Supplementary Material for "Focused information criterion and model averaging based on weighted composite quantile regression"

GANGGANG XU, SUOJIN WANG and JIANHUA Z. HUANG

Department of Statistics, Texas A&M University

1 Proof of Lemma A.1

Proof. Let $R_{ni}^* = R_{ni} - \mathbf{Z}_i^T \boldsymbol{\delta}_n$ be the approximation bias caused by the use of spline basis function and by the local mis-specification in the linear part. The key of the proof is to show that the local mis-specification in the linear part, $\mathbf{Z}_i^T \boldsymbol{\delta}_n$, is sufficiently small so that it does not affect the convergence rate of the nonlinear estimator. It suffices to show that

$$J_n^{\boldsymbol{\delta}_0/10} L_n^2 \max_{1 \le i \le n} |R_{n_i}^*| = o_p(1) \tag{1}$$

and

$$\frac{1}{J_n} \sum_{i=1}^n R_{ni}^{*2} = o_p(1). \tag{2}$$

Note that $|R_{ni}^*| \le |R_{n_i}| + |\mathbf{Z}_i^T \boldsymbol{\delta}_n|$ and $\sup_{\mathbf{X} \in [0,1]^p} |R_{ni}| \le C_2 J_n^{-r}$ by (**??**). Hence, for (1), it suffices to show that $J_n^{\boldsymbol{\delta}_0/10} L_n^2 \max_{1 \le i \le n} |\mathbf{Z}_i^T \boldsymbol{\delta}_n| = o_p(1)$. Note that, for any $\epsilon > 0$,

$$\begin{aligned} P(J_n^{\boldsymbol{\delta}_0/10} L_n^2 \max_{1 \le i \le n} |\mathbf{Z}_i^T \boldsymbol{\delta}_n| < \epsilon) &= \prod_{i=1}^n P(J_n^{\boldsymbol{\delta}_0/10} L_n^2 |\mathbf{Z}_i^T \boldsymbol{\delta}_n| < \epsilon) \\ &\ge \prod_{i=1}^n \{1 - P(|\mathbf{Z}_i^T \boldsymbol{\delta}_n| > \frac{\epsilon}{J_n^{3\boldsymbol{\delta}_0/10}})\} \\ &\ge (1 - \frac{J_n^{9\boldsymbol{\delta}_0/10}}{n^{3/2}\epsilon^3} M)^n \\ &\to \lim_{n \to \infty} \exp\{-\frac{J_n^{9\boldsymbol{\delta}_0/10}}{n^{1/2}\epsilon^3} M\} = 1, \end{aligned}$$

where $E|\mathbf{Z}_i^T \boldsymbol{\delta}_n n^{1/2}|^3 \le M_1$ and the second inequality follows from the Chebyshev inequality since $E||\mathbf{Z}_i||^3 < \infty$ by Condition 2.

For (2), by the inequality $(x+y)^2 \le 2x^2 + 2y^2$ and $\sup_{\mathbf{X} \in [0,1]^p} |R_{ni}| \le C_2 J_n^{-r}$, it suffices to show that $\frac{1}{J_n} \sum_{i=1}^n |\mathbf{Z}_i^T \boldsymbol{\delta}_n|^2 = o_p(1)$, which follows from

$$\frac{1}{J_n} \sum_{i=1}^n |\mathbf{Z}_i^T \boldsymbol{\delta}_n|^2 = \frac{1}{J_n} \frac{1}{n} \sum_{i=1}^n \boldsymbol{\delta}^{*T} \mathbf{Z}_i \mathbf{Z}_i^T \boldsymbol{\delta}^* \xrightarrow{p} 0, \quad \text{as} \quad J_n \to \infty,$$

where $\boldsymbol{\delta}^* = n^{1/2}\boldsymbol{\delta}_n = O(1)$. Having obtained (1) and (2), the rest of the proof follows similarly to that used in He & Shi (1994) and is omitted here to save space. □

