Bootstrap Adjustments to Minimum p-Value Method for Predictive Classification

Na Li, Yanglei Song, C. Devon Lin and Dongsheng Tu

Queen's University

Supplementary Material

The supplementary materials include proofs of all theorems and lemmas in the main text, as well as additional discussions and empirical results. Section S1 presents the proof of Theorem 1 and Theorem 2. Section S2 includes the proof of Theorem 3 as well as Lemma S2.6 for Example 1. Section S3 provides the proof of Lemma 1 and Theorem 4. Section S4 contains the proof of Lemma 2. Section S5 provides additional details on related works. Section S6 includes additional simulation and data analysis results.

S1 Proofs for the random design setup

S1.1 Proof of Theorem 1

In this subsection we provide the proof of Theorem 1. First we briefly review the definition of the functional weak convergence.

Let S be an arbitrary index set, and for a function $f : S \to \mathbb{R}$, denote its ℓ_∞ norm by $\|f\|_\infty = \sup_{s\in S} |f(s)|$. Denote by $\ell^\infty(S)$ the space of uniformly

bounded, real-valued functions on S equipped with ℓ_∞ norm. Let $Z_n = \{Z_{n,s} : s \in S\}$ $(n = 1, 2, \ldots)$ be a sequence of random processes indexed by S, for which $\|Z_n\|_\infty < \infty$ for each $n = 1, 2, \ldots$ almost surely, and Z be a tight random element in $\ell^\infty(S)$. Then as in Van Der Vaart and Wellner (1996), we say Z_n $(n = 1, 2, \ldots)$ *converges weakly* to Z in $\ell^\infty(S)$ if $E(g(Z_n)) \to E(g(Z))$, as $n \to \infty$, for any bounded and continuous function $g : \ell^\infty(S) \to \mathbb{R}$.

The random processes Z_n $(n = 1, 2, \ldots)$, viewed as maps from the underlying probability spaces to $\ell^\infty(S)$, are usually not Borel measurable, in which case the expectations are with respect to outer-probabilities. For details and a definitive treatment of functional weak convergence, we refer readers to Van Der Vaart and Wellner (1996).

Proof. Recall that the random process we are interested in, $\{M_{n,c} : c \in [\ell, u]\}$ where $M_{n,c} = \sqrt{n}\hat{\lambda}_c/\hat{v}_c$, is defined in (2.3). We will first establish the functional convergence for $\{\sqrt{n}\hat{\lambda}_c : c \in [\ell, u]\}$ and $\{\hat{v}_c : c \in [\ell, u]\}$ separately, and then apply the Slutsky's theorem (Kosorok, 2007, Theorem 7.15) and the continuous mapping theorem (Van Der Vaart and Wellner, 1996, Theorem 1.3.6) to obtain the limiting distribution of $\{M_{n,c} : c \in [\ell, u]\}$.

First we consider $\{\sqrt{n}\hat{\lambda}_c : c \in [\ell, u]\}$. Denote by $PX = E(X)$ the

expected value of a random vector or matrix X. Recall that $\hat{Q}_{n,c}$ is defined as $n^{-1}\sum_{i=1}^{n} Z_{i,c}Z_{i,c}^T$ in (2.2). We further define

$$Q_{c_1,c_2} = PZ_{1,c_1}Z_{1,c_2}^T, \quad Q_{c_1} = Q_{c_1,c_1}, \text{ for any real numbers } c_1, c_2. \tag{S1.1}$$

For each $c \in [\ell, u]$, we have

$$\begin{aligned}\sqrt{n}\hat{\lambda}_c =& d^T\hat{Q}_{n,c}^{-1}\left\{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(Z_{i,c}Y_i - PZ_{1,c}Y)\right\} + d^T\sqrt{n}\left(\hat{Q}_{n,c}^{-1} - Q_c^{-1}\right)PZ_{1,c}Y \\ &+\sqrt{n}d^TQ_c^{-1}PZ_{1,c}Y \qquad \text{(S1.2)} \\ =& d^T\hat{Q}_{n,c}^{-1}\left\{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(Z_{i,c}Y_i - PZ_{1,c}Y)\right\} + d^T\sqrt{n}\left(\hat{Q}_{n,c}^{-1} - Q_c^{-1}\right)PZ_{1,c}Y,\end{aligned}$$

where the term (S1.2) is equal to zero for any $c \in [\ell, u]$, due to Lemma S1.1 and the fact that $\lambda_0 = 0$ under H_0.

Define two index sets S and S' as follows:

$$S' = \{1, \ldots, 16\} \times [\ell, u], \quad S = \{1, \ldots, 4\} \times S'. \tag{S1.3}$$

By the functional central limit theorem (Van Der Vaart and Wellner, 1996, Theorem 2.5.2), there exists a tight, zero mean Gaussian process $\{G_{1,c}, G_{2,c} : c \in [\ell, u]\}$ such that

$$\left\{\begin{matrix} \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(Z_{i,c}Y_i - PZ_{1,c}Y) \\ \sqrt{n}(\hat{Q}_{n,c} - Q_c)\end{matrix} : c \in [\ell, u]\right\} \rightsquigarrow \left\{\begin{matrix} G_{1,c} \\ G_{2,c}\end{matrix} : c \in [\ell, u]\right\} \quad \text{in } \ell^\infty(S), \tag{S1.4}$$

where $\rightsquigarrow$ represents the weak convergence as defined above, and $G_{1,c}$ and $G_{2,c}$ are respectively a length 4 random vector and a 4×4 random matrix for a given c. Denote $\mathbf{G}_1 = \{G_{1,c} : c \in [\ell, u]\}$ and $\mathbf{G}_2 = \{G_{2,c} : c \in [\ell, u]\}$.

Let $\mathcal{D}_\phi \subset \ell^\infty(S')$ be the collection of all random processes of the form, $\{D_c : c \in [\ell, u]\}$, where D_c is a 4×4 symmetric and invertible matrix, that are entry-wise uniformly (over $c \in [\ell, u]$) bounded. Now define $\phi(\cdot) : \mathcal{D}_\phi \to \mathcal{D}_\phi$ as the "component-wise" matrix inverse map, i.e., for $\{D_c : c \in [\ell, u]\} \in \mathcal{D}_\phi$,

$$\phi(\{D_c : c \in [\ell, u]\}) = \left\{D_c^{-1} : c \in [\ell, u]\right\}. \tag{S1.5}$$

By Lemma S1.4, $\phi(\cdot)$ is Hadamard-differentiable at $\{Q_c : c \in [\ell, u]\}$ with the derivative map $\phi'_Q(\{D_c : c \in [\ell, u]\}) = \{-Q_c^{-1} D_c Q_c^{-1} : c \in [\ell, u]\}$ for any $\{D_c : c \in [\ell, u]\} \in \mathcal{D}_\phi$. Thus we apply the functional delta method (Kosorok, 2007, Theorem 2.8) and have that in $\ell^\infty(S)$,

$$\left\{ \begin{array}{c} \frac{1}{\sqrt{n}} \sum_{i=1}^n (Z_{i,c} Y_i - P Z_{1,c} Y) \\ \sqrt{n}\left(\hat{Q}_{n,c}^{-1} - Q_c^{-1}\right) \end{array} : c \in [\ell, u] \right\} \rightsquigarrow \left\{ \begin{array}{c} \mathbf{G}_1 \\ \phi'_Q(\mathbf{G}_2) \end{array} \right\}. \tag{S1.6}$$

Further, by the Glivenko-Cantelli theorem (Van Der Vaart and Wellner, 1996, Theorem 2.4.1) and the continuous mapping theorem (Van Der Vaart and Wellner, 1996, Theorem 1.3.6), we have that almost surely,

$$\{\hat{Q}_{n,c}^{-1} : c \in [\ell, u]\} \to \{Q_c^{-1} : c \in [\ell, u]\} = \mathbf{Q}^{-1}, \text{ in } \ell^\infty(S'). \tag{S1.7}$$

Combining (S1.6) and (S1.7), together with the Slutsky's theorem (Kosorok, 2007, Theorem 7.15), we obtain

$$\left\{ \begin{array}{c} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (Z_{i,c} Y_i - P Z_{1,c} Y) \\ \sqrt{n} \left(\hat{Q}_{n,c}^{-1} - Q_c^{-1} \right) \\ \hat{Q}_{n,c}^{-1} \end{array} : c \in [\ell, u] \right\} \rightsquigarrow \left\{ \begin{array}{c} \mathbf{G}_1 \\ \phi'_Q(\mathbf{G}_2) \\ \mathbf{Q}^{-1} \end{array} \right\} \quad \text{in} \quad \ell^\infty(S * S').$$

Then by the continuous mapping theorem (Van Der Vaart and Wellner, 1996, Theorem 1.3.6), $\{\sqrt{n}\hat{\lambda}^c : c \in [\ell, u]\}$ converges weakly to the random process $\{G_{3,c} : c \in [\ell, u]\}$ in $\ell^\infty([\ell, u])$, where

$$G_{3,c} = d^T Q_c^{-1} G_{1,c} - d^T Q_c^{-1} G_{2,c} Q_c^{-1} Q_{c,c_0} \theta_0. \tag{S1.8}$$

Second, we consider $\{\hat{v}_c^2 : c \in [\ell, u]\}$. By the Glivenko-Cantelli theorem (Van Der Vaart and Wellner, 1996, Theorem 2.4.1) and the continuous mapping theorem (Van Der Vaart and Wellner, 1996, Theorem 1.3.6), $\{\hat{v}_c^2 : c \in [\ell, u]\}$ converges to $\{v_c^2 : c \in [\ell, u]\}$ almost surely in $\ell^\infty([\ell, u])$, where v_c^2 is deterministic and defined as

$$v_c^2 = d^T Q_c^{-1} d(\theta_0^T Q_{c_0} \theta_0 + \sigma^2 - \theta_0^T Q_{c,c_0}^T Q_c^{-1} Q_{c,c_0} \theta_0). \tag{S1.9}$$

Finally, by the continuous mapping theorem (Van Der Vaart and Wellner, 1996, Theorem 1.3.6), together with the Slutsky's theorem (Kosorok, 2007, Theorem 7.15), we obtain that $\{M_{n,c} : c \in [\ell, u]\}$ converges weakly to a tight, zero mean Gaussian process $\mathbf{G} = \{G_{3,c}/v_c : c \in [\ell, u]\}$ in $\ell^\infty([\ell, u])$. $\square$

S1.2 Proof of Theorem 2

In this subsection we provide the proof of Theorem 2. Recall the definitions of S, S', $G_{1,c}$, $G_{2,c}$, $G_{3,c}$, and v_c in (S1.3), (S1.4), (S1.8) and (S1.9) respectively. Further define $\hat{Q}^*_{n,c} = n^{-1}\sum_{i=1}^n Z^*_{i,c}(Z^*_{i,c})^T$, where recall that $Z^*_{i,c} = (1, U^*_i, X^*_{i,c}, U^*_i X^*_{i,c})^T$ and $X^*_{i,c} = I(X^*_i \leq c)$.

Proof. Consider the random process consisting of the bootstrap test statistics $\{M^*_{n,c} - M_{n,c} : c \in [\ell, u]\}$, as defined in (3.1), and the bootstrap test statistic for each c is:

$$M^*_{n,c} - M_{n,c} = \frac{\sqrt{n}\hat{\lambda}^*_c}{\hat{v}_c} - \frac{\sqrt{n}\hat{\lambda}_c}{\hat{v}_c} = \frac{\sqrt{n}(\hat{\lambda}^*_c - \hat{\lambda}_c)}{\hat{v}_c}.$$

In the proof of Theorem 1, we have shown that conditional on almost all sequence (Y_i, U_i, X_i) $(i = 1, 2, \ldots)$, $\{\hat{v}^2_c : c \in [\ell, u]\}$ converges to $\{v^2_c : c \in [\ell, u]\}$ in $\ell^\infty([\ell, u])$. Thus to establish the first statement, it suffices to show that conditional on almost all sequence (Y_i, U_i, X_i) $(i = 1, 2, \ldots)$, $\{\sqrt{n}(\hat{\lambda}^*_c - \hat{\lambda}_c) : c \in [\ell, u]\}$ converges weakly to $\{G_{3,c} : c \in [\ell, u]\}$ in $\ell^\infty([\ell, u])$, on which we now focus.

For each c, $\sqrt{n}(\hat{\lambda}_c^* - \hat{\lambda}_c)$ can be written as:

$$\begin{aligned}\sqrt{n}(\hat{\lambda}^{c*} - \hat{\lambda}^c) &= d^T \left(\hat{Q}_{n,c}^*\right)^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^n Z_{i,c}^* Y_i^* - d^T \hat{Q}_{n,c}^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^n Z_{i,c} Y_i \\ &= d^T \left(\hat{Q}_{n,c}^*\right)^{-1} \frac{1}{\sqrt{n}} \left(\sum_{i=1}^n Z_{i,c}^* Y_i^* - \sum_{i=1}^n Z_{i,c} Y_i \right) \\ &+ d^T \sqrt{n} \left\{ \left(\hat{Q}_{n,c}^*\right)^{-1} - \hat{Q}_{n,c}^{-1} \right\} \left(n^{-1} \sum_{i=1}^n Z_{i,c} Y_i \right).\end{aligned}$$

By the Donsker's theorem for the paired bootstrap (Van der Vaart, 1998, Theorem 23.7), conditional on (Y_i, U_i, X_i) $(i = 1, 2, \ldots)$, for almost all sequence (Y_i, U_i, X_i) $(i = 1, 2, \ldots)$,

$$\left\{ \begin{array}{c} \frac{1}{\sqrt{n}} \sum_{i=1}^n Z_{i,c}^* Y_i^* - Z_{i,c} Y_i \\ \sqrt{n} \left(\hat{Q}_{n,c}^* - \hat{Q}_{n,c} \right) \end{array} : c \in [\ell, u] \right\} \rightsquigarrow \left\{ \begin{array}{c} G_{1,c} \\ G_{2,c} \end{array} : c \in [\ell, u] \right\} \quad \text{in } \ell^\infty(S).$$

Recall that $\phi(\cdot)$ is the "component-wise" matrix inverse map defined in (S1.5). By Lemma S1.4, there exists a linear map $\phi'_Q(\mathbf{H}) = \{-Q_c^{-1} H_c Q_c^{-1} : c \in [\ell, u]\}$, such that

$$\frac{\phi(\mathbf{D}_n + t_n \mathbf{H}_n) - \phi(\mathbf{D}_n)}{t_n} - \phi'_Q(\mathbf{H}) \to 0, \quad \text{in } \ell^\infty(S'),$$

for all sequence of $t_n \to 0$, $\mathbf{D_n} \to \{Q_c : c \in [\ell, u]\}$ in $\ell^\infty(S')$, $\mathbf{H}_n \to \mathbf{H} = \{H_c : c \in [\ell, u]\}$ in $\ell^\infty(S')$, such that $\mathbf{D_n}$ and $\mathbf{D_n} + t_n \mathbf{H_n} \in \mathcal{D}_\phi$ for every n. By the bootstrap version of the functional delta method (Van Der Vaart

and Wellner, 1996, Theorem 3.9.13), we obtain that in $\ell^\infty(S)$,

$$\left\{ \begin{array}{c} \frac{1}{\sqrt{n}} \sum_{i=1}^n \left(Z_{i,c}^* Y_i^* - Z_{i,c} Y_i \right) \\ \sqrt{n} \left\{ \left(\hat{Q}_{n,c}^* \right)^{-1} - \hat{Q}_{n,c}^{-1} \right\} \end{array} : c \in [\ell, u] \right\} \rightsquigarrow \left\{ \begin{array}{c} \mathbf{G}_1 \\ \phi_Q'(\mathbf{G}_2) \end{array} \right\}, \qquad \text{(S1.10)}$$

conditional on (Y_i, U_i, X_i) $(i = 1, 2, \ldots)$, for almost every sequence (Y_i, U_i, X_i) $(i = 1, 2, \ldots)$.

By the Glivenko-Cantelli theorem for paired bootstrap (Kosorok, 2007, Theorem 10.15) and the continuous mapping theorem (Van Der Vaart and Wellner, 1996, Theorem 1.3.6), for almost all sequence (Y_i, U_i, X_i) $(i = 1, 2, \ldots)$, in $\ell^\infty(S)$,

$$\left\{ \begin{array}{c} \left(\hat{Q}_{n,c}^* \right)^{-1} \\ \frac{1}{n} \sum_{i=1}^n Z_{i,c} Y_i \end{array} : c \in [\ell, u] \right\} \to \left\{ \begin{array}{c} Q_c^{-1} \\ P Z_{1,c} Y \end{array} : c \in [\ell, u] \right\}. \qquad \text{(S1.11)}$$

Therefore, combining (S1.10) and (S1.11), together with the Slutsky's theorem (Kosorok, 2007, Theorem 7.15), we obtain that in $\ell^\infty(S * S)$,

$$\left\{ \begin{array}{c} \frac{1}{\sqrt{n}} \sum_{i=1}^n Z_{i,c}^* Y_i^* - Z_{i,c} Y_i \\ \sqrt{n} \left\{ \left(\hat{Q}_{n,c}^* \right)^{-1} - \hat{Q}_{n,c}^{-1} \right\} \\ (\hat{Q}_{n,c}^*)^{-1} \\ \frac{1}{n} \sum_{i=1}^n Z_{i,c} Y_i \end{array} : c \in (l, u) \right\} \rightsquigarrow \left\{ \begin{array}{c} \mathbf{G}_1 \\ \phi_Q'(\mathbf{G}_2) \\ \{Q_c^{-1} : c \in [\ell, u]\} \\ \{P Z_{1,c} Y : c \in [\ell, u]\} \end{array} \right\},$$

conditional on (Y_i, U_i, X_i) $(i = 1, 2, \ldots)$, given almost all sequences (Y_i, U_i, X_i) $(i = 1, 2, \ldots)$. Finally, by the continuous mapping theorem (Van Der Vaart and

Wellner, 1996, Theorem 1.3.6), we have that, conditional on (Y_i, U_i, X_i) $(i = 1, 2, \ldots)$, given almost all sequence (Y_i, U_i, X_i) $(i = 1, 2, \ldots)$, the process $\{\sqrt{n}(\hat{\lambda}_c^* - \hat{\lambda}_c) : c \in [\ell, , u]\}$ converges weakly to $\{G_{3,c} : c \in [l, u]\}$ in $\ell^\infty([\ell, u])$.

