Statistica Sinica **17**(2007), 341-360

ASYMPTOTIC DISTRIBUTIONS OF THE BUCKLEY-JAMES ESTIMATOR UNDER NONSTANDARD CONDITIONS

Fanhui Kong and Qiqing Yu

Wilkes University and State University of New York

Abstract: The Buckley-James estimator (BJE) is the most appropriate extension of the least squares estimator (LSE) to the right-censored linear regression model. Lai and Ying (1991) established asymptotic normality of the BJE under a set of regularity conditions. The BJE makes use of the product-limit estimator (PLE). Both the LSE and the PLE are asymptotically normally distributed when underlying distributions are either continuous or discontinuous. It is an interesting question whether the BJE is still asymptotic normal when the underlying distributions are discontinuous. In this paper, we show that the BJE has at least four types of asymptotic distributions under various discontinuity assumptions. In particular, we establish certain conditions under which the BJE does (or does not) have an asymptotic normal distribution.

Key words and phrases: Asymptotic normality, identifiability conditions, linear regression model, right-censorship.

1. Introduction

We investigate the asymptotic distributions of the Buckley and James (1979) estimator (BJE) under the linear regression problem with right-censored data, when underlying distributions are discontinuous.

Regression analysis is one of the most widely used statistical techniques. Its applications occur in almost every field, including engineering, economics, physical sciences, management, life and biological sciences, and the social sciences. In particular, one desires to estimate the relationship between a variable Y and one or more independent variables, say a vector $\mathbf{X}$. One relationship is $Y = \beta'\mathbf{X} + \epsilon$, where β' is the transpose of a regression coefficient vector β and ϵ is a random variable with an unknown cdf F_o. $E(\epsilon)$ may or may not be zero, which is not important, as in general $E(\epsilon)$ is not identifiable under right censoring.

This is a semi-parametric set up, as β is a parameter with finite dimension and F_o is arbitrary (continuous or discontinuous). The BJE is an estimator of β under this set-up. The counterpart of the BJE in the uncensored case is the least squares estimator (LSE).

With complete data, the LSE is the common approach. Under right censoring, there are several extensions of the LSE, including the BJE. Miller and Halpern (1982) compared the performance of the BJE with extensions of the least squares method to censored data by Miller (1976) and by Koul, Susarla and Van Ryzin (1981), and with the Cox (1972) regression analysis that assumes a proportional hazards model instead of the linear regression model. From the results of these different methods applied to the Standford heart transplant data, Miller and Halpern concluded that the Cox and the Buckley-James estimators are the "two most reliable regression estimates to use with censored data" and that "the choice between them should depend on the appropriateness of the proportional hazards model or the linear model for the data."

Buckley and James proposed an algorithm to find a solution by an iterative algorithm. The algorithm may not converge to a BJE even when the BJE exists (see Yu and Wong (2002)). When the BJE based on the original definition of Buckley and James does not exist, James and Smith (1984) proposed a modification of the BJE. Yu and Wong (2002) provided the explicit expression for this BJE by proposing a non-iterative algorithm for finding all possible solutions to the BJE.

Under certain smooth assumptions on the underlying distributions, James and Smith (1984) presented a consistency result on the BJE, and Lai and Ying (1991) showed that a modified BJE is asymptotically efficient if F_o is a normal distribution and is asymptotically normally distributed.

It is possible that F_o is not normal and is not even continuous. Since there are studies on the asymptotic properties of the BJE under the assumption that F_o is normal or is continuous, it is of interest to investigate the asymptotic properties of the BJE when F_o is discontinuous. This problem has not been addressed in the literature.

Under certain regularity conditions, it is well known that the LSE without censoring is asymptotically normally distributed when F_o is either continuous or discontinuous, and the product-limit-estimator (PLE) with right censoring is also asymptotically normally distributed when the underlying distributions are either continuous or discontinuous. Since the BJE is an extension of the LSE and the BJE makes use of the PLE, one would think that when F_o is discontinuous or, in particular, when F_o takes on finitely many values, the BJE would also be asymptotically normally distributed. However, this is not true.

Our results show that the BJE $\hat{\beta}_n$ has the following asymptotic properties.

1. Under certain assumptions, with probability 0.5, the BJE does not exist. If it exists, it does not converge to β (see Examples 1 and 2).
2. Under certain assumptions, $\hat{\beta}_n \to \beta$ a.s. and $\sqrt{n}(\hat{\beta}_n - \beta)$ converges in distribution to Z, where Z has a normal distribution with mean zero ($Z \sim N(0, \sigma^2)$) (see Theorem 2).

3. Under certain assumptions, $\hat{\beta}_n \to \beta$ a.s. and $\sqrt{n}(\hat{\beta}_n - \beta)$ converges in distribution to $\min\{Z, 0\}$ or $\max\{Z, 0\}$ (see Theorem 3).
4. Under certain assumptions, $\mathrm{P}\{\hat{\beta}_n = \beta$ for all large enough $n\} = 1$ (see Theorem 4).

It is worth mentioning that in the last two cases, the BJE does not have an asymptotic normal distribution and, unlike the modified PLE proposed by Lai and Ying (1991), we do not make any modification to the BJE.

We first study the asymptotic properties of the BJE assumed discreteness. Discrete assumptions are common in the literature (see, for example, the classical textbooks on survival analysis by Cox and Oakes (1984, p.101) and by Miller (1981, p.61)). Nelson (1973) provided a discrete data set that fits the linear regression model quite well.

Here, for simplicity, most of our proofs are for the discrete case. It is possible that the asymptotic properties of the BJE under discreteness remain if F_o is discontinuous, given certain regularity conditions.

Since the main purpose of the paper is to find possible asymptotic distributions of the BJE, we take b to be a scalar, for simplicity. It is possible that similar results as 1−4 hold in the multiple linear regression setting under discontinuous assumptions.

The BJE is a special case of an M-estimator. Our results suggest that the other M-estimators may have the similar properties under discontinuity assumptions, so our findings contribute to the understanding of asymptotic properties of M-estimators. Even more, note that the BJE under right censoring has been extended to the case of interval censoring. Rabinowitz, Tsiatis and Aragon (1995) proposed a class of score statistics to estimate β. Their approach parallels the construction of the BJE for right-censored data. Li and Pu (1999) considered a generalization of the BJE for interval-censored data that contains exact observations. Zhang and Li (1996) and Li and Zhang (1998), among others, studied M-estimators with doubly-censored data and Case 1 interval-censored data. Our findings should also provide hints to properties of these M-estimators and extensions of the BJE under interval censoring.

The paper is organized as follows. In Section 2, we set the notation and introduce the algorithm for obtaining the BJE. In Section 3 we present the main results. Some detailed proofs are relegated to Section 4.

2. Notations

Consider the model $Y = \beta X + \epsilon$, where β is a scalar and ϵ and X are random variables. Let C be a censoring variable, $M = \min\{Y, C\}$, $\delta = \mathbf{1}_{(Y \leq C)}$ (the indicator function of the event $\{Y \leq C\}$), $W = C - \beta X$, $U = \min\{W, \epsilon\}$, $T =$

$T(b) = M - bX$, $\mu_x = E(X)$ and let $\mathcal{I}$ be the observable random interval, *i.e.*, $\mathcal{I} = [Y, Y]$ if $\delta = 1$, and $\mathcal{I} = (C, \infty)$ otherwise. Let $(M_i, \delta_i, X_i, C_i, \epsilon_i, W_i, U_i, T_i, \mathcal{I}_i)$, $i = 1, \ldots, n$, be i.i.d. copies of $(M, \delta, X, C, \epsilon, W, U, T, \mathcal{I})$. Given a random variable or random vector, say U, let F_U (S_U) be its distribution (survival) function, f_U its density function, and let $\bar{U} = \sum_{i=1}^n U_i/n$. In a similar manner, let $\bar{U^2}$ and $\bar{UX}$, etc. The BJE is a zero crossing of $H(\cdot)$, where

$$H(b) = \sum_{i=1}^n T_i^*(b)(X_i - \bar{X}), \tag{2.1}$$

$$(T_i^*(b), \delta_i^*) = \begin{cases} (T_i(b), 1) & \text{if } T_i(b) = \max_j T_j(b) \text{ or } \delta_i = 1, \\ \left(\frac{\sum_{t > T_i(b)} t\hat{f}_b(t)}{\hat{S}_b(T_i(b))}, \delta_i\right) & \text{otherwise.} \end{cases} \tag{2.2}$$

Here $\hat{S}_b$ is the PLE of the survival function S_o $(= 1 - F_o)$ based on $(T_i(b)^*, \delta_i^*)$s, and $\hat{f}_b$ is defined by $\hat{f}_b(t) = \hat{S}_b(t-) - \hat{S}_b(t)$, which is the PLE of f_o, the density of F_o. Thus $T_i^*(b)$ can be viewed as an estimate of $E(\epsilon_i | Y_i \in \mathcal{I}_i))$. The motivation for using T_i^* is as follows. If the largest $T_i(b)$ is right censored, $(\sum_{t > T_i(b)} t\hat{f}_b(t))/\hat{S}_b(T_i(b))$ in (2.2) is not defined and is treated as an exact observation in $H(b)$. Hence $\hat{S}_b$ is modified so that it moves the tail probability to the largest observation among the $T_i(b)$'s.

