Supplementary Material to "Regularized projection score estimation of treatment effects in high-dimensional quantile regression"

Chao Cheng[a], Xingdong Feng[a], Jian Huang[b], Xu Liu[a]

[a]*Shanghai University of Finance and Economics*

[b]*University of Iowa*

Supplementary Material

This Supplementary Material includes the detailed proofs of Theorems 1 and 2, and related Lemmas 1–2 in Section 1. We prove Theorems 3 and 4 and related Lemmas in Sections 2 and 3, respectively. In Section 4, the comparison of final sample performance between standard lasso and group lasso is reported by some simulation examples.

S1 Supplementary Material A: Proofs of Theorem 1 and 2

To facilitate expression, we introduce some additional notation. Let $\|M\|_{2,1} = \sum_{j=1}^{p} \|m_j\|$ for any matrix $M \in \mathbb{R}^{n\times p}$, where m_j is the jth column of M, and $\|v\|$ is the standard L_2 norm for any vector $v \in \mathbb{R}^p$. For an index set $S \in \{1, \cdots, p\}$ and a matrix $M \in \mathbb{R}^{n\times p}$, M_S denotes the submatrix of M containing columns of M with indices in S. For a vector v, v_S denotes the subvector of v containing elements of v with indices in S.

Lemma 1. *In the event* $\Omega = \{\max_{1\le j\le q} \|n^{-1}\sum_{i=1}^{n} z_{ij}(x_i - H_0 z_i)\| \le \lambda_2\}$,

$$\frac{1}{2n}\sum_{i=1}^{n} \|(\tilde{H} - H_0)z_i\|^2 \le 2\lambda_2 \sum_{j=1}^{q} \|h_{0j}\|,$$

where h_{0j} *is the* j*th column of* H_0 *in (6) in the main paper.*

Proof of Lemma 1. By the definition of $\tilde{H}$, we have

$$\frac{1}{2n}\sum_{i=1}^{n} \|x_i - \tilde{H}z_i\|^2 + \lambda_2 \sum_{j=1}^{q} \|\tilde{h}_j\| \le \frac{1}{2n}\sum_{i=1}^{n} \|x_i - H_0 z_i\|^2 + \lambda_2 \sum_{j=1}^{q} \|h_{0j}\|.$$

Let $\tilde{\Delta} = (\tilde{H} - H_0)$. After some algebra, this inequality can be written as

$$\frac{1}{2n}\sum_{i=1}^{n} \|\tilde{\Delta}z_i\|^2 \le \frac{1}{n}\sum_{i=1}^{n} z_i'\tilde{\Delta}(x_i - H_0 z_i) + \lambda_2 \sum_{j=1}^{q}(\|h_{0j}\| - \|\tilde{h}_j\|).$$

Note that

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n} z_i'\tilde{\Delta}'(x_i - H_0 z_i) &= \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{q} \tilde{\delta}_j'(x_i - H_0 z_i) z_{ij} \\ &\le \sum_{j=1}^{q} \|\tilde{\delta}_j\| \|\frac{1}{n}\sum_{i=1}^{n} z_{ij}(x_i - H_0 z_i)\|,\end{aligned}$$

where $\tilde{\delta}_j = \tilde{h}_j - h_{0j}$ is the jth column of $\tilde{\Delta}$. Therefore, on the set Ω

$$\frac{1}{2n}\sum_{i=1}^{n}\|\tilde{\Delta}z_i\|^2 \le \lambda_2\sum_{j=1}^{q}(\|\tilde{h}_j\| + \|h_{0j}\|) + \lambda_2\sum_{j=1}^{q}(\|h_{0j}\| - \|\tilde{h}_j\|) = 2\lambda_2\sum_{j=1}^{q}\|h_{0j}\|.$$

This proves Lemma 1. $\diamond$

Lemma 2. *If Assumptions (A1)–(A5) hold, and* $\lambda_2 \ge 2\rho(\Sigma)\sqrt{(2+\delta)\log(2dq)/n}$*, then we have, with probability approaching one,*

$$\max_{1\le k\le d}\|(\tilde{H} - H_0)^{(k)}\|_1 \lesssim \sqrt{d}s_h\lambda_2,$$

$$\max_{1\le k\le d}\|(\tilde{H} - H_0)^{(k)}\| \lesssim \sqrt{s_h}\lambda_2,$$

$$\|\tilde{\eta} - \eta_0\| \lesssim \sqrt{s}\lambda_1,$$

where $\rho(\Sigma) = \max_{1\le j\le d+q}\Sigma_{jj}$*, and* $\delta > 0$ *is a constant.*

Proof of Lemma 2. We show the first two inequalities in both cases, that is, $\tilde{H}$ is obtained from (2.8) and (2.9) in the main paper. Then we show the last inequality.

By Theorem 1 of Raskutti et al. (2010), we have, with probability at least $1 - c'\exp(-cn)$,

$$(v'\hat{\Sigma}_z v)^{1/2} \ge \frac{1}{4}\|\Sigma_z^{1/2}v\| - 9\rho(\Sigma_z)\|v\|\sqrt{\log(q)/n}, \quad \text{for all } v \in \mathbb{R}^q,$$

where c and c' are constants, Σ_z is the covariance of z, $\hat{\Sigma}_z$ is the sample covariance of z, and $\rho(\Sigma) = \max_{1\le j\le q}\Sigma_{jj}$.

For an index set $S \subset \{1, \cdots, q\}$ and a constant $\alpha > 1$, define the cone $\mathcal{C}(S,\alpha) = \{\theta \in \mathbb{R}^q : \|\theta_{S^c}\|_1 \le \alpha\|\theta_S\|_1\}$. We say a symmetric matrix M satisfies the restricted eigenvalue (RE) condition over S with parameters $(\alpha, \gamma) \in [1, \infty) \times (0, \infty)$ if

$$v' M v \geq \gamma^2 \|v\|^2, \quad \text{for all } v \in \mathcal{C}(S, \alpha).$$

Similar to Corollary 1 of Raskutti et al. (2010), when sample size

$$n > c'' \frac{\rho^2(\Sigma)(1+\alpha)^2}{c_\Lambda} |S| \log(q),$$

the matrix $\hat{\Sigma}_z$ satisfies the RE condition with parameters $(\alpha, c_\Lambda/8)$, that is,

$$(v'\hat{\Sigma}_z v)^{1/2} \geq \frac{c_\Lambda}{8} \|v\|, \quad \text{for all } v \in \mathcal{C}(S, \alpha),$$

where c_Λ is given in the assumption (A1).

Case 1: $\tilde{H}$ is obtained from (2.9) in the main paper. Note that the conditional variance $\mathrm{Var}(\sum_{i=1}^n z_{ij}(x_i - H_0 z_i)^{(k)} | z_1, \cdots, z_n) = \Sigma_{kk} \sum_{i=1}^n z_{ij}^2$. We have

$$\begin{aligned}
&P(\max_{1\leq j \leq q} \|\frac{1}{n}\sum_{i=1}^n z_{ij}(x_i - H_0 z_i)\| > t | z_1, \cdots, z_n) \\
&\quad \leq dq \max_{1\leq k\leq d} \max_{1\leq j\leq q} P\left(\left| \frac{1}{n}\sum_{i=1}^n z_{ij}(x_i - H_0 z_i)^{(k)} \right| > t/\sqrt{d} | z_1, \cdots, z_n \right) \\
&\quad \leq 2dq \max_{1\leq k\leq d} \max_{1\leq j\leq q} \exp\left(- \frac{n^2 t^2}{2d\Sigma_{kk}\sum_{i=1}^n z_{ij}^2} \right) \\
&\quad \leq 2dq \exp\left(-\frac{n^2t^2}{2dv_0} \right) \\
&\quad = \exp\left(-\frac{n^2 t^2}{2dv_0} + \log(2dq) \right),
\end{aligned}$$

where $v_0 = \max_{1\leq k\leq d} \max_{1\leq j\leq q} \Sigma_{kk} \sum_{i=1}^n z_{ij}^2 \leq \rho(\Sigma) \sum_{i=1}^n z_{ij}^2$, and $\rho(\Sigma) = \max_{1\leq j\leq d+p} \Sigma_{jj}$.

Since $z_{ij}^2/\Sigma_{jj} - 1$ is sub-exponential with mean 0, we have

$$P(|n^{-1}\sum_{i=1}^n (z_{ij}^2/\Sigma_{jj} - 1)| > t) \leq 2\exp\left(-cn \min\left(\frac{t^2}{K^2}, \frac{t}{K} \right) \right),$$

where $K = \max_{1\le i\le n}\|z_{ij}^2/\Sigma_{jj} - 1\|_{\psi}$. This implies that $\frac{1}{n}\sum_{i=1}^{n}(z_{ij}^2/\Sigma_{jj} - 1) \le \frac{1}{n}$ with probability approaching one, which results in $v_0 \le (n+1)\rho^2(\Sigma)$ with probability approaching one. Thus, we have

$$\max_{1\le j\le q}\|\frac{1}{n}\sum_{i=1}^{n} z_{ij}(x_i - H_0 z_i)\| \le \rho(\Sigma)\sqrt{\frac{(2+\delta)\log(2dq)}{n}} \le \frac{1}{2}\lambda_2 \tag{S1.1}$$

with probability approaching one, where $\delta > 0$ is a constant. By Cauchy-Schwarz inequality, we have

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n} z_i'\tilde{\Delta}'(x_i - H_0 z_i) \le& \sum_{j=1}^{q}\|\delta_j\|\|\frac{1}{n}\sum_{i=1}^{n} z_{ij}(x_i - H_0 z_i)\| \\ \le& \|\tilde{\Delta}\|_{2,1}\max_{1\le j\le q}\|\frac{1}{n}\sum_{i=1}^{n} z_{ij}(x_i - H_0 z_i)\| \\ \le& \frac{1}{2}\lambda_2\|\tilde{\Delta}\|_{2,1}.\end{aligned}$$

Recall $S_h = \{j : h_{0j} \ne 0, j = 1, \cdots, q\}$, which is defined in the assumption (A3). Since $H_{0S_h^c} = 0$, we have $\|H_0\|_{2,1} = \|H_{0S_h}\|_{2,1}$, and

$$\|H_0 + \tilde{\Delta}\|_{2,1} = \|H_{0S} + \tilde{\Delta}_S\|_{2,1} + \|\tilde{\Delta}_{S^c}\|_{2,1} \ge \|H_{0S}\|_{2,1} - \|\tilde{\Delta}_S\|_{2,1} + \|\tilde{\Delta}_{S^c}\|_{2,1}.$$

