DATA INTEGRATION IN HIGH DIMENSION
WITH MULTIPLE QUANTILES

Guorong Dai\textsuperscript{a}, Ursula U. Müller\textsuperscript{b} and Raymond J. Carroll\textsuperscript{b}

\textsuperscript{a}Department of Statistics and Data Science, School of Management, Fudan University
Shanghai 200433, China
\textsuperscript{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

\[ \tilde{\beta}_{nkm} = n^{1/2}(X_{k-a}^T B_{nkm} X_{k-a})^{-1} X_{k-a}^T \psi_{nkm}(\varepsilon) \]

for \( k = 1, \ldots, K \) and \( m = 1, \ldots, M \). Then, provided Assumptions 1, 2, 3 and 4 are satisfied, we have \( \|\tilde{\beta}_{nkm}\| = O_p\{q_n \log n\}^{1/2} \).
Proof of Lemma 1: We calculate
\[
\|\tilde{\beta}_{nkm}\|^2 = n\psi_{nkm}(\varepsilon)^T X_{k:a}(X_{k:a}^T B_{nkm}X_{k:a})^{-2} X_{k:a}^T \psi_{nkm}(\varepsilon)
\leq \lambda_{\min}(n^{-1}X_{k:a}^T B_{nkm}X_{k:a})^{-2} n^{-1}\psi_{nkm}(\varepsilon)^T X_{k:a} X_{k:a}^T \psi_{nkm}(\varepsilon)
\leq Cn^{-1}\psi_{nkm}(\varepsilon)^T X_{k:a} X_{k:a}^T \psi_{nkm}(\varepsilon)
\leq Cn^{-1}q_n(\max_{1 \leq j \leq q_n}|\psi_{nkm}(\varepsilon)^T X_{k:j}|)^2
= Cn^{-1}q_n(\max_{1 \leq j \leq q_n}|\sum_{i=1}^n \psi_{kmi}(\varepsilon)X_{kij}|)^2,
\] (S1.1)
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
\[
\Pr\{|\sum_{i=1}^n \psi_{kmi}(\varepsilon)X_{kij}| \geq L_n(n \log n)^{1/2}\} \leq 2 \exp\{-CL_n^2 \log n\}
\]
for any positive sequence \(L_n \to \infty\). It follows that
\[
\Pr\{\max_{1 \leq j \leq q_n}|\sum_{i=1}^n \psi_{kmi}(\varepsilon)X_{kij}| \geq L_n(n \log n)^{1/2}\}
\leq \sum_{j=1}^{q_n} \Pr\{|\sum_{i=1}^n \psi_{kmi}(\varepsilon)X_{kij}| \geq L_n(n \log n)^{1/2}\}
\leq 2q_n \exp\{-CL_n^2 \log n\} = 2q_n n^{-CL_n^2} \to 0,
\] (S1.2)
where the last step holds true because \(q_n = o(n^{1/2})\); see Assumption 4. Therefore
\[
\max_{1 \leq j \leq q_n}|\sum_{i=1}^n \psi_{kmi}(\varepsilon)X_{kij}| = O_p\{(n \log n)^{1/2}\}.
\]
This combined with (S1.1) gives \(\|\tilde{\beta}_{nkm}\|^2 = O_p(q_n \log n)\), which completes the proof.
Lemma 2. Set $\mathcal{M}_1^* = \{D : D \in \mathcal{M}, D^* \subset D\}$ 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 $1 \leq L_n (\log n)^{1/2} \leq n^{1/10 - c_4/5}$,

$$
\lim_{L_n \to \infty} \lim_{n \to \infty} \Pr\{\left|\sum_{i=1}^{n} \{\rho_m(Y_{ki} - X_{ki}^T \hat{\theta}_{kmD}) - \rho_m(\varepsilon_{kmi})\}\right| \leq L_n |D| \log n, \text{ for any } D \in \mathcal{M}_1^*\} = 1.
$$