Next we establish the second statement, i.e. (3.3), in Theorem 2. Define $\mathfrak{L} = \sup_{c\in[\ell,u]} |G_c|$, and denote by $F_{\mathfrak{L}}$ its distribution function (CDF). Further, for any CDF F, denote its quantile function by F^{-1}, i.e. $F^{-1}(q) = \inf\{x : F(x) \geq q\}$ for $q \in [0, 1]$. By the continuous mapping theorem (Van Der Vaart and Wellner, 1996, Theorem 1.3.6), we have, conditional on almost all sequence (Y_i, U_i, X_i) $(i = 1, 2, \ldots)$,

$$M_n \rightsquigarrow \mathfrak{L}, \quad M_n^* \rightsquigarrow \mathfrak{L}. \tag{S1.12}$$

As $\mathbf{G}$ is a tight, zero mean Gaussian process, $F_{\mathfrak{L}}$ is absolutely continuous on $(0, \infty)$, and $F_{\mathfrak{L}}^{-1}$ is continuous and strictly increasing on $(0, 1)$ (Ledoux and Talagrand, 1991; Davydov et al., 1998). The latter property, together with (Van der Vaart, 1998, Lemma 21.2), implies that conditional on almost all sequence (Y_i, U_i, X_i) $(i = 1, 2, \ldots)$,

$$(F_{n,pb}^*)^{-1}(\xi) \to F_{\mathfrak{L}}^{-1}(\xi), \quad \text{for each } \xi \in (0, 1). \tag{S1.13}$$

Now for a given significant level $\xi \in (0, 1)$, by the definition of $p_{n,pb}^*$ in

(3.2), we have

$$\{p^*_{n,pb} \le \xi\} = \{1 - F^*_{n,pb}(M_n) \le \xi\} = \{F^*_{n,pb}(M_n) \ge 1 - \xi\}$$
$$= \left\{M_n - (F^*_{n,pb})^{-1}(1-\xi) \ge 0\right\}.$$

By (S1.12) and (S1.13), $M_n - (F^*_{n,pb})^{-1}(1-\xi)$ converges weakly to $\mathfrak{L} - F_{\mathfrak{L}}^{-1}(1-\xi)$ as $n \to \infty$. Therefore, by the definition of weak convergence, as $F_{\mathfrak{L}}$ is continuous,

$$\mathrm{pr}(p^*_{n,pb} \le \xi) = \mathrm{pr}\left(M_n - (F^*_{n,pb})^{-1}(1-\xi) \ge 0\right) \to \mathrm{pr}(\mathfrak{L} - F_{\mathfrak{L}}^{-1}(1-\xi) \ge 0) = \xi,$$

as $n \to \infty$. This provides the asymptotic validity of the paired bootstrap adjusted p-values. □

S1.3 Proof of several lemmas

Lemma S1.1. *Let $\{b_i, 1 \le i \le 3\}$ and $\{d_i, 1 \le i \le 4\}$ be real numbers. Assume that $b_i \in (0,1)$ for $1 \le i \le 3$. Define*

$$A = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}, B = \begin{bmatrix} 1 & b_1 & b_2 & b_1b_2 \\ b_1 & b_1 & b_1b_2 & b_1b_2 \\ b_2 & b_1b_2 & b_2 & b_1b_2 \\ b_1b_2 & b_1b_2 & b_1b_2 & b_1b_2 \end{bmatrix}, C_k = \begin{bmatrix} 1 & b_1 & b_3 & b_1b_3 \\ b_1 & b_1 & b_1b_3 & b_1b_3 \\ b_2 & b_1b_2 & b_k & b_1b_k \\ b_1b_2 & b_1b_2 & b_1b_k & b_1b_k \end{bmatrix},$$

and $D = [d_1, d_2, d_3, d_4]^T$, where $k = 2, 3$. Then we have

$$A^TB^{-1}C_2D = d_4(1-b_3)/(1-b_2), \quad A^TB^{-1}C_3D = d_4b_3/b_2.$$

Proof. By elementary matrix multiplication. □

In the next lemma, we show some facts related to $\{Q_c : c \in [\ell, u]\}$ in (S1.1) needed in the proof of Theorems 1 and 2.

Lemma S1.2. *There exists a constant $\tau > 0$ such that for each $c \in [\ell, u]$, the smallest eigenvalue of Q_c is larger than τ, and the absolute value of each entry of Q_c^{-1}, is upper bounded by $1/\tau$.*

Proof. Recall the definition of Q_c in (S1.1), which takes the following form:

$$Q_c = E\left(Z_{1,c} Z_{1,c}^T\right) = \begin{bmatrix} 1 & p & F(c) & pF(c) \\ p & p & pF(c) & pF(c) \\ F(c) & pF(c) & F(c) & pF(c) \\ pF(c) & pF(c) & pF(c) & pF(c) \end{bmatrix},$$

Therefore Q_c is semi-positive definite for each $c \in [\ell, u]$. Denote $e_{1,c} \geq 0$ the smallest eigenvalue for Q_c. Clearly, it suffices to show that there exists $\tau > 0$, such that $e_{1,c} \geq \tau$ for each $c \in [\ell, u]$.

The determinant of Q_c is $p^2 F(c)^2 (F(c) - 1)^2 (p - 1)^2$. Since in Section 3 we assume $0 < F(\ell) < F(u) < 1$ and $0 < p < 1$, there exists $\tau' > 0$ such that

$$\det(Q_c) \geq \tau', \text{ for each } c \in [\ell, u].$$

Since the absolute value of every entry of Q_c is smaller or equal to 1, all of

its eigenvalues are smaller or equal to 4. As a result, $e_{1,c} \geq \tau'/4^3 = \tau > 0$ for each $c \in [\ell, u]$, which completes the proof. $\square$

Lemma S1.3. *For each integer $n = 1, 2, \ldots$ and $c \in [\ell, u]$, let $D_{n,c}$ be a 4×4 symmetric, deterministic matrix and assume that $\{D_{n,c} : c \in [\ell, u]\}$ converges to $\{Q_c : c \in [\ell, u]\}$ in $\ell^\infty(S')$ as $n \to \infty$. There exists $N > 0$ and $\tau > 0$ such that for each $c \in [\ell, u]$ and $n \geq N$, the smallest eigenvalue of $D_{n,c}$ is larger than τ, and the absolute value of each entry of $D_{n,c}^{-1}$, is upper bounded by $1/\tau$. Further,*

$$\{D_{n,c}^{-1} : c \in [\ell, u]\} \text{ converges to } \{Q_c^{-1} : c \in [\ell, u]\} \text{ in } \ell^\infty(S').$$

Proof. For the first statement, we use the variational characterization for the minimal eigenvalue of a symmetric matrix. Specifically, the minimum eigenvalue of $D_{n,c}$ equals to

$$\min_{||v||=1} v^T D_{n,c} v \geq \min_{||v||=1} v^T Q_c v + \min_{||v||=1} v^T \Delta_{n,c} v,$$

where $\Delta_{n,c} = D_{n,c} - Q_c$. By Lemma S1.2 there exists $\tau > 0$ such that $\min_{||v||=1} v^T Q_c v > 2\tau$ for each $c \in [\ell, u]$. Since $\{\Delta_{n,c} : c \in [\ell, u]\} \to 0$ in $\ell^\infty(S')$, there exists $N > 0$ such that for $n \geq N$,

$$\min_{||v||=1} v^T \Delta_{n,c} v \geq -\tau, \text{ for each } c \in [\ell, u],$$

and thus $\min_{||v||=1} v^T D_{n,c} v \geq \tau$ for each $c \in [\ell, u]$. Then the proof for the first statement is complete.

Now we prove the second statement. For large enough n, $D_{n,c}^{-1}$ is invertible for each $c \in [\ell, u]$. Then for large enough n and any $c \in [\ell, u]$, we have

$$D_{n,c}^{-1} - Q_c^{-1} = D_{n,c}^{-1}(D_{n,c} - Q_c)Q_c^{-1}.$$

Then the proof is complete since Q_c^{-1} and $D_{n,c}^{-1}$ is entry-wise uniformly (over $c \in [\ell, u]$) bounded. □

Recall that $\phi(\cdot)$ is the "component-wise" matrix inverse map defined in (S1.5) in Section S1.1.

In the following lemma, we show that $\phi(\cdot)$ satisfies the uniform Hadamard-differentiability (Van Der Vaart and Wellner, 1996, Equation 3.9.12), which is a sufficient condition to apply the functional delta method and the almost sure functional delta method for paired bootstrap method, which are required in the proofs of Theorems 1 and 2, respectively.

Recall that D_ϕ is defined in the proof of Theorem 1 in Subsection S1.1 (above equation (S1.5)).

Lemma S1.4. *The map ϕ satisfies the uniform Hadamard-differentiability in the sense that for the linear map $\phi'_Q(\mathbf{H}) = \{-Q_c^{-1}H_cQ_c^{-1} : c \in [\ell, u]\}$,*

$$\frac{\phi(\mathbf{D_n} + t_n\mathbf{H}_n) - \phi(\mathbf{D_n})}{t_n} \to \phi'_Q(\mathbf{H}), \quad \textit{as } n \to \infty, \quad \textit{in } \ell^\infty(S'),$$

for any $t_n \to 0$, $\mathbf{H_n} = \{H_{n,c} : c \in [\ell, u]\} \to \mathbf{H} = \{H_c : c \in [\ell, u]\}$ *in* $\ell^\infty(S')$, *and* $\mathbf{D_n} = \{D_{n,c} : c \in [\ell, u]\} \to \{Q_c : c \in [\ell, u]\}$ *in* $\ell^\infty(S')$, *where* $\mathbf{D_n}, \mathbf{D_n} + t_n\mathbf{H_n} \in \mathcal{D}_\phi$ *for all* n.

Proof. From Lemma S1.3, for large n, we have that $D_{n,c}$ and $D_{n,c} + t_n H_{n,c}$ are invertible for each $c \in [\ell, u]$, and that the absolute value of each entry of $D_{n,c}^{-1}$ and $(D_{n,c} + t_n H_{n,c})^{-1}$ is upper bounded, uniformly each $c \in [\ell, u]$. Using the Neumann's series (Neumann, 1867), we have

$$
\begin{aligned}
&(D_{n,c} + t_n H_{n,c})^{-1} = D_{n,c}^{-1}(I + t_n H_{n,c} D_{n,c}^{-1})^{-1} \\
=&D_{n,c}^{-1}\left[\sum_{k=0}^{\infty}(-1)^k t_n^k (H_{n,c}D_{n,c}^{-1})^k\right] \\
=&D_{n,c}^{-1} - t_n D_{n,c}^{-1} H_{n,c} D_{n,c}^{-1} + D_{n,c}^{-1}\sum_{k=2}^{\infty}(-1)^k t_n^k (H_{n,c}D_{n,c}^{-1})^k \\
=&D_{n,c}^{-1} - t_n D_{n,c}^{-1} H_{n,c} D_{n,c}^{-1} + t_n^2 (H_{n,c}D_{n,c}^{-1})^2 D_{n,c}^{-1}\left[\sum_{k=0}^{\infty}(-1)^k t_n^k (H_{n,c}D_{n,c}^{-1})^k\right] \\
=&D_{n,c}^{-1} - t_n D_{n,c}^{-1} H_{n,c} D_{n,c}^{-1} + t_n^2 (H_{n,c}D_{n,c}^{-1})^2 (D_{n,c} + t_n H_{n,c})^{-1} \qquad \text{(S1.14)} \\
=&D_{n,c}^{-1} - t_n D_{n,c}^{-1} H_{n,c} D_{n,c}^{-1} + O(t_n^2),
\end{aligned}
$$

where the third term of (S1.14) is $O(t_n^2)$ by Lemma S1.3, which is uniform over $c \in [\ell, u]$. Finally, again by Lemma S1.3 and the continuous mapping theorem, we have that in $\ell^\infty(S')$,

$$\{-D_{n,c}^{-1} H_{n,c} D_{n,c}^{-1} : c \in [\ell, u]\} \to \{-Q_c^{-1} H_c Q_c^{-1} : c \in [\ell, u]\} = \phi'_Q(\mathbf{H}).$$

Hence we finish the proof. □

S2 Proof of the Asymptotic Validity of p-values Generated from the Multiplier Residual Bootstrap and Discussion on the assumptions

S2.1 Proof of Theorem 3 under polynomial moments condition

In this section we discuss the asymptotic validity of the p-values generated from the proposed multiplier residual bootstrap under the fixed design, which is stated in Theorem 3.

Recall that $\hat{Q}_{n,c} = n^{-1}\sum_{i=1}^{n} Z_{i,c}Z_{i,c}^T$ is defined in (2.2). Due to Assumption (A.2), we have that as $n \to \infty$, $\{\hat{Q}_{n,c} : c \in [\ell, u]\}$ converges uniformly, component-wise, to $\{Q_c : c \in [\ell, u]\}$, where

$$Q_c = \begin{bmatrix} 1 & p & F(c) & pF(c) \\ p & p & pF(c) & pF(c) \\ F(c) & pF(c) & F(c) & pF(c) \\ pF(c) & pF(c) & pF(c) & pF(c) \end{bmatrix}, \tag{S2.15}$$

where $0 < F(\ell) \leq F(c) \leq F(u) < 1$ for $c \in [\ell, u]$ and $p \in (0, 1)$. Note that although the interpretation of $F(\cdot)$ and p in the fixed design differs from that in the random design, Lemmas S1.2 and S1.3 still apply. Thus, as our results are asymptotic, without loss of generality, we can assume that there

exists $\tau > 0$, such that for each $n \geq 4$ and $c \in [\ell, u]$,

$$\hat{Q}_{n,c} \text{ is invertible}, \quad \tau < |\hat{Q}_{n,c}| \leq 1, \quad \tau < \left|\hat{Q}_{n,c}^{-1}\right| < 1/\tau, \tag{S2.16}$$

where the inequalities and absolute values are interpreted component-wise.

Proof for the case $r < \infty$. Recall in Sections 2.2 and 4.1, for each $c \in [\ell, u]$, $\sqrt{n}\hat{\lambda}_c$ and $\sqrt{n}\hat{\lambda}_c^*$ can be respectively decomposed as

$$\begin{aligned}
\sqrt{n}\hat{\lambda}_c &= d^T \hat{Q}_{n,c}^{-1}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} Z_{i,c} Z_{i,c_0}^T \theta_0\right) + d^T \hat{Q}_{n,c}^{-1}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} Z_{i,c}\epsilon_i\right) \\
&= I_{n,c} + II_{n,c}, \\
\sqrt{n}\hat{\lambda}_c^* &= d^T \hat{Q}_{n,c}^{-1}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} Z_{i,c} Z_{i,\hat{c}_0}^T \hat{\theta}_0\right) + \hat{\sigma} d^T \hat{Q}_{n,c}^{-1}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} Z_{i,c}\zeta_i\right) \\
&= I_{n,c}^* + II_{n,c}^*,
\end{aligned} \tag{S2.17}$$

where $d = (0, 0, 0, 1)^T$, $\hat{\theta}_0 = (\hat{\alpha}_0, \hat{\beta}_0, \hat{\gamma}_0, 0)^T$, and $\zeta_1, \ldots, \zeta_n$ are independently and identically distributed standard normal random variables. Further, under H_0, $\lambda_0 = 0$.

As discussed in Remark 1, taking the supremum over $c \in [\ell, u]$ is equivalent to taking the supremum over $c \in \mathcal{C}_n = \{X_1, \ldots, X_n\} \cap [\ell, u]$. Conditional on $\epsilon_1, \ldots, \epsilon_n$, the distribution function $F_{n,mrb}^*$ of M_n^* is continuous

and strictly increasing on $[0,\infty)$, and we have

$$\sup_{\xi\in(0,1)}\left|\mathrm{pr}(p^*_{n,mrb}\le\xi)-\xi\right| = \sup_{\xi\in(0,1)}\left|\mathrm{pr}(M_n\ge(F^*_{n,mrb})^{-1}(1-\xi))-\xi\right|$$

$$= \sup_{\xi\in(0,1)}\left|\mathrm{pr}(M_n\ge(F^*_{n,mrb})^{-1}(1-\xi))-\mathrm{pr}_{|\epsilon}(M^*_n\ge(F^*_{n,mrb})^{-1}(1-\xi))\right|$$

$$\le \sup_{t\in(0,\infty)}\left|\mathrm{pr}\left(\sup_{c\in\mathcal{C}_n}\left|\frac{I_{n,c}+II_{n,c}}{\hat{v}_c}\right|\le t\right)-\mathrm{pr}\left(\sup_{c\in\mathcal{C}_n}\left|\frac{I_{n,c}+\frac{\sigma}{\hat{\sigma}}II^*_{n,c}}{\hat{v}_c}\right|\le t\right)\right| \tag{S2.18}$$

$$+\sup_{t\in(0,\infty)}\left|\mathrm{pr}\left(\sup_{c\in\mathcal{C}_n}\left|\frac{I_{n,c}+\frac{\sigma}{\hat{\sigma}}II^*_{n,c}}{\hat{v}_c}\right|\le t\right)-\mathrm{pr}_{|\epsilon}\left(\sup_{c\in\mathcal{C}_n}\left|\frac{I_{n,c}+II^*_{n,c}}{\hat{v}_c}\right|\le t\right)\right| \tag{S2.19}$$

$$+\sup_{t\in(0,\infty)}\left|\mathrm{pr}_{|\epsilon}\left(\sup_{c\in\mathcal{C}_n}\left|\frac{I_{n,c}+II^*_{n,c}}{\hat{v}_c}\right|\le t\right)-\mathrm{pr}_{|\epsilon}\left(\sup_{c\in\mathcal{C}_n}\left|\frac{I^*_{n,c}+II^*_{n,c}}{\hat{v}_c}\right|\le t\right)\right|, \tag{S2.20}$$

where $\mathrm{pr}_{|\epsilon}$ represents the conditional probability given $\epsilon_i,\ i\in[n]$, which is equivalent to the probability conditional on $Y_i,\ i\in[n]$, in the fixed design.

Let $q<\min\{1/6-1/(3r),1/3-4/(3r),\eta_1-1/2,1/2-1/r-\eta_0\}$ be arbitrary. Then it is sufficient to show that $n^q\times$(S2.18) $\to 0$ as $n\to\infty$, and that $n^q\times$(S2.19) and $n^q\times$(S2.20) converge to zero as $n\to\infty$ almost surely.

