The Annals of Statistics
1996, Vol. 24, No. 6, 2720–2743

LINEAR REGRESSION WITH DOUBLY CENSORED DATA

BY CUN-HUI ZHANG[1] AND XIN LI

Rutgers University

Linear regression with doubly censored responses is considered. Buckley–James–Ritov-type estimators are proposed. Semiparametric information and projective scores are discussed. An expansion of the estimating equations is obtained under fairly general assumptions. Sufficient conditions are given for the asymptotic consistency and normality of the estimators.

1. Introduction. Consider the problem of estimating the slope in the simple linear model

$$Y_i = \beta X_i + \epsilon_i, \qquad i = 1, \dots, n, \tag{1.1}$$

where ϵ_i are independent identically distributed (iid) random variables with an unknown common survival function S_ϵ, and $X_1, \dots, X_n$ are (random or degenerate) design variables or covariates. When (Y_i, X_i) are completely observable, the least squares method is commonly used to estimate the unknown β. However, in biometry, engineering and other applications, the responses Y_i are often not completely observable due to censoring, truncation or other forms of sampling bias. Among such linear regression problems, the right-censoring case (with possible left truncation) appears to be the best understood one, which has been investigated by Buckley and James (1979), James and Smith (1984), Koul, Susarla and Van Ryzin (1981), Lai and Ying (1991), Miller and Halpern (1982), Prentice (1978), Ritov (1990), Tsiatis (1990) and Ying (1993) among others. In most other incomplete data models, we no longer have explicit expressions for nonparametric maximum likelihood estimates of S_ϵ or a natural martingale structure as in the right-censoring case, and different methods have to be used to analyze estimates of β in (1.1). Alternatively to (1.1), one may also consider the proportional hazards model of Cox (1972), but similar difficulties also arise when right censorship does not describe the observation scheme.

In this paper we consider linear regression when the response variables Y_i are subject to double censoring. Suppose vectors (Z_i, δ_i, X_i) are observed instead of (Y_i, X_i) in (1.1), where, for some (possibly degenerate) censoring

Received October 1994; revised April 1996.

[1]Research partially supported by the National Security Agency and the Army Research Office.

AMS 1991 *subject classifications*. Primary 62G05, 62G30; secondary 62E20, 62G20.

Key words and phrases. Censored data, estimating equation, *M*-estimator, asymptotic linearity, consistency, normality, information operator, efficiency, survival function, weak convergence.

variables $-\infty \le V_i \le U_i \le \infty$,

$$Z_i = \max(\min(Y_i, U_i), V_i),$$

$$\delta_i = \begin{cases} 1, & \text{if } V_i < Z_i = Y_i \le U_i, \\ 2, & \text{if } Z_i = U_i < Y_i, \\ 3, & \text{if } Z_i = V_i \ge Y_i, \end{cases} \tag{1.2}$$

such that ϵ_i is independent of (X_i, U_i, V_i). This is called double censoring, as Y_i is censored from the right- and left-hand sides when $\delta_i = 2$ and $\delta_i = 3$, respectively. The usual right-censoring model is a special case of (1.2) with $V_i = -\infty$ for all i. Gehan (1965), Peto (1973) and Turnbull (1974) gave examples in which double censoring might arise in medical and other applications. A real doubly censored data set was considered in Leiderman, Babu, Kagia, Kraemer and Leiderman (1973). For the case of known $\beta = 0$, asymptotic properties of nonparametric likelihood estimators of S_ϵ were considered in Tsai and Crowley (1985), Chang and Yang (1987), Chang (1990), Gu and Zhang (1993) and van der Laan (1993). We shall derive estimating equations which extend those of the Buckley–James type in Ritov (1990) from the right-censoring case to the double-censoring case in Section 2. In Section 3 we discuss the semiparametric information for the estimation of β, projective score functions and related operators. Under the compactness condition on the support of a score function, an expansion of our estimating equations is given in Section 4, which is analogous to the results of Ritov (1990) and Lai and Ying (1991). Asymptotic consistency and normality of the estimators are obtained in Section 5. Section 6 contains some discussion, including an alternative smoothing method of estimating β, the choice and estimation of nearly efficient score functions and the estimation of the error distribution. The basic results in this paper can be easily extended to the case of multidimensional β as in the right-censoring case.

We shall consider stochastic processes in the Banach spaces

$$D = \Big\{ h(t) \colon \lim_{s \to t+} h(s) = h(t),\ t \ge -\infty,\ \lim_{s \to t-} h(s) \text{ exists},\ t \le \infty \Big\},$$

$$D_o = \Big\{ h \in D \colon \lim_{t \to \pm\infty} h(t) = 0 \Big\},$$

$$D^3 = \left\{ h(t, j) \colon h(\cdot, j) \in D,\ j = 1, 2, 3,\ \sum_{j=1}^{3} h(t, j) \in D_o \right\},$$

all equipped with the supreme norm $\|h\| = \sup_x |h(x)|$ for functions of any vector x [e.g., $x = t$, $x = (t, j)$]. Convergence in distribution (weak convergence) here is always defined in the sense of Hoffmann-Jorgensen (1984) as described in Dudley (1985).

2. Estimating equations for β. In this section we provide estimating equations for β under (1.1) and (1.2). Our derivation is analogous to that of Ritov (1990), who considered the right-censoring case, but the estimating equations here are given in the form of linear operators acting on stochastic processes, which facilitates our calculations and greatly shortens the expressions.

To motivate our estimators, let us assume the existence of the error density $f_\epsilon = -S'_\epsilon$. The log-likelihood function for the data (Z_i, δ_i, X_i) is

$$\sum_{i=1}^{n}\left[\delta_{i,1}\log f_\epsilon(Z_{b,i}) + \delta_{i,2}\log S_\epsilon(Z_{b,i}) + \delta_{i,3}\log\{1 - S_\epsilon(Z_{b,i})\}\right],$$

where $\delta_{i,j} = I\{\delta_i = j\}$ and $Z_{b,i} = Z_i - bX_i$. As in Ritov (1990), a score function is obtained by differentiating the log-likelihood with respect to b, centering the X_i at $\overline{X}_n$ and then dividing each term by the sample size n, which can be written by calculus and algebra as

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}_n\right)\Bigg[\delta_{i,1}\phi_\epsilon(Z_{b,i}) - \delta_{i,2}\frac{\int_{z>Z_{b,i}}\phi_\epsilon(z)S_\epsilon(dz)}{S_\epsilon(Z_{b,i})}&\\ - \delta_{i,3}\frac{\int_{z\le Z_{b,i}}\phi_\epsilon(z)S_\epsilon(dz)}{1 - S_\epsilon(Z_{b,i})}\Bigg]&,\end{aligned}\tag{2.1}$$

where $\phi_\epsilon = -f'_\epsilon/f_\epsilon$. Since ϕ_ϵ and S_ϵ are unknown, we cannot estimate β by directly setting (2.1) to 0. In the right-censoring case, Buckley and James (1979) proposed replacing ϕ_ϵ by $\phi_{N(0,1)}(t) = t$ and S_ϵ by its product-limit estimator, while Ritov (1990) considered a predetermined $\phi(\cdot)$ satisfying some general conditions. We shall extend their estimating equations to the double-censoring case.

For $k = 0, 1, 2$, $j = 1, 2, 3$ and real numbers t, define

$$\hat{Q}^{(k)}_{b,n} = \hat{Q}^{(k)}_{b,n}(t, j) = n^{-1}\sum_{i=1}^{n}\left(X_i - \overline{X}_n\right)^k I\{Z_i - bX_i \ge t,\ \delta_i = j\}.\tag{2.2}$$

Here, $\hat{Q}^{(k)}_{b,n}$ are considered to be functions as well as signed measures which put mass $-(X_i - \overline{X}_n)^k/n$ at $(Z_i - bX_i, \delta_i)$, $1 \le i \le n$. For survival functions S, define linear operators B_S acting on bounded Borel functions $h = h(z, j)$ by

$$\begin{aligned}(B_S h)(t) = \sum_{j=1}^{3} h(t, j) - \int_{z\le t}\frac{S(t)}{S(z)}h(dz, 2)&\\ + \int_{z>t}\frac{1 - S(t)}{1 - S(z)}h(dz, 3)&,\end{aligned}\tag{2.3}$$

where $\int_{z\le t}\{S(t)/S(z)\}h(dz, 2) = 0$ when $S(t) = 0$, $\int_{z>t}\{(1 - S(t))/(1 - S(z))\}h(dz, 3) = 0$ when $S(t) = 1$ and integration by parts is used when $h(z, j)$ are not of bounded variation in z. Set

$$B_S(t|z, j) = I\{z > t\} + I\{j = 2,\ z \le t\}\frac{S(t)}{S(z)} - I\{j = 3,\ z > t\}\frac{1 - S(t)}{1 - S(z)},$$

which is the conditional survival function $P\{\epsilon_i > t | Z_{\beta,i} = z,\ \delta_i = j\}$ when $S = S_\epsilon$. By (2.3), $(B_S h)(t) = -\int B_S(t|\cdot)\, dh$ when $\sum_{j=1}^3 h(\infty, j) = 0$, so that

$$E\left[n^{-1}\sum_{i=1}^n I\{\epsilon_i > t\}\middle|\text{data}\right] = -\int B_{S_\epsilon}(t|\cdot)\, d\hat{Q}^{(0)}_{\beta,n} = \left(B_{S_\epsilon}\hat{Q}^{(0)}_{\beta,n}\right)(t) \tag{2.4}$$

and B_S can be viewed as a score operator for the estimation of S_ϵ. Furthermore, (2.1) can be written as

$$\frac{-1}{n}\sum_{i=1}^n \left(X_i - \overline{X}_n\right)\int \phi_\epsilon(t) B_{S_\epsilon}(dt|Z_{b,i}, \delta_i) = \int\int \phi_\epsilon(t) B_{S_\epsilon}(dt|\cdot)\, d\hat{Q}^{(1)}_{b,n}.$$

Since $B_{S_\epsilon}(\infty|\cdot) = 0$ and $-\int B_{S_\epsilon}(-\infty|\cdot)\, d\hat{Q}^{(1)}_{b,n} = \Sigma(X_i - \overline{X}_n)/n = 0$, we find through integrating by parts that (2.1) can be further written as

$$-\int\int B_{S_\epsilon}(t|\cdot)\, d\hat{Q}^{(1)}_{b,n}\phi_\epsilon(dt) = \int \left(B_{S_\epsilon}\hat{Q}^{(1)}_{b,n}\right)(t)\,\phi_\epsilon(dt)$$

subject to regularity conditions. Substituting ϕ_ϵ by a fixed score function ϕ, we may estimate β by

$$\begin{aligned} &\xi_n\left(\hat{\beta}_n\right) = o(n^{-1/2}),\\ &\xi_n(b) = \xi_n(b;\phi) = \int \left(B_{\hat{S}_{b,n}}\hat{Q}^{(1)}_{b,n}\right)(t)\, d\phi(t), \end{aligned} \tag{2.5}$$

where $\hat{S}_{b,n}$ satisfies the self-consistency equation $\hat{S}_{b,n} = B_{\hat{S}_{b,n}}\hat{Q}^{(0)}_{b,n}$ in (2.6) below [cf., e.g., Tsai and Crowley (1985) and Gu and Zhang (1993)]. This is called a Buckley–James–Ritov-type estimating equation, since it becomes that of Ritov (1990) when $P\{V_i = -\infty\} = 1$ and that of Buckley and James (1979) when, in addition, $\phi(t) = t$. The $\hat{\beta}_n$ in (2.5) is also called an M-estimator.

