The Annals of Statistics
1994, Vol. 22, No. 3, 1222–1255

A MISSING INFORMATION PRINCIPLE AND M-ESTIMATORS IN REGRESSION ANALYSIS WITH CENSORED AND TRUNCATED DATA[1]

BY TZE LEUNG LAI AND ZHILIANG YING

Stanford University and Rutgers University

A general missing information principle is proposed for constructing M-estimators of regression parameters in the presence of left truncation and right censoring on the observed responses. By making use of martingale central limit theorems and empirical process theory, the asymptotic normality of M-estimators is established under certain assumptions. Asymptotically efficient M-estimators are also developed by using data-dependent score functions. In addition, robustness properties of the estimators are discussed and formulas for their influence functions are derived for the robustness analysis.

1. Introduction. Consider the linear regression model

$$y_i = \alpha + \beta^T x_i + \varepsilon_i, \qquad i = 1, 2, \dots, \tag{1.1}$$

where the ε_i are i.i.d. random variables with a common continuous distribution function having a finite mean (not necessarily 0) and the x_i are independent $\nu \times 1$ random vectors independent of $\{\varepsilon_i\}$. Taking the location parameter α in (1.1) to be a minimizer of the function $R(a) = E\rho(y_1 - \beta^T x_1 - a)$, Huber's M-estimators [Huber (1973)] $\widehat{\alpha}$ and $\widehat{\beta}$ of α and β based on $(x_1, y_1), \dots, (x_n, y_n)$ are defined as a solution vector to the minimization problem

$$\sum_{i=1}^{n} \rho(y_i - a - b^T x_i) \left(= \int \rho(y - a)\, dF^*_{n,b}(y) \right) = \min!, \tag{1.2}$$

where $F^*_{n,b}$ is the empirical distribution constructed from $y_i(b) = y_i - b^T x_i$, $i = 1, \dots, n$. In particular, when $\rho(u) = u^2$, $\widehat{\alpha}$ and $\widehat{\beta}$ reduce to the classical least squares estimates, and they reduce to the maximum likelihood estimates of α and β when $\rho(u) = -\log h(u)$, where h is the density function of ε_i. When ρ is differentiable, the M-estimators $\widehat{\alpha}$ and $\widehat{\beta}$ are also defined as a solution of the estimating equations

$$\sum_{i=1}^{n} \rho'(y_i - a - b^T x_i) = 0, \qquad \sum_{i=1}^{n} x_i \rho'(y_i - a - b^T x_i) = 0. \tag{1.3}$$

Received February 1991; revised November 1993.

[1]Research supported by the National Science Foundation, the National Security Agency and the Air Force Office of Scientific Research.

AMS 1991 *subject classifications.* Primary 62J99; secondary 60F05, 62E20, 62G05, 62G35.

Key words and phrases. M-estimator, censoring, truncation, self-consistency, linear regression, martingale, asymptotic normality, influence function, robustness.

Suppose that the responses y_i in (1.1) are not completely observable due to left truncation and right censoring by random variables t_i and c_i such that $\infty > t_i \geq -\infty$ and $-\infty < c_i \leq \infty$. Let $\widetilde{y}_i = y_i \wedge c_i$ and $\delta_i = I_{\{y_i \leq c_i\}}$, where we use $\wedge$ and $\vee$ to denote minimum and maximum, respectively. In addition to right censorship of the responses y_i by c_i, we shall also assume left truncation in the sense that $(\widetilde{y}_i, \delta_i, x_i)$ can be observed only when $\widetilde{y}_i \geq t_i$. The data, therefore, consist of n observations $(\widetilde{y}_i^o, t_i^o, \delta_i^o, x_i^o)$ with $\widetilde{y}_i^o \geq t_i^o$, $i = 1, \dots, n$. Unless stated otherwise, it will be assumed that (t_i, c_i, x_i^T) are independent random vectors that are independent of the sequence $\{\varepsilon_n\}$. The special case $t_i \equiv -\infty$ corresponds to the "censored regression model," for which Miller (1976) suggested an extension of the least squares estimate $[\rho(u) = u^2/2]$ by substituting $F^*_{n,b}$ in (1.2) by the corresponding Kaplan–Meier estimator $\widehat{F}_{n,b}$ constructed from $(\widetilde{y}_i - b^T x_i, \delta_i)$, $i = 1, \dots, n$. An alternative extension of the least squares estimate $[\rho'(u) = u]$ to the censored regression model, proposed by Buckley and James (1979), is to modify the estimating equations (1.3) for censored data. A similar extension of the least squares estimate to truncated (but uncensored) data by modifying the estimating equations (1.3) was recently proposed by Tsui, Jewell and Wu (1988). An alternative to (1.2) or (1.3) in the case of complete data is the method of rank estimators. The rank approach was first extended to truncated (but uncensored) data by Bhattacharya, Chernoff and Yang (1983), and Lai and Ying (1991b) recently gave a complete generalization of the method to left truncated and right censored data.

A general method for constructing M-estimators in left truncated and right censored regression models is given in Section 2. Influence functions of these M-estimators based on censored and truncated data are derived in Section 3, which also studies their robustness properties. A complete asymptotic theory of M-estimators in left truncated and right censored regression models is developed in Section 4. The asymptotic theory is applied in Section 5 to construct confidence regions and to study the asymptotic efficiency of these M-estimators in the left truncated and right censored regression model. In this connection it is also shown how fully efficient M-estimators can be constructed by a data-dependent choice of score functions.

2. A missing information principle and M-estimators in left truncated and right censored regression models. To begin with, consider the case of known $\beta = 0$ so that (1.1) reduces to a location parameter model. Suppose that the data are only subject to right censorship by i.i.d. random variables c_i, so that the i.i.d. random vectors $(\widetilde{y}_1, \delta_1), \dots, (\widetilde{y}_n, \delta_n)$ are observed for the estimation of the location parameter α. Since the y_i are not completely observable, it is natural to replace the criterion (1.2) defining an M-estimator of α by

$$(2.1) \qquad \sum_{i=1}^{n} E\big[\rho(y_i - a) \,|\, \widetilde{y}_i,\, \delta_i\big] = \min!.$$

Moreover, since the conditional expectation in (2.1) involves the unknown common distribution function F of y_i, it is natural to replace F by its nonparametric maximum likelihood estimator $\widehat{F}_n$, which is the Kaplan–Meier (product-limit) estimator based on the censored data. This leads to the minimization problem

$$\int_{-\infty}^{\infty} \rho(z-a)\, d\widehat{F}_n(z) \left(= \sum_{i=1}^{n} \widehat{w}_i \rho(\widetilde{y}_i - a) \right) = \min!, \tag{2.2}$$

where the weights $\widehat{w}_i$ are the jumps of the Kaplan–Meier curve. It is well known that

$$\sup_{t<\tau} |\widehat{F}_n(t) - F(t)| \to_P 0, \quad \text{where } \tau = \inf\left\{t\colon \left(1-F(t)\right)P[c_1 > t] = 0\right\} \tag{2.3}$$

[cf. Wang (1987)]. Hence, assuming suitable regularity conditions on ρ and that $F(\tau) = 1$, we have

$$\begin{aligned} \inf_a \int_{-\infty}^{\infty} \rho(z-a)\, d\widehat{F}_n(z) &\to_P \inf_a \int_{-\infty}^{\tau} \rho(z-a)\, dF(z) \\ &= \inf_a \int_{-\infty}^{\infty} \rho(z-a)\, dF(z), \end{aligned} \tag{2.4}$$

and therefore the M-estimator defined by the minimization problem (2.2) is a consistent estimator of the minimizer α of $\int_{-\infty}^{\infty} \rho(z-a)\, dF(z)$.

Without assuming β to be known, an additional complication arises since F is the common distribution function of $y_i - \beta^T x_i$, which involves the unknown parameter β. First note that the $\widehat{F}_n$ in (2.2) is the same as $\widehat{F}_{n,\beta}$, where $\widehat{F}_{n,b}$ denotes the Kaplan–Meier curve based on $(\widetilde{y}_i - b^T x_i, \delta_i)_{i \le n}$. Since β in $\widehat{F}_{n,\beta}$ is unknown, an obvious approach is to replace $\widehat{F}_n$ $(= \widehat{F}_{n,\beta})$ in (2.2) by $\widehat{F}_{n,b}$, as suggested by Miller (1976). A difficulty with this approach is that unlike (2.3) we now have under certain regularity conditions that, for every $0 < \varepsilon < 1$ and $K > 0$,

$$\sup\left\{ |\widehat{F}_{n,b}(t) - F_{n,b}(t)|\colon -\infty < t < \infty, \|b\| \le K, \sum_{i=1}^{n} I_{\{\widetilde{y}_i - b^T x_i \ge t\}} \ge n^{1-\varepsilon} \right\} \to 0 \quad \text{a.s.}, \tag{2.5a}$$

where

$$F_{n,b}(t) = 1 - \exp\left(-\sum_{i=1}^{n} \int_{u \le t} \frac{dP\{\widetilde{y}_i - b^T x_i \le u,\ y_i \le c_i\}}{\sum_{j=1}^{n} P\{\widetilde{y}_j - b^T x_j \ge u\}} \right), \tag{2.5b}$$

which may be very different from $F(t)$ for $b \ne \beta$ [cf. Lai and Ying (1991c), Lemma 2]. In particular, this implies that unlike (2.4) the minimizer $(\widehat{a}, \widehat{b})$ of $\int_{-\infty}^{\infty} \rho(z - a)\, d\widehat{F}_{n,b}(z)$ need no longer be consistent and may differ substantially from (α, β).

2.1. *M-estimators in the censored regression model.* Instead of directly modifying the minimization problem (1.2) for censored data to define M-estimators, we can modify the estimating equations (1.3) for censored data to the form

$$
\begin{aligned}
&\sum_{i=1}^{n} E\left[\psi(y_i - a - b^T x_i) \,\middle|\, x_i, \tilde{y}_i, \delta_i\right] = 0, \\
&\sum_{i=1}^{n} x_i E\left[\psi(y_i - a - b^T x_i) \,\middle|\, x_i, \tilde{y}_i, \delta_i\right] = 0,
\end{aligned}
\tag{2.6}
$$

where

$$
\begin{aligned}
&E\left[\psi(y_i - a - b^T x_i) \,\middle|\, x_i, \tilde{y}_i, \delta_i\right] \\
&\quad = \delta_i \psi(\tilde{y}_i - a - b^T x_i) \\
&\qquad + (1 - \delta_i) \frac{\int_{\tilde{y}_i - \beta^T x_i}^{\infty} \psi\big(u - a - (b - \beta)^T x_i\big)\, dF(u)}{1 - F(\tilde{y}_i - \beta^T x_i)},
\end{aligned}
\tag{2.7}
$$

in which $\psi = \rho'$. Since the infinite-dimensional parameter F in (2.7) is unknown, Buckley and James (1979) and Ritov (1990) proposed replacing it by $\widehat{F}_{n,b}$. We shall clarify the role of $\widehat{F}_{n,b}$ in order to extend the approach to the much more difficult setting of left truncated and right censored data.

To do this, we first introduce a general principle to construct estimating equations for the regression parameters when the y_i in (1.1) are not completely observable. Let $\mathbf{z}$ denote the vector of observations and let $\mathbf{y}$ denote the associated unobservable complete data vector. Letting $\gamma = (\alpha, \beta, F)$, suppose that an estimate of γ based on $\mathbf{y}$ is defined by

$$
T(\mathbf{y}; \gamma) = 0. \tag{2.8}
$$

For example, the M-estimators $\widehat{\alpha}$ and $\widehat{\beta}$ of α and β in (1.3) and the empirical distribution function of $y_i - \widehat{\beta}^T x_i$ correspond to (2.8) with

$$
\begin{aligned}
T(\mathbf{y}; \alpha, \beta, F) = \Bigg(&\sum_{i=1}^{n} \psi(y_i - \alpha - \beta^T x_i), \sum_{i=1}^{n} x_i \psi(y_i - \alpha - \beta^T x_i), \\
&F - n^{-1} \sum_{i=1}^{n} \Delta_{y_i - \beta^T x_i} \Bigg),
\end{aligned}
\tag{2.9}
$$

where $\psi = \rho'$ and $\Delta_{y_i - \beta^T x_i}(u) = I_{\{y_i - \beta^T x_i \le u\}}$. Since we do not observe $\mathbf{y}$, we replace the unobservable function $T(\mathbf{y}; \gamma)$ in (2.8) by $E_\gamma[T(\mathbf{y}; \gamma) \mid \mathbf{z}]$, leading to the estimating equation

$$
E_\gamma\left[T(\mathbf{y}; \gamma) \mid \mathbf{z}\right] = 0. \tag{2.10}
$$

In applying this missing information principle to the case where the y_i are subject to right censorship, first note that

$$
E_{\widehat{F}_{n,\beta}}\left[n^{-1} \sum_{i=1}^{n} \Delta_{y_i - \beta^T x_i} \,\middle|\, \tilde{y}_1, \delta_1, x_1, \ldots, \tilde{y}_n, \delta_n, x_n \right] = \widehat{F}_{n,\beta}, \tag{2.11}
$$

which is the "self-consistency" property of the Kaplan–Meier estimator $\widehat{F}_{n,\beta}$ [cf. Efron (1967) and Tsai and Crowley (1985)]. This implies that a solution of (2.9)–(2.10) in the censored regression model is of the form $(\widehat{\alpha}, \widehat{\beta}, \widehat{F}_{n,\beta})$, where $\widehat{\alpha}$ and $\widehat{\beta}$ are defined by the estimating equations (2.6) with F replaced by $\widehat{F}_{n,b}$ in (2.7). Equation (2.10) can be regarded as a generalization of the missing information principle of Orchard and Woodbury (1972), who considered the case in which $T(\mathbf{y};\gamma)$ is the derivative of the log-likelihood of $\gamma = (\alpha,\beta)$, assuming F to be known and to have a smooth density. In (2.10), however, F is not assumed to be known so that $\gamma = (\alpha,\beta,F)$, and $T(\mathbf{y};\gamma)$ is not connected to likelihood maximization.