Upper bounding (S2.18). Recall some definitions in Section 4.1: $n_c\le n$ is the cardinality of $\mathcal{C}_n$, and $II_n=\{II_{n,c}:c\in\mathcal{C}_n\}$, $II^*_n=\{II^*_{n,c}:c\in\mathcal{C}_n\}$ are two n_c-dimensional random vectors. $\mathcal{A}^{re}$ is the collection of all hyper-

rectangles in $\mathbb{R}^{n_c}$; that is, $\mathcal{A}^{re}$ consists of all sets A of the form

$$A = \{x \in \mathbb{R}^{n_c} : t_i \leq x_i \leq s_i, \text{ for all } 1 \leq i \leq n_c\} = \bigotimes_{1 \leq i \leq n_c} [t_i, s_i],$$

for some $-\infty \leq t_i \leq s_i \leq \infty$ for $1 \leq i \leq n_c$. Then

$$\begin{aligned}
\text{(S2.18)} = \sup_{t \in (0,\infty)} \Bigg| & \text{pr}\left(II_n \in \bigotimes_{c \in \mathcal{C}_n} [-\hat{v}_c t - I_{n,c}, \hat{v}_c t - I_{n,c}] \right) \\
& - \text{pr}\left(\frac{\sigma}{\hat{\sigma}} II_n^* \in \bigotimes_{c \in \mathcal{C}_n} [-\hat{v}_c t - I_{n,c}, \hat{v}_c t - I_{n,c}] \right) \Bigg| \\
\leq \sup_{A \in \mathcal{A}^{re}} & \left| \text{pr}\left(II_n \in A \right) - \text{pr}\left(\frac{\sigma}{\hat{\sigma}} II_n^* \in A \right) \right|. \qquad \text{(S2.21)}
\end{aligned}$$

The random vector $\sigma II_n^*/\hat{\sigma}$ is a zero mean Gaussian vector of length n_c, with the same mean and co-variance structure as II_n; we apply the high-dimensional central limit theorem (Chernozhukov et al., 2017, Proposition 2.1). Specifically, due to (S2.16), Assumption (A.1), and the fact that each component of $Z_{i,c} \in [0,1]$, there exits some constant $K > 0$, that only depends on r, τ and $E(|\epsilon_1|^r)$, such that for each $n \geq 4$ and $c \in \mathcal{C}_n$,

$$\begin{aligned}
& \frac{1}{n} \sum_{i=1}^{n} E\left[\left(d^T Q_{n,c}^{-1} Z_{i,c} \epsilon_i \right)^2 \right] = \frac{E[\epsilon_1^2]}{n} \sum_{i=1}^{n} d^T \hat{Q}_{n,c}^{-1} Z_{i,c} Z_{i,c}^T \hat{Q}_{n,c}^{-1} d = \sigma^2 d^T \hat{Q}_{n,c}^{-1} d \in [\tau\sigma^2, \sigma^2/\tau], \\
& \frac{1}{n} \sum_{i=1}^{n} E\left(\left| d^T \hat{Q}_{n,c}^{-1} Z_{i,c} \epsilon_i \right|^3 \right) = \frac{1}{n} \sum_{i=1}^{n} \left| d^T \hat{Q}_{n,c}^{-1} Z_{i,c} \right|^3 E(|\epsilon_1|^3) \leq K, \\
& \frac{1}{n} \sum_{i=1}^{n} E\left(\left| d^T \hat{Q}_{n,c}^{-1} Z_{i,c} \epsilon_i \right|^4 \right) = \frac{1}{n} \sum_{i=1}^{n} \left| d^T \hat{Q}_{n,c}^{-1} Z_{i,c} \right|^4 E(|\epsilon_1|^4) \leq K, \\
& E\left(\max_{c \in \mathcal{C}_n} \left(\left| d^T \hat{Q}_{n,c}^{-1} Z_{i,c} \epsilon_i \right|^r / B_n^r \right) \right) \leq 2, \quad i \in [n],
\end{aligned}$$

where $B_n = Kn^{1/r}$. Therefore by (Chernozhukov et al., 2017, Proposition 2.1), $n^q \times$ (S2.18) $\to 0$.

Upper bounding (S2.19). Similar to (S2.21), we have

$$\text{(S2.19)} \leq \sup_{A \in \mathcal{A}^{re}} \left| \text{pr}_{|\epsilon} \left(II_n^* \in A \right) - \text{pr} \left(\frac{\sigma}{\hat{\sigma}} II_n^* \in A \right) \right|.$$

The $\sigma/\hat{\sigma} II_n^*$ is a centered Gaussian random vectors of length n_c with the covariance matrix $\sigma^2 \Sigma_{n_c}$, where $\Sigma_{n_c}(c_1, c_2) = n^{-1} \sum_{i=1}^n d^T \hat{Q}_{n,c_1}^{-1} Z_{i,c_1} Z_{i,c_2}^T \hat{Q}_{n,c_2}^{-1} d$ for $c_1, c_2 \in \mathcal{C}_n$. Due to (S2.16),

$$\begin{aligned} &|\Sigma_{n_c}(c_1, c_2)| \leq 16/\tau^2, \quad \text{for } c_1, c_2 \in \mathcal{C}_n, \\ &\Sigma_{n_c}(c, c) \geq \tau \sigma^2, \quad \text{for } c \in \mathcal{C}_n. \end{aligned} \tag{S2.22}$$

Further, as $\zeta_1, \ldots, \zeta_n$ are independent and identically distributed standard normal random variables, conditional on the data $\epsilon_1, \ldots, \epsilon_n$, II_n^* is a centered Gaussian random vector of length n_c with the covariance matrices $\hat{\sigma}^2 \Sigma_{n_c}$. Then by the Gaussian comparison lemma (Chernozhukov et al., 2019, Corollary 5.1), there exists a constant $K > 0$, that only depends on τ and σ^2, such that

$$\text{(S2.19)} \leq K \|\hat{\sigma}^2 \Sigma_{n_c} - \sigma^2 \Sigma_{n_c}\|_\infty^{1/2} \log(n) \leq K \sqrt{\frac{16}{\tau^2}} \left| \hat{\sigma}^2 - \sigma^2 \right|^{1/2} \log(n).$$

where $\| \cdot \|_\infty$ denotes the ℓ^∞ norm of a matrix, and the last inequality is due to (S2.22). By Lemma S2.3, $n^{q_1} |\hat{\sigma}^2 - \sigma^2| \to 0$ almost surely for any $q_1 < 1/2 - 1/r$. Thus $n^q \times$ (S2.19) $\to 0$ converges to zero almost surely

since $r > 4$.

Upper bounding (S2.20). For the last term (S2.20), we have

$$
\begin{aligned}
(\text{S2.20}) &= \sup_{t\in(0,\infty)} \left| \mathrm{pr}_{|\epsilon}\left(\sup_{c\in\mathcal{C}_n} \left| \frac{I_{n,c} + II^*_{n,c}}{\hat{v}_c} \right| \le t \right) - \mathrm{pr}_{|\epsilon}\left(\sup_{c\in\mathcal{C}_n} \left| \frac{I^*_{n,c} + II^*_{n,c}}{\hat{v}_c} \right| \le t \right) \right| \\
&\le \sup_{t\in(0,\infty)} \Bigg| \mathrm{pr}_{|\epsilon}(II^*_n \in \bigotimes_{c\in\mathcal{C}_n} [-I_{n,c} - t\hat{v}_c, -I_{n,c} + t\hat{v}_c]) \\
&\qquad\qquad -\mathrm{pr}_{|\epsilon}(II^*_n \in \bigotimes_{c\in\mathcal{C}_n} [-I^*_{n,c} - t\hat{v}_c, -I^*_{n,c} + t\hat{v}_c]) \Bigg| \\
&\le \sup_{s_1,s_2\in\mathbb{R}^{n_c}} \left(\mathrm{pr}_{|\epsilon}(-s_2 - \Delta_n \le II^*_n \le s_1 + \Delta_n) - \mathrm{pr}_{|\epsilon}(-s_2 \le II^*_n \le s_1) \right) \\
&= \sup_{s_1,s_2\in\mathbb{R}^{n_c}} \left(\mathrm{pr}_{|\epsilon}\left(II^*_n \le s_1 + \Delta_n, -II^*_n \le s_2 + \Delta_n\right) - \mathrm{pr}_{|\epsilon}\left(II^*_n \le s_1 \ , \ -II^*_n \le s_2\right) \right), \\
&= \sup_{s_1,s_2\in\mathbb{R}^{n_c}} \Bigg(\mathrm{pr}_{|\epsilon}\left(\frac{\sigma}{\hat{\sigma}} II^*_n \le s_1 + \frac{\sigma}{\hat{\sigma}}\Delta_n, -\frac{\sigma}{\hat{\sigma}} II^*_n \le s_2 + \frac{\sigma}{\hat{\sigma}}\Delta_n \right) \\
&\qquad\qquad -\mathrm{pr}_{|\epsilon}\left(\frac{\sigma}{\hat{\sigma}} II^*_n \le s_1 \ , \ -\frac{\sigma}{\hat{\sigma}} II^*_n \le s_2 \right) \Bigg),
\end{aligned}
$$

where $\Delta_n = \sup_{c\in\mathcal{C}_n} \left| I_{n,c} - I^*_{n,c} \right|$ and the inequalities and the scalar addition involving vectors are interpreted component-wise.

Due to (S2.22), and by applying the Nazarov's inequality (Chernozhukov et al., 2017, Lemma A.1) (and also Nazarov (2003)) to the Gaussian vector $\sigma/\hat{\sigma}(-II^*_n, II^*_n)$ of length $2n_c$, there exists a constant $K > 0$, that only depends on τ, σ^2, such that

$$
(\text{S2.20}) \ \le \ K \frac{\sigma^2}{\hat{\sigma}^2} \Delta_n \sqrt{\log(2n)}.
$$

Due to Lemma S2.1, for any $q_1 < \min\{\eta_1 - 1/2, 1/2 - 1/r - \eta_0\}$, $n^{q_1}\Delta_n \to 0$

almost surely, which, together with Lemma S2.3, implies that $n^q \times$(S2.20) $\to$ 0 as $n \to \infty$ almost surely. $\square$

For $c_1, c_2 \in [\ell, u]$, denote

$$Q_{c_1,c_2} = \begin{bmatrix} 1 & p & F(c_2) & pF(c_2) \\ p & p & pF(c_2) & pF(c_2) \\ F(c_1) & pF(c_1) & F(c_1 \wedge c_2) & pF(c_1 \wedge c_2) \\ pF(c_1) & pF(c_1) & pF(c_1 \wedge c_2) & pF(c_1 \wedge c_2) \end{bmatrix}$$

where $c_1 \wedge c_2 = \min\{c_1, c_2\}$. Clearly $Q_c = Q_{c,c}$ from the definition of Q_c in (S2.15).

Lemma S2.1. *Suppose the assumptions in Theorem 3 hold with $r < \infty$. Recall that $I_{n,c}$ and $I^*_{n,c}$ are defined in* (S2.17)*. For any $q < \min\{\eta_1 - 1/2, 1/2 - 1/r - \eta_0\}$, almost surely,*

$$n^q \Delta_n \to 0 \quad \textit{as} \;\; n \to \infty, \qquad \textit{where} \;\; \Delta_n = \sup_{c \in \mathcal{C}_n} \left| I_{n,c} - I^*_{n,c} \right|.$$

Proof. By the assumption, H_0 holds, i.e., $\lambda_0 = 0$. We will assume that $\gamma_0 \neq 0$, i.e., c_0 is identified in Model (2.1), and both Assumptions (A.2) and (A.3) hold. We indicate the minor modifications in a remark after this proof that are needed for the case that $\gamma_0 = 0$.

Recall the definition of $\hat{Q}_{n,c}$ in (2.2). Δ_n can be upper bounded by

$$\begin{aligned}
\sup_{c\in\mathcal{C}_n}\left|I_{n,c}-I^*_{n,c}\right| &= \sup_{c\in\mathcal{C}_n}\left|d^T\hat{Q}_{n,c}^{-1}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n Z_{i,c}(Z_{i,c_0}^T\theta_0 - Z_{i,\hat{c}_0}^T\hat{\theta}_0)\right)\right| \\
&\le \sup_{c\in\mathcal{C}_n}\left|d^TQ_c^{-1}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n Z_{i,c}(Z_{i,c_0}^T\theta_0 - Z_{i,\hat{c}_0}^T\hat{\theta}_0)\right)\right| \\
&+ \sup_{c\in\mathcal{C}_n}\left|d^T\sqrt{n}\left(\hat{Q}_{n,c}^{-1}-Q_c^{-1}\right)\left(\frac{1}{n}\sum_{i=1}^n Z_{i,c}(Z_{i,c_0}^T\theta_0 - Z_{i,\hat{c}_0}^T\hat{\theta}_0)\right)\right| \\
&= O_{1,n}+O_{2,n}.
\end{aligned}$$

Then it suffices to show that both $n^qO_{1,n}$ and $n^qO_{2,n}$ converge to zero almost surely.

Upper bounding $O_{1,n}$. The term $O_{1,n}$ can be upper bounded by

$$\begin{aligned}
O_{1,n} &= \sup_{c\in\mathcal{C}_n}\left|d^TQ_c^{-1}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n Z_{i,c}\left(Z_{i,c_0}^T\theta_0 - Z_{i,\hat{c}_0}^T\theta_0 + Z_{i,\hat{c}_0}^T\theta_0 - Z_{i,\hat{c}_0}^T\hat{\theta}_0\right)\right)\right| \\
&= \sup_{c\in\mathcal{C}_n}\left|d^TQ_c^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^n Z_{i,c}\left((Z_{i,c_0}-Z_{i,\hat{c}_0})^T\theta_0\right) + d^TQ_c^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^n Z_{i,c}Z_{i,\hat{c}_0}^T(\theta_0-\hat{\theta}_0)\right| \\
&\le \sup_{c\in\mathcal{C}_n}\left|d^TQ_c^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^n\left(Z_{i,c}(Z_{i,c_0}-Z_{i,\hat{c}_0})^T\theta_0 - (Q_{c,c_0}-Q_{c,\hat{c}_0})\theta_0\right)\right| \qquad \text{(S2.23)} \\
&+ \sup_{c\in\mathcal{C}_n}\left|d^TQ_c^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^n\left(Z_{i,c}Z_{i,\hat{c}_0}^T - Q_{c,\hat{c}_0}\right)(\theta_0-\hat{\theta}_0)\right| \qquad \text{(S2.24)} \\
&+ \sup_{c\in\mathcal{C}_n}\left|d^TQ_c^{-1}\sqrt{n}Q_{c,c_0}\theta_0 - d^TQ_c^{-1}\sqrt{n}Q_{c,\hat{c}_0}\hat{\theta}_0\right|. \qquad \text{(S2.25)}
\end{aligned}$$

Note that (S2.25) is equal to zero due to Lemma S1.1, and the facts that $\hat{\theta}_0 = (\hat{\alpha}_0, \hat{\beta}_0, \hat{\gamma}_0, 0)^T$ and $\lambda_0 = 0$ under H_0. Next we deal with terms (S2.23)

and (S2.24).

For the term (S2.23), by considering three cases: $c > \max\{c_0, \hat{c}_0\}$, $c < \min\{c_0, \hat{c}_0\}$, and $|c - c_0| \leq |\hat{c}_0 - c_0|$, we have

$$\sup_{c \in \mathcal{C}_n} \left\| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (Z_{i,c}((Z_{i,c_0} - Z_{i,\hat{c}_0})^T - (Q_{c,c_0} - Q_{c,\hat{c}_0}) \right\|_\infty \leq \tag{S2.26}$$

$$\sup_{c:|c-c_0| \leq |\hat{c}_0 - c_0|} \sqrt{n}(|s_{n,c} - s_{n,c_0} - (F(c) - F(c_0))| + |q_{n,c} - q_{n,c_0} - p(F(c) - F(c_0))|),$$

where $\|\cdot\|_\infty$ denotes the ℓ^∞ norm of a matrix. From Lemma S2.3, $n^{\eta_1}|\hat{c}_0 - c_0|$ converges to zero almost surely, then due to Assumption (A.3) and Lemma S1.2, we have

$$n^{\eta_1 - \frac{1}{2}} \times \text{(S2.23)} \to 0, \quad \text{as } n \to \infty.$$

Now we consider (S2.24). From Lemma S2.3, $n^{q_1}\|\theta_0 - \hat{\theta}_0\|$ converges to 0 almost surely for any $q_1 < 1/2 - 1/r$, where $\|\cdot\|$ denotes the Euclidean norm (or equivalently operator norm). Further,

$$\begin{aligned} & n^{-\eta_0} \sup_{c \in \mathcal{C}_n} \left\| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (Z_{i,c} Z_{i,\hat{c}_0}^T - Q_{c,\hat{c}_0}) \right\|_\infty \\ & \leq n^{1/2 - \eta_0}(|p_n - p| + \sup_{c \in \mathcal{C}_n}(|s_{n,c} - F(c)| + |q_{n,c} - pF(c)|)), \end{aligned}$$

converges to zero by Assumption (A.2), which, together with (S2.16), implies that for any $q_1 < 1/2 - 1/r$, almost surely

$$n^{q_1 - \eta_0} \times \text{(S2.24)} \to 0, \quad \text{as } n \to \infty.$$

Combing above two results, for any $q < \min\{\eta_1 - 1/2, 1/2 - 1/r - \eta_0\}$, we have almost surely

$$n^q O_{1,n} \to 0, \quad \text{as } n \to \infty. \tag{S2.27}$$

Upper bounding $O_{2,n}$. The term $O_{2,n}$ can be decomposed in the same way as $O_{1,n}$:

$$\begin{aligned} O_{2,n} &= \sup_{c\in\mathcal{C}_n} \left| d^T \sqrt{n} \left(\hat{Q}_{n,c}^{-1} - Q_c^{-1} \right) \left(\frac{1}{n} \sum_{i=1}^n (Z_{i,c}(Z_{i,c_0}^T \theta_0 - Z_{i,\hat{c}_0}^T \hat{\theta}_0)) \right) \right| \\ &\le \sup_{c\in\mathcal{C}_n} \left| d^T \sqrt{n} \left(\hat{Q}_{n,c}^{-1} - Q_c^{-1} \right) \left(\frac{1}{n} \sum_{i=1}^n Z_{i,c}(Z_{i,c_0}^T - Z_{i,\hat{c}_0}^T)\theta_0 \right) \right| \qquad \text{(S2.28)} \\ &+ \sup_{c\in\mathcal{C}_n} \left| d^T \sqrt{n} \left(\hat{Q}_{n,c}^{-1} - Q_c^{-1} \right) \left(\frac{1}{n} \sum_{i=1}^n Z_{i,c} Z_{i,\hat{c}_0}^T \right) (\theta_0 - \hat{\theta}_0) \right|. \qquad \text{(S2.29)} \end{aligned}$$

For a matrix (or vector) A, denote $\|A\|$ its operator norm. Then by Assumption (A.2) and (S2.16),

$$\begin{aligned} & n^{-\eta_0} \sup_{c\in\mathcal{C}_n} \left\| d^T \sqrt{n} \left(\hat{Q}_{n,c}^{-1} - Q_c^{-1} \right) \right\| = \sup_{c\in\mathcal{C}_n} \left\| \hat{Q}_{n,c}^{-1} n^{1/2-\eta_0} (\hat{Q}_{n,c} - Q_c) Q_c^{-1} \right\| \\ \le & \left(\sup_{c\in\mathcal{C}_n} \left\| \hat{Q}_{n,c}^{-1} \right\| \right) \left(n^{1/2-\eta_0} \sup_{c\in\mathcal{C}_n} \left\| \hat{Q}_{n,c} - Q_c \right\| \right) \left(\sup_{c\in\mathcal{C}_n} \|Q_c\| \right) \\ \lesssim & n^{1/2-\eta_0} (|p_n - p| + \sup_{c\in\mathcal{C}_n} (|s_{n,c} - F(c)| + |q_{n,c} - pF(c)|)) \to 0, \qquad \text{(S2.30)} \end{aligned}$$

as $n \to \infty$, where $\lesssim$ here means up to a constant that only depends on τ in (S2.16).

