Technical Report on
"A Modification to the Buckley-James Estimate"

Zifan Huang[a] and Qiqing Yu[a]

[a] Department of Mathematics and Statistics, Binghamton University, NY 13902, USA

ABSTRACT
The Buckley-James estimator (BJE) is an estimator of β for the semi-parametric linear regression model $Y = \beta'\mathbf{X} + W$ with right-censored data. Several iterative algorithms for the BJE have been proposed so far. However, they may either converge to a value far away from the BJE, or fail to converge at all. On the other hand, Yu and Wong (2002) introduced a non-iterative algorithm for finding all solutions of the BJE. While theoretically appealing, this approach becomes computationally intensive if $\beta \in \mathcal{R}^p$ with $p > 1$ and with a large sample size n. This paper presents a modification to the BJE with a non-iterative algorithm. It yields the exact BJE if $\beta \in \mathcal{R}$; otherwise it consistently approximates both $\hat{\beta}$ and the true parameter β as n increases. We compare its performance against the BJE through simulation studies and illustrative examples. We also carry out a data analysis of a real-world data set.

KEYWORDS
Linear regression; survival analysis; right censorship model; semiparametric model; Buckley-James Estimate

1. Introduction.

This paper studies the problem of parameter estimation under the semiparametric multiple linear regression model with right-censored (RC) data.
Regression analysis constitutes a central tool in statistical inference, with broad applicability in disciplines such as economics, engineering, biomedical research, and the social sciences. Specifically, we consider the following semiparametric multiple linear regression (LR) model.

(AS) $Y = \beta'\mathbf{X} + W$, where Y and W are random variables, $\mathbf{X}$ is a p-dimensional random vector, $\beta \in \mathcal{R}^p$, the p-dimensional Euclidean space, Y is subject to right censoring with censoring variable C, $P(Y \le C) \in (0, 1]$, and C, $\mathbf{X}$ and W are independent. Both β and the survival function $S_W(t)$ $(= P(W > t))$ are unknown. The observations $(\mathbf{X}_1, M_1, \delta_1)$, ..., $(\mathbf{X}_n, M_n, \delta_n)$ are i.i.d. from $(\mathbf{X}, M, \delta)$, where $\delta = \mathbf{1}(Y \le C)$ (the indicator function) and $M = Y \wedge C$.

The LR model is commonly expressed as $Y = \alpha + \beta'\mathbf{X} + \epsilon$, where $E(\epsilon) = 0$ and $\mathbf{X} \perp \epsilon$. However, $E(Y)$ may not exist such as in the case that W follows a Cauchy distribution. Even if α $(=E(W))$ exists, it is often non-identifiable under censoring, such as in the case that $P(C < \tau) = 1$ and $S_W(\tau) > 0$ for some $\tau \in \mathcal{R}$. Moreover,

CONTACT Zifan Huang. Email: zhuang91@binghamton.edu

researchers are typically more interested in the effect β of $\mathbf{X}$ on Y. Therefore, (AS) is more appropriate, general, and practical.

The least squares estimator (LSE) remains the predominant method for analyzing complete data. In the presence of right censoring, several extensions of the LSE have been proposed. Among the early contributions, Miller (1976) introduced an adaptation of the least squares principle, while Buckley and James (1979) proposed the Buckley-James estimator (BJE) of β as a solution to a modified normal equation of the sum of squares represented by $H(\mathbf{b})$ (see Eq. (2.1) in Section 2 (§2)). Other modifications to the LSE methodology were proposed by Chatterjee and McLeish (1986), and Leurgans (1987). Notably, Hillis (1993) conducted a comparative study and concluded that the BJE performs favorably compared to alternative estimators in the regression setting.

Buckley and James also proposed an iterative algorithm for computing the BJE. They pointed out the potential non-uniqueness and non-existence of the BJE as the root of the function $H(\mathbf{b})$. In subsequent work, James and Smith (1984) proposed refinements to the BJE, one of which redefined the estimator as a zero-crossing point (ZC) of $H(\mathbf{b})$. This ZC-based definition has since been widely adopted in the literature (see *e.g.*, Lai and Ying (1991)). The asymptotic properties of the ZC-based BJE have been rigorously established under various regularity conditions (see, *e.g.*, Lai and Ying (1991) and Wang *et al.* (2010)). However, the recent work by Yu (2023) demonstrated that the ZC-based BJE can be inconsistent, even when a consistent estimator exists in the form of a strict zero-crossing point (SZC) (see Definition 3 in §2). This observation leads to the refined definition of the BJE as the SZC of $H(\mathbf{b})$. Yu and Huang (2024) further established the existence of the BJE as a ZC even under degenerate settings such as $\mathbf{X} \equiv$ a constant and proved that an SZC always exists whenever β is identifiable.

The original iterative algorithm developed by Buckley and James has been adapted for the ZC and is integrated into the R function bj() within the current R package rms (Stare, *et al.* (2001)). However, this iterative approach may yield no solution at all. Jin, Lin, and Ying (2006) as well as Wang, Zhao, and Fu (2016) attempted to modify the iterative algorithm to reduce the instability of the iteration. Nevertheless, these efforts have not completely overcome the drawback that the procedure may oscillate indefinitely between multiple values for some data sets (see Table 5.1 in §5). Yu and Wong (2002) proposed a non-iterative algorithm capable of identifying all possible ZCs (see §2). While this approach guarantees identification of all ZCs, its computational burden grows rapidly with the dimension p of β ($\in \mathcal{R}^p$) and with the sample size, rendering it inconvenient in higher dimensions.

We propose a modified version of the BJE along with a non-iterative and computationally efficient algorithm. Our method is built upon the non-iterative approach of Yu and Wong (2002), with key modifications that enhance scalability. Specifically, the modification is exactly the same as the ZC in the case $p = 1$, or in some cases (see Examples 1 and 2 in §2). In general, simulation studies suggest that the proposed method is a consistent approximation to the BJE and β.

Section 2 introduces the proposed algorithm and provides illustrative examples. Section 3 presents simulation results to demonstrate the effectiveness of the new algorithm and the approximation to the BJE $\hat{\beta}$ and β. Section 4 illustrates the practical utility of our estimator via a real data analysis. Section 5 provides a data example in which the three existing iterative algorithms fail to produce a solution, highlighting the advantage of our approach.

2. Preliminary results.
We first introduce the BJE of $\beta \in \mathcal{R}^p$. Let $\mathbf{b} \in \mathcal{R}^p$, $T_i(\mathbf{b}) = M_i - \mathbf{b}'\mathbf{X}_i$, $(T_i^*(\mathbf{b}), \delta_i^*(\mathbf{b}))$

$$= \begin{cases} (T_i(\mathbf{b}), 1) & \text{if } \delta_i = 1 \text{ or } T_i(\mathbf{b}) = T_{(n)}(\mathbf{b}) \ (= \max_j T_j(\mathbf{b})), \\ (\frac{\sum_{t > T_i(\mathbf{b})} t \hat{f}_{\mathbf{b}}(t)}{\hat{S}_{\mathbf{b}}(T_i(\mathbf{b}))}, \delta_i) & \text{otherwise,} \end{cases} \tag{2.1}$$

and $\hat{S}_{\mathbf{b}}$ is the product-limit-estimator (PLE) of $F_{Y-\mathbf{b}'\mathbf{X}}$ based on $(T_i(\mathbf{b}), \delta_i^*(\mathbf{b}))$'s.

$$\text{Then } H(\mathbf{b}) = (H_1(\mathbf{b}), ..., H_p(\mathbf{b}))' = \sum_{i=1}^{n} (\mathbf{X}_i - \overline{\mathbf{X}}) T_i^*(\mathbf{b}). \tag{2.2}$$

If the largest $T_i(\mathbf{b})$ is right censored, $\frac{\sum_{t > T_i(\mathbf{b})} t \hat{f}_{\mathbf{b}}(t)}{\hat{S}_{\mathbf{b}}(T_i(\mathbf{b}))}$ in (2.1) is not defined and it is treated as an exact observation in $H(\mathbf{b})$. Abusing notations, write $\delta_i^* = \delta_i^*(\mathbf{b})$. The BJE was originally defined as a root of $H(\mathbf{b})$, but the root may not exist, thus the BJE $\hat{\beta}$ is defined as a ZC of $H(\mathbf{b})$ (see Lai and Ying (1991) p.1371) until recently. Notice that $H(\mathbf{b})$ is piecewise linear in $\mathbf{b}$ (see Remark 3.1 in Yu and Wong (2002)).

Definition 1. Let $p = 1$. A point b is called a ZC of $H(\cdot)$ if $H(b-)H(b+) \leq 0$ or $H(b) = 0$.

So a root of $H(\mathbf{b})$ is also a ZC. Notice that if $\delta_i \equiv 1$ and X_i's are not identical, then the ZC and BJE is the LSE. To emphasize the distinction from the complete-data setting, we henceforth assume that the data are subject to right-censoring, i.e., $\min_i \delta_i = 0$ (unless being specified otherwise). When the dimension $p \geq 2$, Yu and Huang (2024) highlighted a fundamental limitation of the naive extension of the zero-crossing (ZC) definition based on Definition 1, and proposed a refined formulation to address this issue. The modified definition, formalized in Definition 2, ensures better theoretical and computational properties of the estimator in higher-dimensional settings.

Definition 2. Let $\beta \in \mathcal{R}^p$. $\hat{\beta}$ is a ZC of $H(\mathbf{b})$ if $\forall$ $(i,j) \in \{1,...,p\} \times \{1,2\}$, $\exists$ $\mathbf{b}_{ikj} \in \mathcal{R}^p$ such that (1) $\lim_{j\to\infty} H_k(\mathbf{b}_{i2j}) \cdot \lim_{j\to\infty} H_k(\mathbf{b}_{i1j}) \leq 0$ or $H(\mathbf{b}) = \mathbf{0}$, and (2) $\lim_{j\to\infty} \mathbf{b}_{ikj} \to \hat{\beta}$, $k \in \{1,2\}$.

In the same paper, they established that the BJE as a ZC always exists, however, it may happen that each point in an interval is a BJE. To address this issue, they redefined the BJE as the strict zero-crossing (SZC) (see Definition 3 below). The SZC criterion excludes such degenerate solutions and provides a more meaningful and robust characterization of the BJE. Furthermore, they demonstrated that an SZC always exists as long as the parameter β is identifiable.

Definition 3. If $\beta \in \mathcal{R}^p$, a point $\hat{\beta}$ is a strict ZC (SZC) if $\exists$ 2 sequences $\{b_i\}_{i\geq 1}$ and $\{b_j^o\}_{j\geq 1}$ such that they converge to $\hat{\beta}$ and $\lim_{b_i\to b} sign(H(b_i)) \neq \lim_{b_j^o\to b} sign(H(b_j^o))$, where $sign(0) \stackrel{def}{=} 0$.

Based on earlier definition of the BJE as a ZC, Yu and Wong (2002) developed a comprehensive algorithm for deriving all ZCs in finite many steps. The algorithm is described as follows:

Algorithm for the ZC with $p = 1$ (by Yu and Wong (2002)):

(1) Compute the slope $b_{ij} = \frac{M_i - M_j}{X_i - X_j}$ for each pair (i, j) with $X_i < X_j$. Let $q_1 < q_2 < \cdots < q_{m_b}$ be all the distinct b_{ij}'s and let $(q_0, q_{m_b+1}) = (-\infty, \infty)$. For each $h = 0, 1, ..., m_b$, first 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 $b = q_{m_b} + 1$ if $h = m_b$. With the given b, denoted by a_h, compute $H(a_h)$. If $H(a_h) = 0$ then a_h is a BJE, otherwise let $\hat{b}_h = \frac{\sum_{j=1}^n (X_j - \overline{X}) M_j^*(a_h)}{\sum_{k=1}^n (X_k - \overline{X}) X_k^*(a_h)}$ (if $\hat{b}_h$ exists) and check whether $\hat{b}_h \in (q_h, q_{h+1})$, where

$$M_i^*(b) = M_i\delta_i + (1 - \delta_i) \sum_{t > T_i(b)} \frac{\hat{f}_b(t)}{\hat{S}_b(T_i(b))} \frac{\sum_{j=1}^n M_j \mathbf{1}_{(T_j(b)=t, \delta_j^*=1)}}{\sum_{k=1}^n \mathbf{1}_{(T_k(b)=t, \delta_k^*=1)}} \text{ and} \tag{2.3}$$

$$X_i^*(b) = X_i\delta_i + (1 - \delta_i) \sum_{t > T_i(b)} \frac{\hat{f}_b(t)}{\hat{S}_b(T_i(b))} \frac{\sum_{j=1}^n X_j \mathbf{1}_{(T_j(b)=t, \delta_j^*=1)}}{\sum_{k=1}^n \mathbf{1}_{(T_k(b)=t, \delta_k^*=1)}}, \quad i = 1, ..., n. \tag{2.4}$$

If so, then $\hat{b}_h$ is a solution to equation $H(b) = 0$ and thus a ZC of H.

(2) Compute $H(q_i-)$, $H(q_i)$ and $H(q_i+)$, $i = 1, ..., m_b$, where $H(\hat{b}+)$ and $H(\hat{b}-)$ are the right- and left-hand limits of H, respectively. By (2.1) and (2.3) we have

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

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

(3) If $H(q_i+) = H(q_{i+1}-) = 0$, then each $b \in (q_i, q_{i+1})$ is also a ZC.

