Supplementary Material for "Multi-round smoothed composite quantile regression for distributed data"

Fengrui Di[1], Lei Wang[1*]

[1] *School of Statistics and Data Science & LPMC, Nankai University*

No. 94 Weijin Road, Tianjin 300071, China.

Proof of Proposition 1

Let

$$\boldsymbol{C} = \boldsymbol{P}_N - \boldsymbol{P} = \frac{1}{N}\sum_{k=1}^{K}\sum_{i=1}^{N}\boldsymbol{x}_i\Big\{H\Big(\frac{y_i - \tilde{b}_{\tau_k} - \boldsymbol{x}_i^T\tilde{\boldsymbol{\beta}}}{h}\Big) - I\{y_i - b_{\tau_k} - \boldsymbol{x}_i^T\boldsymbol{\beta} \geq 0\} + \frac{y_i - \tilde{b}_{\tau_k} - \boldsymbol{x}_i^T\boldsymbol{\beta}}{h}H'\Big(\frac{y_i - \tilde{b}_{\tau_k} - \boldsymbol{x}_i^T\tilde{\boldsymbol{\beta}}}{h}\Big)\Big\}.$$

Note that $\|\boldsymbol{C}\|_2 = \sup_{v\in\mathbb{R}^p\|v\|_2=1}|v^T\boldsymbol{C}|$. Let $S_{1/2}^{p-1}$ be a 1/2 net of the unit sphere S^{p-1} in the Euclidean distance in $\mathbb{R}^p$. According to the proof of Lemma 3 in Cai et al. (2010),

$$d_p := Card(S_{1/2}^{p-1}) \leq 5^p.$$

Let $v_1, \ldots, v_{d_p}$ be the centers of the d_p elements in the net. Therefore, for any v in S^{p-1}, it can be shown that $\|v - v_j\|_2 \leq 1/2$ for some j and $\|\boldsymbol{C}\|_2 \leq \sup_{j\leq d_p}|v_j^T\boldsymbol{C}| + \|\boldsymbol{C}\|_2/2$. Above all lead to

$$\|\boldsymbol{C}\|_2 \leq 2\sup_{j\leq d_p}|v_j^T\boldsymbol{C}|.$$

For $\alpha \in \mathbb{R}^p$, denote $\boldsymbol{C}_j(\alpha) = \frac{1}{N}\sum_{k=1}^{K}\sum_{i=1}^{N}v_j^T\boldsymbol{x}_iH(\alpha)$ and

$$H(\alpha) = H\Big(\frac{y_i - \tilde{b}_{\tau_k} - \boldsymbol{x}_i^T\alpha}{h}\Big) - I\{y_i - b_{\tau_k} - \boldsymbol{x}_i^T\boldsymbol{\beta} \geq 0\} + \frac{y_i - \tilde{b}_{\tau_k} - \boldsymbol{x}_i^T\boldsymbol{\beta}}{h}H'\Big(\frac{y_i - \tilde{b}_{\tau_k} - \boldsymbol{x}_i^T\alpha}{h}\Big).$$

*Address for correspondence: Lei Wang, School of Statistics and Data Science & LPMC, Nankai University. E-mail: lwangstat@nankai.edu.cn.

When $\|\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}\|_2 \le a_n$, it leads to

$$\|\boldsymbol{C}\|_2 \le 2 \sup_{j \le d_p} \sup_{\|\alpha - \beta\|_2 \le a_n} |\boldsymbol{C}_j(\alpha)|. \tag{1}$$