From the proof of Lemma 1, we have

$$\begin{aligned}0 \le \frac{1}{2n}\sum_{i=1}^{n}\|\tilde{\Delta} z_i\|^2 \le& \frac{1}{n}\sum_{i=1}^{n} z_i'\tilde{\Delta}(x_i - H_0 z_i) + \lambda_2(\|H_0\|_{2,1} - \|\tilde{H}\|_{2,1}) \\ \le& \frac{1}{n}\sum_{i=1}^{n} z_i'\tilde{\Delta}(x_i - H_0 z_i) + \lambda_2\|\tilde{\Delta}_{S_h}\|_{2,1} - \lambda_2\|\tilde{\Delta}_{S_h^c}\|_{2,1} \\ \le& \frac{1}{2}\lambda_2\|\tilde{\Delta}\|_{2,1} + \lambda_2\|\tilde{\Delta}_{S_h}\|_{2,1} - \lambda_2\|\tilde{\Delta}_{S_h^c}\|_{2,1} \\ \le& \frac{3}{2}\lambda_2\|\tilde{\Delta}_{S_h}\|_{2,1} - \frac{1}{2}\lambda_2\|\tilde{\Delta}_{S_h^c}\|_{2,1},\end{aligned} \tag{S1.2}$$

which implies that $\|\tilde{\Delta}_{S_h^c}\|_{2,1} \le 3\|\tilde{\Delta}_{S_h}\|_{2,1}$ and $\tilde{\Delta} \in \mathcal{C}(S_h, 3)$, and consequently that

$$\begin{aligned}
\frac{dc_\Lambda^2}{64}\|\tilde{\Delta}\|^2 \le & \frac{1}{2n}\sum_{i=1}^n \|\tilde{\Delta} z_i\|^2 \\
\le & \frac{1}{2}\lambda_2\|\tilde{\Delta}\|_{2,1} + \lambda_2(\|H_0\|_{2,1} - \|\tilde{H}\|_{2,1}) \\
\le & \frac{1}{2}\lambda_2\|\tilde{\Delta}\|_{2,1} + \lambda_2\|\tilde{\Delta}_{S_h}\|_{2,1} - \lambda_2\|\tilde{\Delta}_{S_h^c}\|_{2,1} \\
\le & \frac{3}{2}\lambda_2\|\tilde{\Delta}_{S_h}\|_{2,1} - \frac{1}{2}\lambda_2\|\tilde{\Delta}_{S_h^c}\|_{2,1} \\
\le & \frac{3}{2}\lambda_2\sqrt{s_h}\|\tilde{\Delta}\|.
\end{aligned}$$

This concludes that

$$\|\tilde{\Delta}\| \le \frac{96}{dc_\Lambda^2}\sqrt{s_h}\lambda_2,$$

and further

$$\|\tilde{\Delta}\|_{2,1} \le 4\|\tilde{\Delta}_{S_h}\|_{2,1} \le 4\sqrt{s_h}\|\tilde{\Delta}\| \lesssim s_h\lambda_2,$$

which implies that with probability approaching one,

$$\max_{1\le k\le d}\|(H_0 - \tilde{H})^{(k)}\|_1 \le \sum_{k=1}^d \|(H_0 - \tilde{H})^{(k)}\|_1 \le \sqrt{d}\|\tilde{\Delta}\|_{2,1} \lesssim \sqrt{d}s_h\lambda_2,$$

and

$$\max_{1\le k\le d}\|(H_0 - \tilde{H})^{(k)}\|^2 \le \sum_{j=1}^q \|h_{0j} - \tilde{h}_j\|^2 = \|\tilde{\Delta}\|^2 \lesssim s_h\lambda_2^2.$$

Case 2: $\tilde{H}$ is obtained by (2.8) in the main paper, which implies that $\tilde{H}$ consists of d standard lasso estimators. Similar to (S1.1), we have

$$\max_{1\le k\le d}\max_{1\le j\le q}|\frac{1}{n}\sum_{i=1}^n z_{ij}(x_{ik} - H_0^{(k)} z_i)| \le \rho(\Sigma)\sqrt{\frac{(2+\delta)\log(2q)}{n}} \le \frac{1}{2}\lambda_2.$$

Recall that $S_{h,k} = \{j : h_{0kj} \neq 0, 1 \leq j \leq q\}$ for $1 \leq k \leq d$, which is defined in the assumption (A3), where h_{0kj} is the (k,j)th element of H_0. Similar to Lemma 1, it is easy to see that by Cauchy-Schwarz inequality,

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n}\tilde{\Delta}^{(k)}z_i(x_{ik}-H_0^{(k)}z_i) \leq& \|\tilde{\Delta}^{(k)}\|_1 \max_{1\leq j\leq q}\|\frac{1}{n}\sum_{i=1}^{n}z_{ij}(x_{ik}-H_0^{(k)}z_i)\|\\ \leq& \frac{1}{2}\lambda_2\|\tilde{\Delta}^{(k)}\|_1,\end{aligned}$$

and

$$\begin{aligned}\frac{1}{2n}\sum_{i=1}^{n}(\tilde{\Delta}^{(k)}z_i)^2 \leq& \frac{1}{n}\sum_{i=1}^{n}\tilde{\Delta}^{(k)}z_i(x_{ik}-H_0^{(k)}z_i)+\lambda_2(\|H_0^{(k)}\|_1-\|\tilde{H}^{(k)}\|_1)\\ \leq& \frac{1}{2}\lambda_2\|\tilde{\Delta}^{(k)}\|_1+\lambda_2\|\tilde{\Delta}^{(k)}_{S_{h,k}}\|_1-\lambda_2\|\tilde{\Delta}^{(k)}_{S^c_{h,k}}\|_1\\ \leq& \frac{3}{2}\lambda_2\|\tilde{\Delta}^{(k)}_{S_{h,k}}\|_1-\frac{1}{2}\lambda_2\|\tilde{\Delta}^{(k)}_{S^c_{h,k}}\|_1.\end{aligned}$$

which implies that $\tilde{\Delta}^{(k)} \in \mathcal{C}(S_{h,k}, 3)$. and consequently

$$\frac{c_\Lambda^2}{64}\|\tilde{\Delta}\|^2 \leq \frac{1}{2n}\sum_{i=1}^{n}(\tilde{\Delta}^{(k)}z_i)^2.$$

It follows from above that

$$\begin{aligned}\frac{c_\Lambda^2}{64}\|\tilde{\Delta}^{(k)}\|^2 \leq& \frac{1}{n}\sum_{i=1}^{n}\tilde{\Delta}^{(k)}z_i(x_{ik}-H_0^{(k)}z_i)+\lambda_2(\|H_0^{(k)}\|_1-\|\tilde{H}^{(k)}\|_1)\\ \leq& \frac{1}{2}\lambda_2\|\tilde{\Delta}^{(k)}\|_1+\lambda_2\|\tilde{\Delta}^{(k)}_{S_{h,k}}\|_1-\lambda_2\|\tilde{\Delta}^{(k)}_{S^c_{h,k}}\|_1\\ \leq& \frac{3}{2}\lambda_2\|\tilde{\Delta}^{(k)}_{S_{h,k}}\|_1\\ \leq& \frac{3}{2}\lambda_2\sqrt{s_{h,k}}\|\tilde{\Delta}^{(k)}\|,\end{aligned}$$

and then

$$\|\tilde{\Delta}^{(k)}\| \leq \frac{96}{c_\Lambda^2}\lambda_2\sqrt{s_{h,k}},$$

and further

$$\|\tilde{\Delta}^{(k)}\|_1 \leq 4\|\tilde{\Delta}^{(k)}_{S_{h,k}}\|_1 \leq 4\sqrt{s_{h,k}}\|\tilde{\Delta}^{(k)}\| \leq s_{h,k}\lambda_2.$$

It is obvious that we have, with probability approaching one,

$$\max_{1\leq k\leq d}\|(H_0 - \tilde{H})^{(k)}\|_1 = \max_{1\leq k\leq d}\|\tilde{\Delta}^{(k)}\|_1 \lesssim s_h\lambda_2,$$

and

$$\max_{1\leq k\leq d}\|(H_0 - \tilde{H})^{(k)}\| = \max_{1\leq k\leq d}\|\tilde{\Delta}^{(k)}\| \lesssim \sqrt{s_h}\lambda_2.$$

For the last inequality, we need to verify Conditions D1–D4 of Theorem 2 of Belloni and Chernozhukov (2011). Condition D1 is satisfied under Assumption (A4). Since β is just some constants, we only need to verify the sparsity condition, which holds under Assumption (A2). Condition D3 is satisfied because we consider $\hat{\sigma}_j^2$ needed in Belloni and Chernozhukov (2011) as a constant in (2.2) and (2.8) of our manuscript. Condition D4 holds under Assumption (A1). Thus, by Theorem 2 of Belloni and Chernozhukov (2011), we have

$$\|\tilde{\eta} - \eta_0\| \lesssim \sqrt{s}\lambda_1.$$

This completes the proof.⋄

Proof of Theorem 1.

Let $\mathbb{P}_n$ be the empirical measure of $(y_i, x_i, z_i), 1 \le i \le n$, $\omega = (y, x, z)$, $\mathcal{U}_\eta = \{\eta \in \mathbb{R}^q : \|\eta - \eta_0\| \le C(s\log(q)/n)^{1/2}, \|\eta\|_0 = O(s)\}$, and $\mathcal{U}_H = \{H \in \mathbb{R}^{d\times q} : \|(H - H_0)^{(j)}\|_1 \le Cs_h(\log(q)/n)^{1/2}, \|(H - H_0)^{(j)}\| \le C'(s_h\log(q)/n)^{1/2}, j = 1, \cdots, d\}$, where $M^{(j)}$ denotes the jth row of a matrix M. Write $\psi(\omega; \beta, \eta, H) = \psi_\tau\{y - x'\beta - z'\eta\}(x - Hz)$. By Theorem 2.10 of Kosorok (2008), it suffices to show that

$$n^{-1} \sup_{\beta\in\mathbb{R}^d, \eta\in\mathcal{U}_\eta, H\in\mathcal{U}_H} \left\| \sum_{i=1}^n \psi(\omega_i; \beta, \eta, H) - \sum_{i=1}^n \mathrm{P}\psi(\omega_i; \beta, \eta, H) \right\| \xrightarrow{p} 0. \quad \text{(S1.3)}$$

