ASYMPTOTIC NORMALITY OF A CLASS OF ADAPTIVE STATISTICS WITH APPLICATIONS TO SYNTHETIC DATA METHODS FOR CENSORED REGRESSION

by

Tze Leung Lai, Zhiliang Ying and Zukang Zheng

TECHNICAL REPORT NO. 4
MARCH 1993

Prepared Under Grant MDA 909-92-H-3092
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
ASYMPTOTIC NORMALITY OF A CLASS OF ADAPTIVE STATISTICS WITH APPLICATIONS TO SYNTHETIC DATA METHODS FOR CENSORED REGRESSION

by

Tze Leung Lai, Zhiliang Ying and Zukang Zheng

TECHNICAL REPORT NO. 4
MARCH 1993

PREPARED UNDER GRANT MDA 909-92-H-3092
FOR THE NATIONAL SECURITY AGENCY

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
Asymptotic Normality of a Class of Adaptive Statistics with Applications to Synthetic Data Methods for Censored Regression

TZE LEUNG LAI*, ZHILIANG YING† AND ZUKANG ZHENG

Stanford University, University of Illinois and Fudan University

Abstract

Motivated by regression analysis of censored survival data, we develop herein a general asymptotic distribution theory for estimators defined by estimating equations of the form \( \sum_{i=1}^{n} \xi(w_i, \theta, \hat{G}_n) = 0 \), in which the \( w_i \) represent observed data, \( \theta \) is an unknown parameter to be estimated, and \( \hat{G}_n \) represents an estimate of some unknown underlying distribution. This general theory is used to establish asymptotic normality of synthetic least squares estimates in censored regression models and to evaluate the covariance matrices of the limiting normal distributions.

AMS 1980 subject classifications: primary 62J99, secondary 62G05, 62P10, 60G44.

Key words and phrases: censored data, regression, asymptotic distribution, adaptive estimating equation, von Mises calculus, martingales.

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

† Research supported by the National Science Foundation and the National Security Agency.
Running title: CLT FOR ADAPTIVE STATISTICS

Address for correspondence and galley proofs

Tze Leung Lai
Department of Statistics
Stanford University
Stanford, CA 94305
1. INTRODUCTION AND SUMMARY

Suppose that in the linear regression model

\[ E(y_i|x_i) = \beta'x_i, \]

(1.1a)

the \( y_i \) are not completely observable and the observations are \((x_i, \bar{y}_i, \delta_i), i = 1, \ldots, n\), where \( \beta \) and \( x_i \) are \( p \times 1 \) vectors,

\[ \bar{y}_i = y_i \wedge c_i, \quad \delta_i = I_{\{y_i \leq c_i\}}, \]

(1.1b)

in which \( \wedge \) denotes minimum. Letting \( \epsilon_i = y_i - \beta'x_i \) and assuming the \( \epsilon_i \) to be i.i.d. random variables with a common distribution function \( F \) and the \((c_i, x'_i)\) to be independent random vectors that are independent of \( \{\epsilon_i\} \), Buckley and James [1] introduced the following adaptive approach to estimate \( \theta \). Suppose first that \( F \) is known. Since

\[ E(y_i|\delta_i, \bar{y}_i, x_i) = \bar{y}_i + (1 - \delta_i) \int_{\bar{y}_i - \beta'x_i}^{\infty} (1 - F(s)) \, ds / (1 - F(\bar{y}_i - \beta'x_i)), \]

and since \( E[E(y_i|\delta_i, \bar{y}_i, x_i)|x_i] = E(y_i|x_i) = \beta'x_i \), least squares regression of \( E(y_i|\delta_i, \bar{y}_i, x_i) \) on \( x_i \) suggests estimating \( \beta \) by solving the equation

\[ \sum_{i=1}^{n} x_i \zeta(\bar{y}_i - \theta'x_i, \delta_i; F) = 0, \quad \text{where} \]

(1.2a)

\[ \zeta(x, \delta; F) = z + (1 - \delta) \int_{z}^{\infty} (1 - F(s)) \, ds / (1 - F(z)). \]

When \( F \) is unknown, replacing it in (1.2a) by the Kaplan-Meier estimator \( \hat{F}_{n, \theta} \) of \( F \) based on \((\bar{y}_i - \theta'x_i, \delta_i)_{1 \leq i \leq n}\) yields the Buckley-James estimator of \( \beta \) defined by the estimating equation

\[ \sum_{i=1}^{n} x_i \zeta(\bar{y}_i - \theta'x_i, \delta_i; \hat{F}_{n, \theta}) = 0. \]

(1.2b)

This approach has recently been extended to construct M-estimators of \( \beta \) based on the censored data (1.1b) by Ritov [11] and Lai and Ying [8]. The estimating equations defining these M-estimators are again of the form (1.2b) but with a different definition of \( \zeta \).

An alternative adaptive approach to estimate \( \beta \) was proposed by Koul, Susarla and Van Ryzin [6], who assumed the censoring variables \( c_i \) to be i.i.d. with a common distribution
function $G$ and such that $\{c_i\}$ and $\{(x_i, y_i)\}$ are independent. If $G$ is continuous and known, they showed that an unbiased estimator of $\beta$ is given by the linear estimating equation

$$\sum_{i=1}^{n} x_i \{\eta(y_i, \delta_i; G) - \theta' x_i\} = 0,$$

where $\eta(y_i, \delta_i; G) = \delta_i y_i / (1 - G(y_i))$. \hspace{1cm} (1.3)

When $G$ is unknown, they proposed to replace $G$ in (1.3) by a product-limit estimator $\hat{G}_n$. This approach to estimate $\theta$ was subsequently refined and extended by Leurgans [9] and Zheng [14], and a detailed discussion will be given in Section 2.

Note that both the Buckley-James and the Koul-Susarla-Van Ryzin estimators are adaptive statistics defined by estimating equations of the form

$$\sum_{i=1}^{n} \xi(w_i, \theta, \hat{H}_{n, \theta}) = 0,$$ \hspace{1cm} (1.4)

where the $w_i$ represent the observed data (with $w_i = (x_i, y_i, \delta_i)$ for censored regression data), $\theta$ is an unknown parameter and $\hat{H}_{n, \theta}$ is an estimator of an unknown distribution $H$. The idea behind these adaptive statistics is that one begins by assuming $H$ to be known and arrives at an estimating equation of the form $\sum_{i=1}^{n} \xi(w_i, \theta, H) = 0$. Replacing the unknown $H$ by an estimator $\hat{H}_{n, \theta}$ (which may also depend on $\theta$ as in the Buckley-James case but not in the Koul-Susarla-Van Ryzin case) gives the estimating equation (1.4).

In this paper we study the problem of asymptotic normality of solutions of estimating equations of the form (1.4) in which $\theta$ is $p \times 1$ and the $w_i$ are $q \times 1$ vectors, $\xi(w, \theta, H)$ is a $p \times 1$ vector and $H$ is a univariate or multivariate distribution function. As in [7] for the Buckley-James estimator, we establish the desired asymptotic normality in two steps. The first step is to show that the random function $S_n : \mathbf{R}^p \rightarrow \mathbf{R}^p$ defined by $S_n(\theta) = \sum_{i=1}^{n} \xi(w_i, \theta, \hat{H}_{n, \theta})$ is asymptotically linear in some neighborhood of the true parameter $\theta_0$ as $n \rightarrow \infty$, with the diameter of the neighborhood approaching 0. Letting $\hat{H}_n = \hat{H}_{n, \theta_0}$, the second step is to show that the adaptive statistic $\sum_{i=1}^{n} \xi(w_i, \theta_0, \hat{H}_n)$ is asymptotically normal by using an approximation of the form

$$\sum_{i=1}^{n} \xi(w_i, \theta_0, \hat{H}_n) = \sum_{i=1}^{n} \xi(w_i, \theta_0, H) + \sum_{i=1}^{n} w_i^* + o_p(\sqrt{n}),$$ \hspace{1cm} (1.5)
where \((w_1, w^*_1), (w_2, w^*_2), \ldots\) are independent and
\[
E \xi(w_i, \theta_0, H) = 0 = E w^*_i \text{ for all } i, \text{ and for every } \delta > 0,
\]
\[
\lim_{n \to \infty} n^{-1} \sum_{i=1}^{n} E \{ ||\xi(w_i, \theta_0, H) + w^*_i||^2 I_{\{||\xi(w_i, \theta_0, H) + w^*_i||^2 \geq \delta n\}} \} = 0.
\] (1.6)