For every i, we divide the interval $[\beta_i - a_n, \beta_i + a_n]$ into n^M small subintervals and each has length $2a_n/n^M$, where M is a large positive number. Therefore, there exist a set of points $\left\{\alpha_l, 1 \le l \le n^{Mp}\right\} \in \mathbb{R}^p$ such that for any α in the ball $\|\alpha - \boldsymbol{\beta}\|_2 \le a_n$,

$$\|\alpha - \alpha_l\|_2 \le 2\sqrt{p} a_n / n^M,$$

for some $1 \le l \le n^{Mp}$. Let $\Delta(\alpha) = \alpha - \boldsymbol{\beta}$ and it can be proved that

$$\begin{aligned} \frac{y_i - \tilde{b}_{\tau_k} - \boldsymbol{x}_i^T \boldsymbol{\beta}}{h} H'\Big(\frac{y_i - \tilde{b}_{\tau_k} - \boldsymbol{x}_i^T \alpha}{h}\Big) &= \frac{\boldsymbol{x}_i^T \Delta(\alpha)}{h} H'\Big(\frac{y_i - \tilde{b}_{\tau_k} - \boldsymbol{x}_i^T \boldsymbol{\beta} - \boldsymbol{x}_i^T \Delta(\alpha)}{h}\Big) \\ &+ \frac{y_i - \tilde{b}_{\tau_k} - \boldsymbol{x}_i^T \boldsymbol{\beta} - \boldsymbol{x}_i^T \Delta(\alpha)}{h} H'\Big(\frac{y_i - \tilde{b}_{\tau_k} - \boldsymbol{x}_i^T \boldsymbol{\beta} - \boldsymbol{x}_i^T \Delta(\alpha)}{h}\Big). \end{aligned} \tag{2}$$

Note that $xH'(x)$ and $H'(x)$ are Lipschitz continuous. When $\|\alpha - \boldsymbol{\beta}\|_2 \le a_n$ and $\|\alpha - \alpha_l\|_2 \le 2\sqrt{p} a_n / n^M$ for some $1 \le l \le n^{Mp}$,

$$\begin{aligned} &\Big|\frac{y_i - \tilde{b}_{\tau_k} - \boldsymbol{x}_i^T \boldsymbol{\beta}}{h} H'\Big(\frac{y_i - \tilde{b}_{\tau_k} - \boldsymbol{x}_i^T \alpha}{h}\Big) - \frac{y_i - \tilde{b}_{\tau_k} - \boldsymbol{x}_i^T \boldsymbol{\beta}}{h} H'\Big(\frac{y_i - \tilde{b}_{\tau_k} - \boldsymbol{x}_i^T \alpha_l}{h}\Big)\Big| \\ &\le Ch^{-1} \|\boldsymbol{x}_i\|_2 \|\alpha - \alpha_l\|_2 + Ch^{-2} \|\boldsymbol{x}_i\|_2^2 \|\Delta(\alpha)\|_2 \|\alpha - \alpha_l\|_2 \\ &\le Ch^{-1} \sqrt{p} a_n n^{-M} \|\boldsymbol{x}_i\|_2 + Ch^{-2} \sqrt{p} a_n^2 n^{-M} \|\boldsymbol{x}_i\|_2^2. \end{aligned}$$

Similarly, we can prove that

$$\Big|H\Big(\frac{y_i - \tilde{b}_{\tau_k} - \boldsymbol{x}_i^T \alpha}{h}\Big) - H\Big(\frac{y_i - \tilde{b}_{\tau_k} - \boldsymbol{x}_i^T \alpha_l}{h}\Big)\Big| \le Ch^{-1} \sqrt{p} a_n n^{-M} \|\boldsymbol{x}_i\|_2.$$

Therefore, it leads to

$$\sup_j \sup_{\|\alpha - \beta\|_2 \le a_n} |\boldsymbol{C}_j(\alpha)| - \sup_j \sup_l |\boldsymbol{C}_j(\alpha_l)| \le C \frac{\sqrt{p} a_n K}{n^M h N} \sum_{i=1}^N \|\boldsymbol{x}_i\|_2^2 + C \frac{\sqrt{p} a_n^2 K}{n^M h^2 N} \sum_{i=1}^N \|\boldsymbol{x}_i\|_2^3.$$

Since $\max_{ij} E|x_{ij}|^3 < \infty$, by letting M large enough, we have

$$\sup_j \sup_{\|\alpha - \beta\|_2 \le a_n} |\boldsymbol{C}_j(\alpha)| - \sup_j \sup_l |\boldsymbol{C}_j(\alpha_l)| = O_P(n^{-\gamma}), \tag{3}$$

for any $\gamma > 0$. It is enough to show that $\sup_j \sup_l |\boldsymbol{C}_j(\alpha_l)|$ satisfies the bound in the Proposition 1.

