Columbia University
Columbia University Biostatistics Technical Report Series

Year 2007 *Paper* 8

A Least Squares Estimation in Truncated Linear Regression

Wei Yann Tsai* Xinhua Liu†
Xiaodong Luo‡

*Columbia University, wt5@columbia.edu
†Columbia University
‡Mount Sinai School of Medicine

A Least Squares Estimation in Truncated Linear Regression

Wei Yann Tsai, Xinhua Liu, and Xiaodong Luo

Abstract

We investigate least squares estimation for regression coefficients of the covariates in the multiple linear regression model with truncated data and propose an alternative consistent least squares type estimator to the existing ones. The estimator is proved to have an asymptotic normal distribution with the same asymptotic variance matrix as the estimator proposed by Lai and Ying (1992b). However, the estimator is much simpler in computation than Lai and Ying's estimator. The estimation procedure does not require calculation of the nonparametric estimate of the error distribution. A simulation study shows that the estimator performs well even with a moderate sample size.

A Least Squares Estimation in Truncated Linear Regression

Abbreviated Title: Truncated Linear Regression

By Wei-Yann Tsai[1], Xinhua Liu and Xiaodong Luo

Columbia University

We investigate a least squares estimation for regression coefficients of the covariates in the multiple linear regression model with truncated data and propose an alternative consistent least squares type estimator to the existing ones. The estimator is proved to have an asymptotic normal distribution with the same asymptotic variance matrix as the estimator proposed by Lai and Ying (1992b). However, the estimator is much simpler in computation than Lai and Ying's estimator. The estimation procedure does not require to calculate the nonparametric estimate of the error distribution. A simulation study shows that the estimator performs well even with a moderate sample size.

[1]Partial work of this paper was done while the first author visited National Cheng Kung University, Taiwan.

AMS 1991 subject classifications. 62J05, 93C41.
Key words and phrases. Estimating function, Least squares estimator, Linear Regression, Truncation.

1. INTRODUCTION

The linear regression model considered here is $Y^* = \boldsymbol{\beta}^T \mathbf{X}^* + \varepsilon^*$, where $\mathbf{X}^*$ is a p-vector of covariates, and $\boldsymbol{\beta}$ is the p-vector of parameters of interest. Let V^* be a truncation random variable. We assume that the random vector $(V^*, \mathbf{X}^*)$ is independent of the error term ε^*, and ε^* has a distribution F (which needs not have mean 0) with probability density function f and finite variance σ^2. A right truncated regression model assumes that $(Y^*, V^*, \mathbf{X}^*)$ is observed if and only if $Y^* \leq V^*$. Suppose that $(y_i, v_i, \mathbf{x}_i), i = 1, \cdots, n$, are observed data. Note that $y_i \leq v_i$ for all i. By multiplying each variable by -1, it can be converted to a left truncated regression with data $(y_i', v_i', \mathbf{x}_i)$ observable only if $y_i' \geq v_i'$ for all i. Subsequently, we will only consider a right truncated regression model and estimate $\boldsymbol{\beta}$ based on the n independent observations $(y_i, v_i, \mathbf{x}_i), i = 1, \cdots, n$, subject to right truncation. Notice that, conditioned on n, we can treat $(y_i, v_i, \mathbf{x}_i), i = 1, \cdots, n$ as i.i.d. random vectors.

It is well known that the ordinary least squares estimate of $\boldsymbol{\beta}$ based on observed data $(y_i, \mathbf{x}_i), i = 1, \cdots, n$, is inconsistent. When V^* is degenerate and $(Y^*, \mathbf{X}^*)$ has multivariate Gaussian distribution with $E(\mathbf{X}^*) = \mathbf{0}$, the least squares slope estimate will be attenuated toward null, although it may not be true in general. By assuming that the underlying error distribution belongs to a parametric distribution family, $\boldsymbol{\beta}$ can be estimated by the maximum conditional likelihood method. There is extensive literature on the subject. In recent years, several nonparametric estimation methods for $\boldsymbol{\beta}$ have been proposed and their large sample properties have been discussed as well. Bhattacharya, Chernoff, and Yang (1983) proposed rank estimators of $\boldsymbol{\beta}$ in the simple linear model. Subsequently, a general rank-type approach was developed by Lai and Ying (1992a). Powell (1986) considered a method for estimating $\boldsymbol{\beta}$ assuming error distribution is symmetric. Tsui, Jewell and Wu (1988) introduced a bias-corrected least squares estimator. Lai and Ying (1992b) established asymptotic theory for the bias-corrected least squares estimator with slight modification. Chen, Tsai and Chao (1996) (hereafter denoted by CTC) obtained a least squares type estimator of $\boldsymbol{\beta}$ based on an extension of ordinary least squares estimation. CTC's estimator is consistent, easy in computation, and invariant when dependent and/or independent variables shift by a constant, but it is not as efficient as Tsui-Jewell-Wu's estimator or Lai and Ying's modified estimator when error has a Gaussian distribution.

In this paper, we first review the estimating equations considered by Tsui, Jewell and Wu (1988) as well as Lai and Ying (1992b). Then an alternative consistent least squares type estimator is proposed based on an unbiased estimating equation. This estimator, as a weighted version of CTC's estimator, takes all the advantages of CTC's method, yet is more efficient when error has a normal distribution since it is proven to share the same asymptotic normal distribution as proposed by Lai and Ying (1992b). A simulation study is conducted to examine the finite sample performance of the proposed estimator and the result is satisfactory.

2 LEAST SQUARE TYPE ESTIMATION

2.1 Bias corrected estimating equation for $\boldsymbol{\beta}$

Given $\mathbf{b}$, a fixed value of $\boldsymbol{\beta}$, we can calculate the residual $e_i(\mathbf{b}) = y_i - \mathbf{b}^T\mathbf{x}_i$ and truncation points $t_i(\mathbf{b}) = v_i - \mathbf{b}^T\mathbf{x}_i$, for $i = 1, \cdots, n$. The ordinary least squares estimator $\hat{\boldsymbol{\beta}}_{OLS}$ of $\boldsymbol{\beta}$ is the solution of $\sum_{i=1}^n \mathbf{x}_i(e_i(\mathbf{b}) - \bar{e}(\mathbf{b})) = \mathbf{0}$, where $\bar{e}(\mathbf{b}) = \sum_{i=1}^n e_i(\mathbf{b})/n$. It is biased in the presence of truncation since $E(e_i(\boldsymbol{\beta}) - \bar{e}(\boldsymbol{\beta})|\mathbf{x}_i) \neq 0$. To correct the bias, Tsui, Jewell and Wu (1988) considered the estimating equation

$$\sum_{i=1}^n \mathbf{x}_i \left\{ e_i(\mathbf{b}) - E(e_i(\mathbf{b})|\mathbf{x}_i) \right\} = 0 \tag{2.1}$$

where for right truncation and true $\boldsymbol{\beta}$,

$$E(e_i(\boldsymbol{\beta})|\mathbf{x}_i) \quad = \quad E(\varepsilon^*|\varepsilon^* \leq t_i(\boldsymbol{\beta})) = \frac{\int_{-\infty}^{t_i(\boldsymbol{\beta})} uF(du)}{F(t_i(\boldsymbol{\beta}))} \ .$$

When error distribution F is unknown, the bias corrected estimating equation, first considered by Tsui, Jewell and Wu (1988), is in the form of

$$\xi(\mathbf{b}) = \sum_{i=1}^n \mathbf{x}_i \left\{ e_i(\mathbf{b}) - \hat{E}(e_i(\mathbf{b})|\mathbf{x}_i) \right\} \tag{2.2}$$

where $\hat{E}(e_i(\mathbf{b})|\mathbf{x}_i) = \int_{-\infty}^{t_i(\mathbf{b})} d\hat{F}_b(du)/\hat{F}_b(t_i(\mathbf{b}))$ and $\hat{F}_b$ is the product-limit estimator of F based on truncated data $(e_i(\mathbf{b}), t_i(\mathbf{b})), i = 1, ..., n$. The argument is that when $\hat{E}(e_i(\boldsymbol{\beta})|\mathbf{x}_i)$ converges to $E(e_i(\boldsymbol{\beta})|\mathbf{x}_i)$ for the true $\boldsymbol{\beta}$, a consistent estimator of $\boldsymbol{\beta}$ can be obtained by solving the estimating equation $\xi(\mathbf{b}) = \mathbf{0}$. To get the estimate of $\boldsymbol{\beta}$, the authors proposed iterative algorithms. The Tsui-Jewell-Wu's estimator, however, involves estimation of the product-limit estimator of the error distribution that could cause instability, especially in view of its presence as a denominator in the calculation of the expectation of residuals conditioned on $\mathbf{x}_i$. Lai and Ying (1992b) have established asymptotic theory for the least squares estimator satisfying (2.2) with the estimator $\hat{F}$ of error distribution function F being slightly modified. The two estimators have good statistical properties but the estimation procedures are complicated and computationally intensive because they involve non-parametric estimation of underlying error distribution function with truncated data.

