Marginal Quantile Regression for Varying Coefficient Models with Longitudinal Data

Weihua Zhao Weiping Zhang and Heng Lian

Supplementary Material: proofs and additional tables

Appendix A. Notations

For any real-valued function f on [0,1], let $\|f\|_\infty = \sup_{x\in[0,1]} |f(x)|$. Let $\|\cdot\|_2$ be the L_2 norm for functions and l_2 norm for vectors. For any matrix $\boldsymbol{A}$, $\|\boldsymbol{A}\|$ denotes the modulus of the largest singular value of $\boldsymbol{A}$. We use c, c_1, c_2 to denote generic positive constants which can take different values at different places. Note that by Condition (C3) and the result of De Boor (2001) (P. 149), there exists a constant c such that

$$\sup_{t\in[0,1]} |\alpha_l(t) - \boldsymbol{B}_l^T(t)\boldsymbol{\gamma}_l^0| \le ch^r, \quad l = 1, \ldots, p, \tag{S.1}$$

where $\boldsymbol{\gamma}_l^0$ can be viewed as the best approximation coefficient vector for $\alpha_l(t)$.

Let $\boldsymbol{\gamma}_0 = (\boldsymbol{\gamma}_1^{0T}, \ldots, \boldsymbol{\gamma}_p^{0T})$, $\boldsymbol{\mu}_i^0 = (\mu_{i1}^0, \ldots, \mu_{im_i}^0)^T$ with $\mu_{ij}^0 = \boldsymbol{X}_{ij}^T\boldsymbol{\alpha}(t_{ij})$ and $\boldsymbol{\epsilon}_i = \boldsymbol{y}_i - \boldsymbol{\mu}_i^0$. Denote $\boldsymbol{S}_k(\boldsymbol{\gamma}) = \frac{1}{n}\sum_{i=1}^n \boldsymbol{S}_{ik}(\boldsymbol{\gamma})$, where $\boldsymbol{S}_{ik}(\boldsymbol{\gamma}) = \boldsymbol{U}_i^T\boldsymbol{\Gamma}_i\boldsymbol{M}_{ki}\psi_\tau(\boldsymbol{y}_i - \boldsymbol{U}_i\boldsymbol{\gamma})$.

Similarly write $\boldsymbol{S}^0(\boldsymbol{\gamma}) = (\boldsymbol{S}_1^{0T}(\boldsymbol{\gamma}), \ldots, \boldsymbol{S}_v^{0T}(\boldsymbol{\gamma}))^T$ with $\boldsymbol{S}_k^0(\boldsymbol{\gamma}) = \frac{1}{n}\sum_{i=1}^n \boldsymbol{S}_{ik}^0(\boldsymbol{\gamma})$, where

$$\boldsymbol{S}_{ik}^0(\boldsymbol{\gamma}) = \boldsymbol{S}_{ik1}^0 + \boldsymbol{S}_{ik2}^0 + \boldsymbol{S}_{ik3}^0(\boldsymbol{\gamma}), \quad \boldsymbol{S}_{ik1}^0 = \boldsymbol{U}_i^T\boldsymbol{\Gamma}_i\boldsymbol{M}_{ki}\psi_\tau(\boldsymbol{\epsilon}_i),$$

$$\boldsymbol{S}_{ik2}^0 = \boldsymbol{U}_i^T\boldsymbol{\Gamma}_i\boldsymbol{M}_{ki}\boldsymbol{\Gamma}_i(\boldsymbol{\mu}_i^0 - \boldsymbol{U}_i\boldsymbol{\gamma}_0), \quad \boldsymbol{S}_{ik3}^0(\boldsymbol{\gamma}) = \boldsymbol{U}_i^T\boldsymbol{\Gamma}_i\boldsymbol{M}_{ki}\boldsymbol{\Gamma}_i(\boldsymbol{U}_i\boldsymbol{\gamma}_0 - \boldsymbol{U}_i\boldsymbol{\gamma}).$$

Then we denote

$$\boldsymbol{S}_{k1}^0 = \frac{1}{n}\sum_{i=1}^n \boldsymbol{S}_{ik1}^0, \quad \boldsymbol{S}_{k2}^0 = \frac{1}{n}\sum_{i=1}^n \boldsymbol{S}_{ik2}^0 \text{ and } \boldsymbol{S}_{k3}^0(\boldsymbol{\gamma}) = \frac{1}{n}\sum_{i=1}^n \boldsymbol{S}_{ik3}^0(\boldsymbol{\gamma}).$$

Define the set $\Theta_n(C) = \{\boldsymbol{\gamma} : \|\boldsymbol{B}^T(\boldsymbol{\gamma} - \boldsymbol{\gamma}_0)\|_2 = C/\sqrt{nh}\}$ for a sufficiently large constant C, where $\boldsymbol{B}(t) = (\boldsymbol{B}_1^T(t), \ldots, \boldsymbol{B}_p^T(t))^T$ and

$$\|\boldsymbol{B}^T(\boldsymbol{\gamma} - \boldsymbol{\gamma}_0)\|_2 = \left(\int [\sum_{l=1}^{p} \sum_{k=1}^{K_l} B_{lk}(t)(\gamma_{lk} - \gamma_{lk}^0)]^2 dt\right)^{1/2}.$$

