DATA INTEGRATION IN HIGH DIMENSION WITH MULTIPLE QUANTILES

Guorong Dai[a], Ursula U. Müller[b] and Raymond J. Carroll[b]

[a]*Department of Statistics and Data Science, School of Management, Fudan University Shanghai 200433, China*

[b]*Department of Statistics, Texas A&M University College Station, TX 77843, USA*

Supplementary Material

S1 Lemmas

Lemma 1. *Use the notation from Section 2 and write*

$$\widetilde{\beta}_{nkm} = n^{1/2}(X_{k\cdot a}^{\mathrm{T}} B_{nkm} X_{k\cdot a})^{-1} X_{k\cdot a}^{\mathrm{T}} \psi_{nkm}(\varepsilon)$$

for $k = 1, \ldots, K$ *and* $m = 1, \ldots, M$*. Then, provided Assumptions 1, 2, 3 and 4 are satisfied, we have* $\|\widetilde{\beta}_{nkm}\| = O_p\{(q_n \log n)^{1/2}\}$.

Proof of Lemma 1: We calculate

$$\begin{aligned}\|\widetilde{\beta}_{nkm}\|^2 &= n\psi_{nkm}(\varepsilon)^{\mathrm{T}}X_{k\cdot a}(X_{k\cdot a}^{\mathrm{T}}B_{nkm}X_{k\cdot a})^{-2}X_{k\cdot a}^{\mathrm{T}}\psi_{nkm}(\varepsilon)\\ &\leq \lambda_{\min}(n^{-1}X_{k\cdot a}^{\mathrm{T}}B_{nkm}X_{k\cdot a})^{-2}n^{-1}\psi_{nkm}(\varepsilon)^{\mathrm{T}}X_{k\cdot a}X_{k\cdot a}^{\mathrm{T}}\psi_{nkm}(\varepsilon)\\ &\leq Cn^{-1}\psi_{nkm}(\varepsilon)^{\mathrm{T}}X_{k\cdot a}X_{k\cdot a}^{\mathrm{T}}\psi_{nkm}(\varepsilon)\\ &\leq Cn^{-1}q_n(\max_{1\leq j\leq q_n}|\psi_{nkm}(\varepsilon)^{\mathrm{T}}X_{k\cdot j}|)^2\\ &= Cn^{-1}q_n(\max_{1\leq j\leq q_n}|\textstyle\sum_{i=1}^{n}\psi_{kmi}(\varepsilon)X_{kij}|)^2, \qquad \text{(S1.1)}\end{aligned}$$

where the third step uses Assumptions 2 and 3. Since $\psi_{kmi}(\varepsilon)X_{kij}$ has mean zero and is bounded by Assumption 1, Hoeffding's inequality gives

$$\mathrm{pr}\{|\textstyle\sum_{i=1}^{n}\psi_{kmi}(\varepsilon)X_{kij}| \geq L_n(n\log n)^{1/2}\} \leq 2\exp\{-CL_n^2\log n\}$$

for any positive sequence $L_n \to \infty$. It follows that

$$\begin{aligned}&\mathrm{pr}\{\max_{1\leq j\leq q_n}|\textstyle\sum_{i=1}^{n}\psi_{kmi}(\varepsilon)X_{kij}| \geq L_n(n\log n)^{1/2}\}\\ &\quad\leq \textstyle\sum_{j=1}^{q_n}\mathrm{pr}\{|\sum_{i=1}^{n}\psi_{kmi}(\varepsilon)X_{kij}| \geq L_n(n\log n)^{1/2}\}\\ &\quad\leq 2q_n\exp\{-CL_n^2\log n\} = 2q_n n^{-CL_n^2} \to 0, \qquad \text{(S1.2)}\end{aligned}$$

where the last step holds true because $q_n = o(n^{1/2})$; see Assumption 4. Therefore

$$\max_{1\leq j\leq q_n}|\textstyle\sum_{i=1}^{n}\psi_{kmi}(\varepsilon)X_{kij}| = O_p\{(n\log n)^{1/2}\}.$$

This combined with (S1.1) gives $\|\widetilde{\beta}_{nkm}\|^2 = O_p(q_n\log n)$, which completes the proof.

Lemma 2. *Set $\mathcal{M}_1^* = \{\mathcal{D} : \mathcal{D} \in \mathcal{M}, \mathcal{D}^* \subset \mathcal{D}\}$ and use the notation from Section 3. Let Assumptions 1, 3, 6 and 7 be satisfied. Let c_4 be the constant from Assumption 7. Then we have, for $k = 1, \ldots, K$, $m = 1, \ldots, M$, and any positive sequence L_n that tends to infinity and satisfies $L_n \to \infty$ and $1 \leq L_n (log\, n)^{1/2} \leq n^{1/10 - c_4/5}$,*

$$\lim_{L_n \to \infty} \mathrm{pr}\{|\textstyle\sum_{i=1}^n \big\{\rho_m(Y_{ki} - X_{ki\mathcal{D}}^{\mathrm{T}} \widehat{\theta}_{km\mathcal{D}}) - \rho_m(\varepsilon_{kmi})\}|$$
$$\leq L_n |\mathcal{D}| log\, n,\ \textit{for any } \mathcal{D} \in \mathcal{M}_1^*\big\} = 1.$$

Proof of Lemma 2: Under Assumptions 1, 3, 6 and 7, Lemma A.2 in the supplement to Lee et al. (2014) gives

$$\lim_{L\to\infty} \lim_{n\to\infty} \mathrm{pr}\{\|\widehat{\theta}_{km\mathcal{D}} - \theta^*_{km\mathcal{D}}\| \leq L n^{-1/2} (|\mathcal{D}| \log p_n)^{1/2},$$
$$\text{for any } \mathcal{D} \in \mathcal{M}_1^*\} = 1. \tag{S1.3}$$

Then, as $L_n \to \infty$,

$$\mathrm{pr}\{\|\widehat{\theta}_{km\mathcal{D}} - \theta^*_{km\mathcal{D}}\| \leq L_n n^{-1/2} (|\mathcal{D}| \log p_n)^{1/2},\ \text{for any } \mathcal{D} \in \mathcal{M}_1^*\} \to 1. \tag{S1.4}$$