Throughout the paper, we make use of the following assumptions.

A1. ϵ and (X, C) are independent.
A2. (ϵ, C, X) takes on finitely many values.

Under A2, by the Strong Law of Large Numbers (SLLN), we have

$$\frac{H(\beta)}{n} \to E(E((X - E(X))\epsilon^* | \mathcal{I})) \text{ a.s., where } \epsilon^* = \begin{cases} \epsilon & \text{if} \quad \epsilon < \tau, \\ \tau & \text{if} \quad \epsilon \geq \tau, \end{cases} \tag{2.3}$$

and $\tau = \sup\{t : \ P\{U > t\} > 0\}$. The foregoing discussion establishes the following lemma.

Lemma 1. *Under* A1 *and* A2, $H(\beta)/n \to 0$ *a.s., as* $n \to \infty$.

Details of the proof of the lemma are given in Kong (2005).

Remark 2.1. Note that if $F_{\epsilon^*} = F_o$, then

$$\inf\{t : \ F_W(t) = 1\} \geq \inf\{s : \ F_o(s) = 1\}, \tag{2.4}$$

and $T_i^*(\beta) = T_i(\beta)$ for all i, at least when n is large enough. Thus, (2.4) is the justification for the modification at (2.2). Without loss of generality (WLOG), we can assume that (2.4) holds. Otherwise, hereafter replace ϵ by ϵ^* (see (2.3)).

Let $e_1 < \cdots < e_{m_o}$ be the possible values of ϵ such that $e_{m_o} \le \tau$ (see (2.3)), $x_1 < \cdots < x_{m_x}$ be the possible values of X , $c_1 < \cdots < c_{m_c}$ be the possible values of C, and $u_1 < \cdots < u_{m_u}$ be the possible values of U. For all possible j and k, let $t_{1jk}(b) = e_j + (\beta - b)x_k$, $t_{2jk}(b) = c_j - bx_k$, $d_j = \sum_{h=1}^n \mathbf{1}_{(\epsilon_h = e_j, \delta_h = 1)}$, $r_j = \sum_{h=1}^n \mathbf{1}_{(T_h(\beta) \ge e_j)}$, $d_{1jk}(b) = \sum_{i=1}^n \mathbf{1}_{(T_i(b) = t_{1jk}(b), \delta_i = 1)}$, $r_{1jk}(b) = \sum_{i=1}^n \mathbf{1}_{(T_i(b) \ge t_{1jk}(b))}$, $n_{1jk}(b) = \sum_{i=1}^n \mathbf{1}_{(T_i(b) = t_{1jk}(b), X_i = x_k, \delta_i = 1)}$, $n_{2jk}(b) = \sum_{i=1}^n \mathbf{1}_{(T_i(b) = t_{2jk}(b), X_i = x_k, \delta_i = 0)}$, $p_{1jk} = \lim_{n\to\infty} n_{1jk}/n$ a.s. and $p_{2jk} = \lim_{n\to\infty} n_{2jk}/n$ a.s.. Abusing notation, we write $d_{1jk} = d_{1jk}(b)$ and $r_{1jk} = r_{1jk}(b)$, etc.

In order to derive all solutions to the BJE, Yu and Wong (2002) made use of the following notation. Let b_{ij} be the solution to an equation $T_i(b) = T_j(b)$, where $X_i \ne X_j$. Let $b_1 < \cdots < b_{m_\beta}$ be all the distinct values of the b_{ij}'s. Let $\mathcal{B}_o = \{b_1, \ldots, b_{m_\beta}\}$. Let $b_0 = -\infty$ and $b_{m_\beta+1} = \infty$. Let $\mathcal{B}$ be the subset of $\mathcal{B}_o$ such that each element of $\mathcal{B}$ is the solution to an equation $T_i(b) = T_j(b)$, where $X_i \ne X_j$ and $\delta_i \cdot \delta_j = 0$. Let $\mathcal{B} = \{q_1, \ldots, q_{m_b}\}$. Let $q_0 = -\infty$ and $q_{m_b+1} = \infty$.

Lemma 2.(Yu and Wong (2002, Remark 3.1)) *If β is a scalar then, given j,*
1. *for each i, the rank (or order) of $T_i(b)$ remains the same if $b \in (b_j, b_{j+1})$;*
2. *for each i, $\hat{S}_b(T_i(b))$ is constant in b on (b_j, b_{j+1});*
3. *$H(\cdot)$ is linear (in b) on (b_j, b_{j+1}).*

Remark 2.2. Statement (3) in Lemma 2 can be modified as follows: $H(\cdot)$ is linear in b on the interval (q_i, q_{i+1}) for each i (see Kong (2005)).

Based on Remark 2.2, the original algorithm proposed by Yu and Wong (2002) for finding all BJEs can be improved as follows.

The algorithm for the BJE.
1. For each $q_h \in \mathcal{B}$, compute the PLE $\hat{S}_b$ for a $b \in (q_h, q_{h+1})$. For example, let b be the midpoint of the interval (q_h, q_{h+1}) if $0 < h < m_b$, $b = q_1 - 1$ if $h = 0$, and let $b = q_{m_b} + 1$ if $h = m_b$. Denote such b by a_h and compute
$$\hat{b}_h = \frac{\sum_{j=1}^n (X_j - \bar{X}) M_j^*(a_h)}{\sum_{k=1}^n (X_k - \bar{X}) X_k^*(a_h)}, \tag{2.5}$$
where $(M_i^*(\cdot), X_i^*(\cdot))$ is an estimate of $E((M_i, X_i)|Y_i \in I_i)$, or
$$\begin{aligned} M_i^*(b) &= M_i\delta_i + (1-\delta_i) \sum_{t > T_i(b)} \frac{\hat{f}_b(t)}{\hat{S}_b(T_i(b))} \frac{\sum_{j=1}^n M_j \mathbf{1}_{(T_j(b)=t, \delta_j=1)}}{\sum_{k=1}^n \mathbf{1}_{(T_k(b)=t, \delta_k=1)}}, \\ X_i^*(b) &= X_i\delta_i + (1-\delta_i) \sum_{t > T_i(b)} \frac{\hat{f}_b(t)}{\hat{S}_b(T_i(b))} \frac{\sum_{j=1}^n X_j \mathbf{1}_{(T_j(b)=t, \delta_j=1)}}{\sum_{k=1}^n \mathbf{1}_{(T_k(b)=t, \delta_k=1)}}, \ i = 1, \ldots, n. \end{aligned} \tag{2.6}$$
2. If $\hat{b}_h \in (q_h, q_{h+1})$ then $\hat{b}_h$ is a solution to equation $H(b) = 0$ and is a BJE of β.

3. Compute $H(q_i-)$, $H(q_i)$ and $H(q_i+)$, $i = 1, ..., m_b$, where $H(q_i+)$ and $H(q_i-)$ are the right- and left-hand limits of H, respectively. By (2.1), (2.2) and (2.6) we have

$$H(b) = \begin{cases} \sum_{j=1}^{n} (X_j - \bar{X})(M_j^*(q_i) - bX_j^*(q_i)) & \text{if} \quad b = q_i,\ 1 \le i \le m_b, \\ \sum_{j=1}^{n} (X_j - \bar{X})(M_j^*(a_i) - bX_j^*(a_i)) & \text{if} \quad b \in (q_i, q_{i+1}),\ 0 \le i \le m_b. \end{cases} \tag{2.7}$$

If $H(q_i-)H(q_i+) \le 0$, or $H(q_i-)H(q_i) \le 0$, or $H(q_i)H(q_i+) \le 0$, then q_i is a BJE.

3. Main Results

We investigate the asymptotic properties of the BJE in this section. The proofs of the lemmas are given in Section 4. Some detailed and tedious proofs of the statements in examples, and certain statements in the proofs of lemmas and theorems, are given in more detail in Kong (2005).