Let

$$\boldsymbol{\Sigma}_{kk'}(\boldsymbol{\gamma}) = \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{S}_{ik}(\boldsymbol{\gamma}) \boldsymbol{S}_{ik'}^T(\boldsymbol{\gamma}) \quad \text{and} \quad \boldsymbol{\Sigma}_{kk'}^0 = \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{S}_{ik1}^0 \boldsymbol{S}_{ik'1}^{0T}$$

be the (k, k')-block of

$$\boldsymbol{\Sigma}_n(\boldsymbol{\gamma}) = \{\boldsymbol{\Sigma}_{kk'}(\boldsymbol{\gamma})\}_{k,k'=1}^{v} \quad \text{and} \quad \boldsymbol{\Sigma}_n^0 = \{\boldsymbol{\Sigma}_{kk'}^0\}_{k,k'=1}^{v}, \tag{S.2}$$

respectively. Finally, we define

$$\boldsymbol{Q}_n^0(\boldsymbol{\gamma}) = n(\boldsymbol{S}^0(\boldsymbol{\gamma}))^T \{\boldsymbol{\Sigma}_n^0\}^{-1} \boldsymbol{S}^0(\boldsymbol{\gamma}). \tag{S.3}$$

Appendix B. Lemmas

The following lemmas are needed in preparation for the proof of the theorems.

Lemma 1. Assume Conditions (C1), (C3), (C5) and (C6) hold. There exist two constants $0 < c_1 \leq c_2$, such that, except in an event whose probability tends to zero as $n \to \infty$,

$$c_1\|\boldsymbol{\gamma}\|_2^2 \leq \|\boldsymbol{B}^T\boldsymbol{\gamma}\|_2^2 \leq c_2\|\boldsymbol{\gamma}\|_2^2,$$

for any vector $\boldsymbol{\gamma}$ with length $\sum_{l=1}^{p} J_l$, where $\|\boldsymbol{B}^T\boldsymbol{\gamma}\|_2^2 = \int(\sum_{l=1}^{p}\sum_{k=1}^{K_l} B_{lk}(t)\gamma_{lk})^2 dt$. Furthermore, there also exist two constants $0 < c_1' \leq c_2'$, such that, except in an event whose probability tends to zero as $n \to \infty$,

$$c_1'\|\boldsymbol{\gamma}\|_2^2 \leq \|\boldsymbol{B}^T\boldsymbol{\gamma}\|_n^2 \leq c_2'\|\boldsymbol{\gamma}\|_2^2,$$

where $\|\boldsymbol{B}^T\boldsymbol{\gamma}\|_n^2 = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{m_i}\sum_{j=1}^{m_i}(\boldsymbol{B}^T(t_{ij})\boldsymbol{\gamma})^2$.

Proof. The results can be proved along the lines of Lemmas 3 and 6 in Xue et al. (2010). □

Lemma 2. Assume that Conditions (C1)-(C7) hold, then one has

$$\|\boldsymbol{S}_{k1}^0\|_2 = O_p(1/\sqrt{n}),\ \|\boldsymbol{S}_{k2}^0\|_2 = O_p(h^r).$$

Furthermore, there exist two constants $0 < c_1 \le c_2$ such that for any $\boldsymbol{\gamma} \in \Theta_n(C)$,

$$c_1 C/\sqrt{nh} \le \|\boldsymbol{S}_{k3}^0(\boldsymbol{\gamma})\|_2 \le c_2 C/\sqrt{nh},$$

except in an event whose probability goes to 0 as $n \to \infty$.

Proof. By (C2), (C4) and (C7), we have

$$\mathrm{E}(\boldsymbol{S}_{k1}^0) = 0 \ \text{ and } \ \mathrm{tr}\left(\mathrm{E}(\boldsymbol{S}_{ik1}^0 \boldsymbol{S}_{ik1}^{0T})\right) = \mathrm{tr}\left(\mathrm{E}\boldsymbol{U}_i^T \boldsymbol{\Gamma}_i \boldsymbol{M}_{ki} \psi_\tau(\boldsymbol{\epsilon}_i) \psi_\tau^T(\boldsymbol{\epsilon}_i) \boldsymbol{M}_{ki}^T \boldsymbol{\Gamma}_i \boldsymbol{U}_i)\right) = O(1).$$

Therefore, we obtain $E\|\boldsymbol{S}_{k1}^0\|_2^2 = \mathrm{tr}(E(\boldsymbol{S}_{ik1}^0 \boldsymbol{S}_{ik1}^{0T}))/n = O(1/n)$ which implies $\|\boldsymbol{S}_{k1}^0\|_2 = O_p(1/\sqrt{n})$.

For $\boldsymbol{S}_{k2}^0$, by Lemma 1, equation (S.1) and the Conditions (C4) and (C7), applying the Cauchy-Schwartz inequality, we have

$$\begin{aligned}
|\boldsymbol{\nu}^T \boldsymbol{S}_{k2}^0| &= \left|\tfrac{1}{n}\textstyle\sum_{i=1}^n \boldsymbol{\nu}^T \boldsymbol{U}_i^T \boldsymbol{\Gamma}_i \boldsymbol{M}_{ki} \boldsymbol{\Gamma}_i (\boldsymbol{\mu}_i^0 - \boldsymbol{U}_i \boldsymbol{\gamma}_0)\right| \\
&\le \left(\tfrac{1}{n}\textstyle\sum_i \boldsymbol{\nu}^T \boldsymbol{U}_i^T \boldsymbol{\Gamma}_i \boldsymbol{M}_{ki} \boldsymbol{M}_{ki}^T \boldsymbol{\Gamma}_i \boldsymbol{U}_i \boldsymbol{\nu}\right)^{1/2} \left(\tfrac{1}{n}\textstyle\sum_i (\boldsymbol{\mu}_i^0 - \boldsymbol{U}_i \boldsymbol{\gamma}_0)^T \boldsymbol{\Gamma}_i \boldsymbol{\Gamma}_i^T (\boldsymbol{\mu}_i^0 - \boldsymbol{U}_i \boldsymbol{\gamma}_0)\right)^{1/2} \\
&\le c\left(\boldsymbol{\nu}^T \boldsymbol{\nu}\right)^{1/2} \|\boldsymbol{B}^T(t)\boldsymbol{\gamma}_0 - \boldsymbol{\alpha}(t)\|_\infty \\
&\le ch^r.
\end{aligned}$$

It follows that $\|\boldsymbol{S}_{k2}^0\|_2 = O_p(h^r)$.

