PARTIALLY FUNCTIONAL LINEAR QUANTILE REGRESSION
WITH MEASUREMENT ERRORS

Mengli Zhang, Lan Xue, Carmen D. Tekwe, Yang Bai and Annie Qu

Shanghai University of Finance and Economics, Oregon State University,
Indiana University School of Public Health and University of California, Irvine

Supplementary Material

This supplement material contains necessary lemmas, detailed proofs of the main theorems and additional simulation results.

S1 Technical Lemmas and Proofs

Lemma 1. Under assumptions (A1)-(A3), (A6) \( m^{2\alpha_x+2} \leq cn, \, nm^{2\alpha_x} \Delta^\nu_0 = O(1) \)
as \( n \to \infty \), we have, for \( i = 1, \ldots, n \), \( j = 1, \ldots, m \),

\[
\|\| \hat{K}_x - K_x \|\|^2 = O_p(\Delta^\nu_n),
\]

\[
\|\| \hat{\phi}_j - \phi_j \|\|^2 = O_p(\Delta^\nu_n + j^2 n^{-1}),
\]

\[
(\hat{W}_i - W_i)^T b_0 = O_p(\frac{\sqrt{m}}{\sqrt{n}}),
\]

\[
b_0^T (\hat{\Sigma}_u - \Sigma_u) b_0 = O_p(\frac{\sqrt{m}}{\sqrt{n}}).
\]
Proof. Here we assume \(m^{2\alpha_x+2} \leq cn\), which plays the same role as \(m = n^{1/(\alpha_x+2\beta)}\) in [Kato, 2012]. Thus, following the same calculations as in supplement of [Kato, 2012], the first two equations in our lemma hold. The difference between \(\hat{W}_i\) and \(W_i\) consists of two parts. One is caused by smoothing observed functions \(W_i(t)\). The other is cause by projecting \(W_i(t)\) on estimated \(\hat{\phi}_j(t)\) instead of \(\phi_j(t)\). After simple calculations, the third equation is proved. Similarly, we can get the fourth equation. \(\square\)

Lemma 2. Let \(\rho_h^*(W_i, \gamma) = \rho_h^*(Y_i - W_i^T b(\tau) - Z_i^T \theta(\tau), b^T \Sigma_u b)\) and \(l_i^*(\gamma) = \rho_h^*(W_i, \gamma) - E\rho_h^*(W_i, \gamma)\), where \(\gamma = (b^T(\tau), \theta^T(\tau))^T\). Under assumptions (A1)-(A2), (A4)-(A5), (A7)-(A8) and \(m^{2\alpha_x+2} \leq cn\), we have, for any constant \(c > 0\),

\[
\sup_{\|b\| \leq c, \theta \in \Theta} \frac{1}{n} \left| \sum_{i=1}^{n} l_i^*(\gamma) \right| = O_p \left\{ h^{-1} \exp(ch^{-2}) n^{-1/2} \sqrt{m \log(m+n)} \right\}.
\]

Proof. Let \(\Gamma = \{\gamma : \|b\| \leq c, \theta \in \Theta\}\). Assumptions required entail that, for any \(\gamma \in \Gamma\),

\[
E[\rho_h^*(W_i, \gamma)]^2 \leq c E \left\{ \int_0^{1/h} \left( e_i^2 + \sigma^2 \right) \exp(t^2 \sigma^2/2) dt \right\}^2 
\leq ch^{-2} \exp(\sigma^2 h^{-2}) E[e_i^2 + \sigma^2]^2 \leq ch^{-2} \exp(ch^{-2}),
\]

where \(\sigma^2 = b^T \Sigma_u b\), \(e_i = Y_i - W_i^T b - Z_i^T \theta\). Let \(V_n = \sqrt{\sum_{i=1}^n E[l_i^*(\gamma)]^2 + \ldots}\)
\[ \sum_{i=1}^{n}[l^*_i(\gamma)]^2. \] Then \( V_n = O_p(h^{-1}\exp(ch^{-2})n^{-1/2}). \) To prove Lemma 2, it suffices to show that, for sufficiently large \( c > 0, \)

\[ P \left( \sup_{\gamma \in \Gamma} \frac{\sum_{i=1}^{n} l^*_i(\gamma)}{V_n \sqrt{\log(m+n)}} \geq c \right) \to 0, \text{ as } n \to +\infty. \]  \((S1.1)\)

The proof of equation \((S1.1)\) makes use of the continuity property of \( l^*_i(\gamma) \) that there exists \( r_1 > 0 \) such that for every \( \gamma_1 \in \Gamma, 0 < d \leq 1, 1 \leq i \leq n, \)

\[ E \sup_{\|\gamma_2 - \gamma_1\| \leq d} |l^*_i(\gamma_2) - l^*_i(\gamma_1)| \leq n^{r_1}d, \]

which is easily derived by calculating upper bound of \( \|\frac{\partial \rho^*_i(W, \gamma)}{\partial \gamma}\|. \) Now consider a partition of \( \Gamma \) with cubes \( S_1, \ldots, S_M \) such that \( \bigcup_{j=1}^{M} S_j = \Gamma \) and the length of edge of each cube, \( \delta_n, \) satisfies \( \delta_n = n^{-(10+r_1)}/(m + p) \) or \( M \leq \left( \frac{2c}{\delta_n} \right)^{m+p}. \) Then

\[ P \left( \sup_{\gamma \in \Gamma} \frac{\sum_{i=1}^{n} l^*_i(\gamma)}{V_n \sqrt{\log(m+n)}} \geq c \right) \]

\[ \leq \sum_{j=1}^{M} P \left( \sup_{\gamma \in S_j} \frac{\sum_{i=1}^{n} l^*_i(\gamma)}{V_n \sqrt{\log(m+n)}} \geq c \right). \]  \((S1.2)\)

