A MISSING INFORMATION PRINCIPLE AND $M$-ESTIMATORS IN
REGRESSION ANALYSIS WITH CENSORED AND TRUNCATED DATA

by

Tze Leung Lai and Zhiliang Ying

TECHNICAL REPORT NO. 14
FEBRUARY 1991

Prepared Under Grant MDA 904-89-H-2040
For The National Security Agency

Reproduction in Whole or in Part is Permitted
for any Purpose of the United States Government

Approved for public release; distribution unlimited

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
A Missing Information Principle and $M$-Estimators in Regression Analysis with Censored and Truncated Data

by

Tze Leung Lai and Zhiliang Ying

TECHNICAL REPORT NO. 14
FEBRUARY 1991

PREPARED UNDER GRANT MDA 904-89-H-2040
FOR THE NATIONAL SECURITY AGENCY

Also supported by National Science Foundation Grant DMS87-15614 and issued as Technical Report No. 20, Department of Statistics, Stanford University.

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
A MISSING INFORMATION PRINCIPLE AND M-ESTIMATORS IN
REGRESSION ANALYSIS WITH CENSORED AND TRUNCATED DATA

TZE LEUNG LAI and ZHILIANG YING

Stanford University and University of Illinois

Abstract

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.

Abbreviated Title: M-ESTIMATORS FOR CENSORED/TRUNCATED DATA


Key words and phrases. EM algorithm, M-estimator, censoring, truncation, self consistency, linear regression, martingale, asymptotic normality.

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

Consider the linear regression model

\[(1.1) \quad y_i = \alpha + \beta^T x_i + \epsilon_i \quad (i = 1, 2, \cdots)\]

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

\[(1.2) \quad \sum_{i=1}^{n} \rho(y_i - a - b^T x_i) = \min!\]

In particular, when \(\rho(u) = u^2\), \(\hat{\alpha}\) and \(\hat{\beta}\) reduce to the classical least squares estimates, and they reduce to the maximum likelihood estimates of \(\alpha\) and \(\beta\) when \(\rho(u) = \log h(u)\), where \(h\) is the density function of \(\epsilon_i\). When \(\rho\) is differentiable, the M-estimators \(\hat{\alpha}, \hat{\beta}\) are also defined as a solution of the system of estimating equations

\[(1.3) \quad \sum_{i=1}^{n} \rho'(y_i - a - b^T x_i) = 0, \quad \sum_{i=1}^{n} x_i \rho'(y_i - a - b^T x_i) = 0.\]

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 \(\bar{y}_i = y_i \wedge c_i\) 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 \((\bar{y}_i, \delta_i, x_i)\) can be observed only when \(\bar{y}_i \geq t_i\). The data, therefore, consist of \(n\) observations \((\bar{y}_i, t_i, \delta_i, x_i)\) with \(\bar{y}_i \geq t_i\), \(i = 1, \cdots, n\).

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 \(\{\epsilon_i\}\). The special case \(t_i \equiv -\infty\) corresponds to the “censored regression model”, while the case \(c_i \equiv \infty\) corresponds to the “truncated regression model”. For the case \(\rho(u) = u^2\), Miller (1976) proposed the following extension of (1.2) to the censored regression model. Noting that the left hand side of (1.2) can be expressed in the form \(n \int_{-\infty}^{\infty} (z-a)^2 dF_{n,b}^\ast(z)\), where \(F_{n,b}^\ast\) is the empirical distribution function of \(y_i - b^T x_i\) \((i = 1, \cdots, n)\), he suggested replacing \(F_{n,b}^\ast\) by the Kaplan-Meier curve \(\hat{F}_{n,b}\) based on \((\bar{y}_i - b^T x_i, \delta_i)_{1 \leq i \leq n}\), and defined \(\hat{\alpha}, \hat{\beta}\) as a solution to the minimization problem

\[(1.4) \quad \sum_{i=1}^{n} \hat{\omega}_i(b)(\bar{y}_i - a - b^T x_i)^2 = \int_{-\infty}^{\infty} (z-a)^2 d\hat{F}_{n,b}(z) = \min!, \]
where the weights \( \hat{\omega}_i(b) \) are the jumps of \( \hat{F}_{n,b} \). Setting the derivative of (1.4) with respect to \( a \) equal to 0 gives \( \hat{\alpha} = \sum_{i=1}^{n} \hat{\omega}_i(\beta)(\bar{y}_i - \beta^T x_i) \) and therefore \( \beta \) can be determined as a solution to

\[
(1.5) \quad \sum_{i=1}^{n} \hat{\omega}_i(b)(\bar{y}_i - b^T x_i)^2 - \{ \sum_{i=1}^{n} \hat{\omega}_i(b)(\bar{y}_i - b^T x_i) \}^2 = \text{min}!
\]

Throughout the sequel we shall let \( F \) denote the common distribution function of \( \epsilon_i + \alpha \). An alternative extension of the classical least squares estimate to the censored regression model was proposed by Buckley and James (1979). Instead of (1.2), they started with (1.3) which reduces in the case \( \rho(u) = u^2 \) to the normal equations

\[
(1.6) \quad \sum_{i=1}^{n} (x_i - \bar{x}_n)(y_i - b^T x_i) = 0, \quad a = n^{-1} \sum_{i=1}^{n} (y_i - b^T x_i),
\]

where \( \bar{x}_n = n^{-1} \sum_{i=1}^{n} x_i \). For censored data \( (\bar{y}_i, \delta_i, x_i) \), noting that

\[
(1.7) \quad E(y_i | \bar{y}_i, \delta_i) = \delta_i \bar{y}_i + (1 - \delta_i) \{ \beta^T x_i + \int_{\bar{y}_i - b^T x_i}^{\infty} u dF(u) / (1 - F(\bar{y}_i - \beta^T x_i)) \},
\]

they proposed to replace the \( y_i - b^T x_i \) in (1.6) by the following estimate of \( E(y_i | \bar{y}_i, \delta_i) - b^T x_i \):

\[
(1.8a) \quad y_i^* (b) = \delta_i(\bar{y}_i - b^T x_i) + (1 - \delta_i) \{ \int_{\bar{y}_i - b^T x_i}^{\infty} u d\hat{F}_{n,b}(u) \} / \{ 1 - \hat{F}_{n,b}(\bar{y}_i - b^T x_i) \}.
\]

They therefore defined \( \beta \) as a solution to the estimating equation

\[
(1.8b) \quad \sum_{i=1}^{n} (x_i - \bar{x}_n)y_i^* (b) = 0,
\]

or more precisely, as a minimizer of \( \| \sum_{i=1}^{n} (x_i - \bar{x}_n)y_i^* (b) \| \), where \( \| a \| = \sqrt{a^T a} \). Once \( \hat{\beta} \) is defined, \( \hat{\alpha} \) is determined explicitly from (1.6) (with \( y_i^* (b) \) in place of \( y_i - b^T x_i \)) as \( \hat{\alpha} = n^{-1} \sum_{i=1}^{n} y_i^* (\hat{\beta}) \).

Obviously Miller’s approach can be extended to general loss functions \( \rho \) by simply replacing \( (\bar{y}_i - a - b^T x_i)^2 \) in (1.4) by \( \rho(\bar{y}_i - a - b^T x_i) \). The Buckley-James approach can also be extended to general score functions \( \psi \) without any difficulty. As a generalization of (1.8a), let

\[
(1.9a) \quad \psi_i^* (a, b) = \delta_i \psi(\bar{y}_i - a - b^T x_i) + (1 - \delta_i) \{ \int_{\bar{y}_i - b^T x_i}^{\infty} \psi(u - a) d\hat{F}_{n,b}(u) \} / \{ 1 - \hat{F}_{n,b}(\bar{y}_i - b^T x_i) \}.
\]

Define \( \hat{\alpha}, \hat{\beta} \) as a solution to the estimating equations

\[
(1.9b) \quad \sum_{i=1}^{n} (x_i - \bar{x}_n)\psi_i^* (a, b) = 0, \quad \sum_{i=1}^{n} \psi_i^* (a, b) = 0,
\]

\]

2
or more precisely, as a minimizer of \( \| \sum_{i=1}^{n} (x_i - x_i) \|_{\psi^*_{i}(a,b)} + \| \sum_{i=1}^{n} \psi^*_{i}(a,b) \| \). In the case \( \psi(u) = u \), \( \hat{\alpha} \) and \( \hat{\beta} \) reduce to the Buckley-James estimators, noting that \( \psi^*_{i}(a,b) = y^*_i(b) - a \) in this case.

It is well known that the approaches of Miller and of Buckley and James to generalize the classical least squares estimates to censored data lead to very different estimators, and empirical and simulation studies have shown the Buckley-James estimator of \( \beta \) to be superior to Miller's estimator, cf. Miller and Halpern (1982). Therefore the generalization of the Buckley-James approach in (1.9) appears to be a better way to construct \( M \)-estimators in censored regression models than the extension of Miller's approach described above. By an asymptotic analysis of both approaches, we shall show that this is indeed the case.

Tsui, Jewell and Wu (1988) proposed an analog of the Buckley-James approach to estimate \( \beta \) in the truncated regression model. In Section 2 we discuss a general heuristic principle for constructing \( M \)-estimators in left truncated and right censored regression models. This general principle will also be shown to be closely related to the missing information principle of Orchard and Woodbury (1972), the self-consistency concept introduced by Efron (1967), and the EM algorithm of Dempster, Laird and Rubin (1977).

In Section 3, we introduce a slight modification of the \( M \)-estimators constructed from this general heuristic principle to make them more tractable. A complete asymptotic theory of this class of \( M \)-estimators is then developed by extending the arguments introduced in Lai and Ying (1989a, 1990) for the analysis of the Buckley-James estimator of \( \beta \) in the censored regression model and the analogous Tsui-Jewell-Wu estimator in the truncated regression model. The asymptotic theory is applied in Section 4 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 based on censored and truncated data.

