Online Supplementary Appendix C for "Variable Selection in Heteroscedastic Single Index Quantile Regression" by Eliana Christou and Michael G. Akritas.

Lemma C.1 *Assume that for some $r > 2$, $\mathrm{E}|Q_\tau(Y|\mathbf{X})|^r < \infty$ and $\sup_{t\in\mathfrak{T}_{\mathbf{b}}} \mathrm{E}[|Q_\tau(Y|\mathbf{X})|^r|\mathbf{b}_1^\top\mathbf{X} = t]f_{\mathbf{b}}(t) < \infty$ holds for all $\mathbf{b} \in \Theta$, where $\mathfrak{T}_{\mathbf{b}} = \{t : t = \mathbf{b}_1^\top\mathbf{x}, \mathbf{x} \in \mathcal{X}_0\}$, $\mathcal{X}_0$ is the compact support of $\mathbf{X}$, and $f_{\mathbf{b}}$ is the density of $\mathbf{b}_1^\top\mathbf{X}$. Moreover, assume that $Q_\tau(Y|\mathbf{x})$ is in $H_s(\mathcal{X}_0)$ for some s with $[s] \leq k$, where $H_s(\mathcal{X}_0)$ is defined in Appendix A and k is the order of the local polynomial conditional quantile estimators $\widehat{Q}_\tau(Y|\mathbf{X}_i)$ and $\widehat{Q}_\tau^{VS}(Y|\mathbf{X}_i)$ (used in (2.3) and (2.6), respectively).*

1. *Under Assumptions GS1-GS2 and Assumptions A1-A5 given in Appendix A,*

$$\sup_{\mathbf{b}\in\Theta, t\in\mathfrak{T}_{\mathbf{b}}} \left|\widehat{g}^{NW}(t|\mathbf{b}) - g(t|\mathbf{b})\right| = O_p\left(a_n^* + a_n + h^2\right),$$

where $\widehat{g}^{NW}(t|\mathbf{b})$ is defined in (2.3), $a_n^* = (\log n/n)^{s/(2s+d)}$, and $a_n = [\log n/(nh)]^{1/2}$.

2. *Under the sparsity assumption, Assumptions GS1-GS3, Assumptions A1-A7 given in Appendix A, and the conditions $nh^4 = o(1)$, where h is the bandwidth used in (2.6), $\lambda_1 \to 0$ and $\sqrt{n}\lambda_1 \to \infty$ as $n \to \infty$, where λ_1 is the tuning parameter used in (2.4),*

$$\sup_{\mathbf{b}\in\Theta, t\in\mathfrak{T}_{\mathbf{b}}} \left|\widehat{g}_{VS}^{NW}(t|\mathbf{b}) - g(t|\mathbf{b})\right| = O_p\left(a_n^{**} + a_n + h^2\right),$$

where $\widehat{g}_{VS}^{NW}(t|\mathbf{b})$ is defined in (2.6), and $a_n^{**} = (\log n/n)^{s/(2s+d^*)}$.

Proof. The proof uses the same steps as those in the proof of Proposition 3.1 of Christou and Akritas (2016). We outline here the basic steps.

Let $\widehat{g}(t|\mathbf{b})$ denote either $\widehat{g}^{NW}(t|\mathbf{b})$, defined in (2.3), or $\widehat{g}_{VS}^{NW}(t|\mathbf{b})$, defined in (2.6). Also, let $\widehat{Q}_\tau^*(Y|\mathbf{x})$ denote either $\widehat{Q}_\tau(Y|\mathbf{x})$ or $\widehat{Q}_\tau^{VS}(Y|\mathbf{x})$; see Section 2. Set $K_h(\cdot) = K(\cdot/h)$, and write $\widehat{g}(t|\mathbf{b}) = \widehat{\Psi}(t|\mathbf{b})/\widehat{f}_{\mathbf{b}}(t)$, where $\widehat{\Psi}(t|\mathbf{b}) = (nh)^{-1}\sum_{i=1}^n \widehat{Q}_\tau^*(Y|\mathbf{X}_i)K_h\left(t - \mathbf{b}_1^\top\mathbf{X}_i\right)$ and $\widehat{f}_{\mathbf{b}}(t) = (nh)^{-1}\sum_{i=1}^n K_h\left(t - \mathbf{b}_1^\top\mathbf{X}_i\right)$.

For the denominator, we use Theorem 6 of Hansen (2008) [take his $\beta = \infty$ and the mixing coefficients as $\alpha_m = 0$] to obtain

$$\sup_{\mathbf{b}\in\Theta, t\in\mathfrak{T}_{\mathbf{b}}} \left|\widehat{f}_{\mathbf{b}}(t) - f_{\mathbf{b}}(t)\right| = O_p\left[\left(\frac{\log n}{nh}\right)^{1/2} + h^2\right] = O_p(a_n + h^2). \tag{C.1}$$

For the numerator, we show that $\widehat{\Psi}(t|\mathbf{b})$ is consistent estimator of $\Psi(t|\mathbf{b}) = g(t|\mathbf{b})f_{\mathbf{b}}(t)$, uniformly in $\mathbf{b} \in \Theta$ and $t \in \mathfrak{T}_{\mathbf{b}}$. By letting $\Psi^*(t|\mathbf{b}) = (nh)^{-1}\sum_{i=1}^n Q_\tau(Y|\mathbf{X}_i)K_h\left(t - \mathbf{b}_1^\top\mathbf{X}_i\right)$, we can show that

$$\begin{aligned}|\widehat{\Psi}(t|\mathbf{b}) - \Psi^*(t|\mathbf{b})| &= \left|\frac{1}{nh}\sum_{i=1}^n \left[\widehat{Q}_\tau^*(Y|\mathbf{X}_i) - Q_\tau(Y|\mathbf{X}_i)\right] K_h\left(t - \mathbf{b}_1^\top\mathbf{X}_i\right)\right| \\ &\leq \sup_{1\leq i\leq n}\left|\widehat{Q}_\tau^*(Y|\mathbf{X}_i) - Q_\tau(Y|\mathbf{X}_i)\right| \frac{1}{nh}\sum_{i=1}^n K_h\left(t - \mathbf{b}_1^\top\mathbf{X}_i\right),\end{aligned}$$

