Statistica Sinica 2(1992), 519-539

ASYMPTOTIC THEORY OF A BIAS-CORRECTED LEAST SQUARES ESTIMATOR IN TRUNCATED REGRESSION

Tze Leung Lai and Zhiliang Ying

Stanford University and University of Illinois

Abstract: Let $y = \beta^T x + \epsilon$ denote the intrinsic relation between the response y and a covariate vector x, where ϵ represents an unobservable random variable. A truncated regression model assumes the existence of another (truncation) variable t so that (x, y, t) is observed if and only if $t \leq y$ and that nothing is observed if $t > y$. Tsui, Jewell and Wu (1988) have proposed a bias-corrected method to extend the classical least squares approach to regression analysis with truncated data and have found the method to perform well in an extensive simulation study. To develop an asymptotic theory for this approach, we first introduce a slight modification of their estimator to make it more tractable and then establish the consistency and asymptotic normality of the modification under certain regularity conditions. By making use of the asymptotic normality result, approximate confidence regions for β are also given.

Key words and phrases: Asymptotic normality, bias-corrected estimator, consistency, counting process, linear regression, martingale, product-limit estimator, truncation.

1. Introduction

A regression model with incomplete (and therefore potentially biased) observations is the "truncated regression model" in the econometrics literature (cf. Tobin (1958), Goldberger (1981), Amemiya (1985)), in astronomy (cf. Segal (1975), Nicoll and Segal (1980)), and in biomedical studies (cf. Cox and Oakes (1984)). The model assumes the usual regression structure

$$y_j^* = \beta^T x_j^* + \epsilon_j^*, \; j = 1, 2, \dots, \tag{1.1}$$

between the response y_j^* and a $p \times 1$ vector x_j^* of covariates, with i.i.d. random errors ϵ_j^* having a common distribution function F (which need not have mean 0) such that $\int_{-\infty}^{\infty} t^2 dF(t) < \infty$. Suppose that (x_i^T, y_i, t_i) can be observed only when $y_i \geq t_i$, where (t_i, x_i^T) are independent random vectors that are independent of $\{\epsilon_i\}$. The t_i are called *(left) truncation variables.* A right truncated regression model can be similarly defined, with (x_i^T, y_i, t_i) observable only when $y_i \leq t_i$, but can be converted to a left truncated regression model by multiplying each variable by -1. Hence in the sequel we shall only consider a left truncated model, for which the observations are

$$(x_i^T, y_i, t_i) \text{ with } y_i \geq t_i, \; i = 1, \dots, n. \tag{1.2}$$

By assuming the underlying error distribution F to belong to a parametric family of distributions, β can be estimated by the method of maximum likelihood; and there is an extensive literature on the subject (cf. Amemiya (1985)). Without such parametric assumptions, Bhattacharya, Chernoff and Yang (1983) introduced, for the case of univariate x_i, an extension of Adichie's (1967) rank estimator of the slope parameter β for the linear regression model (1.1), based on the Wilcoxon score function, to the truncated regression model, and showed that, under certain regularity conditions, the estimator is consistent and asymptotically normal. A general rank-type approach was subsequently developed by Lai and Ying (1992a), where counting processes and their associated martingales are used as natural vehicles for studying rank estimators based on truncated data, in the same way as they have been used for censored data (cf. Gill (1980)).

As is well known, the classical least squares estimator of β is biased in the presence of truncation. A bias-corrected modification of the least squares estimator was recently introduced by Tsui, Jewell and Wu (1988). Their idea is to first construct $\hat{y}_i$ in such a way that $E(\hat{y}_i|x_i) \approx \beta^T x_i$, and then to regress the $\hat{y}_i$ on x_i to obtain an estimator of β. A more detailed description of their method will be given in the next section. In an extensive simulation study, they found the estimator to perform better than the Bhattacharya-Chernoff-Yang estimator. On the other hand, while there is a relatively complete asymptotic theory for the Bhattacharya-Chernoff-Yang estimator and for more general rank estimators based on truncated data (cf. Lai and Ying (1992a)), a corresponding theory for the bias-corrected least squares estimator is lacking, and it is therefore not possible to compare the asymptotic properties of the two approaches.

A similar bias-corrected modification of the least squares estimate has been introduced by Buckley and James (1979) for the case where the y_j^* in (1.1) are subject to censorship by (right) censoring variables c_j (so that the observed responses are $\min(y_j^*, c_j)$, $I_{\{y_j^* \leq c_j\}}$, $j = 1, \dots, n$). Although James and Smith (1984) have shown the Buckley-James estimator to be consistent under certain assumptions, there is a gap in their proof, as pointed out recently by Lai and Ying (1991b). To get around the difficulties with the Buckley-James estimator caused by the instability at the upper tail of the associated Kaplan-Meier estimator of the underlying error distribution, Lai and Ying (1991b) introduced a simple modification of the estimator and were able to prove that the modified estimator is indeed consistent and asymptotically normal under certain regularity conditions.

By extending the ideas developed in Lai and Ying (1991b) for censored data, we develop, in Section 3 below, a parallel large sample theory for a slight modification of the bias-corrected least squares estimator of Tsui, Jewell and Wu (1988). The modification, introduced in Section 2, enables us to get around certain difficulties with the product-limit estimator of the underlying error distribution used in the bias-corrected approach. Section 2 also provides basic stochastic integral representations of certain statistics involved, which enable us to apply martingale theory and empirical process theory to analyze these statistics. Another useful idea which makes these tools applicable is to regard the observations (x_i^T, y_i, t_i), $i = 1, \ldots, n$, as generated from a larger sample of independent random vectors (x_j^{*T}, y_j^*, t_j^*), $j = 1, 2, \ldots$, with random sample size

$$n^* = \inf\{m : \sum_{j=1}^{m} I_{\{t_j^* \le y_j^*\}} = n\}, \tag{1.3}$$

such that the observations (x_i^T, y_i, t_i) correspond to the (x_j^{*T}, y_j^*, t_j^*) with $t_j^* \le y_j^*$ (cf. Lai and Ying (1991a,1992a)). Section 4 makes use of the asymptotic theory developed for the bias-corrected least squares estimator to construct approximate confidence regions for β and also to make comparisons with the rank estimators of Bhattacharya, Chernoff and Yang (1983) and of Lai and Ying (1992a).

2. A Modification of the Bias-Corrected Least Squares Estimator and Related Stochastic Integrals

For notational convenience, let $y_i(b) = y_i - b^T x_i$, $t_i(b) = t_i - b^T x_i$, $y_i^*(b) = y_i^* - b^T x_i^*$ and $t_i^*(b) = t_i^* - b^T x_i^*$, regarding the observed $(x_i^T, y_i, t_i), i = 1, \ldots, n$, as generated from independent random vectors $(x_i^{*T}, y_i^*, t_i^*), i = 1, 2, \ldots$. The usual least squares estimator of β is the solution of

$$\sum_{i=1}^{n} x_i y_i(b) = 0, \tag{2.1}$$

and is biased in the presence of truncation since $E(y_i(\beta)|x_i) \ne 0$. Tsui, Jewell and Wu (1988) noted that

$$E(y_i(\beta)|t_i, x_i) = K_i(\beta) \quad \text{where} \tag{2.2}$$

$$K_i(b) = \int_{t_i(b)}^{\infty} u dF(u) / [1 - F(t_i(b))]. \tag{2.3}$$

They therefore proposed to replace (2.1) by

$$\sum_{i=1}^{n} x_i(y_i(b) - K_i(b)) = 0, \tag{2.4}$$

and to replace the unknown F in (2.3) by a product-limit estimator defined below.