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

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

Then, as $L_n \to \infty$,

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

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

\[
\max_{D \in \mathcal{M}_1^*} |D|^{-1/2} \left| \hat{V}_{kmD} - E(\hat{V}_{kmD} \mid X_k^D) \right|
\]

\[
+ 2 \sum_{i=1}^n X_{kiD}^T (\hat{\theta}_{kmD} - \theta_{kmD}^*) \psi_{kmi}(\varepsilon) \right| = o_p(1) \quad (S1.5)
\]

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

\[
\left| \sum_{i=1}^n X_{kiD}^T (\hat{\theta}_{kmD} - \theta_{kmD}^*) \psi_{kmi}(\varepsilon) \right|
\]

\[
\leq \| \hat{\theta}_{kmD} - \theta_{kmD}^* \| \| \sum_{i=1}^n X_{kiD} \psi_{kmi}(\varepsilon) \|
\]

\[
\leq \| \hat{\theta}_{kmD} - \theta_{kmD}^* \| |D|^{1/2} \max_{1 \leq j \leq p_n} \left| \sum_{i=1}^n X_{kij} \psi_{kmi}(\varepsilon) \right|
\]

\[
\leq L_n n^{-1/2} (|D| \log p_n)^{1/2} |D|^{1/2} L_n (n \log n)^{1/2}
\]

\[
= L_n^2 |D| \log n \quad (S1.6)
\]

for any \( D \in \mathcal{M}_1^* \). 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 \( \theta_D \in \mathbb{R}^{|D|} \) satisfying \( \| \theta_D - \theta_{kmD}^* \| \leq \]
\[ L_n n^{-1/2} (|\mathcal{D}| \log p_n)^{1/2}, \]

\[
|\sum_{i=1}^{n} E\{\rho_m(Y_{ki} - X_{ki}^T \theta^*_{\mathcal{D}}) - \rho_m(\varepsilon_{kmi}) \mid X_{ki}\}| \\
= \sum_{i=1}^{n} E \left\{ \int_{0}^{X_{ki}^T (\theta^*_{\mathcal{D}} - \theta^*_{kmD})} F_{km}(s \mid X_{ki}) - F_{km}(0 \mid X_{ki}) \, ds \right\} \\
= \sum_{i=1}^{n} \int_{0}^{X_{ki}^T (\theta^*_{\mathcal{D}} - \theta^*_{kmD})} s f_{km}(\bar{s} \mid X_{ki}) \, ds \\
\leq C (\theta^*_{\mathcal{D}} - \theta^*_{kmD})^T \sum_{i=1}^{n} (X_{ki} X_{ki}^T) (\theta^*_{\mathcal{D}} - \theta^*_{kmD}) \\
\leq C n \lambda_{\max}(n^{-1} X_{kD} X_{kD}^T) \|\theta^*_{\mathcal{D}} - \theta^*_{kmD}\|^2 \\
\leq C n \|\theta^*_{\mathcal{D}} - \theta^*_{kmD}\|^2 \leq C L_n^2 |\mathcal{D}| \log p_n. \quad \text{(S1.7)}
\]

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 \( \varepsilon_{km} \) given \( X_k \). The third step uses a Taylor expansion with some \( \bar{s} \) between 0 and \( X_{ki}^T (\theta^*_{\mathcal{D}} - \theta^*_{kmD}) \). The fourth step holds true because of Assumption 3 and the fact that

\[
\sup_{1 \leq i \leq n} |X_{ki}^T (\theta^*_{\mathcal{D}} - \theta^*_{kmD})| \leq \sup_{1 \leq i \leq n} \|X_{ki}\| \|\theta^*_{\mathcal{D}} - \theta^*_{kmD}\| \\
\leq C L_n d_n n^{-1/2} (\log n)^{1/2} \\
\leq C n^{4c_4/5 - 2/5} (\log n)^{1/2} \to 0
\]

from Assumptions 1 and 7. Combining (S1.4), (S1.5), (S1.6) and (S1.7)
yields that, for any $D \in M_1^*$,

$$
\hat{V}_{kmD} \leq |E(\hat{V}_{kmD} \mid X_{kD})| + 2|\sum_{i=1}^{n} X_{kiD}^T(\hat{\theta}_{kmD} - \theta^*_{kmD})\psi_{kmi}(\varepsilon)| + |D|o_p(1)
$$

$$
\leq CL^2_n|D|\log p_n + L^2_n|D|\log n + |D|o_p(1) \leq CL^2_n|D|\log n
$$

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

This finishes the proof.