Let $\mathcal{F}_1 = \{\psi_\tau(y_i - x_i'\beta - z_i'\eta) : \beta \in \mathbb{R}^d, \eta \in \mathcal{U}_\eta\}$, $\mathcal{F}_2^{(j)} = \{(x - Hz)^{(j)} : H \in \mathcal{U}_H\}$, and $\mathcal{F} = \mathcal{F}_1 \cdot (\cup_{j=1}^d \mathcal{F}_2^{(j)})$, where $\xi^{(j)}$ denotes the jth component of a vector ξ. Since $\psi_\tau(\cdot)$ is monotone and bounded, the VC index of $\mathcal{F}_1$ is $O(d + s)$ by Lemma 2.6.15 of van der Vaart and Wellner (1996), which implies that the uniform entropy numbers is

$$\sup_Q \log N(\varepsilon\|F_1\|_{Q,2}, \mathcal{F}_1, \|\cdot\|_{Q,2}) \le C(d + s)\log(1/\varepsilon),$$

for all $0 < \varepsilon \le 1$, where $F_1 \equiv 1$ is the envelope of $\mathcal{F}_1$. Let $F_2 = \|x - Hz\|_\infty$ be the envelope function of $\mathcal{F}_2^{(j)}$. The VC index of $\mathcal{F}_2^{(j)}$ is $O(s_h)$, for all $j = 1, \cdots, d$, and then the uniform entropy numbers obey

$$\sup_Q \log N(\varepsilon\|F_2\|_{Q,2}, \mathcal{F}_2^{(j)}, \|\cdot\|_{Q,2}) \le C(d + s_h)\log(1/\varepsilon),$$

which results in that the uniform entropy numbers of $\mathcal{F}$ obeys

$$\begin{aligned}\sup_Q \log N(\varepsilon\|F_2\|_{Q,2}, \mathcal{F}, \|\cdot\|_{Q,2}) \leq& \sup_Q \log N\left(\frac{\varepsilon}{2}\|F_1\|_{Q,2}, \mathcal{F}_1, \|\cdot\|_{Q,2}\right) \\ &+ \sup_Q \log N\left(\frac{\varepsilon}{2}\|F_2\|_{Q,2}, \cup_{j=1}^d \mathcal{F}_2^{(j)}, \|\cdot\|_{Q,2}\right) \\ \leq& \sup_Q \log N\left(\frac{\varepsilon}{2}\|F_1\|_{Q,2}, \mathcal{F}_1, \|\cdot\|_{Q,2}\right) \\ &+ \log(d) \sup_Q \log N\left(\frac{\varepsilon}{2}\|F_2\|_{Q,2}, \mathcal{F}_2^{(j)}, \|\cdot\|_{Q,2}\right) \\ \leq& C(d + s \vee s_h)\log(1/\varepsilon).\end{aligned}$$

It is obvious by Assumption (A5) that

$$\|x - Hz\|_\infty \leq \|x - H_0 z\|_\infty + \max_{1\leq j\leq d} \|(H - H_0)^{(j)}\|_1 \|z\|_\infty \lesssim \|x - H_0 z\|_\infty + \|z\|_\infty s_h \lambda_2,$$

which implies that by Assumption (A1)

$$\|F_2\|_{\mathrm{P},2} \lesssim (E[\|x - H_0 z\|^2])^{1/2} + \zeta_n s_h \lambda_2 \lesssim \zeta_n s_h \lambda_2. \tag{S1.4}$$

And it follows Assumption (A1) that

$$\begin{aligned}\sup_{f\in\mathcal{F}} \|f\|_{\mathrm{P},2}^2 \leq& \sup_{H\in\mathcal{U}_H} E[\|x - Hz\|^2] \\ \leq& E[\|x - H_0 z\|^2] + \sup_{H\in\mathcal{U}_H} E[\|(H - H_0)z\|^2] \\ \leq& C + C' s_h^2 \log(q)/n \\ \lesssim& 1,\end{aligned} \tag{S1.5}$$

where C and C' are some constants.

Therefore, by Assumption (A5) and Theorem 5.2 of Chernozhukov et al. (2014) with $\sigma^2 = \|F_2\|^2_{\mathrm{P},2}$, we have

$$\begin{aligned}&\frac{1}{\sqrt{n}}E\left\{\sup_{\beta\in\mathbb{R}^d,\eta\in\mathcal{U}_\eta,H\in\mathcal{U}_H}\left\|\sum_{i=1}^n\psi(\omega_i;\beta,\eta,H)-\sum_{i=1}^n\mathrm{P}\psi(\omega_i;\beta,\eta,H)\right\|\right\}\\&\lesssim\zeta_n s_h\lambda_2(s\vee s_h)^{1/2}+n^{-1/2}(s\vee s_h)\\&=o(n^{1/2}).\end{aligned}$$

Thus, combining this with Theorem 5.1 of Chernozhukov et al. (2014), we have

$$n^{-1}\sup_{\beta\in\mathbb{R}^d,\eta\in\mathcal{U}_\eta,H\in\mathcal{U}_H}\left\|\sum_{i=1}^n\psi(\omega_i;\beta,\eta,H)-\sum_{i=1}^n\mathrm{P}\psi(\omega_i;\beta,\eta,H)\right\|=o_p(1).$$

This completes the proof. ⋄

Lemma 3. *If Assumptions (A1)–(A5) hold, then with probability approaching one,*

$$\|\hat\beta-\beta_0\|\lesssim\sqrt{s\vee s_h}(\lambda_1\vee\lambda_2).$$

Proof of Lemma 3. Note that

$$\begin{aligned}\mathrm{P}[\psi(w;\hat\beta,\tilde\eta,\tilde H)-\psi(w;\beta_0,\eta_0,H_0)]&=-\mathbb{P}_n\psi(w;\beta_0,\eta_0,H_0)\\&+\mathbb{P}_n\psi(w;\hat\beta,\tilde\eta,\tilde H)+(\mathbb{P}_n-\mathrm{P})\{\psi(w;\beta_0,\eta_0,H_0)-\psi(w;\hat\beta,\tilde\eta,\tilde H)\}.\quad\text{(S1.6)}\end{aligned}$$

Since $\mathbb{P}_n\psi(w;\beta_0,\eta_0,H_0)=O_p(n^{-1/2})$ and $\mathbb{P}_n\psi(w;\hat\beta,\tilde\eta,\tilde H)\approx 0$, and we have shown in Theorem 1 that

$$(\mathbb{P}_n-\mathrm{P})\{\psi(w;\beta_0,\eta_0,H_0)-\psi(w;\hat\beta,\tilde\eta,\tilde H)\}\lesssim n^{-1/2}(s\vee s_h)^{1/2}\zeta_n s_h\lambda_2,$$

we obtain that

$$\mathrm{P}[\psi(w;\hat{\beta},\tilde{\eta},\tilde{H})-\psi(w;\beta_0,\eta_0,H_0)]\lesssim n^{-1/2}(s\vee s_h)^{1/2}\zeta_n s_h\lambda_2.$$

It is clear by Lemma 2 and Assumption (A4) that

$$\begin{aligned}&\mathrm{E}_{x,z}\{\psi(w;\hat{\beta},\tilde{\eta},\tilde{H})-\psi(w;\beta_0,\eta_0,H_0)\}\\&=f(0|x,z)\{x'(\hat{\beta}-\beta_0)+z'(\tilde{\eta}-\eta_0)+(\tilde{H}-H_0)z\}(x-H_0z)\{1+o_p(1)\}\\&=f(0|x,z)x'(\hat{\beta}-\beta_0)(1+o_p(1))+O_p(\sqrt{s}\lambda_1)+O_p(\sqrt{s_h}\lambda_2),\end{aligned}$$

which implies the result of Lemma 3, where $\mathrm{E}_{x,z}$ denotes the conditional expectation of w given x and z. $\diamond$

Proof of Theorem 2.

Let

$$\psi(\omega;\beta,\eta,H)=\psi_\tau\{y-x'\beta-z'\eta\}(x-Hz).$$

With common notations in the empirical process literature, we can write

$$\tilde{\Psi}_n(\beta)=\mathbb{P}_n\psi(\cdot;\beta,\tilde{\eta},\tilde{H}),$$

and

$$\Psi(\beta,\eta,H)=\mathrm{P}\psi(\cdot;\beta,\eta,H).$$

We first need the claim that

$$(\mathbb{P}_n-\mathrm{P})\{\psi(\cdot;\hat{\beta},\tilde{\eta},\tilde{H})-\psi(\cdot;\beta_0,\eta_0,H_0)\}=o_p(n^{-1/2}).\tag{S1.7}$$

It is clear that

$$\|(\mathbb{P}_n - \mathrm{P})\{\psi(\cdot;\hat{\beta},\tilde{\eta},\tilde{H}) - \psi(\cdot;\beta_0,\eta_0,H_0)\}\| \leq \sup_{f\in\mathcal{F}_3} \|(\mathbb{P}_n - \mathrm{P})(f)\|,$$

where

$$\mathcal{F}_3 = \cup_{1\leq j\leq d}\, \mathcal{F}_3^{(j)}, \text{ and}$$
$$\mathcal{F}_3^{(j)} = \{\psi(\cdot;\beta,\eta,H)^{(j)} - \psi(\cdot;\beta_0,\eta_0,H_0)^{(j)} : \|\beta-\beta_0\| \leq C\tau_n, \eta\in\mathcal{U}_\eta, H\in\mathcal{U}_H\}$$

for sufficiently large constant C, and $\tau_n = \sqrt{s\vee s_h}(\lambda_1\vee\lambda_2)$. We need to prove that

$$\sup_{f\in\mathcal{F}_3} \mathrm{P}\|f\|^2 \lesssim \tau_n. \tag{S1.8}$$

Recall

$$\Psi(\beta,\eta,H) = \mathrm{P}\psi_\tau(y - x'\beta - z'\eta)(x - Hz) = \mathrm{E}\left\{\mathrm{E}[\psi_\tau\{y - x'\beta - z'\eta\}|x,z](x-Hz)\right\}.$$

Let $g(x,z;\beta,\eta) = \mathrm{E}[\psi_\tau(y - x'\beta - z'\eta)|x,z]$, $\Delta = x'(\beta-\beta_0) + z'(\eta-\eta_0)$. Since

$$g(x,z;\beta,\eta) = \mathrm{E}\{\psi_\tau(y - x'\beta_0 - z'\eta_0 - \Delta)|x,z\} = f(0|x,z)\Delta\{1+o_p(1)\},$$

and

$$\mathrm{E}\{I_{\{y-x'\beta-z'\eta\leq 0\}} I_{\{y-x'\beta_0-z'\eta_0\leq 0}|x,z\} = \mathrm{E}\{I_{\{\varepsilon\leq\Delta\}} I_{\{\varepsilon\leq 0\}}|x,z\} = \min\{F(\Delta),F(0)\},$$

it then follows from Lemma 2 and Theorem 1 that

$$\begin{aligned}
&\mathrm{P}\|\psi(\cdot;\beta,\eta,H_0)-\psi(\cdot;\beta_0,\eta_0,H_0)\|^2\\
=&\mathrm{P}\{\psi_\tau(\varepsilon-\Delta)-\psi_\tau(\varepsilon)\}^2\|x-H_0z\|^2\\
=&\{\mathrm{P}I_{\{\varepsilon\le\Delta\}}+\mathrm{P}I_{\{\varepsilon\le 0\}}-2\mathrm{P}I_{\{\varepsilon\le\Delta\}}I_{\{\varepsilon\le 0\}}\}\|x-H_0z\|^2\\
=&\mathrm{P}\left[F_\varepsilon(\Delta)+\tau-2\min\{F_\varepsilon(\Delta),\tau\}\right]\|x-H_0z\|^2\\
\le&\mathrm{P}\{f(0|x,z)(\|x'(\beta-\beta)\|+\|z'(\eta-\eta_0)\|)\}\times\|x-H_0z\|^2(1+o_p(1))\\
\lesssim&\tau_n.
\end{aligned}\tag{S1.9}$$