To define the product-limit estimator and to analyze functionals thereof in the sequel, it is convenient to introduce the following empirical processes. Let

$$N(b,u) = \sum_{i=1}^{n} I_{\{y_i(b)\ge u\}}, \quad N^x(b,u) = \sum_{i=1}^{n} x_i I_{\{y_i(b)\ge u\}}; \tag{2.5}$$

$$J(b,u) = \sum_{i=1}^{n} I_{\{t_i(b)\le u\le y_i(b)\}}, \; J^x(b,u) = \sum_{i=1}^{n} x_i I_{\{t_i(b)\le u\le y_i(b)\}}; \tag{2.6}$$

$$S(b,u) = \sum_{i=1}^{n} I_{\{t_i(b)\le u\}}, \; S^x(b,u) = \sum_{i=1}^{n} x_i I_{\{t_i(b)\le u\}}. \tag{2.7}$$

Define n^* as in (1.3) and note that $N(b,u) = \sum_{i=1}^{n^*} I_{\{y_i^*(b)\ge u,\, t_i^*\le y_i^*\}}$, etc. The product-limit estimator of F based on the truncated residuals $y_i(b)$ is defined by

$$\hat{F}(b,u) = 1 - \prod_{i:y_i(b)\le u} \left(1 - \frac{1}{J(b, y_i(b))}\right). \tag{2.8}$$

Instead of replacing $F(u)$ by $\hat{F}(b,u)$ in both the numerator and denominator of (2.3), one can also estimate (2.3) directly by $\hat{K}_i(b) = \int_{v\ge t_i(b)} v d\hat{F}(b, v|t_i(b))$, where

$$\hat{F}(b,v|u) = 1 - \prod_{i:u<y_i(b)\le v} \left(1 - \frac{1}{J(b, y_i(b))}\right), \; v > u, \tag{2.9}$$

is the product-limit estimator of the conditional probability

$$F(v|u) = P(\epsilon_1^* \le v | \epsilon_1^* > u), \; v > u \tag{2.10}$$

(cf. Wang, Jewell and Tsai (1986), Lai and Ying (1991a)). Tsui, Jewell and Wu (1988) therefore suggested replacing the left hand side of (2.4) by

$$\begin{aligned}\tilde{\xi}(b) &= \sum_{i=1}^{n} x_i \Big\{ y_i(b) - \hat{K}_i(b) \Big\} \\ &= \sum_{i=1}^{n} x_i \Big\{ y_i(b) - \int_{v\ge t_i(b)} v d\hat{F}(b, v|t_i(b)) \Big\}. \end{aligned} \tag{2.11}$$

In view of (2.6) and (2.7), we can rewrite (2.11) in the form

$$\begin{aligned}\tilde{\xi}(b) &= \sum_{i=1}^{n} x_i \Big\{ y_i(b) - t_i(b) - \int_{t_i(b)}^{\infty} (1 - \hat{F}(b, v|t_i(b)))dv \Big\} \\ &= -\int_{-\infty}^{\infty} u dJ^x(b,u) - \int_{-\infty}^{\infty} \Big\{ \int_u^{\infty} (1 - \hat{F}(b,v|u))dv \Big\} dS^x(b,u). \quad (2.12)\end{aligned}$$

Since $\tilde{\xi}(b)$ is, in general, a discontinuous function of b, we cannot define, in analogy with (2.4), the bias-corrected least squares estimator $\tilde{\beta}$ simply as a root of $\tilde{\xi}(b)$. It will be assumed in the sequel that an upper bound $\rho > \|\beta\|$ is known, where $\|\cdot\|$ denotes the Euclidean norm. A natural analogue of (2.4), therefore, is to define $\tilde{\beta}$ as a minimizer of $\|\tilde{\xi}(b)\|$ with $\|b\| \le \rho$. For the univariate case $p = 1$, it is also possible to define $\tilde{\beta}$ as a zero-crossing of $\tilde{\xi}(b)$. An iterative algorithm for computing $\tilde{\beta}$ has been provided by Tsui, Jewell and Wu (1988).

The discontinuous random function (2.12) appears to be quite intractable. Even at $b = \beta$, the product-limit estimator $\hat{F}(\beta, v|u)$ of $F(v|u)$ is known to be rather unstable when the relative frequency $n^{-1}J(\beta, v)$ is small. Similar difficulties also arise in the analysis of the Buckley-James (1979) estimator for censored regression data. To get around these difficulties, Lai and Ying (1991b) introduced a slight modification of the Buckley-James estimator by using a smooth weight function to dampen the instability at the upper tail of the Kaplan-Meier estimate of F based on the censored residuals.

To define an analogous modification in the case of truncated data, let p be a twice continuously differentiable and nondecreasing function on the real line such that

$$p(u) = 0 \text{ for } u \le 0, \; p(u) = 1 \text{ for } u \ge 1. \quad (2.13)$$

With $S(b,u)$ and $N(b,u)$ defined in (2.7) and (2.5), define weight functions

$$p_1(b,u) = p\Big((\frac{S(b,u)}{n} - \frac{c_1}{\log n}) \log n \Big), \; p_2(b,u) = p\Big((\frac{N(b,u)}{n} - \frac{c_2}{\log n}) \log n \Big), \quad (2.14)$$

where c_1, c_2 are positive constants. As an extension of the ideas of Lai and Ying (1991b) for censored data to be explained below, we modify $\tilde{\xi}(b)$ in (2.12) by

$$\begin{aligned}\hat{\xi}(b) = &-\sum_{i=1}^{n} x_i \int_{-\infty}^{\infty} u p_1(b, t_i(b)) d\big[I_{\{t_i(b) \le u \le y_i(b)\}} p_2(b,u) \big] \\ &- \int_{-\infty}^{\infty} \Big\{ \int_u^{\infty} (1 - \hat{F}(b,v|u)) p_2(b,v) dv \Big\} p_1(b,u) dS^x(b,u). \quad (2.15)\end{aligned}$$

In analogy with (2.6), let

$$J_{p_1}^x(b,u)=\sum_{i=1}^n x_i p_1(b,t_i(b))I_{\{t_i(b)\le u\le y_i(b)\}}. \tag{2.16}$$

Then the first term on the right hand side of (2.15) is $-\int_{-\infty}^{\infty} ud[p_2(b,u)J_{p_1}^x(b,u)]$. By letting $\Delta p_2(b,u)=p_2(b,u+)-p_2(b,u)$, we can express (2.15) in the form

$$\begin{aligned}\hat{\xi}(b)=&\sum_{i=1}^n x_i p_1(b,t_i(b))\Big\{y_i(b)p_2(b,y_i(b))-t_i(b)p_2(b,t_i(b))\\&-\int_{t_i(b)}^{\infty}[1-\hat{F}(b,v|t_i(b))]p_2(b,v)dv\Big\}\\&-\sum_{i=1}^n y_i(b)J_{p_1}^x(b,y_i(b))\Delta p_2(b,y_i(b)),\end{aligned} \tag{2.17}$$

which shows more clearly how the Tsui-Jewell-Wu statistics (2.11) are modified.

The basic idea underlying (2.15) is the same as that introduced in Lai and Ying (1991b) to modify the Buckley-James statistics for censored data, but the details are quite different. In the censored case where the y_j^* in (1.1) are censored by censoring variables c_j, the risk set size $Z(b,u)=\sum_{i=1}^n I_{\{y_i^*(b)\wedge(c_i-bx_i^*)\ge u\}}$ is small relative to n only at the upper tail of the Kaplan-Meier curve based on the residuals. Here, for truncated data, the product-limit estimator is unstable at both the left and right tails corresponding to small values of $n^{-1}J(b,u)$. Hence, instead of a single weight function of the form $p^*(b,u)=p(n^\lambda\{n^{-1}Z(b,u)-cn^{-\lambda}\})$ with $c>0$, $0<\lambda<1$, used in Lai and Ying (1991b) for censored data, we use, here, two weight functions p_1 and p_2 for truncated data. As explained in Lai and Ying (1991b), in view of (2.13), the weight function $p^*(b,u)$ is a smooth analog of the standard trimming function $I_{\{Z(b,u)\ge cn^{1-\lambda}\}}$; and, likewise, the weight functions p_1 and p_2 in (2.15) are introduced to restrict the range of integration only to u and v for which the risk set sizes $J(b,u)$ and $J(b,v)$ are not too small relative to n.