and

$$\sup_{\mathbf{b}\in\Theta, t\in\mathfrak{T}_{\mathbf{b}}} |\widehat{\Psi}(t|\mathbf{b}) - \Psi^*(t|\mathbf{b})| = \begin{cases} O_p(a_n^*), & \text{if } \widehat{Q}_\tau^*(Y|\mathbf{X}_i) = \widehat{Q}_\tau(Y|\mathbf{X}_i) \\ O_p(a_n^{**}), & \text{if } \widehat{Q}_\tau^*(Y|\mathbf{X}_i) = \widehat{Q}_\tau^{VS}(Y|\mathbf{X}_i), \end{cases} \tag{C.2}$$

where the last equality follows from relation (C.1), Assumption A2, and the uniform consistency results for $\widehat{Q}_\tau(Y|\mathbf{X}_i)$ (cf. Guerre and Sabbah 2012), and for $\widehat{Q}_\tau^{VS}(Y|\mathbf{X}_i)$ (see Proposition 3.2). Next, Theorem 2 of Hansen (2008) yields $\sup_{\mathbf{b}\in\Theta, t\in\mathfrak{T}_{\mathbf{b}}} |\Psi^*(t|\mathbf{b}) - E[\Psi^*(t|\mathbf{b})]| = O_p(a_n)$, where $a_n = [\log n/(nh)]^{1/2}$, and recalling the notation $\Psi(t|\mathbf{b}) = g(t|\mathbf{b})f_{\mathbf{b}}(t)$, and using Assumption A4, $\mathrm{E}[\Psi^*(t|\mathbf{b})] = \Psi(t|\mathbf{b}) + O(h^2)$. Thus, $\sup_{\mathbf{b}\in\Theta, t\in\mathfrak{T}_{\mathbf{b}}} |\Psi^*(t|\mathbf{b}) - \Psi(t|\mathbf{b})| = O_p(a_n + h^2)$ which, together with (C.2) yields

$$\sup_{\mathbf{b}\in\Theta, t\in\mathfrak{T}_{\mathbf{b}}} \left|\widehat{\Psi}(t|\mathbf{b}) - \Psi(t|\mathbf{b})\right| = \begin{cases} O_p\left(a_n^* + a_n + h^2\right) & \text{if } \widehat{Q}_\tau^*(Y|\mathbf{X}_i) = \widehat{Q}_\tau(Y|\mathbf{X}_i) \\ O_p\left(a_n^{**} + a_n + h^2\right), & \text{if } \widehat{Q}_\tau^*(Y|\mathbf{X}_i) = \widehat{Q}_\tau^{VS}(Y|\mathbf{X}_i). \end{cases} \tag{C.3}$$

Therefore, using (C.1), (C.3) and Assumption A2, we get

$$\left|\frac{\widehat{\Psi}(t|\mathbf{b})}{\widehat{f}_{\mathbf{b}}(t)} - g(t|\mathbf{b})\right| = \begin{cases} O_p\left(a_n^* + a_n + h^2\right) & \text{if } \widehat{Q}_\tau^*(Y|\mathbf{X}_i) = \widehat{Q}_\tau(Y|\mathbf{X}_i) \\ O_p\left(a_n^{**} + a_n + h^2\right), & \text{if } \widehat{Q}_\tau^*(Y|\mathbf{X}_i) = \widehat{Q}_\tau^{VS}(Y|\mathbf{X}_i) \end{cases}$$

uniformly in $\mathbf{b} \in \Theta$ and $t \in \mathfrak{T}_{\mathbf{b}}$.

Note: For what follows, $\Pr(\cdot|\mathbb{X})$ and $\mathrm{E}(\cdot|\mathbb{X})$ will denote the conditional probability and conditional expectation, respectively, on the design matrix $\mathbb{X}$.

Lemma C.2 *Let $\widehat{g}(t|\mathbf{b})$ denote either $\widehat{g}^{NW}(t|\mathbf{b})$, defined in (2.3), or $\widehat{g}^{NW}_{VS}(t|\mathbf{b})$, defined in (2.6). Define, for any $\boldsymbol{\gamma} \in \mathbb{R}^{d-1}$,*

$$\widetilde{A}_n(\tau, \boldsymbol{\gamma}) = \sum_{i=1}^{n} \left\{ \rho_\tau[Y_i^* - \widetilde{g}(\mathbf{X}_i|\boldsymbol{\gamma}/\sqrt{n} + \boldsymbol{\beta}, \boldsymbol{\beta})] - \rho_\tau(Y_i^*) \right\}, \tag{C.4}$$

where $Y_i^* = Y_i - \widehat{g}(\boldsymbol{\beta}_1^\top \mathbf{X}_i|\boldsymbol{\beta})$ and $\widetilde{g}(\mathbf{X}_i|\boldsymbol{\gamma}/\sqrt{n} + \boldsymbol{\beta}, \boldsymbol{\beta}) = \widehat{g}[(\boldsymbol{\gamma}/\sqrt{n} + \boldsymbol{\beta})_1^\top \mathbf{X}_i|\boldsymbol{\gamma}/\sqrt{n} + \boldsymbol{\beta}] - \widehat{g}(\boldsymbol{\beta}_1^\top \mathbf{X}_i|\boldsymbol{\beta})$, for $(\boldsymbol{\gamma}/\sqrt{n} + \boldsymbol{\beta})_1 = (1, (\boldsymbol{\gamma}/\sqrt{n} + \boldsymbol{\beta})^\top)^\top$. Then, under the assumptions of Lemma C.1, Assumptions A6 and A7 given in Appendix A, and the condition $nh^4 = o(1)$, the following quadratic approximation holds uniformly in $\boldsymbol{\gamma}$ in a compact set, $\widetilde{A}_n(\tau, \boldsymbol{\gamma}) = (1/2)\boldsymbol{\gamma}^\top \mathbb{V} \boldsymbol{\gamma} + \mathbf{W}_n^\top \boldsymbol{\gamma} + o_p(1)$, where