It then follows from Lemma 2 that

$$\mathrm{P}\|\psi(\cdot;\beta,\eta,H)-\psi(\cdot;\beta,\eta,H_0)\|^2\le 2\mathrm{P}\|(H-H_0)z\|^2\lesssim\tau_n^2. \tag{S1.10}$$

Therefore, by (S1.9) and (S1.10), and the triangle inequality, (S1.8) holds. We define the envelope function of $\mathcal{F}_3$ as

$$F_3=\sup_{1\le j\le d}\ \sup_{\|\beta-\beta_0\|\le C\tau_n,\eta\in\mathcal{U}_\eta,H\in\mathcal{U}_H}2|\psi(\cdot;\beta,\eta,H)^{(j)}|.$$

By Assumption (A5) and Lemma 2, we have, for all $H\in\mathcal{U}_H$,

$$|(x-Hz)^{(j)}|\le|(x-H_0z)^{(j)}|+\|(H_0-H)^{(j)}\|_1\|z\|_\infty\le|(x-H_0z)^{(j)}|+s_h\lambda_2\|\|z\|_\infty,$$

which implies that

$$\|F_3\|_{\mathrm{P},2}\lesssim\zeta_n s_h\lambda_2.$$

Since $\mathcal{F}_3\subset\mathcal{F}-\mathcal{F}$ with $\mathcal{F}$ defined in the proof of Theorem 1, we then obtain that

$$\sup_Q\log N(\varepsilon\|F_3\|_{Q,2},\mathcal{F}_3,\|\cdot\|_{Q,2})\lesssim(s\vee s_h)^{1/2}\log(1/\varepsilon). \tag{S1.11}$$

Combining the above inequality with Assumption (A5) and Theorems 5.1 and 5.2 of Chernozhukov et al. (2014) by applying $\sigma^2 = \tau_n$, we have that

$$\begin{aligned}\sup_{f\in\mathcal{F}_3}\|(\mathbb{P}_n - \mathrm{P})(f)\| \lesssim & \sqrt{\tau_n(s\vee s_h)\log(1/\delta_n)/n} + n^{-1}(s\vee s_h)^{1/2}\log(1/\delta_n) \\ =& o_p(n^{-1/2}),\end{aligned} \tag{S1.12}$$

which implies that (S1.7) holds, where $\delta_n = \sqrt{\tau_n}/(\zeta_n s_h \lambda_2)$.

With (S1.7), we rewrite the equation $\tilde{\Psi}_n(\hat{\beta}) = \mathbb{P}_n\psi(\cdot;\hat{\beta},\tilde{\eta},\tilde{H}) \approx 0$ as follows.

$$\begin{aligned}0 =& (\mathbb{P}_n - \mathrm{P})\{\psi(\cdot;\hat{\beta},\tilde{\eta},\tilde{H}) - \psi(\cdot;\beta_0,\eta_0,H_0)\} + \mathbb{P}_n\psi(\cdot;\beta_0,\eta_0,H_0) \\ &+ \mathrm{P}\{\psi(\cdot;\hat{\beta},\tilde{\eta},\tilde{H}) - \psi(\cdot;\beta_0,\eta_0,H_0)\} \\ =& o_p(n^{-1/2}) + \mathbb{P}_n\psi(\cdot;\beta_0,\eta_0,H_0) + \mathrm{P}\{\psi(\cdot;\hat{\beta},\tilde{\eta},\tilde{H}) - \psi(\cdot;\beta_0,\eta_0,H_0)\}.\end{aligned}$$

Then we have

$$\Psi(\hat{\beta},\tilde{\eta},\tilde{H}) - \Psi(\beta_0,\eta_0,\tilde{H}) = -\mathbb{P}_n\psi(\cdot;\beta_0,\eta_0,H_0) + o_p(n^{-1/2}).$$

Let $V_n(\Delta) = n^{-1/2}\sum_{i=1}^n (F(\Delta_i) - \tau)(x_i - Hz_i) - \Psi(\beta,\eta,H)$. Since $\mathrm{P}V_n = 0$ and

$$\begin{aligned}\mathrm{Var}(V_n(\Delta)) =& \mathrm{Var}\{(F(\Delta_i) - \tau)(x_i - Hz_i) - \Psi(\beta,\eta,H)\} \\ \le& \mathrm{E}[f^2(\bar{w}|x_i,z_i)(x_i - Hz_i)(x_i - Hz_i)^T\Delta_i^2] \\ =& O(\|\beta - \beta_0\| + \|\eta - \eta_0\|),\end{aligned}$$

where $\bar{w}_i$ is between Δ_i and 0, we have by Assumption (A5), $V_n(\Delta) = o(1)$, for any $\|\beta - \beta_0\| + \|\eta - \eta_0\| = o_p(1)$ and $\|H - H_0\| = o_p(1)$, and furthermore,

$$\Psi(\hat{\beta},\tilde{\eta},\tilde{H}) = \frac{1}{n}\sum_{i=1}^n\{F(\hat{\Delta}_i) - \tau\}(x_i - \tilde{H}z_i) + o_p(n^{-1/2}).$$

According to Taylor's expansion, we have

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n}\{F(\hat{\Delta}_i)-\tau\}(x_i-\tilde{H}z_i) =&\frac{1}{n}\sum_{i=1}^{n} f(\hat{w}_i|x_i,z_i)(x_i-\tilde{H}z_i)x_i'(\hat{\beta}-\beta_0)\\ &+\frac{1}{n}\sum_{i=1}^{n} f(\hat{w}_i|x_i,z_i)(x_i-\tilde{H}z_i)z_i'(\tilde{\eta}-\eta_0),\end{aligned}$$

where $\hat{w}_i$ is between $\hat{\Delta}_i$ and 0. The second term on the right is $o_p(1)$. In fact, by Assumption (A5) and Theorem 2 of Belloni and Chernozhukov (2011), we have

$$\begin{aligned}\left\|\frac{1}{n}\sum_{i=1}^{n} f(\hat{w}_i|x_i,z_i)(x_i-\tilde{H}z_i)z_i'(\tilde{\eta}-\eta_0)\right\| \le&\|\tilde{\eta}-\eta_0\| \max_{1\le j\le q}\left\|\frac{1}{n}\sum_{i=1}^{n} f(\hat{w}_i|x_i,z_i)(x_i-\tilde{H}z_i)z_{ij}\right\|\\ =&o_p(s\lambda_1 s^{-1}\{\log(q)\}^{-1/2})\\ =&o_p(n^{-1/2}).\end{aligned}$$

This implies that

$$\frac{1}{n}\sum_{i=1}^{n} f(\hat{w}_i|x_i,z_i)(x_i-\tilde{H}z_i)x_i'(\hat{\beta}-\beta_0) = \Psi(\beta_0,\eta_0,\tilde{H})+o_p(n^{-1/2}),$$

and by the law of large numbers and the continuity of $f(u|x,z)$,

$$\mathrm{E}\{f(0|x_i,z_i)(x_i-\tilde{H}z_i)x_i'\}(\hat{\beta}-\beta_0) = \Psi(\beta_0,\eta_0,\tilde{H})+o_p(n^{-1/2}).$$

Also, under Assumptions (A1), (A2), (A4), and (A6), and Lemma 2, we have

$$\begin{aligned}\|\mathrm{E}\{f(0|x,z)(x-\tilde{H}z)x'\}-\mathrm{E}\{f(0|x,z)(x-H_0z)x'\}\| \le& C\mathrm{E}\{\|(\tilde{H}-H_0)zx'\|\}\\ \le& C\mathrm{E}\{\|(\tilde{H}-H_0)z\|\}\mathrm{E}\{\|x\|\}\\ =&o(1),\end{aligned}$$

so it follows that

$$\hat{\beta}-\beta_0 = -n^{-1}\left[\mathrm{E}\left\{f(0|x,z)(x-H_0z)x'\right\}\right]^{-1}\sum_{i=1}^{n}\psi_\tau(\varepsilon_i)(x_i-H_0z_i)+o_p(n^{-1/2}), \quad \text{(S1.13)}$$

where $\varepsilon_i = y_i - \beta_0' x_i - \eta_0' z_i$. Therefore, the result holds under Assumptions (A4) and (A6). $\diamond$

S2 Supplementary Material B: Refitted wild bootstrap

Let $n_1 = [n/2]$ and $n_2 = n - n_1$, where $[u]$ denotes the integer not greater than the positive number u. We randomly split the original dataset into two even parts V_1 and V_2. Without loss of generality, we assume that $V_1 = \{(y_i, x_i, z_i) : n_1 + 1 \le i \le n\}$ and $V_2 = \{(y_i, x_i, z_i) : 1 \le i \le n_1\}$. Let $\tilde{\beta}_1$ and $\tilde{\eta}_1$ be the estimator from

$$(\tilde{\beta}_1, \tilde{\eta}_1) = \underset{\beta,\eta}{\operatorname{argmin}} \frac{1}{n_2} \sum_{i=n_1+1}^{n} \rho_\tau(y_i - x_i'\beta - z_i'\eta) + \lambda_1 \|\eta\|_1, \tag{S2.1}$$

which is similar to (2.2) in the main paper where the original sample is replaced by its first part of the dataset V_1.

Let $\hat{S}_1 = \{j : \tilde{\eta}_{1j} \neq 0, 0 \le j \le q, \tilde{\eta}_1 = (\tilde{\eta}_{11}, \cdots, \tilde{\eta}_{1q})^T\}$, $\hat{s}_1 = |\hat{S}_1|$, and $\hat{T}_1 = \{v \in \mathbb{R}^q : v_j = 0, \forall j \in \hat{S}_1^c\}$. Let $\hat{\beta}_2$ and $\tilde{\eta}_2$ be the estimator from

$$(\hat{\beta}_2, \tilde{\eta}_2) = \underset{\beta \in \mathbb{R}^d, \eta \in \hat{T}_1}{\operatorname{argmin}} \frac{1}{n_1} \sum_{i=1}^{n_1} \rho_\tau(y_i - x_i'\beta - z_i'\eta), \tag{S2.2}$$

which is the regular quantile regression estimation based on the second part of the dataset V_2.

