

REGRESSION M -ESTIMATORS WITH DOUBLY CENSORED DATA

BY JIAN-JIAN REN¹ AND MINGGAO GU²

Tulane University and McGill University

The M -estimators are proposed for the linear regression model with random design when the response observations are doubly censored. The proposed estimators are constructed as some functional of a Campbell-type estimator \hat{F}_n for a bivariate distribution function based on data which are doubly censored in one coordinate. We establish strong uniform consistency and asymptotic normality of \hat{F}_n and derive the asymptotic normality of the proposed regression M -estimators through verifying their Hadamard differentiability property. As corollaries, we show that our results on the proposed M -estimators also apply to other types of data such as uncensored observations, bivariate observations under univariate right censoring, bivariate right-censored observations, and so on. Computation of the proposed regression M -estimators is discussed and the method is applied to a doubly censored data set, which was encountered in a recent study on the age-dependent growth rate of primary breast cancer.

1. Introduction. When statisticians are interested in modeling the lifetime distribution under consideration as a function of some covariate, the following linear regression model is one of the most widely used tools in statistical analysis:

$$(1.1) \quad X_i = \alpha + T_i \beta + e_i, \quad i = 1, \dots, n,$$

where X_i are the lifetime random variables (r.v.), T_i are the covariate variables which are independent and identically distributed (i.i.d.) with d.f. F_T , e_i are the i.i.d. error variables with zero mean, T_i and e_i are independent and $(\alpha, \beta) \in \mathbb{R}^2$ is the regression parameter to be estimated. One may note that in model (1.1), X_i 's are i.i.d. random variables with a common d.f. F_X . There are many well-developed theories for this model and computer software is available when complete data are observed. However, in medical follow-up and reliability studies, incomplete data are frequently encountered, which demand new methods so that regression models can be properly used to analyze lifetime data. The right-censored linear regression model has been studied by Buckley and James (1979), Koul, Susarla and Van Ryzin (1981),

Received February 1996; revised May 1997.

¹ Research supported in part by NSF Grants DMS-95-10376 and DMS-96-26532.

² Research supported in part by Natural Science and Engineering Research Council of Canada.

AMS 1991 subject classifications. Primary 62G05, 62J05; secondary 62E20.

Key words and phrases. Asymptotic normality, bivariate distribution function, bivariate right-censored data, consistency, Hadamard differentiability, linear regression model, M -estimators, statistical functional, weak convergence.

Leurgans (1987), Ritov (1990), Lai and Ying (1991), Zhou (1992) and others. In Lai and Ying (1994), the linear regression model with left-truncated and right-censored response variables is considered. Recently, Zhang and Li (1996) extended Buckley–James–Ritov-type regression estimators from the right-censored case to the doubly censored case. In this paper, we consider the *doubly censored linear regression model*; that is, the response variables X_i 's in model (1.1) are doubly censored, and we construct our regression estimators in a different way from that in Zhang and Li (1996).

To be precise, in this study one does not observe $\{X_i\}$ in model (1.1), but a doubly censored sample:

$$(1.2) \quad V_i = \begin{cases} X_i, & \text{if } Z_i < X_i \leq Y_i, & \delta_i = 1, \\ Y_i, & \text{if } X_i > Y_i, & \delta_i = 2, \\ Z_i, & \text{if } X_i \leq Z_i, & \delta_i = 3, \end{cases} \quad i = 1, \dots, n,$$

where independent from X_i , (Z_i, Y_i) are i.i.d. realizations of (Z, Y) with $P\{Z < Y\} = 1$, and Y_i and Z_i are called *right* and *left censoring variables*, respectively. Examples of the doubly censored sample (1.2) encountered in practice have been given by Gehan (1965), Turnbull (1974) and others. In particular, doubly censored data (1.2) occurred in a recent study on the age-dependent growth rate of primary breast cancer [Peer, Van Dijck, Hendriks, Holland and Verbeek (1993)]. In our study of the linear regression model (1.1), we consider the case that the covariate r.v.'s T_i are observable and they are independent from the censoring variables (Y_i, Z_i) . The problem considered here is to estimate (α, β) in (1.1) based on (V_i, δ_i, T_i) , $1 \leq i \leq n$.

To construct an M -estimator of (α, β) , we note that when there is no censoring, the robust M -estimator (α_n, β_n) for model (1.1) is given as the solution of the following equations:

$$(1.3) \quad \sum_{i=1}^n \psi(X_i - \theta_1 - T_i \theta_2) = 0 \quad \text{and} \quad \sum_{i=1}^n T_i \psi(X_i - \theta_1 - T_i \theta_2) = 0,$$

where ψ is the *score function* [Huber (1981)]. In particular, if $\psi(x) = x$, the solution of (1.3) is the *least squares estimator* (LSE). If we denote the empirical d.f. of (X_i, T_i) , $1 \leq i \leq n$, as

$$(1.4) \quad F_n(x, t) = \frac{1}{n} \sum_{i=1}^n I\{X_i \leq x, T_i \leq t\},$$

then (1.3) is equivalent to

$$(1.5) \quad \begin{aligned} \iint \psi(x - \theta_1 - \theta_2 t) dF_n(x, t) &= 0 \quad \text{and} \\ \iint t \psi(x - \theta_1 - \theta_2 t) dF_n(x, t) &= 0. \end{aligned}$$

Hence, if we define a functional $\tau(\cdot)$ at F_n as the solution of (1.5), then we have $(\alpha_n, \beta_n) = \tau(F_n)$. Naturally, if an estimator \hat{F}_n for the joint d.f. F of

(X_i, T_i) based on (V_i, δ_i, T_i) , $1 \leq i \leq n$, is available, then the generalized M -estimator for (1.1) may be constructed by $(\hat{\alpha}_n, \hat{\beta}_n) = \tau(\hat{F}_n)$.

In this context and for its importance in its own right, a Campbell-type estimator \hat{F}_n [Campbell (1981)] for the bivariate distribution function of (X_i, T_i) based on data (V_i, δ_i, T_i) , $1 \leq i \leq n$, is constructed and studied in Section 2, where we also establish strong uniform consistency and asymptotic normality of \hat{F}_n , with the proofs deferred to Section 5. In Section 3, we show that the functional $\tau(\cdot)$ defined by (1.5) is Hadamard differentiable (the proofs are deferred to Section 6) and that the asymptotic normality of the proposed M -estimator $\tau(\hat{F}_n)$ follows from the asymptotic normality of \hat{F}_n . As corollaries, we also show in Section 3 that our results on the proposed M -estimators apply to other types of data, such as uncensored data, bivariate observations under univariate right censoring [Lin and Ying (1993)], bivariate right-censored observations [Dabrowska (1988)], and so on. Section 4 discusses the computation of the proposed M -estimator and applies the proposed regression M -estimators to a doubly censored data set encountered in the study of primary breast cancer (Peer, Van Dijck, Hendriks, Holland and Verbeek, 1993).

One may note that with some modifications in the proofs, the results established in this paper can be extended to p -dimensional ($p > 1$) linear regression models when the covariate variables are observable and the response variables are doubly censored.

One may also note that the independence condition between the covariate variable T_i and the censoring variable (Y_i, Z_i) is not required in Zhang and Li (1996). This condition is needed here because we construct our regression estimators through some functional of a bivariate distribution estimator \hat{F}_n for the distribution of (X_i, T_i) . Usually, when one wants to estimate the bivariate distribution with censored data, for identifiability reasons it has to be assumed that the censored vector is independent of (X_i, T_i) . For reference, see Stute (1993) who considered such an estimation problem when X_i is right censored. The advantage of our functional plug-in method for constructing the regression estimators is that it is easily applicable to different types of censored data; this will be discussed in Section 3.

2. Bivariate distribution function estimator. The distribution of the underlying lifetime is often of special interest when incomplete data are observed. In the right-censored case, the product limit estimator of Kaplan and Meier (1958) has been generally accepted as a substitute for the empirical distribution function, since it is the nonparametric maximum likelihood estimator (NPMLE) [Cox and Oakes (1984), page 48] and possesses the properties of self-consistency [Efron (1967)], asymptotic normality [Breslow and Crowley (1974); Gill (1983)], and asymptotic efficiency [Wellner (1982)]. In the doubly censored case, it has been shown that all these properties are also possessed by the NPMLE or the self-consistent estimators. See Mykland and Ren (1996) on the NPMLE and the self-consistent estimators, see Chang (1990) or Gu and Zhang (1993) on the asymptotic normality and see Gu and

Zhang (1993) on the asymptotic efficiency. Other related work can be found in Groeneboom (1987), Samuelsen (1989) and Ren (1995), among others.

However, for the problems using bivariate observations which may be incomplete either in one coordinate or in both, the direct use of the self-consistent estimator often leads to computationally and analytically intractable problems. One way to handle such a problem is to use the conditional distribution approach, which was applied by Campbell (1981) to estimate the bivariate distribution when the bivariate observations are possibly right, censored in both coordinates. In our study here, we consider the problem of estimating the bivariate distribution when one coordinate is subject to double censoring as expressed in (1.2). An immediate application of this study is the linear regression model (1.1), which is discussed in Section 1 and Section 3. In the following, we will construct our estimator using the conditional distribution approach and will establish the strong uniform consistency and the asymptotic normality of the proposed estimator.

Using observations (V_i, δ_i, T_i) , $1 \leq i \leq n$, which are described in Section 1, we construct the estimator \hat{F}_n for the bivariate d.f. F of (X_i, T_i) through the conditional self-consistent estimating equation for doubly censored data. First, we observe that for any fixed t and $j = 1, 2, 3$, if we denote $Q_t^{(j)}(x)$ as the conditional distribution $P\{V \leq x, \delta = j | T \leq t\}$ and $F_t(x)$ as the conditional distribution $P\{X \leq x | T \leq t\}$, then from (2.7) of Gu and Zhang (1993) we have

$$(2.1) \quad F_t(x) = Q_t^{(0)}(x) - \int_{u \leq x} \frac{1 - F_t(x)}{1 - F_t(u)} dQ_t^{(2)}(u) + \int_{x < u} \frac{F_t(x)}{F_t(u)} dQ_t^{(3)}(u),$$

where $Q_t^{(0)}(x) = P\{V \leq x | T \leq t\} = \sum_{j=1}^3 Q_t^{(j)}(x)$. By multiplying $P\{T \leq t\}$ on both sides of (2.1), we obtain

$$(2.2) \quad F(x, t) = Q^{(0)}(x, t) - \int_{u \leq x} \frac{S(x, t)}{S(u, t)} Q^{(2)}(du, t) + \int_{x < u} \frac{F(x, t)}{F(u, t)} Q^{(3)}(du, t),$$

where $S(x, t) = F(\infty, t) - F(x, t)$ and $Q^{(j)}(x, t) = P\{V \leq x, \delta = j, T \leq t\}$, $j = 1, 2, 3$, with $Q^{(0)}(x, t) = P\{V \leq x, T \leq t\} = \sum_{j=1}^3 Q^{(j)}(x, t)$. Thus, if we denote

$$(2.3) \quad Q_n^{(j)}(x, t) = \frac{1}{n} \sum_{i=1}^n I\{V_i \leq x, \delta_i = j, T_i \leq t\}, \quad j = 1, 2, 3,$$

$$Q_{n,t}^{(j)}(x) = Q_n^{(j)}(x, t) / \hat{G}_n(t), \quad j = 1, 2, 3 \text{ with } \hat{G}_n(t) > 0,$$

$$Q_n^{(0)}(x, t) = \sum_{j=1}^3 Q_n^{(j)}(x, t) \quad \text{and} \quad Q_{n,t}^{(0)}(x) = \sum_{j=1}^3 Q_{n,t}^{(j)}(x),$$

where $\hat{G}_n(t) = n^{-1} \sum_{i=1}^n I\{T_i \leq t\}$ is the empirical d.f. of T_1, \dots, T_n , (2.1) implies

that for each fixed t , an estimator $\hat{F}_{n,t}$ for the conditional distribution F_t is given by a solution of the following equation:

$$(2.4) \quad \begin{aligned} \hat{F}_{n,t}(x) &= Q_{n,t}^{(0)}(x) - \int_{u \leq x} \frac{1 - \hat{F}_{n,t}(x)}{1 - \hat{F}_{n,t}(u)} dQ_{n,t}^{(2)}(u) \\ &\quad + \int_{x < u} \frac{\hat{F}_{n,t}(x)}{\hat{F}_{n,t}(u)} dQ_{n,t}^{(3)}(u). \end{aligned}$$

Naturally, an estimator \hat{F}_n for F is given by

$$(2.5) \quad \hat{F}_n(x, t) = \hat{F}_{n,t}(x) \hat{G}_n(t),$$

which, based on (2.2), is equivalent to a solution of the following equation:

$$(2.6) \quad \begin{aligned} \hat{F}_n(x, t) &= Q_n^{(0)}(x, t) - \int_{u \leq x} \frac{S_n(x, t)}{S_n(u, t)} Q_n^{(2)}(du, t) \\ &\quad + \int_{x < u} \frac{\hat{F}_n(x, t)}{\hat{F}_n(u, t)} Q_n^{(3)}(du, t), \end{aligned}$$

where

$$(2.7) \quad S_n(x, t) = \hat{G}_n(t) - \hat{F}_n(x, t),$$

and the convention $\int_{u \leq x} = 0$ ($\int_{x < u} = 0$) if $\hat{G}_n(t) = \hat{F}_n(x, t)$ [$\hat{F}_n(x, t) = 0$] is adopted.