2.2. *M-estimators in the left truncated and right censored regression model.* We now apply the above missing information principle to derive M-estimators of regression parameters when the data are subject to both right censorship and left truncation. As in Turnbull (1976), assume in this derivation that the t_i^o, c_i^o and x_i^o are nonrandom, with $c_i^o \ge t_i^o$, and regard the observed sample $\mathbf{z} = (\widetilde{y}_i^o, t_i^o, \delta_i^o, x_i^o)_{1\le i\le n}$ as having been generated from a larger sample $\mathbf{y} = (y_i, t_i, c_i, x_i)_{1\le i\le m(n)}$, where

$$\begin{aligned} &\sigma_0 = 0, \qquad \sigma_j = \inf\{i > \sigma_{j-1}: \widetilde{y}_i \ge t_j^o\}, \qquad m(n) = \sigma_n, \\ &(t_i, c_i, x_i) = (t_j^o, c_j^o, x_j^o) \quad \text{for } \sigma_{j-1} < i \le \sigma_j. \end{aligned} \tag{2.12}$$

Let $\gamma = (\alpha,\beta,F)$ as before. If $\mathbf{y}$ were observed, then we would be able to estimate γ by (2.8) with

$$T(\mathbf{y};\gamma) = \left(\sum_{i=1}^{m(n)} \psi(y_i(\beta)-\alpha), \sum_{i=1}^{m(n)} x_i\psi(y_i(\beta)-\alpha), m(n)F - \sum_{i=1}^{m(n)} \Delta_{y_i(\beta)}\right), \tag{2.13}$$

where (and in the sequel) $y_i(b) = y_i - b^T x_i$, $c_i(b) = c_i - b^T x_i$, $\widetilde{y}_i^o(b) = \widetilde{y}_i^o - b^T x_i^o$ and $t_i^o(b) = t_i^o - b^T x_i^o$. Note that

$$\sum_{i=1}^{m(n)} \psi(y_i(\beta)-\alpha) = \int_{-\infty}^{\infty} \psi(t-\alpha)\, d\left[\sum_{i=1}^{m(n)} \Delta_{y_i(\beta)}(t)\right], \tag{2.14}$$

$$\begin{aligned} E_F\left[\sum_{i=1}^{m(n)} \Delta_{y_i(\beta)}(u) \,\middle|\, \mathbf{z}\right] &= E_F\left[\sum_{j=1}^{n} \sum_{\sigma_{j-1}<i\le\sigma_j} I_{\{y_i(\beta)\le u\}} \,\middle|\, \mathbf{z}\right] \\ &= \sum_{j=1}^{n} \delta_j^o I_{\{\widetilde{y}_j^o(\beta)\le u\}} + \sum_{j=1}^{n} (1-\delta_j^o) I_{\{\widetilde{y}_j^o(\beta)\le u\}} \\ &\quad \times \frac{F(u) - F(\widetilde{y}_j^o(\beta))}{1 - F(\widetilde{y}_j^o(\beta))} + \sum_{j=1}^{n} \frac{F(u \wedge t_j^o(\beta))}{1 - F(t_j^o(\beta))}, \end{aligned} \tag{2.15}$$

$$E_F[m(n)\,|\,\mathbf{z}] = E_F\left[\sum_{j=1}^{n} (\sigma_j - \sigma_{j-1}) \,\middle|\, \mathbf{z}\right] = \sum_{j=1}^{n} \frac{1}{1 - F(t_j^o(\beta))}. \tag{2.16}$$

Turnbull (1976) and Lai and Ying (1991a, b) studied the following extension of the Kaplan–Meier estimator of F to the left truncated and right censored regression model. Let $\tilde{y}^o_{(1)}(b) \le \cdots \le \tilde{y}^o_{(k)}(b)$ denote all the ordered uncensored residuals. For $i = 1, \dots, k$, let

$$Z_n(b,u) = \#\{j \le n\colon t^o_j(b) \le u \le \tilde{y}^o_j(b)\}, \tag{2.17}$$

where we use the notation $\#A$ to denote the number of elements of a set A. Define

$$\widehat{F}_{n,b}(u) = 1 - \prod_{i:\tilde{y}^o_i(b) \le u,\ \delta^o_i = 1} \left(1 - \frac{1}{Z_n\big(b, \tilde{y}^o_i(b)\big)}\right), \tag{2.18}$$

$$\begin{aligned} H_n(b,u) &= \sum_{j=1}^n \delta^o_j I_{\{\tilde{y}^o_j(b) \le u\}}, \qquad K_n(b,u) = \sum_{j=1}^n (1 - \delta^o_j) I_{\{\tilde{y}^o_j(b) \le u\}}, \\ L_n(b,u) &= \sum_{j=1}^n I_{\{t^o_j(b) \le u\}}. \end{aligned} \tag{2.19}$$

In the case where β is known, $\widehat{F}_{n,\beta}$ is the nonparametric maximum likelihood estimator of the common continuous distribution function F of the $y_i - \beta^T x_i$ and

$$E_{\widehat{F}_{n,\beta}}\big[m(n) \,|\, \mathbf{z}\big] = \int_{-\infty}^{\infty} \big(1 - \widehat{F}_{n,\beta}(t)\big)^{-1} dL_n(\beta, t) = \widehat{m}_\beta, \quad \text{say}, \tag{2.20}$$

by (2.16). Furthermore, $\widehat{F}_{n,\beta}$ satisfies the self-consistency property

$$E_{\widehat{F}_{n,\beta}}\left[\sum_{i=1}^{m(n)} \Delta_{y_i(\beta)}(u) \,\middle|\, \mathbf{z}\right] = \widehat{m}_\beta \widehat{F}_{n,\beta}(u) \tag{2.21}$$

[cf. Turnbull (1976)], noting in view of (2.15) that the self-consistency property (2.21) is equivalent to

$$\begin{aligned} \widehat{m}_\beta \widehat{F}_{n,\beta}(u) = H_n(\beta,u) &+ \int_{-\infty}^{u} \frac{\widehat{F}_{n,\beta}(u) - \widehat{F}_{n,\beta}(t)}{1 - \widehat{F}_{n,\beta}(t)} \, dK_n(\beta,t) \\ &+ \int_{-\infty}^{\infty} \frac{\widehat{F}_{n,\beta}(u \wedge t)}{1 - \widehat{F}_{n,\beta}(t)} \, dL_n(\beta,t), \end{aligned}$$

which can also be expressed in the form

$$\begin{aligned} \widehat{m}_\beta \widehat{F}_{n,\beta}(u) &= H_n(\beta,u) \\ &+ \int_{-\infty}^{u} \left[\int_{-\infty}^{s} \frac{dK_n(\beta,t)}{1 - \widehat{F}_{n,\beta}(t)} + \int_{s}^{\infty} \frac{dL_n(\beta,t)}{1 - \widehat{F}_{n,\beta}(t)}\right] d\widehat{F}_{n,\beta}(s). \end{aligned} \tag{2.22}$$

From (2.13), (2.20) and (2.21), it follows that a solution of (2.10) is of the form $(\widehat{\alpha}, \widehat{\beta}, \widehat{F}_{n,\widehat{\beta}})$. By (2.22),

$$\begin{aligned}\widehat{m}_\beta \int_{-\infty}^{\infty} \psi(u-a)\, d\widehat{F}_{n,\beta}(u) &= \int_{-\infty}^{\infty} \psi(u-a)\, dH_n(\beta,u) \\ &\quad + \int_{-\infty}^{\infty} \psi(u-a)\left[\int_{-\infty}^{u} \frac{dK_n(\beta,t)}{1-\widehat{F}_{n,\beta}(t)} + \int_{u}^{\infty} \frac{dL_n(\beta,t)}{1-\widehat{F}_{n,\beta}(t)}\right] d\widehat{F}_{n,\beta}(u).\end{aligned} \tag{2.23}$$

In view of (2.10), (2.14), (2.21) and (2.23), $\widehat{\alpha}$ and $\widehat{\beta}$ are given by the estimating equations

$$\sum_{i=1}^{n} \psi_i^*(a,b) = 0, \qquad \sum_{i=1}^{n} x_i^o \psi_i^*(a,b) = 0, \tag{2.24}$$

or more precisely, $(\widehat{\alpha}, \widehat{\beta})$ is defined as a minimizer of $|\Sigma_{i=1}^n \psi_i^*(a,b)| + \|\Sigma_{i=1}^n x_i^o \times \psi_i^*(a,b)\|$, where

$$\begin{aligned}\psi_i^*(a,b) &= \delta_i^o \psi\big(\widetilde{y}_i^o(b)-a\big) + (1-\delta_i^o)\frac{\int_{\widetilde{y}_i^o(b)}^{\infty} \psi(u-a)\, d\widehat{F}_{n,b}(u)}{1-\widehat{F}_{n,b}(\widetilde{y}_i^o(b))} \\ &\quad + \frac{\int_{-\infty}^{t_i^o(b)} \psi(u-a)\, d\widehat{F}_{n,b}(u)}{1-\widehat{F}_{n,b}(t_i^o(b))}.\end{aligned} \tag{2.25}$$

Note that the first equation in (2.24) is equivalent to $\int_{-\infty}^{\infty} \psi(u-a)\, d\widehat{F}_{n,b}(u) = 0$, which implies that $\int_{-\infty}^{t_i^o(b)} \psi(u-a)\, d\widehat{F}_{n,b}(u) = -\int_{t_i^o(b)}^{\infty} \psi(u-a)\, d\widehat{F}_{n,b}(u)$. Hence we can rewrite (2.25) in the form

$$\begin{aligned}\psi_i^*(a;b) &= \delta_i^o \psi\big(\widetilde{y}_i^o(b)-a\big) + \big(1-\delta_i^o\big)\frac{\int_{\widetilde{y}_i^o(b)}^{\infty} \psi(u-a)\, d\widehat{F}_{n,b}(u)}{1-\widehat{F}_{n,b}(\widetilde{y}_i^o(b))} \\ &\quad - \frac{\int_{t_i^o(b)}^{\infty} \psi(u-a)\, d\widehat{F}_{n,b}(u)}{1-\widehat{F}_{n,b}(t_i^o(b))}.\end{aligned} \tag{2.26}$$

3. Influence functions and robustness. The influence function, or influence curve (IC), provides an important heuristic tool for studying robustness and is the cornerstone of the infinitesimal approach to robust statistics [cf. Hampel, Ronchetti, Rousseeuw and Stahel (1986)]. Expressing an estimator based on i.i.d. observations as a functional T of their empirical distribution, the infinitesimal approach studies robustness of the estimator via the Gâteaux or other derivatives of T. In particular, the influence curve of T is

$$\text{IC}(x; T, G) = \lim_{t \downarrow 0} t^{-1}\left\{T\big((1-t)G + t\delta_x\big) - T(G)\right\} \tag{3.1}$$

at a distribution G, where δ_x denotes the distribution that puts mass 1 to x. It measures the normalized influence of adding one more observation x to a very large sample on the value of the estimate.

In this section we first evaluate the influence curve of the M-estimator (2.24), assuming that

$$\big(t_i, c_i, x_i^T, \varepsilon_i\big) \text{ are i.i.d., with } \varepsilon_i \text{ independent of } \big(t_i, c_i, x_i^T\big). \tag{3.2}$$

We shall regard the observed data $(\widetilde{y}_i^o, t_i^o, \delta_i^o, x_i^o)$, $i = 1, \ldots, n$, as having been generated from a larger sample of i.i.d. random vectors $(t_i, c_i, x_i^T, \varepsilon_i)$, $i = 1, \ldots, m$, with $n = n(m)$ and with t_i, x_i, $\widetilde{y}_i = (\alpha + \beta^T x_i + \varepsilon_i) \wedge c_i$ and $\delta_i = I_{\{\alpha + \beta^T x_i + \varepsilon_i \le c_i\}}$ observed only when $\widetilde{y}_i \ge t_i$. The M-estimator defined by (2.24) and (2.26) can be expressed as a function of this larger sample $\{(t_i, c_i, x_i^T, \varepsilon_i)\colon 1 \le i \le m\}$, where the function involves only $(t_i, x_i^T, \widetilde{y}_i, \delta_i) I_{\{\widetilde{y}_i \ge t_i\}}$, and therefore we can represent the M-estimator as a functional of the empirical distribution G_m of $\{(t_i, c_i, x_i^T, \varepsilon_i)\colon 1 \le i \le m\}$.

The following "implicit function formula" for the influence curve of a statistical functional T that is defined implicitly by a functional equation will be applied to the M-estimator defined implicitly by (2.24). Letting $\mathcal{M}$ denote the space of probability measures on some measurable space, suppose that $\Phi\colon \mathcal{M} \times \mathbf{R}^p \to \mathbf{R}^p$ is a von Mises functional and that $T\colon \mathcal{M} \to \mathbf{R}^p$ is well defined by

$$T(G) = \theta \quad \Leftrightarrow \quad \Phi(G, \theta) = 0. \tag{3.3}$$

For any fixed $\theta \in \mathbf{R}^p$, define $\Phi_\theta\colon \mathcal{M} \to \mathbf{R}^p$ by $\Phi_\theta(G) = \Phi(G, \theta)$. Then

$$\left[\frac{\partial}{\partial \theta} \Phi(G, \theta)\right]_{\theta = T(G)} \text{IC}(x; T, G) = -\,\text{IC}\big(x;\, \Phi_{T(G)}, G\big), \tag{3.4}$$

as can be shown by a standard argument similar to that in Hampel, Ronchetti, Rousseeuw and Stahel [(1986), page 101]. The notation $(\partial/\partial\theta)\Phi(G, \theta)$ in (3.4) denotes the Jacobian matrix, recalling that both θ and Φ are p-dimensional.

To define the functional Φ so that (2.24) can be expressed as $\Phi(G_m, \theta) = 0$ with $\theta = (a, b^T)$, we first represent the product-limit estimator $\widehat{F}_{n,b}$ in terms of the empirical subsurvival functions

$$\begin{aligned}
\widehat{S}_{b,1}(u) &= m^{-1} \sum_{i=1}^{m} I_{\{\alpha + \varepsilon_i \ge (b - \beta)^T x_i + u,\ c_i(\beta) \ge \alpha + \varepsilon_i \ge t_i(\beta)\}}, \\
\widehat{S}_{b,2}(u) &= m^{-1} \sum_{i=1}^{m} I_{\{c_i(\beta) \ge (b - \beta)^T x_i + u,\ \alpha + \varepsilon_i > c_i(\beta) \ge t_i(\beta)\}}, \\
\widehat{S}_{b,3}(u) &= m^{-1} \sum_{i=1}^{m} I_{\{(\alpha + \varepsilon_i) \wedge c_i(\beta) \ge t_i(\beta) > (b - \beta)^T x_i + u\}}.
\end{aligned} \tag{3.5}$$

By (2.17), $Z_n(b,u) = m[\widehat{S}_{b,1}(u) + \widehat{S}_{b,2}(u) - \widehat{S}_{b,3}(u)]$, recalling that $\widetilde{y}_i(b) = (y_i \wedge c_i) - b^T x_i = [(\alpha + \varepsilon_i) \wedge c_i(\beta)] - (b - \beta)^T x_i$. Therefore (2.18) defines $\widehat{F}_{n,b}(u)$ as a functional J_u of $\widehat{S}_{b,1}$, $\widehat{S}_{b,2}$ and $\widehat{S}_{b,3}$, where

$$J_u(S_1, S_2, S_3) = 1 - \prod_{t \le u} \left\{ 1 - \frac{S_1(t) - S_1(t+)}{S_1(t) + S_2(t) - S_3(t)} \right\} \qquad \left(\frac{0}{0} = 0 \right). \tag{3.6}$$

For $b = \beta$, since the $\widehat{S}_{b,j}$ in (3.5) are functionals of the empirical distribution G_m of $\{(t_i, c_i, x_i^T, \varepsilon_i): 1 \le i \le m\}$, the functional (3.6) can be used to express $\widehat{F}_{n,\beta}(u)$ as a functional of G_m, say, $\widehat{F}_{n,\beta}(u) = J_u^*(G_m)$. The influence curve of J_u^* at the underlying distribution G of $(t_1, c_1, x_1^T, \varepsilon_1)$ can be evaluated by modifying Reid's arguments [Reid (1981)] in the censored case. Letting

$$\Gamma_0(s) = P\{t_1(\beta) \le s \le c_1(\beta)\}, \qquad \tau_0 = \inf\{s: \Gamma_0(s) > 0\}, \tag{3.7}$$

and noting that $\widehat{F}_{n,\beta}$ only estimates consistently the conditional distribution of $\alpha + \varepsilon_1$ given $\alpha + \varepsilon_1 > \tau_0$ [cf. Lai and Ying (1991a)], the influence curve of the functional J_u^* is given by

$$\begin{aligned} \mathrm{IC}\Big((t, c, x^T, \varepsilon); J_u^*, G \Big) &= \frac{1 - F(u)}{1 - F(\tau_0)} I_{\{(\alpha + \beta^T x + \varepsilon) \wedge c \ge t\}} \\ &\quad \times \left\{ \frac{I_{\{\alpha + \varepsilon \le u \wedge (c - \beta^T x)\}}}{\big(1 - F(\alpha + \varepsilon)\big)\Gamma_0(\alpha + \varepsilon)} \right. \\ &\quad \left. - \int_{t - \beta^T x}^{(\alpha + \varepsilon) \wedge (c - \beta^T x) \wedge u} \frac{dF(s)}{\big(1 - F(s)\big)^2 \Gamma_0(s)} \right\}, \end{aligned} \tag{3.8}$$

if $\tau_0 < t - \beta^T x < u$ and $F(u) < 1$. For the special case $\alpha = \beta = 0$ and $t_i = t = -\infty$, $\tau_0 = -\infty$ and (3.8) agrees with IC_1 or IC_2 in Reid [(1981), formula (2.2)] according as $\varepsilon \le c$ or $\varepsilon > c$.