$$\mathbb{V} = \mathrm{E}\left\{ [g'(\boldsymbol{\beta}_1^\top \mathbf{X}|\boldsymbol{\beta})]^2 [\mathbf{X}_{-1} - \mathrm{E}(\mathbf{X}_{-1}|\boldsymbol{\beta}_1^\top \mathbf{X})][\mathbf{X}_{-1} - \mathrm{E}(\mathbf{X}_{-1}|\boldsymbol{\beta}_1^\top \mathbf{X})]^\top f_{\epsilon|\mathbf{X}}(0|\mathbf{X}) \right\}, \tag{C.5}$$

and

$$\mathbf{W}_n = -n^{-1/2} \sum_{i=1}^{n} \rho'_\tau(Y_i^*) g'(\boldsymbol{\beta}_1^\top \mathbf{X}_i|\boldsymbol{\beta})[\mathbf{X}_{i,-1} - \mathrm{E}(\mathbf{X}_{-1}|\boldsymbol{\beta}_1^\top \mathbf{X})], \tag{C.6}$$

for $g'(t|\mathbf{b}) = (\partial/\partial t) g(t|\mathbf{b})$, and $\mathbf{X}_{-1}$ the $(d-1)$-dimensional vector consisting of coordinates $2, \ldots, d$ of $\mathbf{X}$.

Proof. The proof uses the same steps as those in the proof of Lemma C.6 of Christou and Akritas (2016). We outline here the basic steps.

Define H to be a class of bounded functions $\eta : \mathbb{R}^d \to \mathbb{R}$, whose value at $(t, \boldsymbol{\beta}^\top)^\top \in \mathbb{R}^d$ can be written as $\eta(t|\boldsymbol{\beta})$, in the non-separable space $l^\infty(t, \boldsymbol{\beta}) = \{(t, \boldsymbol{\beta}^\top)^\top : \mathbb{R}^d \to \mathbb{R} : \|\eta\|_{(t,\boldsymbol{\beta})} := \sup_{(t,\boldsymbol{\beta}^\top)^\top \in \mathbb{R}^d} |\eta(t|\boldsymbol{\beta})| < \infty\}$, and having bounded and continuous partial derivatives, where the first and second derivatives with respect to t exist and are bounded. Thus, H includes $g(t|\boldsymbol{\beta})$, as well as $\widehat{g}(t|\boldsymbol{\beta})$ for n large enough, almost surely. Define $\widetilde{A}_n(\eta, \tau, \boldsymbol{\gamma}) = \sum_{i=1}^{n} \left\{ \rho_\tau[e_i(\boldsymbol{\beta}, \eta) - \widetilde{\eta}(\mathbf{X}_i|\boldsymbol{\gamma}/\sqrt{n} + \boldsymbol{\beta}, \boldsymbol{\beta})] - \right.$

$\rho_\tau[e_i(\boldsymbol{\beta},\eta)]\big\}$, where $e_i(\boldsymbol{\beta},\eta) = Y_i - \eta(\boldsymbol{\beta}_1^\top \mathbf{X}_i|\boldsymbol{\beta})$ and $\widetilde{\eta}(\mathbf{X}_i|\boldsymbol{\gamma}/\sqrt{n}+\boldsymbol{\beta},\boldsymbol{\beta}) = \eta[(\boldsymbol{\gamma}/\sqrt{n}+\boldsymbol{\beta})_1^\top \mathbf{X}_i|\boldsymbol{\gamma}/\sqrt{n}+\boldsymbol{\beta}] - \eta(\boldsymbol{\beta}_1^\top \mathbf{X}_i|\boldsymbol{\beta})$, and write $\widetilde{A}_n(\eta,\tau,\boldsymbol{\gamma})$ as

$$\mathrm{E}\left[\widetilde{A}_n(\eta,\tau,\boldsymbol{\gamma})\,|\mathbb{X}\right] - \sum_{i=1}^n \left\{\rho_\tau'[e_i(\boldsymbol{\beta},\eta)] - \mathrm{E}\{\rho_\tau'[e_i(\boldsymbol{\beta},\eta)]|\mathbb{X}\}\right\}\widetilde{\eta}\left(\mathbf{X}_i|\boldsymbol{\gamma}/\sqrt{n}+\boldsymbol{\beta},\boldsymbol{\beta}\right) + R_n(\eta,\tau,\boldsymbol{\gamma}), \quad \text{(C.7)}$$

where $\mathbb{X}$ denotes the design matrix, and $R_n(\eta,\tau,\boldsymbol{\gamma})$ is the remainder term defined by (C.7). Using the same steps as in the proof of Lemma C.6 of Christou and Akritas (2016), we can show that

$$\begin{aligned}\mathrm{E}\left[\widetilde{A}_n(\eta,\tau,\boldsymbol{\gamma})\,|\mathbb{X}\right] &= -\sum_{i=1}^n \mathrm{E}\{\rho_\tau'[e_i(\boldsymbol{\beta},\eta)]|\mathbb{X}\}\widetilde{\eta}\left(\mathbf{X}_i|\boldsymbol{\gamma}/\sqrt{n}+\boldsymbol{\beta},\boldsymbol{\beta}\right)\\ &\quad+\frac{1}{2}\sum_{i=1}^n[\widetilde{\eta}(\mathbf{X}_i|\boldsymbol{\gamma}/\sqrt{n}+\boldsymbol{\beta},\boldsymbol{\beta})]^2\varphi''\left[g(\boldsymbol{\beta}_1^\top \mathbf{X}_i|\boldsymbol{\beta}) - \eta(\boldsymbol{\beta}_1^\top \mathbf{X}_i|\boldsymbol{\beta})\big|\mathbb{X}\right] + o_p(1), \quad \text{(C.8)}\end{aligned}$$

uniformly in $\eta \in H$. Following, using the Uniform Law of Large Numbers for Triangular Arrays (Jennrich, 1969), we can show that $\sup_{\eta\in H}|R_n(\eta,\tau,\boldsymbol{\gamma})| = o_p(1)$, where $R_n(\eta,\tau,\boldsymbol{\gamma})$ is defined in (C.7).

