\tau)} \frac{dF(\tau)}{1 - F(\tau-)} \\
&= \int_{(-\infty, \xi]} s(u) \frac{dF(u)}{1 - F(u-)} \\
&\quad - \iint \frac{s(u) 1_{\{\tau \leq \xi\}} 1_{\{\tau < u\}}}{(1 - F(\tau))(1 - F(\tau-))}\, dF(u)\, dF(\tau).
\end{aligned}$$

Apply Fubini's theorem to the second integral on the r.h.s. to obtain

$$\int_{(-\infty,\xi]} W_F s(u)\frac{dF(u)}{1-F(u-)} = -\int_{(\xi,\infty)} s(u)\frac{dF(u)}{1-F(\xi)} = -\frac{\int_{(\xi,\infty)} s(u)\,dF(u)}{1-F(\xi)}.$$

Hence

$$\int_{(\infty,t]} W_F s(u)\, d\mathbb{M}(u;\varepsilon|F)$$
$$= s(\varepsilon)1_{\{\varepsilon\le t\}} - \frac{\int_{(\varepsilon,\infty)} s(u)\,dF(u)}{1-F(\varepsilon)}1_{\{\varepsilon\le t\}} - \int_{(-\infty,t\wedge\varepsilon]} W_F s(u)\frac{dF(u)}{1-F(u-)}$$
$$= s(\varepsilon)1_{\{\varepsilon\le t\}} + \frac{\int_{(t,\infty)} s(u)\,dF(u)}{1-F(t)}1_{\{\varepsilon> t\}}$$
$$= \mathbb{Q}_s(t;\varepsilon|F).$$ □

This proposition enables us to establish the following result.

PROPOSITION 3.2. *For all $\beta\in R^m$ and s, w: $R\to R$ and* w.p.1.,

(i) $$\Psi_n(\beta;s) = n^{-1/2}\sum_{i=1}^{n}(Z_i-\bar{Z})\int_{(-\infty,\zeta_i^\beta]} W_{\mathbb{F}_n^\beta}s(u)\,d\mathbb{M}(u;\varepsilon_i^\beta|\mathbb{F}_n^\beta) = \Gamma_n(\beta;W_{\mathbb{F}_n^\beta}s).$$

(ii) $$\Gamma_n(\beta;w) = \Psi_n(\beta;S_{\mathbb{F}_n^\beta}w).$$

PROOF. By Lemma 3.1 and the fact that the Z_i are centered it is enough to prove the first part. [The centering of the Z_i's makes $\Psi_n(\beta;s)\equiv\Psi_n(\beta;s+c)$ for any constant c. In particular, we can assume that $\int s\,dF=0$.] It follows from (3.1) and Proposition 3.1 that

$$\Psi_n(\beta;s) = n^{-1/2}\sum_{i=1}^{n}(Z_i-\bar{Z})\int_{(-\infty,\zeta_i^\beta]} W_{\mathbb{F}_n^\beta}s(u)\,d\mathbb{M}(u;\varepsilon_i^\beta|\mathbb{F}_n^\beta). \tag{3.3}$$

$\mathbb{F}_n^\beta$ has atoms only on uncensored observations and

$$\frac{d\mathbb{F}_n^\beta(u)}{1-\mathbb{F}_n^\beta(u-)} = \left[\sum_{j=1}^{n}1_{\{\varepsilon_j^\beta\wedge\zeta_j^\beta\ge u\}}\right]^{-1}\sum_{i=1}^{n}\Delta_i 1_{\{u=\varepsilon_i^\beta\}}.$$

Hence, for any function h,

$$\sum_{i=1}^{n}\int_{(-\infty,\zeta_i^\beta]} h(u)\,d\mathbb{M}(u;\varepsilon_i^\beta|\mathbb{F}_n^\beta)$$
$$= \sum_{i=1}^{N}\Delta_i h(\varepsilon_i^\beta) - \int\sum_{i=1}^{n}1_{\{\varepsilon_i^\beta\wedge\zeta_i^\beta\ge u\}}h(u)\frac{d\mathbb{F}^{\beta_n}(u)}{1-\mathbb{F}_n^\beta(u-)} \tag{3.4}$$
$$= 0 \quad \text{w.p.1.}$$

Now combine (3.2), (3.3) and (3.4) with

$$h(u) = \frac{\sum_{j=1}^{n}\left(Z_j - \bar{Z}\right)1_{\{\varepsilon_j^{\beta} \wedge \zeta_j^{\beta} \geq u\}}}{\sum_{j=1}^{n} 1_{\{\varepsilon_j^{\beta} \wedge \zeta_j^{\beta} \geq u\}}} W_{\mathbb{F}_n^{\beta}} s(u)$$

to establish the first part. The second part is immediate by Lemma 3.1. □

4. Asymptotic equivalence of the two estimators. In view of Proposition 3.2, we see that the Buckley–James family of estimating equations is equivalent to the Tsiatis family except that in one of the sets of equations there is a random function. In this section we show that $\Psi_n(\beta_n; s) = \Gamma_n(\beta_n; W_{F_0}s) + o_p(1)$, where $\{\beta_n\}$ is nonstochastic and $\limsup n^{1/2}\|\beta_n - \beta\| < \infty$. The tightness of $\Psi_n(\cdot\,; s)$ is proved in Lemma 5.2, and see Tsiatis (1987) for the tightness of $\Gamma_n(\cdot\,; w)$ (the tightness of the latter is not proved by us), so the result can be extended to stochastic sequences $\{\beta_n\}$ as well. We begin by claiming that $W_{\mathbb{F}_n^{\beta_n}}s$ converges to $W_{F_0}s$.

LEMMA 4.1. *Suppose $s \in \mathscr{S}_0$ and $\{\beta_n\}$ is a nonstochastic sequence such that $\limsup\|\beta_n - \beta_0\| < \infty$. Then*

$$\left\| W_{\mathbb{F}_n^{\beta_n}} s - W_{F_0} s \right\|_{\infty} = O_p(n^{-1/2}) \quad \text{as } n \to \infty.$$

The lemma follows from Lemmas 7.1(ii) and 7.2(iii).

PROPOSITION 4.1. *Suppose the conclusion of Lemma* 4.1 *holds. Then*

$$\Psi_n(\beta_n; s) = \Gamma_n\left(\beta_n; W_{F_0}s\right) + o_p(1).$$

PROOF. By (3.1), (3.2) and Proposition 3.2 it is enough to prove that

$$n^{-1/2}\sum_{i=1}^{n} \int Z_i h(u) 1_{\{\zeta_i^{\beta_n} \geq u\}}\, d\mathbb{M}\left(u; \varepsilon_i^{\beta_n} | \mathbb{F}_n^{\beta_n}\right) \to_p 0, \tag{4.1}$$

where $h = W_{\mathbb{F}_n^{\beta_n}}s - W_{F_0}s$. For the proof of (4.1) we consider the process

$$L^{\beta_n}(t) = n^{-1/2} \int_{(t,\infty)} \sum_{i=1}^{n} h(u) 1_{\{\zeta_i^{\beta_n} \geq u\}}\left\{Z_i\, d\mathbb{M}\left(u; \varepsilon_i^{\beta_n} | \mathbb{F}_n^{\beta_n}\right) - \bar{D}^{\beta_n}(u)\, dN_i^{\beta_n}(u)\right\},$$

where

$$\bar{D}^{\beta_n}(u) = \frac{R^{\beta_n}(u+)}{R^{\beta_n}(u)} E\left(Z | Y - \beta_n^T Z = u, \Delta = 1\right) - \frac{1}{R^{\beta_n}(u+)} \sum_{i=1}^{n} Z_i 1_{\{\zeta_i^{\beta_n} \wedge \varepsilon_i^{\beta_n} > u\}}$$

and

$$R^{\beta_n}(u) = \sum_{i=1}^{n} 1_{\{\zeta_i^{\beta_n} \wedge \varepsilon_i^{\beta_n} \geq u\}}.$$

Note that as $R^{\beta_n}(u) = O_p(n)$ for all n values of interest, and

$$E\left\{Z | Y - \beta_0^T Z = u, \Delta = 1\right\} = E\left\{Z | C - \beta_0^T Z \geq u\right\} = E\left\{Z | Y \wedge C - \beta_0^T Z \geq u\right\}.$$

Thus $\bar{D}^{\beta_n}(u)$ is essentially the difference between a mean and its estimate. In Lemma 7.3(ii) it is shown that $\bar{D}^{\beta_n}(\varepsilon_1^{\beta_n})1_{\{\varepsilon_1^{\beta_n}\le c_0\}} \to_p 0$. Since $|\bar{D}^{\beta_n}(\cdot)| \le 2\,\mathrm{ess\,sup}\|Z\|$, we find that it converges in the mean, and hence

$$E\left\{n^{-1}\sum_{i=1}^{n}|\bar{D}^{\beta_n}(\varepsilon_i^{\beta_n})|1_{\{\varepsilon_i^{\beta_n}\le c_0\}}\right\} \to 0.$$