Algorithm with $p \geq 2$:

(1) First find all solutions to the system of p equations $M_{i_k} - \mathbf{b}'\mathbf{X}_{i_k} = M_{j_k} - \mathbf{b}'\mathbf{X}_{j_k}$,

$$i.e., \ \mathbf{b} = \mathcal{Q}(M_{i_1} - M_{j_1}, ..., M_{i_p} - M_{j_p})' \ \ (\in \mathcal{R}^p), \tag{2.5}$$

if $\mathcal{Q} = ((\mathbf{X}_{i_1} - \mathbf{X}_{j_1}, \cdots, \mathbf{X}_{i_p} - \mathbf{X}_{j_p})')^{-1}$ exists. It becomes $b = \frac{M_i - M_j}{X_i - X_j}$ if $p = 1$. Let $\mathbf{b}_1$, ..., $\mathbf{b}_{m_o}$ be all distinct solutions of Eq. (2.5). As vertices, these $\mathbf{b}_i$'s can form finitely many disjoint convex sets, say Q_h's, each with non-empty open interior in $\mathcal{R}^p$, just like the roles of q_1, ..., q_{m_b} in Item (1) of the algorithm for the ZC with $p = 1$). Let $\mathbf{c}$ be an interior point of Q_h. Note that now, given j, $\hat{S}_{\mathbf{b}}(T_j)$ does not change for each $\mathbf{b}$ in the subset Q_h $\forall$ j. Check whether $H(\mathbf{c}) = \mathbf{0}$. If so, $\mathbf{c}$ is a ZC, otherwise, solve $\sum_{j=1}^n (M_j^*(\mathbf{c}) - \mathbf{b}'\mathbf{X}_j^*(\mathbf{c}))(\mathbf{X}_j - \overline{\mathbf{X}}) = \mathbf{0}$ for $\mathbf{b}$, *i.e.*,

$$\mathbf{b} = ((\sum_{i=1}^n (\mathbf{X}_i - \overline{\mathbf{X}})(\mathbf{X}_i^*(\mathbf{c}))')^{-1} \sum_{j=1}^n M_j^*(\mathbf{c})(\mathbf{X}_j - \overline{\mathbf{X}}), \tag{2.6}$$

if $\mathcal{Q}$ exists, where $M_j^*(\mathbf{c})$ is the same as in Eq. (2.3), but $\mathbf{X}_i(\mathbf{c})$ in Eq. (2.4) is replaced by $\mathbf{X}_i^* = (\mathbf{X}_{i1}^*, ..., \mathbf{X}_{ip}^*)'$ and $\mathbf{X}_{ik}^*(\mathbf{c}) = \mathbf{X}_{ik}\delta_i + (1-\delta_i)\sum_{t>T_i(\mathbf{c})} \frac{\hat{f}_b(t)}{\hat{S}_b(T_i(b))} \frac{\sum_{j=1}^n \mathbf{X}_{jk}\mathbf{1}_{(T_j(b)=t,\delta_j^*=1)}}{\sum_{k=1}^n \mathbf{1}_{(T_k(b)=t,\delta_k^*=1)}}$, $k = 1, ...,$ p and $i = 1, ..., n$, as $\mathbf{X}_i \in \mathcal{R}^p$, $p \geq 2$. If $\mathbf{b}$ given by (2.6) exists (*i.e.*, $((\sum_{i=1}^n (\mathbf{X}_i - \overline{\mathbf{X}})(\mathbf{X}_i^*(\mathbf{c}))')^{-1}$ exists) and satisfies $\mathbf{b} \in Q_h$, then $\mathbf{b}$ is a ZC.

(2) It is similar to the case $p = 1$ and is skipped.

(3) If $H(\mathbf{b}) = 0$ for two distinct $\mathbf{b} \in Q_i$, then each $\mathbf{b} \in Q_i$ is a ZC.

While the algorithm performs well in low-dimensional settings, its computational cost increases substantially as the dimension p increases, as seen from Example 1.

Example 1 (Example 2 (Yu (2023))). There are 4 $(M_i, \delta_i, X_{i1}, X_{i2})$: $(1,0,1,0)$, $(1,0,0,1)$, $(0.6,1,0,0)$ and $(0.1,1,0,0)$, generated from the model $Y = X_1 + X_2 + W$, $C = 1$ and $W \sim U(0,1)$ (the uniform distribution on $(0,1)$). Write $H(\mathbf{b}) = (H_1(\mathbf{b}), H_2(\mathbf{b}))'$, where $\mathbf{b} = (b_1, b_2)'$. Then

$$H(\mathbf{b}) = \begin{cases} (0,0)' & \text{if } \mathbf{b} \in \{(x,y):\ x \wedge y > 0.9\}, \\ (+,+)' & \text{if } \mathbf{b} \in \{(x,y):\ x \vee y \leq 0.9\}, \\ (+,0)' & \text{if } \mathbf{b} \in \{(x,y):\ x < 0.9 < y\}, \\ (0,+)' & \text{if } \mathbf{b} \in \{(x,y):\ x > 0.9 > y\}, \end{cases}$$

where "+" stands for "> 0". Thus each $\mathbf{b} \in \{(x,y)' :\ x \wedge y \geq 0.9\}$ is a ZC. In particular, $\mathbf{b} = (0.9, 0.9)'$ is a ZC and is not a root of $H(\mathbf{b})$, whereas $\beta = (1,1)'$ is the true parameter value. This can be derived by Algorithm of Yu and Wong (2002) as follows. The possible solutions to $M_i - M_k = b_1(X_{i1} - X_{k1}) + b_2(X_{i2} - X_{k2})$ are

(i,k) :	$(1,2)$	$(1,3)$	$(1,4)$	$(2,3)$	$(2,4)$	$(3,4)$
(b_1,b_2) :	$b_1 = b_2$	$b_1 = 0.4$	$b_1 = 0.9$	$b_2 = 0.4$	$b_2 = 0.9$	$\emptyset$

Then the possible solutions to $\begin{cases} M_i - M_k = b_1(X_{i1} - X_{k1}) + b_2(X_{i2} - X_{k2}) \\ M_h - M_j = b_1(X_{h1} - X_{j1}) + b_2(X_{h2} - X_{j2}) \end{cases}$ are $(b_1, b_2) \in \{(0.9,0.4), (0.4,0.9), (0.9,0.9), (0.4,0.4)\}$. Denote these 4 points by $\mathbf{b}_1, \ldots, \mathbf{b}_4$. These 4 points yield 9 open rectangles as shown in Figure 1 below, denoted by $Q1, \ldots, Q9$ (see Figure 1), as well as their open boundary line segments. Then $(\mathbf{X}_i - \overline{\mathbf{X}}) = \begin{pmatrix} (1,0,0,0) - 0.25 \\ (0,1,0,0) - 0.25 \end{pmatrix}$, $T_i = M_i - b_1 X_{i1} - b_2 X_{i2}$ and T_i^* is as in Eq. (2.2). $H(\mathbf{b})$ is linear in each of these 49 disjoint sets by Yu and Wong (2002). Thus check whether $H(\mathbf{b})$ has a ZC in $Q1, \ldots, Q9$ one by one, as well as at $\mathbf{b}_1, \ldots, \mathbf{b}_4$ and on those boundary line segments. For instance, check $H(\mathbf{b})$ on Q7 as follows.

Q7$= \{(x,y) :\ x \vee y < 0.4\}$: Let $\mathbf{c} = -(1,1) \in Q7$, then $T(\mathbf{b}) = (2,2,0.1,0.6)$, $T^*(\mathbf{b}) = T(\mathbf{b})$, $X^* = X$ and $M^* = M$, where $\mathbf{b}$, M_i^* and X_i^* are as defined in Eq. (2.6), (2.3) and (2.4). And $M_j^*(\mathbf{c})$ is the same as in Eq. (2.3), but $\mathbf{X}_i(\mathbf{c})$ in Eq. (2.4) is replaced by $\mathbf{X}_i^* = (\mathbf{X}_{i1}^*, \ldots, \mathbf{X}_{ip}^*)'$ and $\mathbf{X}_{ik}^*(\mathbf{c}) = \mathbf{X}_{ik}\delta_i + (1-\delta_i)\sum_{t>T_i(\mathbf{c})} \frac{\hat{f}_b(t)}{\hat{S}_b(T_i(b))} \frac{\sum_{j=1}^n \mathbf{X}_{jk}\mathbf{1}_{(T_j(b)=t,\ \delta_j^*=1)}}{\sum_{k=1}^n \mathbf{1}_{(T_k(b)=t,\ \delta_k^*=1)}}$, $k = 1, \ldots, p,\ i = 1, \ldots, n$, as $\mathbf{X}_i \in \mathcal{R}^p$, $p \geq 2$. It is further shown (see Yu (2023)) that Eq. (2.6) yields

$$\mathbf{b} = \begin{pmatrix} 1-\frac{1}{4} & -\frac{1}{4} \\ -\frac{1}{4} & 1-\frac{1}{4} \end{pmatrix}^{-1} \begin{pmatrix} 1-\frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} \\ -\frac{1}{4} & 1-\frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 0.1 \\ 0.6 \end{pmatrix} = \tfrac{1.3}{2}(1,1)' \notin Q7,$$

and it implies that there is no BJE in Q7.

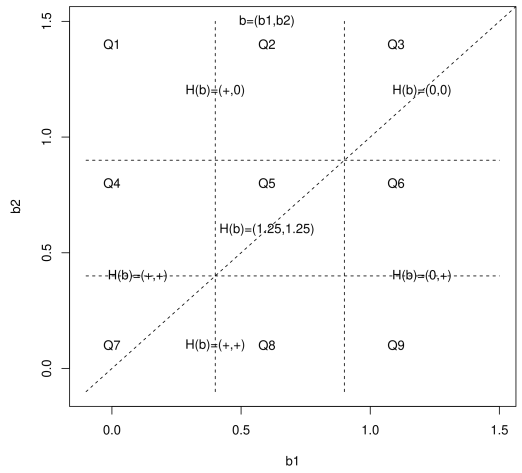

Figure 1. $H(b)$ in Various Rectangles

We refer the proof of the other regions to Example 2 in Yu (2023). □

Remark 1. The algorithm proposed by Yu and Wong (2002) yields all possible BJEs for a given set, but it is time-consuming. Is it possible to modify the algorithm so that it is faster ? A possible new algorithm for the BJE of $\beta \in \mathcal{R}^p$, $p \geq 1$ is as follows. Suppose that $(\mathbf{X}_1, ..., \mathbf{X}_n)$ are orthogonal, that is, they can be treated as p $n-$dimensional orthogonal vectors, say $\mathbf{F}_i = (\mathbf{X}_{1i}, \ldots, \mathbf{X}_{ni})'$, $i = 1, ..., p$. Then apply the algorithm of Yu and Wong (2002) to solve the BJE $\hat{\beta}_i$ with $(\mathbf{M}, \vec{\delta}, \mathbf{F}_i)$, where $\mathbf{M} = (M_1, \ldots, M_n)'$, $\vec{\delta} = (\delta_1, \ldots, \delta_n)$ and $\mathbf{F}_i$, $i = 1, ..., p$. Then hopefully, the BJE of β is $\hat{\beta} = (\hat{\beta}_1, \ldots, \hat{\beta}_p)'$.

Example 1 (continued). $\mathbf{B}' = (\mathbf{X}_1, \ldots, \mathbf{X}_4) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}$ and $\mathbf{B}'\mathbf{B} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$.
Using the approach in Remark 1, first solve $\hat{\beta}_1$ with $(M_i, \delta_i, \mathbf{X}_{i1})$'s. It is shown in Huang and Yu (2025) (see §A.1) that each $b \in [0.9, \infty)$ is a ZC. Since $(M_1, M_2) = (1, 1)$ and $(\delta_1, \delta_2) = (0, 0)$, by symmetry, each $b \geq 0.9$ is a ZC of β_2. Thus each point in $\{(x, y) : \ x \wedge y \geq 0.9\}$ is a BJE of $\hat{\beta}$. This derivation is much simpler than the derivation given in Example 1 using the original algorithm of Yu and Wong (2002). In that approach, one needs to check whether $\exists$ ZC of $H(\mathbf{b})$ in the following 49 sets (instead of 7 sets twice in the new algorithm):
(1) 9 open rectangle sets $Q1, \ldots, Q9$;
(2) 24 open line segments of the boundaries for $Q1, \ldots, Q9$;
(3) 16 singletons of the endpoints of the above 24 line segments.
But the new approach only checks 2 points and 3 intervals twice instead. In general, the time saving is from m^p to $p \times m$. Notice that $(\mathbf{X}_1, ..., \mathbf{X}_4)$ can be viewed as two orthogonal row vectors.

In view of Example 1 and Remark 1, we propose a modification of the definition of the BJE together with a more efficient algorithm that remains computationally feasible in higher dimensions and is capable of consistently approximating the true parameter value β when the sample size n is sufficiently large.

A modification of the BJE: Let $\mathbf{B}' \stackrel{def}{=} (\mathbf{X}_1, ..., \mathbf{X}_n)_{p\times n}$ with its transpose $(\mathbf{B}_1, ..., \mathbf{B}_p) \stackrel{def}{=} \mathbf{B}$. The full singular value decomposition (SVD) of $\mathbf{B}$ is $\mathbf{B} = \mathbf{U}_{n\times n} D_{n\times p} \mathbf{V}'_{p\times p}$. Let $\tilde{\mathbf{B}} \stackrel{def}{=} \mathbf{U}D \stackrel{def}{=} (\tilde{\mathbf{X}}_1, ..., \tilde{\mathbf{X}}_n)' \stackrel{def}{=} (\tilde{\mathbf{B}}_1, ..., \tilde{\mathbf{B}}_p)$, where $\tilde{\mathbf{X}}_i = (\tilde{\mathbf{X}}_{i1}, ..., \tilde{\mathbf{X}}_{ip})'$. Let $\hat{\beta}$ be the BJE based on $(\mathbf{M}, \vec{\delta}, \mathbf{B})$, where $\mathbf{M} = (M_1, ..., M_n)$ and $\vec{\delta} = (\delta_1, ..., \delta_n)$. Let $\tilde{\beta} = \mathbf{V}\hat{\gamma}$, where $\hat{\gamma}' = (\hat{\gamma}_1, ..., \hat{\gamma}_p)$, $\hat{\gamma}_j$ is the BJE based on $(\mathbf{M}, \vec{\delta}, \tilde{\mathbf{B}}_j)$ ($\tilde{\mathbf{B}}_j = (\tilde{\mathbf{X}}_{1j}, \cdots, \tilde{\mathbf{X}}_{nj})'$), $j = 1$, ..., p. We call $\tilde{\beta}$ the modified BJE.

Remark 2. We expect that $\hat{\beta} \approx \tilde{\beta}$ in the sense that (1) $\lim_{n\to\infty} \hat{\beta} = \lim_{n\to\infty} \tilde{\beta}$ and (2) $Cov(\hat{\beta})(Cov(\tilde{\beta}))^{-1} \to I$, the identity matrix, provided that (AS) holds and β is identifiable.