S2 Proofs of the Theorems


$$
\|n^{1/2}(\hat{\theta}_{km} - \theta^*_{km}) - \tilde{\beta}_{nkm}\| = o_p(1) \quad (S2.1)
$$

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

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

It follows that for every $k$ and $m$,

$$
\max_{1 \leq j \leq q_n}|\hat{\theta}_{kmj} - \theta^*_{kmj}| \leq \|\hat{\theta}_k - \theta^*_k\| = O_p\{n^{-1/2}(q_n\log n)^{1/2}\}
$$

$$
= O_p\{n^{(c_1-1)/2}(\log n)^{1/2}\}.
$$
Hence
\[
\max_{1 \leq j \leq q_n} \| \hat{\theta}^{(j)} - \theta^*(j) \|_1 \leq KM \max_{1 \leq k \leq K} \max_{1 \leq m \leq M} \max_{1 \leq j \leq q_n} | \hat{\theta}_{kmj} - \theta^*_{kmj} |
= O_p \{ n^{(c_1 - 1)/2} (\log n)^{1/2} \},
\]
which, combined with Assumption 5, yields
\[
\min_{1 \leq j \leq q_n} \| \hat{\theta}^{(j)} \|_1 \geq \min_{1 \leq j \leq q_n} \| \theta^*(j) \|_1 - \max_{1 \leq j \leq q_n} \| \hat{\theta}^{(j)} - \theta^*(j) \|_1 \geq C n^{(c_2 - 1)/2} - \{ n^{(c_1 - 1)/2} (\log n)^{1/2} \} = O_p \{ n^{(c_2 - 1)/2} \}.
\]
We assume \( \lambda_n = o \{ n^{(c_2 - 1)/2} \} \), which implies
\[
\text{pr} \{ \min_{1 \leq j \leq q_n} \| \hat{\theta}^{(j)} \|_1 \geq a \lambda_n \} \to 1. \quad \text{(S2.3)}
\]
The subderivative of the objective function (2.2) with respect to \( \theta^{(j)} \) is
\[
\frac{\partial \Gamma \lambda_n(\theta)}{\partial \theta^{(j)}} = \begin{cases} 
\frac{\partial \ell_n(\theta)}{\partial \theta^{(j)}} + \lambda_n S(\theta^{(j)}), & \| \theta^{(j)} \|_1 \leq \lambda_n, \\
\frac{\partial \ell_n(\theta)}{\partial \theta^{(j)}} + S(\theta^{(j)}) \left( \frac{a \lambda_n - \| \theta^{(j)} \|_1}{a-1} \right), & \lambda_n < \| \theta^{(j)} \|_1 < a \lambda_n, \\
\frac{\partial \ell_n(\theta)}{\partial \theta^{(j)}}, & a \lambda_n \leq \| \theta^{(j)} \|_1,
\end{cases} \quad \text{(S2.4)}
\]
with
\[
S(\theta^{(j)}) = (\text{Sign}(\theta_{11j}), \ldots, \text{Sign}(\theta_{1Mj}), \ldots, \text{Sign}(\theta_{K1j}), \ldots, \text{Sign}(\theta_{KMj}))^T,
\]
where \( \text{Sign}(x) = x/|x| \) for \( x \neq 0 \) and \( \text{Sign}(0) = [-1, 1] \). Thus (S2.3) implies that, with probability tending to one, \( \hat{\theta}^{(j)} \) \( (1 \leq j \leq q_n) \) belongs to the third case in (S2.4). Combined with the fact that \( \hat{\theta} \) is a local minimizer of \( \ell_n(\theta) \),
it gives that
\[ 0 \in \partial \ell(\theta) / \partial \theta^{(j)}|_{\theta = \hat{\theta}} = \partial \Gamma_{\lambda_n}(\theta) / \partial \theta^{(j)}|_{\theta = \hat{\theta}}. \]  
(S2.5)