We first prove the proposition under (C3)*. Let $\hat{x}_{ij} = x_{ij} I\{|x_{ij}| \le N^\kappa\}$, $\hat{\boldsymbol{x}}_i = (\hat{x}_{i1}, \ldots, \hat{x}_{ip})^T$, $\epsilon_{ik} = y_i - b_{\tau_k} - \boldsymbol{x}_i^T \boldsymbol{\beta}$, $\Delta(b_k) = \tilde{b}_{\tau_k} - b_{\tau_k}$ and $|\tilde{b}_{\tau_k} - b_{\tau_k}| \le b_n$ for $i = 1, \ldots, N$ and $k = 1, \ldots, K$. Denote $\widehat{\boldsymbol{C}}_j(\alpha) = \frac{1}{N} \sum_{k=1}^K \sum_{i=1}^N v_j^T \hat{\boldsymbol{x}}_i \widehat{H}(\alpha)$ and

$$\widehat{H}(\alpha) = H\Big(\frac{\epsilon_{ik} - \Delta(b_k) - \hat{\boldsymbol{x}}_i^T \Delta(\alpha)}{h}\Big) - I\{\epsilon_{ik} \ge 0\} + \frac{\epsilon_{ik} - \Delta(b_k)}{h} H'\Big(\frac{\epsilon_{ik} - \Delta(b_k) - \hat{\boldsymbol{x}}_i^T \Delta(\alpha)}{h}\Big).$$

Using Chebyshev's inequality, we have

$$P\Big(\sup_j \sup_l |\boldsymbol{C}_j(\alpha_l)| \ne \sup_j \sup_l |\widehat{\boldsymbol{C}}_j(\alpha_l)|\Big) \le \sum_{i=1}^N \sum_{j=1}^p P(|x_{ij}| \ge N^\kappa) = O(p/N) = o(1).$$

Denote $f_1(\cdot|\boldsymbol{x}), \ldots, f_K(\cdot|\boldsymbol{x})$ as the corresponding conditional density functions of $\epsilon_1, \ldots, \epsilon_K$ given $\boldsymbol{x}$, which are local-scale functions for $f(\cdot|\boldsymbol{x})$ in Condition (C1). Since the conditional density functions $f_1(\cdot|\boldsymbol{x}), \ldots, f_K(\cdot|\boldsymbol{x})$ are Lipschitz continuous, they are also bounded. By Condition (C2) that $H(x)$ is bounded,

$$\begin{aligned} & E\Big[H\Big(\frac{\epsilon_{ik} - \Delta(b_k) - \hat{\boldsymbol{x}}_i^T \Delta(\alpha)}{h}\Big) - I\{\epsilon_{ik} \ge 0\}\Big|\boldsymbol{x}_i\Big]^2 \\ = h \int_{-\infty}^{\infty} & [H(x) - I\{x \ge -\hat{\boldsymbol{x}}_i^T \Delta(\alpha)/h - \Delta(b_k)/h\}]^2 f_k(hx + \hat{\boldsymbol{x}}_i^T \Delta(\alpha) + \Delta(b_k)|\boldsymbol{x}_i) dx \\ & \le Ch\Big(|\hat{\boldsymbol{x}}_i^T \Delta(\alpha)/h + \Delta(b_k)/h| + 1\Big), \end{aligned} \tag{4}$$

and

$$E\Big[\frac{\epsilon_{ik} - \Delta(b_k)}{h} H'\Big(\frac{\epsilon_{ik} - \Delta(b_k) - \hat{\boldsymbol{x}}_i^T \Delta(\alpha)}{h}\Big)\Big|\boldsymbol{x}_i\Big]^2 \le Ch\Big(|\hat{\boldsymbol{x}}_i^T \Delta(\alpha)/h|^2 + 1\Big). \tag{5}$$