Formally, letting \(\xi_0(w, H) = \xi(w, \theta_0, H)\), one can derive the approximation (1.5) by using (i) the functional delta method that yields
\[
n^{-1} \sum_{i=1}^{n} \{ \xi_0(w_i, \hat{H}_n) - \xi_0(w_i, H) \} \overset{d}{=} \{ n^{-1} \sum_{i=1}^{n} D\xi_0(w_i) \} (\hat{H}_n - H),
\] (1.7)
where \(D\xi_0(w)\) denotes the partial (Hadamard) derivative of \(\xi_0(w, H)\) with respect to \(H\) (cf. [4]), and (ii) the law of large numbers for the independent random variables \(D\xi(w_i)\) taking values in some separable normed vector space of linear operators, that yields
\[
n^{-1} \sum_{i=1}^{n} D\xi_0(w_i) \overset{P}{\to} \ell,
\] (1.8)
where \(\ell\) is a linear functional (taking values in \(\mathbb{R}^p\)). For a rigorous justification of these formal calculations, one needs assumptions on Hadamard differentiability of \(\xi_0\) at \(H\) and on finiteness of moments of certain norms of \(D\xi_0(w_i)\) that will be discussed in Section 3. In applications to particular problems, however, one can often bypass these assumptions by checking (1.5) directly without going through (1.7) and (1.8), as we shall do in Section 2 for a class of synthetic estimators in censored regression models (including the Koul-Susarla-Van Ryzin [6] and Leurgans [9] estimators as special cases).

Concerning the asymptotic linearity of \(S_n(\theta)\) as \(n \to \infty\) and \(\theta \to \theta_0\), formal differentiation (with respect to \(H\)) and the law of large numbers as in (1.7) and (1.8) yield
\[
n^{-1} \sum_{i=1}^{n} \xi(w_i, \theta, \hat{H}_{n, \theta}) - n^{-1} \sum_{i=1}^{n} \xi(w_i, \theta, \hat{H}_n) \overset{d}{=} \ell(\hat{H}_{n, \theta} - \hat{H}_n)
\] as \(n \to \infty\) and \(\theta \to \theta_0\). Under suitable regularity conditions, we can further approximate \(\ell(\hat{H}_{n, \theta} - \hat{H}_n)\) by \(A_1(\theta - \theta_0)\), where \(A_1\) is some nonrandom \(p \times p\) matrix, recalling that \(\hat{H}_n = \hat{H}_{n, \theta_0}\). Likewise, formal differentiation with respect to \(\theta\) yields
\[
n^{-1} \sum_{i=1}^{n} \xi(w_i, \theta, \hat{H}_n) - n^{-1} \sum_{i=1}^{n} \xi(w_i, \theta_0, \hat{H}_n) \overset{d}{=} \{ n^{-1} \sum_{i=1}^{n} \frac{\partial}{\partial \theta} \xi(w_i, \theta_0, \hat{H}_n) \}(\theta - \theta_0),
\]
and by the law of large numbers, the matrix
\[ n^{-1} \sum_{i=1}^{n} \frac{\partial}{\partial \theta} \xi(w_i, \theta_0, H) \] should converge to some nonrandom matrix \( A_2 \) under suitable assumptions. Hence, to establish the asymptotic linearity of \( S_n(\theta) \), we shall check the following two conditions for some \( \delta > 0 \):

\[
\sup_{\|\theta - \theta_0\| \leq n^{-1/2+\delta}} \| \sum_{i=1}^{n} (\xi(w_i, \theta, \hat{H}_n) - \xi(w_i, \theta, \hat{H}_n)) - nA_1(\theta - \theta_0) \| / (\sqrt{n} + n\|\theta - \theta_0\|) \to 0, \tag{1.9}
\]

\[
\sup_{\|\theta - \theta_0\| \leq n^{-1/2+\delta}} \| \sum_{i=1}^{n} (\xi(w_i, \theta, \hat{H}_n) - \xi(w_i, \theta_0, \hat{H}_n)) - nA_2(\theta - \theta_0) \| / (\sqrt{n} + n\|\theta - \theta_0\|) \to 0, \tag{1.10}
\]
as \( n \to \infty \), for some nonrandom \( p \times p \) matrices \( A_1 \) and \( A_2 \).

The following theorem establishes the asymptotic normality of solutions of estimating equations of the form (1.4) under conditions (1.5), (1.6), (1.9), (1.10). Strictly speaking, (1.4) may not have a solution, in which case we shall take \( \hat{\theta}_n \) that minimizes \( n^{-1} \| \sum_{i=1}^{n} \xi(w_i, \theta, \hat{H}_n) \| \) over some set of \( \theta \)-values to within an \( o_p(n^{-1/2}) \) order to be an estimator of \( \theta \) defined by (1.4), cf. [8]. The theorem will be proved in Section 3, where we shall also discuss (1.5)-(1.10) further and use von Mises calculus for the statistical functional \( \sum_{i=1}^{n} \xi(w_i, \theta_0, \hat{H}_n) \) to obtain an approximation of the functional by \( U \)-statistics. Section 2 gives an application of the theorem to derive the asymptotic distribution of a general class of synthetic least squares estimators in censored regression models.

**THEOREM 1.** (i) Suppose that (1.5) holds, in which \((w_1, w_1^*), (w_2, w_2^*), \ldots\) are independent and satisfy (1.6) and \( \hat{H}_n = \hat{H}_{n, \theta_0} \). Then \( n^{-1/2} \sum_{i=1}^{n} \xi(w_i, \theta_0, \hat{H}_n) \) converges in distribution as \( n \to \infty \) to a multivariate normal distribution with mean 0 and covariance matrix

\[
V = \lim_{n \to \infty} n^{-1} \sum_{i=1}^{n} \text{Cov}(w_i^* + \xi(w_i, \theta_0, H)), \tag{1.11}
\]

assuming that the limit on the right hand side of (1.11) exists.

(ii) Suppose that (1.5), (1.6), (1.9) and (1.10) hold for some \( \delta > 0 \), in which \( A_1 \) and \( A_2 \) are nonrandom \( p \times p \) matrices such that \( A = A_1 + A_2 \) is nonsingular. Let \( \Theta_n = \{ \theta : \|\theta - \theta_0\| \leq n^{-1/2+\delta} \} \) and let \( \hat{\theta}_n \in \Theta_n \) be such that

\[
n^{-1} \| \sum_{i=1}^{n} \xi(w_i, \theta_n, \hat{H}_{n, \hat{\theta}_n}) \| - \inf_{\theta \in \Theta_n} n^{-1} \| \sum_{i=1}^{n} \xi(w_i, \theta, \hat{H}_{n, \theta}) \| = o_p(n^{-1/2}). \tag{1.12}
\]

4
Define $V$ by \((1.11)\), in which the limit is assumed to exist. Then as $n \to \infty$, $\sqrt{n}(\hat{\theta}_n - \theta_0)$ has a limiting normal distribution with mean 0 and covariance matrix $A^{-1}V(A^{-1})'$. 