As pointed out at the end of Section 2, the first equation in (2.24) is equivalent to $\int_{-\infty}^{\infty} \psi(u - a)\, d\widehat{F}_{n,b}(u) = 0$, which upon multiplying both sides by $1 - F(\tau_0)$ yields

$$\big(1 - F(\tau_0)\big) \int_{-\infty}^{\infty} \psi(u - a)\, d\widehat{F}_{n,b}(u) = 0. \tag{3.9a}$$

We now express the second equation in (2.24) in terms of similar stochastic integrals with respect to empirical measures on $\mathbf{R}^{\nu+1}$. For any $u \in \mathbf{R}$ and Borel subset $B \subset \mathbf{R}^\nu$, define

$$\widetilde{S}_{b,1}(u,B) = m^{-1} \sum_{i=1}^{m} I_{\{\alpha+\varepsilon_i \geq (b-\beta)^T x_i + u,\ c_i(\beta) \geq \alpha+\varepsilon_i \geq t_i(\beta)\}} I_{\{x_i \in B\}}$$

and define $\widetilde{S}_{b,2}(u,B)$ and $\widetilde{S}_{b,3}(u,B)$ similarly by multiplying the corresponding summands in (3.5) by $I_{\{x_i \in B\}}$. In view of (2.26), the second equation in (2.24) can be written in the form

$$\text{(3.9b)} \quad \int_{\mathbf{R}^{\nu+1}} x\psi(u-a)\, d\widetilde{S}_{b,1} + \int_{\mathbf{R}^{\nu+1}} \frac{\int_u^\infty \psi(t-a)\, d\widehat{F}_{n,b}(t)}{1-\widehat{F}_{n,b}(u)} x\, d(\widetilde{S}_{b,2} - \widetilde{S}_{b,3}) = 0,$$

in which the integrands are functions of u, x and the $\widetilde{S}_{b,j}$ are regarded as distributions of (u, x^T). Combine the left-hand sides of (3.9a) and (3.9b) into a $(1+\nu)\times 1$ vector, which can be expressed as a functional $\Phi(G_m, \theta)$ of $\theta = (a, b^T)$ and the empirical distribution G_m of $\{(t_i, c_i, x_i^T, \varepsilon_i)$: $1 \leq i \leq m\}$. Thus,

$$\text{(3.10)} \quad \Phi(G_m, \theta) = \left((1-F(\tau_0)) \int_{-\infty}^{\infty} \psi(u-a)\, d\widehat{F}_{n,b}(u), \right. \\ \left. \text{left-hand side of (3.9b)} \right)^T.$$

Let $\theta_0 = (\alpha, \beta^T)$, where α is defined as the solution of the equation $\int_{\tau_0}^{\tau} \psi(u-a)\, dF(u) = 0$ and $\tau = \inf\{s > \tau_0\colon (1-F(s))\Gamma_0(s) = 0\}$. We integrate over $[\tau_0, \tau]$ because all jumps of $\widehat{F}_{n,\beta}$ occur inside this interval. Under (3.2), the expectation of $\widetilde{S}_{b,j}$ does not depend on m and is denoted by $S_{b,j}$, for example,

$$\text{(3.11)} \quad S_{b,1}(u,B) = E\left\{ \left[F\big(c_1(\beta)\big) - F\Big(t_1(\beta) \vee \big(u + (b-\beta)^T x_1\big)\Big) \right] I_{\{x_1 \in B\}} \right\}.$$

Replacing $\widetilde{S}_{b,j}$ by $S_{\beta,j}$ and $\widehat{F}_{n,b}$ by F in the left-hand side of (3.9b) gives the value 0, so $\Phi(G, \theta_0) = 0$, where G is the underlying distribution of $(t_1, c_1, x_1^T, \varepsilon_1)$. Therefore, by (3.4), we can evaluate thé influence curve of our M-estimator via that of the functional $\Phi_{\theta_0} = \Phi(\cdot, (\alpha, \beta^T))$.

To find the influence curves of Φ_{θ_0} at G, we shall evaluate the two influence curves of the functionals associated with (3.9a) and (3.9b) separately and make use of Reid's chain rule [Reid (1981)] for influence curves. The integral in (3.9a) with $a = \alpha$ and $b = \beta$ is a functional T_1 of $\widehat{F}_{n,\beta}$, defined by $T_1(\widehat{F}_{n,\beta}) = (1-F(\tau_0)) \times \int_{\tau_0}^{\tau} \psi(u-\alpha)\, d\widehat{F}_{n,\beta}(u)$. Since $\int_{\tau_0}^{\tau} \psi(u-\alpha)\, dF(u) = 0$, $\mathrm{IC}(u; T_1, F) = (1-F(\tau_0))\psi(u-\alpha) I_{\{\tau_0 \leq u \leq \tau\}}$. Since $\widehat{F}_{n,\beta}(u) = J_u^*(G_m)$, we can make use of (3.8) on the influence curve of J_u^* together with Reid [(1981), formula (3.2) (in which a minus sign is missing)] to conclude that the influence curve of the functional $T_1 \circ J^*$ associated with the left-hand side of (3.9a), in which $a = \alpha$ and $b = \beta$, is

$$\begin{aligned}
&\mathrm{IC}\big((t,c,x^T,\varepsilon); T_1 \circ J^*, G\big)\\
&= -\int_{\tau_0}^{\tau} \mathrm{IC}\big((t,c,x^T,\varepsilon); J_u^*, G\big)\left\{\frac{d}{du}\psi(u-\alpha)\right\} du\\
&= -\int_{\tau_0}^{\tau}\left\{\int_{t-\beta^T x}^{u} \frac{I_{\{s \le (\alpha+\varepsilon)\wedge(c-\beta^T x)\}}}{(1-F(s))^2\Gamma_0(s)}\, dF(s)\right\}\\
&\qquad \times d\left\{\int_u^{\tau} (1-F(v))\, d\psi(v-\alpha)\right\}\\
&\quad - \frac{I_{\{\alpha+\varepsilon \le c-\beta^T x\}}}{(1-F(\alpha+\varepsilon))\Gamma_0(\alpha+\varepsilon)} \int_{\alpha+\varepsilon}^{\tau} (1-F(u))\, d\psi(u-\alpha),
\end{aligned} \tag{3.12}$$

if $\tau > (\alpha+\varepsilon)\wedge(c-\beta^T x) \ge t-\beta^T x \ge \tau_0$. Here and in the sequel it will be assumed that ψ is absolutely continuous with respect to Lebesgue measure. Note that the last term in (3.12) represents an uncensored observation and can be written as $-\delta r_0(\alpha+\varepsilon)$, where

$$r_0(u) = \frac{\left\{\int_u^{\tau}(1-F(v))\, d\psi(v-\alpha)\right\}}{\left\{(1-F(u))\Gamma_0(u)\right\}}, \qquad \delta = I_{\{\alpha+\varepsilon \le c-\beta^T x\}}. \tag{3.13}$$

Integration by parts then simplifies (3.12) to

$$\begin{aligned}
&\mathrm{IC}\big((t,c,x^T,\varepsilon); T_1 \circ J^*, G\big)\\
&= \int_{t-\beta^T x}^{(\alpha+\varepsilon)\wedge(c-\beta^T x)} \frac{r_0(u)}{1-F(u)}\, dF(u) - \delta r_0(\alpha+\varepsilon).
\end{aligned} \tag{3.14}$$

Similar calculations can be used to evaluate the influence curve of the functional associated with the left-hand side of (3.9b) with $a = \alpha$ and $b = \beta$, which we can rewrite as

$$\begin{aligned}
&\int_{\mathbf{R}^{\nu+1}} x\psi(u-\alpha)\, d(\widetilde{S}_{\beta,1} + \widetilde{S}'_{\beta,2} - \widetilde{S}_{\beta,3})\\
&\quad + \int_{\mathbf{R}^{\nu+1}} \frac{\int_u^{\infty}(1-\widehat{F}_{n,\beta}(t))\, d\psi(t-\alpha)}{1-\widehat{F}_{n,\beta}(u)}\, x\, d(\widetilde{S}_{\beta,2} - \widetilde{S}_{\beta,3}),
\end{aligned}$$

after applying integration by parts to $-\int_u^{\infty}\psi(t-\alpha)\, d(1-\widehat{F}_{n,\beta}(t))$. The influence curve of the functional T_2 associated with the second summand above can be simplified by repeated use of integration by parts and leads to an expression that is completely analogous to (3.14). The influence curve of the functional T_3

associated with the first summand above is easy to compute and leads to the last term in the following:

$$\begin{aligned}&\mathrm{IC}\Big(\big(t,c,x^T,\varepsilon\big);T_2+T_3,G\Big)\\ &\quad=\int_{t-\beta^Tx}^{(\alpha+\varepsilon)\wedge(c-\beta^Tx)}\frac{r_1(u)}{1-F(u)}\,dF(u)-\delta r_1(\alpha+\varepsilon)\\ &\qquad -x\Big\{\delta\psi(\varepsilon)+(1-\delta)\Psi\big(c-\beta^Tx\big)-\Psi\big(t-\beta^Tx\big)\Big\},\end{aligned}\tag{3.15}$$

if $\tau > (\alpha+\varepsilon)\wedge(c-\beta^Tx) \geq t-\beta^Tx \geq \tau_0$, where

$$\begin{aligned}&r_1(u)=\frac{\Gamma_1(u)}{\Gamma_0(u)}\frac{\int_u^\tau\big(1-F(v)\big)d\psi(v-\alpha)}{1-F(u)},\qquad \Psi(u)=\frac{\int_u^\tau\psi(v-\alpha)\,dF(v)}{1-F(u)},\\ &\Gamma_1(u)=E\big\{x_iI_{\{t_i(\beta)\leq u\leq c_i(\beta)\}}\big\}.\end{aligned}\tag{3.16}$$

The column vector formed from (3.14) and (3.15) gives the influence curve of the functional $\Phi_{\theta_0} = \Phi(\cdot, (\alpha, \beta^T))$, where Φ is defined in (3.10). To apply (3.4) to evaluate the influence curve of the functional T associated with our M-estimator, it remains to find $(\partial/\partial\theta)\Phi(G,\theta)$. First note that since G is the underlying distribution of $(t_1, c_1, x_1^T, \varepsilon_1)$,

$$\begin{aligned}\Phi\Big(G,\big(a,b^T\big)\Big)=\Bigg(&\big(1-F(\tau_0)\big)\int_{-\infty}^{\infty}\psi(u-a)\,dF_{n,b}(u),\ \int_{\mathbf{R}^{\nu+1}}x\psi(u-a)\,dS_{b,1}\\ &+\int_{\mathbf{R}^{\nu+1}}\bigg\{\int_u^\infty\frac{\psi(t-a)\,dF_{n,b}(t)}{1-F_{n,b}(u)}\bigg\}x\,d(S_{b,2}-S_{b,3})\Bigg)^T,\end{aligned}$$

where the $S_{b,j}$ are defined as in (3.11) and

$$F_{n,b}(t)=1-\exp\bigg\{-\int_{u\leq t}\frac{dP\{t_1(b)\leq y_1(b)\leq c_1(b)\wedge u\}}{P\{\tilde{y}_1(b)\geq u\geq t_1(b)\}}\bigg\},$$

which agrees with (2.5b) in the censored case. Formal differentiation with respect to a and b under the integral signs, assuming the functions involved to be differentiable, gives

$$\bigg[\frac{\partial}{\partial\theta}\Phi(G,\theta)\bigg]_{\theta=(\alpha,\beta^T)}=-\begin{pmatrix}\psi_\alpha & -g_\alpha^T\\ 0 & C_\alpha\end{pmatrix}.\tag{3.17}$$

The $(\nu\times\nu)$ matrix C_α, $(\nu\times 1)$ vector g_α and scalar ψ_α in (3.17) are given explicitly in (4.27) and (4.38) of Section 4, where we give a rigorous derivation of these quantities via an asymptotic linearity argument under precisely stated regularity conditions (given in Theorem 2) which allow the (t_i, c_i, x_i^T) to be nonidentically distributed but such that

$$\lim_{m\to\infty} E\left\{ m^{-1} \sum_{i=1}^{m} x_i^r I_{\{t_i(\beta) \le s \le c_i(\beta)\}} \right\} = \Gamma_r(s) \tag{3.18}$$

$$\text{for } r = 0, 1, 2 \text{ and } s < F^{-1}(1),$$

where $x_i^2 = x_i x_i^T$ and $\Gamma_r(s) = \int x_1^r I_{\{t_1(\beta) \le s \le c_1(\beta)\}}\, dG$.

Summarizing, we obtain from (3.4) and (3.17) the influence function formula

$$\mathrm{IC}\Big((t, c, x^T, \varepsilon);\, T, G\Big) = \begin{pmatrix} \psi_\alpha & -g_\alpha^T \\ 0 & C_\alpha \end{pmatrix}^{-1} \begin{pmatrix} (3.14) \\ (3.15) \end{pmatrix}, \tag{3.19}$$

for the functional T associated with the M-estimator of (α, β^T) defined by (2.24), at the underlying distribution G of $(t_1, c_1, x_1^T, \varepsilon_1)$ under assumption (3.2). Moreover, we can also apply (3.19) to study robustness in the case of independent and nonidentically distributed (t_i, c_i, x_i) such that (3.18) holds with $\Gamma_r(s)$ being the corresponding expectation under some limiting distribution G.

The preceding discussion can be easily extended to the following situation. Suppose that only the slope parameter β in (1.1) is of interest. Clearly we can absorb the α into $\alpha + \varepsilon_i$ and thereby assume that $\alpha = 0$ in (1.1), noting that F in the preceding discussion is in fact the distribution function of $\alpha + \varepsilon_i$. We can fix any a in the second equation of (2.24) and use it as an estimating equation for β. Denoting the functional associated with this M-estimator of β by $T^{(\beta)}$, its influence curve at G is

$$\mathrm{IC}\Big((t, c, x^T, \varepsilon);\, T^{(\beta)}, G\Big) = C_a^{-1} \cdot (3.15), \tag{3.20}$$

where we replace the $\alpha + \varepsilon$ in (3.15) by ε, and the α in (3.16) by a.