The identifiability assumption made under the uncensored case is $P\{X_1 \ne X_2\} > 0$. In order to understand a modification of the identifiability condition to right censoring, we first look at the following example.

Example 1. Let $\beta = 1$. Suppose that ϵ and $X \sim bin(1, 1/2)$ and $C \equiv 0.5$. Then it can be shown (see Kong (2005)) that

$$\begin{aligned} &\text{with probability (w.p.) approximately } 1/2 \text{ there is no BJE} \\ &\text{and w.p. approximately } 1/2\ \hat{\beta}_n = 0.5 \text{ is a BJE.} \end{aligned} \tag{3.1}$$

Thus, the BJE, if it exists, is not consistent and is not normally distributed.

In Example 1, we have the naive extension of assumption $P\{X_1 \ne X_2\} > 0$ to the censoring case. Example 1 indicates that it is not the identifiability condition in the censored regression. The main feature in this example is that $P\{\delta_1 = \delta_2 = 1 \text{ and } X_1 \ne X_2\} = 0$. This justifies the following identifiability condition for the simple linear regression model.

A3. $P\{\delta_1 = \delta_2 = 1 \text{ and } X_1 \ne X_2\} > 0$.

Note that in Example 1, the random variables are all discrete. The same phenomenon will occur under continuous cases. The following is such an example.

Example 2. Suppose that $\beta = 1$, $C \sim U(0, 0.5)$, $\epsilon \sim U((-0.1, 0) \cup (1, 1.1))$, $X \sim bin(1, 1/2)$, and C and X are independent. Then (3.1) still holds. Thus there is no consistent BJE. The proof is similar to that for (3.1) in Example

1. Here A3 does not hold, and therefore has nothing to do with the continuity assumption on ϵ.

If one assumes that ϵ and (C, X) are continuous and independent, and assumes that $\mathrm{P}\{\delta = 1\} > 0$, then $P\{X_1 \neq X_2\} > 0$ implies A3. Thus A3 is not even mentioned in Lai and Ying (1991) or in James and Smith (1984).

Under A1 and A2, the LSE in the case of complete data and the PLE are both asymptotically normally distributed. Consequently, one would expect that the BJE would also be asymptotically normally distributed. However the following is a counterexample.

Example 3. Let $\beta = 1$. Suppose that ϵ and $X \sim bin(1, 1/2)$ and $C \equiv 1$. It can be shown (see Kong (2005)) that the BJE $\hat{\beta}_n$ is consistent and

$$\begin{aligned}&\sqrt{n}(\hat{\beta}_n - \beta) \text{ converges in distribution to } \min\{Z, 0\},\\ &\text{where } Z \sim N(0, \sigma^2) \text{ and } \sigma > 0.\end{aligned} \tag{3.2}$$

The main feature in Example 3 is that $C - \beta X = \epsilon$ if $(C, X, \epsilon) = (1, 1, 0)$. That is,

A4. $\mathrm{P}\{C - \beta X = \epsilon < \tau\} > 0$, where τ is given in (2.3).

A4 says that the cdfs of $C - \beta X$ and F_o share a common discontinuity point. We establish a theorem that if A4 does not hold then the BJE may still be asymptotic normal if F_o is not continuous.

Before we present the theorem, we need to establish some preliminary results.

Remark 3.1. Since we consider the regression model, WLOG, we can assume that $\beta \neq 0$. Furthermore, we can assume $\beta = 1$. Otherwise, replace X by X/β. Moreover, we can assume that $X > 0$. The reason is as follows. A2 implies that X is bounded. By subtracting a lower bound d of X from X, Y and C, respectively, resulting in $X^{(n)}$, $Y^{(n)}$ and $C^{(n)}$, the model becomes $(Y - d) = \beta(X - d) + \epsilon$ $(Y^{(n)} = \beta X^{(n)} + \epsilon)$, where $\beta = 1$ and the observable random vector becomes $(M - d, \delta)$ $(= (M^{(n)}, \delta))$ with $M - d = \min\{Y - d, C - d\}$ $(M^{(n)} = \min\{Y^{(n)}, C^{(n)}\})$ and $\delta = \mathbf{1}_{((Y-d) \leq (C-d))}$ $(= \mathbf{1}_{(Y^{(n)} \leq C^{(n)})})$. Now $X^{(n)} = X - d > 0$.

Under A1 and A3, we can assume that $\beta \in \mathcal{B}_o$. Denote $b_{i_o} = \beta$. Under A2, by letting n be large enough, we can further assume that both $\mathcal{B}$ and $\mathcal{B}_o$ do not depend on n. Notice that the PLE is

$$\hat{S}_b(t) = \prod_{\substack{(j,k): t_{1jk}(b) \leq t \\ \text{over all distinct } t_{1jk}(b)s}} \Big(1 - \frac{d_{1jk}(b)}{r_{1jk}(b)}\Big) \text{ and } \hat{S}_\beta(t) = \prod_{i:\ e_i \leq t} (1 - \frac{d_i}{r_i}). \tag{3.3}$$

Take

$$f^-_{hm} = \lim_{n\to\infty} \hat{f}^-_{hm}, \text{ where } \hat{f}^-_{hm} = \hat{f}_b(t_{1hm}(b)), \text{ if } b \in (b_{i_o-1}, \beta);$$
$$f_{hm} = \lim_{n\to\infty} \hat{f}_{hm}, \text{ where } \hat{f}_{hm} = \hat{f}_\beta(e_h)\frac{\sum_{j=1}^n \mathbf{1}_{(T_j(\beta)=e_h,\delta_j=1,X_j=x_m)}}{\sum_{k=1}^n \mathbf{1}_{(T_k(\beta)=e_h,\delta_k=1)}}; \tag{3.4}$$
$$f^+_{hm} = \lim_{n\to\infty} \hat{f}^+_{hm}, \text{ where } \hat{f}^+_{hm} = \hat{f}_b(t_{1hm}(b)), \text{ if } b \in (\beta, b_{i_o+1}).$$

Remark 3.2. By 2 of Lemma 2, for $b \in (b_i, b_{i+1})$, $\hat{f}_b(t_{1hm}(b))$ is constant in b, so $\hat{f}^-_{hm}$, $\hat{f}^+_{hm}$ and $\hat{f}_{hm}$ do not depend on b and neither do f^-_{hm}, f_{hm} and f^+_{hm}. Moreover, verify that for $b \approx \beta$ and $b \neq \beta$, $d_{1jk}(b)$ is constant in b.

Lemma 3. *Suppose that* A2 *holds.*

1. *If for each triple (i,j,k), $c_i - \beta x_k \neq e_j$, except perhaps for one triple (i,j,k) with $k = m_x$, then $\hat{f}^-_{hm} = \hat{f}_\beta(e_h)d_{1hm}/d_h$ and $f^-_{hm} = P\{X = x_m|e_h \leq W\}f_o(e_h)$.*
2. *If for each triple (i,j,k), $c_i - \beta x_k \neq e_j$, except perhaps for one triple (i,j,k) with $k = 1$, then $\hat{f}^+_{hm} = \hat{f}_\beta(e_h)d_{1hm}/d_h$ and $f^+_{hm} = P\{X = x_m|e_h \leq W\}f_o(e_h)$.*
3. *If $\beta \notin \mathcal{B}$, then $\hat{f}^+_{hm} = \hat{f}^-_{hm} = \hat{f}_{hm} = \hat{f}_\beta(e_h)d_{1hm}/d_h$ and $f^-_{hm} = f_{hm} = f^+_{hm} = P\{X = x_m|e_h \leq W\}f_o(e_h)$.*

Mimicking the expression of $H(b)$ in (2.7), define

$$\mathcal{H}(b) = \sum_{j=1}^n (X_j - \bar{X})(M^*_j - bX^*_j), \text{ where } M^*_j = M^*_j(\beta) \text{ and } X^*_j = X^*_j(\beta) \tag{3.5}$$

(see (2.6)). In general, $\mathcal{H}(b) = H(b)$ may not be true.

$$\text{Let } \hat{b} \text{ be the solution to } \mathcal{H}(b) = 0, \text{ that is, } \hat{b} = \frac{\sum_{j=1}^n (X_j - \bar{X})M^*_j}{\sum_{k=1}^n (X_k - \bar{X})X^*_k}. \tag{3.6}$$