2 Proof of Lemma A.2

Proof. Using the identity

$$\rho_\tau(r - v) - \rho_\tau(r) = -v\{I(r < 0) - \tau\} + \int_0^v \{I(r \le t) - I(r \le 0)\}dt,$$

we can show that

$$E\{U_{i,k}(L_n J_n^{1/2}\eta_k, L_n J_n^{1/2}\boldsymbol{\theta})\} = \int_{R^*_{ni}}^{R^*_{ni}+L_n J_n^{1/2}(n^{-1/2}\eta_k+\mathbf{T}_i^T\boldsymbol{\theta})} \{G(\xi_k + t) - G(\xi_k)\}\, dt. \tag{3}$$

Let $c_n = \max_i(L_n J_n^{1/2}||\mathbf{T}_i|| + |R^*_{ni}|)$. Note that $E||T_i|| = O_p(n^{-1/2}J_n^{1/2})$. Using (1) it is straightforward to show that $c_n = o_p(1)$. Hence we can use the Taylor expansion $G(\xi_k+t)-G(\xi_k) = \{g(\xi_k^*) - g(\xi_k)\}t + g(\xi_k)t$ for some $\xi_k^* \in [R^*_{ni}, R^*_{ni} + L_n J_n^{1/2}(n^{-1/2}\eta_k + \mathbf{T}_i^T\boldsymbol{\theta})]$. Plugging this back to (3), we have

$$\begin{aligned} &E\{U_{i,k}(L_n J_n^{1/2}\eta_k, L_n J_n^{1/2}\boldsymbol{\theta})\} \\ &\quad = \frac{g(\xi_k)}{2}[\{R^*_{ni} + L_n J_n^{1/2}(n^{-1/2}\eta_k + \mathbf{T}_i^T\boldsymbol{\theta})\}^2 - R^{*2}_{ni}] + D_{i,k}(\eta_k, \boldsymbol{\theta}), \end{aligned}$$

where $D_{i,k}(\eta_k, \boldsymbol{\theta}) = \int_{R^*_{ni}}^{R^*_{ni}+L_n J_n^{1/2}(n^{-1/2}\eta_k+\mathbf{T}_i^T\boldsymbol{\theta})}\{g(\xi_k^*) - g(\xi_k)\}t\, dt$. Since $g(\cdot)$ is Lipschitz in a neighborhood of ξ_k by Condition 4, there exists some $M_k > 0$ such that $|g(\xi_k^*) - g(\xi_k)| \le M_k|\xi_k^* - \xi_k| \le M_k|L_n J_n^{1/2}(n^{-1/2}\eta_k + \mathbf{T}_i^T\boldsymbol{\theta})|$, which implies that

$$\begin{aligned} &\left|\frac{1}{J_n}\sum_{k=1}^K\sum_{i=1}^n w_k E\{U_{i,k}(L_n J_n^{1/2}\eta_k, L_n J_n^{1/2}\boldsymbol{\theta})\}\right| \\ &\quad \ge \Bigg|\underbrace{\frac{L_n^2}{2}\sum_{k=1}^K\sum_{i=1}^n w_k g(\xi_k)(\mathbf{T}_i^{*T}\boldsymbol{\theta}_k)^2}_{I_1} + \underbrace{\frac{L_n}{J_n^{1/2}}\sum_{k=1}^K\{w_k g(\xi_k)\}\sum_{i=1}^n R^*_{ni}\mathbf{T}_i^{*T}\boldsymbol{\theta}_k}_{I_2}\Bigg| \\ &\qquad \underbrace{- c_n L_n^2 \sum_{k=1}^K\sum_{i=1}^n w_k M_k(\mathbf{T}_i^{*T}\boldsymbol{\theta}_k)^2}_{I_3}, \end{aligned}$$

where $\mathbf{T}_i^* = (n^{-1/2}, \mathbf{T}_i^T)^T$ and $\boldsymbol{\theta}_k = (\eta_k, \boldsymbol{\theta}^T)^T$ for $k = 1, \ldots, K$. By the definition of $\mathbf{T}_i^*$ and the fact that all covariates $\mathbf{Z}$ and spline basis are centered, one has