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 \((\tilde{y}_1, \delta_1), \ldots, (\tilde{y}_n, \delta_n)\) are observed for the estimation of \( \alpha \). Since the \( y_i \) are not completely observable, it is natural to replace the criterion (1.2) defining an \( M \)-estimator of \( \alpha \) by

\[
\sum_{i=1}^{n} E[\rho(y_i - a)|\tilde{y}_i, \delta_i] = \min
\]
Moreover, since the conditional expectation in (2.1) involves the unknown common distribution function $F$ the $y_i$, it is natural to replace $F$ by its nonparametric maximum likelihood estimator $\hat{F}_n$, which is the Kaplan-Meier (product-limit) estimator based on the censored data. This leads to the minimization problem

\begin{equation}
\int_{-\infty}^{\infty} \rho(z - a) d\hat{F}_n(z) = \sum_{i=1}^{n} \tilde{w}_i \rho(\tilde{y}_i - a) = \min!,
\end{equation}

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

\begin{equation}
\sup_{t < \tau} |\hat{F}_n(t) - F(t)| \xrightarrow{P} 0, \quad \text{where } \tau = \inf\{t : (1 - F(t))P[c_i > t] = 0\},
\end{equation}

cf. Wang(1987). Therefore under suitable regularity conditions on $\rho$,

\begin{equation}
\inf_a \int_{-\infty}^{\infty} \rho(z - a) d\hat{F}_n(z) \xrightarrow{P} \inf_a \int_{-\infty}^{\tau} \rho(z - a) dF(z) = \inf_a \int_{-\infty}^{\infty} \rho(z - a) dF(z),
\end{equation}

if $F(\tau) = 1$. Note that for the criterion (1.2) (with $b = \beta = 0$) when the $y_i$ are fully observable, we also have

\begin{equation}
\inf_a n^{-1} \sum_{i=1}^{n} \rho(y_i - a) \xrightarrow{P} \inf_a \int_{-\infty}^{\infty} \rho(z - a) dF(z).
\end{equation}

Without assuming $\beta$ 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 $\beta$. First note that the $\hat{F}_n$ in (2.2) is the same as $\hat{F}_{n,\beta}$, where as in (1.4), $\hat{F}_{n,\beta}$ is the Kaplan-Meier curve based on $(\tilde{y}_i - \beta^T x_i, \delta_i)_{i \leq n}$. Since $\beta$ in $\hat{F}_{n,\beta}$ is unknown, Miller's(1976) approach is to replace $\hat{F}_n(= \hat{F}_{n,\beta})$ in (2.2) by $\hat{F}_{n,b}$, leading to the minimization problem

\begin{equation}
\int_{-\infty}^{\infty} \rho(z - a) d\hat{F}_{n,b}(z) = \sum_{i=1}^{n} \tilde{w}_i (b) \rho(\tilde{y}_i - a - b^T x_i) = \min!,
\end{equation}

which reduces to (1.4) in the case $\rho(u) = u^2$. A difficulty with this approach is that unlike (2.3), we now have under certain regularity conditions that

\begin{equation}
\sup \{|\hat{F}_{n,b}(t) - \hat{F}_{n,b}(t)| : \|b\| \leq K, \sum_{i=1}^{n} I(\tilde{y}_i - b^T x_i \geq t) \geq n^{1-\epsilon}\} \rightarrow 0 \quad \text{a.s.},
\end{equation}

for every $0 < \epsilon < 1$ and $K > 0$, where

\begin{equation}
\hat{F}_{n,b}(t) = 1 - \exp\left(-\sum_{i=1}^{n} \int_{y_i \leq} dP\{\tilde{y}_i - b^T x_i \leq u, \gamma_i \leq c_i\}\right) / \sum_{j=1}^{n} P\{\tilde{y}_j - b^T x_j \geq u\},
\end{equation}

4
which may be very different from \( F(t) \) for \( b \neq \beta \), cf. Lemma 2 of Lai and Ying (1989). Therefore unlike (2.4), one now has under suitable assumptions on \( \rho \) that

\[
\inf_{a,b} \int_{-\infty}^{\infty} \rho(z - a) d\hat{F}_{n,b}(z) - \inf_{a,b} \int_{z \leq \tau} \rho(z - a) dF_{n,b}(z) \xrightarrow{P} 0.
\]

Since \( \inf_{a} \int_{z \leq \tau} \rho(z - a) dF_{n,b}(z) \) may be minimized at \( b \) far from \( \beta \), (2.9) explains the inconsistency and unsatisfactory performance of Miller’s estimates noted previously in empirical and simulation studies.

### 2.1. An EM algorithm in the censored regression model.

Instead of following Miller to modify the minimization problem (1.2) for censored data to define M-estimators, we can extend the Buckley-James (1979) idea and modify the estimating equations (1.3) for censored data as follows. Letting \( \psi = \rho' \), first modify (1.3) to the form

\[
\sum_{i=1}^{n} E[\psi(y_i - a - b^T x_i) | \bar{y}_i, \delta_i] = 0,
\]

and note that

\[
E[\psi(y_i - a - b^T x_i) | \bar{y}_i, \delta_i] = \delta_i \psi(y_i - a - b^T x_i) + \frac{\int_{\bar{y}_i - \beta^T x_i}^{\infty} \psi(u - a - (b - \beta)^T x_i) dF(u)}{1 - F(\bar{y}_i - \beta^T x_i)}.
\]

In view of (2.3), we shall replace the unknown infinite dimensional parameter \( F \) in (2.10b) by \( \hat{F}_{n,\beta} \).

The unrealizable estimating equations (2.10) (even with \( F \) replaced by \( \hat{F}_{n,\beta} \) in (2.10b)), suggest the following iterative scheme to estimate \( \alpha \) and \( \beta \):

**Step 1. Initialize \( \hat{\alpha}^{(1)} \) and \( \hat{\beta}^{(1)} \)**

**Step k.** Replace the unknown \( \beta \) and \( F \) in (2.10b) by \( \hat{\beta}^{(k-1)} \) and \( \hat{F}_{n,\hat{\beta}^{(k-1)}} \), leading to the approximation \( \hat{E}^{(k)}[\psi(y_i - a - b^T x_i) | \bar{y}_i, \delta_i] \) of (2.10b). Put this approximation into (2.10a) to determine an approximate solution \( \hat{\alpha}^{(k)} \), \( \hat{\beta}^{(k)} \) of the resultant estimating equation.

Iterating these steps until convergence to determine the estimates of \( \alpha \) and \( \beta \) is tantamount to defining these estimates by the estimating equations (1.9b). This iterative scheme also suggests that instead of the Miller-type extension (2.2) of the minimization criterion (1.2) to define M-estimators, a more appropriate extension of (1.2) to the censored regression model is the following analog of the EM algorithm (cf. Dempster, Laird and Rubin, 1977):

**kth E-step.** Replacing \( \psi \) by \( \rho \) in (2.10b), substitute the unknown \( \beta \) and \( F \) there by \( \hat{\beta}^{(k-1)} \) and \( \hat{F}_{n,\hat{\beta}^{(k-1)}} \). This gives \( \hat{E}^{(k)}[\rho(y_i - a - b^T x_i) | \bar{y}_i, \delta_i] \).
\(k^{th} M\)-step. Define \(\hat{\alpha}^{(k)}\) and \(\hat{\beta}^{(k)}\) as a minimizer of

\[
(2.11) \quad \sum_{i=1}^{n} E^{(k)}[\rho(y_i - a - b^T x_i)|\tilde{y}_i, \delta_i].
\]

When \(\rho\) is differentiable, the partial derivatives of (2.11) with respect to \(a\) and \(b\) are \(\sum_{i=1}^{n} E^{(k)}[\rho'(y_i - a - b^T x_i)|\tilde{y}_i, \delta_i]\) and \(\sum_{i=1}^{n} x_i E^{(k)}[\rho'(y_i - a - b^T x_i)|\tilde{y}_i, \delta_i]\), and therefore the \(k^{th}\) M-step reduces to the estimating equations (2.10a) with \(E\) replaced by \(E^{(k)}\) and with \(\psi = \rho'\).

2.2. A general missing information principle. We next describe a general heuristic principle to construct estimating equations for the regression parameters when the \(y_i\) in (1.1) are not completely observable. Let \(z\) denote the vector of observations and let \(y\) denote the associated unobservable complete data vector. Letting \(\gamma = (\alpha, \beta, F)\), suppose that an estimate of \(\gamma\) based on \(y\) is given by the estimating equation

\[
(2.12) \quad T(y; \gamma) = 0.
\]

For example, the M-estimators \(\hat{\alpha}, \hat{\beta}\) of \(\alpha\) and \(\beta\) in (1.3) and the empirical distribution function of \(y_i - \beta^T x_i\) correspond to (2.12) with

\[
(2.13) \quad T(y; \alpha, \beta, F) = \left( \sum_{i=1}^{n} \psi(y_i - \alpha - \beta^T x_i), \sum_{i=1}^{n} x_i \psi(y_i - \alpha - \beta^T x_i), F - n^{-1} \sum_{i=1}^{n} \Delta_{y_i - \beta^T x_i} \right),
\]

where \(\psi = \rho'\) and \(\Delta_{y_i - \beta^T x_i}(u) = I_{(y_i - \beta^T x_i \leq u)}\). Since we do not observe \(y\), we replace the unobservable estimating function \(T(y; \gamma)\) in (2.12) by \(E_{\gamma}[T(y; \gamma)|z]\), leading to the estimating equation

\[
(2.14) \quad E_{\gamma}[T(y; \gamma)|z] = 0.
\]

In applying this missing information principle to (2.13) when the \(y_i\) are subject to right censorship, first note that

\[
(2.15) \quad E_{\hat{F}_{n, \beta}} \left[ n^{-1} \sum_{i=1}^{n} \Delta_{y_i - \beta^T x_i}|\tilde{y}_1, \delta_1, x_1, \ldots, \tilde{y}_n, \delta_n, x_n \right] = \hat{F}_{n, \beta},
\]

which is the “self-consistency” property of the Kaplan-Meier estimator \(\hat{F}_{n, \beta}\), cf. Efron (1967) and Tsai and Crowley (1985). This implies that a solution of (2.13)–(2.14) in the censored regression model is of the form \((\hat{\alpha}, \hat{\beta}, \hat{F}_{n, \beta})\), where \(\alpha\) and \(\hat{\beta}\) are given by the estimating equation (1.9b). Hence this general principle includes the extension (1.9) of the Buckley-James estimator as a special case.
The heuristic principle (2.14) can be regarded as a generalization of the missing information principle of Orchard and Woodbury (1972) who considered the case in which \( T(y; \gamma) \) is the derivative of the log likelihood of \( \gamma = (\alpha, \beta) \), assuming a known density so that there is no infinite dimensional nuisance parameter \( F \). For such \( T \), Dempster, Laird and Rubin (1977) showed that under certain assumptions, their EM algorithm for solving iteratively the estimating equation \( E_{\alpha, \beta}[T(y; \gamma) | z] = 0 \) indeed converges to the maximum likelihood estimate of \( \alpha, \beta \) based on \( z \).