Under Assumptions 1, 3, 6 and 7, and since $1 \leq L_n (\log n)^{1/2} \leq n^{1/10 - c_4/5}$, we can apply Lemma A.1 in the supplement to Lee et al. (2014), which gives

$$\max_{\mathcal{D}\in\mathcal{M}_1^*}\big|\,|\mathcal{D}|^{-1}[\widehat{V}_{km\mathcal{D}} - E(\widehat{V}_{km\mathcal{D}} \mid X_{k\cdot\mathcal{D}}) \\ + 2\textstyle\sum_{i=1}^n X_{ki\mathcal{D}}^{\mathrm{T}}(\widehat{\theta}_{km\mathcal{D}} - \theta_{km\mathcal{D}}^*)\psi_{kmi}(\varepsilon)]\,\big| = o_p(1) \qquad \text{(S1.5)}$$

with $\widehat{V}_{km\mathcal{D}} = \sum_{i=1}^n\{\rho_m(Y_{ki} - X_{ki\mathcal{D}}^{\mathrm{T}}\widehat{\theta}_{km\mathcal{D}}) - \rho_m(\varepsilon_{kmi})\}$. Then we have, on an event that has probability tending to one,

$$\begin{aligned}
&|\textstyle\sum_{i=1}^n X_{ki\mathcal{D}}^{\mathrm{T}}(\widehat{\theta}_{km\mathcal{D}} - \theta_{km\mathcal{D}}^*)\psi_{kmi}(\varepsilon)| \\
&\quad\le \|\widehat{\theta}_{km\mathcal{D}} - \theta_{km\mathcal{D}}^*\|\|\textstyle\sum_{i=1}^n X_{ki\mathcal{D}}\psi_{kmi}(\varepsilon)\| \\
&\quad\le \|\widehat{\theta}_{km\mathcal{D}} - \theta_{km\mathcal{D}}^*\||\mathcal{D}|^{1/2}\max_{1\le j\le p_n}|\textstyle\sum_{i=1}^n X_{kij}\psi_{kmi}(\varepsilon)| \\
&\quad\le L_n n^{-1/2}(|\mathcal{D}|\log p_n)^{1/2}|\mathcal{D}|^{1/2}L_n(n\log n)^{1/2} \\
&\quad= L_n^2|\mathcal{D}|\log n \qquad \text{(S1.6)}
\end{aligned}$$

for any $\mathcal{D} \in \mathcal{M}_1^*$. The last but one step uses (S1.2) and (S1.4). From Assumption 7 we have $p_n = O(n^{c_3})$. Hence (S1.2) holds true when q_n is substituted by p_n. We also have, for any $\theta_{\mathcal{D}} \in \mathbb{R}^{|\mathcal{D}|}$ satisfying $\|\theta_{\mathcal{D}} - \theta_{km\mathcal{D}}^*\| \le$

$L_n n^{-1/2}(|\mathcal{D}|\log p_n)^{1/2}$,

$$\begin{aligned}
&|\textstyle\sum_{i=1}^n E\{\rho_m(Y_{ki} - X_{ki\mathcal{D}}^{\mathrm{T}}\theta_{\mathcal{D}}) - \rho_m(\varepsilon_{kmi}) \mid X_{ki}\}| \\
&\quad = \textstyle\sum_{i=1}^n E\{\int_0^{X_{ki\mathcal{D}}^{\mathrm{T}}(\theta_{\mathcal{D}}-\theta_{km\mathcal{D}}^*)} I(\varepsilon_{kmi} \le s) - I(\varepsilon_{kmi} \le 0) ds \mid X_{ki}\} \\
&\quad = \textstyle\sum_{i=1}^n \int_0^{X_{ki\mathcal{D}}^{\mathrm{T}}(\theta_{\mathcal{D}}-\theta_{km\mathcal{D}}^*)} F_{km}(s \mid X_{ki}) - F_{km}(0 \mid X_{ki}) ds \\
&\quad = \textstyle\sum_{i=1}^n \int_0^{X_{ki\mathcal{D}}^{\mathrm{T}}(\theta_{\mathcal{D}}-\theta_{km\mathcal{D}}^*)} s f_{km}(\bar{s} \mid X_{ki}) ds \\
&\quad \le C(\theta_{\mathcal{D}} - \theta_{km\mathcal{D}}^*)^{\mathrm{T}} \textstyle\sum_{i=1}^n (X_{ki\mathcal{D}} X_{ki\mathcal{D}}^{\mathrm{T}})(\theta_{\mathcal{D}} - \theta_{km\mathcal{D}}^*) \\
&\quad \le Cn\lambda_{\max}(n^{-1} X_{k\cdot\mathcal{D}}^{\mathrm{T}} X_{k\cdot\mathcal{D}}) \|\theta_{\mathcal{D}} - \theta_{km\mathcal{D}}^*\|^2 \\
&\quad \le Cn\|\theta_{\mathcal{D}} - \theta_{km\mathcal{D}}^*\|^2 \le CL_n^2 |\mathcal{D}|\log p_n. \qquad \text{(S1.7)}
\end{aligned}$$

The first step in the above results is from Knight's identity (Knight, 1998). In the second step, $F_{km}(\cdot \mid X_k)$ is the conditional distribution function of ε_{km} given X_k. The third step uses a Taylor expansion with some $\bar{s}$ between 0 and $X_{ki\mathcal{D}}^{\mathrm{T}}(\theta_{\mathcal{D}} - \theta_{km\mathcal{D}}^*)$. The fourth step holds true because of Assumption 3 and the fact that

$$\begin{aligned}
\sup_{1\le i\le n}|X_{ki\mathcal{D}}^{\mathrm{T}}(\theta_{\mathcal{D}} - \theta_{km\mathcal{D}}^*)| &\le \sup_{1\le i\le n}\|X_{ki\mathcal{D}}\|\|\theta_{\mathcal{D}} - \theta_{km\mathcal{D}}^*\| \\
&\le CL_n d_n n^{-1/2}(\log n)^{1/2} \\
&\le Cn^{4c_4/5-2/5}(\log n)^{1/2} \to 0
\end{aligned}$$

from Assumptions 1 and 7. Combining (S1.4), (S1.5), (S1.6) and (S1.7) yields that, for any $\mathcal{D} \in \mathcal{M}_1^*$,

$$\widehat{V}_{km\mathcal{D}} \le |E(\widehat{V}_{km\mathcal{D}} \mid X_{k\cdot\mathcal{D}})| + 2|\textstyle\sum_{i=1}^n X_{ki\mathcal{D}}^{\mathrm{T}}(\widehat{\theta}_{km\mathcal{D}} - \theta_{km\mathcal{D}}^*)\psi_{kmi}(\varepsilon)| + |\mathcal{D}|o_p(1)$$
$$\le CL_n^2|\mathcal{D}|\log p_n + L_n^2|\mathcal{D}|\log n + |\mathcal{D}|o_p(1) \le CL_n^2|\mathcal{D}|\log n$$

with probability approaching one, where the $o_p(1)$ term comes from (S1.5). This finishes the proof.