Now we consider (S2.28). It is clear that

$$\sup_{c\in\mathcal{C}_n}\left\|\frac{1}{n}\sum_{i=1}^{n}Z_{i,c}(Z_{i,c_0}^T-Z_{i,\hat{c}_0}^T)\right\|_\infty$$
$$\leq\sup_{c\in\mathcal{C}_n}\|Q_{c,c_0}-Q_{c,\hat{c}_0}\|_\infty+\sup_{c\in\mathcal{C}_n}\left\|\frac{1}{n}\sum_{i=1}^{n}Z_{i,c}(Z_{i,c_0}^T-Z_{i,\hat{c}_0}^T)-(Q_{c,c_0}-Q_{c,\hat{c}_0})\right\|_\infty.$$

By Lemma S2.3, $n^{\eta_1}|\hat{c}_0-c_0|$ converges to zero almost surely. Since $F(\cdot)$ is differentiable at c_0 (Assumption (A.3)), we have that almost surely, $n^{\eta_1}\sup_{c\in\mathcal{C}_n}\|Q_{c,c_0}-Q_{c,\hat{c}_0}\|_\infty\ \to\ 0$. Further, in (S2.26), we showed that $n^{\eta_1}\sup_{c\in\mathcal{C}_n}\left\|\frac{1}{n}\sum_{i=1}^{n}Z_{i,c}(Z_{i,c_0}^T-Z_{i,\hat{c}_0}^T)-(Q_{c,c_0}-Q_{c,\hat{c}_0})\right\|_\infty\to 0$. Thus almost surely,

$$n^{\eta_1}\sup_{c\in\mathcal{C}_n}\left\|\frac{1}{n}\sum_{i=1}^{n}Z_{i,c}(Z_{i,c_0}^T-Z_{i,\hat{c}_0}^T)\right\|_\infty\to 0,\quad\text{as } n\to\infty, \tag{S2.31}$$

Combining this with (S2.30), we have that almost surely,

$$n^{\eta_1-\eta_0}\ \times\ (\text{S2.28})\ \to\ 0,\quad\text{as } n\to\infty.$$

Now consider the term (S2.29). All entries of $n^{-1}\sum_{i=1}^{n}Z_{i,c}Z_{i,\hat{c}_0}^T$ are in $[0,1]$. By Lemma S2.3, $n^{q_1}\|\hat{\theta}_0-\theta_0\|\to 0$ almost surely for any $q_1<1/2-1/r$. Thus, together with (S2.30), we conclude that for any $q_1<1/2-1/r$, almost surely

$$n^{q_1-\eta_0}\ \times\ (\text{S2.29})\ \to\ 0,\quad\text{as } n\to\infty.$$

Combing the bound for (S2.28) and (S2.29), for any $q<\min\{\eta_1-$

$1/2, 1/2 - 1/r - \eta_0\}$, we have almost surely

$$n^q \times O_{2,n} \to 0, \quad \text{as } n \to \infty. \tag{S2.32}$$

Finally, combining equations (S2.27) and (S2.32), we finish the proof.

□

Remark S2.1. If $\gamma_0 = \lambda_0 = 0$, the almost sure convergence rate of the least squares estimators $\hat{\theta}_0$ still holds but the consistency of $\hat{c}_0$, which is required to show that (S2.23) and (S2.28) converge to 0, does not hold anymore (Lemma S2.3). However, (S2.23) and (S2.28) identically equal 0 when $\gamma_0 = \lambda_0 = 0$, for all $n = 1, 2, \ldots$, which indicates that $n^q \Delta_n$ converges to zero almost surely as well due to the fact that $q < 1/2 - 1/r - \eta_0$.

S2.2 Proof of Theorem 3 under the sub-Gaussian condition

In this subsection, we focus on the case $r = \infty$ in Theorem 3. The proof is almost identical to the proof for the case $r < \infty$ in the previous Subsection S2.1, with the essential difference being that we replace the high-dimensional central limit theorem (Chernozhukov et al., 2017, Proposition 2.1) by (Chernozhukov et al., 2019, Theorem 2.1), which has improved rates of convergence.

Proof for the case $r = \infty$. Let $q < \min\{1/4, \eta_1 - 1/2, 1/2 - \eta_0\}$ be arbitrary. Recall that in Subsection S2.1, we showed that

$$\sup_{\xi \in (0,1)} \left| \mathrm{pr}(p^*_{n,mrb} \leq \xi) - \xi \right| \ \leq \ (\text{S2.18}) + (\text{S2.19}) + (\text{S2.20}).$$

Since $r = \infty$, i.e., ϵ_1 has a sub-Gaussian tail with parameter $\rho > 0$, for any $r_1 \in (4, \infty)$, $E(|\epsilon_1|^{r_1}) < \infty$. As a result, by the same argument as in the proof for the $r < \infty$ case, and by letting $r_1 \to \infty$, we have

$$n^q \left((\text{S2.19}) + (\text{S2.20}) \right) \to 0, \quad \text{almost surely.}$$

Now in (S2.21), we showed

$$(\text{S2.18}) \leq \sup_{A \in \mathcal{A}^{re}} \left| \mathrm{pr}\left(II_n \in A \right) - \mathrm{pr}\left(\frac{\sigma}{\hat{\sigma}} II^*_n \in A \right) \right|.$$

Due to (S2.16), Assumption (A.1) with $r = \infty$, and the fact that each component of $Z_{i,c} \in [0,1]$, there exits some constant $K > 0$, that only depends on τ and ρ, such that for each $n \geq 4$ and $c \in \mathcal{C}_n$,

$$\begin{aligned}
&\frac{1}{n} \sum_{i=1}^{n} E\left[\left(d^T Q^{-1}_{n,c} Z_{i,c} \epsilon_i \right)^2 \right] = \frac{E[\epsilon_1^2]}{n} \sum_{i=1}^{n} d^T \hat{Q}^{-1}_{n,c} Z_{i,c} Z^T_{i,c} \hat{Q}^{-1}_{n,c} d = \sigma^2 d^T \hat{Q}^{-1}_{n,c} d \in [\tau\sigma^2, \sigma^2/\tau], \\
&\frac{1}{n} \sum_{i=1}^{n} E\left(\left| d^T \hat{Q}^{-1}_{n,c} Z_{i,c} \epsilon_i \right|^4 \right) = \frac{1}{n} \sum_{i=1}^{n} \left| d^T \hat{Q}^{-1}_{n,c} Z_{i,c} \right|^4 E(|\epsilon_1|^4) \leq K E(|\epsilon_1|^4), \\
&E\left(\exp\left(\left| d^T \hat{Q}^{-1}_{n,c} Z_{i,c} \epsilon_i \right|^2 / K^2 \right) \right) \leq 2, \quad i \in [n]. \qquad (\text{S2.33})
\end{aligned}$$

Then by (Chernozhukov et al., 2019, Theorem 2.1), there exists some

constant $K > 0$, that only depends on τ, σ^2, ρ, such that

$$\text{(S2.18)} \leq \sup_{A \in \mathcal{A}^{re}} \left| \text{pr}\left(II_n \in A\right) - \text{pr}\left(\frac{\sigma}{\hat{\sigma}} II_n^* \in A\right) \right| \leq K \left(\frac{\log^5(n^2)}{n} \right)^{1/4}.$$

Thus, $n^q \times$ (S2.18) $\to 0$ as $n \to \infty$. □

S2.3 Almost sure convergence of the profile least squares estimators

In this subsection we discuss the strong consistency of the profile least squares estimators $\hat{c}_0$ defined in (4.1) and the corresponding least squares regression parameters estimator $\hat{\theta}_{\hat{c}_0}$ in (2.2). It is a key result for removing the bias part $\{I_{n,c} : c \in [\ell, u]\}$ under the fixed design. For this purpose, we state and prove a general theorem to find an upper bound on the rate of the almost sure convergence for M-estimators, which is in the same spirit as the rate of convergence in probability (Van der Vaart, 1998, Theorem 3.2.5).

For a random variable V and $r \in (0, \infty)$, denote by $\|V\|_r = E\left[|V|^r\right]^{1/r}$ its L_r norm. Let $\{\mathbf{D}_n(g) : g \in \mathcal{G}\}$ be a sequence of stochastic processes each indexed by a space $\mathcal{G}$ equipped with a semimetric d_n, and $D_n = \{D_n(g) : g \in \mathcal{G}\}$ a sequence of deterministic functions indexed by $\mathcal{G}$. For each $n = 1, 2, \ldots$ let $\hat{g}_n$ and g_0 be the point of maximum of the "criterion function" $g \mapsto \mathbf{D}_n(g)$

and the centering function $g \mapsto D_n(g)$, respectively, i.e.,

$$\hat{g}_n = \operatorname*{argmax}_{g \in \mathcal{G}} \mathbf{D}_n(g), \qquad g_0 = \operatorname*{argmax}_{g \in \mathcal{G}} D_n(g).$$

Lemma S2.2. *Let $r \in (2, \infty)$. Suppose that for each $g \in \mathcal{G}$:*

$$D_n(g) - D_n(g_0) \leq -d_n^2(g, g_0), \tag{S2.34}$$

and that there exist positive constants K and a such that for every $n = 1, 2, \ldots$ and $\delta \geq n^{-1/2}$,

$$\left\| \sup_{d_n(g,g_0) \leq \delta} \sqrt{n} \left| (\mathbf{D}_n - D_n)(g) - (\mathbf{D}_n - D_n)(g_0) \right| \right\|_r \leq K \delta^{1-2/r} (\log(\delta^{-1} + \delta))^a, \tag{S2.35}$$

Then $n^q d_n(\hat{g}_n, g_0)$ converges to zero almost surely for any $q \in (0, 1/2 - 1/r)$.

Proof. Fix $q \in (0, 1/2 - 1/r)$. Let $t > 0$ be arbitrary. For each $n, j \in \{1, 2, \ldots\}$, define

$$S_{j,n} = \{g \in \mathcal{G} : 2^{j-1} t < n^q d_n(g, g_0) \leq 2^j t\}.$$

By definition,

$$\{n^q d_n(\hat{g}_n, g_0) > t\} = \bigcup_{j \geq 1} \{\hat{g}_n \in S_{j,n}\} \subset \bigcup_{j \geq 1} \left\{ \sup_{g \in S_{j,n}} (\mathbf{D}_n(g) - \mathbf{D}_n(g_0)) \geq 0 \right\}.$$

As a result, we have

$$\operatorname{pr}(n^q d_n(\hat{g}_n, g_0) > t) \leq \sum_{j \geq 1} \operatorname{pr}\left(\sup_{g \in S_{j,n}} (\mathbf{D}_n(g) - \mathbf{D}_n(g_0)) \geq 0 \right). \tag{S2.36}$$

For $j \geq 1$ and any $g \in S_{j,n}$,

$$D_n(g) - D_n(g_0) \leq -d_n^2(g, g_0) \leq -t^2 2^{2j-2} n^{-2q},$$

and thus the event $\left\{\sup_{g \in S_{j,n}} (\mathbf{D}_n(g) - \mathbf{D}_n(g_0)) \geq 0\right\}$ implies that:

$$\sup_{g \in S_{j,n}} |(\mathbf{D}_n(g) - \mathbf{D}_n(g_0)) - (D_n(g) - D_n(g_0))| \geq t^2 2^{2j-2} n^{-2q}.$$

Note that $q < 1/2 - 1/r$, i.e., $r(1/2 - q) > 1$. By the Markov's inequality, the series in (S2.36) is bounded by

$$\begin{aligned}
&\sum_{j \geq 1} \mathrm{pr}\left(\sup_{g \in S_{j,n}} |(\mathbf{D}_n(g) - \mathbf{D}_n(g_0)) - (D_n(g) - D_n(g_0))| \geq t^2 2^{2j-2} n^{-2q}\right) \\
\leq &\sum_{j \geq 1} \frac{E\left(\sup_{d_n(g,g_0) \leq t 2^j / n^q} \sqrt{n}\, |(\mathbf{D}_n(g) - \mathbf{D}_n(g_0)) - (D_n(g) - D_n(g_0))|^r\right)}{t^{2r} n^{r/2} 2^{2jr-2r} n^{-2rq}} \\
\leq &\sum_{j \geq 1} \frac{K (t 2^j n^{-q})^r (\log^{ar}(2^j n^q (t + t^{-1}))}{t^{2r} n^{r/2} 2^{2jr-2r} n^{-2rq}} \qquad \text{by (S2.35)} \\
\leq &\frac{K n^{rq} \log^{ar}(n^q (t + t^{-1}))}{t^r n^{r/2}} \sum_{j \geq 1} j^{ar} 2^{2r-rj} \leq \frac{K \log^{ar}(n^q (t + t^{-1})}{t^r n^{r/2 - rq}},
\end{aligned}$$

where the constant K above may depend on a, r, t, and may vary from line to line. Since $q < 1/2 - 1/r$, i.e., $r(1/2 - q) > 1$, we have

$$\sum_{n=1}^{\infty} \mathrm{pr}(n^q d_n(\hat{g}_n, g_0) > t) \leq \sum_{i=1}^{\infty} \frac{K \log^{ar}(n^q (t + t^{-1})}{t^r n^{r/2 - rq}} < \infty.$$

Then the proof is complete by the Borel–Cantelli lemma. □

Recall that $\hat{\theta}_{\hat{c}_0}$ is the least squares estimator (2.2) associated with the profile least squares estimator $\hat{c}_0$ in (4.1). For a vector, $\|\cdot\|$ denotes its Euclidean norm.

Lemma S2.3. *Let $r \in (4, \infty)$. Suppose that Assumptions* (A.1) *and* (A.2) *hold. For any $q_1 < 1/2 - 1/r$,*

$$n^{q_1} \left\| \hat{\theta}_{\hat{c}_0} - \theta_0 \right\| \to 0, \quad \textit{almost surely,} \tag{S2.37}$$

$$n^{q_1} |\hat{\sigma}^2 - \sigma^2| \to 0, \quad \textit{almost surely.} \tag{S2.38}$$

Furthermore, if either $\gamma_0 \neq 0$ or $\lambda_0 \neq 0$, and Assumption (A.3) *holds, then*

$$n^{\eta_1} |\hat{c}_0 - c_0| \to 0, \quad \textit{almost surely.} \tag{S2.39}$$

Proof. We will first apply Lemma S2.2 to prove (S2.37) and (S2.39). Let $\delta \geq n^{-1/2}$ be arbitrary. In this proof, K denotes a constant that may depend on $\delta, \theta_0, r, E(|\epsilon_1|^r)$, and may vary from line to line. Some statements only hold for large enough n, but it does not affect the final conclusion.

For $\theta = (\alpha, \beta, \gamma, \lambda)^T \in \mathbb{R}^4, c \in [\ell, u], x \in \mathbb{R}, t \in \{0, 1\}, \epsilon \in \mathbb{R}$, denote $x_c = I(x \leq c)$, and

$$g_{\theta,c}(x, t) = \alpha + \beta t + \gamma x_c + \lambda x_c t, \quad \tilde{g}_{\theta,c}(x, t, \epsilon) = g_{\theta,c}(x, t)\epsilon.$$

For $\theta_1, \theta_2 \in \mathbb{R}^4$ and $c_1, c_2 \in [\ell, u]$, define

$$d_n^2(g_{\theta_1,c_1}, g_{\theta_2,c_2}) = \frac{1}{n} \sum_{i=1}^{n} (g_{\theta_1,c_1}(R_i) - g_{\theta_2,c_2}(R_i))^2 = \|g_{\theta_1,c_1} - g_{\theta_2,c_2}\|_{L_2(\mathbb{P}_n)}^2,$$

where $R_i = (U_i, X_i),\ i \in [n]$, and $\mathbb{P}_n = n^{-1} \sum_{i=1}^{n} \delta_{R_i}$ is the discrete empirical measure, and δ_{R_i} is the Dirac measure charged at R_i.

Finally we denote $g_0 = g_{\theta_0,c_0}$, and

$$\mathcal{G}_\delta = \left\{g_{\theta,c} :\ d_n(g_{\theta,c}, g_0) \le \delta,\ \theta \in \mathbb{R}^4,\ c \in [\ell, u]\right\}, \quad \tilde{\mathcal{G}}_\delta = \{\tilde{g}_{\theta,c} :\ g_{\theta,c} \in \mathcal{G}_\delta\}.$$

By Lemma S2.4, $d_n^2(g_{\theta,c}, g_0) \le \delta^2$ implies that $\|\theta - \theta_0\|^2$ is bounded by $K\delta^2$ for some $K > 0$. Thus $g_{\theta,c}(\cdot) \in \mathcal{G}_\delta$ is uniformly bounded, and we consider the following envelope functions for $\mathcal{G}_\delta$ and $\tilde{\mathcal{G}}_\delta$ respectively,

$$G_\delta(x,t) = K(1+\delta), \quad \tilde{G}_\delta(x,t,\epsilon) = K(1+\delta)|\epsilon|.$$

The model in (2.1) can be equivalently written as

$$Y_i = g_0(R_i) + \epsilon_i, \quad i \in [n].$$

Let $\delta > 0$ be arbitrary. Define the "criterion function" $g_{\theta,c} \in \mathcal{G}_\delta \mapsto \mathbf{D}_n(g_{\theta,c})$ and its centering function $g_{\theta,c} \in \mathcal{G} \to D_n(g_{\theta,c})$ as follows:

$$\mathbf{D}_n(g_{\theta,c}) = \frac{2}{n}\sum_{i=1}^n (g_{\theta,c}(R_i) - g_0(R_i))\epsilon_i - \frac{1}{n}\sum_{i=1}^n (g_{\theta,c}(R_i) - g_0(R_i))^2,$$
$$D_n(g_{\theta,c}) = -\frac{1}{n}\sum_{i=1}^n (g_{\theta,c}(R_i) - g_0(R_i))^2.$$

By the definition of $\hat{c}_0$ in (4.1) and $\hat{\theta}_{\hat{c}_0}$ in (2.2),

$$\hat{g}_n = g_{\hat{\theta}_{\hat{c}_0},\hat{c}_0} = \operatorname*{argmin}_{g_{\theta,c}\in\mathcal{G}} \frac{1}{n}\sum_{i=1}^n (Y_i - g(R_i))^2 = \operatorname*{argmax}_{g_{\theta,c}\in\mathcal{G}} \mathbf{D}_n(g_{\theta,c}).$$

The first condition (S2.34) in Lemma S2.2 is satisfied with equality, since for any $g_{\theta,c} \in \mathcal{G}_\delta$,

$$D_n(g_{\theta,c}) - D_n(g_0) = -\frac{1}{n}\sum_{i=1}^n (g_{\theta,c}(R_i) - g_0(R_i))^2 = -d_n^2(g_{\theta,c}, g_0).$$

Verifying (S2.35) in Lemma S2.2. Define $W_n(g_{\theta,c}) = \sqrt{n}\,(\mathbf{D}_n(g_{\theta,c}) - D_n(g_{\theta,c}))$ for $g_{\theta,c} \in \mathcal{G}_\delta$. Observe that

$$W_n(g_{\theta,c}) - W_n(g_0) = \frac{2}{\sqrt{n}} \sum_{i=1}^{n} (g_{\theta,c}(R_i) - g_0(R_i))\epsilon_i.$$

By (Van Der Vaart and Wellner, 1996, Proposition A.1.6) and since $g_{\theta,c} \in G_\delta$ is uniformly bounded, we have

$$\begin{aligned}
&\| \sup_{g_{\theta,c}\in\mathcal{G}_\delta} (W_n(g_{\theta,c}) - W_n(g_0))\|_r \\
\leq & K\| \sup_{g_{\theta,c}\in\mathcal{G}_\delta} (W_n(g_{\theta,c}) - W_n(g_0))\|_1 + Kn^{-1/2}\| \sup_{i\in[n],\, g_{\theta,c}\in\mathcal{G}_\delta} |(g_{\theta,c}(R_i) - g_0(R_i))\epsilon_i|\|_r \\
\leq & K\| \sup_{g_{\theta,c}\in\mathcal{G}_\delta} (W_n(g_{\theta,c}) - W_n(g_0))\|_1 + Kn^{-(1/2-1/r)} E(|\epsilon_1|^r).
\end{aligned}$$

