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

Appendix C. SOME LEMMAS

LEMMA C.1. *(Corollary 1 (ii)-*Guerre and Sabbah *[3]): 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 $\widehat{Q}_\tau(Y|\mathbf{X}_i)$ defined in connection with (2.5). Under the Assumptions GS1-GS3, given in the Appendix A, and for all τ in an interval $[\underline{\tau}, \overline{\tau}]$,*

$$\sup_{\mathbf{x}\in\mathcal{X}_0} |\widehat{Q}_\tau(Y|\mathbf{x}) - Q_\tau(Y|\mathbf{x})| = O_p\left(\frac{\ln n}{n}\right)^{s/(2s+d)} = O_p(a_n^*),$$

if h_1 (used in (2.5)) is asymptotically proportional to $(\ln n/n)^{1/(2s+d)}$.

Proof: See Guerre and Sabbah [3]. ■

LEMMA C.2. *(Convexity Lemma -* Pollard *[6]): Let $\{A_n(\mathbf{u}) : \mathbf{u} \in U\}$ be a sequence of real valued random convex functions defined on a convex, open subset $U \in \mathbb{R}^d$. Suppose $A(\mathbf{u})$ is a real-valued function on U for which $A_n(\mathbf{u}) \to A(\mathbf{u})$ in probability, for each $\mathbf{u} \in U$. Then, for each compact subset K of U,*

$$\sup_{\mathbf{u}\in K} |A_n(\mathbf{u}) - A(\mathbf{u})| \xrightarrow{p} 0.$$

The function $A(\cdot)$ is necessarily convex on U.

Proof: See Pollard [6]. ■

LEMMA C.3. *(Quadratic Approximation Lemma -* Hjort and Pollard *[4]): Suppose $A_n(\mathbf{u})$ is convex and can be represented as $(1/2)\mathbf{u}^\top\mathbb{V}\mathbf{u} + \mathbf{U}_n^\top\mathbf{u} + C_n + r_n(\mathbf{u})$, where $\mathbb{V}$ is symmetric and positive definite, $\mathbf{U}_n$ is stochastically bounded, C_n is arbitrary, and $r_n(\mathbf{u})$ goes to zero in probability for each $\mathbf{u}$. Then, $\boldsymbol{\alpha}_n$, the argmin of A_n is only $o_p(1)$ away from $\boldsymbol{\beta}_n = -\mathbb{V}^{-1}\mathbf{U}_n$, the argmin of $(1/2)\mathbf{u}^\top\mathbb{V}\mathbf{u} + \mathbf{U}_n^\top\mathbf{u} + C_n$. If also $\mathbf{U}_n \xrightarrow{d} \mathbf{U}$, then $\boldsymbol{\alpha}_n \xrightarrow{d} -\mathbb{V}^{-1}\mathbf{U}$.*

Proof: See Hjort and Pollard [4]. ■

LEMMA C.4. *(Uniform Law of Large Numbers for Triangular Arrays): Suppose (a) Θ is compact, (b) $g(\mathbf{X}_{ni}, \theta)$ is continuous at each $\theta \in \Theta$ with probability one, (c) $g(\mathbf{X}_{ni}, \theta)$ is dominated by a function $G(\mathbf{X}_{ni})$, i.e. $|g(\mathbf{X}_{ni}, \theta)| \leq G(\mathbf{X}_{ni})$, and (d) $\sup_n \mathrm{E}G(\mathbf{X}_{ni}) < \infty$. Then,*

$$\sup_{\theta\in\Theta}\left|\frac{1}{n}\sum_{i=1}^{n}\{g(\mathbf{X}_{ni}, \theta) - \mathrm{E}g(\mathbf{X}_{ni}, \theta)\}\right| \xrightarrow{p} 0.$$

Proof: See Jennrich [5]. ■

LEMMA C.5. *Let $\widehat{g}^{NW}(t|\mathbf{b})$ be as defined in (2.4). Then, under the assumptions of Proposition 3.1, and the condition $nh_2^4 = o(1)$, where h is the bandwidth in (2.4), we have that for any $\mathbf{b}$,*

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{\widehat{g}^{NW}(\mathbf{b}_1^\top\mathbf{X}_i|\mathbf{b}) - g(\mathbf{b}_1^\top\mathbf{X}_i|\mathbf{b})\right\} = o_p(1).$$

Proof: Let $K_{2,h_2}(\cdot) = K_2(\cdot/h_2)$, $\widehat{f}_{\mathbf{b}}(t) = (nh_2)^{-1}\sum_{i=1}^n K_{2,h_2}\left(t - \mathbf{b}_1^\top\mathbf{X}_i\right)$, and write

$$\widehat{T}_n = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\widehat{g}^{NW}(\mathbf{b}_1^\top\mathbf{X}_i|\mathbf{b}) = \frac{1}{n^{1.5}h_2}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\widehat{Q}_\tau(Y|\mathbf{X}_j)K_{2,h_2}\{\mathbf{b}_1^\top(\mathbf{X}_i - \mathbf{X}_j)\}}{\widehat{f}_{\mathbf{b}}(\mathbf{b}_1^\top\mathbf{X}_i)}.$$

The proof consists of two steps. In the first step it will be shown that

$$\widehat{T}_n - T_n = o_p(1), \tag{C.1}$$

where

$$T_n = \frac{1}{n^{1.5}h_2}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{Q_\tau(Y|\mathbf{X}_j)K_{2,h_2}\{\mathbf{b}_1^\top(\mathbf{X}_i - \mathbf{X}_j)\}}{f_{\mathbf{b}}(\mathbf{b}_1^\top\mathbf{X}_i)},$$

and in the second step it will be shown that

$$T_n - \frac{1}{\sqrt{n}}\sum_{i=1}^{n} g(\mathbf{b}_1^\top\mathbf{X}_i|\mathbf{b}) = o_p(1). \tag{C.2}$$

To show (C.1) write

$$\begin{aligned}
\widehat{T}_n - T_n &= \widehat{T}_n - T_n \pm \frac{1}{n^{1.5}h_2}\sum_{i=1}^{n}\sum_{j=1}^{n} Q_\tau(Y|\mathbf{X}_j)\frac{K_{2,h_2}\{\mathbf{b}_1^\top(\mathbf{X}_i - \mathbf{X}_j)\}}{\widehat{f}_{\mathbf{b}}(\mathbf{b}_1^\top\mathbf{X}_i)} \\
&= \frac{1}{n^{1.5}h_2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left\{\widehat{Q}_\tau(Y|\mathbf{X}_j) - Q_\tau(Y|\mathbf{X}_j)\right\}\frac{K_{2,h_2}\{\mathbf{b}_1^\top(\mathbf{X}_i - \mathbf{X}_j)\}}{\widehat{f}_{\mathbf{b}}(\mathbf{b}_1^\top\mathbf{X}_i)} \\
&\quad - \frac{1}{n^{1.5}h_2}\sum_{i=1}^{n}\sum_{j=1}^{n} Q_\tau(Y|\mathbf{X}_j)K_{2,h_2}\{\mathbf{b}_1^\top(\mathbf{X}_i - \mathbf{X}_j)\}\left\{\frac{1}{f_{\mathbf{b}}(\mathbf{b}_1^\top\mathbf{X}_i)} - \frac{1}{\widehat{f}_{\mathbf{b}}(\mathbf{b}_1^\top\mathbf{X}_i)}\right\}
\end{aligned} \tag{C.3}$$