S2 Proofs of the Theorems

Proof of Theorem 1: Under Assumptions 1-4, Lemma 6 of Sherwood and Wang (2016) gives

$$\|n^{1/2}(\widehat{\theta}_{km} - \theta_{km}^*) - \widetilde{\beta}_{nkm}\| = o_p(1) \tag{S2.1}$$

for every k and m, with $\widetilde{\beta}_{nkm}$ defined in Lemma 1. Therefore

$$\|\widehat{\theta}_{km} - \theta_{km}^*\| = O_p\{n^{-1/2}(q_n\log n)^{1/2}\}. \tag{S2.2}$$

It follows that for every k and m,

$$\max\nolimits_{1\le j\le q_n}|\widehat{\theta}_{kmj} - \theta_{kmj}^*| \le \|\widehat{\theta}_k - \theta_k^*\| = O_p\{n^{-1/2}(q_n\log n)^{1/2}\}$$
$$= O_p\{n^{(c_1-1)/2}(\log n)^{1/2}\}.$$

Hence

$$\max_{1\le j\le q_n}\|\widehat{\theta}^{(j)}-\theta^{*(j)}\|_1 \le KM \max_{1\le k\le K}\max_{1\le m\le M}\max_{1\le j\le q_n}|\widehat{\theta}_{kmj}-\theta^{*}_{kmj}|$$
$$= O_p\{n^{(c_1-1)/2}(\log n)^{1/2}\},$$

which, combined with Assumption 5, yields

$$\begin{aligned}\min\nolimits_{1\le j\le q_n}\|\widehat{\theta}^{(j)}\|_1 &\ge \min\nolimits_{1\le j\le q_n}\|\theta^{*(j)}\|_1 - \max\nolimits_{1\le j\le q_n}\|\widehat{\theta}^{(j)}-\theta^{*(j)}\|_1\\ &\ge Cn^{(c_2-1)/2}-\{n^{(c_1-1)/2}(\log n)^{1/2}\} = O_p\{n^{(c_2-1)/2}\}.\end{aligned}$$

We assume $\lambda_n = o\{n^{(c_2-1)/2}\}$, which implies

$$\mathrm{pr}\{\min\nolimits_{1\le j\le q_n}\|\widehat{\theta}^{(j)}\|_1 \ge a\lambda_n\}\to 1. \tag{S2.3}$$

The subderivative of the objective function (2.2) with respect to $\theta^{(j)}$ is

$$\frac{\partial\Gamma_{\lambda_n}(\theta)}{\partial\theta^{(j)}} = \begin{cases} \frac{\partial\ell_n(\theta)}{\partial\theta^{(j)}} + \lambda_n\mathbb{S}(\theta^{(j)}), & \|\theta^{(j)}\|_1 \le \lambda_n,\\ \frac{\partial\ell_n(\theta)}{\partial\theta^{(j)}} + \mathbb{S}(\theta^{(j)})\frac{(a\lambda_n-\|\theta^{(j)}\|_1)}{a-1}, & \lambda_n < \|\theta^{(j)}\|_1 < a\lambda_n,\\ \frac{\partial\ell_n(\theta)}{\partial\theta^{(j)}}, & a\lambda_n \le \|\theta^{(j)}\|_1,\end{cases} \tag{S2.4}$$

with

$$\mathbb{S}(\theta^{(j)}) = (\mathrm{Sign}(\theta_{11j}),\ldots,\mathrm{Sign}(\theta_{1Mj}),\ldots,\mathrm{Sign}(\theta_{K1j}),\ldots,\mathrm{Sign}(\theta_{KMj}))^{\mathrm{T}},$$

where $\mathrm{Sign}(x) = x/|x|$ for $x\ne 0$ and $\mathrm{Sign}(0)=[-1,1]$. Thus (S2.3) implies that, with probability tending to one, $\widehat{\theta}^{(j)}$ $(1\le j\le q_n)$ belongs to the third case in (S2.4). Combined with the fact that $\widehat{\theta}$ is a local minimizer of $\ell_n(\theta)$, it gives that

$$0 \in \partial \ell(\theta)/\partial \theta^{(j)}|_{\theta=\widehat{\theta}} = \partial \Gamma_{\lambda_n}(\theta)/\partial \theta^{(j)}|_{\theta=\widehat{\theta}}. \tag{S2.5}$$

Under Assumptions 1-5, the equation (3.5) in Lemma 1 of Sherwood and Wang (2016) yields that for every k and m,

$$\mathrm{pr}\{\max_{q_n<j\le p_n}|\partial \ell(\theta)/\partial \theta_{kmj}|_{\theta=\widehat{\theta}}| > \lambda_n\} \to 0. \tag{S2.6}$$

Since $\|\widehat{\theta}^{(j)}\|_1 = 0$ for $q_n < j \le p_n$, which belongs to the first case in (S2.4), we have

$$\partial \Gamma_{\lambda_n}(\theta)/\partial \theta^{(j)}|_{\theta=\widehat{\theta}} = \partial \ell(\theta)/\partial \theta^{(j)}|_{\theta=\widehat{\theta}} + \lambda_n \mathbb{S}(\mathbf{0}) \tag{S2.7}$$

Since $\mathbb{S}(\mathbf{0}) = \{(u_1, \ldots, u_K) : |u_k| \le 1, k = 1 \ldots, K\}$, (S2.6) and (S2.7) imply that for $q_n < j \le p_n$,

$$\mathrm{pr}\{0 \in \partial \Gamma_{\lambda_n}(\theta)/\partial \theta^{(j)}|_{\theta=\widehat{\theta}}\} \to 1. \tag{S2.8}$$

Combining (S2.5) and (S2.8) completes the proof.

