Supplementary Materials for "Generalized Varying Coefficient Partially Linear Measurement Errors Models" [1]

Jun Zhang[1*], Zhenghui Feng[2], Peirong Xu[3], Hua Liang[4]

APPENDIX

We present the conditions, prepare a preliminary lemma, and give the proofs of the main results.

A.1. Conditions.

The following are the regularity conditions for our asymptotic results.

(C1) $q_2(x,y) < 0$ for $x \in \mathbb{R}$ and y in the range of the response variable.

(C2) The functions $T''(\cdot)$ and $g'''(\cdot)$ are continuous.

(C3) The random variable U has bounded support $\mathcal{U}$. The elements of the function $\alpha''(u)$ are continuous in $u \in \mathcal{U}$.

(C4) The density functions $f_U(u)$, $f_V(v)$ of U, V are Lipschitz continuous and bounded away from 0 and infinite on their supports, respectively. Moreover, the joint density function $f_{U,V}(u,v)$ of (U,V) is continuous on the support $\mathcal{U} \times \mathcal{V}$.

(C5) $E\big[q_l^s(Z,Y)N^{\otimes 2}\big|U=u\big]$, $E\big[q_l^s(Z,Y)N^{\otimes 2}\big|V=v\big]$ and $E\big[q_l^s(Z,Y)N^{\otimes 2}\big|U=u,V=v\big]$ for $l=1,2$, $s=1,2$ are Lipschitz continuous and twice differentiable on $u \in \mathcal{U}$ and $v \in \mathcal{V}$. Moreover, $E\{q_2^2(Z,Y)\} \le \infty$, $E\{q_1^{2+\delta}(Z,Y)\} \le \infty$ for some $\delta > 2$ and $E\big[\rho_2(Z)N^{\otimes 2}\big|U=u\big]$ is nonsingular for each $u \in \mathcal{U}$.

(C6) The kernel functions $K(\cdot)$, $L(\cdot)$ are univariate bounded, continuous and symmetric density functions satisfying that $\int t^2K(t)dt \ne 0$, $\int t^2L(t)dt \ne 0$, and $\int |t|^jK(t)dt < \infty$, $\int |t|^jL(t)dt < \infty$. for $j=1,2,3,4$. Moreover, the second derivatives of $K(\cdot)$ and $L(\cdot)$ are bounded on $\mathbb{R}^1$.

(C7) The bandwidths h and b satisfy:
(i) $b=b_k$, $k=1,\dots,d$, $b \asymp c_bh_o$ for some constant $c_b>0$; $h \asymp c_hh_o$ for some constant $c_h>0$.
(ii) $h_o \to 0$ as $n \to \infty$, $nh_o^2/(\log h_o^{-1})^4 \to \infty$, $nh_o^4 \to 0$.

(C8) For all λ_{1j}, λ_{2s}, $j=1,\dots,d$, $s=1,\dots,r$, $\lambda_{1j} \to 0$, $\sqrt{n}\lambda_{1j} \to \infty$, $\lambda_{2s} \to 0$, $\sqrt{n}\lambda_{2s} \to \infty$, and $\lim\inf_{n\to\infty} \lim\inf_{u\to 0^+} p'_{\lambda_{1j}}(u)/\lambda_{1j} > 0$, $\lim\inf_{n\to\infty} \lim\inf_{u\to 0^+} p'_{\lambda_{2s}}(u)/\lambda_{2s} > 0$.

[1]Corresponding Author: Jun Zhang. Jun Zhang(E-mail: zhangjunstat@gmail.com), Shen Zhen-Hong Kong Joint Research Centre for Applied Statistical Sciences, College of Mathematics and Computational Science, Institute of Statistical Sciences at Shenzhen University, Shenzhen University, Shenzhen, China. Zhenghui Feng (E-mail: zhfengwise@gmail.com), School of Economics, and the Wang Yanan Institute for Studies in Economics, Xiamen University, Xiamen, China. Peirong Xu (E-mail: xupeirong@seu.edu.cn), Department of Mathematics, Southeast University, Nanjing, China. Hua Liang(E-mail: hliang@email.gwu.edu) Department of Statistics, George Washington University, Washington, D.C., USA.

A.2. Preliminary Lemmas

Lemma A.1. *Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be i.i.d. random vectors, where the Y's are scalar random variables. Assume that $E|Y|^r < \infty$ and that $\sup_x \int |y|^r f(x,y)dy < \infty$, where f denotes the joint density of (X,Y). Let $K(\cdot)$ be a bounded positive function with bounded support, satisfying a Lipschitz condition. Then*

$$\sup_x \Big| \frac{1}{n} \sum_{i=1}^n K_h(X_i - x)Y_i - E[K_h(X_i - x)Y_i] \Big| = O_P\big[\{nh/\log(1/h)\}^{-1/2}\big]. \tag{A.1}$$

provided that $n^{2\epsilon-1}h \to \infty$ for some $\epsilon < 1 - r^{-1}$.

Proof. Lemma A.1 follows a direct result of Mack and Silverman (1982). □

In the following, for notational simplicity, we absorb σ^2 into $T(\cdot)$, i.e., $\mathrm{Var}(Y|U,\xi,W,X) = T\{\mu(U,\xi,W,X)\}$. Write $Z = \beta^T\xi + \theta^T W + \alpha(U)^T X$, $\tilde{Z}_i(u) = \beta^T\xi_i + \theta^T W_i + X_i^T\alpha(u) + (U_i - u)X_i^T\alpha'(u)$, $N_i = (X_i^T, \xi_i^T, W_i^T)^T$, $R_i^u = \big(N_i^T, h^{-1}(U_i - u)X_i^T\big)^T$ and further denote that $L_b(\cdot) = \mathrm{diag}\big(L_{b_1}(\cdot), \ldots, L_{b_d}(\cdot)\big)$. Denote the local quasi-likelihood estimators from (2.2.2) by $\hat{\alpha}_*(u)$, $\hat{B}_*(u)$, $\hat{\beta}_*$, $\hat{\theta}_*$ and let

$$\hat{\delta} = \sqrt{nh}\left\{(\hat{\alpha}_*(u) - \alpha(u))^T, (\hat{\beta}_* - \beta))^T, (\hat{\theta}_* - \theta)^T, h(\hat{B}_*(u) - \alpha'(u))^T\right\}.$$

Denote $v_{K_0} = \int K^2(s)ds$ and $v_{K_2} = \int s^2K^2(s)ds$, $\mu_{K_2} = \int s^2K(s)dt$, $\mu_{K_4} = \int s^4K(s)ds$. Let $C_v(\xi,W,X) = \mathrm{diag}\big(v_{K_0}\tilde{N}^{\otimes 2}, v_{K_2}X^{\otimes 2}\big)$, $S_\mu(\xi,W,X) = \mathrm{diag}\big(\tilde{N}^{\otimes 2}, \mu_{K_2}X^{\otimes 2}\big)$.

Lemma A.2. *Under Conditions (C1)-(C7), and the identifiability condition $E(Y|U,\xi(V), W,X) = E(Y|U,V,W,X)$, if $h_o \to 0$, $nh_o/\log(nh_o) \to \infty$, uniformly in $u \in \mathcal{U}$, we have*

$$\hat{\delta} = A_u^{-1}B_n^u + O_P\big(h_o^2\sqrt{nh_o} + \{\{\log h_o^{-1}\}/\{nh_o\}\}^{1/2}\big), \tag{A.2}$$

where $A_u = f_U(u)E\{\rho_2(Z)S_\mu(\xi,W,X)\big|U=u\}$, and B_n^u are defined as

$$\begin{aligned} B_n^u \ &\stackrel{\text{def}}{=}\ \sqrt{\frac{h}{n}}\sum_{i=1}^n q_1(\tilde{Z}_i(u),Y_i)R_i^u K_h(U_i-u) \\ &\quad -\sqrt{\frac{h}{n}}\sum_{i=1}^n \frac{f_{U,V}(u,V_i)}{f_V(V_i)}E\{\rho_2(Z)(N^T,\mathbf{0})^T\big|U=u,V=V_i\}e_i^T\beta. \end{aligned}$$

Proof. Let $r_n = (nh)^{-1/2}$, $\hat{\omega}_i = (\beta_* - \beta)^T(\hat{\xi}_i - \xi_i) + \beta^T(\hat{\xi}_i - \xi_i)$. From (2.2.2), $\hat{\delta}$ maximizes

$$\ell_n(\delta_*) = h\sum_{i=1}^n \Big[\mathcal{Q}\{g^{-1}(r_n\delta_*^T R_i^u + \hat{\omega}_i + \tilde{Z}_i(u)), Y_i\} - \mathcal{Q}\{g^{-1}(\tilde{Z}_i(u)), Y_i\}\Big] K_h(U_i-u)$$

with respective δ_*. Taking Taylor expansion of $\mathcal{Q}(g^{-1}(\cdot),Y_i)$ at $\tilde{Z}_i(u)$, we have

$$\begin{aligned} \ell_n(\delta_*) \quad &= \quad \Big[r_n h\sum_{i=1}^n\Big\{q_1(\tilde{Z}_i(u),Y_i) + q_2(\tilde{Z}_i(u),Y_i)\hat{\omega}_i\Big\}K_h(U_i-u)R_i^{uT}\Big]\delta_* \\ &\quad +\delta_*^T\Big[\frac{r_n^2 h}{2}\sum_{i=1}^n q_2(\tilde{Z}_i(u),Y_i)\big(R_i^u\big)^{\otimes 2}K_h(U_i-u)\Big]\delta_* \\ &\quad +\Big[h\sum_{i=1}^n\{q_1(\tilde{Z}_i(u),Y_i)\hat{\omega}_i + 0.5q_2(\tilde{Z}_i(u),Y_i)\hat{\omega}_i^2\}K_h(U_i-u)\Big] \\ &\stackrel{\text{def}}{=} \quad \hbar_{n1}^T\delta_* + \delta_*^T\hbar_{n2}\delta_* + \hbar_{n3}. \end{aligned}$$

Step 1.1. Note that $\hbar_{n3}$ is irrelevant to δ_*. We show that $\hbar_{n3} = O_P(1)$.

$$\begin{aligned}
\hbar_{n3} &= h\sum_{i=1}^{n} q_1(\tilde{Z}_i(u), Y_i)\hat{\omega}_i K_h(U_i - u) + h\sum_{i=1}^{n} 0.5q_2(\tilde{Z}_i(u), Y_i)\hat{\omega}_i^2 K_h(U_i - u) \\
&\stackrel{\text{def}}{=} \hbar_{n3}^{[1]} + \hbar_{n3}^{[2]}.
\end{aligned}$$

By using (2.1.3), we obtain that

$$\begin{aligned}
\hbar_{n3}^{[1]} &= \frac{hb^2\mu_{L2}}{2}\sum_{i=1}^{n} q_1(\tilde{Z}_i(u), Y_i)\xi^{(2)}(V_i)^T\beta K_h(U_i - u)(1 + o_P(1)) \\
&\quad + \frac{h}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{q_1(\tilde{Z}_i(u), Y_i)}{f_V(V_i)}\left\{L_b(V_j - V_i)e_j\right\}^T\beta K_h(U_i - u)(1 + o_P(1)) \\
&\stackrel{\text{def}}{=} \hbar_{n3}^{[1]}(1) + \hbar_{n3}^{[1]}(2).
\end{aligned}$$