The first term on the right hand side of (C.3) can be shown to be $o_p(1)$ through the Bahadur representation of $\widehat{Q}_\tau(Y|\mathbf{X}_j) - Q_\tau(Y|\mathbf{X}_j)$ (see Guerre and Sabbah [3]). Indeed, it is straightforward to see that the order of the remainder term of the representation is sufficiently small, while the part resulting from the leading term of the representation can be shown to be $o_p(1)$. Moreover, by writing the second term on the right hand side of (C.3) as

$$\begin{aligned}
&\frac{1}{n^{1.5}h_2}\sum_{i=1}^{n}\sum_{j=1}^{n} Q_\tau(Y|\mathbf{X}_j)K_{2,h_2}\{\mathbf{b}_1^\top(\mathbf{X}_i - \mathbf{X}_j)\}\left\{\frac{\widehat{f}_{\mathbf{b}}(\mathbf{b}_1^\top\mathbf{X}_i) - f_{\mathbf{b}}(\mathbf{b}_1^\top\mathbf{X}_i)}{f_{\mathbf{b}}(\mathbf{b}_1^\top\mathbf{X}_i)\widehat{f}_{\mathbf{b}}(\mathbf{b}_1^\top\mathbf{X}_i)} \pm \frac{\widehat{f}_{\mathbf{b}}(\mathbf{b}_1^\top\mathbf{X}_i) - f_{\mathbf{b}}(\mathbf{b}_1^\top\mathbf{X}_i)}{f_{\mathbf{b}}^2(\mathbf{b}_1^\top\mathbf{X}_i)}\right\} \\
= &\frac{1}{n^{1.5}h_2}\sum_{i=1}^{n}\sum_{j=1}^{n} Q_\tau(Y|\mathbf{X}_j)K_{2,h_2}\{\mathbf{b}_1^\top(\mathbf{X}_i - \mathbf{X}_j)\}\frac{\{\widehat{f}_{\mathbf{b}}(\mathbf{b}_1^\top\mathbf{X}_i) - f_{\mathbf{b}}(\mathbf{b}_1^\top\mathbf{X}_i)\}^2}{f_{\mathbf{b}}^2(\mathbf{b}_1^\top\mathbf{X}_i)\widehat{f}_{\mathbf{b}}(\mathbf{b}_1^\top\mathbf{X}_i)} \\
&+\frac{1}{n^{1.5}h_2}\sum_{i=1}^{n}\sum_{j=1}^{n} Q_\tau(Y|\mathbf{X}_j)K_{2,h_2}\{\mathbf{b}_1^\top(\mathbf{X}_i - \mathbf{X}_j)\}\frac{\widehat{f}_{\mathbf{b}}(\mathbf{b}_1^\top\mathbf{X}_i) - f_{\mathbf{b}}(\mathbf{b}_1^\top\mathbf{X}_i)}{f_{\mathbf{b}}^2(\mathbf{b}_1^\top\mathbf{X}_i)},
\end{aligned}$$

it is easily seen that the first term on the right hand side is $o_p(1)$, while straightforward second moment calculations and the condition $\sqrt{n}h_2^2 = o(1)$ reveal that the second term is $o_p(1)$. This shows (C.1). Next, to show (C.2), we write its left hand side as

$$
\begin{aligned}
& T_n - \frac{1}{\sqrt{n}} \sum_{i=1}^n g(\mathbf{b}_1^\top \mathbf{X}_i|\mathbf{b}) \pm \frac{1}{n^{1.5}h_2} \sum_{i=1}^n \sum_{j=1}^n g(\mathbf{b}_1^\top \mathbf{X}_i|\mathbf{b}) \frac{K_{2,h_2}\{\mathbf{b}_1^\top(\mathbf{X}_i - \mathbf{X}_j)\}}{f_{\mathbf{b}}(\mathbf{b}_1^\top \mathbf{X}_i)} \\
= \; & \frac{1}{n^{1.5}h_2} \sum_{i=1}^n \sum_{j=1}^n \left\{Q_\tau(Y|\mathbf{X}_j) - g(\mathbf{b}_1^\top \mathbf{X}_i|\mathbf{b})\right\} \frac{K_{2,h_2}\{\mathbf{b}_1^\top(\mathbf{X}_i - \mathbf{X}_j)\}}{f_{\mathbf{b}}(\mathbf{b}_1^\top \mathbf{X}_i)} \\
& + \frac{1}{\sqrt{n}} \sum_{i=1}^n \left[\frac{1}{nh_2} \sum_{j=1}^n g(\mathbf{b}_1^\top \mathbf{X}_i|\mathbf{b}) \frac{K_{2,h_2}\{\mathbf{b}_1^\top(\mathbf{X}_i - \mathbf{X}_j)\}}{f_{\mathbf{b}}(\mathbf{b}_1^\top \mathbf{X}_i)} - g(\mathbf{b}_1^\top \mathbf{X}_i|\mathbf{b}) \right] \\
= \; & \frac{1}{n^{1.5}} \binom{n}{2} U_n \\
& + \frac{1}{n^{1.5}h_2} \sum_{i=1}^n \sum_{j=1}^n g(\mathbf{b}_1^\top \mathbf{X}_i|\mathbf{b}) K_{2,h_2}\{\mathbf{b}_1^\top(\mathbf{X}_i - \mathbf{X}_j)\} \left\{ \frac{1}{f_{\mathbf{b}}(\mathbf{b}_1^\top \mathbf{X}_i)} - \frac{1}{\widehat{f}_{\mathbf{b}}(\mathbf{b}_1^\top \mathbf{X}_i)} \right\}, \quad \text{(C.4)}
\end{aligned}
$$

where

$$
U_n = \frac{1}{\binom{n}{2}} \sum_{i<j} \phi(\mathbf{X}_i, \mathbf{X}_j),
$$

for

$$
\begin{aligned}
\phi(\mathbf{X}_i, \mathbf{X}_j) \; = \; & \frac{1}{h_2} \left\{Q_\tau(Y|\mathbf{X}_j) - g(\mathbf{b}_1^\top \mathbf{X}_i|\mathbf{b})\right\} \frac{K_{2,h_2}\{\mathbf{b}_1^\top(\mathbf{X}_i - \mathbf{X}_j)\}}{f_{\mathbf{b}}(\mathbf{b}_1^\top \mathbf{X}_i)} \\
& + \frac{1}{h_2} \left\{Q_\tau(Y|\mathbf{X}_i) - g(\mathbf{b}_1^\top \mathbf{X}_j|\mathbf{b})\right\} \frac{K_{2,h_2}\{\mathbf{b}_1^\top(\mathbf{X}_i - \mathbf{X}_j)\}}{f_{\mathbf{b}}(\mathbf{b}_1^\top \mathbf{X}_j)}.
\end{aligned}
$$

The second term of relation (C.4) can be shown to be $o_p(1)$ using similar steps as those in the proof of the second term of relation (C.3), while for the first term we use U-statistic techniques and we get

$$
\begin{aligned}
\phi_1(\mathbf{X}_i) \; = \; & E\Big[\mathrm{E}\{\phi(\mathbf{X}_i, \mathbf{X}_j)|\mathbf{b}_1^\top \mathbf{X}_j, \mathbf{b}_1^\top \mathbf{X}_i\} \Big| \mathbf{X}_i \Big] \\
= \; & \mathrm{E}\left[\frac{1}{h_2} \left\{g(\mathbf{b}_1^\top \mathbf{X}_j|\mathbf{b}) - g(\mathbf{b}_1^\top \mathbf{X}_i|\mathbf{b})\right\} \frac{K_{2,h_2}\{\mathbf{b}_1^\top(\mathbf{X}_i - \mathbf{X}_j)\}}{f_{\mathbf{b}}(\mathbf{b}_1^\top \mathbf{X}_i)} \Big| \mathbf{X}_i \right] \\
& + \mathrm{E}\left[\frac{1}{h_2} \left\{g(\mathbf{b}_1^\top \mathbf{X}_i|\mathbf{b}) - g(\mathbf{b}_1^\top \mathbf{X}_j|\mathbf{b})\right\} \frac{K_{2,h_2}\{\mathbf{b}_1^\top(\mathbf{X}_i - \mathbf{X}_j)\}}{f_{\mathbf{b}}(\mathbf{b}_1^\top \mathbf{X}_j)} \Big| \mathbf{X}_i \right] \\
= \; & \frac{1}{h_2} \int \left\{g(\mathbf{b}_1^\top \mathbf{X}_j|\mathbf{b}) - g(\mathbf{b}_1^\top \mathbf{X}_i|\mathbf{b})\right\} \frac{K_{2,h_2}\{\mathbf{b}_1^\top(\mathbf{X}_i - \mathbf{x}_j)\}}{f_{\mathbf{b}}(\mathbf{b}_1^\top \mathbf{X}_i)} f_{\mathbf{b}}(\mathbf{b}_1^\top \mathbf{X}_j) d(\mathbf{b}_1^\top \mathbf{X}_j) \\
& + \frac{1}{h_2} \int \left\{g(\mathbf{b}_1^\top \mathbf{X}_i|\mathbf{b}) - g(\mathbf{b}_1^\top \mathbf{X}_j|\mathbf{b})\right\} K_{2,h_2}\{\mathbf{b}_1^\top(\mathbf{X}_i - \mathbf{x}_j)\} d(\mathbf{b}_1^\top \mathbf{X}_j) \\
= \; & \int \left\{g(uh_2 + \mathbf{b}_1^\top \mathbf{X}_i|\mathbf{b}) - g(\mathbf{b}_1^\top \mathbf{X}_i|\mathbf{b})\right\} \frac{K_2(u)}{f_{\mathbf{b}}(\mathbf{b}_1^\top \mathbf{X}_i)} f_{\mathbf{b}}(uh_2 + \mathbf{b}_1^\top \mathbf{X}_i) du \\
& + \int \left\{g(\mathbf{b}_1^\top \mathbf{X}_i|\mathbf{b}) - g(uh_2 + \mathbf{b}_1^\top \mathbf{X}_i|\mathbf{b})\right\} K_2(u) du \\
= \; & O_p(h_2^2),
\end{aligned}
$$