Analogous to Lemma 1 of Lai and Ying (1991b), it will be shown in the Appendix that for any nonrandom functions g_1 and g_2 of bounded variation on the real line,

$$\begin{aligned}E\Big\{&\int_{-\infty}^{\infty} ug_1(t_i^*(\beta))d\big[g_2(u)I_{\{t_i^*(\beta)\le u\le y_i^*(\beta)\}}\big]\\+&\int_{-\infty}^{\infty}\big[\int_u^{\infty}(1-F(v|u))g_2(v)dv\big]g_1(u)dI_{\{t_i^*(\beta)\le u,t_i^*(\beta)\le y_i^*(\beta)\}}|x_i^*\Big\}=0,\end{aligned} \tag{2.18}$$

for $i=1,2,\ldots$; and this suggests, as in Lai and Ying (1991b), that the modification (2.15) does not introduce bias through the trimming operation (at least when $b = \beta$), (cf. (3.8) of Section 3). As pointed out in Lai and Ying (1991b), an important idea to remove the thorny issues of bias in performing such trimming is to compensate the bias introduced in the second term on the right hand side of (2.15), for which such trimming is needed, by a corresponding adjustment in the first term. Without having to worry about bias, we can in fact trim quite substantially to ensure a moderate risk set size $J(b,u)$; and here, we require it to be at least of the order $n/\log n$, although for the censored case treated in Lai and Ying (1991b), a risk set size $Z(b,u)$ of the order $n^{1-\lambda}$ suffices.

As explained in Lai and Ying (1988,1991b), the use of a smooth weight function instead of straightforward trimming is analogous to the kernel method in density estimation. Since a key idea in the analysis of $\hat{\xi}(b)$ is to approximate it by a nonrandom function $\xi_m(b)$, our use of smooth trimming functions p_1, p_2 leads to a smooth $\xi_m(b)$ which is essential to the analysis of asymptotic properties in the next section.

3. Large Sample Properties

In this section we discuss the asymptotic properties of the modified bias-corrected statistics $\hat{\xi}(b)$, defined by (2.15), and the corresponding estimator $\hat{\beta}$, defined as a minimizer of $\|\hat{\xi}(b)\|$ with $\|b\| \leq \rho$. Specifically, our results show that under certain regularity conditions, $\hat{\beta}$ is consistent and asymptotically normal. The proof of these results uses the same ideas as those developed in Lai and Ying (1991b) for the censored case. Our development consists of two main steps in the analysis of $\hat{\xi}(b)$, represented by Theorems 1 and 2 below. These two steps are then combined to provide the desired consistency and asymptotic normality of $\hat{\beta}$ in Theorem 3.

Theorem 1 establishes the asymptotic normality of $\hat{\xi}(\beta)$ by using martingale central limit theorems and a decomposition of $\hat{\xi}(\beta)$ in Lemma 2 below. Lemma 1 first reviews a basic martingale structure for truncated data that has been introduced in Lai and Ying (1991a,1992a). This martingale structure enables us to apply martingale central limit theorems to prove the asymptotic normality of $\hat{\xi}(\beta)$. As pointed out in Section 1, we shall regard the observed (x_i^T, y_i, t_i), $i = 1, \ldots, n$, as generated by a larger sample $(x_i^{*T}, y_i^*, t_i^*), i = 1, \ldots, n^*$, where n^* is defined in (1.3).

Lemma 1. *Let $\mathcal{F}_s$ be the complete σ-field generated by*

$$I_{\{t_i^* \leq y_i^*\}},\ I_{\{t_i^* \leq u \leq y_i^*\}},\ I_{\{t_i^* \leq y_i^* \leq u\}},\ t_i^*,\ x_i^*\ (u \leq s, i = 1, 2, \ldots).$$

Let $Y(s) = \sum_{i=1}^{n^*} I_{\{t_i^*(\beta) \leq y_i^*(\beta) \leq s\}}$ and $\Lambda(u) = -\log(1 - F(u))$, where F is the common continuous distribution function of $y_i^*(\beta)$. Define

$$W(s) = Y(s) - \int_{-\infty}^{s} J(\beta, u) d\Lambda(u), \quad -\infty < s < \infty. \tag{3.1}$$

Then $\{W(s), \mathcal{F}_s, -\infty < s < \infty\}$ *is a martingale with predictable variation process*

$$\langle W \rangle (s) = \int_{-\infty}^{s} J(\beta, u) d\Lambda(u).$$

Lemma 2. *Define* $\hat{\xi}(b)$ *as in* (2.15) *and* $J_{p_1}^x$ *as in* (2.16). *Then*

$$P\{\hat{\xi}(\beta) = \hat{\xi}_1 + \hat{\xi}_2 \text{ for all large } n\} = 1, \tag{3.2}$$

where

$$\begin{aligned}
\hat{\xi}_1 = &- \int_{-\infty}^{\infty} u d[p_2(\beta, u) J_{p_1}^x(\beta, u)] \\
&- \int_{-\infty}^{\infty} \left\{ \int_{u}^{\infty} (1 - F(v|u)) p_2(\beta, v) dv \right\} p_1(\beta, u) dS^x(\beta, u), \\
\hat{\xi}_2 = &\int_{-\infty}^{\infty} \left\{ \int_{-\infty}^{u-} (1 - \hat{F}(\beta, u - |v)) p_1(\beta, v) dS^x(\beta, v) \right\} \\
&\cdot \left\{ \int_{u}^{\infty} (1 - F(s|u)) p_2(\beta, s) ds \right\} \frac{dW(u)}{J(\beta, u)}.
\end{aligned}$$

The proof of Lemma 2 is given in the Appendix. As in Lai and Ying (1991b) for the censored case, certain regularity conditions are needed for a rigorous justification of the results given in Theorems 1 and 2 below. These conditions are listed as follows:

C1. $\|x_i^*\| \le B$ for all i and some nonrandom constant B.
C2. F has a twice continuously differentiable density f such that $\int_{-\infty}^{\infty} u^2 dF(u) < \infty$ and $\int_{-\infty}^{\infty} (f'(u)/f(u))^2 dF(u) < \infty$.
C3. $\int_{-\infty}^{\infty} \sup_{|h| \le \delta} \{|f'(u+h)| + |f''(u+h)|\} du < \infty$ for some $\delta > 0$.
C4. $\sup_n E|t_n^*| < \infty$.
C5. $\sup_{\|b\| \le \rho, -\infty < u < \infty} \sum_1^m P\{u \le t_i^* - b^T x_i^* \le u + h\} = O(mh)$ as $h \to 0$ and $mh \to \infty$.
C6. $\lim_{m \to \infty} m^{-1} \sum_{i=1}^m E\{x_i^{*k} I_{\{t_i^* - \beta^T x_i^* \le u\}}\} = \Gamma_k^*(u)$ exists for all $u < F^{-1}(1)$ and $k = 0, 1, 2$, and $\int_{-\infty}^{\infty} \Gamma_0^*(u) dF(u) > 0$, where $a^0 = 1, a^1 = a$ and $a^2 = aa^T$ for vectors a.