2. CENSORED REGRESSION ANALYSIS VIA SYNTHETIC DATA

In this section we apply Theorem 1(i) to prove asymptotic normality of a class of estimators of the parameter $\beta$ in the censored regression model \((1.1a)-(1.1b)\). A basic assumption behind these estimators is that the censoring variables $c_i$ are i.i.d. with a common distribution $G$ and such that \(\{c_i\}\) is independent of the sequence of independent random vectors \(\{(x'_i,y_i)\}\). Under this assumption, Koul, Susarla and Van Ryzin [6] noted that if $G$ is known then it is possible to adjust $\tilde{y}_i$ to yield unbiased modifications $y_{i,G}^*$ with $E(y_{i,G}^*|x_i) = E(y_i|x_i)$. Specifically, assuming $G$ to be continuous, note that

\[
y_{i,G}^* = \frac{\delta_i \tilde{y}_i}{(1 - G(\tilde{y}_i))} \tag{2.1}
\]

has the same conditional mean (given $x_i$) as $y_i$, and therefore $(x'_i, y_{i,G}^*)$, $i = 1, \cdots, n$, are independent random vectors with $E(y_{i,G}^*|x_i) = \beta'x_i$. Hence, if $G$ is known, then one can use the least squares estimate

\[
\hat{\beta}_G = \left(\sum_{i=1}^{n} x_i x'_i\right)^{-1} \sum_{i=1}^{n} x_i y_{i,G}^*. \tag{2.2}
\]

When $G$ is unknown, Koul, Susarla and Van Ryzin [6] proposed to replace $G$ by a product-limit estimator $\hat{G}_n$. This is the background behind their estimator

\[
\hat{\beta}_n = \left(\sum_{i=1}^{n} x_i x'_i\right)^{-1} \sum_{i=1}^{n} x_i y_{i,G}^*, \tag{2.3}
\]

in which $y_{i,G}^*$ is defined by \((2.1)\).

Note that \((2.1)\) sets a censored response to be 0 and multiplies an uncensored response by a factor $(1 - G(\tilde{y}_i))^{-1}$, which can be very large if $G(\tilde{y}_i)$ is near 1. It seems more reasonable to prolong the censored response instead of curtailing it to 0. When $G$ is continuous and has support in $[0, \infty)$,

\[
y_{i,G}^* = \int_{0}^{\tilde{y}_i} \frac{ds}{1 - G(s)} \tag{2.4}
\]
also satisfies $E(y^\star_{i,G}|x_i) = E(y_i|x_i)$, as will be shown below. Leurgans [9] proposed to use (2.3) in which $y^\star_{i,G}$ is given by (2.4) instead of (2.1), and called the $y^\star_{i,G}$ in (2.3) "synthetic data", since such data are "synthesized" from the raw data $(\tilde{y}_i, \delta_i)$ to fit the regression model $E(y^\star_{i,G}|x_i) = \beta'x_i$.

Independently of Leurgans, Zheng [14] developed a general method to construct such synthetic data for regression analysis in the presence of right censoring of the response variable. Starting with the case of known $G$, he considered synthetic data of the form

$$y^\star_{i,G} = \delta_i \psi_G(\tilde{y}_i) + (1 - \delta_i) \Psi_G(\tilde{y}_i), \quad (2.5)$$

where $\psi_G$, $\Psi_G$ are Borel functions on the real line such that

$$(1 - G(y-))\psi_G(y) + \int_{t<y} \Psi_G(t)dG(t) = y \quad \text{for all } y. \quad (2.6)$$

In particular, if $G$ has a density function $g$ and $\psi_G$, $\Psi_G$ are differentiable, then differentiating both sides of (2.6) yields the differential equation

$$(1 - G(y)) \frac{d}{dy} \psi_G(y) - g(y) \psi_G(y) + g(y) \Psi_G(y) = 1. \quad (2.7)$$

Since $c_i$ is independent of $(x_i', y_i)$, it follows from (2.5) and (2.6) that

$$E(y^\star_{i,G}|y_i, x_i) = (1 - G(y_i-))\psi_G(y_i) + \int_{t<y_i} \Psi_G(t)dG(t) = y_i,$$

which in turn implies that

$$E(y_i|x_i) = E\{E(y^\star_{i,G}|y_i, x_i)|x_i\} = E(y^\star_{i,G}|x_i). \quad (2.8)$$

Without assuming $G$ to be known, Zheng [14] proposed to use the synthetic least squares estimator (2.3) in which $\hat{G}_n$ is the product-limit estimator and $y^\star_{i,G}$ is given by (2.5) and (2.6).

Let $\alpha: \mathbb{R} \to \mathbb{R}$ be a continuous function such that $\int_{-\infty}^t |\alpha(s)|dG(s)/(1 - G(s-)) < \infty$ for all $t$ and define

$$\psi_G(y) = \frac{y}{1 - G(y-)} - \int_{t<y} \frac{\alpha(t)dG(t)}{1 - G(t-)}, \quad \Psi_G(y) = \alpha(y) - \int_{t<y} \frac{\alpha(t)dG(t)}{1 - G(t-)}. \quad (2.9)$$
Then \((\psi_G, \Psi_G)\) is a solution to (2.6). When \(G\) is continuous, the case \(\alpha \equiv 0\) corresponds to the Koul-Susarla-Van Ryzin method (2.1), while integration by parts can be used to show that the case \(\alpha(t) = t/(1 - G(t))\) corresponds to Leurgans’ method (2.4) if \(G(t) = 0\) for \(t < 0\). We shall focus on synthetic data of the explicit form (2.5) and (2.9). Moreover, without assuming \(G\) to be known, we shall replace \(G\) in (2.5) and (2.9) by the following product-limit estimator \(\hat{G}_n\) of \(G\):

\[
1 - \hat{G}_n(t) = \prod_{s \leq t} (1 - \Delta N_n(s)/Y_n(s)), \quad \text{where}
\]

\[
Y_n(s) = \sum_{i=1}^{n} I_{\{\tilde{y}_i \geq s\}}, \quad N_n(s) = \sum_{i=1}^{n} I_{\{\tilde{y}_i \leq s, \delta_i = 0\}},
\]

in which we use the notation \(\Delta N(s) = N(s) - N(s-)\) and the convention \(0/0 = 0\). Unlike the complicated nonlinear estimating equation (1.2b) in the Buckley-James method, this synthetic data approach has the advantage of giving a linear estimating equation (1.4), in which \(w_i = (x_i, \tilde{y}_i, \delta_i), \hat{H}_{\theta} = \hat{G}_n\) and

\[
\xi(w_i, \theta, H) = x_i\{\delta_i\psi_H(\tilde{y}_i) + (1 - \delta_i)\Psi_H(\tilde{y}_i) - \theta' x_i\},
\]

which is a linear function of \(\theta\). We can solve this linear estimating equation explicitly, giving the unique solution (2.3) when \(\sum_{i=1}^{n} x_i x'_i\) is nonsingular.

The following theorem establishes consistency and asymptotic normality of (2.3) under certain conditions. Let \(\mu_n\) denote the average distribution of \((x_i, y_i), 1 \leq i \leq n\), i.e.,

\[
\mu_n(B, y) = n^{-1} \sum_{i=1}^{n} P\{x_i \in B, y_i \leq y\}
\]

for any Borel subset \(B\) of \(\mathbb{R}^p\) and \(y \in \mathbb{R}\), and let

\[
\mu_{n,y}(y) = \mu_n(\mathbb{R}^p, y) = n^{-1} \sum_{i=1}^{n} P\{y_i \leq y\}.
\]

In addition to the assumptions on \((x_i, y_i, c_i)\) stated at the beginning of this section, we shall also assume the following:

(C1) The limits \(\Gamma_0(t) = \lim_{n \to \infty} n^{-1} \sum_{i=1}^{n} P\{y_i > t\}, \Gamma_1(t) = \lim_{n \to \infty} n^{-1} \sum_{i=1}^{n} E(x_i I_{\{y_i > t\}})\) exist for every \(t\).
(C2) For all $s \leq \tau = \inf\{t : \Gamma_0(t)(1 - G(t)) = 0\}$, $\Delta G(s) \cdot \Delta \Gamma_i(s) = 0$ for $i = 0, 1$, and

$$\int_{-\infty}^{\tau} \left\{ |t| \int_{t > y} t d \Gamma_1(t) ||\Gamma_1(t)||^2 / [(1 - G(y))\Gamma_0(y-)]^2 + \alpha^2(y)||\Gamma_1(y)||^2 / \Gamma_0^2(y-) \right\} \Gamma_0(y)dG(y) < \infty.$$ 