Based on (4), (5) and the definition of $\widehat{H}(\alpha)$, it leads to

$$\begin{aligned} E\Big(v_j^T \hat{\boldsymbol{x}}_i \widehat{H}(\alpha)\Big)^2 \le & 2E\Big(v_j^T \hat{\boldsymbol{x}}_i\Big)^2 \Big[H\Big(\frac{\epsilon_{ik} - \hat{\boldsymbol{x}}_i^T \Delta(\alpha) - \Delta(b_k)}{h}\Big) - I\{\epsilon_{ik} \ge 0\}\Big]^2 \\ & + 2E\Big(v_j^T \hat{\boldsymbol{x}}_i\Big)^2 \Big[\frac{\epsilon_{ik} - \Delta(b_k)}{h} H'\Big(\frac{\epsilon_{ik} - \Delta(b_k) - \hat{\boldsymbol{x}}_i^T \Delta(\alpha)}{h}\Big)\Big]^2 \\ \le & Ch\Big(\sup_{\|\theta\|_2 = 1} E(\theta^T \hat{\boldsymbol{x}}_i)^2 + \sup_{\|\theta\|_2 = 1} E|\theta^T \hat{\boldsymbol{x}}_i|^3 \|\alpha - \boldsymbol{\beta}\|_2/h \\ & + \sup_{\|\theta\|_2 = 1} E|\theta^T \hat{\boldsymbol{x}}_i|^2 |\Delta(b_k)|/h + \sup_{\|\theta\|_2 = 1} E(\theta^T \hat{\boldsymbol{x}}_i)^4 \|\alpha - \boldsymbol{\beta}\|_2^2/h^2\Big) \\ \le & Ch\left(1 + \|\alpha - \boldsymbol{\beta}\|_2/h + |\Delta(b_k)|/h + \|\alpha - \boldsymbol{\beta}\|_2^2/h^2\right), \end{aligned}$$

where we use the inequalities that

$$\sup_{\|\theta\|_2=1} E(\theta^T \hat{\boldsymbol{x}}_i)^4 \leq 8 \sup_{\|\theta\|_2=1} E(\theta^T \boldsymbol{x}_i)^4 + 8 \sup_{\|\theta\|_2=1} (E \sum_{j=1}^{p} |\theta_j x_{ij}| I\{|x_{ij}| \geq N^\kappa\})^4$$

$$\leq 8 \sup_{\|\theta\|_2=1} E(\theta^T \boldsymbol{x}_i)^4 + 8p \sum_{j=1}^{p} E(x_{ij})^4 I\{|x_{ij}| \geq N^\kappa\} \leq C,$$

$$\sup_{\|v\|=1, \|u\|=1} E[(v^T \hat{\boldsymbol{x}}_i)^2 |u^T \hat{\boldsymbol{x}}_i|] \leq \sup_{\|\theta\|=1} E|\theta^T \hat{\boldsymbol{x}}_i|^3 \leq C,$$

$$\sup_{\|v\|=1, \|u\|=1} E[(v^T \hat{\boldsymbol{x}}_i)^2 (u^T \hat{\boldsymbol{x}}_i)^2] \leq \sup_{\|\theta\|=1} E|\theta^T \hat{\boldsymbol{x}}_i|^4 \leq C.$$

According to (2) and noting that $H(x)$, $H'(x)$ and $xH'(x)$ are bounded, we can obtain that

$$|v_j^T \hat{\boldsymbol{x}}_i \widehat{H}(\alpha)| \leq C \|\hat{\boldsymbol{x}}_i\|_2 (1 + \|\hat{\boldsymbol{x}}_i\|_2 \|\alpha - \boldsymbol{\beta}\|_2 / h) \leq C\sqrt{p} N^\kappa + CpN^{2\kappa} \|\alpha - \boldsymbol{\beta}\|_2 / h.$$

Under Condition (C3)*,

$$p \log(KN) = o\Big(\frac{\sqrt{Nph \log(kN)}}{pN^{2\kappa}}\Big).$$

Using Benstein's inequality, for any $\gamma > 0$, there exists a constant C such that

$$\sup_j \sup_l P\Big(|\widehat{\boldsymbol{C}}_j(\alpha_l) - E\widehat{\boldsymbol{C}}_j(\alpha_l)| \geq C\sqrt{\frac{ph \log(KN)}{N}}\Big) = O\left((kN)^{-\gamma p}\right).$$