Let \( c_j \) be the center of \( S_j. \) Consider \( \bar{\delta}_i(j) = \{ \sup_{\gamma \in S_j} |l^*_i(\gamma) - l^*_i(c_j)| \leq n^{-4} \}, \)

\[ l^*_i(j) = l^*_i(c_j)I\{\bar{\delta}_i(j)\}, l^{**}_i(j) = l^*_i(j) - El^*_i(j). \] For \( \gamma \in S_j, \) Hölder inequality
and the continuity of $l_i^*(\gamma)$ entail that

$$\left| E[l_i^*(\gamma)I\{\bar{\delta}_i^c(j)\}] \right| \leq \left\{ E[l_i^*(\gamma)]^2 \right\}^{1/2} \left\{ \mathbb{P}[\bar{\delta}_i^c(j)] \right\}^{1/2} \leq \left\{ n^{-6}E[l_i^*(\gamma)]^2 \right\}^{1/2}, \quad (S1.3)$$

and observe that

$$\frac{|\sum_{i=1}^n l_i^*(\gamma)|}{V_n} \leq \frac{|\sum_{i=1}^n l_i^*(\gamma)I\{\bar{\delta}_i(j)\}|}{V_n} + \frac{|\sum_{i=1}^n l_i^*(\gamma)I\{\bar{\delta}_i(j)\}|}{V_n} + \sqrt{\sum_{i=1}^n I\{\bar{\delta}_i^c(j)\} + \sum_{i=1}^n \left| E[l_i^*(\gamma)I\{\bar{\delta}_i^c(j)\}] \right|} \leq \sqrt{\sum_{i=1}^n I\{\bar{\delta}_i^c(j)\} + n^{-5/2} + \sum_{i=1}^n \left| E[l_i^*(\gamma)I\{\bar{\delta}_i(j)\}] \right|}. \quad (S1.4)$$

Moreover, notice that $l_i^*(\gamma) = l_i^*(\gamma) - l_i^*(c_j) + l_i^*(c_j)$ and we have

$$\left| \sum_{i=1}^n l_i^*(\gamma)I\{\bar{\delta}_i(j)\} - E[l_i^*(\gamma)I\{\bar{\delta}_i(j)\}] \right| \leq \left| \sum_{i=1}^n [l_i^*(\gamma) - l_i^*(c_j)]I\{\bar{\delta}_i(j)\} - E[l_i^*(\gamma) - l_i^*(c_j)]I\{\bar{\delta}_i(j)\} \right| + \left| \sum_{i=1}^n l_i^*(c_j)I\{\bar{\delta}_i(j)\} - E[l_i^*(c_j)I\{\bar{\delta}_i(j)\}] \right| \leq 2n^{-3} + \left| \sum_{i=1}^n l_i^{**}(j) \right|. \quad (S1.5)$$

By $(S1.3)$, $(S1.4)$ and $(S1.5)$, we have

$$\frac{|\sum_{i=1}^n l_i^*(\gamma)|}{V_n} \leq 2n^{-3} + \frac{\sum_{i=1}^n l_i^{**}(j)}{V_n} + \sqrt{\sum_{i=1}^n I\{\bar{\delta}_i^c(j)\} + n^{-5/2}}.$$
Similarly for $V_n$ in the denominator, one has $V_n^*[1 + o_p(1)] \leq 4\sqrt{3}V_n$, where

$$V_n^* = \sqrt{\sum_i E[l_i^{**}(j)]^2} + \sqrt{\sum_i l_i^{**}(j)^2}$$

and $V_n^* = O_p(V_n)$. Finally, we have

$$\left| \frac{\sum_{i=1}^n l_i^*(\gamma)}{V_n} \right| \leq \frac{|\sum_{i=1}^n l_i^{**}(j)| + 2n^{-3}}{(4\sqrt{3})^{-1}V_n^*[1 + o_p(1)]} + \sqrt{\sum_{i=1}^n I\{\delta_i^C(j)\}} + n^{-5/2}. \quad (S1.6)$$