uniformly in $\mathbf{X}_i$ by the condition that the support of $\mathbf{X}$ is bounded. Thus, the expected value of the first term on the right hand side of (C.4) is $\binom{n}{2}n^{-1.5}O_p(h_2^2) = O_p(\sqrt{n}h_2^2) = o_p(1)$. Moreover, the general formula for the variance of U-statistics gives

$$\mathrm{Var}\left\{\frac{1}{n^{1.5}}\binom{n}{2}U_n\right\} = O(n)\mathrm{Var}(U_n) = O(n)\mathrm{Var}\left[\binom{n}{2}^{-1}\left\{2(n-2)\sigma_1^2 + \sigma_2^2\right\}\right] = o(1),$$

where $\sigma_1^2 = \mathrm{Var}\{\phi_1(\mathbf{X})\}$ and $\sigma_2^2 = \mathrm{Var}\{\phi(\mathbf{X}_1, \mathbf{X}_2)\}$. Therefore, (C.2) follows. ■

LEMMA C.6. *Let $A_n(\tau, \boldsymbol{\gamma})$ defined in (B.2) for $\boldsymbol{\gamma} \in \mathbb{R}^{d-1}$ such that $\boldsymbol{\gamma} + \boldsymbol{\beta} \in \Theta$. Then, under the assumptions of Proposition 3.2, we have the following quadratic approximation, uniformly in $\boldsymbol{\gamma}$ in a compact set,*

$$A_n\left(\tau, \boldsymbol{\gamma}\right) = \frac{1}{2}\boldsymbol{\gamma}^\top \mathbb{V}\boldsymbol{\gamma} + \mathbf{W}_n^\top \boldsymbol{\gamma} + o_p(1), \tag{C.5}$$

where $\mathbb{V}$ and $\mathbf{W}_n$ are defined in (3.1) and (3.2) respectively.

Proof: For the proof 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 $g(t|\boldsymbol{\beta})$, and, according to Proposition 3.1, includes $\widehat{g}^{NW}(t|\boldsymbol{\beta})$ for n large enough, almost surely. Define,

$$A_n(\eta, \tau, \boldsymbol{\gamma}) = \sum_{i=1}^{n}\left[\rho_\tau\{e_i(\boldsymbol{\beta}, \eta) - \widetilde{\eta}(\mathbf{X}_i|\boldsymbol{\gamma}/\sqrt{n} + \boldsymbol{\beta}, \boldsymbol{\beta})\} - \rho_\tau\{e_i(\boldsymbol{\beta}, \eta)\}\right], \tag{C.6}$$

where $e_i(\boldsymbol{\beta}, \eta) = Y_i - \eta(\boldsymbol{\beta}_1^\top \mathbf{X}_i|\boldsymbol{\beta})$ and $\widetilde{\eta}(\mathbf{X}_i|\boldsymbol{\gamma}/\sqrt{n} + \boldsymbol{\beta}, \boldsymbol{\beta}) = \eta\{(\boldsymbol{\gamma}/\sqrt{n} + \boldsymbol{\beta})_1^\top \mathbf{X}_i|\boldsymbol{\gamma}/\sqrt{n} + \boldsymbol{\beta}\} - \eta(\boldsymbol{\beta}_1^\top \mathbf{X}_i|\boldsymbol{\beta})$, where, according to the convention used, $(\boldsymbol{\gamma}/\sqrt{n}+\boldsymbol{\beta})_1 = (1, (\boldsymbol{\gamma}/\sqrt{n}+\boldsymbol{\beta})^\top)^\top$. Showing that the quadratic approximation holds uniform in $\eta \in H$, it also holds for replacing η with $\widehat{g}^{NW}$, where, after the replacement, (C.6) reduces to the objective function defined in (B.2).

Write $A_n\left(\eta, \tau, \boldsymbol{\gamma}\right)$ as

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

where $\mathbb{X}$ denotes the design matrix, and $R_n(\eta, \tau, \boldsymbol{\gamma})$ is the remainder term defined by (C.7). Let

$\varphi(t|\mathbf{x}) = \mathrm{E}\{\rho_\tau(\epsilon+t)|\mathbf{X}=\mathbf{x}\}$ as defined in Assumption A6, and observe that

$$
\begin{aligned}
&\mathrm{E}\left\{A_n(\eta,\tau,\boldsymbol{\gamma})\,|\mathbb{X}\right\} = \sum_{i=1}^n \mathrm{E}\left[\rho_\tau\left\{e_i(\boldsymbol{\beta},\eta) - \widetilde{\eta}\left(\mathbf{X}_i|\boldsymbol{\gamma}/\sqrt{n}+\boldsymbol{\beta},\boldsymbol{\beta}\right)\right\} - \rho_\tau\{e_i(\boldsymbol{\beta},\eta)\}\Big|\mathbb{X}\right] \\
= &\sum_{i=1}^n \mathrm{E}\left[\rho_\tau\left\{\epsilon_i + g(\boldsymbol{\beta}_1^\top\mathbf{X}_i|\boldsymbol{\beta}) - \eta(\boldsymbol{\beta}_1^\top\mathbf{X}_i|\boldsymbol{\beta}) - \widetilde{\eta}\left(\mathbf{X}_i|\boldsymbol{\gamma}/\sqrt{n}+\boldsymbol{\beta},\boldsymbol{\beta}\right)\right\}\right. \\
&\left.-\rho_\tau\{\epsilon_i + g(\boldsymbol{\beta}_1^\top\mathbf{X}_i|\boldsymbol{\beta}) - \eta(\boldsymbol{\beta}_1^\top\mathbf{X}_i|\boldsymbol{\beta})\}\Big|\mathbb{X}\right] \\
= &\sum_{i=1}^n \left[\varphi\left\{g(\boldsymbol{\beta}_1^\top\mathbf{X}_i|\boldsymbol{\beta}) - \eta(\boldsymbol{\beta}_1^\top\mathbf{X}_i|\boldsymbol{\beta}) - \widetilde{\eta}\left(\mathbf{X}_i|\boldsymbol{\gamma}/\sqrt{n}+\boldsymbol{\beta},\boldsymbol{\beta}\right)|\mathbb{X}\right\}\right. \\
&\quad\left.-\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\}\right] \\
= &-\sum_{i=1}^n \widetilde{\eta}(\mathbf{X}_i|\boldsymbol{\gamma}/\sqrt{n}+\boldsymbol{\beta},\boldsymbol{\beta})\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\} \\
&+\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\} \\
&-\frac{1}{6}\sum_{i=1}^n \{\widetilde{\eta}(\mathbf{X}_i|\boldsymbol{\gamma}/\sqrt{n}+\boldsymbol{\beta},\boldsymbol{\beta})\}^3\varphi'''\left\{g(\boldsymbol{\beta}_1^\top\mathbf{X}_i|\boldsymbol{\beta}) - \eta(\boldsymbol{\beta}_1^\top\mathbf{X}_i|\boldsymbol{\beta}) + u_i|\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) \\
&+\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), \qquad \text{(C.8)}
\end{aligned}
$$