The requirement that the covariates be uniformly bounded in C1 is a technical assumption, which enables us to simply apply the results developed in Lai and Ying (1988). Condition C2 requires that the error density have a finite second moment and finite Fisher information with respect to location shift. C3–C6, which correspond to (3.3)–(3.6) of Lai and Ying (1991b), are needed to derive a basic asymptotic linearity result for the bias-corrected statistics $\hat{\xi}(b)$ and to ensure that $n^{1/2}(\hat{\beta}-\beta)$ has a limiting normal distribution.

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

$$\frac{1}{m}\sum_{i=1}^{m} I_{\{t_i^* \le y_i^*\}} - \frac{1}{m}\sum_{i=1}^{m}\int_{-\infty}^{\infty} G_i(u)dF(u) \to 0 \quad \text{a.s.,}$$

where G_i is the distribution function of $t_i^* - \beta^T x_i^*$. Therefore (1.3), C6 and the dominated convergence theorem imply that

$$\frac{n}{n^*} \to \int_{-\infty}^{\infty} \Gamma_0^*(u)dF(u) \quad \text{a.s.} \tag{3.3}$$

Assume C6 and define

$$\Gamma_k(u) = \Gamma_k^*(u) \Big/ \int_{-\infty}^{\infty} \Gamma_0^*(u)dF(u), \quad k = 0, 1, 2. \tag{3.4}$$

From the strong law of large numbers together with (3.3) and C6, it follows that $n^{-1}\sum_{i=1}^{n^*} x_i^{*k} I_{\{t_i^* - \beta^T x_i^* \le u\}} \to \Gamma_k(u)$ a.s. for $k = 0, 1, 2$.

Theorem 1. *Suppose that the conditions* C1–C6 *hold. Define* $\hat{\xi}(\beta)$ *by* (2.15) *and* $\Gamma_k(u)$ *by* (3.4). *Then* $n^{-1/2}\hat{\xi}(\beta)$ *converges, in distribution, to a multivariate normal random variable with mean vector* 0 *and covariance matrix*

$$V = \int_{-\infty}^{\infty} \left\{\Gamma_2(u) - \frac{\Gamma_1(u)\Gamma_1^T(u)}{\Gamma_0(u)}\right\}\left\{\frac{\int_u^{\infty}(1-F(v))dv}{1-F(u)}\right\}^2 dF(u). \tag{3.5}$$

The proof of Theorem 1 is similar to that of Lemmas 4, 6 and Theorem 2 of Lai and Ying (1991b). Here we make use of the martingale structure in Lemma 1 in place of Lemma 5 of that paper. An outline of the proof is given in the Appendix. The next theorem provides two approximation results for the random function $\hat{\xi}(b)$. The first result, which will be used to establish the strong consistency of $\hat{\beta}$, shows that in any bounded region of b, $\hat{\xi}(b)$ can be approximated by $\xi_{n^*}(b)$, where $\xi_m(b)$ is a nonrandom function defined in (3.7) below. The second result, which together with Theorem 1 implies the asymptotic normality of $\hat{\beta}$, shows that $\hat{\xi}(b) - \hat{\xi}(\beta)$ is asymptotically linear in $b - \beta$ as $n \to \infty$ and $b \to \beta$. To define $\xi_m(b)$, first define, for $m \ge 1$,

$$\begin{aligned}
\bar{N}_m(b,u) &= \sum_{i=1}^{m} EI_{\{t_i^*(b)\le y_i^*(b),\, u\le y_i^*(b)\}},\\
\bar{N}_m^x(b,u) &= \sum_{i=1}^{m} E\big[x_i^* I_{\{t_i^*(b)\le y_i^*(b),\, u\le y_i^*(b)\}}\big],\\
\bar{J}_m(b,u) &= \sum_{i=1}^{m} EI_{\{t_i^*(b)\le u\le y_i^*(b)\}},\\
\bar{J}_m^x(b,u) &= \sum_{i=1}^{m} E\big[x_i^* I_{\{t_i^*(b)\le u\le y_i^*(b)\}}\big], \qquad (3.6)\\
\bar{S}_m(b,u) &= \sum_{i=1}^{m} EI_{\{t_i^*(b)\le \min(u, y_i^*(b))\}},\\
\bar{S}_m^x(b,u) &= \sum_{i=1}^{m} E\big[x_i^* I_{\{t_i^*(b)\le \min(u, y_i^*(b))\}}\big],\\
\bar{n}(m) &= \sum_{i=1}^{m} P\{t_i^* \le y_i^*\}.
\end{aligned}$$

Also let

$$\begin{aligned}
q_1(b,u) &= p\bigg(\Big(\frac{\bar{S}_m(b,u)}{\bar{n}(m)} - \frac{c_1}{\log \bar{n}(m)}\Big)\log \bar{n}(m)\bigg),\\
q_2(b,u) &= p\bigg(\Big(\frac{\bar{N}_m(b,u)}{\bar{n}(m)} - \frac{c_2}{\log \bar{n}(m)}\Big)\log \bar{n}(m)\bigg),\\
F_m(b,v|u) &= 1 - \exp\Big\{-\int_{u<s\le v} \frac{d\bar{N}_m(b,s)}{\bar{J}_m(b,s)}\Big\},\\
\bar{J}_{m,q_1}^x(b,x) &= \sum_{i=1}^{m} E\big[x_i^* q_1(b, t_i^*(b)) I_{\{t_i^*(b)\le u\le y_i^*(b)\}}\big],
\end{aligned}$$

and define

$$\begin{aligned}
\xi_m(b) = &- \int_{-\infty}^{\infty} u d\big[q_2(b,u)\bar{J}_{m,q_1}^x(b,u)\big]\\
&- \int_{-\infty}^{\infty}\Big\{\int_u^{\infty}(1-F_m(b,v|u))q_2(b,v)dv\Big\} q_1(b,u) d\bar{S}_m^x(b,u). \qquad (3.7)
\end{aligned}$$

Analogous to Lemma 3(ii) of Lai and Ying (1991b), application of (2.18) shows that

$$\xi_m(\beta) = 0. \qquad (3.8)$$

Theorem 2. *Suppose that conditions* C1–C6 *are satisfied. Then for any* $\epsilon > 0$,

$$\sup_{\|b\| \leq \rho} \|\hat{\xi}(b) - \xi_{n^*}(b)\| = o(n^{1/2+\epsilon}) \quad \text{a.s.} \tag{3.9}$$

Let $\tau_0 = \inf\{s : \Gamma_0(s) > 0\}$ *and assume that*

$$\lim_{m\to\infty} \frac{\log m}{m} \sum_{i=1}^{m} P\{t_i^* - \beta^T x_i^* < \tau_0 - \epsilon\} = 0 \text{ for every } \epsilon > 0 \text{ if } F(\tau_0) > 0. \tag{3.10}$$

Then, with probability 1,

$$\hat{\xi}(b) = \hat{\xi}(\beta) - An(b - \beta) + o(\max\{n^{1/2}, n\|b - \beta\|\}) \tag{3.11}$$

uniformly in $\|b - \beta\| \leq n^{-\epsilon}$ *for every* $\epsilon > 0$, *where*

$$A = \int_{\tau_0}^{\infty} \left\{\Gamma_2(u) - \frac{\Gamma_1(u)\Gamma_1^T(u)}{\Gamma_0(u)}\right\}\left\{\int_u^{\infty} (1 - F(v|u))dv\right\} \cdot \left\{\frac{f'(u)}{f(u)} + \frac{f(u)}{1 - F(u)}\right\} dF(u). \tag{3.12}$$

The proof of Theorem 2 uses essentially the same arguments as those used in the proofs of Lemmas 2, 3 and Theorem 1 of Lai and Ying (1991b) for the modified Buckley-James statistics. The details are omitted. By combining Theorems 1 and 2, we obtain the consistency and asymptotic normality of the minimizer $\hat{\beta}$ of $\|\hat{\xi}(b)\|$ in the following.