(C3) Letting $\tau_n = \inf\{t : G(t) = 1 \text{ or } \mu_{n,Y}(y) = 1\}$, there exists $y(\epsilon) < \tau$ for every $\epsilon > 0$ such that for all large $n$,

$$\int_{x \in \mathbb{R}^p} \int_{y = y(\epsilon)}^{\tau_n} \frac{||x||||y||d\mu_n(x,y)}{(1 - G(y))(1 - \mu_{n,Y}(y))^1/2} + \int_{y(\epsilon)}^{\tau_n} \frac{\alpha^2(t)dG(t)}{(1 - G(t-))(1 - \mu_{n,Y}(t-))} < \epsilon.$$ 

(C4) The random vectors \( z_i = x_i\{\delta_i\psi_G(\tilde{y}_i) + (1 - \delta_i)\Psi_G(\tilde{y}_i) - \beta'x_i \} \) satisfy the Lindeberg condition \( \lim_{n \to \infty} n^{-1} \sum_{i=1}^{n} E||z_i||^2 I_{\{||z_i|| \geq \epsilon n\}} = 0 \) for every $\epsilon > 0$.

(C5) \( n^{-1} \sum_{i=1}^{n} E\{[\delta_i\psi_G(\tilde{y}_i) + (1 - \delta_i)\Psi_G(\tilde{y}_i) - \beta'x_i]^2 x_i x_i' \} \) converges as $n \to \infty$ to a positive definite matrix $V_1$.

(C6) \( n^{-1} \sum_{i=1}^{n} x_i x_i' \xrightarrow{P} A \) for some nonrandom positive definite matrix $A$.

**THEOREM 2.** In the censored regression model (1.1a)-(1.1b), assume (C1)-(C6) and let $S_n(\theta) = \sum_{i=1}^{n} x_i\{\delta_i\psi_G(\tilde{y}_i) + (1 - \delta_i)\Psi_G(\tilde{y}_i) - \theta'x_i \}$, where $\psi_G$ and $\Psi_G$ are given by (2.9) and $\hat{G}_n$ is the product-limit estimator (2.10). Let $\hat{\beta}_n$ be the solution of the linear estimating equation $S_n(\theta) = 0$.

(i) As $n \to \infty$, $n^{-1/2}S_n(\beta)$ has a limiting multivariate normal distribution with mean 0 and covariance matrix

$$V = V_1 - \int_{-\infty}^{\tau} \left\{ |t| \int_{t > y} t d \Gamma_1(t) (1 - G(y))\Gamma_0(y-)^{-1} \right\} \frac{\alpha^2(y)\Gamma_1(y)(1 - G(y-))}{\Gamma_0(y-)} dG(y), \quad (2.14)$$

where $\Gamma_0, \Gamma_1, V_1$ are given by (C1) and (C5) and $x^\otimes$ denotes the $p \times p$ matrix $xx'$ for $x \in \mathbb{R}^p$.

(ii) $\hat{\beta}_n \xrightarrow{p} \beta$; in fact, $\sqrt{n}(\hat{\beta}_n - \beta)$ has a limiting normal distribution with mean 0 and covariance matrix $A^{-1}V A^{-1}$, where $A$ is given by (C6).

In Theorem 2, the function $\alpha(t)$ in (2.9) is assumed to be fixed in advance and therefore cannot involve the unknown $G$. Hence we cannot set $\alpha(t) = t/(1 - G(t))$ in Theorem 2 to derive from it the asymptotic normality of Leurgans’ estimator (2.3)-(2.4) even when $G$ is assumed to be continuous and to have support in $[0, \infty)$. To derive the asymptotic normality of Leurgans’
estimator, we shall apply Theorem 3 below, whose proof is a straightforward modification of that of Theorem 2. Assuming that \( G(t) = 0 \) if \( t < 0 \), a solution to (2.6) is \((\psi_G, \Psi_G)\) defined by

\[
\psi_G(y) = \int_0^y \frac{dt}{1 - G(t-)} - \int_{t < y} \frac{\alpha(t)dG(t)}{1 - G(t-)}, \quad \Psi_G(y) = \alpha(y) + \int_0^y \frac{dt}{1 - G(t-)} - \int_{t \leq y} \frac{\alpha(t)dG(t)}{1 - G(t-)},
\]

(2.15)

where \( \alpha : \mathbb{R} \to \mathbb{R} \) is a continuous function such that \( \int_{-\infty}^t |\alpha(s)|dG(s)/(1 - G(s-)) < \infty \) for all \( t \), as is assumed in (2.9). The special case \( \alpha \equiv 0 \) in (2.15) corresponds to Leurgans’ method (2.4). Replacing (2.9) by (2.15) entails the following change in Condition (C2):

(D2) For all \( s \leq \tau = \inf \{ t : \Gamma_0(t)(1 - G(t)) = 0 \} \), \( \Delta G(s) \cdot \Delta \Gamma_i(s) = 0 \) for \( i = 0, 1 \), and

\[
\int_{-\infty}^\tau \left\{ \left[ \int_{t > y} \Gamma_1(t)dt \right]^2 / [(1 - G(y))\Gamma_0(y-)]^2 + \alpha^2(y)\| \Gamma_1(y) \|^2 / \Gamma_0^2(y-) \right\} \Gamma_0(y)dG(y) < \infty.
\]

THEOREM 3. Suppose that \( G(t) = 0 \) for \( t < 0 \) and that in Theorem 2 we replace (2.9) by (2.15) and Condition (C2) by (D2). Then the conclusions of Theorem 2 still hold with (2.14) replaced by

\[
V = V_1 - \int_{-\infty}^\tau \left\{ \int_{t > y} \Gamma_1(t)dt \right\} \left( \frac{\alpha(y)\Gamma_1(y)}{\Gamma_0^2(y-)} \right)^{\otimes 2} \left( \frac{(1 - G(y))\Gamma_0(y)}{1 - G(y-) - } \right) dG(y).
\]

(2.16)

When \((x_i, y_i)\) are i.i.d., (C1) is automatically satisfied since the \( y_i \) have a common distribution function \( 1 - \Gamma_0 \), and (C3) can be simplified to

(D3) \[
E \left\{ \frac{\| x_1 \| y_1 \}}{(1 - G(y_1))\Gamma_0^{1/2}(y)} \right\} + \int_{-\infty}^\tau \frac{\alpha^2(t)dG(t)}{(1 - G(t-))\Gamma_0(t-)} < \infty,
\]

noting that \( \tau_n = \tau \) and \( \mu_{n,Y} = 1 - \Gamma_0 \) in this case. Moreover, (C4), (C5) and (C6) can be replaced by the simple condition

(D4) \[
E(\| z_1 \|^2 + \| x_1 \|^2) < \infty, \quad E(z_1z_1^*) \quad \text{and} \quad E(x_1x_1^*) \quad \text{are positive definite}.
\]

The expression (2.14) or (2.16) for the covariance matrix \( V \) of the limiting normal distribution of \( n^{-1/2}S_n(\beta) \) is much simpler than those given by Koul, Susarla and Van Ryzin [6] and Srinivasan and Zhou [12] for the particular case \( \psi_G(y) = y/(1 - G(y-)), \Psi_G(y) = 0 \).
and \( x_i = (1, u_i)' \) and by Zhou [16] for Leurgans' estimator (2.3)–(2.4). Furthermore, even for these special cases, [6], [12] and [16] were only able to establish that \( \sqrt{n}(\hat{\beta}_n - \beta_n) \) has a limiting normal distribution for some specified \( \beta_n \), with \( \beta_n \) converging to \( \beta \) at a rate that may be slower than \( n^{-1/2} \). Theorems 2 and 3, however, give the limiting normal distribution of \( \sqrt{n}(\hat{\beta}_n - \beta) \), which can be therefore used to construct approximate confidence regions for \( \beta \), as will be discussed further in Remark (ii) following the proof of Theorem 2.