Actually, estimation of the conditional expectation $E(e_i(\mathbf{b})|\mathbf{x}_i)$ does not need to rely on estimation of the error distribution and complicated calculations can be avoided. Let

$$\bar{e}(\mathbf{b}, t) = \frac{\sum_{j=1}^n e_j(\mathbf{b}) I(e_j(\mathbf{b}) \le t \le t_j(\mathbf{b}))}{\sum_{j=1}^n I(e_j(\mathbf{b}) \le t \le t_j(\mathbf{b}))}.$$

By strong law of large number, $\bar{e}(\boldsymbol{\beta}, t)$ converges almost surely to

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \varepsilon I(\varepsilon \le t \le s) F(d\varepsilon) G(ds)/[F(t)(1-G(t))]$$
$$= \int_{-\infty}^{\infty} \varepsilon I(\varepsilon \le t) F(d\varepsilon)/F(t) = E(\varepsilon^*|\varepsilon^* \le t) \quad (2.3)$$

where $G(.)$ is distribution function of $T^*(\boldsymbol{b}) = V^* - \boldsymbol{b}^T\mathbf{X}^*$. Define $r_i(\mathbf{b}) = e_i(\mathbf{b}) - \bar{e}(\mathbf{b}, t_i(\mathbf{b})), i = 1, ..., n$. It is easy to prove that $E(r_i(\boldsymbol{\beta})) = 0$. Therefore, $r_i(\mathbf{b})$ can be used to construct an optimal estimating function for $\boldsymbol{\beta}$. According to Godambe and Thompson (1984), an optimal estimating function takes the form of

$$\mathbf{U}_0(\mathbf{b}) = \sum_{i=1}^n \frac{\partial E(r_i(\mathbf{b})|\mathbf{x}_i)}{\partial \mathbf{b}} Var^{-1}(r_i(\mathbf{b})|\mathbf{x}_i) r_i(\mathbf{b})$$

At true $\boldsymbol{\beta}$, the conditional variance of $r_i(\boldsymbol{\beta})$, $Var(r_i(\boldsymbol{\beta})|\mathbf{x}_i) = Var(\varepsilon^*|\varepsilon^* \le t_i(\boldsymbol{\beta}))$, and

$$\frac{\partial E(r_i(\mathbf{b})|\mathbf{x}_i)}{\partial \mathbf{b}}|_{\mathbf{b}=\boldsymbol{\beta}} = -\Big\{\mathbf{x}_i - \mathbf{u}(\boldsymbol{\beta}, t_i(\boldsymbol{\beta}))\Big\} d(t_i(\boldsymbol{\beta}))$$

where $\mathbf{u}(\mathbf{b}, t) = E(\mathbf{X}^*|T^*(\mathbf{b}) \ge t),\ T^*(\mathbf{b})$ is defined in (2.3) and

$$d(t) = 1 + \Big\{E(\varepsilon^*|\varepsilon^* \le t) - t\Big\} f(t)/F(t).$$

The estimating function involves density function f of error ε^*, which may not be well estimated non-parametrically, especially when sample size n is not very large. However, when error ε^* has a Gaussian distribution, $d(t)$ is proportional to $Var(\varepsilon^*|\varepsilon^* \le t)$

hence the estimating function $\mathbf{U}_0(\mathbf{b})$ can be simplified.

Since $\mathbf{u}(\boldsymbol{\beta}, t)$ can be consistently estimated by $\bar{\mathbf{x}}(\boldsymbol{\beta}, t)$, where

$$\bar{\mathbf{x}}(\mathbf{b}, t) = \frac{\sum_{j=1}^{n} \mathbf{x}_j I(e_j(\mathbf{b}) \leq t \leq t_j(\mathbf{b}))}{\sum_{j=1}^{n} I(e_j(\mathbf{b}) \leq t \leq t_j(\mathbf{b}))},$$

we propose the following estimating function,

$$\mathbf{U}_1(\mathbf{b}) = \sum_{i=1}^{n} \Big\{ \mathbf{x_i} - \bar{\mathbf{x}}(\mathbf{b}, t_i(\mathbf{b})) \Big\} \Big\{ e_i(\mathbf{b}) - \bar{e}(\mathbf{b}, t_i(\mathbf{b})) \Big\}. \tag{2.4}$$

Alternatively, CTC considered the estimation based on the quasi-independence of $e_i(\boldsymbol{\beta})$ and $t_i(\boldsymbol{\beta})$. For the right truncated data of $(y_i, v_i, \mathbf{x}_i)$ with $y_i \leq v_i$, their estimating function for $\boldsymbol{\beta}$ is

$$\mathbf{U}(\mathbf{b}) = \sum_{i=1}^{n} \sum_{j=1}^{n} (\mathbf{x}_i - \mathbf{x}_j)(e_i(\mathbf{b}) - e_j(\mathbf{b}))\, I[max(e_i(\mathbf{b}), e_j(\mathbf{b})) \leq min(t_i(\mathbf{b}), t_j(\mathbf{b}))] \tag{2.5}$$

which has a symmetric form with mean zero at true $\boldsymbol{\beta}$. It is also an extension of estimating function for ordinary least squares estimation. By solving the estimating equation $\mathbf{U}(\mathbf{b}) = \mathbf{0}$, a least squares type estimator $\hat{\boldsymbol{\beta}}_{CTC}$ can be obtained at convergence of an iterative scheme. The estimation procedure does not involve non-parametric estimation of error distribution function and is computationally easy. With initial guess of $\boldsymbol{\beta} = \mathbf{0}$, the ordinary least squares estimator is one step iteration. Since the estimating function is a U-statistic, relevant theory has been applied to obtain the asymptotic distribution of the estimator. However, the authors showed by simulation that their estimator is less efficient than Lai and Ying's estimator in the case of Gaussian error.

Since

$$\begin{aligned}\mathbf{U}(\mathbf{b}) &= 2\sum_{i=1}^{n} m(\mathbf{b}, t_i(\mathbf{b}))[\{\mathbf{x_i} - \bar{\mathbf{x}}(\mathbf{b}, t_i(\mathbf{b}))\}\{e_i(\mathbf{b}) - \bar{e}(\mathbf{b}, t_i(\mathbf{b}))\} + \\ &\quad \{\bar{c}(\mathbf{b}, t_i(\mathbf{b})) - \bar{\mathbf{x}}(\mathbf{b}, t_i(\mathbf{b}))\bar{e}(\mathbf{b}, t_i(\mathbf{b}))\}] \\ &\approx 2\sum_{i=1}^{n} m(\mathbf{b}, t_i(\mathbf{b}))\{\mathbf{x_i} - \bar{\mathbf{x}}(\mathbf{b}, t_i(\mathbf{b}))\}\{e_i(\mathbf{b}) - \bar{e}(\mathbf{b}, t_i(\mathbf{b}))\},\end{aligned}$$

where $m(\mathbf{b}, t_i(\mathbf{b})) = \sum_{j=1}^{n} I(e_j(\mathbf{b}) \le t_i(\mathbf{b}) \le t_j(\mathbf{b}))$ and

$$\bar{c}(\mathbf{b}, t) = \frac{\sum_{j=1}^{n} \mathbf{x}_j e_j(\mathbf{b}) I(e_j(\mathbf{b}) \le t \le t_j(\mathbf{b}))}{\sum_{j=1}^{n} I(e_j(\mathbf{b}) \le t \le t_j(\mathbf{b}))}.$$

$\mathbf{U}(\mathbf{b})$ can be considered as a weighted version of $\mathbf{U}_1(\mathbf{b})$. Like $\mathbf{U}(\mathbf{b}), \mathbf{U}_1(\mathbf{b})$ involves only moment type estimators of conditional expectations.