$$\sum_{i=1}^n \mathbf{T}_i^*\mathbf{T}_i^{*T} = \mathbf{I}_{J_n+p_S+1}, \tag{4}$$

where p_S is the number of linear covariates in the Sth submodel. Since $||\boldsymbol{\theta}||^2 + ||\boldsymbol{\eta}||^2 = 1$, at least one $||\boldsymbol{\theta}_k||^2$ is of order $O(1)$. Hence, we have

$$I_1 = \frac{L_n^2}{2}\sum_{k=1}^{K} w_k g(\xi_k)||\boldsymbol{\theta}_k||^2 = O(L_n^2).$$

Similarly, since $c_n \sim o_p(1)$, it follows that $I_3 = o_p(L_n^2)$. Recalling that $R_{ni}^* = R_{ni} - \mathbf{Z}_i\boldsymbol{\delta}_n$ and $R_{ni} \sim O(J_n^{-r})$, we have

$$\begin{aligned}
|I_2| &\leq L_n \sum_{k=1}^{K}\{w_k g(\xi_k)\}\{\sum_{i=1}^{n}|\frac{1}{J_n^{1/2}} R_{ni}^* \mathbf{T}_i^{*T}\boldsymbol{\theta}_k|\} \\
&\leq \frac{L_n}{2}\sum_{k=1}^{K}\{w_k g(\xi_k)\}\{\sum_{i=1}^{n}(\mathbf{T}_i^{*T}\boldsymbol{\theta}_k)^2 + \frac{1}{J_n}\sum_{i=1}^{n} R_{ni}^{*2}\} \\
&= \frac{L_n}{2}\sum_{k=1}^{K}\{w_k g(\xi_k)\}\{||\boldsymbol{\theta}_k||^2 + o_p(1)\} = O_p(L_n).
\end{aligned}$$

The second last equation follows from Equation (2). Therefore, I_1 is the dominate term, which implies

$$\left|\frac{1}{J_n}\sum_{k=1}^{K}\sum_{i=1}^{n} w_k E\{U_{i,k}(L_n J_n^{1/2}\eta_k, L_n J_n^{1/2}\boldsymbol{\theta})\}\right| \geq |I_1 + I_2| - I_3 = O_p(L_n^2). \tag{5}$$

One the other hand, a straightforward application of the Chebychev inequality shows that for any $\epsilon > 0$, there exists $M > 0$ such that for all n

$$\begin{aligned}
\sup_{||\boldsymbol{\theta}||^2+||\boldsymbol{\eta}||^2=1} & P\left\{\left|\sum_{k=1}^{K}\sum_{i=1}^{n} w_k(n^{-1/2}\eta_k + \mathbf{T}_i^T\boldsymbol{\theta})\phi_{\tau_k}(\varepsilon_i - \xi_k)\right| > M\right\} \\
&\leq \frac{1}{M^2}\sum_{k=1}^{K}\sum_{l=1}^{K} w_k w_l\{\min(\tau_k, \tau_l) - \tau_k\tau_l\}(||\boldsymbol{\theta}||^2 + \eta_k\eta_l) < \epsilon,
\end{aligned}$$

which implies that when $||\boldsymbol{\theta}||^2 + ||\boldsymbol{\eta}||^2 = 1$,

$$\sum_{k=1}^{K}\sum_{i=1}^{n} w_k[L_n J_n^{-1/2}(n^{-1/2}\eta_k + \mathbf{T}_i^T\boldsymbol{\theta})\phi_{\tau_k}(\varepsilon_i - \xi_k)] = O_p(L_n J_n^{-1/2}). \tag{6}$$

Combining (5) and (6), since the first part of (A.5) dominates the second part and is of order $O_p(L_n^2)$, the proof of Lemma A.2 is completed. □