Since $\delta \geq n^{-1/2}$, we have $n^{-(1/2-1/r)} \leq \delta^{1-2/r}$. Thus, due to Assumption (A.1), to verify (S2.35) in Lemma S2.2, it suffices to show that

$$\| \sup_{g_{\theta,c}\in\mathcal{G}_\delta} (W_n(g_{\theta,c}) - W_n(g_0))\|_1 \leq K\delta^{1-2/r}(\log(\delta^{-1} + \delta))^K. \tag{S2.40}$$

Let $\xi_i, i \in [n]$ be i.i.d. Rademacher variables independent from data $Y_i, i \in [n]$, i.e. $\mathrm{pr}(\xi_1 = 1) = \mathrm{pr}(\xi_1 = -1) = 0.5$. Then by (Van Der Vaart and Wellner, 1996, Lemma 2.3.1) and the law of the total expectation,

$$\begin{aligned}
& E\left[\sup_{g_{\theta,c}\in\mathcal{G}_\delta} |W_n(g_{\theta,c}) - W_n(g_0)|\right] \\
\leq & E\left[\sup_{\tilde{g}_{\theta,c}\in\tilde{\mathcal{G}}_\delta} \left|\frac{2}{\sqrt{n}} \sum_{i=1}^{n} (\tilde{g}_{\theta,c}(R_i,\epsilon_i) - \tilde{g}_0(R_i,\epsilon_i))\xi_i\right|\right] \\
= & E\left[E\left[\sup_{\tilde{g}_{\theta,c}\in\tilde{\mathcal{G}}_\delta} \left|\frac{2}{\sqrt{n}} \sum_{i=1}^{n} (\tilde{g}_{\theta,c}(R_i,\epsilon_i) - \tilde{g}_0(R_i,\epsilon_i))\xi_i\right| \,\middle|\, \epsilon_1,\ldots,\epsilon_n\right]\right].
\end{aligned}$$

By (Van Der Vaart and Wellner, 1996, Corollary 2.2.8),

$$E\left[\sup_{\tilde{g}_{\theta,c}\in\tilde{\mathcal{G}}_\delta}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\tilde{g}_{\theta,c}(R_i,\epsilon_i)-\tilde{g}_0(R_i,\epsilon_i))\right|\right]$$
$$\leq KE\left[\int_0^{v_n/\|\tilde{G}_\delta\|_{L_2(\mathbb{P}_n)}}\sup_Q\sqrt{1+\log N(\zeta\|\tilde{G}_\delta\|_{L_2(Q),2},\tilde{\mathcal{G}}_\delta,L_2(Q))}d\zeta\,\|\tilde{G}_\delta\|_{L_2(\mathbb{P}_n)}\right]$$
$$=KE[J(v_n/\|\tilde{G}_\delta\|_{L_2(\mathbb{P}_n)})\|\tilde{G}_\delta\|_{L_2(\mathbb{P}_n)}],$$

where $v_n^2=\sup_{g_{\theta,c}\in\mathcal{G}_\delta}\mathbb{P}_n[(g_{\theta,c}-g_0)^2\epsilon^2]$, the $\sup_Q$ is taken over all discretely supported probability measures, $N(\zeta,\tilde{\mathcal{G}}_\delta,L_2(Q))$ denotes the ζ-covering number for the set $\tilde{\mathcal{G}}_\delta$ under the semimetric $L_2(Q)$, and $J(\cdot)$ denotes the uniform entropy integral. By the Holder's inequality, since $g_{\theta,c}\in\mathcal{G}_\delta$ is uniformly bounded, we have

$$\begin{aligned}v_n^2&=\sup_{g_{\theta,c}\in\mathcal{G}_\delta}\mathbb{P}_n[(g_{\theta,c}-g_0)^{2/p}(g_{\theta,c}-g_0)^{2-2/p}\epsilon^2]\\&\leq\sup_{g_{\theta,c}\in\mathcal{G}_\delta}\mathbb{P}_n[(g_{\theta,c}-g_0)^{2/p}(K(1+\delta))^{2-2/p}\epsilon^2]\\&\leq\sup_{g_{\theta,c}\in\mathcal{G}_\delta}K\mathbb{P}_n[(g_{\theta,c}-g_0)^2]^{1/p}\mathbb{P}_n[\epsilon^r]^{2/r}\leq K\delta^{2/p}\|\epsilon\|^2_{L_r(\mathbb{P}_n)},\end{aligned}$$

where $1/p=1-2/r$. Due to (Chernozhukov et al., 2014, Lemma A.2), since $\|\epsilon\|_{L_r(\mathbb{P}_n)}\geq\|\epsilon\|_{L_2(\mathbb{P}_n)}$, $g_{\theta,c}\in\mathcal{G}_\delta$ is uniformly bounded,

$$E\left[J\left(\frac{v_n}{\|\tilde{G}_\delta\|_{L_2(\mathbb{P}_n)}}\right)\|\tilde{G}_\delta\|_{L_2(\mathbb{P}_n)}\right]\leq E\left[J\left(K\delta^{1-2/r}\frac{\|\epsilon\|_{L_r(\mathbb{P}_n)}}{\|\epsilon\|_{L_2(\mathbb{P}_n)}}\right)\|\epsilon\|_{L_2(\mathbb{P}_n)}\right]$$
$$\leq KE[J(\delta^{1-2/r})\|\epsilon\|_{L_r(\mathbb{P}_n)}]\leq KJ(\delta^{1-2/r})\|\epsilon_1\|_r. \tag{S2.41}$$

By (Van Der Vaart and Wellner, 1996, Theorem 2.6.7) and (Chernozhukov

et al., 2014, Lemma A.6 and Corollary A.1), $(\tilde{\mathcal{G}}_\delta, \tilde{G}_\delta)$ is a VC-type class, and thus $J(\delta^{1-2/r}) \leq K\delta^{1-2/r}\left(\log(1/\delta+\delta)\right)^K$, which proves (S2.40).

As the two conditions have been satisfied, by Lemma S2.2, $n^{q_1} d_n(\hat{g}_n, g_0)$ converges to zero almost surely for any $q_1 \in (0, 1/2-1/r)$. By Lemma S2.4, we have that for any $q_1 \in (0, 1/2-1/r)$, (S2.37) holds. By Lemma S2.5, if either $\gamma_0 \neq 0$ or $\lambda_0 \neq 0$, (S2.39) holds.

Next we consider (S2.38) regarding the variance estimator $\hat{\sigma}^2$. Without loss of generality we use n instead of $n-k$ in $\hat{\sigma}^2$ hereafter. Expanding the expression of $\hat{\sigma}^2$ yields

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^n \epsilon_i^2 + \frac{2}{n}\sum_{i=1}^n (g_0(R_i) - \hat{g}_n(R_i))\epsilon_i + d_n^2(\hat{g}_n, g_0).$$

For the first term, since $r > 4$, by the law of the iterated logarithm, for any $q_1 \in (0, 1/2)$, almost surely $n^{q_1}\left|\frac{1}{n}\sum_{i=1}^n \epsilon_i^2 - \sigma^2\right| \to 0$ as $n \to \infty$. For the third term, we showed before that for any $q_1 \in (0, 1/2-1/r)$, almost surely $n^{2q_1} d_n^2(\hat{g}_n, g_0) \to 0$ as $n \to \infty$.

Now, for the second term,

$$\left|\frac{n^{q_1}}{n}\sum_{i=1}^n (g_0(R_i) - \hat{g}_n(R_i))\epsilon_i\right| \leq \sup_{g_{\theta,c}\in\mathcal{G}_1}\left|\frac{n^{q_1}}{n}\sum_{i=1}^n (g_0(R_i) - g_{\theta,c}(R_i))\epsilon_i\right| + \left|n^{q_1-1}\sum_{i=1}^n (g_0(R_i) - \hat{g}_n(R_i))\epsilon_i\right| I(d_n(\hat{g}_n, g_0) > 1), \tag{S2.42}$$

where (S2.42) converges to zero almost surely as $d_n(\hat{g}_n, g_0)$ converges to zero

as $n \to \infty$ almost surely. Further, since $r(1/2 - q_1) > 1$, for any $t > 0$, by the Markov's inequality,

$$\sum_{n=1}^{\infty} \mathrm{pr}\left(\sup_{g_{\theta,c} \in \mathcal{G}_1} \left| n^{q_1 - 1} \sum_{i=1}^{n} (g_0(R_i) - g_{\theta,c}(R_i))\epsilon_i \right| > t \right)$$
$$\leq \sum_{n=1}^{\infty} \frac{E\left(\left\{ \sup_{g_{\theta,c} \in \mathcal{G}_1} |W_n(g_{\theta,c})| \right\}^r \right)}{n^{r(1/2 - q_1)} t^r} \leq \sum_{n=1}^{\infty} \frac{K}{n^{r(1/2 - q_1)} t^r} < \infty,$$

where the second inequality is due to (S2.41). By the Borel–Cantelli lemma, we conclude that $n^{q_1} \left| \frac{2}{n} \sum_{i=1}^{n} (g_0(R_i) - \hat{g}_n(R_i))\epsilon_i \right|$ converges to zero almost surely for any $q_1 \in (0, 1/2 - 1/r)$. Then the proof is complete. □

Recall the function class $\mathcal{G} = \{g_{\theta,c} : \theta \in \mathbb{R}^4, c \in [\ell, u]\}$ and its semi-metric d_n in the previous proof.

Lemma S2.4. *Let $r \in (4, \infty)$. Suppose Assumption* (A.2) *holds. There exist $K > 0$ and $N > 0$ such that for $n \geq N$,*

$$\|\theta - \theta_0\|^2 \leq K d_n^2(g_{\theta,c}, g_0), \quad \textit{for any } \theta \in \mathbb{R}^4,\ c \in [\ell, u].$$

Proof. Expanding $d_n^2(g_{\theta,c}, g_0)$ yields

$$\frac{1}{n} \sum_{i=1}^{n} ((\alpha - \alpha_0) + (\beta - \beta_0)U_i + (\gamma X_{i,c} - \gamma_0 X_{i,c_0}) + (\lambda X_{i,c} U_i - \lambda_0 X_{i,c_0} U_i))^2.$$

First, considering the i's such that $U_i = X_{i,c} = X_{i,c_0} = 0$, it is clear that

$$d_n^2(g_{\theta,c}, g_0) \geq (\alpha - \alpha_0)^2 \frac{1}{n} \sum_{i=1}^{n} I(U_i = X_{i,c} = X_{i,c_0} = 0)$$
$$\geq (\alpha - \alpha_0)^2 \frac{1}{n} \sum_{i=1}^{n} I(U_i = 0, X_{i,u} = 0),$$

where $n^{-1}\sum_{i=1}^{n} I(U_i = 0, X_{i,u} = 0)$ is lower bounded for large enough n, based on Assumption (A.2). Therefore there exists a constant K_1 such that for large enough n,

$$(\alpha - \alpha_0)^2 \leq K_1 d_n^2(g_{\theta,c}, g_0).$$

Second, considering the i's such that $U_i = 1, X_{i,c} = X_{i,c_0} = 0$, we have

$$\begin{aligned} d_n^2(g_{\theta,c}, g_0) &\geq (\alpha - \alpha_0 + \beta - \beta_0)^2 \frac{1}{n}\sum_{i=1}^{n} I(U_i = 1, X_{i,c} = X_{i,c_0} = 0) \\ &\geq (\alpha - \alpha_0 + \beta - \beta_0)^2 \frac{1}{n}\sum_{i=1}^{n} I(U_i = 1, X_{i,u} = 0), \end{aligned}$$

which indicates there exists $K_2 > 0$ such that for large enough n,

$$(\beta - \beta_0)^2 \leq 2(\alpha - \alpha_0)^2 + 2(\alpha - \alpha_0 + \beta - \beta_0)^2 \leq K_2 d_n^2(g_{\theta,c}, g_0),$$

as $n^{-1}\sum_{i=1}^{n} I(U_i = 1, X_{i,u} = 0)$ is lower bounded for large enough n due to Assumption (A.2), and by the fact that $2(a+b)^2 + 2a^2 \geq b^2$.

Similar, by considering i's such that $U_i = 0, X_{i,c} = X_{i,c_0} = 1$ and $U_i = X_{i,c} = X_{i,c_0} = 1$, respectively, we can find a constant $K_3, K_4 > 0$ such that for large enough n,

$$(\gamma - \gamma_0)^2 \leq K_3 d_n^2(g_{\theta,c}, g_0), \quad (\lambda - \lambda_0)^2 \leq K_4 d_n^2(g_{\theta,c}, g_0).$$

Then the proof is complete. □

Lemma S2.5. *Let $r \in (4, \infty)$. Suppose that either $\gamma_0 \neq 0$ or $\lambda_0 \neq 0$, and that Assumptions* (A.2) *and* (A.3) *hold. For $n = 1, 2, \ldots$, let $(\theta_n, c_n) \in$*

$\mathbb{R}^4 \times [\ell, u]$ *be a sequence of deterministic vectors. If for any* $q < 1 - 2/r$, $n^q d_n^2(g_{\theta_n, c_n}, g_0) \to 0$ *as* $n \to \infty$, *then* $n^{\eta_1}|c_n - c_0| \to 0$ *as* $n \to \infty$.

Proof. We will prove it by contradiction. Assume the contrary that $n^{\eta_1}|c_n - c_0|$ does not converge to zero. By arguing along a subsequence, without loss of generality, we can assume that there exists $K > 0$ such that either "Case I: $n^{\eta_1}(c_n - c_0) \geq K^{-1}$ for each $n = 1, 2, \ldots$" or "Case II: $n^{\eta_1}(c_n - c_0) < -K^{-1}$ for each $n = 1, 2, \ldots$". As the argument for both cases are similar, we focus on Case I and show that the Case I leads to a contradiction.

Recall the definition of $s_{n,c}, q_{n,c}$ before Assumption (A.2). Denote $(\alpha_n, \beta_n, \gamma_n, \lambda_n)$ the components of θ_n. Since $n^{\eta_1}(c_n - c_0) \geq K^{-1}$, we have

$$
\begin{aligned}
d_n^2(g_{\theta_n, c_n}, g_0) \geq{}& (\alpha_n - \alpha_0 + \gamma_n)^2 \left(\frac{1}{n} \sum_{i=1}^n I\left(U_i = 0, X_{i,c_n} = 1, X_{i,c_0} = 0\right) \right) \\
&+ ((\alpha_n - \alpha_0) + (\beta_n - \beta_0) + (\gamma_n) + (\lambda_n))^2 \left(\frac{1}{n} \sum_{i=1}^n I(U_i = 1, X_{i,c_n} = 1, X_{i,c_0} = 0) \right)
\end{aligned}
$$

Denote $\bar{c}_n = c_0 + K^{-1} n^{-\eta_1}$. By definition,

$$
\begin{aligned}
&\frac{1}{n} \sum_{i=1}^n I\left(U_i = 0, X_{i,c_n} = 1, X_{i,c_0} = 0\right) \\
\geq &\frac{1}{n} \sum_{i=1}^n (1 - U_i)(X_{i,\bar{c}_n} - X_{i,c_0}) = s_{n,\bar{c}_n} - s_{n,c_0} - (q_{n,\bar{c}_n} - q_{n,c_0}).
\end{aligned}
$$

Similarly, $\frac{1}{n} \sum_{i=1}^n I(U_i = 1, X_{i,c_n} = 1, X_{i,c_0} = 0) \geq q_{n,\bar{c}_n} - q_{n,c_0}$.

Due to Assumptions (A.2) and (A.3),

$$\liminf_{n\to\infty} n^{\eta_1}(s_{n,\bar{c}_n} - s_{n,c_0} - (q_{n,\bar{c}_n} - q_{n,c_0}))$$
$$\geq \lim_{n\to\infty} n^{\eta_1}\left(F(\bar{c}_n) - F(c_0) - p(F(\bar{c}_n) - F(c_0)) - o(1)\right) = K^{-1}(1-p)F'(c_0) > 0,$$

where $o(1)$ represents a sequence of numbers converging to zero. By a similar way we can show $\liminf_{n\to\infty} n^{\eta_1}(q_{n,\bar{c}_n} - q_{n,c_0}) \geq K^{-1}pF'(c_0) > 0$.

Now by Lemma S2.4, $\theta_n \to \theta_0$. Thus if either $\gamma_0 \neq 0$ or $\lambda_0 \neq 0$, at least one of the following holds:

$$\liminf_{n\to\infty}(\alpha_n - \alpha_0) + (\beta_n - \beta_0) + (\gamma_n))^2 > 0,$$
$$\liminf_{n\to\infty}((\alpha_n - \alpha_0) + (\beta_n - \beta_0) + (\gamma_n) + (\lambda_n))^2 > 0.$$

Now combining above results, we have

$$\liminf_{n\to\infty} n^{\eta_1} d_n^2(g_{\theta_n,c_n}, g_0) > 0,$$

which is a contradiction as $\eta_1 < 1 - 2/r$. □

S2.4 Discussions on the assumptions with Example 1

In this subsection we show that Assumptions (A.2) and (A.3) are satisfied under the setup of the Example 1.

Lemma S2.6. *Let $r \in (4, \infty]$ and (U_i, X_i) $(i = 1, 2, \ldots)$ be an independent and identically distributed sequence of random variables. Assume X_1 and*

U_1 are independent. Denote by F_0 the distribution of X_1. If $0 < E(U_1) < 1$, $0 < F_0(\ell) < F_0(u) < 1$, and F_0 is differentiable at c_0 with $F_0'(c_0) > 0$, then Assumptions (A.2) *and* (A.3) *hold almost surely with $F = F_0$, $p = E(U_1)$, any $\eta_0 \in (0, 1/2 - 1/r)$, and any $\eta_1 \in (1/2, 1 - 2/r)$.*

Proof. It suffices to show the case that $r < \infty$. We start with Assumption (A.2). For independent and identically distributed (U_i, X_i) $(i = 1, 2, \ldots)$, with X_1 and U_1 being independent, by the law of the iterated logarithm, we have

$$\limsup_{n\to\infty} \frac{|\sum_{i=1}^n (X_{i,c} - X_{i,c_0}) - (F(c) - F(c_0))|}{\sqrt{2n\log(\log n)}} = 1, \quad \text{almost surely,}$$
$$\limsup_{n\to\infty} \frac{|\sum_{i=1}^n U_i(X_{i,c} - X_{i,c_0}) - p(F(c) - F(c_0))|}{\sqrt{2n\log(\log n)}} = 1, \quad \text{almost surely,}$$

for any $c \in [\ell, u]$. Since $\log\log(n) = o(n^q)$ for any $q > 0$, Assumption (A.2) holds for any $\eta_0 \in (0, 1/2 - 1/r)$ almost surely.

Next we consider Assumption (A.3). Let $\eta_1 \in (1/2, 1 - 2/r)$ be arbitrary. We will show that for any $K_0 > 0$, almost surely,

$$n^{\eta_1 - 1} \sup_{|c-c_0| \le K_0 n^{-\eta_1}} \left| \sum_{i=1}^n ((X_{i,c} - X_{i,c_0}) - (F(c) - F(c_0))) \right| \to 0, \quad \text{(S2.43)}$$
$$n^{\eta_1 - 1} \sup_{|c-c_0| \le K_0 n^{-\eta_1}} \left| \sum_{i=1}^n (U_i(X_{i,c} - X_{i,c_0}) - p(F(c) - F(c_0))) \right| \to 0. \quad \text{(S2.44)}$$

We only provide the proof of (S2.43). The proof of (S2.44) is similar and thus omitted.