Theorem 1. *If* A1, A2 *and* A3 *hold, then*

1. $H(\beta)/\sqrt{n} \xrightarrow{D} N(0, \sigma^2_H)$, *where* $\sigma^2_H = \mathrm{Var}\,(\mathcal{T})$, $\mathcal{T} = g_1(\epsilon, X)\delta + g_2(C, X)(1 - \delta)$,

$$g_1(e_s, x_t) = (x_t - \mu_x)e_s + \sum_{(j,k):t_{2jk}<e_s} p_{2jk}(x_k - \mu_x)\sigma_{sjk} - (x_t - x_{m_x})E(\epsilon^*) \ (\textit{see } (2.3))$$
$$g_2(c_s, x_t) = -(x_t - x_{m_x})E(\epsilon^*) + (x_t - x_{m_x})\nu_{st},$$
$$\nu_{jk} = \sum_{h:e_h>t_{2jk}} \frac{e_h \sum_{k'} p_{1hk'} \prod_{m:t_{2jk}<e_m<e_h}\left(1 - \frac{\sum_{k'} p_{1mk'}}{\sum_{(i,k'):i\geq m} p_{1ik'} + \sum_{(j',k'):t_{2j'k'}\geq e_m} p_{2j'k'}}\right)}{\sum_{(i,k'):i\geq h} p_{1ik'} + \sum_{(j',k'):t_{2j'k'}\geq e_h} p_{2j'k'}},$$

$$\sigma_{jks} = \sum_{h:t_{2jk}<e_h\le e_s} \frac{e_h f_o(e_h)}{S_U(e_h-)S_o(t_{2jk})}\Big[S_U(e_h-)\sum_{m:t_{2jk}<e_m<e_h}\frac{f_o(em)}{S_o(e_m)S_U(e_m-)}-1\Big] + \frac{e_s}{S_W(e_s-)S_o(t_{2jk})} + \sum_{h:h>s}\frac{e_h f_o(e_h)}{S_o(t_{2jk})}\sum_{m:t_{2jk}<e_m\le e_s}\frac{f_o(e_m)}{S_o(e_m)S_U(e_m-)};$$

2. $\sqrt{n}(\hat{b}-\beta) \xrightarrow{D} N(0,\sigma_\beta^2)$, *where* $\sigma_\beta^2 = \sigma_H^2/\sigma_2^2$ *and* $\sigma_2 = \lim_{n\to\infty}(X\bar{X}^* - \bar{X}\cdot\bar{X}^*) > 0$ *a.s.*.

Let $\mu_{\epsilon^*} = E(\epsilon^*)$. Since $\epsilon = M - \beta X$ if $\delta = 1$ and $W = M - \beta X$ if $\delta = 0$, $\mathcal{T}$ is a function of $(M, \delta, X, \mu_x, \mu_{\epsilon^*}, \beta)$, say $\mathcal{T} = \mathcal{T}(M, \delta, X, \mu_x, \mu_{\epsilon^*}, \beta)$. Consequently, an estimate of σ_H^2 is $\tilde{\sigma}_H^2 = \overline{\hat{\mathcal{T}}^2} - (\overline{\hat{\mathcal{T}}})^2$, where $\hat{\mathcal{T}}_i = \mathcal{T}(M_i, \delta_i, X_i, \hat{\mu}_x, \hat{\mu}_{\epsilon^*}, \hat{\beta}_n)$, $\hat{\mu}_x$ and $\hat{\mu}_{\epsilon^*}$ are empirical estimates of μ_x and μ_{ϵ^*}, respectively, and $\hat{\beta}_n$ is the BJE.

Proof of Theorem 1. The main idea of the proof is under A2, both H and $\hat{b}$ are algebraic functions of the sample mean of a random vector with finite dimension. Thus by the Central Limit Theorem (CLT) and the delta method, we can establish the asymptotic normality of $H(\beta)/n$ and $\hat{b}$.

We first prove Statement 1. For simplicity, write $t_{2jk} = t_{2jk}(\beta)$, etc. By (2.6), (2.7) and (3.3),

$$\begin{aligned}\frac{H(\beta)}{n} &= \frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})\Big(\delta_i\epsilon_i + (1-\delta_i)\frac{\sum_{(h,m):\ e_h>W_i} e_h\hat{f}_\beta(e_h)\frac{n_{1hm}}{r_h}}{\hat{S}_\beta(W_i)}\Big)\\ &= \sum_{i,k}\frac{n_{1ik}}{n}(x_k-\bar{X})e_i\\ &\quad + \sum_{j,k}\frac{n_{2jk}}{n}(x_k-\bar{X})\sum_{h:\ e_h>t_{2jk}} e_h\frac{d_h}{r_h}\prod_{m:\ t_{2jk}<e_m<e_h}(1-\frac{d_m}{r_m}) \qquad (3.7)\end{aligned}$$

(as $\hat{f}_\beta(e_h) = \hat{S}_\beta(e_h-)d_h/r_h$). Details of the proof of (3.7) are given in Kong (2005). By the SLLN, $p_{1jk} = P\{\epsilon = e_j \le W, X = x_k\}$ and $p_{2jk} = P\{\epsilon > W, C = c_j, X = x_k\}$. Let $p_h = \lim_{n\to\infty} d_h/n$ a.s.. The existence of these (almost sure) limits is guaranteed by the SLLN. Verify that $\lim_{n\to\infty} r_h/n = S_U(e_h-)$ a.s.. The (almost sure) limit of $H(\beta)/n$ is

$$\begin{aligned}H_o &= \sum_{i,k} p_{1ik}(x_k-\mu_x)e_i\\ &\quad + \sum_{j,k} p_{2jk}(x_k-\mu_x)\sum_{h:e_h>t_{2jk}}\frac{e_h p_h}{S_U(e_h-)}\prod_{m:t_{2jk}<e_m<e_h}(1-\frac{p_m}{S_U(e_m-)})\end{aligned}$$

(by (3.7)). Verify that for each t, $S_U(t)$ is a function of p_{ijk} and

$$p_h = \sum_k p_{1hk},\ \mu_x = \sum_{i,k} p_{1ik}x_k + \sum_{j,k} p_{2jk}x_k \text{ and } \sum_{i,j,k} p_{ijk} = 1. \qquad (3.8)$$

Hence, we can write $H_o = H_o(v)$, where v is a finite-dimensional (say $m_v \times 1$ dimensional) vector whose components are p_{ijk}, for all possible (i, j, k), except for one p_{2jk}, say $p_{2m_c m_x}$, provided that for this particular (j, k), we have

$$P\{1_{(\epsilon > W, C=c_j, X=x_k)} = 1\} \in (0, 1). \qquad (3.9)$$

WLOG, we can assume that (3.9) holds for $(j, k) = (m_c, m_x)$. Let $\hat{v}$ be the estimator of v that estimates the components of v by $\hat{p}_{ijk} = n_{ijk}/n$. Since $\hat{p}_{1jk} = (1/n)\sum_{i=1}^n \mathbf{1}_{(\epsilon_i = e_j \le W_i, X_i = x_k)}$, $\hat{p}_{2jk} = (1/n)\sum_{i=1}^n \mathbf{1}_{(C_i = c_j, W_i < \epsilon_i, X_i = x_k)}$, one can write $\hat{v} = \bar{V}$, where V is a random vector of finite dimension with components $\mathbf{1}_{(\epsilon = e_j \le W, X = x_k)}$, $\mathbf{1}_{(\epsilon > W, C = c_i, X = x_k)}$, $j = 1, \ldots, m_o$, $i = 1, \ldots, m_c$, $k = 1, \ldots, m_x$, except for the term $1_{(\epsilon > W, C = c_{m_c}, X = x_{m_x})}$. By Lemma 1, $H_o(v) = \lim_{n\to\infty} H(\beta)/n = 0$ a.s., thus, $H(\beta)/n = H_o(\hat{v}) = H_o(\hat{v}) - H_o(v)$. Since $\hat{v}$ is a sample mean of $m_v \times 1$ dimensional random vector, m_v is a finite integer independent of n, and H_o has continuous partial derivatives, the CLT yields the asymptotic normality of $H(\beta)/n$. The derivation of σ_H^2 is based on the delta method, thus it is a trivial but tedious calculation. Its proof can be found in Kong (2005). It completes the proof of Statement 1.