The proof of Theorem 3 will be omitted since it is exactly analogous to that of Theorem 2. To prove Theorem 2, we first introduce the following lemma, whose proof will be given at the end of this section.

**Lemma 1.** With the same notation and assumptions as in Theorem 2, let

\[
S_{n,G} = \sum_{i=1}^{n} x_i \{ \delta_i \psi_G(\bar{y}_i) + (1 - \delta_i)\Psi_G(\bar{y}_i) - \beta' x_i \},
\]

\[
\Lambda(t) = \int_{-\infty}^{t} (1 - G(s-))^{-1} dG(s), \quad M_i(t) = (1 - \delta_i)I_{\{\bar{y}_i \leq t\}} - \int_{-\infty}^{t} I_{\{c_i \geq s, y_i > s\}} d\Lambda(s).
\]

Then \( \{M_i(t), -\infty < t < \infty\} \) is a martingale with predictable variation process \( \langle M_i \rangle(t) = \int_{-\infty}^{t} I_{\{c_i \geq s, y_i > s\}} (1 - \Delta \Lambda(s)) d\Lambda(s) \) and

\[
S_{n,G} + \sum_{i=1}^{n} \int_{-\infty}^{t} \left\{ -\frac{\int_{s>y} sd\Gamma_1(s)}{(1 - G(y))\Gamma_0(y-)} - \frac{\alpha(y)\Gamma_1(y)}{\Gamma_0(y-)} \right\} dM_i(y) = S_0(\beta) + o_p(\sqrt{n}). \tag{2.18}
\]

**Proof of Theorem 2.** Let \( U_n \) denote the second summand on the left hand side of (2.18) and let

\[
H(y) = -\{ \int_{s>y} sd\Gamma_1(s) \} / \{(1 - G(y))\Gamma_0(y-)\} - \alpha(y)\Gamma_1(y) / \Gamma_0(y-).
\]

Let \( U_n(t) = \sum_{i=1}^{n} \int_{-\infty}^{t} H(y) dM_i(y) \), so that \( U_n = U_n(\tau) \). Then \( \{n^{-1/2}U_n(t), -\infty < t \leq \tau\} \) is a vector martingale with predictable variation process

\[
\langle n^{-1/2}U_n \rangle(t) = \int_{-\infty}^{t} (H(y))^{\otimes 2} (n^{-1} \sum_{i=1}^{n} I_{\{c_i \geq y, y_i > y\}})(1 - \Delta \Lambda(y)) d\Lambda(y)
\]

\[
P \left[ \int_{-\infty}^{t} (H(y))^{\otimes 2} \Gamma_0(y)(1 - \Delta \Lambda(y)) dG(y) \right] = V_2(t), \tag{2.19}
\]

10
by the law of large numbers, and therefore \( \{n^{-1/2}U_n(t), -\infty < t \leq \tau \} \) converges weakly to a Gaussian martingale by Rebolloed's martingale central limit theorem (cf. [2], p. 17). Hence \( U_n/\sqrt{n} = U_n(\tau)/\sqrt{n} \) has a limiting normal distribution with mean 0 and covariance matrix \( V_2(\tau) \) in view of (C2), where \( V_2(\cdot) \) is defined in (2.17).

Fix any \( t < \tau \) such that \( G \) is continuous at \( t \), and consider the joint distribution of \( n^{-1/2}(S_n, G, U_n(t)) \). Note that \( \| \int_{-\infty}^{t} H(y)dM_j(y) \| \) is uniformly bounded with \( E \int_{-\infty}^{t} H(y)dM_j(y) = 0 \), and that \( \delta_i(1 - \delta_i) = 0 \), from which we obtain

\[
\frac{1}{n} E \left\{ \sum_{i=1}^{n} \frac{\delta_i x_i \bar{y}_i}{1 - G(\bar{y}_i -)} U_n(t)' \right\} = - \int_{-\infty}^{t} E \left\{ \frac{1}{n} \sum_{i=1}^{n} \frac{\delta_i x_i \bar{y}_i}{1 - G(\bar{y}_i -)} I_{\{\delta_i \geq y, y_i > y\}} (H(y))' d\Lambda(y) \right\}
\]

\[
= \int_{-\infty}^{t} \left\{ \sum_{s > y} s d \Gamma_1(s) \right\} (H(y))'(1 - G(y-))^{-1} dG(y),
\]

by (C1). Similarly it follows from (C1) that

\[
\frac{1}{n} E \left\{ \sum_{i=1}^{n} (1 - \delta_i)x_i \alpha(\bar{y}_i) U_n(t)' \right\} - \int_{-\infty}^{t} \left\{ \alpha(y) \Gamma_1(y) + \frac{\int_{s > y} \alpha(s) \Gamma_1(s) dG(s)}{1 - G(y-)} \right\} (H(y))' dG(y).
\]

Moreover, using (C1) and integration by parts, it can be shown that

\[
- \frac{1}{n} E \sum_{i=1}^{n} \left\{ \delta_i \int_{s < \bar{y}_i} \frac{\alpha(s) dG(s)}{1 - G(s-)} + (1 - \delta_i) \int_{s \geq \bar{y}_i} \frac{\alpha(s) dG(s)}{1 - G(s-)} \right\} x_i U_n(t)'
\]

\[
= - \int_{-\infty}^{t} \left\{ \int_{u \geq y} \alpha(u) \Gamma_1(u) dG(u) \right\} (H(y))' d\Lambda(y) - \int_{-\infty}^{t} \alpha(y) \Gamma_1(y) \Delta G(y)(H(y))' d\Lambda(y).
\]

In view of (2.9), adding (2.20), (2.21) and (2.22) yields

\[
\frac{1}{n} E(S_n, G U_n(t)') = \int_{-\infty}^{t} \left\{ \int_{s > y} s d \Gamma_1(s) \right\} (H(y))' \frac{1 - G(y)}{1 - G(y-)} dG(y)
\]

\[
= - \int_{-\infty}^{t} H(y)(H(y))' \Gamma_0(y-) (1 - \Delta \Lambda(y)) dG(y) = - V_2(t),
\]

noting that \( \Gamma_0(y-) \Delta G(y) = \Gamma_0(y) \Delta G(y) \) by (C2). Since \( (S_n', G, U_n(t)')' \) is a sum of independent zero-mean random vectors satisfying the Lindeberg condition in view of (C4) and the uniform boundedness of \( \int_{-\infty}^{t} H(y)dM_j(y) \), the central limit theorem for sums of independent random
vectors gives the limiting normal distribution of $n^{-1/2}(S_{n,G}, U_n(t))'$. Letting $t \to \tau$ then gives the limiting joint distribution of $n^{-1/2}(S_{n,G} + U_n(\tau))$ with covariance matrix

$$
\lim_{n \to \infty} E(n^{-1/2}S_{n,G}) \otimes E(n^{-1/2}U_n(\tau)) + 2 \lim_{n \to \infty} n^{-1} E(S_{n,G}U_n(\tau)')
= V_1 + V_2(\tau) - 2V_2(\tau) = V_1 - V_2(\tau), \text{ by } (2.23).
$$

(2.24)

In view of Lemma 1, the desired conclusion on the limiting distribution of $S_n(\beta)/\sqrt{n}$ follows. This and (C6) in turn give the desired limiting normal distribution for $\sqrt{n}(\hat{\beta}_n - \beta) = (n^{-1} \sum_{i=1}^{n} x_ix_i')^{-1} S_n(\beta)/\sqrt{n}$. 