The modification is actually motivated by the property of the LSE, as proved in Proposition 1.

Proposition 1. Suppose that $p = 1$ or $\delta_i = 1$ for all i. Then $\tilde{\beta} = \hat{\beta}$.

The proof for $p = 1$ is trivially true and the proof for the 2nd case is in Huang and Yu (2025) (see §A.2).

Hereafter, let $\hat{\beta}$ be the BJE and $\tilde{\beta}$ be the estimator based on the new method. Even though $\tilde{\beta}$ does not always coincide with the BJE $\hat{\beta}$, it is expected that they are close if the sample size is large, such as the in next example.

Example 2. Suppose that $P(\mathbf{X} = (0, 0)') = P(\mathbf{X} = (0.5, 0)') = P(\mathbf{X} = (0.5, 0.5)') = 1/3$, $W \sim U(0, 1)$, $C \equiv c \in [0, 1]$, $\mathbf{X} \perp W$ and $\beta = (1, 1)'$. Let $M_{(i)}$'s be the order statistics of M_i's, $M_{(m)}$ the largest value for M_i's that less than c, and $\hat{\beta} = (\hat{\beta}_1, \hat{\beta}_2) = (2c - 2M_{(1)}, 2M_{(m)} - 2M_{(1)})$. Then

(1) If $c \in (0.5, 1]$, then $\exists!$ BJE of $\mathbf{b}$ near $(1, 1)$.
(2) If $c = 0.5$, then $\hat{\beta}$ is the unique SZC of $H(\mathbf{b})$ and $H(\mathbf{b}) = \mathbf{0}$ if $\mathbf{b} \in (\hat{\beta}_1, \infty) \times (\hat{\beta}_2, \infty)$.

The proof of these two statements are given in Huang and Yu (2025) (see §A.3).

Theorem 1. Under the assumptions of Example 2 with $c = 0.5$,
(1) $\hat{\beta} = (\hat{\beta}_1, \hat{\beta}_2) = (2c - 2M_{(1)}, 2M_{(m)} - 2M_{(1)}) \to (1, 1)$ a.s..
(2) $\tilde{\beta} = \mathbf{V}\hat{\gamma}$, where $\hat{\gamma}_1 = (M_{(1)} - 0.5)/0.43$ and $\hat{\gamma}_2 = (0.5 - M_{(1)})/0.16$ (if n is large).

The proof of Theorem 1 is given in Huang and Yu (2025) (see §A.4). When n is large enough, $M_1 \to 0$, $\hat{\gamma} \approx (-1.16, 3.13)'$, $\tilde{\beta} = \mathbf{V}\hat{\gamma} \approx (1, 1)$. Moreover, the comparison of the BJE to our modified estimator for this special case through simulation studies is given in §3.1.b.

Remark 3. (The estimation of the standard deviations (SD) of the BJE under the assumption of Example 2). Since $W \sim U(0, 1)$, $\hat{\beta}_1 = 2c - 2M_{(1)}$, then $\sigma^2_{\hat{\beta}_1} = 4\sigma^2_{M_{(1)}} =$

$4\frac{n}{(n+1)^2(n+2)} \approx \frac{4}{n^2}$, and $\sigma_{\hat\beta_1} \approx \frac{2}{n}$. $\sigma^2_{\hat\beta_2} = 4(\sigma^2_{M_{(m)}} + \sigma^2_{M_{(1)}} - 2Cov(M_{(m)}, M_{(1)})) = 4[\frac{m(n-m+1)}{(n+1)^2(n+2)} + \frac{n}{(n+1)^2(n+2)} - \frac{2(n-m+1)}{(n+1)^2(n+2)}]$. The proof is given in Huang and Yu (2025) (see §A.5).

Remark 4. (The estimation of the SD of $\tilde\beta$ under the assumption of Example 2). $\sigma_{\tilde\beta_1} = \sigma_{\tilde\beta_2} \approx \frac{2}{n}$. The proof is given in Huang and Yu (2025) (see §A.6).

3. Simulation results. We shall present results of two simulation studies in this section. In §3.1, $W \sim U(0,1)$ and in §3.2, $W \sim N(\mu, \sigma^2)$. Each result is based on $M = 1000$ replications.

3.1 Simulation Results under the assumption in Example 2. Let $W \sim U(0,1)$ and $\beta = (1,1)'$. Let $\hat\Sigma$ and $\check\Sigma$ be the covariance matrices of $\hat\beta$ and $\tilde\beta$, respectively.
In Table 3.1, we report the empirical means, SD and covariance matrices of both the BJE $\hat\beta$ and the modified estimator $\tilde\beta$ across different sample sizes n, with $c = 0.51$. The reference rate in the tables is given by Remarks 3 and 4 in §2.

Table 3.1. Simulation results under $W \sim U(0,1)$ with $\beta = (1,1)'$

	$n = 40$	$n = 160$	$n = 640$	Reference Rate
(a) Exact estimator $\hat\beta$				
Mean	(0.962, 0.949)	(0.986, 0.982)	(0.991, 0.990)	–
SD	(0.0398, 0.0537)	(0.0120, 0.0403)	(0.0028, 0.0104)	$(2/n, \sqrt{5/(36n)})$
(b) Modified estimator $\tilde\beta$				
Mean	(0.957, 0.944)	(0.985, 0.984)	(0.995, 0.995)	–
SD	(0.0412, 0.0430)	(0.0110, 0.0121)	(0.0025, 0.0026)	$(2/n, 2/n)$
(c) Covariance matrices				
$\hat\Sigma$	$\begin{pmatrix}0.0144 & 0.0155\\0.0155 & 0.0141\end{pmatrix}$	$\begin{pmatrix}0.00037 & 0.00105\\0.00105 & 0.00174\end{pmatrix}$	$\begin{pmatrix}1\times10^{-6} & 2.1\times10^{-5}\\2.1\times10^{-5} & 1.2\times10^{-5}\end{pmatrix}$	–
$\tilde\Sigma$	$\begin{pmatrix}0.0156 & 0.0188\\0.0188 & 0.0163\end{pmatrix}$	$\begin{pmatrix}0.00051 & 0.00112\\0.00112 & 0.00202\end{pmatrix}$	$\begin{pmatrix}0.000002 & 0.000033\\0.000033 & 0.000014\end{pmatrix}$	–

Table 3.2. Mean of $\tilde\beta$ and the mean/SD of $(\hat\beta - \tilde\beta)$ under $W \sim U(0,1)$ with $\beta = (1,1)'$

Setting	n	Mean of $\tilde\beta$	Mean of $(\hat\beta - \tilde\beta)$	SD of $(\hat\beta - \tilde\beta)$
Case I	45	(0.964, 0.954)	(0.0158, 0.0156)	(0.025, 0.023)
	150	(0.984, 0.984)	(0.0048, 0.0048)	(0.0054, 0.0054)
	640	(0.992, 0.993)	(0.0013, 0.0013)	(0.0013, 0.0014)
Case II	40	(0.973, 0.957)	(0.014, 0.014)	(0.025, 0.026)
	160	(0.990, 0.990)	(0.0037, 0.0039)	(0.0058, 0.0058)
	640	(0.995, 0.999)	(0.0010, 0.0010)	(0.0014, 0.0015)

In Table 3.2, we summarize the replication-based comparisons of $\hat\beta$ and $\tilde\beta$ for Example 3 with $c = 0.5$. Two uniform-error scenarios are considered:
Case I. The sample consists of $3n$ observations with n replicates of each covariate triplet $(0,0)$, $(0,0.5)$, and $(0.5,0.5)$, again with $W \sim U(0,1)$ independent of X_i. This

corresponds to the special design described in Example 3.
Case II. The covariates X_i are drawn independently from $\{(0,0),(0.5,0),(0.5,0.5)\}$ with equal probability 1/3, and $W \sim U(0,1)$ is independent of X_i.

3.2 Simulation Results under the assumption $W \sim N(\mu, \sigma^2)$ and $c = 0.51$. We consider two assumptions: (a) $W \sim N(0,1)$ with $\beta = (1,1)'$, and (b) $W \sim N(2,4)$ with $\beta = (2,3)'$.

Table 3.3. Results under Normal errors $W \sim N(0,1)$ and $W \sim N(2,4)$

	$n = 40$	$n = 160$	$n = 640$	Reference Rate
(a) $W \sim N(0,1)$**,** $\beta = (1,1)'$				
Mean	(0.893, 0.875)	(0.930, 0.908)	(0.984, 0.979)	–
SD	(0.205, 0.223)	(0.097, 0.107)	(0.051, 0.058)	$\sqrt{1/n}$
$\hat{\Sigma}$	$\begin{pmatrix} 0.0089 & 0.0106 \\ 0.0106 & 0.0073 \end{pmatrix}$	$\begin{pmatrix} 0.0026 & 0.0028 \\ 0.0028 & 0.0026 \end{pmatrix}$	$\begin{pmatrix} 0.00057 & 0.00055 \\ 0.00055 & 0.00059 \end{pmatrix}$	–
$\tilde{\Sigma}$	$\begin{pmatrix} 0.0092 & 0.0122 \\ 0.0122 & 0.0099 \end{pmatrix}$	$\begin{pmatrix} 0.0030 & 0.0031 \\ 0.0031 & 0.0029 \end{pmatrix}$	$\begin{pmatrix} 0.00056 & 0.00056 \\ 0.00056 & 0.00059 \end{pmatrix}$	–
(b) $W \sim N(2,4)$**,** $\beta = (2,3)'$				
Mean	(1.642, 2.490)	(1.854, 2.775)	(1.975, 2.957)	–
SD	(0.303, 0.337)	(0.138, 0.153)	(0.062, 0.063)	–
$\hat{\Sigma}$	$\begin{pmatrix} 0.0090 & 0.0108 \\ 0.0107 & 0.0108 \end{pmatrix}$	$\begin{pmatrix} 0.0021 & 0.0026 \\ 0.0026 & 0.0021 \end{pmatrix}$	$\begin{pmatrix} 0.00049 & 0.00048 \\ 0.00049 & 0.00049 \end{pmatrix}$	–
$\tilde{\Sigma}$	$\begin{pmatrix} 0.0103 & 0.0110 \\ 0.0110 & 0.0105 \end{pmatrix}$	$\begin{pmatrix} 0.0029 & 0.0030 \\ 0.0030 & 0.0027 \end{pmatrix}$	$\begin{pmatrix} 0.00048 & 0.00048 \\ 0.00049 & 0.00049 \end{pmatrix}$	–

4. Data Analysis. The *Olympic* data set (available as `Olympic.NH4.df` in the **EnvStats** R package) contains weekly or biweekly measurements of ammonium (NH_4) concentrations (in mg/L) in wet atmospheric deposition, recorded from January 2009 to December 2011 at the Hoh Ranger Station in Olympic National Park, Washington. These measurements are part of the National Atmospheric Deposition Program/National Trends Network (NADP/NTN).

Table 4.1. Weekly measurements across months

	Week 1	Week 2	Week 3	Week 4	Week 5
Month 1		< 0.006			
Month 2	0.006		0.016	< 0.006	
Month 3	0.015	0.023	0.034	0.022	
Month 4	0.007	0.021	0.012	< 0.006	
Month 5	0.021	0.015	0.088	0.058	
Month 6	< 0.006	< 0.006			
Month 7	< 0.006	< 0.006	0.074		
Month 8	0.011	0.121	< 0.006		
⋮	⋮	⋮	⋮	⋮	⋮
Month 35	0.036	< 0.008	0.012	0.030	0.022
Month 36	0.008				

Table 4.1 presents data from the first eight and the last two months of the sampling

period. In total, the data set contains 56 uncensored (observed) values and 46 RC values, corresponding to multiple detection limits. Among these, four distinct detection limits were used, although only two of them appear in the subset shown in the table.

In addition to ammonium concentrations, the *Olympic* data set also contains covariates including real-time air temperature, denoted as X_1 (in Kelvin), and real-time humidity, denoted as X_2 (in oz/cu. yd). Based on a proportional hazards (PH) model applied to this data set (see references therein), a normal approximation yields a 95% confidence interval for the mean NH_4 deposition of (0.014, 0.028) mg/L, with the sample mean of the observed (uncensored) values being 0.020 mg/L. We also fit the data to a log-linear regression model of the form $\log(Y) = \beta'\mathbf{X} + W$, where $\mathbf{X} = (X_1, X_2)$. Using our proposed estimation algorithm, the resulting mean estimate for NH_4 deposition is 0.019 mg/L, which lies within the 95% confidence interval based on the PH model and is close to its sample mean.

To assess model adequacy, several classical diagnostic tools are available for linear regression analysis. In particular, Q–Q plots can be used—albeit in a less conventional manner—to evaluate the goodness-of-fit of the residuals under specific distributional assumptions.

Dong and Yu (2019), as well as Yu and Liu (2020), introduced marginal distribution (MD) approaches for model diagnostics under semiparametric regression frameworks. These MD-based methods allow flexible model checking for a variety of regression models, including the linear regression (LR) model, the proportional hazards model, the generalized LR model, and others. The MD framework incorporates both diagnostic plots and formal statistical tests and can be readily adapted to parametric settings. In this paper, we first introduce the parametric versions of MD diagnostics and subsequently apply them to the *Olympic* data set.

Suppose that $(X_1, Y_1), ..., (X_n, Y_n)$ are i.i.d. observations from $F_{X,Y}$ with density function $f_{X,Y}$, where X is a p-dimensional random vector and Y is a response variable. Let $F_{Y|X}$ be the conditional cumulative distribution function (cdf) with density function $f_{Y|X}$. Denote $F_o = F_{Y|X}(\cdot|0)$, which is called the baseline cdf of $F_{Y|X}$. Given $F_{X,Y} \in \Theta$, the family of all joint cdfs of (X, Y), F_o is well defined, even if (X, Y) does not satisfy the LR model and $E(W)$ may not exist. In particular, under assumption (AS), $F_o = F_W$. We consider the test of $H_0 : F_{X,Y} \in \Theta_0$, where $\Theta_0 = \{F_{X,Y} : (X, Y, W) \text{ satisfies (1.1)}\}$. Define $Y^* = \beta X + W^*$, where $F_{W^*}(\cdot) = F_{Y|X}(\cdot|0)$ and $X \perp W^*$.

