International Journal of Statistics and Management System
Vol. 5 No. 1 (January-June, 2020)

Submitted: 13th March 2020 Revised: 25th April 2020 Accepted: 29th May 2020

Asymptotic distributions of BJE in linear regression models with mixed interval–censored data *

Cuixian Chen †

Abstract

We consider the estimation problem with mixed interval–censored (MIC) data under the multiple linear regression model. The Buckley–James(1979) type of estimator (BJE) has been extended from right–censored data to interval-censored data by Rabinowitz *et. al.* (1995). We establish that the BJE has an asymptotic normal distribution under certain discrete regularity conditions. The real data examples of discrete data are given and various non-normal asymptotic distributions of the BJE are also presented when the regularity conditions are violated.

1. Introduction This article provides the asymptotic properties of the Buckley–James (1979) estimator under certain discrete regularity conditions and under the multiple linear regression model with MIC data (see Yu, Wong and Li (2001)). This study is motivated by the data analysis of the marriage data set with covariates from the National Longitudinal Survey of Youth (NLSY, 1979-1998). The variables of interests in this data set are the first marriage ages of a couple. For each variable, there are exact observations, as well as right-censored (RC), left-censored (LC) and strictly interval-censored (SIC) observations. Since all records are discretized, the data are discrete.

For univariate interval-censored data, Groeneboom & Wellner (1992) proposed a case 2 interval censorship model. The main assumption in the model is that there are exactly two

---

Key words and phrases: Linear regression, mixed interval censorship, consistency, asymptotic normality.
AMS 2000 subject classifications. Primary 62J05; secondary 62N01.
†*Mailing Address:* Department of Mathematical Sciences, SUNY, Binghamton, NY 13902, USA

follow-ups for each individual. Schick and Yu (2000) pointed out that the assumption in the case 2 model is not realistic and proposed a mixed case interval censorship model, which allows the number of follow-ups to be random. These models do not allow uncensored observations. Yu, Wong and Li (2001) extended the mixed case model to a MIC model, which allows RC, LC, SIC and uncensored observations and established strong consistency and asymptotic normality of the self-consistent estimator (SCE) under mild conditions.

Linear regression models with censored data were first studied in 1970's (see Miller (1981)). Miller (1976) and Buckley and James (1979) proposed two extensions of the least squares estimator to the case of univariate right censorship regression model. It turns out that the Buckley-James estimator is an optimal extension of the least squares estimator under certain smoothness regularity conditions (see Lai and Ying (1991)). The Buckley-James type of estimator (BJE) has been extended from RC data to IC data by Rabinowitz *et. al*. (1995). Li and Pu (1999) considered generalization of the BJE for IC data that contained exact observations. However, both papers did not give proofs of the asymptotic properties of the BJE under interval censoring.

Kong and Yu (2006) showed that the BJE has an asymptotic normal distributions under certain discontinuity assumptions and under the univariate right censorship regression model. The BJE for the linear regression model with MIC data makes use of the SCE of the underlying error distribution. It is an interesting question whether the BJE still has asymptotic normality under certain discontinuity assumptions on the error distribution and under our model set-ups. This note settles the problem.

The paper is organized as follows. A multiple mixed interval-censorship linear regression model and notations are introduced in section 2. Main results are presented in section 3. It is shown that under certain regularity conditions, the BJE has an asymptotic normal distribution. We also present examples that the BJE has different asymptotic distributions when the regularity conditions are violated. The proofs of lemmas are relegated in section 4. The proofs of the statements in the examples are given in a technical report (Chen (2006)), as they are elementary and tedious.

2. Model description. Consider a multiple linear regression model:

$$X = \beta'\mathbf{Z} + \epsilon,$$

where $\mathbf{Z}$ is a vector of p-dimensional covariate, X a monotonically transformed failure time from a known transformation and ϵ has an unknown distribution function F_0. Moreover, assume that X is subject to mixed interval censoring, that is, a mixture of various case k models and a right censorship model. Let K be a non-negative random integer. If $K = 0$, then X is subject to a right censorship model with censoring variable $Y_{0,0}$. If $K > 0$, then on the event $\{K = k\}$, X is subject to a case k interval censorship model with random variables $\{Y_{k,j} : j = 1, \cdots, k\}$ such that $Y_{k,1} < \cdots < Y_{k,k}$. Denote

$$\mathbf{Y} = \begin{pmatrix} Y_{0,0} & & & & \\ Y_{1,1} & & & & \\ Y_{2,1} & Y_{2,2} & & & \\ \vdots & \vdots & \ddots & & \\ Y_{k,1} & Y_{k,2} & \cdots & Y_{k,k} & \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}.$$

For convenience, let $Y_{k,0} = -\infty$ and $Y_{k,k+1} = \infty$ for $k \geq 1$.

We assume that on the event $\{K = k\}$, the observable extended random vector is (L, R), where

$$(L, R) = \begin{cases} (X, X) & \text{if } k = 0 \text{ and } X \leq Y_{0,0}, \\ (Y_{0,0}, \infty) & \text{if } k = 0\text{and } X > Y_{0,0}, \\ (Y_{k,j-1}, Y_{k,j}) & \text{if } k \geq 1, Y_{k,j-1} < X \leq Y_{k,j}, j = 1, ..., k+1. \end{cases}$$

Note that L and R are not ordinary random variables as they may take $\pm\infty$.

We make use of the following assumptions throughout the paper.

A1 ϵ and $(\mathbf{Z}, K, \mathbf{Y})$ are independent.

A2 $(\epsilon,\ \mathbf{Z}, K,\ \mathbf{Y})$ takes on finitely many values.

Let $(X_i, \mathbf{Z}_i, \epsilon_i, L_i, R_i)$, $i = 1, \cdots, n$ be *i.i.d.* copies of $(X, \mathbf{Z}, \epsilon, L, R)$. $(L_i, R_i, \mathbf{Z}_i)$s are called MIC data. Denote the observed intervals by

$$\mathcal{I}_i = \mathcal{I}_i(\mathbf{b}) = \begin{cases} (L_i - \mathbf{b}'\mathbf{Z}_i, R_i - \mathbf{b}'\mathbf{Z}_i] & \text{if } L_i < R_i, \\ [L_i - \mathbf{b}'\mathbf{Z}_i, R_i - \mathbf{b}'\mathbf{Z}_i] & \text{if } L_i = R_i. \end{cases}$$

The BJE is a zero-crossing of the modified score function $H(\mathbf{b})$,

$$H(\mathbf{b}) = \sum_{i=1}^{n} (\mathbf{Z}_i - \overline{\mathbf{Z}})(\hat{X}_i^* - \mathbf{b}'\mathbf{Z}_i)$$

where

$$\hat{X}_i^* = X_i\delta_i + (1-\delta_i)\left(\mathbf{b}'\mathbf{Z}_i + \frac{\sum_{t\in\mathcal{I}_i(\mathbf{b})} t\hat{f}_{\mathbf{b}}(t)}{\sum_{t\in\mathcal{I}_i(\mathbf{b})} \hat{f}_{\mathbf{b}}(t)}\right), \tag{2.1}$$