2.2 Asymptotic Results

The main results are stated in the following theorems. Theorem 1 claims asymptotic normality of estimating function $\mathbf{U}_1(\boldsymbol{\beta})$. The corresponding estimator $\hat{\boldsymbol{\beta}}$, defined as a minimizer of $||\mathbf{U}_1(\mathbf{b})||$ with $||\mathbf{b}|| \le C$, where C is a known large positive constant and $||.||$ denotes the Euclidean norm, is proved to be consistent and asymptotically normal based on the normality of $\mathbf{U}_1(\mathbf{b})$ and theorem 2. By Theorem 3, we claim that the proposed estimator has the same asymptotic variance as that of Lai and Ying.

As recommended by Lai and Ying (1992a, 1992b), conditioned on n, we shall regard the observed $(y_i, v_i, \mathbf{x}_i), i = 1, \cdots, n$, as generated by a large sample $(y_i^*, v_i^*, \mathbf{x}_i^*), i = 1, \cdots, n^*$, where $n^* = \inf\{m : \sum_{i=1}^{m} I(y_i^* \le v_i^*) = n\}$. Define $t_i^*(\mathbf{b}) = v_i^* - \mathbf{b}\mathbf{x}_i^*$, $e_i^*(\mathbf{b}) = y_i^* - \mathbf{b}\mathbf{x}_i^*, i = 1, \cdots, n^*$. The following lemma reviews a basic martingale structure for truncated data, see Lai and Ying (1992a, 1992b).

<u>Lemma 1</u> Let F_s be the complete $\sigma-$field generated by $I(e_i^*(\boldsymbol{\beta}) \le t_i^*(\boldsymbol{\beta}))$, $I(u \le$

$t_i^*(\boldsymbol{\beta}))$, $I(u \leq y_i^*)$, $I(t_i^*(\boldsymbol{\beta}) \leq s)$, $\mathbf{x}_i^*, u \leq s, i = 1, 2, \cdots$. Let $N_i(s) = I(e_i(\boldsymbol{\beta}) \leq t_i(\boldsymbol{\beta}) \leq s)$ and $\Lambda(u) = -ln(1 - G(u))$, where G is defined in (2.3). Define $M_i(s) = N_i(s) - \int_{-\infty}^{s} Y_i(\boldsymbol{\beta}, u) d\Lambda(u)$, $-\infty < s < \infty$, with $Y_i(\mathbf{b}, t) = I(e_i(\mathbf{b}) \leq t \leq t_i(\mathbf{b}))$. Then $M_i(s), F_s, -\infty < s < \infty, i = 1, 2, ..., n$, are martingale with predictable variation process $< M_i > (s) = \int_{-\infty}^{s} Y_i(\boldsymbol{\beta}, u) d\Lambda(u)$, and the predictable covariation process $< M_i, M_j > (s) = 0$ for $i \neq j$.

As Lai and Ying (1992b), we need certain regularity conditions for the results in the theorems below. These conditions are listed as following,

C1. $||\mathbf{x}_i^*|| \leq B$ for all i and some constant B.

C2. F has a continuously differentiable density f such that $\int_{-\infty}^{\infty} u^2 F(du) < \infty$.

Theorem 1 If C1 and C2 are satisfied, then $n^{-1/2}\mathbf{U}_1(\boldsymbol{\beta}) \xrightarrow{D} N(\mathbf{0}, \boldsymbol{\Sigma})$, where the $p \times p$ matrix $\boldsymbol{\Sigma} = E\Big\{[\mathbf{x}_i - \mathbf{u}(\boldsymbol{\beta}, t_i(\boldsymbol{\beta}))][\mathbf{x}_i - \mathbf{u}(\boldsymbol{\beta}, t_i(\boldsymbol{\beta}))]^T Var(\varepsilon^* | \varepsilon^* \leq t_i(\boldsymbol{\beta}))\Big\}$. It can be consistently estimated by

$$\hat{\boldsymbol{\Sigma}} = n^{-1} \sum_{i=1}^{n} [\mathbf{x}_i - \bar{\mathbf{x}}(\tilde{\boldsymbol{\beta}}, t_i(\tilde{\boldsymbol{\beta}}))][\mathbf{x}_i - \bar{\mathbf{x}}(\tilde{\boldsymbol{\beta}}, t_i(\tilde{\boldsymbol{\beta}}))]^T \widehat{Var}(\varepsilon^* | \varepsilon^* \leq t_i(\tilde{\boldsymbol{\beta}})),$$

where $\tilde{\boldsymbol{\beta}}$ is any consistent estimator of $\boldsymbol{\beta}$, and

$$\widehat{Var}(\varepsilon^* | \varepsilon^* \leq t_i(\tilde{\boldsymbol{\beta}})) = \frac{\sum_{j=1}^{n} (e_j(\tilde{\boldsymbol{\beta}}) - \bar{e}(\tilde{\boldsymbol{\beta}}, t_i(\tilde{\boldsymbol{\beta}})))^2 I(e_j(\tilde{\boldsymbol{\beta}}) \leq t_i(\tilde{\boldsymbol{\beta}}) \leq t_j(\tilde{\boldsymbol{\beta}}))}{\sum_{j=1}^{n} I(e_j(\tilde{\boldsymbol{\beta}}) \leq t_i(\tilde{\boldsymbol{\beta}}) \leq t_j(\tilde{\boldsymbol{\beta}}))}.$$

The proof is outlined in the appendix.

Theorem 2 Suppose that conditions C1 and C2 are satisfied. Define $\hat{\boldsymbol{\beta}}$ by $||\mathbf{U}_1(\hat{\boldsymbol{\beta}})|| = \min_{\mathbf{b}:||\mathbf{b}|| \leq C}\{||\mathbf{U}_1(\mathbf{b})||\}$. Assume that for every $\delta > 0$, $\inf_{||\mathbf{b}|| \leq C, ||\mathbf{b} - \boldsymbol{\beta}|| \geq \delta}$ $\{||\bar{\mathbf{U}}_1(\mathbf{b})||\} > 0$, where $\bar{\mathbf{U}}_1(\mathbf{b}) = \lim_{n \to \infty} \mathbf{U}_1(\mathbf{b})/n$. By straight calculation, we have

$$\mathbf{A} = \frac{\partial}{\partial \mathbf{b}} \bar{\mathbf{U}}_1(\mathbf{b})|_{\mathbf{b}=\boldsymbol{\beta}} = E[\{\mathbf{x_i} - \mathbf{u}(\boldsymbol{\beta}, t_i(\boldsymbol{\beta}))\}\{\mathbf{x_i} - \mathbf{u}(\boldsymbol{\beta}, t_i(\boldsymbol{\beta}))\}^T d(t_i(\boldsymbol{\beta}))].$$

Then (i) $\hat{\boldsymbol{\beta}} \longrightarrow \boldsymbol{\beta}$ $a.s.$

(ii) If $\mathbf{A}$ is nonsingular, then $\sqrt{n}$ $(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{D} N(0, \mathbf{A}^{-1}\boldsymbol{\Sigma}\mathbf{A}^{-1})$, where $\boldsymbol{\Sigma}$ is defined in Theorem 1.

The proof of Theorem 2 may use the same arguments in the proof of Theorem 3 in Lai and Ying (1992b) and is omitted here.

Because $\mathbf{A}^{-1}\boldsymbol{\Sigma}\mathbf{A}^{-1}$ is a function of the probability density function f of the random error ε^*, a consistent estimator of the asymptotic variance of $\hat{\boldsymbol{\beta}}$ cannot be obtained without using smoothing methods. In practice, to construct a confidence set for the true $\boldsymbol{\beta}$, we may adopt Chen, Tsai and Chao's approach via the asymptotic behavior of $n^{-1/2}\mathbf{U}_1(\boldsymbol{\beta})$. It follows that an approximate (1-α)% confidence set could be $\{\boldsymbol{\beta} : n^{-1}\mathbf{U}_1(\boldsymbol{\beta})^T\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{U}_1(\boldsymbol{\beta}) \leq \chi^2_\alpha(p)\}$, where $\chi^2_\alpha(p)$ is the upper 100α percentile of the chi-square distribution with p degrees of freedom.