Denote by $\mathcal{G}$ the class of functions $\{g_c : |c - c_0| \leq K_0 n^{-\eta_1}\}$, where $g_c(x) = I(x \leq c) - I(x \leq c_0)$. We use the following envelop function for the class $\mathcal{G}$:

$$F(x) = I(c_0 - K_0 n^{-\eta_1} \leq x \leq c_0 + K_0 n^{-\eta_1}), \quad \text{for } x \in \mathbb{R}.$$

Clearly $E(F^2(X_1)) = F(c_0 + K_0 n^{-\eta_1}) - F(c_0 - K_0 n^{-\eta_1})$.

Denote

$$\begin{aligned}
\|\mathbb{G}_n\|_{\mathcal{G}} &= \sup_{g_c \in \mathcal{G}} \frac{1}{\sqrt{n}} \left| \sum_{i=1}^n g_c(X_i) - E g_c(X_1) \right| \\
&= \sup_{|c-c_0| \leq K_0 n^{-\eta_1}} \frac{1}{\sqrt{n}} \left| \sum_{i=1}^n \left((X_{i,c} - X_{i,c_0}) - (F(c) - F(c_0)) \right) \right|, \\
M_i &= \sup_{|c-c_0| \leq K_0 n^{-\eta_1}} |(X_{i,c} - X_{i,c_0}) - (F(c) - F(c_0))| \leq 1, \quad \text{for } 1 \leq i \leq n.
\end{aligned}$$

Since $\eta_1 \in (1/2, 1 - 2/r)$, i.e., $r(1-\eta_1)/2 > 1$, for any fixed $t > 0$,

$$\begin{aligned}
&\sum_{n=1}^{\infty} \text{pr}\left(n^{\eta_1 - 1/2} \|\mathbb{G}_n\|_{\mathcal{G}} \geq t\right) \leq_{(1)} \sum_{n=1}^{\infty} \frac{E\left((\|\mathbb{G}_n\|_{\mathcal{G}})^r\right)}{n^{-(\eta_1 - 1/2)r} t^r} \\
\leq_{(2)} &\sum_{n=1}^{\infty} \frac{K\left[E(\|\mathbb{G}_n\|_{\mathcal{G}}) + \frac{1}{\sqrt{n}} \left(E(\max_{1 \leq i \leq n} M_i^r)\right)^{1/r}\right]^r}{n^{-(\eta_1 - 1/2)r} t^r} \\
\leq &\sum_{n=1}^{\infty} \frac{K\left[E(\|\mathbb{G}_n\|_{\mathcal{G}}) + n^{-1/2}\right]^r}{n^{-(\eta_1 - 1/2)r} t^r} = \sum_{n=1}^{\infty} \frac{K}{t^r} \left[\frac{E(\|\mathbb{G}_n\|_{\mathcal{G}}) + n^{-1/2}}{n^{-(\eta_1 - 1/2)}}\right]^r, \quad \text{(S2.45)}
\end{aligned}$$

where K is a constant that only depends on r, the inequality (1) is due to the Markov's inequality, and the inequality (2) is due to the Hoffmann-Jorgensen's inequality (Van Der Vaart and Wellner, 1996, Proposition A.1.6).

By (Van Der Vaart and Wellner, 1996, Theorem 2.14.1), there exists an

absolute constant $K > 0$,

$$E(\|\mathbb{G}_n\|_{\mathcal{G}}) \leq K \sup_Q \int_0^1 \sqrt{1 + \log N(\xi \|F\|_{Q,2}, \mathcal{G}, L_2(Q))}\, d\xi \sqrt{E(F^2(X_1))}$$

where the supremum is taken over all discretely supported probability measures Q, $L_2(Q)$ is the L_2 norm for $\mathcal{G}$ under the probability measure Q, $\|F\|_{Q,2}$ is the $L_2(Q)$ norm of F, and $N(\xi\|F\|_{Q,2}, \mathcal{G}, L_2(Q))$ is the $\xi\|F\|_{Q,2}$ covering number for $\mathcal{G}$ under the $L_2(Q)$ norm. Then by (Van Der Vaart and Wellner, 1996, Theorem 2.6.7) and since F differentiable at c_0, there exists some absolute constant $K > 0$ such that for large enough n,

$$E(\|\mathbb{G}_n\|_{\mathcal{G}}) \leq K n^{-\eta_1/2}.$$

Plugging the above inequality into (S2.45), for some constant $K > 0$ depending only on r,

$$\sum_{n=1}^{\infty} \text{pr}\left(n^{\eta_1 - 1/2} \|\mathbb{G}_n\|_{\mathcal{G}} \geq t\right) \leq \frac{K}{t^r} \sum_{n=1}^{\infty} \left[\frac{n^{-\eta_1/2} + n^{-1/2}}{n^{-(\eta_1 - 1/2)}}\right]^r \leq \frac{2^r K}{t^r} \sum_{n=1}^{\infty} n^{\frac{-(1-\eta_1)r}{2}} < \infty.$$

Finally, the proof of (S2.43) is complete due to Borel-Cantelli lemma. $\square$

S3 Proof of the Consistency under Alternative Hypothesis

S3.1 Proof of the convergence of $\tilde{c}_0$ under the alternative hypothesis

In this section we establish the weak consistency of the minimum p-value estimator $\tilde{c}_0$ under some conditions, when the alternative hypothesis holds. Define for each c,

$$H_c = \frac{|\lambda_c|}{v_c}, \quad c_b = \operatorname*{argmax}_{c\in[\ell,u]} H_c,$$

$$\text{where } \lambda_c = d^T Q_c^{-1} Q_{c,c_0}\theta_0 = \begin{cases} \frac{\lambda_0 F(c)}{F(c_0)}, & \text{if } c \le c_0, \\ \frac{\lambda_0(1-F(c))}{1-F(c_0)}, & \text{otherwise,} \end{cases}$$

$$v_c^2 = [\sigma^2 + \theta_0^T Q_{c_0}\theta_0 - \theta_0^T Q_{c,c_0}^T Q_c^{-1} Q_{c,c_0}\theta_0]/[F(c)(1-F(c))]. \quad \text{(S3.46)}$$

Lemma S3.1. *Suppose for any $t > 0$, $H_{c_b} > \inf_{|c-c_b|>t} H_c$. Under Assumptions* (A.1) *and* (A.2)*, the minimum p-value estimator $\tilde{c}_0$ converges in probability to c_b, if the alternative hypothesis holds, i.e., $\lambda_0 \neq 0$.*

Proof. From the definition of $\tilde{c}_0$, for any $t > 0$, we have that

$$\begin{aligned}
&\mathrm{pr}(|\tilde{c}_0 - c_b| \geq t) \leq \mathrm{pr}\left(\frac{\sqrt{n}|\hat{\lambda}_{c_b}|}{\hat{v}_{c_b}} \leq \sup_{|c-c_b|\geq t} \frac{\sqrt{n}|\hat{\lambda}_c|}{\hat{v}_c}\right) \\
&\leq \mathrm{pr}\left(\frac{\sqrt{n}|\lambda_{c_b}|}{\hat{v}_{c_b}} - \frac{\sqrt{n}|\hat{\lambda}_{c_b} - \lambda_{c_b}|}{\hat{v}_{c_b}} \leq \sup_{|c-c_b|\geq t} \frac{\sqrt{n}|\lambda_c|}{\hat{v}_c} + \sup_{|c-c_b|\geq t} \frac{\sqrt{n}|\hat{\lambda}_c - \lambda_c|}{\hat{v}_c}\right) \\
&\leq \mathrm{pr}\left(n^{1/2-\eta_0} \inf_{|c-c_b|\geq t} \frac{|\lambda_{c_b}|}{\hat{v}_{c_b}} - \frac{|\lambda_c|}{\hat{v}_c} \leq \sup_{|c-c_b|\geq t} \frac{n^{1/2-\eta_0}|\hat{\lambda}_c - \lambda_c|}{\hat{v}_c} + \frac{n^{1/2-\eta_0}|\hat{\lambda}_{c_b} - \lambda_{c_b}|}{\hat{v}_{c_b}}\right) \\
&\leq \mathrm{pr}\left(n^{1/2-\eta_0} \inf_{|c-c_b|\geq t} \frac{|\lambda_{c_b}|}{\hat{v}_{c_b}} - \frac{|\lambda_c|}{\hat{v}_c} \leq 1\right) \qquad \text{(S3.47)} \\
&+ \mathrm{pr}\left(\sup_{|c-c_b|\geq t} \frac{n^{1/2-\eta_0}|\hat{\lambda}_c - \lambda_c|}{\hat{v}_c} + \frac{n^{1/2-\eta_0}|\hat{\lambda}_{c_b} - \lambda_{c_b}|}{\hat{v}_{c_b}} \geq 1\right). \qquad \text{(S3.48)}
\end{aligned}$$

We first consider (S3.47). Due to Assumptions (A.1) and (A.2), by a similar argument as in Lemma S2.3, $v_c/\hat{v}_c \to 1$ as $n \to \infty$ uniformly over $c \in \mathcal{C}_n$ almost surely, and v_c^2 is lower bounded away from zero uniformly over $c \in \mathcal{C}_n$. Thus

$$\begin{aligned}
&\inf_{|c-c_b|\geq t} \frac{|\lambda_{c_b}|}{\hat{v}_{c_b}} - \frac{|\lambda_c|}{\hat{v}_c} = \inf_{|c-c_b|\geq t} \left(H_{c_b}\frac{v_{c_b}}{\hat{v}_{c_b}} - H_c\frac{v_c}{\hat{v}_c}\right) \\
&\geq \inf_{|c-c_b|\geq t} (H_{c_b} - H_c) - \left|H_{c_b}\left(\frac{v_{c_b}}{\hat{v}_{c_b}} - 1\right)\right| - \sup_{|c-c_b|\geq t}\left|H_c\left(\frac{v_c}{\hat{v}_c} - 1\right)\right| \\
&\geq \inf_{|c-c_b|\geq t} (H_{c_b} - H_c) - o_{a.s.}(1),
\end{aligned}$$

where $o_{a.s.}(1)$ represents a sequence of random variables converging to zero almost surely. Since $\inf_{|c-c_b|\geq t}(H_{c_b} - H_c) > 0$, we have

$$\mathrm{pr}\left(n^{1/2-\eta_0} \inf_{|c-c_b|\geq t} \frac{|\lambda_{c_b}|}{\hat{v}_{c_b}} - \frac{|\lambda_c|}{\hat{v}_c} \leq 1\right) \to 0.$$

Now we consider (S3.48). From Lemma S3.2, for large enough n, $\hat{v}_c^2$ is lower bounded away from zero uniformly over $c \in [\ell, u]$ almost surely. As a result, it is sufficient to show that $n^{1/2-\eta_0} \sup_{|c-c_b|\geq t} |\hat{\lambda}_c - \lambda_c|$ converges to zero in probability. Observe that

$$
\begin{aligned}
&n^{1/2-\eta_0} \sup_{c\in\mathcal{C}_n} |\hat{\lambda}_c - \lambda_c| = n^{1/2-\eta_0} \sup_{c\in\mathcal{C}_n} \left| d^T \hat{Q}_{n,c}^{-1} \frac{1}{n} \sum_{i=1}^n Z_{i,c} Y_i - d^T Q_c^{-1} Q_{c,c_0} \theta_0 \right| \\
&\leq n^{1/2-\eta_0} \sup_{c\in\mathcal{C}_n} \left| d^T (\hat{Q}_{n,c}^{-1} - Q_c^{-1}) \frac{1}{n} \sum_{i=1}^n Z_{i,c} Z_{i,c_0}^T \theta_0 \right| \\
&+ n^{1/2-\eta_0} \sup_{c\in\mathcal{C}_n} \left| d^T Q_c^{-1} \left(\frac{1}{n} \sum_{i=1}^n Z_{i,c} Z_{i,c_0} - Q_{c,c_0} \right) \theta_0 \right| \\
&+ n^{1/2-\eta_0} \sup_{c\in\mathcal{C}_n} \left| d^T \hat{Q}_{n,c}^{-1} \frac{1}{n} \sum_{i=1}^n Z_{i,c} \epsilon_i \right| .
\end{aligned}
$$

Due to Assumption (A.2) and Lemma S1.2, and since all entries of $n^{-1} \sum_{i=1}^n Z_{i,c} Z_{i,c_0}^T$ are between 0 and 1, the first two terms converge to zero.

Further, since all entries of $\hat{Q}_{n,c}^{-1}$ and $Z_{i,c}$ are bounded uniformly over $c \in \mathcal{C}_n$ (cf. (S2.16)), there exists a constant K, that may depend on $\tau, r, E(|\epsilon_1|^r)$, such that $\sup_{c\in\mathcal{C}_n} \sum_{i=1}^n E\left(\left| d^T \hat{Q}_{n,c}^{-1} \frac{1}{\sqrt{n}} Z_{i,c} \epsilon_i \right|^2 \right) \leq K$ and

$$
E\left(\sup_{1\leq i\leq n, c\in\mathcal{C}_n} \left| d^T \hat{Q}_{n,c}^{-1} \frac{1}{\sqrt{n}} Z_{i,c} \epsilon_i \right|^2 \right) \leq K n^{-1+2/r},
$$

where the second inequality is due to (Van Der Vaart and Wellner, 1996, Corollary 2.2.2). Then by (Chernozhukov et al., 2017, Lemma E.1), there

exists a constant K, that may depend on τ, $E(|\epsilon_1|^r)$, such that

$$E\left(\sup_{c\in\mathcal{C}_n}\left|d^T\hat{Q}_{n,c}^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}Z_{i,c}\epsilon_i\right|\right)\leq K\sqrt{\log(n)}. \tag{S3.49}$$

As a result, $n^{-\eta_0-1/2}\sup_{c\in\mathcal{C}_n}\left|d^T\hat{Q}_{n,c}^{-1}\sum_{i=1}^{n}Z_{i,c}\epsilon_i\right|$ converges to zero in probability, which implies that (S3.48) converges to zero as $n\to\infty$.

Combining the results of (S3.47) and (S3.48), we finish the proof. □

As the conditions in Lemma S3.1 is difficult to verify, we provide a simple sufficient condition in Lemma 1 in Section 4.1, under which $\tilde{c}_0$ converges weakly to c_0. Next we provide the proof of Lemma 1.

Proof. For each c, denote $\kappa_c^2=F(c)(1-F(c))\lambda_c^2$ and $u_c^2=F(c)(1-F(c))v_c^2$. From the definition of H_c, it is clear that $H_c=|\kappa_c|/u_c$. Next we consider u_c^2 and κ_c^2 separately.

For simplicity, denote $Z_c=(Z_{1,c}^T,\ldots,Z_{n,c}^T)^T$ to be the design matrix for a given $c\in[\ell,u]$. By Assumption (A.2) and Lemma S1.3, $Z_c^TZ_c/n\to Q_c$ and $Z_c^TZ_{c_0}/n\to Q_{c,c_0}$, $(Z_c^TZ_c/n)^{-1}\to Q_c^{-1}$ for all $c\in[\ell,u]$ component-wise. As a result, by the definition of v_c^2 in (S3.46), we have the following almost sure equation:

$$\begin{aligned}u_c^2&=\sigma^2+\theta_0^TQ_{c_0}\theta_0-\theta_0^TQ_{c,c_0}^TQ_c^{-1}Q_{c,c_0}\theta_0\\&=\sigma^2+\lim_{n\to\infty}\frac{1}{n}\theta_0^TZ_{c_0}^T(I-Z_c(Z_cZ_c^T)^{-1}Z_c^T)Z_{c_0}\theta_0,\end{aligned}$$

where the second term is always nonnegative since $I - Z_c(Z_c^T Z_c)^{-1} Z_c^T$ is an idempotent matrix, and achieves 0 when $c = c_0$. Therefore $u_{c_0}^2 = \sigma^2 \leq u_c^2$ for $c \in [\ell, u]$.

Next we deal with κ_c^2. Note that

$$\frac{\partial \kappa_c^2}{\partial F(c)} = \begin{cases} (3F(c)^2 - 4F(c)^3) \times \frac{\lambda_0^2}{F(c_0)^2}, & \text{if } c \leq c_0, \\ (1 - F(c))^2(1 - 4F(c)) \times \frac{\lambda_0^2}{(1-F(c_0))^2}, & \text{otherwise.} \end{cases}$$

Under the alternative, $\lambda_0 \neq 0$. Since $F(c_0) \in [1/4, 3/4]$ and F has a positive derivative at c_0, $F(c) < 3/4$ for $c < c_0$ and $F(c) > 1/4$ for $c > c_0$. Thus, for any $t > 0$,

$$\kappa_{c_0}^2 > \max_{c:|c-c_0|\geq t} \kappa_c^2,$$

which implies that

$$\inf_{|c-c_0|\geq t} (H_{c_0} - H_c) = \inf_{|c-c_0|\geq t} \left(\frac{|\kappa_{c_0}|}{\sigma} - \frac{|\kappa_c|}{u_c} \right) \geq \inf_{|c-c_0|\geq t} \left(\frac{|\kappa_{c_0}|}{\sigma} - \frac{|\kappa_c|}{\sigma} \right) > 0.$$

Then the proof is complete due to Lemma S3.1. □

S3.2 Proof of the consistency of multiplier residual bootstrap under the alternative hypothesis

In this subsection we provide the proof of Theorem 4.

Proof. If $r = \infty$, we have $E(|\epsilon_1|^{r_1}) < \infty$ for any $r_1 \in (4, \infty)$. Thus the case for $r = \infty$ can be recovered by first proving the result for $r_1 < \infty$, and then letting $r_1 \to \infty$. Thus in this proof we focus on the case that $r < \infty$.