δ_i is an indicator function that X_i is exact, $\hat{f}_{\mathbf{b}}$ is a generalized MLE (GMLE) of the density function of ϵ based on the data $\mathcal{I}_i(\mathbf{b})s$ and a zero-crossing of H is a point $\mathbf{v} \in \mathcal{R}^p$ such that for each $c > 0$, there exists a point $\mathbf{b}^-$ with coordinate $b_j^- \in (v_j - c, v_j]$ for each j and there exists a point $\mathbf{b}^+$ with coordinate $b_j^+ \in [v_j, v_j + c)$ for each j satisfying $H_j(\mathbf{b}^+)H_j(\mathbf{b}^-) \leq 0$ for each j, where $\mathbf{v}' = (v_1, ..., v_p)$ and $H(\mathbf{b}) = (H_1(\mathbf{b}), ..., H_p(\mathbf{b}))'$. Moreover, the GMLE maximizes the generalized likelihood function, Λ_n.

Define *innermost intervals* A_j, $j = 1, \cdots, m$ induced by $\mathcal{I}_1, \cdots, \mathcal{I}_n$ to be all the disjoint intervals which are non-empty intersections of these $\mathcal{I}_i s$ such that

$$A_j \cap \mathcal{I}_i = \emptyset \text{ or } A_j \text{ for any } i \text{ and } j.$$

Peto (1973) showed that the GMLE of F_0 assigns weights, say $s_1, \cdots, s_m$, to the corresponding innermost intervals $A_1, \cdots, A_m$ only. Using an argument similar to Hanley and Parnes (1983), it can be shown that the GMLE of $F_0(\mathbf{x})$ must assign all the probability masses $s_1, ..., s_m$ to the sets $A_1, ..., A_m$. Thus the generalized likelihood function is as following:

$$\Lambda_n = \prod_{i=1}^n \mu_F(\mathcal{I}_i) = \prod_{i=1}^n [\sum_{j=1}^m \mathbf{1}(A_j \subset \mathcal{I}_i)s_j] = \prod_{i=1}^n [\sum_{j=1}^m \eta_{ij}s_j]$$

where $\mathbf{1}(A)$ is an indicator function of an event A, $\eta_{ij} = \mathbf{1}(A_j \subset I_i)$, $\mathbf{s}$ $(= (s_1, ..., s_{m-1})')$ $\in D_s$, $D_s = \{\mathbf{s};\ s_i \geq 0, s_1 + \cdots + s_{m-1} \leq 1\}$ and $s_m = 1 - s_1 - \cdots - s_{m-1}$.

We make use of the following notations: let $T_{2i-1}(\mathbf{b})$ and $T_{2i}(\mathbf{b})$ be the endpoints of the interval $\mathcal{I}_i(\mathbf{b})$, *i.e.*, $T_{2i-1}(\mathbf{b}) = L_i - \mathbf{b}'\mathbf{Z}_i$ and $T_{2i}(\mathbf{b}) = R_i - \mathbf{b}'\mathbf{Z}_i$. Take $\xi_{2i-1} = 1$ and $\xi_{2i} = 1$ if $T_{2i-1}(\mathbf{b}) = T_{2i}(\mathbf{b})$; Otherwise, define $\xi_{2i-1} = 0$ and $\xi_{2i} = 0$, $i = 1, 2, \cdots, n$. Let $M_k(\mathbf{b})$ be the right endpoint of the k-th innermost interval, $k = 1, \cdots, m$. Obviously, for each k, $M_k(\mathbf{b}) = R_{i_k} - \mathbf{b}'\mathbf{Z}_{i_k}$ for some $i_k \in \{1, \cdots, n\}$. Denote

$$R_k^s = R_{i_k} \text{ and } \mathbf{Z}_k^s = \mathbf{Z}_{i_k}. \tag{2.2}$$

Since a GMLE $\hat{F}(t)$ of $F_0(t)$ is not uniquely defined for t in an open innermost interval (see Peto (1973) and Turnbull (1976)), we make use of the convention that we only assign probability masses to the right endpoints of innermost intervals, except perhaps at the innermost interval A_m with $\sup\{A_m\} = \infty$. Then

$$\begin{aligned}\hat{X}_i^* &= X_i\delta_i + (1-\delta_i)\left(\mathbf{b}'\mathbf{Z}_i + \frac{\sum_{j=1}^m \eta_{ij}\hat{s}_j M_j(\mathbf{b})}{\sum_{k=1}^m \eta_{ik}\hat{s}_k}\right) \\ &= X_i\delta_i + (1-\delta_i)\left(\mathbf{b}'\mathbf{Z}_i + \frac{\sum_{j=1}^m \eta_{ij}\hat{s}_j (R_j^s - \mathbf{b}'\mathbf{Z}_j^s)}{\sum_{k=1}^m \eta_{ik}\hat{s}_k}\right) \qquad \text{(see(2.2))} \\ &= X_i\delta_i + (1-\delta_i)\left(\frac{\sum_{j=1}^m \eta_{ij}\hat{s}_j R_j^s}{\sum_{k=1}^m \eta_{ik}\hat{s}_k}\right) - (1-\delta_i)\left(\frac{\sum_{j=1}^m \eta_{ij}\hat{s}_j (\mathbf{Z}_j^s)'}{\sum_{k=1}^m \eta_{ik}\hat{s}_k} - \mathbf{Z}_i{}'\right)\mathbf{b}.\end{aligned}$$

Therefore,

$$H(\mathbf{b}) = \sum_{i=1}^n (\mathbf{Z}_i - \overline{\mathbf{Z}})(\hat{X}_i^* - \mathbf{Z}_i{}'\mathbf{b}) = \mathcal{A}(\mathbf{b}) - \mathcal{B}(\mathbf{b})\mathbf{b}, \tag{2.3}$$

where $\mathcal{A}(\mathbf{b}) = \sum_{i=1}^n \mathcal{A}_i(\mathbf{b})$ and $\mathcal{B}(\mathbf{b}) = \sum_{i=1}^n \mathcal{B}_i(\mathbf{b})$,

$$\mathcal{A}_i(\mathbf{b}) = (\mathbf{Z}_i - \overline{\mathbf{Z}})\left(X_i\delta_i + (1-\delta_i)\frac{\sum_{j=1}^m \eta_{ij}\hat{s}_j R_j^s}{\sum_{k=1}^m \eta_{ik}\hat{s}_k}\right), \tag{2.4}$$

$$\mathcal{B}_i(\mathbf{b}) = (\mathbf{Z}_i - \overline{\mathbf{Z}})\left(\mathbf{Z}_i{}'\delta_i + (1-\delta_i)\frac{\sum_{j=1}^m \eta_{ij}\hat{s}_j (\mathbf{Z}_j^s)'}{\sum_{k=1}^m \eta_{ik}\hat{s}_k}\right). \tag{2.5}$$