**Remarks.** (i) If $G$ is known and one uses (2.5) and (2.9) to form the synthetic least squares estimator $\hat{\beta}_G$ defined in (2.2), then $\sqrt{n}(\hat{\beta}_G - \beta) = (n^{-1} \sum_{i=1}^{n} x_ix_i')^{-1} S_{n,G}/\sqrt{n}$ and therefore under (C4) and (C6), $\sqrt{n}(\hat{\beta}_G - \beta)$ has a limiting normal distribution with mean 0 and covariance matrix $A^{-1}V_1A^{-1}$, since $V_1 = \lim_{n \to \infty} E(n^{-1/2}S_{n,G}) \otimes$. The adaptive estimator (2.3) that replaces $G$ in $\hat{\beta}_G$ by $\hat{G}_n$ turns out to have a smaller covariance matrix because of (2.24). Moreover, an analogous result holds if one uses (2.15) instead of (2.9), as in Theorem 3. In the special case of simple linear regression, this phenomenon of variance reduction by substituting $G$ in $\hat{\beta}_G$ with $\hat{G}_n$ was also noted by Koul, Susarla and Van Ryzin [6] and by Srinivasan and Zhou [12] for the case $\alpha \equiv 0$, and by Zhou [16] for Leurgans’ estimator (2.3)–(2.4).

(ii) Since $\hat{\beta}_n$ is a consistent estimator of $\beta$ and since $S_n(\theta)$ is linear in $\theta$, the limiting covariance matrix $V$ of $n^{-1/2}S_n(\beta)$ can be consistently estimated under the assumptions of Theorem 2 or Theorem 3 by $n^{-1} \Sigma_n$, where

$$
\Sigma_n = \sum_{i=1}^{n} \{\delta_i \psi_{\hat{G}_n}(\tilde{y}_i) + (1 - \delta_i) \psi_{\hat{G}_n}(\tilde{y}_i) - \hat{\beta}_n x_i)^2 x_ix_i'.
$$

Suppose that $V$ is positive definite and let $X_n = \sum_{i=1}^{n} x_ix_i'$. Then by Theorem 2 (or 3), a large-sample $(1 - \alpha)$-level confidence region for $\beta$ is $\{\beta : (\hat{\beta}_n - \beta)'X_n \Sigma_n^{-1} X_n(\hat{\beta}_n - \beta) \leq \chi^2_{p,\alpha}\}$, where $\chi^2_{p,\alpha}$ is the $(1 - \alpha)$th quantile of the chi-square distribution with $p$ degrees of freedom.

**Proof of Lemma 1.** For the assertion on $M_i$, see [2], pages 12 and 26. Let $\hat{\lambda}_n(t) = \int_{s \leq t} dN_n(s)/Y_n(s)$, where $N_n$ and $Y_n$ are defined in (2.10). Then $S_n(\beta) - S_{n,G}$ can be expressed
\[
\sum_{i=1}^{n} x_i \delta_i \bar{y}_i \{ \hat{G}_n(\bar{y}_i -) - G(\bar{y}_i -) \}/((1 - \hat{G}_n(\bar{y}_i -))(1 - G(\bar{y}_i -))) \\
- \sum_{i=1}^{n} x_i \int_{-\infty}^{\infty} \{ \delta_i I_{\{ t < \bar{y}_i \}} + (1 - \delta_i) I_{\{ t \leq \bar{y}_i \}} \} \alpha(t) d(\hat{\Lambda}_n(t) - \Lambda(t)) = (I) - (II),
\]

where by Eq. (3.2.13) of [2] and by the definitions of \( \hat{\Lambda}_n, \Lambda \) and \( M_j \),

\[(I) = \sum_{i=1}^{n} x_i \delta_i \bar{y}_i \int_{t \leq \bar{y}_i} \frac{1 - \hat{G}_n(\bar{y}_i -)}{1 - G(t)} \sum_{j=1}^{n} dM_j(t) \frac{dM_j(t)}{Y_n(t)}, \tag{2.25}
\]

\[(II) = \sum_{j=1}^{n} \sum_{i=1}^{n} \int_{-\infty}^{\infty} \{ \delta_i x_i I_{\{ \bar{y}_i > t \}} + (1 - \delta_i) x_i I_{\{ \bar{y}_i \geq t \}} \}(\alpha(t)/Y_n(t)) dM_j(t). \tag{2.26}
\]

Fix \( u < \tau \). Since \( \sup_{t \leq u} |n^{-1} Y_n(t) - (1 - G(t-)) \Gamma_0(t-) - G(t-) \Gamma_0(t-) = O_p(n^{-1/2}) \) and since \( Y_n(s) \) is left continuous while \( \{(\delta_i, x_i, \bar{y}_i), i \neq j\} \) is independent of \( \{ M_j(t), -\infty < t < \infty \} \), it follows from (C1) and Lenglart's inequality (cf. [2], p. 18) that

\[
\sum_{j=1}^{n} \int_{-\infty}^{u} \left\{ \sum_{1 \leq i \leq n, i \neq j} \delta_i x_i I_{\{ \bar{y}_i > t \}} + (1 - \delta_i) x_i I_{\{ \bar{y}_i \geq t \}} \right\} \left( \frac{\alpha(t)}{Y_n(t)} \right) dM_j(t)
\]

\[= \sum_{j=1}^{n} \int_{-\infty}^{u} \left\{ - \int_{s > t} (1 - G(s)) d\Gamma_1(s) + \int_{s \geq t} \Gamma_1(s) dG(s) \right\} \alpha(t)((1 - G(t-)) \Gamma_0(t-) - G(t-))^{-1} dM_j(t)
\]

\[+ o_p(\sqrt{n})
\]

\[= \sum_{j=1}^{n} \int_{-\infty}^{u} \{ (1 - G(t)) \Gamma_1(t) + \Gamma_1(t) \Delta G(t) \} \{ (1 - G(t-)) \Gamma_0(t-) - G(t-))^{-1} \alpha(t) dM_j(t) + o_p(\sqrt{n})
\]

\[= \sum_{j=1}^{n} \int_{-\infty}^{u} \{ \alpha(t) \Gamma_1(t)/\Gamma_0(t-) \} dM_j(t) + o_p(\sqrt{n}),
\]

where we have used the assumption in (C2) that \( G \) and \( \Gamma_1 \) have no common jumps. Furthermore, note that \( dM_j(t) = (1 - \delta_j) dI_{\{ \bar{y}_i \leq u \}} - I_{\{ \bar{y}_i > t, c_j \geq t \}} d\Lambda(t) \) and that by (C1),

\[
\sum_{j=1}^{n} (1 - \delta_j) x_j \alpha(\bar{y}_j) I_{\{ \bar{y}_j \leq u \}}/Y_n(\bar{y}_j) - \sum_{j=1}^{n} \int_{-\infty}^{u} x_j I_{\{ y_j > t, c_j \geq t \}}(\alpha(t)/Y_n(t)) d\Lambda(t)
\]

\[\overset{P}{\to} \int_{-\infty}^{u} \alpha(t) \Gamma_1(t)\{ (1 - G_0(t-)) \Gamma_0(t-) \}^{-1} dG(t) - \int_{-\infty}^{u} \alpha(t) \Gamma_1(t) (\Gamma_0(t-))^{-1} d\Lambda(t) = 0.
\]
Suppose we replace \( \int_{-\infty}^{\infty} \) in (II) by \( \int_{-\infty}^{u} \) and denote the result by (II)_u. The preceding argument has shown that
\[
(II)_u = \sum_{j=1}^{n} \int_{-\infty}^{u} \left\{ \alpha(t) \Gamma_1(t)/\Gamma_0(t-) \right\} dM_j(t) + o_P(\sqrt{n}).
\]

Similarly, since \( \sup_{t \leq u} \left| (1 - \hat{G}_n(t))/(1 - G(t)) - 1 \right| \to 0 \) (cf. [2], p. 56), it can be shown that
\[
\sum_{j=1}^{n} \int_{-\infty}^{u} \left\{ \sum_{i=1}^{n} \frac{x_i \delta_i \hat{y}_i}{1 - \hat{G}_n(\hat{y}_i-)} I_{\{\hat{y}_i > t\}} \right\} \left\{ \frac{1 - \hat{G}_n(t-)}{1 - G(t)} \right\} dM_j(t) Y_n(t)
\]
\[
= - \sum_{j=1}^{n} \int_{-\infty}^{u} \left\{ \int_{s > t} s d\Gamma_1(s) \right\} \{(1 - G(t))\Gamma_0(t-)\}^{-1} dM_j(t) + o_P(\sqrt{n}).
\]