Thus, the inequality (S1.2) follows

$$P\left( \sup_{\gamma \in \Gamma} \left| \frac{\sum_{i=1}^n l_i^*(\gamma)}{V_n \sqrt{\text{mlog}(m+n)}} \right| \geq c \right) \leq \sum_{j=1}^M P \left( \frac{|\sum_{i=1}^n l_i^{**}(j)| + 2n^{-3}}{(4\sqrt{3})^{-1}V_n^*[1 + o_p(1)]} + n^{-5/2} \geq c\sqrt{\text{mlog}(m+n)} \right)$$

$$+ \sum_{j=1}^M P \left( \sqrt{\sum_{i=1}^n I\{\delta_i^C(j)\}} \geq c\sqrt{\text{mlog}(m+n)} \right)$$

$$\leq \sum_{j=1}^M P \left( \frac{|\sum_{i=1}^n l_i^{**}(j)|}{(4\sqrt{3})^{-1}V_n^*} \geq c\sqrt{\text{mlog}(m+n)} \right)$$

$$+ \sum_{j=1}^M P \left( \sqrt{\sum_{i=1}^n I\{\delta_i^C(j)\}} \geq c\sqrt{\text{mlog}(m+n)} \right),$$

where the last equation holds for large $n$. Following Lemma 3.1 in (He and Shao, 2000) and exponential inequality for binomial distribution (Ledoux and Talagrand, 1991, p.51), we have

$$\sum_{j=1}^M P \left( \frac{|\sum_{i=1}^n l_i^{**}(j)|}{(4\sqrt{3})^{-1}V_n^*} \geq c\sqrt{\text{mlog}(m+n)} \right)$$

$$\leq \sum_{j=1}^M C \exp(-16C^{-1}c^2\text{mlog}(m+n)), \quad (S1.7)$$
where \( C \) is a constant, and

\[
\sum_{j=1}^{M} P \left( \sqrt{\sum_{i=1}^{n} I\{ \delta_i^C(j) \}} \geq c\sqrt{\log(m+n)} \right) \\
\leq \sum_{j=1}^{M} \left[ \frac{3\sum_{i=1}^{n} P(\delta_i^C(j))}{c^2 m \log(m+n)} \right]^{c^2 m \log(m+n)}. \tag{S1.8}
\]

The upper bounds of inequality (S1.7) and (S1.8) go to zero after some calculations when \( c \) is sufficiently large and \( n \) goes to infinity, which proves equation (S1.1) and completes our proof.

**Lemma 3.** Under the same assumptions of Lemma 2, similarly denote \( l_i(\gamma) = \frac{[\rho_h(X_i, \gamma) - E\rho_h(X_i, \gamma)]}{n} \), where \( \rho_h(X, \gamma) = \rho_h(Y_i - X_i^T b(\tau) - Z_i^T \theta(\tau)) \).

As \( n \to \infty \), we have

\[
\sup_{\gamma \in \Theta} \left| \sum_{i=1}^{n} l_i(\gamma) \right| = O_p(n^{-1/2}\sqrt{m \log(m+n)}).
\]

**Proof.** The proof of Lemma 3 is very similar to proof of Lemma 2. So we omit it here.
S1. TECHNICAL LEMMAS AND PROOFS

S1.1 Proof of Theorem 1

One notes that

$$\int_0^1 \left\{ \hat{\beta}(t, \tau) - \beta_0(t, \tau) \right\}^2 dt = \int_0^1 \left\{ \sum_{j=1}^m (\hat{b}_j - b_{0j}(\tau))\hat{\phi}_j(t) + \sum_{j=1}^m b_{0j}(\tau)(\hat{\phi}_j(t) - \phi_j(t)) - \sum_{j=m+1}^{+\infty} b_{0j}(\tau)\phi_j(t) \right\}^2 dt$$

$$\leq 3\|\hat{b}(\tau) - b_0(\tau)\|^2 + 3m \sum_{j=1}^m b_{0j}^2(\tau)\|\hat{\phi}_j - \phi_j\|^2 + 3 \sum_{j=m+1}^{+\infty} b_{0j}^2(\tau)$$

$$= I + II + III. \quad (S1.9)$$

Suppose $m^{2\alpha_x+2}n^{-1} = O(1)$ and $nm^{2\alpha_x}\Delta^{\nu_0} = O(1)$. Then assumption (A5) and Lemma I entail that

$$\begin{align*}
II &= O_p(m^{-2\beta+2\alpha_x+4}\Delta^{\nu_0} + mn^{-1}) = O_p(mn^{-1}), \\
III &= O(m^{-2\beta+1}).
\end{align*}$$

Let $\gamma = (b^T(\tau), \theta^T(\tau))^T$ be the vector of unknown parameters at $\tau$th quantile. Correspondingly, denote the vector of true coefficients as $\gamma_0 = (b_0^T(\tau), \theta_0^T(\tau))^T$.