The proposed estimator \hat{F}_n may be obtained numerically using the method in Mykland and Ren (1996). Detailed discussion on this is given in Section 4, where we show that a solution of (2.6) satisfies

$$(2.8) \quad \hat{F}_n(x, t) = \sum_{k=1}^n \sum_{i=1}^k \left(\frac{k}{n} a_{ki} - \frac{k-1}{n} a_{k-1,i} \right) I\{V_i \leq x, T_k \leq t\},$$

where $T_1 \leq \dots \leq T_n$ and $a_{ki} \geq 0$ are constants determined by the sample (V_i, δ_i, T_i) , $i = 1, \dots, n$. In this work, we will always impose the following condition on the solution \hat{F}_n of (2.6): for any t ,

$$(2.9) \quad \begin{aligned} Q_n^{(2)}(a_n(t), t) &= \hat{G}_n(t) \int_{u < a_n(t)} \frac{Q_n^{(2)}(du, t)}{S_n(u, t)}, \\ Q_n^{(3)}(\infty, t) - Q_n^{(3)}(b_n(t), t) &= \hat{G}_n(t) \int_{b_n(t) < u} \frac{Q_n^{(3)}(du, t)}{\hat{F}_n(u, t)}, \end{aligned}$$

where $a_n(t) = \min\{V_i; \delta_i = 1 \text{ or } 3, T_i \leq t\}$ and $b_n(t) = \max\{V_i; \delta_i = 1 \text{ or } 2, T_i \leq t\}$. This condition is motivated by the conditional NPMLE for F_t , and one may see (2.5) of Gu and Zhang (1993) for a similar condition for the NPMLE of F_X .

To state our asymptotic results on the proposed estimator \hat{F}_n , we introduce some notation. Denote F_X, F_Y, F_Z and F_T as the d.f.'s of X, Y, Z and T , respectively, and denote

$$(2.10) \quad \begin{aligned} K(x) &= F_Z(x) - F_Y(x), \\ a &= \sup\{x | F_X(x) = 0\} \quad \text{and} \quad b = \inf\{x | F_X(x) = 1\}, \end{aligned}$$

then from (1.2) we have

$$(2.11) \quad \begin{aligned} K(x-) &= P\{\delta = 1 | X = x\}, & Q^{(1)}(dx, t) &= K(x-)F(dx, t), \\ Q^{(2)}(dx, t) &= S(x, t) dF_Y(x), & Q^{(3)}(dx, t) &= F(x, t) dF_Z(x). \end{aligned}$$

Throughout this paper, $\|\cdot\|$ stands for the supremum norm and $\|\cdot\|_2$ stands for the Euclidean norm in \mathbb{R}^2 , where $\mathbb{R} = (-\infty, \infty)$. The following theorem establishes the strong uniform consistency of \hat{F}_n under the assumption

$$(2.12) \quad \begin{aligned} K(x-) &= P\{\delta = 1 | X = x\} > 0 \\ &\text{for } x \in \{x | F_X(x) > 0, F_X(x-) < 1\}, \end{aligned}$$

with the proof deferred to Section 5, where the results in the one-dimensional case by Gu and Zhang (1993) are used.

THEOREM 2.1. *Suppose that (2.12) holds. Then for a solution \hat{F}_n of (2.6) satisfying (2.9), $\|\hat{F}_n - F\| \rightarrow 0$ a.s., as $n \rightarrow \infty$.*

To establish the weak convergence of the bivariate distribution estimator \hat{F}_n , we denote for $a_j \geq -\infty$ and $b_j \leq \infty$, $j = 1, 2$,

$$(2.13) \quad \begin{aligned} &\mathbb{M}([a_1, b_1] \times [a_2, b_2]) \\ &= \{H | H: [a_1, b_1] \times [a_2, b_2] \rightarrow \mathbb{R} \text{ corresponds to} \\ &\quad \text{a finite signed measure on } \mathbb{R}^2\}, \end{aligned}$$

and consider the Banach space $(\overline{\mathbb{M}}([a_1, b_1] \times [a_2, b_2]), \|\cdot\|)$, where $\overline{\mathbb{M}}([a_1, b_1] \times [a_2, b_2])$ is the closure of $\mathbb{M}([a_1, b_1] \times [a_2, b_2])$. One may note that since $\hat{F}_{n,t}$ is not necessarily a proper d.f. for a fixed t [Mykland and Ren (1996)], \hat{F}_n given by (2.5) is not necessarily a proper bivariate d.f., and that based on (2.8), $\mathbb{M}([a_1, b_1] \times [a_2, b_2])$ contains \hat{F}_n as an element. For our study, we further define the following Banach spaces:

$$(2.14) \quad \begin{aligned} &(D_0([a, b] \times \mathbb{R}), \|\cdot\|) \\ &= \{h \in \overline{\mathbb{M}}([a, b] \times \mathbb{R}) | F(x, t) = 0 \Rightarrow h(x, t) = 0, \\ &\quad S(x-, t) = 0 \Rightarrow h(x-, t) = 0\}, \\ &(D_K([a, b] \times \mathbb{R}), \|\cdot\|_K) \\ &= \{h | Kh \in \overline{\mathbb{M}}([a, b] \times \mathbb{R})\}, \|h\|_K = \|Kh\|, \\ &(D_0^3, \|\cdot\|_3) \\ &= \{h \in \overline{\mathbb{M}} \otimes \overline{\mathbb{M}} \otimes \overline{\mathbb{M}} | B_F h \in D_0([a, b] \times \mathbb{R})\}, \|(h_1, h_2, h_3)\|_3 \\ &= \sum_{j=1}^3 \|h_j\|, \end{aligned}$$

where for any H satisfying

$$(2.15) \quad \begin{aligned} H \in \mathbb{S} = \{ & H \in \mathbb{M}([a, b] \times \mathbb{R}) \mid \text{there exists a d.f. } H_2(t) \\ & \text{such that for any fixed } t \text{ and} \\ & H_t(x) = H(x, t)/H_2(t), H_t(x)/H_t(\infty) \text{ is a d.f.} \} \end{aligned}$$

and

$$(2.16) \quad S_H(x, t) = H_2(t) - H(x, t)$$

and linear operators A_H, B_H, R_H and \mathbb{K} are defined by

$$(2.17) \quad \begin{aligned} (A_H h)(x, t) &= - \int_{u \leq x} \frac{S_H(x, t)}{S_H(u, t)} h(u, t) dF_Y(u) \\ &\quad - \int_{x < u} \frac{H(x, t)}{H(u, t)} h(u, t) dF_Z(u), \\ R_H &= \mathbb{K} - A_H, \quad (\mathbb{K}h)(x, t) = K(x)h(x, t), \\ (B_H(h_1, h_2, h_3))(x, t) &= \sum_{j=1}^3 h_j(x, t) - \int_{u \leq x} \frac{S_H(x, t)}{S_H(u, t)} h_2(du, t) \\ &\quad + \int_{x < u} \frac{H(x, t)}{H(u, t)} h_3(du, t). \end{aligned}$$

One may note that integration by parts should be used above whenever necessary, and that the domains of these operators include all bounded measurable functions, while those of A_H and R_H will be extended under the condition of our Theorem 2.2. In this work, all Banach spaces are equipped with the σ -field generated by all open balls, and random elements and weak convergence are defined as in Pollard [(1984), page 65].

Based on (2.15) and (2.16), it is easy to see that we have $F \in \mathbb{S}$ with $S_F(x, t) = S(x, t) = F_T(t) - F(x, t)$ and $\hat{F}_n \in \mathbb{S}$ with $S_{\hat{F}_n}(x, t) = S_n(x, t) = \hat{G}_n(t) - \hat{F}_n(x, t)$. Thus, (2.2) and (2.6) can be expressed as $F = B_F \mathbf{Q}$ and $\hat{F}_n = B_{\hat{F}_n} \mathbf{Q}_n$, respectively, where

$$\mathbf{Q} = (Q^{(1)}, Q^{(2)}, Q^{(3)}) \quad \text{and} \quad \mathbf{Q}_n = (Q_n^{(1)}, Q_n^{(2)}, Q_n^{(3)}).$$

From some tedious calculation, we obtain

$$\hat{F}_n - F = B_{\hat{F}_n}(\mathbf{Q}_n - \mathbf{Q}) + A_{\hat{F}_n}(\hat{F}_n - F) + (1 - \mathbb{K})(\hat{F}_n - F) + \frac{\eta_n}{\sqrt{n}},$$

where

$$\eta_n(x, t) = \sqrt{n} \left[\hat{G}_n(t) - F_T(t) \right] \int_{u \leq x} \left[\frac{S_n(x, t)}{S_n(u, t)} - 1 \right] dF_Y(u).$$

Hence, we have

$$(2.18) \quad R_{\hat{F}_n} \xi_n = B_{\hat{F}_n} W_n + \eta_n,$$

where

$$(2.19) \quad \xi_n = \sqrt{n}(\hat{F}_n - F) \quad \text{and} \quad W_n = \sqrt{n}(\mathbf{Q}_n - \mathbf{Q}).$$

Because W_n is the empirical process, by (2.9), we have $B_F(\mathbf{Q}_n - \mathbf{Q}) \in D_0([a, b] \times \mathbb{R})$ and as $n \rightarrow \infty$,

$$(2.20) \quad W_n \rightarrow_D W \quad \text{where } W \text{ is a centered Gaussian process in } D_0^3.$$

It is also easy to see that

$$(2.21) \quad \eta_n \rightarrow_D \eta \quad \text{as } n \rightarrow \infty,$$

where $\eta(x, t) = \mathbb{G}_T(t) \int_{u \leq x} [(S(x, t)/S(u, t)) - 1] dF_Y(u)$ and \mathbb{G}_T is the limiting Gaussian process of $\sqrt{n}[\hat{G}_n - F_T]$.

THEOREM 2.2. *Let \hat{F}_n be a solution of (2.6) such that either (2.9) holds or $(\hat{F}_n - F) \in D_0([a, b] \times \mathbb{R})$. Suppose that (2.12) holds and*

$$(2.22) \quad \lim_{t \rightarrow 0} \left\{ \int_{0 < F(u, \infty) < t} \frac{dF_Y(u)}{K(u)} + \int_{0 < S(u, \infty) < t} \frac{dF_Z(u)}{K(u)} \right\} = 0.$$

Then R_F^{-1} , the inverse of R_F , exists and is a bounded linear operator from $D_0([a, b] \times \mathbb{R})$ to $D_K([a, b] \times \mathbb{R})$, and as $n \rightarrow \infty$,

$$\sqrt{n}(\hat{F}_n - F) = \xi_n \rightarrow_D \xi = R_F^{-1}(B_F W + \eta) \quad \text{in } D_K([a, b] \times \mathbb{R}),$$

where W and η are given in (2.20) and (2.21), respectively, and

$$P\{B_F W + \eta = R_F \xi \in D_0([a, b] \times \mathbb{R})\} = 1.$$

The proof of Theorem 2.2 is given in Section 5.