Hence the desired conclusion follows if it can be shown that for every \( a > 0 \), as \( u \to \tau^- \) and \( n \to \infty \),
\[
P \left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\delta_i ||x_i|| \|\hat{y}_i\|}{1 - \hat{G}_n(\hat{y}_i-)} \left| \int_{u < t < \hat{y}_i} \frac{1 - \hat{G}_n(t-)}{1 - G(t)} \sum_{j=1}^{n} \frac{dM_j(t)}{Y_n(t)} \right| \geq a \right\} \to 0, \quad (2.27)
\]
\[
P \left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{||x_i||}{\sqrt{n}} \left[ \left| \int_{u < t < \hat{y}_i} \frac{\alpha(t)}{Y_n(t)} \sum_{j=1}^{n} dM_j(t) \right| + \left| \int_{u < t < \hat{y}_i} \frac{\alpha(t)}{Y_n(t)} \sum_{j=1}^{n} dM_j(t) \right| \right] \geq a \right\} \to 0. \quad (2.28)
\]

To prove (2.27), let \( T_n = \max_{1 \leq i \leq n} \hat{y}_i \) and let \( 0 < \eta < \epsilon^{-2}/5 \). By Theorem 1.1.1 and Corollary 1.3.1 of [13],
\[
P\{n^{-1}Y_n(s) \leq \eta^{-1}(1 - G(s-))(1 - \mu_{n,Y}(s-)) \text{ for all } s \} \geq 1 - \frac{2}{3} \pi^2 \eta(1 - \eta)^{-4}, \quad (2.29)
\]
\[
P\{n^{-1}Y_n(s) \geq \eta(1 - G(s-))(1 - \mu_{n,Y}(s-)) \text{ for all } s \leq T_n \} \geq 1 - \frac{2}{3} \pi^2 \eta^2(1 - \eta)^{-4}. \quad (2.30)
\]

By Theorem 3.2.1 of [2] and Theorem 2.2 of [15] (whose continuity assumption on the underlying distribution functions can in fact be dropped),
\[
P\{1 - \hat{G}_n(s) \leq \eta^{-1}(1 - G(s)) \text{ for all } s \leq T_n \} \geq 1 - \eta. \quad (2.31)
\]
\[
P\{1 - \hat{G}_n(s) \geq \eta(1 - G(s)) \text{ for all } s < T_n \} \geq 1 - 358|\log(5\eta)|^{-2/3}. \quad (2.32)
\]

14
Since $1 - \hat{G}_n(t-)$ and $Y_n(t)$ are left continuous, we can make use of Lenglart's inequality (cf. [2], p. 18) together with (2.29)-(2.32) and Lemma 2.9 of [3] to show that for $\tau' < \tau$,

$$
\sup_{T_n \geq u > v \geq \tau'} (1 - G(v)) (1 - \mu_n, Y(v))^{1/2} \left| \int_{u < t < v} \frac{1 - \hat{G}_n(t-)}{1 - G(t)} \sum_{j=1}^n \frac{dM_j(t)}{Y_n(t)} \right| = O_p(n^{-1/2}). \tag{2.33}
$$

By (2.12) and (C3), for $u \geq y(\varepsilon)$ and all large $n$,

$$
E \sum_{i=1}^n \frac{\delta_i \| x_i \| \| \hat{y}_i \|}{\sqrt{n} (1 - G(\hat{y}_i - ))} \left\{ \frac{I_{\{ \hat{y}_i > u \}}}{\sqrt{n} (1 - G(\hat{y}_i))(1 - \mu_n, Y(\hat{y}_i))^{1/2}} \right\} < \varepsilon. \tag{2.34}
$$

From (2.32)-(2.34), (2.27) follows.

Using (C3), (2.30) and Lenglart’s inequality, it can be shown that for every $\delta > 0$ there exists $u_\delta < \tau$ such that for all $u \geq u_\delta$ and all large $n$,

$$
P \left\{ \sup_{u < v \leq T_n} \left[ \left| \int_{u < t < v} \frac{\alpha(t)}{Y_n(t)} \sum_{j=1}^n dM_j(t) \right| + \left| \int_{u < t \leq v} \frac{\alpha(t)}{Y_n(t)} \sum_{j=1}^n dM_j(t) \right| \right] > \frac{\delta}{\sqrt{n}} \right\} < \delta. \tag{2.35}
$$

Since $n^{-1} \sum_{i=1}^n \| x_i \| \leq \{n^{-1} \text{tr}(\sum_{i=1}^n x_i x_i')\}^{1/2}$, (2.28) follows from (C6) and (2.35).

3. VON MISES CALCULUS AND ADAPTIVE STATISTICS

In this section we first prove Theorem 1 and then consider adaptive statistics of the form $\sum_{i=1}^n \xi_0(w_i, \hat{H}_n)$ in the framework of von Mises calculus for statistical functionals, which will be shown to lead directly to an approximation of these statistics by much simpler $U$-statistics.

Proof of Theorem 1. In view of (1.6), we can apply the central limit theorem for sums of independent random vectors to conclude that $n^{-1/2} \sum_{i=1}^n \{\xi(w_i, \theta_0, H) + w_i^*\}$ has a limiting normal distribution with mean 0 and covariance matrix $V$. From this and (1.5), (i) follows.

By (1.9) and (1.10),

$$
\sum_{i=1}^n \xi(w_i, \theta, \hat{H}_{n, \theta}) = \sum_{i=1}^n \xi(w_i, \theta_0, \hat{H}_n) + nA(\theta - \theta_0) + e_n(\theta), \tag{3.1}
$$

where $\sup_{\| \theta - \theta_0 \| \leq n^{-1/2 + \delta}} \| e_n(\theta) \| / (\sqrt{n} + n\| \theta - \theta_0 \|) \to 0$. From (3.1) and (1.12), it follows that $\| \hat{\theta}_n - \theta_0 \| = O_p(n^{-1/2})$ and that

$$
\sqrt{n}(\theta - \theta_0) = -A^{-1} \sum_{i=1}^n \xi(w_i, \theta_0, \hat{H}_n) / \sqrt{n} + o_p(n^{-1/2}). \tag{3.2}
$$
Combining (3.2) with (i) yields (ii).

In Section 2, we have applied Theorem 1(i) with $\xi$ given by (2.11) to establish the asymptotic normality of the synthetic estimator defined by (2.3), (2.9) (or (2.15)) and (2.10). Since the estimating equation is linear in $\theta$, it has the explicit solution (2.3) and we do not need Theorem 1(ii) that gives the asymptotic linearity property (3.1). In [7], Lai and Ying established the asymptotic normality of a modified version of the Buckley-James estimator by using both asymptotic linearity (as in Theorem 1(ii)) and the asymptotic normality result of Theorem 1(i). In the sequel we shall let $\mathcal{F}$ denote the vector space of right continuous functions $F : \mathbb{R} \to \mathbb{R}$ of bounded variation and such that $\lim_{y \to x^-} F(y)$ exists at every $x$, and define the norm $\|F\| = \sup_x |F(x)|$. Note that the $F_n, \theta$ in (1.2b) and the $\hat{G}_n$ in (2.10) belong to $\mathcal{F}$. For a linear functional $\ell$ on $\mathcal{F}$ taking values in $\mathbb{R}^p$, we shall also denote $\ell(F)$ by $\ell \cdot F$.