We next give a numerical example to illustrate and discuss some insights provided by the influence function formula (3.20) into the robustness properties of M-estimators of the slope β in (1.1). Consider the regression model $y_i = \beta x_i + \varepsilon_i$, $i = 1, \dots, 50$, in which the ε_i are i.i.d. random variables whose common distribution function F is a mixture, $0.6N(0,1) + 0.4N(0, 10^2)$, of two normal distribution functions with respective standard deviations 1 and 10. The 50 design points x_i are evenly spaced in the interval $[-1, 1]$ with $x_1 = -1$ and $x_{50} = 1$. The y_i are subject to right censorship by independent random variables c_i which are normal with means $2.5 + \beta x_i$ and standard deviation 5. The censoring probability $P\{\varepsilon_i > c_i - \beta x_i\}$ is 0.35, and the observations are $(x_i, \tilde{y}_i, \delta_i)$ with $\tilde{y}_i = y_i \wedge c_i$, $\delta_i = I_{\{y_i \le c_i\}}$, $1 \le i \le n = 50$. The second equation of (2.24) with $a = 0$ and $t_i \equiv -\infty$ reduces to

$$\sum_{i=1}^{n} x_i \left\{ \delta_i \psi(\tilde{y}_i - b x_i) + (1 - \delta_i) \int_{\tilde{y}_i - b x_i}^{\infty} \frac{\psi(u)\, d\widehat{F}_{n,b}(u)}{1 - \widehat{F}_{n,b}(\tilde{y}_i - b x_i)} \right\} = 0, \tag{3.21}$$

TABLE 1
Influence curves (IC) *associated with the Buckley–James* (BJ) *and Huber* (H) *score functions for a censored sample of* 50 *observations* $(x_i, \tilde{y}_i, \delta_i)$

x_i	$\tilde{y}_i$	δ_i	$\tilde{y}_i - \beta x_i$	IC(BJ)	IC(H)
−1.0000	−1.79	1	−0.79	6.79	6.79
−0.9592	−1.39	1	−0.43	3.56	3.56
−0.9184	−29.17	1	−28.25†	222.36*	7.87
−0.8876	−0.50	0	0.39	−34.35*	−8.28
−0.8367	−20.91	1	−20.07†	143.95*	7.17
−0.7959	−0.44	1	0.36	−2.45	−2.45
−0.7551	−4.68	0	−3.92	−11.11	0.00
−0.7143	−2.10	0	−1.39	−13.26	0.00
−0.6735	1.12	1	1.79	−10.36	−5.77
−0.6327	−22.17	1	−21.54†	116.81*	5.42
−0.5918	1.98	1	2.57	−13.03	−5.07
−0.5510	−3.68	0	−3.13	−8.46	0.00
−0.5102	−0.13	1	0.38	−1.68	−1.68
−0.4694	0.83	1	1.30	−5.24	−4.02
−0.4286	−0.26	1	0.16	−0.60	−0.60
−0.3878	−2.62	0	−2.24	−6.33	0.00
−0.3469	−0.51	1	−0.17	0.50	0.50
−0.3061	−2.68	1	−2.37	6.23	2.62
−0.2653	−0.45	1	−0.19	0.42	0.42
−0.2245	2.93	1	3.15	−6.06	−1.92
−0.1837	−1.87	0	−1.69	−3.21	0.00
−0.1429	−5.89	1	−5.74	7.03	1.22
−0.1020	0.00	1	0.11	−0.09	−0.09
−0.0612	−1.30	1	−1.24	0.65	0.52
−0.0204	−0.52	1	−0.50	0.09	0.09
0.0204	−1.83	1	−1.85	−0.32	−0.17
0.0612	−11.42	0	−11.48†	0.46	0.00
0.1020	0.15	1	0.05	0.04	0.04
0.1429	−0.04	1	−0.18	−0.22	−0.22
0.1837	−4.33	0	−4.52	2.61	0.00
0.2245	−10.06	0	−10.29†	1.93	0.00
0.2653	−2.42	1	−2.68	−6.10	−2.27
0.3061	−0.11	1	−0.42	−1.10	−1.10
0.3469	−0.05	1	−0.40	−1.19	−1.19
0.3878	−4.74	0	−5.13	5.29	0.00
0.4286	−0.45	1	−0.88	−3.21	−3.21
0.4694	−0.24	1	−0.71	−2.85	−2.85
0.5102	−0.33	1	−0.84	−3.68	−3.68
0.5510	0.52	1	−0.03	−0.16	−0.16
0.5918	3.75	0	3.16	51.08*	5.07
0.6327	−3.43	0	−4.06	9.23	0.00
0.6735	−0.44	1	−1.12	−6.44	−5.77
0.7143	−3.42	0	−4.14	10.38	0.00
0.7551	2.12	1	1.36	8.82	6.47
0.7959	0.12	1	−0.68	−4.62	−4.62
0.8367	−4.33	0	−5.16	11.40	0.00
0.8876	−3.37	1	−4.25	32.36*	−7.61
0.9184	−7.97	0	−8.88†	9.15	0.00
0.9592	−2.54	0	−3.50	14.45	0.00
1.0000	1.72	1	0.72	6.15	6.15

which need not have a solution. An M-estimator is defined as a zero-crossing of the left-hand side of (3.21).

The special case $\psi(u) = u$ in (3.21), which we denote by ψ_{BJ}, gives the Buckley–James estimator. Table 1 tabulates the values of $(x_i, \tilde{y}_i, \delta_i)$ of a random sample of size 50 from the above censored regression model with $\beta = 1$. It also tabulates the influence curve $\mathrm{IC}((x, \tilde{y}, \delta); \mathrm{BJ}, G)$ of the Buckley–James estimator, for the observed values of $(x, \tilde{y}, \delta)$, at the distribution G specified by (i) the normal mixture F of the ε_i, (ii) the $N(2.5, 5^2)$ distribution of $c_i - \beta x_i$ and (iii) a uniform distribution on $[-1, 1]$ for the design vectors x_i such that ε_i, x_i and $c_i - \beta x_i$ are independent.

Most of the absolute values of $\tilde{y}_i - \beta x_i$ in Table 1 are less than 6 and there are six exceptions, which are marked by a dagger. Most of the tabulated values of the influence curve of the Buckley–James estimator are less than 15 in absolute value; the six exceptions are marked by asterisks with three of these greater than 100. To avoid these exceptionally large influences, (3.20) suggests replacing ψ_{BJ} by bounded score functions ψ such that ψ' vanishes outside some bounded set. In particular, we chose Huber's score function $\psi_{\mathrm{H}}(u) = (-1) \vee (u \wedge 1)$, and we computed by numerical integration the influence curve $\mathrm{IC}((x, \tilde{y}, \delta); \mathrm{H}, G)$ of the M-estimator defined by (3.21) with $\psi = \psi_{\mathrm{H}}$. The results, which are given in the last column of Table 1, show a much smaller range (between -8.3 and 8) of influence function values.

Figure 1 gives plots of the data and of (i) the true regression line (the solid line) with slope $\beta = 1$, (ii) the regression line with slope 4.795 (the broken line labelled BJ) fitted by the Buckley–James method and (iii) the broken line with slope 0.79 fitted by using Huber's score function ψ_{H}. The few large negative outliers appear to be quite influential in tilting the Buckley–James line toward them but do not have much influence on the M-estimator using Huber's score function. The data in Table 1 and Figure 1 represent one of 100 samples from the above censored regression model in a simulation study of robustness properties of the M-estimators (3.21) with $\psi = \psi_{\mathrm{BJ}}$ or ψ_{H}. The results of this simulation study are summarized in Table 2, which also considers two other censoring distributions $N(\beta x_i, 5^2)$, $N(20 + \beta x_i, 5^2)$ for the c_i (with respective censoring probabilities 0.5 and 0.015) and the case of noncontaminated $N(0, 1)$ distribution for the ε_i. These results show that the score function ψ_{H} gives a more robust M-estimator of β than ψ_{BJ} when there is contamination. Moreover, even when contamination is absent, the cases with $p = 0$ in Table 2 show that the M-estimator associated with ψ_{H} is only slightly less efficient than the Buckley–James estimator.

Table 3 reports a more extensive simulation study in which the response $y_i = \beta x_i + \varepsilon_i$ (with $\beta = 1$) is subject to both left truncation and right censoring by i.i.d. random vectors (t_i, c_i) that are independent of the ε_i, for $i = 1, \dots, 100$. The x_i are evenly spaced in the interval $[-1, 3]$ and the ε_i are i.i.d. random variables whose common distribution F is contaminated standard normal of the form F_1 or F_2 or F_3, where $F_1 = 0.6N(0, 1) + 0.4N(0, 2^2)$, $F_2 = 0.8N(0, 1) + 0.2N(0, 6^2)$ and $F_3(x) = 0.5P\{N(0, 1) \le x\} + 0.5 \exp\{-\max(1 - x, 0)\}$. Note that each F_i has mean 0 and that and F_1 and F_2 are symmetric but F_3 is not. Each c_i has the

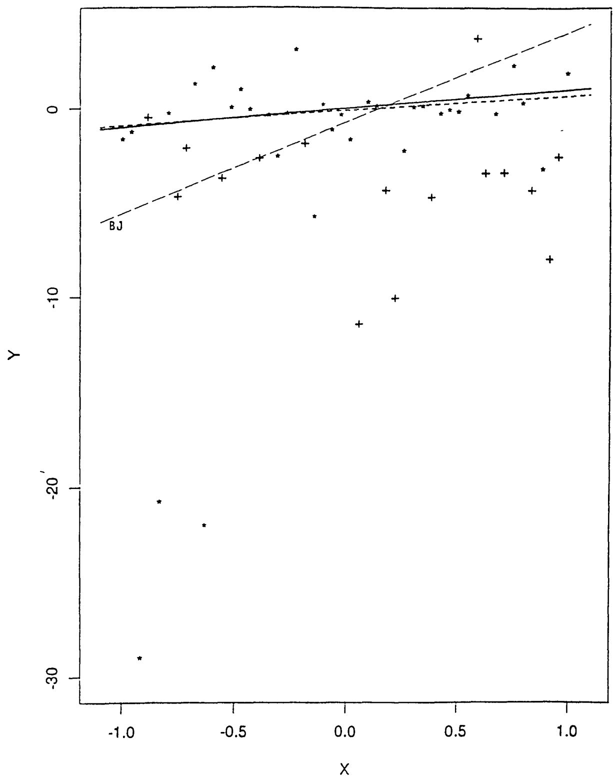

FIG. 1. *Censored data* (+) *and uncensored data* (∗).

TABLE 2

Mean, standard deviation (SD) *and five-number summary of the simulated sampling distribution of M-estimator of* β (= 1) *based on* ψ *and a censored random sample of size* 50 *from* $\varepsilon_i \sim (1-p)N(0,1) + pN(0,10^2)$ *and* $c_i \sim N(\mu + x_i, 5^2)$. *Each simulation result is based on* 100 *replicates*

(p, μ)	ψ	**Mean**	**SD**	**Min.**	**First quartile**	**Median**	**Third quartile**	**Max.**
(0.4,2.5)	BJ	1.03	1.44	−3.09	0.19	0.93	1.94	4.90
	H	0.90	0.52	−0.37	0.54	0.91	1.27	2.11
(0.4, 0)	BJ	0.97	1.80	−3.61	−0.14	0.99	2.29	5.75
	H	1.00	0.65	−0.98	0.56	0.93	1.46	2.45
(0.4, 20)	BJ	1.32	1.52	−2.12	0.49	1.19	2.45	5.19
	H	1.06	0.47	−0.14	0.75	1.09	1.37	2.13
(0, 2.5)	BJ	1.02	0.29	0.23	0.81	1.03	1.21	1.67
	H	1.00	0.33	0.15	0.75	0.99	1.27	1.78
(0, 0)	BJ	1.00	0.32	−0.20	0.83	1.00	1.21	1.88
	H	0.94	0.41	−0.33	0.77	0.97	1.18	2.06
(0, 20)	BJ	0.97	0.22	0.42	0.82	0.99	1.10	1.60
	H	0.97	0.24	0.34	0.83	0.96	1.13	1.61

TABLE 3
Mean and mean squared error (MSE) of M-estimator $\hat\beta_{\rm BJ}$ or $\hat\beta_{\rm H}$ and of naive estimator $\hat\beta_{\rm NA}$ of β (= 1) based on randomly censored and truncated data, with censoring and truncation probabilities p_C and p_T for different error distributions F and parameters θ_C and θ_T of the censoring and truncation distributions

F	θ_C	θ_T	p_C	p_T	$E(\hat\beta_{\rm NA})$	$E(\hat\beta_{\rm BJ})$	$E(\hat\beta_{\rm H})$	MSE $(\hat\beta_{\rm NA})$	MSE $(\hat\beta_{\rm BJ})$	MSE $(\hat\beta_H)$
F_1	2.5	0.4	0.24	0.15	0.606	0.997	0.999	0.166	0.031	0.029
F_1	2.5	1.2	0.24	0.26	0.568	0.991	0.993	0.200	0.045	0.043
F_1	1.5	0.4	0.41	0.17	0.450	1.011	1.010	0.313	0.048	0.046
F_1	1.5	1.2	0.41	0.31	0.430	1.002	1.001	0.336	0.064	0.061
F_1	0.6	0.4	0.58	0.24	0.304	0.993	0.992	0.494	0.075	0.068
F_1	0.6	1.2	0.58	0.42	0.317	0.970	0.967	0.480	0.112	0.109
F_2	2.5	0.4	0.25	0.17	0.592	0.994	0.995	0.179	0.031	0.025
F_2	2.5	1.2	0.25	0.28	0.545	0.991	0.993	0.223	0.043	0.035
F_2	1.5	0.4	0.41	0.19	0.440	1.005	1.006	0.325	0.052	0.040
F_2	1.5	1.2	0.41	0.32	0.421	0.996	0.997	0.348	0.072	0.056
F_2	0.6	0.4	0.58	0.25	0.296	1.001	0.998	0.506	0.091	0.064
F_2	0.6	1.2	0.58	0.43	0.308	0.970	0.961	0.492	0.141	0.111
F_3	2.5	0.4	0.22	0.12	0.701	0.995	0.995	0.095	0.014	0.014
F_3	2.5	1.2	0.21	0.24	0.685	0.996	0.995	0.107	0.017	0.016
F_3	1.5	0.4	0.41	0.14	0.504	1.000	0.998	0.253	0.025	0.023
F_3	1.5	1.2	0.40	0.29	0.507	0.992	0.992	0.252	0.027	0.026
F_3	0.6	0.4	0.60	0.21	0.328	0.989	0.987	0.458	0.047	0.044
F_3	0.6	1.2	0.60	0.39	0.352	0.969	0.966	0.428	0.071	0.066

$N(\theta_C, 1)$ distribution, and $(\theta_T - t_i)/2$ has the exponential distribution with mean 1, where θ_C and θ_T assume different values given in Table 3. These different values of θ_C and θ_T provide a wide range of censoring probabilities $p_C = \sum_{i=1}^{100} P\{y_i > c_i\}/100$ and truncation probabilities $p_T = \sum_{i=1}^{100} P\{\tilde y_i < t_i\}/100$, with p_C ranging from 21 to 60% and p_T ranging from 12 to 43%. Note that $100(1 - p_T)$, which ranges from 57 to 88, is the expected size of the observable sample $\{(\tilde y_i^o, t_i^o, \delta_i^o, x_i^o), 1 \le i \le n\}$ in which $\tilde y_i^o \ge t_i^o$. From this sample, we first compute the "naive estimator"

$$\hat\beta_{\rm NA} = \frac{\left\{\sum_{i=1}^n \left(x_i^o - \bar x^o\right)\tilde y_i^o\right\}}{\left\{\sum_{i=1}^n \left(x_i^o - \bar x_i^o\right)^2\right\}}.$$

We use $\hat\beta_{\rm NA}$ as a convenient starting value for the iterative search of a zero-crossing for the second equation in (2.24) (with $a = 0$) that defines the M-estimator $\hat\beta_{\rm BJ}$ [with ψ in (2.24) given by $\psi(u) = u$] or $\hat\beta_{\rm H}$ [with ψ given by Huber's score function $\psi(u) = (-2) \vee (u \wedge 2)$]. Table 3 gives the mean $E\hat\beta$ and mean squared error $E(\hat\beta - \beta)^2$ for $\hat\beta = \hat\beta_{\rm NA}, \hat\beta_{\rm BJ}, \hat\beta_{\rm H}$, and each result is based on 1000 simulations. These results show that $\hat\beta_{\rm BJ}$ and $\hat\beta_{\rm H}$ have negligible bias except when p_T reaches its highest levels 0.42, 0.43 and 0.39, that $\hat\beta_{\rm NA}$ has a substantial bias whose square gives most of its mean squared error and that $\hat\beta_{\rm H}$ has a smaller mean squared error than $\hat\beta_{\rm BJ}$.