When error has a Gaussian distribution of $N(0, \sigma^2), d(t) = Var(\varepsilon^*|\varepsilon^* \leq t)/\sigma^2$. Therefore the asymptotic variance of $\hat{\boldsymbol{\beta}}$ can be simplified to $\sigma^4\boldsymbol{\Sigma}^{-1}$.

Lai and Ying (1992b) considered a modified version of the estimating function $\xi(\mathbf{b})$ in (2.2). They showed that under some regularity conditions, their estimating function $\xi(\mathbf{b})$ has good large sample property that $n^{-1/2}\xi(\mathbf{b}) \rightarrow N(\mathbf{0}, \mathbf{W})$ and the estimator $\tilde{\boldsymbol{\beta}}_{LY}$, the solution of $\xi(\mathbf{b}) = \mathbf{0}$, has $n^{1/2}(\tilde{\boldsymbol{\beta}}_{LY} - \boldsymbol{\beta}) \rightarrow N(\mathbf{0}, \mathbf{B}^{-1}\mathbf{W}\mathbf{B}^{-1})$, when $\mathbf{B}$ is non-singular, where

$$\begin{aligned}
\mathbf{W} &= \int_{-\infty}^{\infty} (t - \int_{-\infty}^{t} uF(du)/F(t))^2\, Var(\mathbf{X}^*|T^*(\boldsymbol{\beta}) \geq t)(1 - G(t))F(dt)/K, \\
\mathbf{B} &= \int_{-\infty}^{\infty} (t - \int_{-\infty}^{t} uF(du)/F(t))\, Var(\mathbf{X}^*|T^*(\boldsymbol{\beta}) \geq t)(\frac{f'(t)}{f(t)} - \frac{f(t)}{F(t)})\frac{(1 - G(t))}{K}F(dt), \\
K &= \int_{-\infty}^{\infty} F(t)G(dt)
\end{aligned}$$

and $T^*(\boldsymbol{\beta})$ is defined in (2.3).

The following theorem, to be proved in the appendix, establishes that $\mathbf{B}^{-1}\mathbf{W}\mathbf{B}^{-1} = \mathbf{A}^{-1}\boldsymbol{\Sigma}\mathbf{A}^{-\mathbf{1}}$.

<u>Theorem 3</u> If C2 holds, then $\boldsymbol{\Sigma} = \mathbf{W}$ and $\mathbf{A} = \mathbf{B}$.

Note that $\mathbf{A}$ is a function of the probability function f, but $\mathbf{B}$ is a function of f and f'. $\mathbf{A}$ is much easier to be estimated.

2.3 Estimating procedure

Similar to Chen, Tsai and Chao's procedure, we may obtain a least squares estimator $\boldsymbol{\beta}$, as a solution to the estimating equation $\mathbf{U}_1(\mathbf{b}) = \mathbf{0}$, at convergence of an iterative scheme. Suppose that $\mathbf{b}^{(k)}$ is the estimate of $\boldsymbol{\beta}$ obtained at the kth iteration. To update the estimate, we have

$$\begin{aligned} \mathbf{b}^{(k+1)} = & \quad \{\sum_{i=1}^{n}(\mathbf{x}_i - \bar{\mathbf{x}}(\mathbf{b}^{(k)}, t_i(\mathbf{b}^{(k)})))(\mathbf{x}_i - \bar{\mathbf{x}}(\mathbf{b}^{(k)}, t_i(\mathbf{b}^{(k)})))^T\}^{-1} \\ \times & \quad \sum_{i=1}^{n}(\mathbf{x}_i - \bar{\mathbf{x}}(\mathbf{b}^{(k)}, t_i(\mathbf{b}^{(k)})))(y_i - \bar{y}(\mathbf{b}^{(k)}, t_i(\mathbf{b}^{(k)}))) \end{aligned} \tag{2.6}$$

where

$$\bar{y}(\mathbf{b}, t) = \frac{\sum_{i=1}^{n} y_i I(e_i(\mathbf{b}) \le t \le t_i(\mathbf{b}))}{\sum_{i=1}^{n} I(e_i(\mathbf{b}) \le t \le t_i(\mathbf{b}))}.$$

The iteration starts with an initial guess of $\boldsymbol{\beta}$, and ends when certain convergence criterion is met.

It is obvious that the estimating procedure is much simpler than Tsui *et al* and Lai and Ying's methods.

3. A SIMULATION STUDY

To examine the finite sample performance of the proposed estimators, we conducted a simulation study. The parameter settings of the following nine cases are equivalent to the ones in Table 1 of Lai and Ying (1992b). For these nine cases, the random variable V^* is a fixed constant taking values either 0.5, 0.75, or 1, and σ is chosen to be either 0.3, 0.5, or 1. Five hundred and fifty data sets were generated under the simple regression model $Y^* = X^* + \varepsilon^*$, i.e. $\beta = 1$, with X^* from the uniform distribution of $U(0, 1)$ and ε^* from the Gaussian distribution of $N(0, \sigma^2)$. In each data set of triplets (y_i^*, v_i^*, x_i^*) with $y_i^* \leq v_i^*$, 50 observations were randomly selected for analysis.

In the simulation study, the criterion for the convergence of the iterative scheme is $|b^{(k)} - b^{(k+1)}| < 10^{-4}$. It produced a convergent solution of β estimate for the CTC estimator $\hat{\beta}_{CTC}$ and the proposed estimator $\hat{\beta}$ by (2.6) within 10 iterations on average. There is no divergent problem with CTC's procedure in any of the nine cases. The proposed procedure produces few ($\leq$ 1.09%) oscillated estimates.

As pointed by CTC for their estimator, when $p = 1$, for a sufficient large number M, if $||\mathbf{b}|| > M$, $I(e_j(\mathbf{b}) \leq t_i(\mathbf{b}) \leq t_j(\mathbf{b})) = 0$ for all $i \neq j$, the estimating function in (2.5) will have value of zero. Therefore, some estimates may converge to the extreme values that are far away from the true value of $\boldsymbol{\beta}$. In our simulation this phenomenon occurs for large σ, especially when $\sigma = 1$. To reduce the effect of these extreme estimates on the summary statistics, we present the median, the interquartile range, the range between 10% and 90% percentiles, the range between 5% and 95% percentiles, and the range between 1% and 99% percentiles on the converged estimates in table 1. In table 2 we present the bias and the mean square error for the converged estimates, which are trimmed for 2% extreme values with large absolute numbers.

Table 1 and Table 2 are about here

As expected, the ordinary least estimator $\hat{\beta}_{OLS}$ underestimates the true $\beta\,(\beta = 1)$. The $\hat{\beta}_{OLS}$ estimates spread tightly around its average. The median and the mean are far from the true value of β in all of the nine cases. The absolute bias and mean square error increase as truncation proportion $P(Y^* > V^*)$ or error standard deviation σ increases. The CTC estimator $\hat{\beta}_{CTC}$, however, always has biases close to zero while its mean square error increases with $P(Y^* > V^*)$ or σ increasing. The proposed estimator $\hat{\beta}$ can compete well with $\hat{\beta}_{CTC}$. Like $\hat{\beta}_{CTC}$, $\hat{\beta}$ always has negligible bias but the variation increases with $P(Y^* > V^*)$ or σ increasing. Compared with $\hat{\beta}_{CTC}$, $\hat{\beta}$ estimates spread tighter around its average. Table 1 shows that in the nine cases, $\hat{\beta}$ and $\hat{\beta}_{CTC}$ have similar interquartile ranges and similar ranges between 10% and 90% percentiles, while the range between 5% and 95 % percentiles for $\hat{\beta}$ are narrower than that for $\hat{\beta}_{CTC}$. The range between 1% and 99% percentiles for $\hat{\beta}$ is smaller than that for $\hat{\beta}_{CTC}$ in all the cases. Besides, the $\hat{\beta}$ estimates rarely have extreme values even when σ is large. The fact that $\hat{\beta}$ has a mean square error always smaller than $\hat{\beta}_{CTC}$ indicates that $\hat{\beta}$ is more efficient than $\hat{\beta}_{CTC}$.