In particular,

$$n^{-1}\sum_{i=1}^{n}|\bar{D}^{\beta_n}(\varepsilon_i^{\beta_n})|1_{\{\varepsilon_i^{\beta_n}\le c_0\}} \to_p 0.$$

Now use this result together with Lemma 4.1 to show that $L^{\beta_n}(-\infty)$ differs from the l.h.s. of (4.1) by

$$\begin{aligned}\left|n^{-1/2}\int\sum_{i=1}^{n}h(u)\bar{D}^{\beta_n}(u)\,dN_i^{\beta_n}(u)\right| &= \left|n^{-1/2}\sum_{i=1}^{n}\Delta_i h(\varepsilon_i^{\beta_n})\bar{D}^{\beta_n}(\varepsilon_i^{\beta_n})\right|\\ &\le \left(n^{1/2}\|h\|_\infty\right)\left(n^{-1}\sum_{i=1}^{n}\Delta_i|\bar{D}^{\beta_n}(\varepsilon_i^{\beta_n})|1_{\{\varepsilon_i^{\beta_n}\le c_0\}}\right)\\ &\to_p 0.\end{aligned}$$

Hence $L^{\beta_n}(-\infty)$ is an approximation to the l.h.s. of (4.1), and it is enough to prove that

$$L^{\beta_n}(-\infty) \to_p 0. \tag{4.2}$$

We will consider this process together with the family $\mathscr{F}^{\beta_n} = \{\mathscr{F}_t^{\beta_n}\}$ of decreasing σ-fields where $\mathscr{F}$ stands for future observations and

$$\mathscr{F}_t^{\beta_n} = \sigma\left\{\left[1_{\{\varepsilon_i^{\beta_n}\wedge\zeta_i^{\beta_n}\ge t\}}\left(Z_i, \varepsilon_i^{\beta}\wedge\zeta_i^{\beta}, \Delta_i\right), 1_{\{\varepsilon_i^{\beta_n}\wedge\zeta_i^{\beta_n}\ge t\}}\right] : i = 1,\dots,n\right\}.$$

(Given $\mathscr{F}_t^{\beta_n}$, we know all X_i with $\varepsilon_i^{\beta_n}\wedge\zeta_i^{\beta_n}\ge t$ and only the indices of those X_i with $\varepsilon_i^{\beta_n}\wedge\zeta_i^{\beta_n} < t$.)

Note that $W_{\mathbb{F}_n^{\beta_n}}s(t)$ depends on the data through $\{\mathbb{F}_n^{\beta_n}(u): u > t\}$, which is a function only of those observations X_i with $\varepsilon_i^{\beta}\wedge\zeta_i^{\beta} > t$. Hence $h(t)$ is $\mathscr{F}_{t+}^{\beta_n}$-measurable. Likewise it is easy to see that $L^{\beta_n}(t)$ and $R^{\beta_n}(t)$ are $\mathscr{F}_t^{\beta_n}$-measurable.

We now claim that $L^{\beta_n}(t)$ is an $\mathscr{F}^{\beta_n}$-reversed martingale, that is, $E\{L^{\beta_n}(\tau)|\mathscr{F}_t^{\beta_n}\} = L^{\beta_n}(t)$ whenever $\tau \le t$. To see this, note first that $L^{\beta_n}(\cdot)$ is a pure jump process with jumps only on uncensored observations. Next we obtain from the definitions of $\mathbb{F}_n^{\beta_n}$, $\bar{D}^{\beta_n}(u)$ and $R^{\beta_n}(u)$,

$$\begin{aligned}&dL^{\beta_n}(\varepsilon_i^{\beta_n})\\ &\quad= n^{-1/2}h(\varepsilon_i^{\beta_n})\left[Z_i - \bar{D}^{\beta_n}(\varepsilon_i^{\beta_n}) - \sum_{j=1}^{n}Z_j1_{\{\varepsilon_j^{\beta_n}\wedge\zeta_j^{\beta_n}>\varepsilon_i^{\beta_n}\}}\frac{d\mathbb{F}_n^{\beta_n}(\varepsilon_i^{\beta_n})}{1-\mathbb{F}_n^{\beta_n}(\varepsilon_i^{\beta_n}-)}\right]\Delta_i\\ &\quad= n^{-1/2}h(\varepsilon_i^{\beta_n})\left[Z_i - \bar{D}^{\beta_n}(\varepsilon_i^{\beta_n}) - \frac{1}{R^{\beta_n}(\varepsilon_i^{\beta_n})}\sum_{j=1}^{n}Z_j1_{\{\varepsilon_j^{\beta_n}\wedge\zeta_j^{\beta_n}\ge\varepsilon_i^{\beta_n}\}}\right]\Delta_i\\ &\quad= n^{-1/2}h(\varepsilon_i^{\beta_n})\frac{R^{\beta_n}(\varepsilon_i^{\beta_n})-1}{R^{\beta_n}(\varepsilon_i^{\beta_n})}\left[Z_i - E(Z|Y-\beta_n^TZ = \varepsilon_i^{\beta_n}, \Delta_i = 1)\right]\Delta_i.\end{aligned} \tag{4.3}$$

Now, $h(t)(R^{\beta_n}(t) - 1)/R^{\beta_n}(t)$ is $\mathscr{F}_{t+}^{\beta_n}$-measurable and

$$E\left(\Delta_i Z_i | \mathscr{F}_{t+}^{\beta_n}\right) = E\left\{\Delta_i E\left(Z | Y - \beta_n^T Z = \varepsilon_i^{\beta_n}, \Delta_i = 1\right) | \mathscr{F}_t^{\beta_n}\right\}$$

for any $t > \varepsilon_i^{\beta_n}$. Hence $E\{dL^{\beta_n}(u)|\mathscr{F}_t^{\beta_n}\} = 0$ for all $t > u$, and $L^{\beta_n}(\cdot)$ is a reversed martingdale.

The reversed martingale L^{β_n} has at most n jumps. Actually we could describe it as a martingale on the set $1, 2, \dots, n$, representing the order statistics of the residuals. By (4.3) each jump is bounded by

$$2n^{-1/2}|h(\varepsilon_i^{\beta_n})| \operatorname{ess\,sup}\|Z\| = 2n^{-1/2}|W_{\mathbb{F}_n^{\beta_n}} s(\varepsilon_i^{\beta_n}) - W_{F_0} s(\varepsilon_i^{\beta_n})|.$$

We obtain from (4.3) and Lemmas 4.1 and 4.2 that the sum of the conditional expectations of the squares of the jumps, given what is known just before them, converges to 0. Saying it otherwise, we obtained that if K_n is as defined in Lemma 4.2 then

$$\begin{aligned}
\langle L^{\beta_n}(-\infty)\rangle &= \sum E\left\{\left[dL^{\beta_n}(u)\right]^2 | \mathscr{F}_{u+}^{\beta_n}\right\} \\
&= \sum E\left\{\left[dL^{\beta_n}(u) | \mathscr{F}_{u+}^{\beta_n}, dL^{\beta_n}(u) > 0\right\} P\left\{dL^{\beta_n}(u) > 0 | \mathscr{F}_{u+}^{\beta_n}\right\} \\
&\le 4n^{-1} K_n \sum P\left\{dL^{\beta_n}(u) > 1 | \mathscr{F}_{u+}^{\beta}\right\} \\
&\to_p 0,
\end{aligned}$$

since $K_n \to_p 0$ by Lemma 4.2 and $E[\sum P\{dL^{\beta_n}(u) > 1 | \mathscr{F}_{u+}^{\beta}\}] \le n$. Now use an inequality of Lenglart (1977) [see also Gill (1980), pages 18 and 19] or the martingale CLT [cf. Rebolledo (1980)] to conclude that (4.2) holds. □

LEMMA 4.2. *Suppose $s \in \mathscr{S}_0$ and $\{\beta_n\}$ is a nonstochastic sequence such that* $\limsup n^{1/2}\|\beta_n - \beta_0\| < \infty$. *Then for some random variables $\{K_n\}$,*

$$E\left\{\left[W_{\mathbb{F}_n^{\beta_n}} s(t) - W_{F_0} s(t)\right]^2 | \mathscr{F}_u^{\beta_n}\right\} \le K_n \to_p 0 \quad \textit{for all } t < u < c_0.$$

PROOF. The claim follows Lemmas 4.1 and 7.2(iii) since $1 - \mathbb{F}_n^{\beta_n}(t) \ge 1 - \mathbb{F}_n^{\beta_n}(c_0)$ while $1 - \mathbb{F}_n^{\beta_n}(c_0)$ is $\mathscr{F}_u^{\beta_n}$-measurable and $P\{1 - \mathbb{F}_n^{\beta_n}(c_0) > \frac{1}{2}(1 - F_0(c_0))\} \to 1$. □

5. Asymptotic properties of the Buckley–James type estimators. We now turn to investigate the asymptotic properties of the Buckley–James family of estimators. Our main result is that $\Psi_n(\beta_0; s)$ is asymptotically normal with mean 0, and in a neighborhood of β_0 of order $n^{-1/2}$, $\Psi_n(\beta; s)$ is close to a linear function in $n^{1/2}(\beta - \beta_0)$. This implies that with probability converging to 1 there will be a "solution" of

$$\Psi_n(\beta; s) = o_p(1), \tag{5.1}$$

which is $\sqrt{n}$-consistent. Moreover, if the gradient of this linear function is invertible, then any $\sqrt{n}$-consistent solution of (5.1) is an asymptotically normal mean zero random variable.