Under the identifiability condition $E(Y|U, \xi(V), W, X) = E(Y|U, V, W, X)$, we have $E\big\{Y\xi^{(2)}(V)\big\} = E\big\{\xi^{(2)}(V)\mu(U, \xi, W, X)\big\}$. Thus, a direct calculation of the argument $E\big\{\sum_{i=1}^{n} q_1(\tilde{Z}_i(u), Y_i)\xi^{(2)}(V_i)^T \beta K_h(U_i - u)\big\}$ and Lemma A.1 entail that

$$\begin{aligned}
\hbar_{n3}^{[1]}(1) &= nh^3b^2\mu_{L2}f_U(u)\alpha''(u)^T E\big\{X\xi^{(2)}(V)^T\beta\rho_2(Z)\big\}/4 \\
&\quad + O_P\left(nh^4b^2\right) + O_P\big(nhb^2\big\{\log h^{-1}/nh\big\}^{1/2}\big). \qquad \text{(A.3)}
\end{aligned}$$

For the second argument $\hbar_{n3}^{[1]}(2)$, we can use the projection of U-statistics and obtain

$$\hbar_{n3}^{[1]}(2) = h^3\sum_{i=1}^{n}\frac{\mu_{K2}\alpha''(u)^T e_i^T\beta f_{U,V}(u, V_i)}{2f_V(V_i)}E\big\{X\rho_2(Z)\big|U = u, V = V_i\big\}(1 + o_P(1)). \text{(A.4)}$$

This indicates that $\hbar_{n3}^{[1]}(2) = O_P(\sqrt{nh^6})$. This statement with (A.3) indicates that, under Condition (C7), $\hbar_{n3}^{[1]} = o_P(1)$.

Next, we consider the second argument $\hbar_{n3}^{[2]}$.

$$\begin{aligned}
\hbar_{n3}^{[2]} &= \frac{h}{2}\sum_{i=1}^{n}(Y_i - g^{-1}(\tilde{Z}_i(u)))\rho_1'(\tilde{Z}_i(u))K_h(U_i - u)\hat{\omega}_i^2 - \frac{h}{2}\sum_{i=1}^{n}\rho_2(\tilde{Z}_i(u))K_h(U_i - u)\hat{\omega}_i^2 \\
&\stackrel{\text{def}}{=} \hbar_{n3}^{[2]}(1) - \hbar_{n3}^{[2]}(2).
\end{aligned}$$

Similar to $\hbar_{n3}^{[1]}$, we have $\hbar_{n3}^{[2]}(1) = o_P(1)$. For the argument $\hbar_{n3}^{[2]}(2)$, invoking (2.1.3), we have

$$\begin{aligned}
\hbar_{n3}^{[2]}(2) &= \frac{hb^4\mu_{L2}^2}{2}\sum_{i=1}^{n}\rho_2(\tilde{Z}_i(u))K_h(U_i - u)\big(\xi^{(2)}(V_i)^T\beta\big)^2 \\
&\quad + \frac{hb^2\mu_{L2}}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\rho_2(\tilde{Z}_i(u))}{f_V(V_i)}\big(\xi^{(2)}(V_i)^T\beta\big)K_h(U_i - u)\left\{L_b(V_j - V_i)e_j\right\}^T\beta \\
&\quad + \frac{h}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{s=1}^{n}\frac{\rho_2(\tilde{Z}_i(u))}{f_V^2(V_i)}K_h(U_i - u)\left\{L_b(V_j - V_i)e_j\right\}^T\beta\left\{L_b(V_s - V_i)e_s\right\}^T\beta \\
&\stackrel{\text{def}}{=} \hbar_{n3}^{[2]}(2)_1 + \hbar_{n3}^{[2]}(2)_2 + \hbar_{n3}^{[2]}(2)_3.
\end{aligned}$$

Note that $\hbar_{n3}^{[2]}(2)_1$ is a sum of independent and identically distributed random variables. We have $\hbar_{n3}^{[2]}(2)_1 = O_P\left(nhb^4\right) = o_P(1)$. Furthermore, similar to (A.4), we have $\hbar_{n3}^{[2]}(2)_2 = O_P\big(\sqrt{nh^2b^4}\big) = o_P(1)$. For the argument $\hbar_{n3}^{[2]}(2)_3$, we can use the projection of U-statistics with second order (Serfling; 1980, Section 5.5.2) to show that $\hbar_{n3}^{[2]}(2)_3 = O_P(1)$. As a result, we obtain that $\hbar_{n3}^{[2]} = \hbar_{n3}^{[2]}(1) + \hbar_{n3}^{[2]}(2) = O_P(1)$, yielding $\hbar_{n3} = \hbar_{n3}^{[1]} + \hbar_{n3}^{[2]} = O_P(1)$.

Step 1.2. We investigate the argument $\hbar_{n2}$. By the definition of $\rho_2(\cdot,\cdot)$, we have

$$\begin{aligned}\hbar_{n2} \ &\stackrel{\text{def}}{=}\ \hbar_{n2}^{[1]} + \hbar_{n2}^{[2]} = \Big[\frac{1}{2n}\sum_{i=1}^{n}\{Y_i - g^{-1}(\tilde Z_i(u))\}\rho_1'(\tilde Z_i(u))\,(R_i^u)^{\otimes 2}\,K_h(U_i-u)\Big] \\ &\qquad + \Big[-\frac{1}{2n}\sum_{i=1}^{n}\rho_2(\tilde Z_i(u))\,(R_i^u)^{\otimes 2}\,K_h(U_i-u)\Big].\end{aligned}$$

Similar to (A.3), by Lemma A.1 and Condition (C7), we obtain that $\hbar_{n2}^{[1]} = o_P(1)$. For $\hbar_{n2}^{[2]}$, by Lemma A.1, we have

$$\begin{aligned}\hbar_{n2}^{[2]} &= -0.5E\Big[\frac{1}{n}\sum_{i=1}^{n}\rho_2(\tilde Z_i(u))\,(R_i^u)^{\otimes 2}\,K_h(U_i-u)\Big] + O_P\big(\{\log h^{-1}/nh\}^{1/2}\big) \\ &= -0.5f_U(u)E\big[\rho_2(Z)S_\mu(\xi,W,X)\big|U=u\big] + O_P\big(h^2 + \{\log h^{-1}/nh\}^{1/2}\big).\end{aligned}$$

Thus, we have

$$\hbar_{n2} = \hbar_{n2}^{[1]} + \hbar_{n2}^{[2]} = -0.5A_u + O_P\big(h_o^2 + \{\log 1/h_o/nh_o\}^{1/2}\big). \tag{A.5}$$

Step 1.3. From Step 1.1, we know that δ_* is irrelevant to $\hbar_{n3}$ and we had $\hbar_{n3} = O_P(1)$. Thus, we obtain that $\ell_n(\delta_*) - \hbar_{n3} = \hbar_{n1}^T\delta_* + \delta_*^T\hbar_{n2}\delta_*$. As claimed in the proof of Lemma A.2 in Li and Liang (2008), each element of $\hbar_{n2}$ is a sum of i.i.d. random variables of kernel form, and Lemma A.1 entails the uniform convergence property. Thus, with arguments similar to Li and Liang (2008), Convexity Lemma in Pollard (1991) and Lemma A.1 of Carroll et al. (1997), there exists a consistent estimator $\tilde\delta_*$ such that $\sup_{u\in\mathcal{U}}\big|\tilde\delta_* - A_u^{-1}\hbar_{n1}\big| = o_P(1)$. So, the derivative of $\ell_n(\delta_*)$ or $\{\ell_n(\delta_*) - \hbar_{n3}\}$ with respect to δ_* further guarantees a consistent estimator of δ_*, uniformly in $u \in \mathcal{U}$. That is

$$\frac{\partial}{\partial\delta_*}\ell_n(\delta_*)\Big|_{\delta_*=\hat\delta_*} = r_nh\sum_{i=1}^{n}q_1\big(\tilde Z_i(u) + r_n\hat\delta_*^TR_i^u + \hat\omega_i, Y_i\big)R_i^uK_h(U_i-u) = 0. \tag{A.6}$$

Taking Taylor expansion on (A.6) with respect to $\tilde Z_i(u)$ yields

$$\hbar_{n1} + 2\hbar_{n2}\hat\delta_* + \frac{r_n^3h}{2}\sum_{i=1}^{n}q_3\big(\tilde Z_i(u) + \hat\zeta_i, Y_i\big)\big(\hat\delta_*^TR_i^u\big)^2\big(R_i^u\big)K_h(U_i-u) = 0, \tag{A.7}$$

where $\hat\zeta_i$ is between 0 and $r_n\hat\delta_*^TR_i^u + \hat\omega_i$, and $q_3(x,y) = \frac{\partial^3}{\partial x^3}\mathcal{Q}\{g^{-1}(x),y\}$. Similar to the proof of Lemma 2 in Li and Liang (2008), we know the last term in (A.7) is of order $O_P(nr_n^3h) = O_P(1/\sqrt{nh})$. From (A.5) in Step 1.2, we know $\hbar_{n2} = -0.5A_u + O_P\big(h_o^2 + \{\log 1/h_o/nh_o\}^{1/2}\big)$. From (A.7), we obtain that

$$\hat\delta_* = -(2\hbar_{n2})^{-1}\hbar_{n1} + O_P(1/\sqrt{nh}) = A_u^{-1}\hbar_{n1} + O_P\big(h_o^2 + \{\log 1/h_o/nh_o\}^{1/2}\big). \tag{A.8}$$

For the argument $\hbar_{n1}$, similar to $\hbar_{n3}^{[1]}$ in Step 1.1, by using (2.1.3), we have

$$\begin{aligned}
\hbar_{n1} &= \sqrt{\frac{h}{n}}\sum_{i=1}^{n} q_1(\tilde{Z}_i(u),Y_i)R_i^u K_h(U_i-u) + \sqrt{\frac{hb^4}{n}}\sum_{i=1}^{n} q_2(\tilde{Z}_i(u),Y_i)\xi^{(2)}(V_i)^T\beta R_i^u K_h(U_i-u) \\
&\quad + \sqrt{\frac{h}{n^3}}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{q_2(\tilde{Z}_i(u),Y_i)}{f_V(V_i)}\{L_b(V_j-V_i)e_j\}^T\beta R_i^u K_h(U_i-u) \\
&\stackrel{\text{def}}{=} \hbar_{n1}^{[1]}+\hbar_{n1}^{[2]}+\hbar_{n1}^{[3]}.
\end{aligned}$$