In Tables 2 and 3, we have used two different choices of c for Huber's score function $\psi(u) = (-c) \vee (u \wedge c)$, with $c = 2$ in Table 3 and $c = 1$ in the $\psi_{\rm H}$ of

Table 2. For complete data (without censoring and truncation), a well-known robust regression method is to use ψ_{H} in conjunction with suitably standardized residuals in defining M-estimators. Such standardization requires some estimate of the location and scale of the underlying error distribution, typically carried out by including a location parameter in the model [as in (1.1)] and by concomitant estimation of a scale parameter [cf. Huber (1973) and Hampel, Ronchetti, Rousseuw and Stahel (1986)]. An extension of this approach that involves concomitant scale estimation will be presented elsewhere. An alternative approach, which involves adaptive estimation of an asymptotically optimal score function (instead of an appropriate scale parameter to be used in conjunction with Huber's score function ψ_{H}), is given in Section 5.

4. Consistency and asymptotic normality. Throughout the sequel we shall assume that

$$\psi \text{ is continuously differentiable}, \limsup_{|t|\to\infty} |\psi'(t)| < \infty \quad \text{and} \int_{-\infty}^{\infty} \psi^2(t)\, dF(t) < \infty. \tag{4.1}$$

Because of the instability of the product-limit estimator (2.18) at points u for which the "risk set size" $Z_n(b, u)$ is small compared with n, some modification of (2.26) is needed to make the associated M-estimator more tractable. In the formal influence curve calculations in Section 3, we have tacitly assumed $n^{-1}Z_n(b, u)$ to be bounded away from 0, since the functionals in (3.8), (3.10) and (3.17) would become singular otherwise. Without artificially restricting u within such a range, we shall use ideas similar to those introduced in Lai and Ying (1991c, 1992b) for the special case ψ_{BJ} to modify (2.26) and to establish consistency and asymptotic normality of the corresponding M-estimators. Letting

$$\begin{aligned} Z_n^x(b, u) &= \sum_{i=1}^n x_i^o I_{\{t_i^o(b) \le u \le \tilde{y}_i^o(b)\}}, \\ J_n^x(b,u) &= \sum_{i=1}^n x_i^o \Big\{ (1-\delta_i^o) I_{\{\tilde{y}_i^o(b) \ge u\}} + I_{\{t_i^o(b) \le u\}} \Big\}, \qquad -\infty < u < \infty, \end{aligned} \tag{4.2}$$

it will be convenient to express the second equation in (2.24) coupled with (2.26) as

$$\begin{aligned} &-\int_{u\in\mathbf{R}} \psi(u-a)\, dZ_n^x(b,u) \\ &\quad - \int_{u\in\mathbf{R}} \left\{ \int_u^{\infty} \Big(1 - \widehat{F}_{n,b}(v \mid u)\Big) \psi'(v-a)\, dv \right\} dJ_n^x(b,u) = 0, \end{aligned} \tag{4.3}$$

where

$$1 - \widehat{F}_{n,b}(v \mid u) = \frac{1 - \widehat{F}_{n,b}(v)}{1 - \widehat{F}_{n,b}(u)} = \prod_{i:u < \tilde{y}_i^o(b) \le v,\ \delta_i^o = 1} \left(1 - \frac{1}{Z_n(b, \tilde{y}_i^o(b))}\right) \quad \text{for } v \ge u. \tag{4.4}$$

This follows from integration by parts applied to the last two terms of (2.26) if $t_i^o(b) > -\infty$, noting that the last term of (2.26) vanishes if $t_i^o(b) = -\infty$.

4.1. *Smoothing kernels to dampen the instability due to small risk set sizes.* In view of its definition in (2.17), the risk set size $Z_n(b, u)$ is small if either $L_n(b, u)$ or $N_n(b, u)$ is small, where

$$L_n(b,u) = \sum_{i=1}^n I_{\{t_i^o(b) \le u\}}, \qquad N_n(b,u) = \sum_{i=1}^n I_{\{\tilde{y}_i^o(b) \ge u\}}. \tag{4.5}$$

We shall use two smoothing kernels p_{n1} and p_{n2} to down-weight those points u with small $L_n(b,u)/n$ or $N_n(b,u)/n$. Specifically, let p be a twice continuously differentiable function on the real line such that

$$p(u) = \begin{cases} 0, & \text{for } u \le 0, \\ 1, & \text{for } u \ge 1. \end{cases} \tag{4.6a}$$

Let c and d be positive constants and define, for $n \ge 2$,

$$\begin{aligned} p_{n,1}(b,u) &= p\left(\left(\frac{L_n(b,u)}{n} - \frac{c}{\log n}\right)\log n\right), \\ p_{n,2}(b,u) &= p\left(\left(\frac{N_n(b,u)}{n} - \frac{d}{\log n}\right)\log n\right). \end{aligned} \tag{4.6b}$$

Define, for $-\infty < u < \infty$,

$$Z_{n,p}^x(b,u) = \sum_{i=1}^n x_i^o p_{n,1}(b, t_i^o(b)) p_{n,2}(b,u) I_{\{t_i^o(b) \le u \le \tilde{y}_i^o(b)\}}, \tag{4.7}$$

which is a step function with jumps at $t_i^o(b)$ and $\tilde{y}_i^o(b)$, $i = 1, \dots, n$. Moreover, define

$$\begin{aligned} J_{n,p}^x(b,u) = {} & \sum_{i=1}^n x_i^o p_{n,1}(b, t_i^o(b)) \\ & \times \left\{(1 - \delta_i^o) I_{\{\tilde{y}_i^o(b) \ge u\}} + I_{\{t_i^o(b) \le u\}}\right\}, \qquad -\infty < u < \infty, \end{aligned} \tag{4.8}$$

as a modification of J_n^x in (4.2). In analogy with the left-hand side of (4.3), define

$$\begin{aligned}\widehat{\xi}_n(a,b) = &-\int_{u\in\mathbf{R}} \psi(u-a)\,dZ_{n,p}^x(b,u)\\ &-\int_{u\in\mathbf{R}}\left\{\int_u^\infty \left(1-\widehat{F}_{n,b}(v\mid u)\right)\psi'(v-a)p_{n,2}(b,v)\,dv\right\}dJ_{n,p}^x(b,u).\end{aligned} \tag{4.9}$$

We propose to replace (4.3) by the equation $\widehat{\xi}_n(a,b)=0$, which is analogous to the modified Buckley–James estimator of β introduced by Lai and Ying (1991c). In this connection, Lemma 1 of Lai and Ying (1991c), which shows that such modification of the Buckley–James estimator does not produce bias, can be extended as follows. Define

$$F(v\mid u) = \frac{F(v)-F(u)}{1-F(u)} \quad \text{for } v\ge u. \tag{4.10}$$

LEMMA 1. *Let g_1 be a bounded function and let g_2 be a function of bounded variation on* $\mathbf{R}$.

(i) *If ε_i is independent of (t_i,c_i,x_i) and has distribution function F, then for every a,*

$$\begin{aligned} E\Bigg\{&-\int_{-\infty}^{\infty}\psi(u-a)g_1\big(t_i(\beta)\big)d\big[g_2(u)I_{\{t_i(\beta)\le u\le \varepsilon_i\wedge c_i(\beta)\}}\big]\\ &-\int_{-\infty}^{\infty}\left[\int_u^\infty\left(1-F(v|u)\right)\psi'(v-a)g_2(v)\,dv\right]\\ &\times g_1\big(t_i(\beta)\big)\,d\big[-I_{\{t_i(\beta)\le c_i(\beta)\le u,\ c_i(\beta)<\varepsilon_i\}}\\ &\qquad +I_{\{t_i(\beta)\le c_i(\beta)\wedge\varepsilon_i\wedge u\}}\big]\,\Big|\,x_i,c_i,t_i\Bigg\}=0.\end{aligned} \tag{4.11}$$

(ii) *If $c_i\ge t_i$ a.s., then* (4.11) *still holds with ε_i replaced by ε_i^*, whose conditional distribution given (t_i,c_i,x_i) is*

$$P\{\varepsilon_i^*\le u\mid t_i,c_i,x_i\} = \frac{F(u)-F\big(t_i(\beta)\big)}{1-F\big(t_i(\beta)\big)},\qquad u\ge t_i(\beta).$$

Lemma 1 can be proved by applying integration by parts to the left-hand side of (4.11). The choice of the weight functions $p_{n,1}$ and $p_{n,2}$ in our modification of the M-estimator is quite flexible in practice. The basic idea is to restrict the range of integration in (4.3) only to u for which the risk set size $Z_n(b,u)$ is not too small. The thorny issue of bias created by such trimming does not arise because of Lemma 1, which has led to the definition (4.9). Without having to worry about bias, we can in fact trim quite substantially to ensure at least a moderate risk set size $Z_n(b,u)$, although our experience from simulation studies suggests that discarding u with $L_n(b,u) \le 2$ or $N_n(b,u) \le 2$ is often adequate to avoid potential difficulties with $\widehat{F}_{n,b}(u)$. Instead of straightforward deletion of such u, we use in (4.6b) a smooth version analogous to the kernel method in density estimation. This enables us to establish below asymptotic linearity of the random function $\widehat{\xi}_n(a,b)$ near $(\alpha,\ \beta)$ in Theorems 1 and 2.

4.2. *Consistency, asymptotic linearity and asymptotic normality of slope estimate under independent* (t_i, c_i, x_i^T). Suppose that (t_i, c_i, x_i^T) are independent random vectors that are independent of $\{\varepsilon_n\}$ and such that *either*

$$\sup_n E\big(|t_n|^\delta + (c_n^-)^\delta\big) < \infty \quad \text{for some } \delta > 0, \tag{4.12a}$$

which precludes the censored regression model with $t_i \equiv -\infty$, *or*

$$t_i \equiv -\infty \quad \text{and} \quad E\exp(\theta\varepsilon_1^-) + \sup_n E\exp(\theta c_n^-) < \infty \quad \text{for some } \theta > 0, \tag{4.12b}$$

where $a^- = |a| I_{\{a \le 0\}}$. Letting $t_i(b) = t_i - b^T x_i$ and $c_i(b) = c_i - b^T x_i$, suppose that the following hold:

$$\|x_i\| \le K \quad \text{for all } i \text{ and some nonrandom constant } K; \tag{4.13}$$

$$\begin{aligned} \sup_{\|b\| \le \rho,\ -\infty < u < \infty} \sum_{i=1}^m \Big[& P\{u \le t_i(b) \le u + h\} \\ & + P\{u \le c_i(b) \le u + h\}\Big] = O(mh) \\ & \text{as } h \to 0 \text{ and } m \to \infty \text{ with } mh \to \infty; \end{aligned} \tag{4.14}$$

F has a twice continuously differentiable density f such that

$$\begin{aligned} & \int_{-\infty}^{\infty} \left(\frac{f'}{f}\right)^2 dF < \infty \\ & \text{and} \quad \int_{-\infty}^{\infty} \sup_{|h| \le \eta} \{|f'(t+h)| + |f''(t+h)|\}\, dt < \infty \quad \text{for some } \eta > 0; \end{aligned} \tag{4.15}$$

$$\begin{aligned} & m^{-1} \sum_{i=1}^m P\{t_i(\beta) \le s \le c_i(\beta)\} \to \Gamma_0(s), \\ & m^{-1} \sum_{i=1}^m E\{x_i I_{\{t_i(\beta) \le s \le c_i(\beta)\}}\} \to \Gamma_1(s), \\ & m^{-1} \sum_{i=1}^m E\{x_i x_i^T I_{\{t_i(\beta) \le s \le c_i(\beta)\}}\} \to \Gamma_2(s) \quad \text{for } -\infty < s < F^{-1}(1). \end{aligned} \tag{4.16}$$

Assumptions (4.12)–(4.16) are essentially the same as those made by Lai and Ying (1991c, 1992b) in their analysis of the modified Buckley–James and

Tsui–Jewell–Wu estimators of β. The assumption of independent (t_i, c_i, x_i^T) in the present setting does not hold for the model (2.12), and an asymptotic theory of M-estimators in the model (2.12) will be given in Section 4.4. A basic idea in our analysis of (4.9) under these assumptions is to regard the observed sample $(t_i^o, x_i^o, \tilde{y}_i^o, \delta_i^o)$, $i = 1, \dots, n$, of left truncated and right censored observations as having been generated by a larger, randomly stopped sample of independent random vectors (t_i, c_i, x_i^T, y_i), $i = 1, \dots, m(n)$, where

$$m(n) = \inf\left\{ m\colon \sum_{i=1}^{m} I_{\{t_i \le y_i \wedge c_i\}} = n \right\}. \tag{4.17}$$

By the strong law of large numbers, as $m \to \infty$,

$$m^{-1} \sum_{i=1}^{m} I_{\{t_i \le y_i \wedge c_i\}} - m^{-1} \sum_{i=1}^{m} P\{t_i \le y_i \wedge c_i\} \to 0 \quad \text{a.s.} \tag{4.18}$$

Hence, under (4.16) and the assumption

$$\begin{aligned} &\tau_0 := \inf\{s\colon \Gamma_0(s) > 0\} < \tau := \inf\Big\{s > \tau_0\colon (1 - F(s))\Gamma_0(s) = 0\Big\}, \\ &\lim_{n \to \infty} m^{-1} \sum_{i=1}^{m} P\{t_i(\beta) \le c_i(\beta) < s\} = G(s) \quad \text{exists for every } s \in (\tau_0, \tau), \end{aligned} \tag{4.19}$$

it follows from (4.17) and (4.18) that

$$\begin{aligned} \frac{m(n)}{n} \to \Delta \quad \text{a.s.,} \quad \text{where } \frac{1}{\Delta} &= \int_{-\infty}^{\infty} \{\Gamma_0(s) + G(s)\} dF(s) \\ &= \lim_{m \to \infty} m^{-1} \sum_{i=1}^{m} P\{t_i \le y_i \wedge c_i\}. \end{aligned} \tag{4.20}$$