As pointed out in Section 1, to find the independent random variables $w^*_i$ in (1.5), one can proceed by formally differentiating $\xi(w, \theta_0, H)$ with respect to $H$, as in (1.7), and by replacing $n^{-1} \sum_{i=1}^n D\xi(w_i)$ by a limiting linear operator, as in (1.8). For example, for $\xi$ given by (2.11) and (2.9) with $\sup \{\alpha(t)\} < \infty$, letting $\xi(w, \beta, H) = \xi_0((x_i, y_i, \delta_i), H)$, the derivative $D\xi_0((x_i, y_i, \delta_i))$ at $G$ is the linear functional on $\mathcal{F}$ defined by

$$D\xi_0((x_i, y_i, \delta_i)) \cdot F = x_i \delta_i \left\{ \frac{y_i F(y_i)}{1-G(y_i)} - \int_{s \leq y_i} \frac{\alpha(s)F(s)-}{(1-G(s))} dG(s) \right\}$$

$$- x_i (1 - \delta_i) \int_{s \leq y_i} \left( \frac{\alpha(s)F(s)}{1-G(s)} \right)^2 dG(s) + \frac{\alpha(s)}{1-G(s)} dF(s), \quad F \in \mathcal{F},$$

as can be seen by expressing $\{\xi_0(w, G + tF) - \xi_0(w, G)\}/t$ in terms of (2.9) and letting $t \to 0$.

Proceeding as in (1.7) then yields the approximation

$$\sum_{i=1}^n \xi_0((x_i, y_i, \delta_i), \hat{G}_n) - \sum_{i=1}^n \xi_0((x_i, y_i, \delta_i), G) \approx \sum_{i=1}^n D\xi_0((x_i, y_i, \delta_i)) \cdot (\hat{G}_n - G).$$

(3.4)

Suppose that $(x_i, y_i, c_i)$ are i.i.d. Proceeding as in (1.8) is tantamount to replacing the right hand side of (3.3) by its expectation $\ell \cdot F$, where $\ell$ is some linear functional. Using i.i.d. representations of $\hat{G}_n - G$ given by Lo and Singh [10] and Gu and Lai [5], we can represent $n \ell \cdot (\hat{G}_n - G)$ as a sum of i.i.d. random variables, which give the $w^*_i$ in (1.5).

It is difficult to give a rigorous justification of the preceding arguments and to extend them to the case of non-identically distributed $(x_i, y_i, c_i)$ because $\psi_G(y)$ and $\Psi_G(y)$ are not
well-behaved functionals of $G$ when $G(y-)$ is near 1 and because $\hat{G}_n(y)$ is also not a well-behaved functional of empirical sub-distribution functions when $y$ is near $\tau$. For censored data, a much more powerful technique to analyze the product-limit estimator and functionals thereof is to use martingale representations such as (2.25). Therefore in Lemma 1 we have derived the approximation (1.5) from these martingale representations instead of trying to justify (1.7) and (1.8) which would require much more stringent assumptions.

For the functional $\zeta((z, \delta); F)$ in (1.2a) defining the Buckley-James estimator, the formal derivative $D\zeta((z, \delta))$ at $F$ is the linear functional defined by

$$
D\zeta((z, \delta)) \cdot K = (1 - \delta) \left\{ \frac{K(z)}{(1 - F(z))^2} \int_z^\infty (1 - F(s)) ds - \frac{1}{1 - F(z)} \int_z^\infty K(s) ds \right\},
$$

for $K(\in \mathcal{F})$ that is integrable with respect to Lebesgue measure on $(z, \infty)$ for every $z$. The analysis of the Buckley-James estimator in [7] gets around the singularities of the functional (1.2a) of $F$ when $F(z)$ is near 1 by using martingale representations similar to those in the proof of Lemma 1. Furthermore, the proof of the asymptotic linearity property (3.1) in this case relies on using modern empirical process theory to first approximate $\hat{H}_{n, \theta}(= \hat{F}_{n, \theta})$ and $\check{H}_{n}(= \check{F}_{n, \theta})$ by nonrandom functions $H_{n, \theta}$ and $H$, and then to approximate $\sum_{i=1}^n \xi(w_i, \theta, H_{n, \theta})$ and $\sum_{i=1}^n \xi(w_i, \theta, H)$ by nonrandom functions that have the desired asymptotic linearity property, from which the conclusion (3.1) follows.

For complete (instead of censored) data, a typical choice of $\hat{H}_n$ in the adaptive statistic $\sum_{i=1}^n \xi_0(w_i, \hat{H}_n)$ is the empirical distribution function based on observations $u_1, \ldots, u_n$ such that $(w_1, u_1), \ldots, (w_n, u_n)$ are independent. To fix the ideas, we shall assume that the $u_i$ are one-dimensional random variables, as extension to the multivariate case is straightforward. Proceeding heuristically as in (1.7) suggests that the first step in proving asymptotic normality of $\sum_{i=1}^n \xi_0(w_i, \hat{H}_n)$ is to check the approximation

$$
\sum_{i=1}^n \xi_0(w_i, \hat{H}_n) = \sum_{i=1}^n \xi_0(w_i, H) + \sum_{i=1}^n \lambda(w_i) \cdot (\hat{H}_n - H) + o_p(\sqrt{n}), \tag{3.5}
$$

where $\lambda(w_i)$ is a linear operator taking values in $\mathbb{R}^p$. The form of the linear operator $\lambda(w_i)$ can be obtained by formal differentiation of $\xi_0(w_i, H)$ with respect to $H$, i.e., evaluating

$$
\lambda(w_i) \cdot F = \lim_{t \to 0} \{ \xi_0(w_i, H + tF) - \xi_0(w_i, H) \} / t \tag{3.6}
$$

17
for suitably chosen $F$. However, instead of embedding $H$ and $\hat{H}_n$ into some normed vector space $B$ and extending $\xi_0$ to a function from $B$ into $\mathbb{R}^p$ that is Hadamard differentiable at $H$, which requires the convergence in (3.6) to be uniform in $F \in C$ for every compact subset $C$ of $B$ (cf. [4]), it is usually more convenient to check (3.5) directly once the form of $\lambda(w_i)$ is obtained via formal von Mises calculus. We need only check that $\sum_{i=1}^n \xi_0(w_i, \hat{H}_n) - \sum_{i=1}^n \xi_0(w_i, H) - \sum_{i=1}^n \lambda(w_i) \cdot (\hat{H}_n - H)$ is of the order $o_p(\sqrt{n})$, and can do this by using properties not only of $\xi_0$ but also of $\hat{H}_n - H$, e.g., $\sup_t |\hat{H}_n(t) - H(t)| = O_p(n^{-1/2})$. This direct approach also avoids the complication that in (3.5) we actually have $n$ functions $\xi_0(w_1, \cdot), \ldots, \xi_0(w_n, \cdot)$ from $B$ into $\mathbb{R}^p$, which would presumably require the convergence in (3.6) to be also uniform in the $w_i$ if we should use the functional analytic approach based on Hadamard differentiability.

In the case where $\hat{H}_n = n^{-1} \sum_{j=1}^n \delta(u_j)$ is the empirical distribution, where $\delta(u)$ puts unit mass at $u$, we can write

$$n \sum_{i=1}^n \lambda(w_i) \cdot (\hat{H}_n - H) = \sum_{i=1}^n \sum_{j=1}^n \lambda(w_i) \cdot (\delta(u_j) - H),$$

which can be expressed as a $U$-statistic of the form $\sum_{i=1}^n \sum_{j=1}^n h(v_i, v_j)$, with $v_i = (w_i, u_i)$ and $h(v_i, v_j) = \lambda(w_i) \cdot (\delta(u_j) - H)$. We can therefore apply directly the asymptotic theory of $U$-statistics without going through (1.8) that appeals to a law of large numbers to replace $n^{-1} \sum_{i=1}^n \lambda(w_i)$ by a limiting linear operator $\ell$. As pointed out in Section 1, much more stringent conditions on $\lambda(w_i)$ are needed in order to apply such law of large numbers. In particular, the $\lambda(w_i)$ would be required to belong to some separable normed vector space of linear operators and their norms would have to satisfy certain moment conditions. Therefore, to establish the asymptotic normality of $\sum_{i=1}^n \xi_0(w_i, \hat{H}_n)$, it is usually much easier to check (3.5) directly and to apply $U$-statistic theory to (3.7) than to proceed by a rigorous justification of (1.7) and (1.8).

REFERENCES


18