Recall $E(Y|U,\xi(V),W,X)=E(Y|U,V,W,X)$. Similar to the analysis for (A.3), we have $E\big\{\sum_{i=1}^n q_2 (\tilde{Z}_i(u),Y_i)\xi^{(2)}(V_i)^T\beta R_i^u K_h(U_i-u)\big\}=O(nh^2)+O(n)$. Lemma A.1 entails that $\hbar_{n1}^{[2]}=O_P\big(\sqrt{nhb^4}+\big\{\log h^{-1}/nh\big\}^{1/2}\big)$. Appealing to (A.8), thus,

$$\hat{\delta}_* = -A_u^{-1}\big\{\hbar_{n1}^{[1]}+\hbar_{n1}^{[3]}\big\}+O_P\big(h_o^2+\sqrt{nh_o^5}+\big\{\log h_o^{-1}/nh_o\big\}^{1/2}\big). \tag{A.9}$$

A direct calculation yields

$$\begin{aligned}
E\big(\hbar_{n1}^{[1]}\big) &= 0.5\sqrt{nh^5}\mu_{K2}\alpha''(u)^T f_U(u)E\big\{\rho_2(Z)(N^T,\mathbf{0})^T X\big|U=u\big\}+O_P\big(\sqrt{nh^7}\big), \\
\mathrm{Var}\big(\hbar_{n1}^{[1]}\big) &= f_U(u)E\big\{\rho_2(Z)C_v(\xi,W,X)\big|U=u\big\}+O_P(h^5).
\end{aligned}$$

For the argument $\hbar_{n1}^{[3]}$, similar to (A.4), we use the projection of U-statistics and obtain that

$$\hbar_{n1}^{[3]} = -\sqrt{\frac{h}{n}}\sum_{i=1}^{n}\frac{f_{U,V}(u,V_i)}{f_V(V_i)}E\big[\rho_2(Z)(N^T,\mathbf{0})^T\big|U=u,V=V_i\big]e_i^T\beta+O_P\big(\sqrt{nh}(h^2+b^2)\big).$$

Together with (A.9) and the asymptotic expression of $\hbar_{n1}^{[3]}$, invoking Condition (C7), we obtain the asymptotic expression for B_n^u and complete the proof. $\square$

Lemma A.3. *Under the conditions of Theorem 2, with probability approaching to 1, for any given* $\|\beta_1^*-\beta_1\|=O_P(n^{-1/2})$, $\|\theta_1^*-\theta_1\|=O_P(n^{-1/2})$ *and any constant* C_0, *we have*

$$\mathcal{L}_P\left\{\begin{pmatrix}\beta_1\\ \mathbf{0}\end{pmatrix},\begin{pmatrix}\theta_1\\ \mathbf{0}\end{pmatrix}\right\} = \max_{\|\beta_2^*\|\le C_0 n^{-1/2},\|\theta_2^*\|\le C_0 n^{-1/2}}\mathcal{L}_P\left\{\begin{pmatrix}\beta_1\\ \beta_2^*\end{pmatrix},\begin{pmatrix}\theta_1\\ \theta_2^*\end{pmatrix}\right\}$$

Proof. This result follows an argument similar to that of Lemma A.3 in Li and Liang (2008). We omit the details. $\square$

A.3. Proof of Theorem 1

Write $\hat{\sigma}=\sqrt{n}\big(\hat{\beta}-\beta,\hat{\theta}-\theta\big)$, $Q_i=\big(\xi_i^T,W_i^T\big)^T$, $\tilde{Z}_i=\beta^T\xi_i+\theta^T W_i+\hat{\alpha}_*(U_i)^T X_i$, $Z_i=\beta^T\xi_i+\theta^T W_i+\alpha(U_i)^T X_i$ $i=1,\dots,n$. If $\big(\hat{\beta}^T,\hat{\theta}^T\big)^T$ maximizes (2.2.3), then $\hat{\sigma}$ maximizes

$$\Im_n(\sigma_*)=\sum_{i=1}^{n}\Big[\mathcal{Q}\{g^{-1}(n^{-1/2}\sigma_*^T Q_i+\hat{\omega}_i+\tilde{Z}_i),Y_i\}-\mathcal{Q}\{g^{-1}(\tilde{Z}_i),Y_i\}\Big]$$

with respect to σ_*. Taylor expansion entails that

$$\begin{aligned}
\Im_n(\sigma_*) &= \Big[n^{-1/2}\sum_{i=1}^n \{q_1(\tilde Z_i, Y_i)Q_i + q_2(\tilde Z_i, Y_i)Q_i\hat\omega_i\}^T\Big]\sigma_* + \sigma_*^T\Big[\frac{1}{2n}\sum_{i=1}^n q_2(\tilde Z_i, Y_i)Q_i^{\otimes 2}\Big]\sigma_* \\
&\quad + \Big[\sum_{i=1}^n\Big\{q_1(\tilde Z_i, Y_i)\hat\omega_i + 0.5 q_2(\tilde Z_i, Y_i)\hat\omega_i^2\Big\}\Big] \\
&\stackrel{\text{def}}{=} \vartheta_{n1}^T\sigma_* + \sigma_*^T\vartheta_{n2}\sigma_* + \vartheta_{n3}.
\end{aligned}$$

The proof of this theorem is similar to that of Lemma A.2. Note that ϑ_{n3} is irrelevant to σ_*. As such, we consider only ϑ_{n1} and ϑ_{n2}.

Step 2.1. For the argument ϑ_{n2}, we have

$$\vartheta_{n2} \stackrel{\text{def}}{=} \vartheta_{n2}^{[1]} - \vartheta_{n2}^{[2]} = \Big[\frac{1}{2n}\sum_{i=1}^n \big(Y_i - g^{-1}(\tilde Z_i)\big)\rho_1'(\tilde Z_i)Q_i^{\otimes 2}\Big] - \Big[\frac{1}{2n}\sum_{i=1}^n \rho_2(\tilde Z_i)Q_i^{\otimes 2}\Big].$$

We now consider $\vartheta_{n2}^{[1]}$. Taylor expansion entails that

$$\begin{aligned}
\vartheta_{n2}^{[1]} &= \frac{1}{2n}\sum_{i=1}^n \big(Y_i - g^{-1}(Z_i)\big)\rho_1'(Z_i)Q_i^{\otimes 2} + \frac{1}{2n}\sum_{i=1}^n \big(Y_i - g^{-1}(Z_i)\big)\rho_1''(Z_i)(\tilde Z_i - Z_i)Q_i^{\otimes 2} \\
&= \frac{1}{2n}\sum_{i=1}^n \frac{dg^{-1}(t)}{dt}\Big|_{t=Z_i}\rho_1'(Z_i)(\tilde Z_i - Z_i)Q_i^{\otimes 2} + \frac{1}{2n}\sum_{i=1}^n \frac{dg^{-1}(t)}{dt}\Big|_{t=Z_i}\rho_1''(Z_i)(\tilde Z_i - Z_i)^2 Q_i^{\otimes 2}.
\end{aligned}$$

Note that $E(Y_i - g^{-1}(Z_i)|U_i, \xi_i, W_i, X_i) = 0$. Then the first term above is $o_P(1)$ by the law of large numbers. Furthermore, as $\tilde Z_i - Z_i = (\hat\alpha_*(U_i) - \alpha(U_i))^T X_i$, by using Lemma A.2 and Condition (C7), we know that $\max_i|\hat\alpha_{*j}(U_i) - \alpha_j(U_i)| = O_P((nh_o)^{-1/2}) = o_P(n^{-1/4})$ for $j = 1, \ldots, p$. This convergence together with Lemma A.1 in Li and Liang (2008) indicate that $\frac{1}{2n}\sum_{i=1}^n \big(Y_i - g^{-1}(Z_i)\big)\rho_1''(Z_i)(\tilde Z_i - Z_i)Q_i^{\otimes 2} = o_P(n^{-1/2}) = o_P(1)$. Using Lemma A.2 again, we know that each elements of $\frac{1}{2n}\sum_{i=1}^n \frac{dg^{-1}(t)}{dt}\big|_{t=Z_i}\rho_1'(Z_i)(\tilde Z_i - Z_i)Q_i^{\otimes 2}$ is $O_P((nh_o)^{-1/2}) = o_P(1)$. Similarly, we have $\frac{1}{2n}\sum_{i=1}^n \frac{dg^{-1}(t)}{dt}\big|_{t=Z_i}\rho_1''(Z_i)(\tilde Z_i - Z_i)^2 Q_i^{\otimes 2} = o_P(1)$. As a consequence, we obtain that $\vartheta_{n2}^{[1]} = o_P(1)$. For the argument $\vartheta_{n2}^{[2]}$, Taylor expansion entails that $\vartheta_{n2}^{[2]} = \frac{1}{2n}\sum_{i=1}^n \rho_2(Z_i)Q_i^{\otimes 2} + \frac{1}{2n}\sum_{i=1}^n \rho_2'(Z_i)Q_i^{\otimes 2}(\tilde Z_i - Z_i) = E\left[\rho_2(Z)Q^{\otimes 2}\right] + o_P(1)$. Thus, we obtain that

$$\vartheta_{n2} = E\left\{\rho_2(Z)Q^{\otimes 2}\right\} + o_P(1) \stackrel{\text{def}}{=} \Sigma + o_P(1). \tag{A.10}$$

Step 2.2. We analyze the asymptotic expansion of ϑ_{n1}.

$$\vartheta_{n1} \stackrel{\text{def}}{=} \vartheta_{n1}^{[1]} - \vartheta_{n1}^{[2]} = \Big[\frac{1}{\sqrt n}\sum_{i=1}^n \big(Y_i - g^{-1}(\tilde Z_i)\big)\{\rho_1(\tilde Z_i) + \rho_1'(\tilde Z_i)\hat\omega_i\}Q_i\Big] - \Big[\frac{1}{\sqrt n}\sum_{i=1}^n \rho_2(\tilde Z_i)Q_i\hat\omega_i\Big].$$

Step 2.2.1. Note that

$$\begin{aligned}
\vartheta_{n1}^{[1]} &= \frac{1}{\sqrt n}\sum_{i=1}^n Q_i \times \Big(Y_i - g^{-1}(Z_i) - \frac{dg^{-1}(t)}{dt}\Big|_{Z=Z_i}(\tilde Z_i - Z_i)\Big) \\
&\qquad \times\big(\rho_1(Z_i) + \rho_1'(Z_i)(\tilde Z_i - Z_i) + \rho_1'(Z_i)\hat\omega_i + \rho_1''(Z_i)\hat\omega_i(\tilde Z_i - Z_i)\big) \\
&= \frac{1}{\sqrt n}\Big[\sum_{i=1}^n Q_i q_1(Z_i, Y_i) + \sum_{i=1}^n \big(Y_i - g^{-1}(Z_i)\big)\,Q_i\rho_1'(Z_i)\hat\omega_i - \sum_{i=1}^n (\tilde Z_i - Z_i)Q_i\rho_2(Z_i)\Big] + R_n.
\end{aligned}$$

Similar to $\vartheta_{n2}^{[1]}$ in Step 2.1, as $nh_o^2 \to \infty$, Lemma A.2 entails that $\max_i |\hat{\alpha}_{*j}(U_i) - \alpha_j(U_i)| = o_P(n^{-1/4})$ for $j = 1, \ldots, p$. Using Lemma A.1 in Li and Liang (2008), we can show that $\sum_{i=1}^n (Y_i - g^{-1}(Z_i))$ $\{\rho'(Z_i)(\tilde{Z}_i - Z_i) + \rho''(Z_i)(\tilde{Z}_i - Z_i)\hat{\omega}_i\}Q_i = o_P(n^{1/2})$ and $\sum_{i=1}^n Q_i \frac{dg^{-1}(t)}{dt}\big|_{Z=Z_i}(\tilde{Z}_i - Z_i)\{\rho_1'(Z_i)(\tilde{Z}_i - Z_i) + \rho_1'(Z_i)\hat{\omega}_i + \rho_1''(Z_i)\hat{\omega}_i(\tilde{Z}_i - Z_i)\} = o_P(n^{1/2})$. As a result, $R_n = o_P(1)$.