Theorem 3. *Suppose that conditions* C1–C6 *are satisfied. Define* $\hat{\beta}$ *by*

$$\|\hat{\xi}(\hat{\beta})\| = \min_{b: \|b\| \leq \rho} \{\|\hat{\xi}(b)\|\}. \tag{3.13}$$

(i) *Assume that for every* $\delta > 0$,

$$\lim_{m\to\infty} m^{-1} \inf_{\|b\| \leq \rho, \|b-\beta\| \geq \delta} \{\|\xi_m(b)\|\} > 0. \tag{3.14}$$

Then $\hat{\beta} \to \beta$ *a.s.*
(ii) *Assume that* (3.10) *holds and that there exists* $\epsilon \in (0, 1/2)$ *for which*

$$\lim_{m\to\infty} \frac{1}{m^{1/2+\epsilon}} \inf_{\|b\| \leq \rho, \|b-\beta\| \geq m^{-\epsilon}} \|\xi_m(b)\| = \infty, \tag{3.15}$$

and that the matrix A *defined in* (3.12) *is nonsingular. Then* $\|\hat{\beta} - \beta\| = O(n^{-\epsilon})$ *a.s. and* $\sqrt{n}(\hat{\beta} - \beta)$ *has a limiting normal distribution with mean* 0 *and covariance matrix* $A^{-1}VA^{-1}$, *where* V *is defined in* (3.5).

Proof. Since $\xi_m(\beta) = 0$, (i) follows from (3.9) and (3.13). Moreover, from (3.15) and (3.9), we similarly conclude that

$$\lim_{n\to\infty} \inf_{\|b\|\le\rho, \|b-\beta\|\ge n^{-\epsilon}} n^{-1/2-\epsilon}\|\hat{\xi}(b)\| = \infty \quad \text{a.s.,}$$

which implies, in view of (3.11) and (3.13), that $\|\hat{\beta}-\beta\| = O(n^{-\epsilon})$ a.s. Therefore, we can apply (3.11) (with $b = \hat{\beta}$) and Theorem 1 to conclude that $\sqrt{n}(\hat{\beta}-\beta)$ has a limiting normal distribution with mean 0 and covariance matrix $A^{-1}VA^{-1}$. This completes the proof of (ii).

Suppose that (x_i^{*T}, t_i^*) are i.i.d. and that $\int_{-\infty}^{\infty} P\{t_1^* - \beta^T x_1^* \le u\} dF(u) > 0$. Then C6 holds with $\Gamma_k^*(u) = E(x_1^{*k} I_{\{t_i^* - \beta^T x_i^* \le u\}})$ and (3.10) is obviously satisfied. Under assumptions C1–C5, the conclusion (3.11) of Theorem 2 can be shown to hold as $n \to \infty$ and $b \to \beta$ (without having to restrict to $\|b-\beta\| \le n^{-\epsilon}$) for this i.i.d. setting. Moreover, $m^{-1}\xi_m(b) \to \ell(b)$, the convergence being uniform in $\|b\| \le \rho$, where the limit function $\ell(b)$ is defined below. Let $H(x) = P\{x_1^* \le x\}$ and $G_x(s) = P\{t_1^* \le s | x_1^* = x\}$. Define

$$F(b,v|u) = 1 - \exp\left\{-\int_u^v \frac{\int_{-\infty}^{\infty} G_x(s+b^Tx) f(s+(b-\beta)^T x) dH(x)}{\int_{-\infty}^{\infty} G_x(s+b^Tx)[1-F(s+(b-\beta)^Tx)]dH(x)} ds\right\},$$
$$v \ge u, \tag{3.16}$$

$$\ell(b) = -\int_{-\infty}^{\infty} x\left\{\int_{-\infty}^{\infty} u d[G_x(u+b^Tx)(1-F(u+(b-\beta)^Tx))]\right.$$
$$+\int_{-\infty}^{\infty}[\int_u^{\infty}(1-F(b,v|u))dv][1-F(u+(b-\beta)^Tx)]$$
$$\left. dG_x(u+b^Tx)\right\} dH(x). \tag{3.17}$$

Note that $\ell(b)$ is continuous for $\|b\| \le \rho$ and that $\ell(\beta) = 0$ by (2.18). Hence, if

$$\ell(b) \ne 0 \text{ for } b \ne \beta \text{ (with } \|b\| \le \rho), \tag{3.18}$$

then (3.14) is satisfied and $\hat{\beta}$ is strongly consistent. If, furthermore, A is nonsingular, then (3.18) holds for every $0 < \epsilon < 1/2$ in view of (3.11) (with $b \to \beta$) and (3.14), and therefore $\hat{\beta}$ is asymptotically normal by Theorem 3(ii).

By assuming x_i to be i.i.d. and $t_i \equiv c$ (some constant), Tsui (1988) established the strong consistency of the Tsui-Jewell-Wu estimator under certain additional assumptions which are considerably more restrictive than C1–C5 and (3.18). While our proof follows basically the same lines developed in Lai and Ying (1991b) for the Buckley-James estimator based on censored data, Tsui's proof uses a different argument that is based on approximating the estimating equation (2.4) by a more tractable minimization problem. His approach, however, requires knowledge of a finite set to which β belongs so that the minimization is performed over this finite set to make the problem tractable.

4. Confidence Regions and Asymptotic Efficiency

To make use of Theorem 3(ii) to construct approximate $(1-\alpha)$-level confidence regions for β, we first consider consistent estimation of the matrix V defined in (3.5), which can be rewritten as

$$V = \int_{-\infty}^{\infty} \Big\{(1-F(u))\Gamma_2(u) - \frac{[(1-F(u))\Gamma_1(u)][(1-F(u))\Gamma_1(u)]^T}{(1-F(u))\Gamma_0(u)}\Big\} \times \Big\{\int_u^{\infty}(1-F(v|u))dv\Big\}^2 d\Lambda(u). \tag{4.1}$$

An obvious estimator of (4.1) is therefore

$$\hat V = \sum_{r=1}^{n}\Big\{n^{-1}\sum_{i=1}^{n} x_i x_i^T I_{\{t_i(\hat\beta)\le y_r(\hat\beta)\le y_i(\hat\beta)\}} - \frac{J^x(\hat\beta, y_r(\hat\beta))(J^x(\hat\beta, y_r(\hat\beta)))^T}{nJ(\hat\beta, y_r(\hat\beta))}\Big\} \times \Big\{\int_{y_r(\hat\beta)}^{\infty}[1-\hat F(\hat\beta, v|y_r(\hat\beta))]p_1(\hat\beta, v)p_2(\hat\beta, v)dv\Big\}^2 / J(\hat\beta, y_r(\hat\beta)), \tag{4.2}$$

which can be shown to converge a.s. to V under the conditions C1–C6 and (3.14).