for u_i in a neighborhood of $-|\widetilde{\eta}(\mathbf{X}_i|\boldsymbol{\gamma}/\sqrt{n}+\boldsymbol{\beta},\boldsymbol{\beta})|$ and $|\widetilde{\eta}(\mathbf{X}_i|\boldsymbol{\gamma}/\sqrt{n}+\boldsymbol{\beta},\boldsymbol{\beta})|$, where the last equality uses $\varphi'\{g(\boldsymbol{\beta}_1^\top\mathbf{X}_i|\boldsymbol{\beta}) - \eta(\boldsymbol{\beta}_1^\top\mathbf{X}_i|\boldsymbol{\beta})|\mathbb{X}\} = \mathrm{E}[\rho'_\tau\{e_i(\boldsymbol{\beta},\eta)\}|\mathbb{X}]$. The last term of relation (C.8) is $o_p(1)$ uniformly in $\eta \in H$, which follows by noting that

$$
\begin{aligned}
\widetilde{\eta}(\mathbf{X}_i|\boldsymbol{\gamma}/\sqrt{n}+\boldsymbol{\beta},\boldsymbol{\beta}) &= \eta\{(\boldsymbol{\gamma}/\sqrt{n}+\boldsymbol{\beta})_1^\top\mathbf{X}_i|\boldsymbol{\gamma}/\sqrt{n}+\boldsymbol{\beta}\} - \eta(\boldsymbol{\beta}_1^\top\mathbf{X}_i|\boldsymbol{\beta}) \\
&= \frac{\boldsymbol{\gamma}^\top}{\sqrt{n}}\nabla_{\mathbf{b}}\eta(\mathbf{b}_1^\top\mathbf{X}_i|\mathbf{b})\Big|_{\mathbf{b}=\boldsymbol{\beta}+\mathbf{t}_n}, \qquad \text{(C.9)}
\end{aligned}
$$

for $\|\mathbf{t}_n\|$ been between $-\|\boldsymbol{\gamma}\|/\sqrt{n}$ and $\|\boldsymbol{\gamma}\|/\sqrt{n}$, and writing

$$
\begin{aligned}
&\sup_{\eta\in H}\left|\sum_{i=1}^n \{\widetilde{\eta}(\mathbf{X}_i|\boldsymbol{\gamma}/\sqrt{n}+\boldsymbol{\beta},\boldsymbol{\beta})\}^3\varphi'''\left\{g(\boldsymbol{\beta}_1^\top\mathbf{X}_i|\boldsymbol{\beta}) - \eta(\boldsymbol{\beta}_1^\top\mathbf{X}_i|\boldsymbol{\beta}) + u_i|\mathbb{X}\right\}\right| \\
&\le \sum_{i=1}^n \sup_{\eta\in H}\left|\frac{\boldsymbol{\gamma}^\top}{\sqrt{n}}\nabla_{\mathbf{b}}\eta(\mathbf{b}_1^\top\mathbf{X}_i|\mathbf{b})\Big|_{\mathbf{b}=\boldsymbol{\beta}+t_n}\right|^3 \left|\varphi'''\left\{g(\boldsymbol{\beta}_1^\top\mathbf{X}_i|\boldsymbol{\beta}) - \eta(\boldsymbol{\beta}_1^\top\mathbf{X}_i|\boldsymbol{\beta}) + u_i|\mathbb{X}\right\}\right| \\
&= o_p(1),
\end{aligned}
$$

where the last equality follows by the bounded partial derivatives of η and the boundedness of $\varphi'''(\cdot|\mathbf{x})$.

Following, we will 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). Obviously $R_n(\eta,\tau,\boldsymbol{\gamma})$ is centered and can be written as

$$R_n(\eta,\tau,\boldsymbol{\gamma}) = \sum_{i=1}^{n}\left[R_{ni}(\eta,\tau,\boldsymbol{\gamma}) - \mathrm{E}\{R_{ni}(\eta,\tau,\boldsymbol{\gamma})|\mathbb{X}\}\right],$$

where

$$\begin{aligned} R_{ni}(\eta,\tau,\boldsymbol{\gamma}) &= \rho_\tau\{e_i(\boldsymbol{\beta},\eta) - \widetilde{\eta}(\mathbf{X}_i|\boldsymbol{\gamma}/\sqrt{n}+\boldsymbol{\beta},\boldsymbol{\beta})\} - \rho_\tau\{e_i(\boldsymbol{\beta},\eta)\} \\ &\quad + \rho'_\tau\{e_i(\boldsymbol{\beta},\eta)\}\widetilde{\eta}(\mathbf{X}_i|\boldsymbol{\gamma}/\sqrt{n}+\boldsymbol{\beta},\boldsymbol{\beta}). \end{aligned}$$

The uniformity will follow from the Uniform Law of Large Numbers for Triangular Arrays, restated in Lemma C.4, by considering $nR_{ni}(\eta,\tau,\boldsymbol{\gamma})$ which is continuous in η with probability one. Note that, by using the definition of $\rho_\tau(\cdot)$ and $\rho'_\tau(\cdot)$, $R_{ni}(\eta,\tau,\boldsymbol{\gamma})$ can be equivalently written as

$$\begin{aligned} R_{ni}(\eta,\tau,\boldsymbol{\gamma}) &= \left\{e_i(\boldsymbol{\beta},\eta) - \widetilde{\eta}(\mathbf{X}_i|\boldsymbol{\gamma}/\sqrt{n}+\boldsymbol{\beta},\boldsymbol{\beta})\right\} \\ &\quad \times \left[I\left\{e_i(\boldsymbol{\beta},\eta) < 0\right\} - I\left\{e_i(\boldsymbol{\beta},\eta) < \widetilde{\eta}(\mathbf{X}_i|\boldsymbol{\gamma}/\sqrt{n}+\boldsymbol{\beta},\boldsymbol{\beta})\right\}\right] \end{aligned}$$

which is bounded by

$$\begin{aligned} |R_{ni}(\eta,\tau,\boldsymbol{\gamma})| &\le \left|e_i(\boldsymbol{\beta},\eta) - \widetilde{\eta}(\mathbf{X}_i|\boldsymbol{\gamma}/\sqrt{n}+\boldsymbol{\beta},\boldsymbol{\beta})\right| I\{|e_i(\boldsymbol{\beta},\eta)| \le |\widetilde{\eta}(\mathbf{X}_i|\boldsymbol{\gamma}/\sqrt{n}+\boldsymbol{\beta},\boldsymbol{\beta})|\} \\ &\le 2|\widetilde{\eta}(\mathbf{X}_i|\boldsymbol{\gamma}/\sqrt{n}+\boldsymbol{\beta},\boldsymbol{\beta})| I\{|e_i(\boldsymbol{\beta},\eta)| \le |\widetilde{\eta}(\mathbf{X}_i|\boldsymbol{\gamma}/\sqrt{n}+\boldsymbol{\beta},\boldsymbol{\beta})|\}, \end{aligned}$$

where the first inequality follows from the fact that

$$\begin{aligned} &\left|I\left\{e_i(\boldsymbol{\beta},\eta) < 0\right\} - I\left\{e_i(\boldsymbol{\beta},\eta) < \widetilde{\eta}(\mathbf{X}_i|\boldsymbol{\gamma}/\sqrt{n}+\boldsymbol{\beta},\boldsymbol{\beta})\right\}\right| \\ &\le I\{-|\widetilde{\eta}(\mathbf{X}_i|\boldsymbol{\gamma}/\sqrt{n}+\boldsymbol{\beta},\boldsymbol{\beta})| \le e_i(\boldsymbol{\beta},\eta) < |\widetilde{\eta}(\mathbf{X}_i|\boldsymbol{\gamma}/\sqrt{n}+\boldsymbol{\beta},\boldsymbol{\beta})|\}. \end{aligned}$$