Next, we consider the second term in $\vartheta_{n1}^{[1]}$. By using (2.1.3), we have

$$\begin{aligned}
\frac{1}{\sqrt{n}} \sum_{i=1}^n \left(Y_i - g^{-1}(Z_i)\right) Q_i \rho_1'(Z_i)\hat{\omega}_i &= \Big[\frac{b^2 \mu_{L2}}{2\sqrt{n}} \sum_{i=1}^n \left(Y_i - g^{-1}(Z_i)\right) Q_i \rho_1'(Z_i) \xi^{(2)}(V_i)^T \beta\Big] \\
+ \Big[\frac{1}{\sqrt{n^3}} \sum_{i=1}^n \sum_{j=1}^n \frac{Q_i \rho_1'(Z_i)}{f_V(V_i)} \left(Y_i - g^{-1}(Z_i)\right) \{L_b(V_j - V_i) e_j\}^T \beta\Big] &\stackrel{\text{def}}{=} \Upsilon_{n1} + \Upsilon_{n2}. \qquad \text{(A.11)}
\end{aligned}$$

Under the identifiability condition $E(Y|U, \xi(V), W, X) = E(Y|U, V, W, X)$, we have $E\{(Y - g^{-1}(Z))$ $\xi^{(2)}(V)\} = 0$, as $b \to 0$. That is $\Upsilon_{n1} = o_P(1)$. For the argument Υ_{n2}, we can use the projection of U-statistics with second order (Serfling; 1980, Section 5.5.2) and show that $\sqrt{n}\Upsilon_{n2}$ converges in distribution to a weighted sum of independent χ_1^2 random variables, i.e., $\Upsilon_{n2} = O_P(n^{-1/2}) = o_P(1)$.

Now, we consider the third term in $\vartheta_{n1}^{[1]}$. Let $\pi = (I_p, \mathbf{0}, \mathbf{0}, \mathbf{0})_{p \times (2p+d)}$. We have $(nh)^{-1/2}\pi\hat{\delta}_* = \hat{\alpha}_*(u) - \alpha(u)$. By using Lemma A.2, we have

$$\begin{aligned}
\frac{1}{\sqrt{n}} \sum_{i=1}^n Q_i \times (\tilde{Z}_i - Z_i)\rho_2(Z_i) &= \frac{1}{\sqrt{n}} \sum_{i=1}^n Q_i X_i^T (\hat{\alpha}_*(U_i) - \alpha(U_i))\rho_2(Z_i) \qquad \text{(A.12)} \\
&= \frac{1}{n\sqrt{h}} \sum_{i=1}^n Q_i \rho_2(Z_i) X_i^T \pi A_{u_i}^{-1} \left[\hbar_{n1}^{[1]} + \hbar_{n1}^{[3]}\right] + o_P(1).
\end{aligned}$$

Write $\varpi(U_i) = E\{\rho_2(Z) S_\mu(\xi, W, X) \big| U = U_i\}$ and let $\kappa_k(U_i)$ be the k-th element of $E\{\rho_2(Z)$ $N^{\otimes 2} \big| U = U_i\}^{-1} N_i$. Using the asymptotic expression of $\hbar_{n1}^{[1]}$ obtained in Step 1.3 and the projection of U-statistics (Serfling; 1980, Section 5.5.1), as $nh^4 \to 0$, we have

$$\begin{aligned}
&\frac{1}{n\sqrt{h}} \sum_{i=1}^n Q_i \rho_2(Z_i) X_i^T \pi A_{u_i}^{-1} \hbar_{n1}^{[1]} \qquad \text{(A.13)} \\
&= \frac{1}{\sqrt{n^3}} \sum_{i=1}^n \sum_{j=1}^n Q_i X_i^T \frac{\pi}{f_U(U_i)} \varpi(U_i)^{-1} R_j^{u_i} q_1(\tilde{Z}_j(U_i), Y_j) \rho_2(Z_i) K_h(U_j - U_i) \\
&= \frac{1}{\sqrt{n}} \sum_{i=1}^n E\big[Q X^T \rho_2(Z) \big| U = U_i\big] \pi \varpi(U_i)^{-1} (\tilde{N}_i^T, \mathbf{0}^T)^T q_1(Z_i, Y_i) + o_P(1) \\
&= \frac{1}{\sqrt{n}} \sum_{i=1}^n \sum_{k=1}^p E\big[Q X_k \rho_2(Z) \big| U = U_i\big] \kappa_k(U_i) q_1(Z_i, Y_i) + o_P(1).
\end{aligned}$$

Furthermore, let $\tilde{\varpi}(u, v) = E\{\rho_2(Z)(N^T, \mathbf{0})^T \big| U = u, V = v\}$, $\iota_k(u, v)$ be the k-th element of $\varpi(u)^{-1}\tilde{\varpi}(u, v)$, and $\varkappa_k(v) = E\left[Q X_k \rho_2(Z) \iota_k(U, v) \frac{f_{U,V}(U,v)}{f_U(U) f_V(v)}\right]$. Using Lemma A.2 and the projection of U-statistics (Serfling; 1980, Section 5.5.1), we have

$$\frac{1}{n\sqrt{h}}\sum_{i=1}^{n} Q_i\rho_2(Z_i)X_i^T\pi A_{u_i}^{-1}\hbar_{n1}^{[3]} \tag{A.14}$$
$$= -\frac{1}{\sqrt{n^3}}\sum_{i=1}^{n}\sum_{j=1}^{n} Q_i\rho_2(Z_i)X_i^T\pi\frac{f_{U,V}(U_i,V_j)}{f_U(U_i)f_V(V_j)}\varpi(U_i)^{-1}\tilde{\varpi}(U_i,V_j)e_j^T\beta + o_P(1)$$
$$= -\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{k=1}^{p}\varkappa_k(V_i)e_i^T\beta + o_P(1).$$

As a consequence, together with $R_n = o_P(1)$, the expressions (A.12), (A.13), (A.14) and that $\Upsilon_{n1} = o_P(1)$, $\Upsilon_{n2} = o_P(1)$ yield

$$\begin{aligned}\vartheta_{n1}^{[1]} &= -\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{k=1}^{p}\Big(E\big\{QX_k\rho_2(Z)\big|U=U_i\big\}\kappa_k(U_i)q_1(Z_i,Y_i) - \varkappa_k(V_i)e_i^T\beta\Big)\\ &\quad+\frac{1}{\sqrt{n}}\sum_{i=1}^{n} Q_iq_1(Z_i,Y_i) + o_P(1).\end{aligned} \tag{A.15}$$

Step 2.2.2. In this step, we analyze the argument $\vartheta_{n1}^{[2]}$. As $nb^4 \to 0$, the expression (2.1.3) and the projection of U-statistics (Serfling; 1980, Section 5.5.1) entail that

$$\begin{aligned}\vartheta_{n1}^{[2]} &= \frac{1}{\sqrt{n^3}}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\rho_2(Z_i)Q_i}{f_V(V_i)}\big\{L_b(V_j-V_i)e_j\big\}^T\beta + o_P(1)\\ &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n} E\big\{\rho_2(Z)Q\big|V=V_i\big\}e_i^T\beta + o_P(1).\end{aligned} \tag{A.16}$$

Thus, together with (A.15) and (A.16), we have

$$\begin{aligned}\vartheta_{n1} &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Big\{Q_i - \sum_{k=1}^{p}E\big\{QX_k\rho_2(Z)\big|U=U_i\big\}\kappa_k(U_i)\Big\}q_1(Z_i,Y_i)\\ &+ \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Big\{\sum_{k=1}^{p}\varkappa_k(V_i) - E\big\{\rho_2(Z)Q\big|V=V_i\big\}\Big\}e_i^T\beta + o_P(1).\end{aligned} \tag{A.17}$$

Recalling the definition of $\Im_n(\sigma_*)$, together with (A.10) and (A.17), Convexity Lemma of Pollard (1991) and Lemma A.1 of Carroll et al. (1997) indicate that the maximizer $\hat{\sigma}$ of $\Im_n(\sigma_*)$ with respect to σ_* has the asymptotic expression $\hat{\sigma} = \Sigma^{-1}\vartheta_{n1} + o_P(1)$, where Σ is defined in (A.10). We complete the proof. □

A.4. Proof of Theorem 2

Step 3.1. First, we establish asymptotic orders of the estimators $\hat{\beta}_{\lambda_1}, \hat{\theta}_{\lambda_2}$. Let $q_n = n^{-1/2} + a_n^* + b_n^*$, $S_1 = (s_{11},\ldots,s_{1d})^T$, $S_2 = (s_{21},\ldots,s_{2r})^T$ and $\|S_1\| = \|S_2\| = C$ for some constant C. Furthermore, define $\beta(n) = \beta + q_nS_1$, $\theta(n) = \theta + q_nS_2$, and $\sigma(n) = (\beta^T(n),\theta^T(n))^T$, $\sigma = (\beta^T,\theta^T)^T$. Let $\mathcal{S} = \{(S_1,S_2) : \|S_1\| = C, \|S_2\| = C\}$. It suffices to show that, for any given $\varphi > 0$, there exists a large constant C such that

$$P\Big\{\sup_{\mathcal{S}}\mathcal{L}_P(\sigma(n)) < \mathcal{L}_P(\sigma)\Big\} \geq 1-\varphi. \tag{A.18}$$