We will begin by proving that there is a linear function $\phi(\cdot;s)$ such that $\Psi_n(\beta_n;s) = \Psi_n(\beta_0;s) + \phi(n^{1/2}(\beta_n - \beta_0);s) + o_p(1)$ for any deterministic sequence $\{\beta_n\}$ such that $n^{1/2}(\beta_n - \beta_0)$ is bounded. This in itself will be enough to prove that if there is a $\sqrt{n}$-consistent estimator of β_0, then there is a $\sqrt{n}$-consistent estimator that is a solution of (5.1). Alternatively, one can begin with an auxiliary $\sqrt{n}$-consistent estimator, truncate it to a grid of $O_p(n^{-1/2})$ spacing and do a one-step Newton–Raphson improvement similar to (2.2), to find an approximate solution of $\Psi_n(\beta;s) = 0$ (see Remark 6.2). In Lemma 5.2 the tightness of the family of random functions $\Psi_n(\beta_0 + tn^{-1/2};s)$, $t \in (-M, M)$, $M < \infty$, is proved. Hence there is no need to do the unpleasant (but nonharmful) discretization.

The main result of the section is as follows. Let

$$A = \int \operatorname{Var}(Z|C - \beta_0^T Z \geq u) W_{F_0}s(u) W_{F_0}s_0(u)\bar{B}(u)\, dF_0(u)$$

and

$$V = \int \operatorname{Var}(Z|C - \beta_0^T Z \geq u)\{W_{F_0}s(u)\}^2 \bar{B}(u)\, dF_0(u),$$

where $s_0 \equiv -f_0'/f_0$ and $\bar{B}(u) = P(C - \beta_0^T Z \geq u)$.

THEOREM 5.1. *Suppose Assumptions* (A.1)–(A.3) *hold and A is nonsingular. Then*

(i) $\Psi_n(\beta_0;s)$ *is asymptotically distributed as a* $\mathcal{N}(0, V)$ *random variable.*
(ii) *For any* $M < \infty$,

$$\sup_{\|\beta - \beta_0\| \leq Mn^{-1/2}} |\Psi_n(\beta;s) - \Psi_n(\beta_0;s) - n^{1/2}A(\beta - \beta_0)| \to_p 0.$$

(iii) *There is a* $\hat{\beta}_n$ *satisfying* $\Psi_n(\hat{\beta}_n;s) = o_p(1)$ *which is* $\sqrt{n}$*-consistent and asymptotically has a* $\mathcal{N}(0, A^{-1}VA)$ *law.*

(iv) *If* $s = s_0 \equiv -f_0'/f_0$, *then* $A = V$, *and* $\hat{\beta}_n$ *has the asymptotic distribution of the best regular estimator.*

The assumed nonsingularity of the matrix A may seem difficult to check. Note, however, that the matrix is nonsingular if both s and $f_0/(1 - F_0)$ are monotone, strictly monotone increasing in an interval (u, v), and the distribution of Z given $C \geq u$ does not concentrate on a hyperplane.

The theorem will be proved in the following discussion.

In the method of the proof one can distinguish three main steps. The first is the replacement of the random $\mathbb{F}_n^{\beta_n}$ by a nonrandom term. This seems to be easier to do with the use of the counting process martingale theory. The second step is proving the almost linearity of the estimating equation. The last step is the consideration of tightness. We do not know a simple way to do the two last steps with counting process devices. The difficulty follows mainly from the fact that a change in β causes a change in the *order* of the observations and it seems that there is no family of increasing σ-fields such that the Tsiatis (1987) estimator can be described conveniently for two different values of β.

Define for all $t \in R$,

$$\lambda_\beta(t) = \lim_{\nu \searrow t} \nu^{-1} \Pr\{\Delta = 1,\ t < \varepsilon^\beta \le t + \nu | \varepsilon^\beta \ge t\}$$

$$= \frac{E\left[f_0(t + \beta^T Z)\Pr(C > t + \beta^T Z|Z)\right]}{E\left[\{1 - F_0(t + \beta^T Z)\}\Pr(C > t + \beta^T Z|Z)\right]}$$

[the limit exists a.e. in view of the finite information assumption (A.3)]. Let F_β be the distribution function with hazard function λ_β and let f_β be its density. It is reasonable to assume that F_β converges in some sense to F_0 as $\beta \to 0$; see Lemma 7.2 for details.

In Proposition 4.1 it was proved that $\Psi_n(\beta_n; s) = \Gamma_n(\beta; W_{F_0}s) + o_p(1)$. It will be useful to have a slight modification of this result.

PROPOSITION 5.1. *Let $\{\beta_n\}$ be any deterministic sequence such that*

$$\limsup n^{1/2}\|\beta_n - \beta_0\| < \infty.$$

Then

$$\Psi_n(\beta_n; s) = \Gamma_n\left(\beta_n; W_{F_{\beta_n}}s\right) + o_p(1).$$

The proof of the proposition relies on the convergence of $W_{F_{\beta_n}}s$ to $W_{F_0}s$, which is proved in Lemma 7.2. It is essentially the same as the proof of Proposition 4.1 and is omitted.

We next show that the random mean of the covariates of the subjects at risk at time u appearing in the expression (2.5) of $\Gamma_n(\beta; w)$ can be replaced by the nonrandom quantity

$$D^\beta(t) \equiv E(Z|C \wedge Y - \beta^T Z > t).$$

LEMMA 5.1. *Let $\{\beta_n\}$ be any nonrandom sequence such that*

$$\limsup n^{1/2}\|\beta_n - \beta_0\| < \infty.$$

Then

$$\Psi_n(\beta_n; s) = n^{-1/2}\sum_{i=1}^n \int 1_{\{\zeta_i^{\beta_n} \ge u\}}\{Z_i - D^{\beta_n}(u)\} W_{F_{\beta_n}}s(u)\, d\mathbb{M}\left(u; \varepsilon_i^{\beta_n}|F_{\beta_n}\right) + o_p(1).$$

PROOF. By (2.5) and Proposition 5.1 we have

$$\Psi_n(\beta_n; s) = n^{-1/2}\sum_{i=1}^n \int 1_{\{\zeta_i^{\beta_n} \ge u\}}\{Z_i - \hat{D}^{\beta_n}(u)\} W_{F_{\beta_n}}s(u)\, d\mathbb{M}\left(u; \varepsilon_i^{\beta_n}|F_{\beta_n}\right) + o_p(1), \tag{5.1}$$

where

$$\hat{D}^{\beta_n}(t) = \frac{\sum_{i=1}^n Z_i 1_{\{\varepsilon_i^{\beta_n} \wedge \zeta_i^{\beta_n} \geq t\}}}{\sum_{i=1}^n 1_{\{\varepsilon_i^{\beta_n} \wedge \zeta_i^{\beta_n} \geq t\}}}$$

$$= \frac{\sum_{i=1}^n Z_i}{R^{\beta_n}(t)} - \frac{\sum_{i=1}^n Z_i 1_{\{\varepsilon_i^{\beta_n} \wedge \zeta_i^{\beta_n} < t\}}}{R^{\beta_n}(t)}$$

$$= \frac{n\bar{Z}}{R^{\beta_n}(t)} + \frac{\sum_{i=1}^n Z_i 1_{\{\varepsilon_i^{\beta_n} \wedge \zeta_i^{\beta_n} < t\}}}{R^{\beta_n}(t)}.$$

Define the process

$$M^{\beta_n}(t) = n^{-1/2} \sum_{i=1}^n \int_{(-\infty, t]} \left\{ \hat{D}^{\beta_n}(u) - D^{\beta_n}(u) + \frac{n(EZ - \bar{Z})}{R^{\beta_n}(u)} \right\}$$

$$\times W_{F_{\beta_n}} s(u)\, d\mathbb{M}\left(u; \varepsilon_i^{\beta_n} | F_{\beta_n}\right)$$

and the family of increasing σ-fields $\mathscr{H}^{\beta_n} = \{\mathscr{H}_t^{\beta_n}: t \in R\}$, where

$$\mathscr{H}_t^{\beta_n} = \sigma\left\{ \left[1_{\{\varepsilon_i^{\beta_n} \wedge \zeta_i^{\beta_n} \leq t\}} \left(Z_i, \varepsilon_i^{\beta_n} \wedge \zeta_i^{\beta_n}, \Delta_i \right), 1_{\{\varepsilon_i^{\beta_n} \wedge \zeta_i^{\beta_n} \leq t\}} \right] : i = 1, \dots, n \right\}.$$

($\mathscr{H}$ stands for history: Given $\mathscr{H}_t^{\beta_n}$, we know all X_i with $\varepsilon_i^{\beta_n} \wedge \zeta_i^{\beta_n} \leq t$ and only the indices of those X_i with $\varepsilon_i^{\beta_n} \wedge \zeta_i^{\beta_n} > t$.)