### 2.3. M-estimators in the left truncated and right censored regression model

We now apply the general heuristic principle to derive M-estimators of the regression parameters when the data are subject to both right censorship and left truncation. As in Turnbull (1976), we assume in this derivation that the \( t^*_i, c^*_i \) and \( x^*_i \) are nonrandom with \( c^*_i \geq t^*_i \) and regard the observed sample \( z = (\tilde{y}^*_i, t^*_i, \delta^*_i, x^*_i)_{1 \leq i \leq n} \) as having been generated from a larger sample \( y = (y_i, t_i, c_i, x_i)_{1 \leq i \leq m(n)} \), where we define

\[
\sigma_0 = 0, \quad \sigma_j = \inf\{i > \sigma_{j-1} : \tilde{y}_i \geq t^*_j\}, \quad m(n) = \sigma_n, \\
(t_i, c_i, x_i) = (t^*_i, c^*_i, x^*_i) \quad \text{for} \quad \sigma_{j-1} < i \leq \sigma_j.
\]

Let \( \gamma = (\alpha, \beta, F) \) as before. If \( y \) were observed, then we would be able to estimate \( \gamma \) by the estimating equation (2.12) with

\[
T(y; \gamma) = \left( \sum_{i=1}^{m(n)} \psi(y_i(\beta) - \alpha), \sum_{i=1}^{m(n)} x_i \psi(y_i(\beta) - \alpha), m(n)F - \sum_{i=1}^{m(n)} \Delta_y(\beta) \right).
\]

Here and in the sequel, we use the notation \( y_i(b) = y_i - b^T x_i \), \( \tilde{y}^*_i(b) = \tilde{y}^*_i - b^T x^*_i \) and \( t^*_i(b) = t^*_i - b^T x^*_i \). Note that

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

\[
E_F[\sum_{i=1}^{m(n)} \Delta_{y_i(\beta)}(u) | z] = E_F[\sum_{j=1}^{n} \sum_{\sigma_{j-1} < i \leq \sigma_j} I(y_i(\beta) \leq u) | z]
\]

\[
= \sum_{j=1}^{n} \delta^*_j I(\tilde{y}^*_j(\beta) \leq u) + \sum_{j=1}^{n} (1 - \delta^*_j) I(\tilde{y}^*_j(\beta) \leq u) \frac{F(u) - F(\tilde{y}^*_j(\beta))}{1 - F(\tilde{y}^*_j(\beta))} + \sum_{j=1}^{n} \frac{F(u \land t^*_j(\beta))}{1 - F(t^*_j(\beta))},
\]

\[
E_F[m(n) | z] = E_F[\sum_{i=1}^{n} (\sigma_j - \sigma_{j-1}) | z] = \sum_{i=1}^{n} \frac{1}{1 - F(t^*_j(\beta))}.
\]

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 \( \tilde{y}^*_i(b) \leq \cdots \leq \tilde{y}^*_k(b) \) denote all the ordered uncensored residuals. For \( i = 1, \ldots, k \), let

\[
Z_n(b, u) = \#\{j \leq n : t^*_j(b) \leq u \leq \tilde{y}^*_j(b)\},
\]
where we use the notation \( \# A \) to denote the number of elements of a set \( A \). Define

\[
\hat{F}_{n,b}(u) = 1 - \prod_{i: \hat{y}_i^* (b) \leq u, \delta_i = 1} \left( 1 - \frac{1}{\hat{Z}_n(b, \hat{y}_i^* (b))} \right),
\]

\[
H_n(b, u) = \sum_{j=1}^{n} \delta_j I_{(\hat{y}_j^* (b) \leq u)}, \quad K_n(b, u) = \sum_{j=1}^{n} (1 - \delta_j) I_{(\hat{y}_j^* (b) \leq u)},
\]

\[
L_n(b, u) = \sum_{j=1}^{n} I_{(\hat{y}_j^* (b) \leq u)}.
\]

In the case where \( \beta \) is known, \( \hat{F}_{n,\beta} \) is the nonparametric maximum likelihood estimator of \( F \) and by (2.20),

\[
E_{\hat{F}_{n,\beta}} [m(n) \mid z] = \int_{-\infty}^{\infty} (1 - \hat{F}_{n,\beta}(t))^{-1} dL_n(\beta, t) = \hat{m}_\beta, \quad \text{say}.
\]

Furthermore, \( \hat{F}_{n,\beta} \) satisfies the self-consistency property

\[
E_{\hat{F}_{n,\beta}} \left[ \sum_{i=1}^{m(n)} \Delta_{y_i(\beta)}(u) \mid z \right] = \hat{m}_\beta \hat{F}_{n,\beta}(u),
\]

cf. Turnbull (1976), noting in view of (2.19) that the self-consistency property (2.25) is equivalent to

\[
\hat{m}_\beta \hat{F}_{n,\beta}(u) = H_n(\beta, u) + \int_{-\infty}^{u} \frac{\hat{F}_{n,\beta}(u) - \hat{F}_{n,\beta}(t)}{1 - \hat{F}_{n,\beta}(t)} dK_n(\beta, t) + \int_{-\infty}^{\infty} \frac{\hat{F}_{n,\beta}(u \wedge t)}{1 - \hat{F}_{n,\beta}(t)} dL_n(\beta, t),
\]

which can also be expressed in the form

\[
\hat{m}_\beta \hat{F}_{n,\beta}(u) = \left\{ H_n(\beta, u) + \int_{-\infty}^{u} \left[ \int_{-\infty}^{s} \frac{dK_n(\beta, t)}{1 - \hat{F}_{n,\beta}(t)} + \int_{s}^{\infty} \frac{dL_n(\beta, t)}{1 - \hat{F}_{n,\beta}(t)} \right] d\hat{F}_{n,\beta}(s) \right\}.
\]

From (2.17), (2.24) and (2.25), it follows that a solution of the estimating equation (2.14) is of the form \((\alpha, \beta, \hat{F}_{n,\beta})\). Moreover, by (2.26),

\[
\hat{m}_\beta \int_{-\infty}^{\infty} \psi(u-a) d\hat{F}_{n,\beta}(u) = \int_{-\infty}^{\infty} \psi(u-a) dH_n(\beta, u)
\]

\[
+ \int_{-\infty}^{\infty} \psi(u-a) \left[ \int_{-\infty}^{u} \frac{dK_n(\beta, t)}{1 - \hat{F}_{n,\beta}(t)} + \int_{u}^{\infty} \frac{dL_n(\beta, t)}{1 - \hat{F}_{n,\beta}(t)} \right] d\hat{F}_{n,\beta}(u).
\]

In view of (2.14), (2.18), (2.25) and (2.27), \( \hat{\alpha} \) and \( \hat{\beta} \) are given by the estimating equations

\[
\sum_{i=1}^{n} \psi_i^*(a, b) = 0,
\]

\[
\sum_{i=1}^{n} x_i^o \psi_i^*(a, b) = 0,
\]

8
or more precisely, \((\hat{\alpha}, \hat{\beta})\) is defined as a minimizer of \(\sum_{i=1}^{n} \psi_i^*(a, b) + \| \sum_{i=1}^{n} x_i \psi_i^*(a, b) \|\), where

\[
\psi_i^*(a, b) = \delta_i \psi(y_i^*(b) - a) + (1 - \delta_i) \frac{\int_{\tilde{F}_i^*(b)}^{\infty} \psi(u - a) d\hat{F}_{n,b}(u)}{1 - \hat{F}_{n,b}(y_i^*(b))} + \frac{\int_{\hat{t}_i^*(b)}^{\infty} \psi(u - a) d\hat{F}_{n,b}(u)}{1 - \hat{F}_{n,b}(t_i^*(b))}.
\]

For the censored regression model \(t_i = -\infty\), the estimating equations (2.28) are equivalent to the Buckley-James-type estimating equations (1.9). For the truncated regression model \(c_i = \infty\), the estimating equation of Tsui, Jewell and Wu (1988) for the slope parameter \(\beta\) is closely related to the second equation of (2.28) with \(\psi(u) = u\).

Note that the first equation in (2.28) is equivalent to \(\int_{-\infty}^{\infty} \psi(u - a) d\hat{F}_{n,b}(u) = 0\), which in turn implies that

\[
\int_{-\infty}^{t_i^*(b)} \psi(u - a) d\hat{F}_{n,b}(u) = -\int_{t_i^*(b)}^{\infty} \psi(u - a) d\hat{F}_{n,b}(u).
\]

Hence we can express (2.29) alternatively as

\[
\psi_i^*(b; a) = \delta_i \psi(y_i^*(b) - a) + (1 - \delta_i) \frac{\int_{\tilde{F}_i^*(b)}^{\infty} \psi(u - a) d\hat{F}_{n,b}(u)}{1 - \hat{F}_{n,b}(y_i^*(b))} - \frac{\int_{t_i^*(b)}^{\infty} \psi(u - a) d\hat{F}_{n,b}(u)}{1 - \hat{F}_{n,b}(t_i^*(b))} - \psi(t_i^*(b) - a)
\]

assuming \(\psi\) to be continuous. This alternative form of \(\phi_i^*\) will be used in the sequel.

3. Consistency and asymptotic normality of a class of M-estimators in the left truncated and right censored regression model.

Throughout the sequel we shall assume that

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

Because of the instability of the product-limit estimator (2.22) at points \(u\) for which the "risk set size" \(Z_n(b, u)\) is small compared with \(n\), some modification of (2.30) is needed to make the associated M-estimator more tractable. We shall use ideas similar to those introduced in Lai and Ying (1989, 1990) for the special case \(\psi(u) = u\) to modify (2.30) and to prove consistency and asymptotic normality of the corresponding class of M-estimators. Letting

\[
Z_n^*(b, u) = \sum_{i=1}^{n} x_i \mathbb{I}_{(t_i^*(b) \leq u \leq \tilde{y}_i^*(b))},
\]

\[
J_n^*(b, u) = \sum_{i=1}^{n} x_i \{(1 - \delta_i) \mathbb{I}_{\tilde{y}_i^*(b) \geq u} + \mathbb{I}_{t_i^*(b) \leq u}\},
\]

(3.2)
it will be convenient to express the second equation in (2.28) coupled with (2.30) as

\[
(3.3) \quad - \int_{u=-\infty}^{\infty} \psi(u-a)dZ_n^x(b,u) - \int_{u=-\infty}^{\infty} \{ \int_{u}^{\infty} (1 - \hat{F}_{n,b}(v|u))\psi'(v-a)dv \}dJ_n^x(b,u) = 0,
\]

where

\[
(3.4) \quad 1 - \hat{F}_{n,b}(v|u) = \frac{1 - \hat{F}_{n,b}(v)}{1 - \hat{F}_{n,b}(u)} = \prod_{i: u < \delta_i^p(b) \leq v, \delta_i^p = 1} (1 - \frac{1}{Z_n(b, \delta_i^p(b))}) \quad \text{for} \ v \geq u.
\]

Note that for $\psi(u) = u$ and $a = 0$ in the censored case with $t_i \equiv -\infty$, (3.3) reduces to the estimating equation (2.7) of Lai and Ying (1989) defining the modified Buckley-James slope estimator, which multiplies $Z_n^x(b, \cdot)$ and $1 - \hat{F}_{n,b}(\cdot|u)$ by a smooth weight function $p_n(\cdot)$ to remove the instability in $\hat{F}_{n,b}(\cdot|u)$ when the risk set size $Z_n(b, \cdot)$ is small compared with $n$.