In general, the random function $\xi_n(b)$ is neither monotone nor continuous, and, for a given measure $Q^{(0)}$ of the same form as $\hat{Q}^{(0)}_{b,n}$ or its exception, the self-consistency equation

$$\begin{aligned} S(t) &= B_S Q^{(0)}(t)\\ &= \sum_{j=1}^3 Q^{(0)}(t,j) - \int_{z\le t}\frac{S(t)}{S(z)}Q^{(0)}(dz,2)\\ &\quad + \int_{z>t}\frac{1-S(t)}{1-S(z)}Q^{(0)}(dz,3) \end{aligned} \tag{2.6}$$

may not have a unique solution [cf., e.g., Gu and Zhang (1993), page 612]. The nonuniqueness of (2.6) can always be handled by choosing the NPMLE with the same support as $\sum_{j=1}^3 Q^{(0)}(dz,j)$. It will be shown in Section 4 that, under certain regularity conditions, all consistent solutions of (2.5) are asymptotically equivalent up to $o(n^{-1/2})$. In practice, the estimator $\hat{\beta}_n$ in (2.5) could be the location of the infimum of $|\xi_n(b)|$ or a zero-crossing of the process $\xi_n(\cdot)$ [cf. (5.3) in Section 5], and the nonuniqueness of $\hat{\beta}_n$ does not seem to be a serious problem. The solutions of (2.6) can be computed by the EM algorithm, so that the evaluation of $\xi_n(b)$ is relatively easy at each point b. The uniqueness of the NPMLE under double censoring was discussed by Zhan and Wellner (1995).

3. Information and projective scores. A general asymptotic theory of semiparametric estimation was considered by Bickel, Klaassen, Ritov and Wellner (1993), with a detailed discussion of the right-censoring case. See also Stein (1956), Bickel (1982), Begun, Hall, Huang and Wellner (1983), Ritov and Wellner (1988), and Ritov (1990) among others. In this section we study the minimum Fisher information, score functions and related linear operators, for the estimation of β with unknown error distribution. We shall consider throughout the section a single random vector (Z, δ, X), related to another vector (ϵ, X, U, V) in the same manner as in (1.1) and (1.2). The results here certainly apply when a sample from (Z, δ, X) is observed.

Let P be a fixed probability measure under which β is the true regression coefficient in (1.1) and f_ϵ the true error density. Suppose $E|X| < \infty$. Consider a subparametric family of our model such that

$$\begin{aligned}(d/dt)P_\theta\{Y - \theta X \le t|X, U, V\} &= f_\theta(t)\\ &= (f_\epsilon\{1 - (\theta-\beta)\phi\})(t - (\theta - \beta)EX),\end{aligned}$$

where $P = P_\beta$, $\int\phi(t)f_\epsilon(t)\,dt = 0$ and $\|\phi'\| + \|\phi\| < \infty$. The log-likelihood function for the parameter θ based on (Z, δ, X) is

$$\begin{aligned}l(\theta) = {}& I_{\{\delta=1\}}\log f_\theta(Z - \theta X) + I_{\{\delta=2\}}\log S_\theta(Z - \theta X)\\ &+ I_{\{\delta=3\}}\log F_\theta(Z - \theta X),\end{aligned}$$

where F_θ and S_θ are the distribution and survival functions of f_θ. As in (2.1) and (2.4), the score function for this parametric family at $\theta = \beta$ is

$$l'_\phi = \frac{\partial l(\theta)}{\partial\theta}\bigg|_{\theta=\beta} = E^*\big[(X - EX)\phi_\epsilon(\epsilon) - \phi(\epsilon)\big],$$

where ϕ_ϵ is as in (2.1) and $E^*[\cdot] = E[\cdot|Z, \delta, X]$ is the conditional expectation given the observable variables. Suppose $E(l'_0)^2 = E((X - EX)E^*\phi_\epsilon(\epsilon))^2 < \infty$. Let

$$\mathscr{G}_k = \Big\{\phi\colon E\big[(X - EX)^k\phi(\epsilon)\big] = 0,\ E\big[(X - EX)^k E^*\phi(\epsilon)\big]^2 < \infty\Big\}.$$

The minimum Fisher information for the estimation of θ at $\theta = \beta$ is

$$I_* = I_*(\beta, S_\epsilon) = \inf\Big\{E(l'_\phi)^2\colon \phi \in \mathscr{G}_0\Big\}. \tag{3.1}$$

Define $S^{(*)} = \{S_\epsilon, S_U^{(k)}, S_V^{(k)}, k = 0, 1, 2\}$ by

$$\begin{aligned}S_U^{(k)}(t) &= E(X - EX)^k I\{U - \beta X > t\},\\ S_V^{(k)}(t) &= E(X - EX)^k I\{V - \beta X > t\}.\end{aligned} \tag{3.2}$$

For $\phi \in \mathscr{G}_1$, define the projective scores

$$
\begin{aligned}
\rho(z,j,x;\phi) &= \rho(z,j,x;\phi,S^{(*)}) \\
&= (x-EX)E[\phi(\epsilon)|Z-\beta X=z,\ \delta=j] \\
&\quad - \psi(z,j;\phi,S^{(*)}),
\end{aligned}
\tag{3.3}
$$

where $\psi(Z-\beta X,\delta;\phi)=\psi(Z-\beta X,\delta;\phi,S^{(*)})$ is the projection of $(X-EX)E^*\phi(\epsilon)$ to $\mathscr{H}^*=\overline{\{E^*\phi(\epsilon):\phi\in\mathscr{G}_0\}}$. Here, the closure and projection are both in the sense of $L_2(P)$. Note that $\psi(z,j;c)=0$ for all constants c. By (2.4), $E^*\phi$ is a function of $(Z-\beta X,\delta)$. It will be shown in Theorem 3.1 that the projection $\psi(\cdot;\phi)$ depends on P only through $S^{(*)}$. The minimum in (3.1) is achieved at the efficient score

$$
\begin{aligned}
l'_* &= \rho(Z-\beta X,\delta,X;\phi_\epsilon) \\
&= (X-EX)E^*\phi_\epsilon(\epsilon) - \psi(Z-\beta X,\delta;\phi_\epsilon).
\end{aligned}
\tag{3.4}
$$

In Sections 4 and 5 we provide sufficient conditions under which the asymptotic distribution of the $\hat{\beta}_n$ in (2.5) is expressed in terms of the covariance

$$
\begin{aligned}
A(\phi_1,\phi_2) &= A(\phi_1,\phi_2;S^{(*)}) \\
&= E\{\rho(Z-\beta X,\delta,X;\phi_1)\rho(Z-\beta X,\delta,X;\phi_2)\}.
\end{aligned}
\tag{3.5}
$$

In the rest of this section, we consider certain families of linear operators closely related to the projective scores in (3.3). For any functions μ and ν of bounded variation on the real line and survival function S, define linear operators $K_{\mu,\nu}$ by

$$
(K_{\mu,\nu}h)(t) = K_{\mu,\nu}(t)h(t), \qquad K_{\mu,\nu}(t)=\mu(t)-\nu(t)
\tag{3.6}
$$

for all Borel functions h, and define $R_{S,\mu,\nu}$ by

$$
\begin{aligned}
(R_{S,\mu,\nu}h)(t) &= (K_{\mu,\nu}h)(t) - \int_{z\le t}\frac{S(t)}{S(z)}h(z)\mu(dz) \\
&\quad - \int_{z>t}\frac{1-S(t)}{1-S(z)}h(z)\nu(dz)
\end{aligned}
\tag{3.7}
$$

in the sense of Lebesgue–Stieltjes integration. Suppose $E|X|^k<\infty$. Set

$$
\begin{aligned}
K^{(k)} &= K_{S_U^{(k)},S_V^{(k)}}, \qquad K=K^{(0)}, \\
R^{(k)} &= R_{S_\epsilon,S_U^{(k)},S_V^{(k)}}, \qquad R=R^{(0)}.
\end{aligned}
\tag{3.8}
$$

For $h\in D_o$, define functionals

$$
\begin{aligned}
A_h(\phi) &= A_h(\phi;S^{(*)}) \\
&= -\int\phi(t)\big[\{R^{(2)}-R^{(1)}R^{-1}R^{(1)}\}h\big](dt).
\end{aligned}
\tag{3.9}
$$

Let $b_{z,j}(t)$ be the centered $B_{S_\epsilon}(t|\cdot)$ in (2.4), given by

$$
\begin{aligned}
b_{z,j}(t) &= P\{\epsilon > t | Z - \beta X = z,\ \delta = j\} - S_\epsilon(t) \\
&= B_{S_\epsilon}(t|z,j) - S_\epsilon(t).
\end{aligned} \tag{3.10}
$$

The following theorem describes the connection of the operators $R^{(k)}$ to the projection $\psi(z,j;\phi)$ and therefore to the projective scores in (3.3) and their covariance (3.5).

THEOREM 3.1. *Suppose $E|X|^2 < \infty$. Let $\psi(z,j;\phi)$ be the projection in* (3.3), *R and $R^{(k)}$ be given by* (3.8), *$b_{z,j}$ by* (3.10), *$h(t) = \int_{z>t}\phi_h(z)F_\epsilon(dz)$ for some $\phi_h \in \mathscr{G}_0$ and $A(\cdot,\cdot)$ and $A_h(\cdot)$ be given by* (3.5) *and* (3.9), *respectively.*

(i) *If $E\{(X-EX)^k E^*\phi(\epsilon)\}^2 < \infty$, then*

$$
-\int \phi(t)(R^{(k)}h)(dt) = E\big\{(X-EX)^k E^*\phi_h(\epsilon)E^*\phi(\epsilon)\big\}. \tag{3.11}
$$

In particular, $(R^{(k)}h)(t) = E(X-EX)^k E^*\phi_h(\epsilon)E^*I\{\epsilon > t\}$. Consequently, the projection $\psi(\cdot;\phi)$ and the projective score $\rho(\cdot;\phi)$ in (3.3) *depend on P only through $S^{(*)}$.*

(ii) *Suppose $K(t-) > 0$ for all real t. Then the operator R is invertible, and, for $\phi \in \mathscr{G}_1$,*

$$
\psi(z,j;\phi,S^{(*)}) = -\int \phi(t)\big\{R^{(1)}R^{-1}b_{z,j}\big\}(dt), \tag{3.12}
$$

$$
\begin{aligned}
&-\int \phi(t)\{R^{(1)}R^{-1}R^{(1)}h\}(dt) \\
&\qquad = E\psi(Z-\beta X,\delta;\phi)\psi(Z-\beta X,\delta;\phi_h).
\end{aligned} \tag{3.13}
$$

Consequently, $A_h(\phi) = A(\phi,\phi_h)$ if $\phi \in \mathscr{G}_1$ and $E(X-EX)^2|E^*\phi(\epsilon)E^*\phi_h(\epsilon)| < \infty$.

REMARK 3.1. The operator R is crucial in the analysis of nonparametric estimation of S_ϵ with known $\theta = \beta$ in Gu and Zhang (1993) and is the $D_o \to D_o$ version of the information operator for that problem. The right-hand side of (3.11) with $k = 0$ can be written as $\langle I\phi_h, \phi\rangle$, where I is the $L_2(P) \to L_2(P)$ version of R. It seems that the $D_o \to D_o$ version of R is convenient to use here in view of our results in Section 4, where additional discussion about $R^{(k)}$ can be found.

REMARK 3.2. In the right-censoring case $K^{(k)} = S_U^{(k)}$,

$$
d\frac{(R^{(k)}h)(t)}{S_\epsilon(t)} = K^{(k)}(t-)d\frac{h(t)}{S_\epsilon(t)},
$$

so that we have explicit formulas for R^{-1}, $\psi(\cdot;\phi)$, $\rho(\cdot;\phi)$, $A_h(\phi)$ and $A(\phi_1,\phi_2)$. When there is no censoring $A(\phi_1,\phi_2) = \mathrm{Var}(X)\mathrm{Cov}(\phi_1(\epsilon),\phi_2(\epsilon))$.