It is easy to see that $\mathbb{M}(u; \varepsilon_i^{\beta_n} | F_{\beta_n})$ is an $\mathscr{H}^{\beta_n}$-martinglae. Actually F_{β_n} was defined so as to make this statement true. Now $\hat{D}^{\beta_n}(t) + n(EZ - \bar{Z})/R_{\beta_n}(t)$ is $\mathscr{H}_t^{\beta_n}$-measurable. Hence $M^{\beta_n}(\cdot)$ is an integral of a predictable process with respect to a martingale.

We are going now to show that the predictable variation process of M^{β_n} at its endpoint,

$$\langle M^{\beta_n} \rangle = n^{-1} \int_{(-\infty, c_0]} \left\{ \hat{D}^{\beta_n}(u) - D^{\beta_n}(u) + \frac{n(EZ - \bar{Z})}{R^{\beta_n}(u)} \right\}^2$$

$$\times \left\{ W_{F_{\beta_n}} s(u) \right\}^2 R^{\beta_n}(u) \lambda_{\beta_n}(u)\, du$$

$$\leq \int_{(-\infty, c_0]} \left\{ \tilde{D}^{\beta_n}(u) - D^{\beta_n}(u) + \frac{nEZ}{R^{\beta_n}(u)} \right\}^2 \left\{ W_{F_{\beta_n}} s(u) \right\}^2 \lambda_{\beta_n}(u)\, du,$$

converges to 0. First note that the measure $\lambda_{\beta_n}(u) 1_{\{u \leq c_0\}}$ is equivalent to $f_{\beta_n}(u) 1_{\{u \leq c_0\}}\, du$. Then, since F_{β_n} is continuous, the L_2 norm of $W_{F_{\beta_n}} s$ is the same as the norm of s (see Lemma 3.1). But the denominator of the r.h.s. of the definition of λ_β is bounded away from 0 with probability converging to 1 and Z is bounded. Therefore, with probability converging to 1, $f_{\beta_n}(u) \leq 2 \sup\{ f(u + t)$:

$|t| \le \|\beta_n\|\}$. It follows from Assumptions (A.1) and (A.2) that

$$\begin{aligned}
\frac{1}{2}\int_{(-\infty,\, c_0)} & s^2(t) f_{\beta_n}(t)\, dt \\
&\le \int_{(-\infty,\, c_0)} s^2(t) f_0(t)\, dt \\
&\quad + \int_{(-\infty,\, c_0)} s^2(t) \int_{(-\|\beta_n\|,\, \|\beta_n\|)} |f'(t+u)|\, du\, dt \\
&\le \int_{(-\infty,\, c_0)} s^2(t) f_0(t)\, dt \\
&\quad + 2\|\beta_n\| \int_{(-\infty,\, c_0 + \|\beta_n\|)} \sup\{s^2(t+u)\colon |u| \le |\beta_n|\} |f_0'(t)|\, dt
\end{aligned}$$

and the second term in the r.h.s. is of order $n^{-1/2}$ in view of Assumptions (A.1)(i) and (A.3) and Cauchy inequality. We conclude that $W_{\beta_n} s$ is square-integrable.

Now, by Lemma 7.3(i),

$$\left\| \{\hat{D}^{\beta_n}(u) - D^{\beta_n}(u)\} 1_{\{u \le c_0\}} \right\|_\infty \to_p 0.$$

Certainly $n(EZ - \bar{Z})/R^{\beta_n}(c_0) \to_p 0$. Hence

$$\|\hat{D}^{\beta_n}(u) - D^{\beta_n}(u) + n(EZ - \bar{Z})/R^{\beta_n}(u)\|_\infty \to_p 0.$$

We therefore obtain that the predictable variation process, $\langle M^{\beta_n}\rangle(\infty)$, converges to 0 and hence we conclude from the Lenglart (1977) inequality that

$$M^{\beta_n}(\infty) \to_p 0.$$

We are not interested in $M^{\beta_n}(\infty)$ but in a similar expression without the term involving $\bar{Z}$ and EZ. But since for $t \le c_0$, $n(\bar{Z} - EZ)/R^{\beta_n}(t) = O_p(n^{-1/2})$ and $\bar{Z} - EZ$ can be taken outside the integral, a similar argument implies that

$$n^{-1/2} \sum_{i=1}^{n} \int_{(-\infty,\, t)} 1_{\{\zeta_i^{\beta_n} \ge u\}} \{\hat{D}^{\beta_n}(u) - D^{\beta_n}(u)\} W_{F_{\beta_n}} s(u)\, d\mathbb{M}\left(u; \varepsilon_i^{\beta_n} | F_{\beta_n}\right) \to_p 0.$$

Together with (5.1) this proves the lemma. □

The next step is moving back from the martingale representation of the estimating equation to a "standard" M-estimator form. The following result is an immediate consequence of Proposition 3.1 and Lemma 5.1.

PROPOSITION 5.2. *Let $\{\beta_n\}$ be any sequence such that*

$$\limsup n^{1/2} \|\beta_n - \beta_0\| < \infty.$$

Then

$$\Psi_n(\beta_n; s) = n^{-1/2} \sum_{i=1}^{n} \int 1_{\{\zeta_i^{\beta_n} \ge u\}}\{Z_i - D^{\beta_n}(u)\}\, d\mathbb{Q}_s\big(u; \varepsilon_i^{\beta_n} | F_{\beta_n}\big) + o_p(1)$$

$$= n^{-1/2} \sum_{i=1}^{n} \left[Z_i \left\{ \Delta_i s\big(\varepsilon_i^{\beta_n}\big) + (1 - \Delta_i) \frac{\int 1_{\{u \ge \zeta_i^{\beta_n}\}} s(u)\, dF_{\beta_n}(u)}{1 - F_{\beta_n}\big(\zeta_i^{\beta_n}\big)} \right\} - \int 1_{\{\zeta_i^{\beta_n} \ge u\}} D(u)\, d\mathbb{Q}_s\big(u; \varepsilon_i^{\beta_n} | F_{\beta_n}\big) \right] + o_p(1).$$

The final preliminary needed result is the tightness of $\Psi_n(\cdot\,; s)$. The following lemma is proved in Section 7.

LEMMA 5.2. *Let $\{\beta_n\}$ be any deterministic sequence such that*

$$\limsup n^{1/2} \|\beta_n - \beta_0\| < \infty.$$

Then for any sequence $\{\nu_n\}$, $\nu_n \ge 0$, $\nu_n n^{1/2} \to 0$,

$$\sup\{|\Psi_n(\beta; s) - \Psi_n(\beta_n; s)|: \|\beta - \beta_n\| \le \nu_n\} = o_p(1).$$

Given Proposition 5.2 and Lemma 5.2, the proof of Theorem 5.1 is standard and is done in Section 7.

6. Additional remarks.

6.1. *The "favorite" estimators.* It is clear that if F_0 has a second moment then $s(t) \equiv t \wedge c_0 \in \mathscr{S}_0$ and the Buckley–James estimator are covered by our results. The Tsiatis (1987) estimator with $w(t) \equiv -1_{\{t < c_0\}}$ is equivalent to the Buckley–James estimator with

$$s(t) \equiv -S_{F_0} 1_{\{\cdot < c_0\}}(t) \equiv -1 - \log[1 - F_0(t)], \qquad t < c_0.$$

It is efficient if $1 - F_0(t) = \exp\{-e^t\}$ and then $s(t) = e^t - 1$. Other possible functions w can be derived from the rank tests suggested by Gehan (1965), Peto and Peto (1972), Prentice (1978) and Harrington and Fleming (1982).

6.2. *Isolating a $\sqrt{n}$-consistent root.* James and Smith (1984) show that for $Z \in R$ and $s(t) \equiv t \wedge c_0$, any solution of $\Psi_n(\beta; s) = o_p(1)$ is consistent. This result is not strong enough for our purpose since we should be able to isolate a $\sqrt{n}$-consistent root. Clearly, there is no problem when all the roots of $\Psi_n(\beta; s)$ are $O_p(n^{-1/2})$ apart. Otherwise, we need an initial $\sqrt{n}$-consistent estimator. Moreover, given a $\sqrt{n}$-consistent estimator, a one-step Newton–Raphson correction gives an estimator that is equivalent to order $o_p(n^{-1/2})$ to the proper root of

$\Psi_n(\beta; s)$. Thus, if $\beta_0 \in R$, $\tilde{\beta}_n$ is $\sqrt{n}$-consistent and

$$(6.1)\qquad \hat{\beta}_n = \frac{(\tilde{\beta}_n - d_n)\Psi_n(\tilde{\beta}_n + d_n; s) - (\tilde{\beta}_n + d_n)\Psi_n(\tilde{\beta}_n - d_n; s)}{\Psi_n(\tilde{\beta}_n + d_n; s) - \Psi_n(\tilde{\beta}_n - d_n; s)},$$

then $\hat{\beta}$ has the properties given in Theorem 5.1 whenever $\{n^{1/2}d_n\}$ is bounded away from 0 and ∞.