Recall in Sections 2.2 and 4.1, for each $c \in [\ell, u]$, $\sqrt{n}\hat{\lambda}_c$ and $\sqrt{n}\hat{\lambda}_c^*$ can be respectively decomposed as

$$\sqrt{n}\hat{\lambda}_c = d^T \hat{Q}_{n,c}^{-1}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} Z_{i,c} Z_{i,c_0}^T \theta_0\right) + d^T \hat{Q}_{n,c}^{-1}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} Z_{i,c}\epsilon_i\right) = I_{n,c} + II_{n,c},$$
$$\sqrt{n}\hat{\lambda}_c^* = d^T \hat{Q}_{n,c}^{-1}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} Z_{i,c} Z_{i,\hat{c}_0}^T \hat{\theta}_0\right) + \hat{\sigma} d^T \hat{Q}_{n,c}^{-1}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} Z_{i,c}\zeta_i\right) = I_{n,c}^* + II_{n,c}^*,$$

where recall that $d = (0,0,0,1)^T$, $\hat{\theta}_0 = (\hat{\alpha}_0, \hat{\beta}_0, \hat{\gamma}_0, 0)^T$, $\theta_0 = (\alpha_0, \beta_0, \gamma_0, \lambda_0)^T$, and $\zeta_1, \ldots, \zeta_n$ are independent and identically distributed standard normal random variables. Further, $H_1 : \lambda_0 = \lambda_{0,n}$ with $\liminf_{n\to\infty} n^{1/2-\eta_0}|\lambda_{0,n}| > 0$ and $\limsup_{n\to\infty} |\lambda_{0,n}| < \infty$, is assumed. We would like to show the power of the test based on the multiplier residual bootstrap converges to 1 for any fixed significant level $\xi \in (0,1)$. Note that the probability of rejecting the null is lower bounded by

$$\begin{aligned}
\mathrm{pr}(p_{n,mrb}^* \le \xi) &= \mathrm{pr}(M_n \ge (F_{n,mrb}^*)^{-1}(1-\xi)) \\
&= \mathrm{pr}(n^{-\eta_0} M_n \ge n^{-\eta_0}(F_{n,mrb}^*)^{-1}(1-\xi)) \\
&\ge \mathrm{pr}(n^{-\eta_0} M_n \ge K, n^{-\eta_0}(F_{n,mrb}^*)^{-1}(1-\xi) \le K) \\
&\ge \mathrm{pr}(n^{-\eta_0} M_n \ge K) + \mathrm{pr}(n^{-\eta_0}(F_{n,mrb}^*)^{-1}(1-\xi) \le K) - 1
\end{aligned}$$

for any constant $K > 0$.

First we consider $n^{-\eta_0}M_n$. From the definition of M_n, we have

$$\begin{aligned} n^{-\eta_0}M_n &= n^{-\eta_0}\sup_{c\in[\ell,u]}|M_{n,c}| \geq n^{-\eta_0}|M_{n,c_0}| \\ &= n^{-\eta_0}\left|\frac{\sqrt{n}\lambda_0}{\hat{v}_{c_0}} + \frac{d^T\hat{Q}_{n,c_0}^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^n Z_{i,c_0}\epsilon_i}{\hat{v}_{c_0}}\right| \\ &\geq n^{-\eta_0}\left|\frac{\sqrt{n}\lambda_0}{\hat{v}_{c_0}}\right| - n^{-\eta_0}\left|\frac{d^T\hat{Q}_{n,c_0}^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^n Z_{i,c_0}\epsilon_i}{\hat{v}_{c_0}}\right|. \end{aligned}$$

From Lemma S3.2 and the definition of (4.5), there exist some K such that $\limsup_{n\to\infty}\hat{v}_{c_0} < K$ as well as $\liminf_{n\to\infty}\hat{v}_{c_0} > 0$ almost surely, and $\liminf_{n\to\infty} n^{1/2-\eta_0}|\lambda_0| > 0$. Further, due to (S2.16), as $n\to\infty$,

$$n^{-2\eta_0}E\left(\left(d^T\hat{Q}_{n,c_0}^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^n Z_{i,c_0}\epsilon_i\right)^2\right) = \frac{d^T\hat{Q}_{n,c_0}^{-1}d\sigma^2}{n^{2\eta_0}} \to 0.$$

As a result, there exits $K_0 > 0$ such that

$$\mathrm{pr}(n^{-\eta_0}M_n \geq K_0) \to 1, \text{ as } n\to\infty.$$

Now for the K_0 above, it is sufficient to show that

$$\mathrm{pr}(n^{-\eta_0}(F_{n,mrb}^*)^{-1}(1-\xi) > K_0) \to 0, \text{ as } n\to\infty.$$

In turn, it suffices to show that conditional on almost all sequence $\epsilon_1, \epsilon_2, \ldots$, $n^{-\eta_0}M_n^* \to 0$ in probability, i.e., for any $t > 0$,

$$\mathrm{pr}_{|\epsilon}(n^{-\eta_0}M_n^* > t) \to 0, \text{ as } n\to\infty, \text{ almost surely,} \tag{S3.50}$$

on which we now focus.

Let $t > 0$ be arbitrary. Due to (S2.16) and the definition of $\hat{\sigma}^2$,

$$\begin{aligned}
&\mathrm{pr}_{|\epsilon}\left(n^{-\eta_0} M_n^* > t\right) = \mathrm{pr}_{|\epsilon}\left(n^{-\eta_0} \sup_{c\in\mathcal{C}_n} \left|\frac{I_{n,c}^* + II_{n,c}^*}{\hat{v}_c}\right| > t\right) \\
\leq &\mathrm{pr}_{|\epsilon}\left((\tau\hat{\sigma}^2)^{-1/2}\left(n^{-\eta_0}\sup_{c\in\mathcal{C}_n}|I_{n,c}^*| + n^{-\eta_0}\sup_{c\in\mathcal{C}_n}|II_{n,c}^*|\right) > t\right) \\
\leq &\mathrm{pr}_{|\epsilon}\left(n^{-\eta_0}\sup_{c\in\mathcal{C}_n}|I_{n,c}^*| > \sqrt{\tau}\hat{\sigma}t/2\right) \qquad \text{(S3.51)} \\
+&\mathrm{pr}\left(n^{-\eta_0}\sup_{c\in\mathcal{C}_n}|II_{n,c}^*/\hat{\sigma}| > \sqrt{\tau}t/2\right). \qquad \text{(S3.52)}
\end{aligned}$$

For (S3.51), we have

$$\begin{aligned}
n^{-\eta_0}\sup_{c\in\mathcal{C}_n}|I_{n,c}^*| &= n^{-\eta_0}\sup_{c\in\mathcal{C}_n}\left|d^T\hat{Q}_{n,c}^{-1}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n Z_{i,c}Z_{i,\hat{c}_0}\hat{\theta}_0\right)\right| \\
&\leq n^{-\eta_0}\sup_{c\in\mathcal{C}_n}\left|d^T\hat{Q}_{n,c}^{-1}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n Z_{i,c}Z_{i,\hat{c}_0}^T - Q_{c,\hat{c}_0}\right)\hat{\theta}_0\right| \qquad \text{(S3.53)} \\
&+ n^{-\eta_0}\sup_{c\in\mathcal{C}_n}\left|d^T\sqrt{n}(\hat{Q}_{n,c}^{-1} - Q_c^{-1})Q_{c,\hat{c}_0}\hat{\theta}_0\right| \qquad \text{(S3.54)} \\
&+ n^{-\eta_0}\sup_{c\in\mathcal{C}_n}\left|d^T\sqrt{n}Q_c^{-1}Q_{c,c_0}\hat{\theta}_0\right|,
\end{aligned}$$

where the last term is zero by Lemma S1.1 and the fact that $\hat{\theta}_0 = (\hat{\alpha}_0, \hat{\beta}_0, \hat{\gamma}_0, 0)^T$. Due to (S2.16), Assumption (A.2), Lemmas S1.3 and S2.3 (so that $\|\hat{\theta}_0\|$ is bounded almost surely), (S3.53) and (S3.54) converge to zero almost surely. Thus $n^{-\eta_0}\sup_{c\in\mathcal{C}_n}|I_{n,c}^*| \to 0$ as $n \to \infty$ almost surely. Finally, by Lemma S2.3, we have (S3.51) $= 0$, for large enough n, almost surely..

For (S3.52), as $\zeta_1, \ldots, \zeta_n$ are independently and identically distributed standard normal random variables, $\{II_{n,c}^*/\hat{\sigma} : c \in \mathcal{C}_n\}$ is a multivariate

Gaussian vector with mean 0 and covariance matrix Σ_{n_c} (cf. (S2.22)), where all entries of Σ_{n_c} are bounded in absolute value by $16/\tau^2$ and τ appears in (S2.16). By (Van Der Vaart and Wellner, 1996, Lemma 2.2.2), as $n \to 0$,

$$n^{-\eta_0} E\left(\sup_{c\in\mathcal{C}_n} |II^*_{n,c}/\hat{\sigma}|\right) \leq K n^{-\eta_0}\sqrt{\log(1+n)} \to 0,$$

where K is a constant that may depend on τ. Thus (S3.52) $\to 0$ as $n \to 0$, and (S3.50) holds. Then the proof is complete. □

Lemma S3.2. *Let $r \in (4, \infty)$. Under Assumption* (A.2)*, for large enough n, $\hat{v}_c^2$ is lower bounded away from zero uniformly over $c \in [\ell, u]$ almost surely.*

Proof. Recall the definition of $\hat{v}_c^2$ in (2.2): $\hat{v}_c^2 = d^T \hat{Q}_{n,c}^{-1} d\, \mathrm{RSS}_c(\hat{\theta}_c)/(n-4)$. Note that $d^T \hat{Q}_{n,c}^{-1} d$ is lower bounded away from zero by τ uniformly over $c \in \mathcal{C}_n$, from (S2.16), based on Assumption (A.2). By the definition of $\hat{\sigma}^2$,

$$\hat{\sigma}^2 = \frac{1}{n-4}\mathrm{RSS}_{\hat{c}_0}(\hat{\theta}_{\hat{c}_0}) = \min_{c\in\mathcal{C}_n} \frac{1}{n-4}\mathrm{RSS}_c(\hat{\theta}_c) \leq \frac{1}{n-4}\mathrm{RSS}_c(\hat{\theta}_c),$$

for all $c \in \mathcal{C}_n$. Then the proof is complete due to Lemma S2.3. □

S4 Proof of the Asymptotic Validity of the Profile Least squares Estimation based Test

In this Subsection we provide the proof for Lemma 2.

Proof. By the Lindeberg-Feller central limit theorem with the Lyapunov condition, and the weak law of large number, as $n \to \infty$, under $H_0 : \lambda_0 = 0$, we have $\sqrt{n}\hat{\lambda}_{c_0}/\hat{v}_{c_0} \to N(0,1)$ in distribution , where $N(0,1)$ denotes the standard normal distribution. Besides,

$$\frac{\sqrt{n}\hat{\lambda}_{\hat{c}_0}}{\hat{v}_{\hat{c}_0}} - \frac{\sqrt{n}\hat{\lambda}_{c_0}}{\hat{v}_{c_0}} = \frac{\sqrt{n}(\hat{\lambda}_{\hat{c}_0} - \hat{\lambda}_{c_0})}{\hat{v}_{\hat{c}_0}} + \sqrt{n}\hat{\lambda}_{c_0}\left(\frac{1}{\hat{v}_{\hat{c}_0}} - \frac{1}{\hat{v}_{c_0}}\right).$$

By Lemma S3.2, $\hat{v}_c$ is lower bounded by a positive constant uniformly over $c \in [\ell, u]$ for large enough n, almost surely. Since $n^{1/2}\hat{\lambda}_{c_0}$ is bounded in probability and $\hat{v}_c^2$ converges to v_c^2 uniformly over $c \in [\ell, u]$ almost surely, it suffices to show that as $n \to 0$, $\sqrt{n}\left(\hat{\lambda}_{\hat{c}_0} - \hat{\lambda}_{c_0}\right) \xrightarrow{p} 0$, where $\xrightarrow{p}$ means convergence in probability.

Recall that taking the minimum of $\mathrm{RSS}_c(\hat{\theta}_c)$ is equivalent to taking the minimum over $\mathcal{C}_n$. Observe that

$$\begin{aligned}
&\sqrt{n}\left|\hat{\lambda}_{\hat{c}_0} - \hat{\lambda}_{c_0}\right| \\
&= \left| d^T \hat{Q}_{n,\hat{c}_0}^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^n Z_{i,\hat{c}_0} Z_{i,c_0}^T \theta_0 + d^T \hat{Q}_{n,\hat{c}_0}^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^n Z_{i,\hat{c}_0}\epsilon_i - d^T \hat{Q}_{n,c_0}^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^n Z_{i,c_0}\epsilon_i \right| \\
&\le \left| d^T \hat{Q}_{n,\hat{c}_0}^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^n Z_{i,\hat{c}_0}(Z_{i,c_0} - Z_{i,\hat{c}_0})^T \theta_0 \right| + \left| d^T (\hat{Q}_{n,\hat{c}_0}^{-1} - \hat{Q}_{n,c_0}^{-1}) \frac{1}{\sqrt{n}} \sum_{i=1}^n Z_{i,\hat{c}_0}\epsilon_i \right| \\
&+ \left| d^T \hat{Q}_{n,c_0}^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^n (Z_{i,\hat{c}_0} - Z_{i,c_0})\epsilon_i \right|.
\end{aligned}$$

As a result, $\sqrt{n}\left|\hat{\lambda}_{\hat{c}_0} - \hat{\lambda}_{c_0}\right|$ is upper bounded by

$$\sup_{c\in\mathcal{C}_n}\left|d^T\hat{Q}_{n,c}^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^n Z_{i,c}(Z_{i,c_0}-Z_{i,\hat{c}_0})^T\theta_0\right| + \left|d^T\hat{Q}_{n,c_0}^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^n(Z_{i,\hat{c}_0}-Z_{i,c_0})\epsilon_i\right|$$
$$+\sup_{c\in\mathcal{C}_n}\left|d^T\hat{Q}_{n,\hat{c}_0}^{-1}(\hat{Q}_{n,\hat{c}_0}-\hat{Q}_{n,c_0}-(Q_{\hat{c}_0}-Q_{c_0}))\hat{Q}_{n,c_0}^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^n Z_{i,c}\epsilon_i\right|$$
$$+\sup_{c\in\mathcal{C}_n}\left|d^T\hat{Q}_{n,\hat{c}_0}^{-1}(Q_{\hat{c}_0}-Q_{c_0})\hat{Q}_{n,c_0}^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^n Z_{i,c}\epsilon_i\right|,$$

where the first term converges to zero in probability due to (S2.31) and (S2.16).

Further, due to (S3.49), $n^{-1/2-q}\sup_{c\in\mathcal{C}_n}|\sum_{i=1}^n Z_{i,c}\epsilon_i|$ converges to zero in probability for any $q>0$. Then due to Lemma S2.3, (S2.16), and the fact that $F'(c_0)>0$, the third term converges to zero in probability. In addition, since F is differentiable at c_0, and due to (S2.16), the fourth term converges to zero in probability.

It remains to show that the second term, $\left|d^T\hat{Q}_{n,c_0}^{-1}n^{-1/2}\sum_{i=1}^n(Z_{i,\hat{c}_0}-Z_{i,c_0})\epsilon_i\right|$, converges to zero in probability. Due to (S2.16) and Lemma S2.3, it suffices to show that as $n\to\infty$,

$$\sup_{c:|c-c_0|\le K_0 n^{-\eta_1}}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^n(X_{i,c}-X_{i,c_0})\epsilon_i\right| \xrightarrow{p} 0,$$
$$\sup_{c:|c-c_0|\le K_0 n^{-\eta_1}}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^n U_i(X_{i,c}-X_{i,c_0})\epsilon_i\right| \xrightarrow{p} 0. \tag{S4.55}$$

We focus on the proof of the second statement (S4.55), as the proof for the first is similar and simpler.

Define $\mathcal{H}_n = \{h_c : c \in [\ell, u] \cap |c - c_0| \leq K_0 n^{-\eta_1}\}$ where $h_c(x, t, \epsilon) = (I(x \leq c) - I(x \leq c_0))t\epsilon$ for $x, \epsilon \in \mathbb{R}$ and $t \in \{0, 1\}$. Let $H_n(x, t, \epsilon) = I\{|x - c_0| \leq K_0 n^{-\eta_1}\}|\epsilon|$ be an envelope function for $\mathcal{H}_n$.

Similar to the proof of Lemma S2.3, by first using symmetrization and then by (Van Der Vaart and Wellner, 1996, Lemma 2.3.1), we have

$$E\left(\sup_{c:|c-c_0|\leq K_0 n^{-\eta_1}} \left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n} U_i(X_{i,c} - X_{i,c_0})\epsilon_i\right|\right)$$
$$\leq K\int_0^1 \sup_Q \sqrt{1 + \log N(\zeta\|H_n\|_{Q,2}, \mathcal{H}_n, L_2(Q))}d\zeta\, E\left[\|H_n\|_{L_2(\mathbb{P}_n)}\right]$$

where $\|H_n\|^2_{L_2(\mathbb{P}_n)} = \frac{1}{n}\sum_{i=1}^{n} I\{|X_i - c_0| \leq K_0 n^{-\eta_1}\}\epsilon_i^2$. By (Chernozhukov et al., 2014), the uniform entropy integral is bounded in n, since $(\mathcal{H}_n, H_n)$ is a VC-type class. Further, by Jensen's inequality and due to (A.1) and (A.3),

$$E\left[\|H_n\|_{L_2(\mathbb{P}_n)}\right] \leq \|\epsilon_1\|_2\sqrt{\frac{1}{n}\sum_{i=1}^{n} I\{|X_i - c_0| \leq K_0 n^{-\eta_1}\}} \to 0, \quad \text{as } n \to \infty,$$

which completes the proof of (S4.55). □

S5 Discussions on Related Work

In this section we discuss the related work in Seijo and Sen (2011) mentioned in Section 1, and explain why the failure of the conventional bootstrap methods, such as the paired bootstrap and the residual bootstrap, does not contradict with our work.

Specifically, Seijo and Sen (2011) considers the following change-point model under the *random design*, assuming that the model is identifiable, i.e., $\gamma_0 \neq 0$:

$$Y = \alpha_0 + \gamma_0 I(X \leq c_0) + \epsilon. \tag{S5.56}$$

Let $(\hat{\alpha}_0, \hat{\gamma}_0, \hat{c}_0)$ be the least squares estimator:

$$(\hat{\alpha}_0, \hat{\gamma}_0, \hat{c}_0) = \underset{(\alpha,\gamma,c)\in\mathbb{R}^2*[\ell,u]}{\operatorname{argmin}} \; n^{-1} \sum_{i=1}^{n} \left(Y_i - \alpha - \gamma I(X_i \leq c)\right)^2.$$

It is well known (Koul et al., 2003; Seijo and Sen, 2011) that $n(\hat{c}_0 - c_0)$ converges in distribution to the minimizer of a compound Poisson process. But remarkably in Seijo and Sen (2011) and more explicitly in Yu (2014), it is shown that the bootstrap statistics, under either the paired or residual bootstrap, do not converge weakly. As a result, Seijo and Sen (2011) proposed smoothed bootstrap methods to construct valid confidence intervals for c_0.

However, we want to note that by standard arguments, both the paired and residual bootstrap, which are inconsistent for c_0, are in fact consistent for (α_0, γ_0). We conduct the following simulation study to help understand the failure of the paired bootstrap for c_0 and the success for α_0, γ_0. For each repetition, we generate $X_1, \ldots, X_n$ independently from the uniform distribution on $(0, 1)$, and the responses $Y_1, \ldots, Y_n$ are generated using (S5.56),

where $\epsilon_1, \ldots, \epsilon_n$ are generated independently from some distribution F_ϵ. We vary the sample size n, parameters $(\alpha_0, \gamma_0, c_0)$, and consider the following noise distributions: $F_\epsilon^{(1)}$ is the normal distribution with mean 0 and variance 4; $F_\epsilon^{(2)} = 2^{1/2}t(4)$, where $t(4)$ is the t-distribution with 4 degrees of freedom. Table S1 provides the estimated coverage proportions of the nominal 90% and 95% confidence intervals obtained by the paired bootstrap with repetitions $R = 1000$. From Table S1, although the paired bootstrap fails in constructing the confidence intervals for c_0, it constructs valid confidence intervals for α_0, γ_0. This illustrates the fact that for the same model, bootstrap methods may *work for some statistics, but fail for others.*

In contrast, our goal is to test whether the interaction term in (2.1) is significant, i.e., $H_0 : \lambda_0 = 0$, by minimal p-value methods (2.3), under both the identifiable and non-identifiable case, for both the random and fixed design. For the random design, we propose using the paired bootstrap, which does not explicitly estimate c_0. For the fixed design, we propose a residual bootstrap, which uses the least squares estimator $\hat{c}_0$ (4.1) for c_0. However, in the proof the key property for the identifiable case is that $n^{\eta_1}|\hat{c}_0 - c_0| \to 0$ almost surely, where η_1 appears in Assumption (A.3), and can be arbitrarily close to 1 for the examples in Section 4.3. That is, we

Table S1: The estimated coverage probabilities of nominal 90% and 95% CIs for the regression and change-point parameters.