REMARK 3.3. Integrating by parts can be applied in (3.11)–(3.13) when ϕ is of bounded variation. In particular, for the $b_{z,j}$ in (3.10) and $\phi_t(\epsilon) = I\{\epsilon > t\} - S_\epsilon(t)$,

$$\rho\big(z, j, x; \phi_t, S^{(*)}\big) = (x - EX) b_{z,j}(t) - \big(R^{(1)} R^{-1} b_{z,j}\big)(t), \tag{3.14}$$

$$A_h(\phi_t) = A_h\big(\phi_t; S^{(*)}\big) = (R^{(2)} h)(t) - (R^{(1)} R^{-1} R^{(1)} h)(t), \tag{3.15}$$

and, for $E\phi(\epsilon) = 0$,

$$\begin{aligned} \rho(z, j, x; \phi) &= \int \rho(z, j, x; \phi_t) \phi(dt), \\ A_h(\phi) &= \int A_h(\phi_t) \phi(dt). \end{aligned} \tag{3.16}$$

PROOF. (i) Let $E^*_{z,j}[\cdot] = E[\cdot \mid Z - \beta X = z,\ \delta = j]$, $F = F_\epsilon$ and $S = S_\epsilon$. By definition, $E^*_{z,2}\phi_h(\epsilon) = h(z)/S(z)$ and $E^*_{z,3}\phi_h(\epsilon) = -h(z)/F(z)$. By (3.7), $(R^{(k)}h)(dt)$ equals

$$K(t-) h(dt) - \int_{t>z} S(dt)(E^*_{z,2}\phi_h) S_U^{(k)}(dz) + \int_{t \le z} F(dt)(E^*_{z,3}\phi_h) S_V^{(k)}(dz).$$

Let $Q^{(k)}(z, j) = E(X - EX)^k I\{Z - \beta X > z,\ \delta = j\}$. Since $S(z)S_U^{(k)}(dz) = Q^{(k)}(dz, 2)$ and $F(z)S_V^{(k)}(dz) = Q^{(k)}(dz, 3)$ by (1.2),

$$\begin{aligned} -\int \phi(t)(R^{(k)}h)(dt) &= \int \phi(t) K^{(k)}(t-) \phi_h(t) F(dt) \\ &\quad + \int\int_{t>z} \phi(t) S(dt)(E^*_{z,2}\phi_h) S_U^{(k)}(dz) \\ &\quad - \int\int_{t \le z} \phi(t) F(dt)(E^*_{z,3}\phi_h) S_V^{(k)}(dz) \\ &= E\big\{(X - EX)^k E^*\phi(\epsilon) E^*\phi_h(\epsilon)\big\}. \end{aligned}$$

The projection depends only on $S^{(*)}$ since the norm $\{E|E^*(X\phi_1 + \phi_2)|^2\}^{1/2}$ does by (3.11).

(ii) The invertibility of R is given in Gu and Zhang (1993). Let $\psi(z, j)$ be the right-hand side of (3.12). Since $E[S(t)\phi_h(\epsilon)] = 0$, $(Rh)(t) = E\{b_{Z-\beta X,\delta}(t) E^*\phi_h(\epsilon)\}$ by (i), so that

$$E\big\{R^{(1)} R^{-1} b_{Z-\beta X,\delta} E^*\phi_h(\epsilon)\big\} = R^{(1)} R^{-1} R h = R^{(1)} h.$$

Integrating with $\phi(t)$, we find

$$E\psi(Z - \beta X, \delta) E^*\phi_h = E\psi\big(Z - \beta X, \delta; \phi, S^{(*)}\big) E^*\phi_h$$

for all $\phi_h \in \mathscr{G}_0$ by (i). In particular, for the ϕ_t in (3.14) and (3.15), $E^*\phi_t = b_{Z-\beta X,\delta}(t)$ and

$$\begin{aligned} &R^{(1)} R^{-1} E b_{Z-\beta X,\delta} \psi(Z - \beta X, \delta) \\ &\quad = R^{(1)} R^{-1} E b_{Z-\beta X,\delta} \psi\big(Z - \beta X, \delta; \phi, S^{(*)}\big), \end{aligned}$$

which implies

$$E\psi^2(Z-\beta X,\delta)=E\psi(Z-\beta X,\delta;\phi,S^{(*)})\psi(Z-\beta X,\delta).$$

Therefore, (3.12) holds as $\psi(\cdot;\phi,S^{(*)})$ is the projection to $\mathscr{H}^*$. For (3.13), we simply apply $R^{(1)}R^{-1}$ to both sides of $R^{(1)}h=Eb_{Z-\beta X,\delta}\psi(Z-\beta X,\delta;\phi_h,S^{(*)})$. The final assertion follows from (3.3), (3.11) with $k=2$ and (3.13). □

4. Asymptotic linearity of $\xi_n(b)$. The main result of this section is Theorem 4.2, which asserts that, under certain regularity conditions, the process $B_{\hat{S}_{b,n}}\hat{Q}^{(1)}_{b,n}$ in (2.5) is asymptotically linear in b on compact intervals as random elements in D_o. When (Z_i,δ_i,X_i) are iid random vectors, it implies

$$\begin{aligned}\xi_n(b;\phi)=&-(b-\beta)A_{f_\epsilon}(\phi)\\&+n^{-1}\sum_{i=1}^n\rho(Z_i-\beta X_i,\delta_i,X_i;\phi-E\phi(\epsilon))+\cdots\\&\qquad\text{as }(b,n)\to(\beta,\infty),\end{aligned}$$

where $A_h(\phi)$ is given by (3.9) and $\rho(\cdot;\phi)$ is the projective score in (3.3). Unless otherwise stated, our results cover the "double array" case where the observations $(Z_{i,n},\delta_{i,n},X_{i,n})$ depend on both i and n, as we are interested in convergence in probability and in distribution. In the sequel we shall always assume that $\epsilon_{i,n}$ is independent of $(X_{i,n},U_{i,n},V_{i,n})$ for each $1\le i\le n$, and that $\epsilon_{1,n},\dots,\epsilon_{n,n}$ are identically distributed with a common survival function which may depend on n. The independence among the vectors $(Z_{i,n},\delta_{i,n},X_{i,n})$ is not explicitly assumed so that our methods may still be useful in sequential designs.

Define $S^{(*)}_{\beta,n}=\{S_{\beta,n},S^{(k)}_{U,n},S^{(k)}_{V,n},\ k=0,1,2\}$ by $S_{\beta,n}(t)=P\{\epsilon_{i,n}>t\}$,

$$S^{(k)}_{U,n}=S^{(k)}_{U,n}(t)=E\left[n^{-1}\sum_{i=1}^n\left(X_{i,n}-E\overline{X}_n\right)^kI\{U_{i,n}-\beta X_{i,n}>t\}\right],\tag{4.1}$$

$$S^{(k)}_{V,n}=S^{(k)}_{V,n}(t)=E\left[n^{-1}\sum_{i=1}^n\left(X_{i,n}-E\overline{X}_n\right)^kI\{V_{i,n}-\beta X_{i,n}>t\}\right],\tag{4.2}$$

with $\overline{X}_n=\sum_{i=1}^nX_{i,n}/n$, and, for $k=0,1,2$, $j=1,2,3$ and real t, define

$$\begin{aligned}Q^{(k)}_{b,n}&=Q^{(k)}_{b,n}(t,j)\\&=E\left[n^{-1}\sum_{i=1}^n\left(X_{i,n}-E\overline{X}_n\right)^kI\{Z_{i,n}-bX_{i,n}>t,\ \delta_{i,n}=j\}\right].\end{aligned}\tag{4.3}$$

Suppose there exists a random vector (Z,δ,X), related to (ϵ,X,U,V) via (1.1) and (1.2) as in Section 3, such that, for some $\varepsilon_0>0$,

$$\lim_{n\to\infty}\|Q^{(k)}_{b,n}-Q^{(k)}_b\|=0,\qquad k=0,1,2,|b-\beta|<\varepsilon_0,\tag{4.4}$$

where $Q_b^{(k)}(t,j) = E(X-EX)^k I\{Z - bX > t,\ \delta = j\}$. Let $S^{(*)} = \{S_\beta, S_U^{(k)}, S_V^{(k)}, k = 0,1,2\}$ be given by (3.2) with $S_\beta = S_\epsilon$. Suppose further that

$$(4.5)\qquad S_{U,n}^{(0)}(t-) > S_{V,n}^{(0)}(t-) \quad \text{and} \quad S_U^{(0)}(t-) > S_V^{(0)}(t-) \qquad \forall\, t.$$

PROPOSITION 4.1. *Let B_S be given by* (2.3). *Then*

$$(4.6)\qquad \left(B_{S_{\beta,n}} Q_{\beta,n}^{(k)}\right)(t) = S_{\beta,n}(t) E\left[n^{-1}\sum_{i=1}^{n}\left(X_{i,n} - E\overline{X}_n\right)^k\right] \qquad \forall\, t.$$

In particular, for $k=1$, $B_{S_{\beta,n}} Q_{\beta,n}^{(1)} = 0$. Furthermore, if (4.5) *holds, then $S_{\beta,n}$ and S_β are, respectively, the unique solutions of* (2.6) *with $Q^{(0)} = Q_{\beta,n}^{(0)}$ and $Q^{(0)} = Q_\beta^{(0)}$.*

PROOF. Set $q_{i,n} = q_{i,n}(t,j) = I\{Z_{i,n} - \beta X_{i,n} > t,\ \delta_{i,n} = j\}$. By (4.3), (2.3) and (2.4),

$$\begin{aligned} B_{S_{\beta,n}} Q_{\beta,n}^{(k)} &= E\left[n^{-1}\sum_{i=1}^{n}\left(X_{i,n} - E\overline{X}_n\right)^k B_{S_{\beta,n}} q_{i,n}\right] \\ &= E\left[n^{-1}\sum_{i=1}^{n}\left(X_{i,n} - E\overline{X}_n\right)^k P\{\epsilon_{i,n} > t | Z_{i,n}, \delta_{i,n}, X_{i,n}\}\right]. \end{aligned}$$

This implies (4.6) due to the independence of $\epsilon_{i,n}$ and $X_{i,n}$. For $k=1$, we obtain $B_{S_{\beta,n}} Q_{\beta,n}^{(1)} = 0$, as $E\sum_{i=1}^n (X_i - E\overline{X}_n) = 0$. The uniqueness of (2.6) follows from Gu and Zhang (1993), proof of Theorem 1, or Lemma 4.6 below. □

Let $K_{\mu,\nu}$, $R_{S,\mu,\nu}$, K, $K^{(k)}$, R and $R^{(k)}$ be the linear operators in (3.6) and (3.8). Set

$$(4.7)\qquad K_n^{(k)} = K_{S_{U,n}^{(k)}, S_{V,n}^{(k)}}, \qquad R_{S,n}^{(k)} = R_{S, S_{U,n}^{(k)}, S_{V,n}^{(k)}}, \qquad \hat{R}_{b,n}^{(k)} = R_{\hat{S}_{b,n},n}^{(k)},$$

with $K_n = K_n^{(0)}$, $R_{S,n} = R_{S,n}^{(0)}$ and $\hat{R}_{b,n} = \hat{R}_{b,n}^{(0)}$. For the expansion of $B_{\hat{S}_{b,n}} \hat{Q}_{b,n}^{(1)}$, we assume

$$(4.8)\qquad \begin{aligned} &\int_{z\le t} \frac{dS_{U,n}^{(0)}(z)}{K_n(z)} \to \int_{z\le t} \frac{dS_U^{(0)}(z)}{K(z)}, \\ &\int_{z>t} \frac{dS_{V,n}^{(0)}(z)}{K_n(z)} \to \int_{z>t} \frac{dS_V^{(0)}(z)}{K(z)} \end{aligned}$$

as $n \to \infty$, with the limits being finite for all t, and there exist functions $h^{(k)} = h^{(k)}(t,j) \in D^3$, $k = 0,1$, such that, as $(b,n) \to (\beta, \infty)$,

$$(4.9)\qquad \|Q_{b,n}^{(k)} - Q_{\beta,n}^{(k)} - (b-\beta)h^{(k)}\| = o\left(|b-\beta| + n^{-1/2}\right).$$

Define $W_{b,n}^{(k)} = W_n(b,t,j,k)$ by $W_{b,n}^{(k)} = \sqrt{n}\,(\hat{Q}_{b,n}^{(k)} - Q_{b,n}^{(k)})$, $k = 0,1$. We shall further assume

$$(4.10)\qquad W_{b,n}^{(k)}(t,j) = W_n(b,t,j,k) \to_{\mathscr{D}} W(b,t,j,k)$$

in the supreme norm, as processes of $(b, t, j, k) \in (\beta - \varepsilon_0, \beta + \varepsilon_0) \times (-\infty, \infty) \times \{1, 2, 3\} \times \{0, 1\}$ for some $\varepsilon_0 > 0$ such that

$$\lim_{\varepsilon'' \to 0+} \lim_{\varepsilon' \to 0+} P\left\{ \sup_{|b-\beta| \le \varepsilon'} \sup_{t, j, k} |W(b, t, j, k) - W(\beta, t, j, k)| > \varepsilon'' \right\} = 0. \tag{4.11}$$