Similarly, by the Cauchy-Schwartz inequality, on the set $\Theta_n(C)$, we can prove that there exist a constant c_2 such that $\|\boldsymbol{S}_{k3}^0(\boldsymbol{\gamma})\|_2 \le c_2 C/\sqrt{nh}$.

On the other hand, taking $\boldsymbol{\nu} = (\boldsymbol{\gamma} - \boldsymbol{\gamma}_0)/\|\boldsymbol{\gamma} - \boldsymbol{\gamma}_0\|_2$, Conditions (C4)-(C7) and Lemma 1 entail that there exists a positive constant c_1 such that $|\boldsymbol{\nu}^T \boldsymbol{S}_{k3}^0(\boldsymbol{\gamma})| \ge c_1 C/\sqrt{nh}$. Therefore, we obtain $\|\boldsymbol{S}_{k3}^0(\boldsymbol{\gamma})\|_2 \ge c_1 C/\sqrt{nh}$. This completes the proof. □

Lemma 3. Assume that Conditions (C1)-(C7) hold, the eigenvalues of $\boldsymbol{\Sigma}_n^0$ are bounded from 0 and infinity with probability approaching 1. And, on the set $\Theta_n(C)$, we have $\|\boldsymbol{S}^0(\boldsymbol{\gamma})\|_2 = O_p(1/\sqrt{nh})$ and $\boldsymbol{Q}_n^0(\boldsymbol{\gamma}) = O_p(1/h)$.

Proof. By Lemma 1 and the definition of $\boldsymbol{\Sigma}_n^0$ in (S.2), we can easily conclude that $\boldsymbol{\Sigma}_n^0$ has eigenvalues bounded away from 0 and infinity. In addition, $\boldsymbol{S}^0(\boldsymbol{\gamma}) = O_p(1/\sqrt{nh})$ follows directly from Lemma 2, and $\boldsymbol{Q}_n^0(\boldsymbol{\gamma}) = O_p(1/h)$ follows from the above two results. □

Lemma 4. Assume that Conditions (C1)-(C7) hold. On the set $\Theta_n(C)$ for some C sufficiently large, we have

$$\sup_{\boldsymbol{\gamma}\in\Theta_n(C)} \|\boldsymbol{S}(\boldsymbol{\gamma}) - \boldsymbol{S}^0(\boldsymbol{\gamma})\|_2 = o_p(1/\sqrt{nh}), \tag{S.4}$$

$$\sup_{\boldsymbol{\gamma}\in\Theta_n(C)} |\boldsymbol{Q}_n(\boldsymbol{\gamma}) - \boldsymbol{Q}_n^0(\boldsymbol{\gamma})| = o_p(1/h). \tag{S.5}$$

Proof. We only need to consider the components $\boldsymbol{S}_k(\boldsymbol{\gamma}) - \boldsymbol{S}_k^0(\boldsymbol{\gamma})$ of $\boldsymbol{S}(\boldsymbol{\gamma}) - \boldsymbol{S}^0(\boldsymbol{\gamma})$, $k = 1, \ldots, \upsilon$.

For any $\boldsymbol{\nu} \in \mathbb{R}^{\sum_{l=1}^p K_l}$ with $\boldsymbol{\nu}^T\boldsymbol{\nu} = 1$, we write

$$\boldsymbol{\nu}^T(\boldsymbol{S}_k(\boldsymbol{\gamma}) - \boldsymbol{S}_k^0(\boldsymbol{\gamma})) = \boldsymbol{\nu}^T\boldsymbol{\zeta} + \eta, \tag{S.6}$$

where

$$\boldsymbol{\zeta} = \frac{1}{n}\sum_{i=1}^n \boldsymbol{U}_i^T\boldsymbol{\Gamma}_i\boldsymbol{M}_{ki}\left\{\psi_\tau(\boldsymbol{y}_i - \boldsymbol{U}_i\boldsymbol{\gamma}) - \psi_\tau(\boldsymbol{\epsilon}_i) + [F_i(\boldsymbol{U}_i\boldsymbol{\gamma} - \boldsymbol{\mu}_i^0) - F_i(0)]\right\},$$

$$\eta = \frac{1}{n}\sum_{i=1}^n \boldsymbol{\nu}^T\boldsymbol{U}_i^T\boldsymbol{\Gamma}_i\boldsymbol{M}_{ki}\left\{-\boldsymbol{\Gamma}_i(\boldsymbol{\mu}_i^0 - \boldsymbol{U}_i\boldsymbol{\gamma}) - [F_i(\boldsymbol{U}_i\boldsymbol{\gamma} - \boldsymbol{\mu}_i^0) - F_i(0)]\right\},$$

and $F_i(\boldsymbol{U}_i\boldsymbol{\gamma} - \boldsymbol{\mu}_i^0) = (F_{i1}(\boldsymbol{U}_{i1}^T\boldsymbol{\gamma} - \mu_{i1}^0), \ldots, F_{im_i}(\boldsymbol{U}_{im_i}^T\boldsymbol{\gamma} - \mu_{in_i}^0))^T$ with $F_{ij}(\cdot)$ being the conditional cdf of ϵ_{ij}.