Theorem 1 follows if we can show that $\|\hat{\gamma} - \gamma_0\|^2 = O_p(a_n^2)$ with

$$a_n^2 = m^{-2\beta+1} + h^{-1}\exp(ch^{-2})m^{\alpha_x+1/2}[\log(m+n)]^{1/2}n^{-1/2} + hm^{\alpha_x}. \quad (S1.10)$$
MENGLI ZHANG, LAN XUE, CARMEN D. TEKWE, YANG BAI AND ANNIE QU

It is sufficient to show for some constant $C > 0$,

$$P\left( \inf_{\|\gamma - \gamma_0\| = C_{an}} E_n\rho_h^*(W_i, \gamma) > E_n\rho_h^*(W_i, \gamma_0) \right) \to 1, \quad (S1.11)$$

where $E_n\rho_h^*(W_i, \gamma)$ is the objective function. Here with some abuse of notation $\rho_h^*(W_i, \gamma)$ is used to emphasize the dependence of the function $\rho_h^*$ on $W_i$ and $\gamma$. We have following decomposition,

$$\inf_{\|\gamma - \gamma_0\| = C_{an}} E_n\rho_h^*(W_i, \gamma) - E_n\rho_h^*(W_i, \gamma_0) = \inf_{\|\gamma - \gamma_0\| = C_{an}} \{ E_n\rho_h^*(W_i, \gamma) - E_n\rho_T(X_i, \gamma) \} + \inf_{\|\gamma - \gamma_0\| = C_{an}} E_n\rho_T(X_i, \gamma) - E_n\rho_T(X_i, \gamma_0)$$

$$+ E_n\rho_T(X_i, \gamma_0) - E_n\rho_h^*(W_i, \gamma_0) = IV + V + VI. \quad (S1.12)$$

In addition, one notes that

$$\sup_{\|\gamma - \gamma_0\| \leq C_{an}} |E_n\rho_h^*(W_i, \gamma) - E_n\rho_T(X_i, \gamma)| \leq \sup_{\|\gamma - \gamma_0\| \leq C_{an}} |E_n\rho_h^*(W_i, \gamma) - E_n\rho_h(X_i, \gamma)| + \sup_{\|\gamma - \gamma_0\| \leq C_{an}} |E_n\rho_h(X_i, \gamma) - E_n\rho_T(X_i, \gamma)|$$

$$= O_p(h^{-1}exp(ch^{-2})\sqrt{mlog(m+n)n^{-1/2}}) + O_p(h), \quad (S1.13)$$

by Lemmas 2 and 3, and equations (A9) and (A10) in [Wang et al., 2012]. Thus, the terms IV and VI in (S1.12) are of order $O_p \left\{ h^{-1}exp(ch^{-2})\sqrt{mlog(m+n)n^{-1/2}} + h \right\}$. 
Let $\Delta \varepsilon_i = (X_i^T, Z_i^T) (\gamma - \gamma_0)$. Now for term V in (S1.12), Knight’s identity gives that,

$$V = -E_n \{ \Delta \varepsilon_i [\tau - I(\varepsilon_{i0} < 0)] \} + E_n \left\{ \int_0^{\Delta \varepsilon_i} [I(\varepsilon_{i0} < s) - I(\varepsilon_{i0} < 0)] ds \right\}$$

$$= \left\{ -E[\Delta \varepsilon_i (\tau - I(\varepsilon_{i0} < 0))] + E \int_0^{\Delta \varepsilon_i} [I(\varepsilon_{i0} < s) - I(\varepsilon_{i0} < 0)] ds \right\} [1 + o_p(1)]$$

$$= (E_1 + E_2)[1 + o_p(1)].$$

Let $Q_i = Q_{\tau}(Y_i|X_i(t), Z_i)$ and $T_i = \sum_{j=m+1}^{\infty} b_{0j} X_{ij}$. Assumptions (A2) and (A5) entail that $E(T_i)^2 \leq cm^{-2\beta - \alpha_x + 1}$. We have

$$E_1 = -E \left\{ \Delta \varepsilon_i [F_{Y|X(t),Z}(Q_i) - F_{Y|X(t),Z}(Q_i - T_i)] \right\}$$

$$= -E \left[ \Delta \varepsilon_i f_{Y|X(t),Z}(Q_i) T_i \right] (1 + o(1))$$

$$\leq c \sqrt{\text{E}(\Delta \varepsilon_i)^2} \sqrt{\text{ET}_i^2}$$

$$\leq c \sqrt{m^{-2\beta - \alpha_x + 1} \text{E}(\Delta \varepsilon_i)^2}. \quad (\text{S1.14})$$

For $E_2$, we have

$$E_2 = E \int_0^{\Delta \varepsilon_i} [F_{Y|X(t),Z}(s + Q_i - T_i) - F_{Y|X(t),Z}(Q_i - T_i)] ds$$

$$= \left\{ E \int_0^{\Delta \varepsilon_i} f_{Y|X(t),Z}(Q_i - T_i) ds \right\} [1 + o(1)]$$
\[ \geq cE \left[ f_{\gamma_i X(t),Z}(Q_i - T_i)(\Delta \varepsilon_i)^2 \right] \]
\[ \geq cE(\Delta \varepsilon_i)^2. \quad \text{(S1.15)} \]

Combining results in (S1.13), (S1.14) and (S1.15) into error decomposition in (S1.12), one has

\[
\inf_{\|\gamma - \gamma_0\| = C a_n} E_n \rho_h^*(W_i, \gamma) - E_n \rho_h^*(W_i, \gamma_0) \\
\geq c\sqrt{E(\Delta \varepsilon_i)^2 - c\sqrt{m^{-2\beta - \alpha_x + 1}}} \\
+ \mathcal{O}_p(h^{-1} \exp(ch^{-2}) \sqrt{m \log(m + n)n^{-1/2} + h}).
\]