Following, note that $nR_{ni}(\eta,\tau,\boldsymbol{\gamma})$ is dominated by

$$\begin{aligned} |nR_{ni}(\eta,\tau,\boldsymbol{\gamma})| &\le 2n\left|\widetilde{\eta}(\mathbf{X}_i|\boldsymbol{\gamma}/\sqrt{n}+\boldsymbol{\beta},\boldsymbol{\beta})\right| I\{|e_i(\boldsymbol{\beta},\eta)| \le |\widetilde{\eta}(\mathbf{X}_i|\boldsymbol{\gamma}/\sqrt{n}+\boldsymbol{\beta},\boldsymbol{\beta})|\} \\ &\overset{(C.9)}{=} 2n\left|\frac{\boldsymbol{\gamma}^\top}{\sqrt{n}}\nabla_{\mathbf{b}}\eta(\mathbf{b}_1^\top\mathbf{X}_i|\mathbf{b})\Big|_{\mathbf{b}=\boldsymbol{\beta}+\boldsymbol{t}_n}\right| \\ &\quad \times I\left\{|e_i(\boldsymbol{\beta},\eta)| \le \left|\frac{\boldsymbol{\gamma}^\top}{\sqrt{n}}\nabla_{\mathbf{b}}\eta(\mathbf{b}_1^\top\mathbf{X}_i|\mathbf{b})\Big|_{\mathbf{b}=\boldsymbol{\beta}+\boldsymbol{t}_n}\right|\right\} \\ &\le 2\sqrt{n}\sup_{\eta\in H}\left|\boldsymbol{\gamma}^\top\nabla_{\mathbf{b}}\eta(\mathbf{b}_1^\top\mathbf{X}_i|\mathbf{b})\Big|_{\mathbf{b}=\boldsymbol{\beta}+\boldsymbol{t}_n}\right| \\ &\quad \times I\left\{|e_i(\boldsymbol{\beta},\eta)| \le \sup_{\eta\in H}\left|\frac{\boldsymbol{\gamma}^\top}{\sqrt{n}}\nabla_{\mathbf{b}}\eta(\mathbf{b}_1^\top\mathbf{X}_i|\mathbf{b})\Big|_{\mathbf{b}=\boldsymbol{\beta}+\boldsymbol{t}_n}\right|\right\} \\ &\le 2\sqrt{n}\sup_{\eta\in H}\left|\boldsymbol{\gamma}^\top\nabla_{\mathbf{b}}\eta(\mathbf{b}_1^\top\mathbf{X}_i|\mathbf{b})\Big|_{\mathbf{b}=\boldsymbol{\beta}+\boldsymbol{t}_n}\right| \\ &\quad \times I\left\{\left|\epsilon_i - \eta_i^*(\boldsymbol{\beta}) + g(\boldsymbol{\beta}_1^\top\mathbf{X}_i|\boldsymbol{\beta})\right| \le \sup_{\eta\in H}\left|\frac{\boldsymbol{\gamma}^\top}{\sqrt{n}}\nabla_{\mathbf{b}}\eta(\mathbf{b}_1^\top\mathbf{X}_i|\mathbf{b})\Big|_{\mathbf{b}=\boldsymbol{\beta}+\boldsymbol{t}_n}\right|\right\} \\ &= \widetilde{R}_{ni}(\tau,\boldsymbol{\gamma}), \end{aligned}$$

where $\eta_i^*(\boldsymbol{\beta}) = \arg\inf_{\eta\in H}\left|\epsilon_i - \eta(\boldsymbol{\beta}_1^\top \mathbf{X}_i|\boldsymbol{\beta}) + g(\boldsymbol{\beta}_1^\top \mathbf{X}_i|\boldsymbol{\beta})\right|$. Then, using the fact that

$$\begin{aligned}
& \mathrm{E}\left[I\left\{|\epsilon_i - \eta_i^*(\boldsymbol{\beta}) + g(\boldsymbol{\beta}_1^\top \mathbf{X}_i|\boldsymbol{\beta})| \leq \sup_{\eta\in H}\left|\frac{\boldsymbol{\gamma}^\top}{\sqrt{n}}\nabla_{\mathbf{b}}\eta(\mathbf{b}_1^\top \mathbf{X}_i|\mathbf{b})\Big|_{\mathbf{b}=\boldsymbol{\beta}+\boldsymbol{t}_n}\right|\right\}\Big|\mathbb{X}\right] \\
= \;& F_{\epsilon|\mathbf{X}_i}\left\{\eta_i^*(\boldsymbol{\beta}) - g(\boldsymbol{\beta}_1^\top \mathbf{X}_i|\boldsymbol{\beta}) + \sup_{\eta\in H}\left|\frac{\boldsymbol{\gamma}^\top}{\sqrt{n}}\nabla_{\mathbf{b}}\eta(\mathbf{b}_1^\top \mathbf{X}_i|\mathbf{b})\Big|_{\mathbf{b}=\boldsymbol{\beta}+\boldsymbol{t}_n}\right|\Big|\mathbf{X}_i\right\} \\
& - F_{\epsilon|\mathbf{X}_i}\left\{\eta_i^*(\boldsymbol{\beta}) - g(\boldsymbol{\beta}_1^\top \mathbf{X}_i|\boldsymbol{\beta}) - \sup_{\eta\in H}\left|\frac{\boldsymbol{\gamma}^\top}{\sqrt{n}}\nabla_{\mathbf{b}}\eta(\mathbf{b}_1^\top \mathbf{X}_i|\mathbf{b})\Big|_{\mathbf{b}=\boldsymbol{\beta}+\boldsymbol{t}_n}\right|\Big|\mathbf{X}_i\right\} \\
= \;& O(n^{-1/2}),
\end{aligned}$$

almost surely, it is easy to see that $\sup_n \mathrm{E}\{\widetilde{R}_{ni}(\tau,\boldsymbol{\gamma})|\mathbb{X}\} = O(1)$, almost surely, and therefore,

$$\begin{aligned}
& \sup_{\eta\in H}\left|\frac{1}{n}\sum_{i=1}^n [nR_{ni}(\eta,\tau,\boldsymbol{\gamma}) - \mathrm{E}\{nR_{ni}(\eta,\tau,\boldsymbol{\gamma})|\mathbb{X}\}]\right| \\
= \;& \sup_{\eta\in H}\left|\sum_{i=1}^n [R_{ni}(\eta,\tau,\boldsymbol{\gamma}) - \mathrm{E}\{R_{ni}(\eta,\tau,\boldsymbol{\gamma})|\mathbb{X}\}]\right| = o_p(1).
\end{aligned}$$

Next, substituting the expression of $\mathrm{E}\{A_n(\eta,\tau,\boldsymbol{\gamma})|\mathbb{X}\}$ 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 $\eta\in H$,

$$\begin{aligned}
A_n(\eta,\tau,\boldsymbol{\gamma}) &= \frac{1}{2}\sum_{i=1}^n \{\widetilde{\eta}(\mathbf{X}_i|\boldsymbol{\gamma}/\sqrt{n}+\boldsymbol{\beta},\boldsymbol{\beta})\}^2 \varphi''\{g(\boldsymbol{\beta}_1^\top \mathbf{X}_i|\boldsymbol{\beta}) - \eta(\boldsymbol{\beta}_1^\top \mathbf{X}_i|\boldsymbol{\beta})|\mathbb{X}\} \\
& \quad - \sum_{i=1}^n \rho_\tau'\{e_i(\boldsymbol{\beta},\eta)\}\widetilde{\eta}(\mathbf{X}_i|\boldsymbol{\gamma}/\sqrt{n}+\boldsymbol{\beta},\boldsymbol{\beta}) + o_p(1). \qquad \text{(C.10)}
\end{aligned}$$

Since expression (C.10) holds uniformly in $\eta\in H$, where the class H includes $\widehat{g}^{NW}$, we substitute η with $\widehat{g}^{NW}$. Using (a) the fact that $A_n(\widehat{g}^{NW},\tau,\boldsymbol{\gamma})$ reduces to $A_n(\tau,\boldsymbol{\gamma})$ defined in (B.2), (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, and (c) relation