Remark 1 Arrange those m distinct innermost intervals which are induced by the n $\mathcal{I}_i(\mathbf{b})s$ in increasing order. If the largest observation $(L_{(n)}(\mathbf{b}), R_{(n)}(\mathbf{b}))$ is right censored, then the largest innermost interval will also be right censored. In the latter case, we use $(L_{(n)}(\mathbf{b}), L_{(n)}(\mathbf{b})+1]$ to replace the role of the innermost interval $(L_{(n)}(\mathbf{b}), \infty)$. Denote

$$F_0^*(t) = \begin{cases} F_0(t) & \text{if } t \le \tau, \\ F_0(\tau) & \text{if } t \in (\tau, \tau+1), \\ 1 & \text{if } t \ge \tau+1, \end{cases}$$

$$\tau = \inf\{t : P\{V(\beta) \le t\} = 1\}, \text{ where } V(\beta) = L - \beta'\mathbf{Z}\ .$$

Without loss of generality (WLOG), we assume $F_0^* = F_0$, otherwise, replace F_0 by F_0^*.

It is well known that the GMLE of F_0 only depends on the ranks of $2n$ observations $T_j(\mathbf{b})s$. Note that the ranks of these $T_j(\mathbf{b})s$ will only change at the solutions of each pair of linearly independent equations of a form of

$$T_i(\mathbf{b}) = T_j(\mathbf{b}), \text{ where } \mathbf{Z}_i \ne \mathbf{Z}_j. \tag{2.6}$$

The latter equations are of forms:

$$L_i - \mathbf{b}'\mathbf{Z}_i = L_j - \mathbf{b}'\mathbf{Z}_j,\ L_i - \mathbf{b}'\mathbf{Z}_i = R_j - \mathbf{b}'\mathbf{Z}_j \text{ or } R_i - \mathbf{b}'\mathbf{Z}_i = R_j - \mathbf{b}'\mathbf{Z}_j\ ,$$

where L_i, R_i, L_j and R_j are finite, $i \neq j$ and $\mathbf{Z}_i \neq \mathbf{Z}_j$.

Since there are at most $(n(2n-1)-n)$ equations of the form of (2.6), and the solutions to these equations are hyperplanes in $\mathbb{R}^p$, there are at most $(n(2n-1)-n)$ distinct hyperplanes which partition $\mathbb{R}^p$ into finitely many disjoint regions, say $\mathcal{O}_1, \cdots, \mathcal{O}_{m_0}$.

For example, if $p = 2$, for each pair of linearly independent equations of the form of (2.6), there is a unique solution, which represents the intersection of two lines in $\mathbb{R}^2$. Let $\mathcal{B}_1$ be the collection of all such distinct solutions. Each line is further partitioned into finitely many disjoint open line segments by the points belonging to both $\mathcal{B}_1$ and that line. Denote the collection of all such open line segments in all distinct lines by $\mathcal{B}_2$. Take $\mathscr{C}_1 = \{\mathcal{O} : \mathcal{O} = \{\mathbf{b}\}, \mathbf{b} \in \mathcal{B}_1\}$. Notice that $\mathscr{C}_1$ and $\mathcal{B}_2$ are disjoint. These points in $\mathcal{B}_1$ and the open line segments in $\mathcal{B}_2$ will further partition $\mathbb{R}^2 \setminus \{\cup_{\mathbf{b}\in\mathcal{B}_1}\{\mathbf{b}\} \cup \cup_{\mathcal{O}\in\mathcal{B}_2}\mathcal{O}\}$ into finitely many disjoint open regions, which is denoted by $\mathcal{B}_3$. Then an $\mathcal{O}_i$ has one of the forms as follows:

1. $\{\mathbf{b}\}$ where $\mathbf{b} \in \mathcal{B}_1$;
2. an open line segment that belongs to $\mathcal{B}_2$;
3. an open region in $\mathbb{R}^2$ that belongs to $\mathcal{B}_3$.

Take $\mathscr{C}_2 = \mathcal{B}_2 \cup \mathcal{B}_3$. We shall prove in section 4 the following statement.

Lemma 1 *Given $\mathcal{O}_j \in \mathscr{C}_2$,*

1. *for each i, the rank of $T_i(\mathbf{b})$ remains the same if $\mathbf{b} \in \mathcal{O}_j$;*
2. *for each k, the probability mass $\hat{s}_k$ assigned to the k-th innermost interval is constant in $\mathbf{b}$ on $\mathcal{O}_j$;*
3. *$H(\cdot)$ is linear (in $\mathbf{b}$) on $\mathcal{O}_j$.*

3. Main results. We establish in this section the asymptotic normality of the BJE under certain discrete regularity conditions. We also show that if these conditions are violated, then the BJE has various non-normal asymptotic distributions.

In order to establish the asymptotic properties of the BJE, we further assume that:

A3 $P\{\delta_1 = \cdots = \delta_{p+1} = 1,\ rank(\mathbf{Z}_1 - \mathbf{Z}_{p+1}, ..., \mathbf{Z}_p - \mathbf{Z}_{p+1}) = p\} > 0$.

A4 $P\{\epsilon_1 = Y_2^* - \beta'\mathbf{Z}_2\} = 0$ and $P\{Y_3^* - \beta'\mathbf{Z}_3 = Y_4^* - \beta'\mathbf{Z}_4$ and $Y_3^* \neq Y_4^*\} = 0$, where $(\epsilon_i, \mathbf{Z}_i, K_i, \mathbf{Y}_i)$s are i.i.d. from $(\epsilon, \mathbf{Z}, K, \mathbf{Y})$ and $Y_i^* \in \{Y_{i,0,0}, Y_{i,K_i,1}, Y_{i,K_i,2}, \cdots, Y_{i,K_i,K_i}\}$, $i = 2, 3, 4$.

A3 is an identifiability condition, which is introduced in Kong and Yu (2006). A4 is a modification of A4 in Kong and Yu (2006).

Lemma 2 *Under assumptions A1 through A4, given an* $\omega \in \Omega$, *there exists a neighborhood of* β, *say* $O(\beta, c) = \{\mathbf{b} : \| \mathbf{b} - \beta \| < c\}$ *such that* $(\mathcal{A}_i(\mathbf{b}), \mathcal{B}_i(\mathbf{b}))$ *is constant in* $\mathbf{b}$ *for each* i, *where* Ω *is the sample space.*

Lemma 3 *If assumptions A1 through A4 hold, with probability 1 (w.p.1),* $\hat{\mathbf{b}} = (\mathcal{B}(\mathbf{b}))^{-1}\mathcal{A}(\mathbf{b})$ *is a BJE if* n *is large enough.*

For the later development, we need to assume

A5 $\mu_{F_0}(A_j(\beta)) > 0$ for each innermost interval $A_j(\beta)$.