Let $\tilde{\beta}_2^*$ and $\tilde{\eta}_2^*$ be the estimates which satisfy

$$(\tilde{\beta}_2^*, \tilde{\eta}_2^*) = \underset{\beta,\eta}{\operatorname{argmin}} \frac{1}{n_1} \sum_{i=1}^{n_1} \rho_\tau(y_i^* - x_i'\beta - z_i'\eta) + \lambda_1 \|\eta\|_1, \tag{S2.3}$$

which is similar to (S2.2) where y_i is replaced by its bootstapped sample provided in (B4). Then the regularized projection score based on the boostrapped sample for β is given by

$$\tilde{\Psi}_n^*(\beta) \equiv \Psi_n^*(\beta, \tilde{\eta}_2^*, \tilde{H}_2) = \frac{1}{n_1} \sum_{i=1}^{n_1} \psi_\tau(y_i^* - x_i'\beta - z_i'\tilde{\eta}_2^*)(x_i - \tilde{H}_2 z_i), \tag{S2.4}$$

which is similar to (2.10) of the main paper where the original sample and $\tilde{\eta}$ are replaced by its resample and $\tilde{\eta}_2^*$ from (S2.3), respectively. Note that $\tilde{H}_2$ in (S2.4) is estimated from (2.8) in the main paper by only using data V_2. The estimator $\hat{\beta}^*$ based on bootstrapped dataset is the solution to the equation

$$\tilde{\Psi}_n^*(\beta) = 0. \tag{S2.5}$$

Lemma 4. *If Assumptions (A1)-(A5) hold, then with probability approaching one, we have*

$$\|\tilde{\eta}_2 - \eta_0\| \lesssim \sqrt{s\log(p)/n},$$
$$\|\hat{\beta}_2 - \beta_0\| \lesssim n_1^{-1/2}.$$

Proof of Lemma 4. According to Theorems 2 and 3 of Belloni and Chernozhukov (2011), we have, with probability approaching one, $\hat{s} \lesssim s$, and then

$$\|\tilde{\eta}_2 - \tilde{\eta}_1| \lesssim \sqrt{s\log(p)/n_1}.$$

Along the lines of Lemma 2, we have

$$\|\tilde{\eta}_1 - \eta_0\| = \lesssim \sqrt{s\log(p)/n_2}.$$

By triangle inequality, it holds that

$$\|\tilde{\eta}_2 - \eta_0\| \leq \|\tilde{\eta}_2 - \tilde{\eta}_1\| + \|\tilde{\eta}_1 - \eta_0\| \lesssim \sqrt{\hat{s}\log(p)/n_1} + \sqrt{s\log(p)/n_2},$$

which implies that $\|\tilde{\eta}_2 - \eta_0\| \lesssim \sqrt{s\log(p)/n}$ by noting that $\hat{s} \lesssim s$. This leads to the first equation.

Now we prove the second equation. With the similar argument of the proof of Lemma 2, we have

$$\max_{1\le j\le d}\|(\tilde{H}_2 - H_0)^{(j)}\| \lesssim \sqrt{s_h \log(q)/n_1},$$

and consequently, $\hat{\beta}_2 - \beta_0$ is normally distributed with the similar argument of the proof of Theorem 2. This gives the second equation. $\diamond$

Lemma 5. *If Assumptions (A1)-(A5) hold, then with probability approaching one, we have*

$$\|\tilde{\eta}^* - \tilde{\eta}_2\| \lesssim \sqrt{s \log(q)/n_1}.$$

Proof of Lemma 5. Note that P^* takes expectation on bootstrapped sample $\{(y_i^*, x_i, z_i) : 1 \le i \le n_1\}$. It can be shown by Theorem 3 of Belloni and Chernozhukov (2011) that with probability approaching one we have $\hat{s} \lesssim s$. The results follow with the similar argument of the proof of Lemma 2. $\diamond$

Let $\mathbb{P}_n^*$ be the empirical measure of $(y_i^*, x_i, z_i), 1 \le i \le n$, and $\psi(w^*; \beta, \eta, H) = \psi_\tau\{y^* - x'\beta - z'\eta\}(x - Hz)$, where $w^* = (y^*, x, z)$.

Lemma 6. *If $S \subseteq \hat{S}_1$ and $\hat{s}_1 = O(s)$, we have*

$$\hat{\beta}^* - \hat{\beta}_2 = -\left[E\left\{f(0|x,z)(x - H_0 z)x'\right\}\right]^{-1} \frac{1}{n_1}\sum_{i=1}^{n_1} \psi_\tau(\zeta_i|\hat{r}|)(x_i - H_0 z_i) + o_{p^*}(n_1^{-1/2}).$$

Proof of Lemma 6. Consider the following identity

$$\mathrm{P}^*\{\psi(w^*; \hat{\beta}^*, \tilde{\eta}_2^*, \tilde{H}_2) - \psi(w^*; \hat{\beta}_2, \tilde{\eta}_2, H_0)\} = -\mathbb{P}_n^*\{\psi(w^*; \hat{\beta}_2, \tilde{\eta}_2, H_0)\}$$

$$+ \mathbb{P}_n^*\{\psi(w^*; \hat{\beta}^*, \tilde{\eta}_2^*, \tilde{H}_2)\} + (\mathbb{P}_n^* - \mathrm{P}^*)\{\psi(w^*; \hat{\beta}_2, \tilde{\eta}_2, H_0) - \psi(w^*; \hat{\beta}^*, \tilde{\eta}_2^*, \tilde{H}_2)\}. \quad \text{(S2.6)}$$

To show that $\sqrt{n_1}(\mathbb{P}_n^* - \mathrm{P}^*)\{\psi(w^*; \hat{\beta}_2, \tilde{\eta}_2, H_0) - \psi(w^*; \hat{\beta}^*, \tilde{\eta}_2^*, \tilde{H}_2)\} = o_{p^*}(1)$, it suffices by Theorems 5.1 and 5.2 of Chernozhukov et al. (2014) to show that

$$\mathrm{P}^* \|\psi(w^*; \hat{\beta}_2, \tilde{\eta}_2, H_0) - \psi(w^*; \hat{\beta}^*, \tilde{\eta}_2^*, \tilde{H}_2)\|^2 \lesssim \tau_n, \tag{S2.7}$$

and the uniform entropy numbers of $\mathcal{F}_3$ under the bootstrapped sample are not greater than $C \log\{(s \vee s_h)/\varepsilon\}$.

Let $\varepsilon_i^* = \zeta_i |\hat{r}_i|$, for $i = 1, \cdots, n$, where the random weights ζ_i are independently drawn from the distribution G. Let $g^*(x, z; \beta, \eta) = \mathrm{E}\{\psi_\tau(y^* - x'\beta - z'\eta)|\hat{r}, x, z\}$, $\Delta^* = x'(\beta - \hat{\beta}_2) + z'(\eta - \tilde{\eta}_2)$, and $\hat{\Delta}^* = x'(\hat{\beta}^* - \hat{\beta}_2) + z'(\tilde{\eta}_2^* - \tilde{\eta}_2)$.

Since

$$g^*(x, z; \hat{\beta}^*, \tilde{\eta}_2^*) = \mathrm{E}^*\{\psi_\tau(y^* - x'\hat{\beta}_2 - z'\tilde{\eta}_2 - \hat{\Delta}^*)|\hat{r}, x, z\} = G'(0|\hat{r}, x, z)\frac{\hat{\Delta}^*}{|\hat{r}|}\{1 + o_{p^*}(1)\},$$

and

$$\begin{aligned}\mathrm{E}\{I_{\{y^* - x'\hat{\beta}^* - z'\tilde{\eta}^* \le 0\}} I_{\{y^* - x'\hat{\beta}_2 - z'\tilde{\eta}_2 \le 0\}}|x, z\} =& \mathrm{E}\{I_{\{\varepsilon^* \le \hat{\Delta}^*\}} I_{\{\varepsilon^* \le 0\}}|x, z\} \\ =& \min\{G(\hat{\Delta}^*/|\hat{r}|), G(0)\},\end{aligned}$$

we have, by Lemmas 4 and 5,

$$\begin{aligned}
&\mathrm{P}^*\|\psi(w^*;\hat\beta^*,\tilde\eta_2^*,H_0)-\psi(w^*;\hat\beta_2,\tilde\eta_2,H_0)\|^2\\
&=\mathrm{P}^*\{\psi_\tau(\varepsilon^*-\hat\Delta^*)-\psi_\tau(\varepsilon^*)\}^2\|x-H_0z\|^2\\
&=\{\mathrm{P}^*I_{\{\varepsilon^*\le\hat\Delta^*\}}+\mathrm{P}^*I_{\{\varepsilon^*\le 0\}}-2\mathrm{P}^*I_{\{\varepsilon^*\le\hat\Delta^*\}}I_{\{\varepsilon^*\le 0\}}\}\|x-H_0z\|^2\\
&=\mathrm{P}^*\left[G(\hat\Delta^*/|\hat r|)+\tau-2\min\{G(\hat\Delta^*/|\hat r|),\tau\}\right]\|x-H_0z\|^2\\
&\le\mathrm{P}^*\{f(0|x,z)(\|x'(\hat\beta^*-\hat\beta_2)\|+\|z'(\tilde\eta_2^*-\tilde\eta_2)\|)\}\times\|x-H_0z\|^2(1+o_p(1))\\
&\lesssim\tau_n.
\end{aligned}\tag{S2.8}$$

Lemma 2 indicates that

$$\mathrm{P}^*\|\psi(w^*;\hat\beta^*,\tilde\eta_2^*,\tilde H_2)-\psi(w^*;\hat\beta^*,\tilde\eta_2^*,H_0)\|^2\le 2\mathrm{P}\|(\tilde H_2-H_0)z\|^2\lesssim\tau_n^2.\tag{S2.9}$$

By the triangle inequality together with (S2.8) and (S2.9), the inequality (S2.7) holds when conditions in Lemma 2 are satisfied. With the same lines along (S1.11), we can show that the uniform entropy numbers of $\mathcal{F}_3$ under the bootstrapped sample are not greater than $C\log\{(s\vee s_h)/\varepsilon\}$, and $\sqrt{n_1}(\mathbb{P}_n^*-\mathrm{P}^*)\{\psi(w^*;\hat\beta_2,\tilde\eta_2,H_0)-\psi(w^*;\hat\beta^*,\tilde\eta_2^*,\tilde H_2)\}=o_{p^*}(1)$.