		$F_\epsilon^{(1)}$						$F_\epsilon^{(2)}$					
		90%			95%			90%			95%		
n	$(\alpha_0, \gamma_0, c_0)$	α_0	γ_0	c_0	α_0	γ_0	c_0	α_0	γ_0	c_0	α_0	γ_0	c_0
300	$(0, 3, 0.3)$	89.9	88.5	69.1	94.6	93.4	77.8	89.7	88.9	63.9	93.7	93.5	73.1
	$(0, 3, 0.5)$	88.5	89.5	67.4	94.0	94.5	76.9	90.4	87.9	65.9	95.5	94.1	74.3
	$(2, 2, 0.3)$	90.3	90.9	77.3	95.7	95.2	85.8	90.1	88.3	76.8	94.5	93.4	84.6
	$(2, 2, 0.5)$	90.6	90.3	76.7	95.6	95.4	84.7	87.7	91.4	64.6	93.6	94.7	74.6
500	$(0, 3, 0.3)$	91.1	89.5	65.1	95.3	94.9	74.4	87.7	91.4	64.6	93.6	94.7	74.6
	$(0, 3, 0.5)$	89.6	90.0	66.0	94.4	94.6	75.7	89.3	90.0	61.0	94.6	95.5	70.6
	$(2, 2, 0.3)$	90.2	90.1	77.9	94.5	95.7	85.0	90.2	90.6	75.3	95.7	95.0	82.8
	$(2, 2, 0.5)$	90.2	90.1	76.3	95.1	95.0	84.2	88.5	89.5	74.7	94.1	94.5	81.6

do not require the bootstrap consistency for c_0, and the convergence result for $\hat{c}_0$ is similar to those in (Seijo and Sen, 2011), with the difference due to the fixed design.

S6 More Simulation Results

S6.1 Simulation results under the random design

In this subsection we conduct simulation studies to compare the performance of the tests including, the paired bootstrap test with $p^*_{n,pb}$ in (3.2), the profile least squares estimation based test with $p_{n,pf}$ in (4.6), and the

unadjusted minimum p-value test with $p_{n,mp}$ in (1.3), under the random design setup. In the following tables, "PB", "PF" and "MP" stand for the tests based on $p^*_{n,pb}$, $p_{n,pf}$ and $p_{n,mp}$, respectively.

Table S2: The empirical sizes (in percentage) for testing $H_0 : \lambda_0 = 0$ at level 5% under the random design. Here, $\theta^{(1)} = (0, 1, 3, 0)^T$ and $\theta^{(2)} = (2, 1.5, 1, 0)^T$. The bootstrap repetition for $p^*_{n,pb}$ is $B = 2000$

			$F_\epsilon^{(1)}$			$F_\epsilon^{(2)}$			$F_\epsilon^{(3)}$		
n	θ_0	c_0	PB	PF	MP	PB	PF	MP	PB	PF	MP
100	$\theta^{(1)}$	0.3	6.2	5.6	31.1	4.9	6.7	32.4	5.5	5.9	32.4
		0.5	5.5	5.3	29.2	4.7	6.2	36.8	5.3	5.8	28.5
		0.7	6.0	6.1	32.9	4.9	6.1	32.7	5.2	6.7	31.7
	$\theta^{(2)}$	0.3	4.7	18.7	34.4	4.6	16.6	35.0	5.3	20.0	34.7
		0.5	4.9	16.9	34.3	5.2	17.1	34.6	4.8	15.9	33.9
		0.7	4.8	18.6	34.8	4.9	18.9	32.7	4.9	16.7	34.6
300	$\theta^{(1)}$	0.3	5.3	5.1	30.9	4.7	5.6	31.2	5.0	4.8	32.9
		0.5	5.0	4.9	29.6	4.7	5.3	29.9	5.2	4.6	31.1
		0.7	5.4	5.2	33.1	5.2	5.6	34.5	5.9	5.2	33.3
	$\theta^{(2)}$	0.3	5.1	10.0	37.4	5.6	10.1	38.4	4.3	10.3	36.9
		0.5	4.9	9.5	36.6	5.3	8.7	37.9	5.0	8.2	36.9
		0.7	5.2	9.4	35.6	4.9	9.7	34.6	4.3	8.9	37.5

Here we consider the same combinations of $n,\ c_0,\ \theta_0$ and noise distribu-

Table S3: The empirical powers (in percentage) for testing $H_0 : \lambda_0 = 0$ at level 5% under the random design. Here, $\theta^{(3)} = (0, 1, 3, 2)^T$ and $\theta^{(4)} = (2, 1.5, 1, 2)^T$. The bootstrap repetition for $p^*_{n,pb}$ is $B = 2000$

			$F_\epsilon^{(1)}$			$F_\epsilon^{(2)}$			$F_\epsilon^{(3)}$		
n	θ_0	c_0	PB	PF	MP	PB	PF	MP	PB	PF	MP
100	$\theta^{(3)}$	0.3	47.9	59.7	77.6	50.2	62.4	79.6	47.9	62.5	79.5
		0.5	58.0	69.1	83.2	61.8	71.6	85.5	59.7	69.2	82.7
		0.7	49.5	59.1	79.7	51.3	64.2	81.7	49.1	61.6	81.2
	$\theta^{(4)}$	0.3	43.4	60.7	77.8	46.7	64.1	80.6	44.0	64.2	79.1
		0.5	55.1	70.6	83.9	55.5	70.9	85.4	54.8	68.7	84.5
		0.7	43.8	61.9	80.0	49.1	62.9	82.3	44.1	64.4	79.5
300	$\theta^{(3)}$	0.3	93.1	97.5	98.3	92.8	96.8	99.4	93.1	98.0	99.3
		0.5	97.2	99.1	99.6	96.8	98.4	99.6	96.5	99.1	99.4
		0.7	92.5	97.6	98.8	92.3	97.4	98.9	92.5	97.4	99.1
	$\theta^{(4)}$	0.3	92.3	97.4	99.5	92.5	96.7	99.1	92.4	96.6	98.8
		0.5	96.3	99.3	99.4	95.2	98.9	99.6	96.3	98.6	99.7
		0.7	91.9	97.5	98.4	91.5	97.4	98.8	92.1	96.6	99.5

tions with 2000 repetitions as in the simulation study for the fixed design. Under a random design, the data generation is similar to that under a fixed design except that a new sample $(X_i, U_i, \epsilon_i),\ i \in [n]$ is independently and identically generated in each repetition of a simulation.

From Tables S2 and S3, the simulation results under the random design is similar to that under the fixed design. The empirical sizes of the tests based on the bootstrap adjustment, $p^*_{n,pb}$, are close to the nominal 5% level, even when F_ϵ has a heavy tail or is non-symmetric. In comparison, there is an almost 7 times inflation if the unadjusted version, $p_{n,mp}$, is used. Further, when the effect of $1\{X_i \leq c_0\}$ (i.e. γ_0 in (2.1)) is small, the profile least squares estimation based tests, $p_{n,pf}$, controls poorly the Type I error (cf. $\theta^{(2)}$ in Table S2). The empirical powers of tests based on the bootstrap adjustments, $p^*_{n,pb}$, are not as large as the other two, but the gap is mild when the sample size is moderate (say ~ 300).

Table S4 presents the empirical sizes and powers of the three tests under the *non-identifiable* case. From the table, the tests based on $p_{n,pf}$ and $p_{n,mp}$ lose control of the empirical sizes whereas the test based on the bootstrap adjustments $p^*_{n,pb}$ behaves satisfactorily as in the identifiable case. The empirical powers are close for the three tests.

Table S4: The empirical sizes and powers (in percentage) for testing $H_0 : \lambda_0 = 0$ at the level 5%. Here, $\theta^{(5)} = (0, 1, 0, 0)^T$, $\theta^{(6)} = (2, 1.5, 0, 0)^T$, $\theta^{(7)} = (0, 1, 0, 2)^T$, $\theta^{(8)} = (2, 1.5, 0, 2)^T$.

			$F_\epsilon^{(1)}$			$F_\epsilon^{(2)}$			$F_\epsilon^{(3)}$		
n	θ_0	c_0	PB	PF	MP	PB	PF	MP	PB	PF	MP
sizes											
300	$\theta^{(5)}$	0.3	4.7	29.9	36.7	5.4	30.7	39.2	6.1	32.1	38.5
		0.5	6.1	30.7	40.0	4.5	33.4	38.6	5.7	29.7	38.2
		0.7	5.0	28.6	39.6	4.7	30.4	37.2	4.7	29.4	37.5
	$\theta^{(6)}$	0.3	6.1	29.6	38.7	4.9	31.5	38.7	4.7	30.6	37.5
		0.5	5.0	32.2	38.1	4.9	30.0	36.6	4.9	30.7	38.8
		0.7	6.8	30.3	40.7	5.4	29.3	36.8	5.2	29.7	38.2
powers											
300	$\theta^{(7)}$	0.3	93.4	97.4	99.3	92.2	97.3	99.2	92.8	97.2	99.4
		0.5	96.2	98.9	99.7	96.2	98.8	99.6	96.3	99.1	99.5
		0.7	93.0	97.4	99.3	91.5	96.8	98.9	92.2	97.0	99.3
	$\theta^{(8)}$	0.3	92.4	96.7	99.3	92.0	96.9	99.2	93.1	97.6	99.1
		0.5	96.4	98.9	99.9	94.8	98.6	99.8	96.1	98.9	99.7
		0.7	93.0	97.1	99.3	91.6	96.9	98.9	91.7	97.2	99.3

S6.2 Simulation results with heavy tailed error distributions

In this subsection we present simulation results to compare the performance of the tests based on $p^*_{n,mrb}$, $p_{n,pf}$ and $p_{n,mp}$ for some heavy-tailed error distributions under the fixed design setup, where "MRB", "PF" and "MP" stand for the tests based on $p^*_{n,mrb}$, $p_{n,pf}$ and $p_{n,mp}$, respectively. Specifically we consider the following: $F_\epsilon^{(4)}$ is Frechet distribution with the scale and shape parameters being 5 and 5.5; $F_\epsilon^{(5)}$ is the log-logistic distribution with the scale and shape parameters being $\log(6)$ and $1/6$; $F_\epsilon^{(6)}$ is the log-normal distribution with the location parameter 0 and scale parameter 1. Among these heavy-tailed distributions, the moments for $F_\epsilon^{(6)}$ of all orders exist, whereas any moment of order greater than 5 does not exist for $F_\epsilon^{(4)}$ and $F_\epsilon^{(5)}$. Note that although $F_\epsilon^{(2)} = \sqrt{2}t(4)$ has an infinite fourth moment, which violates Assumption (A.1), the performance of the proposed MRB method is satisfactory (Section 5 in the main text and Section S6.1).

From Tables S5 and S6, the proposed MRB test obtains empirical sizes close to the nominal level, whereas the size of the MP method is significantly inflated, under both the identifiable and non-identifiable cases. Further, although the PF method behaves reasonably well in the identifiable case with a large main effect (large $|\gamma_0|$), its empirical sizes are not well controlled under the non-identifiable case or when $|\gamma_0|$ is small.

Table S5: The empirical sizes and powers (in percentage) for testing $H_0 : \lambda_0 = 0$ at the level 5% under the identifiable case. Here, $\theta^{(1)} = (0, 1, 3, 0)^T$, $\theta^{(2)} = (2, 1.5, 1, 0)^T$ are for sizes, and $\theta^{(3)} = (0, 1, 3, 2)^T$, $\theta^{(4)} = (2, 1.5, 1, 2)^T$ are for powers.

			$F_\epsilon^{(4)}$			$F_\epsilon^{(5)}$			$F_\epsilon^{(6)}$		
n	θ_0	c_0	MRB	PF	MP	MRB	PF	MP	MRB	PF	MP
100	$\theta^{(1)}$	0.3	7.0	6.0	29.9	6.8	6.6	28.3	6.4	5.6	32.5
		0.5	5.9	6.1	28.8	6.7	5.3	30.2	5.5	4.9	39.4
	$\theta^{(2)}$	0.3	6.0	18.9	35.4	6.8	20.8	31.3	6.2	23.0	28.6
		0.5	5.8	17.6	28.0	7.3	18.9	36.4	5.2	15.3	32.2
300	$\theta^{(1)}$	0.3	6.3	4.4	40.0	4.6	4.9	37.1	5.3	5.7	46.5
		0.5	4.4	5.7	33.4	5.8	5.2	32.1	5.7	4.8	42.1
	$\theta^{(2)}$	0.3	5.8	10.9	39.5	4.9	10.8	35.1	5.3	9.6	36.1
		0.5	5.0	7.2	34.1	5.0	8.9	36.0	4.5	8.2	34.0
100	$\theta^{(3)}$	0.3	57.6	69.8	76.4	58.5	65.8	77.8	65.8	71.7	84.9
		0.5	70.4	70.4	88.5	67.3	69.3	83.0	73.1	75.5	84.5
	$\theta^{(4)}$	0.3	62.6	66.5	86.1	57.2	61.3	82.3	64.0	72.8	84.8
		0.5	64.1	71.9	83.8	62.0	72.5	83.3	71.4	74.8	87.9
300	$\theta^{(3)}$	0.3	95.3	96.1	99.5	93.9	97.1	98.6	89.7	95.7	99.7
		0.5	97.1	98.2	99.5	97.7	99.1	99.8	96.3	97.5	99.3
	$\theta^{(4)}$	0.3	94.5	95.9	99.4	93.0	97.8	99.2	92.1	96.4	98.6
		0.5	96.6	97.7	99.5	97.3	98.1	99.6	95.2	98.2	99.0

Table S6: The empirical sizes and powers (in percentage) for testing $H_0 : \lambda_0 = 0$ at the level 5% under the non-identifiable case. Here, $\theta^{(5)} = (0, 1, 0, 0)^T$, $\theta^{(6)} = (2, 1.5, 0, 0)^T$ are for sizes, and $\theta^{(7)} = (0, 1, 0, 2)^T$, $\theta^{(8)} = (2, 1.5, 0, 2)^T$ are for powers.

			$F_\epsilon^{(4)}$			$F_\epsilon^{(5)}$			$F_\epsilon^{(6)}$		
n	θ_0	c_0	MRB	PF	MP	MRB	PF	MP	MRB	PF	MP
300	$\theta^{(5)}$	0.3	5.4	30.9	37.6	6.4	30.4	36.7	4.9	29.8	37.0
		0.5	5.1	33.5	35.6	5.6	27.4	38.6	5.4	30.9	34.7
	$\theta^{(6)}$	0.3	5.0	29.4	38.8	5.6	29.8	37.4	4.4	30.2	35.5
		0.5	5.5	30.5	36.0	5.2	29.9	36.5	5.3	29.5	37.7
	$\theta^{(7)}$	0.3	94.7	95.3	99.5	91.3	97.1	99.2	91.0	96.0	98.5
		0.5	95.5	98.5	99.5	96.2	98.9	99.6	94.3	98.1	99.2
	$\theta^{(8)}$	0.3	92.2	97.0	98.9	93.2	97.6	99.2	91.4	96.4	98.4
		0.5	96.1	97.8	99.7	96.4	98.7	99.8	95.6	97.3	99.3

S6.3 More results on the application

We fit Model 2.1 with c_0 being the estimated cutpoints respectively for EREG, LDH, and ALKPH and results are presented in Table S7. It is known the p-values for λ_0 in this table are the p-values from the minimum p-values method.

Bibliography

Chernozhukov, V., D. Chetverikov, and K. Kato (2014). Gaussian approximation of suprema of empirical processes. *Ann. Statist. 42*(4), 1564–1597.

Chernozhukov, V., D. Chetverikov, and K. Kato (2017). Central limit theorems and bootstrap in high dimensions. *Ann. Prob. 45*(4), 2309–2352.

Chernozhukov, V., D. Chetverikov, K. Kato, and Y. Koike (2019). Improved central limit theorem and bootstrap approximations in high dimensions. *arXiv*, 1912.10529.

Davydov, Y. A., M. A. Lifshits, and N. V. Smorodina (1998). *Local Properties of Distributions of Stochastic Functionals.* American Mathematical Soc.

Table S7: Fitted interaction threshold models with, respectively, EREG, LDH and ALKPH as the potential predictive biomarker based on the estimated cutoff points derived from the minimum p-value method

	Estimate	**Std.Err**	t **value**	p**-value**
EREG				
α_0	0.0346	0.0677	0.512	0.6089
β_0	-0.0877	0.0464	-1.893	0.0595
γ_0	-0.5473	0.1015	-5.391	7.01×10^{-8}
λ_0	-0.2980	0.0681	-4.375	1.22×10^{-5}
LDH				
α_0	-0.3778	0.1073	-3.519	0.0004
β_0	-0.3158	0.0731	-4.321	2.5×10^{-5}
γ_0	0.2527	0.1178	2.146	0.0326
λ_0	0.1420	0.0799	1.777	0.0765
ALKPH				
α_0	-0.1323	0.0662	-1.998	0.0465
β_0	-0.1738	0.0465	-3.738	0.0002
γ_0	-0.0863	0.0876	-0.985	0.3253
λ_0	-0.0505	0.0599	-0.842	0.4001

Kosorok, M. R. (2007). *Introduction to Empirical Processes and Semiparametric Inference*. Springer Science & Business Media.

Koul, H. L., L. Qian, and D. Surgailis (2003). Asymptotics of M-estimators in two-phase linear regression models. *Stoch. Proces. Appl. 103*(1), 123–154.

Ledoux, M. and M. Talagrand (1991). *Probability in Banach Spaces: Isoperimetry and Processes*, Volume 23. Springer Science & Business Media.

Nazarov, F. (2003). On the maximal perimeter of a convex set in $\mathbb{R}^n$ with respect to a Gaussian measure. In M. V.D. and S. G. (Eds.), *Geometric Aspects of Functional Analysis: Israel Seminar 2001-2002*, Volume 1807, pp. 169–187. Springer Berlin Heidelberg.

Neumann, C. G. (1867). *Theorie der Besselschen Funktionen*. Teubner Verlag.

Seijo, E. and B. Sen (2011). Change-point in stochastic design regression and the bootstrap. *Ann. Statist. 39*(3), 1580–1607.

Van Der Vaart, A. and J. Wellner (1996). *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer-Verlag New York.

Van der Vaart, A. W. (1998). *Asymptotic Statistics*, Volume 3. Cambridge university press.

Yu, P. (2014). The bootstrap in threshold regression. *Econometric Theory 30*(3), 676–714.