A5 is a modification of assumption 3 in Yu *et. al.* (1998) under the multiple regression model.

Theorem 1 *Suppose that A1 through A4 hold, then the BJE* $\hat{\beta}_n = (\mathcal{B}(\mathbf{b}))^{-1}\mathcal{A}(\mathbf{b})$ *is consistent. Moreover, if A5 holds, then the BJE is asymptotically normally distributed.*

The proof of the theorem is given in section 4. We present an example as follows which satisfies A1-A5 and thus the BJE does have an asymptotic normal distribution. For simplicity, we consider $p = 1$.

Example 1 Let $\beta = 1$. Suppose that (1) ϵ is a random variable which takes values 0 and 2 *w.p.* 0.5 respectively, $Z \sim Bin(1, 1/2)$, K is a random integer which takes values 0 and 2 *w.p.* 0.5 respectively, $Y_{0,0} \equiv 1.5$, $Y_{2,1} \equiv 0.3$, $Y_{2,2} \equiv 1.4$, $\mathbf{Y} = \{Y_{0,0}, Y_{2,1}, Y_{2,2}\}$; (2) ϵ, Z, $\mathbf{Y}$ are independent. It is easy to check that A1 through A5 hold in this example. It is shown in the technical report (Chen (2006)) that

$$\hat{\beta}_n \text{ is asymptotically normal.} \tag{3.1}$$

Assumption A4 is critical in the proof of Theorem 1. Notice that A4 is always true if $(\epsilon, \mathbf{Y}, \mathbf{Z})$ is a continuous random vector. The condition that $\mu_{F_0}(A_j(\beta)) > 0$ for $j = 1$,

$\cdots$, m is a sufficient condition for the asymptotic normality of the SCE of the probability masses assigned to each innermost interval $A_j(\beta)$ (see Yu *et. al.* (1998)). We shall establish an example in the following that if assumptions A5 is violated, the BJE may still has an asymptotic normal distribution under assumptions A1 through A4.

Example 2 Let $\beta = 1$. Suppose that (1) ϵ and $Z \sim Bin(1, 1/2)$, K is a random integer which takes values 0 and 2 *w.p.* 0.5 respectively, $Y_{0,0} \equiv 1.5$, $Y_{2,1} \equiv 0.3$, $Y_{2,2} \equiv 1.8$, $\mathbf{Y} = \{Y_{0,0}, Y_{2,1}, Y_{2,2}\}$; (2) ϵ, Z, $\mathbf{Y}$ are independent. It is shown in the technical report (Chen (2006)) that

$$\hat{\beta}_n \text{ is asymptotically normal.} \tag{3.2}$$

We shall establish examples in the following such that if A3 or A4 do not hold, then the BJE may have various non-normal asymptotic distributions if F_0 is not continuous. The proof of these examples are given in the technical report (Chen (2006)).

Example 3 Let $\beta = 1$. Suppose that (1) ϵ and $Z \sim Bin(1, 1/2)$, K is a random integer which takes values 0 and 2 *w.p.* 0.5 respectively, $Y_{0,0} \equiv 0.5$, $Y_{2,1} \equiv 0.5$, $Y_{2,2} \equiv 1$, $\mathbf{Y} = \{Y_{0,0}, Y_{2,1}, Y_{2,2}\}$; (2) ϵ, Z, $\mathbf{Y}$ are independent. In this case, it can be shown that if n is large,

$$BJE = \begin{cases} 0.5 & w.p.\ 0.5, \\ 1 & w.p.\ 0.5. \end{cases} \tag{3.3}$$

Thus, the BJE is not consistent and not asymptotically normally distributed.

Example 3 justifies the identifiability condition A3 for multiple linear regression model with MIC data.

Example 4 Let $\beta = 1$. Suppose that (1) ϵ and $Z \sim Bin(1, 1/2)$, K is a random integer which takes values 0 and 2 *w.p.* 0.5 respectively, $Y_{0,0} \equiv 1.5$, $Y_{2,1} \equiv 0.5$, $Y_{2,2} \equiv 1$, $\mathbf{Y} = \{Y_{0,0}, Y_{2,1}, Y_{2,2}\}$; (2) ϵ, Z, $\mathbf{Y}$ are independent. It can be shown that the BJE $\hat{\beta}_n$ is consistent and

$$\sqrt{n}(\hat{\beta}_n - \beta) \xrightarrow{D} \min\{W, 0\}, \tag{3.4}$$

where $W \sim N(0, \sigma^2)$ and $\sigma > 0$.

In other words, the BJE does not have an asymptotic normal distribution.

The main character in Example 4 is that

$$Y_{2,2} - \beta Z = \epsilon \text{ if } (Y_{2,2}, Z, \epsilon) = (1, 0, 1) \text{ or } (Y_{2,2}, Z, \epsilon) = (1, 1, 0),$$

$$Y_{0,0} - \beta Z_1 = Y_{2,1} - \beta Z_2 \text{ if } (Y_{0,0}, Z_1, Y_{2,1}, Z_2) = (1.5, 1, 0.5, 0).$$

It can be rewritten as:

$$P\{\epsilon_1 = Y_2^* - \beta'\mathbf{Z}_2\} > 0 \text{ or } P\{Y_3^* - \beta'\mathbf{Z}_3 = Y_4^* - \beta'\mathbf{Z}_4 \text{ and } Y_3^* \neq Y_4^*\} > 0$$

where $Y_i^* \in \{Y_{i,0,0}, Y_{i,2,1}, Y_{i,2,2}\}$, $i = 2, 3, 4$, and $(\epsilon_i, \mathbf{Z}_i, Y_{i,0,0}, Y_{i,2,1}, Y_{i,2,2})$s are i.i.d.

Example 4 justifies that condition A4 is needed for asymptotic normality of the BJE.

4. Proofs of lemmas and the theorem.

Proof of Lemma 1 (1) Given $\mathbf{b}$, it is well known that $\hat{F}_{\mathbf{b}}$ only depends on the ranks of $2n$ $T_j(\mathbf{b})s$(see Turnbull (1974)). That is, $\hat{F}_{\mathbf{b}_1}(T_j(\mathbf{b}_1)) = \hat{F}_{\mathbf{b}_2}(T_j(\mathbf{b}_2))$ if the rank of $T_j(\mathbf{b}_1)$ is the same as the rank of $T_j(\mathbf{b}_2)$ for each j.

Note that the ranks of these $T_j(\mathbf{b})s$ will only change at the solution of each pair of linearly independent equations of the form of (2.6). The solutions to these equations together with the hyperplanes corresponding to these equations partition $\mathbb{R}^p$ into $\mathcal{O}_1, \cdots,$ $\mathcal{O}_{m_0}$, so that for each $\mathbf{b} \in \mathcal{O}_k$, $\mathcal{O}_k \in \mathscr{C}_2$, the ranks of the $2n$ $T_i(\mathbf{b})s$ do not change.