We now prove Statement 2. Recall that $\mathcal{H}(\hat{b}) = 0$ and $H(\beta) = \mathcal{H}(\beta)$. It follows that

$$\sqrt{n}\frac{H(\beta)}{n} = \sqrt{n}\frac{\mathcal{H}(\beta) - \mathcal{H}(\hat{b})}{n} = \sqrt{n}(\hat{b} - \beta)(\bar{XX^*} - \bar{X} \cdot \bar{X^*}) \text{ (by (3.5))}. \qquad (3.10)$$

It can be shown (see Kong (2005)) that $\sigma_2 = \lim_{n\to\infty}(\bar{XX^*} - \bar{X} \cdot \bar{X^*}) > 0$ a.s.. As a consequence of (3.10), $\sqrt{n}(\hat{b} - \beta) = H(\beta)/((\bar{XX^*} - \bar{X} \cdot \bar{X})\sqrt{n})$. By Statement 1 and Slutsky's Theorem, $\sqrt{n}(\hat{b} - \beta) \xrightarrow{D} N(0, \sigma_\beta^2)$, where $\sigma_\beta^2 = \sigma_H^2/\sigma_2^2$.

Theorem 2. *If* A1, A2 *and* A3 *hold and* $P\{C - \beta X = \epsilon\} = 0$, *then* $\hat{\beta}_n = \hat{b}$ *given by* (3.6) *is a BJE if* n *is large enough,* $\hat{\beta}_n$ *is consistent and* $\sqrt{n}(\hat{\beta}_n - \beta) \xrightarrow{D} N(0, \sigma_\beta^2)$.

It is worth mentioning that σ_β^2 (see Theorem 1) derived under the discontinuous assumptions is not the same as the expression for the variance of the BJE under the smoothness assumptions given in Lai and Ying (1991, (4.4)). Notice that both the limiting variance in Lai and Ying and that in Theorems 1 and 2 involve the error density f_o. Lai and Ying made strong assumptions on f_o, while under A2 in our set-up, f_o is essentially a finite-dimensional parameter. Of course, it is also a very strong assumption on f_o.

Proof of Theorem 2. Our notation has $b_{i_o} = \beta$. It is known (see Kong (2005)) that

$$\text{if } P\{C - \beta X = \epsilon\} = 0 \text{ then } H(b) = \mathcal{H}(b) \text{ for each } b \in (b_{i_o - 1}, b_{i_o + 1}). \qquad (3.11)$$

Let $\hat{b}$ be a root of $H(b)$ given by (3.6). By (3.11) and the definition of the BJE, $\hat{b}$ is a BJE if $\hat{b} \in (b_{i_o-1}, b_{i_o+1})$. Under the assumptions of Theorem 2, by taking a large sample size and by (3.11), we can assume that $\hat{b} \in (b_{i_o-1}, b_{i_o+1})$ and thus $\hat{\beta}_n = \hat{b}$ is a BJE. The asymptotic normality follows from Theorem 1. The consistency follows from the equation $\hat{\beta}_n - \beta = (X\bar{X}^* - \bar{X} \cdot \bar{X}^*)^{-1} H(\beta)/n \to 0$ a.s. (see (3.10)) and from Theorem 1 and Lemma 1. This completes the proof of Theorem 2.

In the proof of (3.2) (see Kong (2005)), it is proved that, w.p.1, $H(\beta-) = H(\beta)$ but $\lim_{n\to\infty} H(\beta+)/n < 0$. In fact, we can establish the following lemma.

Lemma 4. *Assume that* A1, A2, A3 *and* A4 *hold.*

1. *If for each* (i, j, k), $c_i - \beta x_k \neq e_j$, *except for only one triple* (i, j, k) *with* $k = m_x$, *then w.p.1,* $\lim_{n\to\infty} H(\beta+)/n < 0$ *but* $H(\beta-) = H(\beta)$.
2. *If for each* (i, j, k), $c_i - \beta x_k \neq e_j$, *except for only one triple* (i, j, k) *with* $k = 1$, *then w.p.*1, $\lim_{n\to\infty} H(\beta-)/n > 0$ *but* $H(\beta+) = H(\beta)$.

Theorem 3. *Suppose that* A1, A2, A3 *and* A4 *hold.*

1. *If the condition in Statement* 1 *of Lemma* 4 *holds, then there is a BJE* $\hat{\beta}_n$ *which is consistent and* $\sqrt{n}(\hat{\beta}_n - \beta)$ *converges in distribution to* $\min\{0, Z\}$, *where* $Z \sim N(0, \sigma_\beta^2)$.
2. *If the condition in Statement* 2 *of Lemma* 4 *holds, then there is a BJE* $\hat{\beta}_n$ *that is consistent and* $\sqrt{n}(\hat{\beta}_n - \beta)$ *converges in distribution to* $\max\{0, Z\}$.

Proof of Theorem 3. The two statements in the theorem are symmetric. Since Example 3 is a special case of Statement 1, we only prove the second statement. Now assume that the assumption in Statement 2 holds. By Lemma 4, $H(\beta+) = H(\beta)$. Hereafter, unless we mention, otherwise, we assume $b \in (\beta, b_{i_o+1})$ $(b_{i_o} = \beta)$. By (2.7),

$$H(b) = \sum_{i=1}^{n} (X_i - \bar{X})(M_i^*(a_1) - bX_i^*(a_1)), \tag{3.12}$$

where $a_1 = (\beta + b_{i_o+1})/2$. It can be shown (see Kong (2005)) that

$$M_i^*(a_1) = M_i^*(\beta) \ (= M_i^*) \text{ and } X_i^*(a_1) = X_i^*(\beta) \ (= X_i^*). \tag{3.13}$$

By (3.5), (3.12) and (3.13),

$$H(b) = \mathcal{H}(b) \text{ for each } b \in [\beta, b_{i_o+1}), \tag{3.14}$$

where $\mathcal{H}$ is given in (3.5). By (3.5) and (3.14), with probability 1,

$$\lim_{n\to\infty} \frac{H(b) - H(\beta)}{(\beta - b)n} = \lim_{n\to\infty} \frac{\mathcal{H}(b) - \mathcal{H}(\beta)}{(\beta - b)n} = \lim_{n\to\infty} \{X\bar{X}^* - \bar{X} \cdot \bar{X}^*\} > 0, \ b \in (\beta, b_{i_o+1}).$$

The last inequality follows from Statement 2 of Theorem 1. It implies that $\lim_{n\to\infty} H(b_{i_o+1}-)/n < \lim_{n\to\infty} H(\beta)/n = 0$ a.s. (by Lemma 1). Let $\hat{b}$ be given by (3.6). Then by (3.10), (3.12) - (3.14), $(\hat{b}-\beta)(X\bar{X}^* - \bar{X}\cdot\bar{X}^*) = H(\beta)/n$. Hence

$$H(\beta) > 0 \text{ iff } \hat{b} > \beta \text{ (as } \hat{b} \text{ is the root of } \mathcal{H}(b)). \tag{3.15}$$

By Theorem 1, $H(\beta)/\sqrt{n} \to N(0, \sigma_H^2)$ in distribution. Thus,

$$P\{H(\beta) > 0\} \to \tfrac{1}{2} \text{ and } \mathrm{P}\{H(\beta) < 0\} \to \tfrac{1}{2}. \tag{3.15}$$

By Statement 2 of Theorem 1, if n is large enough, we have

$$\hat{b} \in (b_{i_o-1}, b_{i_o+1}) \text{ as } \beta \in (b_{i_o-1}, b_{i_o+1}). \tag{3.16}$$

It follows that

(a) with approximate probability 1/2, $H(\beta) > 0$ (by (3.16)) and there is a root of $H(b)$ $(= \mathcal{H}(b))$, say $\hat{b} \in (\beta, b_{i_o+1})$ (by (3.17) and (3.15)).

(b) with approximate probability 1/2, $H(\beta) < 0$ (by (3.16)) and there is a root of $\mathcal{H}(b)$, say $\hat{b} \in (b_{i_o-1}, \beta)$ (by (3.17) and (3.15)).

(c) $P(\sqrt{n}(\hat{b}-\beta) \le t)$ is approximately the same as the cdf of $N(0, \sigma_\beta^2)$ for each t by Statement 2 of Theorem 1, where σ_β^2 is given in Theorem 1.

As a consequence, in (a), $\hat{b}$ is a BJE by definition, denoted by $\hat{\beta}_n$. Moreover, by (c), the BJE $\hat{\beta}_n$ satisfies that $P(\sqrt{n}(\hat{\beta}_n - \beta) \le t)$ is approximately the same as the cdf of $N(0, \sigma_\beta^2)$ for each $t > 0$. On the other hand in case (b), $H(\beta) < 0$ (when $\hat{b} < \beta$). Since $H(\beta-) > 0$ w.p.1 by Statement 2 of Lemma 4, $\hat{\beta}_n = \beta$ is the BJE of β approximately w.p.1/2, as it is a zero crossing of $H(\cdot)$. Consistency is obvious. In summary, Statement 2 of the theorem holds.