COROLLARY 2.1. *Let \hat{F}_n be a solution of (2.6). If $\inf_{x \in [a, b]} K(x -) > 0$, then as $n \rightarrow \infty$,*

$$\xi_n \rightarrow_D \xi = R_F^{-1}(B_F W + \eta) \quad \text{in } \overline{\mathbb{M}}([a, b] \times \mathbb{R}).$$

3. Regression M -estimators. In Section 1, we used the functional plug-in method to construct an M -estimator $(\hat{\alpha}_n, \hat{\beta}_n) = \tau(\hat{F}_n)$ for the regression parameter (α, β) in model (1.1), where \hat{F}_n given by (2.6) is the bivariate d.f. estimator for the d.f. F of (X_i, T_i) based on doubly censored observations (V_i, δ_i, T_i) , $1 \leq i \leq n$, and $\tau(\cdot)$ is a statistical functional defined by (1.5). To be precise, we consider the case that the covariate variable T_i in model (1.1) has a compact support $[0, c]$, $0 < c < \infty$, and for $\mathbb{M}_0 = \mathbb{M}([a, b] \times [0, c])$, the functional $\tau: \mathbb{M}_0 \rightarrow \mathbb{R}^2$ is defined as the root of the following equations:

$$\iint \psi(x - \theta_1 - \theta_2 t) dH(x, t) = 0,$$

$$\iint t \psi(x - \theta_1 - \theta_2 t) dH(x, t) = 0, \quad H \in \mathbb{M}_0,$$

which can be denoted equivalently as

$$(3.1) \quad \Psi(\boldsymbol{\theta}, H) \equiv \iint \psi(x - \boldsymbol{\theta}^T \mathbf{t}) \mathbf{t} dH(x, t) = \mathbf{0}$$

for $\mathbf{t} = (1, t)^T$, $\boldsymbol{\theta} = (\theta_1, \theta_2)^T$, $H \in \mathbb{M}_0$,

where the integration is defined on $(x, t) \in [a, b] \times [0, c]$ for a, b given by (2.10), and this applies in this section and in Section 6 unless the region of the integration is specified. As follows, we derive the asymptotic normality of the regression M -estimator $\tau(\hat{F}_n)$ through the Hadamard differentiability property of the functional $\tau(\cdot)$.

The asymptotic normality of a statistical functional via the Hadamard derivative for univariate observations has been studied by Reeds (1976) and Fernholz (1983) and for multivariate observations by Ren and Sen (1995). In these studies, the empirical distribution functions are used. A more general limiting distribution theory based on the weak convergence of the random elements in Banach space is given in Andersen, Borgan, Gill and Keiding (1993). In our current study, since we consider the incomplete data, the empirical d.f.'s are not applicable. Thus, we will derive the asymptotic normality of $\tau(\hat{F}_n)$ using the general limiting theory given in Andersen, Borgan, Gill and Keiding (1993). Specifically, we will verify the Hadamard differentiability condition of $\tau(\cdot)$, derive its Hadamard derivative τ'_F and obtain the asymptotic normality of $\tau(\hat{F}_n)$ from τ'_F and the weak convergence of \hat{F}_n .

First, we need to investigate the existence of the solution of (3.1) for our bivariate d.f. estimator \hat{F}_n given by (2.6). We note that if the score function ψ is the derivative of some nonnegative convex function ρ , that is, $\rho' = \psi$, then for any bivariate d.f. F , (3.1) is equivalent to the minimization problem

$$(3.2) \quad \min_{\boldsymbol{\theta} \in \mathbb{R}^2} \iint \rho(x - \boldsymbol{\theta}^T \mathbf{t}) dF(x, t),$$

because

$$(3.3) \quad R(\boldsymbol{\theta}) = \iint \rho(x - \boldsymbol{\theta}^T \mathbf{t}) dF(x, t)$$

is a convex function. However, our bivariate d.f. estimator \hat{F}_n is not a proper bivariate d.f. [see (2.8)]; thus (3.1) and (3.2) are not necessarily equivalent when F is replaced by \hat{F}_n . In the next two lemmas, we show the existence of the solution of (3.1) in a neighborhood of F . Some of the following conditions are imposed in each theorem of this section.

ASSUMPTIONS.

(A1) ψ is nondecreasing, bounded, continuous, piecewise differentiable with bounded derivative ψ' such that $\psi'(x) = 0$ for x outside of some finite interval $[d_1, d_2]$, and for x in some neighborhood of 0, $\psi(x)$ has a range

including positive and negative values and $\psi'(x) \geq m > 0$ for a constant $0 < m < \infty$;

(A2) ψ' is of bounded variation;

(A3) $\Psi(\boldsymbol{\beta}, F) = \mathbf{0}$ where $\boldsymbol{\beta} = (\alpha, \beta)^T$.

REMARK 1. (A1) is usually required for Hadamard differentiability property of M -estimators [see Fernholz (1983) for location M -estimator], and Huber's score given in Section 4 satisfies (A1). Conditions (A1) and (A2) are needed in Lemma 3.1 below for the result on integration by parts. Note that $E\{e_i\} = 0$ and (A1) implies $E\{\psi'(e_i)\} > 0$.

REMARK 2. (A3) is implied by $E\{\psi(e_i)\} = 0$ for our model (1.1), and is needed for the consistency of the M -estimator. If e_i in model (1.1) has a symmetric distribution with zero mean, then we have $E\{\psi(e_i)\} = 0$ for Huber's score.

LEMMA 3.1. Under assumptions (A1) and (A2), we have that, for a fixed $\boldsymbol{\theta} \in \mathbb{R}^2$,

$$\begin{aligned}
 & \iint \psi(x - \boldsymbol{\theta}^T \mathbf{t}) \mathbf{t} \, dH(x, t) \\
 (3.4) \quad &= \int \left\{ \int H(x - \cdot, t - \cdot) \boldsymbol{\psi}'_{x, \boldsymbol{\theta}}(dt) \right\} dx + \mu(H) \psi(b - \boldsymbol{\theta}^T \mathbf{c}) \\
 & \quad - \mathbf{c} \int H(x - \cdot, \infty) \, d\psi(x - \boldsymbol{\theta}^T \mathbf{c}) \\
 & \quad - \int H(\infty, t - \cdot) \boldsymbol{\psi}_0(dt) \quad \text{for } H \in \mathbb{M}_0,
 \end{aligned}$$

where $\mathbf{t} = (1, t)^T$, $\mathbf{c} = (1, c)^T$, $\boldsymbol{\psi}_0(t) = \psi(b - \boldsymbol{\theta}^T \mathbf{t}) \mathbf{t}$, $\boldsymbol{\psi}'_{x, \boldsymbol{\theta}}(t) = \psi'(x - \boldsymbol{\theta}^T \mathbf{t}) \mathbf{t}$, and $\mu(H) \equiv \mu_H([a, b] \times [0, c])$ for μ_H denoting the (signed) measure corresponding to H in $[a, b] \times [0, c]$.

LEMMA 3.2. (i) Under assumption (A1), if F is a bivariate d.f. and $R(\boldsymbol{\theta})$ given by (3.3) is defined for any $\boldsymbol{\theta}$, then $\Psi(\boldsymbol{\theta}, F) = \mathbf{0}$ has a unique solution. (ii) Under assumptions (A1)–(A3), for any sufficiently large $B > \|\boldsymbol{\beta}\|_2$, there exists $\eta > 0$ such that for any $H \in \mathbb{M}_0$ and $\|H - F\| \leq \eta$, $\Psi(\boldsymbol{\theta}, H) = \mathbf{0}$ has a solution $\boldsymbol{\theta}_H$ with $\|\boldsymbol{\theta}_H\|_2 \leq B$, and any solution $\boldsymbol{\theta}_H$ of such satisfies $\|\boldsymbol{\theta}_H - \boldsymbol{\beta}\|_2 \rightarrow 0$, as $\|H - F\| \rightarrow 0$.

The proofs of Lemma 3.1 and Lemma 3.2 are given in Section 6. Lemma 3.2 shows that the functional $\tau(\cdot)$ is defined in the neighborhood of any bivariate d.f. F . One may note that although it may not be a proper bivariate d.f., \hat{F}_n given by (2.6) corresponds to a finite signed measure on \mathbb{R}^2 , thus $\hat{F}_n \in \mathbb{M}_0$. Hence, from Theorem 2.1, $\tau(\cdot)$ is defined asymptotically for our bivariate d.f. estimator \hat{F}_n based on (V_i, δ_i, T_i) , $1 \leq i \leq n$. One may also note that for \hat{F}_n in

the neighborhood of F , if there are multiple roots for $\Psi(\boldsymbol{\theta}, \hat{F}_n) = \mathbf{0}$ on a large compact set, the asymptotic results established in Theorem 3.1 below still hold because of Lemma 3.2(ii).

Before stating our asymptotic normality results on the regression M -estimators with doubly censored observations, we give the definition of Hadamard differentiability (or compact differentiability) as follows [Gill (1989)]. Let \mathfrak{B}_1 and \mathfrak{B}_2 be two Banach spaces and $\mathfrak{L}(\mathfrak{B}_1, \mathfrak{B}_2)$ be the set of continuous linear transformation from \mathfrak{B}_1 to \mathfrak{B}_2 .

DEFINITION 3.1. Let \mathfrak{D} be an open set of \mathfrak{B}_1 . A functional $\tau: \mathfrak{D} \rightarrow \mathfrak{B}_2$ is Hadamard differentiable (or compact differentiable) at $F \in \mathfrak{D}$ if there exists $\tau'_F \in \mathfrak{L}(\mathfrak{B}_1, \mathfrak{B}_2)$ such that for any sequence $H_n \in \mathfrak{B}_1$ and $t_n \in \mathbb{R}$ which satisfy $H_n \rightarrow H \in \mathfrak{B}_1$ and $t_n \rightarrow 0$, as $n \rightarrow \infty$,

$$(3.5) \quad \lim_{n \rightarrow \infty} \frac{\tau(F + t_n H_n) - \tau(F) - \tau'_F(t_n H_n)}{t_n} = 0.$$

The linear functional τ'_F is called the Hadamard derivative of $\tau(\cdot)$ at F .

In Theorem 3.1, we show that the functional $\tau(\cdot)$ defined by (3.1) is Hadamard differentiable at the bivariate d.f. F of (X_i, T_i) with the proof deferred to Section 6. One may note that our functional $\tau(\cdot)$ is implicitly defined by (3.1). The implicit function theorem through *Compact Preserving* by Fernholz (1993) is used in our proofs. Some detailed discussions on implicit function theorems can be found in Gill (1989).

THEOREM 3.1. Under assumptions (A1)–(A3), the functional $\tau: \mathbb{M}_0 \rightarrow \mathbb{R}^2$, defined by (3.1), is Hadamard differentiable at F with Hadamard derivative

$$(3.6) \quad \tau'_F(H) = \mathbf{A}^{-1} \iint \psi(x - \boldsymbol{\beta}^T \mathbf{t}) \mathbf{t} dH(x, t),$$

where $H \in \mathbb{M}_0$ [if $H \notin \mathbb{M}_0$, the integration in (3.6) is defined by (3.4)], and

$$(3.7) \quad \mathbf{A} = \iint \psi'(x - \boldsymbol{\beta}^T \mathbf{t}) \begin{bmatrix} 1 & t \\ t & t^2 \end{bmatrix} dF(x, t).$$

Therefore, under the conditions of Corollary 2.1, the M -estimator $\tau(\hat{F}_n) = (\hat{\alpha}_n, \hat{\beta}_n)^T$ for linear regression model (1.1) based on doubly censored data (V_i, δ_i, T_i) , $1 \leq i \leq n$, given by (1.2), satisfies

$$(3.8) \quad \sqrt{n} [\tau(\hat{F}_n) - \tau(F)] \rightarrow_D N_2(\mathbf{0}, \boldsymbol{\Sigma}) \quad \text{as } n \rightarrow \infty,$$

where \hat{F}_n is given by (2.6) and $N_2(\mathbf{0}, \boldsymbol{\Sigma})$ denotes a zero-mean bivariate normal distribution with a covariance matrix $\boldsymbol{\Sigma}$.

REMARK 3. The M -estimators constructed in this paper are motivated by their robustness properties, and condition (A1) on the score function in Theorem 3.1 is satisfied by Huber’s score function. When there is no censor-

ing, M -estimators with Huber's score lose some efficiency, but limit the influence of outliers [Serfling (1980), page 247], which is also expected here for our proposed M -estimators with censored data. However, since the conditional distribution estimator $\hat{F}_{n,t}$ is efficient [Gu and Zhang (1993)] and \hat{F}_n is given as the product of $\hat{F}_{n,t}$ and \hat{G}_n [see (2.5)], with an estimated score ψ_n , better efficiency of our M -estimators may be achieved. The investigation of this will be discussed in this current paper.

One may note that the functional $\tau(\cdot)$ defined by (3.1) and the Hadamard differentiability of $\tau(\cdot)$ at F do not depend on observations in the sample. Hence, this functional plug-in method used to establish (3.8) for doubly censored data also applies to other types of censored data. Next, we give these immediate results as corollaries of Theorem 3.1.