(2) Consequently, $\hat{F}_{\mathbf{b}}(T_j(\mathbf{b}))$ is constant on each region $\mathcal{O}_k$, $\mathcal{O}_k \in \mathscr{C}_2$, that is, $\hat{F}_{\mathbf{b}_1}(T_j(\mathbf{b}_1)) = \hat{F}_{\mathbf{b}_2}(T_j(\mathbf{b}_2))$ for $\mathbf{b}_1, \mathbf{b}_2 \in \mathcal{O}_k$, $\mathcal{O}_k \in \mathscr{C}_2$, for $j = 1, \cdots, 2n$ and $k = 1, \cdots, m_0$. That is, the probability mass $\hat{s}_i$ assigned to i^{th} innermost interval is constant in $\mathbf{b}$ on each $\mathcal{O}_k$, $\mathcal{O}_k \in \mathscr{C}_2$.

(3) From (2.4) and (2.5), it is obvious that $\mathcal{A}_i(\mathbf{b})$ and $\mathcal{B}_i(\mathbf{b})$ only depend on the probability masses assigned to innermost intervals and the right endpoints of innermost intervals, $M_k(\mathbf{b}) = R_k^s - \mathbf{b}'\mathbf{Z}_k^s$ (see(2.2)). They are constant functions of $\mathbf{b}$ on $\mathcal{O}_j$, $\mathcal{O}_j \in \mathscr{C}_2$, by (2). Therefore, by the definition of $H(\mathbf{b})$, it is a linear function of $\mathbf{b}$ on $\mathcal{O}_j$, $\mathcal{O}_j \in \mathscr{C}_2$, as $\mathcal{A}_i(\mathbf{b})$ and $\mathcal{B}_i(\mathbf{b})$ are constant functions of $\mathbf{b}$. ▫

Proof of Lemma 2 Hereafter, we fix an $\omega \in \Omega$. Under assumption A4, it is easy to verify that $\mathbf{b} = \beta$ is not a solution to any pair of linearly independent equations of a form of

$$T_i(\mathbf{b}) = T_j(\mathbf{b}),\ \xi_i \cdot \xi_j = 0 \text{ and } \mathbf{Z}_i \neq \mathbf{Z}_j. \tag{4.1}$$

Notice that by assumption A2, there are only finitely many equations of the form (4.1) whose solutions are hyperplanes in $\mathbb{R}^p$. Thus there are finitely many distinct hyperplanes, which will not change as long as n is large enough. All such distinct hyperplanes partition $\mathbb{R}^p$ into finitely many disjoint regions, say $\mathcal{O}_1^*, \cdots, \mathcal{O}_{m_1}^*$. For each i, it satisfies:

$$\mathcal{O}_i^* = \bigcup_{j \in E_i} \mathcal{O}_j, \text{where } E_i = \{j : \mathcal{O}_j \subset \mathcal{O}_i^*\}.$$

Since β does not belong to the hyperplanes obtained by solving (4.1), the distance from β to each of such hyperplanes is positive. So there exists a neighborhood of β, say $O(\beta, c)$, with $c > 0$ such that:

$$O(\beta, c) \subset \mathcal{O}^*_{k_0} = \bigcup_{j \in E_{k_0}} \mathcal{O}_j, \text{ for some } 1 \le k_0 \le m_1;$$

For any $\mathbf{b}_1 \in O(\beta, c) \cap \mathcal{O}$ with arbitrary $\mathcal{O} \in \mathscr{C}_2$, and for any i and j, with $\mathbf{Z}_i \ne \mathbf{Z}_j$, it satisfies $T_i(\mathbf{b}_1) \ne T_j(\mathbf{b}_1)$, which means that $T_i(\mathbf{b})$s do not change their ranks at $\mathbf{b}_1$. Furthermore, for any $\mathbf{b}_2 \in O(\beta, c) \cap \mathcal{B}_1$, there may exist i_0 and j_0 such that $T_{i_0}(\mathbf{b}_2) = T_{j_0}(\mathbf{b}_2)$, $\xi_{i_0} \cdot \xi_{j_0} = 1$, and $\mathbf{Z}_{i_0} \ne \mathbf{Z}_{j_0}$, which means that only those exact observations may tie together at $\mathbf{b}_2$. Therefore, for any $\mathbf{b} \in O(\beta, c)$, there is no non-singleton observation between any two exact observations. Moreover, only those exact observations $T_i(\mathbf{b})$ may change the ranks.

For each $\mathbf{b} \in O(\beta, c)$, let $\psi_j = \mathbf{1}(A_j \text{ is a singleton})$, $G_i = G_i(\mathbf{b}) = \{j : A_j(\mathbf{b}) \subset \mathcal{I}_i(\mathbf{b})\}$, then we have:

$$\begin{aligned} \sum_{j \in G_i} \eta_{ij} \hat{s}_j M_j(\mathbf{b}) &= \sum_{j \in G_i} \eta_{ij} [\psi_j + (1 - \psi_j)] \hat{s}_j M_j(\mathbf{b}) \\ &= \sum_{j \in G_i} \eta_{ij} \psi_j \hat{s}_j \frac{\sum_{k=1}^n \mathbf{1}(T_{2k}(\mathbf{b}) = M_j(\mathbf{b}), \xi_k = 1) T_{2k}(\mathbf{b})}{\sum_{h=1}^n \mathbf{1}(T_{2h}(\mathbf{b}) = M_j(\mathbf{b}), \xi_h = 1)} \\ &\quad + \sum_{j \in G_i} \eta_{ij} (1 - \psi_j) \hat{s}_j M_j(\mathbf{b}), \end{aligned}$$

Notice that $T_{2i}(\mathbf{b}) = R_i - \mathbf{b}'\mathbf{Z}_i$ and $M_k(\mathbf{b}) = R^s_k - \mathbf{b}'\mathbf{Z}^s_k$ (see(2.2)), then

$$H(\mathbf{b}) = \sum_{i=1}^n (\mathbf{Z}_i - \overline{\mathbf{Z}})(\hat{X}^*_i - \mathbf{Z}_i{}'\mathbf{b}) = \mathcal{A}(\mathbf{b}) - \mathcal{B}(\mathbf{b})\mathbf{b},$$

where $\mathcal{A}(\mathbf{b}) = \sum_{i=1}^n \mathcal{A}_i(\mathbf{b})$ and $\mathcal{B}(\mathbf{b}) = \sum_{i=1}^n \mathcal{B}_i(\mathbf{b})$,