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

where the last equality follows under the Single Index model, we get

$$A_n(\tau,\boldsymbol{\gamma}) = \frac{1}{2}\boldsymbol{\gamma}^\top \mathbb{V}_n \boldsymbol{\gamma} + \mathbf{W}_n^\top \boldsymbol{\gamma} + r_n(\tau,\boldsymbol{\gamma}), \qquad \text{(C.11)}$$

where $r_n(\tau,\boldsymbol{\gamma}) = o_p(1)$,

$$\begin{aligned}
\mathbb{V}_n &= \frac{1}{n}\sum_{i=1}^n \{g'(\boldsymbol{\beta}_1^\top \mathbf{X}_i|\boldsymbol{\beta})\}^2 \left\{\mathbf{X}_{i,-1} - \mathrm{E}\left(\mathbf{X}_{-1}|\boldsymbol{\beta}_1^\top \mathbf{X}\right)\right\}\left\{\mathbf{X}_{i,-1} - \mathrm{E}\left(\mathbf{X}_{-1}|\boldsymbol{\beta}_1^\top \mathbf{X}\right)\right\}^\top \\
& \quad \times \varphi''\left\{g(\boldsymbol{\beta}_1^\top \mathbf{X}_i|\boldsymbol{\beta}) - \widehat{g}^{NW}(\boldsymbol{\beta}_1^\top \mathbf{X}_i|\boldsymbol{\beta})|\mathbb{X}\right\},
\end{aligned}$$

and $\mathbf{W}_n$ is defined in (3.2). Next, showing that $(1/2)\boldsymbol{\gamma}^\top \mathbb{V}_n \boldsymbol{\gamma} = (1/2)\boldsymbol{\gamma}^\top \mathbb{V}\boldsymbol{\gamma} + o_p(1)$, where $\mathbb{V}$ is defined in (3.1), relation (C.11) can be written as,

$$A_n(\tau, \boldsymbol{\gamma}) = \frac{1}{2}\boldsymbol{\gamma}^\top \mathbb{V}\boldsymbol{\gamma} + \mathbf{W}_n^\top \boldsymbol{\gamma} + o_p(1).$$

This is easy to prove by noting that $\varphi'(0|\mathbf{x}) = \tau - F_{\epsilon|\mathbf{X}}(0|\mathbf{x}) = 0$, $\varphi''(0|\mathbf{x}) = f_{\epsilon|\mathbf{X}}(0|\mathbf{x})$ (see the note below for a proof), and

$$\varphi''\{g(\boldsymbol{\beta}_1^\top \mathbf{X}|\boldsymbol{\beta}) - \widehat{g}^{NW}(\boldsymbol{\beta}_1^\top \mathbf{X}|\boldsymbol{\beta})|\mathbf{x}\} = \varphi''(0|\mathbf{x}) + o_p(a_n^* + a_n + h_2^2),$$

where the last equality follows from Proposition 3.1. Therefore,

$$\begin{aligned}
\mathbb{V}_n &= \frac{1}{n}\sum_{i=1}^n \{g'(\boldsymbol{\beta}_1^\top \mathbf{X}_i|\boldsymbol{\beta})\}^2 \{\mathbf{X}_{i,-1} - \mathrm{E}(\mathbf{X}_{-1}|\boldsymbol{\beta}_1^\top \mathbf{X})\}\{\mathbf{X}_{i,-1} - \mathrm{E}(\mathbf{X}_{-1}|\boldsymbol{\beta}_1^\top \mathbf{X})\}^\top \varphi''(0|\mathbf{X}_i) + o_p(1) \\
&= \frac{1}{n}\sum_{i=1}^n \{g'(\boldsymbol{\beta}_1^\top \mathbf{X}_i|\boldsymbol{\beta})\}^2 \{\mathbf{X}_{i,-1} - \mathrm{E}(\mathbf{X}_{-1}|\boldsymbol{\beta}_1^\top \mathbf{X})\}\{\mathbf{X}_{i,-1} - \mathrm{E}(\mathbf{X}_{-1}|\boldsymbol{\beta}_1^\top \mathbf{X})\}^\top f_{\epsilon|\mathbf{X}_i}(0|\mathbf{X}_i) + o_p(1) \\
&= \mathbb{V} + o_p(1).
\end{aligned}$$

Finally, noting that $\mathbf{W}_n$ has bounded second moment (see Lemma C.7), and hence is stochastically bounded, the convex function $A_n(\tau, \boldsymbol{\gamma}) - \mathbf{W}_n^\top \boldsymbol{\gamma}$ converges in probability to the convex function $(1/2)\boldsymbol{\gamma}^\top \mathbb{V}\boldsymbol{\gamma}$. Therefore, it follows from the convexity lemma (Pollard [6]), restated in Lemma C.2, 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 $A_n(\tau, \boldsymbol{\gamma})$ holds uniformly for $\boldsymbol{\gamma}$ in a compact set. ■

Note: The fact that $\varphi'(0|\mathbf{x}) = \tau - F_{\epsilon|\mathbf{X}}(0|\mathbf{x}) = 0$ and $\varphi''(0|\mathbf{x}) = f_{\epsilon|\mathbf{X}}(0|\mathbf{x})$ can be easily proved using simple calculations and Assumption A6,

$$\begin{aligned}
\varphi'(0|\mathbf{x}) &= \frac{\partial}{\partial s}\left\{\int \rho_\tau(z+s) f_{\epsilon|\mathbf{X}}(z|\mathbf{x})dz\right\}\Big|_{s=0} = \int \left\{\frac{\partial}{\partial s}\rho_\tau(z+s)\Big|_{s=0} f_{\epsilon|\mathbf{X}}(z|\mathbf{x})\right\} dz \\
&= \int \rho_\tau'(z) f_{\epsilon|\mathbf{X}}(z|\mathbf{x})dz = \int \{\tau - I(z<0)\} f_{\epsilon|\mathbf{X}}(z|\mathbf{x})dz = \tau - F_{\epsilon|\mathbf{X}}(0|\mathbf{x})
\end{aligned}$$

and

$$\varphi''(0|\mathbf{x}) = \frac{\partial}{\partial s}\varphi'(s|\mathbf{x})\Big|_{s=0} = \frac{\partial}{\partial s}\{\tau - F_{\epsilon|\mathbf{X}}(-s|\mathbf{x})\}\Big|_{s=0} = f_{\epsilon|\mathbf{X}}(0|\mathbf{x}).$$

LEMMA C.7. *Let $\mathbf{W}_n^* = -n^{-1/2}\mathbf{W}_n$, where $\mathbf{W}_n$ is defined in (3.2). Then, under the assumptions of Proposition 3.2,*

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

where Σ is defined in (3.3), and $\Phi(t)$ denotes the standard normal cumulative distribution function. where $\Phi(t)$ denotes the standard normal cumulative distribution function.

Proof: For the proof, we will use the same technique as in the proof of Lemma C.6, that is, 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 $e_i(\boldsymbol{\beta}, \eta) = Y_i - \eta(\boldsymbol{\beta}_1^\top \mathbf{X}_i|\boldsymbol{\beta})$ for