Remark 3.3. In order to simplify the proof of the conclusion, in Theorem 3 we make use of the assumption that there is just one value of (C, X, ϵ) satisfying $C - X = \epsilon$. However, this restriction can be relaxed. For example, assume that $\beta = 1$, ϵ takes three values -1, 0 and 1 with equal probability 1/3, X takes three values -1, 1 and 2 with equal probability 1/3, and $C \equiv 1$. Verify that there are two values of (C, X, ϵ) satisfying $C - \beta X = \epsilon$ and statement (3.2) holds. The proof is given in Kong (2005).

By the definition of zero crossing, a BJE $\hat{\beta}_n$ equals β if n is large enough in the following case:

A5. $\lim_{n\to\infty}(H(\beta-)/n)(H(\beta+)/n) < 0$ w.p.1.

In the following theorem, we provide a situation under which A5 holds.

Theorem 4. *Assume* A1, A2 *and* A3 *hold. If for each* i, j, k, $c_i - \beta x_k \neq e_j$, *except for only two triples, say* (i_o, j_o, k_o) *and* (i'_o, j_o, k'_o) *with* $k_o = 1, k'_o = m_x$

and $c_{i_o} - \beta x_{k_o} = c_{i'_o} - \beta x_{k'_o} = e_{j_o} < e_{m_o}$*, then A5 holds and* $P\{\hat\beta_n = \beta$ *if* n *is large enough*$\} = 1$.

The proof of Theorem 4 is pretty long, see Kong (2005). Instead, we prove the theorem in the special case given in Example 4 below.

Example 4. Let $\beta = 1$. Suppose that (1) $\epsilon \sim bin(1, 1/2)$, (2) X and C take values $-2, 1$, with equal probability 1/2, respectively, and (3) ϵ, X and C are independent. We show (see Section 4) that $(H(\beta-), H(\beta), H(\beta+))/n \to (1/24, 0, -1/8)$ $w.p.1$. Thus A5 holds and $P\{\hat\beta_n = \beta$ if n is large enough$\} = 1$.

Remark 3.4. The assumption A2 can be relaxed. For example, in Example 3, we can assume that ϵ takes on countably many values: 0, ± 1, ... Moreover, in Theorem 2, the assumptions can be reduced to that A1 and A3 hold, and A4 holds.

Remark 3.5. It is possible that under multiple linear regression ($p > 1$) with discrete assumptions, one can still establish the four theorems with some modifications on the assumptions. For instance, Theorem 2 is valid if A3 is replace by A3*.

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

We skip the details.

Appendix

Proofs of most lemmas and some statements in Section 3 are here.

Proof of Lemma 3. It is obvious that either (1) $\mathrm{P}\{C - \beta X = \epsilon\} = 0$, or (2) $\mathrm{P}\{C - \beta X = \epsilon\} > 0$. We give the proof in both cases.

Case (1). We first prove Statement 1. Assume $b \in (b_{i_o-1}, \beta)$. We can assume $x_i > 0$ (see Remark 3.1). It is easy to see that for $b \in (b_{i_o-1}, \beta)$,

$$\prod_{k=1}^{m_x}(1 - \frac{d_{1ik}(b)}{r_{1ik}(b)}) = \frac{r_{1i2}}{r_{1i1}}\frac{r_{1i3}}{r_{1i2}}\frac{r_{1i4}}{r_{1i3}}\cdots\frac{r_i - d_i}{r_{1im_x}} = \frac{r_i - d_i}{r_{1i1}} = 1 - \frac{d_i}{r_i}. \tag{A.1}$$

Hereafter, abusing notations, we may suppress (b) in $d_{hik}(b)$, etc. Then for b in (b_{i_o-1}, β),

$$\begin{aligned}\hat f_b(t_{1hm}(b)) &= \hat S_b(t_{1hm}(b)-)\frac{d_{1hm}(b)}{r_{1hm}(b)} && \text{(by (3.3))}\\ &= (\prod_{i:\ e_i<e_h}\prod_{k=1}^{m_x}(1 - \frac{d_{1ik}}{r_{1ik}}))\prod_{k=1}^{m-1}(1 - \frac{d_{1hk}}{r_{1hk}})\frac{d_{1hm}}{r_{1hm}} && \text{(by (3.3))}\end{aligned}$$

$$= \prod_{i:\ e_i<e_h}(1-\frac{d_i}{r_i})\prod_{k=1}^{m-1}(1-\frac{d_{1hk}}{r_{1hk}})\frac{d_{1hm}}{r_{1hm}} \qquad \text{(by (A.1))}$$
$$= \hat{S}_\beta(e_h-)\frac{d_{1hm}}{r_{1h1}} \qquad \text{(by (3.3))}$$
$$= \hat{S}_\beta(e_h-)\frac{d_h}{r_h}\frac{d_{1hm}}{d_h}$$
$$= \hat{f}_\beta(e_h)\frac{d_{1hm}}{d_h} \qquad \text{(by (3.3))} \tag{A.2}$$
$$\stackrel{a.s.}{\to} f_o(e_h)P\{X=x_m|e_h\le W\}.$$

By Remark 3.2 and 3.4, taking the limit as $b\uparrow\beta$ yields

$$f^-_{hm} = f_o(e_h)P\{X=x_m|e_h\le W\}. \tag{A.3}$$

Thus Statement 1 of Lemma 3 holds in case (1).

Now assume that (1) is true and $b\in(\beta,b_{i_o+1})$. Then $e_j < t_{1jm_x}(b) < \cdots t_{1j1}(b) < e_{j+1}$ and there is no censoring in (e_j,e_{j+1}). Statement 2 can be proved in a similar manner as Statement 1, with minor modifications due to the fact that $\beta-b<0$ when $b>\beta$. Thus if (1) is true then we have

$$f^+_{hm} = f_o(e_h)P\{X=x_m|e_h\le W\}. \tag{A.4}$$

By (A.3), (A.4) and the definition of f_{hm}, Statement 3 follows. This completes the proof of the lemma if (1) is true.

Case (2). Suppose that the assumption in Statement 1 holds and $c_i-\beta x_{m_x}=e_j$. Then $t_{1j1}(b) < t_{1j2}(b) < \cdots < t_{1jm_x}(b) = t_{2im_x}(b)$ for each $b\in(b_{i_o-1},b_{i_o})$, as $t_{2im_x}(b) = c_i - bx_{m_x} = e_j + (\beta-b)x_{m_x} = t_{1jm_x}(b) > \cdots > e_j + (\beta-b)x_1 = t_{1j1}(b)$. By assumption, there is no censoring in $(t_{1j1}(b), t_{1jm_x}(b))$. Thus the proof parallels the arguments after (A.1). We skip the details.

Statement 2 can be proved in a similar manner. We skip the details. Statement 3 is not relevant in case (2). This completes the proof of the lemma.

Proof of Lemma 4. By symmetry, it suffices to establish Statement 2 of Lemma 4. It follows from Statement 2 of Lemma 3 that $H(\beta+)=H(\beta)$ under the assumption of Statement 2 in Lemma 3. We shall now prove $\lim_{n\to\infty} H(\beta-)/n > 0$ a.s.. By the assumption, there is just one value of (C,X,ϵ) such that $C-X=\epsilon$ and $X=x_1$. WLOG, assume that $c_i-x_1=e_i$, where $(C,X,\epsilon)=(c_i,x_1,e_i)$. That is, there is censoring at $t_{2i1}(b)$ and $\mathrm{P}\{W=e_i, X>x_1\}=0$. We have $n_{2i1}(b)+d_{1i1}(b)+d_{1i2}(b)+\cdots+d_{1im_x}(b) = r_i - r_i^+$, where $r_i^+=\sum_{h=1}^n \mathbf{1}_{(T_h(b)>e_i)}$. Assume $b\in(b_{i_o-1},\beta)$. Then

$$r_{1i1}(b)=r_i,\ r_{1i2}(b)=r_{1i1}(b)-d_{1i1}(b)-n_{2i1}(b), \tag{A.5}$$

$r_{1i3}(b) = r_{1i2}(b) - d_{1i2}(b)$, $r_{1i4}(b) = r_{1i3}(b) - d_{1i3}(b), \ldots$. Furthermore, it can be shown (see Kong (2005)) that

$$\prod_k (1 - \frac{d_{1ik}(b)}{r_{1ik}(b)}) = (1 - \frac{d_i}{r_i})\phi_n, \tag{A.6}$$

where $\phi_n = (r_i - d_i - n_{2i1}(b))/(r_i - d_i)(r_{1i2}(b) + n_{2i1}(b))/r_{1i2}(b) < 1$ a.s. by (A.5). Let $A_1 = \{W = e_i < \epsilon \text{ and } X = x_1\}$, $A_2 = \{W \geq e_i = \epsilon \text{ and } X \geq x_2\}$ and $A_3 = \{\epsilon \wedge W > e_i\}$. Verify that they are mutually exclusive events. Then by the SLLN we have

$$\phi_n \stackrel{a.s.}{\to} \phi = \frac{P(A_3)P(A_3 \cup A_2 \cup A_1)}{P(A_3 \cup A_1)P(A_2 \cup A_3)} = \frac{P(A_3)[P(A_3) + P(A_2) + P(A_1)]}{[P(A_3) + P(A_1)][P(A_2) + P(A_3)]} < 1.$$