With the approximated equation $\mathbb{P}_n\psi(\cdot;\hat\beta^*,\tilde\eta_2^*,\tilde H_2)\approx 0$, we have

$$\begin{aligned}
&\sqrt{n_1}\mathrm{P}^*\{\psi(w^*;\hat\beta^*,\tilde\eta_2^*,\tilde H_2)-\psi(w^*;\hat\beta_2,\tilde\eta_2,H_0)\}\\
&\qquad\qquad=-\sqrt{n_1}\mathbb{P}_n^*\{\psi(w^*;\hat\beta_2,\tilde\eta_2,H_0)\}+o_{p^*}(1).
\end{aligned}\tag{S2.10}$$

It is clear that

$$\begin{aligned}&\mathrm{P}^*\{\psi(w^*;\hat\beta^*,\tilde\eta_2^*,H)-\psi(w^*;\hat\beta_2,\tilde\eta_2,H)\}\\&=\mathrm{P}^*[\psi_\tau\{\zeta|\hat r|-x'(\hat\beta^*-\hat\beta_2)-z'(\tilde\eta_2^*-\tilde\eta_2)\}-\psi_\tau(\zeta|\hat r|)](x-Hz).\end{aligned}\qquad\text{(S2.11)}$$

For any vector $u_1,u_2\in\mathbb{R}^d$ and $v_1,v_2\in\mathbb{R}^q$, we have, by Assumptions (A4) and (B3),

$$\begin{aligned}&\mathrm{P}^*\left\{\psi_\tau(\zeta|\varepsilon-x'u_1-z'v_1|-x'u_2-z'v_2)-\psi_\tau(\zeta|\varepsilon-x'u_1-z'v_1|)\right\}\\
&=\int\int I_{\{0\le\zeta|\varepsilon-x'u_1-z'v_1|<x'u_2+z'v_2\}}dF(\varepsilon)dG(\zeta)I_{\{x'u_2+z'v_2\ge0\}}\\
&\quad-\int\int I_{\{x'u_2+z'v_2\le\zeta|\varepsilon-x'u_1-z'v_1|<0\}}dF(\varepsilon)dG(\zeta)I_{\{x'u_2+z'v_2<0\}}\\
&=\int_0^{+\infty}[F\{x'u_1+z'v_1+\zeta^{-1}(x'u_2+z'v_2)\}-F\{x'u_1+z'v_1-\zeta^{-1}(x'u_2+z'v_2)\}]\\
&\quad\times dG(\zeta)I_{\{x'u_2+z'v_2>0\}}\\
&\quad-\int_{-\infty}^0[F\{x'u_1+z'v_1+\zeta^{-1}(x'u_2+z'v_2)\}-F\{x'u_1+z'v_1-\zeta^{-1}(x'u_2+z'v_2)\}]\\
&\quad\times dG(\zeta)I_{\{x'u_2+z'v_2\le0\}}\\
&=2\int_0^{+\infty}\zeta^{-1}dG(\zeta)f(0|x,z)(x'u_2+z'v_2)I_{\{x'u_2+z'v_2>0\}}\\
&\quad-2\int_{-\infty}^0\zeta^{-1}dG(\zeta)f(0|x,z)(x'u_2+z'v_2)I_{\{x'u_2+z'v_2\le0\}}\\
&\quad+O(|x'u_2+z'v_2|(|x'u_1+z'v_1|+|x'u_2+z'v_2|)^{1/2})\\
&=f(0|x,z)(x'u_2+z'v_2)+O(|x'u_2+z'v_2|(|x'u_1+z'v_1|+|x'u_2+z'v_2|)^{1/2}),\end{aligned}$$

where the last equation holds by Assumption (A4). Note that $\mathrm{P}^*\{\psi(w^*;\hat\beta_2,\tilde\eta_2,\tilde H_2)-$

$\psi(w^*;\hat\beta_2,\tilde\eta_2,H_0)\} = 0$. Thus, it follows from (S2.11) that for the left side of (S2.10)

$$
\begin{aligned}
&\mathrm{P}^*\{\psi(w^*;\hat\beta^*,\tilde\eta_2^*,\tilde H_2)-\psi(w^*;\hat\beta_2,\tilde\eta_2,H_0)\}\\
=&\mathrm{P}^*\{\psi(w^*;\hat\beta^*,\tilde\eta_2^*,\tilde H_2)-\psi(w^*;\hat\beta_2,\tilde\eta_2,\tilde H_2)\}+\mathrm{P}^*\{\psi(w^*;\hat\beta_2,\tilde\eta_2,\tilde H_2)-\psi(w^*;\hat\beta_2,\tilde\eta_2,H_0)\}\\
=&\mathrm{P}^*\{\psi(w^*;\hat\beta^*,\tilde\eta_2^*,\tilde H_2)-\psi(w^*;\hat\beta_2,\tilde\eta_2,\tilde H_2)\}\\
=&\mathrm{E}\{f(0|x,z)x'(\hat\beta^*-\hat\beta_2)(x-\tilde H_2 z)\}+\mathrm{E}\{f(0|x,z)z'(\tilde\eta_2^*-\tilde\eta_2)(x-\tilde H_2 z)\}\\
&\quad+O((|x'(\hat\beta_2-\beta_0)+z'(\tilde\eta_2-\eta_0)|+|x'(\hat\beta^*-\hat\beta_2)+z'(\tilde\eta_2^*-\tilde\eta_2)|)^{1/2})\\
&\quad\times O(|x'(\hat\beta^*-\hat\beta_2)+z'(\tilde\eta_2^*-\tilde\eta_2)|\|x-\tilde H_2 z\|)\\
=&\mathrm{E}\{f(0|x,z)x'(\hat\beta^*-\hat\beta_2)(x-\tilde H_2 z)\}+\mathrm{E}\{f(0|x,z)(\tilde\eta_2^*-\tilde\eta_2)'z(x-\tilde H_2 z)\}+o_{p^*}(n_1^{-1/2}),
\end{aligned}
\tag{S2.12}
$$

where the last equation holds by Lemmas 2, 4 and 5.

Let $\mathcal{U}_v = \{v\in\mathbb{R}^q : \|v\| = O_{p^*}(s\lambda_1)\}$. For any $v\in\mathcal{U}_v$ and $H\in\mathcal{U}_H$, we have, by Assumption (A5),

$$
\begin{aligned}
\mathrm{E}\{f(0|x,z)(x-Hz)z'v\} =&\frac{1}{n}\sum_{i=1}^n f(0|x_i,z_i)(x_i-Hz_i)z_i'v+O_{p^*}(n^{-1}\|v\|)\\
=&o_{p^*}(s\lambda_1 s^{-1}\{\log(q)\}^{-1/2})+O_{p^*}(s\log(q)n^{-2}).
\end{aligned}
$$

Combining this and (S2.12), we have

$$
\begin{aligned}
&\sqrt{n_1}\mathrm{P}^*\{\psi(w^*;\hat\beta^*,\tilde\eta_2^*,\tilde H_2)-\psi(w^*;\hat\beta_2,\tilde\eta_2,H_0)\}\\
&\quad=\sqrt{n_1}E\{f(0|x,z)x'(\hat\beta^*-\hat\beta_2)(x-\tilde H_2 z)\}+o_{p^*}(1).
\end{aligned}
$$

Inserting above equation to (S2.6), we further have

$$\begin{aligned}\sqrt{n_1}\mathrm{E}\{f(0|x,z)x'(\hat{\beta}^* - \hat{\beta}_2)(x - \tilde{H}_2 z)\} =& - \mathbb{P}_n^*\{\psi(w^*; \hat{\beta}_2, \tilde{\eta}_2, H_0)\} + o_{p^*}(1) \\ =& \frac{1}{n_1}\sum_{i=1}^{n_1} \psi_\tau(\zeta_i|\hat{r}|)(x_i - H_0 z_i) + o_{p^*}(1).\end{aligned} \tag{S2.13}$$

Since $\|\tilde{H}_2 - H_0\| = o_p(1)$ by (S1.8), it then follows from (S2.13) and Assumption (A6) that

$$\hat{\beta}^* - \hat{\beta}_2 = -\left[\mathrm{E}\left\{f(0|x,z)(x - H_0 z)x'\right\}\right]^{-1} \frac{1}{n_1}\sum_{i=1}^{n_1} \psi_\tau(\zeta_i|\hat{r}_i|)(x_i - H_0 z_i) + o_{p^*}(n^{-1/2}).$$

This completes the proof of Lemma 6. $\diamond$

Proof of Theorem 3.

Lemma 6 implies that

$$\sqrt{n_1}(\hat{\beta}^* - \hat{\beta}_2) \xrightarrow{L} N(0, Q^{-1}DQ^{'-1}),$$

where $Q = \mathrm{E}\left\{f(0|x,z)(x - H_0 z)x'\right\}$, and $D = \tau(1-\tau)\mathrm{E}\left\{(x - H_0 z)(x - H_0 z)'\right\}$. This completes the proof of Theorem 3 by Theorem 2 and the asymptotic result above. $\diamond$

S3 Supplementary Material C: One-step estimator

Lemma 7. *If Assumptions (A1), (A2) and (A4) hold , we have, with probability approaching one,*

$$\|\tilde{\beta} - \beta_0\| \lesssim \sqrt{s}\lambda_1.$$

Proof of Lemma 7. By (S1.3) and Theorem 2.10 of Kosorok (2008), we have

$$\tilde{\beta} \xrightarrow{p} \beta_0. \tag{S3.1}$$

Since $\mathrm{P}\psi_\tau(y - x'\beta_0 - z'\eta_0)x = 0$, by Lemma 2 and Assumption (A4), we have with probability approaching one,

$$\begin{aligned}
&\mathrm{E}_{x,z}\{\psi_\tau(y - x'\tilde{\beta} - z'\tilde{\eta})x - \psi_\tau(y - x'\beta_0 - z'\eta_0)x\} \\
&=f(0|x,z)x'(\tilde{\beta} - \beta_0)\{(1 + o(1)\} + f(0|x,z)z'(\tilde{\eta} - \eta_0)\{1 + o(1)\} \\
&=f(0|x,z)x'(\tilde{\beta} - \beta_0)\{1 + o(1)\} + O(\sqrt{s}\lambda_1).
\end{aligned} \tag{S3.2}$$