$$\begin{aligned} \mathcal{A}_i(\mathbf{b}) &= (\mathbf{Z}_i - \overline{\mathbf{Z}}) \Bigg(X_i \delta_i + (1 - \delta_i) \frac{1}{\sum_{j \in G_i} \eta_{ij} \hat{s}_j} \sum_{j \in G_i} \Big(\eta_{ij} (1 - \psi_j) \hat{s}_j R^s_j \Big) \\ &\quad + (1 - \delta_i) \frac{1}{\sum_{j \in G_i} \eta_{ij} \hat{s}_j} \sum_{j \in G_i} \Big(\eta_{ij} \psi_j \hat{s}_j \frac{\sum_{k=1}^n \mathbf{1}(T_{2k}(\mathbf{b}) = M_j(\mathbf{b}), \xi_k = 1) R_k}{\sum_{h=1}^n \mathbf{1}(T_{2h}(\mathbf{b}) = M_j(\mathbf{b}), \xi_h = 1)} \Big) \Bigg), \end{aligned}$$

$$\begin{aligned}\mathcal{B}_i(\mathbf{b}) &= (\mathbf{Z}_i - \overline{\mathbf{Z}})\Bigg(\mathbf{Z}_i'\delta_i + (1-\delta_i)\frac{1}{\sum_{j\in G_i}\eta_{ij}\hat{s}_j}\sum_{j\in G_i}\Big(\eta_{ij}(1-\psi_j)\hat{s}_j\mathbf{Z}_j^{s'}\Big)\\ &\quad +(1-\delta_i)\frac{1}{\sum_{j\in G_i}\eta_{ij}\hat{s}_j}\sum_{j\in G_i}\Big(\eta_{ij}\psi_j\hat{s}_j\frac{\sum_{k=1}^n \mathbf{1}(T_{2k}(\mathbf{b})=M_j(\mathbf{b}),\xi_k=1)\mathbf{Z}_k'}{\sum_{h=1}^n \mathbf{1}(T_{2h}(\mathbf{b})=M_j(\mathbf{b}),\xi_h=1)}\Big)\Bigg).\end{aligned}$$

Now we need to prove for any $\mathbf{b}_1, \mathbf{b}_2 \in O(\beta, c)$, $\mathcal{A}_i(\mathbf{b}_1) = \mathcal{A}_i(\mathbf{b}_2)$, $\mathcal{B}_i(\mathbf{b}_1) = \mathcal{B}_i(\mathbf{b}_2)$. Consider two cases: (1) $\delta_i = 1$ and (2) $\delta_i = 0$.
Case (1), $\delta_i = 1$. Since $(\mathcal{A}_i(\mathbf{b}), \mathcal{B}_i(\mathbf{b})) = ((\mathbf{Z}_i - \overline{\mathbf{Z}})\mathbf{X}_i, (\mathbf{Z}_i - \overline{\mathbf{Z}})\mathbf{Z}_i')$, it is trivially true.
Case (2), $\delta_i = 0$. We have:

$$\begin{aligned}\mathcal{A}_i(\mathbf{b}) &= (\mathbf{Z}_i - \overline{\mathbf{Z}})\frac{1}{\sum_{j\in G_i}\eta_{ij}\hat{s}_j}\sum_{j\in G_i}\eta_{ij}\\ &\quad\times\left((1-\psi_j)\hat{s}_jR_j^s + \psi_j\hat{s}_j\frac{\sum_{k=1}^n \mathbf{1}(T_{2k}(\mathbf{b})=M_j(\mathbf{b}),\xi_k=1)R_k}{\sum_{h=1}^n \mathbf{1}(T_{2h}(\mathbf{b})=M_j(\mathbf{b}),\xi_h=1)}\right).\end{aligned}$$

Take

$$s_j^* = \frac{s_j}{\sum_{h=1}^n \mathbf{1}(T_{2h}(\mathbf{b})=M_j(\mathbf{b}),\xi_h=1)}. \tag{4.2}$$

Remark 2 The motivation of defining s_i^* is as follows. Note that each distinct $\{T_i(\mathbf{b})\}$ with $\xi_i = 1$ is a (singleton) innermost interval. If there is a tie of several exact observations $T_j(\mathbf{b})$s at $T_i(\mathbf{b})$, we can pretend that each exact observation is a distinct singleton innermost interval. It is similar to the case of the empirical distribution function, that is, in the case that there is a tie, the weight at that tie is the accumulation of weights assigned to each observation at that tie which is $1/n$. By defining s_i^*, WLOG, we can assume that all exact observations within $\mathcal{I}_i(\mathbf{b})$ are distinct, and thus their ranks are irrelevant as far as s_j^* is concerned.

In order to prove that for any $\mathbf{b}_1, \mathbf{b}_2 \in O(\beta, c)$, $\mathcal{A}_i(\mathbf{b}_1) = \mathcal{A}_i(\mathbf{b}_2)$, it suffices to show that:
1. $G_i(\mathbf{b}_1) = G_i(\mathbf{b}_2)$, where $G_i(\mathbf{b}) = \{j : A_j(\mathbf{b}) \subset \mathcal{I}_i(\mathbf{b})\}$,
2. $s_j^*(\mathbf{b}_1) = s_j^*(\mathbf{b}_2)$,
3. $\sum_{j\in G_i(\mathbf{b}_1)}\eta_{ij}(\mathbf{b}_1)\hat{s}_j(\mathbf{b}_1) = \sum_{j\in G_i(\mathbf{b}_2)}\eta_{ij}(\mathbf{b}_2)\hat{s}_j(\mathbf{b}_2)$.

Notice that the GMLE of s_j is an SCE. According to self-consistent equation $s_j = \sum_{i=1}^n \frac{1}{n} \frac{\eta_{ij} s_j}{\sum_{k\in G_i} \eta_{ik} s_k}$, for any $\mathbf{b} \in O(\beta, c)$, we have:

$$\begin{aligned} s_j &= \psi_j \sum_{h=1}^n \mathbf{1}(T_{2h}(\mathbf{b}) = M_j(\mathbf{b}), \xi_h = 1) s_j^* + (1-\psi_j) s_j \\ &= \sum_{i=1}^n \frac{1}{n} \frac{\eta_{ij}[\psi_j \sum_{h=1}^n \mathbf{1}(T_{2h}(\mathbf{b}) = M_j(\mathbf{b}), \xi_h = 1) s_j^* + (1-\psi_j) s_j]}{\sum_{k\in G_i} \eta_{ik}[\psi_k \sum_{h=1}^n \mathbf{1}(T_{2h}(\mathbf{b}) = M_k(\mathbf{b}), \xi_h = 1) s_k^* + (1-\psi_k) s_k]}, \end{aligned} \tag{4.3}$$

In view of the foregoing expression of self-consistent equation and Remark 2, WLOG, we can assume that all exact observations within $\mathcal{I}_i(\mathbf{b})$ are distinct.