Write

$$
\begin{aligned}
F_{n1} &= \sum_{i=1}^{n}\big[\mathcal{Q}\{g^{-1}(\hat{\alpha}_*(U_i)^T X_i+\hat{\xi}_i^T\beta(n)+W_i^T\theta(n)),Y_i\} \\
&\qquad\qquad -\mathcal{Q}\{g^{-1}(\hat{\alpha}_*(U_i)^T X_i+\hat{\xi}_i^T\beta+W_i^T\theta),Y_i\}\big], \\
F_{n2} &= -n\sum_{j=1}^{d_1}\{p_{\lambda_{1j}}(|\beta_j+q_n s_{1j}|)-p_{\lambda_{1j}}(|\beta_j|)\}-n\sum_{l=1}^{r_1}\{p_{\lambda_{2l}}(|\theta_l+q_n s_{2l}|)-p_{\lambda_{2l}}(|\theta_l|)\},
\end{aligned}
$$

where d_1, r_1 are the numbers of elements of β_1, θ_1, respectively. Recall $p_\lambda(0)=0$ and $p_\lambda(|t|)\geq 0$ for all t. Thus, $\mathcal{L}_P(\sigma(n))-\mathcal{L}_P(\sigma)\leq F_{n1}+F_{n2}$. We now consider F_{n1}, F_{n2}. Let $\hat{Z}_i=\hat{\alpha}_*(U_i)^T X_i+\hat{\xi}_i^T\beta+W_i^T\theta$. Taylor expansion entails that

$$
\begin{aligned}
F_{n1} &= \sum_{i=1}^{n} q_1\left(\hat{Z}_i,Y_i\right) q_n\times\big(\hat{\xi}_i^T S_1+W_i^T S_2\big)+\frac{1}{2}\sum_{i=1}^{n} q_2\big(\hat{Z}_i+\hat{\zeta}_i^*,Y_i\big)q_n^2\times\big(\hat{\xi}_i^T S_1+W_i^T S_2\big)^2 \\
&\overset{\text{def}}{=} F_{n1}^{[1]}-F_{n1}^{[2]},
\end{aligned}
$$

where $|\hat{\zeta}_i^*|$ is between in 0 and $|\hat{Z}_i-Z_i|$. For the argument F_{n1}, we have

$$
\begin{aligned}
F_{n1}^{[1]} &= q_n\sum_{i=1}^{n} q_1(Z_i,Y_i)\big(\hat{\xi}_i^T S_1+W_i^T S_2\big)+q_n\sum_{i=1}^{n} q_2(Z_i,Y_i)(\hat{Z}_i-Z_i)\big(\hat{\xi}_i^T S_1+W_i^T S_2\big) \\
&\quad +\frac{q_n}{2}\sum_{i=1}^{n} q_3\big(Z_i+\hat{\zeta}_i^{**},Y_i\big)(\hat{Z}_i-Z_i)^2\big(\hat{\xi}_i^T S_1+W_i^T S_2\big) \\
&\overset{\text{def}}{=} \Delta_{n1}^{[1]}+\Delta_{n1}^{[2]}+\Delta_{n1}^{[3]},
\end{aligned}
$$

where $|\hat{\zeta}_i^{**}|$ is between in 0 and $|\hat{Z}_i-Z_i|$. Under the condition that $a_n^*=O(n^{-1/2})$, $b_n=O(n^{-1/2})$, we have $q_n=O(n^{-1/2})$. Recall $\|S_1\|=\|S_2\|=C$. Using (2.1.3) and similar to the analysis for (A.11), we can obtain that $\Delta_{n1}^{[1]}=q_n\sum_{i=1}^n q_1(Z_i,Y_i)Q_i^T(S_1^T,S_2^T)^T+o_P(1)C$. Using (2.1.3) again, similar to (A.12) and (A.16), we can show that $\Delta_{n1}^{[2]}=q_n\sum_{i=1}^n \Xi_i^T(S_1^T,S_2^T)^T+o_P(1)C$, where Ξ_i's are $i.i.d$ random variables with mean zero and finite variance. For the argument $\Delta_{n1}^{[3]}$, using the results of Lemma A.2 and (2.1.3), we know that $(\hat{Z}_i-Z_i)^2=O_P\big(1/nh+b^4+\log b^{-1}/nb\big)$. Thus, under Condition (C7), we have $\Delta_{n1}^{[3]}=o_P(1)C$.

Next, we consider $F_{n1}^{[2]}$. Taylor expansion entails that

$$
\begin{aligned}
F_{n1}^{[2]} &= \frac{q_n^2}{2}\sum_{i=1}^{n}(Y_i-g^{-1}(Z_i))\rho_1'(Z_i)\big(\hat{\xi}_i^T S_1+W_i^T S_2\big)^2-\frac{q_n^2}{2}\sum_{i=1}^{n}\rho_2(Z_i)\big(\hat{\xi}_i^T S_1+W_i^T S_2\big)^2 \\
&\quad +\frac{q_n^2}{2}\sum_{i=1}^{n}\Big\{\frac{dg^{-1}(t)}{dt}\Big|_{t=Z_i}\rho_1'(Z_i)+(Y_i-g^{-1}(Z_i))\rho_1''(Z_i)-\rho_2'(Z_i)\Big\}\big(\hat{\xi}_i^T S_1+W_i^T S_2\big)^2 \\
&\qquad\qquad \times\big[(Z_i-\hat{Z}_i)+\hat{\zeta}_i^*\big] \\
&\overset{\text{def}}{=} \Omega_{n1}^{[1]}-\Omega_{n1}^{[2]}+\Omega_{n1}^{[3]}.
\end{aligned}
$$

Note that $q_n^2=O(n^{-1})$. Using (2.1.3) and $E\left[(Y-g^{-1}(Z))\rho'(Z)\middle|U,\xi,W,X\right]=0$, we can obtain that $\Omega_{n1}^{[1]}=\frac{q_n^2}{2}\sum_{i=1}^n(Y_i-g^{-1}(Z_i))\rho_1'(Z_i)\left(\xi_i^T S_1+W_i^T S_2\right)^2+o_P(1)C^2=o_P(1)C^2$. Moreover, $|\hat{\zeta}_i^*|$ is between in 0 and $|\hat{Z}_i-Z_i|$. Then (2.1.3) and Lemma A.2 guarantee $|\hat{\zeta}_i^*|=O_P\big(1/\sqrt{nh}+$

$b^2 + \sqrt{\log b^{-1}/nb}\big)$. Hence, we have $\Omega_{n1}^{[3]} = o_P(1)C^2$. For the argument $\Omega_{n1}^{[2]}$, we have $\Omega_{n1}^{[2]} = \frac{q_n^2}{2}\sum_{i=1}^n \rho_2(Z_i)\left(\xi_i^T S_1 + W_i^T S_2\right)^2 + o_P(1)C^2 = (S_1, S_2)\Sigma(S_1^T, S_2^T)^T + o_P(1)C^2$. As a result, these arguments for $F_{n1}^{[1]}$ and $F_{n1}^{[2]}$ mean

$$F_{n1} = q_n \sum_{i=1}^n \left\{q_1(Z_i, Y_i)Q_i^T + \Xi_i^T\right\}(S_1^T, S_2^T)^T - 0.5(S_1, S_2)\Sigma(S_1^T, S_2^T)^T + o_P(1)C^2. \quad \text{(A.19)}$$

Using Taylor expansion and Cauchy-Schwartz inequality, we conclude that F_{n2} is bounded by $n\sqrt{d_1}q_n a_n^* \| S_1\| + nq_n^2 a_n^{**}\|S_1\|^2 + n\sqrt{r_1}q_n b_n^*\|S_2\| + nq_n^2 b_n^{**}\|S_2\|^2 = O(\sqrt{d_1} + \sqrt{r_1})C + O(a_n^{**}C^2 + b_n^{**}C^2)$. As $a_n^{**} \to 0$ and $b_n^{**} \to 0$, we obtain that $|F_{n2}| = O_P(1)C$. From (A.19), we know that its first term is $O_P(1)C$ by the central limit theorem. Provided C is sufficiently large, $(S_1, S_2)\Sigma(S_1^T, S_2^T)^T$ dominates the first term of (A.19) and F_{n2}. As a consequence, (A.18) holds for sufficiently large C. We then achieve $\hat{\beta}_{\lambda_1} = \beta + O_P(n^{-1/2})$, $\hat{\theta}_{\lambda_2} = \theta + O_P(n^{-1/2})$. □

A.5. Proof of Theorem 3

From Lemma A.3, we know that $\hat{\beta}_{\lambda_1(2)} = \mathbf{0}_{(d-d_1)\times 1}$, and $\hat{\theta}_{\lambda_2(2)} = \mathbf{0}_{(r-r_1)\times 1}$. We next investigate asymptotic normality of $\hat{\beta}_{\lambda_1(1)}, \hat{\theta}_{\lambda_2(1)}$. Define $\hat{\sigma}_{\lambda(1)} = \sqrt{n}\big((\hat{\beta}_{\lambda_1(1)} - \beta_{(1)})^T, (\hat{\theta}_{\lambda_2(1)} - \theta_{(1)})^T\big)^T$, $Q_{i(1)} = (\xi_{i(1)}^T, W_{i(1)}^T)^T$, $\tilde{Z}_{i(1)} = \beta_{(1)}^T\xi_{i(1)} + \theta_{(1)}^T W_{i(1)} + \hat{\alpha}_*(U_i)^T X_i$, $Z_{i(1)} = \beta_{91)}^T\xi_{i(1)} + \theta_{(1)}^T W_{i(1)} + \alpha(U_i)^T X_i$, $\hat{\omega}_{i(1)} = \beta_{(1)}^T(\hat{\xi}_{i(1)} - \xi_{i(1)}) + (\beta_{(1)*} - \beta_{(1)})^T(\hat{\xi}_{i(1)} - \xi_{i(1)})$. Then $\hat{\sigma}_{\lambda(1)}$ maximizes

$$\begin{aligned}
&\sum_{i=1}^n \left[\mathcal{Q}\{g^{-1}\left(\tilde{Z}_{i(1)} + \hat{\omega}_{i(1)} + n^{-1/2}\sigma_{*(1)}^T Q_{i(1)}\right)\} - \mathcal{Q}\{g^{-1}(\tilde{Z}_{i(1)})\}\right] - n\sum_{j=1}^{d_1} p_{\lambda_{1j}}(|\beta_{j*}|) \\
&-n\sum_{s=1}^{r_1} p_{\lambda_{2s}}(|\theta_{s*}|) \stackrel{\text{def}}{=} \Im_n(\sigma_{*1}) - n\sum_{j=1}^{d_1} p_{\lambda_{1j}}(|\beta_{j*}|) - n\sum_{s=1}^{r_1} p_{\lambda_{2s}}(|\theta_{s*}|), \qquad \text{(A.20)}
\end{aligned}$$

with respect to $\sigma_{*(1)}$.