Denote the edf of $F_Y(t)$ by $\hat{F}_Y(t) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{(Y_i \le t)}$. We call the 95% pointwise confidence interval of $F_Y(t)$ the confidence band (CB) of F_Y. Figures 2 and 3 present the 95% confidence bands of log-transformed edf of the NH4 wet deposition data within the log-transformed distribution of Y in our model and within the log-transformed distribution of Y using Cox Proportional-Hazards Model, called the parametric marginal distribution (MD) plots.

The MD plot is to plot $y = \hat{F}_{Y^*}(t)$ and $y = \hat{F}_Y(t)$ together with the 95% CB of F_Y, where $\hat{F}_{Y^*}(t) = \frac{1}{n}\sum_{i=1}^{n} \hat{F}_o(t - \hat{\beta}X_i)$, and $\hat{F}_o$ is the parametric MLE of F_o. If the two curves are close, e.g, the curve of $y = \hat{F}_{Y^*}(t)$ lies within the CB of F_Y, then it suggests that the model does fit the data. If most of the curve of $y = \hat{F}_{Y^*}(t)$ lie outside the CB of F_Y, then it suggests that the model does not fit the data.

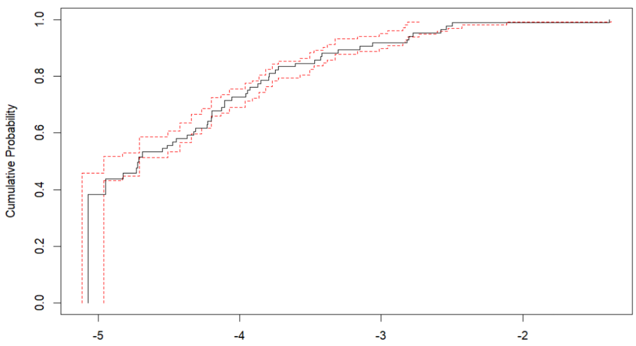

Figure 2. MD plot under log-LR model

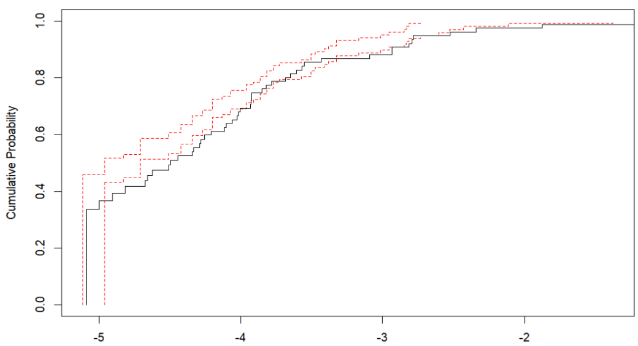

Figure 3. MD plot under Cox Proportional-Hazards model with log-transformed response

As the figures shown, the LR model with the log-transformed response fits the data better. The MD plotting method leads to a class of tests called MD tests for $H_0 : F_{X,Y} \in \Theta_0$, as follows. $T_1 = \sum_t |\hat{F}_Y(t) - \hat{F}_{Y^*}(t)|\hat{f}_Y(t)$ or $T_2 = \sup_t |\hat{F}_Y(t) - \hat{F}_{Y^*}(t)|$, $T_3 = \int \mathcal{W}(t)(\hat{F}_Y(t) - \hat{F}_{Y^*}(t))dG(t)$, or $T_4 = \int \mathcal{W}(t)|\hat{F}_Y(t) - \hat{F}_{Y^*}(t)|^k dG(t)$, where $\mathcal{W}(t)$ is a weight function, $k \geq 1$ is a constant, and dG is a measure, e.g., dt, $d\hat{F}_o$, $d\hat{F}_Y$ and $d\hat{F}_{Y^*}(t)$. The p-value can be obtained by the resampling method. Under the parametric assumption, in particular under Model (AS), the distributions of these statistics T_j can be approximated by making use of the procedure as follows.

[B0] Generate the $\{W_i^*\}_{i=1}^n$ from $U(\hat{a}, \hat{b})$, let $Y_i^* = \hat{\beta}X_i + W_i^*$, and evaluate a T_j above, say $T_1 = \sum_t |\hat{F}_Y(t) - \hat{F}_{Y^*}(t)|\hat{f}_Y(t)$.
[B1] Obtain the MLE $(\hat{a}^*, \hat{b}^*, \hat{\beta}^*)$ based on (Y_i^*, X_i)'s.
[B2] Generate a random sample of size n from $U(\hat{a}^*, \hat{b}^*)$, say, $W_1^{(1)}$, ..., $W_n^{(1)}$.
[B3] Resample X_i's of size n with replacement, say, $X_1^{(1)}$, ..., $X_n^{(1)}$.
[B4] Let $Y_i^{(1)} = \hat{\beta}^* X_i^{(1)} + W_i^{(1)}$, $i = 1, ..., n$.
[B5] Obtain the MLE $(\hat{a}^{*(1)}, \hat{b}^{*(1)}, \hat{\beta}^{*(1)})$ based on $(Y_i^{(1)}, X_i^{(1)})$'s.

[B6] Generate $W_i^{*(1)}$'s using $(\hat{a}^{*(1)}, \hat{b}^{*(1)})$ and let $Y_i^{*(1)} = \hat{\beta}^{*(1)} X_i^{(1)} + W_i^{*(1)}$, $i = 1, ..., n$.

[B7] Now, obtain a value of T_1, say $T_1^{(1)}$, based on $(Y_i^{(1)}, Y_i^{*(1)}, X_i^{(1)})$'s.

[B8] Repeat the steps B2, ..., B7 a large number of times, say 100 times, obtain $T_1^{(s)}$, $s = 2, ..., 100$. Thus the desired percentile can be estimated by the edf of these $T_1^{(s)}$'s.

If we use the MD test for our model with $T_2 = \sup_t |\hat{F}_Y(t) - \hat{F}_{Y^*}(t)|$ at significance level $\alpha = 0.05$, and repeat for 100 times, the p-value is 0.44, which suggests that the data fit the log- LR model. In contrast, for the Cox proportional-hazards model, the p-value is 0.04, which suggests that the data do not fit that model.

Section 5. Comments on other existing iterative algorithms for the approximation of the BJE. There are at least three existing iterative algorithms in the literature for approximating the BJE. Based on the original iterative algorithm by Buckley and James, Lin, Jin, and Ying (2006) developed a resampling-based estimator that involves iterative procedures similar in spirit to those proposed by Jin et al. (2003). Separately, Wang, Zhao, and Fu (2016) proposed the smoothed Buckley-James estimator (SBJ), which incorporates the induced smoothing framework originally introduced by Brown and Wang (2005). While these methods often perform well, they may fail to yield a stable solution for certain datasets. For example, consider a simulated dataset with $n = 40$, generated under the assumption $Y = \beta'\mathbf{X} + W$ with $\beta' = (0.5, 0.5)$, $X_1 \sim \text{Exp}(1) - 1$, $X_2 \sim \text{Exp}(1) - 2$, $W \sim N(3, 1)$, $C \sim \text{Exp}(1)$, where Y is subject to right-censoring by C, X_1, X_2 and W are independent, and the observed data consist of $T = \min(Y, C)$ and $\delta = \mathbf{1}\{Y \leq C\}$.

Table 5.1. Comparison of four BJE estimation methods under simulated data with `set.seed(100)` and $n = 40$.

Method	**Output**	**Remarks**
`bj()` (R package)	No convergence in 80 steps	Failed to converge
Lin, Jin, and Ying (2006)	Algorithm did not converge	Warning issued
SBJ (Wang, Zhao & Fu, 2016)	Algorithm did not converge	Last 4 steps oscillate
Proposed method	$\tilde{\beta} = (0.38, 0.31)$, SD=0.31	our modified BJE
Yu and Wong (2002)	$\hat{\beta} = (0.30, 0.32)$, SD=0.31	the true value of the BJE

Using `set.seed(100)`, we generated a dataset (given in Huang and Yu (2025) (see §A.9)) and applied these three iterative methods to this data set. None of them produced a convergent estimate: the `bj()` function which is the iterative algorithm by Buckley and James failed to converge within 80 iterations, the method of Lin, Jin, and Ying terminated with a convergence warning, and the SBJ estimator exhibited oscillatory behavior in its final steps. In contrast, our proposed method successfully converged to the correct solution.

A comparison of the four methods is summarized in Table 5.1. The SDs of our proposed estimator $\tilde{\beta}$ are $(0.162, 0.170)$, whereas for the algorithm of Yu and Wong (2002) the SDs of $\hat{\beta}$ are $(0.251, 0.257)$. The intervals based on each estimate and its SD cover the true parameter value. These results highlight the numerical instability and convergence issues that may arise with existing iterative methods. In contrast, our proposed approach demonstrates superior robustness and reliability, even in challenging settings where traditional algorithms fail.

Conflict Statement. The authors state that there is no conflict of interest.

References

[1] Buckley, J. and James, I. (1979). Linear regression with censored data. *Biometrika*, 66(3), 429–436. DOI: 10.1093/biomet/66.3.429.

[2] Chatterjee, S. and McLeish, D. L. (1986). Fitting linear regression models to censored data by least squares and maximum likelihood methods. *Communications in Statistics—Theory and Methods*, 15, 3227–3243. DOI: 10.1080/03610928608829305.

[3] Hillis, S. L. (1991). Extending M-estimation to include censored data via James's method. *Communications in Statistics—Simulation and Computation*, 20(1), 121–128. DOI: 10.1080/03610919108812943.

[4] Huang, Z. F. and Yu, Q. Q. (2025). Technical Report of "A Modification to the Buckley–James Estimate". Available at: `https://people.math.binghamton.edu/qyu/ftp/approx.pdf`.

[5] James, I. R. and Smith, P. J. (1984). Consistency results for linear regression with censored data. *The Annals of Statistics*, 12(2), 590–600. DOI: 10.1214/aos/1176346507.

[6] Kong, F. H. and Yu, Q. Q. (2007). Asymptotic distributions of the Buckley–James estimator under non-standard conditions. *Statistica Sinica*, 17, 341–360.

[7] Lai, T. L. and Ying, Z. L. (1991). Large sample theory of a modified Buckley–James estimator for regression-analysis with censored data. *The Annals of Statistics*, 19(3), 1370–1402. DOI: 10.1214/aos/1176348253.

[8] Leurgans, S. (1987). Linear models, random censoring and synthetic data. *Biometrika*, 74(2), 301–309. DOI: 10.1093/biomet/74.2.301.

[9] Miller, R. G. (1976). Least squares regression with censored data. *Biometrika*, 63(3), 449–464. DOI: 10.1093/biomet/63.3.449.

[10] Miller, R. and Halpern, J. (1982). Regression with censored data. *Biometrika*, 69(3), 521–531. DOI: 10.1093/biomet/69.3.521.

[11] Stare, J., Harrell, F. E. and Heinzl, H. (2001). BJ: an S-Plus program to fit linear regression models to censored data using the Buckley–James method. *Computer Methods and Programs in Biomedicine*, 64(1), 45–52. DOI: 10.1016/S0169-2607(00)00083-3.

[12] Wang, Y., Chen, C. and Kong, F. (2011). Variance estimation of the Buckley–James estimator under discrete assumptions. *Journal of Statistical Computation and Simulation*, 81(4), 481–496. DOI: 10.1080/00949650903421085.

[13] Yu, Q. Q. and Wong, G. Y. C. (2002). How to find all Buckley–James estimates instead of just one? *Journal of Statistical Computation and Simulation*, 72(6), 451–460. DOI: 10.1080/00949650213701.

[14] Yu, Q. Q. (2024). A proper selection among multiple Buckley–James estimates. *Metrika*, 87(6), 713–728. DOI: 10.1007/s00184-023-00933-1.

[15] Jin, Z., Lin, D. Y. and Ying, Z. (2006). On least-squares regression with censored data. *Biometrika*, 93(1), 147–161. DOI: 10.1093/biomet/93.1.147.

[16] Jin, Z., Lin, D. Y., Wei, L. J. and Ying, Z. (2003). Rank-based inference for the accelerated failure time model. *Biometrika*, 90(2), 341–353. DOI: 10.1093/biomet/90.2.341.

[17] Wang, Y.-G., Zhao, Y. and Fu, L. (2016). The Buckley–James estimator and induced smoothing. *Australian & New Zealand Journal of Statistics*, 58(2), 211–225. DOI: 10.1111/anzs.12155.

[18] Brown, B. M. and Wang, Y. G. (2005). Standard errors and covariance matrices for smoothed rank estimators. *Biometrika*, 92(1), 149–158. DOI: 10.1093/biomet/92.1.149.

[19] Zeng, D. and Lin, D. Y. (2008). Efficient resampling methods for nonsmooth estimating functions. *Biostatistics*, 9(2), 355–363. DOI: 10.1093/biostatistics/kxm034.

Appendix.

§A.1. Proof of Example 1 (continued)

⊢: If $X = (1,0,0,0)$, $\delta = (0,0,1,1)$, $M = (1,1,0.6,0.1)$, then each $b \in [0.9,\infty)$ is a BJE.