Under Assumptions 1-5, the equation (3.5) in Lemma 1 of Sherwood and Wang (2016) yields that for every \( k \) and \( m \),
\[ \Pr\{\max_{q_n < j \leq p_n} |\partial \ell(\theta) / \partial \theta^{kmj}|_{\theta = \hat{\theta}} > \lambda_n\} \to 0. \]  
(S2.6)

Since \( \|\hat{\theta}^{(j)}\|_1 = 0 \) for \( q_n < j \leq p_n \), which belongs to the first case in (S2.4), we have
\[ \partial \Gamma_{\lambda_n}(\theta) / \partial \theta^{(j)}|_{\theta = \hat{\theta}} = \partial \ell(\theta) / \partial \theta^{(j)}|_{\theta = \hat{\theta}} + \lambda_n \mathbb{S}(0) \]  
(S2.7)

Since \( \mathbb{S}(0) = \{ (u_1, \ldots, u_K) : |u_k| \leq 1, k = 1 \ldots, K \} \), (S2.6) and (S2.7) imply that for \( q_n < j \leq p_n \),
\[ \Pr\{0 \in \partial \Gamma_{\lambda_n}(\theta) / \partial \theta^{(j)}|_{\theta = \hat{\theta}}\} \to 1. \]  
(S2.8)

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

**Proof of Theorem 2:** Set \( \hat{\beta}_n = n^{1/2}(\hat{\theta}_a - \theta^*_a) \), \( \tilde{\beta}_n = n^{-1/2}R_n^{-1}X_a^T\psi_n(\varepsilon) \) and write \( A_n \Sigma_n^{-1/2}\tilde{\beta}_n = \sum_{i=1}^n D_{ni} \), where \( D_{ni} = n^{-1/2}A_n \Sigma_n^{-1/2}R_n^{-1}d_{ni} \), \( d_{ni} = \{ \psi_{1,i}(\varepsilon)^T \otimes X_{1ia}^T, \ldots, \psi_{K,i}(\varepsilon)^T \otimes X_{Kia}^T \}^T \) and, for every \( k \) and \( i \), \( \psi_{k,i}(\varepsilon) = \ldots \)
\(\{\psi_{k1i}(\varepsilon), \ldots, \psi_{kMi}(\varepsilon)\}\). We have \(E(D_{ni}) = 0\) since \(E(\delta_{ni}) = 0\) and
\[\sum_{i=1}^{n} E(D_{ni}D_{ni}^T) = n^{-1} E[A_n \Sigma_n^{-1/2} R_n^{-1} \{\sum_{i=1}^{n} E(\delta_{ni} \delta_{ni}^T | \mathcal{X})\} R_n^{-1} \Sigma_n^{-1/2} A_n^T]\]
\[= E\{A_n \Sigma_n^{-1/2} R_n^{-1} (n^{-1} X_a H_n X_a) R_n^{-1} \Sigma_n^{-1/2} A_n^T\}\]
\[= E(A_n \Sigma_n^{-1/2} R_n^{-1} S_n R_n^{-1} \Sigma_n^{-1/2} A_n^T) = A_n A_n^T \to G.\]