Since $|\mu_{ij}^0 - \boldsymbol{U}_{ij}^T\boldsymbol{\gamma}| = O_p(1/\sqrt{nh})$ on the set $\Theta_n(C)$, by the Taylor expansion, we get

$$F_i(\boldsymbol{U}_i\boldsymbol{\gamma} - \boldsymbol{\mu}_i^0) - F_i(0) = -\boldsymbol{\Gamma}_i(\boldsymbol{\mu}_i^0 - \boldsymbol{U}_i\boldsymbol{\gamma}) + O_p(1/(nh)). \tag{S.7}$$

By Lemma 6.2 of He and Shi (1996) together with Condition (C4), we have

$$\sup_{1\le i\le n} \|\boldsymbol{U}_i\| = O_p(h^{-1/2}). \tag{S.8}$$

Applying Lemma 19.24 of Van der Vaart (2000), we have

$$\sup_{\boldsymbol{\gamma}\in\Theta_n(C)} \left|\frac{1}{n}\sum_{i=1}^n \left\{\psi_\tau(\boldsymbol{y}_i - \boldsymbol{U}_i\boldsymbol{\gamma}) - \psi_\tau(\boldsymbol{\epsilon}_i) + [F_i(\boldsymbol{U}_i\boldsymbol{\gamma} - \boldsymbol{\mu}_i^0) - F_i(0)]\right\}\right| = o_p(n^{-1/2}). \tag{S.9}$$

Therefore, by Conditions (C2) and (C7) together with (S.8), we obtain

$$\sup_{\boldsymbol{\nu}^T\boldsymbol{\nu}=1, \boldsymbol{\gamma}\in\Theta_n(C)} \boldsymbol{\nu}^T\boldsymbol{\zeta} = O_p(h^{-1/2}) \times o_p(n^{-1/2}) = o_p(1/\sqrt{nh}). \tag{S.10}$$

On the other hand, for the term η in (S.6), by (S.7) and (S.8), we get

$$\sup_{\boldsymbol{\nu}^T\boldsymbol{\nu}=1,\gamma\in\Theta_n(C)} |\eta| = O_p(h^{-1/2}) \times O_p(1/(nh)) = o_p(1/\sqrt{nh}). \tag{S.11}$$

Hence, combining (S.6)-(S.11), we obtain

$$\sup_{\boldsymbol{\nu}^T\boldsymbol{\nu}=1,\gamma\in\Theta_n(C)} \boldsymbol{\nu}^T(\boldsymbol{S}_k(\boldsymbol{\gamma}) - \boldsymbol{S}_k^0(\boldsymbol{\gamma})) = o_p(1/\sqrt{nh}),$$

and it follows that (S.4) holds.

Next, we prove (S.5). Notice that

$$\begin{aligned} \|\boldsymbol{\Sigma}_n^{-1}(\boldsymbol{\gamma}) - \{\boldsymbol{\Sigma}_n^0\}^{-1}\| &= \|\boldsymbol{\Sigma}_n^{-1}(\boldsymbol{\gamma})[\boldsymbol{\Sigma}_n(\boldsymbol{\gamma}) - \boldsymbol{\Sigma}_n^0]\{\boldsymbol{\Sigma}_n^0\}^{-1}\| \\ &\leq \|\boldsymbol{\Sigma}_n^{-1}(\boldsymbol{\gamma})\|\|\boldsymbol{\Sigma}_n(\boldsymbol{\gamma}) - \boldsymbol{\Sigma}_n^0\|\|\{\boldsymbol{\Sigma}_n^0\}^{-1}\|. \end{aligned}$$

Consider the (k,k')-block of $\{\boldsymbol{\Sigma}_n(\boldsymbol{\gamma}) - \boldsymbol{\Sigma}_n^0\}$, by the definitions of $\boldsymbol{\Sigma}_{kk'}(\boldsymbol{\gamma})$ and $\boldsymbol{\Sigma}_{kk'}^0$, as in the proof of the first part of the lemma, by some careful calculations, we can show that

$$\|\boldsymbol{\Sigma}_{kk'}(\boldsymbol{\gamma}) - \boldsymbol{\Sigma}_{kk'}^0\| = o_p(1).$$

Therefore, we have

$$\|\boldsymbol{\Sigma}_n(\boldsymbol{\gamma}) - \boldsymbol{\Sigma}_n^0\| = o_p(1).$$

Based on Lemma 3, the eigenvalues of $\boldsymbol{\Sigma}_n^0$ are bounded from 0 and infinity, therefore we can conclude that the eigenvalues of $\boldsymbol{\Sigma}_n(\boldsymbol{\gamma})$ are also bounded from 0 and infinity, which leads to

$$\|\boldsymbol{\Sigma}_n^{-1}(\boldsymbol{\gamma}) - \{\boldsymbol{\Sigma}_n^0\}^{-1}\| = o_p(1). \tag{S.12}$$

Finally, by the definitions of $\boldsymbol{Q}_n(\boldsymbol{\gamma})$ and $\boldsymbol{Q}_n^0(\boldsymbol{\gamma})$, (S.5) holds by (S.4), (S.12) and Lemma 2. □

Lemma 5. For any small positive ε, there exist constants ϱ and C such that

$$P\left\{\inf_{\gamma\in\Theta_n(C)} \boldsymbol{Q}_n^0(\boldsymbol{\gamma}) > \boldsymbol{Q}_n^0(\boldsymbol{\gamma}_0) + \frac{\varrho}{h}\right\} > 1 - \varepsilon,$$

for n large enough.