<u>Proof of Theorem 2</u>: Set $\widehat{\beta}_n = n^{1/2}(\widehat{\theta}_a - \theta_a^*)$, $\widetilde{\beta}_n = n^{-1/2}R_n^{-1}X_a^{\mathrm{T}}\psi_n(\varepsilon)$ and write $A_n\Sigma_n^{-1/2}\widetilde{\beta}_n = \sum_{i=1}^n D_{ni}$, where $D_{ni} = n^{-1/2}A_n\Sigma_n^{-1/2}R_n^{-1}\delta_{ni}$, $\delta_{ni} = \{\psi_{1\cdot i}(\varepsilon)^{\mathrm{T}} \otimes X_{1ia}^{\mathrm{T}}, \ldots, \psi_{K\cdot i}(\varepsilon)^{\mathrm{T}} \otimes X_{Kia}^{\mathrm{T}}\}^{\mathrm{T}}$ and, for every k and i, $\psi_{k\cdot i}(\varepsilon) =$

$\{\psi_{k1i}(\varepsilon), \ldots, \psi_{kMi}(\varepsilon)\}^{\mathrm{T}}$. We have $E(D_{ni}) = \mathbf{0}$ since $E(\delta_{ni}) = \mathbf{0}$ and

$$\begin{aligned}\textstyle\sum_{i=1}^n E(D_{ni}D_{ni}^{\mathrm{T}}) &= n^{-1}E[A_n\Sigma_n^{-1/2}R_n^{-1}\{\textstyle\sum_{i=1}^n E(\delta_{ni}\delta_{ni}^{\mathrm{T}} \mid \mathcal{X})\}R_n^{-1}\Sigma_n^{-1/2}A_n^{\mathrm{T}}] \\ &= E\{A_n\Sigma_n^{-1/2}R_n^{-1}(n^{-1}X_a^{\mathrm{T}}H_nX_a)R_n^{-1}\Sigma_n^{-1/2}A_n^{\mathrm{T}}\} \\ &= E(A_n\Sigma_n^{-1/2}R_n^{-1}S_nR_n^{-1}\Sigma_n^{-1/2}A_n^{\mathrm{T}}) = A_nA_n^{\mathrm{T}} \to G.\end{aligned}$$

For any $\eta > 0$ we obtain

$$\begin{aligned}&\textstyle\sum_{i=1}^n E\{\|D_{ni}\|^2 I(\|D_{ni}\| > \eta)\} \\ &\le \eta^{-2}\textstyle\sum_{i=1}^n E(\|D_{ni}\|^4) \\ &= (n\eta)^{-2}\textstyle\sum_{i=1}^n E\{(\delta_{ni}^{\mathrm{T}}R_n^{-1}\Sigma_n^{-1/2}A_n^{\mathrm{T}}A_n\Sigma_n^{-1/2}R_n^{-1}\delta_{ni})^2\} \\ &\le (n\eta)^{-2}\lambda_{\max}^2(A_n^{\mathrm{T}}A_n)\textstyle\sum_{i=1}^n E\{(\delta_{ni}^{\mathrm{T}}R_n^{-1}\Sigma_n^{-1}R_n^{-1}\delta_{ni})^2\} \\ &\le Cn^{-2}\textstyle\sum_{i=1}^n E\{(\delta_{ni}^{\mathrm{T}}S_n^{-1}\delta_{ni})^2\} \\ &\le Cn^{-2}\textstyle\sum_{i=1}^n E\{\lambda_{\min}(S_n)^{-2}\|\delta_{ni}\|^4\} \\ &\le Cn^{-2}\textstyle\sum_{i=1}^n E(\|\delta_{ni}\|^4) \\ &= Cn^{-2}\textstyle\sum_{i=1}^n E\{(\textstyle\sum_{k=1}^K\textstyle\sum_{m=1}^M \psi_{kmi}(\varepsilon)^2\|X_{kia}\|^2)^2\} \\ &\le Cn^{-2}\textstyle\sum_{i=1}^n E\{(\max_{1\le k\le K}\|X_{kia}\|)^4\} \\ &\le Cn^{-1}E\{(\max_{1\le i\le n}\max_{1\le k\le K}\|X_{kia}\|)^4\} \\ &\le Cn^{-1}q_n^2 = o(1),\end{aligned}$$

with $\lambda_{\max}(\cdot)$ being the largest eigenvalue of a square matrix. The fourth step in the above display results from the fact that $\lambda_{\max}(A_n^{\mathrm{T}}A_n) \to C$. The sixth step uses the condition that $\lambda_{\min}(S_n)$ is uniformly bounded away from zero. The last but one step holds true because of Assumption 1, and the last step uses Assumption 4. This shows that the Lindeberg-Feller condition for the central limit theorem is satisfied, i.e. we have

$$A_n\Sigma_n^{-1/2}\widetilde{\beta}_n = \textstyle\sum_{i=1}^n D_{ni} \to N(\mathbf{0}, G) \text{ in distribution } (n \to \infty). \qquad \text{(S2.9)}$$

It is obvious that $\widetilde{\beta}_n = (\widetilde{\beta}_{n11}^{\mathrm{T}}, \ldots, \widetilde{\beta}_{n1M}^{\mathrm{T}}, \ldots, \widetilde{\beta}_{nK1}^{\mathrm{T}}, \ldots, \widetilde{\beta}_{nKM}^{\mathrm{T}})^{\mathrm{T}}$ with $\widetilde{\beta}_{nkm}$ defined in Lemma 1. Hence, using (S2.1), we have

$$\|\widehat{\beta}_n - \widetilde{\beta}_n\| \leq \textstyle\sum_{k=1}^K \sum_{m=1}^M \|\widehat{\beta}_{nkm} - \widetilde{\beta}_{nkm}\| = o_p(1).$$