Since $T_1(b) = 1-b$, $T_2(b) = 1$, $T_3(b) = 0.6$, $T_4(b) = 0.1$ by the algorithm of Yu and Wong (2002), the possible solutions to $M_i - M_k = b(X_i - X_k)$ are $b \in \{0, 0.4, 0.9\}$, and they break $\mathcal{R}$ into 4 open intervals, denoted by $Q1, \ldots, Q4$. We shall show that

$$H(b) = \sum_i (X_i - \overline{X})(M_i - bX_i)$$

$$= \begin{cases} \left(\frac{3}{4}\right)(1-b) - \frac{1}{4}(1+0.6+0.1) = 0.325 - 0.75\,b > 0.325 & \forall\ b\ in\ Q1 = (-\infty,0), \\ \left(\frac{3}{4}\right)(1) - \frac{1}{4}(1+0.6+0.1) = 0.325 & \forall\ b\ in\ Q2 = (0,0.4), \\ \left(\frac{3}{4}\right)(\frac{1+0.6}{2}) - \frac{1}{4}(1+0.6+0.1) = 0.175 & \forall\ b\ in\ Q3 = (0.4,0.9), \\ \left(\frac{3}{4}\right)(\frac{1+0.6+0.1}{3}) - \frac{1}{4}(1+0.6+0.1) = 0 & \forall\ b\ in\ Q4 = (0.9,\infty). \end{cases}$$

$$H(0) = 0.325,\ H(0.4) = 0.175 \text{ and } H(0.9) = 0. \tag{a.1}$$

Q1: $b \leq 0$. Here $T_1(b) = 1-b > 1$ so the orders of T_i's are $0.1 < 0.6 < 1 < T_1(b)$. The largest time is $T_1(b)$ with $\delta_1 = 0$, it forces $\delta_1^* = 1$; hence all four are events and $T_i^* = T_i$, $M_i^* = M_i$, $X_i^* = X_i$. Thus $H(b) = \sum_i (X_i - \overline{X})(M_i - bX_i) = \left(\frac{3}{4}\right)(1-b) - \frac{1}{4}(1+0.6+0.1) = 0.325 - 0.75\,b > 0.325 > 0$, so no root of $H(b)$ in $(-\infty, 0)$.

Q2: $0 < b < 0.4$. Here $T_1(b) = 1-b \in (0.6,1)$ and $0.1 < 0.6 < T_1(b) < 1$. The largest time is $T_2 = 1$ with $\delta_2 = 0$, it forces $\delta_2^* = 1$. Conditional on $T > T_1(b)$, the only remaining event time is 1, so unit 1 has $T_1^*(b) = T_1(b)$, $M_1^*(b) = 1$, $X_1^*(b) = 0$. Units 2–4 are events: $M_2^* = 1$, $M_3^* = 0.6$, $M_4^* = 0.1$, $X_2^* = X_3^* = X_4^* = 0$. Thus $H(b) = \left(\frac{3}{4}\right) \cdot 1 - \frac{1}{4}(1+0.6+0.1) = 0.325 > 0$, so no root of $H(b)$ in $(0, 0.4)$.

Q3: $0.4 < b < 0.9$. Here $T_1(b) = 1-b \in (0.1,0.6)$ so the order is $0.1 < T_1(b) < 0.6 < 1$. Again it forces $\delta_2^* = 1$. Unit 1 lies between the events at 0.6 and 1, hence $T_1^*(b) = T_1(b)$, $M_1^*(b) = \frac{1}{2}(0.6) + \frac{1}{2}(1) = 0.8$, $X_1^*(b) = 0$. Units 2–4 remain events: $M_2^* = 1$, $M_3^* = 0.6$, $M_4^* = 0.1$, with $X_2^* = X_3^* = X_4^* = 0$. Thus $H(b) = \left(\frac{3}{4}\right) \cdot 0.8 - \frac{1}{4}(1+0.6+0.1) = 0.175 > 0$, hence no root in $(0.4, 0.9)$.

Q4: $b > 0.9$. Here $T_1(b) = 1-b \in (0,0.1)$ so $T_1(b) < 0.1 < 0.6 < 1$. Since unit 1 is censored before the first event, $\widehat{S}(0.1) = \frac{2}{3}$, $\widehat{S}(0.6) = \frac{1}{3}$, $\widehat{S}(1) = 0$, so conditional on $T > T_1(b)$, $P(T = 0.1) = P(T = 0.6) = P(T = 1) = \frac{1}{3}$. Thus $T_1^*(b) = T_1(b)$, $M_1^*(b) = \frac{1}{3}(0.1 + 0.6 + 1) = \frac{17}{30}$, $X_1^*(b) = 0$. Units 2–4 are events with $M_2^* = 1$, $M_3^* = 0.6$, $M_4^* = 0.1$. Hence $H(b) = \left(\frac{3}{4}\right) \cdot \frac{17}{30} - \frac{1}{4}(1+0.6+0.1) = 0$, so $H(b) \equiv 0$ on $(0.9, \infty)$.

Boundary behavior. Similarly, it can be shown that $H(0) = 0.325$, $H(0.4) = 0.175$, $H(0.9) = 0$.

Thus (a.1) holds and $b = 0.9$ is the SZC, and every $b \in [0.9, \infty)$ is a ZC (hence a BJE solution). □

§A.2. Proof of Proposition 1.

⊢: Suppose that $\delta_i = 1$ for all i. Then $\tilde{\beta} = \hat{\beta}$.

For simplicity, assume that $\delta_i \equiv 1$ and $E(Y) = \mathbf{B}'\beta$. Then we have $\mathbf{M}' = \mathbf{Y}$. Let

$\gamma = \mathbf{V}^{-1}\beta$. Then $E(Y) = \mathbf{B}'\beta = \mathbf{U}D\mathbf{V}'\beta = \mathbf{U}D\gamma$, as $\gamma = \mathbf{V}'\beta$.

$$\begin{aligned}\hat\gamma =&((\mathbf{U}D)'(\mathbf{U}D))^{-1}(\mathbf{U}D)'Y = (D'\mathbf{U}'\mathbf{U}D)^{-1}(\mathbf{U}D)'Y\\ =&(D'D)^{-1}D'\mathbf{U}'Y = D^{-1}\mathbf{U}'Y \quad => \qquad \hat\gamma' = (U_1'\mathbf{Y}/d_1, ..., U_p'\mathbf{Y}/d_p).\\ \hat\beta =&(\mathbf{B}'\mathbf{B})^{-1}\mathbf{B}'\mathbf{Y} = ((\mathbf{U}D\mathbf{V}')'(\mathbf{U}D\mathbf{V}'))^{-1}(\mathbf{U}D\mathbf{V}')'Y\\ =&(\mathbf{V}D\mathbf{U}'\mathbf{U}D\mathbf{V}')^{-1}(\mathbf{V}D\mathbf{U}')'Y\\ =&(\mathbf{V}D^2\mathbf{V}')^{-1}(\mathbf{V}D\mathbf{U}')'Y = (\mathbf{V}D^{-2}\mathbf{V}')\mathbf{V}D\mathbf{U}'Y = \mathbf{V}D^{-1}\mathbf{U}'Y = \mathbf{V}\hat\gamma\\ =&\tilde\beta. \square\end{aligned}$$

§A.3. Proof of Example 2.

For convenience, restate Example 2 as follows.

Suppose that $P(\mathbf{X} = (0,0)') = P(\mathbf{X} = (0.5,0)') = P(\mathbf{X} = (0.5,0.5)') = 1/3$, $W \sim U(0,1)$, $C \equiv c \in [0,1]$, $\mathbf{X} \perp W$ and $\beta = (1,1)'$. Let $M_{(i)}$'s be the order statistics of M_i's, $M_{(m)}$ be the largest value among the M_i's that is less than c, and $\hat\beta = (\hat\beta_1, \hat\beta_2) = (2c - 2M_{(1)}, 2M_{(m)} - 2M_{(1)})$.

Then

(1) If $c \in (0.5, 1]$, then $\exists!$ BJE of $\mathbf{b}$ near $(1,1)$.

(2) If $c = 0.5$, then $\hat\beta$ is the unique SZC of $H(\mathbf{b})$ and $H(\mathbf{b}) = (0,0)'$ if $\mathbf{b} \in (\hat\beta_1, \infty) \times (\hat\beta_2, \infty)$.

The proof is given in 2 steps.

Step 1. $\vdash$: Statement (2) holds.

Assume $c = 0.5$. WLOG we can assume $M_1 \le M_2 \le \cdots \le M_n$. In particular, $M_1 \le M_2 \le \cdots \le M_m < c = M_{m+1} = \cdots M_n$. The key step for finding all BJE's is to find the solutions to $\begin{cases} M_i - M_k = b_1(X_{i1} - X_{k1}) + b_2(X_{i2} - X_{k2}) \\ M_h - M_j = b_1(X_{h1} - X_{j1}) + b_2(X_{h2} - X_{j2}), \end{cases}$ i.e.,

$$\begin{cases} T_i(\mathbf{b}) = T_k(\mathbf{b}) \\ T_h(\mathbf{b}) = T_j(\mathbf{b}), \end{cases} \quad \text{where } T_i(\mathbf{b}) = \begin{cases} M_i (\in (0,c)) & \text{if } i = 1, ..., m \\ c & \text{if } i = m+1, ..., s \\ c - 0.5b_1 & \text{if } i = s+1, ..., t \\ c - 0.5(b_1 + b_2) & \text{if } i = t+1, ..., n. \end{cases}$$

Then the solutions are

$(b_1, b_2) = \{(0,0), (2c - 2M_j, -(2c - 2M_j)), (0, 2c - 2M_j), (2c - 2M_j, 0), (2c - 2M_j, 2M_i - 2M_j)\}$. If the largest $T_i(b)$ is right censored, it is treated as an exact observation in $H(b)$.

Modify δ_i as $\delta^*(b_1, b_2) = \mathbf{1}(i \in \{1, ..., m\}) + \mathbf{1}(i \notin \{1, ..., m\}, b_1 < 0, b_2 \ge 0, b_1 + b_2 \le 2(c - M_m)) + \mathbf{1}(i \in \{s+1, ..., t\}, b_1 < 0, b_2 \ge 0, b_1 + b_2 > 2(c - M_m)) + \mathbf{1}(i \notin \{1, ..., m\}, b_2 < 0, b_1 \le 2(c - M_m)) + \mathbf{1}(i \in \{t+1, ..., n\}, b_2 < 0, b_1 > 2(c - M_m)) + \mathbf{1}(i \in \{m+1, ..., n\}, 0 \le b_1 \le 2(c - M_m), 0 \le b_1 + b_2 \le 2(c - M_m))$. Then $T_i^*(b) =$

$$\begin{cases} M_i(M_i \in (0,c)) & \text{if } i = 1,...,m \\ c & \text{if } i \notin \{1,...,m\}, 0 \le b_1 \le 2(c-M_m); 0 \le b_1+b_2 \le 2(c-M_m) \\ c-0.5b_1 & \text{if } i \notin \{1,...,m\}; b_1 < 0 \le b_2, b_1+b_2 \le 2(c-M_m) \\ c-0.5(b_1+b_2) & \text{if } i \notin \{1,...,m\}; b_2 < 0, b_1 \le 2(c-M_m) \\ c-0.5b_1 & \text{if } i \in \{m+1,...,s,t+1,...,n\}; b_1 < 0 \le b_2, b_1+b_2 > 2(c-M_m) \\ c-0.5(b_1+b_2) & \text{if } i \in \{t+1,...,n\}; b_2 < 0, b_1 > 2(c-M_m) \\ G_i(\cdot) & \text{if } i \notin \{1,...,s\}; b_1, b_1+b_2 > 0, b_1 > 2(c-M_m) \text{ or } b_1+b_2 > 2(c-M_m), \end{cases}$$

$$\text{where } G_i(\cdot) = \begin{cases} \frac{\sum \mathbf{1}_{(M_j > T_i(\mathbf{b}))} M_j + (c-0.5b_1)(s-m)}{(\sum \mathbf{1}_{(M_j > T_i(\mathbf{b}))} + s - m)} & \text{if } i = s+1,...,t,\ b_1+b_2 \text{ and } b_1 > 2(c-M_m) \\ \frac{\sum \mathbf{1}_{(M_j > T_i(\mathbf{b}))} M_j + (c-0.5(b_1+b_2))(t-s)}{(\sum \mathbf{1}_{(M_j > T_i(\mathbf{b}))} + t - s)} & \text{if } i = t+1,...,n,\ b_1+b_2 \text{ and } b_1 > 2(c-M_m) \\ \frac{\sum \mathbf{1}_{(M_j > T_i(\mathbf{b}))} M_j + (c-0.5b_1)(s-m)}{(\sum \mathbf{1}_{(M_j > T_i(\mathbf{b}))} + s - m)} & \text{if } i = s+1,...,t, b_1 > 2(c-M_m) \ge b_1+b_2 \\ \frac{\sum \mathbf{1}_{(M_j > T_i(\mathbf{b}))} M_j + c(t-s)}{(\sum \mathbf{1}_{(M_j > T_i(\mathbf{b}))} + t - s)} & \text{if } i = t+1,...,n, b_1 > 2(c-M_m) \ge b_1+b_2 \\ \frac{\sum \mathbf{1}_{(M_j > T_i(\mathbf{b}))} M_j + c(s-m)}{(\sum \mathbf{1}_{(M_j > T_i(\mathbf{b}))} + s - m)} & \text{if } i = s+1,...,t, b_1 \le 2(c-M_m) < b_1+b_2 \\ \frac{\sum \mathbf{1}_{(M_j > T_i(\mathbf{b}))} M_j + (c-0.5(b_1+b_2))(t-s)}{(\sum \mathbf{1}_{(M_j > T_i(\mathbf{b}))} + t - s)} & \text{if } i = t+1,...,n, b_1 \le 2(c-M_m) < b_1+b_2. \end{cases}$$

Note that $\overline{\mathbf{X}} = (\overline{X}_1, \overline{X}_2)'$, thus by Eq. (2.1), $H(\mathbf{b}) = (H_1(\mathbf{b}), H_2(\mathbf{b}))'$ (where $\mathbf{b} = (b_1, b_2)'$). For $b_1 < 0 \le b_2$, $b_1 + b_2 \le 2(c - M_m)$, $H_1(\mathbf{b})$ and $H_2(\mathbf{b})$ are linear functions of b_1, say, $H_1^*(b_1) \overset{def}{=} H_1(\mathbf{b})$ and $H_2^*(b_1) \overset{def}{=} H_2(\mathbf{b})$, and for $j = 1, 2$,
$H_j^*(b_1) = \sum_{i=1}^m M_i(0 - \overline{X}_j) + \sum_{i>m} c(1 - b_1)(0.5 - \overline{X}_j)$;
$H_j^*(b_1)$ is decreasing in $b_1 \in (-\infty, 0)$;
$H_j^*(b_1) \ge \sum_{i=1}^m M_i(-\overline{X}_j) + \sum_{i>m} c(0.5 - \overline{X}_j) > 0$ a.s., as $m \approx n/3$ if n is large when $b_1 < 0$.

