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


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 $g''\alpha(\cdot)$ 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) $E[q_l^s(Z,Y)N^2|U = u], E[q_l^s(Z,Y)N^2|V = v]$ and $E[q_l^s(Z,Y)N^2|U = u, V = v]$ for $l = 1, 2$, $s = 1, 2$ are Lipschitz continuous and twice differentiable on $u \in \mathcal{U}$ and $v \in \mathcal{V}$. Moreover, $E\{q_2^3(Z,Y)\} \leq \infty, E\{q_1^{2+\delta}(Z,Y)\} \leq \infty$ for some $\delta > 2$ and $E[\rho_2(Z)N^2|U = u]$ 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 $\int t^2 K(t) dt \neq 0$, $\int t^2 L(t) dt \neq 0$, and $\int |t|^2 K(t) dt < \infty$, $\int |t|^2 L(t) dt < \infty$, 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) $b = b_k, k = 1, \ldots, d, b \asymp c_b h_o$ for some constant $c_b > 0$; $h \asymp c_h h_o$ for some constant $c_h > 0$.

(ii) $h_o \to 0$ as $n \to \infty$, $nh_o^2/(\log h_o)^4 \to \infty$, $nh^4 \to 0$.

(C8) For all $\lambda_{1j}, \lambda_{2s}, j = 1, \ldots, d$, $s = 1, \ldots, r$, $\lambda_{1j} \to 0, \sqrt{n}\lambda_{1j} \to \infty$, $\lambda_{2s} \to 0, \sqrt{n}\lambda_{2s} \to \infty$, and

$$\lim_{n \to \infty} \inf_{u \to 0+} \lim_{s \to 0+} \lambda_{1j}^{-1}(u)/\lambda_{1j} > 0, \lim_{n \to \infty} \inf_{u \to 0+} \lim_{s \to 0+} \lambda_{2s}^{-1}(u)/\lambda_{2s} > 0.$$
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

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

provided that \(n^{2r-1}h \to \infty\) 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 \(\sigma^2\) into \(T(\cdot)\), i.e., \(\text{Var}(Y|U,\xi,W,X) = T(\mu,\xi,W,X)\). Write \(Z = \beta^T\xi + \theta^TW + \alpha(U)^TX, \tilde{Z}_i(u) = \beta^T\xi_i + \theta^TW_i + X_i^T\alpha(u) + (U_i - u)X_i^T\alpha'(u), N_i = (X_i^T, \xi_i^T,W_i^T)^T, R_i^u = (N_i^T, h^{-1}(U_i - u)X_i^T)^T\) and further denote that \(L_b(\cdot) = \text{diag}(L_{b_1}(\cdot), \ldots, L_{b_k}(\cdot))\). Denote the local quasi-likelihood estimators from (2.2.2) by \(\hat{\alpha}_*(u), \hat{B}_*(u), \hat{\beta}_*, \hat{\theta}_*\) and let

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

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

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

\[
\delta = A_u^{-1}B_u^u + O_P\left(h_o^{2}\sqrt{nh_o} + \{\log h_o^{-1}\}/\{nh_o\}\right)^{1/2}, \tag{A.2}
\]

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

\[
B_u^{u} \overset{\text{def}}{=} \sqrt{\frac{n}{h}} \sum_{i=1}^{n} q_1(\tilde{Z}_i(u), Y_i)R_i^u K_h(U_i - u) - \sqrt{\frac{n}{h}} \sum_{i=1}^{n} f_U(V_i) f_U(V_i) E\{\rho_2(Z)(N^T, 0)^T|U = u, V = V_i\} e_i^T \beta.
\]

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

\[
\ell_n(\delta) = h \sum_{i=1}^{n} \left[ Q\{g^{-1}(r_n\delta_i^T R_i^u + \hat{\omega}_i + \tilde{Z}_i(u), Y_i)\} - Q\{g^{-1}(\tilde{Z}_i(u), Y_i)\} \right] K_h(U_i - u)
\]