With the $\rho(\cdot; \cdot, \cdot)$ given by (3.3) and $S^{(*)}_{\beta, n}$ by (4.1) and (4.2), define

$$\begin{aligned}\eta_n(\phi) &= \eta_n\big(\phi; S^{(*)}_{\beta, n}\big) \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \rho\big(Z_{i,n} - \beta X_{i,n}, \delta_{i,n}, X_{i,n}; \phi, S^{(*)}_{\beta, n}\big),\end{aligned} \tag{4.12}$$

and, for the processes $W^{(k)}(t, j) = W(\beta, t, j, k)$ in (4.10), define

$$\eta(\phi) = \int \eta(z) \phi(dz), \qquad \eta = B_{S_\beta} W^{(1)} - R^{(1)} R^{-1} B_{S_\beta} W^{(0)}. \tag{4.13}$$

We shall consider expansions of stochastic processes such that the remainder is dominated by $|\varepsilon_n(b)|(|b - \beta| + n^{-1/2})$ satisfying

$$\lim_{\varepsilon'' \to 0+} \lim_{\varepsilon' \to 0+} \limsup_{n \to \infty} P\left\{ \sup_{|b-\beta| \le \varepsilon'} |\varepsilon_n(b)| > \varepsilon'' \right\} = 0. \tag{4.14}$$

THEOREM 4.2. *Let $S^{(*)}_\beta$ be given by* (3.2), *$A_h(\cdot) = A_h(\cdot; S^{(*)})$ by* (3.9), *$\eta_n(\cdot)$ by* (4.12) *and $\eta(\cdot)$ by* (4.13). *Suppose $P\{\sup_{1 \le i \le n} |X_{i,n}| > M_1\} = 0$ for all n and some $M_1 < \infty$, $0 < S_\beta(t) < 1$ for all t, $\overline{X}_n - E\overline{X}_n = o_P(1)$ and the functions $f_\beta = -S'_\beta$ and $Q^{(0)}_\beta(\cdot, j)$, $j = 1, 2, 3$, are all continuous. Suppose* (4.4), (4.5) *and* (4.8)–(4.11) *hold.*

(i) *Let $B_{\hat{S}_{b,n}} \hat{Q}^{(1)}_{b,n}$ be as in* (2.5). *Then*

$$\begin{aligned}B_{\hat{S}_{b,n}} \hat{Q}^{(1)}_{b,n}(t) &= n^{-1/2} \eta_n(\phi_{t,n}) - (b - \beta) A_{f_\beta}(\phi_t) \\ &\quad + \tilde{\varepsilon}_n(b, t)\big(|b - \beta| + n^{-1/2}\big)\end{aligned} \tag{4.15}$$

such that

$$\eta_n(\phi_{t,n}) I_{\{-M_1^* \le t < M_1^*\}} \to_{\mathscr{D}} \eta(\phi_t) I_{\{-M_1^* \le t < M_1^*\}}$$

as processes in D_o and (4.14) *holds for $\varepsilon_n(b) = \sup_{|t| \le M_1^*} |\tilde{\varepsilon}_n(b, t)|$ for every $M_1^* < \infty$, where $\phi_{t,n}(\cdot) = I\{\cdot > t\} - S_{\beta,n}(t)$ and $\phi_t(\cdot) = I\{\cdot > t\} - S_\beta(t)$.*

(ii) *Let $A(\cdot, \cdot) = A(\cdot, \cdot; S^{(*)})$ be given by* (3.5) *and ϕ and ϕ_j be functions of bounded variation. Then $A_{f_\beta}(\phi) = A_{f_\beta}(\phi - E\phi(\epsilon)) = A(\phi, \phi_\epsilon)$ if $\phi_\epsilon = -f'_\beta / f_\beta \in L_2(f_\beta)$, and $\eta(\phi)$ are normal variables with $E\eta(\phi) = 0$ and $\mathrm{Cov}(\eta(\phi_1), \eta(\phi_2)) = A(\phi_1 - E\phi_1(\epsilon), \phi_2 - E\phi_2(\epsilon))$ if $(Z_{i,n}, \delta_{i,n}, X_{i,n})$, $1 \le i \le n$, are independent vectors.*

REMARK 4.1. Suppose the observations are iid vectors from (Z, δ, X) of Section 3 such that (ϵ, X) are independent of each other and of (U, V). It will be shown in the proof of Theorem 5.2 that conditions (4.9)–(4.11) hold when X, U and V all have uniformly continuous marginal densities.

REMARK 4.2. Clearly, by (4.12) and (3.16), $\eta_n(\phi + \int \phi \, dS_{\beta, n}) = \int \eta_n(\phi_{t,n})\phi(dt)$ and $A_{f_\beta}(\phi - E\phi(\epsilon)) = \int A_{f_\beta}(\phi_t)\phi(dt)$. By (4.13), $\eta(\phi) = \int \eta(\phi_t)\phi(dt)$.

A real number $b = b_0$ is a zero-crossing of a function $h(b)$ if

$$\lim_{\Delta \to 0+} \inf\{h(b) : |b - b_0| \le \Delta\} \le 0$$

and

$$\lim_{\Delta \to 0+} \sup\{h(b) : |b - b_0| \le \Delta\} \ge 0.$$

COROLLARY 4.3. *Suppose the conditions of Theorem* 4.2(*i*) *hold and* $\phi_\epsilon = -f'_\beta/f_\beta \in L_2(f_\beta)$. *Let* $\xi_n(b; \phi)$ *be given by* (2.5) *for some* ϕ *with* $\int |\phi(dt)| < \infty$ *and* $\phi(dt) = 0$ *outside a compact interval. Then*

$$\xi_n(b; \phi) = n^{-1/2}\eta_n(\phi - E_n\phi) - (b - \beta)A(\phi, \phi_\epsilon) + \varepsilon_n(b)(|b - \beta| + n^{-1/2}) \tag{4.16}$$

such that $\eta_n(\phi - E_n\phi) \to_{\mathscr{D}} \eta(\phi)$ *and* $\varepsilon_n(b)$ *satisfies* (4.14), *where* $E_n\phi = -\int \phi \, dS_{\beta, n}$. *Furthermore, if* $A(\phi, \phi_\epsilon) \neq 0$, *then there exist* $\hat{\beta}_n$ *and events* Ω_n *with* $P\{\Omega_n\} \to 1$ *such that* $\hat{\beta}_n$ *is a zero-crossing of* $\xi_n(b; \phi)$ *with* $\xi_n(\hat{\beta}_n; \phi) = o(n^{-1/2})$ *on* Ω_n, *and* $\sqrt{n}(\hat{\beta}_n - \beta) \to_{\mathscr{D}} \eta(\phi)/A(\phi, \phi_\epsilon)$. *In addition, if* $\tilde{\beta}_n$ *is a sequence of zero-crossings of* $\xi_n(b; \phi)$ *converging in probability to* β, *then* $\sqrt{n}(\hat{\beta}_n - \tilde{\beta}_n) \to 0$ *in probability.*

REMARK 4.3. If ϕ is close to ϕ_ϵ in $L_2(f_\beta)$, then $A(\phi, \phi_\epsilon) > 0$ by (3.5). For technical reasons, our conditions on ϕ may exclude the efficient scores ϕ_ϵ for some densities [e.g., $\phi_\epsilon(t) = t$ for $\epsilon \sim N(\mu, \sigma^2)$], but they allow a choice of ϕ close to ϕ_ϵ. See Remarks 5.1 and 5.3 and Section 6 for further discussion. In particular, Theorem 6.1 allows the use of an estimated ϕ.

REMARK 4.4. In the right-censoring case, Ritov (1990) obtained the expansion (4.16) for $b - \beta = O(n^{-1/2})$, Lai and Ying (1991) covered the neighborhoods $b - \beta = O(n^{-\lambda})$ with some $\lambda > 0$ for $\phi(t) = t$, while Ying (1993) established an expansion over $b - \beta = o(1)$ in the sense of (4.14) for rank-based estimating functions.

The proof of Theorem 4.2 is based on strong continuity and invertibility of the information operators and the following identity.

PROPOSITION 4.4. *Let* S_1 *and* S_2 *be survival functions and* $q = q(t, j)$ *be such that the functions* $\mu(t) = -\int_{z > t}\{S_1(z)\}^{-1}q(dz, 2)$ *and* $\nu(t) = -\int_{z > t}\{1 - S_1(z)\}^{-1}q(dz, 3)$ *are both of bounded variation. Then*

$$(B_{S_2} - B_{S_1})q - \mu(-\infty)(S_2 - S_1) = -R_{S_2, \mu, \nu}(S_2 - S_1).$$

REMARK 4.5. For $q = Q^{(k)}$ and $S_1 = S_\epsilon$, Proposition 4.4 describes the relationship between the information operators and the expectation of score differences. For $k = 0$ and $S_2 = \hat{S}$, this identity is (2.11) of Gu and Zhang (1993). See also (2.5) of Vardi and Zhang (1992) and (21) of Tsai and Zhang (1995) for analogous key identities in some other models with incomplete data. van der Laan (1993) considered a similar identity for linear parameters in general convex models.

Throughout the sequel we shall denote by M any finite positive constant, by $\varepsilon_n(b)$ any processes of b satisfying (4.14), by $\varepsilon_n(b, x)$ any processes of (b, x) satisfying $\sup_x |\varepsilon_n(b, x)| = \varepsilon_n(b)$, where the second variable x can be any vector [e.g., $\varepsilon_n(b, t, j)$ for $x = (t, j)$ and $\varepsilon_n(b, t, j, k)$ for $x = (t, j, k)$]. The definitions of M, $\varepsilon_n(b)$ and $\varepsilon_n(b, \cdot)$ may change from one place to another. We need three lemmas to describe properties of the score and information operators.

LEMMA 4.5 (Uniqueness and continuity of the self-consistency equation). *Suppose the second part of* (4.5) *holds. Then the equation* $S = B_S Q^{(0)}$ *in* (2.6) *has a unique solution* $S = S_\beta$ *for* $Q^{(0)} = Q^{(0)}_\beta$ *and the solution of* (2.6) *is continuous at* $Q^{(0)}_\beta$ *in the sense that* $\|S - S_\beta\| \to 0$ *as* $\|Q^{(0)} - Q^{(0)}_\beta\| \to 0$.

LEMMA 4.6 [Properties of the score operators B_S in (2.3)]. *Suppose the second part of* (4.5) *holds,* $\|S_n - S_\beta\| = o(1)$ *and* $0 < S_\beta < 1$ *for all* t. *Let* $h_n = h_n(t, j) \in D^3$ *and* $h = h(t, j) \in D^3$.

(i) $\|B_S\| \le 5$ *for all survival functions* S.
(ii) *If* $\|h_n - h\| = o(1)$, *then* $\|B_{S_n} h_n - B_{S_\beta} h\| = o(1)$.

LEMMA 4.7 [Properties of the information operators $R^{(k)}_{S,n}$ in (4.7)]. *Suppose the conditions of Theorem* 4.2 *hold. Let* K_n *and* K *be given by* (4.7) *and* (3.8), *respectively, and let* $h_n = h_n(t)$ *and* $h = h(t)$ *be members of* D_o. *Suppose* $\|h_n - h\| = o(1)$ *and* $\|S_n - S_\beta\| = o(1)$.

(i) *The operator* $R_{S,n}$ *is invertible for all survival functions* S.
(ii) *The operator* KR^{-1} *is bounded from* D_o *to* D_o *and*

$$\lim_{n\to\infty} \|K_n R^{-1}_{S_n, n} h_n - KR^{-1} h\| = 0.$$

(iii) *Let* $M > 0$. *The operator* $I_{[-M, M)} R^{(1)} R^{-1}$ *is bounded from* D_o *to* D_o *and*

$$\lim_{n\to\infty} \|I_{[-M, M)}\{R^{(1)}_{S_n, n} R^{-1}_{S_n, n} h_n - R^{(1)} R^{-1} h\}\| = 0.$$

Lemma 4.5 follows from the proof of Theorem 1 in Gu and Zhang (1993). The proof of Proposition 4.4 is essentially algebra and is omitted. Lemma 4.6 is step 3 of the proof of Theorem 2 in Gu and Zhang (1993), Section 4. Lemma 4.7 is proved in the Appendix. Notice that conditions (4.9)–(4.11) are actually not needed for Lemma 4.7.