For particular submodels one can use the estimator suggested by Miller (1976) (for the model $C - \beta_0^T Z$ independent of Z), Koul, Susarla and Van Rayzin (1981) (C independent of Z) or Powell (1984) (type I censoring, i.e., C is always observed) as appropriate simple but $\sqrt{n}$-consistent estimators.

Ritov (1984, 1986) suggested a (not too practical) estimator that seems to be $\sqrt{n}$-consistent under the minimal assumption of positive finite information for estimating the slope. The estimator is relatively simple when Z has positive masses at points $z_1, \dots, z_k$ which are not on a proper hyperplane. Then we can consider only that part of the data with Z equal to one of $z_1, \dots, z_k$. We can find a q such that $f_0(t_q) > 0$ and $\min_j \Pr(C - \beta_0^T z_j > t_q | Z = z_j) > 0$, where $F_0(t_q) = q$. Now construct estimates $\hat{t}_{q1}, \dots, t_{qk}$ of t_q based on the subsamples $\{X_i\colon Z_i = z_j\}$, $j = 1, \dots, k$, respectively, and estimate $\tilde{\beta}_n$ by a robust estimator of regression based on $\{(\hat{t}_{qj}, z_j)\colon j = 1, \dots, k\}$. If Z is continuous and $\beta \in R$, the the following program can be carried out. Divide the range of Z into intervals small enough $[O_p(n^{-1/2})]$ that Z does not vary much in each, but large enough that a sufficient number of them contain $O_p(n^{1/2})$ observations. Continue, as in the discrete case, as if Z is constant in each interval. If one chooses the intervals without any prior knowledge of the distribution of X, then the detailed construction is tedious and is described in Ritov (1986).

6.3. *Information bound.* The information bound for the estimation of β was derived in Ritov and Wellner (1988) and Bickel, Klaassen, Ritov and Wellner (1989). It is given by the matrix V in Theorem 5.1 with $s = -f_0'/f_0$. It may be of interest to note that the information is greater then 0 if, and only if, $\Pr(Y \le C) > 0$ and the support of Z given $Y \le C$ is not a proper subset of R^m [Ritov (1984, 1986)].

6.4. *Efficient estimators.* If $s = -f_0'/f_0$ then the estimator is efficient. Now consider the problem of constructing an efficient estimator without prior knowledge of F_0. We need to consider two problems: (i) How to replace c_0 by $c_n \to \infty$. (ii) How to estimate $-f_0'/f_0$. Note that the first problem does not arise when there is a c_0 such that $\Pr(Y - \beta_0^T Z > c_0) > 0$ and $\Pr(C - \beta_0^T Z > c_0) = 0$. When $f_0'(t)/f_0(t)$ is estimated using only residuals greater than t, the martingale structures used in the proof are not destroyed. Moreover, the rate of convergence of the estimated score function is not important for the asymptotic efficiency. Again, these problems can be solved, but the calculations are tedious. See Tsiatis (1987) and Ritov (1986).

6.5. *The Tsiatis estimator.* Suppose Assumptions (A.2) and (A.3) hold and (A.1) is satisfied by $S_{F_0}w$. It was proved in Proposition 4.1 that $\Gamma_n(\beta; w) = \Psi_n(\beta; S_{F_0}) + o_p(1)$, uniformly for $\beta \in B_n$, where $B_1, B_2, \dots$ are deterministic finite sets such that $\limsup \operatorname{card}(B_n) < \infty$. Thus, as a corollary of this paper, we get an alternative proof to Tsiatis' (1987) main result restricted to discretized estimators, for example, estimators that take values on the grid $\{0, \pm n^{-1/2}, \pm 2n^{-1/2}, \dots\}$. This paper is irrelevant to the tightness considerations in Tsiatis' paper and therefore to the uniform behavior of his estimator on small intervals.

6.6. *Consistent estimator of the variance.* Tsiatis (1987) suggested a consistent estimate of the variance of $\hat{\beta}$. The bootstrap may be another alternative. If $\beta \in R$ then the one-step estimator given by (6.1) seems to provide an easy algorithm. For any pseudosample calculate (6.1) with $\tilde{\beta}_n$, the initial point, always equal to the estimator derived from the original sample.

7. Lemmas and proofs. In this section we assume that (A.1)–(A.3) hold, and w.l.o.g. that $\beta_0 = 0$, $\Pr(|Z| \le 1) = 1$ and $s(t)1_{\{t \ge c_0\}} = 0$ for all $s \in \mathscr{S}_0$. We denote by $\{\beta_n\}$ any deterministic sequence such that $n^{1/2}\beta_n$ is bounded. Let F_β be as in Section 5. Define

$$\mathbb{H}_u^\beta(t) \equiv n^{-1}\#\left\{X_i\colon \Delta_i = 1,\ \varepsilon_i^\beta \le t\right\},$$

$$\mathbb{H}^\beta(t) \equiv n^{-1}\#\left\{X_i\colon \varepsilon_i^\beta \wedge \zeta_i^\beta \le t\right\}.$$

In the following lemma the convergence of $n^{1/2}(\mathbb{F}_n^\beta - F_\beta)$ is investigated. Its proof is standard—the fact that $Y - \beta^T Z$ is not independent of $C - \beta^T Z$ is immaterial to the standard proofs of the convergence of the KM estimator. The proof is given by completeness.

LEMMA 7.1.

(i) $$\sup_{t \le c_0} |\mathbb{F}_n^{\beta_n}(t) - F_{\beta_n}(t)| = O_p(n^{-1/2}).$$

(ii) *For any* $s \in \mathscr{S}_0$,

$$\sup_{t \le c_0} n^{-1/2}\left|\int_{(t, c_0)} s(u)\left\{d\mathbb{F}_n^{\beta_n}(u) - dF_{\beta_n}(u)\right\}\right| \to 0 \quad \textit{in second mean}$$

and

$$\left\| W_{\mathbb{F}_n^{\beta_n}} s - W_{F_{\beta_n}} s \right\|_\infty = O_p(n^{-1/2}).$$

PROOF. First note that $\mathbb{F}_n^{\beta_n}(\cdot) \ge \mathbb{H}^{\beta_n}(\cdot)$. Hence by Assumption (A.1)

$$\Pr\left\{\mathbb{F}_n^{\beta_n}(c_0) < 1 - \gamma\right\} \to 1, \tag{7.1}$$

for some $\gamma > 0$. Let

$$\hat{\Lambda}^{\beta_n}(t) = \int_{(-\infty, t]} \left(1 - \mathbb{F}_n^{\beta_n}(u-)\right)^{-1} d\mathbb{F}_n^{\beta_n}(u).$$

By (7.1) all increments of $\hat{\Lambda}^{\beta_n}$ on $(-\infty, c_0]$ are $O_p(n^{-1})$. Hence

$$\begin{aligned} 1 - \mathbb{F}_n^{\beta_n}(t) &= \exp\{-\hat{\Lambda}^{\beta_n}(t)\} + O_p(n^{-1}) \\ &= \left(1 - F_{\beta_n}(t)\right)\exp\left\{-\hat{\Lambda}^{\beta_n}(t) + \int_{(-\infty, t]} \lambda_{\beta_n}(u)\, du\right\} + O_p(n^{-1}), \end{aligned}$$

where the $O_p(n^{-1})$ term is uniform for $t \le c_0$ [it is a function of $\mathbb{H}^{\beta_n}(c_0)$ only]. But $\hat{\Lambda}^{\beta_n}(\cdot) - \int_{(-\infty,\cdot]}\lambda_{\beta_n}(u)\,du$ is an $\mathscr{H}^{\beta_n}$-martingale, where the family $\mathscr{H}^{\beta_n}$ of σ-fields is as defined in the proof of Lemma 5.2. Now the value of the predictable variation process of this martingale evaluated at the endpoint c_0 is given by

$$\left\langle \hat{\Lambda}^{\beta_n}(c_0) - \int_{(-\infty, c_0]} \lambda_{\beta_n}(u)\, du \right\rangle = n^{-1} \int_{(-\infty, c_0]} \left(1 - \mathbb{H}^{\beta_n}(u-)\right)^{-1} \lambda_{\beta_n}(u)\, du,$$

which is $O_p(n^{-1})$ by (7.1). This proves the first part by the Lenglart (1977) inequality.