Since

$$\begin{aligned}
&\mathrm{P}\{\psi_\tau(y - x'\tilde{\beta} - z'\tilde{\eta})x - \psi_\tau(y - x'\beta_0 - z'\eta_0)x\} = -\mathbb{P}_n\psi_\tau(y - x'\beta_0 - z'\eta_0)x \\
&+ \mathbb{P}_n\psi_\tau(y - x'\tilde{\beta} - z'\tilde{\eta})x + (\mathbb{P}_n - \mathrm{P})\{\psi_\tau(y - x'\beta_0 - z'\eta_0)x - \psi_\tau(y - x'\tilde{\beta} - z'\tilde{\eta})x\},
\end{aligned} \tag{S3.3}$$

and $\mathbb{P}_n\psi_\tau(y - x'\beta_0 - z'\eta_0) = O_p(n^{-1/2})$ and $\mathbb{P}_n\psi_\tau(y - x'\tilde{\beta} - z'\tilde{\eta}) \approx 0$, it suffices to show that

$$(\mathbb{P}_n - \mathrm{P})\{\psi_\tau(y - x'\beta_0 - z'\eta_0)x - \psi_\tau(y - x'\tilde{\beta} - z'\tilde{\eta})x\} \lesssim \sqrt{s}\lambda_1. \tag{S3.4}$$

Let $\mathcal{F}_4 = \cup_{1\le j\le d}\mathcal{F}_4^{(j)}$ with $\mathcal{F}_4^{(j)} = \{(\psi_\tau(y_i - x_i'\beta - z_i'\eta) - \psi_\tau(y_i - x_i'\beta_0 - z_i'\eta_0))x_i^{(j)} : \beta \in \mathbb{R}^d, \eta \in \mathcal{U}_\eta\}$, where $x_i^{(j)}$ is the jth component of x_i. As the same lines along the proof of Theorem 1, we have

$$\sup_Q \log N(\varepsilon\|F_4\|_{Q,2}, \mathcal{F}_4, \|\cdot\|_{Q,2}) \le C(d+s)\log(1/\varepsilon),$$

where $F_4 = 2\|x\|_\infty$ is the envelope of $\mathcal{F}_4$ with $\|F_4\|_{\mathrm{P},2} \lesssim 1$ by Assumption (A1), and

$$\sup_{f\in\mathcal{F}_4} \|f\|_{\mathrm{P},2}^2 \lesssim 1.$$

Therefore, as the same lines along the proof of Theorem 1 and 2, by Theorem 5.1 and 5.2 of Chernozhukov et al. (2014) with $\sigma^2 = \|F_4\|_{\mathrm{P},2}^2$, we have

$$\sup_{f\in\mathcal{F}_4} \|(\mathbb{P}_n - \mathrm{P})(f)\| = o_p(n^{-1/2}),$$

which implies that (S3.4) holds. We complete the proof of Lemma 7 by combining (S3.2) with (S3.3) and (S3.4). $\diamond$

Lemma 8. *If Assumptions (A1)–(A4) hold , we have*

$$\hat{\beta}_{one} \xrightarrow{p} \beta_0.$$

Proof of Lemma 8. This Lemma can be shown with the similar argument of the proof of Theorem 1.$\diamond$

Lemma 9. *If conditions of Theorem 4 hold , we have*

$$\hat{\beta}_{one} - \hat{\beta}_0 = -\left[E\left\{f(0|x,z)(x - H_0 z)(x - H_0 z)'\right\}\right]^{-1} \frac{1}{n}\sum_{i=1}^n \psi_\tau(\varepsilon_i)(x_i - H_0 z_i) + o_p(n^{-1/2}).$$

Proof of Lemma 9. Let $\tilde{\psi}(w;\beta,\tilde{\beta},\eta,H)=\psi_\tau\{y-(x-Hz)'\beta-(Hz)'\tilde{\beta}-z'\eta\}(x-Hz)$, where $w=(y,x',z')'$. It is obvious that $\tilde{\psi}(w;\beta,\beta,\eta,H)=\psi(w;\beta,\eta,H)$, where $\psi(w;\beta,\eta,H)=\psi_\tau\{y-x'\beta-z'\eta\}(x-Hz)$ is defined in the proof of Theorem 2. As similarly considered in the proof of Lemma 6, we use the following identity

$$\begin{aligned}&\mathrm{P}\{\tilde{\psi}(w;\hat{\beta}_{one},\tilde{\beta},\tilde{\eta},\tilde{H})-\psi(w;\beta_0,\eta_0,H_0)\}=-\mathbb{P}_n\psi(w;\beta_0,\eta_0,H_0)\\&\quad+\mathbb{P}_n\tilde{\psi}(w;\hat{\beta}_{one},\tilde{\beta},\tilde{\eta},\tilde{H})+(\mathbb{P}_n-\mathrm{P})\{\psi(w;\beta_0,\eta_0,H_0)-\tilde{\psi}(w;\hat{\beta}_{one},\tilde{\beta},\tilde{\eta},\tilde{H})\}.\end{aligned}\tag{S3.5}$$

We first show $\sqrt{n}(\mathbb{P}_n-\mathrm{P})\{\psi(w;\beta_0,\eta_0,H_0)-\tilde{\psi}(w;\hat{\beta}_{one},\tilde{\beta},\tilde{\eta},\tilde{H})\}=o_p(1)$, which results in that

$$\begin{aligned}&\mathrm{P}\{\tilde{\psi}(w;\hat{\beta}_{one},\tilde{\beta},\tilde{\eta},\tilde{H})-\psi(w;\beta_0,\eta_0,H_0)\}=-\mathbb{P}_n\psi(w;\beta_0,\eta_0,H_0)\\&\qquad\qquad+\mathbb{P}_n\tilde{\psi}(w;\hat{\beta}_{one},\tilde{\beta},\tilde{\eta},\tilde{H})+o_p(n^{-1/2}).\end{aligned}\tag{S3.6}$$

As the same lines along the proof of Theorem 2, by Theorem 5.1 and 5.2 of Chernozhukov et al. (2014), it suffices to show that

$$\mathrm{P}\|\psi(w;\beta_0,\eta_0,H_0)-\tilde{\psi}(w;\hat{\beta}_{one},\tilde{\beta},\tilde{\eta},\tilde{H})\|^2\lesssim\tau_n,\tag{S3.7}$$

and

$$\sup_Q\log N(\varepsilon\|F_5\|_{Q,2},\mathcal{F}_5,\|\cdot\|_{Q,2})\lesssim(d+s\vee s_h)\log(1/\varepsilon),\tag{S3.8}$$

where

$$\begin{aligned}\mathcal{F}_5&=\cup_{1\le j\le d}\mathcal{F}_5^{(j)},\ \text{and}\\ \mathcal{F}_5^{(j)}&=\{\psi(\cdot;\beta,\tilde{\beta},\eta,H)^{(j)}-\psi(\cdot;\beta_0,\eta_0,H_0)^{(j)}:\|\beta-\beta_0\|\vee\|\tilde{\beta}-\beta_0\|\le C\tau_n,\eta\in\mathcal{U}_\eta,H\in\mathcal{U}_H\}.\end{aligned}$$

Since (S3.8) can be shown by the similar argument as that of (S1.11), we only give the proof of (S3.7) below. By triangle inequality, we have

$$\begin{aligned}
&\mathrm{P}\|\psi(w;\beta_0,\eta_0,H_0)-\tilde{\psi}(w;\hat{\beta}_{one},\tilde{\beta},\tilde{\eta},\tilde{H})\|^2\\
&\le\mathrm{P}\|\psi(w;\beta_0,\eta_0,H_0)-\tilde{\psi}(w;\hat{\beta}_{one},\tilde{\beta},\tilde{\eta},H_0)\|^2\\
&\quad+\mathrm{P}\|\tilde{\psi}(w;\hat{\beta}_{one},\tilde{\beta},\tilde{\eta},\tilde{H})-\tilde{\psi}(w;\hat{\beta}_{one},\tilde{\beta},\tilde{\eta},H_0)\|^2 \qquad \text{(S3.9)}\\
&\le\mathrm{P}\|\psi(w;\beta_0,\eta_0,H_0)-\tilde{\psi}(w;\hat{\beta}_{one},\tilde{\beta},\tilde{\eta},H_0)\|^2+2\mathrm{P}\|(\tilde{H}-H_0)z\|^2\\
&=\mathrm{P}\|\psi(w;\beta_0,\eta_0,H_0)-\tilde{\psi}(w;\hat{\beta}_{one},\tilde{\beta},\tilde{\eta},H_0)\|^2+\tau_n^2,
\end{aligned}$$

where the last equation follows from Lemma 2. Let $g(x,z;\beta,\tilde{\beta},\eta)=\mathrm{E}[\psi_\tau\{y-(x-H_0z)'\beta-(H_0z)'\tilde{\beta}-z'\eta\}|x,z]$, $\Delta=x'(\beta-\beta_0)+z'(\eta-\eta_0)+(H_0z)'(\tilde{\beta}-\beta)$, and $\hat{\Delta}=x'(\hat{\beta}-\beta_0)+z'(\tilde{\eta}-\eta_0)+(H_0z)'(\tilde{\beta}-\hat{\beta})$. We have that

$$g(x,z;\hat{\beta},\tilde{\beta},\tilde{\eta})=\mathrm{E}\{\psi_\tau(\varepsilon-\hat{\Delta})|x,z\}=f(0|x,z)\hat{\Delta}+o_p(1),$$

and

$$\begin{aligned}
\mathrm{E}\{I_{\{y-(x-H_0z)'\hat{\beta}-(H_0z)'\tilde{\beta}-z'\tilde{\eta}\le 0\}}I_{\{y-x'\beta_0-z'\eta_0\le 0\}}|x,z\} =&\mathrm{E}\{I_{\{\varepsilon\le\hat{\Delta}\}}I_{\{\varepsilon\le 0\}}|x,z\}\\
=&\min\{F(\hat{\Delta}),F(0)\}.
\end{aligned}$$

Therefore, by Assumptions (A1) - (A4) and Lemmas 2 and 8, we have

$$\begin{aligned}
&\mathrm{P}\|\psi(w;\hat\beta_{one},\tilde\beta,\tilde\eta,H_0)-\psi(w;\beta_0,\eta_0,H_0)\|^2\\
&=\mathrm{P}\{\psi_\tau(\varepsilon-\hat\Delta)-\psi_\tau(\varepsilon)\}^2\|x-H_0z\|^2\\
&=\{\mathrm{P}I_{\{\varepsilon\le\hat\Delta\}}+\mathrm{P}I_{\{\varepsilon\le0\}}-2\mathrm{P}I_{\{\varepsilon\le\hat\Delta\}}I_{\{\varepsilon\le0\}}\}\|x-H_0z\|^2\\
&=\mathrm{P}\left[F_\varepsilon(\hat\Delta)+\tau-2\min\{F_\varepsilon(\hat\Delta),\tau\}\right]\|x-H_0z\|^2 \qquad \text{(S3.10)}\\
&\le\mathrm{P}\{f(0|x,z)(\|x'(\hat\beta_{one}-\beta)\|+\|z'(\tilde\eta-\eta_0)\|+(H_0z)'(\hat\beta_{one}-\tilde\beta))\}\\
&\quad\times\|x-H_0z\|^2(1+o_p(1))\\
&\lesssim\tau_n.
\end{aligned}$$

Thus, (S3.7) holds by substituting (S3.10) into (S3.9).