It follows that

$$\begin{aligned} \|A_n\Sigma_n^{-1/2}(\widehat{\beta}_n - \widetilde{\beta}_n)\|^2 &= (\widehat{\beta}_n - \widetilde{\beta}_n)^{\mathrm{T}}\Sigma_n^{-1/2}A_nA_n^{\mathrm{T}}\Sigma_n^{-1/2}(\widehat{\beta}_n - \widetilde{\beta}_n) \\ &\leq \lambda_{\max}(A_nA_n^{\mathrm{T}})\lambda_{\min}(\Sigma_n)^{-1}\|\widehat{\beta}_n - \widetilde{\beta}_n\|^2 = o_p(1). \end{aligned}$$

In the last step we used $\lambda_{\max}(A_nA_n^{\mathrm{T}}) \to C$, Assumption 2 and the condition that $\lambda_{\min}(S_n)$ is uniformly bounded away from zero. This combined with (S2.9) yields

$$n^{1/2}A_n\Sigma_n^{-1/2}(\widehat{\theta}_a - \theta_a^*) = A_n\Sigma_n^{-1/2}\widehat{\beta}_n \to N(\mathbf{0}, G) \text{ in distribution } (n \to \infty).$$

Proof of Theorem 3: Consider the set of overfitted models $\mathcal{M}_1 = \{\mathcal{D} \in \mathcal{M} : \mathcal{D}^* \subset \mathcal{D}, \mathcal{D} \neq \mathcal{D}^*\}$ and the set of underfitted models $\mathcal{M}_2 = \{\mathcal{D} \in \mathcal{M} : \mathcal{D}^* \not\subset \mathcal{D}\}$. Since $\mathcal{M}_1 \cup \mathcal{M}_2 = \mathcal{M} \backslash \{\mathcal{D}^*\}$ it suffices to show

$$\lim_{n\to\infty}\text{pr}\{\min_{\mathcal{D}\in\mathcal{M}_1}\text{MQBIC}(\mathcal{D}) > \text{MQBIC}(\mathcal{D}^*)\} = 1, \quad \text{(S2.10)}$$

$$\lim_{n\to\infty}\text{pr}\{\min_{\mathcal{D}\in\mathcal{M}_2}\text{MQBIC}(\mathcal{D}) > \text{MQBIC}(\mathcal{D}^*)\} = 1. \quad \text{(S2.11)}$$

We first prove (S2.10). Write $\widehat{W}_{\mathcal{D}} = n^{-1}\sum_{k=1}^{K}\sum_{m=1}^{M}\sum_{i=1}^{n}\rho_m(Y_{ki} - X_{ki\mathcal{D}}^{\mathrm{T}}\widehat{\theta}_{km\mathcal{D}})$ and $W^* = n^{-1}\sum_{k=1}^{K}\sum_{m=1}^{M}\sum_{i=1}^{n}\rho_m(\varepsilon_{kmi})$. From Lemma 2 we know that we can choose some sequence L_n that does not depend on $\mathcal{D}$ and satisfies $L_n \to \infty$, $L_n = o(T_n)$ and $n^{-1}L_n d_n \log n \to 0$ such that for $k = 1, \ldots, K$ and $m = 1, \ldots, M$,

$$\text{pr}\{|\textstyle\sum_{i=1}^{n}\{\rho_m(Y_i - X_{ki\mathcal{D}}^{\mathrm{T}}\widehat{\theta}_{km\mathcal{D}}) - \rho_m(\varepsilon_{kmi})\}| \leq (MK)^{-1}L_n|\mathcal{D}|\log n, \text{ for any } \mathcal{D} \in \mathcal{M}_1^*\} \to 1. \quad \text{(S2.12)}$$

Since

$$|\widehat{W}_{\mathcal{D}} - W^*| \leq n^{-1}\textstyle\sum_{k=1}^{K}\sum_{m=1}^{M}|\sum_{i=1}^{n}\{\rho_m(Y_i - X_{ki\mathcal{D}}^{\mathrm{T}}\widehat{\theta}_{km\mathcal{D}}) - \rho_m(Y_i - X_{ki\mathcal{D}^*}^{\mathrm{T}}\theta_{km\mathcal{D}^*}^*)\}|,$$

we have

$$\text{pr}\{|\widehat{W}_{\mathcal{D}} - W^*| \leq n^{-1}L_n|\mathcal{D}|\log n, \text{ for any } \mathcal{D} \in \mathcal{M}_1^*\} \to 1.$$

It follows that

$$\text{pr}\{|\widehat{W}_{\mathcal{D}} - \widehat{W}_{\mathcal{D}^*}| \leq n^{-1}L_n(|\mathcal{D}| + |\mathcal{D}^*|)\log n, \quad \text{for any } \mathcal{D} \in \mathcal{M}_1^*\} \to 1 \tag{S2.13}$$

and that

$$\text{pr}\{\widehat{W}_{\mathcal{D}^*} \geq C, \text{ for any } \mathcal{D} \in \mathcal{M}_1^*\} \to 1. \tag{S2.14}$$

Here we used Assumption 9 and the fact that $n^{-1}L_n|\mathcal{D}^*|\log n \to 0$ (Assumption 7). Therefore, with probability tending to one,

$$\begin{aligned}
&\min_{\mathcal{D}\in\mathcal{M}_1}\text{MQBIC}(\mathcal{D}) - \text{MQBIC}(\mathcal{D}^*) \\
&= \min_{\mathcal{D}\in\mathcal{M}_1}[\log\{1 + \widehat{W}_{\mathcal{D}^*}^{-1}(\widehat{W}_{\mathcal{D}} - \widehat{W}_{\mathcal{D}^*})\} + (2n)^{-1}T_n(|\mathcal{D}| - |\mathcal{D}^*|)\log n] \\
&\geq \min_{\mathcal{D}\in\mathcal{M}_1}\{-2\widehat{W}_{\mathcal{D}^*}^{-1}|\widehat{W}_{\mathcal{D}} - \widehat{W}_{\mathcal{D}^*}| + (2n)^{-1}T_n(|\mathcal{D}| - |\mathcal{D}^*|)\log n\} \\
&\geq \min_{\mathcal{D}\in\mathcal{M}_1}\{-Cn^{-1}L_n(|\mathcal{D}| + |\mathcal{D}^*|)\log n + \\
&\qquad (2n)^{-1}T_n(|\mathcal{D}| - |\mathcal{D}^*|)\log n\}.
\end{aligned} \tag{S2.15}$$

The first inequality in the above derivation comes from the fact that $\log(1+x) \geq -2|x|$ for any $|x| \in (-1/2, 1/2)$, from equation (S2.13) combined with $n^{-1}L_n d_n \log n \to 0$, and from (S2.14). The last step holds true because of (S2.13) and (S2.14). Then (S2.15) implies (S2.10) because $L_n = o(T_n)$ and $|\mathcal{D}| > |\mathcal{D}^*|$.