Here \( E(\Delta \varepsilon_i)^2 = (\gamma - \gamma_0)^T \Lambda \Lambda (\gamma - \gamma_0) \), with \( \Lambda = \text{diag}\left(\kappa_{x1}^{1/2}, \ldots, \kappa_{xm}^{1/2}, 1, \ldots, 1\right) \) is a diagonal matrix with dimension \( m + p \). Therefore assumptions (A2) and (A3) give that \( E(\Delta \varepsilon_i)^2 \geq m^{-\alpha_x} \|\gamma - \gamma_0\|^2 \lambda_{\text{min}}(A)^{-1} \). Therefore, (S1.11) follows by noting that the first term is the dominating term uniformly in \( \|\gamma - \gamma_0\| = C a_n \) as long as \( C \) is large enough.

\section{S1.2 Proof of Theorem 2}

For simplicity, we replaced \( \hat{W}_i \) and \( \hat{\Sigma}_u \) with their true values \( W_i \) and \( \Sigma_u \) respectively in the objective function. However, Theorems 1 and 2 still hold if the estimated values are used by \( (\hat{W}_i - W_i)^T b_0 = \mathcal{O}_p(\frac{\sqrt{m}}{\sqrt{n}}) \) and \( b_0^T (\hat{\Sigma}_u - \Sigma_u) b_0 = \mathcal{O}_p(\frac{\sqrt{m}}{\sqrt{n}}) \) in Lemma [1]. Denote \( \hat{\theta}_1^* = \sqrt{n} \{ E_n[Z_i^T \frac{\partial^2}{\partial \varepsilon_i \partial \varepsilon_i^T} (\varepsilon_i - U_i^T b_0 + b_0^T \Sigma b_0)]\}^{-1} \)
$E_n [Z_i \frac{\partial \rho^*}{\partial \epsilon} (\epsilon_{i0} - U_i^T b_0, b_0^T \Sigma u b_0)]$. Obviously, it converges to $N(0, D^{-1}B D^{-1})$ according to Theorem 4. in (Wang et al., 2012). Thus we only need to prove $\sqrt{n}(\hat{\theta} - \theta_0) = \hat{\theta}_1^* + o_p(1)$. One notes that the objective function can be rewritten as

$$E_n [\rho^*_h(\epsilon_{i0} + T_i - \frac{1}{\sqrt{n}} Z_i^T \theta_1 - X_i^T (b - b_0) - (Z_i - Z_i^*)^T (\theta - \theta_0) - U_i^T \hat{b}, \hat{b}^T \Sigma u \hat{b})]$$

where $\theta_1 = \sqrt{n}(\theta - \theta_0)$, $T_i = \sum_{j=m+1}^{\infty} b_{0j} X_{ij}$. Note that $\hat{\theta}_1 = \sqrt{n}(\hat{\theta} - \theta_0)$ is the minimum of $E_n [\rho^*_h(\epsilon_{i0} + T_i - \frac{1}{\sqrt{n}} Z_i^T \theta_1 - X_i^T (b - b_0) - (Z_i - Z_i^*)^T (\theta - \theta_0) - U_i^T \hat{b}, \hat{b}^T \Sigma u \hat{b})]$. For simplicity, we define $D_i = (X_i^T, (Z_i - Z_i^*)^T)^T$ and $d(b, \theta) = ((b - b_0)^T, (\theta - \theta_0)^T)^T$. To prove our theorem, it suffices to prove that, for any $\delta > 0$,

$$P \left( \inf_{\|\theta_1 - \hat{\theta}_1^*\| = \delta} E_n [\rho^*_h(\epsilon_{i0} + T_i - \frac{1}{\sqrt{n}} Z_i^T \theta_1 - D_i^T d(b, \hat{\theta}) - U_i^T \hat{b}, \hat{b}^T \Sigma u \hat{b})] \right) \rightarrow 1.$$

As proved in Theorem 1, $\|\hat{\theta} - \theta_0\| = O_p(a_n)$ and $\|\hat{b} - b_0\| = O_p(a_n)$ with $a_n$ defined in (S1.10). Thus, by assumption (A10), equation (S1.16) follows from
following equation

\[
\sup_{\|\theta_1 - \hat{\theta}_1\| = \delta} I(\hat{\theta}, \hat{b}) \left| E_n[\rho_h(\varepsilon_{i0} + T_i - \frac{1}{\sqrt{n}} Z_i^T \theta_1 - D_i^T d(\hat{b}, \hat{\theta}) - U_i^T \hat{b}, \hat{b}^T \Sigma_u \hat{b}) - 2^{1/2} \| \hat{\rho}_h(\varepsilon_{i0} + T_i - \frac{1}{\sqrt{n}} Z_i^T \hat{\theta}_1 - D_i^T d(\hat{b}, \hat{\theta}) - U_i^T \hat{b}, \hat{b}^T \Sigma_u \hat{b})]\right] - 1/n \right|
\]