$\eta \in H$, and let $\mathbf{T}_i(\eta) = \mathbf{Z}_i(\eta) - \mathrm{E}\{\mathbf{Z}_i(\eta)|\mathbb{X}\}$. Using the Berry-Esseen theorem (Berry [1], and Esseen [2]), we will show that $n^{-1/2}\sum_{i=1}^n \mathbf{T}_i(\eta)$ converges to a multivariate normal distribution, uniformly in $\eta \in H$. To do this, use the Cramer-Wald theorem and, for any $\mathbf{t} \in \mathbb{R}^{d-1}$, consider $\mathbf{t}^\top \mathbf{T}_i(\eta) = \mathbf{t}^\top[\mathbf{Z}_i(\eta) - \mathrm{E}\{\mathbf{Z}_i(\eta)|\mathbb{X}\}]$, where

$$\begin{aligned}
\mathrm{E}\{\mathbf{Z}_i(\eta)|\mathbb{X}\} &= \mathrm{E}\left[\rho'_\tau\{e_i(\boldsymbol{\beta},\eta)\}|\mathbb{X}\right] g'(\boldsymbol{\beta}_1^\top \mathbf{X}_i|\boldsymbol{\beta})\{\mathbf{X}_{i,-1} - \mathrm{E}(\mathbf{X}_{-1}|\boldsymbol{\beta}_1^\top \mathbf{X})\} \\
&= \left[\tau - F_{\epsilon|\mathbf{X}_i}\{\eta(\boldsymbol{\beta}_1^\top \mathbf{X}_i|\boldsymbol{\beta}) - g(\boldsymbol{\beta}_1^\top \mathbf{X}_i|\boldsymbol{\beta})|\mathbf{X}_i\}\right] g'(\boldsymbol{\beta}_1^\top \mathbf{X}_i|\boldsymbol{\beta})\{\mathbf{X}_{i,-1} - \mathrm{E}(\mathbf{X}_{-1}|\boldsymbol{\beta}_1^\top \mathbf{X})\}.
\end{aligned}$$

Following, conditionally on the design matrix $\mathbb{X}$, $\mathbf{t}^\top \mathbf{T}_1(\eta), \ldots, \mathbf{t}^\top \mathbf{T}_n(\eta)$ are independent random variables, with $\mathrm{E}\{\mathbf{t}^\top \mathbf{T}_i(\eta)|\mathbb{X}\} = 0$ by definition, and

$$\begin{aligned}
\mathrm{Var}\{\mathbf{t}^\top \mathbf{T}_i(\eta)|\mathbb{X}\} &= \mathbf{t}^\top\{\mathbf{X}_{i,-1} - \mathrm{E}(\mathbf{X}_{-1}|\boldsymbol{\beta}_1^\top \mathbf{X})\}\mathrm{Var}[\rho'_\tau\{e_i(\boldsymbol{\beta},\eta)\}|\mathbb{X})]\{g'(\boldsymbol{\beta}_1^\top \mathbf{X}_i|\boldsymbol{\beta})\}^2 \\
&\quad \{\mathbf{X}_{i,-1} - \mathrm{E}(\mathbf{X}_{-1}|\boldsymbol{\beta}_1^\top \mathbf{X})\}^\top \mathbf{t} \\
&= \mathbf{t}^\top\{\mathbf{X}_{i,-1} - \mathrm{E}(\mathbf{X}_{-1}|\boldsymbol{\beta}_1^\top \mathbf{X})\}F_{\epsilon|\mathbf{X}_i}\{\eta(\boldsymbol{\beta}_1^\top \mathbf{X}_i|\boldsymbol{\beta}) - g(\boldsymbol{\beta}_1^\top \mathbf{X}_i|\boldsymbol{\beta})|\mathbf{X}_i\}\{g'(\boldsymbol{\beta}_1^\top \mathbf{X}_i|\boldsymbol{\beta})\}^2 \\
&\quad \times[1 - F_{\epsilon|\mathbf{X}_i}\{\eta(\boldsymbol{\beta}_1^\top \mathbf{X}_i|\boldsymbol{\beta}) - g(\boldsymbol{\beta}_1^\top \mathbf{X}_i|\boldsymbol{\beta})|\mathbf{X}_i\}]\{\mathbf{X}_{i,-1} - \mathrm{E}(\mathbf{X}_{-1}|\boldsymbol{\beta}_1^\top \mathbf{X})\}^\top \mathbf{t} \\
&= \sigma_i^2(\eta).
\end{aligned}$$

Moreover,

$$\begin{aligned}
\mathrm{E}\{|\mathbf{t}^\top \mathbf{T}_i(\eta)|^3|\mathbb{X}\} &= \mathrm{E}\left[\left|\mathbf{t}^\top[\mathbf{Z}_i(\eta) - \mathrm{E}\{\mathbf{Z}_i(\eta)|\mathbb{X}\}]\right|^3 \Big|\mathbb{X}\right] \\
&= \mathrm{E}\Big[\Big|\left[\rho'_\tau\{e_i(\boldsymbol{\beta},\eta)\} - \tau + F_{\epsilon|\mathbf{X}_i}\{\eta(\boldsymbol{\beta}_1^\top \mathbf{X}_i|\boldsymbol{\beta}) - g(\boldsymbol{\beta}_1^\top \mathbf{X}_i|\boldsymbol{\beta})|\mathbf{X}_i\}\right] g'(\boldsymbol{\beta}_1^\top \mathbf{X}_i|\boldsymbol{\beta}) \\
&\qquad \times \mathbf{t}^\top\{\mathbf{X}_{i,-1} - \mathrm{E}(\mathbf{X}_{-1}|\boldsymbol{\beta}_1^\top \mathbf{X})\}\Big|^3\Big|\mathbb{X}\Big] \\
&= \mathrm{E}\left[\left|F_{\epsilon|\mathbf{X}_i}\{\eta(\boldsymbol{\beta}_1^\top \mathbf{X}_i|\boldsymbol{\beta}) - g(\boldsymbol{\beta}_1^\top \mathbf{X}_i|\boldsymbol{\beta})|\mathbf{X}_i\} - I\{e_i(\boldsymbol{\beta},\eta) < 0\}\right|^3 \Big|\mathbb{X}\right] \\
&\quad \times\left|g'(\boldsymbol{\beta}_1^\top \mathbf{X}_i|\boldsymbol{\beta})\mathbf{t}^\top\{\mathbf{X}_{i,-1} - \mathrm{E}(\mathbf{X}_{-1}|\boldsymbol{\beta}_1^\top \mathbf{X})\}\right|^3 \\
&= \left[F^3_{\epsilon|\mathbf{X}_i}\{\eta(\boldsymbol{\beta}_1^\top \mathbf{X}_i|\boldsymbol{\beta}) - g(\boldsymbol{\beta}_1^\top \mathbf{X}_i|\boldsymbol{\beta})|\mathbf{X}_i\} + F_{\epsilon|\mathbf{X}_i}\{\eta(\boldsymbol{\beta}_1^\top \mathbf{X}_i|\boldsymbol{\beta}) - g(\boldsymbol{\beta}_1^\top \mathbf{X}_i|\boldsymbol{\beta})|\mathbf{X}_i\}\right] \\
&\quad \left[1 - F_{\epsilon|\mathbf{X}_i}\{\eta(\boldsymbol{\beta}_1^\top \mathbf{X}_i|\boldsymbol{\beta}) - g(\boldsymbol{\beta}_1^\top \mathbf{X}_i|\boldsymbol{\beta})|\mathbf{X}_i\}\right]\left|g'(\boldsymbol{\beta}_1^\top \mathbf{X}_i|\boldsymbol{\beta})\mathbf{t}^\top\{\mathbf{X}_{i,-1} - \mathrm{E}(\mathbf{X}_{-1}|\boldsymbol{\beta}_1^\top \mathbf{X})\}\right|^3 \\
&= \rho_i(\eta) < \infty,
\end{aligned}$$