Next, substituting the expression of $\mathrm{E}[\widetilde{A}_n(\eta,\tau,\boldsymbol{\gamma})|\mathbb{X}]$ derived in (C.8), to relation (C.7) and using the fact that $\sup_{\eta\in H}|R_n(\eta,\tau,\boldsymbol{\gamma})| = o_p(1)$, we get, uniformly in $\eta\in H$,

$$\begin{aligned}\widetilde{A}_n(\eta,\tau,\boldsymbol{\gamma}) &= \frac{1}{2}\sum_{i=1}^n[\widetilde{\eta}(\mathbf{X}_i|\boldsymbol{\gamma}/\sqrt{n}+\boldsymbol{\beta},\boldsymbol{\beta})]^2\varphi''[g(\boldsymbol{\beta}_1^\top \mathbf{X}_i|\boldsymbol{\beta}) - \eta(\boldsymbol{\beta}_1^\top \mathbf{X}_i|\boldsymbol{\beta})|\mathbb{X}]\\ &\quad-\sum_{i=1}^n \rho_\tau'[e_i(\boldsymbol{\beta},\eta)]\widetilde{\eta}(\mathbf{X}_i|\boldsymbol{\gamma}/\sqrt{n}+\boldsymbol{\beta},\boldsymbol{\beta}) + o_p(1). \quad \text{(C.9)}\end{aligned}$$

Since expression (C.9) holds uniformly in $\eta\in H$, where the class H includes $\widehat{g}$, we substitute η with $\widehat{g}$. Using (a) the fact that $\widetilde{A}_n(\widehat{g},\tau,\boldsymbol{\gamma})$ reduces to $\widetilde{A}_n(\tau,\boldsymbol{\gamma})$ defined in (C.4), (b) relation

$$\sum_{i=1}^n \widetilde{g}(\mathbf{X}_i|\boldsymbol{\gamma}/\sqrt{n}+\boldsymbol{\beta},\boldsymbol{\beta}) = \sum_{i=1}^n \left\{g[(\boldsymbol{\gamma}/\sqrt{n}+\boldsymbol{\beta})_1^\top \mathbf{X}_i|\boldsymbol{\gamma}/\sqrt{n}+\boldsymbol{\beta}] - g(\boldsymbol{\beta}_1^\top \mathbf{X}_i|\boldsymbol{\beta})\right\} + o_p(n^{-1/2}),$$

follows from Lemma C.5 of Christou and Akritas (2016), and (c) relation

$$\begin{aligned}g[(\boldsymbol{\gamma}/\sqrt{n}+\boldsymbol{\beta})_1^\top \mathbf{X}_i|\boldsymbol{\gamma}/\sqrt{n}+\boldsymbol{\beta}] - g(\boldsymbol{\beta}_1^\top \mathbf{X}_i|\boldsymbol{\beta}) &= \frac{\boldsymbol{\gamma}}{\sqrt{n}}\nabla_{\mathbf{b}}g(\mathbf{b}_1^\top \mathbf{X}_i|\mathbf{b})\Big|_{\mathbf{b}=\boldsymbol{\beta}} + O_p(n^{-1})\\ &= \frac{\boldsymbol{\gamma}}{\sqrt{n}}g'(\boldsymbol{\beta}_1^\top \mathbf{X}_i|\boldsymbol{\beta})[\mathbf{X}_{i,-1} - E(\mathbf{X}_{-1}|\boldsymbol{\beta}_1^\top \mathbf{X})],\end{aligned}$$

where the last equality follows under the Single Index model, we get $\widetilde{A}_n(\tau,\boldsymbol{\gamma}) = (1/2)\boldsymbol{\gamma}^\top\mathbb{V}\boldsymbol{\gamma} + \mathbf{W}_n^\top\boldsymbol{\gamma} + r_n(\tau,\boldsymbol{\gamma})$, where $r_n(\tau,\boldsymbol{\gamma}) = o_p(1)$. Finally, noting that $\mathbf{W}_n$ has bounded second moment (see Lemma C.3) and hence is stochastically bounded, the convex function $\widetilde{A}_n(\tau,\boldsymbol{\gamma}) - \mathbf{W}_n^\top\boldsymbol{\gamma}$ converges in probability to the convex function $(1/2)\boldsymbol{\gamma}^\top\mathbb{V}\boldsymbol{\gamma}$. Therefore, it follows from the convexity lemma (Pollard, 1991) that for any compact set K, $\sup_{\boldsymbol{\gamma}\in K}|r_n(\tau,\boldsymbol{\gamma})| = o_p(1)$. Thus, the quadratic approximation to the convex function $\widetilde{A}_n(\tau,\boldsymbol{\gamma})$ holds uniformly for $\boldsymbol{\gamma}$ in a compact set.

Lemma C.3 *Let $\mathbf{W}_n^* = -n^{-1/2}\mathbf{W}_n$, where $\mathbf{W}_n$ defined in (C.6). Then, under the assumptions of Lemma C.2,*

$$\Pr\left\{\sqrt{n}\{[\tau(1-\tau)]^{-1/2}\Sigma^{-1}\mathbf{W}_n^*\} \le t|\mathbb{X}\right\} = \Phi(t) + o_p(1),$$

where $\Sigma = \mathrm{E}\left\{[g'(\boldsymbol{\beta}_1^\top\mathbf{X}|\boldsymbol{\beta})]^2[\mathbf{X}_{-1} - \mathrm{E}(\mathbf{X}_{-1}|\boldsymbol{\beta}_1^\top\mathbf{X})][\mathbf{X}_{-1} - \mathrm{E}(\mathbf{X}_{-1}|\boldsymbol{\beta}_1^\top\mathbf{X})]^\top\right\}$ and $\Phi(t)$ denotes the standard normal cumulative distribution function.

Proof. The proof uses the same steps as those in the proof of Lemma C.7 of Christou and Akritas (2016). We outline here the basic steps.