Complete i.i.d. sample case. Suppose that for the linear regression model (1.1), a complete i.i.d. sample (X_i, T_i) , $i = 1, \dots, n$, is observed. Then, the empirical d.f. F_n given by (1.4) can be used to construct the M -estimator $\tau(F_n)$ for $\boldsymbol{\beta} = (\alpha, \beta)^T$, where $\tau(\cdot)$ is defined by (3.1). Since F_n is a proper bivariate d.f., by Lemma 3.2(i) we know that $\tau(F_n)$ is well defined. Since, by Theorem 3.1, $\tau(\cdot)$ is Hadamard differentiable at F , from (3.2) of Theorem 3.1 in Ren and Sen (1995) and from (3.6), we know that for a continuous F ,

$$\begin{aligned} \sqrt{n} [\tau(F_n) - \tau(F)] &= \tau'_F(\sqrt{n} [F_n - F]) + o_p(1) \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^n (\zeta_i - E_F\{\zeta_i\}) + o_p(1), \end{aligned}$$

where $\zeta_i = \mathbf{A}^{-1}(\psi(X_i - \alpha - \beta T_i), \psi(X_i - \alpha - \beta T_i)T_i)^T$. Since ζ_i , $1 \leq i \leq n$, are i.i.d. observations with zero mean, by the Central Limit Theorem, we obtain the asymptotic normality of the regression M -estimator $\tau(F_n)$. We state this result in the following corollary.

COROLLARY 3.1. *Assume (A1)–(A3), and assume that F is continuous. Let F_n be given by (1.4) and $\tau(F_n)$, defined by (3.1), be the M -estimator for linear regression model (1.1) with complete i.i.d. sample (X_i, T_i) , $1 \leq i \leq n$. Then,*

$$\sqrt{n} [\tau(F_n) - \tau(F)] \rightarrow_D N_2(\mathbf{0}, \boldsymbol{\Sigma}_0) \quad \text{as } n \rightarrow \infty,$$

where $\boldsymbol{\Sigma}_0 = \text{cov}_F(\zeta_i)$.

Bivariate observations under the univariate right censoring case. For any real numbers x and t , we denote $x \vee t = \max\{x, t\}$ and $x \wedge t = \min\{x, t\}$. Suppose that for the linear regression model (1.1), the following i.i.d. bivariate observations under univariate right censoring are observed:

$$(3.9) \quad \begin{aligned} \tilde{X}_i &= X_i \wedge C_i, & \tilde{T}_i &= T_i \wedge C_i, \\ \delta_i^x &= I\{X_i \leq C_i\}, & \delta_i^t &= I\{T_i \leq C_i\}, \end{aligned}$$

where $i = 1, \dots, n$, and C_i is the right censoring variable which is independent from (X_i, T_i) . This type of censoring is considered by Lin and Ying (1993). From (2.3) and the appendix of their work, a bivariate d.f. estimator for bivariate d.f. F of (X_i, T_i) can be obtained as

$$\hat{F}_n^c(x, t) = \frac{1}{n} \sum_{i=1}^n \frac{I\{\tilde{X}_i \leq x, \tilde{T}_i \leq t\}}{\hat{H}_n(x \vee t)} - \frac{1 - \hat{H}_n(x \vee t)}{\hat{H}_n(x \vee t)},$$

where \hat{H}_n is the product-limit survival function estimator based on $\tilde{C}_i = \tilde{X}_i \vee \tilde{T}_i$, $\delta_i^c = 1 - \delta_i^x \delta_i^t$ for $1 \leq i \leq n$, and the weak convergence of $\sqrt{n}[\hat{F}_n^c - F]$ to a centered Gaussian process on some compact set $[a_0, b_0] \times [0, c]$ can be obtained. Now, define a functional $\tau_0(\cdot)$ as the root of the equations

$$(3.10) \quad \Psi_0(\boldsymbol{\theta}, H) \equiv \int_{a_0}^{b_0} \int_0^c \psi(x - \boldsymbol{\theta}^T \mathbf{t}) \mathbf{t} dH(x, t) = \mathbf{0}.$$

Then, the regression M -estimator in model (1.1) based on data (3.9) can be constructed as $\tau_0(\hat{F}_n^c)$. From our Lemma 3.2(ii), we know that $\tau_0(\hat{F}_n^c)$ is defined when \hat{F}_n^c is close to F . By Theorem 3.1, we know that $\tau_0(\cdot)$ is Hadamard differentiable at F and its Hadamard derivative τ'_{0_F} is given by (3.6) with integration region $[a_0, b_0] \times [0, c]$. From the weak convergence of \hat{F}_n^c and from Theorem II.8.1. of Andersen, Borgan, Gill and Keiding (1993), we know that

$$\sqrt{n}[\tau_0(\hat{F}_n^c) - \tau_0(F)] = \tau'_{0_F}(\sqrt{n}[\hat{F}_n^c - F]) + o_p(1).$$

From $\tau'_{0_F}(\cdot)$, Lemma 3.1, and the weak convergence of $\sqrt{n}[\hat{F}_n^c - F]$, we obtain that $\tau'_{0_F}(\sqrt{n}[\hat{F}_n^c - F])$ converges in distribution to a bivariate normal distribution. We state this result in the following corollary.

COROLLARY 3.2. *Assume (A1) and (A2). Under the conditions that $\sqrt{n}[\hat{F}_n^c - F]$ weakly converges to a centered Gaussian process on a compact set $[a_0, b_0] \times [0, c]$, the regression M -estimator $\tau_0(\hat{F}_n^c)$, defined by (3.10), for model (1.1) based on data (3.9) satisfies*

$$\sqrt{n}[\tau_0(\hat{F}_n^c) - \tau_0(F)] \rightarrow_D N_2(\mathbf{0}, \boldsymbol{\Sigma}_c) \quad \text{as } n \rightarrow \infty,$$

where $\boldsymbol{\Sigma}_c$ is the covariance matrix determined by τ'_{0_F} and the limiting covariance of $\sqrt{n}[\hat{F}_n^c - F]$, which can be derived from (2.4) of Lin and Ying (1993).

The bivariate right-censored sample case. Suppose that for the linear regression model (1.1), the following i.i.d. bivariate right-censored sample is observed:

$$(3.11) \quad \begin{aligned} \tilde{X}_i &= X_i \wedge C_i, & \tilde{T}_i &= T_i \wedge D_i, \\ \delta_i^x &= I\{X_i \leq C_i\}, & \delta_i^t &= I\{T_i \leq D_i\}, \end{aligned}$$

where $i = 1, \dots, n$, and (C_i, D_i) is the bivariate right censoring variable which is independent of (X_i, T_i) . This type of censoring is considered by Dabrowska (1988, 1989), among others. From the bivariate survival function estimator of Dabrowska (1988), page 1484, a bivariate d.f. estimator \hat{F}_n^{cd} using data (3.11) can be obtained, and from Dabrowska (1989), the weak convergence of $\sqrt{n}[\hat{F}_n^{cd} - F]$ to a centered Gaussian process on some compact set $[a_0, b_0] \times [0, c]$ can be obtained. Thus, the regression M -estimator in model (1.1) based on data (3.11) can be constructed as $\tau_0(\hat{F}_n^{cd})$, where τ_0 is defined by (3.10). The asymptotic normality of this estimator $\tau_0(\hat{F}_n^{cd})$ follows from the proof of Corollary 3.2 discussed above. We state this result in Corollary 3.3.

COROLLARY 3.3. *Assume (A1) and (A2). Under the conditions that $\sqrt{n}[\hat{F}_n^{cd} - F]$ converges weakly to a centered Gaussian process on a compact set $[a_0, b_0] \times [0, c]$, the regression M -estimator $\tau_0(\hat{F}_n^{cd})$, defined by (3.10), for model (1.1) based on data (3.11) satisfies*

$$\sqrt{n}[\tau_0(\hat{F}_n^{cd}) - \tau_0(F)] \rightarrow_D N_2(\mathbf{0}, \Sigma_{cd}) \quad \text{as } n \rightarrow \infty$$

where Σ_{cd} is the covariance matrix determined by τ_0' and the limiting covariance of $\sqrt{n}[\hat{F}_n^{cd} - F]$.

REMARK 4. In Corollary 3.2 and 3.3, $\tau_0(F)$ is well defined, but may not be equal to β . They are almost the same if (A3) holds and $[a_0, b_0] \times [0, c]$ is sufficiently close to the support of (X, T) .

4. Computation and example. In this section, we consider the computation of the regression M -estimator $\tau(\hat{F}_n)$ for $\tau(\cdot)$ defined by (3.1) and \hat{F}_n given by (2.6), and its application to a doubly censored data set encountered in the study of primary breast cancer [Peer, Van Dijck, Hendriks, Holland and Verbeek (1993)].

Without loss of generality, assume that $T_1 < \dots < T_n$ and all V_1, \dots, V_n are distinct. Then, for $t = T_k$ in (2.3) we have that $\hat{G}_n(T_k) = k/n$ and

$$Q_{n, T_k}^{(j)}(x) = \frac{Q_n^{(j)}(x, T_k)}{\hat{G}_n(T_k)} = \frac{1}{k} \sum_{i=1}^k I\{V_i \leq x, \delta_i = j\}, \quad j = 1, 2, 3.$$

Thus, (2.4) is equivalent to (2.2) of Mykland and Ren (1996) and can be computed by their algorithm (2.5) which gives

$$(4.1) \quad \hat{F}_{n, T_k}(x) = \sum_{i=1}^k a_{ki} I\{V_i \leq x\},$$

where $a_{ki} \geq 0$ with $\sum_{i=1}^k a_{ki} \leq 1$. One may note that condition (2.9) can be satisfied if a proper initial point in the algorithm is chosen. For detailed discussion, see Mykland and Ren (1996). Since $\hat{G}_n(T_k) = k/n$, by (2.5) we have

$$(4.2) \quad \hat{F}_n(x, T_k) = \frac{k}{n} \sum_{i=1}^k a_{ki} I\{V_i \leq x\}, \quad k = 1, \dots, n.$$

From (2.4), we can see easily that the equation changes only according to $T_k \leq t < T_{k+1}$. Thus, we have that for any x and $t \geq T_1$,

$$\begin{aligned}
 \hat{F}_n(x, t) &= \sum_{k=1}^n \hat{F}_n(x, T_k) I\{T_k \leq t < T_{k+1}\} \\
 (4.3) \quad &= \sum_{k=1}^n \sum_{i=1}^k \left(\frac{k}{n} a_{ki} - \frac{k-1}{n} a_{k-1,i} \right) I\{V_i \leq x, T_k \leq t\} \\
 &= \sum_{k=1}^n \sum_{i=1}^k b_{ki} I\{V_i \leq x, T_k \leq t\},
 \end{aligned}$$

where $T_{n+1} = \infty$, $a_{k-1,k} = a_{0,i} = 0$ and $b_{ki} = (k/n)a_{ki} - (k-1/n)a_{k-1,i}$.

Since (2.4) does not have a unique solution [see Mykland and Ren (1996) or Gu and Zhang (1993)], the solution of (2.6) given by (4.3) is not unique. Nonetheless, since the asymptotic properties of \hat{F}_n established in Section 2 apply to any solution of (2.6) satisfying (2.9), then any solution of (2.6) satisfying (2.9) may be used to construct the regression M -estimator $\tau(\hat{F}_n)$ for the linear model (1.1) when the sample size is large.

For an \hat{F}_n given by (4.3), to find the regression M -estimator $\tau(\hat{F}_n)$ defined by (3.1), we need to solve the following equations:

$$\begin{aligned}
 (4.4) \quad &\sum_{k=1}^n \sum_{i=1}^k b_{ki} \psi(V_i - \theta_1 - \theta_2 T_k) = 0, \\
 &\sum_{k=1}^n \sum_{i=1}^k b_{ki} T_k \psi(V_i - \theta_1 - \theta_2 T_k) = 0.
 \end{aligned}$$

This is a system of nonlinear equations and can be solved using the Newton–Raphson method [Press, Teukolsky, Vetterling and Flannery (1992), pages 372–378]. To illustrate our proposed method, we apply the regression M -estimator $\tau(\hat{F}_n)$ defined by (3.1) to a real data set below.