\[\leq o_p(1), \quad (S1.17)\]

where \(I(\hat{\theta}, \hat{b}) = I(\|\hat{\theta} - \theta_0\| \leq c_1 a_n, \|\hat{b} - b_0\| \leq c_2 a_n)\). By the definition of \(\hat{\theta}_1^*\), the left side of equation (S1.17) is less than or equal to

\[
2 \sup_{\|\theta_1\| \leq c_{1+n}} \left| E_n[\rho_h(\varepsilon_{i0} + T_i - \frac{1}{\sqrt{n}} Z_i^T \theta_1 - D_i^T d(\hat{b}, \theta) - U_i^T b, b^T \Sigma_u b)] - E_n[\rho_h(\varepsilon_{i0} + T_i - D_i^T d(\hat{b}, \theta) - U_i^T b, b^T \Sigma_u b)] \right. \\
+ \frac{1}{2n} \theta_i^T E_n[Z_i^* Z_i^T \frac{\partial^2 \rho_h}{\partial \varepsilon^2} \varepsilon_{i0} - U_i^T b_0, b_0^T \Sigma_u b_0]) \theta_1 \left| \right. \\
- \frac{1}{2n} \theta_i^T E_n[Z_i^* Z_i^T \frac{\partial^2 \rho_h}{\partial \varepsilon^2} \varepsilon_{i0} - U_i^T b_0, b_0^T \Sigma_u b_0]) \theta_1 \right| \leq o_p(1) \quad (S1.18)
\]
The equation (S1.18) holds by (A10), $E(\tilde{Z}_i - Z_i^*) = 0$ and the fact that $\rho_h$ is the conditional expectation of $\rho_h^*$ at any points with taking expectation of measurement errors, $U_i^T b$. Furthermore, applying Taylor expansion to $\rho_h(\cdot)$ in formula (S1.18) at $\varepsilon_{i0}$ and ignoring terms higher than 2-order, we can get the dominant term of (S1.18) as

$$\sup_{\|\theta_1\| \leq c} \frac{2}{\sqrt{n}} \left| \theta_1^T E_n \left\{ Z_i^T [T_i - D_i^T d(b, \theta)] \frac{\partial^2 \rho_h}{\partial \varepsilon^2} (\varepsilon_{i0}) \right\} \right|$$

$$= \sup_{\|\theta_1\| \leq c} \frac{2}{\sqrt{n}} \left| \theta_1^T E_n \left[ Z_i^* T_i \frac{\partial^2 \rho_h}{\partial \varepsilon^2} (\varepsilon_{i0}) \right] + \theta_1^T E_n \left[ Z_i^* X_i^T \frac{\partial^2 \rho_h}{\partial \varepsilon^2} (\varepsilon_{i0}) \right] (b - b_0) \right|$$

$$+ \theta_1^T E_n \left[ Z_i^* (Z_i - Z_i^*)^T \frac{\partial^2 \rho_h}{\partial \varepsilon^2} (\varepsilon_{i0}) \right] (\theta - \theta_0) \right|$$

$$\leq \frac{c}{\sqrt{n}} \| E_n [Z_i^* T_i \frac{\partial^2 \rho_h}{\partial \varepsilon^2} (\varepsilon_{i0})] \| + a_n \frac{c}{\sqrt{n}} \| E_n [Z_i^* X_i^T \frac{\partial^2 \rho_h}{\partial \varepsilon^2} (\varepsilon_{i0})] \|$$

$$+ a_n \frac{c}{\sqrt{n}} \| E_n [Z_i^* (Z_i - Z_i^*)^T \frac{\partial^2 \rho_h}{\partial \varepsilon^2} (\varepsilon_{i0})] \| \quad \text{(S1.19)}$$

Notice that $E\|Z_i^*\|^4$ is bounded and that each of the three summation terms in formula (S1.19) is zero mean by assumption (A9). The first term is of order $o_p(\frac{1}{n})$ plus assumptions (A2), (A5) and (A10). The second and third terms are also $o_p(\frac{1}{n})$ by $a_n \sqrt{m} = o(1)$. So far, we have completed the proof of $\sqrt{n}(\hat{\theta} - \theta_0) = \hat{\theta}_1^* + o_p(1)$ and the result follows straightly.
Table 1: Average number of basis selected by BIC criterion for data generated in Case 1. The values in the parentheses are Monte Carlo standard deviations.