Since the limiting covariance matrix of $\sqrt{n}(\hat\beta-\beta)$ in Theorem 3(ii) is $A^{-1}VA^{-1}$, the problem of estimating A consistently has also to be addressed. In view of (3.12), A involves both the density f and its derivative f' of the underlying error distribution. Although in principle one can resort to kernel and other methods to estimate these quantities, there are practical difficulties of estimating them well enough to provide a reliable estimate of A. It is, however, possible to avoid these difficulties by extending the method of Wei, Ying and Lin (1990) for censored data to completely bypass estimation of A. First note that by Theorem 1, $n^{-1}\hat\xi^T(\beta)V^{-1}\hat\xi(\beta)$ converges in distribution to a $\chi^2(p)$ random variable. Since $\hat V \to V$ a.s., it then follows that $n^{-1}\hat\xi^T(\beta)\hat V^{-1}\hat\xi(\beta)$ also has a limiting $\chi^2(p)$ distribution, and therefore

$$\{b : \|b\| \le \rho, n^{-1}\hat\xi^T(b)\hat V^{-1}\hat\xi(b) \le \chi^2_{\alpha;p}\} \tag{4.3}$$

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

If we are only interested in some components of the parameter vector β, we can also use the procedure of Wei, Ying and Lin (1990) to construct an approximate confidence region for these components. Let $\beta = (\beta_1^T, \beta_2^T)^T$. Suppose that

β_1 is the parameter of interest and that we want to construct a $(1-\alpha)$-level confidence region only for β_1. Choose constants $\delta \in (0, 1/2)$ and $c > 0$. Let $\hat{\beta}_1$, $\hat{\beta}_2$ denote the corresponding components in the bias-corrected least squares estimator $\hat{\beta}$. Let $Q(b_1, b_2) = n^{-1}\hat{\xi}^T(b)\hat{V}^{-1}\hat{\xi}(b)$. Then it follows from Theorems 2 and 3 and an argument similar to Appendix B of Wei, Ying and Lin (1990) that $D(\beta_1) = \min\{Q(\beta_1, b_2) : \|b_2 - \hat{\beta}_2\| \le cn^{-\delta}\}$ is asymptotically $\chi^2(p_1)$ distributed, where p_1 is the dimension of β_1. Therefore an approximate $(1-\alpha)$-level confidence region for β_1 is the set $\{b_1 : \|b_1\| \le \rho, D(b_1) \le \chi^2_{\alpha;p_1}\}$.

The asymptotic variance formula of Theorem 3 also enables us to compare large sample performance of the bias-corrected least squares estimator with other available methods. In particular, we can compare it with the estimator of Bhattacharya et al. (1983). By applying this variance formula to the truncated regression model (1.1)–(1.2) with $\beta = 1$, $t_i = t$ and x_i being i.i.d. uniform on $(-1, 0)$, Table 1 gives the variances of the limiting normal distributions of three estimators, namely, the bias-corrected least squares estimator $\hat{\beta}$, the rank estimator β^* of Bhattacharya et al. (1983) based on an extension of the Wilcoxon rank statistics to truncated data, and the rank estimator β^{**} of Lai and Ying (1992a) based on an alternative extension of the Wilcoxon statistics to truncated data. In each case, the error distribution is assumed to be normal with mean 0 and its standard deviation σ is specified. Note that assumptions C1–C6 are satisfied and that $A \ne 0$ in the examples considered in Table 1. Moreover, as will be shown in Lemma 3 below, condition (3.18) is satisfied. The limiting variance $A^{-2}V$ of $\sqrt{n}(\hat{\beta} - \beta)$ as $n \to \infty$ is evaluated from (3.5) and (3.12) in Table 1 by numerical integration. As shown by Bhattacharya et al. (1983) and Lai and Ying (1992a), under certain regularity conditions, $\sqrt{n}(\beta^* - \beta)$ has a limiting normal distribution with mean 0 and variance

$$\frac{\int(\Gamma_2(u) - \Gamma_1^2(u)/\Gamma_0(u))\Gamma_0^2(u)(1-F(u))^2 dF(u)}{\{\int(\Gamma_2(u) - \Gamma_1^2(u)/\Gamma_0(u))\Gamma_0(u)(1-F(u))[f'(u)/f(u) + f(u)/(1-F(u))]dF(u)\}^2} \tag{4.4}$$

while $\sqrt{n}(\beta^{**} - \beta)$ has a limiting normal distribution with mean 0 and variance

$$\frac{\int(\Gamma_2(u) - \Gamma_1^2(u)/\Gamma_0(u))(1-F(u))^2 dF(u)}{\{\int(\Gamma_2(u) - \Gamma_1^2(u)/\Gamma_0(u))(1-F(u))[f'(u)/f(u) + f(u)/(1-F(u))]dF(u)\}^2}. \tag{4.5}$$

Both (4.4) and (4.5) are also evaluated by numerical integration in Table 1.

In Table 2 we replace the underlying normal distribution $N(0, \sigma^2)$ considered in Table 1 by a logistic distribution with density $f(u) = \lambda e^{-\lambda u}/(1 + e^{-\lambda u})^2$, $-\infty < u < \infty$, so its standard deviation is $\sigma = \lambda^{-1}\pi/\sqrt{3}$. The examples in Table 2 again satisfy assumptions C1–C6 and the condition $A \ne 0$. Moreover, condition (3.18) is satisfied in view of the following lemma, whose proof is given

Table 1. Variances of limiting normal distributions of three estimators. Normal error distributions.

σ	t	$\hat{\beta}$	β^*	β^{**}
.3	−.50	6.94	7.66	8.46
	−.75	3.53	3.80	3.91
	−1.00	2.07	2.22	2.19
.5	−.50	15.27	16.67	18.22
	−.75	10.06	10.88	11.23
	−1.00	7.00	7.55	7.53
1.0	−.50	45.86	50.60	52.11
	−.75	37.96	41.57	42.21
	−1.00	31.78	34.55	34.70

in the Appendix. Note that both the normal and the logistic distributions considered in Tables 1 and 2 have increasing failure rates (IFR), i.e., the hazard function $f/(1-F)$ is increasing. The following lemma shows that the function $\ell(b)$ defined in (3.17) has a unique zero at $b=\beta$ for IFR distributions F.

Lemma 3. *Define* $F(b,v|u)$ *by* (3.16) *and* $\ell(b)$ *by* (3.17), *in which* F, H *and* G_x *are distribution functions on the real line such that* F *has a bounded density function* f *and* $\int_{-\infty}^{\infty} u^2 dF(u) + \sup_x \int_{-\infty}^{\infty} |u| dG_x(u) < \infty$. *Suppose that the hazard function* $f/(1-F)$ *is strictly increasing and positive everywhere, and that the support of the probability measure associated with* H *is compact and contains more than one point. Then* $\ell(b)$ *has a unique zero at* $b=\beta$.

Table 2. Variances of limiting normal distributions of three estimators. Logistic error distributions.

σ	t	$\hat{\beta}$	β^*	β^{**}
.3	−.50	13.72	10.29	9.50
	−.75	4.33	3.73	3.55
	−1.00	1.94	1.78	1.74
.5	−.50	36.56	28.96	26.35
	−.75	14.09	12.14	11.43
	−1.00	7.04	6.35	6.14
1.0	−.50	101.27	84.57	79.75
	−.75	59.43	50.82	48.63
	−1.00	38.28	33.45	32.41

Tables 1 and 2 show that the limiting variances change quite substantially for different values of t. Note that as t increases the heavier the truncation becomes. Table 1 shows that for normally distributed ϵ_i, the bias-corrected least squares estimator $\hat\beta$ is asymptotically more efficient than the rank estimators β^* and β^{**} that are based on extensions of the Wilcoxon statistics to truncated data. This is in agreement with the simulation results reported by Tsui et al. (1988) comparing their bias-corrected least squares estimator $\tilde\beta$ (which has been reviewed in Section 2) with the rank estimator β^* in a setting similar to that of Table 1 but with right instead of left truncation. This is also consistent with the case of complete data, for which the least squares estimator is asymptotically most efficient when the underlying error distribution is normal. When the underlying error distribution is non-normal, it is well known that rank estimators can substantially outperform least squares estimators in the case of complete data. Table 2 shows that for truncated data the bias-corrected least squares estimator $\hat\beta$ is less efficient than the rank estimators β^{**} and β^* when the underlying error distribution is logistic.