EXAMPLE. In a recent study of the age-dependent growth rate of primary breast cancer (Peer, Van Dijck, Hendriks, Holland and Verbeek (1993); Ren and Peer (1997)], a doubly censored sample is encountered. The age X (in months), at which a tumor volume was developed, was observed among 236 women aged 41–84 years. From 1981 to 1990, serial screening mammograms with a mean screening interval of two years were obtained. Among the tumor volumes detected by the screening mammograms, 45 women had tumor volumes observed at the first screening mammograms, yielding left-censored observations; 79 did not have tumor volumes observed at the last screening mammograms, yielding right-censored observations and 112 were observed with tumor growth during the period of the serial screening mammograms, yielding uncensored observations. For each woman, the age T (in months) at which she started the first screening mammogram was recorded. To study the relation between X and T , which is an important issue in breast cancer research, we use the linear regression model (1.1) with data (V_i, δ_i, T_i) ,

$1 \leq i \leq 236$. In Figure 1, we display the scatterplot of (V_i, T_i) , $1 \leq i \leq 236$, which indicates that the linear model (1.1) might be appropriate for this data set. Using Huber's score function ψ given by

$$\psi(x) = \begin{cases} c, & \text{if } x > c, \\ x, & \text{if } -c \leq x \leq c, \\ -c, & \text{if } x < -c, \end{cases}$$

where $c = 330$, the regression M -estimator constructed for model (1.1) in Theorem 3.1 is calculated as $(\hat{\alpha}_n, \hat{\beta}_n) = (36.4, 1.03)$ by the methods discussed above using (V_i, δ_i, T_i) , $1 \leq i \leq 236$. The fitted regression line $\hat{y} = \hat{\alpha}_n + \hat{\beta}_n x$ is plotted in Figure 1 (for a different choice of c the fitted regression line does not appear to be very much different). Our experience shows that computation is efficient for a reasonable sample size. In Figure 1, we also plot the fitted regression line by the usual least squares estimate (LSE) method [i.e., the solution of (1.3) with score function $\psi(x) = x$] using (V_i, T_i) , $1 \leq i \leq 236$, which ignores censoring in the data. One may note that the fitted regression line by the proposed M -estimate method is located above that by the LSE method. This may very well be expected, since the proposed method takes the

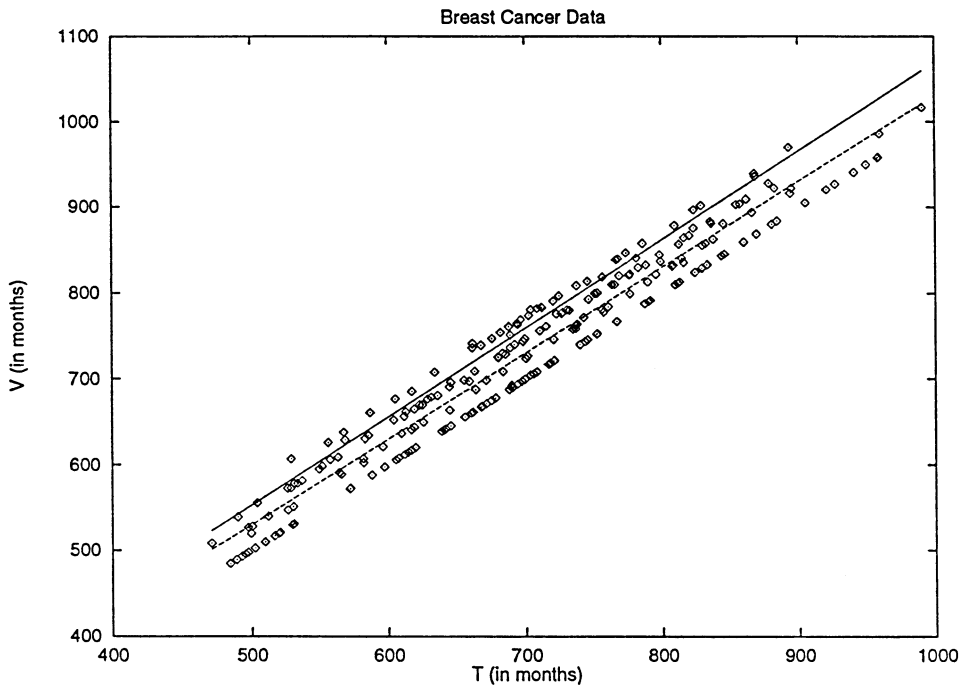


FIG. 1. —, fitted regression line by proposed method using (V_i, δ_i, T_i) , $1 \leq i \leq 236$; ----, fitted regression line by LSE method using (V_i, T_i) , $1 \leq i \leq 236$.

censoring mechanism of the data into account and the data set is more heavily right censored than left censored.

5. Proofs of Theorems 2.1 and 2.2.

PROOF OF THEOREM 2.1. Suppose that for each n , $\hat{F}_n(x, t)$ is a bivariate function given by (2.5) satisfying (2.9). Then, conditions (2.5) and (2.6) of Gu and Zhang (1993) are satisfied by (2.9) and (2.12). Applying Theorem 1 of Gu and Zhang (1993), we have that for each t with $F_T(t) > 0$, $\|\hat{F}_{n,t} - F_t\| \rightarrow 0$ almost surely as $n \rightarrow \infty$. Since

$$|\hat{F}_n(x, t) - F(x, t)| \leq |\hat{F}_{n,t}(x) - F_t(x)| |\hat{G}_n(t) + F_t(x)| |\hat{G}_n(t) - F_T(t)|,$$

we have that for each t , $\sup_x |\hat{F}_n(x, t) - F(x, t)| \rightarrow 0$ and $\sup_x |\hat{F}_n(x, t-) - F(x, t-)| \rightarrow 0$ almost surely as $n \rightarrow \infty$.

In the next step, we prove that the convergence is uniform in t . For any $\varepsilon > 0$, let $-\infty = t_0 \leq t_1 \leq \dots \leq t_k = \infty$ be a sequence of points such that $F_T(t_i) - F_T(t_{i-1}) \leq \varepsilon$, $i = 1, \dots, k$. From the first step, for almost all ω in the sample space, we can choose N_ω such that $\sup_x |\hat{F}_n(x, t_i) - F(x, t_i)| \leq \varepsilon$ and $\sup_x |\hat{F}_n(x, t_i-) - F(x, t_i-)| \leq \varepsilon$ for $n \geq N_\omega$ and $i = 0, 1, \dots, k$. Since we have

$$\begin{aligned} |\hat{F}_n(x, t) - F(x, t)| &\leq \max_{0 \leq i \leq k} |\hat{F}_n(x, t_i) - F(x, t_i)| \\ &\quad + \max_{0 \leq i \leq k} |\hat{F}_n(x, t_i-) - F(x, t_i-)| \\ &\quad + \max_{1 \leq i \leq k} |F(x, t_i-) - F(x, t_i)|, \end{aligned}$$

we see that $|\hat{F}_n(x, t) - F(x, t)|$ is bounded by 3ε on the same ω when $n \geq N_\omega$. This shows that the convergence is uniform in t almost surely. \square

Before proving Theorem 2.2, we need to define some notation. With F reserved for the true bivariate distribution function of (X, T) , we denote F'_m , $F'_{Y,m}$, $F'_{Z,m}$, $m \geq 1$ and F' as distribution functions such that

$$F' - F \in D_0([a, b] \times \mathbb{R}),$$

$$F'(x, t) = 0 \quad \Rightarrow \quad F'_m(x, t) = 0; \quad S'(x, t) = 0 \quad \Rightarrow \quad S'_m(x, t) = 0,$$

$$\|F'_m - F'\| \rightarrow 0, \quad \|F'_{Y,m} - F'_Y\| \rightarrow 0, \quad \|F'_{Z,m} - F'_Z\| \rightarrow 0,$$

$$K_m = F'_{Z,m} - F'_{Y,m},$$

where $S'_m(x, t) = F'_m(\infty, t) - F'_m(x, t)$ and $S'(x, t) = F'(\infty, t) - F'(x, t)$. With these definitions, we have a lemma similar to Lemma 2 of Gu and Zhang (1993).

LEMMA 5.1. Let $h_m, g_m, m \geq 1$, and g be functions in $D_0([a, b] \times \mathbb{R})$ such that $\|g_m - g\| \rightarrow 0$ and $R_m h_m = g_m$ and A_m, R_m and \mathbb{K}_m are defined as in

(2.17) with (F, F_Y, F_Z) replaced by $(F'_m, F_{Y,m}, F_{Z,m})$. Suppose that the conditions of Theorem 2.2 hold and for all $t > t_0$, where t_0 is a fixed number such that $F(\infty, t_0) > 0$,

$$(5.1) \quad \lim_{\tau \rightarrow 0^+} \sup_{m, t \geq t_0} \left\{ \int_{0 < F(u, t) < \tau} \frac{dF_{Y,m}(u)}{K_m(u)} + \int_{0 < S(u, t) < \tau} \frac{dF_{Z,m}(u)}{K_m(u)} \right\} = 0.$$

Let R_F be given by (2.17). Then there exists $h \in D_K([a, b] \times \mathbb{R})$ such that $\|\mathbb{K}_m h_m - \mathbb{K}h\| \rightarrow 0$, as $m \rightarrow \infty$, and $R_F h = g$.

PROOF. First, we show that if $\|\mathbb{K}_m h_m\| \leq 1$, then $\{\mathbb{K}_m h_m, m \geq 1\}$ is totally bounded on the space $D_0([a, b] \times [t_0, \infty))$. The proof of this is split into three steps.

Step 1. Define

$$(5.2) \quad v_m^+(x; t', t) = \int_{x < u} \left\{ \frac{F'_m(x, t')}{F'_m(u, t')} - \frac{F'_m(x, t)}{F'_m(u, t)} \right\} h_m(u, t) dF_{Z,m}(u),$$

$$(5.3) \quad v_m^-(x; t', t) = \int_{u \leq x} \left\{ \frac{S'_m(x, t')}{S'_m(u, t')} - \frac{S'_m(x, t)}{S'_m(u, t)} \right\} h_m(u, t) dF_{Y,m}(u).$$

We are going to show

$$(5.4) \quad \begin{aligned} & \lim_{\delta \rightarrow 0} \sup_{x, t', t} \{ |v_m^+(x; t', t)|; t_0 \leq t < t', |t' - t| \leq \delta \} = 0, \\ & \lim_{\delta \rightarrow 0} \sup_{x, t', t} \{ |v_m^-(x; t', t)|; t_0 \leq t < t', |t' - t| \leq \delta \} = 0. \end{aligned}$$

The argument of Step 2 in the proof of Lemma 2 of Gu and Zhang (1993) can be used to show that

$$\lim_{\tau_0 \rightarrow 0} \sup_{x, t \geq t_0} \left\{ \left| \int_{x < u} \frac{F'_m(x, t')}{F'_m(u, t')} h_m(u, t) dF_{Z,m}(u) \right|; F(x, t') < \tau_0 \right\} = 0,$$

and the same equation holds, with the argument (x, t') and (u, t') in the integration replaced by (x, t) and (u, t) , respectively, since $F(x, t') > F(x, t)$.

To prove the first half of (5.4), we are left with the case $F(x, t') \geq \tau_0$. We have

$$|v_m^+(x; t', t)| \leq \|F'_m(\cdot, t') - F'_m(\cdot, t)\| \frac{2}{\tau_0} \int_{x < u} \frac{dF_{Z,m}(u)}{K_m(u)}$$

since

$$\begin{aligned} \frac{F'_m(x, t')}{F'_m(u, t')} - \frac{F'_m(x, t)}{F'_m(u, t)} &= \frac{F'_m(x, t') - F'_m(x, t)}{F'_m(u, t')} \\ &\quad - \frac{F'_m(x, t)}{F'_m(u, t)} \frac{F'_m(u, t') - F'_m(u, t)}{F'_m(u, t')}. \end{aligned}$$

The second half of (5.4) can be proved in the same way. Details are omitted.

Step 2. We will show

$$(5.5) \quad \lim_{\delta \rightarrow 0} \sup_{x, t', t} \{ |\mathbb{K}_m h_m(x, t') - \mathbb{K}_m h_m(x, t)|; \tau \leq t < t', |t' - t| \leq \delta \} = 0.$$

Simple calculation shows that

$$(5.6) \quad \begin{aligned} R_{m, F'_m(\cdot, t)}(h_m(\cdot, t') - h_m(\cdot, t))(x) \\ = -v_m^+(x; t', t) - v_m^-(x; t', t) + g_m(x, t') - g_m(x, t), \end{aligned}$$

where $R_{m, F'_m(\cdot, t)}$ is a one-dimensional operator as in (2.9) of Gu and Zhang (1993) with $(F'_m(\cdot, t), F_{Y, m}, F_{Z, m})$. Lemma 2 of Gu and Zhang (1993) shows that the operator R_{m, F'_m}^{-1} is continuous in terms of its defining function $(F'_m, F_{Y, m}, F_{Z, m})$ (with supremum norm). Thus, (5.5) follows from (5.4) and (5.6). Moreover, we have

$$(5.7) \quad \sup_{t \in [t_0, \infty)} \|R_{m, F'_m(\cdot, t)}^{-1}\| < \infty.$$

Hence, (5.5) follows from (5.6) and (5.7) combining with (5.4) of Step 1.