<table>
<thead>
<tr>
<th>τ</th>
<th>α</th>
<th>n</th>
<th>Oracle</th>
<th>Naive</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5</td>
<td>1.1</td>
<td>200</td>
<td>1.730(0.468)</td>
<td>1.500(0.522)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>400</td>
<td>1.980(0.141)</td>
<td>1.830(0.403)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>600</td>
<td>2.030(0.171)</td>
<td>2.000(0.201)</td>
</tr>
<tr>
<td>0.5</td>
<td>2</td>
<td>200</td>
<td>1.530(0.521)</td>
<td>1.280(0.494)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>400</td>
<td>1.800(0.402)</td>
<td>1.520(0.502)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>600</td>
<td>1.920(0.273)</td>
<td>1.680(0.469)</td>
</tr>
<tr>
<td>0.75</td>
<td>1.1</td>
<td>200</td>
<td>1.810(0.506)</td>
<td>1.500(0.560)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>400</td>
<td>2.000(0.318)</td>
<td>1.880(0.433)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>600</td>
<td>2.070(0.256)</td>
<td>2.000(0.284)</td>
</tr>
<tr>
<td>0.75</td>
<td>2</td>
<td>200</td>
<td>1.520(0.611)</td>
<td>1.200(0.402)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>400</td>
<td>1.740(0.485)</td>
<td>1.520(0.522)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>600</td>
<td>1.920(0.273)</td>
<td>1.800(0.426)</td>
</tr>
</tbody>
</table>
Table 2: The value of selected $\hat{h}$ based on SIMEX for data generated in Case 1.

<table>
<thead>
<tr>
<th>$\alpha$</th>
<th>n</th>
<th>$\tau = 0.5$</th>
<th>$\tau = 0.75$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>200</td>
<td>2.147</td>
<td>1.854</td>
</tr>
<tr>
<td></td>
<td>400</td>
<td>1.714</td>
<td>1.526</td>
</tr>
<tr>
<td></td>
<td>600</td>
<td>1.480</td>
<td>1.372</td>
</tr>
<tr>
<td>2</td>
<td>200</td>
<td>2.288</td>
<td>1.705</td>
</tr>
<tr>
<td></td>
<td>400</td>
<td>1.808</td>
<td>1.569</td>
</tr>
<tr>
<td></td>
<td>600</td>
<td>1.552</td>
<td>1.405</td>
</tr>
</tbody>
</table>

Table 3: Bias, Variance and MSE of parametric coefficient estimators in three different data generating models.

<table>
<thead>
<tr>
<th>$\tau$</th>
<th>Method</th>
<th>Bias($\hat{\theta}_0$)</th>
<th>Bias($\hat{\theta}_1$)</th>
<th>Bias($\hat{\theta}_2$)</th>
<th>Var($\hat{\theta}_0$)</th>
<th>Var($\hat{\theta}_1$)</th>
<th>Var($\hat{\theta}_2$)</th>
<th>MSE($\hat{\theta}_0$)</th>
<th>MSE($\hat{\theta}_1$)</th>
<th>MSE($\hat{\theta}_2$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5</td>
<td>CL</td>
<td>-0.0269</td>
<td>-0.0064</td>
<td>0.0361</td>
<td>0.0535</td>
<td>0.0088</td>
<td>0.0389</td>
<td>0.0536</td>
<td>0.0088</td>
<td>0.0398</td>
</tr>
<tr>
<td></td>
<td>NAIVE</td>
<td>-0.1975</td>
<td>0.0013</td>
<td>0.2005</td>
<td>0.0209</td>
<td>0.0130</td>
<td>0.0100</td>
<td>0.0597</td>
<td>0.0129</td>
<td>0.0501</td>
</tr>
<tr>
<td></td>
<td>ORACLE</td>
<td>-0.0230</td>
<td>-0.0048</td>
<td>0.0275</td>
<td>0.0272</td>
<td>0.0105</td>
<td>0.0177</td>
<td>0.0275</td>
<td>0.0105</td>
<td>0.0183</td>
</tr>
<tr>
<td>0.75</td>
<td>CL</td>
<td>-0.0050</td>
<td>0.0080</td>
<td>0.0287</td>
<td>0.0389</td>
<td>0.0125</td>
<td>0.0323</td>
<td>0.0385</td>
<td>0.0124</td>
<td>0.0328</td>
</tr>
<tr>
<td></td>
<td>NAIVE</td>
<td>-0.1625</td>
<td>0.0016</td>
<td>0.1738</td>
<td>0.0245</td>
<td>0.0186</td>
<td>0.0086</td>
<td>0.0507</td>
<td>0.0184</td>
<td>0.0387</td>
</tr>
<tr>
<td></td>
<td>ORACLE</td>
<td>-0.0300</td>
<td>0.0049</td>
<td>0.0281</td>
<td>0.0316</td>
<td>0.0182</td>
<td>0.0203</td>
<td>0.0322</td>
<td>0.0180</td>
<td>0.0209</td>
</tr>
<tr>
<td>0.5</td>
<td>CL</td>
<td>-0.0283</td>
<td>0.0191</td>
<td>0.0226</td>
<td>0.0345</td>
<td>0.0105</td>
<td>0.0371</td>
<td>0.0350</td>
<td>0.0108</td>
<td>0.0372</td>
</tr>
<tr>
<td></td>
<td>NAIVE</td>
<td>-0.2253</td>
<td>0.0184</td>
<td>0.2018</td>
<td>0.0114</td>
<td>0.0139</td>
<td>0.0080</td>
<td>0.0620</td>
<td>0.0141</td>
<td>0.0486</td>
</tr>
<tr>
<td></td>
<td>ORACLE</td>
<td>-0.0238</td>
<td>0.0115</td>
<td>0.0162</td>
<td>0.0175</td>
<td>0.0120</td>
<td>0.0105</td>
<td>0.0179</td>
<td>0.0120</td>
<td>0.0107</td>
</tr>
<tr>
<td>0.75</td>
<td>CL</td>
<td>0.0334</td>
<td>0.0106</td>
<td>0.0242</td>
<td>0.0465</td>
<td>0.0126</td>
<td>0.0388</td>
<td>0.0472</td>
<td>0.0126</td>
<td>0.0390</td>
</tr>
<tr>
<td></td>
<td>NAIVE</td>
<td>-0.2922</td>
<td>0.0014</td>
<td>0.2894</td>
<td>0.0172</td>
<td>0.0177</td>
<td>0.0065</td>
<td>0.1024</td>
<td>0.0176</td>
<td>0.0901</td>
</tr>
<tr>
<td></td>
<td>ORACLE</td>
<td>-0.0284</td>
<td>0.0010</td>
<td>0.0224</td>
<td>0.0248</td>
<td>0.0138</td>
<td>0.0158</td>
<td>0.0254</td>
<td>0.0136</td>
<td>0.0162</td>
</tr>
</tbody>
</table>