Since $\mathbb{P}_n\psi(w;\hat\beta_{one},\tilde\beta,\tilde\eta,\tilde H)\approx0$, (S3.6) can be rewritten as

$$\mathrm{P}\{\tilde\psi(w;\hat\beta_{one},\tilde\beta,\tilde\eta,\tilde H)\ -\ \psi(w;\beta_0,\eta_0,H_0)\}\ =\ -\mathbb{P}_n\psi(w;\beta_0,\eta_0,H_0)\ +\ o_p(n^{-1/2}). \qquad \text{(S3.11)}$$

Revoking the definitions of Δ and $\hat\Delta$, as the same way as the proof of Theorem 2, we have by Lemmas 2 and 7 that

$$\begin{aligned}
\mathrm{P}\tilde{\psi}(w;\hat{\beta}_{one},\tilde{\beta},\tilde{\eta},\tilde{H}) =& \frac{1}{n}\sum_{i=1}^{n} f(\hat{w}_i|x_i,z_i)(x_i-\tilde{H}z_i)(x_i-\tilde{H}z_i)'(\hat{\beta}_{one}-\beta_0) \\
&+ \frac{1}{n}\sum_{i=1}^{n} f(\hat{w}_i|x_i,z_i)(x_i-\tilde{H}z_i)(\tilde{H}z_i)'(\tilde{\beta}-\beta_0) \\
&+ \frac{1}{n}\sum_{i=1}^{n} f(\hat{w}_i|x_i,z_i)(x_i-\tilde{H}z_i)z_i'(\tilde{\eta}-\eta_0) + o_p(n^{-1/2}) \\
=& \frac{1}{n}\sum_{i=1}^{n} f(\hat{w}_i|x_i,z_i)(x_i-\tilde{H}z_i)(x_i-\tilde{H}z_i)'(\hat{\beta}_{one}-\beta_0) + o_p(n^{-1/2}),
\end{aligned}$$

where $\hat{w}_i$ is between $\hat{\Delta}_i$ and 0. This combing (S3.11) implies that

$$\frac{1}{n}\sum_{i=1}^{n} f(\hat{w}_i|x_i,z_i)(x_i-\tilde{H}z_i)(x_i-\tilde{H}z_i)'(\hat{\beta}_{one}-\beta_0) = -\mathbb{P}_n\psi(w;\beta_0,\eta_0,H_0)+o_p(n^{-1/2}).$$

The rest of the proof of Lemma 9 can be completed as the lines along the proof of Theorem 2. $\diamond$

Proof of Theorem 4.

Lemma 9 implies that

$$\sqrt{n_1}(\hat{\beta}_{one}-\beta_0) \xrightarrow{L} N(0,\tilde{Q}^{-1}D\tilde{Q}^{-1}),$$

where $\tilde{Q} = \mathrm{E}\left\{f(0|x,z)(x-H_0z)(x-H_0z)'\right\}$, and $D = \tau(1-\tau)\mathrm{E}\left\{(x-H_0z)(x-H_0z)'\right\}$.

This completes the proof of Theorem 4. $\diamond$

S4 Simulation Study

Two sample sizes $n = 50, 100$ and two penalties are used, and two quantile levels $\tau = 0.5$ and $\tau = 0.75$ are considered.

We simulate data from the model

$$y_i = \mu + \sum_{j=1}^{3} x_{ij}\beta_j + \sum_{k=1}^{199} z_{ik}\eta_k + e_i, \quad i = 1, \cdots, n,$$

where the covariate (x_i, z_i), and the model error e_i are independently generated from the multivariate normal distribution with mean zero and covariance Σ, and the standard normal distribution, respectively. We consider a sparsity structure with coefficients given as

$$(\mu, \beta_1, \beta_2, \beta_3, \eta_1, \eta_2, \eta_3, \cdots, \eta_{199}) = (3, 3, 3, 3, 3, 3, 0, \cdots, 0).$$

We refer to Section 7.1 of the main paper for the method used here.

We conduct a simulation according to two estimators of H_0 obtained from (2.8) and (2.9) in the main paper. The settings are the same as those of Section 7.1 in our main paper except the covariance of covariate (x, z) is Σ, which may not be the identity matrix. We generate 1000 bootstrap samples for each among 1000 replicates to estimate covariance matrix. Table 1 and 2 report the simulation results. The biases, the estimated relative efficiencies and coverage probabilities of the method with those two penalties are similar as indicated in Tables 1–2. If the covariate is correlated, the biases of parameter estimates increase, but larger sample size can dramatically reduce these biases.

Bibliography

Belloni, A. and V. Chernozhukov (2011, feb). $\ell 1$-penalized quantile regression in high-dimensional sparse models. *The Annals of Statistics 39*(1), 82–130.

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

Kosorok, M. R. (2008). *Introduction to Empirical Processes and Semiparametric Inference*. Springer.

Raskutti, G., M. J. Wainwright, and B. Yu (2010). Restricted eigenvalue properties for correlated gaussian designs. *Journal of Machine Learning Research 11*(Aug), 2241–2259.

van der Vaart, A. W. and J. A. Wellner (1996). *Weak Convergence and Empirical Processes*. Springer New York.

Table 1: Estimated coverage probability (CP) at 95% confidence level, and the estimated relative efficiencies (RE) and biases (Bias) of the proposed estimator (EC) and the oracle estimator (Oracle), where correlation $\mathrm{Cov}(x, z) = (\Sigma_{ij})$ with $\Sigma_{ij} = 0.2$ if $i \neq j$ and $\Sigma_{ij} = 1$ if $i = j$, and H is estimated by column-wise lasso penalty (glasso) given in (9) of main paper and element-wise lasso penalty (lasso) given in (8) of this response.

penality	(n, τ)	Parameter	Bias of EC ($\times 10^{-3}$)	Bias of Oracle ($\times 10^{-3}$)	RE	CP ($\times 100\%$)
glasso	$(50, 0.5)$	β_1	-9.608	-0.971	0.811	95.9
		β_2	0.945	1.993	0.701	95.8
		β_3	-3.486	-8.541	0.813	96.2
	$(50, 0.75)$	β_1	-2.744	-6.880	0.697	99.0
		β_2	2.891	-0.413	0.617	97.8
		β_3	-4.957	-12.802	0.690	98.7
		Parameter	Bias of EC ($\times 10^{-3}$)	Bias of Oracle ($\times 10^{-3}$)	RE	CP ($\times 100\%$)
	$(100, 0.5)$	β_1	-2.245	-1.684	0.992	95.6
		β_2	-2.913	0.455	0.919	96.5
		β_3	-7.060	-6.316	0.948	96.2
	$(100, 0.75)$	β_1	-5.663	-0.863	0.854	97.2
		β_2	1.615	1.790	0.927	97.7
		β_3	-8.156	-2.809	0.938	97.8
lasso	$(50, 0.5)$	β_1	-9.663	-0.971	0.811	95.8
		β_2	0.792	1.993	0.701	95.6
		β_3	-3.232	-8.541	0.813	96.6
	$(50, 0.75)$	β_1	1.462	-5.257	0.580	98.3
		β_2	4.563	0.439	0.620	98.5
		β_3	-2.773	-11.627	0.688	98.7
		Parameter	Bias of EC ($\times 10^{-3}$)	Bias of Oracle ($\times 10^{-3}$)	RE	CP ($\times 100\%$)
	$(100, 0.5)$	β_1	-2.250	-1.676	1.007	95.8
		β_2	-3.421	-0.511	0.918	96.7
		β_3	-6.641	-5.651	0.937	96.5
	$(100, 0.75)$	β_1	-4.736	-0.079	0.864	97.1
		β_2	1.661	1.978	0.934	97.6
		β_3	-8.247	-2.895	0.934	97.7

Table 2: Estimated coverage probability (CP) at 95% confidence level, and the estimated relative efficiencies (RE) and biases (Bias) of the proposed estimator (EC) and the oracle estimator (Oracle), where correlation $\mathrm{Cov}(x, z) = (\Sigma_{ij}) = I$, and H is estimated by column-wise lasso penalty (glasso) given in (9) of main paper and element-wise lasso penalty (lasso) given in (8) of this response.

penality	(n, τ)	Parameter	Bias of EC ($\times 10^{-3}$)	Bias of Oracle ($\times 10^{-3}$)	RE	CP ($\times 100\%$)
glasso	$(50, 0.5)$	β_1	46.196	2.490	0.809	95.9
		β_2	59.432	5.938	0.680	95.1
		β_3	47.222	-6.525	0.758	95.9
	$(50, 0.75)$	β_1	72.762	-3.514	0.717	97.8
		β_2	71.994	1.025	0.566	97.1
		β_3	60.511	-11.403	0.702	98.2
		Parameter	Bias of EC ($\times 10^{-3}$)	Bias of Oracle ($\times 10^{-3}$)	RE	CP ($\times 100\%$)
	$(100, 0.5)$	β_1	27.805	-0.900	0.913	95.6
		β_2	29.157	0.346	0.852	95.2
		β_3	21.411	-5.985	0.913	96.3
	$(100, 0.75)$	β_1	30.019	-0.062	0.866	97.5
		β_2	35.170	-0.247	0.877	97.7
		β_3	28.490	-2.543	0.891	98.0
lasso	$(50, 0.5)$	β_1	46.759	1.019	0.807	96.0
		β_2	54.662	4.431	0.658	95.8
		β_3	43.509	-7.917	0.584	96.8
	$(50, 0.75)$	β_1	74.170	-2.546	0.707	98.2
		β_2	71.452	6.504	0.542	97.8
		β_3	57.725	-10.903	0.624	98.3
		Parameter	Bias of EC ($\times 10^{-3}$)	Bias of Oracle ($\times 10^{-3}$)	RE	CP ($\times 100\%$)
	$(100, 0.5)$	β_1	26.810	-0.824	0.917	96.2
		β_2	28.900	1.417	0.851	96.2
		β_3	21.657	-6.674	0.908	96.1
	$(100, 0.75)$	β_1	30.275	0.784	0.869	97.6
		β_2	35.051	1.044	0.872	97.5
		β_3	26.188	-3.872	0.893	98.3