Step 3. Equation (5.5) shows the total boundedness of $\mathbb{K}_m h_m(x, t)$ with respect to t . The total boundedness of $\mathbb{K}_m h_m$ is established if we show it is totally bounded with respect to x . The arguments in Step 1 and Step 2 of the proof of Lemma 2 of Gu and Zhang (1993) can be used with the observation that the inequalities and limits there are all uniform in t with $t \geq t_0$.

With (5.7), the proof of the total boundedness of $\mathbb{K}_m h_m$ on the space $D_0([a, b] \times [t_0, \infty))$ follows exactly the argument in the proof of Lemma 2 of Gu and Zhang (1993). We omit the details. \square

PROOF OF THEOREM 2.2. We first observe that since for $t < t'$, $F(u, t') \leq \tau$ implies $F(u, t) \leq \tau$ and $S(u, t') \leq \tau$ implies $S(u, t) \leq \tau$, condition (2.22) implies that for any t_0 with $F(\infty, t_0) > 0$,

$$\lim_{\tau \rightarrow 0} \sup_{t \geq t_0} \left\{ \int_{0 < F(u, t) < \tau} \frac{dF_Y(u)}{K(u)} + \int_{0 < S(u, t) < \tau} \frac{dF_Z(u)}{K(u)} \right\} = 0,$$

which in turn, implies the condition of Lemma 5.1 if we discretize the distribution F , F_Y and F_Z . The proof for the weak convergence of $\sqrt{n}(\hat{F}_n - F)$ on the set $[a, b] \times [t_0, \infty)$ follows from the one for Theorem 2 of Gu and Zhang (1993) with Lemma 2 there replaced by Lemma 5.1 in this paper. The weak convergence of $\sqrt{n}(\hat{F}_n - F)$ on the set $[a, b] \times (-\infty, t_0]$ can be deduced in the same way as above by noting that

$$\hat{F}_n(x, t) - F(x, t) = (\hat{F}_n(x, \infty) - F(x, \infty)) - (\hat{\bar{F}}_n(x, t) - \bar{F}(x, t)),$$

where $\bar{F}(x, t) = P\{X \leq x, T > t\}$ and $\hat{\bar{F}}_n$ is the corresponding estimator of \bar{F} . We observe that if \hat{F}_n satisfies (2.6), then $\hat{\bar{F}}_n$ satisfies (2.6) with the corresponding changes in the definitions for $Q_n^{(j)}$, $j = 0, 1, 2, 3$; thus $\hat{\bar{F}}_n$ satisfies the corresponding equation (2.18). Finally, we note that $F(\infty, t_0) \leq \delta$ implies that $\bar{F}(\infty, t_0) \geq 1 - \delta$. Therefore Lemma 5.1 again can be applied. The details are omitted. \square

6. Proofs of Lemmas 3.1, 3.2 and Theorem 3.1.

PROOF OF LEMMA 3.1. From (A1) and (A2), we know that for any fixed x and $\boldsymbol{\theta}$, each of $\boldsymbol{\psi}_{\boldsymbol{\theta}}(t) = \psi(b - \boldsymbol{\theta}^T \mathbf{t})$ and $\boldsymbol{\psi}'_{x, \boldsymbol{\theta}}(t) = \psi'(x - \boldsymbol{\theta}^T \mathbf{t})$ is of bounded variation on $[0, c]$ and

$$(6.1) \quad \int_0^c |\boldsymbol{\psi}_{\boldsymbol{\theta}}(dt)| < \infty.$$

Since $\psi' = 0$ outside of $[d_1, d_2]$, then for a fixed $\boldsymbol{\theta}$, there exist $-\infty < a' < b' < \infty$ such that for any $H \in \mathbb{M}_0$,

$$(6.2) \quad \begin{aligned} & \int \left\{ \int H(x - \cdot, t - \cdot) \boldsymbol{\psi}'_{x, \boldsymbol{\theta}}(dt) \right\} dx \\ &= \int_{a'}^{b'} \left\{ \int_0^c H(x - \cdot, t - \cdot) \boldsymbol{\psi}'_{x, \boldsymbol{\theta}}(dt) \right\} dx, \\ & \iint \psi(x - \boldsymbol{\theta}^T \mathbf{t}) \mathbf{t} dH(x, t) \\ &= \int_{a'}^{b'} \int_0^c \psi(x - \boldsymbol{\theta}^T \mathbf{t}) \mathbf{t} dH(x, t) \\ & \quad + C_1 \int_a^{a'} \int_0^c \mathbf{t} dH(x, t) + C_2 \int_{b'}^b \int_0^c \mathbf{t} dH(x, t), \end{aligned}$$

where C_1 and C_2 are constants, and

$$(6.3) \quad \int_{a'}^{b'} \int_0^c |\boldsymbol{\psi}'_{x, \boldsymbol{\theta}}(dt)| dx < \infty.$$

It suffices to show (3.4) for all bivariate d.f. H . First, it is easy to check that (3.4) holds for $H(x, t) = I\{A \leq x, B \leq t\}$, where $A \in [a, b]$ and $B \in [0, c]$. This implies that for any bivariate d.f. H , (3.4) holds for an empirical d.f. H_N based on a random sample of size N from H . Letting $N \rightarrow \infty$, the proof follows from (6.1)–(6.3). \square

PROOF OF LEMMA 3.2(i). From (A1), we know that ρ is nonnegative, continuous and convex with $\lim_{x \rightarrow \pm\infty} \rho(x) = \infty$. Thus, for any bivariate d.f. F , $R(\boldsymbol{\theta})$ given by (3.3) is convex and continuous, and by (A1), it is twice differentiable. From Bazaraa, Sherali and Shetty [(1993), page 118], we know that if $R(\boldsymbol{\theta})$ attains its global minimum at some point $\boldsymbol{\theta}_0$, then its gradient

$$(6.4) \quad \nabla R(\boldsymbol{\theta}) = - \iint \psi(x - \boldsymbol{\theta}^T \mathbf{t}) \mathbf{t} dF(x, t)$$

must satisfy $\nabla R(\boldsymbol{\theta}_0) = \mathbf{0}$. Thus $\Psi(\boldsymbol{\theta}_0, F) = \mathbf{0}$, because $\nabla R(\boldsymbol{\theta}) = -\Psi(\boldsymbol{\theta}, F)$. Hence, to show the existence of a solution of $\Psi(\boldsymbol{\theta}, F) = \mathbf{0}$, it suffices to show that $R(\boldsymbol{\theta})$ has a global minimum. Since $R(\boldsymbol{\theta})$ is continuous, it suffices to show that

$$\lim_{\|\boldsymbol{\theta}\|_2 \rightarrow \infty} R(\boldsymbol{\theta}) = \infty,$$

which is equivalent to

$$(6.5) \quad \lim_{\lambda \rightarrow \infty} \inf_{\|\mathbf{e}\|_2=1} R(-\lambda \mathbf{e}) = \infty.$$

Let $\mathbf{e} = (e_1, e_2)^T$ with $e_1^2 + e_2^2 = 1$. Suppose $e_1 \geq 0$. Since $\rho(x) \rightarrow \infty$, as $|x| \rightarrow \infty$, then for any $M > 0$, there exists $A_M > 0$ such that $\rho(x) \geq M$ for $|x| \geq A_M$. Denote μ_F as the measure corresponding to F on \mathbb{R}^2 , without loss of generality, we may assume that $\mu_F\{|x| \leq 1, t \geq 1\} > 0$ and $\mu_F\{|x| \leq 1, 0 \leq t \leq \frac{1}{2}\} > 0$. If $e_2 \geq 0$, then for $|x| \leq 1$ and $t \geq 1$, we have

$$x + \lambda \mathbf{e}^T \mathbf{t} = x + \lambda(e_1 + e_2 t) \geq -1 + \lambda(e_1 + e_2) \geq -1 + \lambda$$

because $(e_1 + e_2)^2 = 1 + 2e_1 e_2 \geq 1$. Hence, for large enough λ , we have $x + \lambda \mathbf{e}^T \mathbf{t} \geq A_M$ and

$$(6.6) \quad \begin{aligned} \iint \rho(x + \lambda \mathbf{e}^T \mathbf{t}) dF &\geq \iint_{|x| \leq 1, t \geq 1} \rho(x + \lambda \mathbf{e}^T \mathbf{t}) dF \\ &\geq M \mu_F\{|x| \leq 1, t \geq 1\}. \end{aligned}$$

If $e_2 \leq 0$, then $e_2 \leq -\sqrt{\frac{3}{4}}$ when $0 \leq e_1 \leq \frac{1}{2}$, and $-\sqrt{\frac{3}{4}} \leq e_2 \leq 0$ when $e_1 \geq \frac{1}{2}$. For $0 \leq e_1 \leq \frac{1}{2}$, $e_2 \leq -\sqrt{\frac{3}{4}}$, $|x| \leq 1$, $t \geq 1$, we have

$$x + \lambda \mathbf{e}^T \mathbf{t} = x + \lambda(e_1 + e_2 t) \leq 1 + \frac{1 - \sqrt{3}}{2} \lambda;$$

thus for large enough λ , we have $x + \lambda \mathbf{e}^T \mathbf{t} \leq -A_M$ and (6.6). For $e_1 \geq \frac{1}{2}$, $-\sqrt{\frac{3}{4}} \leq e_2 \leq 0$, $|x| \leq 1$, $0 \leq t \leq \frac{1}{2}$, we have

$$x + \lambda \mathbf{e}^T \mathbf{t} = x + \lambda(e_1 + e_2 t) \geq -1 + \lambda\left(\frac{1}{2} - \sqrt{\frac{3}{4}} t\right) \geq -1 + \frac{2 - \sqrt{3}}{4} \lambda;$$

thus for large enough λ , we have $x + \lambda \mathbf{e}^T \mathbf{t} \geq A_M$ and

$$\begin{aligned} \iint \rho(x + \lambda \mathbf{e}^T \mathbf{t}) dF &\geq \iint_{|x| \leq 1, 0 \leq t \leq 1/2} \rho(x + \lambda \mathbf{e}^T \mathbf{t}) dF \\ &\geq M \mu_F\{|x| \leq 1, 0 \leq t \leq \frac{1}{2}\}. \end{aligned}$$

This completes the proof for (6.5) when $e_1 \geq 0$. Similarly, (6.5) can be shown for the case of $e_1 \leq 0$.

Suppose that $\Psi(\boldsymbol{\theta}, F) = \mathbf{0}$ has two different solutions $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$. Then from Bazaraa, Sherali and Shetty [(1993), page 118], we know that $R(\boldsymbol{\theta})$ attains its global minimum at $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$. From convexity of $R(\boldsymbol{\theta})$, we know that $h(\lambda) = R(\lambda \boldsymbol{\theta}_1 + (1 - \lambda) \boldsymbol{\theta}_2) \equiv$ “the minimum value of $R(\boldsymbol{\theta})$,” for $0 \leq \lambda \leq 1$, thus, for any $0 \leq \lambda \leq 1$, $\lambda \boldsymbol{\theta}_1 + (1 - \lambda) \boldsymbol{\theta}_2$ is a solution of $\Psi(\boldsymbol{\theta}, F) = \mathbf{0}$. Hence, we have

$$\begin{aligned} 0 \equiv h''(\lambda) &= \iint \psi'(x - (\lambda \boldsymbol{\theta}_1 + (1 - \lambda) \boldsymbol{\theta}_2)^T \mathbf{t}) \{(\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2)^T \mathbf{t}\}^2 dF(x, t), \\ &\lambda \in (0, 1), \end{aligned}$$

which implies $\psi'(x - \frac{1}{2}(\boldsymbol{\theta}_1 + \boldsymbol{\theta}_2)^T \mathbf{t}) \equiv 0$ for any $x \in [a, b]$, $t \in [0, c]$. This means $\psi(x - \frac{1}{2}(\boldsymbol{\theta}_1 + \boldsymbol{\theta}_2)^T \mathbf{t}) \equiv \zeta$ for any $x \in [a, b]$, $t \in [0, c]$. From (A1), we know that there exists a unique point x_0 in \mathbb{R} such that $\psi(x_0) = 0$. Hence, we must have $\zeta \neq 0$. Since $\frac{1}{2}(\boldsymbol{\theta}_1 + \boldsymbol{\theta}_2)$ is a solution of $\Psi(\boldsymbol{\theta}, F) = \mathbf{0}$, we have

$$\mathbf{0} = \iint \Psi(x - \frac{1}{2}(\boldsymbol{\theta}_1 + \boldsymbol{\theta}_2)^T \mathbf{t}) \mathbf{t} dF(x, t) = \zeta \iint \mathbf{t} dF(x, t) \neq \mathbf{0},$$

a contradiction. Therefore, the solution of $\Psi(\boldsymbol{\theta}, F) = \mathbf{0}$ is unique. \square

PROOF OF LEMMA 3.2(ii). Without loss of generality, we consider the case of $-\infty < a < \infty$ and $b = \infty$, because other cases can be shown similarly.