PROOF OF THEOREM 4.2. By (4.1) and (4.2), $S^{(k)}_{U,n}(-\infty) = 1 - k$ for $k = 0, 1$, so that, by (4.7) and Proposition 4.4 with $q = Q^{(k)}_{\beta,n}$, $\mu = S^{(k)}_{U,n}$ and $\nu = S^{(k)}_{V,n}$,

$$\left(B_{\hat{S}_{b,n}} - B_{S_{\beta,n}}\right)Q^{(k)}_{\beta,n} = -\hat{R}^{(k)}_{b,n}\left(\hat{S}_{b,n} - S_{\beta,n}\right) + (1-k)\left(\hat{S}_{b,n} - S_{\beta,n}\right).$$

By Proposition 4.1, $B_{S_{\beta,n}}Q^{(1)}_{\beta,n} = 0$. It follows that

$$B_{\hat{S}_{b,n}}\hat{Q}^{(1)}_{b,n} = B_{\hat{S}_{b,n}}\left(\hat{Q}^{(1)}_{b,n} - Q^{(1)}_{\beta,n}\right) - \hat{R}^{(1)}_{b,n}\left(\hat{S}_{b,n} - S_{\beta,n}\right).$$

Similarly by the self-consistency (2.6) of $\hat{S}_{b,n}$ and $S_{\beta,n}$, we obtain

$$\hat{R}_{b,n}\left(\hat{S}_{b,n} - S_{\beta,n}\right) = B_{\hat{S}_{b,n}}\left(\hat{Q}^{(0)}_{b,n} - Q^{(0)}_{\beta,n}\right). \tag{4.17}$$

Since $\hat{R}_{b,n}$ is invertible by Lemma 4.7(i),

$$B_{\hat{S}_{b,n}}\hat{Q}^{(1)}_{b,n} = B_{\hat{S}_{b,n}}\left(\hat{Q}^{(1)}_{b,n} - Q^{(1)}_{\beta,n}\right) - \hat{R}^{(1)}_{b,n}\left[\hat{R}_{b,n}\right]^{-1}B_{\hat{S}_{b,n}}\left(\hat{Q}^{(0)}_{b,n} - Q^{(0)}_{\beta,n}\right).$$

Since $\hat{Q}^{(k)}_{b,n} - Q^{(k)}_{\beta,n} = n^{-1/2}W^{(k)}_{b,n} + Q^{(k)}_{b,n} - Q^{(k)}_{\beta,n}$, we have

$$B_{\hat{S}_{b,n}}\hat{Q}^{(1)}_{b,n} = -(b-\beta)\hat{A}_{b,n} + n^{-1/2}\hat{\eta}_{b,n}, \tag{4.18}$$

where $\hat{A}_{b,n} = \hat{A}_{b,n}(t)$ and $\hat{\eta}_{b,n} = \hat{\eta}_{b,n}(t)$ are given by

$$-(b-\beta)\hat{A}_{b,n} = B_{\hat{S}_{b,n}}\left(Q^{(1)}_{b,n} - Q^{(1)}_{\beta,n}\right) - \hat{R}^{(1)}_{b,n}\left[\hat{R}_{b,n}\right]^{-1}B_{\hat{S}_{b,n}}\left(Q^{(0)}_{b,n} - Q^{(0)}_{\beta,n}\right),$$

$$\hat{\eta}_{b,n} = B_{\hat{S}_{b,n}}W^{(1)}_{b,n} - \hat{R}^{(1)}_{b,n}\left[\hat{R}_{b,n}\right]^{-1}B_{\hat{S}_{b,n}}W^{(0)}_{b,n}.$$

Letting $n \to \infty$ first in (4.9), we find $\|Q^{(k)}_b - Q^{(k)}_\beta - (b-\beta)h^{(k)}\| = o(|b-\beta|)$ by (4.4), so that by direct computation

$$\begin{aligned} B_{S_\beta}h^{(k)} &= (\partial/\partial b)B_{S_\beta}Q^{(k)}_b|_{b=\beta} \\ &= -R^{(k+1)}_{S_\beta}f_\beta - EXR^{(k)}_{S_\beta}f_\beta, \end{aligned} \tag{4.19}$$

as the exchange between $\partial/\partial b$ and the integrations of $B_{S_\beta}Q^{(k)}_b$ in (2.3) is allowed by the continuity assumption. This and (3.15) imply

$$\begin{aligned} B_{S_\beta}h^{(1)} - R^{(1)}R^{-1}B_{S_\beta}h^{(0)} &= -\left\{R^{(2)}_{S_\beta}f_\beta - R^{(1)}R^{-1}R^{(1)}f_\beta\right\} \\ &= -A_{f_\beta}(\phi_t). \end{aligned} \tag{4.20}$$

Define $\tilde{\eta}_n = B_{S_{\beta,n}}W^{(1)}_{\beta,n} - R^{(1)}_{S_{\beta,n},n}R^{-1}_{S_{\beta,n},n}B_{S_{\beta,n}}W^{(0)}_{\beta,n}$ and $\tilde{W}^{(1)}_{\beta,n} = \tilde{W}^{(1)}_{\beta,n}(t,j)$ by

$$\tilde{W}^{(1)}_{\beta,n}(t,j) = n^{-1/2}\sum_{i=1}^{n}\left(X_{i,n} - E\bar{X}_n\right) \times\left[I\{Z_{i,n} - \beta X_{i,n} > t,\ \delta_{i,n} = j\} - Q^{(0)}_{\beta,n}(t,j)\right].$$

By (4.12), (2.3), (2.4) and (3.14),

$$\eta_n(\phi_{t,n}) = B_{S_{\beta,n}}\tilde{W}^{(1)}_{\beta,n} - R^{(1)}_{S_{\beta,n},n}R^{-1}_{S_{\beta,n},n}B_{S_{\beta,n}}W^{(0)}_{\beta,n}. \tag{4.21}$$

By (2.2), $W^{(1)}_{\beta,n} - \tilde W^{(1)}_{\beta,n} = (E\overline{X}_n - \overline{X}_n)W^{(0)}_{\beta,n} - \sqrt{n}\,Q^{(1)}_{\beta,n}$, so that, by (4.6) with $k=1$, Lemma 4.6(i), (4.10) and the condition $\overline{X}_n - E\overline{X}_n = o_P(1)$,

$$\begin{aligned}\|\eta_n(\phi_{t,n}) - \tilde\eta_n\| &= \big\|\big(E\overline{X}_n - \overline{X}_n\big)B_{S_\beta}W^{(0)}_{\beta,n} - \sqrt{n}\,B_{S_\beta}Q^{(1)}_{\beta,n}\big\| \\ &\le 5|E\overline{X}_n - \overline{X}_n|\,\|W^{(0)}_{\beta,n}\| \\ &= O_P(1)|E\overline{X}_n - \overline{X}_n| = o_P(1).\end{aligned}\tag{4.22}$$

It follows from (4.10), (4.13) and the continuous mapping theorem that

$$I_{[-M_1^*,M_1^*)}\Big\{B_{S_\beta}W^{(1)}_{\beta,n} - R^{(1)}R^{-1}B_{S_\beta}W^{(0)}_{\beta,n}\Big\} \to_{\hat{\mathscr{D}}} I_{[-M_1^*,M_1^*)}\eta$$

in D_o, as B_{S_β} and $I_{[-M_1^*,M_1^*)}R^{(1)}R^{-1}B_{S_\beta}$ are both bounded linear operators from $D^3 \to D_o$ by Lemmas 4.6(i) and 4.7(iii). This and (4.18), (4.22) and (4.20) imply that the expansion (4.15) is a consequence of

$$I_{[-M_1^*,M_1^*)}\hat\eta_{b,n} = I_{[-M_1^*,M_1^*)}\Big\{B_{S_\beta}W^{(1)}_{\beta,n} - R^{(1)}R^{-1}B_{S_\beta}W^{(0)}_{\beta,n}\Big\} + \varepsilon_n(b,t),\tag{4.23}$$

$$I_{[-M_1^*,M_1^*)}\tilde\eta_n = I_{[-M_1^*,M_1^*)}\Big\{B_{S_\beta}W^{(1)}_{\beta,n} - R^{(1)}R^{-1}B_{S_\beta}W^{(0)}_{\beta,n}\Big\} + \varepsilon_n(b,t),\tag{4.24}$$

$$-I_{[-M_1^*,M_1^*)}\hat A_{b,n} = I_{[-M_1^*,M_1^*)}\Big\{B_{S_\beta}h^{(1)} - R^{(1)}R^{-1}B_{S_\beta}h^{(0)}\Big\} + \varepsilon_n(b,t).\tag{4.25}$$

We shall only prove (4.23), as the proofs of (4.24) and (4.25) are similar and simpler in view of the definition of $\tilde\eta_n$ in the previous paragraph and (4.9).

By (4.10) and (4.11), $W^{(k)}_{b,n}(t,j) - W^{(k)}_{\beta,n}(t,j) = \varepsilon_n(b,t,j)$. This and (4.4) and (4.9) imply $\hat Q^{(0)}_{b,n} - Q^{(0)}_\beta = \varepsilon_n(b,t,j)$, so that $\hat S_{b,n} - S_\beta = \varepsilon_n(b,t)$ by Lemma 4.5. These and Lemmas 4.6(ii) and 4.7(iii) with $h_n = \varepsilon_n$ and $h=0$ imply

$$I_{[-M_1^*,M_1^*)}\hat\eta_{b,n} = I_{[-M_1^*,M_1^*)}\Big\{B_{\hat S_{b,n}}W^{(1)}_{\beta,n} - \hat R^{(1)}_{b,n}\hat R^{-1}_{b,n}B_{\hat S_{b,n}}W^{(0)}_{\beta,n}\Big\} + \varepsilon_n(b,t).\tag{4.26}$$

Since $W^{(k)}_{\beta,n} \to_{\hat{\mathscr{D}}} W(\beta,t,j,k)$ and $P\{W^{(k)}_{\beta,n} \in D^3\} = 1$, for a given $\varepsilon > 0$, there exist a compact set C_ε in D^3 and $\varepsilon'_n \to 0+$ such that $P\{W^{(k)}_{\beta,n} \notin C_\varepsilon(\varepsilon'_n)\} \le \varepsilon$ for $k=0,1$, where $C_\varepsilon(\varepsilon'_n) = \bigcup_{h'\in C_\varepsilon}\{h \in D^3: \|h-h'\| \le \varepsilon'_n\}$. By Lemmas 4.6(ii) and 4.7(iii) and the fact that $\hat S_{b,n} - S_\beta = \varepsilon_n(b,t)$, we have

$$\begin{aligned}\sup_{h,h'\in C_\varepsilon(\varepsilon'_n)}&\Big\|I_{[-M_1^*,M_1^*)}\Big\{\big(B_{\hat S_{b,n}} - B_{S_\beta}\big)h' - \Big(\hat R^{(1)}_{b,n}\hat R^{-1}_{b,n}B_{\hat S_{b,n}} - R^{(1)}R^{-1}B_{S_\beta}\Big)h\Big\}\Big\| \\ &= \varepsilon_n(b).\end{aligned}$$

This and (4.26) imply (4.23) on $\bigcap_{k=0}^1\{W^{(k)}_{\beta,n} \in C_\varepsilon(\varepsilon'_n)\}$ and therefore on the entire probability space as $P\{W^{(k)}_{\beta,n} \notin C_\varepsilon(\varepsilon'_n)\} \le \varepsilon$ for arbitrarily small $\varepsilon > 0$.