Let H define the class of functions as described in the proof of Lemma C.2 and define $\mathbf{Z}_i(\eta) = \rho'_\tau[e_i(\boldsymbol{\beta},\eta)]g'(\boldsymbol{\beta}_1^\top\mathbf{X}_i|\boldsymbol{\beta})[\mathbf{X}_{i,-1} - \mathrm{E}(\mathbf{X}_{-1}|\boldsymbol{\beta}_1^\top\mathbf{X})]$, where $e_i(\boldsymbol{\beta},\eta) = Y_i - \eta(\boldsymbol{\beta}_1^\top\mathbf{X}_i|\boldsymbol{\beta})$, and let $\mathbf{T}_i(\eta) = \mathbf{Z}_i(\eta) - \mathrm{E}[\mathbf{Z}_i(\eta)|\mathbb{X}]$. Using the Berry-Esseen theorem (Berry 1941, and Esseen 1942), we can show that $n^{-1/2}\sum_{i=1}^n \mathbf{T}_i(\eta)$ converges to a multivariate normal distribution, uniformly in $\eta \in H$; see Christou and Akritas (2016) for details. Specifically, for any $\mathbf{t} \in \mathbb{R}^{d-1}$, and conditionally on the design matrix $\mathbb{X}$,

$$\left|\Pr\left[\frac{\sum_{i=1}^n \mathbf{t}^\top\mathbf{T}_i(\eta)}{\sqrt{\sum_{i=1}^n \sigma_i^2(\eta)}} \le t \middle| \mathbb{X}\right] - \Phi(t)\right| \le C_0\left[\sum_{i=1}^n \sigma_i^2(\eta)\right]^{-3/2}\sum_{i=1}^n \rho_i(\eta),$$

where $\sigma_i^2(\eta) = \mathrm{Var}[\mathbf{t}^\top\mathbf{T}_i(\eta)|\mathbb{X}]$ and $\rho_i(\eta) = \mathrm{E}[|\mathbf{t}^\top\mathbf{T}_i(\eta)|^3|\mathbb{X}] < \infty$. Noting that

$$\sup_{\eta\in H}\left|\frac{1}{n\sqrt{n}}\sum_{i=1}^n \rho_i(\eta)\right| \le \frac{1}{n\sqrt{n}}\sum_{i=1}^n \sup_{\eta\in H}|\rho_i(\eta)| = o(1) \quad \text{(C.10)}$$

a.s., and

$$\sup_{\eta \in H} \left| \frac{1}{n} \sum_{i=1}^{n} \sigma_i^2(\eta) - \widetilde{v}^2(\eta) \right| = o(1) \tag{C.11}$$

a.s., where

$$\begin{aligned} \widetilde{v}^2(\eta) &= \mathbf{t}^\top \mathrm{E}\Big\{ F_{\epsilon|\mathbf{X}}[\eta(\boldsymbol{\beta}_1^\top \mathbf{X}|\boldsymbol{\beta}) - g(\boldsymbol{\beta}_1^\top \mathbf{X}|\boldsymbol{\beta})|\mathbf{X}]\{1 - F_{\epsilon|\mathbf{X}}[\eta(\boldsymbol{\beta}_1^\top \mathbf{X}|\boldsymbol{\beta}) - g(\boldsymbol{\beta}_1^\top \mathbf{X}|\boldsymbol{\beta})|\mathbf{X}]\} \\ & \quad [g'(\boldsymbol{\beta}_1^\top \mathbf{X}|\boldsymbol{\beta})]^2 [\mathbf{X}_{-1} - \mathrm{E}(\mathbf{X}_{-1}|\boldsymbol{\beta}_1^\top \mathbf{X})][\mathbf{X}_{-1} - \mathrm{E}(\mathbf{X}_{-1}|\boldsymbol{\beta}_1^\top \mathbf{X})]^\top \Big\} \mathbf{t}, \end{aligned}$$

we have, conditionally on $\mathbb{X}$,

$$\left| \Pr\left[\frac{\sum_{i=1}^{n} \mathbf{t}^\top \mathbf{T}_i(\eta)}{\sqrt{\sum_{i=1}^{n} \sigma_i^2(\eta)}} \le t \Big| \mathbb{X} \right] - \Phi(t) \right| = o_p(1), \tag{C.12}$$

uniformly in $\eta \in H$. Since (C.12) holds uniformly in $\eta \in H$, it also holds for $\eta = \widehat{g}$, where

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathrm{E}\left[\mathbf{t}^\top \mathbf{Z}_i(\widehat{g}) \big| \mathbb{X} \right] = o_p(1) \text{ and } \frac{1}{n} \sum_{i=1}^{n} \sigma_i^2(\widehat{g}) = \mathbf{t}^\top \tau(1-\tau)\Sigma \mathbf{t} + o_p(1). \tag{C.13}$$

Therefore, using (C.12), (C.13), and Slutsky's theorem, we get that, conditionally on $\mathbb{X}$, $\sqrt{n}\mathbf{W}_n^* \xrightarrow{d} \mathcal{N}(0, \tau(1-\tau)\Sigma)$, where the unconditional case follows from the Dominated Convergence theorem and the almost sure convergence of (C.10) and (C.11).

Lemma C.4 *Let $\widehat{g}(t|\mathbf{b})$ denote either $\widehat{g}^{NW}(t|\mathbf{b})$, defined in (2.3), or $\widehat{g}_{VS}^{NW}(t|\mathbf{b})$, defined in (2.6), and let $\widehat{\boldsymbol{\beta}}$ to be*

$$\widehat{\boldsymbol{\beta}} = \arg\min_{\mathbf{b} \in \Theta} \left\{ \sum_{i=1}^{n} \rho_\tau[Y_i - \widehat{g}(\mathbf{b}_1^\top \mathbf{X}_i|\mathbf{b})] + n \sum_{j=2}^{d} p_\lambda(|b_j|) \right\}, \tag{C.14}$$

where $\mathbf{b}_1 = (1, \mathbf{b}^\top)^\top = (1, b_2, \ldots, b_d)^\top$, and $\lambda \to 0$ as $n \to \infty$. Then, under the assumptions of Lemma C.2, $\widehat{\boldsymbol{\beta}}$ is $\sqrt{n}$-consistent estimator of $\boldsymbol{\beta}$. Moreover, for $\widehat{\boldsymbol{\beta}} = (\widehat{\boldsymbol{\beta}}_{11}^\top, \widehat{\boldsymbol{\beta}}_{12}^\top)^\top$, where $\widehat{\boldsymbol{\beta}}_{11}$ is of cardinality $(\widehat{d}^* - 1) = card(\{j \in (2, \ldots, d) : \widehat{\beta}_j \neq 0\})$, and for $\sqrt{n}\lambda \to \infty$ as $n \to \infty$, we have that, with probability tending to one,