For any \(\eta > 0\) we obtain
\[
\sum_{i=1}^{n} E\{\|D_{ni}\|^2 I(\|D_{ni}\| > \eta)\}
\leq \eta^{-2} \sum_{i=1}^{n} E(\|D_{ni}\|^4)
\leq (n \eta)^{-2} \sum_{i=1}^{n} E\{(\delta_{ni} R_n^{-1} \Sigma_n^{-1/2} A_n^T A_n \Sigma_n^{-1/2} R_n^{-1} \delta_{ni})^2\}
\leq (n \eta)^{-2} \lambda_{\text{max}}^2 (A_n^T A_n) \sum_{i=1}^{n} E\{(\delta_{ni} R_n^{-1} \Sigma_n^{-1} R_n^{-1} \delta_{ni})^2\}
\leq C n^{-2} \sum_{i=1}^{n} E\{(\delta_{ni} S_n^{-1} \delta_{ni})^2\}
\leq C n^{-2} \sum_{i=1}^{n} E\{\lambda_{\text{min}}(S_n)^{-2} \|\delta_{ni}\|^2\}
\leq C n^{-2} \sum_{i=1}^{n} E(\|\delta_{ni}\|^4)
= C n^{-2} \sum_{i=1}^{n} E\{(\sum_{k=1}^{K} \sum_{m=1}^{M} \psi_{kmi}(\varepsilon)^2 \|X_{kia}\|^2)^2\}
\leq C n^{-2} \sum_{i=1}^{n} E\{(\max_{1 \leq k \leq K} \|X_{kia}\|)^4\}
\leq C n^{-1} E\{(\max_{1 \leq i \leq n} \max_{1 \leq k \leq K} \|X_{kia}\|)^4\}
\leq C n^{-1} q_n^2 = o(1),
\]

with \(\lambda_{\text{max}}(\cdot)\) being the largest eigenvalue of a square matrix. The fourth step in the above display results from the fact that \(\lambda_{\text{max}}(A_n^T A_n) \to C\). 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

\[
A_n \Sigma_n^{-1/2} \tilde{\beta}_n = \sum_{i=1}^n D_{ni} \rightarrow N(0, G) \text{ in distribution } (n \rightarrow \infty). \tag{S2.9}
\]

It is obvious that \( \tilde{\beta}_n = (\tilde{\beta}_{n1}, \ldots, \tilde{\beta}_{nM}, \ldots, \tilde{\beta}_{nK1}, \ldots, \tilde{\beta}_{nKM})^T \) with \( \tilde{\beta}_{nkm} \) defined in Lemma 1. Hence, using (S2.1), we have

\[
\| \tilde{\beta}_n - \hat{\beta}_n \| \leq \sum_{k=1}^K \sum_{m=1}^M \| \tilde{\beta}_{nkm} - \hat{\beta}_{nkm} \| = o_p(1).
\]

It follows that

\[
\| A_n \Sigma_n^{-1/2} (\tilde{\beta}_n - \hat{\beta}_n) \|^2 = (\tilde{\beta}_n - \hat{\beta}_n)^T \Sigma_n^{-1/2} A_n A_n^T \Sigma_n^{-1/2} (\tilde{\beta}_n - \hat{\beta}_n) 
\leq \lambda_{\max}(A_n A_n^T) \lambda_{\min}(\Sigma_n)^{-1} \| \tilde{\beta}_n - \hat{\beta}_n \|^2 = o_p(1).
\]

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

\[
n^{1/2} A_n \Sigma_n^{-1/2} (\tilde{\theta}_a - \theta_a^*) = A_n \Sigma_n^{-1/2} \tilde{\beta}_n \rightarrow N(0, G) \text{ in distribution } (n \rightarrow \infty).
\]
Proof of Theorem 3: Consider the set of overfitted models $\mathcal{M}_1 = \{D \in \mathcal{M} : D^* \subset D, D \neq D^*\}$ and the set of underfitted models $\mathcal{M}_2 = \{D \in \mathcal{M} : D^* \not\subset D\}$. Since $\mathcal{M}_1 \cup \mathcal{M}_2 = \mathcal{M}\setminus\{D^*\}$ it suffices to show

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

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

We first prove (S2.10). Write $\hat{W}_D = n^{-1}\sum_{k=1}^K \sum_{m=1}^M \sum_{i=1}^n \rho_m(Y_{ki} - X_{ki}^T \hat{\theta}_{kmD})$ 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 $D$ 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 $m = 1, \ldots, M$,