Moreover, we can show, by an argument similsr to (A.1) and (A.6), that

$$\prod_k (1 - \frac{d_{1jk}(b)}{r_{1jk}(b)}) = (1 - \frac{d_j}{r_j}) \text{ for } j \neq i, \text{ if } b \in (b_{i_o-1}, \beta).$$

Thus by (3.3)

$$\lim_{b \uparrow \beta} \hat{S}_b(t) = \begin{cases} \hat{S}_\beta(t) & \text{if } t < e_i, \\ \phi_n \hat{S}_\beta(t) & \text{if } t \geq e_i. \end{cases} \tag{A.7}$$

Note that $t_{1i1}(b) = c_i - bx_1 < t_{1ik}(b)$, $\forall\ k > 1$, but $t_{1i1}(\beta) = c_i - \beta x_1 = t_{1ik}(\beta)$, $\forall\ k > 1$. Furthermore, let $\hat{S}_{\beta,k}(e_i) = \lim_{b\uparrow\beta} \hat{S}_b(t_{1ik}(b))$. By (3.3),

$$\hat{S}_{\beta,1}(e_i) = \hat{S}_\beta(e_i-)(1 - \frac{n_{1i1}}{r_i}) \text{ and } \hat{S}_\beta(e_i) = \hat{S}_\beta(e_i-)(1 - \frac{d_i}{r_i}).$$

It follows from the foregoing equations and (A.7) that

$$\begin{aligned} \frac{\hat{S}_{\beta,1}(e_i) - \hat{S}_{\beta,m_x}(e_i)}{\hat{S}_{\beta,1}(e_i)} &= \frac{\hat{S}_\beta(e_i)\frac{r_i - n_{1i1}}{r_i - d_i} - \phi_n \hat{S}_\beta(e_i)}{\hat{S}_\beta(e_i)\frac{r_i - n_{1i1}}{r_i - d_i}} \\ &= 1 - \frac{\phi_n(r_i - d_i)}{r_i - n_{1i1}} (\stackrel{def}{=} 1 - \tilde{\phi}_n). \end{aligned} \tag{A.8}$$

Recall that $t_{2hm}(b) = c_h - bx_m$. Let $w_{hm} = t_{2hm}(\beta)$. Then it can be shown (see

Kong (2005)) that with probability 1,

$$\lim_{b\uparrow\beta}\frac{\sum_k \mathbf{1}_{(t=t_{1jk}(b)>t_{2hm}(b))}t\hat{f}_b(t)}{\hat{S}_b(t_{2hm}(b))}$$

$$=\begin{cases} e_i(1-\tilde{\phi}_n) & \text{if} \quad w_{hm}=e_i=e_j \text{ (by A4 and (A.8))}, \\ \frac{\tilde{\phi}_n e_j\hat{f}_\beta(e_j)}{\hat{S}_\beta(e_i)} & \text{if} \quad w_{hm}=e_i<e_j, \\ \frac{e_j\hat{f}_\beta(e_j)}{\hat{S}_\beta(w_{hm})} & \text{if} \quad e_i<w_{hm}<e_j \text{ or if } w_{hm}<e_j<e_i, \\ \frac{e_j(\hat{S}_\beta(e_j-)-\phi_n\hat{S}_\beta(e_j))}{\hat{S}_\beta(w_{hm})} & \text{if} \quad w_{hm}<e_i=e_j \text{ (by A4)}, \\ \frac{\phi_n e_j\hat{f}(e_j)}{\hat{S}_\beta(w_{hm})} & \text{if} \quad w_{hm}<e_i<e_j, \end{cases} \tag{A.9}$$

and with probability 1,

$$\lim_{n\to\infty}\Big[\frac{1}{n}H(\beta-)-\frac{1}{n}H(\beta)\Big]$$
$$=\lim_{n\to\infty}\frac{1}{n}\sum_{j=1}^{n}(1-\delta_j)\Big[\lim_{b\uparrow\beta}\frac{\sum_{t>T_j(b)}t\hat{f}_b(t)}{\hat{S}_b(T_j(b))}-\frac{\sum_{t>T_j(\beta)}t\hat{f}_\beta(t)}{\hat{S}_\beta(T_j(\beta))}\Big](X_j-\bar{X}) \quad \text{(by (2.1) and (2.2))}$$
$$=\lim_{n\to\infty}\bigg(\frac{n_{2i1}}{n}\frac{(1-\tilde{\phi}_n)\big(e_i\hat{S}_\beta(e_i)-\sum_{j>i}e_j\hat{f}_\beta(e_j)\big)}{S_\beta(e_i)}(x_1-\bar{X})$$
$$+\sum_{(h,m):w_{hm}<e_i}\frac{n_{2hm}}{n}\frac{(1-\phi_n)\big(e_i\hat{S}_\beta(e_i)-\sum_{j>i}e_j\hat{f}_\beta(e_j)}{\hat{S}_\beta(w_{hm})}(x_m-\bar{X})\bigg) \text{ (by (A.9))}$$
$$=P(W=e_i<\epsilon,X=x_1)(1-\tilde{\phi})\frac{e_iS_o(e_i)-\sum_{j>i}e_jf_o(e_j)}{S_o(e_i)}(x_1-\mu_x)$$
$$+\sum_{(h,m):w_{hm}<e_i}P(W=w_{hm}<\epsilon,X=x_m)(1-\phi)\frac{e_iS_o(e_i)-\sum_{j>i}e_jf_o(e_j)}{S_o(w_{hm})}(x_m-\mu_x)$$
$$=f_{W,X}(e_i,x_1)(1-\tilde{\phi})\big(e_iS_o(e_i)-\sum_{j>i}e_jf_o(e_j)\big)(x_1-\mu_x)$$
$$+\sum_{(h,m):w_{hm}<e_i}P(W=w_{hm},X=x_m)(1-\phi)\big(e_iS_o(e_i)-\sum_{j>i}e_jf_o(e_j)\big)(x_m-\mu_x)$$
$$=[f_{W,X}(e_i,x_1)(x_1-\mu_x)+\frac{1-\phi}{1-\tilde{\phi}}\sum_{(h,m):w_{hm}<e_i}P(W=w_{hm},X=x_m)(x_m-\mu_x)]\cdot\nu$$
$$(\nu=(1-\tilde{\phi})\big(e_iS_o(e_i)-\textstyle\sum_{j>i}e_jf_o(e_j)\big))$$