This leads to

$$\sup_j \sup_l |\boldsymbol{C}_j(\alpha_l) - E\boldsymbol{C}_j(\alpha_l)| = O_P\Big(\sqrt{\frac{ph \log(KN)}{N}}\Big). \tag{6}$$

It remains to give a bound for $E\boldsymbol{C}_j(\alpha)$. Let $F_k(\cdot|\boldsymbol{x})$ be the conditional distribution of ϵ_k given $\boldsymbol{x}$. Denote $t = hx + \hat{\boldsymbol{x}}_i^T \Delta(\alpha) + \Delta(b_k)$ and we have

$$EH\Big(\frac{\epsilon_{ik} - \Delta(b_k) - \hat{\boldsymbol{x}}_i^T \Delta(\alpha)}{h}\Big|\boldsymbol{x}_i\Big) = \int_{-\infty}^{\infty} H\Big(\frac{t - \Delta(b_k) - \hat{\boldsymbol{x}}_i^T \Delta(\alpha)}{h}\Big) f_k(t|\boldsymbol{x}_i) dt$$

$$= \int_{-\infty}^{\infty} H(x) f_k(t|\boldsymbol{x}_i) dt = -\int_{-\infty}^{\infty} H(x) d(1 - F_k(t|\boldsymbol{x}_i))$$

$$= \int_{-\infty}^{\infty} (1 - F_k(t|\boldsymbol{x}_i)) H'(x) dx = 1 - \tau_k - f_k(0|\boldsymbol{x}_i) \int_{-\infty}^{\infty} (t) H'(x) dx + O(1) \int_{-\infty}^{\infty} (t)^2 \left|H'(x)\right| dx$$

$$= 1 - \tau_k - f_k(0|\boldsymbol{x}_i)\Big(\hat{\boldsymbol{x}}_i^T \Delta(\alpha) + \Delta(b_k) + h\int_{-\infty}^{\infty} xH'(x) dx\Big) + O\Big(h^2 + \Big(\hat{\boldsymbol{x}}_i^T \Delta(\alpha) + \Delta(b_k)\Big)^2\Big).$$

Similarly,

$$E\Big[\frac{\epsilon_{ik}-\Delta(b_k)}{h}H'\Big(\frac{\epsilon_{ik}-\Delta(b_k)-\hat{\boldsymbol{x}}_i^T\Delta(\alpha)}{h}\Big)\Big|\boldsymbol{x}_i\Big] = \int_{-\infty}^{\infty}(t-\Delta(b_k)/h)H'(x)f_k(t|\boldsymbol{x}_i)dx$$

$$= f_k(0|\boldsymbol{x}_i)\Big(\hat{\boldsymbol{x}}_i^T\Delta(\alpha)+h\int_{-\infty}^{\infty}xH'(x)dx\Big)+O(1)\int_{-\infty}^{\infty}(hx+\hat{\boldsymbol{x}}_i^T\Delta(\alpha)+\Delta(b_k))^2|H'(x)|dx$$

$$= f_k(0|\boldsymbol{x}_i)\Big(\hat{\boldsymbol{x}}_i^T\Delta(\alpha)+h\int_{-\infty}^{\infty}xH'(x)dx\Big)+O(h^2+(\hat{\boldsymbol{x}}_i^T\Delta(\alpha)+\Delta(b_k))^2).$$

These two equalities imply that uniformly in α and j,

$$|E\boldsymbol{C}_j(\alpha)| \le C(K_1h^2+K_2\|\Delta(\alpha)\|_2^2+K_3\Delta(b_1)^2).$$

Hence

$$\sup_i\sup_l|E\boldsymbol{C}_j(\alpha_l)| = O_P(h^2+a_n^2+b_n^2). \tag{7}$$

Combining that with (1), (3) and (6), this completes the proof of the proposition under Condition (C3)*.