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

Since

$$|\hat{W}_D - W^*| 
\leq n^{-1}\sum_{k=1}^K \sum_{m=1}^M |\sum_{i=1}^n \{\rho_m(Y_i - X_{ki}^T \hat{\theta}_{kmD}) - \rho_m(Y_i - X_{ki}^T \theta_{kmD}^*)\}|,$$

we have

$$\Pr\{|\hat{W}_D - W^*| \leq n^{-1}L_n|D|\log n, \text{ for any } D \in \mathcal{M}_1^* \} \to 1.$$
It follows that

$$\Pr\{|\hat{W}_D - \hat{W}_{D^*}| \leq n^{-1}L_n(|D| + |D^*|)\log n,$$

for any $D \in \mathcal{M}_1^*$ \to 1 \quad (S2.13)$$

and that

$$\Pr\{\hat{W}_{D^*} \geq C, \text{ for any } D \in \mathcal{M}_1^*\} \to 1. \quad (S2.14)$$

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

$$\min_{D \in \mathcal{M}_1} \text{MQBIC}(D) - \text{MQBIC}(D^*)$$

$$= \min_{D \in \mathcal{M}_1} \{\log\{1 + \hat{W}_D^{-1}(\hat{W}_D - \hat{W}_{D^*})\} + (2n)^{-1}T_n(|D| - |D^*|)\log n\}$$

$$\geq \min_{D \in \mathcal{M}_1} \{-2\hat{W}_D^{-1}|\hat{W}_D - \hat{W}_{D^*}| + (2n)^{-1}T_n(|D| - |D^*|)\log n\}$$

$$\geq \min_{D \in \mathcal{M}_1} \{-Cn^{-1}L_n(|D| + |D^*|)\log n +$$

$$\quad (2n)^{-1}T_n(|D| - |D^*|)\log n\}. \quad (S2.15)$$

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 $|D| > |D^*|$.

To prove equation (S2.11) we introduce $D' = D \cup D^*$ for any $D \in$
$\mathcal{M}_2$. Since $q$ is fixed by Assumption 7, there is a parameter with minimum absolute value $\nu > 0$, i.e. $\nu = \min_{1 \leq k \leq K} \min_{1 \leq m \leq M} \min_{j \in D^*} |\theta^*_{kmj}| > 0$.

Since (S1.3) still holds for any set in $\mathcal{M}_2^* = \{D \subset \{1, \ldots, p\} : |D| \leq 2d_n, D^* \subset D\}$, we have

$$\text{pr}\left\{\max_{D \in \mathcal{M}_2} \left\|\hat{\theta}_{kmD'} - \theta^*_{kmD'}\right\| \leq \nu\right\} \to 1. \quad (S2.16)$$

For $k = 1, \ldots, K$, $m = 1, \ldots, M$ and any $D \in \mathcal{M}_2$, let $\tilde{\theta}_{kmD'}$ be a $|D'| \times 1$ vector, i.e. the dimension of $\tilde{\theta}_{kmD'}$ is given by the number of indices in the set $D' = D \cup D^*$. We define it as an extended version of $\hat{\theta}_{kmD}$: the components of $\tilde{\theta}_{kmD'}$ that correspond to the index set $D$ coincide with the components of $\hat{\theta}_{kmD}$; the remaining components are filled with zeros. For example, if $D = \{1, 3\}$, $D^* = \{1, 2\}$ and $\hat{\theta}_{kmD} = \{1.4, 0.7\}$, then $D' = \{1, 2, 3\}$, $|D'| = 3$ and $\tilde{\theta}_{kmD'} = (1.4, 0.7)^T$. Since $D^* \not\subset D$, there exist some $k_0$ and $m_0$ such that $\|\tilde{\theta}_{km_0D'} - \theta^*_{km_0D'}\| \geq \nu$. Combined with (S2.16) and since the check function is convex, this implies that there exists a $|D'| \times 1$ vector $\tilde{\theta}_{D'}$ such that $\|\tilde{\theta}_{D'} - \theta^*_{km_0D'}\| = \nu$ and

$$\sum_{i=1}^n \rho_{m_0} (Y_{k_0i} - X_{k_0iD'}^T \tilde{\theta}_{D'}) \leq \sum_{i=1}^n \rho_{m_0} (Y_{k_0i} - X_{k_0iD}^T \hat{\theta}_{km_0D'}) = \sum_{i=1}^n \rho_{m_0} (Y_{k_0i} - X_{k_0iD}^T \tilde{\theta}_{km_0D}).$$