3.1. Smoothing kernels to dampen the instability due to small risk set sizes. In view of its definition in (2.21), the risk set size $Z_n(b, u)$ is small if either $L_n(b, u)$ or $N_n(b, u)$ is small, where

\[
(3.5) \quad L_n(b, u) = \sum_{i=1}^{n} I\{t_i^p(b) \leq u\}, \quad N_n(b, u) = \sum_{i=1}^{n} I\{\delta_i^p(b) \geq u\}.
\]

We shall use two smoothing kernels $p_{n1}$, $p_{n2}$ 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

\[
(3.6a) \quad p(u) = 0 \ \text{for} \ u \leq 0, \quad p(u) = 1 \ \text{for} \ u \geq 1.
\]

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

\[
(3.6b) \quad p_{n1}(b, u) = p\left(\frac{L_n(b, u)}{n} - \frac{c}{\log n} \log n\right), \quad p_{n2}(b, u) = p\left(\frac{N_n(b, u)}{n} - \frac{d}{\log n} \log n\right).
\]

Define

\[
(3.7) \quad Z_{n,p}^x(b, u) = \sum_{i=1}^{n} x_i^p p_{n1}(b, t_i^p(b)) p_{n2}(b, u) I\{t_i^p(b) \leq u \leq \delta_i^p(b)\},
\]

which is a step function with jumps at $t_i^p(b), \delta_i^p(b)$ ($i = 1, \cdots, n$) and is a modified version of $Z_n^x$ in (3.2). Moreover, define the step function

\[
(3.8) \quad J_{n,p}^x(b, u) = \sum_{i=1}^{n} x_i^p p_{n1}(b, t_i^p(b)) \{(1 - \delta_i^p)I\{\delta_i^p(b) \geq u\} + I\{t_i^p(b) \leq u\}\}
\]
as a modification of $J_1^r$ in (3.2). In analogy with the left hand side of (3.3), define

$$(3.9) \quad \hat{\xi}_n(a,b) = - \int_{u=-\infty}^{\infty} \psi(u-a) dZ_{n,p}^x(b,u) - \int_{u=-\infty}^{\infty} \left\{ \int_{u}^{\infty} (1 - \hat{F}_{n,b}(v|u)) \psi'(v-a) p_{n,2}(b,v) dv \right\} dJ_{n,p}^x(b,u).$$

We propose to replace (3.3) by the equation $\hat{\xi}_n(a,b) = 0$, which is completely analogous to the modified Buckley-James/Tsui-Jewel-Wu estimator of $\beta$ introduced by Lai and Ying (1989/1990) for $\psi(u) = u$ and $a = 0$ in the censored/truncated regression model. In this connection, Lemma 1 of Lai and Ying (1989) on how to introduce smoothing kernels into the Buckley-James equation without producing bias (at least when $b = \beta$) can be extended to the following lemma, whose proof is given in the Appendix.

**LEMMA 1.** Let $g_1$ be a bounded function and $g_2$ be a function of bounded variation on the real line.

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

$$(3.10) \quad E \left\{ - \int_{-\infty}^{\infty} \psi(u-a) g_1(t_i(\beta)) d [g_2(u) I_{t_i(\beta) \leq u \leq c_i(\beta)}] - \int_{-\infty}^{\infty} \left[ \int_{u}^{\infty} (1 - F(v|u)) \times \psi'(v-a) g_2(v) dv \right] g_1(t_i(\beta)) d [I_{t_i(\beta) \leq c_i(\beta) \leq u, c_i(\beta) \leq c_i(\beta)] + I_{t_i(\beta) \leq c_i(\beta) \leq c_i(\beta)}] \right\} = 0,$$

where

$$(3.11) \quad F(v|u) = \frac{(F(v) - F(u))}{(1 - F(u))} \quad \text{for} \quad v \geq u.$$

(ii) If $c_i \geq t_i$, then (3.10) still holds with $\epsilon_i$ replaced by $\epsilon_i^*$ whose conditional distribution given $(t_i, c_i, x_i)$ is

$$(3.12) \quad P\{\epsilon_i^* \leq u \mid t_i, c_i, x_i\} = \{F(u) - F(t_i(\beta))\}/\{1 - F(t_i(\beta))\}, \quad u \geq t_i(\beta).$$

3.2. 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 $\{\epsilon_n\}$ and such that either

$$(3.13a) \quad \sup_n E(|t_n|^\delta + (c_n^-)^\delta) < \infty \quad \text{for some} \quad \delta > 0, \quad \text{where} \quad a^- \text{ denotes the negative part of} \quad a \quad \text{(i.e.,} \quad a^- = |a| I_{a \leq 0}),$$

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

$$(3.13b) \quad t_i \equiv -\infty \quad \text{and} \quad E \exp(\theta \epsilon_1^-) + \sup_n E \exp(\theta c_n^-) < \infty \quad \text{for some} \quad \theta > 0.$$
Letting \( t_i(b) = t_i - b^T x_i \) and \( c_i(b) = c_i - b^T x_i \), suppose that

\begin{equation}
\|x_i\| \leq K \quad \text{for all } i \text{ and some nonrandom constant } K;
\end{equation}

\begin{equation}
\sup_{\|b\| \leq \rho, -\infty < u < \infty} \sum_{i=1}^{m} \left[ P\{u \leq t_i(b) \leq u + h\} + P\{u \leq c_i(b) \leq u + h\} \right] = O(mh)
\end{equation}
as \( h \to 0 \) and \( m \to \infty \) with \( mh \to \infty \);

\begin{equation}
F \text{ has a twice continuously differentiable density } f \text{ such that}
\int_{-\infty}^{\infty} (f'/f)^2 dF < \infty \quad \text{and} \quad \int_{-\infty}^{\infty} \sup_{|h| \leq \eta} \{|f'(t + h)| + |f''(t + h)|\} dt < \infty \quad \text{for some } \eta > 0;
\end{equation}

\begin{equation}
m^{-1} \sum_{i=1}^{m} P\{t_i(\beta) \leq s \leq c_i(\beta)\} \to \Gamma_0(s), m^{-1} \sum_{i=1}^{m} E\{x_i I_{\{t_i(\beta) \leq s \leq c_i(\beta)\}}\} \to \Gamma_1(s),
\end{equation}

\begin{equation}
m^{-1} \sum_{i=1}^{m} E\{x_i x_i^T I_{\{t_i(\beta) \leq s \leq c_i(\beta)\}}\} \to \Gamma_2(s) \quad \text{for } s < F^{-1}(1).
\end{equation}

The assumptions (3.13)–(3.17) are essentially the same as those made by Lai and Ying (1989, 1990) in their analysis of the modified Buckley-James and Tsui-Jewel-Wu estimators of \( \beta \).

A basic idea in the subsequent analysis is to regard the observed sample \((t^o_i, c^o_i, x^o_i, \bar{y}^o_i), i = 1, \ldots, n\), of left truncated and right censored observations as having been generated by a larger, randomly stopped sample of independent random vectors \((t_i, c_i, x_i^T, y_i), i = 1, \ldots, m(n)\), where

\begin{equation}
m(n) = \inf\{m : \sum_{i=1}^{m} I_{\{t_i \leq y_i \land c_i\}} = n\}.
\end{equation}

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

\begin{equation}
m^{-1} \sum_{i=1}^{m} I_{\{t_i \leq y_i \land c_i\}} - m^{-1} \sum_{i=1}^{m} P\{t_i \leq y_i \land c_i\} \to 0 \quad \text{a.s.}
\end{equation}

Hence, under (3.17) and the assumption

\begin{equation}
\tau_0 := \inf\{s : \Gamma_0(s) > 0\} < \tau := \inf\{s > \tau_0 : (1 - F(s))\Gamma_0(s) = 0\} \quad \text{and}
\end{equation}

\begin{equation}
\lim_{n \to \infty} m^{-1} \sum_{i=1}^{m} P\{t_i(\beta) \leq c_i(\beta) < s\} = G(s) \quad \text{exists for every } s \in (\tau_0, \tau),
\end{equation}

12
it follows from (3.18) and (3.19) that

\[(3.21)\quad m(n)/n \to \Delta \quad \text{a.s., where} \quad 1/\Delta = \int_{-\infty}^{\infty} \{G(s) + G(s)\} dF(s) = \lim_{m \to \infty} m^{-1} \sum_{i=1}^{m} P\{t_i \leq y_i \wedge c_i\}.\]

Another important idea in the subsequent analysis is to approximate \(\hat{\xi}_n(a, b)\) by certain nonrandom functions. 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 (1989).

**LEMMA 2.** Under the assumptions (3.1) and (3.13) – (3.16), let

\[Z_m(b, u) = \sum_{i=1}^{m} P\{t_i \leq u \leq \bar{y}_i(b)\}, \quad L_m(b, u) = \sum_{i=1}^{m} P\{t_i \leq \bar{y}_i(b) \wedge u\},\]

\[\bar{N}_m(b, u) = \sum_{i=1}^{m} P\{\bar{y}_i(b) \geq u \vee t_i(b)\}, \quad \bar{n}(m) = \sum_{i=1}^{m} P\{t_i \leq \bar{y}_i\},\]

\[\bar{p}_{m,1}(b, u) = p\left(\frac{L_m(b, u)}{\bar{n}(m)} - \frac{c}{\log \bar{n}(m)}\right) \log \bar{n}(m),\]

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

\[Z_{m,p}^x(b, u) = \sum_{i=1}^{m} E\{x_i \bar{p}_{m,1}(b, t_i(b)) \bar{p}_{m,2}(b, u) I\{t_i(b) \leq u \leq \bar{y}_i(b)\}\},\]

\[J_{m,p}^x(b, u) = \sum_{i=1}^{m} E\{x_i \bar{p}_{m,1}(b, t_i(b))(-I\{c_i < y_i, t_i(b) \leq \bar{y}_i(b) \leq u\} + I\{t_i(b) \leq \bar{y}_i(b) \wedge u\}\}.$

Letting \(\bar{H}_m(b, u) = \sum_{i=1}^{m} P\{y_i \leq c_i, t_i(b) \leq \bar{y}_i(b) \leq u\},\) define for \(v \geq u\)

\[(3.22)\quad F_{m,b}(v|u) = 1 - \exp\{-\int_{u < s \leq v} \frac{d\bar{H}_m(b, s)}{Z_m(b, s)}\},\]

\[(3.23)\quad \xi_m(a, b) = -\int_{-\infty}^{\infty} \psi(u - a) dZ_{m,p}^x(b, u)\]

\[-\int_{-\infty}^{\infty} \int_{u}^{\infty} \{1 - F_{m,b}(v|u))\psi'(v - a) \bar{p}_{m,2}(b, v)dv\} dJ_{m,p}^x(b, u).\]

Suppose that \(\lim_{m \to \infty} \bar{n}(m)/m\) exists and is positive. Then for every \(\epsilon > 0\) and \(0 < \gamma < 1\) and for any positive numbers \(A\) and \(B\),

\[(3.24)\quad \sup_{|a| \leq A, \|b\| \leq B} ||\hat{\xi}_n(a, b) - \xi_m(n)(a, b)|| = o(n^{1/2+\epsilon}) \quad \text{a.s.,}\]

\[(3.25)\quad \sup \{||\hat{\xi}_n(a, b) - \xi_n(a', b') - \xi_m(n)(a, b) + \xi_m(n)(a', b')||: \]

\[|a - a'| + \|b - b'\| \leq n^{-\gamma}, |a| \vee |a'| \leq A, \|b\| \vee ||b'|| \leq B\} = o(n^{(1-\gamma)/2+\epsilon}) \quad \text{a.s.}\]

13
Since \( F_{m,\beta}(v|u) = F(v|u) \) by (3.22), putting \( b = \beta \) in (3.23) and applying Lemma 1 yield

\[
\xi_m(a, \beta) = 0 \quad \text{for every } a.
\]

This suggests that we can choose any \( a \) and estimate \( \beta \) 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 \( \{ b : \|b\| \leq B \} \). More precisely, an estimator \( \hat{\beta}_n(a) \) of \( \beta \) will be defined as a minimizer of \( \|\hat{\xi}_n(a, b)\| \) over the region \( \{ b : \|b\| \leq B \} \). The following theorem, which is an extension of Theorems 1 and 3 of Lai and Ying (1989) on the modified Buckley-James estimator, establishes the asymptotic normality of \( \hat{\beta}_n(a) \) via the asymptotic linearity of \( \hat{\xi}_n(a, b) \) in the neighborhood \( \{ b : \|b\| \leq n^{-\epsilon} \} \) of \( \beta \) and the asymptotic normality of \( n^{-1/2} \hat{\xi}_n(a, \beta) \).