Choose some number b' such that $a < b' < \infty$ and denote

$$(6.7) \quad R'_H(\boldsymbol{\theta}) = \int_a^{b'} \int_0^c \rho(x - \boldsymbol{\theta}^T \mathbf{t}) dH(x, t), \quad H \in \mathbb{M}_0.$$

From the proof of (6.5), we know that $R'_F(\boldsymbol{\theta}) \rightarrow \infty$, as $\|\boldsymbol{\theta}\|_2 \rightarrow \infty$. Let $R_F(\boldsymbol{\beta}) = M_\beta$ for $R_F(\boldsymbol{\theta})$ given by (3.3), then there exists $\boldsymbol{\theta}_0 \in \mathbb{R}^2$ such that $R'_F(\boldsymbol{\theta}_0) = M > M_\beta$, and there exists $A_M > \|\boldsymbol{\beta}\|_2$ such that

$$(6.8) \quad R'_F(\boldsymbol{\theta}) \geq M, \quad \text{for } \|\boldsymbol{\theta}\|_2 \geq A_M.$$

We choose a real number ζ such that

$$(6.9) \quad 0 < \zeta < \frac{1}{2}(M - M_\beta).$$

Since $\psi' = 0$ outside of $[d_1, d_2]$, there exists b'' such that $b' \leq b'' < \infty$ with $\psi(x - \boldsymbol{\theta}^T \mathbf{t}) \equiv C_1 = \psi(\infty)$ for $\|\boldsymbol{\theta}\|_2 \leq 2A_M$, $x \geq b''$, $t \in [0, c]$, and

$$(6.10) \quad \left| C_1 \int_{b''}^\infty \int_0^c \boldsymbol{\theta}^T \mathbf{t} dF(x, t) \right| \leq \zeta/2 \quad \text{for } \|\boldsymbol{\theta}\|_2 \leq 2A_M.$$

Note that for $H \in \mathbb{M}_0$ and $\|\boldsymbol{\theta}\|_2 \leq 2A_M$,

$$(6.11) \quad \Psi(\boldsymbol{\theta}, H) = \int_a^{b''} \int_0^c \psi(x - \boldsymbol{\theta}^T \mathbf{t}) \mathbf{t} dH(x, t) + C_1 \int_{b''}^\infty \int_0^c \mathbf{t} dH(x, t),$$

which is the negative gradient of

$$(6.12) \quad \varphi_H(\boldsymbol{\theta}) = \int_a^{b''} \int_0^c \rho(x - \boldsymbol{\theta}^T \mathbf{t}) dH(x, t) - C_1 \int_{b''}^\infty \int_0^c \boldsymbol{\theta}^T \mathbf{t} dH(x, t).$$

Denote

$$(6.13) \quad R''_H(\boldsymbol{\theta}) = \int_a^{b''} \int_0^c \rho'(x - \boldsymbol{\theta}^T \mathbf{t}) dH(x, t), \quad H \in \mathbb{M}_0,$$

then for $\|\boldsymbol{\theta}\|_2 \leq 2A_M$,

$$(6.14) \quad \begin{aligned} \varphi_H(\boldsymbol{\theta}) &= R''_F(\boldsymbol{\theta}) + R''_{H-F}(\boldsymbol{\theta}) - C_1 \int_{b''}^\infty \int_0^c \boldsymbol{\theta}^T \mathbf{t} d[H - F] \\ &\quad - C_1 \int_{b''}^\infty \int_0^c \boldsymbol{\theta}^T \mathbf{t} dF(x, t). \end{aligned}$$

Since $\rho' = \psi$ and ψ is bounded, we know that ρ is of bounded variation.

From the proof of Lemma 3.1, we can show that for any $H \in \mathbb{M}_0$,

$$\begin{aligned}
 R''_H(\boldsymbol{\theta}) &= \int_a^{b''} \left\{ \int_0^c H(x - \cdot, t -) d\psi(x - \boldsymbol{\theta}^T \mathbf{t}) \right\} dx \\
 (6.15) \quad &+ \mu'(H) \rho(b'' - \boldsymbol{\theta}^T \mathbf{c}) - \int_a^{b''} H(x - \cdot, \infty) d\rho(x - \boldsymbol{\theta}^T \mathbf{c}) \\
 &- \int_0^c H(\infty, t -) d\rho(b'' - \boldsymbol{\theta}^T \mathbf{t}),
 \end{aligned}$$

where $\mu'(H) = \mu_H([a, b''] \times [0, c])$ for μ_H given in (3.4), and similarly we also can show that for any $H \in \mathbb{M}_0$,

$$\int_{b''}^\infty \int_0^c t dH(x, t) = c[H(\infty, \infty) - H(b'', \infty)] - \int_0^c [H(\infty, t -) - H(b'', t -)] dt.$$

Hence, for any $H \in \mathbb{M}_0$, there exists $B_M > 0$ such that

$$\begin{aligned}
 |R''_{H-F}(\boldsymbol{\theta})| &\leq B_M \|H - F\|, \\
 (6.16) \quad \left| C_1 \int_{b''}^\infty \int_0^c \boldsymbol{\theta}^T \mathbf{t} d[H - F] \right| &\leq B_M \|H - F\| \quad \text{for } \|\boldsymbol{\theta}\|_2 \leq 2A_M.
 \end{aligned}$$

Let $0 < \eta < \zeta/(4B_M)$, then noting that ρ is nonnegative, by (6.14), (6.16) and (6.10), we have that for $H \in \mathbb{M}_0$ satisfying $\|H - F\| \leq \eta$,

$$\begin{aligned}
 (6.17) \quad \varphi_H(\boldsymbol{\theta}) &\geq R'_F(\boldsymbol{\theta}) - B_M \|H - F\| - B_M \|H - F\| - \zeta/2 \\
 &\geq M - \zeta, \quad \text{for } A_M \leq \|\boldsymbol{\theta}\|_2 \leq 2A_M,
 \end{aligned}$$

and by (6.14), (6.16), (6.10) and (6.9),

$$\begin{aligned}
 (6.18) \quad \varphi_H(\boldsymbol{\beta}) &\leq R'_F(\boldsymbol{\beta}) + B_M \|H - F\| + B_M \|H - F\| + \zeta/2 \\
 &\leq M_\beta + \zeta < M - \zeta \quad \text{for } \|\boldsymbol{\beta}\|_2 \leq A_M.
 \end{aligned}$$

Since it is continuous, $\varphi_H(\boldsymbol{\theta})$ must have a local minimum in $\|\boldsymbol{\theta}\|_2 < 2A_M$. Hence, from (6.11) and (6.12), we have that for $H \in \mathbb{M}_0$ satisfying $\|H - F\| \leq \eta$, $\Psi(\boldsymbol{\theta}, H) = \mathbf{0}$ has a solution in $\|\boldsymbol{\theta}\|_2 \leq A_M < 2A_M$.

Moreover, from Lemma 3.1 we know that $\Psi(\boldsymbol{\theta}, H)$ converges to $\Psi(\boldsymbol{\theta}, F)$ uniformly on any compact set of $\boldsymbol{\theta}$ when $\|H - F\| \rightarrow 0$. Thus, for any solution $\boldsymbol{\theta}_H$ satisfying $\Psi(\boldsymbol{\theta}_H, H) = \mathbf{0}$ and $\|\boldsymbol{\theta}_H\|_2 \leq A_M$, we have

$$\Psi(\boldsymbol{\theta}_H, F) = \Psi(\boldsymbol{\theta}_H, F) - \Psi(\boldsymbol{\theta}_H, H) \rightarrow \mathbf{0} \quad \text{as } \|H - F\| \rightarrow 0.$$

From the dominated convergence theorem and the uniqueness of the solution for $\Psi(\boldsymbol{\theta}, F) = \mathbf{0}$, we have that $\|\boldsymbol{\theta}_H - \boldsymbol{\beta}\|_2 \rightarrow 0$, as $\|H - F\| \rightarrow 0$. \square

PROOF OF THEOREM 3.1. First, we show that $\Psi(\boldsymbol{\theta}, H)$ given by (3.1) is Hadamard differentiable at $(\boldsymbol{\beta}, F)$ with Hadamard derivative

$$\begin{aligned}
 (6.19) \quad \Psi'_{(\boldsymbol{\beta}, F)}(\boldsymbol{\theta}, H) &= - \iint \psi'(x - \boldsymbol{\beta}^T \mathbf{t}) \boldsymbol{\theta}^T \mathbf{t} dF(x, t) \\
 &+ \iint \psi(x - \boldsymbol{\beta}^T \mathbf{t}) \mathbf{t} dH(x, t),
 \end{aligned}$$

where $H \in \mathbb{M}_0$. From Definition 3.1 of the Hadamard derivative, we need to show that for $t_n \rightarrow 0$, $\zeta_n \rightarrow \zeta \in \mathbb{R}^2$, $H_n \rightarrow H \in \mathbb{M}_0$, as $n \rightarrow \infty$, with $F + t_n H_n \in \mathbb{M}_0$,

$$(6.20) \quad \left\| \left\{ \Psi(\boldsymbol{\beta} + t_n \zeta_n, F + t_n H_n) - \Psi(\boldsymbol{\beta}, F) - t_n \Psi'_{(\boldsymbol{\beta}, F)}(\zeta_n, H_n) \right\} / t_n \right\|_2 \rightarrow 0 \quad \text{as } n \rightarrow \infty.$$

Note that from Lemma 3.1, we have

$$(6.21) \quad \begin{aligned} & \left\| \iint \left\{ \psi(x - \boldsymbol{\beta}^T \mathbf{t} - t_n \zeta_n^T \mathbf{t}) - \psi(x - \boldsymbol{\beta}^T \mathbf{t}) \right\} \mathbf{t} dH_n(x, t) \right\|_2 \\ & \leq M_\psi \|H_n - H\| + 4 \|H_n\| \|\boldsymbol{\psi}_n(\mathbf{c})\|_2 \\ & + \left\| \int \left\{ \int H(x - \cdot, t - \cdot) \boldsymbol{\psi}'_{x, n}(dt) \right\} dx \right. \\ & \quad \left. - \mathbf{c} \int H(x - \cdot, \infty) d \left[\psi(x - \boldsymbol{\beta}^T \mathbf{c} - t_n \zeta_n^T \mathbf{c}) - \psi(x - \boldsymbol{\beta}^T \mathbf{c}) \right] \right. \\ & \quad \left. - \int H(\infty, t - \cdot) \boldsymbol{\psi}_n(dt) \right\|_2, \end{aligned}$$

where $M_\psi > 0$ is a constant and

$$\begin{aligned} \boldsymbol{\psi}_n(t) &= [\psi(b - \boldsymbol{\beta}^T \mathbf{t} - t_n \zeta_n^T \mathbf{t}) - \psi(b - \boldsymbol{\beta}^T \mathbf{t})] \mathbf{t}, \\ \boldsymbol{\psi}'_{x, n}(t) &= [\psi'(x - \boldsymbol{\beta}^T \mathbf{t} - t_n \zeta_n^T \mathbf{t}) - \psi'(x - \boldsymbol{\beta}^T \mathbf{t})] \mathbf{t}. \end{aligned}$$