The findings in the simulation study are consistent with our large sample results discussed in the previous section. In summary, the estimator $\hat{\beta}$ is better than $\hat{\beta}_{CTC}$ in the finite sample performance and it behaves well even with a moderate sample size of $n = 50$.

4. DISCUSSION

To estimate the coefficients of the covariates in linear regression model with truncated data, CTC considered an unbiased estimating equation to obtain a consistent least squares estimator. Under Gaussian error assumption, their estimator, however, is less efficient than those proposed by Tsui et al (1988). We modify CTC's estimating equation to obtain a consistent least squares estimator which is efficient under Gaussian error assumption. As CTC's estimator, the proposed estimation procedure has advantages that Tsui-Jewell-Wu's estimator and Lai-Ying's estimator do not have: the computational simplicity and the invariance for a location shift of dependent variable and/or independent variables. Furthermore, the estimator we propose has the same asymptotic distribution as the estimators proposed by Lai and Ying (1992b).

It is interesting to note that by replacing $E(\varepsilon^*|\varepsilon^* \leq t_i(\mathbf{b}))$ in (2.1) with the consistent moment estimator $\bar{e}_i(\mathbf{b}, t_i(\mathbf{b}))$, we can get another unbiased estimating function,

$$\mathbf{U}_2(\mathbf{b}) = \sum_{i=1}^{n} \mathbf{x}_i \{e_i(\mathbf{b}) - \bar{e}_i(\mathbf{b}, t_i(\mathbf{b}))\}.$$

By solving the equation of $\mathbf{U}_2(\mathbf{b}) = \mathbf{0}$, we can also obtain a least squares type estimator $\hat{\boldsymbol{\beta}}_2$ at convergence of an iterative scheme. Although the estimation procedure is easy to perform, the estimator $\hat{\boldsymbol{\beta}}_2$ is not as good as $\hat{\boldsymbol{\beta}}$ (the solution of $\mathbf{U}_1(\mathbf{b}) = \mathbf{0}$) in several aspects. It is clear that $\hat{\boldsymbol{\beta}}_2$ is no longer invariant when a independent variable has a constant shift. Moreover, it is not efficient when error has Gaussian distribution. In our simulation study, the $\hat{\boldsymbol{\beta}}_2$ estimates have medians close to the true $\boldsymbol{\beta}$, but have larger mean square errors than $\hat{\boldsymbol{\beta}}$ estimates in all nine cases. The results are not presented in the tables.

We note that the proposed estimating function $\mathbf{U}_1(\mathbf{b})$ is, in general, a discontinuous function of $\mathbf{b}$ and that the solution of $\mathbf{U}_1(\mathbf{b}) = \mathbf{0}$ may not always exist. It explains why we have some oscillated estimates in the simulation study.

APPENDIX

Throughout the appendix, for any p-vector $\mathbf{x}$, we use $\mathbf{x}^{(k)}$ to denote the k-th component of $\mathbf{x}$, $k = 1, \cdots, p$. For any $p \times p$ matrix $\mathbf{X}$, we use $\mathbf{X}_{ij}$ for the element at the i-th row and the j-th column.

<u>Sketch proof of Theorem 1.</u>

Define $Y_j(\mathbf{b}, t) = I(e_j(\mathbf{b}) \le t \le t_j(\mathbf{b}))$, $m(\mathbf{b}, t) = \sum_{j=1}^{n} Y_j(\mathbf{b}, t)$ and $K = \int_{-\infty}^{\infty} F(t)G(dt)$.

For $k = 1$, ..., p, let

$$H_k^e(e, x, t) = \frac{1}{n} \sum_{i=1}^{n} I(e_i(\boldsymbol{\beta}) \le e, \mathbf{x}_i^{(k)} \le x, t_i(\boldsymbol{\beta}) \le t),$$

$$H_k(e, x, t) = P(e_i(\boldsymbol{\beta}) \le e, \mathbf{x}_i^{(k)} \le x, t_i(\boldsymbol{\beta}) \le t),$$

$$\mathcal{L}_k(x, t) = P(\mathbf{x}_i^{(k)} \le x, t_i(\boldsymbol{\beta}) \le t),$$

$$\mathcal{L}_k^*(x, t) = P(\mathbf{X}^{*(k)} \le x, T^*(\boldsymbol{\beta}) \le t).$$

Since $H_k(de, dx, dt) = I(e \le t) f(e) \mathcal{L}_k^*(dx, dt)/K$, we have

$$\int_{e \in \mathbb{R}} H_k(de, dx, dt) = \mathcal{L}_k(dx, dt) = F(t)\mathcal{L}_k^*(dx, dt)/K.$$

With

$$\gamma_1(t) = E(\sum_{i=1}^{n} e_i(\boldsymbol{\beta}) Y_i(\boldsymbol{\beta}, t)/n) = \int_{-\infty}^{t} uF(du)(1 - G(t))/K$$

and

$$\gamma_2(t) = E(m(\boldsymbol{\beta}, t)/n) = \frac{F(t)(1 - G(t))}{K},$$

we have

$$\gamma(t) = \gamma_1(t)/\gamma_2(t) = \int_{-\infty}^{t} u\, F(du)/F(t) = E(\varepsilon^* | \varepsilon^* \le t).$$

By Lemma 1 and the uniform boundedness of $\mathbf{x}_i^*$,

$$\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\bar{\mathbf{x}}(\boldsymbol{\beta},t_i(\boldsymbol{\beta}))(e_i(\boldsymbol{\beta})-\bar{e}(\boldsymbol{\beta},t_i(\boldsymbol{\beta})))\\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_{-\infty}^{\infty}\bar{\mathbf{x}}(\boldsymbol{\beta},t)(e_i(\boldsymbol{\beta})-\bar{e}(\boldsymbol{\beta},t))dN_i(t)\\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_{-\infty}^{\infty}\bar{\mathbf{x}}(\boldsymbol{\beta},t)(e_i(\boldsymbol{\beta})-\bar{e}(\boldsymbol{\beta},t))dM_i(t)\\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_{-\infty}^{\infty}\mathbf{u}(\boldsymbol{\beta},t)(e_i(\boldsymbol{\beta})-\bar{e}(\boldsymbol{\beta},t))dM_i(t)+o_p(1)\\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathbf{u}(\boldsymbol{\beta},t_i(\boldsymbol{\beta}))(e_i(\boldsymbol{\beta})-\bar{e}(\boldsymbol{\beta},t_i(\boldsymbol{\beta})))+o_p(1).
\end{aligned}$$

For $k=1,\cdots,\ p$, we have

$$\begin{aligned}
\frac{1}{\sqrt{n}}\mathbf{U}_1^{(k)}(\boldsymbol{\beta}) &= \sqrt{n}\int_{I\!R^3}(x-\mathbf{u}^{(k)}(\boldsymbol{\beta},t))(e-\bar{e}(\boldsymbol{\beta},t))H_k^e(de,dx,dt)+o_p(1)\\
&= \sqrt{n}\int_{I\!R^3}(x-\mathbf{u}^{(k)}(\boldsymbol{\beta},t))(e-\frac{\gamma_1(t)}{\gamma_2(t)})\{(H_k^e-H_k)(de,dx,dt)\}\\
&\quad -\sqrt{n}\int_{I\!R^3}(x-\mathbf{u}^{(k)}(\boldsymbol{\beta},t))(\frac{\sum_{i=1}^n e_i(\boldsymbol{\beta})Y_i(\boldsymbol{\beta},t)/n-\gamma_1(t)}{\gamma_2(t)})H_k(de,dx,dt)\\
&\quad +\sqrt{n}\int_{I\!R^3}(x-\mathbf{u}^{(k)}(\boldsymbol{\beta},t))(\frac{\gamma_1(t)}{\gamma_2(t)}\ \frac{m(\boldsymbol{\beta},t)/n-\gamma_2(t)}{\gamma_2(t)})H_k(de,dx,dt)+o_p(1)\\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\{(\boldsymbol{\eta}_{1i}^{(k)}-E(\boldsymbol{\eta}_{1i}^{(k)})-(\boldsymbol{\eta}_{2i}^{(k)}-E(\boldsymbol{\eta}_{2i}^{(k)})+(\boldsymbol{\eta}_{3i}^{(k)}-E(\boldsymbol{\eta}_{3i}^{(k)})\}+o_p(1),
\end{aligned}$$