To prove equation (S2.11) we introduce $\mathcal{D}' = \mathcal{D} \cup \mathcal{D}^*$ for any $\mathcal{D} \in$

$\mathcal{M}_2$. Since q is fixed by Assumption 7, there is a parameter with minimum absolute value $\nu > 0$, i.e. $\nu = \min_{1\le k\le K}\min_{1\le m\le M}\min_{j\in\mathcal{D}^*}|\theta^*_{kmj}| > 0$. Since (S1.3) still holds for any set in $\mathcal{M}_2^* = \{\mathcal{D}\subset\{1,\ldots,p_n\} : |\mathcal{D}| \le 2d_n, \mathcal{D}^*\subset\mathcal{D}\}$, we have

$$\mathrm{pr}\{\max\nolimits_{\mathcal{D}\in\mathcal{M}_2}\|\widehat{\theta}_{km\mathcal{D}'} - \theta^*_{km\mathcal{D}'}\| \le \nu\} \to 1. \tag{S2.16}$$

For $k = 1,\ldots,K$, $m = 1,\ldots,M$ and any $\mathcal{D}\in\mathcal{M}_2$, let $\widetilde{\theta}_{km\mathcal{D}'}$ be a $|\mathcal{D}'|\times 1$ vector, i.e. the dimension of $\widetilde{\theta}_{km\mathcal{D}'}$ is given by the number of indices in the set $\mathcal{D}' = \mathcal{D}\cup\mathcal{D}^*$. We define it as an extended version of $\widehat{\theta}_{km\mathcal{D}}$: the components of $\widetilde{\theta}_{km\mathcal{D}'}$ that correspond to the index set $\mathcal{D}$ coincide with the components of $\widehat{\theta}_{km\mathcal{D}}$; the remaining components are filled with zeros. For example, if $\mathcal{D} = \{1,3\}$, $\mathcal{D}^* = \{1,2\}$ and $\widehat{\theta}_{km\mathcal{D}} = \{1.4, 0.7\}$, then $\mathcal{D}' = \{1,2,3\}$, $|\mathcal{D}'| = 3$ and $\widetilde{\theta}_{km\mathcal{D}'} = (1.4, 0, 0.7)^{\mathrm{T}}$. Since $\mathcal{D}^*\not\subset\mathcal{D}$, there exist some k_0 and m_0 such that $\|\widetilde{\theta}_{k_0m_0\mathcal{D}'} - \theta^*_{k_0m_0\mathcal{D}'}\| \ge \nu$. Combined with (S2.16) and since the check function is convex, this implies that there exists a $|\mathcal{D}'|\times 1$ vector $\bar{\theta}_{\mathcal{D}'}$ such that $\|\bar{\theta}_{\mathcal{D}'} - \theta^*_{k_0m_0\mathcal{D}'}\| = \nu$ and

$$\begin{aligned}\textstyle\sum_{i=1}^n \rho_{m_0}(Y_{k_0i} - X^{\mathrm{T}}_{k_0i\mathcal{D}'}\bar{\theta}_{\mathcal{D}'}) &\le \textstyle\sum_{i=1}^n \rho_{m_0}(Y_{k_0i} - X^{\mathrm{T}}_{k_0i\mathcal{D}'}\widetilde{\theta}_{k_0m_0\mathcal{D}'})\\ &= \textstyle\sum_{i=1}^n \rho_{m_0}(Y_{k_0i} - X^{\mathrm{T}}_{k_0i\mathcal{D}}\widehat{\theta}_{k_0m_0\mathcal{D}}).\end{aligned}$$

Now set $G_{\mathcal{D}'}(\omega) = n^{-1}\sum_{i=1}^n\{\rho_{m_0}(\varepsilon_{k_0m_0i} - X^{\mathrm{T}}_{k_0i\mathcal{D}'}\omega) - \rho_{m_0}(\varepsilon_{k_0m_0i})\}$ and

$B_\nu(\mathcal{D}') = \{\omega \in \mathbb{R}^{|\mathcal{D}'|} : \|\omega\| = \nu\}$. Then we have, for any $\mathcal{D} \in \mathcal{M}_2$,

$$
\begin{aligned}
& n^{-1}\textstyle\sum_{i=1}^n \{\rho_{m_0}(Y_{k_0 i} - X_{k_0 i\mathcal{D}}^{\mathrm{T}}\widehat{\theta}_{k_0 m_0 \mathcal{D}}) - \rho_{m_0}(Y_{k_0 i} - X_{k_0 i\mathcal{D}'}^{\mathrm{T}}\widehat{\theta}_{k_0 m_0 \mathcal{D}'})\} \\
& \geq n^{-1}\textstyle\sum_{i=1}^n \{\rho_{m_0}(Y_{k_0 i} - X_{k_0 i\mathcal{D}}^{\mathrm{T}}\bar{\theta}_{\mathcal{D}'}) - \rho_{m_0}(Y_{k_0 i} - X_{k_0 i\mathcal{D}'}^{\mathrm{T}}\widehat{\theta}_{k_0 m_0 \mathcal{D}'})\} \\
& = G_{\mathcal{D}'}(\bar{\theta}_{\mathcal{D}'} - \theta^*_{k_0 m_0 \mathcal{D}'}) - G_{\mathcal{D}'}(\widehat{\theta}_{k_0 m_0 \mathcal{D}'} - \theta^*_{k_0 m_0 \mathcal{D}'}) + \\
& \quad E\{G_{\mathcal{D}'}(\bar{\theta}_{\mathcal{D}'} - \theta^*_{k_0 m_0 \mathcal{D}'}) \mid X_{k_0 \cdot \mathcal{D}'}\} - E\{G_{\mathcal{D}'}(\bar{\theta}_{\mathcal{D}'} - \theta^*_{k_0 m_0 \mathcal{D}'}) \mid X_{k_0 \cdot \mathcal{D}'}\} \\
& \geq \inf_{\omega \in B_\nu(\mathcal{D}')} E\{G_{\mathcal{D}'}(\omega) \mid X_{k_0 \cdot \mathcal{D}}\} \\
& \quad - \sup_{\omega \in B_\nu(\mathcal{D}')} |G_{\mathcal{D}'}(\omega) - E\{G_{\mathcal{D}'}(\omega) | X_{k_0 \cdot \mathcal{D}'}\}| - G_{\mathcal{D}'}(\widehat{\theta}_{k_0 m_0 \mathcal{D}'} - \theta^*_{k_0 m_0 \mathcal{D}'}).
\end{aligned} \tag{S2.17}
$$