3 Proof of Lemma A.3

Proof. Using arguments similar to those in the proof of Lemma A.2, we can show that

$$\begin{aligned}
E\{V_{i,k}(\eta_k, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2)\} = \frac{g(\xi_k)}{2}\{&(R_{ni}^* + n^{-1/2}\eta_k + \mathbf{T}_{1i}^T\boldsymbol{\theta}_1 + \mathbf{T}_{2i}^T\boldsymbol{\theta}_2)^2 \\
&- (R_{ni}^* + n^{-1/2}\eta_k + \mathbf{T}_{2i}^T\boldsymbol{\theta}_2)^2\} + D_{i,k}(\eta_k, \boldsymbol{\theta}),
\end{aligned}$$

where $D_{i,k}(\eta_k, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = \int_{R_{ni}^*+n^{-1/2}\eta_k+\mathbf{T}_{2i}^T\boldsymbol{\theta}_2}^{R_{ni}^*+n^{-1/2}\eta_k+\mathbf{T}_{1i}^T\boldsymbol{\theta}_1+\mathbf{T}_{2i}^T\boldsymbol{\theta}_2} \{g(\xi_k^*) - g(\xi_k)\}t\,dt$ and ξ_k^* is some point in a small neighborhood of ξ_k. Then

$$\begin{aligned}\sum_{k=1}^{K}\sum_{i=1}^{n} E\{V_{i,k}(\eta_k, \boldsymbol{\theta})\} &= \frac{\sum_{k=1}^{K} w_k g(\xi_k)}{2}\left\{\boldsymbol{\theta}_1^T(\sum_{i=1}^{n}\mathbf{T}_{1i}\mathbf{T}_{1i}^T)\boldsymbol{\theta}_1 - \boldsymbol{\theta}_1^T(\sum_{i=1}^{n}\mathbf{T}_{1i}\mathbf{Z}_i^T)\boldsymbol{\delta}_n\right\} \\ &\quad + \sum_{k=1}^{K} w_k g(\xi_k)\boldsymbol{\theta}_1^T\sum_{i=1}^{n}\mathbf{T}_{1i}R_{ni} + \sum_{k=1}^{K}\sum_{i=1}^{n} w_k D_{i,k}(\eta_k, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2) \\ &= \frac{b}{2}\boldsymbol{\theta}_1^T\boldsymbol{\theta}_1 - bn^{1/2}\boldsymbol{\theta}_1^T\mathbf{H}_{1n}^-\Pi_s\hat{\boldsymbol{\Sigma}}\begin{pmatrix} 0 \\ \boldsymbol{\delta}\end{pmatrix} + r_n(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2).\end{aligned}$$

Using arguments similar to those in the proof of Lemma A.2, when $||\boldsymbol{\theta}_1|| \leq M$ and $||\boldsymbol{\theta}_2||^2 + ||\boldsymbol{\eta}||^2 \leq L^2 J_n$, we can show that

$$\sum_{k=1}^{K}\sum_{i=1}^{n} w_k D_{i,k}(\eta_k, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = o_p(1)||\boldsymbol{\theta}_1||^2 = o_p(1), \tag{7}$$

and by the Cauchy-Schwartz inequality and recall that $n^{1/(2r)} \ll J_n$ by Condition 3,

$$(\sum_{i=1}^{n}\boldsymbol{\theta}_1^T\mathbf{T}_{1i}R_{ni})^2 \leq \boldsymbol{\theta}_1^T(\sum_{i=1}^{n}\mathbf{T}_{1i}\mathbf{T}_{1i}^T)\boldsymbol{\theta}_1 O(nJ_n^{-2r}) = o(1)||\boldsymbol{\theta}_1||^2. \tag{8}$$

Hence, $\sup_{||\boldsymbol{\theta}_1||\leq M, ||\boldsymbol{\theta}_2||^2+||\boldsymbol{\eta}||^2\leq L^2J_n} |r_n(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)| = o_p(1)$. □