where

$$\begin{aligned}
\boldsymbol{\eta}_{1i}^{(k)} &= (\mathbf{x}_i^{(k)}-\mathbf{u}^{(k)}(\boldsymbol{\beta},t_i(\boldsymbol{\beta})))(e_i(\boldsymbol{\beta})-\gamma(t_i(\boldsymbol{\beta})),\\
\boldsymbol{\eta}_{2i}^{(k)} &= \int_{I\!R^3}(x-\mathbf{u}^{(k)}(\boldsymbol{\beta},t))\frac{e_i(\boldsymbol{\beta})Y_i(\boldsymbol{\beta},t)}{\gamma_2(t)}H_k(de,dx,dt)\\
&= \int_{I\!R^2}(x-\mathbf{u}^{(k)}(\boldsymbol{\beta},t))\frac{e_i(\boldsymbol{\beta})Y_i(\boldsymbol{\beta},t)}{1-G(t)}\mathcal{L}_k^*(dx,dt),\\
\boldsymbol{\eta}_{3i}^{(k)} &= \int_{I\!R^3}(x-\mathbf{u}^{(k)}(\boldsymbol{\beta},t))\frac{\gamma_1(t)}{\gamma_2^2(t)}Y_i(\boldsymbol{\beta},t)H_k(de,dx,dt)\\
&= \int_{I\!R^2}(x-\mathbf{u}^{(k)}(\boldsymbol{\beta},t))\frac{\gamma(t)Y_i(\boldsymbol{\beta},t)}{1-G(t)}\mathcal{L}_k^*(dx,dt).
\end{aligned}$$

Therefore, by Central Limit Theorem, $\frac{1}{\sqrt{n}}\mathbf{U}_1(\boldsymbol{\beta})$ converge to a multivariate normal distribution with mean $\mathbf{0}$ and variance - covariance matrix $Var(\boldsymbol{\eta}_{1i}-\boldsymbol{\eta}_{2i}+\boldsymbol{\eta}_{3i})$. Below we will show that $Var(\boldsymbol{\eta}_{1i} - \boldsymbol{\eta}_{2i} + \boldsymbol{\eta}_{3i}) = \boldsymbol{\Sigma}$.

Since $E(e_i(\boldsymbol{\beta})|\mathbf{x}_i) = \gamma(t_i(\boldsymbol{\beta}))$ and $Var(e_i(\boldsymbol{\beta})|\mathbf{x}_i) = Var(\varepsilon^*|\varepsilon^* \leq t_i(\boldsymbol{\beta}))$, we have

$$\begin{aligned}
Var(\boldsymbol{\eta}_{1i}) &= E\Big\{\Big[\mathbf{x}_i - \mathbf{u}(\boldsymbol{\beta}, t_i(\boldsymbol{\beta}))\Big]\Big[\mathbf{x}_i - \mathbf{u}(\boldsymbol{\beta}, t_i(\boldsymbol{\beta}))\Big]^T \Big[e_i(\boldsymbol{\beta}) - \gamma(t_i(\boldsymbol{\beta}))\Big]^2\Big\} \\
&= E\Big\{\Big[\mathbf{x}_i - \mathbf{u}(\boldsymbol{\beta}, t_i(\boldsymbol{\beta}))\Big]\Big[\mathbf{x}_i - \mathbf{u}(\boldsymbol{\beta}, t_i(\boldsymbol{\beta}))\Big]^T E\Big[(e_i(\boldsymbol{\beta}) - \gamma(t_i(\boldsymbol{\beta})))^2|\mathbf{x}_i\Big]\Big\} \\
&= E\Big\{\Big[\mathbf{x}_i - \mathbf{u}(\boldsymbol{\beta}, t_i(\boldsymbol{\beta}))\Big]\Big[\mathbf{x}_i - \mathbf{u}(\boldsymbol{\beta}, t_i(\boldsymbol{\beta}))\Big]^T Var(e_i(\boldsymbol{\beta})|\mathbf{x}_i)\Big\} \\
&= E\Big\{\Big[\mathbf{x}_i - \mathbf{u}(\boldsymbol{\beta}, t_i(\boldsymbol{\beta}))\Big]\Big[\mathbf{x}_i - \mathbf{u}(\boldsymbol{\beta}, t_i(\boldsymbol{\beta}))\Big]^T Var(\varepsilon^*|\varepsilon^* \leq t_i(\boldsymbol{\beta}))\Big\} \\
&= \boldsymbol{\Sigma}.
\end{aligned}$$

Meanwhile, for $k = 1, \cdots, p$, we have

$$\boldsymbol{\eta}_{2i}^{(k)} - \boldsymbol{\eta}_{3i}^{(k)} = \int_{I\!R^2} \boldsymbol{\zeta}_i^{(k)}(x, s)\mathcal{L}_k^*(dx, ds),$$

where $\boldsymbol{\zeta}_i^{(k)}(x, s) = [x - \mathbf{u}^{(k)}(\boldsymbol{\beta}, s)][e_i(\boldsymbol{\beta}) - \gamma(s)]Y_i(\boldsymbol{\beta}, s)/[1 - G(s)]$.

For $k = 1, ..., p$, $l = 1, ..., p$, because

$$\begin{aligned}
& Cov[\boldsymbol{\zeta}_i^{(k)}(x, s), \boldsymbol{\zeta}_i^{(l)}(z, t)] \\
= \;& \frac{x - \mathbf{u}^{(k)}(\boldsymbol{\beta}, s)}{1 - G(s)} \cdot \frac{z - \mathbf{u}^{(l)}(\boldsymbol{\beta}, t)}{1 - G(t)} E\Big\{[e_i(\boldsymbol{\beta}) - \gamma(s)][e_i(\boldsymbol{\beta}) - \gamma(t)]Y_i(\boldsymbol{\beta}, s)Y_i(\boldsymbol{\beta}, t)\Big\} \\
= \;& \frac{[x - \mathbf{u}^{(k)}(\boldsymbol{\beta}, s)][z - \mathbf{u}^{(l)}(\boldsymbol{\beta}, t)]}{K[1 - G(s \wedge t)]} \int_{-\infty}^{s\wedge t} [e - \gamma(s)][e - \gamma(t)]F(de),
\end{aligned}$$

and $Cov[\boldsymbol{\eta}_{1i}^{(k)}, \boldsymbol{\zeta}_i^{(l)}(x, s)]$

$$\begin{aligned}
&= \frac{x - \mathbf{u}^{(l)}(\boldsymbol{\beta}, s)}{1 - G(s)} E\Big\{[\mathbf{x}_i^{(k)} - \mathbf{u}^{(k)}(\boldsymbol{\beta}, t_i(\boldsymbol{\beta}))][e_i(\boldsymbol{\beta}) - \gamma(t_i(\boldsymbol{\beta})][e_i(\boldsymbol{\beta}) - \gamma(s)]Y_i(\boldsymbol{\beta}, s)\Big\} \\
&= \frac{x - \mathbf{u}^{(l)}(\boldsymbol{\beta}, s)}{1 - G(s)} \int_{I\!R^3} \Big\{[z - \mathbf{u}^{(k)}(\boldsymbol{\beta}, t)][e - \gamma(t)][e - \gamma(s)]I(e \leq t \leq s)\Big\} H_k(de, dz, dt) \\
&= \frac{x - \mathbf{u}^{(l)}(\boldsymbol{\beta}, s)}{K[1 - G(s)]} \int_{I\!R^3} \Big\{[z - \mathbf{u}^{(k)}(\boldsymbol{\beta}, t)][e - \gamma(t)][e - \gamma(s)]I(e \leq t \leq s)\Big\} \mathcal{L}^*(dz, dt)F(de) \\
&= \frac{x - \mathbf{u}^{(l)}(\boldsymbol{\beta}, s)}{K[1 - G(s)]} \int_{I\!R^2} \int_{-\infty}^{t} [e - \gamma(t)][e - \gamma(s)]F(de)I(t \leq s)[z - \mathbf{u}^{(k)}(\boldsymbol{\beta}, t)]\mathcal{L}^*(dz, dt),
\end{aligned}$$