The formula $A_{f_\beta}(\phi) = A(\phi,\phi_\varepsilon)$ is in Theorem 3.1(ii). Suppose $(Z_{i,n},\delta_{i,n},X_{i,n})$ are independent vectors and ϕ is of bounded variation. Since $\int \cdot\, d\phi$ is a continuous functional on D_o, $\eta_n(\phi - E_n\phi) \to_{\hat{\mathscr{D}}} \eta(\phi)$ by (3.16), (4.12) and (4.13). By (4.12), $\eta_n(\phi - E_n\phi)/\sqrt{n}$ is an average of uniformly bounded independent random variables, so that $\eta(\phi)$ is normal and $\mathrm{Var}(\eta_n(\phi - E_n\phi)) \to \mathrm{Var}(\eta(\phi))$. By (2.4) and (3.10), $EX^k_{i,n}b_{Z_{i,n}-\beta X_{i,n},\delta_{i,n}}(t) = \mathrm{Cov}(X^k_{i,n}, I\{\epsilon_{i,n} > t\}) = 0$, which implies by (3.14) that $E\rho(Z_{i,n} - \beta X_{i,n},\delta_{i,n},X_{i,n};\phi - E_n\phi, S^{(*)}_{\beta,n}) = 0$, $1 \le i \le n$. Hence, by (4.12), (3.5) and the continuity of the operators,

$$\begin{aligned}\operatorname{Var}(\eta_n(\phi - E_n\phi)) &= \frac{1}{n}\sum_{i=1}^{n} E\rho^2\left(Z_{i,n} - \beta X_{i,n}, \delta_{i,n}, X_{i,n}; \phi - E_n\phi, S^{(*)}_{\beta,n}\right)\\ &= A\left(\phi - E_n\phi, \phi - E_n\phi; S^{(*)}_{\beta,n}\right)\\ &\to A(\phi - E\phi, \phi - E\phi; S^{(*)}).\end{aligned}$$

The covariance of $\eta(\phi)$ is determined by the variance as $\eta(\phi)$ is linear in ϕ. □

5. Consistency and asymptotic normality. In this section we consider asymptotic properties of estimators of β in (1.1) under double censorship (1.2). Sufficient conditions for asymptotic consistency are given in Theorem 5.1. Corollary 4.3 is used to obtain the asymptotic normality of $\hat{\beta}_n$ in Theorem 5.2 when $(Z_{i,n}, \delta_{i,n}, X_{i,n})$, $1 \le i \le n$, are independent random vectors.

Let $Q^{(k)}_{b,n}$ and $Q^{(k)}_b$ be as in (4.3) and (4.4), and let S_b be a solution of the self-consistency equation (2.6) with $Q^{(0)} = Q^{(0)}_b$. Suppose

$$(5.1)\qquad \sup_{b_1 \le b \le b_2} \|\hat{Q}^{(k)}_{b,n} - Q^{(k)}_b\| = o_P(1), \qquad k = 0, 1, \forall\, -\infty < b_1 < b_2 < \infty,$$

and

$$(5.2)\qquad K_b(t-) = \int_{z \ge t} \frac{Q^{(0)}_b(dz, 3)}{1 - S_b(z)} - \int_{z \ge t} \frac{Q^{(0)}_b(dz, 2)}{S_b(z)} > 0 \qquad \forall\, t, b.$$

Condition (5.2) implies the uniqueness of S_b and is a consequence of $Q^{(0)}_b(dz, 1) > 0$, as $Q^{(0)}_b(dz, 1) = K_b(z-)S_b(dz)$. Let $\tilde{\beta}_n$ be some preliminary (and possibly asymptotically inconsistent) estimates of β such that $\tilde{\beta}_n - \beta = O_P(1)$. For example, $\tilde{\beta}_n$ could be the naive least squares estimator based on those $(Z_{i,n}, X_{i,n})$ with $\delta_{i,n} = 1$, or the one given by (6.1) below. With the $\xi_n(b)$ in (2.5), define

$$(5.3)\qquad \hat{\beta}_n \in C(\xi_n), \qquad |\hat{\beta}_n - \tilde{\beta}_n| = \min\left\{|b - \tilde{\beta}_n| : b \in C(\xi_n)\right\},$$

where $C(h)$ is the (always closed) set of zero-crossings of h (cf. Corollary 4.3).

THEOREM 5.1. *Let $\hat{\beta}_n$ be given by* (5.3) *with $\xi_n(b) = \int B_{\hat{S}_{b,n}} \hat{Q}^{(1)}_{b,n}\, d\phi$ in* (2.5). *Suppose $E|X| < \infty$, $Q^{(0)}_b = Q^{(0)}_b(t, j)$ is continuous in (b, t) and* (5.1) *and* (5.2) *hold. Then*

$$(5.4)\qquad \sup_{b_1 \le b \le b_2} \|B_{\hat{S}_{b,n}} \hat{Q}^{(1)}_{b,n} - B_{S_b} Q^{(1)}_b\| = o_P(1) \qquad \forall\, -\infty < b_1 < b_2 < \infty,$$

and $B_{S_\beta} Q^{(1)}_b$ is a D_o-valued continuous function of b. Moreover, $P\{|\hat{\beta}_n - \beta| > \varepsilon\} \to 0$ for all $\varepsilon > 0$, provided that $\int |\phi(dt)| < \infty$ and

$$(5.5)\qquad \zeta(b_1)\zeta(b_2) < 0 \qquad \forall\, b_1 < \beta < b_2, \qquad \zeta(b) = \int \left(B_{S_\beta} Q^{(1)}_b\right)(t)\,\phi(dt).$$

REMARK 5.1. Suppose $(Z_{i,n}, \delta_{i,n}, X_{i,n})$ are independent vectors. Then (5.1) holds via Glivenko and Cantelli, if

$$\limsup_{n\to\infty} n^{-1} \sum_{i=1}^{n} E|X_{i,n}| I\{|X_{i,n}| > M\} = o(1)$$

as $M \to \infty$ and

$$\sup_{b_1 \le b \le b_2} \|Q_{b,n}^{(k)} - Q_b^{(k)}\| = o(1), \qquad k = 0, 1, \forall\ -\infty < b_1 < b_2 < \infty. \tag{5.6}$$

Condition (5.5) is analogous to (5.30) of Lai and Ying (1991) and the condition $m(b) \ne 0$ in $C_r \setminus \{\beta\}$ of Theorem 4(i) of Ying (1993). In the right-censoring case, Ying (1993) proved that his condition holds for certain ϕ if X is a zero–one variable or f_ϵ has increasing failure rate, and his method may also give $A(\phi, \phi_\epsilon) > 0$. It is not clear if (5.5) holds or $A(\phi, \phi_\epsilon) > 0$ in the case of double censoring under similar conditions. Further discussion of consistent estimation of β can be found in Section 6.

PROOF OF THEOREM 5.1. The asymptotic consistency of $\hat{\beta}_n$ follows easily from (5.4) and (5.5) as $|\zeta(b) - \xi_n(b)| \le \|B_{\hat{S}_{b,n}} \hat{Q}_{b,n}^{(1)} - B_{S_b} Q_b^{(1)}\| \int |\phi(dt)|$ and $B_{S_\beta} Q_\beta^{(1)} = 0$ by (4.6). Taking subsequences if necessary, we may assume (5.4) holds in the sense of almost sure convergence. Let b be a finite number and $b_n \to b$. It suffices to show $\|B_{\hat{S}_{b_n,n}} \hat{Q}_{b_n,n}^{(1)} - B_{S_b} Q_b^{(1)}\| \to 0$. By (5.4) and the boundedness of $\{b_n\}$, $\|\hat{Q}_{b_n,n}^{(k)} - Q_{b_n}^{(k)}\| \to 0$. Since $E|X| < \infty$ and $Q_b^{(0)}$ is decreasing in t and continuous in (b, t), $Q_b^{(k)}(\cdot, j)$ is continuous in b as elements of D for all j and $k = 0, 1$, so that $\|\hat{Q}_{b_n,n}^{(k)} - Q_b^{(k)}\| \to 0$. By Lemma 4.5 and (5.2), $\|\hat{S}_{b_n,n} - S_b\| \to 0$. Hence $\|B_{\hat{S}_{b_n,n}} \hat{Q}_{b_n,n}^{(1)} - B_{S_b} Q_b^{(1)}\| \to 0$ by Lemma 4.6(ii). $\square$

In the rest of this section, we shall assume $(Z_{i,n}, \delta_{i,n}, X_{i,n})$, $1 \le i \le n$, are independent vectors and the common survival function $S_{\beta,n} = S_\beta$ of $\epsilon_{i,n}$ does not depend on n.

THEOREM 5.2. *Let $\hat{\beta}_n$ be given by* (5.3) *with $\xi_n(b) = \xi_n(b; \phi)$ such that $\int |\phi(dt)| < \infty$ and $\phi(dt) = 0$ outside a compact interval. Let $A(\phi, \phi)$ and $\eta(\phi)$ be as in Theorem* 4.2. *Suppose $P\{\sup_{1\le i\le n} |X_{i,n}| > M_1\} = 0$ for all n and some $M_1 < \infty$, $0 < S_\beta(t) < 1$ for all t, $\phi_\epsilon = -f'_\beta/f_\beta \in L_2(f_\beta)$ and $A(\phi, \phi_\epsilon) \ne 0$. Suppose* (4.5), (4.8) *and* (5.6) *hold, and*

$$\lim_{\epsilon\to 0^+} \limsup_{n\to\infty} \sup_{|t-s|\le\varepsilon} \sup_{|b-\beta|\le\varepsilon} \frac{1}{n} \sum_{i=1}^{n} E|f_{j,i,n}(s + bX_{i,n}|X_{i,n}) - f_{j,i,n}(t + \beta X_{i,n}|X_{i,n})| = 0 \tag{5.7}$$

for $j = 2, 3$, where $f_{2,i,n}(t|X_{i,n})$ and $f_{3,i,n}(t|X_{i,n})$ are conditional densities of $U_{i,n}$ and $V_{i,n}$, respectively, given $X_{i,n}$, $1 \le i \le n$. If either (5.2) *and* (5.5) *hold or $\tilde{\beta}_n = \beta + o_P(1)$, then $\sqrt{n}(\hat{\beta}_n - \beta) \to_{\mathscr{D}} N(0, \sigma^2)$ with $\sigma^2 = A(\phi - E\phi, \phi - E\phi)/A^2(\phi, \phi_\epsilon)$, where $E\phi = -\int \phi\, dS_\beta$.*

REMARK 5.2. Suppose $(Z_{i,n}, \delta_{i,n}, X_{i,n})$ are iid vectors from (Z, δ, X) such that X and ϵ are independent of each other and of (U, V). Then (5.7) holds if U and V both have uniformly continuous marginal densities, and (4.8) can be replaced by the finiteness of $\int_{z \le t} K^{-1}(z) S_U^{(0)}(dz)$ and $\int_{z > t} K^{-1}(z) S_V^{(0)}(dz)$ for all t. Condition (5.7) is slightly stronger than (3.5) of Lai and Ying (1991), who essentially assumed the uniform boundedness of the densities of the censoring variables.

REMARK 5.3. For $\phi(t) = a_0 + a_1\phi_\epsilon(t)$ [e.g., $\phi(t) = I\{t > 0\}$ for double-exponential errors], the estimators in Theorem 5.4 are asymptotically semiparametric efficient, as $\sigma^2 = 1/A(\phi_\epsilon, \phi_\epsilon) = 1/I_*$ by (3.1), (3.4) and (3.5).