By using Taylor expansion, $p_{\lambda_{1j}}(|\beta_{j*}|) = p_{\lambda_{1j}}(|\beta_j|) + p'_{\lambda_{1j}}(|\beta_j|)\text{sign}(\beta_j)(\beta_{j*} - \beta_i) + \frac{1}{2}p''_{\lambda_{1j}}(|\beta_j|)(\beta_{j*} - \beta_i)^2$. Similar to the analysis for $\Im_n(\sigma_*)$ in the proof of Theorem 1, together with Taylor expansions of $p_{\lambda_{1j}}(|\beta_{j*}|)$ and $p_{\lambda_{2s}}(|\theta_{s*}|)$, we can re-write (A.20) as

$$\begin{aligned}
&\Im_n(\sigma_{*(1)}) - n\sum_{j=1}^{d_1} p_{\lambda_{1j}}(|\beta_{j*}|) - n\sum_{s=1}^{r_1} p_{\lambda_{2s}}(|\theta_{s*}|) \\
&\quad = \Big[n^{-1/2}\sum_{i=1}^n \left\{q_1(\tilde{Z}_{i(1)}, Y_i)Q_{i(1)} + q_2(\tilde{Z}_{i(1)}, Y_i)Q_{i(1)}\hat{\omega}_{i(1)}\right\}^T + \sqrt{n}\mathcal{R}_{n,\lambda_1,\lambda_2}^T\Big]\sigma_{*(1)} \\
&\quad + \sigma_{*(1)}^T\Big[\frac{1}{2n}\sum_{i=1}^n q_2(\tilde{Z}_{i(1)}, Y_i)Q_{i(1)}^{\otimes 2} + \frac{1}{2}\Sigma_{n,\lambda_1,\lambda_2}\Big]\sigma_{*(1)} + \mathcal{C}_n, \qquad \text{(A.21)}
\end{aligned}$$

where $\mathcal{C}_n$ is irrelevant to $\sigma_{*(1)}$. Define $\Sigma_{(1)} = E\big\{\rho_2(Z_{(1)})Q_{(1)}^{\otimes 2}\big\}$. Similar to the proof of Theorem 1, maximizing (A.21) with respect to $\sigma_{*(1)}$, we have

$$\begin{aligned}
&\sqrt{n}\big(\Sigma_{(1)}+\Sigma_{n,\lambda_1,\lambda_2}\big)\Big\{\big((\hat\beta_{\lambda_1(1)}-\beta_{(1)})^T,(\hat\theta_{\lambda_2(1)}-\theta_{(1)})^T\big)^T+\big(\Sigma_{(1)}+\Sigma_{n,\lambda_1,\lambda_2}\big)^{-1}\mathcal{R}_{n,\lambda_1,\lambda_2}\Big\}\\
&=\frac{1}{\sqrt{n}}\sum_{i=1}^n\Big\{Q_{i(1)}-\sum_{k=1}^p E\big\{Q_{(1)}X_k\rho_2(Z_{(1)})\big|U=U_i\big\}\kappa_{1k}(U_i)\Big\}q_1(Z_{i(1)},Y_i)\\
&\quad+\frac{1}{\sqrt{n}}\sum_{i=1}^n\Big\{\sum_{k=1}^p\varkappa_{1k}(V_i)-E\big\{\rho_2(Z_{(1)})Q_{(1)}\big|V=V_i\big\}\Big\}e_{i(1)}^T\beta_{(1)}+o_P(1).
\end{aligned}$$

We complete the proof.

A.6. Proof of Theorem 4

Write $X_i^u=\big(X_i^T,h^{-1}X_i(U_i-u)^T\big)^T$, $\hat\eta(u)=(nh)^{1/2}\big((\hat\alpha(u)-\alpha(u))^T,h(\hat B(u)-\alpha(u))^T\big)^T$, $\hat Z_i(u)=\alpha(u)^TX_i+\alpha'(u)X_i(U_i-u)+\hat\beta^T\xi_i+\hat\theta_iW_i$. We know that $\hat\eta(u)$ is the maximizer of (A.22) with respect to η_*,

$$\begin{aligned}
\wp_n(\eta_*)=h\sum_{i=1}^n\Big[&\mathcal{Q}\big\{g^{-1}\big(r_n^{-1}\eta_*^TX_i^u+(\hat\xi_i-\xi_i)^T\hat\beta+\hat Z_i(u)\big),Y_i\big\}\\
&-\mathcal{Q}\{g^{-1}(\hat Z_i(u)),Y_i\}\Big]K_h(U_i-u).
\end{aligned}\tag{A.22}$$

Similar to handling $\ell_n(\delta^*)$ in Lemma A.2, we express (A.22) as

$$\begin{aligned}
\wp_n(\eta_*)\;&=\;\Big[r_nh\sum_{i=1}^n\Big\{q_1(\hat Z_i(u),Y_i)+q_2(\hat Z_i(u),Y_i)(\hat\xi_i-\xi_i)^T\hat\beta\Big\}K_h(U_i-u)X_i^{uT}\Big]\eta_*\\
&\quad+\eta_*^T\Big[\frac{r_n^2h}{2}\sum_{i=1}^n q_2(\hat Z_i(u),Y_i)\big(X_i^u\big)^{\otimes2}K_h(U_i-u)\Big]\eta_*\\
&\quad+\Big[h\sum_{i=1}^n\{q_1(\hat Z_i(u),Y_i)(\hat\xi_i-\xi_i)^T\hat\beta+\frac12 q_2(\hat Z_i(u),Y_i)\big((\hat\xi_i-\xi_i)^T\hat\beta\big)^2\}K_h(U_i-u)\Big]\\
&\stackrel{\text{def}}{=}\;\hbar_{n1}^T\eta_*+\eta_*^T\hbar_{n2}\eta_*+\hbar_{n3}.
\end{aligned}$$

Furthermore, it follows from the arguments similar to the proof of Lemma A.2 that

$$\hbar_{n2}=-\frac12 f_U(u)E\big\{\rho_2(Z)\mathrm{diag}\big(X^{\otimes2},\mu_{K_2}X^{\otimes2}\big)\big|U=u\big\}+o_P(1).$$

We now consider $\hbar_{n1}$, which equals to

$$\begin{aligned}
&r_nh\sum_{i=1}^n q_1(\hat Z_i(u),Y_i)K_h(U_i-u)X_i^u+r_nh\sum_{i=1}^n q_2(\hat Z_i(u),Y_i)(\hat\xi_i-\xi_i)^T\hat\beta K_h(U_i-u)X_i^u\\
\stackrel{\text{def}}{=}\;&\hbar_{n1}^{[1]}+\hbar_{n1}^{[2]}.
\end{aligned}$$

Step 4.1. Note that $\hat{Z}_i(u) - \tilde{Z}_i(u) = (\hat{\beta}-\beta)^T\xi_i + (\hat{\theta}-\theta)^T W_i$. Thus, using Taylor expansion yields that

$$\begin{aligned}
\hbar_{n1}^{[1]} &= \sqrt{\frac{h}{n}}\sum_{i=1}^n q_1(\tilde{Z}_i(u), Y_i)K_h(U_i-u)X_i^u \\
&\quad -\sqrt{\frac{h}{n}}\sum_{i=1}^n \rho_2(\tilde{Z}_i(u))K_h(U_i-u)X_i^u\{\xi_i^T(\hat{\beta}-\beta)+W_i^T(\hat{\theta}-\theta)\} \\
&\quad +\sqrt{\frac{h}{n}}\sum_{i=1}^n (Y_i - g^{-1}(\tilde{Z}_i(u)))\rho_1'(\tilde{Z}_i(u))K_h(U_i-u)X_i^u\{\xi_i^T(\hat{\beta}-\beta)+W_i^T(\hat{\theta}-\theta)\} \\
&\quad +\sqrt{\frac{h}{n}}\sum_{i=1}^n \frac{dg^{-1}(t)}{dt}\Big|_{t=\tilde{Z}_i(u)}\rho_1'(\tilde{Z}_i(u))K_h(U_i-u)X_i^u\{\xi_i^T(\hat{\beta}-\beta)+W_i^T(\hat{\theta}-\theta)\}^2 \\
&\overset{\text{def}}{=} \hbar_{n1}^{[1]}(1)+\hbar_{n1}^{[1]}(2)+\hbar_{n1}^{[1]}(3)+\hbar_{n1}^{[1]}(4).
\end{aligned}$$

By Theorem 1, we know that $\hat{\beta}-\beta = O_P(n^{-1/2})$, $\hat{\theta}-\theta = O_P(n^{-1/2})$. Furthermore, $(nh)^{1/2}\{\frac{1}{n}\sum_{i=1}^n \big[\rho_2(\tilde{Z}_i(u))K_h(U_i-u)X_i^u\xi_i^T - E\Big[\rho_2(\tilde{Z}_i(u))K_h(U_i-u)X_i^u\xi_i^T\Big]\ \}\big]$ converges to $N(0, A(u))$ in distribution for some positive definite matrix $A(u)$. Thus, from Lemma A.1 and $\hat{\beta}-\beta = O_P(n^{-1/2})$, we have $\sqrt{\frac{h}{n}}\sum_{i=1}^n \rho_2(\tilde{Z}_i(u))K_h(U_i-u)X_i^u\xi_i^T(\hat{\beta}-\beta) = O_P(h^{1/2}) + O_P(1)(\hat{\beta}-\beta) = o_P(1)$, and then $\hbar_{n1}^{[1]}(2) = o_P(1)$. Similarly, we can show that $\hbar_{n1}^{[1]}(3) = o_P(1)$, $\hbar_{n1}^{[1]}(4) = o_P(1)$. For $\hbar_{n1}^{[1]}(1)$, we have

$$\begin{aligned}
&(nh)^{1/2}\Big[\frac{1}{n}\sum_{i=1}^n \big\{q_1(\tilde{Z}_i(u), Y_i)K_h(U_i-u)X_i^u - E\Big[q_1(\tilde{Z}_i(u), Y_i)K_h(U_i-u)X_i^u\Big]\ \big\}\Big] \\
&\xrightarrow{\mathcal{L}} N(0, \Sigma_{*1}(u)), \qquad \text{(A.23)}
\end{aligned}$$

where $\Sigma_{*1}(u) = f_U(u)E\{\rho_2(Z)\text{diag}\big(v_{K_0}X^{\otimes 2}, v_{K_2}X^{\otimes 2}\big)\big|U=u\}$.

Step 4.2. We deal with $\hbar_{n1}^{[2]}$. In fact, we only need to consider $\sqrt{\frac{h}{n}}\sum_{i=1}^n q_2(\tilde{Z}_i(u), Y_i)(\hat{\xi}_i - \xi_i)^T\beta K_h(U_i - u)X_i^u$. By the definition of $q_2(\cdot,\cdot)$ and (2.1.3), the central limit theorem entails that

$$\begin{aligned}
&\sqrt{nh}\Big[\frac{1}{n}\sum_{i=1}^n \big\{\rho_2(\tilde{Z}_i(u))\xi^{(2)}(V_i)K_h(U_i-u)X_i^u - E\big[\rho_2(\tilde{Z}_i(u))\xi^{(2)}(V_i)K_h(U_i-u)X_i^u\big]\big\}\Big] \\
&\xrightarrow{\mathcal{L}} N(0, A^*(u)) \qquad \text{(A.24)}
\end{aligned}$$

for some matrix $A^*(u)$. From (A.24), we achieve that $0.5\mu_{L2}b^2\sqrt{\frac{h}{n}}\sum_{i=1}^n \rho_2(\tilde{Z}_i(u))K_h(U_i-u)\xi^{(2)}(V_i)^T \beta X_i^u = O_P(\sqrt{nh}b^2) + O_P(b^2) = o_P(1)$. Moreover, similar to the analysis for (A.14), we have $\frac{(nh)^{1/2}}{n^2}\sum_{i=1}^n\sum_{j=1}^n \frac{\rho_2(\tilde{Z}_i(u))}{f_V(V_i)}\left[L_b(V_j - V_i)e_j\right]^T \beta K_h(U_i-u)X_i^u = O_P(h^{1/2}) + O_P(h^{1/2}(b^2+h^2)) = o_P(1)$ by using the projection of U-statistics (Serfling (1980), Section 5.3.1)). As such, $\hbar_{n1}^{[2]} = o_P(1)$.