1. ***Sparsity:*** $\widehat{\boldsymbol{\beta}}_{12} = 0$ *and*

2. ***Asymptotic Normality:*** $\sqrt{n}(\widehat{\boldsymbol{\beta}}_{11} - \boldsymbol{\beta}_{11}) \xrightarrow{d} \mathcal{N}(0, \tau(1-\tau)\mathbb{V}_{11}^{-1}\Sigma_{11}\mathbb{V}_{11}^{-1})$, *where* $\mathbb{V}_{11}$ *and* Σ_{11} *are defined in (3.1) and (3.2) respectively.*

Proof. To study the asymptotic properties of $\widehat{\boldsymbol{\beta}}$ defined in (C.14), we consider an equivalent objective function. Observe that by adding and subtracting the quantity $\widehat{g}(\boldsymbol{\beta}_1^\top \mathbf{X}_i|\boldsymbol{\beta})$ in the first part of the objective function (C.14), we get

$$\sum_{i=1}^n \rho_\tau \left\{Y_i^* - [\widehat{g}(\mathbf{b}_1^\top \mathbf{X}_i|\mathbf{b}) - \widehat{g}(\boldsymbol{\beta}_1^\top \mathbf{X}_i|\boldsymbol{\beta})]\right\} = \sum_{i=1}^n \rho_\tau [Y_i^* - \widetilde{g}(\mathbf{X}_i|\mathbf{b}, \boldsymbol{\beta})], \tag{C.15}$$

where $Y_i^* = Y_i - \widehat{g}(\boldsymbol{\beta}_1^\top \mathbf{X}_i|\boldsymbol{\beta})$ and, for any $\boldsymbol{\gamma} \in \mathbb{R}^{d-1}$ such that $\boldsymbol{\gamma} + \boldsymbol{\beta} \in \Theta$, we define $\widetilde{g}(\mathbf{X}_i|\boldsymbol{\gamma} + \boldsymbol{\beta}, \boldsymbol{\beta}) = \widehat{g}[(\boldsymbol{\gamma}+\boldsymbol{\beta})_1^\top \mathbf{X}_i|\boldsymbol{\gamma}+\boldsymbol{\beta}] - \widehat{g}(\boldsymbol{\beta}_1^\top \mathbf{X}_i|\boldsymbol{\beta})$, where according to the convention used, $(\boldsymbol{\gamma}+\boldsymbol{\beta})_1 = (1, (\boldsymbol{\gamma}+\boldsymbol{\beta})^\top)^\top$. For the sake of convenience in the derivation of the asymptotic results we replace relation (C.15) with $\sum_{i=1}^n \left\{\rho_\tau[Y_i^* - \widetilde{g}(\mathbf{X}_i|\mathbf{b}, \boldsymbol{\beta})] - \rho_\tau(Y_i^*)\right\}$ and we define the new objective function

$$\widehat{A}_n(\tau, \boldsymbol{\gamma}) = \widetilde{A}_n(\tau, \boldsymbol{\gamma}) + n\sum_{j=2}^d p_\lambda(|\gamma_j/\sqrt{n} + \beta_j|),$$

where $\boldsymbol{\gamma} = \sqrt{n}(\mathbf{b} - \boldsymbol{\beta})$, and $\widetilde{A}_n(\tau, \boldsymbol{\gamma})$ is defined in (C.4).

For the proof we use the same strategy as in Wu and Liu (2009). To prove the $\sqrt{n}$-consistency of $\widehat{\boldsymbol{\beta}}$, enough to show that for any given $\delta > 0$, there exists a constant C such that

$$\Pr\left[\inf_{\|\boldsymbol{\gamma}\| \geq C} \widehat{A}_n(\tau, \boldsymbol{\gamma}) > \widehat{A}_n(\tau, \mathbf{0})\right] \geq 1 - \delta, \tag{C.16}$$

since this implies that with probability at least $1-\delta$ there exists a local minimum in the ball $\{\boldsymbol{\gamma}/\sqrt{n} + \boldsymbol{\beta} : \|\boldsymbol{\gamma}\| \leq C\}$. Write

$$\begin{aligned}
\widehat{A}_n(\tau, \boldsymbol{\gamma}) - \widehat{A}_n(\tau, \mathbf{0}) &= \widetilde{A}_n(\tau, \boldsymbol{\gamma}) - \widetilde{A}_n(\tau, \mathbf{0}) + n\sum_{j=2}^d \left[p_\lambda(|\gamma_j/\sqrt{n} + \beta_j|) - p_\lambda(|\beta_j|)\right] \\
&\geq \widetilde{A}_n(\tau, \boldsymbol{\gamma}) - \widetilde{A}_n(\tau, \mathbf{0}) + n\sum_{j=2}^{d^*} \left[p_\lambda(|\gamma_j/\sqrt{n} + \beta_j|) - p_\lambda(|\beta_j|)\right],
\end{aligned}$$

where, for large n,

$$n\sum_{j=2}^{d^*}\left[p_\lambda(|\gamma_j/\sqrt{n}+\beta_j|)-p_\lambda(|\beta_j|)\right]=0. \tag{C.17}$$

This follows from (a) $|\beta_j|>0$ for $j=2,\ldots,d^*$, (b) the SCAD penalty is flat for arguments of magnitude larger than $a\lambda$, and (c) $\lambda\to 0$. Following, Lemma C.2 yields that

$$\widetilde{A}_n(\tau,\boldsymbol{\gamma})-\widetilde{A}_n(\tau,\mathbf{0})=\frac{1}{2}\boldsymbol{\gamma}^\top\mathbb{V}\boldsymbol{\gamma}+\mathbf{W}_n^\top\boldsymbol{\gamma}+o_p(1), \tag{C.18}$$

where $\mathbb{V}$ and $\mathbf{W}_n$ are defined in (C.5) and (C.6) respectively, for any $\boldsymbol{\gamma}$ in a compact subset of $\mathbb{R}^{d-1}$. Therefore, the difference (C.18) is dominated by the quadratic term $(1/2)\boldsymbol{\gamma}^\top\mathbb{V}\boldsymbol{\gamma}$ for $\|\boldsymbol{\gamma}\|$ greater than or equal to sufficiently large C. Using (C.17) and (C.18), the difference $\widehat{A}_n(\tau,\boldsymbol{\gamma})-\widehat{A}_n(\tau,\mathbf{0})$ is also dominated by the quadratic term $(1/2)\boldsymbol{\gamma}^\top\mathbb{V}\boldsymbol{\gamma}$ for $\|\boldsymbol{\gamma}\|$ greater than or equal to sufficiently large C, and (C.16) follows.