Proof. Let $\boldsymbol{\Delta}(\boldsymbol{\gamma}) = \boldsymbol{S}^0(\boldsymbol{\gamma}) - \boldsymbol{S}^0(\boldsymbol{\gamma}_0)$. Then we can write

$$\frac{1}{n}\left\{\boldsymbol{Q}_n^0(\boldsymbol{\gamma}) - \boldsymbol{Q}_n^0(\boldsymbol{\gamma}_0)\right\} = \boldsymbol{\Delta}^T(\boldsymbol{\gamma})(\boldsymbol{\Sigma}_n^0)^{-1}\boldsymbol{\Delta}(\boldsymbol{\gamma}) + 2\boldsymbol{\Delta}^T(\boldsymbol{\gamma})(\boldsymbol{\Sigma}_n^0)^{-1}\boldsymbol{S}^0(\boldsymbol{\gamma}_0).$$

According to the definitions of $\boldsymbol{S}^0(\boldsymbol{\gamma})$ and $\boldsymbol{S}^0(\boldsymbol{\gamma}_0)$, one has $\boldsymbol{\Delta}(\boldsymbol{\gamma}) = \boldsymbol{S}_{k3}^0(\boldsymbol{\gamma})$. By Lemma 3, the eigenvalues of $\{\boldsymbol{\Sigma}_n^0\}^{-1}$ are bounded, hence on the set $\Theta_n(C)$

$$\begin{aligned} |\boldsymbol{\Delta}^T(\boldsymbol{\gamma})(\boldsymbol{\Sigma}_n^0)^{-1}\boldsymbol{S}^0(\boldsymbol{\gamma}_0)| &= O_p\left(\textstyle\sum_{k=1}^{v} \|\boldsymbol{S}_{k3}^0(\boldsymbol{\gamma})\|_2 \left(\|\boldsymbol{S}_{k1}^0\|_2 + \|\boldsymbol{S}_{k2}^0\|_2\right)\right) \\ &= O_p(C/(nh)), \end{aligned}$$

by Lemma 2.

On the other hand,

$$\boldsymbol{\Delta}^T(\boldsymbol{\gamma})(\boldsymbol{\Sigma}_n^0)^{-1}\boldsymbol{\Delta}(\boldsymbol{\gamma}) \asymp \sum_{k=1}^{v} \left\|\boldsymbol{S}_{k3}^0(\boldsymbol{\gamma})\right\|_2^2 \asymp \frac{C^2}{nh},$$

again by Lemma 2. Choosing C sufficiently large, we can obtain the desired result. This completes the proof. □

Lemma 6. For any $\varepsilon > 0$, there exists C such that

$$P\left\{\inf_{\boldsymbol{\gamma}\in\Theta_n(C)} \boldsymbol{Q}_n(\boldsymbol{\gamma}) > \boldsymbol{Q}_n(\boldsymbol{\gamma}_0)\right\} > 1 - \varepsilon.$$

Proof. Note that we can write

$$\boldsymbol{Q}_n(\boldsymbol{\gamma}) - \boldsymbol{Q}_n(\boldsymbol{\gamma}_0) = \boldsymbol{Q}_n(\boldsymbol{\gamma}) - \boldsymbol{Q}_n^0(\boldsymbol{\gamma}) + \boldsymbol{Q}_n^0(\boldsymbol{\gamma}) - \boldsymbol{Q}_n^0(\boldsymbol{\gamma}_0) + \boldsymbol{Q}_n^0(\boldsymbol{\gamma}_0) - \boldsymbol{Q}_n(\boldsymbol{\gamma}_0).$$

The result follows from

$$\sup_{\boldsymbol{\gamma}\in\Theta_n(C)} |\boldsymbol{Q}_n(\boldsymbol{\gamma}) - \boldsymbol{Q}_n^0(\boldsymbol{\gamma})| = o_p(1/h)$$

and

$$|\boldsymbol{Q}_n(\boldsymbol{\gamma}_0) - \boldsymbol{Q}_n^0(\boldsymbol{\gamma}_0)| = o_p(1/h),$$

together with Lemma 5. □

Appendix C. Proofs of Theorems

Proof of Theorem 1

Proof. The theorem follows directly from Lemma 1 and Lemma 6. □

Proof of Theorem 2

Proof. Consider

$$\tilde{\boldsymbol{S}}_k(\boldsymbol{\gamma}) = \mathrm{E}\left\{\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{S}_{ik}(\boldsymbol{\gamma} + (nh)^{-1/2}\boldsymbol{\Omega}^{1/2}\boldsymbol{\delta})\right\}.$$

Let $\boldsymbol{\theta} = \sqrt{nh}(\boldsymbol{\gamma} - \boldsymbol{\gamma}_0)$. By the Taylor expansion, for all $\|\boldsymbol{\theta}\| \leq C$ with some finite constant C, we get

$$\sqrt{nh}\tilde{\boldsymbol{S}}_k(\boldsymbol{\gamma}_0 + \boldsymbol{\theta}/\sqrt{nh}) = \sqrt{nh}\tilde{\boldsymbol{S}}_k(\boldsymbol{\gamma}_0) + \tilde{\boldsymbol{S}}'_k(\boldsymbol{\gamma}_0)\boldsymbol{\theta} + o_p(1).$$