Now set $G_{D'}(\omega) = n^{-1} \sum_{i=1}^n \{\rho_{m_0}(\varepsilon_{k_0m_0i} - X_{k_0iD'}^T \omega) - \rho_{m_0}(\varepsilon_{k_0m_0i})\}$ and
$B_{\nu}(\mathcal{D}') = \{\omega \in \mathbb{R}^{\lvert \mathcal{D}' \rvert} : \|\omega\| = \nu\}$. Then we have, for any $\mathcal{D} \in \mathcal{M}_2$,

$$n^{-1}\sum_{i=1}^{n}\{\rho_{m_0}(Y_{ki} - X_{ki\mathcal{D}}^T\hat{\theta}_{k_0\mathcal{D}0}) - \rho_{m_0}(Y_{ki} - X_{ki\mathcal{D}}^T\hat{\theta}_{k_0\mathcal{D}'0})\}$$

$$\geq n^{-1}\sum_{i=1}^{n}\{\rho_{m_0}(Y_{ki} - X_{ki\mathcal{D}}^T\hat{\theta}_{\mathcal{D}0}) - \rho_{m_0}(Y_{ki} - X_{ki\mathcal{D}}^T\hat{\theta}_{k_0\mathcal{D}'0})\}$$

$$= G_{\mathcal{D}'}(\hat{\theta}_{\mathcal{D}0} - \theta^{*}_{k_0\mathcal{D}0}) - G_{\mathcal{D}'}(\hat{\theta}_{k_0\mathcal{D}'0} - \theta^{*}_{k_0\mathcal{D}'0}) +$$

$$E\{G_{\mathcal{D}'}(\hat{\theta}_{\mathcal{D}0} - \theta^{*}_{k_0\mathcal{D}0}) \mid X_{k_0\mathcal{D}'0}\} - E\{G_{\mathcal{D}'}(\hat{\theta}_{k_0\mathcal{D}'0} - \theta^{*}_{k_0\mathcal{D}'0}) \mid X_{k_0\mathcal{D}'0}\}$$

$$\geq \inf_{\omega \in B_{\nu}(\mathcal{D}')} E\{G_{\mathcal{D}'}(\omega) \mid X_{k_0\mathcal{D}'0}\} - \sup_{\omega \in B_{\nu}(\mathcal{D}')} |G_{\mathcal{D}'}(\omega) - E\{G_{\mathcal{D}'}(\omega) \mid X_{k_0\mathcal{D}'0}\}| - G_{\mathcal{D}'}(\hat{\theta}_{k_0\mathcal{D}'0} - \theta^{*}_{k_0\mathcal{D}'0}).$$

(S2.17)

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

$$E\{G_{\mathcal{D}'}(\omega) \mid X_{k_0\mathcal{D}'0}\}$$

$$= n^{-1}\sum_{i=1}^{n}\int_{0}^{X_{ki\mathcal{D}'}^T\omega} F_{k_0\mathcal{D}0}(s \mid X_{ki\mathcal{D}'0}) - F_{k_0\mathcal{D}0}(0 \mid X_{ki\mathcal{D}'0}) ds$$

$$= n^{-1}\sum_{i=1}^{n}\int_{0}^{X_{ki\mathcal{D}'}^T\omega} sf_{k_0\mathcal{D}0}(s \mid X_{ki\mathcal{D}'0}) ds$$

$$\geq C\omega^T\{n^{-1}\sum_{i=1}^{n}(X_{ki\mathcal{D}'0}X_{ki\mathcal{D}'}^T)\}\omega$$

$$\geq C\lambda_{\min}(n^{-1}X_{k_0\mathcal{D}'0}X_{k_0\mathcal{D}'0})\|\omega\|^2 = C\|\omega\|^2,$$

(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

$$\max_{\mathcal{D} \in \mathcal{M}_2^*} \sup_{\omega \in B_{\nu}(\mathcal{D}')} |G_{\mathcal{D}'}(\omega) - E\{G_{\mathcal{D}'}(\omega) \mid X_{k_0\mathcal{D}'0}\}| = o_p(1).$$

(S2.19)
It is obvious that (S2.12) is still valid when $M^*_1$ is substituted by $M^*_2$.