As a consequence, $\hbar_{n1}^{[2]} = o_P(1)$, together with Condition (C7), i.e., $nh^4 \to 0$, $nb^4 \to 0$, and (A.23), mean that $\hbar_{n1} \xrightarrow{\mathcal{L}} N(0, \Sigma_{*1}(u))$. Similar to the proof of Lemma A.2, the maximizer $\hat{\eta}$ of $\wp_n(\eta_*)$ is asymptotically equivalent to $\hat{\eta} = -2\hbar_{n2}^{-1}\hbar_{n1} + o_P(1)$. Invoking the asymptotic result of $\hbar_{n2}$ above, we complete the proof. □

A.7. *Proof of Theorem 5*

Decompose $\mathcal{H}_1 - \mathcal{H}_0 = \mathcal{I}_{n1} + \mathcal{I}_{n2} + \mathcal{I}_{n3} + \mathcal{I}_{n4}$, where

$$\mathcal{I}_{n1} = \sum_{i=1}^{n} \big[\mathcal{Q}\{g^{-1}\big(\hat\alpha(U_i)^T X_i + \hat\beta^T \hat\xi_i + \hat\theta^T W_i\big)\} - \mathcal{Q}\{g^{-1}\big(\hat\alpha(U_i)^T X_i + \hat\beta^T \xi_i + \hat\theta^T W_i\big)\}\big],$$
$$\mathcal{I}_{n2} = \sum_{i=1}^{n} \big[\mathcal{Q}\{g^{-1}\big(\hat\alpha(U_i)^T X_i + \hat\beta^T \xi_i + \hat\theta^T W_i\big)\} - \mathcal{Q}\{g^{-1}\big(\hat\alpha(U_i)^T X_i + \beta^T \xi_i + \theta^T W_i\big)\}\big],$$
$$\mathcal{I}_{n3} = \sum_{i=1}^{n} \left[\mathcal{Q}\left\{g^{-1}\left(\hat\alpha(U_i)^T X_i + \beta^T \xi_i + \theta^T W_i\right)\right\} - \mathcal{Q}\left\{g^{-1}\left(\beta^T \xi_i + \theta^T W_i\right)\right\}\right],$$
$$\mathcal{I}_{n4} = \sum_{i=1}^{n} \left[\mathcal{Q}\left\{g^{-1}\left(\beta^T \xi_i + \theta^T W_i\right)\right\} - \mathcal{Q}\left\{g^{-1}\left(\bar\beta^T \xi_i + \bar\theta^T W_i\right)\right\}\right].$$

By similar arguments used in the proof of Theorem 3 in Li and Liang (2008), we can show that $2\mathcal{I}_{n2} \xrightarrow{\mathcal{L}} \chi^2_p$ and $2\mathcal{I}_{n4} \xrightarrow{\mathcal{L}} \chi^2_p$. Next, we analyze $\mathcal{I}_{n1}$ and $\mathcal{I}_{n3}$.

Step 5.1. Taylor expansion entails that

$$\begin{aligned}\mathcal{I}_{n1} \quad &\stackrel{\text{def}}{=} \quad \mathcal{I}_{n1}^{[1]} + \mathcal{I}_{n1}^{[2]} = \sum_{i=1}^{n} q_1\big(\hat\alpha(U_i)^T X_i + \beta^T \xi_i + \hat\theta^T W_i, Y_i\big)(\hat\xi_i - \xi)^T \hat\beta \\ &\qquad + \frac{1}{2}\sum_{i=1}^{n} q_2\big(\hat\alpha(U_i)^T X_i + \beta^T \xi_i + \hat\theta^T W_i, Y_i\big)\big((\hat\xi_i - \xi)^T \hat\beta\big)^2.\end{aligned} \tag{A.25}$$

Step 5.1.1. Under H_0, by Theorem 4, we know that $\hat\alpha(u) = O_P((nh)^{-1/2})$, and $Z_i = \beta^T \xi_i + \theta^T W_i$. Define $\Delta Z_i = \hat\alpha(U_i)^T X_i + (\hat\beta - \beta)^T \xi_i + (\hat\theta - \theta)^T W_i$. Thus, Taylor expansion entails that

$$\begin{aligned}\mathcal{I}_{n1}^{[1]} \quad &= \quad \sum_{i=1}^{n} q_1\left(Z_i, Y_i\right)(\hat\xi_i - \xi)^T \hat\beta + \sum_{i=1}^{n}\left[\big(Y_i - g^{-1}(Z_i)\big)\rho'(Z_i) + \rho_2(Z_i)\right](\hat\xi_i - \xi)^T \hat\beta \Delta Z_i \\ &\qquad + \sum_{i=1}^{n}\Big[\Big(\frac{dg^{-1}(t)}{dt}\Big|_{t=Z_i}\Big)\rho'(Z_i)\Big](\hat\xi_i - \xi)^T \hat\beta (\Delta Z_i)^2 \\ &\stackrel{\text{def}}{=} \quad \mathcal{I}_{n1}^{[1]}(1) + \mathcal{I}_{n1}^{[1]}(2) + \mathcal{I}_{n1}^{[1]}(3).\end{aligned}$$

Similar to the arguments for (A.11), as $nh^4 \to 0$, we have $\mathcal{I}_{n1}^{[1]}(1) = O_P(1) = o_P(h^{-1/2})$. For the argument $\mathcal{I}_{n1}^{[1]}(3)$, using Lemma A.1, Theorems 2 and 4, we know that $\mathcal{I}_{n1}^{[1]}(3) = O_P\big(n\big(b^2 + \big(\frac{\log b^{-1}}{nb}\big)^{1/2}\big)/nh\big)$. Under Condition (C7), we can have that $\mathcal{I}_{n1}^{[1]}(3) = o_P(h^{-1/2})$. For the argument $\mathcal{I}_{n1}^{[1]}(2)$, invoking (2.1.3), Theorems 1 and 4, similar to the arguments for (A.11) and (A.12), we have $b^2 \sum_{i=1}^{n}\left[\big(Y_i - g^{-1}(Z_i)\big)\rho'(Z_i) + \rho_2(Z_i)\right]\xi(V_i)^T \hat\beta \times \big(\hat\alpha(U_i)^T X_i + (\hat\beta - \beta)^T \xi_i + (\hat\theta - \theta)^T W_i\big) = O_P(n^{1/2} b^2) + O_P(b^2) = o_P(1)$. Furthermore, using the projection of U-statistics (Serfling; 1980, Section 5.3.1), we have

$$\begin{aligned}&\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\left[\big(Y_i - g^{-1}(Z_i)\big)\rho'(Z_i) + \rho_2(Z_i)\right]\frac{\left[L_b(V_i - V_j)e_j\right]^T \hat\beta}{f_V(V_i)} \times \\ &\qquad \big(\hat\alpha(U_i)^T X_i + (\hat\beta - \beta)^T \xi_i + (\hat\theta - \theta)^T W_i\big) = O_P(1) = o_P(h^{-1/2}).\end{aligned}$$

Thus, we obtain that $\mathcal{I}_{n1}^{[1]}(2) = o_P(h^{-1/2})$, and $\mathcal{I}_{n1}^{[1]} = o_P(h^{-1/2})$.

Step 5.1.2. By the definition of $q_2(\cdot,\cdot)$, we have

$$\begin{aligned}
\mathcal{I}_{n1}^{[2]} \stackrel{\text{def}}{=} \mathcal{I}_{n1}^{[2]}(1) - \mathcal{I}_{n1}^{[2]}(2) &= 0.5\sum_{i=1}^{n}(Y_i - g^{-1}(Z_i+\Delta Z_i))\rho'(Z_i+\Delta Z_i)\big((\hat{\xi}_i-\xi)^T\hat{\beta}\big)^2 \\
&\quad -0.5\sum_{i=1}^{n}\rho_2(Z_i+\Delta Z_i)\big((\hat{\xi}_i-\xi)^T\hat{\beta}\big)^2.
\end{aligned}$$

Taylor expansion of $g^{-1}(Z_i+\Delta Z_i)$ and $\rho'(Z_i+\Delta Z_i)$ with respect to Z_i yields that

$$\begin{aligned}
\mathcal{I}_{n1}^{[2]}(1) &= 0.5\sum_{i=1}^{n}\big(Y_i - g^{-1}(Z_i)\big)\left(\rho'(Z_i)+\Delta Z_i\right)\big((\hat{\xi}_i-\xi)^T\hat{\beta}\big)^2 \\
&\quad +0.5\sum_{i=1}^{n}\Big[\frac{dg^{-1}(t)}{dt}\Big|_{t=Z_i}\Big](\rho'(Z_i)+\Delta Z_i)\big((\hat{\xi}_i-\xi)^T\hat{\beta}\big)^2\Delta Z_i.
\end{aligned}$$

Together with (2.1.3) and Lemma A.1, we know that $\big\{(\hat{\xi}_i-\xi)^T\hat{\beta}\big\}^2 = O_P(b^4+\log b^{-1}/(nb))$. Thus, we have $\sqrt{nh}\big\{(\hat{\xi}_i-\xi)^T\hat{\beta}\big\}^2 = o_P(n^{-1/4})$ since $nh_o^6\to 0$ and $nh_o^2/(\log h_o^{-1})^4\to\infty$. So Lemma A.1 in Li and Liang (2008) entails that $\frac{1}{\sqrt{nh}}\sum_{i=1}^n\big(Y_i-g^{-1}(Z_i)\big)\left(\rho'(Z_i)+\Delta Z_i\right)\big\{\sqrt{nh}\big((\hat{\xi}_i-\xi)^T\hat{\beta}\big)^2\big\} = o_P(h^{-1/2})$. By Theorems 1 and 4, we know that, under H_0, $\Delta Z_i = O_P(1/\sqrt{nh})$. Again together with the result $\big\{(\hat{\xi}_i-\xi)^T\hat{\beta}\big\}^2 = O_P(b^4+\log b^{-1}/(nb))$, we have

$$\begin{aligned}
&\frac{1}{2}\sum_{i=1}^{n}\big[\frac{dg^{-1}(t)}{dt}\big|_{t=Z_i}\big](\rho'(Z_i)+\Delta Z_i)\big\{(\hat{\xi}_i-\xi)^T\hat{\beta}\big\}^2\Delta Z_i \\
&\qquad = nO_P\big(1/\sqrt{nh}\big)O_P\big(b^4+\log b^{-1}/(nb)\big) = o_P(h^{-1/2}),
\end{aligned}\tag{A.26}$$

since $nh_o^6\to 0$ and $nh_o^2/(\log h_o^{-1})^2\to\infty$. Thus, we have $\mathcal{I}_{n1}^{[2]}(1) = o_P(h^{-1/2})$.