By (8) in Section 2.3, we get $\tilde{\boldsymbol{S}}'_k(\boldsymbol{\gamma}_0) = -\frac{1}{n}\sum_i \boldsymbol{U}_i^T\boldsymbol{\Gamma}_i\boldsymbol{M}_{ki}\boldsymbol{\Xi}_i\boldsymbol{U}_i$, where $\boldsymbol{\Xi}_i$ is an $m_i \times m_i$ diagonal matrix with elements $(\boldsymbol{\Xi}_i)_{jj} = \phi\left(\sqrt{nh}\frac{y_{ij} - \boldsymbol{U}_{ij}^T\boldsymbol{\gamma}_0}{r_{ij}}\right)\frac{\sqrt{nh}}{r_{ij}}$. Denote $a_{ij} = \boldsymbol{U}_{ij}^T\boldsymbol{\gamma}_0 - \mu_{ij}^0$, then $(\boldsymbol{\Xi}_i)_{jj} = \phi\left(\sqrt{nh}\frac{\epsilon_{ij} - a_{ij}}{r_{ij}}\right)\frac{\sqrt{nh}}{r_{ij}}$. Furthermore, through calculating the expection and variance of the each element of $\tilde{\boldsymbol{S}}'_k(\boldsymbol{\gamma}_0)$, we obtain

$$\tilde{\boldsymbol{S}}'_k(\boldsymbol{\gamma}_0) = -\frac{1}{n}\sum_i \boldsymbol{U}_i^T\boldsymbol{\Gamma}_i\boldsymbol{M}_{ki}\boldsymbol{\Lambda}_i\boldsymbol{U}_i + o_p(1), \tag{S.13}$$

where $\boldsymbol{\Lambda}_i$ is an $m_i \times m_i$ diagonal matrix with elements $\mathrm{E}_{\epsilon_{ij}}\phi\left(\sqrt{nh}\frac{\epsilon_{ij} - a_{ij}}{r_{ij}}\right)\frac{\sqrt{nh}}{r_{ij}}$.

Note that

$$\begin{aligned}
\mathrm{E}_{\epsilon_{ij}}\phi\left(\sqrt{nh}\tfrac{\epsilon_{ij} - a_{ij}}{r_{ij}}\right)\tfrac{\sqrt{nh}}{r_{ij}} &= \int \tfrac{\sqrt{nh}}{r_{ij}}\phi\left(\sqrt{nh}\tfrac{t - a_{ij}}{r_{ij}}\right) f_{ij}(t)dt \\
&= \int \phi\left(t\right) f_{ij}\left(\tfrac{r_{ij}}{\sqrt{nh}}t + a_{ij}\right) dt \\
&= \int \phi\left(t\right) f_{ij}(0)dt + \tfrac{r_{ij}}{\sqrt{nh}}\int \phi(t) f'_{ij}(o^*)\left(t + \tfrac{\sqrt{nh}}{r_{ij}}a_{ij}\right) dt,
\end{aligned}$$

where o^* lies between 0 and $\frac{r_{ij}}{\sqrt{nh}}t + a_{ij}$. Since $a_{ij} = O_p(h^r)$ and $r_{ij} = O_p(\|\boldsymbol{U}_{ij}\|) = O_p(h^{-1/2})$, by Condition (C2),

$$\left|\frac{r_{ij}}{\sqrt{nh}}\int \phi(t) f'_{ij}(o^*)\left(t + \frac{\sqrt{nh}}{r_{ij}}a_{ij}\right) dt\right| = O_p\left(\frac{r_{ij}}{\sqrt{nh}}\int |t|\phi(t)dt + a_{ij}\right) \longrightarrow 0$$

as $n \to \infty$. Therefore, we obtain $\mathrm{E}_{\epsilon_{ij}}\phi\left(\sqrt{nh}\frac{\epsilon_{ij}-a_{ij}}{r_{ij}}\right)\frac{\sqrt{nh}}{r_{ij}} = f_{ij}(0) + o_p(1)$.

Denote $\boldsymbol{G}_k = \frac{1}{n}\sum_i \boldsymbol{U}_i^T \boldsymbol{\Gamma}_i \boldsymbol{M}_{ki} \boldsymbol{\Gamma}_i \boldsymbol{U}_i$ and $\boldsymbol{G} = (\boldsymbol{G}_1^T, \ldots, \boldsymbol{G}_v^T)^T$. Then, we have

$$\sqrt{nh}\tilde{\boldsymbol{S}}(\boldsymbol{\gamma}_0 + \boldsymbol{\theta}/\sqrt{nh}) = \sqrt{nh}\tilde{\boldsymbol{S}}(\boldsymbol{\gamma}_0) - \boldsymbol{G}\boldsymbol{\theta} + o_p(1). \tag{S.14}$$

On the other hand, it is not difficult to see from the previous proofs for the unsmoothed case that

$$\sqrt{nh}\boldsymbol{S}(\boldsymbol{\gamma}_0 + \boldsymbol{\theta}/\sqrt{nh}) = \sqrt{nh}\boldsymbol{S}(\boldsymbol{\gamma}_0) - \boldsymbol{G}\boldsymbol{\theta} + o_p(1). \tag{S.15}$$

Therefore, if the following equation

$$\sqrt{nh}\|\tilde{\boldsymbol{S}}(\boldsymbol{\gamma}_0) - \boldsymbol{S}(\boldsymbol{\gamma}_0)\|_2 \to 0 \tag{S.16}$$

holds in probability, then by (S.14) and (S.15), we have

$$\sup_{\|\boldsymbol{\gamma}-\boldsymbol{\gamma}_0\|_2 \le C/\sqrt{nh}} \|\tilde{\boldsymbol{S}}(\boldsymbol{\gamma}) - \boldsymbol{S}(\boldsymbol{\gamma})\|_2 = o_p(1/\sqrt{nh})$$

for given C. And, by the similar proof line of Lemma 4, we can get

$$\sup_{\|\boldsymbol{\gamma}-\boldsymbol{\gamma}_0\|_2 \le C/\sqrt{nh}} \|\tilde{\boldsymbol{Q}}(\boldsymbol{\gamma}) - \boldsymbol{Q}(\boldsymbol{\gamma})\|_2 = o_p(1/h).$$

Similar to the proof of Lemma 5, Lemma 6 and Theorem 1, we can obtain the desired result.