For $b_2 < 0$, $b_1 \le 2(c - M_m)$, $H_1(\mathbf{b})$ and $H_2(\mathbf{b})$ are linear functions of $b_1 + b_2$, *i.e.*
$H_1^*(b_1 + b_2) \overset{def}{=} H_1(\mathbf{b})$ and $H_2^*(b_1 + b_2) \overset{def}{=} H_2(\mathbf{b})$, and for $j = 1, 2$,
$H_j^*(b_1 + b_2) = \sum_{i=1}^m M_i(0 - \overline{X}_j) + \sum_{i>m} c(1 - b_1 - b_2)(0.5 - \overline{X}_j)$;
$H_j^*(b_1 + b_2)$ is decreasing in $b_1 + b_2$;
$H_j^*(b_1 + b_2) \ge \sum_{i=1}^m M_i(-\overline{X}_j) + \sum_{i>m} c(0.5 - \overline{X}_j) > 0$ a.s., as $m \approx n/3$ if n is large.

For $b_1 > 0$ and $b_1 + b_2 > 0$,

$$\begin{aligned} H_2(b_1, b_2) &= -\sum_{i=1}^m M_i \overline{X}_2 - \sum_{i>m}^s c\overline{X}_2 + \sum_{i>s}^t T_i^*(b_1, b_2)(-\overline{X}_2) + \sum_{i>t}^n T_i^*(b_1, b_2)(0.5 - \overline{X}_2) \\ &= -(\sum_{i=1}^m M_i + \sum_{i>m}^s c)\overline{X}_2 + \sum_{i>s}^t G(b_1, b_2)(-\overline{X}_2) + \sum_{i>t}^n G(b_1, b_2)(0.5 - \overline{X}_2); \\ H_1(b_1, b_2) &= -\sum_{i=1}^m M_i \overline{X}_1 - \sum_{i>m}^s c\overline{X}_1 + \sum_{i>s}^t T_i^*(b_1, b_2)(0.5 - \overline{X}_1) + \sum_{i>t}^n T_i^*(b_1, b_2)(0.5 - \overline{X}_1) \\ &= -(\sum_{i=1}^m M_i + \sum_{i>m}^s c)\overline{X}_1 + \sum_{i>s}^t G(b_1, b_2)(0.5 - \overline{X}_1) + \sum_{i>t}^n G(b_1, b_2)(0.5 - \overline{X}_1). \end{aligned}$$

$G(\cdot)$ is a left-continuous decreasing step function in $\{b_1 > 0\} \times \{b_1 + b_2 > 0\}$, with drops at $(2c-2M_i, 2c-2M_j)$, where $i = s+1, ..., t$, and $j = t+1, ..., n$. For example, for $0 < b_1 + b_2 \leq 2(c - M_m)$, $0 < b_1 \leq 2(c - M_m)$,

$$G(b_1, b_2) = \frac{\sum_{j=1}^{m} \mathbf{1}_{(b_1+b_2>2c-2M_j)} M_j + c(s-m)}{(\sum_{j=1}^{m} \mathbf{1}_{(b_1+b_2>2c-2M_j)} + s - m)} \qquad \text{(letting } M_0 = -\infty\text{)}$$
$$= \mathbf{1}_{(b_1+b_2 \in (0, 2c-2M_m])} c + \sum_{i=1}^{m} \mathbf{1}_{(b_1+b_2 \in (2c-2M_i, 2c-2M_{i-1}])} \frac{\sum_{j=i}^{m} M_j + c(s-m)}{(m-i+1+s-m)},$$

$G(\hat{\beta}_1, \hat{\beta}_2) = \frac{\sum_{j=2}^{m} M_j + c(s-m)}{(s-1)} > \frac{\sum_{j=1}^{m} M_j + c(s-m)}{s} = G(\hat{\beta}_1+, \hat{\beta}_2+)$, as $M_1 \leq .... \leq M_m < c$. $H_j(b_1, b_2) = 0 < H_j(\hat{\beta}) \leq H_j(t_1, t_2)$ $\forall t_1 \leq \hat{\beta}_1, t_2 \leq \hat{\beta}_2$ and $b_1 > \hat{\beta}_1$, $b_1 + b_2 > \hat{\beta}_2$. Thus $\hat{\beta}$ is the unique SZC and each $\mathbf{b} \in (\hat{\beta}_1, \infty) \times (\hat{\beta}_2, \infty)$ is a root of $H(\mathbf{b})$.

Step 2. $\vdash$: Statement (1) holds:
Let $\delta_i = \mathbf{1}(i \leq m \text{ or } i \in (k, h])$, $M_1 \leq \cdots \leq M_m < c$, and $0.5 < M_{k+1} \leq \cdots \leq M_h < c$, $M_i = \begin{cases} W_i(< c) & \text{if } i = 1, ..., m \\ c(\leq W_i) & \text{if } i = m+1, ..., k \\ W_i + 0.5(< c) & \text{if } i = k+1, ..., h \\ c(\leq W_i + 0.5) & \text{if } i = h+1, ..., n. \end{cases}$

Then $T_i(\mathbf{b}) = \begin{cases} M_i(\in (0, c)) & \text{if } i = 1, ..., m \\ c & \text{if } i = m+1, ..., k \\ M_i - 0.5b_1 & \text{if } i = k+1, ..., t \\ M_i - 0.5(b_1 + b_2) & \text{if } i = t+1, ..., h \\ c - 0.5b_1 & \text{if } i = h+1, ..., s \\ c - 0.5(b_1 + b_2) & \text{if } i = s+1, ..., n. \end{cases}$ Let $\delta_i^*(\mathbf{b}) = \begin{cases} 1 & \text{if } i \leq k \text{ and } b_1 \leq 0, b_1 + b_2 \leq 0, \text{ or } i > k \text{ and } b_1 \geq 0 \text{ or } b_1 + b_2 \geq 0 \\ \delta_i & \text{o.w..} \end{cases}$. Then

$$T_i^*(\mathbf{b}) = \begin{cases} M_i(= W_i) & \text{if } i = 1, ..., m \\ Z(\cdot) & \text{if } i = m+1, ..., k \text{ and } b_1 \leq 0, b_1 + b_2 \leq 0 \\ c & \text{if } i = m+1, ..., k; b_1 > 0; b_1 + b_2 > 0 \\ M_i - 0.5b_1 & \text{if } i = k+1, ..., h; b_1 \leq 0, b_1 + b_2 > 0 \\ c - 0.5b_1 & \text{if } i = h+1, ..., n; b_1 \leq 0, b_1 + b_2 > 0 \\ M_i - 0.5(b_1 + b_2) & \text{if } i = k+1, ..., h; b_1 > 0, b_1 + b_2 \leq 0 \\ c - 0.5(b_1 + b_2) & \text{if } i = h+1, ..., n; b_1 > 0, b_1 + b_2 \leq 0 \\ M_i - 0.5b_1 & \text{if } i = k+1, ..., t; b_1 \leq 0, b_1 + b_2 \leq 0 \\ M_i - 0.5(b_1 + b_2) & \text{if } i = t+1, ..., h; b_1 \leq 0, b_1 + b_2 \leq 0 \\ c - 0.5b_1 & \text{if } i = h+1, ..., s; b_1 \leq 0, b_1 + b_2 \leq 0 \\ c - 0.5(b_1 + b_2) & \text{if } i > s; b_1 \leq 0, b_1 + b_2 \leq 0 \\ G(\cdot) & \text{if } i > h, b_1 > 0, b_1 + b_2 > 0, \end{cases}$$

where $Z(\cdot) = \frac{\sum_{j>k}^{h} \mathbf{1}_{(M_j > T_i(\mathbf{b}))}(M_j - \mathbf{b}'X_j) + (c - \mathbf{b}'X_j)(n-h)}{(\sum \mathbf{1}_{(M_j > T_i(\mathbf{b}))} + n - h)}$, $G(\cdot) = \frac{\sum \mathbf{1}_{(M_j > T_i(\mathbf{b}))} M_j + c(k-m)}{(\sum \mathbf{1}_{(M_j > T_i(\mathbf{b}))} + k - m)}$.

Notice that $-2(c-0.5) \leq 2(M_{m+1} - c) \leq \cdots \leq 2(M_k - c) < 0$, for $b_1 \leq 0$, $b_1 + b_2 \leq 0$. $Z(b_1, b_2) = \frac{\sum_{j>k}^{h} \mathbf{1}_{(M_j > T_i(\mathbf{b}))}(M_j - \mathbf{b}'X_j) + (c - \mathbf{b}'X_j)(n-h)}{(\sum \mathbf{1}_{(M_j > T_i(\mathbf{b}))} + n - h)}$ is a 2-dimensional left-continuous

strictly decreasing function; $G(\cdot)$ is a left-continuous decreasing step function in $\{b_1 > 0\} \times \{b_1 + b_2 > 0\}$, with drops at $(2c - 2M_i, 2c - 2M_j)$, where $i = s+1, ..., t$, and $j = t+1, ..., n$ from the proof of Statement (1). $H(\mathbf{b}) = (H_1(\mathbf{b}), H_2(\mathbf{b}))'$, $\overline{\mathbf{X}} = (\overline{X}_1, \overline{X}_2)'$. If n is large enough, then $\frac{m}{n} \approx 1/6$, $\frac{k}{n} \approx \frac{2c-0.5}{3}$, $\frac{h-k}{n} \approx \frac{c-0.5}{3}$, $\frac{n-h}{n} \approx \frac{1.5-c}{3}$, $\overline{X}_1 \approx 1/3$, $\overline{X}_2 \approx 1/6$.
$\frac{H_1(1,1)}{m/3} = \frac{2}{3}c^2 + (c-0.5)(0.5-\frac{1}{3}) + (1-(c-0.5))\frac{(c-1/6)(0.5)+(1-c)c/6)}{1.5-c} = \frac{c^2}{6} + \frac{7c}{6} - \frac{1}{6} > 0$,
Similarly, $\frac{H_1(1-,1-)}{m/3} = -\frac{c^2}{6} - \frac{1}{12} < 0$, $\frac{H_2(1,1)}{m/6} = \frac{c^2}{12} + \frac{3c}{8} - \frac{1}{12} > 0$, $\frac{H_2(1-,1-)}{m/6} = -\frac{c^2}{12} - \frac{3c}{4} - \frac{1}{3} < 0$.
If $\mathbf{b} = (b_1, b_2)$ where $b_1 < 0$ and $b_2 \leq -b_1$,

$$T_i^*(\mathbf{b}) = \begin{cases} M_i (= W_i) & \text{if } i = 1, ..., m \\ Z(\cdot) = \frac{\sum_{j>k}^{h} \mathbf{1}_{(M_j > T_i(\mathbf{b}))}(M_j - \mathbf{b}'X_j) + (c - \mathbf{b}'X_j)(n-h)}{(\sum \mathbf{1}_{(M_j > T_i(\mathbf{b}))} + n - h)} & \text{if } i = m+1, ..., k \\ M_i - 0.5b_1 & \text{if } i = k+1, ..., t \\ M_i - 0.5(b_1 + b_2) & \text{if } i = t+1, ..., h \\ c - 0.5b_1 & \text{if } i = h+1, ..., s \\ c - 0.5(b_1 + b_2) & \text{if } i > s, \end{cases}$$

$$\begin{aligned} H_1(b_1, b_2) = & \sum_{i=1}^{m} M_i \overline{X}_1 - \sum_{i>m}^{k} (X_{i1} - \overline{X}_1) Z(b_1, b_2) + \sum_{i>k}^{\prime} (M_i - 0.5b_1)(X_{i1} - \overline{X}_1) \\ & + \sum_{i>t}^{h} (M_i - 0.5(b_1 + b_2))(X_{i1} - \overline{X}_1) + \sum_{i>h}^{s} (c - 0.5b_1)(X_{i1} - \overline{X}_1) \\ & + \sum_{i>s}^{n} (c - 0.5(b_1 + b_2))(X_{i1} - \overline{X}_1); \\ H_2(b_1, b_2) = & \sum_{i=1}^{m} M_i \overline{X}_2 - \sum_{i>m}^{k} (X_{i2} - \overline{X}_2) Z(b_1, b_2) + \sum_{i>k}^{\prime} (M_i - 0.5b_1)(X_{i2} - \overline{X}_2) \\ & + \sum_{i>t}^{h} (M_i - 0.5(b_1 + b_2))(X_{i2} - \overline{X}_2) + \sum_{i>h}^{s} (c - 0.5b_1)(X_{i2} - \overline{X}_2) \\ & + \sum_{i>s}^{n} (c - 0.5(b_1 + b_2))(X_{i2} - \overline{X}_2). \end{aligned}$$

If $b_0 < b_1 < 0$ and $b_2 \leq -b_1$, $H_1(b_0, b_2) - H_1(b_1, b_2)$

$$
\begin{aligned}
&= \sum_{i>m}^{k} (X_{i1} - \overline{X}_1)(Z(b_0, b_2) - Z(b_1, b_2)) + \sum_{i>k}^{l} (M_i - 0.5(b_0 - b_1))(X_{i1} - \overline{X}_1) \\
&\quad + \sum_{i>t}^{h} (M_i - 0.5(b_0 - b_1))(X_{i1} - \overline{X}_1) + \sum_{i>h}^{s} (c - 0.5(b_0 - b_1))(X_{i1} - \overline{X}_1) \\
&\quad + \sum_{i>s}^{n} (c - 0.5(b_0 - b_1))(X_{i1} - \overline{X}_1) > 0;
\end{aligned}
$$

$$
\begin{aligned}
H_2(b_0, b_2) - H_2(b_1, b_2)& \\
&= \sum_{i>m}^{k} (X_{i2} - \overline{X}_2)(Z(b_0, b_2) - Z(b_1, b_2)) + \sum_{i>k}^{l} (M_i - 0.5(b_0 - b_1))(X_{i2} - \overline{X}_2) \\
&\quad + \sum_{i>t}^{h} (M_i - 0.5(b_0 - b_1))(X_{i2} - \overline{X}_2) + \sum_{i>h}^{s} (c - 0.5(b_0 - b_1))(X_{i2} - \overline{X}_2) \\
&\quad + \sum_{i>s}^{n} (c - 0.5(b_0 - b_1))(X_{i2} - \overline{X}_2) > 0.
\end{aligned}
$$

If $b_1 < 0$ and $b_0 < b_2 < -b_1$,

$$
\begin{aligned}
&H_1(b_1, b_0) - H_1(b_1, b_2) = \sum_{i>m}^{k} (X_{i1} - \overline{X}_1)(Z(b_1, b_0) - Z(b_1, b_2)) \\
&+ \sum_{i>t}^{h} (M_i - 0.5(b_0 - b_2))(X_{i1} - \overline{X}_1) + \sum_{i>s}^{n} (c - 0.5(b_0 - b_2))(X_{i1} - \overline{X}_1) > 0; \\
&H_2(b_0, b_2) - H_2(b_1, b_2) = \sum_{i>m}^{k} (X_{i2} - \overline{X}_2)(Z(b_1, b_0) - Z(b_1, b_2)) \\
&+ \sum_{i>t}^{h} (M_i - 0.5(b_0 - b_2))(X_{i2} - \overline{X}_2) + \sum_{i>s}^{n} (c - 0.5(b_0 - b_2))(X_{i2} - \overline{X}_2) > 0.
\end{aligned}
$$

Therefore, both $H_1(b_1, b_2)$ and $H_2(b_1, b_2)$ are strictly decreasing when $b_1 < 0$ and $b_2 \leq -b_1$, i.e., over the region $(-\infty, 0) \times (-\infty, -b_1]$.