Since the non-singleton innermost intervals within $\mathcal{I}_i(\mathbf{b})$ will not change at all and the singleton innermost intervals are the "same" according to the discussion in Remark 2 for each $\mathbf{b} \in O(\beta, c)$, $G_i(\mathbf{b}_1) = G_i(\mathbf{b}_2)$ and therefore (1) holds.

Note that by assumption in Remark 2, all exact observations within $\mathcal{I}_i(\mathbf{b})$ are distinct and the ranks are irrelevant. Then the GMLE of the probability mass on each singleton innermost interval within $\mathcal{I}_i(\mathbf{b})$ is the same as s_j^* (see (4.2) and (4.3)), therefore (2) holds.

According to (1) and (2), all possible innermost intervals within $\mathcal{I}_i(\mathbf{b})$ are the same and the weights on each innermost interval are also the same for each $\mathbf{b} \in O(\beta, c)$ and therefore (3) holds.

It implies that for any $\mathbf{b}_1, \mathbf{b}_2 \in O(\beta, c)$, $\mathcal{A}_i(\mathbf{b}_1) = \mathcal{A}_i(\mathbf{b}_2)$. The proof of $\mathcal{B}_i(\mathbf{b}_1) = \mathcal{B}_i(\mathbf{b}_2)$ is similar and is skipped.▫

Proof of Lemma 3 Note that from the proof of lemma 2, under A3 and A4, for n large enough, the distinct hyperplanes obtained by solving $T_i(\mathbf{b}) = T_j(\mathbf{b})$, $\xi_i \cdot \xi_j = 0$ do not change as n increases. Furthermore, β does not belong to these hyperplanes. Thus $O(\beta, c)$ remains the same as n increases. Moreover, $(\mathcal{A}_i(\beta), \mathcal{B}_i(\beta))$ will take finitely many values. WLOG, we can assume that the first m of $(\mathcal{A}_i(\beta), \mathcal{B}_i(\beta))$ are distinct.

$$\frac{H(\mathbf{b})}{n} = \sum_{i=1}^m \overline{p}_i \mathcal{A}_i(\mathbf{b}) - \sum_{i=1}^m \overline{p}_i \mathcal{B}_i(\mathbf{b})\mathbf{b} = \sum_{i=1}^m \overline{p}_i \mathcal{A}_i(\beta) - \sum_{i=1}^m \overline{p}_i \mathcal{B}_i(\beta)\mathbf{b}, \quad \text{for } \mathbf{b} \in O(\beta, c),$$

where

$$\overline{p}_i = \frac{1}{n} \sum_{j=1}^n 1[(\mathcal{A}_j(\mathbf{b}), \mathcal{B}_j(\mathbf{b})) = (\mathcal{A}_i(\mathbf{b}), \mathcal{B}_i(\mathbf{b}))], \quad \text{for } \mathbf{b} \in O(\beta, c).$$

By strong law of large number, $w.p.1$, $\overline{p}_i \to E(\overline{p}_i)$ $(\triangleq p_i)$. For $\mathbf{b} \in O(\beta, c)$, $i = 1, 2, \cdots, m$, $w.p.1$,

$$\mathcal{A}_i(\mathbf{b}) = \mathcal{A}_i(\beta) \to (\mathbf{Z}_i - E(\mathbf{Z}))\{X_i\delta_i + (1-\delta_i)E(X_i|\epsilon_i \in \mathcal{I}_i(\beta))\}(\triangleq \mathcal{A}_{i0}(\beta)),$$

$$\mathcal{B}_i(\mathbf{b}) = \mathcal{B}_i(\beta) \to (\mathbf{Z}_i - E(\mathbf{Z}))\{\mathbf{Z}_i'\delta_i + (1-\delta_i)E(\mathbf{Z}_i'|\epsilon_i \in \mathcal{I}_i(\beta))\}(\triangleq \mathcal{B}_{i0}(\beta)).$$

Then, $w.p.1$

$$\begin{aligned} lim_{n\to\infty}\frac{H(\beta)}{n} &= \sum_{i=1}^{m} p_i(\mathcal{A}_{i0}(\beta) - \mathcal{B}_{i0}(\beta)\beta) \\ &= E[E((\mathbf{Z}-E(\mathbf{Z}))\epsilon|\epsilon \in \mathcal{I})] \\ &= E[(\mathbf{Z}-E(\mathbf{Z}))\epsilon] \\ &= E(\mathbf{Z}-E(\mathbf{Z}))E(\epsilon) \quad \text{(as } \epsilon \text{ and } X \text{ are independent)} \\ &= \mathbf{0}_{p\times 1}. \end{aligned}$$

Claim 1: $\mathcal{B}(\beta) = \sum_{i=1}^{m} \overline{p}_i \mathcal{B}_i(\beta)$ is not singular, provided that n is large enough. Then based on claim 1, $w.p.1$, for $\mathbf{b} \in O(\beta, c)$ and under A4, we have,

$$\begin{aligned} \frac{H(\mathbf{b})}{n} &= \sum_{i=1}^{m} \overline{p}_i \mathcal{A}_i(\beta) - \sum_{i=1}^{m} \overline{p}_i \mathcal{B}_i(\beta)\mathbf{b}, \quad \text{for } \mathbf{b} \in O(\beta, c) \\ &\to \sum_{i=1}^{m} p_i(\mathcal{A}_{i0}(\beta) - \mathcal{B}_{i0}(\beta)\mathbf{b}) = \begin{cases} \mathbf{0} & \text{if } \mathbf{b} = \beta, \\ \neq \mathbf{0} & \text{if } \mathbf{b} \in O(\beta, c)\backslash\{\beta\}. \end{cases} \end{aligned}$$

It follows from the foregoing equation that $w.p.1$,

$$\hat{\mathbf{b}} = (\sum_{i=1}^{m} \overline{p}_i \mathcal{B}_i(\beta))^{-1}(\sum_{i=1}^{m} \overline{p}_i \mathcal{A}_i(\beta)) = (\mathcal{B}(\beta))^{-1}\mathcal{A}(\beta) \to \beta, \text{as } n \to \infty. \tag{4.4}$$

$\hat{\mathbf{b}}$ may not be a BJE. However, by taking a large sample size, WLOG, we can assume, $\hat{\mathbf{b}} \in O(\beta, c)$. Thus, $(\mathcal{A}(\hat{\mathbf{b}}), \mathcal{B}(\hat{\mathbf{b}})) = (\mathcal{A}(\beta), \mathcal{B}(\beta))$ and

$$H(\hat{\mathbf{b}}) = \mathcal{A}(\hat{\mathbf{b}}) - \mathcal{B}(\hat{\mathbf{b}})\hat{\mathbf{b}} = \mathcal{A}(\beta) - \mathcal{B}(\beta)\hat{\mathbf{b}} = \mathbf{0}.$$

Therefore, $\hat{\beta}_n = \hat{\mathbf{b}}$ is a BJE.