The proof of (S.16) can be done directly by calculating its expectation and variance and we omit it here. □

Proof of Theorem 3

Proof. Let $L_n(\boldsymbol{\gamma}) = \boldsymbol{Q}_n(\boldsymbol{\gamma}) + n\sum_{l=1}^p p_{\lambda_l}(\|\boldsymbol{\gamma}_l\|_{\boldsymbol{H}_l})$. We next prove that for large n and any $\varepsilon > 0$, there exists a constant C large enough such that

$$P\left(\inf_{\boldsymbol{\gamma}:\|\boldsymbol{B}^T(\boldsymbol{\gamma}-\boldsymbol{\gamma}^0)\|_2 = C/\sqrt{nh}} L_n(\boldsymbol{\gamma}) > L_n(\boldsymbol{\gamma}_0)\right) \ge 1 - \varepsilon. \tag{S.17}$$

As a result, (S.17) implies that $L_n(\cdot)$ has a local minimum in the ball $\{\boldsymbol{\gamma} : \|\boldsymbol{B}^T(\boldsymbol{\gamma} - \boldsymbol{\gamma}_0)\|_2 = C/\sqrt{nh}\}$, which will imply the desired result.

To prove (S.17), using $p_{\lambda_l}(0) = 0$ and $p_{\lambda_l}(0) \geq 0$, we have

$$L_n(\boldsymbol{\gamma}) - L_n(\boldsymbol{\gamma}_0) \geq \boldsymbol{Q}_n(\boldsymbol{\gamma}) - \boldsymbol{Q}_n(\boldsymbol{\gamma}_0) + n\sum_{l=1}^{d_0}\{p_{\lambda_l}(\|\boldsymbol{\gamma}_l\|_{\boldsymbol{H}_l}) - p_{\lambda_l}(\|\boldsymbol{\gamma}_l^0\|_{\boldsymbol{H}_l})\}. \tag{S.18}$$

It is easy to see that for any $\boldsymbol{\gamma}$ satisfying $\|\boldsymbol{B}^T(\boldsymbol{\gamma}-\boldsymbol{\gamma}_0)\|_2 = O_p(1/\sqrt{nh})$, both $\|\boldsymbol{\gamma}_l\|_{\boldsymbol{H}_l} \geq a\lambda_l$ and $\|\boldsymbol{\gamma}_l^0\|_{\boldsymbol{H}_l} \geq a\lambda_l$ hold when n is sufficiently large if $\lambda_l \to 0$. Hence, by the property of the SCAD penalty function, we have

$$\sum_{l=1}^{d_0}\{p_{\lambda_l}(\|\boldsymbol{\gamma}_l\|_{\boldsymbol{H}_l}) - p_{\lambda_l}(\|\boldsymbol{\gamma}_l^0\|_{\boldsymbol{H}_l})\} = 0 \tag{S.19}$$

when n is sufficiently large.

By Lemma 6, on the set $\Theta = \{\boldsymbol{\gamma} : \|\boldsymbol{B}^T(\boldsymbol{\gamma} - \boldsymbol{\gamma}_0)\|_2 = C/\sqrt{nh}\}$, we have

$$P(\inf_{\boldsymbol{\gamma}\in\Theta} \boldsymbol{Q}_n(\boldsymbol{\gamma}) - \boldsymbol{Q}_n(\boldsymbol{\gamma}_0) > 0) > 1 - \varepsilon \tag{S.20}$$

for any small $\varepsilon > 0$ when n is sufficiently large, which implies (S.17).

□

Proof of Theorem 4

Proof. We prove the sparsity by contradiction. Suppose that there exists a $(d_0 + 1) \leq l \leq p$ such that $\hat{\alpha}_l^P \neq 0$. Let $\boldsymbol{\gamma}^*$ be the vector obtained from $\hat{\boldsymbol{\gamma}}^P$ with l-th component $\hat{\boldsymbol{\gamma}}_l^P$ being replaced by 0. It will be shown that

$$L_n(\hat{\boldsymbol{\gamma}}^P) > L_n(\boldsymbol{\gamma}^*), \tag{S.21}$$

which is a contradiction.

Note that

$$L_n(\hat{\boldsymbol{\gamma}}^P) - L_n(\boldsymbol{\gamma}^*) = \boldsymbol{Q}_n(\hat{\boldsymbol{\gamma}}^P) - \boldsymbol{Q}_n(\boldsymbol{\gamma}^*) + np_{\lambda_l}(\|\hat{\boldsymbol{\gamma}}_l^P\|_{\boldsymbol{H}_l}). \tag{S.22}$$

For the penalty term $np_{\lambda_l}(\|\hat{\boldsymbol{\gamma}}_l^P\|_{\boldsymbol{H}_l})$ in (S.22), since $\|\hat{\boldsymbol{\gamma}}_l^P\|_{\boldsymbol{H}_l} = O_p(1/\sqrt{nh}) = o(\lambda_l)$, by the expression of the SCAD penalty, we have

$$np_{\lambda_l}(\|\hat{\boldsymbol{\gamma}}_l^P\|_{\boldsymbol{H}_l}) = n\lambda_l\|\hat{\boldsymbol{\gamma}}_l^P\|_{\boldsymbol{H}_l},$$

with probabity approaching 1.