If $b_1 > 0$ and $b_2 \leq -b_1$, $T_i^*(\mathbf{b}) = \begin{cases} M_i(= W_i) & \text{if } i = 1, ..., m \\ M_i - 0.5(b_1 + b_2) & \text{if } i = k+1, ..., h \\ c - 0.5(b_1 + b_2) & \text{if } i = h+1, ..., n, \end{cases}$

$$H_1(b_1, b_2) = \sum_{i=1}^{m} M_i \overline{X}_1 + \sum_{i>k}^{h} (M_i - 0.5(b_1 + b_2))(X_{i1} - \overline{X}_1) + \sum_{i>h}^{n} (c - 0.5(b_1 + b_2))(X_{i1} - \overline{X}_1);$$
$$H_2(b_1, b_2) = \sum_{i=1}^{m} M_i \overline{X}_2 + \sum_{i>k}^{h} (M_i - 0.5(b_1 + b_2))(X_{i2} - \overline{X}_2) + \sum_{i>h}^{n} (c - 0.5(b_1 + b_2))(X_{i2} - \overline{X}_2).$$

If $0 < b_0 < b_1$ and $b_2 \leq -b_1$,

$$H_1(b_0, b_2) - H_1(b_1, b_2) = \sum_{i>k}^{h} (M_i - 0.5(b_0 - b_1))(X_{i1} - \overline{X}_1) + \sum_{i>h}^{n} (c - 0.5(b_0 - b_1))(X_{i1} - \overline{X}_1) > 0;$$

$$H_2(b_0, b_2) - H_2(b_1, b_2) = \sum_{i>k}^{h} (M_i - 0.5(b_0 - b_1))(X_{i2} - \overline{X}_2) + \sum_{i>h}^{n} (c - 0.5(b_0 - b_1))(X_{i2} - \overline{X}_2) > 0.$$

If $b_1 > 0$ and $b_0 < b_2 \leq -b_1$,

$$H_1(b_1, b_0) - H_1(b_1, b_2) = \sum_{i>k}^{h} (M_i - 0.5(b_0 - b_2))(X_{i1} - \overline{X}_1) + \sum_{i>h}^{n} (c - 0.5(b_0 - b_2))(X_{i1} - \overline{X}_1) > 0;$$

$$H_2(b_1, b_0) - H_2(b_1, b_2) = \sum_{i>k}^{h} (M_i - 0.5(b_0 - b_2))(X_{i2} - \overline{X}_2) + \sum_{i>h}^{n} (c - 0.5(b_0 - b_2))(X_{i2} - \overline{X}_2) > 0.$$

Then both $H_1(b_1, b_2)$ and $H_2(b_1, b_2)$ are strictly decreasing over the region $(0, \infty) \times (-\infty, -b_1]$.

If $b_1 < 0$ and $b_2 > -b_1$, $T_i^*(\mathbf{b}) = \begin{cases} M_i(= W_i) & \text{if } i = 1, ..., m \\ M_i - 0.5b_1 & \text{if } i = k+1, ..., h \\ c - 0.5b_1 & \text{if } i = h+1, ..., n, \end{cases}$

$$H_1(b_1, b_2) = \sum_{i=1}^{m} M_i \overline{X}_1 + \sum_{i>k}^{h} (M_i - 0.5b_1)(X_{i1} - \overline{X}_1) + \sum_{i>h}^{n} (c - 0.5b_1)(X_{i1} - \overline{X}_1);$$
$$H_2(b_1, b_2) = \sum_{i=1}^{m} M_i \overline{X}_2 + \sum_{i>k}^{h} (M_i - 0.5b_1)(X_{i2} - \overline{X}_2) + \sum_{i>h}^{n} (c - 0.5b_1)(X_{i2} - \overline{X}_2).$$

If $b_1 > 0$ and $b_2 > -b_1$, $T_i^*(\mathbf{b}) = \begin{cases} M_i(= W_i) & \text{if } i = 1, ..., m \\ c & \text{if } i = m+1, ..., k \\ G(\cdot) & \text{if } i = k+1, ..., n, \end{cases}$

$$H_1(b_1, b_2) = \sum_{i=1}^{m} M_i \overline{X}_1 + \sum_{i>m}^{k} c(X_{i1} - \overline{X}_1) + \sum_{i>k}^{n} G(b_1, b_2)(X_{i1} - \overline{X}_1);$$

$$H_2(b_1, b_2) = \sum_{i=1}^{m} M_i \overline{X}_2 + \sum_{i>m}^{k} c(X_{i1} - \overline{X}_2) + \sum_{i>k}^{n} G(b_1, b_2)(X_{i2} - \overline{X}_2).$$

Therefore, we find that $H_1(b_1, b_2)$ is strictly decreasing and $H_2(b_1, b_2)$ is strictly increasing throughout $\mathbb{R}^2$. Hence, the BJE of $\boldsymbol{\beta}$ exists and is unique in a neighborhood of $(1, 1)$. □

§A.4. Proof of Theorem 1.

First look at the simplest case: $\mathbf{B} = (\mathbf{X}_1, \mathbf{X}_2)'$, where $\mathbf{X}_1 = (0, 0.5, 0.5)$ and $\mathbf{X}_2 = (0, 0, 0.5)$. The singular value decomposition of $\mathbf{B}$ is $\mathbf{B} = \mathbf{U}\Sigma\mathbf{V}'$, where $\mathbf{U} \approx \begin{pmatrix} 0 & 0 & -1 \\ -0.53 & -0.85 & 0 \\ -0.85 & 0.53 & 0 \end{pmatrix}$, $\Sigma \approx \begin{pmatrix} 0.81 & 0 \\ 0 & 0.31 \\ 0 & 0 \end{pmatrix}$ and $\mathbf{V} \approx \begin{pmatrix} -0.85 & -0.53 \\ -0.53 & 0.85 \end{pmatrix}$.

Therefore $\mathbf{U}\Sigma = \mathbf{B}\mathbf{V} = \begin{pmatrix} 0 & 0 \\ -0.43 & -0.26 \\ -0.69 & 0.16 \end{pmatrix}$. Let $A = \underbrace{\begin{bmatrix} B \\ \vdots \\ B \end{bmatrix}}_{n \text{ blocks}} \in \mathbb{R}^{3n \times 2}$, where $B = \begin{bmatrix} 0 & 0 \\ 0.5 & 0 \\ 0.5 & 0.5 \end{bmatrix} \in \mathbb{R}^{3 \times 2}$. Equivalently, let $1_n \in \mathbb{R}^n$ be the all-ones vector and I_3 the 3×3 identity, $A = (1_n \otimes I_3)\, B$. We seek a full SVD $A = U \Sigma V'$ with $U \in \mathbb{R}^{3n \times 3n}$ orthogonal, $V \in \mathbb{R}^{2 \times 2}$ orthogonal, and rectangular diagonal $\Sigma \in \mathbb{R}^{3n \times 2}$.

Right singular vectors V (independent of n):
Since $A'A = \sum_{k=1}^{n} B'B = n\,(B'B)$, $A'A$ differs from $B'B$ by a scalar factor n. Hence V is the orthonormal eigenbasis of $B'B$ and does not depend on n, while the singular values scale by $\sqrt{n}$. $B'B = \begin{bmatrix} 0.5 & 0.25 \\ 0.25 & 0.25 \end{bmatrix}$, $\lambda_{1,2} = \frac{3 \pm \sqrt{5}}{8}$. Let $\varphi = \frac{1+\sqrt{5}}{2}$. Solving $(B'B - \lambda I)v = 0$ yields the unnormalized eigenvectors $v_1 \propto \begin{bmatrix} 1 \\ \frac{1}{\varphi} \end{bmatrix}$, $v_2 \propto \begin{bmatrix} 1 \\ -\varphi \end{bmatrix}$. After normalization, $V = \begin{bmatrix} \dfrac{1}{\sqrt{1 + (1/\varphi)^2}} & \dfrac{1}{\sqrt{1 + \varphi^2}} \\ \dfrac{1/\varphi}{\sqrt{1 + (1/\varphi)^2}} & \dfrac{-\varphi}{\sqrt{1 + \varphi^2}} \end{bmatrix}$, $\sigma_{1,2}(A) = \sqrt{n\,\lambda_{1,2}} = \sqrt{\frac{n(3 \pm \sqrt{5})}{8}}$.

Left singular vectors U:
Write the thin SVD of B as $B = U_B \Sigma_B V'$, $U_B = \begin{bmatrix} u_1^B & u_2^B \end{bmatrix} \in \mathbb{R}^{3 \times 2}$, $\Sigma_B = \operatorname{diag}\big(\sigma_1(B), \sigma_2(B)\big)$, with the same V as above and singular values $\sigma_i(B) = \sqrt{\lambda_i}$. Let

$e := \frac{1_n}{\sqrt{n}} \in \mathbb{R}^n$, $E_\perp \in \mathbb{R}^{n\times(n-1)}$ have orthonormal columns spanning $e^\perp$. Then
(i) Columns of U corresponding to the two nonzero singular values.
For $i = 1, 2$, $u_i = \frac{1}{\sigma_i(A)} A v_i = \frac{1}{\sqrt{n}\,\sigma_i(B)} (1_n \otimes I_3) B v_i = \frac{1}{\sqrt{n}} (1_n \otimes u_i^B) = e \otimes u_i^B$.
Thus the first two columns of U are $u_1 = e \otimes u_1^B$, $u_2 = e \otimes u_2^B$.

(ii) The left nullspace (columns of U for zero singular values). Since $\mathrm{rank}(A) = 2$, the left nullspace of A has dimension $3n - 2$ and decomposes as
$\ker(A') = \big(e^\perp \otimes \mathbb{R}^3\big) \oplus \big(\mathrm{span}(e) \otimes \ker(B')\big)$.
Note that $\ker(B')$ is one-dimensional; an explicit unit vector can be chosen as $q_0 = (1, 0, 0)'$, due to $q_0' B = 0$. Therefore, an explicit orthonormal basis for $\ker(A')$ is given by $(E_\perp \otimes e_1,\ E_\perp \otimes e_2,\ E_\perp \otimes e_3) \cup (e \otimes q_0)$, where (e_1, e_2, e_3) is any orthonormal basis of $\mathbb{R}^3$ (e.g., the standard basis; one can also take (u_1^B, u_2^B, q_0)).

(iii) Assembling a full orthogonal U. Stacking the orthonormal columns, we can take $U = [\, e\otimes u_1^B,\ e\otimes u_2^B \mid e\otimes q_0 \mid E_\perp\otimes e_1,\ E_\perp\otimes e_2,\ E_\perp\otimes e_3 \,] \in \mathbb{R}^{3n\times 3n}$. By construction, U is orthogonal, its first two columns satisfy $Av_i = \sigma_i(A) u_i$ ($i = 1, 2$), and the remaining $3n - 2$ columns span $\ker(A')$.
The identities $U'U = I_{3n}$, $U\Sigma V' = A$, $\Sigma = \begin{bmatrix} \mathrm{diag}(\sigma_1(A), \sigma_2(A)) \\ 0 \end{bmatrix} \in \mathbb{R}^{3n\times 2}$ hold up to the usual columnwise sign ambiguity of singular vectors.