PROOF OF THEOREM 5.2. We shall verify the conditions of Corollary 4.3 and Theorem 5.1. Clearly, (5.6) implies (4.4), $\phi_\epsilon \in L_2(f_\beta)$ implies uniform continuity of f_β and (5.6) and the uniform boundedness of $X_{i,n}$ imply (5.1) (Remark 5.1). It suffices to show (4.9)–(4.11) and the continuity of $Q_b^{(k)}$,

Let $S_{j,i,n}(t, b) = \int_{z > t} f_{j,i,n}(z + bX_{i,n} | X_{i,n})\, dz$. By (4.3), $h_{b,n}^{(k)} = (\partial/\partial b) Q_{b,n}^{(k)}$ exist and

$$
\begin{aligned}
h_{b,n}^{(k)}(t,1) &= \frac{-1}{n}\sum_{i=1}^{n} E\Big[\big(X_{i,n} - E\overline{X}_n\big)^k f_\beta\big(t + (b-\beta)X_{i,n}\big) \\
&\qquad\qquad \times\{S_{2,i,n}(t,b) - S_{3,i,n}(t,b)\}\Big],\\
h_{b,n}^{(k)}(t,2) &= \frac{-1}{n}\sum_{i=1}^{n} E\Big[\big(X_{i,n} - E\overline{X}_n\big)^k f_{2,i,n}(t + bX_{i,n}|X_{i,n}) \\
&\qquad\qquad \times S_\beta\big(t + (b-\beta)X_{i,n}\big)\Big],\\
h_{b,n}^{(k)}(t,3) &= \frac{-1}{n}\sum_{i=1}^{n} E\Big[\big(X_{i,N} - E\overline{X}_n\big)^k f_{3,i,n}(t + bX_{i,n}|X_{i,n}) \\
&\qquad\qquad \times\big\{1 - S_\beta\big(t + (b-\beta)X_{i,n}\big)\big\}\Big].
\end{aligned}
$$

It follows from (5.7), the uniform continuity of f_β and the uniform boundedness of $X_{i,n}$ that $\|h_{b,n}^{(k)} - h_{\beta,n}^{(k)}\| = \varepsilon_n(b)$ and $h_{\beta,n}^{(k)}(t,j)$ are uniformly continuous in t. By (5.6) and the fact $S_{\beta,n} = S_\beta$, $S_{2,n} = En^{-1}\sum_i S_{2,i,n}(\cdot,\beta) \to S_U$ and $S_{3,n} = En^{-1}\sum_i S_{3,i,n}(\cdot,\beta) \to S_V$ uniformly. This and the uniform continuity of $(d/dt)S_{j,n}(t)$ imply $\sup_n |(d/dt)S_{j,n}(t)| \to 0$ as $t \to \pm\infty$, which then implies $\sup_n |h_{\beta,n}^{(k)}(t,j)| \to 0$ as $t \to \pm\infty$ for all j and k, in view of the uniform continuity of f_β and the expressions of $h_{\beta,n}^{(k)}$ above. Thus $\overline{\{h_{\beta,n}^{(k)},\ n \ge 1\}}$ are compact in D^3. Letting $\|h_{\beta,n}^{(k)} - h^{(k)}\| \to 0$ along a subsequence of n, we find $h^{(k)} = (\partial/\partial b)Q_b^{(k)}|_{b=\beta}$ by the argument leading to (4.20). This uniqueness of the limit point of $\{h_{\beta,n}^{(k)},\ n \ge 1\}$ gives (4.9). Since the $X_{i,n}$'s are bounded, condition (4.10) follows from standard results in empirical process theory, especially the equicontinuity of the empirical processes with uniformly bounded VC classes of functions [cf. e.g., Giné and Zinn (1986) and Pollard (1984), page 157]. Condition (4.11) follows from (4.10) and the continuity of the covariance function of the limiting process $W(b, t, j, k)$. The proof of the continuity of $Q_b^{(k)}$ is omitted as it is simpler than that of (4.9). □

6. Discussion and some additional results. In this section we discuss the asymptotic consistency and the choice and estimation of a "good" score function ϕ. The estimation of the error distribution is also considered.

As mentioned in Remark 4.1, the expansion (4.15) holds under explicit conditions on the marginal distributions of ϵ, X, U and V in the iid case. However, condition (5.5) is assumed in Theorem 5.1 for asymptotic consistency, $A(\phi, \phi_\epsilon) \neq 0$ is assumed in Corollary 4.3 and Theorem 5.2 for the asymptotic normality and $\phi = a_0 + a_1\phi_\epsilon$ is needed for the asymptotic efficiency by Remark 5.3. As indicated by the results of Miller and Halpern (1982), these may not be serious problems in practice when a suitable preassigned ϕ is chosen. For example, if there is reason to believe $S_\epsilon \approx N(0, \sigma_\epsilon^2)$, we may use $\phi(t) = \phi(t; M) = \max(-M, \min(M, t))$ with some $M > 2\sigma_\epsilon$.

Under strong explicit regularity conditions, smoothing methods might be used to produce (preliminary) estimators $\tilde{\beta}_n$ of β which are asymptotically consistent. For example, we may estimate β by solving

$$\sum_{i=1}^{n} \left(X_i - \overline{X}_n\right) \int B_{\hat{S}(\cdot|X_i)}(t|Z_i, \delta_i)\, d_t\phi(t - bX_i) = 0, \tag{6.1}$$

with a monotone ϕ and a reasonable decreasing estimator $\hat{S}(t|x)$ of $P\{Z > t|X = x\}$. Since $\hat{S}(t|x)$ does not depend on b, the left-hand side of (6.1) is monotone in b, so that the asymptotic consistency is a direct consequence of a Glivenko–Cantelli-type law of large numbers for the estimating equation. One way of estimating $P\{Z > t|X = x\}$ is solving (2.6) with

$$Q^{(0)} = \sum_i k_n(X_i - x) I\{Z_i > t,\, \delta_i = j\} \Big/ \sum_i k_n(X_i - x) \tag{6.2}$$

for some positive kernels $k_n(\cdot)$. In view of Lemma 4.5, we believe that the asymptotic consistency of the $\hat{S}(t|x)$ based on (6.2) and therefore (6.1) can be obtained under some conditional versions of (4.5) given X.

Since the projection $\psi(\cdot; \phi, S^{(*)})$ is a function of $Z - \beta X$ and δ, by (3.5) and (3.8),

$$E\Big[\operatorname{Var}(X|Z - \beta X, \delta)\{E^*\phi(\epsilon)\}^2\Big] \le A(\phi, \phi) \le \operatorname{Var}(X) E\phi^2(\epsilon). \tag{6.3}$$

Since $A(\phi_1, \phi_2)$ is an inner product, by the Cauchy–Schwarz inequality,

$$A(\phi, \phi_\epsilon) \ge A(\phi_\epsilon, \phi_\epsilon)\left\{1 - \sqrt{\frac{A(\phi - \phi_\epsilon, \phi - \phi_\epsilon)}{A(\phi_\epsilon, \phi_\epsilon)}}\right\}. \tag{6.4}$$

We may use (6.3) and (6.4) to bound $A(\phi, \phi)$, $A(\phi, \phi_\epsilon)$ and the relative efficiency $(\sigma^2 I_*)^{-1}$ of the estimators in (2.5) and (5.3), where σ^2 is as in Theorem 5.2 and I_* is given by (3.1).

If we wish to use a score function $\phi(t) = \phi(t; \beta, f_\epsilon, \dots)$ depending on unknown parameters, we may also estimate it. We do not pursue this in Sections 4 and 5 for the sake of space as well as the lack of optimal estimation and selection of ϕ under our regularity conditions. For finite positive M_1^* and M_2^*, let $\Phi(M_1^*, M_2^*)$ be the collection of functions ϕ such that $\phi(dt) = 0$ for $|t| \geq M_1^*$ and $\int|\phi(dt)| \leq M_2^*$. Let $\hat{\phi}_{b,n}$ be such that, for every real t and $\varepsilon'' > 0$,

$$\lim_{\varepsilon' \to 0+} \limsup_{n\to\infty} P\Biggl\{ \bigcup_{|b-\beta|\leq \varepsilon'} \Bigl(\bigl\{ \hat{\phi}_{b,n} \notin \Phi(M_1^*, M_2^*) \bigr\} \cup \bigl\{ |\hat{\phi}_{b,n}(t) - \phi(t)| > \varepsilon'' \bigr\} \Bigr) \Biggr\} = 0. \tag{6.5}$$

THEOREM 6.1. *Suppose* (6.5) *holds. Then the conclusions of Theorem* 4.2, *Corollary* 4.3 *and Theorem* 5.2 *hold under the respective conditions when* $\xi_n(b) = \xi_n(b; \phi)$ *is replaced by* $\hat{\xi}_n(b) = \int (B_{\hat{S}_{b,n}} \hat{Q}_{b,n}^{(1)})(t) \hat{\phi}_{b,n}(dt)$.

REMARK 6.1. Let $0 < \varepsilon_0 < 1$ be a small given number. Since $\bigcup_{c>0} \Phi(c, c)$ is dense in $L_2(f_\beta)$, there exist estimates $\hat{M}_1$, $\hat{M}_2$, $\hat{\phi}_{b,n}$ and $\hat{\phi}_{b,n}^*$ such that $\|\hat{\phi}_{b,n}^* - \phi_\varepsilon\|_2 \leq \varepsilon_n(b)$, $\hat{A}_{b,n}(\hat{\phi}_{b,n} - \hat{\phi}_{b,n}^*, \hat{\phi}_{b,n} - \hat{\phi}_{b,n}^*) \leq \varepsilon_0^2 \hat{A}_{b,n}(\hat{\phi}_{b,n}^*, \hat{\phi}_{b,n}^*)$, $\hat{\phi}_{b,n} \in \Phi(\hat{M}_1, \hat{M}_2)$, $\hat{M}_j = O_P(1)$ and $|\hat{\phi}_{b,n}(t) - \phi(t)| \leq \varepsilon_n(b)$ for every t, where $\varepsilon_n(b)$ satisfies (4.14), $\|\cdot\|_2$ is the norm of $L_2(f_\beta)$ and $\hat{A}_{b,n}(\cdot, \cdot) = A(\cdot, \cdot; \hat{S}_{b,n}^{(*)})$ is given by (3.5) with $\hat{S}_{b,n}^{(*)}$ being the self-consistent estimator of $S^{(*)}$ based on $\hat{Q}_{b,n}^{(k)}$ and $\hat{S}_{b,n}$. These guarantee (6.5) and $A(\phi - \phi_\epsilon, \phi - \phi_\epsilon) \leq \varepsilon_0^2 A(\phi_\epsilon, \phi_\epsilon)$. By (6.4) and Theorem 6.1, $A(\phi, \phi_\epsilon) \geq (1 - \varepsilon_0) A(\phi_\epsilon, \phi_\epsilon)$, and the zero-crossings of $\hat{\xi}_n(b)$ with such $\hat{\phi}_{b,n}$ will be highly efficient under the other conditions of Theorem 5.2. We may also use $\hat{\phi}_{b,n}(t) = \phi(t + b\overline{X}_n)$ to stabilize the center of $\hat{\phi}_{b,n}$ for large $|b - \beta|$. It is possible to derive efficient estimators of β by this method, if one finds suitable upper bounds for the rate of convergence of the remainder in (4.15) when M_1^* (slowly) converges to ∞ along with n.

PROOF. By (6.5), $\phi \in \Phi(M_1^*, M_2^*)$. By Theorem 4.2 it suffices to show

$$\lim_{\varepsilon\to 0+} \lim_{\varepsilon'\to 0+} \limsup_{n\to\infty} P\left\{ \sup_{|b-\beta|\leq\varepsilon'} \sup_{h\in C(\varepsilon)} \left| \int h\, d\bigl(\hat{\phi}_{b,n} - \phi\bigr) \right| > \varepsilon'' \right\} = 0$$

for all compact sets C in D_o and positive ε'', where $C(\varepsilon) = \bigcup_{h'\in C} \{h \in D_o: \|h - h'\| \leq \varepsilon\}$. Since $\hat{\phi}_{b,n} \in \Phi(M_1^*, M_2^*)$ with large probability and step functions h with finite jumps are dense in D_o, this is a consequence of $\hat{\phi}_{b,n}(t) = \phi(t) + \varepsilon_n(b)$. □

The identities (4.17)–(4.19) and (4.21) are also useful for the expansions of other statistics of interest such as estimates of the error survival function.