we have

$$
\begin{aligned}
& Cov[\boldsymbol{\eta}_{2i}^{(k)} - \boldsymbol{\eta}_{3i}^{(k)}, \boldsymbol{\eta}_{2i}^{(l)} - \boldsymbol{\eta}_{3i}^{(l)}] \\
= & \int_{I\!R^2}\int_{I\!R^2} \frac{[x-\mathbf{u}^{(k)}(\boldsymbol{\beta},s)][z-\mathbf{u}^{(l)}(\boldsymbol{\beta},t)]}{K[1-G(s\wedge t)]} \int_{-\infty}^{s\wedge t} [e-\gamma(s)][e-\gamma(t)]F(de)\mathcal{L}_k^*(dx,ds)\mathcal{L}_l^*(dz,dt) \\
= & \; 2\; Cov[\boldsymbol{\eta}_{1i}^{(k)}, \boldsymbol{\eta}_{2i}^{(l)} - \boldsymbol{\eta}_{3i}^{(l)}].
\end{aligned}
$$

<u>Proof of Theorem 3</u>

Let $\phi(t) = F(t)\; Var(\varepsilon^*|\varepsilon^* \le t)$, and $\psi(t) = F(t)d(t)$.

For $k = 1, ..., p$, $l = 1, ..., p$, let

$$
\begin{aligned}
\mathcal{U}_{k,l}(x,z,t) &= P(\mathbf{x}_i^{(k)} \le x, \mathbf{x}_i^{(l)} \le z, t_i(\boldsymbol{\beta}) \le t), \\
\mathcal{U}_{k,l}^*(x,z,t) &= P(\mathbf{X}^{*(k)} \le x, \mathbf{X}^{*(l)} \le z, T^*(\boldsymbol{\beta}) \le t).
\end{aligned}
$$

The relations under condition C2, to be referred, are:

$$
\mathcal{U}_{k,l}(dx,dz,dt) = F(t)\; \mathcal{U}_k^*(dx,dz,dt)/K \qquad (A.1)
$$

$$
\frac{d}{dt}\phi(t) = [t-\gamma(t)]^2 f(t) \qquad (A.2)
$$

$$
\frac{d}{dt}\psi(t) = [t-\gamma(t)]F(t)\frac{d}{dt}\left[\frac{f(t)}{F(t)}\right] \qquad (A.3)
$$

By $(A.1)$ and $(A.2)$, we have

$$
\begin{aligned}
K\cdot\boldsymbol{\Sigma}_{kl} &= KE\Big\{\Big[\mathbf{x}_i^{(k)} - \mathbf{u}^{(k)}(\boldsymbol{\beta}, t_i(\boldsymbol{\beta}))\Big]\Big[\mathbf{x}_i^{(l)} - \mathbf{u}^{(l)}(\boldsymbol{\beta}, t_i(\boldsymbol{\beta}))\Big] Var(\varepsilon^*|\varepsilon^* \le t_i(\boldsymbol{\beta}))\Big\} \\
&= K\int_{I\!R^3}\Big[x-\mathbf{u}^{(k)}(\boldsymbol{\beta},t)\Big]\Big[z-\mathbf{u}^{(l)}(\boldsymbol{\beta},t)\Big] Var(\varepsilon^*|\varepsilon^*\le t)\mathcal{U}_{k,l}(dx,dz,dt) \\
&= \int_{I\!R^3}\Big[x-\mathbf{u}^{(k)}(\boldsymbol{\beta},t)\Big]\Big[z-\mathbf{u}^{(l)}(\boldsymbol{\beta},t)\Big]\phi(t)\mathcal{U}_{k,l}^*(dx,dz,dt),
\end{aligned}
$$

$$
\begin{aligned}
K\cdot\mathbf{W}_{kl} &= \int_{-\infty}^{\infty}\Big[s-\gamma(s)\Big]^2 Cov(\mathbf{X}^{*(k)},\mathbf{X}^{*(l)}|T^*(\boldsymbol{\beta})\ge s)\Big[1-G(s)\Big]F(ds) \\
&= \int_{-\infty}^{\infty}\Big[s-\gamma(s)\Big]^2\Big\{\int_{I\!R^2}\Big[x-\mathbf{u}^{(k)}(\boldsymbol{\beta},s)\Big]\Big[z-\mathbf{u}^{(l)}(\boldsymbol{\beta},s)\Big]\int_{t\ge s}\mathcal{U}_{k,l}^*(dx,dz,dt)\Big\}F(ds)
\end{aligned}
$$

$$
\begin{aligned}
&= \int_{I\!R^3} \left\{ \int_{-\infty}^{t} \Big[x - \mathbf{u}^{(k)}(\boldsymbol{\beta}, s)\Big]\Big[z - \mathbf{u}^{(l)}(\boldsymbol{\beta}, s)\Big] d\phi(s) \right\} \mathcal{U}^*_{k,l}(dx, dz, dt) \\
&= K \cdot \boldsymbol{\Sigma}_{kl} + \mathbf{R}_{kl},
\end{aligned}
$$

where

$$
\begin{aligned}
\mathbf{R}_{kl} &= \int_{I\!R^3} \left\{ \int_{-\infty}^{t} \phi(s) d\Big[x - \mathbf{u}^{(k)}(\boldsymbol{\beta}, s)\Big]\Big[z - \mathbf{u}^{(l)}(\boldsymbol{\beta}, s)\Big] \right\} \mathcal{U}^*_{k,l}(dx, dz, dt) \\
&= -\int_{-\infty}^{\infty} \left\{ \int_{I\!R^3} I(t \geq s)\Big[z - \mathbf{u}^{(l)}(\boldsymbol{\beta}, s)\Big]\mathcal{U}^*_{k,l}(dx, dz, dt) \right\} \phi(s) d\mathbf{u}^{(k)}(\boldsymbol{\beta}, s) \\
&\quad - \int_{-\infty}^{\infty} \left\{ \int_{I\!R^3} I(t \geq s)\Big[x - \mathbf{u}^{(k)}(\boldsymbol{\beta}, s)\Big]\mathcal{U}^*_{k,l}(dx, dz, dt) \right\} \phi(s) d\mathbf{u}^{(l)}(\boldsymbol{\beta}, s) \\
&= -\int_{-\infty}^{\infty} \left\{ \int_{I\!R^2} I(t \geq s)\Big[z - \mathbf{u}^{(l)}(\boldsymbol{\beta}, s)\Big]\mathcal{L}^*_{l}(dz, dt) \right\} \phi(s) d\mathbf{u}^{(k)}(\boldsymbol{\beta}, s) \\
&\quad - \int_{-\infty}^{\infty} \left\{ \int_{I\!R^2} I(t \geq s)\Big[x - \mathbf{u}^{(k)}(\boldsymbol{\beta}, s)\Big]\mathcal{L}^*_{k}(dx, dt) \right\} \phi(s) d\mathbf{u}^{(l)}(\boldsymbol{\beta}, s) \\
&= -\int_{-\infty}^{\infty} \Big[1 - G(s)\Big] E\Big(\mathbf{X}^{*(l)} - \mathbf{u}^{(l)}(\boldsymbol{\beta}, s) | T^*(\boldsymbol{\beta}) \geq s\Big) \phi(s) d\mathbf{u}^{(k)}(\boldsymbol{\beta}, s) \\
&\quad - \int_{-\infty}^{\infty} \Big[1 - G(s)\Big] E\Big(\mathbf{X}^{*(k)} - \mathbf{u}^{(k)}(\boldsymbol{\beta}, s) | T^*(\boldsymbol{\beta}) \geq s\Big) \phi(s) d\mathbf{u}^{(l)}(\boldsymbol{\beta}, s) \\
&= 0.
\end{aligned}
$$

Similarly, by $(A.1)$ and $(A.3)$, we can show

$$
K \cdot \mathbf{A}_{kl} = K \cdot \mathbf{B}_{kl} = \int_{I\!R^3} \Big[x - \mathbf{u}^{(k)}(\boldsymbol{\beta}, t)\Big]\Big[z - \mathbf{u}^{(l)}(\boldsymbol{\beta}, t)\Big] \psi(t) \mathcal{U}^*_{k,l}(dx, dz, dt). \qquad \square
$$

REFERENCE

Bhattacharya, P. K., Chernoff, H., and Yang, S. S. (1983). Nonparametric estimation of the slope of a truncated regression. *Ann. Statist.*, **11**, 505-14.