Another important idea in the analysis of the random function (4.9) is to approximate it using certain nonrandom functions $\xi_m(a, b)$. This is the content of the following lemma, which can be proved by straightforward modifications of the arguments used in the proof of Lemma 2 of Lai and Ying (1991c). To define the functions ξ_m, let

$$\overline{Z}_m(b, u) = \sum_{i=1}^{m} P\{t_i(b) \le u \le \tilde{y}_i(b)\}, \qquad \overline{L}_m(b, u) = \sum_{i=1}^{m} P\{t_i(b) \le \tilde{y}_i(b) \wedge u\},$$

$$\overline{N}_m(b, u) = \sum_{i=1}^{m} P\{\tilde{y}_i(b) \ge u \vee t_i(b)\}, \qquad \overline{n}_m = \sum_{i=1}^{m} P\{t_i \le \tilde{y}_i\},$$

$$\overline{p}_{m,1}(b, u) = p\left(\left(\frac{\overline{L}_m(b, u)}{\overline{n}_m} - \frac{c}{\log \overline{n}_m}\right) \log \overline{n}_m\right),$$

$$\bar{p}_{m,2}(b,u) = p\left(\left(\frac{\overline{N}_m(b,u)}{\bar{n}_m} - \frac{d}{\log \bar{n}_m}\right)\log \bar{n}_m\right),$$

$$\overline{Z}^x_{m,p}(b,u) = \sum_{i=1}^{m} E\left\{x_i \bar{p}_{m,1}(b,t_i(b))\bar{p}_{m,2}(b,u) I_{\{t_i(b)\le u \le \tilde{y}_i(b)\}}\right\},$$

$$\overline{J}^x_{m,p}(b,u) = \sum_{i=1}^{m} E\left\{x_i \bar{p}_{m,1}(b,t_i(b))\left(-I_{\{c_i<y_i,\, t_i(b)\le \tilde{y}_i(b)\le u\}} + I_{\{t_i(b)\le \tilde{y}_i(b)\wedge u\}}\right)\right\}.$$

Letting $\overline{H}_m(b,u) = \sum_{i=1}^m P\{y_i \le c_i,\, t_i(b) \le \tilde{y}_i(b) \le u\}$, define for $v \ge u$,

$$F_{m,b}(v\,|\,u) = 1 - \exp\left\{-\int_{u<s\le v} \frac{d\overline{H}_m(b,s)}{\overline{Z}_m(b,s)}\right\}, \tag{4.21}$$

$$\begin{aligned}\xi_m(a,b) = &-\int_{-\infty}^{\infty} \psi(u-a)\,d\overline{Z}^x_{m,p}(b,u)\\ &-\int_{-\infty}^{\infty}\left\{\int_u^{\infty}\left(1-F_{m,b}(v\,|\,u)\right)\psi'(v-a)\right.\\ &\qquad\left.\times\, \bar{p}_{m,2}(b,v)\,dv\right\} d\overline{J}^x_{m,p}(b,u).\end{aligned} \tag{4.22}$$

LEMMA 2. *Assume* (4.1) *and* (4.12)–(4.15) *and suppose that* $\lim_{m\to\infty}\bar{n}_m/m$ *exists and is positive. Then, for every* $\varepsilon > 0$ *and* $0 < \gamma < 1$ *and for any positive numbers A and B*,

$$\sup_{|a|\le A,\ \|b\|\le B} \|\hat{\xi}_n(a,b) - \xi_{m(n)}(a,b)\| = o\left(n^{1/2+\varepsilon}\right) \quad a.s., \tag{4.23}$$

$$\begin{aligned}&\sup\{\|\hat{\xi}_n(a,b) - \hat{\xi}_n(a',b') - \xi_{m(n)}(a,b) + \xi_{m(n)}(a',b')\|:\\ &\qquad |a-a'| + \|b-b'\| \le n^{-\gamma},\ |a|\vee|a'|\le A,\ \|b\|\vee\|b'\|\le B\}\\ &\quad = o\left(n^{(1-\gamma)/2+\varepsilon}\right) \quad a.s.\end{aligned} \tag{4.24}$$

Since $F_{m,\beta}(v\,|\,u) = F(v\,|\,u)$ by (4.21), setting $b = \beta$ in (4.22) and applying Lemma 1 yield

$$\xi_m(a,\beta) = 0 \quad \text{for every } a. \tag{4.25}$$

This suggests that we can choose any a and estimate β by the estimating equation $\hat{\xi}_n(a,b) = 0$. Throughout the sequel we shall assume knowledge of an upper bound $B > \|\beta\|$ so that we can restrict b to the ball $\{b\colon \|b\| \le B\}$. More precisely, an estimator $\hat{\beta}_n(a)$ of β will be defined as a minimizer of $\|\hat{\xi}_n(a,b)\|$ over the region $\{b\colon \|b\| \le B\|$. In the one-dimensional case $\nu = 1$, we can also define $\hat{\beta}_n(a)$ alternatively as a zero-crossing (in the interval $[-B,B]$) of $\hat{\xi}_n(a,b)$. The following theorem, which is an extension of Theorems 1 and 3 of Lai and Ying (1991c) on the modified Buckley–James estimator, establishes the asymptotic normality of $\widehat{\beta}_n(a)$ via the asymptotic linearity of $\widehat{\xi}_n(a,b)$ in the neighborhood $\{b\colon \|b\| \le n^{-\varepsilon}\}$ of β and the asymptotic normality of $n^{-1/2}\widehat{\xi}_n(a,\beta)$.

THEOREM 1. *Under the assumptions* (4.1), (4.12)–(4.16) *and* (4.19), *define* $\xi_m(a,b)$ *by* (4.22).

(i) *If* $\liminf_{m\to\infty} m^{-1}\{\inf_{\|b\|\le B,\ \|b-\beta\|\ge\delta}\|\xi_m(a,b)\|\} > 0$ *a.s. for every* $\delta > 0$, *then* $\widehat{\beta}_n(a) \to \beta$ *a.s.*

(ii) *Defining* τ_0 *and* τ *as in* (4.19), *assume that*

$$\lim_{m\to\infty}\frac{\log m}{m}\sum_{i=1}^{m}\Big[P\{t_i(\beta) < \tau_0 - \varepsilon\}I_{\{F(\tau_0)>0\}} + P\{c_i(\beta) > \tau + \varepsilon\}I_{\{F(\tau)<1\}}\Big] = 0 \tag{4.26}$$

for every $\varepsilon > 0$. *Define the* $\nu\times\nu$ *matrix*

$$C_a = \int_{\tau_0}^{\tau}\left\{\Gamma_2(u) - \frac{\Gamma_1(u)\Gamma_1^T(u)}{\Gamma_0(u)}\right\}\left\{\frac{\int_u^\tau (1-F(s))\psi'(s-a)\,ds}{1-F(u)}\right\} \times\left\{\frac{f'(u)}{f(u)} + \frac{f(u)}{1-F(u)}\right\}dF(u). \tag{4.27}$$

Then with probability 1, *for every* $\varepsilon > 0$,

$$\widehat{\xi}_n(a,b) = \widehat{\xi}_n(a,\beta) - n\Delta C_a(b-\beta) + o\big(n^{1/2}\vee n\|b-\beta\|\big)\quad \textit{uniformly in } \|b-\beta\| \le n^{-\varepsilon}, \tag{4.28}$$

where Δ *is defined in* (4.20). *Moreover, as* $n\to\infty$, $n^{-1/2}\widehat{\xi}_n(a,\beta)$ *has a limiting normal distribution with mean* 0 *and covariance matrix* ΔV_a, *where*

$$V_a = \int_{\tau_0}^{\tau}\left\{\Gamma_2(u) - \frac{\Gamma_1(u)\Gamma_1^T(u)}{\Gamma_0(u)}\right\}\left\{\frac{\int_u^\tau (1-F(s))\psi'(s-a)\,ds}{1-F(u)}\right\}dF(u). \tag{4.29}$$

(iii) *Suppose that* (4.26) *holds, that* (4.27) *is nonsingular and that for some* $\varepsilon\in(0,\frac{1}{2})$,

$$\lim_{m\to\infty} m^{-(1/2+\varepsilon)}\left\{\inf_{\|b\|\le B,\ \|b-\beta\|\ge m^{-\varepsilon}}\|\xi_m(a,b)\|\right\} = \infty. \tag{4.30}$$

Then $\widehat{\beta}_n(a) = \beta + O(n^{-\varepsilon})$ *a.s. and* $\sqrt{n}(\widehat{\beta}(a)-\beta)$ *has a limiting normal distribution with mean* 0 *and covariance matrix* $\Delta^{-1}C_a^{-1}V_aC_a^{-1}$.

Part (i) of the theorem is an immediate consequence of (4.23) and (4.25). Part (ii) will be proved in the Appendix. To prove part (iii), first note that (4.30), (4.23) and (4.20) imply that $n^{-1/2-\varepsilon} \inf_{\|b\| \le B, \|b-\beta\| \ge n^{-\varepsilon}} \|\widehat{\xi}_n(a,b)\| \to \infty$ a.s., and then apply part (ii) of the theorem.

Suppose that (t_i, c_i, x_i^T) are i.i.d. Then conditions (4.16), (4.19) and (4.26) are trivially satisfied. Moreover, (4.14) reduces to the simple Lipschitz condition

$$\sup_{\|b\| \le B, \ -\infty < u < \infty} \Big[P\{u \le t_1(b) \le u+h\} + P\{u \le c_1(b) \le u+h\}\Big] = O(h)$$

as $h \to 0$. Furthermore, $m^{-1}\xi_m(a,b) \to \xi(a,b)$ as $m \to \infty$, uniformly in $\|b\| \le B$, where ξ is obtained by replacing $\overline{p}_{m,1}$ and $\overline{p}_{m,2}$ in $\overline{Z}^x_{m,p}$, $\overline{J}^x_{m,p}$ and in (4.22) by the indicator function of $[\tau_0, \tau]$ and setting $m = 1$ in (4.22). Note that $\xi(a,b)$ is continuous in $\|b\| \le B$. Hence, if $\xi(b) \ne 0$ for $b \ne \beta$ (with $\|b\| \le B$), then the assumption of Theorem 1(i) is satisfied. Moreover, if $F(\tau) < 1$, then by Lemma 2 and an argument similar to the proof of Lemma 3(iii) of Lai and Ying (1991c), $\xi_m(a,b) \sim -m\Delta C_a(b-\beta)$ as $m \to \infty$ and $b \to \beta$, and therefore, when C_a is nonsingular,

$$\liminf_{m \to \infty} m^{-(1-\varepsilon)} \Big\{ \inf_{m^{-\varepsilon} \le \|b-\beta\| \le \delta} \|\xi_m(a,b)\| \Big\} > 0,$$

for every $\varepsilon > 0$ and some $\delta > 0$. Hence (4.30) holds if $F(\tau) < 1$ and $\xi(b) \ne 0$ for $b \ne \beta$ ($\|b\| \le B$) and if C_a is nonsingular.

4.3. *Asymptotic normality of M-estimators of α and β when ψ has compact support.* In the classical theory of M-estimators based on complete data, the location parameter α in (1.1) is commonly chosen as a zero of the function $E\psi(y_1 - \beta^T x_1 - a) = \int_{-\infty}^{\infty} \psi(u-a)\,dF(u)$, or equivalently, as a minimizer of the function $E\rho(y_1 - \beta^T x_1 - a)$ with $\rho' = \psi$. However, in the present setting of left truncated and right censored data, F can only be consistently estimated within the interval (τ_0, τ) even if β should be assumed known [cf. Lai and Ying (1991a)], and we therefore have to replace $E\psi(y_1 - \beta^T x_1 - a)$ by $\int_{\tau_0}^{\tau} \psi(u-a)\,dF(u)$. Consequently, the location parameter α in (1.1) will be chosen as a solution of

$$\int_{\tau_0}^{\tau} \psi(u-\alpha)\,dF(u) = 0. \tag{4.31}$$

Under (4.1) the left-hand side of (4.31) is continuous in α, and (4.31) has a unique solution if ψ is strictly increasing (or strictly decreasing) and $\lim_{a \to \infty} \psi(a)$ and $\lim_{a \to -\infty} \psi(a)$ have opposite signs.

Suppose that (4.31) has a unique solution α in the interior of some given interval $[A_0, A_1]$ and

(4.32) ψ has support $[s_0, s_1]$ with $s_0 + A_0 > \tau_0$, $s_1 + A_1 < \tau$.

Under the assumption (4.32), we can rewrite (4.31) as $\int_{\tau_0}^{\tau} \psi(u-a)\,dF(u \,|\, s_0 + A_0) = 0$, where $F(\cdot \,|\, \cdot)$ is defined in (4.10). Defining $\widehat{F}_{n,b}(\cdot|\cdot)$ as in (4.4), let

$$\widehat{\zeta}_n(a,b) = n \int_{s_0 + A_0}^{s_1 + A_1} \psi(u-a)\,d\widehat{F}_{n,b}(u \,|\, s_0 + A_0). \tag{4.33}$$

To estimate α and β, we use the following modifications of the estimating equations (2.24) [in which the first equation can be expressed in the form $n\int_{-\infty}^{\infty}\psi(u-a)\,d\widehat{F}_{n,b}(u)=0$]:

$$\widehat{\zeta}_n(a,b)=0,\qquad \widehat{\xi}_n(a,b)=0. \tag{4.34}$$

More precisely, $(\widehat{\alpha}_n,\widehat{\beta}_n)$ is defined as a minimizer of $|\widehat{\zeta}_n(a,b)|+\|\widehat{\xi}_n(a,b)\|$ over the region $\{A_0\le a\le A_1,\ \|b\|\le B\}$. Asymptotic properties of $(\widehat{\alpha}_n,\widehat{\beta}_n)$ are given in the following.