**THEOREM 1.** Under the assumptions (3.1), (3.13) – (3.17) and (3.20), define \( \hat{\xi}_n(a, b) \) by (3.9) and \( \xi_m(a, b) \) by (3.23). For any fixed \( a \), let \( \hat{\beta}_n(a) \) be a minimizer of \( \|\hat{\xi}_n(a, b)\| \) over \( \{ b : \|b\| \leq B \} \).

(i) If \( \liminf_{m \to \infty} m^{-1} \{ \inf_{\|b\| \leq B, \|b-\beta\| \geq \epsilon} \|\xi_m(a, b)\| \} > 0 \) a.s. for every \( \delta > 0 \), then \( \hat{\beta}_n(a) \to \beta \) a.s.

(ii) Defining \( \tau_0 \) and \( \tau \) as in (3.20), assume that

\[
\lim_{m \to \infty} \frac{\log m}{m} \sum_{i=1}^{m} [ P\{ t_i(\beta) < \tau_0 - \epsilon \} I(F(\tau_0) > 0) + P\{ c_i(\beta) > \tau + \epsilon \} I(F(\tau) < 1) ] = 0
\]

for every \( \epsilon > 0 \). Define the \( \nu \times \nu \) matrix

\[
C_a = \Delta \int_{\tau_0}^{\tau} \left\{ \Gamma_2(u) - \frac{\Gamma_1(u)\Gamma_1^T(u)}{\Gamma_0(u)} \right\} \left\{ \int_u^{\tau} (1 - F(s|u))\psi'(s - a)ds \right\} \left\{ \frac{f'(u)}{f(u)} + \frac{f(u)}{1 - F(u)} \right\} dF(u),
\]

where \( \Delta \) is defined in (3.21) and \( 1 - F(s|u) = (1 - F(s)) / (1 - F(u)) \) for \( s \geq u \), as in (3.11). Then with probability 1, for every \( \epsilon > 0 \),

\[
\hat{\xi}_n(a, b) = \hat{\xi}_n(a, \beta) - nC_a(b - \beta) + o(n^{1/2} \nu n\|b - \beta\|) \quad \text{uniformly in } \|b - \beta\| \leq n^{-\epsilon}.
\]

Moreover, as \( n \to \infty \), \( n^{-1/2} \hat{\xi}_n(a, \beta) \) has a limiting normal distribution with mean 0 and covariance matrix

\[
V_a = \Delta \int_{\tau_0}^{\tau} \left\{ \Gamma_2(u) - \frac{\Gamma_1(u)\Gamma_1^T(u)}{\Gamma_0(u)} \right\} \left\{ \int_u^{\tau} (1 - F(s|u))\psi'(s - a)ds \right\}^2 dF(u).
\]

(iii) Suppose that (3.27) holds, that (3.28) is nonsingular, and that there exists \( \epsilon \in (0, 1/2) \) for which

\[
\lim_{m \to \infty} m^{-(1/2+\epsilon)} \inf_{\|b\| \leq B, \|b-\beta\| \geq n^{-\epsilon}} \|\xi_m(a, b)\| = \infty.
\]
Then $\hat{\beta}_n(a) = \beta + O(n^{-r})$ a.s. and $\sqrt{n}(\hat{\beta}(a) - \beta)$ has a limiting normal distribution with mean 0 and covariance matrix $C_{\alpha}^{-1}V_{\alpha}C_{\alpha}^{-1}$.

Part (i) of the theorem is an immediate consequence of (3.24) and (3.26). Part (ii) will be proved in the Appendix. To prove part (iii), first note that (3.31), (3.24) and (3.21) imply that $n^{-1/2-\epsilon}\inf_{\|b\| \leq B, \|b-\beta\| \geq n^{-\epsilon}} \|\hat{\xi}_n(a,b)\| \to \infty$ a.s., and therefore the desired conclusion follows from part (ii) of the theorem.

3.3. Asymptotic normality of $M$-estimators of $\alpha$ and $\beta$ when $\psi$ has compact support. In the classical theory of $M$-estimators based on complete data, the location parameter $\alpha$ 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 $(\tau_0, \tau)$ even if $\beta$ 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 $\alpha$ in (1.1) will be chosen as a solution of the equation

$$\int_{\tau_0}^{\tau} \psi(u-a) dF(u) = 0. \tag{3.32}$$

Note that the left hand side of (3.32) is a continuous function of $\alpha$ under (3.1), and that (3.32) has a unique solution if $\psi$ is strictly increasing (or strictly decreasing) and $\lim_{a \to \infty} \psi(a), \lim_{a \to -\infty} \psi(a)$ have opposite signs.

We shall assume that (3.32) has a unique solution $\alpha$ in the interior of some given interval $[A_0, A_1]$ and that

$$\psi \text{ has support } [s_0, s_1] \text{ with } s_0 + A_0 > \tau_0, s_1 + A_1 < \tau. \tag{3.33}$$

Under the assumption (3.33), we can rewrite (3.32) as $\int_{\tau_0}^{\tau} \psi(u-a) dF(u|s_0+A_0) = 0$, where $F(\cdot | \cdot)$ is defined in (3.11). Defining $\hat{F}_{n,b}(\cdot | \cdot)$ as in (3.4), let

$$\hat{\xi}_n(a,b) = n \int_{s_0+A_0}^{s_1+A_1} \psi(u-a) d\hat{F}_{n,b}(u|s_0+A_0). \tag{3.34}$$

To estimate $\alpha$ and $\beta$, we use the following modifications of the estimating equations (2.28) (in which the first equation can be expressed in the form $n \int_{-\infty}^{\infty} \psi(u-a) d\hat{F}_{n,b}(u) = 0)$:

$$\hat{\nu}_n(a,b) = 0, \quad \hat{\xi}_n(a,b) = 0. \tag{3.35}$$

15
More precisely, \((\hat{\alpha}_n, \hat{\beta}_n)\) is defined as a minimizer of \(|\hat{\zeta}_n(a, b)| + ||\hat{\zeta}_n(a, b)||\) over the region \(\{A_0 \leq a \leq A_1, ||b|| \leq B\}\). Asymptotic properties of \((\hat{\alpha}_n, \hat{\beta}_n)\) are given in the following.

**THEOREM 2.** Under the assumptions (3.1), (3.13) – (3.17), (3.20), (3.27) and (3.33), suppose that (3.32) holds for some \(\alpha \in (A_0, A_1)\). Assume furthermore that

\[
\int_{\tau_0}^{\tau} \psi'(u - \alpha) dF(u) \neq 0, \quad \text{and} \quad \int_{\tau_0}^{\tau} \psi(u - a) dF(u) \neq 0 \quad \text{if} \quad a \neq \alpha \quad \text{and} \quad A_0 \leq a \leq A_1.
\]

Suppose that the matrix \(C_\alpha\) defined in (3.28) is nonsingular, and that there exists \(e \in (0, 1/2)\) such that

\[
\lim_{m \to \infty} m^{-\left(\frac{1}{2} + c\right)} \inf_{\Lambda_0 \leq a \leq A_1, ||b|| \leq B, ||b - \beta|| \geq m^{-e}} ||\xi_m(a, b)|| = \infty,
\]

where \(\xi_m(a, b)\) is defined in (3.23). Then \(|\hat{\alpha}_n - \alpha| + ||\hat{\beta}_n - \beta|| = O(n^{-e})\) a.s. and \(\sqrt{n}(\hat{\alpha}_n - \alpha, \hat{\beta}_n - \beta)^T\) has a limiting normal distribution with mean 0 and covariance matrix

\[
\left(\begin{array}{cc}
\psi_\alpha^{-2}(\nu_1 + g_\alpha C_\alpha^{-1} V_\alpha C_\alpha^{-1} g_\alpha) & \psi_\alpha^{-1} g_\alpha C_\alpha^{-1} V_\alpha C_\alpha^{-1} g_\alpha \\
\psi_\alpha^{-1} g_\alpha C_\alpha^{-1} V_\alpha C_\alpha^{-1} g_\alpha & C_\alpha^{-1} V_\alpha C_\alpha^{-1}
\end{array}\right),
\]

where \(V_\alpha\) is defined in (3.30) and

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

\[
\nu_\alpha = \frac{1}{\Delta} \int_{\tau_0}^{\tau} \int_{\tau_0}^{\tau} (1 - F(u))(1 - F(v)) \psi'(u - \alpha) \psi'(v - \alpha) \frac{dF(u) dF(v)}{\Gamma_0(s)(1 - F(s))^2} du dv,
\]

\[
g_\alpha = \int_{\tau_0}^{\tau} \psi(u - \alpha) \left\{ (1 - F(u)) \int_{\tau_0}^{u} \frac{\Gamma_1(s)}{\Gamma_0(s)} (f'(s) + \frac{f(s)}{1 - F(s)}) \frac{dF(s)}{1 - F(s)} \right\}.
\]

The proof of Theorem 2 will be given in the Appendix, where the arguments used to prove Theorem 1(ii) will be extended to establish similar asymptotic linearity results for \((\hat{\zeta}_n(a, b), \hat{\xi}_n^T(a, b))^T\) and the asymptotic normality of \((\hat{\zeta}_n(\alpha, \beta), \hat{\xi}_n^T(\alpha, \beta))^T\), from which the conclusions of Theorem 2 follow.

### 3.3. Asymptotic theory of M-estimators under independent \((t^*_i, c^*_i, x^*_i)\).

The \((t_i, c_i, x_i^T)\) defined in the truncation-censorship model (2.16) are clearly not independent random vectors. This leads to the alternative setting in which \((t_i^o, c_i^o, x_i^o^T)\) are independent random vectors with \(c_i^o \geq t_i^o > -\infty\) and are independent of \(\{\epsilon_n\}\) and in which

\[
\tilde{g}_i^o = (\sigma_i + \beta^T x_i^o) \wedge c_i^o, \quad \text{where} \quad \sigma_i = \inf\{n > \sigma_{i-1} : \epsilon_n \geq t_i^o - \beta^T x_i^o\},
\]

as in (2.16). Let \(\epsilon_i^* = \epsilon_{\sigma_i}\). Then \((t_i^o, c_i^o, x_i^o^T, \epsilon_i^*)\) are independent random vectors and

\[
P\{\epsilon_i^* \leq u|t_i^o, c_i^o, x_i^o\} = \{F(u) - F(t_i^o(\beta))\}/\{1 - F(t_i^o(\beta))\}, \quad u \geq t_i^o(\beta),
\]

16
as in (3.12). Hence Lemma 1(ii) is applicable.