$$
\begin{aligned}
&= \Big(f_{W,X}(e_i,x_1)(x_1-\mu_x) + \frac{P(A_1)}{P(A_3\cup A_1)}\sum_{m=1}^{m_x} P(W<e_i, X=x_m)(x_m-\mu_x)\Big)\nu \\
&\qquad\qquad\qquad\qquad\qquad\qquad (\text{as } \frac{1-\phi}{1-\tilde\phi} = \frac{P(A_1)}{P(A_1)+P(A_3)}) \\
&= \Big(f_{W,X}(e_i,x_1)(x_1-\mu_x) \\
&\quad + \frac{P(A_1)}{P(A_3\cup A_1)}\sum_{m=1}^{m_x}(f_X(x_m) - P(W\ge e_i, X=x_m))(x_m-\mu_x)\Big)\nu \\
&= \Big(f_{W,X}(e_i,x_1)(x_1-\mu_x) - \frac{P(A_1)}{P(A_3\cup A_1)}\sum_{m=1}^{m_x} P(W\ge e_i, X=x_m))(x_m-\mu_x)\Big)\nu \\
&\qquad\qquad\qquad\qquad\qquad\qquad (\text{as } E(X-\mu_x)=0) \\
&= \Big(f_{W,X}(e_i,x_1)(x_1-\mu_x) - \frac{P(A_1)}{P(A_3\cup A_1)}P(W=e_i, X=x_1)(x_1-\mu_x) \\
&\quad - \frac{P(A_1)}{P(A_3\cup A_1)}\sum_{m=1}^{m_x} P(W>e_i, X=x_m))(x_m-\mu_x)\Big)\nu \\
&\qquad\qquad\qquad\qquad\qquad\qquad (\text{as } P(W=e_i, X>x_1)=0) \\
&> \Big(\frac{P(A_3)}{P(A_3\cup A_1)} f_{W,X}(e_i,x_1)(x_1-\mu_x) \\
&\quad - \frac{P(A_1)}{P(A_3\cup A_1)}\sum_{m=1}^{m_x} P(W>e_i, X=x_m))(x_1-\mu_x)\Big)\nu \\
&\qquad\qquad\qquad\qquad\qquad\qquad (as\ \nu = (1-\tilde\phi)\sum_{j>i}(e_i-e_j)f_o(e_j)<0) \\
&= \Big(\frac{P(A_3)}{P(A_3\cup A_1)} f_{W,X}(e_i,x_1)(x_1-\mu_x) - \frac{P(A_1)}{P(A_3\cup A_1)}P(W>e_i)(x_1-\mu_x)\Big)\nu \\
&= \big(P(A_3)f_{W,X}(e_i,x_1) - P(A_1)P(W>e_i)\big)\frac{x_1-\mu_x}{P(A_3\cup A_1)}\nu \\
&= 0.
\end{aligned}
$$

(the last equality holds by the definitions of A_1 and A_3, and independence of ϵ and (W, X)). It follows from the foregoing inequality that $\lim_{n\to\infty} H(\beta-)/n > \lim_{n\to\infty} H(\beta)/n = 0$ a.s (by Lemma 1). This completes the proof of Statement 2 of the lemma.

By symmetry, Statement 1 of the lemma can be proved in a similar manner.

Proof of the statement in Example 4. By the assumptions in the example, there are 2^3 possible values of (ϵ, X, C). Let n_i be the number of observations of type i, $i = 1, \ldots, 8$. Verify $\mathcal{B} = \{-1/3, 0, 2/3, 1\}$ by the definition of $\mathcal{B}$ in Section

2. In order to show $H(\beta-) \neq H(\beta+)$, since $\beta = 1$, we only need to consider b in two intervals : $(2/3, 1)$ and $(1, \infty)$.

By Lemma 2 and Remark 2.2, in order to evaluate $H(\beta-)/n$, consider $b \in (2/3, 1)$ and compute $T_i^* = T_i^*(1-)$ by (2.2). The quantities in (2.1) and (2.2) are computed in the following table, arranged in ascending orders of $T_i(1-)$s. This arrangement makes the evaluation of $\hat{S}_b$, $\hat{f}_b$, and $T_i^*(b)$ easier (see (3.3) and (2.2)).

i	ϵ	X_i	Y_i	C_i	M_i	δ_i	$T_i(b)$	$T_i(1)$	$T_i(1-)$	$\hat{f}_{1-}(T_i(1-)) \approx$	$T_i^*(1-) \approx$
1	0	1	1	-2	-2	0	$-2-b$	-3	-3	0	4/9
2	1	1	2	-2	-2	0	$-2-b$	-3	-3	0	4/9
3	0	-2	-2	-2	-2	1	$-2+2b$	0	$0-$	1/3	0
4	0	-2	-2	1	-2	1	$-2+2b$	0	$0-$	1/3	0
5	1	-2	-1	-2	-2	0	$-2+2b$	0	$0-$	0	2/3
6	0	1	1	1	1	1	$1-b$	0	$0+$	2/9	0
7	1	1	2	1	1	0	$1-b$	0	$0+$	0	1
8	1	-2	-1	1	-1	1	$-1+2b$	1	1	4/9	1

By (2.1), (2.2) and the foregoing table,

$$\begin{aligned} \frac{H(\beta-)}{n} &= \frac{1}{n}\{(n_1T_1^* + n_2T_2^*)(1-\bar{X}) + n_5T_5^*(-2-\bar{X}) \\ &\quad + n_7T_7^*(1-\bar{X}) + n_8T_8^*(-2-\bar{X})\} \\ &\stackrel{a.s.}{\to} \frac{1.5}{8}\{\frac{4}{9} + \frac{4}{9} - \frac{2}{3} + 1 - 1\} \qquad (\text{as } \bar{X} \stackrel{a.s.}{\to} -0.5) \\ &= \frac{1}{24}. \end{aligned}$$

In order to evaluate $H(\beta+)$, consider $b \in (1, \infty)$ and compute $T_i^* = T_i^*(1+)$ by (2.2). The quantities in (2.1) and (2.2) are computed in the following table. In the table we rearrange the orders of types of observations in ascending orders of $T_i^*(1+)s$.

i	ϵ	X_i	Y_i	C_i	M_i	δ_i	$T_i(b)$	$T_i(1)$	$T_i(1+)$	$\hat{f}_{1+}(T_i(1+)) \approx$	$T_i^*(1+) \approx$
1	0	1	1	-2	-2	0	$-2-b$	-3	-3	0	5/12
2	1	1	2	-2	-2	0	$-2-b$	-3	-3	0	5/12
6	0	1	1	1	1	1	$1-b$	0	$0-$	1/6	0
7	1	1	2	1	1	0	$1-b$	0	$0-$	0	1/2
3	0	-2	-2	-2	-2	1	$-2+2b$	0	$0+$	5/12	0
4	0	-2	-2	1	-2	1	$-2+2b$	0	$0+$	5/12	0
5	1	-2	-1	-2	-2	0	$-2+2b$	0	$0+$	0	1
8	1	-2	-1	1	-1	1	$-1+2b$	1	1	5/12	1

It follows from (2.1), (2.2) and the foregoing table that

$$\frac{H(\beta+)}{n} = \frac{1}{n}\{(n_1T_1^* + n_2T_2^* + n_7T_7^*)(1-\bar{X}) + n_5T_5^*(2+\bar{X}) + n_8T_8^*(2+\bar{X})\} \stackrel{a.s.}{\to} -\frac{1}{8}.$$

Thus A5 holds and w.p.1, a BJE $\hat{\beta}_n = \beta$ if n is large enough.

Acknowledgement

Mr. Yishi Wang first found the partial answer to Example 3 in his solution to a final examination problem given by the second author. This finding led to our paper. The authors thank the Editor, an associate editor and a referee for their helpful comments.

References

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

Cox, D. R. (1972). Regression models and life tables. *J. Roy. Statist. Soc. Ser. B* **34**, 187-220.

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

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

Kong, F. H. (2005). Asymptotic distributions of Buckley-James estimator. Ph.D. thesis, Binghamton University.

Koul, H., Susarla, V. and Van Ryzin, J. (1981). Regression analysis with randomly right-censored data. *Ann. Statist.* **9**, 1276-1288.

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

Li, G. and Zhang, C. H. (1998). Linear regression with interval censored data. *Ann. Statist.* **26**, 1306-1327.

Li, L. X. and Pu, Z. W. (1999). Regression models with arbitrarily interval-censored observations. *Comm. Statist. Theory Methods* **28**, 1547-1563.

Miller, R. G. (1976). Least squares regression with censored data. *Biometrika* **63**, 449-464.

Miller, R. G. (1981). *Survival Analysis*. Wiley, New York.

Miller, R. G. and Halpern, J (1982). Regression with censored data. *Biometrika* **69**, 521-531.

Nelson, W. (1973). Analysis of residuals from censored data. *Technometrics* **15**, 697-715.

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

Yu, Q. Q. and Wong, G. Y. C. (2002). How to find all Buckley-James estimates instead of just one? *J. Statist. Comput. Simulation* **72**, 451-460.

Zhang, C. H. and Li, X. (1996). Linear regression with doubly censored data. *Ann. Statist.* **24**, 2720-2743.

Department of Mathematics and Computer Sciences, Wilkes University, Wilkes-Barre, PA 18766, U.S.A.

E-mail: fanhui.kong@wilkes.edu

Department of Mathematical Sciences, SUNY, Binghamton, NY 13902, U.S.A.

E-mail: qyu@math.binghamton.edu

(Received May 2005; accepted August 2005)