Now we shall prove claim 1: Take G to be a transformation such that:

$$G(\mathbf{Z}_i') = \mathbf{Z}_i'\delta_i + (1-\delta_i)E(\mathbf{Z}_i'|\epsilon_i \in \mathcal{I}_i(\beta)).$$

Let $\mu_{\mathbf{Z}} = E(\mathbf{Z})$, then

$$\begin{aligned}\sum_{i\in G_i} p_i(\mathbf{Z}_i - \mu_{\mathbf{Z}})G(\mu'_{\mathbf{Z}}) &= \sum_{i\in G_i} p_i(\mathbf{Z}_i - \mu_{\mathbf{Z}})\{\mu'_{\mathbf{Z}}\delta_i + (1-\delta_i)E(\mu'_{\mathbf{Z}}|\epsilon_i \in \mathcal{I}_i(\beta))\} \\ &= \sum_{i\in G_i} p_i(\mathbf{Z}_i - \mu_{\mathbf{Z}})\{\mu'_{\mathbf{Z}}\delta_i + (1-\delta_i)\mu'_{\mathbf{Z}}\} \\ &= E(\mathbf{Z} - \mu_{\mathbf{Z}})\mu'_{\mathbf{Z}} \\ &= \mathbf{0_{p\times p}}.\end{aligned}$$

Therefore,

$$\begin{aligned}\sum_{i\in G_i} p_i\mathcal{B}_{i0}(\beta) &= \sum_{i\in G_i} p_i\mathcal{B}_{i0}(\beta) - \sum_{i\in G_i} p_i(\mathbf{Z}_i - \mu_{\mathbf{Z}})G(\mu'_{\mathbf{Z}}) \\ &= \sum_{i\in G_i} p_i(\mathbf{Z}_i - \mu_{\mathbf{Z}})\{(\mathbf{Z}_i{}' - \mu'_{\mathbf{Z}})\delta_i + (1-\delta_i)E[(\mathbf{Z}_i{}' - \mu'_{\mathbf{Z}})|\epsilon_i \in \mathcal{I}_i(\beta))]\} \\ &= \sum_{i\in G_i} p_i(\mathbf{Z}_i - \mu_{\mathbf{Z}})\{(\mathbf{Z}_i{}' - \mu'_{\mathbf{Z}})\delta_i + (1-\delta_i)E[(\mathbf{Z}_i{}' - \mu'_{\mathbf{Z}})]\} \\ &\qquad \text{(as } Z_i \text{ and } \epsilon_i \text{ are independent)} \\ &= \sum_{i\in G_i} p_i\delta_i(\mathbf{Z}_i - \mu_{\mathbf{Z}})(\mathbf{Z}_i - \mu_{\mathbf{Z}})' \\ &= E\{\delta(\mathbf{Z} - \mu_{\mathbf{Z}})(\mathbf{Z} - \mu_{\mathbf{Z}})'\}.\end{aligned}$$

By A3, $E\{(\mathbf{Z} - \mu_{\mathbf{Z}})(\mathbf{Z} - \mu_{\mathbf{Z}})'\}$ is positive definite and therefore $\sum_{i\in G_i} p_i\mathcal{B}_{i0}(\beta)$ is also positive definite. ▫

Proof of Theorem 1 Since $\hat{\mathbf{b}}$ is a BJE if n is large enough by Lemma 3 and $\hat{\mathbf{b}} \to \beta$ $w.p.1$ by (4.4), it follows that the BJE is consistent. Now given $\omega \in \Omega$, in view of the expressions of $\mathcal{A}_i(\beta)$ and $\mathcal{B}_i(\beta)$, it is clear that $(\mathcal{A}_i(\beta), \mathcal{B}_i(\beta))$ is a rational function of the SCE $\hat{s}_j$. Since $\mathcal{A}(\beta)$ and $\mathcal{B}(\beta)$ are the linear functions of $\mathcal{A}_i(\beta)$ and $\mathcal{B}_i(\beta)$ respectively, $(\mathcal{A}(\beta), \mathcal{B}(\beta))$ is a rational function of the SCE $\hat{s}_j$. By using similar arguments in Yu *et. al*. (1998, 2001), it can be proved that under the assumptions A1, A2 and A5, the SCE of probability masses s_i with MIC data is strong consistent and asymptotic normality. By Lemma 3, when n is large enough, $\hat{\beta}_n = \hat{\mathbf{b}} = (\mathcal{B}(\beta))^{-1}\mathcal{A}(\beta)$ is the BJE and it is a rational function of the SCE $\hat{s}_j$. Therefore, $\hat{\beta}_n = \hat{\mathbf{b}}$ is asymptotically normally distributed by using the delta method. ▫

REFERENCES

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

[2] Chen, C. (2006). Technical report to "Asymptotic distributions of BJE in linear regression models with mixed interval-censored data". Mathematics Department, Binghamton University.

[3] Groeneboom, P. and Wellner, J. A. (1992). Information bounds and nonparametric maximum likelihood estimation. *Birkhäuser Verlag, Basel.*

[4] Hanley, J. A. and Parnes, M. N. (1983). Nonparametric estimation of a multivariate distribution in the presence of censoring. *Biometrics.* **39** 129–139.

[5] Kong, F.H. and Yu, Q.Q. (2006). Asymptotic distributions of the Buckley-James estimator under non-standard conditions. *Statistica Sinica*, (to appear).

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

[7] Li, L.X. and Pu, Z.W. (1999). Regression models with arbitrarily interval-censored observations. *Comm. in Statist., Theory and Methods* 1547–1563.

[8] Miller, R.G. (1981). Survival Analysis. *Wiley.*

[9] Miller, R.G. (1976). Least squares regression with censored data. *Biometrika* **63** 449–464.

[10] Peto, R. (1973). Experimental survival curves for interval-censored data. *Applied Statist.* **22** 86–91.

[11] Rabinowitz, D. Tsiatis, A. and Aragon, J. (1995). Regression with interval-censored data. *Biometrika* **82** 501–513.

[12] Schick, A. and Yu, Q. Q. (2000). Consistency of the GMLE with mixed case interval-censored data. *Scan. J. of Statist.* **27** 45–55.

[13] Turnbull, B. W. (1976). The empirical distribution function with arbitrary grouped, censored and truncated data. *J. Roy. Statist. Soc. Ser.* B. **38** 290–295.

[14] Yu, Q. Q., Li, L. X. and Wong, G. Y. C. (1998). Asymptotic variance of the GMLE of a survival function with interval-censored data. *Sankhya.* **60** 184-197.

[15] Yu, Q. Q., Wong, G. Y. C. and Li, L. X. (2001). Asymptotic properties of self-consistent estimators with mixed interval-censored data. *Ann. Inst. Statist. Math..* **53** 469–486.

[16] Yu, Q.Q. and Wong, G.Y.C. (2003). Semi-parametric MLE in simple linear regression with interval–censored data. *Communications in Statistics-Simulation and Computation* **32** 147–164.