Replace \( t_i(b) \) by \( t_i^\circ(b) \), \( c_i(b) \) by \( c_i^\circ(b) \), and \( (t_i, c_i, x_i) \) by \( (t_i^\circ, c_i^\circ, x_i^\circ) \) in the assumptions (3.13a), (3.14)-(3.16) and (3.27). Note that (3.13b) is no longer relevant here since it is assumed that \( t_i^\circ > -\infty \). Moreover, replace the assumptions (3.17) and (3.20) by

\[
\begin{align*}
\sup_i F(t_i^\circ(\beta)) < 1 \quad \text{a.s.,} \quad \lim_{n \to \infty} n^{-1} \sum_{i=1}^{n} E\{I_{t_i^\circ(\beta) \leq s \leq c_i^\circ(\beta)} / (1 - F(t_i^\circ(\beta))) \} &= \Gamma_0(s), \\
\lim_{n \to \infty} n^{-1} \sum_{i=1}^{n} E\{x_i x_i^T I_{t_i^\circ(\beta) \leq s \leq c_i^\circ(\beta)} / (1 - F(t_i^\circ(\beta))) \} &= \Gamma_1(s), \\
\lim_{n \to \infty} n^{-1} \sum_{i=1}^{n} E\{x_i x_i^T I_{t_i^\circ(\beta) \leq s \leq c_i^\circ(\beta)} / (1 - F(t_i^\circ(\beta))) \} &= \Gamma_2(s) \quad \text{for} \quad s < F^{-1}(1).
\end{align*}
\]

Replacing \( t_i(b) \) by \( t_i^\circ(b) \), \( \hat{y}_i(b) \) by \( \hat{y}_i^\circ(b) \), \( (t_i, \hat{y}_i) \) by \( (t_i^\circ, \hat{y}_i^\circ) \) in the various quantities defined in Lemma 2, note that in particular \( n(\cdot) = \sum_{i=1}^{m} P\{t_i^\circ \leq \hat{y}_i^\circ\} = m \) and that (3.24) and (3.25) with these modifications and with \( m(n) \) replaced by \( n \) still hold in the present setting, again by a straightforward modification of the proof of Lemma 2 of Lai and Ying (1989) that uses general tightness results for stochastic integrals of empirical-type processes established in Lai and Ying (1988). Define \( \tau_0 \) and \( \tau \) by (3.20) and set \( \Delta = 1 \). Then Theorems 1 and 2 with these modifications still hold in the present setting of independent \( (t_i^\circ, c_i^\circ, x_i^\circ) \) with \( c_i^\circ \geq t_i^\circ > -\infty \) and with \( \hat{y}_i^\circ \) given by (3.39), as can be shown by an obvious modification of their proofs given in the Appendix and by applying Lemma 1 of Lai and Ying (1991c) together with (3.40).

3.5. **Comparison with rank estimators and remarks on Ritov's related work.** Lai and Ying (1991b) recently developed a class of rank estimators, generated by smooth score functions \( w \), of the slope parameter \( \beta \) in the left truncated and right censored regression model, and established the asymptotic normality of these rank estimators under certain regularity conditions. In the case \( F(\tau) = 1 \), the limiting normal distribution of the \( M \)-estimator \( \hat{\beta}_n(\cdot) \) in Theorem 1 turns out to be the same as that of the rank estimator generated by the score function

\[
w(u) = \psi(u - a) - (1 - F(u))^{-1} \int_{u}^{\infty} \psi(s - a) dF(s) = - \int_{u}^{\tau} (1 - F(s|u)) \psi'(s - u) ds,
\]

cf. Corollary 1 of Lai and Ying (1991b). This shows that the well known asymptotic equivalence between \( M \)-estimators and rank estimators in the classical setting of completely observable \( y_i \) in (1.1) extends to the present framework in which the \( y_i \) are subject to both left truncation and right censoring.

When the \( y_i \) are subject only to right censoring (i.e., \( t_i \equiv -\infty \)) and \( (c_i, x_i^T) \) are i.i.d. random vectors, Ritov (1990) recently used the first equation of (1.9b) with \( a = 0 \) to define "Buckley-James-type" estimators associated with a given score function \( \psi \), which he assumed to satisfy the following
conditions for some prescribed positive constants \( K \) and \( \delta \):

\[
(3.42a) \quad \psi(t) = \psi(t \wedge K) \quad \text{for all} \quad t,
\]

\[
(3.42b) \quad P\{y_1 \wedge c_1 - b^T x_1 < K\} < 1 \quad \text{for} \quad ||b - \beta|| < \delta.
\]

Let \( \Psi_n(b) = n^{-1/2} \sum_{i=1}^{n} (x_i - \bar{x}_n) \psi_i^*(0, b) \), where \( \psi_i^* \) is defined by (1.9a). Under several additional assumptions, he showed that for every sequence of nonrandom constants \( \beta_n \) such that \( \beta_n = \beta + O(n^{-1/2}) \),

\[
(3.43) \quad \Psi_n(\beta_n) = S_n(\beta_n) + o_p(1), \quad \text{where}
\]

\[
S_n(b) = n^{-1/2} \sum_{i=1}^{n} \delta_i w(\tilde{y}_i(b)) \left\{ x_i - \frac{\sum_{j=1}^{n} x_j I\{\tilde{y}_j(b) \geq \tilde{y}_i(b)\}}{\sum_{j=1}^{n} I\{\tilde{y}_j(b) \geq \tilde{y}_i(b)\}} \right\},
\]

\[
w(u) = \psi(u) - (1 - F(u))^{-1} \int_{u}^\infty \psi(t) dF(t).
\]

In particular, for \( \beta_n \equiv \beta \), he obtained the limiting normal distribution of \( \Psi_n(\beta) \) from (3.43) and the limiting normal distribution of the censored rank statistics \( S_n(\beta) \), cf. Gill (1980). Assuming the existence of a preliminary \( \sqrt{n} \)-consistent estimator and defining a "Buckley-James-type" estimator of \( \beta \) as a solution of

\[
(3.44) \quad \Psi_n(b) = o_p(1) \quad \text{with} \quad ||b - \beta|| \leq n^{-1/2 + \epsilon},
\]

for some \( 0 < \epsilon < 1/2 \), he made use of (3.43) and the asymptotic linearity of the censored rank statistics \( S_n(b) \) in the region \( \{b : ||b - \beta|| \leq n^{-1/2 + \epsilon}\} \) to conclude that the Buckley-James-type estimator is asymptotically normal.

Ritov's (1990) proof of (3.43) is based on a simple application of Lenglart's inequality (cf. Gill, 1980) to \( L^{\beta_n}(t) \), defined two lines below (4.1) on page 310, which he claims to be a reverse martingale with respect to the \( \sigma \)-fields \( \mathcal{F}_{i}^{\beta_n} \) defined three lines below (4.2) on page 311 of the paper. However, this claim need not hold in the presence of censoring; in fact, \( L^{\beta_n}(t) \) may even fail to be \( \mathcal{F}_{i}^{\beta_n} \)-measurable. Letting \( \hat{F}_{n,b}^b \) denote the Kaplan-Meier estimator (which is \( \hat{F}_{n,b} \) in our notation) based on the residuals \( \tilde{y}_i(b) \), Ritov concludes that \( L^{\beta_n}(t) \) is \( \mathcal{F}_{i}^{\beta_n} \)-measurable from the observation that

\[
(3.45) \quad \{\hat{F}_{n,b}^b(u) : u > t\} \text{ is } \mathcal{F}_{i}^{\beta_n} \text{ - measurable},
\]

(see the second paragraph after (4.2) on page 311 of his paper). However, (3.45) is false, as will be shown in the Appendix.

There does not appear to be a simple martingale proof of (3.43). We are able to prove (3.43) by using an argument similar to the proof of Theorem 1 in the Appendix. However, the apparent
simplicity of Ritov’s approach to derive the asymptotic normality of Buckley-James-type estimators via (3.43) disappears, since there is no gain in simplicity over a more direct analysis of $\Psi_n(b)$ along the lines similar to those used to prove Theorem 1. Furthermore, the restrictive assumptions (3.42a) and (3.42b), which entail prior knowledge of (i) the upper end-point of the support of $y_1 \wedge c_1$ and (ii) bounds for $|\beta^T x_1|$ so that the constant $K$ can be determined, appear to be crucial in the arguments we use to prove (3.43).

4. Confidence regions and asymptotically efficient $M$-estimators of $\beta$.

Theorem 1 can be used to construct approximate $(1 - \alpha)$-level confidence regions for $\beta$ by a straightforward extension of the ideas of Wei, Ying and Lin (1990) who studied the problem of constructing confidence regions from rank estimators in the censored regression model. Since $a$ is fixed in Theorem 1, we shall simply write $\hat{\beta}_n$ instead of $\hat{\beta}_n(a)$, and $\hat{\xi}_n(\beta)$ instead of $\hat{\xi}_n(a, \beta)$. In view of (3.17) and (3.21), an obvious estimator of the matrix $V_a$ defined in (3.30) is

\begin{equation}
\hat{V}_n = \sum_{j=1}^n \left\{ n^{-1} \sum_{i=1}^n x_i^2 x_i^T I_{\{X_i^2(\hat{\beta}_n) \leq \bar{y}_j^2(\hat{\beta}_n) \leq \bar{y}_j^2(\hat{\beta}_n)\}} - \frac{Z_n(\bar{y}_n(\hat{\beta}_n), \bar{y}_n(\hat{\beta}_n))}{n Z_n(\hat{\beta}_n, \bar{y}_n(\hat{\beta}_n))} \right\} \times \left\{ \int_0^\infty [1 - \hat{F}_{n, \hat{\beta}_n}(s)] \psi'(s - a) p_{n,1}(\hat{\beta}_n, s) p_{n,2}(\hat{\beta}_n, s) ds \right\}^2 / Z_n(\hat{\beta}_n, \bar{y}_n(\hat{\beta}_n)),
\end{equation}

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

\begin{equation}
\{ b : \|b\| \leq B, n^{-1} \hat{\xi}_n^T(\beta) V_a^{-1} \hat{\xi}_n(\beta) \leq \chi^2_{1 - \alpha, p} \}
\end{equation}

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

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

\begin{equation}
\lambda' / \lambda = f' / f + f / (1 - F).
\end{equation}

Assuming that $f$ is twice continuously differentiable and that $F(\tau) = 1$, we obtain from (4.3) and integration by parts that

\begin{equation}
\int_u^\tau [1 - F(s | u)] (\frac{\lambda'}{\lambda} - \lambda') ds = \frac{1}{1 - F(u)} \int_u^\tau (1 - F(s)) d(\frac{f'}{f}) = - \frac{\lambda'(u)}{\lambda(u)}, u < \tau.
\end{equation}
Hence, for the special case $a = 0$ and $\psi = (\lambda' / \lambda) - \lambda$ in (3.28) and (3.30), $C_a^{-1}V_aC_a^{-1} = I_f^{-1}$, where

\begin{equation}
I_f = \Delta \int_{r_0}^{r} \{(\Gamma_2(u) - \frac{\Gamma_1(u)\Gamma_1^T(u)}{\Gamma_0(u)}) \{\frac{\lambda'(u)}{\lambda(u)}\}}^2 dF(u).
\end{equation}