THEOREM 2. *Under the assumptions* (4.1), (4.12)–(4.16), (4.19), (4.26) *and* (4.32), *suppose that* (4.31) *holds for some* $\alpha\in(A_0,A_1)$. *Assume furthermore that*

$$\int_{\tau_0}^{\tau}\psi'(u-a)\,dF(u)\neq 0\quad\text{and}\quad\int_{\tau_0}^{\tau}\psi(u-a)\,dF(u)\neq 0 \tag{4.35}$$
$$\text{if } a\neq\alpha \text{ and } A_0\le a\le A_1.$$

Suppose that the matrix C_α *defined in* (4.27) *is nonsingular, and that there exists* $\varepsilon\in(0,\frac{1}{2})$ *such that*

$$\lim_{m\to\infty} m^{-(1/2+\varepsilon)}\left\{\inf_{A_0\le a\le A_1,\ \|b\|\le B,\ \|b-\beta\|\ge m^{-\varepsilon}}\|\xi_m(a,b)\|\right\}=\infty, \tag{4.36}$$

where $\xi_m(a,b)$ *is defined in* (4.22). *Then* $|\widehat{\alpha}_n-\alpha|+\|\widehat{\beta}_n-\beta\|=O(n^{-\varepsilon})$ *a.s. and* $\sqrt{n}(\widehat{\alpha}_n-\alpha,\widehat{\beta}_n^T-\beta^T)^T$ *has a limiting normal distribution with mean* 0 *and covariance matrix*

$$\frac{1}{\Delta}\begin{pmatrix}\psi_\alpha^{-2}(\nu_\alpha+g_\alpha^TC_\alpha^{-1}V_\alpha C_\alpha^{-1}g_\alpha) & \psi_\alpha^{-1}g_\alpha^TC_\alpha^{-1}V_\alpha C_\alpha^{-1}\\ \psi_\alpha^{-1}C_\alpha^{-1}V_\alpha C_\alpha^{-1}g_\alpha & C_\alpha^{-1}V_\alpha C_\alpha^{-1}\end{pmatrix}, \tag{4.37}$$

where Δ *is defined in* (4.20), V_α *is defined in* (4.29) *and*

$$\psi_\alpha=\int_{\tau_0}^{\tau}\psi'(u-\alpha)\,dF(u),$$

$$g_\alpha=\int_{\tau_0}^{\tau}\psi(u-\alpha)\,d\left\{(1-F(u))\int_{\tau_0}^{u}\frac{\Gamma_1(s)}{\Gamma_0(s)}\left(\frac{f'(s)}{f(s)}+\frac{f(s)}{1-F(s)}\right)\frac{dF(s)}{1-F(s)}\right\}, \tag{4.38}$$

$$\nu_\alpha=\int_{\tau_0}^{\tau}\int_{\tau_0}^{\tau}(1-F(u))(1-F(v))\psi'(u-\alpha)\psi'(v-\alpha)\times\int_{\tau_0}^{u\wedge v}\frac{dF(s)}{\Gamma_0(s)(1-F(s))^2}\,du\,dv.$$

The proof of Theorem 2 will be given in the Appendix. Although the influence curve results in Section 3 suggest via standard heuristic arguments [cf. Hampel, Ronchetti, Rousseeuw and Stahel (1986), page 85] a representation of $m(n)(\widehat{\alpha}_n - \alpha, \widehat{\beta}_n^T - \beta^T)$ as a sum of $m(n)$ independent random variables plus a negligible remainder, such representation is very difficult to establish rigorously, and our proof of Theorem 2 uses more direct martingale representations and asymptotic linearity arguments via empirical process theory.

4.4. *Asymptotic theory of M-estimators under independent* (t_i^o, c_i^o, x_i^o). The (t_i, c_i, x_i^T) defined in the truncation–censorship model (2.12) are clearly not independent random vectors. We now consider the alternative setting in which (t_i^o, c_i^o, x_i^{oT}) are independent random vectors with $c_i^o \geq t_i^o > -\infty$ and are independent of $\{\varepsilon_n\}$ and in which $\widetilde{y}_i^o = (\varepsilon_{\sigma_i} + \beta^T x_i^o) \wedge c_i^o$, where $\sigma_i = \inf\{n > \sigma_{i-1}\colon \varepsilon_n \geq t_i^o - \beta^T x_i^o\}$, as in (2.12). Let $\varepsilon_i^* = \varepsilon_{\sigma_i}$. Then $(t_i^o, c_i^o, x_i^{oT}, \varepsilon_i^*)$ are independent random vectors and

$$P\{\varepsilon_i^* \leq u \mid t_i^o, c_i^o, x_i^o\} = \frac{F(u) - F\big(t_i^o(\beta)\big)}{1 - F\big(t_i^o(\beta)\big)}, \qquad u \geq t_i^o(\beta), \tag{4.39}$$

so Lemma 1(ii) is applicable. Replace $t_i(b)$ by $t_i^o(b)$, $c_i(b)$ by $c_i^o(b)$ and (t_i, c_i, x_i) by (t_i^o, c_i^o, x_i^o) in the assumptions (4.12a), (4.13)–(4.15) and (4.26). Note that (4.12b) is no longer relevant here since it is assumed that $t_i^o > -\infty$. Moreover, replace the assumptions (4.16) and (4.19) by

$$\sup_i F\big(t_i^o(\beta)\big) < 1 \quad \text{a.s.},$$

$$\begin{aligned} &\lim_{n\to\infty} n^{-1} \sum_{i=1}^n E\left\{ \frac{I_{\{t_i^o(\beta) \leq s \leq c_i^o(\beta)\}}}{1 - F\big(t_i^o(\beta)\big)} \right\} = \Gamma_0(s), \\ &\lim_{n\to\infty} n^{-1} \sum_{i=1}^n E\left\{ \frac{x_i^o I_{\{t_i^o(\beta) \leq s \leq c_i^o(\beta)\}}}{1 - F\big(t_i^o(\beta)\big)} \right\} = \Gamma_1(s), \\ &\lim_{n\to\infty} n^{-1} \sum_{i=1}^n E\left\{ \frac{x_i^o x_i^{oT} I_{\{t_i^o(\beta) \leq s \leq c_i^o(\beta)\}}}{1 - F\big(t_i^o(\beta)\big)} \right\} = \Gamma_2(s) \quad \text{for } s < F^{-1}(1). \end{aligned} \tag{4.40}$$

Replacing $t_i(b)$ by $t_i^o(b)$, $\widetilde{y}_i(b)$ by $\widetilde{y}_i^o(b)$, $(t_i, \widetilde{y}_i)$ by $(t_i^o, \widetilde{y}_i^o)$ in the various quantities defined in Lemma 2, note that in particular $\overline{n}_m = \sum_{i=1}^m P\{t_i^o \leq \widetilde{y}_i^o\} = m$ and that (4.23) and (4.24) with these modifications and with $m(n)$ replaced by n still hold in the present setting, again by modifying the proof of Lemma 2 of Lai and Ying (1991c) that uses general tightness results for stochastic integrals of empirical-type processes established in Lai and Ying (1988). Define τ_0 and τ by (4.19) and set $\Delta = 1$. Then Theorems 1 and 2 with these modifications still hold in the present setting of independent (t_i^o, c_i^o, x_i^o) with $c_i^o \geq t_i^o > -\infty$, as can be shown by an obvious modification of their proofs given in the Appendix and by applying Lemma 1 of Lai and Ying (1992a) together with (4.39).

5. Confidence regions and asymptotically efficient M-estimators of β. Theorem 1 can be used to construct approximate $(1-\alpha)$-level confidence regions for β by an extension of the ideas of Wei, Ying and Lin (1990). Since a

is fixed in Theorem 1, we shall simply write $\widehat{\beta}_n$ instead of $\widehat{\beta}_n(a)$ and $\widehat{\xi}_n(\beta)$ instead of $\widehat{\xi}_n(a,\beta)$. In view of (4.16), (4.20) and (4.29), an obvious estimator of the matrix ΔV_a is

$$\widehat{V}_n = \sum_{j=1}^{n} \frac{1}{Z_n(\widehat{\beta}_n, \widetilde{y}_j^o(\widehat{\beta}_n))} \delta_j^o \left\{ n^{-1} \sum_{i=1}^{n} x_i^o x_i^{oT} I_{\{t_i^o(\widehat{\beta}_n) \le \widetilde{y}_j^o(\widehat{\beta}_n) \le \widetilde{y}_i^o(\widehat{\beta}_n)\}} - \frac{Z_n^x(\widehat{\beta}_n, \widetilde{y}_j^o(\widehat{\beta}_n)) \left[Z_n^x(\widehat{\beta}_n, \widetilde{y}_j^o(\widehat{\beta}_n))\right]^T}{n Z_n(\widehat{\beta}_n, \widetilde{y}_j^o(\widehat{\beta}_n))} \right\} \times \left\{ \int_{\widetilde{y}_j^o(\beta)}^{\infty} \left[1 - \widehat{F}_{n,\widehat{\beta}_n}(s \mid \widetilde{y}_j^o(\widehat{\beta}_n))\right] \psi'(s-a) p_{n,1}(\widehat{\beta}_n, s) p_{n,2}(\widehat{\beta}_n, s)\, ds \right\}^2, \tag{5.1}$$

which can be shown by arguments similar to those of Wei, Ying and Lin (1990) to converge a.s. to ΔV_a under the assumptions of Theorem 1. By Theorem 1(i), $n^{-1}\widehat{\xi}_n^T(\beta)V_a^{-1}\widehat{\xi}_n(\beta)$ has a limiting χ^2-distribution with p degrees of freedom. Since $\widehat{V}_n \to \Delta V_a$ a.s., it then follows that

$$\{b: \|b\| \le B,\ n^{-1}\widehat{\xi}_n^T(b)\widehat{V}_n^{-1}\widehat{\xi}_n(b) \le \chi^2_{1-\alpha;\,p}\} \tag{5.2}$$

is an approximate $(1-\alpha)$-level confidence region for β, where $\chi^2_{1-\alpha;\,p}$ denotes the $100(1-\alpha)$-percentile of the χ^2-distribution with p degrees of freedom.

Let $\lambda = f/(1-F)$ be the hazard function of f. Then

$$\frac{\lambda'}{\lambda} = \frac{f'}{f} + \frac{f}{1-F}. \tag{5.3}$$

Assuming that f is twice continuously differentiable and that $F(\tau) = 1$, we obtain from (5.3) and integration by parts that

$$\begin{aligned} \int_u^{\tau} \left[1 - F(s \mid u)\right] \left(\frac{\lambda'}{\lambda} - \lambda\right)'(s)\, ds &= \frac{1}{1-F(u)} \int_u^{\tau} (1 - F(s))\, d\left(\frac{f'}{f}\right) \\ &= -\frac{\lambda'(u)}{\lambda(u)}, \qquad u < \tau. \end{aligned} \tag{5.4}$$

Hence, for the special case $a = 0$ and $\psi = (\lambda'/\lambda) - \lambda$ in (4.27) and (4.29), $C_a^{-1}V_aC_a^{-1} = I_f^{-1}$, where

$$I_f = \int_{\tau_0}^{\tau} \left\{ \Gamma_2(u) - \frac{\Gamma_1(u)\Gamma_1^T(u)}{\Gamma_0(u)} \right\} \left\{ \frac{\lambda'(u)}{\lambda(u)} \right\}^2 dF(u). \tag{5.5}$$

Therefore, for the M-estimator $\widehat{\beta}_n$ associated with the score function $\psi = (\lambda'/\lambda) - \lambda$ (with $a = 0$), the covariance matrix of the limiting normal distribution of $\sqrt{n}(\widehat{\beta}_n - \beta)$ under suitable regularity conditions is $\Delta^{-1}I_f^{-1}$. Moreover, as shown by Lai and Ying (1992a), $\Delta^{-1}I_f^{-1}$ is an asymptotic lower bound for the covariance matrices of regular estimators in the semiparametric problem of estimating β when the density f of ε_1 and the distributions of the independent random vectors (t_i, c_i, x_i^T) are unknown.

Since f is unknown, the optimal score function $\psi = (\lambda'/\lambda) - \lambda$ is not available to form the asymptotically efficient M-estimator. We now extend the ideas of Lai and Ying (1991b), on adaptive choice of score functions in constructing asymptotically efficient rank estimators of β, to M-estimators. Divide the sample into two disjoint subsets, the first of which is $\{(t_i^o, x_i^o, \delta_i^o, \tilde{y}_i^o): i \leq n/2\}$. From the first subsample, define L_{n_1}, N_{n_1}, $Z_{n_1,p}^x$ and $J_{n_1,p}^x$ by (4.5), (4.7) and (4.8), in which n is replaced by n_1. Also define $\widehat{F}_{n_1,b}$ by (2.18) with n replaced by n_1. Let $\widehat{\psi}_{n,2}$ be an estimate of $(\lambda'/\lambda) - \lambda$ based on the second subsample of $n_2 = n - n_1$ observations and define, in analogy with (4.9),

$$\begin{aligned} \widehat{\xi}_{n_1}(b) = &- \int_{u=-\infty}^{\infty} \widehat{\psi}_{n,2}(u)\, dZ_{n_1,p}^x(b,u) \\ &- \int_{u=-\infty}^{\infty} \bigg\{ \int_u^{\infty} \Big(1 - \widehat{F}_{n_1,b}(v \mid u)\Big) \\ &\qquad \times \widehat{\psi}'_{n,2}(b) p_{n_1,2}(b,u)\, dv \bigg\}\, dJ_{n_1,p}^x(b,u). \end{aligned} \tag{5.6}$$

Likewise from the second subsample define

$$\begin{aligned} \widehat{\xi}_{[n,n_1]}(b) = &- \int_{-\infty}^{\infty} \widehat{\psi}_{n,1}(u)\, dZ_{[n,n_1],p}^x(b,u) \\ &- \int_{-\infty}^{\infty} \bigg\{ \int_u^{\infty} \Big(1 - \widehat{F}_{[n,n_1],b}(v \mid u)\Big) \\ &\qquad \times \widehat{\psi}'_{n,1}(v) p_{[n,n_1],2}(b,u)\, dv \bigg\}\, dJ_{[n,n_1]}^x(b,u), \end{aligned} \tag{5.7}$$

where $\widehat{\psi}_{n,1}$ represents an estimate of $(\lambda'/\lambda) - \lambda$ based on the first subsample and we use the notation $L_{[n,n_1]}$, $Z_{[n,n_1]}^x$, $J_{[n,n_1]}^x$ and so forth to denote (4.5), (4.7), (4.8) and so forth, in which the sum $\Sigma_{i=1}^n$ is replaced by $\Sigma_{i=n_1+1}^n$ (i.e., the summands are only from the second subsample). Combining the two subsample statistics (5.6) and (5.7) gives

$$\xi_n^*(b) = \widehat{\xi}_{n_1}(b) + \widehat{\xi}_{[n,n_1]}(b). \tag{5.8}$$

From the jth subsample, starting with preliminary estimates $b_{n,j}$ such that $b_{n,j} - \beta = O_p(n^{-r})$ for some $0 < r \leq \frac{1}{2}$, Lai and Ying (1991b) showed (i) how to construct from the uncensored residuals $\tilde{y}_i^o - b_{n,j}^T x_i^o$ in the jth subsample a smooth consistent estimate $\widehat{\lambda}_{n,j}$ of the hazard function λ and (ii) how to smooth $\widehat{\lambda}'_{n,j}/\widehat{\lambda}_{n,j}$ to obtain a smooth consistent estimate of λ'/λ. Using these smooth consistent estimates to define smooth consistent estimates $\widehat{\psi}_{n,2}$ and $\widehat{\psi}_{n,1}$ of

$(\lambda'/\lambda) - \lambda$ for (5.6) and (5.7), it can be shown by a modification of the proof of Theorem 1 in the Appendix and by the arguments of Lai and Ying [(1991b), Theorem 2] that the adaptive M-estimator β_n^*, defined as a minimizer of $\xi_n^*(b)$ in the sphere $\{b: \|b - (b_{n,1} + b_{n,2})/2\| \le n^{-\varepsilon}\}$ with $0 < \varepsilon < r$, is asymptotically efficient in the sense that

$$\sqrt{n}(\beta_n^* - \beta) \to_{\mathcal{L}} N\big(0, \Delta^{-1} I_f^{-1}\big), \tag{5.9}$$

under (4.12)–(4.16), (4.19) and the additional assumption

$$I_f \text{ is nonsingular, } F(\tau) = 1 \text{ and } f(s) > 0 \text{ for } \tau_0 < s < \tau. \tag{5.10}$$

The discussion in this section has been for the setting of independent random vectors (t_i, c_i, x_i^T) that are independent of $\{\varepsilon_n\}$, as is assumed in Theorem 1. As has been shown in Section 4.4, the conclusions of Theorem 1 can be extended to the setting in which $(t_i^o, c_i^o, x_i^{o\,T})$ are independent random vectors that are independent of $\{\varepsilon_n\}$ and such that $c_i^o \ge t_i^o > -\infty$ and (4.39) holds. Hence we can also extend the above construction of confidence regions and adaptive M-estimators of β to this setting.