Chen, C.-H., Tsai, W.-Y., and Chao, W.-H.(1996). The product-limit correlation coefficient and liner regression for truncated data. *J. Am. Statist. Assoc.* **91**, 1181-86.

Godambe, V. P. and Thompson, M. E. (1984). Robust estimation through estimating equations. *Biometrika* **71**, 115-25.

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). Asymptotic theory of a bias-corrected least squares estimator in truncated regression. *Statistica Sinica* **2**, 519-539.

Powell, J. L. (1986). Symmetrically trimmed least square estimation for Tobit models. *Econometrica* **54**, 1435-60.

Tsui, K. L., Jewell, N. P., and Wu, C. F. J. (1988). A nonparametric approach to the truncated regression problem. *J. Am. Statist. Assoc.* **83**, 785-92.

Table 1. Median and range between percentiles for estimates converged

σ	y_0	$p(y > t_0)$		Median	25-75 %	10-90 %	5-95%	1-99 %
0.3	1	0.1196	$\hat{\boldsymbol{\beta}}_{OLS}$	0.7952	0.1820	0.3480	0.4479	0.5850
			$\hat{\boldsymbol{\beta}}_{CTC}$	0.9822	0.2641	0.5360	0.7081	1.0283
			$\hat{\boldsymbol{\beta}}$	0.9810	0.2684	0.5354	0.7142	0.9729
0.3	0.75	0.2834	$\hat{\boldsymbol{\beta}}_{OLS}$	0.5967	0.1799	0.3398	0.4334	0.6064
			$\hat{\boldsymbol{\beta}}_{CTC}$	0.9952	0.4058	0.8170	1.0642	1.4796
			$\hat{\boldsymbol{\beta}}$	0.9863	0.3988	0.7751	1.0420	1.4176
0.3	0.5	0.5	$\hat{\boldsymbol{\beta}}_{OLS}$	0.4630	0.1560	0.3337	0.4281	0.6055
			$\hat{\boldsymbol{\beta}}_{CTC}$	0.9868	0.4944	1.0437	1.4036	2.3463
			$\hat{\boldsymbol{\beta}}$	0.9767	0.4492	0.9547	1.3142	2.0266
0.5	1	0.1952	$\hat{\boldsymbol{\beta}}_{OLS}$	0.6195	0.2799	0.5126	0.6376	0.8923
			$\hat{\boldsymbol{\beta}}_{CTC}$	0.9612	0.4681	0.9199	1.2418	1.9534
			$\hat{\boldsymbol{\beta}}$	0.9667	0.4805	0.9147	1.1941	1.8132
0.5	0.75	0.3342	$\hat{\boldsymbol{\beta}}_{OLS}$	0.5175	0.2394	0.4561	0.6036	0.8908
			$\hat{\boldsymbol{\beta}}_{CTC}$	0.9955	0.5775	1.1690	1.6116	2.7709
			$\hat{\boldsymbol{\beta}}$	0.9964	0.5281	1.1278	1.5434	2.4958
0.5	0.5	0.5	$\hat{\boldsymbol{\beta}}_{OLS}$	0.4010	0.2413	0.4504	0.5530	0.8637
			$\hat{\boldsymbol{\beta}}_{CTC}$	0.9813	0.7551	1.4238	1.9605	3.2710
			$\hat{\boldsymbol{\beta}}$	0.9565	0.7051	1.4259	1.8212	2.8880
1	1	0.3156	$\hat{\boldsymbol{\beta}}_{OLS}$	0.3594	0.4095	0.7864	1.0230	1.4100
			$\hat{\boldsymbol{\beta}}_{CTC}$	1.0107	1.1111	2.1243	2.7581	4.3901
			$\hat{\boldsymbol{\beta}}$	1.0058	1.1058	2.1302	2.8361	4.3748
1	0.75	0.4052	$\hat{\boldsymbol{\beta}}_{OLS}$	0.4258	0.4444	0.8613	1.1838	1.6940
			$\hat{\boldsymbol{\beta}}_{CTC}$	0.9832	1.1557	2.5663	3.3823	6.0838
			$\hat{\boldsymbol{\beta}}$	0.9827	1.2089	2.3827	3.1368	5.1245
1	0.5	0.5	$\hat{\boldsymbol{\beta}}_{OLS}$	0.5216	0.4894	0.9557	1.1909	1.6995
			$\hat{\boldsymbol{\beta}}_{CTC}$	0.9482	1.2278	2.5233	3.3560	6.9753
			$\hat{\boldsymbol{\beta}}$	0.9297	1.2295	2.4425	3.2322	6.6404

Table 2. Bias and mean square error (MSE) of estimates converged

σ	t_0	$p(y > t_0)$		N^+	$BIAS^*$	MSE^*
0.3	1	0.120	$\hat{\boldsymbol{\beta}}_{OLS}$	550	-0.2099	0.0620
			$\hat{\boldsymbol{\beta}}_{CTC}$	550	-0.0058	0.0462
			$\hat{\boldsymbol{\beta}}$	550	-0.0054	0.0441
0.3	0.75	0.283	$\hat{\boldsymbol{\beta}}_{OLS}$	550	-0.3998	0.1771
			$\hat{\boldsymbol{\beta}}_{CTC}$	550	0.0278	0.1043
			$\hat{\boldsymbol{\beta}}$	550	0.0233	0.0956
0.3	0.5	0.5	$\hat{\boldsymbol{\beta}}_{OLS}$	550	-0.5360	0.3039
			$\hat{\boldsymbol{\beta}}_{CTC}$	550	0.0527	0.2204
			$\hat{\boldsymbol{\beta}}$	548	0.0339	0.1661
0.5	1	0.195	$\hat{\boldsymbol{\beta}}_{OLS}$	550	-0.3666	0.1734
			$\hat{\boldsymbol{\beta}}_{CTC}$	550	0.0109	0.1515
			$\hat{\boldsymbol{\beta}}$	549	0.0059	0.1397
0.5	0.75	0.334	$\hat{\boldsymbol{\beta}}_{OLS}$	550	-0.4838	0.2671
			$\hat{\boldsymbol{\beta}}_{CTC}$	550	0.0490	0.2520
			$\hat{\boldsymbol{\beta}}$	549	0.0358	0.2223
0.5	0.5	0.5	$\hat{\boldsymbol{\beta}}_{OLS}$	550	-0.5972	0.3868
			$\hat{\boldsymbol{\beta}}_{CTC}$	539	0.0332	0.3070
			$\hat{\boldsymbol{\beta}}$	533	0.0126	0.2754
1	1	0.316	$\hat{\boldsymbol{\beta}}_{OLS}$	550	-0.4860	0.3723
			$\hat{\boldsymbol{\beta}}_{CTC}$	539	0.0571	0.6476
			$\hat{\boldsymbol{\beta}}$	539	0.0644	0.6399
1	0.75	0.405	$\hat{\boldsymbol{\beta}}_{OLS}$	550	-0.5834	0.4614
			$\hat{\boldsymbol{\beta}}_{CTC}$	539	0.0527	0.8925
			$\hat{\boldsymbol{\beta}}$	536	0.0158	0.8046
1	0.5	0.5	$\hat{\boldsymbol{\beta}}_{OLS}$	550	-0.6521	0.5216
			$\hat{\boldsymbol{\beta}}_{CTC}$	539	-0.0170	0.9286
			$\hat{\boldsymbol{\beta}}$	537	-0.0102	0.8656

$^+$ Number of converged simulation sample.

* The sample of converged estimates is trimmed 2 % for extreme values.