where the equality before the last follows from the fact that

$$\begin{aligned}
&\mathrm{E}\left[\left|F_{\epsilon|\mathbf{X}_i}\left\{\eta(\boldsymbol{\beta}_1^\top \mathbf{X}_i|\boldsymbol{\beta}) - g(\boldsymbol{\beta}_1^\top \mathbf{X}_i|\boldsymbol{\beta})\big|\mathbf{X}_i\right\} - I\{e_i(\boldsymbol{\beta},\eta) < 0\}\right|^3 \Big|\mathbb{X}\right] \\
&= \int_{\eta(\boldsymbol{\beta}_1^\top \mathbf{X}_i|\boldsymbol{\beta}) - g(\boldsymbol{\beta}_1^\top \mathbf{X}_i|\boldsymbol{\beta})}^{\infty} F^3_{\epsilon|\mathbf{X}_i}\left\{\eta(\boldsymbol{\beta}_1^\top \mathbf{X}_i|\boldsymbol{\beta}) - g(\boldsymbol{\beta}_1^\top \mathbf{X}_i|\boldsymbol{\beta})\big|\mathbf{X}_i\right\} f_{\epsilon|\mathbf{X}_i}(\epsilon|\mathbf{X}_i)d\epsilon \\
&\quad - \int_{-\infty}^{\eta(\boldsymbol{\beta}_1^\top \mathbf{X}_i|\boldsymbol{\beta}) - g(\boldsymbol{\beta}_1^\top \mathbf{X}_i|\boldsymbol{\beta})} \left[F_{\epsilon|\mathbf{X}_i}\{\eta(\boldsymbol{\beta}_1^\top \mathbf{X}_i|\boldsymbol{\beta}) - g(\boldsymbol{\beta}_1^\top \mathbf{X}_i|\boldsymbol{\beta})|\mathbf{X}_i\} - 1\right]^3 f_{\epsilon|\mathbf{X}_i}(\epsilon|\mathbf{X}_i)d\epsilon \\
&= F^3_{\epsilon|\mathbf{X}_i}\left\{\eta(\boldsymbol{\beta}_1^\top \mathbf{X}_i|\boldsymbol{\beta}) - g(\boldsymbol{\beta}_1^\top \mathbf{X}_i|\boldsymbol{\beta})\big|\mathbf{X}_i\right\}\left[1 - F_{\epsilon|\mathbf{X}_i}\{\eta(\boldsymbol{\beta}_1^\top \mathbf{X}_i|\boldsymbol{\beta}) - g(\boldsymbol{\beta}_1^\top \mathbf{X}_i|\boldsymbol{\beta})|\mathbf{X}_i\}\right] \\
&\quad - \left[F_{\epsilon|\mathbf{X}_i}\{\eta(\boldsymbol{\beta}_1^\top \mathbf{X}_i|\boldsymbol{\beta}) - g(\boldsymbol{\beta}_1^\top \mathbf{X}_i|\boldsymbol{\beta})|\mathbf{X}_i\} - 1\right]^3 F_{\epsilon|\mathbf{X}_i}\{\eta(\boldsymbol{\beta}_1^\top \mathbf{X}_i|\boldsymbol{\beta}) - g(\boldsymbol{\beta}_1^\top \mathbf{X}_i|\boldsymbol{\beta})|\mathbf{X}_i\}.
\end{aligned}$$

Then, by ESSEEN [2], conditionally on the design matrix $\mathbb{X}$,

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

where it is easy to see that

$$\sup_{\eta\in H}\left|\frac{1}{n\sqrt{n}}\sum_{i=1}^{n}\rho_i(\eta)\right|\le\frac{1}{n\sqrt{n}}\sum_{i=1}^{n}\sup_{\eta\in H}|\rho_i(\eta)|=o(1), \tag{C.12}$$

a.s. and

$$\sup_{\eta\in H}\left|\frac{1}{n}\sum_{i=1}^{n}\sigma_i^2(\eta)-\widetilde{v}^2(\eta)\right|=o(1) \tag{C.13}$$

a.s., where

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

The uniformity of (C.13) follows from the Uniform Strong Law of Large Numbers (JENNRICH [5]) since $\sigma_i^2(\eta)$ is dominated by $\mathbf{t}^{\top}\{g'(\boldsymbol{\beta}_1^{\top}\mathbf{X}_i|\boldsymbol{\beta})\}^2\{\mathbf{X}_{i,-1}-\mathrm{E}(\mathbf{X}_{-1}|\boldsymbol{\beta}_1^{\top}\mathbf{X})\}\{\mathbf{X}_{i,-1}-\mathrm{E}(\mathbf{X}_{-1}|\boldsymbol{\beta}_1^{\top}\mathbf{X})\}^{\top}\mathbf{t}$, which has a bounded expectation given by $\mathbf{t}^{\top}\Sigma\mathbf{t}$, where Σ is defined in (3.3). Therefore, conditionally on $\mathbb{X}$,

$$\left|\Pr\left\{\frac{\sum_{i=1}^{n}\mathbf{t}^{\top}\mathbf{T}_i(\eta)}{\sqrt{\sum_{i=1}^{n}\sigma_i^2(\eta)}}\le t\Big|\mathbb{X}\right\}-\Phi(t)\right|=o_p(1), \tag{C.14}$$

uniformly in $\eta\in H$. Since (C.14) holds uniformly in $\eta\in H$, it also holds for $\eta=\widehat{g}^{NW}$, where

$$\begin{aligned}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathrm{E}\left\{\mathbf{t}^{\top}\mathbf{Z}_i(\widehat{g}^{NW})|\mathbb{X}\right\} \quad &= \quad \mathbf{t}^{\top}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[\tau-F_{\epsilon|\mathbf{X}_i}\{\widehat{g}^{NW}(\boldsymbol{\beta}_1^{\top}\mathbf{X}_i|\boldsymbol{\beta})-g(\boldsymbol{\beta}_1^{\top}\mathbf{X}_i|\boldsymbol{\beta})|\mathbf{X}_i\}\right] \\ & \qquad \times g'(\boldsymbol{\beta}_1^{\top}\mathbf{X}_i|\boldsymbol{\beta})\{\mathbf{X}_{i,-1}-\mathrm{E}(\mathbf{X}_{-1}|\boldsymbol{\beta}_1^{\top}\mathbf{X})\} \\ &= \quad o_p(1), \end{aligned} \tag{C.15}$$

and the last equality follows form the fact that $F_{\epsilon|\mathbf{X}_i}\{\widehat{g}^{NW}(\boldsymbol{\beta}_1^{\top}\mathbf{X}_i|\boldsymbol{\beta})-g(\boldsymbol{\beta}_1^{\top}\mathbf{X}_i|\boldsymbol{\beta})|\mathbf{X}_i\}=F_{\epsilon|\mathbf{X}_i}(0|\mathbf{X}_i)+o_p(a_n^*+a_n+h_2^2)=\tau+o_p(a_n^*+a_n+h_2^2)$ which holds uniformly (see Proposition 3.1) and that $n^{-1/2}\sum_{i=1}^{n}g'(\boldsymbol{\beta}_1^{\top}\mathbf{X}_i|\boldsymbol{\beta})\{\mathbf{X}_{i,-1}-\mathrm{E}(\mathbf{X}_{-1}|\boldsymbol{\beta}_1^{\top}\mathbf{X})\}=O_p(1)$. Furthermore,

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n}\sigma_i^2(\widehat{g}^{NW}) \quad &= \quad \mathbf{t}^{\top}\frac{1}{n}\sum_{i=1}^{n}\{\mathbf{X}_{i,-1}-\mathrm{E}(\mathbf{X}_{-1}|\boldsymbol{\beta}_1^{\top}\mathbf{X})\}F_{\epsilon|\mathbf{X}_i}\{\widehat{g}^{NW}(\boldsymbol{\beta}_1^{\top}\mathbf{X}_i|\boldsymbol{\beta})-g(\boldsymbol{\beta}_1^{\top}\mathbf{X}_i|\boldsymbol{\beta})|\mathbf{X}_i\} \\ & \qquad \{g'(\boldsymbol{\beta}_1^{\top}\mathbf{X}_i|\boldsymbol{\beta})\}^2[1-F_{\epsilon|\mathbf{X}_i}\{\widehat{g}^{NW}(\boldsymbol{\beta}_1^{\top}\mathbf{X}_i|\boldsymbol{\beta})-g(\boldsymbol{\beta}_1^{\top}\mathbf{X}_i|\boldsymbol{\beta})|\mathbf{X}_i\}] \\ & \qquad \{\mathbf{X}_{i,-1}-\mathrm{E}(\mathbf{X}_{-1}|\boldsymbol{\beta}_1^{\top}\mathbf{X})\}^{\top}\mathbf{t} \\ &= \quad \mathbf{t}^{\top}\tau(1-\tau)\Sigma\mathbf{t}+o_p(1).\end{aligned} \tag{C.16}$$

Therefore, using (C.14), (C.15), (C.16) 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.12) and (C.13).

■

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

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

[4] Hjort, N. L. & D. Pollard (1993) Asymptotics for minimisers of convex processes. Unpublished manuscript, http://arxiv.org/pdf/1107.3806v1.pdf

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

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