(ii) By part (i) it is enough to prove that

$$\sup_{t \le c_0} \left| \int_{(t, c_0)} s(u)\left\{d\mathbb{F}_n^{\beta_n}(u) - dF_{\beta_n}(u)\right\}\right| = O_p(n^{-1/2}).$$

This will follow if

$$\sup_{t \le c_0} \left| \int_{(-\infty, t)} s(u)\left\{d\mathbb{F}_n^{\beta_n}(u) - dF_{\beta_n}(u)\right\}\right| = O_p(n^{-1/2}).$$

But

$$n^{1/2} \int_{(-\infty, \cdot)} s(t)\left\{d\mathbb{F}_n^{\beta_n}(t) - \frac{1 - \mathbb{F}_n^{\beta_n}(t-)}{1 - F_{\beta_n}(t)} dF_{\beta_n}(t)\right\}$$

stopped at c_0 is an $\mathscr{H}^{\beta_n}$-martingale with the predictable variation process

$$\begin{aligned} &\int_{(-\infty,\cdot)} s^2(t)\left(1 - \mathbb{F}_n^{\beta_n}(t-)\right)^2 \left\{\left(1 - F_{\beta_n}(t)\right)\left(1 - \mathbb{H}^{\beta_n}(t-)\right)\right\}^{-1} dF_{\beta_n}(t) \\ &\quad \le \left(1 - F_{\beta_n}(c_0)\right)^{-1} \int_{(-\infty, c_0)} s^2(t)\, dF_{\beta_n}(t). \end{aligned}$$

Hence the supremum of this process is $O_p(1)$ and part (ii) follows from the Lenglart (1977) inequality. □

We now investigate the convergence of F_{β_n} to F_0.

LEMMA 7.2.

(i)
$$\sup_{t \le c_0} |F_\beta(t) - F_0 t)| = O(\|\beta\|) \quad \text{as } \beta \to 0.$$

(ii) *For all* $s \in \mathscr{S}_0$,

$$\left\| \int_{(-\infty, \cdot]} s(t)\{ f_\beta(t) - f_0(t)\}\, dt \right\|_\infty = O(\|\beta\|) \quad \text{as } \beta \to 0.$$

(iii)
$$\left\| W_{F_\beta} s - W_{F_0} s \right\|_\infty = O(\|\beta\|) \quad \text{as } \beta \to 0.$$

PROOF. (i) Note that

$$\begin{aligned}
|\lambda_\beta(t) - \lambda_0(t)| &\le \operatorname*{ess\,sup}_{|\xi| \le \|\beta\|} |\lambda_0(t+\xi) - \lambda_0(t)| \\
&\le \int_{(-\|\beta\|, \|\beta\|)} |\lambda_0'(t+\xi)|\, d\xi \qquad (7.2)\\
&\le \int_{(-\|\beta\|, \|\beta\|)} \left[\frac{|f_0'(t+\xi)|}{1 - F_0(t+\xi)} + \left(\frac{f_0(t+\xi)}{1 - F_0(t+\xi)} \right)^2 \right] d\xi,
\end{aligned}$$

where $\lambda_0 = (1 - F_0)^{-1} f_0$. Hence, for any $t \le c_0$,

$$\begin{aligned}
\left| \log \frac{1 - F_0(t)}{1 - F_\beta(t)} \right| &\le \int_{(-\infty, c_0)} |\lambda_\beta(t) - \lambda_0(t)|\, dt \\
&\le \int_{(-\|\beta\|, \|\beta\|)} \int_{(-\infty, c_0 + \|\beta\|)} |\lambda_0'(t+\xi)|\, dt\, d\xi \\
&\le 2\|\beta\|(a_1 + a_2),
\end{aligned}$$

where

$$a_1 \equiv \left(1 - F_0(c_0 + \|\beta\|)\right)^{-1} \int |f_0'(t)|\, dt$$

and

$$a_2 = \{1 - F_0(c_0 + \|\beta\|)\}^{-2} \operatorname{ess\,sup} f_0.$$

Both a_1 and a_2 are finite, since for $\|\beta\|$ small enough $1 - F_0(c_0 + \|\beta\|) > \gamma$ and

$$\operatorname{ess\,sup} f_0 \le \frac{1}{2} \int |f_0'(t)|\, dt \le \frac{1}{2} \left\{ \int \frac{(f_0'(t))^2}{f_0(t)}\, dt \right\}^{1/2} < \infty$$

by (A.3).

(ii) Let $s \in \mathscr{S}_0$. Then

$$\int s(t)\{f_\beta(t) - f_0(t)\}\,dt = \int s(t)\frac{F_0(t) - F_\beta(t)}{1 - F_0(t)} f_0(t)\,dt$$
$$+ \int s(t)(1 - F_\beta(t))(\lambda_\beta(t) - \lambda_0(t))\,dt$$
$$= L_1 + L_2, \quad \text{say}.$$

Since $s \in L_2(F)$ and $s(t) \equiv s(t)1_{\{t \le c_0\}}$ we conclude from part (i) that $L_1 = O_p(n^{-1/2})$. Now from (7.2) and $s \in \mathscr{S}_0$,

$$|L_2| \le \int_{(-\infty, c_0]} |s(t)| \int_{(t-\|\beta\|, t+\|\beta\|)} \{b_1|f_0'(\xi)| + a_2 f_0(\xi)\}\,d\xi\,dt$$
$$\le 2\|\beta\| \int_{(-\infty, c_0+\|\beta\|]} \{b_1|f_0'(\xi)| + a_2 f_0(\xi)\}\sup\{|s(t)|: |t-\xi| \le \|\beta\|\}\,d\xi$$
$$\le 2\|\beta\| \left[b_1 \left\{ \int \frac{(f_0'(t))^2}{f_0(t)}\,dt \right\}^{1/2} \left\{ \int f_0(t)\sup\{s^2(\xi): |t-\xi| \le \|\beta\|\}\,dt \right\}^{1/2} \right.$$
$$\left. + \left\{ \int f_0(t)\sup\{s^2(\xi): |t-\xi| \le \|\beta\|\}\,dt \right\}^{1/2} \right]$$
$$= O_p(\|\beta\|),$$

where $b_1 = (1 - F_0(c_0 + \|\beta\|))^{-1}$ and a_2 is as above.

(iii) Immediate corollary of parts (i) and (ii). □

LEMMA 7.3. *Define*

$$\hat{D}^\beta(t) = \sum_{i=1}^n Z_i 1_{\{\varepsilon_i^{\beta_n} \wedge \zeta_i^{\beta_n} \ge t\}} \Big/ \sum_{i=1}^n 1_{\{\varepsilon_i^{\beta_n} \wedge \zeta_i^{\beta_n} \ge t\}}.$$

Then

(i) $$\sup_{t \le c_0} |\hat{D}^{\beta_n}(t) - E(Z|Y \wedge C - \beta_n^T Z \ge t)| = O_p(n^{-1/2}).$$

(ii) $$E(Z|Y \wedge C - \beta_n^T Z \ge \varepsilon_1^{\beta_n}) - E(Z|C - \beta_n^T Z \ge Y - \beta_n^T Z = \varepsilon_1^{\beta_n}) \to_p 0.$$

(iii) $$\overline{D}^{\beta_n}(\varepsilon_i^{\beta_n}) \to_p 0,$$

where $\overline{D}$ is as defined in the proof of Proposition 4.1.

PROOF. (i) It is a simple empirical process result [cf. Shorack and Wellner (1986), page 283] that

$$\sup_t \left| n^{-1/2} \sum_{i=1}^{n} \left[Z_i 1_{\{\varepsilon_i^{\beta_n} \wedge \zeta_i^{\beta_n} \geq t\}} - E\left(Z 1_{\{Y \wedge C - \beta_n^T Z \geq t\}}\right)\right] \right| = O_p(1) \tag{7.3}$$

and

$$\sup_t \left| n^{-1/2} \sum_{i=1}^{n} \left[1_{\{\varepsilon_i^{\beta_n} \wedge \zeta_i^{\beta_n} \geq t\}} - E\left(1_{\{Y \wedge C - \beta_n^T Z \geq t\}}\right)\right] \right| = O_p(1).$$

Since, by assumption, the denominator of $\hat{D}^{\beta_n}(t)$ is $O_p(n)$ for all $t \leq c_0$ the first part follows (7.3).

(ii) First note that the Cauchy–Schwarz inequality yields for all t and τ,

$$\begin{aligned} |f_0(t) - f_0(\tau)| &\leq \int_{(t,\tau]} |f_0'(\xi)|\, d\xi \\ &\leq \left\{ \int \frac{(f_0'(\xi))^2}{f_0(\xi)} d\xi \right\}^{1/2} \{|t - \tau| \operatorname{ess\,sup} f_0\}^{1/2} \\ &\leq |t - \tau|^{1/2} \int \frac{(f_0'(\xi))^2}{f_0(\xi)} d\xi. \end{aligned}$$

Hence for any t such that $f_0(t) \geq \|\beta_n\|^{1/4}$,

$$\begin{aligned} &E\left(Z | C - \beta_n^T Z \geq Y - \beta^T Z = t\right) \\ &\quad = \left(1 + O\left(\|\beta_n\|^{1/2}\right)\right) E\left(Z | C - \beta_n^T Z \geq t\right). \end{aligned} \tag{7.4}$$

Now Z is bounded, and $\Pr\{f_0(\varepsilon_1^{\beta_n}) \geq \|\beta_n\|^{1/4}\} \to 1$, and the claim follows.