If $[a, b]$ is not finite, then for any n and $t \in [0, c]$, we have $\psi'_{x, n}(t) \equiv 0$ and $[\psi(x - \boldsymbol{\beta}^T \mathbf{c} - t_n \zeta_n^T \mathbf{c}) - \psi(x - \boldsymbol{\beta}^T \mathbf{c})] \equiv 0$ when $|x|$ is large enough. Hence, the integration region on the right-hand side of the inequality of (6.21) can always be equivalently considered as a compact set, say $[a', b'] \times [0, c]$. Since $H \in \mathbb{M}_0$, by Neuhaus (1971), we know that H can be approximated uniformly by a step function on $[a', b'] \times [0, c]$. Following the last part of the proof of Lemma 3 by Gill [(1989), pages 110–111], we can show that the last term on the right-hand side of the inequality (6.21) converges to 0 as $n \rightarrow \infty$. Thus,

$$(6.22) \quad \left\| \iint \left\{ \psi(x - \boldsymbol{\beta}^T \mathbf{t} - t_n \zeta_n^T \mathbf{t}) - \psi(x - \boldsymbol{\beta}^T \mathbf{t}) \right\} \mathbf{t} dH_n(x, t) \right\|_2 \rightarrow 0 \quad \text{as } n \rightarrow \infty.$$

Since

$$\begin{aligned} & \frac{1}{t_n} \left\{ \Psi(\boldsymbol{\beta} + t_n \zeta_n, F + t_n H_n) - \Psi(\boldsymbol{\beta}, F) - t_n \Psi'_{(\boldsymbol{\beta}, F)}(\zeta_n, H_n) \right\} \\ &= \iint \left\{ \frac{\psi(x - \boldsymbol{\beta}^T \mathbf{t} - t_n \zeta_n^T \mathbf{t}) - \psi(x - \boldsymbol{\beta}^T \mathbf{t})}{t_n} + \psi'(x - \boldsymbol{\beta}^T \mathbf{t}) \zeta_n^T \mathbf{t} \right\} \mathbf{t} dF(x, t) \\ & \quad + \frac{1}{t_n} \iint \left\{ \psi(x - \boldsymbol{\beta}^T \mathbf{t} - t_n \zeta_n^T \mathbf{t}) - \psi(x - \boldsymbol{\beta}^T \mathbf{t}) \right\} \mathbf{t} d(t_n H_n(x, t)), \end{aligned}$$

(6.20) follows from (6.22) and the dominated convergence theorem.

From Lemma 3.2(ii), we know that for $H \in \mathbb{M}_0$, $\Psi(\boldsymbol{\theta}, H) = \mathbf{0}$ has a solution in the neighborhood of $\boldsymbol{\beta}$ when H is in a neighborhood of F . We also know that for a fixed $H \in \mathbb{M}_0$, $\Psi(\boldsymbol{\theta}, H)$ is continuous and differentiable in $\boldsymbol{\theta}$. Thus, by the Implicit Function Theorem on \mathbb{R}^2 , we know that $\Psi(\boldsymbol{\theta}, H) = \mathbf{u}$ has a solution $T(\mathbf{u}, H) = \boldsymbol{\theta}$ for $\mathbf{u} \in \mathbb{R}^2$ in the neighborhood of $\mathbf{0}$, H in the neighborhood of F and $\boldsymbol{\theta}$ in the neighborhood of $\boldsymbol{\beta}$.

The partial derivative of $\Psi(\boldsymbol{\theta}, H)$ with respect to $\boldsymbol{\theta}$ at $(\boldsymbol{\beta}, F)$ is given by the matrix \mathbf{A} in (3.7). From (A3) and Remark 1 in Section 3, we have that for any $\mathbf{u} \in \mathbb{R}^2$,

$$\mathbf{u}^T \mathbf{A} \mathbf{u} = E\{\psi'(e_i)\}E\{(u_1 + u_2 T)^2\} > 0.$$

Hence, \mathbf{A} is positive definite, thus nonsingular.

To use the Implicit Function Theorem of Fernholz (1983), Theorem 3.2.4, to show that the functional $\tau(\cdot)$ defined by (3.1) is Hadamard differentiable at F , it suffices to verify the following compact preserving condition: if Γ is any compact set in $\overline{\mathbb{M}}_0$ and K a compact set in \mathbb{R}^2 , then for any $t_n \rightarrow 0$, as $n \rightarrow \infty$ and $\{(H_n, \boldsymbol{\zeta}_n)\} \subset \Gamma \times K$ with $F + t_n H_n \in \mathbb{M}_0$,

$$(6.23) \quad \{T(t_n \boldsymbol{\zeta}_n, F + t_n H_n) - T(\mathbf{0}, F)\} / t_n$$

is bounded. Let $\Psi(\boldsymbol{\theta}_n, F) = t_n \boldsymbol{\zeta}_n$, $T(t_n \boldsymbol{\zeta}_n, F) = \boldsymbol{\theta}_n$, $\Psi(\boldsymbol{\eta}_n, F + t_n H_n) = t_n \boldsymbol{\zeta}_n$ and $T(t_n \boldsymbol{\zeta}_n, F + t_n H_n) = \boldsymbol{\eta}_n$. Then, from (A1) it can be shown that there exist constants $C > 0$ and $M > 0$ such that for sufficiently large n ,

$$C \|\boldsymbol{\theta}_n - \boldsymbol{\beta}\|_2^2 \leq -t_n (\boldsymbol{\theta}_n - \boldsymbol{\beta})^T \boldsymbol{\zeta}_n \leq t_n \|\boldsymbol{\theta}_n - \boldsymbol{\beta}\|_2 \|\boldsymbol{\zeta}_n\|_2$$

and

$$\begin{aligned} C \|\boldsymbol{\eta}_n - \boldsymbol{\theta}_n\|_2^2 &\leq (\boldsymbol{\eta}_n - \boldsymbol{\theta}_n)^T \int \int \psi(x - \boldsymbol{\eta}_n^T \mathbf{t}) \mathbf{t} d[t_n H_n(x, t)] \\ &\leq t_n \|\boldsymbol{\eta}_n - \boldsymbol{\theta}_n\|_2 \|H_n\| M. \end{aligned}$$

Hence, (6.23) follows from the usual straightforward argument.

The asymptotic normality of $\tau(\hat{F}_n)$ follows from (3.6), (3.4), the weak convergence of $\sqrt{n}[\hat{F}_n - F]$, Theorem II.8.1 of Andersen, Borgan, Gill and Keiding (1993) and Iranpour and Chacon [(1988), pages 154–157]. \square

Acknowledgments. The authors would like to thank Dr. P. G. M. Peer at the Department of Medical Informatics and Epidemiology, University of Nijmegen, The Netherlands, for providing the breast cancer data set for our research. The authors would also like to thank one referee for detailed and helpful comments on the original version of this manuscript, which shortened some of the proofs in this paper.

REFERENCES

- ANDERSEN, P. K., BORGAN, Ø., GILL, R. D. and KIEIDING, N. (1993). *Statistical Models Based on Counting Processes*. Springer, New York.
- BAZARAA, M. S., SHERALI, H. D. and SHETTY, C. M. (1993). *Nonlinear Programming: Theory and Algorithms*. Wiley, New York.

- BILLINGSLEY, P. (1968). *Convergence of Probability Measures*. Wiley, New York.
- BRESLOW, N. and CROWLEY, J. (1974). A large sample study of the life table and product limit estimates under random censorship. *Ann. Statist.* **2** 437–443.
- BUCKLEY, J. and JAMES, I. (1979). Linear regression with censored data. *Biometrika* **66** 429–436.
- CAMPBELL, G. (1981). Nonparametric bivariate estimation with randomly censored data. *Biometrika* **68** 417–422.
- 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. and OAKES, D. (1984). *Analysis of Survival Data*. Chapman & Hall, New York.
- DABROWSKA, D. M. (1988). Kaplan–Meier estimate on the plane. *Ann. Statist.* **16** 1475–1489.
- DABROWSKA, D. M. (1989). Kaplan–Meier estimate on the plane: weak convergence, LIL, and the bootstrap. *J. Multivariate Anal.* **29** 308–325.
- EFRON, B. (1967). The two sample problem with censored data. *Proc. Fifth Berkeley Symp. Math. Statist. Probab.* **4** 831–853. Univ. California Press, Berkeley.
- FERNHOLZ, L. T. (1983). *Von Mises Calculus for Statistical Functional. Lecture Notes in Statist.* **19**. Springer, New York.
- GEHAN, E. A. (1965). A generalized two-sample Wilcoxon test for doubly censored data. *Biometrika* **52** 650–653.
- GILL, R. D. (1983). Large sample behavior of the product-limit estimator on the whole line. *Ann. Statist.* **11** 49–58.
- GILL, R. D. (1989). Non- and semiparametric maximum likelihood estimators and the von Mises method. *Scand. J. Statist.* **16** 97–128.
- GROENEBOOM, P. (1987). Asymptotics for incomplete censored observations. Technical Report 87-18, Dept. Mathematics, Univ. Amsterdam.
- GU, M. G. and ZHANG, C. H. (1993). Asymptotic properties of self-consistent estimators based on doubly censored data. *Ann. Statist.* **21** 611–624.
- HUBER, P. J. (1981). *Robust Statistics*. Wiley, New York.
- IRANPOUR, R. and CHACON, P. (1988). *Basic Stochastic Processes. The Mark Kac Lectures*. MacMillan, New York.
- KAPLAN, E. L. and MEIER, P. (1958). Nonparametric estimation from incomplete observations. *J. Amer. Statist. Assoc.* **53** 457–481.
- 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.
- LAI, T. L. and YING, Z. (1994). A missing information principle and M -estimators in regression analysis with censored and truncated data. *Ann. Statist.* **22** 1222–1255.
- LIN, D. Y. and YING, Z. L. (1993). A simple nonparametric estimator of the bivariate survival function under univariate censoring. *Biometrika* **80** 573–581.
- LEURGANS, S. (1987). Linear models, random censoring and synthetic data. *Biometrika* **74** 301–309.
- MYKLAND, P. A. and REN, J. (1996). Algorithms for computing self-consistent and maximum likelihood estimators with doubly censored data. *Ann. Statist.* **24** 1740–1764.
- NEUHAUS, G. (1971). On weak convergence of stochastic processes with multidimensional time parameter. *Ann. Math. Statist.* **42** 1285–1295.
- PEER, P. G., VAN DIJCK, J. A., HENDRIKS, J. H., HOLLAND, R. and VERBEEK, A. L. (1993). Age-dependent growth rate of primary breast cancer. *Cancer* **71** 3547–3551.
- POLLARD, D. (1984). *Convergence of Stochastic Processes*. Springer, New York.
- PRESS, W. H., TEUKOLSKY, S. A., VETTERLING, W. T. and FLANNERY, B. P. (1992). *Numerical Recipes*. Cambridge Univ. Press.
- REEDS, J. A. (1976). On the definition of Von Mises functional. Ph.D. dissertation, Harvard Univ., Cambridge, MA.

- REN, J. (1995). Generalized Cramér–von Mises tests of goodness of fit for doubly censored data. *Ann. Inst. Statist. Math.* **47** 525–549.
- REN, J. and SEN, P. K. (1995). Hadamard differentiability on $D[0, 1]^p$. *J. Multivariate Anal.* **55** 14–28.
- REN, J. and PEER, P. G. M. (1997). A study on effectiveness of screening mammograms in detection of primary breast cancer. Unpublished manuscript.
- RITOV, Y. (1990). Estimation in a linear regression model with censored data. *Ann. Statist.* **18** 303–328.
- SAMUELSEN, S. O. (1989). Asymptotic theory for non-parametric estimators from doubly censored data. *Scand. J. Statist.* **16** 1–21.
- SERFLING, R. J. (1980). *Approximation Theorems of Mathematical Statistics*. Wiley, New York.
- STUTE, W. (1993). Consistency estimation under random censorship when covariables are present. *J. Multivariate Anal.* **45** 89–103.
- TURNBULL, B. W. (1974). Nonparametric estimation of a survivorship function with doubly censored data. *J. Amer. Statist. Assoc.* **69** 169–173.
- WELLNER, J. A. (1982). Asymptotic optimality of the product limit estimator. *Ann. Statist.* **10** 595–602.
- ZHANG, C.-H. and LI, X. (1996). Linear regression with doubly censored data. *Ann. Statist.* **24** 2720–2743.
- ZHOU, M. (1992). Asymptotic normality of the synthetic data regression estimator for censored survival data. *Ann. Statist.* **20** 1002–1021.

DEPARTMENT OF MATHEMATICS
TULANE UNIVERSITY
NEW ORLEANS, LA 70118
E-MAIL: renj@ultra1.math.tulane.edu

DEPARTMENT OF MATHEMATICS AND STATISTICS
MCGILL UNIVERSITY
805 SHERBROOKE ST. W.
MONTREAL, QUEBEC H3A 2K6
CANADA