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

\[
\ell_n(\delta) = \left[ r_n h \sum_{i=1}^{n} \left\{ q_1(\tilde{Z}_i(u), Y_i) + q_2(\tilde{Z}_i(u), Y_i)\hat{\omega}_i \right\} K_h(U_i - u) R_i^u \right] \delta_* + \delta_* \left[ \frac{r_n^2 h}{2} \sum_{i=1}^{n} q_2(\tilde{Z}_i(u), Y_i)(R_i^u)^\otimes 2 K_h(U_i - u) \right] \delta_* + h \sum_{i=1}^{n} \left\{ q_1(\tilde{Z}_i(u), Y_i)\hat{\omega}_i + 0.5q_2(\tilde{Z}_i(u), Y_i)\hat{\omega}_i^2 \right\} K_h(U_i - u) \right]
\]

\[
\overset{\text{def}}{=} h_n^1 \delta_* + \delta_* h_n^2 \delta_* + h_n^3.
\]
Step 1.1. Note that \( h_{n3} \) is irrelevant to \( \delta_* \). We show that \( h_{n3} = O_P(1) \).

\[
h_{n3} = \frac{h}{n} \sum_{i=1}^{n} q_1(\tilde{Z}_i(u), Y_i)\hat{\omega}_i K_h(U_i - u) + \frac{h}{n} \sum_{i=1}^{n} 0.5q_2(\tilde{Z}_i(u), Y_i)\hat{\omega}_i^2 K_h(U_i - u)
\]

\[
= \defeq h_{n3}^{[1]} + h_{n3}^{[2]}.
\]

By using (2.1.3), we obtain that

\[
h_{n3}^{[1]} = \frac{hb^2\mu_{L_2}}{2} \sum_{i=1}^{n} q_1(\tilde{Z}_i(u), Y_i)\xi^{(2)}(V_i)^T \beta K_h(U_i - u)(1 + o_P(1))
\]

\[
+ \frac{h}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{q_1(\tilde{Z}_i(u), Y_i)}{f_{V}(V_i)} \{L_b(V_j - V_i)e_j\}^T \beta K_h(U_i - u)(1 + o_P(1))
\]

\[
= \defeq h_{n3}^{[1]}(1) + h_{n3}^{[1]}(2).
\]

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

\[
h_{n3}^{[1]}(1) = \frac{nh^3b^2\mu_{L_2}^2 f_U(u)\alpha''(u)^T E\{X\xi^{(2)}(V)^T \beta \rho_2(Z)\}}{4} + O_P(nh^4b^3) + O_P(nh^2\{\log h^{-1}/nh\}^{1/2}).\tag{A.3}
\]

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

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

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

Next, we consider the second argument \( h_{n3}^{[2]} \).

\[
h_{n3}^{[2]} = \frac{h}{2} \sum_{i=1}^{n} (Y_i - g^{-1}(\tilde{Z}_i(u)))\rho_1(\tilde{Z}_i(u))K_h(U_i - u)e_i - \frac{h}{2} \sum_{i=1}^{n} \rho_2(\tilde{Z}_i(u))K_h(U_i - u)e_i^2
\]

\[
= \defeq h_{n3}^{[2]}(1) - h_{n3}^{[2]}(2).
\]

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

\[
h_{n3}^{[2]}(2) = \frac{hb^4\mu_{L_2}}{2} \sum_{i=1}^{n} \rho_2(\tilde{Z}_i(u))K_h(U_i - u)(\xi^{(2)}(V_i)^T \beta)^2
\]

\[
+ \frac{hb^2\mu_{L_2}}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{\rho_2(\tilde{Z}_i(u))}{f_{V}(V_i)} \xi^{(2)}(V_i)^T \beta K_h(U_i - u)\{L_b(V_j - V_i)e_j\}^T \beta
\]

\[
+ \frac{hn^2}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{s=1}^{n} \frac{\rho_2(\tilde{Z}_i(u))}{f_{V}(V_i)} K_h(U_i - u)\{L_b(V_j - V_i)e_j\}^T \beta \{L_b(V_s - V_i)e_s\}^T \beta
\]

\[
= \defeq h_{n3}^{[2]}(1) + h_{n3}^{[2]}(2) + h_{n3}^{[2]}(3).
\]
Note that \( h_{n3}^{[2]}(2) \) is a sum of independent and identically distributed random variables. We have \( h_{n3}^{[2]}(2)_1 = O_P(nh^\delta) = O_p(1) \). Furthermore, similar to (A.4), we have \( h_{n3}^{[2]}(2)_2 = O_P(\sqrt{nh^2b^3}) = O_p(1) \). For the argument \( h_{n3}^{[2]}(2)_3 \), we can use the projection of \( U \)-statistics with second order (Serfling; 1980, Section 5.5.2) to show that \( h_{n3}^{[2]}(2)_3 = O_P(1) \). As a result, we obtain that \( h_{n3}^{[2]} = h_{n3}^{[2]}(1) + h_{n3}^{[2]}(2) = O_P(1) \), yielding \( h_{n3} = h_{n3}^{[1]} + h_{n3}^{[2]} = O_P(1) \).