We now prove the proposition under condition (C3). To bound $\boldsymbol{C}_j(\alpha_l)-E\boldsymbol{C}_j(\alpha_l)$ under condition (C3), we introduce the following exponential inequality from Cai and Liu (2011).

LEMMA 1 *Let $S_1,\ldots,S_N$ be independent random variables with mean zero. Suppose that there exist some $t>0$ and $\bar{B}_N$ such that $\sum_{k=1}^N ES_k^2e^{t|S_k|}\le \bar{B}_N^2$. Then for $0<x<\bar{B}_N$*

$$P\Big(\sum_{k=1}^N S_k > C_t\bar{B}_Nx\Big) \le \exp\left(-x^2\right),$$

where $C_t = t+t^{-1}$.

According to $a_n = O(h)$ and (2), denote $|\sum_{k=1}^K v_j^T\boldsymbol{x}_iH(\alpha)| \le C|v_j^T\boldsymbol{x}_i|(1+|(\alpha-\boldsymbol{\beta})^T\boldsymbol{x}_i|/\|\alpha-\boldsymbol{\beta}\|_2) := S_{ij}$, which implies that

$$E(|\sum_{k=1}^K v_j^T\boldsymbol{x}_iH(\alpha)|)^2e^{\eta_1|\sum_{k=1}^K v_j^T\boldsymbol{x}_iH(\alpha)|} \le E(|\sum_{k=1}^K v_j^T\boldsymbol{x}_iH(\alpha)|)^2e^{\eta_1S_{ij}} \le Ch.$$

Let $\bar{B}_N = C\sqrt{Nh}$ and $x=\sqrt{\gamma p\log(KN)}$. By Lemma 1 and the fact that $\sqrt{p\log(KN)} = o(\sqrt{Nh})$, for sufficiently large C, it can be shown that

$$\sup_j\sup_l P\Big(|\boldsymbol{C}_j(\alpha_l)-E\boldsymbol{C}_j(\alpha_l)| \ge C\sqrt{\frac{ph\log(KN)}{N}}\Big) = O\left((KN)^{-\gamma p}\right).$$

Combining that with (1), (3) and (7), we complete the proof of the proposition under condition (C3).

Proof of Proposition 2

Based on the proof of Lemma 3 in Cai et al. (2010), we have $\|\boldsymbol{Q}_N - \boldsymbol{Q}\| \leq 5\sup_{j\leq b_p}|v_j^T(\boldsymbol{Q}_N - \boldsymbol{Q})v_j|$, where $\{v_i, 1 \leq i \leq b_p\}$ are some non-random vectors with $\|v_i\|_2 = 1$ and $b_p \leq 5^p$. Now let

$$\boldsymbol{Q}_j(\alpha) = \frac{1}{Nh}\sum_{k=1}^{K}\sum_{i=1}^{N}(v_j^T\boldsymbol{x}_i\boldsymbol{x}_i^Tv_j)H'\Big(\frac{y_i - \tilde{b}_{\tau_k} - \boldsymbol{x}_i^T\alpha}{h}\Big) = \frac{1}{Nh}\sum_{k=1}^{K}\sum_{i=1}^{N}(v_j^T\boldsymbol{x}_i)^2H'\Big(\frac{y_i - \tilde{b}_{\tau_k} - \boldsymbol{x}_i^T\alpha}{h}\Big).$$

Therefore, when $\|\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}\|_2 \leq a_n$,

$$\|\boldsymbol{Q}_N - \boldsymbol{Q}\| \leq 5\sup_{j\leq b_p}\sup_{|\alpha-\beta|\leq a_n}|\boldsymbol{Q}_j(\alpha) - v_j^T\boldsymbol{Q}v_j|. \tag{8}$$

Because $H'(x)$ is Lipschitz continuous, we get

$$\Big|\frac{1}{h}H'\Big(\frac{y_i - \tilde{b}_{\tau_k} - \boldsymbol{x}_i^T\alpha}{h}\Big) - \frac{1}{h}H'\Big(\frac{y_i - \tilde{b}_{\tau_k} - \boldsymbol{x}_i^T\alpha_l}{h}\Big)\Big| \leq Ch^{-2}|\boldsymbol{x}_i^T(\alpha - \alpha_l)|.$$