For the argument $\mathcal{I}_{n1}^{[2]}(2)$, Taylor expansion $\rho_2(Z_i+\Delta Z_i) = \rho_2(Z_i)+\rho_2'(Z_i)\Delta Z_i$ and the fact that $\sum_{i=1}^n\rho_2'(Z_i)\big((\hat{\xi}_i-\xi)^T\hat{\beta}\big)^2\Delta Z_i = o_P(h^{-1/2})$ yield

$$\begin{aligned}
\mathcal{I}_{n1}^{[2]}(2) &= \frac{b^4}{2}\sum_{i=1}^{n}\rho_2(Z_i)\big\{\xi^{(2)}(V_i)^T\beta\big\}^2 + \frac{1}{2}\sum_{i=1}^{n}\rho_2(Z_i)\Big\{\frac{1}{nf_V(V_i)}\sum_{j=1}^{n}\left[L_b(V_j-V_i)e_j\right]^T\beta\Big\}^2 \\
&\quad +\frac{b^2}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\rho_2(Z_i)}{f_V(V_j)}\big\{\xi^{(2)}(V_i)^T\beta\big\}\big\{L_b(V_i-V_j)e_j\big\}^T\beta + o_P(h^{-1/2}).
\end{aligned}\tag{A.27}$$

As $nb^4\to 0$, $b^4\sum_{i=1}^n\rho_2(Z_i)\big\{\xi^{(2)}(V_i)^T\beta\big\}^2 = o_P(1)$. By using the projection of U-statistics (Serfling; 1980, Section 5.3.1), the third term of (A.27) is $O_P(n^{1/2}b^2) = o_P(1)$, provided $nb^4\to 0$. Next, we analyze the second term of (A.27) in detail. We re-write it as

$$\begin{aligned}
&\frac{1}{2}\sum_{i=1}^{n}\rho_2(Z_i)\Big\{\frac{1}{nf_V(V_i)}\sum_{j=1}^{n}\left[L_b(V_j-V_i)e_j\right]^T\beta\Big\}^2 \\
&= \frac{1}{2n^2b^2}\sum_{i=1}^{n}\frac{\rho_2(Z_i)}{f_V^2(V_i)}L(0)^2\left\{e_i^T\beta\right\}^2 + \frac{1}{2n^2}\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n}\frac{\rho_2(Z_i)}{f_V^2(V_i)}\left\{\left[L_b(V_j-V_i)e_j\right]^T\beta\right\}^2 \\
&\quad +\frac{1}{2n^2}\sum_{i=1}^{n}\sum_{s=1}^{n}\sum_{j=1,j\neq s}^{n}\frac{\rho_2(Z_i)}{f_V^2(V_i)}\left\{\left[L_b(V_s-V_i)e_s\right]^T\beta\right\}\left\{\left[L_b(V_j-V_i)e_j\right]^T\beta\right\}.
\end{aligned}\tag{A.28}$$

By the law of large numbers, we know that the first term of (A.28) is $O_P(1/(nb^2)) = o_P(1)$, provided $nb^2 \to \infty$. For the second term of (A.28), the convergence of U-statistics entails that

$$\frac{1}{2n^2}\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n}\frac{\rho_2(Z_i)}{f_V^2(V_i)}\left\{[L_b(V_j-V_i)e_j]^T\beta\right\}^2$$
$$= b^{-1}v_{L_0}\beta^T\Sigma_e\beta E\{E(\rho_2(Z_i)|V_i)/f_V(V_i)\} + O_P\left(1/\sqrt{nb^2}\right) + O_P(b). \tag{A.29}$$

For the third term of (A.28), we have

$$\frac{1}{2n^2}\sum_{i=1}^{n}\sum_{s=1}^{n}\sum_{j=1,j\neq s}^{n}\frac{\rho_2(Z_i)}{f_V^2(V_i)}\left\{[L_b(V_s-V_i)e_s]^T\beta\right\}\left\{[L_b(V_j-V_i)e_j]^T\beta\right\}$$
$$= \frac{L(0)}{n^2b}\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n}\frac{\rho_2(Z_i)}{f_V^2(V_i)}e_i^T\beta\left\{[L_b(V_j-V_i)e_j]^T\beta\right\}$$
$$+\frac{1}{2n^2}\sum_{1\leq i\neq j,j\neq s,i\neq s\leq n}\frac{\rho_2(Z_i)}{f_V^2(V_i)}\left\{[L_b(V_s-V_i)e_s]^T\beta\right\}\left\{[L_b(V_j-V_i)e_j]^T\beta\right\}. \tag{A.30}$$

Using the second order of the projection of U-statistics (Serfling; 1980, Section 5.3.2), we know that the first term of (A.30) satisfies $\frac{L(0)}{n^2b}\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n}\frac{\rho_2(Z_i)}{f_V^2(V_i)}e_i^T\beta\left\{[L_b(V_j-V_i)e_j]^T\beta\right\} = O_P(1/nb)$. We directly adopt the arguments used in (Fan et al.; 2001, Pages 182-186) and know that the second term of (A.30)equals

$$\frac{b^{-1/2}}{2}\left\{\frac{\sqrt{b}}{nb}\sum_{j\neq l}\frac{\{E(\rho_2(Z_l)|V_l)\}}{f_V(V_l)}e_j^T\beta e_l^T\beta\left[L*L\big((V_j-V_l)/b\big)\right]\right\} + o_P(b^{-1/2})$$
$$\stackrel{\text{def}}{=} \frac{b^{-1/2}}{2}W_b(n) + o_P(b^{-1/2}). \tag{A.31}$$

Using Proposition 3.2 in de Jong (1987) or Fan et al. (2001, page 186), it can be shown that $W_b(n) \xrightarrow{\mathcal{L}} N(0,\sigma_L^2)$ with

$$\sigma_L^2 = 2\left\{\int[L*L(s)]^2ds\right\}^2 E\left\{\frac{\{E(\rho_2(Z)|V)\}^2}{f_V(V)}\right\}\times(\beta^T\Sigma_e\beta)^2.$$

A combination of (A.27)-(A.31), together with that $\mathcal{I}_{n1}^{[1]} = o_P(h_o^{-1/2})$, $\mathcal{I}_{n1}^{[2]}(1) = o_P(h_o^{-1/2})$, indicates that

$$\mathcal{I}_{n1} = -\frac{v_{L0}\beta^T\Sigma_e\beta}{c_bh_o}E\left\{\frac{E(\rho_2(Z_i)|V_i)}{f_V(V_i)}\right\} - \frac{c_b^{-1/2}h_o^{-1/2}}{2}W_b(n) + o_P(h_o^{-1/2}). \tag{A.32}$$

Step 5.2. Taylor expansion entails that $\mathcal{I}_{n3} = \sum_{i=1}^n q_1(Z_i,Y_i)\hat{\alpha}(U_i)^TX_i + \sum_{i=1}^n q_2(Z_i,Y_i)\{\hat{\alpha}(U_i)^TX_i\}^2$. Recalling the proof of Theorem 4, we have, under H_0, $\sqrt{nh}\hat{\alpha}(u) = -2\hbar_{n2}^{-1}\hbar_{n1}^{[1]} - 2\hbar_{n2}^{-1}\hbar_{n1}^{[2]}$. By the results obtained in Step 4.2, we know that $-2\hbar_{n2}^{-1}\hbar_{n1}^{[2]} = O_P(\sqrt{nh}b^2 + h^{1/2}) = o_P(1)$, as $nh_o^4 \to 0$. As such, we only need to consider the main term $-2\hbar_{n2}^{-1}\hbar_{n1}^{[1]}$. It follows from a direct use of Theorem 10 of Fan et al. (2001) that

$$\mathcal{I}_{n3} = -\frac{p|\mathcal{U}|}{c_hh_o}[K(0) - 0.5v_{K0}] - \frac{c_h^{-1/2}h_o^{-1/2}}{2}W_h(n) + o_P(h_o^{-1/2}), \tag{A.33}$$

where $v_{K_0} = \int K^2(u)du$ and $W_h(n) \xrightarrow{\mathcal{L}} N(0, \sigma_K^2)$ with $\sigma_K^2 = 2p\left\{\int[2K(s) - K * K(s)]^2 ds\right\}^2 |\mathcal{U}|$. Note that $\hbar_{n1}^{[1]}$ only involves the argument $\{Y_i, U_i, \xi_i, W_i, X_i\}$, so does $W_h(n)$. Furthermore, e is independent of (Y, V, W, X). So we have $EW_b(n)W_h(n) = 0$. Recall that $2\mathcal{I}_{n2} \stackrel{\mathcal{L}}{\to} \chi_p^2 = o_P(h_o^{-1/2})$ and $2\mathcal{I}_{n4} \stackrel{\mathcal{L}}{\to} \chi_p^2 = o_P(h_o^{-1/2})$. Thus, together with (A.32) and (A.33), we have that

$$\begin{aligned} \mathcal{H}_1 - \mathcal{H}_0 &= -\frac{v_{L_0}\beta^T\Sigma_e\beta}{c_b h_o} E\left\{\frac{E(\rho_2(Z_i)|V_i)}{f_V(V_i)}\right\} - \frac{p|\mathcal{U}|}{c_h h_o}[K(0) - 0.5v_{K_0}] \\ &\quad -\frac{h_o^{-1/2}}{2}\left\{c_b^{-1/2}W_b(n) + c_h^{-1/2}W_h(n)\right\} + o_P(h_o^{-1/2}). \end{aligned} \tag{A.34}$$

Note that $c_b^{-1/2}W_b(n) + c_h^{-1/2}W_h(n) \xrightarrow{\mathcal{L}} N(0, c_h^{-1}\sigma_K^2 + c_b^{-1}\sigma_L^2)$. By (A.34), we know that $r_{LK}\{\mathcal{H}_0 - \mathcal{H}_1 - \chi_{df_n}^2\} \xrightarrow{\mathcal{L}} 0$. The proof is complete.

REFERENCES

Carroll, R. J., Fan, J., Gijbels, I. and Wand, M. P. (1997). Generalized partially linear single-index models, *Journal of the American Statistical Association* **92**(438): 477–489.

de Jong, P. (1987). A central limit theorem for generalized quadratic forms, *Probability Theory and Related Fields* **75**: 261–277.

Fan, J., Zhang, C. M. and Zhang, J. (2001). Generalized likelihood ratio statistics and wilkes phenomenon, *The Annals of Statistics* **29**: 153–193.

Li, R. and Liang, H. (2008). Variable selection in semiparametric regression modeling, *The Annals of Statistics* **36**: 261–286.

Mack, Y. P. and Silverman, B. W. (1982). Weak and strong uniform consistency of kernel regression estimates, *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* **61**: 405–415.

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

Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics*, John Wiley & Sons Inc., New York.