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

\[
h_{n2} = h_{n2}^{[1]} + h_{n2}^{[2]} = \left[ \frac{1}{2n} \sum_{i=1}^{n} \{ Y_i - g^{-1}(\tilde{Z}_i(u)) \} \rho_1'(\tilde{Z}_i(u)) (R_i^u)^{\otimes 2} K_h(U_i - u) \right] + \left[ -\frac{1}{2n} \sum_{i=1}^{n} \rho_2(\tilde{Z}_i(u)) (R_i^u)^{\otimes 2} K_h(U_i - u) \right].
\]

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

\[
h_{n2}^{[2]} = -0.5E \left[ \frac{1}{n} \sum_{i=1}^{n} \rho_2(\tilde{Z}_i(u)) (R_i^u)^{\otimes 2} K_h(U_i - u) \right] + O_P \left( \{ \log h^{-1}/nh \}^{1/2} \right) = -0.5f_U(u)E \left[ \rho_2(Z)S_\mu(\xi, W, X) | U = u \right] + O_P \left( h^2 + \{ \log h^{-1}/nh \}^{1/2} \right).
\]

Thus, we have

\[
h_{n2} = h_{n2}^{[1]} + h_{n2}^{[2]} = -0.5A_u + O_P \left( h^2 + \{ \log 1/h_o/\on (n) \}^{1/2} \right). \tag{A.5}
\]

**Step 1.3.** From Step 1.1, we know that \( \delta_* \) is irrelevant to \( h_{n3} \) and we had \( h_{n3} = O_P(1) \). Thus, we obtain that \( \ell_n(\delta_*) - h_{n3} = h_{n1}^T \delta_* + \delta_*^T h_{n2} \delta_* \). As claimed in the proof of Lemma A.2 in Li and Liang (2008), each element of \( h_{n2} \) 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 \( \delta_* \) such that \( \sup_{u \in \mathcal{U}} | \delta_* - A_u^{-1} h_{n1} | = O_P(1) \). So, the derivative of \( \ell_n(\delta_*) \) or \( \{ \ell_n(\delta_*) - h_{n3} \} \) with respect to \( \delta_* \) further guarantees a consistent estimator of \( \delta_* \), uniformly in \( u \in \mathcal{U} \). That is

\[
\frac{\partial}{\partial \delta_*} \ell_n(\delta_*) \bigg|_{\delta_* = \delta_*} = r_nh \sum_{i=1}^{n} q_1(\tilde{Z}_i(u) + r_n \delta_*^T R_i^u + \tilde{\omega}_i, Y_i) R_i^u K_h(U_i - u) = 0. \tag{A.6}
\]

Taking Taylor expansion on (A.6) with respect to \( \tilde{Z}_i(u) \) yields

\[
h_{n1} + 2h_{n2}\delta_* + \frac{1}{2} \sum_{i=1}^{n} q_3(\tilde{Z}_i(u) + \tilde{\zeta}_i, Y_i) (\delta_*^T R_i^u)^2 (R_i^u) K_h(U_i - u) = 0, \tag{A.7}
\]

where \( \tilde{\zeta}_i \) is between 0 and \( r_n \delta_*^T R_i^u + \tilde{\omega}_i \), and \( q_3(x, y) = \frac{\partial^3}{\partial x^2 \partial y} Q_\xi^{-1}(g^{-1}(x), y) \). Similar to the proof of Lemma 2 in Li and Liang (2008), we know the last term in (A.7) is of order \( O_P(nr_n^2h) = O_P(1/\sqrt{nh}) \). From (A.5) in Step 1.2, we know \( h_{n2} = -0.5A_u + O_P \left( h^2 + \{ \log 1/h_o/\on (n) \}^{1/2} \right) \). From (A.7), we obtain that

\[
\delta_* = -(2h_{n2})^{-1} h_{n1} + O_P(1/\sqrt{nh}) = A_u^{-1} h_{n1} + O_P \left( h^2 + \{ \log 1/h_o/\on (n) \}^{1/2} \right). \tag{A.8}
\]
For the argument $h_{n1}$, similar to $h_{n1}^{[2]}$ in Step 1.1, by using (2.1.3), we have