In the univariate case $p = 1$, application of the Schwarz inequality to the integral (3.12) defining A gives

$$A^2 \le D \int_{-\infty}^{\infty} \left\{\Gamma_2(u) - \frac{\Gamma_1^2(u)}{\Gamma_0(u)}\right\}\left\{\int_u^{\infty}(1 - F(v|u))dv\right\}^2 dF(u), \tag{4.6}$$

where

$$D = \int_{-\infty}^{\infty} \left\{\Gamma_2(u) - \frac{\Gamma_1^2(u)}{\Gamma_0(u)}\right\}\left\{\frac{f'(u)}{f(u)} + \frac{f(u)}{1 - F(u)}\right\}^2 dF(u). \tag{4.7}$$

Consequently, the variance $A^{-2}V$ of the limiting normal distribution of $\sqrt{n}(\hat\beta - \beta)$ is bounded below by D^{-1}. The inequality in (4.6) is strict unless there exists $\alpha \ne 0$ for which

$$\int_u^{\infty}(1 - F(v|u))dv = \alpha\{f'(u)/f(u) + f(u)/(1 - F(u))\} \text{ a.e. } (F). \tag{4.8}$$

When f is the density of a normal distribution with mean μ and variance σ^2, (4.8) holds with $\alpha = \sigma^2$, and therefore the variance of the limiting normal distribution of $\sqrt{n}(\hat\beta - \beta)$ attains the lower bound D^{-1}. If, in addition, the $t_i^* - \beta x_i^*$ are i.i.d. and are independent of the x_i^*, then the usual maximum likelihood estimator of β is also asymptotically normal with the same asymptotic variance, and therefore the bias-corrected least squares estimator is asymptotically as efficient as the parametric (normal) maximum likelihood estimator.

Without assuming $t_i^* - \beta x_i^*$ and x_i^* to be independent, the general theory of asymptotic lower bounds for the variances of the limiting distributions of regular estimators in semiparametric estimation developed by Begun, Hall, Huang and Wellner (1983) can be applied if the (t_i^*, x_i^*) are assumed to be i.i.d. More generally, for i.i.d. (t_i^*, x_i^{*T}) in the truncated regression model (1.1)–(1.2) with p-dimensional regressors x_i^*, define the $p\times p$ matrix D by (4.7). The results of Begun et al. (1983) show that the limiting distribution of $\sqrt{n}(T_n - \beta)$ for a sequence of regular estimators $\{T_n\}$ is a convolution of $N(0, D^{-1})$ with some distribution. Extension of this theory to the setting of Theorem 3, in which (t_i^*, x_i^{*T}) need not be identically distributed, is given in Lai and Ying (1992b). Note that if (4.8) holds for some $\alpha \neq 0$, as is the case of normal f, then $D^{-1} = A^{-1}VA^{-1} =$ the covariance matrix of the asymptotic distribution of the bias-corrected least squares estimator $\hat{\beta}$.

Acknowledgement

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

Appendix

Proof of (2.18). Let $H_i(u) = P\{t_i^*(\beta) \le u | x_i^*\}$. The left hand side of (2.18) is equal to

$$\begin{aligned}&\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} u g_1(v) d\{I_{\{v\le u\}}(1-F(u))g_2(u)\} dH_i(v)\\ &\quad + \int_{-\infty}^{\infty}\{\int_u^{\infty}(1-F(v))g_2(v)dv\} g_1(u)dH_i(u)\\ &= -\int_{-\infty}^{\infty}\{\int_{-\infty}^{\infty} I_{\{v\le u\}}(1-F(u))g_2(u)du\} g_1(v)dH_i(v)\\ &\quad + \int_{-\infty}^{\infty}\{\int_u^{\infty}(1-F(v))g_2(v)dv\} g_1(u)dH_i(u) = 0,\end{aligned}$$

where the first equality follows from integration by parts.

Proof of Lemma 2. To prove (3.2), it suffices, in view of (2.15), to show that with probability 1, for all large n,

$$\begin{aligned}&\int_{-\infty}^{\infty}\{\int_u^{\infty}(1-F(v|u))p_2(\beta,v)dv\} p_1(\beta,u)dS^x(\beta,u)\\ &\quad - \int_{-\infty}^{\infty}\{\int_u^{\infty}(1-\hat{F}(\beta,v|u))p_2(\beta,v)dv\} p_1(\beta,u)dS^x(\beta,u)\\ &= \int_{-\infty}^{\infty}\{\int_{-\infty}^{u-}(1-\hat{F}(\beta,u-|v))p_1(\beta,v)dS^x(\beta,v)\}\\ &\qquad \{\int_u^{\infty}(1-F(s|u))p_2(\beta,s)ds\}\frac{dW(u)}{J(\beta,u)}. \qquad (A.1)\end{aligned}$$

By Doleans-Dade's exponential formula (cf. Shorack and Wellner (1986, p.897)), we have

$$\frac{1-\hat{F}(\beta,v|u)}{1-F(v|u)} = 1 - \int_{u+}^{v} \frac{1-\hat{F}(\beta,s-|u)}{1-F(s|u)} \frac{dW(s)}{J(\beta,s)} \tag{A.2}$$

if $\inf_{u\le s\le v} J(\beta,s) > 0$, in which case (A.2) implies that

$$1-F(v|u)-(1-\hat{F}(\beta,v|u)) = (1-F(v|u))\int_{u+}^{v} \frac{1-\hat{F}(\beta,s-|u)}{1-F(s|u)} \frac{dW(s)}{J(\beta,s)}. \tag{A.3}$$

Since for large n, $\inf_{u\le s\le v} J(\beta,s) > 0$ if $p_1(\beta,u) > 0$ and $p_2(\beta,v) > 0$, it follows from (A.3) that the left hand side of $(A.1)$ becomes, with probability 1,

$$\int_{-\infty}^{\infty}\Big\{\int_{u}^{\infty}(1-F(v|u))[\int_{u+}^{v}\frac{1-\hat{F}(\beta,s-|u)}{1-F(s|u)}\frac{dW(s)}{J(\beta,s)}]p_2(\beta,v)dv\Big\}p_1(\beta,u)dS^x(\beta,u), \tag{A.4}$$

for all large n. (A.1) then follows from (A.4) by interchanging the order of integration.

Proof of Theorem 1. In view of Lemma 2 it suffices to establish the asymptotic normality of $\hat{\xi}_1+\hat{\xi}_2$. Since $\hat{\xi}_2$ is a stochastic integral with respect to the martingale process W, we can use arguments similar to those in the proof of Lemmas 5 and 6 of Lai and Ying (1991b) to show that with probability 1,

$$\begin{aligned}\hat{\xi}_2 = \int_{-\infty}^{\infty}&\Big\{\int_{-\infty}^{u-}(1-F(u|v))q_1(\beta,v)d\bar{S}_{n^*}(\beta,v)\Big\}\\ &\times\Big\{\int_{u}^{\infty}(1-F(s|u))q_2(\beta,s)ds\Big\}\frac{dW(u)}{\bar{J}_{n^*}(\beta,u)} + o_p(\sqrt{n}).\end{aligned} \tag{A.5}$$

Moreover, analogous to the proof of Lemma 4 of Lai and Ying (1991b), we approximate the random weight functions p_1 and p_2 in $\hat{\xi}_1$ by their nonrandom counterparts q_1 and q_2 and obtain

$$\begin{aligned}\hat{\xi}_1 = &-\sum_{i=1}^{n^*} x_i^* p_1(\beta,t_i^*(\beta))\int_{-\infty}^{\infty} ud[q_2(\beta,u)I_{\{t_i^*(\beta)\le u\le y_i^*(\beta)\}}]\\ &-\sum_{i=1}^{n^*} x_i^*\int_{-\infty}^{\infty}\Big\{\int_{u}^{\infty}(1-F(v|u))q_2(\beta,s)ds\Big\}q_1(\beta,u)dI_{\{t_i^*(\beta)\le u\}} + o_p(\sqrt{n}).\end{aligned} \tag{A.6}$$

By making use of (A.5), (A.6) and (3.3), we can use an argument similar to that in the proof of Theorem 2(ii) of Lai and Ying (1991b) to show that $n^{-1/2}(\hat{\xi}_1+\hat{\xi}_2)$ has a limiting normal distribution with mean 0 and covariance matrix (3.5).