Case 1

Case 2

Case 3
Table 3 (continued table)

<table>
<thead>
<tr>
<th>τ</th>
<th>Method</th>
<th>Bias($\hat{\theta}_0$)</th>
<th>Bias($\hat{\theta}_1$)</th>
<th>Var($\hat{\theta}_0$)</th>
<th>Var($\hat{\theta}_1$)</th>
<th>Var($\hat{\theta}_2$)</th>
<th>MSE($\hat{\theta}_0$)</th>
<th>MSE($\hat{\theta}_1$)</th>
<th>MSE($\hat{\theta}_2$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5</td>
<td>CL</td>
<td>-0.0064</td>
<td>-0.0185</td>
<td>0.0132</td>
<td>0.0337</td>
<td>0.0106</td>
<td>0.0242</td>
<td>0.0334</td>
<td>0.0109</td>
</tr>
<tr>
<td></td>
<td>NAIVE</td>
<td>-0.1761</td>
<td>-0.0177</td>
<td>0.1787</td>
<td>0.0148</td>
<td>0.0148</td>
<td>0.0050</td>
<td>0.0457</td>
<td>0.0150</td>
</tr>
<tr>
<td></td>
<td>ORACLE</td>
<td>-0.0211</td>
<td>-0.0116</td>
<td>0.0175</td>
<td>0.0218</td>
<td>0.0140</td>
<td>0.0111</td>
<td>0.0220</td>
<td>0.0140</td>
</tr>
<tr>
<td>0.75</td>
<td>CL</td>
<td>-0.0234</td>
<td>-0.0077</td>
<td>0.0338</td>
<td>0.0317</td>
<td>0.0096</td>
<td>0.0290</td>
<td>0.0319</td>
<td>0.0096</td>
</tr>
<tr>
<td></td>
<td>NAIVE</td>
<td>-0.1682</td>
<td>-0.0065</td>
<td>0.1869</td>
<td>0.0173</td>
<td>0.0154</td>
<td>0.0074</td>
<td>0.0455</td>
<td>0.0153</td>
</tr>
<tr>
<td></td>
<td>ORACLE</td>
<td>0.0045</td>
<td>0.0012</td>
<td>-0.0050</td>
<td>0.0247</td>
<td>0.0137</td>
<td>0.0165</td>
<td>0.0244</td>
<td>0.0136</td>
</tr>
</tbody>
</table>

Table 4: Bias, Variance and IMSE of functional coefficient estimators in three different data generating models.

<table>
<thead>
<tr>
<th>τ</th>
<th>CL</th>
<th>NAIVE</th>
<th>ORACLE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Bias(^2) Var MSE</td>
<td>Bias(^2) Var MSE</td>
<td>Bias(^2) Var MSE</td>
</tr>
<tr>
<td>Case 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td>0.0141 0.0960 0.1101</td>
<td>0.0907 0.0295 0.1202</td>
<td>0.0134 0.0323 0.0457</td>
</tr>
<tr>
<td>0.75</td>
<td>0.0124 0.0812 0.0936</td>
<td>0.0862 0.0280 0.1142</td>
<td>0.0106 0.0486 0.0592</td>
</tr>
<tr>
<td>Case 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td>0.0123 0.0828 0.0950</td>
<td>0.0939 0.0221 0.1161</td>
<td>0.0093 0.0291 0.0384</td>
</tr>
<tr>
<td>0.75</td>
<td>0.0537 0.0942 0.1479</td>
<td>0.1873 0.0252 0.2125</td>
<td>0.0105 0.0349 0.0454</td>
</tr>
<tr>
<td>Case 3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td>0.0111 0.0608 0.0718</td>
<td>0.0761 0.0176 0.0937</td>
<td>0.0084 0.0273 0.0357</td>
</tr>
<tr>
<td>0.75</td>
<td>0.0154 0.0831 0.0984</td>
<td>0.0885 0.0300 0.1185</td>
<td>0.0080 0.0344 0.0424</td>
</tr>
</tbody>
</table>
Bibliography