APPENDIX

PROOF OF THEOREM 1(ii). The proof of (4.28) makes use of (4.24), (4.25) and an analysis of $\xi_m(a,b) - \xi_m(a,\beta)$ similar to the proof of Lemma 3(ii) of Lai and Ying (1991c). To show the asymptotic normality of $\widehat{\xi}_n(a,\beta)$, let $\Lambda(u) = -\log(1 - F(u))$ and $M_n(s) = \sum_{i=1}^{m(n)} I_{\{t_i(\beta) \le \varepsilon_i \le s \wedge c_i(\beta)\}} - \int_{-\infty}^{s} Z_n(\beta, u)\, d\Lambda(u)$. Analogous to the proof of Lemma 4 of Lai and Ying (1991c), we approximate $p_{n,j}(\beta, u)$ by $\bar{p}_{m(n),j}(\beta, u)$ $(j = 1, 2)$, and use arguments similar to the proofs of Lemma 2 of Lai and Ying (1992b) and Lemmas 5 and 6 of Lai and Ying (1991c) to show that $\widehat{\xi}_n(a,\beta) = \widetilde{\xi}_n^{(1)} + \widetilde{\xi}_n^{(2)} + o_p(\sqrt{n}\,)$, where

$$\begin{aligned} \widetilde{\xi}_n^{(1)} = {} & -\sum_{i=1}^{m(n)} x_i \bar{p}_{m(n),1}\big(\beta, t_i(\beta)\big) \\ & \times \Bigg\{ \int_{-\infty}^{\infty} \psi(u-a)\, d\big[\bar{p}_{m(n),2}(\beta,u) I_{\{t_i(\beta) \le u \le \tilde{y}_i(\beta)\}}\big] \\ & + \int_{-\infty}^{\infty} \bigg[\int_u^{\infty} \Big(1 - F(v \mid u)\Big) \bar{p}_{m(n),2}(\beta, v) \psi'(v-a)\, dv \bigg] \\ & \times d\big[-I_{\{t_i(\beta) \le c_i(\beta) \le u,\; c_i(\beta) < \varepsilon_i\}} + I_{\{t_i(\beta) \le u \wedge \tilde{y}_i(\beta)\}} \big] \Bigg\}, \end{aligned} \tag{A.1}$$

$$\widetilde{\xi}_n^{(2)} = \int_{\tau_0}^{\tau} \frac{\Gamma_1(u)}{\Gamma_0(u)} \bigg\{ \int_u^{\tau} \Big(1 - F(v \mid u)\Big) \psi'(v-a)\, dv \bigg\}\, dM_n(u). \tag{A.2}$$

An argument similar to the proof of Theorem 2(ii) of Lai and Ying (1991c) can be used to show that $n^{-1/2}(\widetilde{\xi}_n^{(1)} + \widetilde{\xi}_n^{(2)})$ has a limiting normal distribution with mean 0 and covariance matrix ΔV_a. □

PROOF OF THEOREM 2. From (4.25) and (4.36), it follows that $\widehat{\beta}_n \to \beta$ a.s. Since $s_0 - A_0 > \tau_0$, it can be shown by an argument similar to the proof (4.4) and (4.5) of Lai and Ying (1988) that

$$\text{(A.3)} \quad \sup_{\|b\| \leq B,\ \tau > u \geq s_0 + A_0} \left|\widehat{F}_{n,b}(u \mid s_0 + A_0) - F_{m(n),b}(u \mid s_0 + A_0)\right| = o\left(n^{-1/2+\delta}\right) \quad \text{a.s.,}$$

$$\text{(A.4)} \quad \begin{aligned} \sup_{\|b-\beta\| \leq n^{-\gamma},\ \tau > u \geq s_0 + A_0} &\left|\widehat{F}_{n,b}(u \mid s_0 + A_0) - \widehat{F}_{n,\beta}(u \mid s_0 + A_0)\right. \\ &\left. - F_{m(n),b}(u \mid s_0 + A_0) + F_{m(n),\beta}(u \mid s_0 + A_0)\right| \\ = o\left(n^{-(1+\gamma)/2+\delta}\right) \quad \text{a.s.,} \end{aligned}$$

for every $\delta > 0$ and $0 < \gamma < 1$, where $F_{m,b}(\cdot \mid \cdot)$ is defined in (4.21). Since $\widehat{\beta}_n \to \beta$ a.s. and since

$$\sup_{\tau > u \geq s_0 + A_0} \left|F_{m,b}(u \mid s_0 + A_0) - F(u \mid s_0 + A_0)\right| \to 0 \quad \text{as } m \to \infty \text{ and } b \to \beta,$$

we obtain from (A.3) that

$$\sup_{\tau > u \geq s_0 + A_0} \left|\widehat{F}_{n,\widehat{\beta}_n}(u \mid s_0 + A_0) - F(u \mid s_0 + A_0)\right| \to 0 \quad \text{a.s.,}$$

and therefore, by (4.1),

$$\text{(A.5)} \quad \begin{aligned} \sup_{A_0 \leq a \leq A_1} &\left| \int_{s_0 + A_0}^{s_1 + A_1} \psi(u - a)\, d\widehat{F}_{n,\widehat{\beta}_n}(u \mid s_0 + A_0) \right. \\ &\left. - \int_{s_0 + A_0}^{s_1 + A_1} \psi(u - a)\, dF(u \mid s_0 + A_0) \right| \to 0 \quad \text{a.s.} \end{aligned}$$

Since $\int_{s_0+A_0}^{s_1+A_1} \psi(u-a)\, dF(u \mid s_0 + A_0) = \{1 - F(s_0 + A_0)\}^{-1} \int_{\tau_0}^{\tau} \psi(u-a)\, dF(u)$ for $a \in [A_0, A_1]$ by (4.32), it follows from (A.5) and (4.35) that $\widehat{\alpha}_n \to \alpha$ a.s.

To simplify matters, multiply both sides of the estimating equation $\widehat{\zeta}_n(a,b) = 0$ by $\Delta(1 - F(s_0 + A_0))$, so that we shall work with $\widetilde{\zeta}_n = \Delta(1 - F(s_0 + A_0))\widehat{\zeta}_n$ instead of $\widehat{\zeta}_n$. In view of (A.4) and (4.32), we can use an argument similar to the proof of Lemma 3(ii) of Lai and Ying (1991c) to extend (4.28) to the following asymptotic linearity property: With probability 1, for every $\delta > 0$,

$$\text{(A.6)} \quad \begin{pmatrix} \widetilde{\zeta}_n(a,b) - \widetilde{\zeta}_n(\alpha,\beta) \\ \widehat{\xi}_n(a,b) - \widehat{\xi}_n(\alpha,\beta) \end{pmatrix} = n\Delta \begin{pmatrix} -\psi_\alpha & g_\alpha^T \\ 0 & -C_\alpha \end{pmatrix} \begin{pmatrix} a - \alpha \\ b - \beta \end{pmatrix} + o\left(\sqrt{n} \vee n\|b - \beta\| \vee n|a - \alpha|\right)$$

uniformly in $|a-\alpha| + \|b-\beta\| \le n^{-\delta}$. We next show that

$$\frac{1}{\sqrt{n}}\begin{pmatrix}\widetilde{\zeta}_n(\alpha,\beta)\\ \widehat{\xi}_n(\alpha,\beta)\end{pmatrix} \to_{\mathcal{L}} N\left(0, \Delta\begin{pmatrix}\nu_\alpha & 0\\ 0 & V_\alpha\end{pmatrix}\right). \tag{A.7}$$

From (4.1) and (4.31)–(4.33), it follows that

$$\begin{aligned} n^{-1}\widehat{\zeta}_n(\alpha,\beta) &= \int_{s_0+A_0}^{s_1+A_1} \psi(u-\alpha)\,d\Big[\widehat{F}_{n,\beta}(u\mid s_0+A_0) - F(u\mid s_0+A_0)\Big]\\ &= -\int_{s_0+A_0}^{s_1+A_1} \Big\{\widehat{F}_{n,\beta}(u\mid s_0+A_0) - F(u\mid s_0+A_0)\Big\}\psi'(u-\alpha)\,du. \end{aligned} \tag{A.8}$$

By Lai and Ying [(1991a), Theorem 5(i)] $\sqrt{n}\{F_{n,\beta}(\cdot\mid s_0+A_0) - F(\cdot|s_0+A_0)\}$ converges weakly in $D(s_0+A_0, s_1+A_1]$ to $(1-F(\cdot|s_0+A_0))W$, where $W(t)$ is a zero-mean Gaussian process with

$$\operatorname{Var}\big(W(t)\big) = \frac{1}{\Delta}\int_{s_0+A_0}^{t} \frac{dF(s)}{\Gamma_0(s)\big(1-F(s)\big)^2}$$

and with independent increments. From this and (A.8), it follows that

$$n^{-1/2}\widehat{\zeta}_n(\alpha,\beta) \to_{\mathcal{L}} N\Big(0, \{1-F(s_0+A_0)\}^{-2}\nu_\alpha/\Delta\Big).$$

Defining $\widetilde{\xi}_n^{(1)}$ and $\widetilde{\xi}_n^{(2)}$ by (A.1) and (A.2) with $a=\alpha$, it can be shown that

$$\begin{aligned} &n^{-1}E\{\widetilde{\xi}_n^{(1)}M_n(t)\}\\ &\quad \to -\Delta\int_{s_0+A_0}^{t} \Gamma_1(s)\left\{\int_s^{s_1+A_1}\Big(1-F(v\mid s)\Big)\psi'(v-\alpha)\,dv\right\}dF(s),\\ &n^{-1}E\{\widetilde{\xi}_n^{(2)}M_n(t)\}\\ &\quad \to \Delta\int_{s_0+A_0}^{t} \Gamma_1(s)\left\{\int_s^{s_1+A_1}\Big(1-F(v\mid s)\Big)\psi'(v-\alpha)\,dv\right\}dF(s). \end{aligned} \tag{A.9}$$

Moreover, by Lai and Ying [(1991a), (4.17)],

$$\begin{aligned} &\widehat{F}_{n,\beta}(u\mid s_0+A_0) - F(u\mid s_0+A_0)\\ &\quad = \Big(1-F(u\mid s_0+A_0)\Big)\int_{(s_0+A_0,u]} \frac{1-\widehat{F}_{n,\beta}(t-)}{1-F(t\mid s_0+A_0)} I_{\{Z_n(\beta,t)>0\}}\frac{dM_n(t)}{Z_n(\beta,t)}. \end{aligned} \tag{A.10}$$

From (A.8)–(A.10) and standard arguments involving Rebolledo's central limit theorem [cf. Gill (1980)], (A.7) follows, recalling that $\widetilde{\zeta}_n = \Delta(1-F(s_0+A_0))\widehat{\zeta}_n$.

From (A.6) and (A.7), we obtain as in Theorem 1(iii) that $|\widehat{\alpha}_n - \alpha| + \|\widehat{\beta}_n - \beta\| = O(n^{-\varepsilon})$ a.s. and that $\sqrt{n}(\widehat{\alpha}_n - \alpha, \widehat{\beta}_n^T - \beta^T)^T$ has a limiting normal distribution with mean 0 and covariance matrix

$$\left[\frac{1}{\Delta}\begin{pmatrix} -\psi_\alpha^{-1} & -\psi_\alpha^{-1} g_\alpha^T C_\alpha^{-1} \\ 0 & -C_\alpha^{-1} \end{pmatrix}\right]\left[\Delta\begin{pmatrix} \nu_\alpha & 0 \\ 0 & V_\alpha \end{pmatrix}\right] \times \left[\frac{1}{\Delta}\begin{pmatrix} -\psi_\alpha^{-1} & 0 \\ -\psi_\alpha^{-1} C_\alpha^{-1} g_\alpha & -C_\alpha^{-1} \end{pmatrix}\right] = (4.37),$$

noting that

$$\begin{pmatrix} -\psi_\alpha & g_\alpha^T \\ 0 & -C_\alpha \end{pmatrix}^{-1} = \begin{pmatrix} -\psi_\alpha^{-1} & -\psi_\alpha^{-1} g_\alpha^T C_\alpha^{-1} \\ 0 & -C_\alpha^{-1} \end{pmatrix}. \qquad \square$$

REFERENCES

BHATTACHARYA, P. K., CHERNOFF, H. and YANG, S.S. (1983). Nonparametric estimation of the slope of a truncated regression. *Ann. Statist.* **11** 505–514.

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

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

GILL, R. D. (1980). *Censoring and Stochastic Integrals. Math. Centre Tracts* **124**. Math. Centrum, Amsterdam.

HAMPEL, F. R., RONCHETTI, E. M., ROUSSEEUW, P. J. and STAHEL, W. A. (1986). *Robust Statistics: The Approach Based on Influence Functions*. Wiley, New York.

HUBER, P. J. (1973). Robust regression: asymptotics, conjectures and Monte Carlo. *Ann Statist.* **1** 799–821.

LAI, T. L. and YING, Z. (1988). Stochastic integrals of empirical-type processes with applications to censored regression. *J. Multivariate Anal.* **27** 334–358.

LAI, T. L. and YING, Z. (1991a). Estimating a distribution function with truncated and censored data. *Ann. Statist.* **19** 417–442.

LAI, T. L. and YING, Z. (1991b). Rank regression methods for left-truncated and right-censored data. *Ann. Statist.* **19** 531–554.

LAI, T. L. and YING, Z. (1991c). 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. (1992a). Asymptotically efficient estimation in censored and truncated regression models. *Statist. Sinica* **2** 17–46.

LAI, T. L. and YING, Z. (1992b). Asymptotic theory of a bias-corrected least squares estimator in truncated regression. *Statist. Sinica* **2** 519–539.

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

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

ORCHARD, T. and WOODBURY, M. A. (1972). A missing information principle: theory and applications. *Proc. Sixth Berkeley Symp. Math. Statist. Probab.* **1** 697–715. Univ. California Press, Berkeley.

REID, N. (1981). Influence functions for censored data. *Ann. Statist.* **9** 78–92.

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

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** 1314–1334.

TSUI, K.-L., JEWELL, N. P. and WU, C. F. J. (1988). A nonparametric approach to the truncated regression problem. *J. Amer. Statist. Assoc.* **83** 785–792.

TURNBULL, B. W. (1976). The empirical distribution function with arbitrarily grouped, censored and truncated data. *J. Roy. Statist. Soc. Ser. B* **38** 290–295.

WANG, J. G. (1987). A note on the uniform consistency of the Kaplan–Meier estimator. *Ann. Statist.* **15** 1313–1316.

WEI, L. J., YING, Z. and LIN, D. Y. (1990). Linear regression analysis of censored survival data based on rank tests. *Biometrika* **77** 845–851.

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
SEQUOIA HALL
STANFORD, CALIFORNIA 94305

DEPARTMENT OF STATISTICS
RUTGERS UNIVERSITY
HILL CENTER, BUSCH CAMPUS
NEW BRUNSWICK, NEW JERSEY 08903