Hence

$$\text{pr}\left\{ \max_{D' \in M^*_2} |G_{D'}(\hat{\theta}_{k,0,m} - \theta^*_{k,0,m})| \leq C n^{-1} d_n \log n \right\} \to 1,$$

which gives $\max_{D' \in M^*_2} |G_{D'}(\hat{\theta}_{k,0,m} - \theta^*_{k,0,m})| = o_P(1)$. This, combined with (S2.17), (S2.18) and (S2.19) implies that, with probability approaching one,

$$n^{-1} \min_{D \in M_2} \sum_{i=1}^n \left\{ \rho_m(Y_{k,i} - X_{k,i,D}^T \hat{\theta}_{k,0,m}) - \rho_m(Y_{k,i} - X_{k,i,D'}^T \hat{\theta}_{k,0,m}'D') \right\} \geq 2C. \quad (S2.20)$$

Since $D \in D'$ we have $\sum_{i=1}^n \left\{ \rho_m(Y_{k,i} - X_{k,i,D}^T \hat{\theta}_{k,m,D}) - \rho_m(Y_{k,i} - X_{k,i,D'}^T \hat{\theta}_{k,m,D'}) \right\} \geq 0$

for any $k$, $m$ and $D \in M_2$. It follows

$$\hat{W}_D - \hat{W}_{D'}$$

$$= n^{-1} \sum_{k=1}^K \sum_{m=1}^M \sum_{i=1}^n \left\{ \rho_m(Y_{k,i} - X_{k,i,D}^T \hat{\theta}_{k,m,D}) - \rho_m(Y_{k,i} - X_{k,i,D'}^T \hat{\theta}_{k,m,D'}) \right\}$$

$$\geq n^{-1} \sum_{i=1}^n \left\{ \rho_m(Y_{k,i} - X_{k,i,D}^T \hat{\theta}_{k,0,m}) - \rho_m(Y_{k,i} - X_{k,i,D'}^T \hat{\theta}_{k,0,m}'D') \right\}.$$

This, combined with (S2.20), gives

$$\text{pr}\left\{ \min_{D \in M_2} (\hat{W}_D - \hat{W}_{D'}) \geq 2C \right\} \to 1. \quad (S2.21)$$
Then, with probability tending to one,

\[
\min_{D \in \mathcal{M}_2} MQBIC(D) - MQBIC(D') = \min_{D \in \mathcal{M}_2} \{MQBIC(D) - MQBIC(D') + MQBIC(D') - MQBIC(D)\} \\
\geq \min_{D \in \mathcal{M}_2} \{MQBIC(D) - MQBIC(D')\} > 0
\] (S2.22)

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 \(D \in (\mathcal{M}^*_2 \setminus \{D^*\})\), we know that, with probability tending to one, \(MQBIC(D') \geq MQBIC(D^*)\) for any \(D' \in \mathcal{M}^*_2\). This and (S2.22) yield

\[
\min_{D \in \mathcal{M}_2} MQBIC(D) - MQBIC(D^*) = \min_{D \in \mathcal{M}_2} \{MQBIC(D) - MQBIC(D') + MQBIC(D') - MQBIC(D^*)\} \\
\geq \min_{D \in \mathcal{M}_2} \{MQBIC(D) - MQBIC(D')\} > 0,
\]

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

\[ Y_{ki} = X_{ki}^T\alpha^*_k + 0.7\xi_{ki}X_{ki3} \quad (k = 1, 2; i = 1, \ldots, n), \]  

(S3.1)

we consider two components, \( \hat{\theta}_{113} \) and \( \hat{\theta}_{15(20)} \), 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 \( X_{1i}^T\alpha^*_1 \) 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


Figure 1: Histogram of $\hat{\theta}_{113}$ 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 $\hat{\theta}_{15(20)}$. 

Figure