Proof of Lemma 3. The change of variables $v = u + bx$ yields

$$\int_{-\infty}^{\infty} ud[G_x(u+bx)(1-F(u+(b-\beta)x)] = \int_{-\infty}^{\infty} vd[G_x(v)(1-F(v-\beta x))],$$

since $bx\int_{-\infty}^{\infty} d[G_x(v)(1-F(v-\beta x))] = 0$. Hence, the first summand on the right hand side of (3.17) is $\int_{-\infty}^{\infty} x\{\int_{-\infty}^{\infty} vd[G_x(v)(1-F(v-\beta x))]\}dH(x)$, which does not depend on b. To show that $\ell(b) \neq 0$ for $b \neq \beta$, it therefore suffices to prove that $\ell^*(b) \neq \ell^*(\beta)$ for $b \neq \beta$, where $\ell^*(b)$ denotes the second summand on the right hand side of (3.17). Let

$$\begin{aligned} \psi_b(s) &= \frac{\int_{-\infty}^{\infty} G_y(s+by)f(s+(b-\beta)y)dH(y)}{\int_{-\infty}^{\infty} G_y(s+by)[1-F(s+(b-\beta)y)]dH(y)}, \\ e_b(u) &= \int_u^{\infty}\{1-F(b,v|u)\}dv = \int_u^{\infty}\exp\{-\int_u^v \psi_b(s)ds\}ds. \end{aligned} \tag{A.7}$$

Integration by parts yields

$$\begin{aligned} \int_{-\infty}^{\infty} e_b(u)[1-F(u+(b-\beta)x)]dG_x(u+bx) &= \int_{-\infty}^{\infty} G_x(u+bx) \\ \times \{[1-F(u+(b-\beta)x)][1-\psi_b(u)e_b(u)] &+ e_b(u)f(u+(b-\beta)x)\}du. \end{aligned}$$

Putting this in the second summand of (3.17) gives

$$\ell^*(b) = -\int_{-\infty}^{\infty} x\int_{-\infty}^{\infty} G_x(w)[1-F(w-\beta x)]dwdH(x) + \int_{-\infty}^{\infty} e_b(u)h_b(u)du,$$

where

$$\begin{aligned} h_b(u) =&\psi_b(u)\int_{-\infty}^{\infty} xG_x(u+bx)[1-F(u+(b-\beta)x)]dH(x) \\ &-\int_{-\infty}^{\infty} xG_x(u+bx)f(u+(b-\beta)x)dH(x). \end{aligned} \tag{A.8}$$

Fix $b > \beta$ and let $\delta = b - \beta (> 0)$. We now show that $h_b(u) \leq 0$ and that strict inequality holds for all large u. From the definition of $\psi_b(u)$ in (A.7), it suffices to show that

$$\left\{\int_{-\infty}^{\infty} f(u+\delta y)G_y(u+by)dH(y)\right\} \cdot \left\{\int_{-\infty}^{\infty} x[1-F(u+\delta x)]G_x(u+bx)dH(x)\right\}$$
$$\leq \left\{\int_{-\infty}^{\infty} [1-F(u+\delta y)]G_y(u+by)dH(y)\right\} \cdot \left\{\int_{-\infty}^{\infty} xf(u+\delta x)G_x(u+bx)dH(x)\right\}, \tag{A.9}$$

with strict inequality for all large u. Since $f(v)/(1-F(v))$ is strictly increasing, it follows that for $x > y$,

$$(x-y)\{f(u+\delta y)[1-F(u+\delta x)] - f(u+\delta x)[1-F(u+\delta y)]\} < 0. \tag{A.10}$$

Expressing both sides of (A.9) as double integrals of the form $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}$ and then representing their difference as a double integral of the form $\iint_{y<x}$, we can use (A.10) to establish (A.9), with strict inequality if $\iint_{y<x} G_y(u+by)G_x(u+bx)dH(x)dH(y) > 0$, which holds for all large u since the support of H has more than one point.

Similarly, it can be shown, for fixed $b < \beta$, that $h_b(u) \geq 0$, with strict inequality for all large u. Since $e_b(u) > 0$, it then follows that

$$\int_{-\infty}^{\infty} e_b(u)h_b(u)du < 0 \text{ if } b > \beta, \int_{-\infty}^{\infty} e_b(u)h_b(u)du > 0 \text{ if } b < \beta. \tag{A.11}$$

By (A.7) and (A.8), we have $h_\beta(u) \equiv 0$. From (A.8) and (A.11), it then follows that $\ell^*(b) < \ell^*(\beta)$ if $b > \beta$ and $\ell^*(b) > \ell^*(\beta)$ if $b < \beta$.

References

Adichie, J. N. (1967). Estimates of regression parameters based on rank tests. *Ann. Math. Statist.* **38**, 894-904.

Amemiya, T. (1985). *Advanced Econometrics.* Basil Brackwell Ltd, Oxford.

Begun, J. M., Hall, W. J., Huang, W. M. and Wellner, J. A. (1983). Information and asymptotic efficiency in parametric nonparametric models. *Ann. Statist.* **11**, 432-452.

Bhattacharya, P. K., Chernoff, H. and Yang, S. S. (1983). Nonparametric estimation of the slope of a truncated regression. *Ann. Statist.* **11**, 505-514.

Buckley, J. and James, I. (1979). Linear regression with censored data. *Biometrika* **66**, 429-436.

Cox, D. R. and Oakes, D. (1984). *Analysis of Survival Data.* Chapman and Hall, London.

Gill, R. D. (1980). *Censoring and Stochastic Integrals* (Mathematical Centre Tracts No. 124). Mathematisch Centrum, Amsterdam.

Goldberger, A. S. (1981). Linear regression after selection *J. Econometrics* **15**, 357-366.

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

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

Lai, T. L. and Ying, Z. (1991a). Estimating a distribution function with truncated and censored data. *Ann. Statist.* **19**, 417-442.

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

Lai, T. L. and Ying, Z. (1992a). Linear rank statistics in regression analysis with censored or truncated data. *J. Multivariate Anal.* **40**, 13-45.

Lai, T. L. and Ying, Z. (1992b). Asymptotically efficient estimation in censored and truncated regression models. *Statistica Sinica* **2**, 17-46.

Nicoll, J. F. and Segal, I. E. (1980). Nonparametric estimation of the observational cutoff bias. *Astron. Astrophys.* **82**, L3-L6.

Segal, I. E. (1975). Observational validation of the chronometric cosmology: I. Preliminaries and the red shift-magnitude relation. *Proc. Natl. Acad. Sci.* USA **72**, 2437-2477.

Shorack, G. R. and Wellner, J. A. (1986). *Empirical Processes with Applications to Statistics.* John Wiley, New York.

Tobin, J. (1958). Estimation of relationships for limited dependent variables. *Econometrica* **26**, 24-36.

Tsui, K. L. (1988). Strong consistency for nonparametric estimation with truncated regression data. AT&T Bell Laboratories Technical Report.

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

Wang, M. C., Jewell, N. P. and Tsai, W. Y. (1986). Asymptotic properties of the product limit estimate under random truncation. *Ann. Statist.* **14**, 1597-1605.

Wei, L. J., Ying, Z. and Lin, D. Y. (1990). Linear regression analysis of censored survival data based on rank tests. *Biometrika* **77**, 845-851.

Department of Statistics, Stanford University, Stanford, CA 94305, U.S.A.

Department of Statistics, University of Illinois at Urbana, IL 61820, U.S.A.

(Received June 1990; accepted August 1991)