Similar to (S1.7), we have, for any $\mathcal{D}' \in \mathcal{M}_2^*$ and $\omega \in B_\nu(\mathcal{D}')$,

$$
\begin{aligned}
& E\{G_{\mathcal{D}'}(\omega) \mid X_{k_0 \cdot \mathcal{D}'}\} \\
& = n^{-1}\textstyle\sum_{i=1}^n \int_0^{X_{k_0 i\mathcal{D}'}^{\mathrm{T}}\omega} F_{k_0 m_0}(s \mid X_{k_0 i\mathcal{D}'}) - F_{k_0 m_0}(0 \mid X_{k_0 i\mathcal{D}'}) ds \\
& = n^{-1}\textstyle\sum_{i=1}^n \int_0^{X_{k_0 i\mathcal{D}'}^{\mathrm{T}}\omega} s f_{k_0 m_0}(\bar{s} \mid X_{k_0 i\mathcal{D}'}) ds \\
& \geq C\omega^{\mathrm{T}}\{n^{-1}\textstyle\sum_{i=1}^n (X_{k_0 i\mathcal{D}'} X_{k_0 i\mathcal{D}'}^{\mathrm{T}})\}\omega \\
& \geq C\lambda_{\min}(n^{-1}X_{k_0 \cdot \mathcal{D}'}^{\mathrm{T}} X_{k_0 \cdot \mathcal{D}'})\|\omega\|^2 = C\|\omega\|^2,
\end{aligned} \tag{S2.18}
$$

where the third step uses Assumption (3) and the last step Assumption (6). Then, under Assumptions 1, 3, 6 and 7, Lemma A.3 in the supplement to Lee et al. (2014) gives

$$
\max_{\mathcal{D}' \in \mathcal{M}_2^*} \sup_{\omega \in B_\nu(\mathcal{D}')} |G_{\mathcal{D}'}(\omega) - E\{G_{\mathcal{D}'}(\omega) \mid X_{k_0 \cdot \mathcal{D}'}\}| = o_p(1). \tag{S2.19}
$$

It is obvious that (S2.12) is still valid when $\mathcal{M}_1^*$ is substituted by $\mathcal{M}_2^*$. Hence

$$\mathrm{pr}\{\max_{\mathcal{D}'\in\mathcal{M}_2^*}|G_{\mathcal{D}'}(\widehat{\theta}_{k_0m_0\mathcal{D}'}-\theta^*_{k_0m_0\mathcal{D}'})|\le Cn^{-1}L_nd_n\log n\}\to 1,$$

which gives $\max_{\mathcal{D}'\in\mathcal{M}_2^*}|G_{\mathcal{D}'}(\widehat{\theta}_{k_0m_0\mathcal{D}'}-\theta^*_{k_0m_0\mathcal{D}'})|=o_p(1)$. This, combined with (S2.17), (S2.18) and (S2.19) implies that, with probability approaching one,

$$n^{-1}\min_{\mathcal{D}\in\mathcal{M}_2}\sum_{i=1}^n\{\rho_m(Y_{k_0i}-X_{k_0i\mathcal{D}}^{\mathrm{T}}\widehat{\theta}_{k_0m_0\mathcal{D}})-\rho_m(Y_{k_0i}-X_{k_0i\mathcal{D}'}\widehat{\theta}_{k_0m_0\mathcal{D}'})\}\ge 2C. \quad \text{(S2.20)}$$

Since $\mathcal{D}\in\mathcal{D}'$ we have $\sum_{i=1}^n\{\rho_m(Y_{ki}-X_{ki\mathcal{D}}^{\mathrm{T}}\widehat{\theta}_{km\mathcal{D}})-\rho_m(Y_{ki}-X_{ki\mathcal{D}'}\widehat{\theta}_{km\mathcal{D}'})\}\ge 0$ for any k, m and $\mathcal{D}\in\mathcal{M}_2$. It follows

$$\begin{aligned}&\widehat{W}_{\mathcal{D}}-\widehat{W}_{\mathcal{D}'}\\&=n^{-1}\sum_{k=1}^K\sum_{m=1}^M\sum_{i=1}^n\{\rho_m(Y_{ki}-X_{ki\mathcal{D}}^{\mathrm{T}}\widehat{\theta}_{km\mathcal{D}})-\rho_m(Y_{ki}-X_{ki\mathcal{D}'}\widehat{\theta}_{km\mathcal{D}'})\}\\&\ge n^{-1}\sum_{i=1}^n\{\rho_m(Y_{k_0i}-X_{k_0i\mathcal{D}}^{\mathrm{T}}\widehat{\theta}_{k_0m_0\mathcal{D}})-\rho_m(Y_{k_0i}-X_{k_0i\mathcal{D}'}\widehat{\theta}_{k_0m_0\mathcal{D}'})\}.\end{aligned}$$

This, combined with (S2.20), gives

$$\mathrm{pr}\{\min_{\mathcal{D}\in\mathcal{M}_2}(\widehat{W}_{\mathcal{D}}-\widehat{W}_{\mathcal{D}'})\ge 2C\}\to 1. \quad \text{(S2.21)}$$