(iii) Since $\inf_{t \leq c_0} R^{\beta_n}(t) = O_p(n)$, we obtain

$$\sup_{t \leq c_0} |\overline{D}^{\beta_n}(t) - E(Z | C - \beta^T Z \geq Y - \beta^T Z = t) + \hat{D}^{\beta_n}(t)| = O_p(n^{-1/2}),$$

and hence part (iii) follows (i) and (ii). □

PROOF OF LEMMA 5.2 [Tightness of $\Psi_n(\cdot; s)$]. Two terms are involved in the tightness considerations: $n^{-1/2}\sum_{i=1}^n \Delta_i Z_i s(\varepsilon_i^\beta)$ and

$$n^{-1/2} \sum_{i=1}^{n} (1 - \Delta_i) Z_i \int_{(\zeta_i^\beta, \infty)} s(t)\, d\mathbb{F}^\beta(t) / \left(1 - \mathbb{F}^{\beta_n}(\zeta_i^\beta)\right).$$

It is easy to see that Assumption (A.1)–(A.3) imply that

$$\sup\left\{ \left| n^{-1/2} \sum_{i=1}^{n} \Delta_i Z_i \left[s\left(\varepsilon_i^\beta\right) - s\left(\varepsilon_i^{\beta_n}\right)\right] \right| : \|\beta - \beta_n\| \leq \nu_n \right\} \to 0.$$

Hence it is enough to consider the family of random variables:

$$\left\{ n^{-1/2} \sum_{i=1}^{n} (1-\Delta_i) Z_i \frac{\int_{(\zeta_i^\beta, \infty)} s(t)\, d\mathbb{F}_n^\beta(t)}{1-\mathbb{F}_n^{\beta_n}(\zeta_i^\beta)} : \|\beta-\beta_n\| \le \nu_n \right\}. \tag{7.5}$$

We begin by investigating the behavior of $\mathbb{F}_n^\beta(\cdot)$ as a function of β. Now, for some $\gamma > 0$, $\Pr\{\mathbb{H}^{\beta_n}(c_0) < 1-\gamma\} \to 1$. Hence

$$\sup_{t,\, \|\beta-\beta_n\| \le \nu_n} \left| \log\{1-\mathbb{F}_n^\beta(t)\} - n^{-1} \sum_{i=1}^{n} \Delta_i 1_{\{\varepsilon_i^\beta \le t\}} \{1-\mathbb{H}^\beta(\varepsilon_i^\beta -)\}^{-1} \right| \tag{7.6}$$
$$= O_p(n^{-1}).$$

Now, for all β,

$$\mathbb{H}^{\beta_n}(u - \|\beta-\beta_n\|) \le \mathbb{H}^\beta(u) \le \mathbb{H}^{\beta_n}(u + \|\beta-\beta_n\|). \tag{7.7}$$

Hence

$$\begin{aligned}
n^{-1/2} \sum_{i=1}^{n} &\Delta_i 1_{\{\varepsilon_i^\beta \le t\}} \{1-\mathbb{H}^\beta(\varepsilon_i^\beta -)\}^{-1} - n^{-1/2} \sum_{i=1}^{n} \Delta_i 1_{\{\varepsilon_i^{\beta_n} \le t\}} \{1-\mathbb{H}^{\beta_n}(\varepsilon_i^{\beta_n} -)\}^{-1} \\
&\le n^{-1/2} \sum_{i=1}^{n} \Delta_i 1_{\{\varepsilon_i^{\beta_n} < t+\nu_n\}} \{1-\mathbb{H}^{\beta_n}(\varepsilon_i^{\beta_n} + 2\nu_n -)\}^{-1} \\
&\quad - n^{-1/2} \sum_{i=1}^{n} \Delta_i 1_{\{\varepsilon_i^{\beta_n} \le t\}} \{1-\mathbb{H}^{\beta_n}(\varepsilon_i^{\beta_n} -)\}^{-1} \\
&= n^{-1/2}\{\mathbb{H}_u^{\beta_n}(t+3\nu_n) - \mathbb{H}_u^{\beta_n}(t)\} \\
&\quad + n^{1/2} \int_{(-\infty, t]} \{1-\mathbb{H}^{\beta_n}(u)\}^{-1} \{d\mathbb{H}_u^{\beta_n}(u+2\nu_n) - d\mathbb{H}_u^{\beta_n}(u)\} \\
&= n^{1/2}\{\mathbb{H}_u^{\beta_n}(t+3\nu_n) - \mathbb{H}_u^{\beta_n}(t)\} \\
&\quad + n^{1/2} \int_{(-\infty, t]} \{\mathbb{H}_u^{\beta_n}(u+2\nu_n) - \mathbb{H}_u^{\beta_n}(u)\}\, d\{1-\mathbb{H}^{\beta_n}(u)\}^{-1} \\
&\to_p 0
\end{aligned}$$

as $n \to \infty$ and $\nu \to 0$ uniformly in $t \in (-\infty, c_0]$.

The last claim follows the uniform continuity of $\mathbb{H}_u^{\beta_n}$ [note that $\mathbb{H}_u^{\beta_n}$ is a sample from a distribution with a bounded density as a conseqeunce of the finite information assumption and see Shorack and Wellner (1986), pages 542–552]. A similar argument applies to the infimum over β. Therefore (7.6) implies that

$$\sup_{t \le c_0,\, \|\beta-\beta_n\| \le \nu_n} |\mathbb{F}_n^\beta(t) - \mathbb{F}_n^{\beta_n}(t)| \to_p \quad \text{as } \nu \to 0. \tag{7.8}$$

To make long equations shorter, define

$$G^\beta(\cdot) = \int_{[\cdot, c_0)} s(u)\{1-\mathbb{F}_n^\beta(u)\}\, d\mathbb{H}_u^\beta(u)$$

and

$$K^{\beta}(\cdot) = \{1 - \mathbb{H}^{\beta}(\cdot)\}^{-1}.$$

Now, for any $t \le c_0$ and β: $\|\beta - \beta_n\| \le \nu_n$,

$$\begin{aligned}\int_{(t,\infty)} s(u)\, d\mathbb{F}_n^{\beta}(u) &= \int_{(t,c_0)} s(u) \frac{1 - \mathbb{F}_n^{\beta}(u-)}{1 - \mathbb{H}^{\beta}(u-)} d\mathbb{H}_u^{\beta}(u) \\ &= \int_{(t,c_0]} s(u)\{1 - \mathbb{F}_n^{\beta}(u-)\} \\ &\quad \times \left[1 + \int_{(-\infty,u)} d\{1 - \mathbb{H}^{\beta}(\xi)\}^{-1}\right] d\mathbb{H}_u^{\beta}(u) \\ &= G^{\beta}(t) + \int_{(-\infty,c_0]} G^{\beta}(\xi)\, dK^{\beta}(\xi).\end{aligned} \tag{7.9}$$

With probability converging to 1, $\{1 - \mathbb{H}^{\beta_n}(\cdot)\}^{-1}$ is a finite positive measure on $(-\infty, c_0]$. Therefore (7.9) implies that

$$\sup_{t \le c_0,\, \|\beta - \beta_n\| \le \nu_n} \left| n^{1/2} \int_{(t,c_0]} s(u)\{d\mathbb{F}_n^{\beta}(u) - d\mathbb{F}_n^{\beta_n}(u)\} \right| \to_p 0 \tag{7.10}$$

if

$$\sup_{t \le c_0,\, \|\beta - \beta_n\| \le \nu_n} n^{1/2} |G^{\beta}(t) - G^{\beta_n}(t)| \to_p 0 \tag{7.11}$$

and

$$\sup_{\|\beta - \beta_n\| \le \nu_n} n^{1/2} \left| \int_{(-\infty,c_0)} G^{\beta_n}(u)\{dK^{\beta_n}(u) - dK^{\beta}(u)\} \right| \to_p 0. \tag{7.12}$$