Therefore, for the $M$-estimator $\hat{\beta}_n$ associated with the score function $\psi = (\lambda' / \lambda) - \lambda$ (with $a = 0$), the covariance matrix of the limiting normal distribution of $\sqrt{n}(\hat{\beta}_n - \beta)$ under suitable regularity conditions is $I_f^{-1}$. Moreover, as shown by Lai and Ying (1991c), $I_f$ corresponds to the limiting Fisher information matrix for certain parametric families containing the unknown $f$, and $I_f^{-1}$ is an asymptotic lower bound for the covariance matrices of regular estimators in the semiparametric problem of estimating $\beta$ when the common density $f$ of the $\epsilon_i$ 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 $\beta$ to $M$-estimators. Divide the sample into two disjoint subsets, the first of which is $\{(t_i^0, x_i^0, \delta_i^0, y_i^0) : i \leq n/2\}$. From the first subsample, define $L_{n_1}, N_{n_1}, Z_{n_1,p}, J_{n_1,p}$ by (3.5), (3.7) and (3.8) in which $n$ is replaced by $n_1$. Also define $\hat{F}_{n_1,b}$ by (2.22) with $n$ replaced by $n_1$. Let $\hat{\phi}_{n,2}$ be an estimate of $(\lambda' / \lambda) - \lambda$ based on the second subsample of $n_2 = n - n_1$ observations and define in analogy with (3.9)

\begin{equation}
\hat{\xi}_{n_1}(b) = -\int_{-\infty}^{\infty} \hat{\psi}_{n,2}(u) dZ_{n_1,p}(b, u) \{-\int_{-\infty}^{\infty} \left\{ \int_{u}^{\infty} (1 - \hat{F}_{n_1,b}(v|u)) \hat{\phi}_{n,2}'(v)p_{n_1,2}(b, u, v) dv \right\} dJ_{n_1,p}(b, u).
\end{equation}

Likewise from the second subsample define

\begin{equation}
\hat{\xi}_{[n,n_1]}(b) = -\int_{-\infty}^{\infty} \hat{\psi}_{n,1}(u) dZ_{[n,n_1],p}(b, u) \{-\int_{-\infty}^{\infty} \left\{ \int_{u}^{\infty} (1 - \hat{F}_{[n,n_1],b}(v|u)) \hat{\psi}_{n,1}'(v)p_{[n,n_1],2}(b, u, v) dv \right\} dJ_{[n,n_1]}(b, u),
\end{equation}

where $\hat{\psi}_{n,1}$ represents an estimate of $(\lambda' / \lambda) - \lambda$ based on the first subsample and we use the notation $L_{[n,n_1]}, Z_{[n,n_1],p}, J_{[n,n_1],p}$, etc., to denote (3.5), (3.7), (3.8), etc., in which the sum $\sum_{i=1}^{n}$ is replaced by $\sum_{i=n_1+1}^{n}$ (i.e., the summands are only from the second subsample). Combining the two subsample statistics (4.6) and (4.7) gives

\begin{equation}
\xi^n(b) = \hat{\xi}_{n_1}(b) + \hat{\xi}_{[n,n_1]}(b).
\end{equation}
From the $j$th subsample, starting with preliminary estimates $b_{n,j}$ such that $b_{n,j} - \beta = O_p(n^{-r})$ for some $r > 0$, 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 $j$th subsample a smooth consistent estimate $\hat{\lambda}_{n,j}$ of the hazard function $\lambda$, and (ii) how to smooth $\hat{\lambda}_{n,j}$ to obtain a smooth consistent estimate of $\lambda' / \lambda$. Using these smooth consistent estimates to define smooth consistent estimates $\hat{\psi}_{n,2}, \hat{\psi}_{n,1}$ of $(\lambda' / \lambda) - \lambda$ for (4.6) and (4.7), it can be shown by a modification of the proof of Theorem 1 (see Remark in the Appendix) that the adaptive $M$-estimator $\beta_n^*$, defined as a minimizer of $\xi_n^*(b)$ in the sphere $\{b : \|b - (b_{n,1} + b_{n,2})/2\| \leq n^{-\epsilon}\}$ with $0 < \epsilon < r \wedge \frac{1}{2}$, is asymptotically efficient in the sense that

$$
(4.9) \quad \sqrt{n}(\beta_n^* - \beta) \xrightarrow{d} N(0, I_f^{-1}),
$$

under (3.13)-(3.17), (3.20) and the additional assumption

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

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 $\{\epsilon_n\}$, as is assumed in Theorem 1. As has been shown in Subsection 3.4, 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 $\{\epsilon_n\}$ and such that $c_i^o \geq t_i^o > -\infty$ and (3.39) holds, under obvious modifications of the assumptions (3.13)-(3.17) and (3.20). Hence we can extend the above construction of confidence regions and adaptive $M$-estimators of $\beta$ to this alternative setting, for which martingale representations and information bounds have also been developed in Lai and Ying (1991c).
APPENDIX

Proof of Lemma 1. For notational simplicity, we shall assume that $a = 0$ and $\beta = 0$. To prove (i), note that the left hand side of (3.10) is equal to 0 if $c_i < t_i$. If $t_i \leq c_i$, integration by parts shows that the left hand side of (3.10) is equal to

$$g_1(t_i) \left\{ \int_{t_i}^{c_i} g_2(u)(1 - F(u))d\psi(u) + \int_{-\infty}^{\infty} (1 - F(u))^{-1} \left[ \int_u^{\infty} (1 - F(v)) g_2(v) d\psi(v) \right] \times [(1 - F(c_i))dI_{\{u \geq c_i\}} - (1 - F(t_i))dI_{\{u \geq t_i\}}] \right\}$$

$$= g_1(t_i) \left\{ \int_{t_i}^{c_i} g_2(u)(1 - F(u))d\psi(u) - \int_{c_i}^{\infty} (1 - F(v)) g_2(v) d\psi(v) \right\} = 0.$$

The proof of (ii) is similar.

Proof of Theorem 1(ii). The proof of (3.29) makes use of (3.25), (3.26) and an analysis of $\xi_m(a, b) - \xi_m(a, \beta)$ similar to the proof of Lemma 3(ii) of Lai and Ying (1989). To prove the asymptotic normality of $\hat{\xi}_n(a, \beta)$, let $\Lambda(u) = -\log(1 - F(u))$ and define

$$(A.1) \quad M_n(s) = \sum_{i=1}^{m(n)} I_{\{t_i(\beta) \leq \bar{s} \wedge c_i(\beta)\}} - \int_{-\infty}^{s} Z_n(\beta, u) d\Lambda(u).$$

Let $G_s$ be the complete $\sigma$-field generated by

$$(A.2) \quad x_i, t_i, I_{\{t_i(\beta) \leq \bar{s}_i(\beta)\}}, \delta_i I_{\{t_i(\beta) \leq \bar{s}_i(\beta) \leq s\}}, I_{\{t_i(\beta) \leq \bar{s}_i(\beta)\}}, I_{\{t_i(\beta) \leq s \wedge \bar{c}_i(\beta)\}}$$

$$(u \leq s, i = 1, 2, \cdots).$$

Then $\{M_n(s), G_s, -\infty < s < \infty\}$ is a martingale with predictable variation process

$$(A.3) \quad \langle M_n \rangle(s) = \int_{-\infty}^{s} Z_n(\beta, u) d\Lambda(u),$$

cf. Lemma 5 of Lai and Ying (1991a). Define

$$(A.4) \quad \tilde{\xi}^{(1)}_n = -\int_{-\infty}^{\infty} \psi(u - a) dZ_{n, p}(\beta, u)$$

$$- \int_{-\infty}^{\infty} \left\{ \int_{u}^{\infty} (1 - F(v|u)) \psi'(v - a) p_{n, 2}(\beta, u) dv \right\} dJ_{n, \beta}(\beta, u),$$

$$\tilde{\xi}^{(2)}_n = \int_{-\infty}^{\infty} \left\{ \int_{-\infty}^{u} (1 - \tilde{F}_{n, \beta}(u - |s|)) p_{n, 2}(\beta, s) dJ_{n, p}(\beta, s) \right\}$$

$$\times \left\{ \int_{u}^{\infty} (1 - F(v|u)) p_{n, 2}(\beta, v) \psi'(v - a) dv \right\} \frac{dM_n(u)}{Z_n(\beta, u)}.$$
Then by the same argument as in the proof of Lemma 2 of Lai and Ying (1990), it can be shown that

\[(A.5)\quad P\{\xi_n(a, \beta) = \hat{\xi}_n^{(1)} + \hat{\xi}_n^{(2)} \text{ for all large } n\} = 1.\]

Analogous to the proof of Lemma 4 of Lai and Ying (1989), we approximate \(p_{n,j}(\beta, u)\) by \(\tilde{p}_{m(n),j}(\beta, u)\) \((j = 1, 2)\) in \(\hat{\xi}_n^{(1)}\) and obtain that

\[(A.6)\quad \hat{\xi}_n^{(1)} = \bar{\xi}_n^{(1)} + o_p(n^{1/2}), \quad \text{where}\]

\[
\bar{\xi}_n^{(1)} = -\sum_{i=1}^{m(n)} x_i \tilde{p}_{m(n),1}(\beta, \theta_i(\beta)) \left\{ \int_{-\infty}^{\infty} \psi(u - a) d[p_{m(n),2}(\beta, u) I_{\{t_i(\beta) \leq u \leq \tilde{\gamma}(\beta)\}}] \right.
\]

\[
+ \int_{-\infty}^{\infty} \left[ \int_u^{\infty} (1 - F(v|u)) \tilde{p}_{m(n),2}(\beta, v) \psi'(v - a) dv \right] d[I_{\{t_i(\beta) \leq c_i(\beta) \leq u, c_i(\beta) < \tilde{c}_i\}} + I_{\{t_i(\beta) \leq u \land \tilde{\gamma}(\beta)\}}].
\]

Moreover, using arguments similar to those in the proof of Lemmas 5 and 6 of Lai and Ying (1989), it can be shown that

\[(A.7)\quad \hat{\xi}_n^{(2)} = \bar{\xi}_n^{(2)} + o_p(1), \quad \text{where}\]

\[
\bar{\xi}_n^{(2)} = \int_{r_0}^{r} \Gamma_1(u) \left\{ \int_u^{r} (1 - F(v|u)) \psi'(v - a) dv / \Gamma_0(u) \right\} dM_n(u).
\]

In view of (A.5)-(A.7), we can use \((3.21), (3.3)\) and the same argument as in the proof of Theorem 2(ii) of Lai and Ying (1989) to show that \(n^{-1/2}\bar{\xi}_n(a, \beta)\) has a limiting normal distribution with mean 0 and covariance matrix \(V_a\).