Then, with probability tending to one,

$$
\begin{aligned}
&\min_{\mathcal{D}\in\mathcal{M}_2}\text{MQBIC}(\mathcal{D}) - \text{MQBIC}(\mathcal{D}') \\
&= \min_{\mathcal{D}\in\mathcal{M}_2}[\log\{1+\widehat{W}_{\mathcal{D}'}^{-1}(\widehat{W}_{\mathcal{D}} - \widehat{W}_{\mathcal{D}'})\} - (2n)^{-1}T_n(|\mathcal{D}'| - |\mathcal{D}|)\log n] \\
&\geq \min_{\mathcal{D}\in\mathcal{M}_2}[\min\{\log 2, \widehat{W}_{\mathcal{D}'}^{-1}(\widehat{W}_{\mathcal{D}} - \widehat{W}_{\mathcal{D}'})/2\} - (2n)^{-1}T_n|\mathcal{D}^*|\log n] \\
&\geq \min_{\mathcal{D}\in\mathcal{M}_2}[\min\{\log 2, \widehat{W}_{\mathcal{D}'}^{-1}C\} - (2n)^{-1}T_n|\mathcal{D}^*|\log n] > 0 \qquad \text{(S2.22)}
\end{aligned}
$$

The first inequality comes from the fact that $\log(1+x) \geq \min\{x/2, \log 2\}$ for any $x \geq 0$. The second inequality uses (S2.21). The last step uses Assumption 8 and the fact that (S2.14) is still valid when $\mathcal{M}_1^*$ is substituted by $\mathcal{M}_2^*$. Since (S2.10) can be easily extended to any $\mathcal{D} \in (\mathcal{M}_2^*\backslash\{\mathcal{D}^*\})$, we know that, with probability tending to one, MQBIC($\mathcal{D}'$) $\geq$MQBIC($\mathcal{D}^*$) for any $\mathcal{D}' \in \mathcal{M}_2^*$. This and (S2.22) yield

$$
\begin{aligned}
&\min_{\mathcal{D}\in\mathcal{M}_2}\text{MQBIC}(\mathcal{D}) - \text{MQBIC}(\mathcal{D}^*) \\
&= \min_{\mathcal{D}\in\mathcal{M}_2}\{\text{MQBIC}(\mathcal{D}) - \text{MQBIC}(\mathcal{D}') + \text{MQBIC}(\mathcal{D}') - \text{MQBIC}(\mathcal{D}^*)\} \\
&\geq \min_{\mathcal{D}\in\mathcal{M}_2}\{\text{MQBIC}(\mathcal{D}) - \text{MQBIC}(\mathcal{D}')\} > 0,
\end{aligned}
$$

with probability tending to one. This proves (S2.11).

S3 Additional Results of Simulations

In this section we check the asymptotic normality stated in Theorem 2 of Section 2 using simulations. Under the setting of Table 2 in Section 4 with

$(n, p) = (200, 1000)$, $T = (\log p)/3$ and the regression model

$$Y_{ki} = X_{ki}^{\mathrm{T}} \alpha_k^* + 0.7 \xi_{ki} X_{ki3} \quad (k = 1, 2;\ i = 1, \dots n), \tag{S3.1}$$

we consider two components, $\widehat{\theta}_{113}$ and $\widehat{\theta}_{15(20)}$, of the estimator generated by our data integration (DI) approach. The corresponding covariates X_{1i3} and $X_{1i(20)}$ affect the response Y_{1i} via the terms $0.7\xi_{1i}X_{1i3}$ and $X_{1i}^{\mathrm{T}}\alpha_1^*$ in (S3.1), respectively. In Figures 1 and 2 we present the histograms of the two components based on $1,000$ simulated data sets. We can see the curves in the plots are unimodal, approximately symmetric and bell-shaped, which confirms the asymptotic normality stated in Theorem 2.

Bibliography

Knight, K. (1998). Limiting distributions for ℓ_1 regression estimators under general conditions. *Annals of Statistics* **18**, 755–770.

Lee, E. R., H. Noh, and B. U. Park (2014). Model selection via Bayesian information criterion for quantile regression models. *Journal of the American Statistical Association* **109**, 216–229.

Sherwood, B. and L. Wang (2016). Partially linear additive quantile regression in ultra-high dimension. *Annals of Statistics* **44**, 288–317.

Figure 1: Histogram of $\widehat{\theta}_{113}$ generated by our data integration (DI) method. The setting is the same as Table 2 in Section 4 with $(n, p) = (200, 1000)$ and $T = (\log p)/3$.

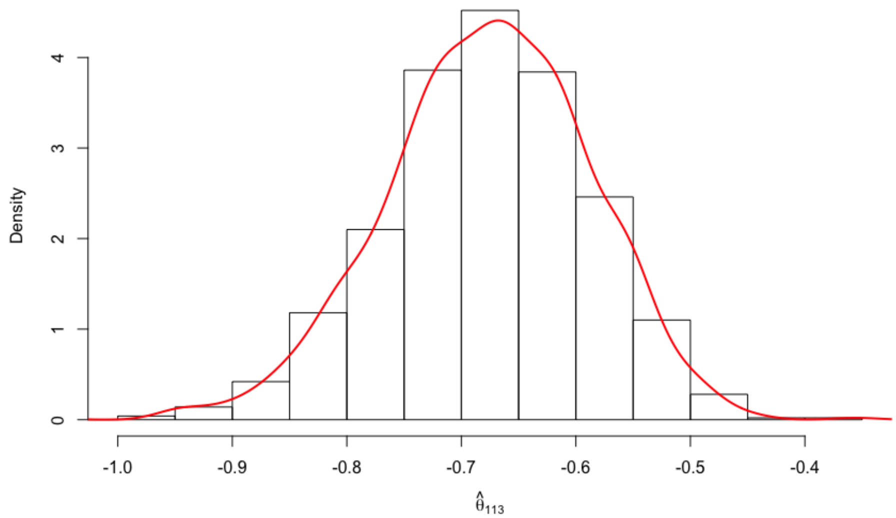

Figure 2: We consider the same scenario as Figure 1 but now investigate $\widehat{\theta}_{15(20)}$.