For $\boldsymbol{Q}_n(\hat{\boldsymbol{\gamma}}^P) - \boldsymbol{Q}_n(\boldsymbol{\gamma}^*)$, by the similar expansion used in Lemmas 4-6, we can prove

$$\begin{aligned} |\boldsymbol{Q}_n(\hat{\boldsymbol{\gamma}}^P) - \boldsymbol{Q}_n(\boldsymbol{\gamma}^*)| &= O_p\left(\sqrt{\tfrac{n}{h}}\|\hat{\boldsymbol{\gamma}}^P - \boldsymbol{\gamma}^*\|_2\right) \\ &= O_p\left(\sqrt{\tfrac{n}{h}}\|\hat{\gamma}_l^P\|_{\boldsymbol{H}_l}\right). \end{aligned}$$

Since $\min_l \lambda_l n^{r/(2r+1)} \to \infty$, the penalty term is the dominant term in (S.22), and (S.22) is positive with probability approaching 1, which is a contradiction. □

Appendix D. Some Tables

References

De Boor, C. (2001). A practical guide to splines, revised edition, vol. 27 of applied mathematical sciences.

He, X. and Shi, P. (1996). Bivariate tensor-product B-splines in a partly linear model, *Journal of Multivariate Analysis* **58**(2): 162–181.

Van der Vaart, A. W. (2000). *Asymptotic statistics*, Cambridge university press, Cambridge, UK.

Xue, L., Qu, A. and Zhou, J. (2010). Consistent model selection for marginal generalized additive model for correlated data, *Journal of the American Statistical Association* **105**(492): 1518–1530.

Table 1: Mean and standard deviation of AIMSE ($\times 100$) with SCAD penalty for divergent dimension $p = [n^{1/2}]$ in Example 2 .

			Penalized estimate			Oracle estimate		
Model	Error	τ	WI	QIF-CS	QIF-AR	WI	QIF-CS	QIF-AR
HM	Normal-CS	0.25	3.7683	3.4452	3.6053	3.3133	1.6467	2.1449
			2.0124	1.5938	2.1150	1.7800	0.8398	1.0814
		0.5	3.2106	2.3316	2.9449	2.9836	1.3332	1.7782
			1.6352	1.1282	1.4577	1.6389	0.5786	0.8623
	Normal-AR	0.25	4.3749	4.3208	4.2865	3.4143	3.2041	2.4425
			2.0746	2.5094	2.1002	1.4687	1.3536	1.0164
		0.5	3.8609	3.8437	3.8255	3.0331	2.6918	2.2028
			1.8659	2.5699	1.9876	1.3823	1.1313	1.0951
	t-CS	0.25	5.9815	5.0874	5.7861	5.5239	2.7862	3.4634
			3.1152	2.5283	3.3309	3.0939	1.6518	1.7747
		0.5	3.7140	2.5541	3.1955	3.3510	1.5983	2.0209
			1.8262	1.1476	1.6550	1.6078	0.7857	0.9673
	t-AR	0.25	6.4317	6.1757	5.9001	5.4388	4.8993	4.0250
			3.2455	3.5592	3.2448	2.6951	2.6941	1.9092
		0.5	4.5197	4.4046	4.3651	3.4973	3.0396	2.3898
			2.3046	3.1533	2.0071	1.7136	1.6474	1.1354
HT	Normal-CS	0.25	9.1378	7.3080	9.0360	6.7065	4.0029	5.3356
			4.5228	3.3541	4.4581	3.1707	1.9201	3.0351
		0.5	7.6052	5.3813	7.2759	6.1978	3.1219	4.2959
			3.6879	2.4015	3.6149	3.2437	1.3238	1.8428
	Normal-AR	0.25	11.1713	10.3638	10.2541	7.5311	7.3759	6.2250
			4.7264	5.2473	5.0741	3.0786	3.0901	2.6846
		0.5	8.9280	8.8312	8.6933	6.7230	6.0828	5.3273
			4.3854	4.8328	3.9472	2.9976	2.4674	2.5820
	t-CS	0.25	18.1109	11.2549	16.2405	11.3302	6.6377	8.3862
			9.3907	5.2600	8.2955	5.6889	3.7501	4.3779
		0.5	8.8348	7.0418	8.0592	7.0082	3.7593	5.0501
			4.3352	3.2325	4.7830	3.2100	1.8540	2.3530
	t-AR	0.25	18.0460	17.1934	17.0211	12.0352	11.2474	9.9787
			8.3855	9.6504	8.3061	5.4686	6.5977	4.6405
		0.5	10.2997	10.0843	9.4915	7.5855	7.1556	6.0257
			4.9041	5.2676	4.5249	3.4021	3.9304	3.0565

Table 2: Variable selection results for Example 2 with divergent dimension $p = [n^{1/2}]$.

			FNR			FPR		
Model	Error	τ	WI	QIF-CS	QIF-AR	WI	QIF-CS	QIF-AR
HM	Normal-CS	0.25	0	0	0	0.040	0.038	0.039
		0.5	0	0	0	0.026	0.023	0.026
	Normal-AR	0.25	0	0	0	0.042	0.039	0.040
		0.5	0	0	0	0.028	0.027	0.027
	t-CS	0.25	0	0	0	0.060	0.055	0.058
		0.5	0	0	0	0.041	0.039	0.040
	t-AR	0.25	0	0	0	0.076	0.075	0.077
		0.5	0	0	0	0.050	0.045	0.047
HT	Normal-CS	0.25	0	0	0	0.104	0.101	0.107
		0.5	0	0	0	0.091	0.089	0.091
	Normal-AR	0.25	0	0	0	0.124	0.121	0.127
		0.5	0	0	0	0.100	0.095	0.099
	t-CS	0.25	0	0.001	0.001	0.183	0.182	0.187
		0.5	0	0	0	0.114	0.110	0.115
	t-AR	0.25	0	0.002	0.002	0.193	0.192	0.193
		0.5	0	0	0	0.124	0.123	0.125