Therefore,

$$\sup_j \sup_{\|\alpha-\beta\|_2\leq a_n}|\boldsymbol{Q}_j(\alpha) - v_j^T\boldsymbol{Q}v_j| - \sup_j\sup_l|\boldsymbol{Q}_j(\alpha_l) - v_j^T\boldsymbol{Q}v_j| \leq \frac{C\sqrt{p}Ka_n}{n^MNh^2}\sum_{i=1}^{N}\|\boldsymbol{x}_i\|_2^3.$$

Since $\max_{ij}\mathbb{E}|x_{ij}|^3 < \infty$, by letting M large enough, we have for any $\gamma > 0$,

$$\sup_j \sup_{\|\alpha-\beta\|_2\leq a_n}|\boldsymbol{Q}_j(\alpha) - v_j^T\boldsymbol{Q}v_j| - \sup_j\sup_l|\boldsymbol{Q}_j(\alpha_l) - v_j^T\boldsymbol{Q}v_j| = O_P(n^{-\gamma}). \tag{9}$$

We now prove the proposition under under (C3)*. Similarly to the proof of Proposition 1, denote $\hat{x}_{ij} = x_{ij}I\{|x_{ij}| \leq N^\kappa\}$ and

$$\widehat{\boldsymbol{Q}}_j(\alpha) = \frac{1}{Nh}\sum_{k=1}^{K}\sum_{i=1}^{N}(v_j^T\hat{\boldsymbol{x}}_i)^2H'\Big(\frac{\epsilon_{ik} - \Delta(b_k) - \hat{\boldsymbol{x}}_i^T\Delta(\alpha)}{h}\Big).$$

We can prove that

$$\begin{aligned} &P\Big(\sup_j\sup_l|\boldsymbol{Q}_j(\alpha_l) - v_j^T\boldsymbol{Q}v_j| \neq \sup_j\sup_l|\widehat{\boldsymbol{Q}}_j(\alpha_l) - v_j^T\boldsymbol{Q}v_j|\Big) \\ &\qquad \leq \sum_{i=1}^{N}\sum_{j=1}^{p}P(|x_{ij}| \geq N^\kappa) = o(1). \end{aligned} \tag{10}$$

As the proof of (5), we have

$$E\Big[H'\Big(\frac{\epsilon_{ik} - \Delta(b_k) - \hat{\boldsymbol{x}}_i^T\Delta(\alpha)}{h}\Big|\boldsymbol{x}_i\Big)\Big]^2 = O(h).$$

Using Bernstein's inequality, for any $\gamma > 0$, there exists a constant C such that

$$\sup_j \sup_l P\Big(|\widehat{\boldsymbol{Q}}_j(\alpha_l) - E(\widehat{\boldsymbol{Q}}_j(\alpha_l))| \geq C\sqrt{\frac{p\log(KN)}{Nh}}\Big) = O((KN)^{-\gamma p}). \tag{11}$$

Moreover,

$$\begin{aligned} E\Big[\frac{1}{h}H'\Big(\frac{\epsilon_{ik} - \Delta(b_k) - \hat{\boldsymbol{x}}_i^T\Delta(\alpha)}{h}\Big|\boldsymbol{x}_i\Big)\Big] &= \int_{-\infty}^{\infty} H'(x) f_k(hx + \hat{\boldsymbol{x}}_i^T\Delta(\alpha) + \Delta(b_k)|\boldsymbol{x}_i)dx \\ &= f_k(0|\boldsymbol{x}_i) + O(h + |\hat{\boldsymbol{x}}_i^T\Delta(\alpha) + \Delta(b_k)|). \end{aligned}$$