**Remark.** The preceding arguments can be modified to establish for the adaptive statistics \(\xi_n^*(b)\) defined in (4.8) the following asymptotic linearity and asymptotic normality properties under the assumptions (3.13)-(3.17), (3.20) and (4.10):

\[(A.8)\quad \xi_n^*(b) = \xi_n^*(\beta) - nI_f(b - \beta) + o(n^{1/2} \vee n\|b - \beta\|) \text{ uniformly in } \|b - \beta\| \leq n^{-\epsilon},\]

with probability 1, for every \(\epsilon > 0\), and

\[(A.9)\quad n^{-1/2}\xi_n^*(\beta) \xrightarrow{\mathcal{D}} N(0, I_f).\]

To prove (A.8), it suffices to prove the corresponding results for \(\hat{\xi}_{n,1}(b)\) and \(\hat{\xi}_{n,n,1}(b)\) by applying Lemma 4 of Lai and Ying (1991b) and arguments similar to the proofs of Lemma 3(ii) of Lai and Ying (1989). Moreover, (A.9) can be proved by a modification of the preceding proof of the asymptotic normality of \(\hat{\xi}_n(a, \beta)\), using ideas similar to those of the proof of Theorem 2 of Lai and Ying (1991b). Since \(I_f\) is assumed to be nonsingular, (4.9) follows from (A.8) and (A.9).
Proof of Theorem 2. From (3.26) and (3.37), it follows that \( \hat{\beta}_n \rightarrow \beta \) 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 for every \( \delta > 0 \) and \( 0 < \gamma < 1 \),

\[
(A.10) \sup_{\|b\| \leq B, \tau > u \geq s_0 + A_0} |\hat{F}_{n,b}(u|s_0 + A_0) - F_{m(n),b}(u|s_0 + A_0)| = o(n^{-1/2+\delta}) \quad \text{a.s.,}
\]

\[
(A.11) \sup_{\|b - \beta\| \leq n^{-\gamma}, \tau > u \geq s_0 + A_0} |\hat{F}_{n,b}(u|s_0 + A_0) - \hat{F}_{n,\beta}(u|s_0 + A_0) - F_{m(n),b}(u|s_0 + A_0) + F_{m(n),\beta}(u|s_0 + A_0)| = o(n^{-1/2+\delta}) \quad \text{a.s.,}
\]

where \( F_{m,b}(\cdot, \cdot) \) is defined in (3.22). Since \( \hat{\beta}_n \rightarrow \beta \) a.s. and since \( \sup_{\tau > u \geq s_0 + A_0} |F_{m,b}(u|s_0 + A_0) - F(u|s_0 + A_0)| \rightarrow 0 \) as \( m \rightarrow \infty \) and \( b \rightarrow \beta \), we obtain from (A.10) that \( \sup_{\tau > u \geq s_0 + A_0} |\hat{F}_{n,b}(u|s_0 + A_0) - F(u|s_0 + A_0)| \rightarrow 0 \) a.s., and therefore by (3.1),

\[
(A.12) \sup_{A_0 \leq a \leq A_1} |\int_{s_0 + A_0}^{s_1 + A_1} \psi(u) \, d\hat{F}_{n,b}(u|s_0 + A_0) - \int_{s_0 + A_0}^{s_1 + A_1} \psi(u) \, dF(u|s_0 + A_0)| \rightarrow 0 \quad \text{a.s.}
\]

Since \( \int_{s_0 + A_0}^{s_1 + A_1} \psi(u-a) \, dF(u|s_0 + A_0) = (1 - F(s_0 + A_0))^{-1} \int_{s_0}^{s_1} \psi(u-a) \, dF(u) \) for \( a \in [A_0, A_1] \) by (3.33), it follows from (A.12) and (3.36) that \( \hat{\alpha}_n \rightarrow \alpha \) a.s.

In view of (A.11) and (3.33), we can use an argument similar to the proof of Lemma 3(ii) of Lai and Ying (1989) to extend (3.29) to the following asymptotic linearity property: With probability 1, for every \( \delta > 0 \),

\[
(A.13) \begin{pmatrix} \hat{\xi}_n(a,b) - \hat{\xi}_n(\alpha,\beta) \\ \hat{\xi}_n(a,b) - \hat{\xi}_n(\alpha,\beta) \end{pmatrix} = n \begin{pmatrix} -(1 - F(s_0 + A_0))^{-1} \psi_\alpha & -C_\alpha \\ 0 & -b - \beta \end{pmatrix} \begin{pmatrix} a - \alpha \\ b - \beta \end{pmatrix} + o(\sqrt{n} \wedge n \|b - \beta\| \wedge n |a - \alpha|) \quad \text{uniformly in } |a - \alpha| + \|b - \beta\| \leq n^{-\delta},
\]

We next make use of (A.5)-(A.7) together with the weak convergence of \( \sqrt{n} \{\hat{F}_{n,b}(\cdot|s_0 + A_0) - F(\cdot|s_0 + A_0)\} \) established in Theorem 5(i) of Lai and Ying (1991a) to show that

\[
(A.14) \frac{1}{\sqrt{n}} \begin{pmatrix} \hat{\xi}_n(\alpha,\beta) \\ \hat{\xi}_n(\alpha,\beta) \end{pmatrix} \overset{L}{\rightarrow} N(0, \begin{pmatrix} (1 - F(s_0 + A_0))^{-2} \psi_\alpha & 0 \\ 0 & V_\alpha \end{pmatrix}).
\]

From (3.1) and (3.32)-(3.34), it follows that

\[
(A.15) n^{-1} \hat{\xi}_n(\alpha, \beta) = \int_{s_0 + A_0}^{s_1 + A_1} \psi(u - \alpha) \, d[\hat{F}_{n,b}(u|s_0 + A_0) - F(u|s_0 + A_0)]
\]

\[
= -\int_{s_0 + A_0}^{s_1 + A_1} \{\hat{F}_{n,b}(u|s_0 + A_0) - F(u|s_0 + A_0)\} \psi'(u - \alpha) \, du.
\]
By Theorem 5(i) of Lai and Ying (1991a), $\sqrt{n}\{F_{n,0}(\cdot|s_0 + A_0) - F(\cdot|s_0 + A_0)\}$ 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 independent increments and variance function

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

From this and (A.15), it follows that $n^{-1/2} \xi_n(\alpha, \beta) \xi N(0, \{1 - F(s_0 + A_0)\}^{-2} \nu_0)$ where $\nu_0$ is given by (3.38).

Defining $M_n(s)$ by (A.1) and $\xi_n^{(1)}, \xi_n^{(2)}$ by (A.6) and (A.7) with $a = \alpha$, it can be shown that

$$n^{-1} E\{\xi_n^{(1)} M_n(t)\} \to -\Delta \int_{s_0 + A_0}^{t} \Gamma_1(s)\left\{\int_{s}^{s_1 + A_1} (1 - F(v)) \psi'(v - \alpha) dv\right\} dF(s),$$

$$n^{-1} E\{\xi_n^{(2)} M_n(t)\} \to \Delta \int_{s_0 + A_0}^{t} \Gamma_1(s)\left\{\int_{s}^{s_1 + A_1} (1 - F(v)) \psi'(v - \alpha) dv\right\} dF(s).$$

Moreover, by (4.17) of Lai and Ying (1991a)

$$F_n,0(u|s_0 + A_0) - F(u|s_0 + A_0) = (1 - F(u|s_0 + A_0))$$

$$\times \int_{(s_0 + A_0, u]} \frac{1 - F_n,0(t)}{1 - F(t|s_0 + A_0)} I\{Z_n(\beta, t) > 0\} dM_n(t).$$

From (A.15), (A.17) and (A.18), together with (A.5)-(A.7) and standard arguments involving Rebolledo's central limit theorem (cf. Gill, 1980), (A.14) follows.

Let $p = 1 - F(s_0 + A_0)$. From (A.13) and (A.14), we obtain as in Theorem 1(iii) that $|\hat{\alpha}_n - \alpha| + ||\hat{\beta}_n - \beta|| = O(n^{-\epsilon})$ a.s. and that $\sqrt{n}(\hat{\alpha}_n - \alpha, \hat{\beta}_n - \beta^T)^T$ has a limiting normal distribution with mean 0 and covariance matrix

$$
\begin{pmatrix}
-p\psi^{-1}_\alpha & -\psi^{-1}_\alpha g^T \alpha C^{-1} \\
0 & -C^{-1}_\alpha
\end{pmatrix} \begin{pmatrix}
p^{-2} \nu_0 & 0 \\
0 & V_\alpha
\end{pmatrix} \begin{pmatrix}
-p\psi^{-1}_\alpha & 0 \\
-\psi^{-1}_\alpha C^{-1}_\alpha & -C^{-1}_\alpha
\end{pmatrix} = (3.38),
$$

noting that

$$
\begin{pmatrix}
-p^{-1} \psi_\alpha & p^{-1} g^T \alpha \\
0 & -C^{-1}_\alpha
\end{pmatrix}^{-1} = \begin{pmatrix}
-p\psi^{-1}_\alpha & -\psi^{-1}_\alpha g^T \alpha C^{-1} \\
0 & -C^{-1}_\alpha
\end{pmatrix}.
$$

Counter-example to (3.45). To see that (3.45) is false in the presence of censoring, consider the case $\beta_n = \beta_0 = 0$ and drop the superscript $\beta_n$ in $F_n^{\beta_n}$. In this special case, note that the Kaplan-Meier curve $F_n^{\beta_n}$ based on the residuals simply corresponds to the classical Kaplan-Meier curve $\hat{F}$ based on $(\epsilon_i \wedge c_i, \delta_i)_{1 \leq i \leq n}$, and Ritov's (1990) $\sigma$-field $\mathcal{F}_t$ is generated by

$$I_{\{\epsilon_i \wedge c_i \geq t\}}, I_{\{\epsilon_i \wedge c_i \geq t\}}(x_i, \epsilon_i \wedge c_i, \delta_i), \quad i = 1, \ldots, n.$$
Note that unlike the increasing family \( \{G_s\} \) defined by (A.2), \( \{\mathcal{F}_i\} \) is a decreasing family of \( \sigma \)-fields. Ritov claims that \( \hat{F}(t) \) is \( \mathcal{F}_1 \)-measurable, but this is not true in view of Efron's (1967) "redistributed-to-the-right" characterization of the Kaplan-Meier estimator. The following simple counter-example suffices to show this point.

Let \( n = 3 \). If \( \hat{F}(2) \) should be \( \mathcal{F}_2 \)-measurable, then by the definition \( \mathcal{F}_2 \), \( \hat{F}(2) \) should be uniquely determined from

\[
I_{\{\epsilon_i \wedge c_i \geq 2\}}, \ I_{\{\epsilon_i \wedge c_i \geq 2\}}(x_i, \epsilon_i \wedge c_i, \delta_i) \quad (i = 1, 2, 3).
\]

(A.20)

Consider the following two cases in which \( \epsilon_1 \wedge c_1 = 1, \epsilon_2 \wedge c_2 = 1.5, \epsilon_3 \wedge c_3 = 3 \), and \( \delta_3 = 1 \). Note that these two cases have the same values of the variables in (A.20).

- **Case 1.** \( \delta_1 = 0, \delta_2 = 1 \). Then \( \hat{F}(2) = 1/2 \).
- **Case 2.** \( \delta_1 = 1, \delta_2 = 0 \). Then \( \hat{F}(2) = 1/3 \).

These two cases therefore have different values of \( \hat{F}(2) \) although they have the same values for the random variables in (A.20) that generate \( \mathcal{F}_2 \). Therefore \( \hat{F}(2) \) is not \( \mathcal{F}_2 \)-measurable, contrary to the claim (3.45).
REFERENCES