But

$$\begin{aligned} G^{\beta}(t) - G^{\beta_n}(t) &= \sum_{i=1}^{n} \Delta_i 1_{\{\varepsilon_i^{\beta} > t\}} s(\varepsilon_i^{\beta})\{1 - \mathbb{F}_n^{\beta}(\varepsilon_i^{\beta}-)\} \\ &\quad - \sum_{i=1}^{n} \Delta_i 1_{\{\varepsilon_i^{\beta_n} > t\}} s(\varepsilon_i^{\beta_n})\{1 - \mathbb{F}_n^{\beta_n}(\varepsilon_i^{\beta_n}-)\} \\ &= \sum_{i=1}^{n} \Delta_i 1_{\{\varepsilon_i^{\beta_n} > t\}} \big[s(\varepsilon_i^{\beta})\{1 - \mathbb{F}_n^{\beta}(\varepsilon_i^{\beta}-)\} \\ &\qquad\qquad - s(\varepsilon_i^{\beta_n})\{1 - \mathbb{F}_n^{\beta_n}(\varepsilon_i^{\beta_n}-)\}\big] \\ &\quad + \sum_{i=1}^{n} \Delta_i 1_{\{\varepsilon_i^{\beta} > t \ge \varepsilon_i^{\beta_n}\}} s(\varepsilon_i^{\beta})\{1 - \mathbb{F}_n^{\beta}(\varepsilon_i^{\beta}-)\} \\ &\quad - \sum_{i=1}^{n} \Delta_i 1_{\{\varepsilon_i^{\beta_n} > t \ge \varepsilon_i^{\beta}\}} s(\varepsilon_i^{\beta})\{1 - \mathbb{F}_n^{\beta}(\varepsilon_i^{\beta}-)\} \\ &\to_p 0, \end{aligned}$$

by (7.8), (A.3) and the bound on the modulus of continuity of $\mathbb{H}_u^{\beta_n}$. This proves (7.11). By (7.7), (7.12) holds if

$$\sup_{t \le c_0,\ |\xi| \le \nu_n} n^{1/2}|G^{\beta_n}(t+\xi) - G^{\beta_n}(t)| \to_p 0. \tag{7.13}$$

The empirical processes argument and (A.1) show that (7.13) holds. Hence (7.12) and therefore (7.10) hold as well.

Finally, (7.7) and (7.10) prove the tightness of (7.5) and the lemma follows. □

PROOF OF THEOREM 5.1. (i) Define

$$\tilde{\Psi}_n(\beta_n; s) = n^{-1/2} \sum_{i=1}^{n} \int 1_{\{\zeta_i^{\beta_n} \ge u\}}\{Z_i - D^{\beta_n}(u)\}\, d\mathbb{Q}_s\big(u; \varepsilon_i^{\beta_n}|F_{\beta_n}\big).$$

Then $\Psi_n(\beta_n; s) = \tilde{\Psi}_n(\beta_n; s) + o_p(1)$ by Proposition 5.2. But $\tilde{\Psi}_n(\beta_0; s)$ is a normalized sum of i.i.d. random variables. Hence $\Psi_n(\beta_0; s)$ is asymptotically normal. Since $\mathbb{Q}_S(\cdot; \varepsilon_1^{\beta_0}|F_0)$ is a martingale and $\varepsilon_1^{\beta_0}$ is independent of $\zeta_1^{\beta_0}$ and Z_1 we obtain

$$\begin{aligned} &E\Big\{1_{\{\zeta_1^{\beta_0} \ge u\}}\{Z_1 - D^{\beta_0}(u)\}\, d\mathbb{Q}_s\big(u; \varepsilon_1^{\beta_0}|F_{\beta_0}\big)\Big\} \\ &\quad = E\Big\{1_{\{\zeta_1^{\beta_0} \ge u\}}\{Z_1 - D^{\beta_0}(u)\}\Big\} E\Big\{d\mathbb{Q}_s\big(u; \varepsilon_1^{\beta_0}|F_{\beta_0}\big)\Big\} \\ &\quad = 0 \end{aligned}$$

[recall that $D^{\beta_0}(u) = E(Z|C - \beta_0^T Z \ge u)$] and

$$\mathrm{Var}\bigg\{\int 1_{\{\zeta_1^{\beta_0} \ge u\}}\{Z_1 - D^{\beta_0}(u)\}\, d\mathbb{Q}_s\big(u; \varepsilon_1^{\beta_0}|F_{\beta_0}\big)\bigg\} = V.$$

Part (i) follows.

(ii) Now define

$$\overline{\Psi}_n(\beta_n; s) = n^{-1/2} \sum_{i=1}^{n} \int 1_{\{\zeta_i^{\beta_n} \ge u\}}\{Z_i - D^{\beta_n}(u)\}\, d\mathbb{Q}_s\big(u; \varepsilon_i^{\beta_n}|F_0\big).$$

We claim that $\overline{\Psi}_n(\beta_n; s) = \tilde{\Psi}_n(\beta_n; s) + o_p(1)$. Considering Lemma 7.2, we need only to check the expectations. But for some function $h(\cdot)$,

$$\begin{aligned} E\big\{\overline{\Psi}_n(\beta_n; s) - \tilde{\Psi}_n(\beta_n; s)\big\} &= E\bigg\{n^{-1/2} \sum_{i=1}^{n} \int 1_{\{\zeta_i^{\beta_n} \wedge \varepsilon_i^{\beta_n} \ge u\}}\{Z_i - D^{\beta_n}(u)\} h(u)\, du\bigg\} \\ &= 0, \end{aligned}$$

by the definition of $D^{\beta_n}(\cdot)$.

By Lemma 7.3 and the dominated convergence theorem

$$\mathrm{Var}\big\{\overline{\Psi}_n(\beta_n; s) - \overline{\Psi}_n(\beta_0; s)\big\} \to 0.$$

Hence part (ii) will follow if the expectations are proper. Now let

$$\chi(X; \beta_n) = \int 1_{\{\zeta^{\beta_n} \ge u\}}\{Z - D^{\beta_n}(u)\}\, d\mathbb{Q}_s\big(u; \varepsilon^{\beta_n}|F_0\big)$$

and let E_{β_n} be the expectation under the measure with density $f_0(y -$

$\beta_n^T z)h(z, c)$. Then $E_{\beta_n}\chi(X, \beta_n) = 0$. Hence

$$E_{\beta_0}\chi(X, \beta_n) = \int\int \chi(y, c, z; \beta_n) f_0(y - \beta_0^T z)\, dy\, h(z, c)\, d\mu$$

$$= \int\int \chi(y, c, z; \beta_n)\left[f_0(y - \beta_0^T z) - f_0(y - \beta_n^T z)\right] dy\, h(z, c)\, d\mu.$$

Taking the limit as $\beta_n \to \beta_0$ and recalling (A.3), we obtain

$$n^{1/2}\left\{ E_{\beta_0}\chi(X; \beta_n) - \operatorname{cov}\left\{\chi(X; \beta_0), Z^T \frac{f_0'(\varepsilon)}{f_0(\varepsilon)}\right\}(\beta_n - \beta_0)\right\} \to 0.$$

(iii) Follows immediately from the first two parts.

(iv) Follows from part (iii) and the information calculations done in Ritov and Wellner (1988) or Bickel, Klaassen, Ritov and Wellner (1989). □

Acknowledgments. Without the encouragement, advice and suggestions of Peter J. Bickel and Jon A. Wellner this revision could not have been written. The author wants to thank Richard D. Gill for a careful reading and many valuable suggestions.

REFERENCES

BICKEL, P. J., KLAASSEN, C. A. J., RITOV, Y. and WELLNER, J. A. (1989). *Efficient and Adaptive Estimation in Semiparametric Models*. Forthcoming Monograph. John Hopkins Univ. Press, Baltimore.

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

EFRON, B. and JOHNSTONE, I. M. (1990). Fisher's information in terms of the hazard rate. *Ann. Statist.* **18** 38–62.

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*, Math Centre Tracts 124. Mathematisch Centrum, Amsterdam.

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

HOPKINS, L. A. (1984). Adaptive maximum likelihood estimation of regression parameters with censored survival. Ph.D. dissertation, Univ. California, Berkeley.

JAMES, I. R. and SMITH, P. J. (1984). Consistency results for linear regression with censored data. *Ann. Statist.* **12** 590–600.

KALBFLEISCH, J. D. and PRENTICE, R. L. (1981). *The Statistical Analysis of Failure Time Data*. Wiley, New York.

KOUL, H., SUSARLA, V. and VAN RAYZIN, J. (1981). Regression analysis with randomly right censored data. *Ann. Statist.* **81** 1276–1288.

LAWLESS, J. F. (1982). *Statistical Models and Methods for Lifetime Data*. Wiley, New York.

LENGLART, E. (1977). Relations de domination entre deux processus. *Ann. Inst. H. Poincaré Sect. B (N.S.)* **13** 171–179.

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

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

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

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

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

REBOLLEDO, R. (1980). Central limit theorem for local martingales. *Z. Wahrsch. Verw. Gebiete* **51** 269–286.

RITOV, Y. (1984). Efficient and unibased estimation in nonparametric linear regression with censored data. Unpublished manuscript.

RITOV, Y. (1986). Efficient estimation in a linear regression with censored data. Unpublished manuscript.

RITOV, Y. and WELLNER, J. A. (1988). Censoring, martingales and the Cox model. *Contemporary Mathematics. Statistical Inference for Stochastic Processes* (N. U. Prabhu, ed.) **80** 191–220. Amer. Math. Soc.

SHORACK, G. R. and WELLNER, J. A. (1986). *Empirical Processes with Application to Statistics.* Wiley, New York.

TSIATIS, A. 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
THE HEBREW UNIVERSITY OF JERUSALEM
JERUSALEM 91905
ISRAEL