Therefore,

$$\begin{aligned} |E(\widehat{\boldsymbol{Q}}_j(\alpha)) - v_j^T\boldsymbol{Q}v_j| &\leq CK\max_i E|(v_j^T\boldsymbol{x}_i)^2 - (v_j^T\hat{\boldsymbol{x}}_i)^2| + CK(h + \|\alpha - \boldsymbol{\beta}\|_2 + \Delta(b_1)) \\ &\leq CK\max_i(\sum_{j=1}^{P} Ex_{ij}^2 I\{|x_{ij}| \geq N^\kappa\})^{1/2} + CK\max_i\sum_{j=1}^{p} Ex_{ij}^2 I\{|x_{ij}| \geq N^\kappa\} \qquad (12) \\ &\quad + CK(h + \|\alpha - \boldsymbol{\beta}\|_2 + \Delta(b_1)) \leq C_1(h + a_n + b_n). \end{aligned}$$

Combining (12) with (8), (9), (10) and (11), we can get the desired inequality under under (C3)*. We now prove the proposition under condition (C3). Let $S_{ij} = \sum_{k=1}^K (v_j^T\boldsymbol{x}_i)^2 H'\Big(\frac{\epsilon_{ik} - \boldsymbol{x}_i^T\Delta(\alpha) - \Delta(b_k)}{h}\Big)$ and we have

$$E(S_{ij})^2 e^{\eta_1|S_{ij}|} \leq E(S_{ij})^2 e^{C\eta_1(v_j^T\boldsymbol{x}_i)^2} = O(h).$$

Let $\bar{B}_N = C\sqrt{Nh}$ and $x = \sqrt{\gamma p\log(KN)}$. By Lemma 1 again, it leads to

$$\sup_j \sup_l P\Big(|\boldsymbol{Q}_j(\alpha_l) - E(\boldsymbol{Q}_j(\alpha_l))| \geq C\sqrt{\frac{p\log(KN)}{Nh}}\Big) = O((KN)^{-\gamma p}). \tag{13}$$

Combining (8), (9), (12) and (13), the desired inequality under under (C3) is obtained.

Proof of Theorems 1-3

For independent random vectors $\{\boldsymbol{x}_i, 1 \leq i \leq N\}$ with $\sup_{ij} E|x_{ij}|^3 = O(1)$, let

$$S_{NK} = \sum_{k=1}^K \sum_{i=1}^N \boldsymbol{x}_i(I\{\epsilon_{ik} \geq 0\} + \tau_k - 1).$$

Since $E\|S_{NK}\|^2 = O(Np)$, it leads to

$$\Big\|\frac{1}{N}S_{NK}\Big\|_2 = O_P\Big(\sqrt{\frac{p}{N}}\Big). \tag{14}$$

By (14) and Propositions 1-2, it is easy to show that the result holds. For $q = 1$, let $a_n = \sqrt{p/n}$, $b_n = \sqrt{1/n}$, since we assume that $\|\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}\|_2 = O_P(\sqrt{p/n})$ and $|\tilde{b}_{\tau_k} - b_{\tau_k}| = O_P(\sqrt{1/n})$ for $k = 1, \ldots, K$. Suppose the theorem holds for $q = g - 1$ with some $g \geq 2$. Noting that $p = O\left(n/(\log(KN))^2\right)$, we have

$$\sqrt{ph^{(g-1)}(\log(KN))/N} = O(\sqrt{p/N}).$$

Then for $q = g$ with initial estimator $(\tilde{b}_{\tau_1}, \ldots, \tilde{b}_{\tau_K}, \tilde{\boldsymbol{\beta}}) = (\hat{b}_{\tau_1}^{(g-1)}, \ldots, \hat{b}_{\tau_K}^{(g-1)}, \hat{\boldsymbol{\beta}}^{(g-1)})$, it can be seen that $a_n = O(h^{(g)})$ and $b_n = O(h^{(g)})$. Hence, we have proved Theorem 1. Theorem 2 follows directly from Theorem 1 by the Lindeberg-Feller central limit theorem. In addition, Theorem 2 is a special situation of Theorem 3, which can be confirmed similarly as Theorem 2.

References

Cai, T. and Liu, W. (2011). Adaptive thresholding for sparse covariance matrix estimation. *Journal of the American Statistical Association*, 106, 672–684.

Cai, T.T., Zhang, C.H. and Zhou, H.H. (2010). Optimal rates of convergence for covariance matrix estimation. *The Annals of Statistics*, 38, 2118–2144.