Given that $P(X = (0, 0)') = P(X = (0.5, 0)') = P(X = (0.5, 0.5)') = 1/3$, when n is large, it suffices to analyze the data that have $3n$ pairs $(\mathbf{X}_i, Y_i)$ with n triples of $\mathbf{X}_i$ equal to $(0, 0)$, $(0, 0.5)$, and $(0.5, 0.5)$.
For this data, we still have
$\mathbf{V} \approx \begin{pmatrix} -0.85 & -0.53 \\ -0.53 & 0.85 \end{pmatrix}$. Thus $\tilde{\mathbf{B}} = \mathbf{U}\Sigma = \mathbf{B}\mathbf{V} \approx \begin{pmatrix} 0 & -0.43 & -0.69 & 0 & -0.43 & -0.69 & ... \\ 0 & -0.26 & 0.16 & 0 & -0.26 & 0.16 & ... \end{pmatrix}'_{2\times 3n}$.
Set $c = 0.5$. By our modification, $\hat{\gamma}_1$ is the BJE based on $(M, \vec{\delta}, \tilde{\mathbf{B}}_1)$, where $\tilde{\mathbf{B}}_1 = (0, -0.43, -0.69, 0, -0.43, -0.69, ...)$, WLOG, we can assume that $M_1 \le M_2 \le \cdots \le M_m < c = M_{m+1} = \cdots M_n$. The key step for finding all BJE's is to find the solutions to $T_i(b) = T_j(b)$ $\forall \tilde{X}_i > \tilde{X}_j$, where
$$T_i(b) = \begin{cases} M_i (= W_i) & \text{if } i = 1, ..., m \\ 0.5 & \text{if } i = m+1, ..., k \\ 0.5 + 0.43b & \text{if } i = k+1, ..., s \\ 0.5 + 0.69b & \text{if } i = s+1, ..., n, \end{cases}$$
Possible solutions to $T_i(b) = T_j(b)$ are $b \in \{0, (M_i - 0.5)/0.43, (M_i - 0.5)/0.69\}$, $T_{(n)} = 0.5 + 0.69b\mathbf{1}(b \ge 0)$, $\delta_i^*(b) = \mathbf{1}(b < 0, m < i \le k) + \mathbf{1}(b \ge 0, k < i \le s)$ for $i > m$. $$T_i^*(b) = \begin{cases} M_i (= W_i) & \text{if } i = 1, ..., m \\ 0.5 & \text{if } i = m+1, ..., k \text{ and } b < 0 \\ 0.5 + 0.69b & \text{if } i = k+1, ..., n \text{ and } b \ge 0 \\ G_i(b) = \frac{\sum \mathbf{1}_{(M_j > T_i(b))} M_j + c(k-m)}{\sum \mathbf{1}_{(M_j > T_i(b))} + k - m} & \text{if } i > k \text{ and } b < 0. \end{cases}$$
If $b_1 < b < 0$, $$T_i^*(b) - T_i^*(b_1) = \begin{cases} 0 & \text{if } i = 1, ..., k \\ G_i(b) - G_i(b_1) & \text{if } i = k+1, ..., n, \end{cases}$$
$T_i(b)$ is strictly increasing when $i = k+1, ..., n$, $M_j < c$, therefore $G_i(b)$ is a left-

continuous increasing step function on $(-\infty, 0)$ with drops at
$b = (M_1 - 0.5)/0.69, ..., (M_m - 0.5)/0.69,\ (M_1 - 0.5)/0.43, ..., (M_m - 0.5)/0.43$,
$H(b_1) - H(b) = \sum_{i=k+1}^{s} -0.05(G(b_1) - G(b)) + \sum_{i=s+1}^{n} -0.32 \cdot (G(b_1) - G(b)) > 0$.
Thus $H(b) \downarrow$ if $b < 0$.

If $0 \geq b_1 < b$, $T_i^*(b) - T_i^*(b_1) = \begin{cases} 0 & \text{if } i = 1, ..., k \\ b - b_1 & \text{if } i = k+1, ..., n, \end{cases}$

$H(b_1) - H(b) = \sum_{i=k+1}^{s} -0.05(b_1 - b) + \sum_{i=s+1}^{n} -0.32(b_1 - b) > 0$. Thus $H(b) \downarrow$ if $b \geq 0$. $H(b) = \sum_1^m (0.37)M_i + \sum_{m+1}^{k}(0.37)c + \sum_{k+1}^{s}(-0.05G_i(b)) + \sum_{s+1}^{n}(-0.32G_i(b))$.
When $b < (M_1 - 0.5)/0.43$, $H(b) = 0$, therefore, $\hat{\gamma}_1 = (M_1 - 0.5)/0.43 \approx -1.16$.

Similarly, $\hat{\gamma}_2$ is the BJE based on $(M, \vec{\delta}, \tilde{\mathbf{B}}_2)$, where $\tilde{\mathbf{B}}_1 = (0, -0.26, 0.16, 0, -0.26, 0.16, ...)$. Assume that $M_1 \leq M_2 \leq \cdots \leq M_m < c = M_{m+1} = \cdots M_n$.

$$T_i(b) = \begin{cases} M_i (= W_i) & \text{if } i = 1, ..., m \\ 0.5 & \text{if } i = m+1, ..., k \\ 0.5 + 0.26b & \text{if } i = k+1, ..., s \\ 0.5 - 0.16b & \text{if } i = s+1, ..., n, \end{cases}$$

Possible solutions to $T_i(b) = T_j(b)$ are $b \in \{0, (M_i - 0.5)/0.26, (0.5 - M_i)/0.16\}$,

$$T_i^*(b) = \begin{cases} M_i (= W_i) & \text{if } i = 1, ..., m \\ 0.5 & \text{if } i = m+1, ..., k \text{ and } b < 0 \\ 0.5 + 0.26b & \text{if } i = k+1, ..., s \text{ and } b \geq 0 \\ 0.5 - 0.16b & \text{if } i = s+1, ..., n \text{ and } b < 0 \\ G_i(b) & \text{if } i = k+1, ..., s \text{ and } b < 0 \text{ or } i = s+1, ..., n \text{ and } b \geq 0, \end{cases}$$

where $G(\cdot) = \begin{cases} \frac{\sum \mathbf{1}_{(M_j > T_i(b))} M_j + (0.5 - 0.16b)(n-s) + 0.5(k-m)}{\sum \mathbf{1}_{(M_j > T_i(b))} + n - s + k - m} & \text{if } i = k+1, ..., s \text{ and } b < 0 \\ \frac{\sum \mathbf{1}_{(M_j > T_i(b))} M_j + (0.5 + 0.26b)(s-k) + 0.5(k-m)}{\sum \mathbf{1}_{(M_j > T_i(b))} + s - m} & \text{if } i = s+1, ..., n \text{ and } b \geq 0. \end{cases}$

$G_i(b)$ is left-continuous increasing on $(-\infty, 0)$ with drops at $b = (M_1 - 0.5)/0.26, ..., (M_m - 0.5)/0.26$, and left-continuous decreasing on $(0, \infty)$ with drops at $b = (0.5 - M_m)/0.16, ..., (0.5 - M_1)/0.16$. $H(b) = \sum_1^m 0.03M_i + \sum_{m+1}^{k}(0.03)0.5 + \sum_{k+1}^{s}(-0.23T_i^*(b)) + \sum_{s+1}^{n}(0.19T_i^*(b))$.
If $b_1 < b < 0$, $H(b_1) - H(b) = \sum_{k+1}^{s}(-0.23(G_i(b_1) - G_i(b))) + \sum_{s+1}^{n} 0.19(b - b_1) > 0$, therefore, $H(b) \downarrow$ if $b < 0$. If $0 \leq b_1 < b$, $H(b_1) - H(b) = \sum_{k+1}^{s}(-0.23(b_1 - b)) + \sum_{s+1}^{n} 0.19(G_i(b_1) - G_i(b)) > 0$. Thus $H(b) \downarrow$ if $b \geq 0$. $H(0) = \sum_1^m 0.03M_i + \sum_{m+1}^{k}(0.03)0.5 + \sum_{k+1}^{s}(-0.23)(0.5) + \sum_{s+1}^{n}(0.19)(0.5) < 0$. When $b > (0.5 - M_1)/0.16$, $H(b) = 0$, thus $\hat{\gamma}_2 = (0.5 - M_1)/0.16 \approx 3.13$. □

§A.5. Derivation of $\sigma^2_{\hat{\beta}_1}$ and $\sigma^2_{\hat{\beta}_2}$ in Remark 3. Recall $\hat{\beta}_2 = 2M_{(m)} - 2M_{(1)}$ and $f_{M_{(1)},M_{(m)}}(x_1, x_m) = \binom{n}{1, m-2, 1, n-m} f(x_1) f(x_m) (F(x_m) - F(x_1))^{m-2} (S(x_m))^{n-m}$. Under $U(0,1)$, $0 < x_1 < x_m < 1$ and $f_{M_{(1)},M_{(m)}}(x_1, x_m) = \binom{n}{1, m-2, 1, n-m}(x_m - x_1)^{m-2}(1 - x_m)^{n-m}$.
Let $w = x_m - x_1$, the probability density function (pdf) $g(w)$ of $M_{(m)} - M_{(1)}$ is

$g(w)=\binom{n}{1,m-2,1,n-m}w^{m-2}\int_0^{1-w}((1-w)-x_1)^{n-m}dx_1$. Let $x_1=z(1-w)$,

$$\begin{aligned}\int_0^{1-w}((1-w)-x_1)^{n-m}dx_1 =&\int_0^{1}((1-w)-z(1-w))^{n-m}(1-w)dz\\ =&(1-w)^{n-m+1}\int_0^{1}(1-z)^{n-m}dz\end{aligned}$$

Thus $g(w)=\binom{n}{1,m-2,1,n-m}w^{m-2}(1-w)^{n-m+1}\int_0^1(1-z)^{n-m}dz$.
Since $\int_0^1(1-z)^{n-m}dz=B(1,n-m+1)$, the beta function,

$$\begin{aligned}g(w)=&\binom{n}{1,m-2,1,n-m}w^{m-2}(1-w)^{n-m+1}B(1,n-m+1)\\ =&\binom{n}{1,m-2,1,n-m}w^{m-2}(1-w)^{n-m+1}\frac{(n-m)!}{(n+1-m)!}\end{aligned}$$

yields that $g(w)=\frac{(n)!}{(n+1-m)!(m-2)!}w^{m-2}(1-w)^{n-m+1}$, which is the pdf of $Beta(m-1,n-m+2)$.
$\frac{1}{2}\hat{\beta}_2\sim Beta(m-1,n-m+2)$ yields that
$\sigma^2_{\hat{\beta}_2}=4(\sigma^2_{M_{(m)}}+\sigma^2_{M_{(1)}}-2Cov(M_{(m)},M_{(1)}))=4[\frac{m(n-m+1)}{(n+1)^2(n+2)}+\frac{n}{(n+1)^2(n+2)}-\frac{2(n-m+1)}{(n+1)^2(n+2)}]$. □

§A.6. Derivation of $\sigma_{\tilde{\beta}_i}$ in Remark 4. We first revisit a key component of our modified algorithm— its transformation of the design matrix into an orthogonal basis through singular value decomposition. Let the design matrix $B\in\mathbb{R}^{n\times p}$ have the singular value decomposition (SVD): $B=U\Sigma V'$, where $U\in\mathbb{R}^{n\times n}$ satisfies $U'U=I_n$, $\Sigma\in\mathbb{R}^{n\times p}$ is a diagonal matrix containing the singular values, $V\in\mathbb{R}^{p\times p}$ satisfies $V'V=I_p$.
The linear model becomes $Y=B\beta+W=U\Sigma V'\beta+W$.
Define $J=U\Sigma$ and $\eta=V'\beta$, thus $Y=J\eta+W$.
To verify that the columns of J are orthogonal, we compute: $J'J=(U\Sigma)'(U\Sigma)=\Sigma'U'U$.
Since $U'U=I_n$, it follows that $J'J=\Sigma'\Sigma$.
Hence, the columns of J are orthogonal, as $J'J$ is diagonal. From the result above, assume $c=0.5$, for $\tilde{\gamma}_2$, since $W\sim U(0,1)$, in a similar way, it can be shown that $\tilde{\gamma}_2=2c-2M_{(1)}$, then $\sigma^2_{\tilde{\gamma}_2}=4\sigma^2_{M_{(1)}}=4\frac{n}{(n+1)^2(n+2)}\approx\frac{4}{n^2}$, $\sigma_{\tilde{\gamma}_2}\approx\frac{2}{n}$. Because the columns of J are orthogonal, $\sigma_{\tilde{\gamma}_1}\approx\frac{2}{n}$. The parameter transformation $\tilde{\beta}=V'\gamma$ represents the projection of the original coefficient vector γ onto the orthonormal basis formed by the right singular vectors of B. Thus $\sigma_{\tilde{\beta}_1}=\sigma_{\tilde{\beta}_2}\approx\frac{2}{n}$.

§A.7. Dataset used in section 5.

```
> set.seed(100)
> X1 = rexp(40,1)-1
> X2 = rexp(40,1)-2
> W = rnorm(40,3,1)
> C = rexp(40,1)
> X1
 [1] -0.07578838 -0.27616282 -0.89535513  2.09736234 -0.37519476  0.17442933 -0.90688281
 [8]  0.74839077 -0.75000705 -0.80567353 -0.47489784 -0.66195658  1.02319166  0.12324659
[15]  0.13104764 -0.61941894 -0.92837769 -0.57839231 -0.92330260 -0.50205555  0.37355298
[22]  0.75765606 -0.45939678 -0.80749656 -0.06515796 -0.60945172 -0.87495123 -0.90969064
```

```
[29] -0.44483112 -0.34539310 -0.79335129 -0.16287652 -0.07850923  0.01526657  0.97572692
[36]  0.57683127 -0.29329020  1.22597879  0.14168355 -0.64720366
> X2
 [1] -1.0588629 -1.1369676 -0.6579047 -0.9959493 -1.8548705 -1.5669925 -1.4504422 -1.3319458
 [9]  0.5436060 -0.8913817 -1.9737972 -0.4682717 -0.8743009 -1.0369141 -1.9650805 -0.6117929
[17]  1.6926827 -0.7489380  1.0862769 -0.8612663 -1.8580438 -1.6302395 -1.9919973 -1.3268571
[25] -1.3862665  0.3743638 -0.5875932 -1.8608325 -1.3195985 -0.7202034 -0.7360438 -1.0625966
[33]  2.1294648 -1.8966976  0.7107049 -0.1019819 -1.2956354 -0.8701737  0.4719963 -0.4650352
> W
 [1] 4.331444 3.609377 1.286996 2.155061 4.449412 2.042647 3.899763 3.799767 3.518908 3.172302
[11] 2.011158 2.016507 3.552303 4.069043 2.790775 3.934464 2.549438 3.393486 3.268965 3.576410
[21] 3.453594 2.609456 1.763391 3.987087 2.841761 4.797495 2.470176 3.252704 3.800993 4.709522
[31] 3.928155 4.163567 3.282857 2.802167 3.679962 2.452839 3.337054 3.655829 1.202088 2.846535
> C
 [1] 0.18675555 2.30298096 2.22000486 1.17447326 0.88155598 0.10215654 1.17270953 0.69137906
 [9] 0.24158141 1.28970727 0.89179329 1.21762653 0.40226987 0.15218058 0.43222072 0.32843593
[17] 0.41250040 0.81165915 0.42450099 0.32188422 3.08499476 2.96650475 1.25242206 1.18030233
[25] 0.68628497 0.09184619 0.23421824 0.58260902 1.04534201 0.13693588 0.50920579 0.05972813
[33] 0.46520084 0.48996440 0.77288643 0.46311903 0.81170865 1.05907050 2.06825870 3.51018993
```