Next, we will show the sparsity part. To prove that, with probability tending to one, $\widehat{\boldsymbol{\beta}}_{12}=0$, we will show that for any given $\widetilde{\boldsymbol{\beta}}_{11}$ satisfying $\left\|\widetilde{\boldsymbol{\beta}}_{11}-\boldsymbol{\beta}_{11}\right\|=O_p(n^{-1/2})$ and any constant C,

$$\widehat{A}_n[\tau,\sqrt{n}((\widetilde{\boldsymbol{\beta}}_{11}-\boldsymbol{\beta}_{11})^\top,\mathbf{0}^\top)^\top]=\min_{\left\|\widetilde{\boldsymbol{\beta}}_{12}\right\|\le Cn^{-1/2}}\widehat{A}_n[\tau,\sqrt{n}((\widetilde{\boldsymbol{\beta}}_{11}-\boldsymbol{\beta}_{11})^\top,\widetilde{\boldsymbol{\beta}}_{12}^\top)^\top]. \tag{C.19}$$

Write

$$\begin{aligned}
&\widehat{A}_n[\tau,\sqrt{n}((\widetilde{\boldsymbol{\beta}}_{11}-\boldsymbol{\beta}_{11})^\top,\mathbf{0}^\top)^\top]-\widehat{A}_n[\tau,\sqrt{n}((\widetilde{\boldsymbol{\beta}}_{11}-\boldsymbol{\beta}_{11})^\top,\widetilde{\boldsymbol{\beta}}_{12}^\top)^\top]\\
=\;&\widetilde{A}_n[\tau,\sqrt{n}((\widetilde{\boldsymbol{\beta}}_{11}-\boldsymbol{\beta}_{11})^\top,\mathbf{0}^\top)^\top]-\widetilde{A}_n[\tau,\sqrt{n}((\widetilde{\boldsymbol{\beta}}_{11}-\boldsymbol{\beta}_{11})^\top,\widetilde{\boldsymbol{\beta}}_{12}^\top)^\top]-n\sum_{j=d^*+1}^{d}p_\lambda(|\widetilde{\beta}_j|)\\
=\;&\frac{1}{2}\sqrt{n}((\widetilde{\boldsymbol{\beta}}_{11}-\boldsymbol{\beta}_{11})^\top,\mathbf{0}^\top)\mathbb{V}\sqrt{n}((\widetilde{\boldsymbol{\beta}}_{11}-\boldsymbol{\beta}_{11})^\top,\mathbf{0}^\top)^\top+\mathbf{W}_n^\top\sqrt{n}((\widetilde{\boldsymbol{\beta}}_{11}-\boldsymbol{\beta}_{11})^\top,\mathbf{0}^\top)^\top\\
&-\frac{1}{2}\sqrt{n}((\widetilde{\boldsymbol{\beta}}_{11}-\boldsymbol{\beta}_{11})^\top,\widetilde{\boldsymbol{\beta}}_{12}^\top)\mathbb{V}\sqrt{n}((\widetilde{\boldsymbol{\beta}}_{11}-\boldsymbol{\beta}_{11})^\top,\widetilde{\boldsymbol{\beta}}_{12}^\top)^\top-\mathbf{W}_n^\top\sqrt{n}((\widetilde{\boldsymbol{\beta}}_{11}-\boldsymbol{\beta}_{11})^\top,\widetilde{\boldsymbol{\beta}}_{12}^\top)^\top\\
&-n\sum_{j=d^*+1}^{d}p_\lambda(|\widetilde{\beta}_j|),
\end{aligned}$$

where the last equality follows from the quadratic approximation derived in Lemma C.2, and $\mathbb{V}$ and

$\mathbf{W}_n$ are defined in (C.5) and (C.6) respectively. Using the facts that $\left\|\widetilde{\boldsymbol{\beta}}_{11} - \boldsymbol{\beta}_{11}\right\| = O_p(n^{-1/2})$ and $0 < \left\|\widetilde{\boldsymbol{\beta}}_{12}\right\| \leq Cn^{-1/2}$, we get that

$$\frac{1}{2}\sqrt{n}((\widetilde{\boldsymbol{\beta}}_{11} - \boldsymbol{\beta}_{11})^\top, \mathbf{0}^\top)\mathbb{V}\sqrt{n}((\widetilde{\boldsymbol{\beta}}_{11} - \boldsymbol{\beta}_{11})^\top, \mathbf{0}^\top)^\top = O_p(1), \tag{C.20}$$

$$\frac{1}{2}\sqrt{n}((\widetilde{\boldsymbol{\beta}}_{11} - \boldsymbol{\beta}_{11})^\top, \widetilde{\boldsymbol{\beta}}_{12}^\top)\mathbb{V}\sqrt{n}((\widetilde{\boldsymbol{\beta}}_{11} - \boldsymbol{\beta}_{11})^\top, \widetilde{\boldsymbol{\beta}}_{12}^\top)^\top = O_p(1) \tag{C.21}$$

and

$$\mathbf{W}_n^\top\sqrt{n}((\widetilde{\boldsymbol{\beta}}_{11} - \boldsymbol{\beta}_{11})^\top, \mathbf{0}^\top)^\top - \mathbf{W}_n^\top\sqrt{n}((\widetilde{\boldsymbol{\beta}}_{11} - \boldsymbol{\beta}_{11})^\top, \widetilde{\boldsymbol{\beta}}_{12}^\top)^\top \;=\; -\sqrt{n}(\mathbf{0}^\top, \widetilde{\boldsymbol{\beta}}_{12}^\top)\mathbf{W}_n = O_p(\sqrt{n}), \tag{C.22}$$

where the last equality follows from the asymptotic normality result derived in Lemma C.3. Therefore, using relation $n\sum_{j=d^*+1}^{d} p_\lambda(|\widetilde{\beta}_j|) \geq n\lambda\left(\sum_{j=d^*+1}^{d}|\widetilde{\beta}_j|\right)[1+o(1)]$, (see Wu and Liu 2009, proof of Lemma 1, online supplement, page S24, for the proof), relations (C.20), (C.21), (C.22) and the facts that (a) $\sqrt{n}\lambda \to \infty$ and (b) the term $n\lambda = \sqrt{n}(\sqrt{n}\lambda)$ is of higher order than $\sqrt{n}$, we get that the difference $\widehat{A}_n[\tau, \sqrt{n}((\widetilde{\boldsymbol{\beta}}_{11} - \boldsymbol{\beta}_{11})^\top, \mathbf{0}^\top)^\top] - \widehat{A}_n[\tau, \sqrt{n}((\widetilde{\boldsymbol{\beta}}_{11} - \boldsymbol{\beta}_{11})^\top, \widetilde{\boldsymbol{\beta}}_{12}^\top)^\top]$ is dominated by $-n\sum_{j=d^*+1}^{d} p_\lambda(|\widetilde{\beta}_j|)$. Hence, (C.19) follows.