THEOREM 6.2. Suppose $0 < S_\beta(t) < 1$ for all t and conditions (4.4), (4.5) and (4.8)–(4.11) hold for $k = 0$. Then, with the notation of Theorems 4.2 and 5.2,

$$\begin{aligned} K_n\Big(\hat{S}_{b,n} - S_{\beta,n}\Big) & \\ &= n^{-1/2}K_n R^{-1}_{S_{\beta,n},n}B_{S_{\beta,n}}W^{(0)}_{\beta,n} - (b-\beta)K(g_\epsilon + EXf_\epsilon) \qquad (6.6)\\ &\quad + \varepsilon_n(b,t)\big(|b-\beta| + n^{-1/2}\big) \end{aligned}$$

such that $K_n R^{-1}_{S_{\beta,n},n}B_{S_{\beta,n}}W^{(0)}_{\beta,n} \to_{\hat{\mathscr{D}}} KH$ as elements in D_o and (4.14) holds for $\varepsilon_n(b) = \sup_t|\varepsilon_n(b,t)|$, where $f_\epsilon = f_\beta$, $g_\epsilon(t) = \psi(t,2;\phi_\epsilon)S_\beta(t)$ for the projections in (3.3) and $H = R^{-1}B_{S_\beta}W^{(0)}$. Furthermore, if all conditions of Theorem 5.2 hold, then

$$\begin{aligned} K_n\sqrt{n}\left(\hat{S}_{\hat{\beta}_n,n} - S_{\beta,n}\right) &= K_n R^{-1}_{S_{\beta,n},n}B_{S_{\beta,n}}W^{(0)}_{\beta,n} \\ &\quad - \frac{\eta_n(\phi - E_n\phi)}{A(\phi,\phi_\epsilon)}K(g_\epsilon + EXf_\epsilon) + \tilde{\varepsilon}_n(t) \qquad (6.7) \end{aligned}$$

such that $\|\tilde{\varepsilon}_n\| = o_P(1)$, $(K_n R^{-1}_{S_{\beta,n},n}B_{S_{\beta,n}}W^{(0)}_{\beta,n}, \eta_n(\phi - E_n\phi)) \to_{\hat{\mathscr{D}}} (KH, \eta(\phi))$ in $D_o \times (-\infty,\infty)$ and $H(t)$ is a Gaussian process independent of $\eta(\phi)$ with $EH(t) = 0$ and $EH(t)H(s) = (R^{-1}v_s)(t)$, where $v_s(t) = S_\beta(\max(s,t)) - S_\beta(s)S_\beta(t)$.

REMARK 6.2. The independence of $H(t)$ and $\eta(\phi)$ is a consequence of Theorem 3 of Gu and Zhang (1993).

PROOF. Similar to the proof of (4.15), by (4.17) and Lemmas 4.6(ii) and 4.7(ii),

$$K_n\Big(\hat{S}_{b,n} - S_{\beta,n}\Big) = n^{-1/2}K_n R^{-1}_{S_{\beta,n},n}W^{(0)}_{\beta,n} + (b-\beta)KR^{-1}B_{S_\beta}h^{(0)} + \cdots .$$

By Theorem 3.1, $(R^{(1)}f_\beta)(t) = E(X-EX)E^*\phi_t E^*\phi_\epsilon = (Rg_\epsilon)(t)$, so that, by (4.19), $-R^{-1}B_{S_\beta}h^{(0)} = R^{-1}R^{(1)}f_\beta + EXf_\beta = g_\epsilon + EXf_\epsilon$. Hence we have (6.6) and (6.7).

Since $\psi(z,j;\phi_{t,n},S^{(*)}_{\beta,n})$ is the projection, by (2.4), (3.3) and (4.12),

$$B_{S_{\beta,n}}W^{(0)}_{\beta,n} = n^{-1/2}\sum_{i=1}^{n} E\big[\phi_{t,n}(\epsilon_{i,n})|\text{data}\big]$$

is uncorrelated to $\eta_n(\phi - E_n\phi)$, so that $EH(t)\eta(\phi) = 0$ for all t. Since $v_s(t) = -\int_t^\infty \phi_s(z)S_\epsilon(dz)$, by Theorem 3.1(i),

$$E\Big(B_{S\beta}W^{(0)}\Big)(t)\Big(B_{S_\beta}W^{(0)}\Big)(s) = EE^*\phi_t E^*\phi_s = (Rv_s)(t),$$

so that $EH(t)(B_{S_\beta}W^{(0)})(s) = (R^{-1}Rv_s)(t) = v_s(t) = v_t(s)$ and $EH(\cdot)H(s) = E(R^{-1}B_{S_\beta}W^{(0)})H(s) = R^{-1}v_s$. □

APPENDIX

PROOF OF LEMMA 4.7. As $0 < S_\beta < 1$ for all t, our (4.5) implies (2.6) of Gu and Zhang (1993) and our (4.8) implies (2.13) and (4.3) of Gu and Zhang (1993). Part (i) follows from step 1 of the proof of Theorem 2 in Gu and Zhang (1993). Part (ii) is an immediate consequence of Lemma 2 of Gu and Zhang (1993) by standard methods in functional analysis. Let $g_n = K_n R_{S_n, n}^{-1} h_n$ and $g = KR^{-1}h$. By part (ii), $g_n \to g$ in D_o, so that part (iii) follows if $I_{[-M,M)} R_{S_n,n}^{(1)} K_n^{-1}$ strongly converge to $I_{[-M,M)} R^{(1)} K^{-1}$ as linear operators in D_o. It suffices to show $\|I_{[-M,M)} R_{S_n,n}^{(1)} K_n^{-1}\| = O(1)$, $\|I_{[-M,M)} R^{(1)} K^{-1}\| < \infty$ and $\|I_{[-M,M)} \{R_{S_n,n}^{(1)} K_n^{-1} - R^{(1)} K^{-1}\} g\| = o(1)$ for all g of bounded variation, as these g are dense in D_o.

It follows from (4.8), (4.1), (4.2) and the condition $P\{|X_{i,n}| \le M_1\} = 1$ that

$$
\begin{aligned}
&\sup_{|t|\le M} \frac{|K_n^{(1)}(t)|}{K_n(t)} + \sup_{|t|\le M} \left\{ \int_{z\le t} \frac{|S_{U,n}^{(1)}(dz)|}{K_n(z)} + \int_{z>t} \frac{|S_{V,n}^{(1)}(dz)|}{K_n(z)} \right\} \\
&\qquad \le 2M_1 + 2M_1 \sup_{|t|\le M} \left| \int_{z\le t} \frac{S_{U,n}^{(0)}(dz)}{K_n(t)} + \int_{z>t} \frac{S_{V,n}^{(0)}(dz)}{K_n(z)} \right| \\
&\qquad \le M^* < \infty
\end{aligned}
\tag{A.1}
$$

as $|S_{U,n}^{(1)}(dz)| \le 2M_1 |S_{U,n}^{(0)}(dz)|$, $|S_{V,n}^{(1)}(dz)| \le 2M_1 |S_{V,n}^{(0)}(dz)|$ and

$$
\begin{aligned}
|K_n^{(k)}(t)| &= \left| \frac{1}{n} \sum_{i=1}^{n} E\big(X_{i,n} - E\overline{X}_n\big)^k I\{V_{i,n} - \beta X_{i,n} \le t < U_{i,n} - \beta X_{i,n}\} \right| \\
&\le (2M_1)^k K_n^{(0)}.
\end{aligned}
$$

By (4.8) and (3.7), $\|I_{[-M,M)} R_{S_n,n}^{(1)} K_n^{-1}\|$ is bounded by (A.1). We can also obtain by the same method $\|I_{[-M,M)} R^{(1)} K^{-1}\| \le M^*$.

Let g be a function of bounded variation in D_o. Since $\|K_n^{(k)} - K^{(k)}\| \to 0$ and $|K_n^{(1)}(t)|/K_n(t) \le 2M_1$, $\|I_{[-M,M)} \{R_{S_n,n}^{(1)} K_n^{-1} - R^{(1)} K^{-1}\} g\|$ is bounded by a sum of a $o(1)$ term and the norms involving two integrations in view of the definitions in (4.7) and (3.7). By symmetry, we shall only deal with the first one, which is bounded by

$$
\begin{aligned}
&\sup_{|t|\le M} \left| \int_{z\le t} g(z) \left(\frac{S_n(t)}{S_n(z)} \frac{S_{U,n}^{(1)}(dz)}{K_n(z)} - \frac{S_\beta(t)}{S_\beta(z)} \frac{S_U^{(1)}(dz)}{K(z)} \right) \right| \\
&\qquad \le o(1) + \sup_{|t|\le M} \left| \int_{z\le t} \frac{g(z)}{S_\beta(z)} \left(\frac{S_{U,n}^{(1)}(dz)}{K_n(z)} - \frac{S_U^{(1)}(dz)}{K(z)} \right) \right|
\end{aligned}
$$

as $S_n(t)/S_n(z)$ converge uniformly to $S_\beta(t)/S_\beta(z)$ for $z < t \le M$. The proof is finished via integrating by parts, as

$$\int_{z\le t} \frac{S_{U,n}^{(1)}(dz)}{K_n(z)} = \int_{z\le t} \frac{Q_{\beta,n}^{(1)}(dz,2)}{S_{\beta,n}(z)K_n(z)} \to \int_{z\le t} \frac{S_U^{(1)}(dz)}{K(z)}$$

uniformly for $t \le M$ by (4.5) and (A.1). □

REFERENCES

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

BICKEL, P. J. (1982). On adaptive estimation. *Ann. Statist.* **10** 647–671.

BICKEL, P. J., KLAASSEN, C. A. J., RITOV, Y. and WELLNER, J. A. (1993). *Efficient and Adaptive Estimation for Semiparametric Models*. Johns Hopkins Univ. Press.

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

CHANG, M. N. (1990). Weak convergence of a self-consistent estimator of the survival function with doubly censored data. *Ann. Statist.* **18** 391–404.

CHANG, M. N. and YANG, G. L. (1987). Strong consistency of a nonparametric estimator of the survival function with doubly censored data. *Ann. Statist.* **15** 1536–1547.

COX, D. R. (1972). Regression models and life-tables (with discussion). *J. Roy. Statist. Soc. Ser. B* **34** 187–202.

DUDLEY, R. M. (1985). An extended Wichura theorem, definitions of Donsker class, and weighted empirical distributions. *Lecture Notes in Math.* **1153** 141–178. Springer, New York.

GEHAN, E. A. (1965). A generalized two-sample Wilcoxon test for doubly censored data. *Biometrika* **52** 650–653.

GINÉ, E. and ZINN, J. (1986). Lectures on the central limit theorem for empirical processes. *Lecture Notes in Math.* **1221** 50–113. Springer, New York.

GU, M. G. and ZHANG, C. H. (1993). Asymptotic properties of self-consistent estimators based on doubly censored data. *Ann. Statist.* **21** 611–624.

HOFFMANN-JORGENSEN, J. (1984). Stochastic processes on polish spaces. Unpublished manuscript.

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

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

LAI, T. L. and YING, Z. (1991). Large sample theory of a modified Buckley–James estimator for regression analysis with censored data. *Ann. Statist.* **19** 1370–1402.

LEIDERMAN, P. H., BABU, B., KAGIA, J., KRAEMER, H. C. and LEIDERMAN, G. F. (1973). African infant precocity and social influence during the first year. *Nature* **242** 247–249.

MILLER, R. G. and HALPERN, J. (1982). Regression with censored data. *Biometrika* **69** 521–531.

PETO, R. (1973). Experimental survival curves for interval censored data. *J. Roy. Statist. Soc. Ser. C* **22** 86–91.

POLLARD, D. (1984). *Convergence of Stochastic Processes*. Springer, New York.

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

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, martingale, and the Cox model. *Contemp. Math.* **80** 191–219.

STEIN, C. (1956). Efficient nonparametric testing and estimation. *Proc. Third Berkeley Symp. Math. Statist. Probab.* **1** 187–195. Univ. California Press, Berkeley.

TSAI, W. Y. and CROWLEY, J. (1985). A large sample study of generalized maximum likelihood estimators from incomplete data via self-consistency. *Ann. Statist.* **13** 1317–1334.

TSAI, W.-Y. and ZHANG, C.-H. (1995). Asymptotic properties of nonparametric maximum likelihood estimator for interval-truncated data. *Scand. J. Statist.* **22** 361–370.

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

TURNBULL, B. W. (1974). Nonparametric estimation of a survivorship function with doubly censored data. *J. Amer. Statist. Assoc.* **69** 169–173.

VAN DER LAAN, M. J. (1996). *Efficient and Inefficient Estimation in Semiparametric Models. CWI Tract* **44**. Centre Math. Comput. Sci., Amsterdam.

VARDI, Y. and ZHANG, C.-H. (1992). Large sample study of empirical distributions in a random-multiplicative censoring model. *Ann. Statist.* **20** 1022–1039.

YING, Z. (1993). A large sample study of rank estimation for censored regression data. *Ann. Statist.* **21** 76–99.

ZHAN, Y. and WELLNER, J. (1995). Double censoring: characterization and computation of the nonparametric maximum likelihood estimator. Technical Report 292, Dept. Statistics, Univ. Washington, Seattle.

DEPARTMENT OF STATISTICS
RUTGERS UNIVERSITY
HILL CENTER FOR THE MATHEMATICAL SCIENCES
BUSCH CAMPUS
PISCATAWAY, NEW JERSEY 08855