Finally, we will show the asymptotic normality part. The $\sqrt{n}$-consistency of $\widehat{\boldsymbol{\beta}}$ yields that there exists a $\sqrt{n}$-consistent minimizer $\widehat{\boldsymbol{\beta}}_{11}$ of $\widehat{A}_n[\tau, \sqrt{n}((\mathbf{b}_{11} - \boldsymbol{\beta}_{11})^\top, \mathbf{0}^\top)^\top]$. Thus, define $\widehat{\boldsymbol{\gamma}}_{11} = \sqrt{n}(\widehat{\boldsymbol{\beta}}_{11} - \boldsymbol{\beta}_{11})$ to be the minimizer of

$$\widehat{A}_n[\tau, (\boldsymbol{\gamma}_{11}^\top, \mathbf{0}^\top)^\top] = \widetilde{A}_n[\tau, (\boldsymbol{\gamma}_{11}^\top, \mathbf{0}^\top)^\top] + n\sum_{j=2}^{d^*} p_\lambda(|\gamma_j/\sqrt{n} + \beta_j|). \tag{C.23}$$

The quadratic approximation derived in Lemma C.2 yields that

$$\begin{aligned}\widetilde{A}_n[\tau, (\boldsymbol{\gamma}_{11}^\top, \mathbf{0}^\top)^\top] \;&=\; \frac{1}{2}(\boldsymbol{\gamma}_{11}^\top, \mathbf{0}^\top)\mathbb{V}(\boldsymbol{\gamma}_{11}^\top, \mathbf{0}^\top)^\top + \mathbf{W}_n^\top(\boldsymbol{\gamma}_{11}^\top, \mathbf{0}^\top)^\top + o_p(1)\\ &=\; \frac{1}{2}\boldsymbol{\gamma}_{11}^\top\mathbb{V}_{11}\boldsymbol{\gamma}_{11} + \mathbf{W}_{n,1}^\top\boldsymbol{\gamma}_{11} + o_p(1),\end{aligned} \tag{C.24}$$

where $\mathbb{V}_{11}$ is defined in (3.1) and

$$\mathbf{W}_{n,1} = -n^{-1/2}\sum_{i=1}^{n}\rho'_\tau(Y_i^*)g'(\boldsymbol{\beta}_1^\top\mathbf{X}_i|\boldsymbol{\beta})[\mathbf{X}_{i1,-1} - \mathrm{E}(\mathbf{X}_{1,-1}|\boldsymbol{\beta}_1^\top\mathbf{X})],$$

for $Y_i^* = Y_i - \widehat{g}(\boldsymbol{\beta}_1^\top\mathbf{X}_i|\boldsymbol{\beta})$. Therefore, for large n, and using relations (C.17) and (C.24), the objective function $\widehat{A}_n[\tau,(\boldsymbol{\gamma}_{11}^\top,\mathbf{0}^\top)^\top]$ in (C.23) can be written as

$$\widehat{A}_n[\tau,(\boldsymbol{\gamma}_{11}^\top,\mathbf{0}^\top)^\top] = \frac{1}{2}\boldsymbol{\gamma}_{11}^\top\mathbb{V}_{11}\boldsymbol{\gamma}_{11} + \mathbf{W}_{n,1}^\top\boldsymbol{\gamma}_{11} + o_p(1) + n\sum_{j=2}^{d^*}p_\lambda(|\beta_j|),$$

where the last term does not depend on $\boldsymbol{\gamma}_{11}$. Thus, for large n, the minimizer $\widehat{\boldsymbol{\gamma}}_{11}$ is only $o_p(1)$ away from $\widehat{\boldsymbol{\gamma}}_{11}^* = \mathbb{V}_{11}^{-1}\mathbf{W}_{n,1}$. Therefore, the asymptotic normality of $\mathbf{W}_{n,1}$, which is a direct consequence of Lemma C.3, yields $\sqrt{n}(\widehat{\boldsymbol{\beta}}_{11} - \boldsymbol{\beta}_{11}) \xrightarrow{d} \mathcal{N}(0,\tau(1-\tau)\mathbb{V}_{11}^{-1}\Sigma_{11}\mathbb{V}_{11}^{-1})$, where Σ_{11} is defined in (3.2).

References

[1] Berry, A. C. (1941). The Accuracy of the Gaussian Approximation to the Sum of Independent Variates. *Transactions of the American Mathematical Society* **49**(1) 122–136.

[2] Christou, E. and Akritas, M. G. (2016). Single index quantile regression for heteroscedastic data. *Journal of Multivariate Analysis* **150** 169–182.

[3] Esseen, C. G. (1942). On the Liapunoff limit of error in the theory of probability. *Arkiv for matematik astronomi och fysik* **A28** 1–19.

[4] Guerre, E. and Sabbah, C. (2012). Uniform bias study and Bahadur representation for local polynomial estimators of the conditional quantile function. *Econometric Theory* **28**(01) 87–129.

[5] Hansen, B. (2008). Uniform Convergence Rates for kernel estimation with dependent data. *Econometric Theory* **24** 726–748.

[6] Jennrich, R. I. (1969). Asymptotic Properties of Non-Linear Least Squares Estimators. *The Annals of Mathematical Statistics* **40** 633–646.

[7] Pollard, D. (1991). Asymptotics for least absolute deviation regression estimators. *Econometric Theory* **7**(2) 186–199.

[8] Wu, Y. and Liu, Y. (2009). Variable Selection in Quantile Regression. *Statistica Sinica* **19** 801–817.