$$h_{n1} = \sqrt{\frac{h}{n}} \sum_{i=1}^{n} q_2(\tilde{Z}_i(u), Y_i)R_i^T K_h(U_i - u) + \sqrt{\frac{h \beta^2}{n}} \sum_{i=1}^{n} \sigma_i \tilde{Z}_i(u) \xi_i^2 V_i^T \beta R_i^T K_h(U_i - u)$$

$$+ \sqrt{\frac{h}{n^2}} \sum_{i=1}^{n} \sum_{j=1}^{n} q_2(\tilde{Z}_i(u), Y_i) \{L_0(V_j - V_i)e_j\}^T \beta R_i^T K_h(U_i - u)$$

$$\overset{\text{def}}{=} h_{n1}^{[1]} + h_{n1}^{[2]} + h_{n1}^{[3]}.$$ 

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

$$\hat{\delta}_n = -A_n^{-1} \{h_{n1}^{[1]} + h_{n1}^{[3]}\} + O_P(h_{n1}^{2} + \sqrt{nh\beta^2} + \{\log h^{-1}/nh\})^{1/2}.$$  \hspace{1cm} (A.9)

A direct calculation yields

$$E(h_{n1}^{[1]}) = 0.5\sqrt{nh\beta^2} \mu_{2K2\alpha''} (u)^T f_U(u)E\{\rho_2(Z)(N^T, 0)^T X|U = u\} + O_P(\sqrt{nh\beta^2}),$$

$$\text{Var}(h_{n1}^{[1]}) = f_U(u)E\{\rho_2(Z)C_{\alpha'}(\xi, W, X)|U = u\} + O_P(h^5).$$

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

$$h_{n1}^{[3]} = -\sqrt{\frac{h}{n}} \sum_{i=1}^{n} \frac{f_U(U_i, V_i)}{f(V_i)} E\{\rho_2(Z)(N^T, 0)^T |U = u, V = V_i\} e_i^T \beta + O_P(\sqrt{nh(h^2 + \beta^2)}).$$

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

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

$$\mathcal{L}_P \left\{ \left( \frac{\beta_1}{\theta_1} \right), \left( \frac{\theta_1}{0} \right) \right\} \leq \max_{\|\beta_2^*\| \leq C_0 n^{-1/2}, \|\theta_2^*\| \leq C_0 n^{-1/2}} \mathcal{L}_P \left\{ \left( \frac{\beta_1}{\beta_2^*} \right), \left( \frac{\theta_1}{\theta_2^*} \right) \right\}$$

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

A.3. Proof of Theorem 1

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

$$\mathcal{L}_n(\sigma) = \sum_{i=1}^{n} \left[ Q\{g^{-1}(n^{-1/2}\sigma_i^T Q_i + \omega_i + \tilde{Z}_i), Y_i\} - Q\{g^{-1}(\tilde{Z}_i), Y_i\} \right]$$
with respect to $\sigma_*$. Taylor expansion entails that
\[
\Theta_n(\sigma_*) = \left[n^{-1/2} \sum_{i=1}^{n} \left\{ q_1(\tilde{Z}_i, Y_i) Q_i + q_2(\tilde{Z}_i, Y_i) Q_i \omega_i \right\}^T \right] \sigma_* + \sigma_*^T \left[\frac{1}{2n} \sum_{i=1}^{n} q_2(\tilde{Z}_i, Y_i) Q_i^{\otimes 2} \right] \sigma_* + \sigma_*^T \left[\frac{1}{2n} \sum_{i=1}^{n} q_2(\tilde{Z}_i, Y_i) Q_i^{\otimes 2} \right] \sigma_*
\]
\[
+ \frac{1}{2n} \sum_{i=1}^{n} \left\{ q_1(\tilde{Z}_i, Y_i) \omega_i + 0.5q_2(\tilde{Z}_i, Y_i) \omega_i^2 \right\}
\]
\[
\overset{\text{def}}{=} \theta_n^T \sigma_* + \sigma_*^T \theta_n \sigma_* + \theta_n^T \sigma_3.
\]

The proof of this theorem is similar to that of Lemma A.2. Note that $\vartheta_n^3$ is irrelevant to $\sigma_*$. As such, we consider only $\vartheta_n^1$ and $\vartheta_n^2$.  

**Step 2.1.** For the argument $\vartheta_n^2$, we have
\[
\vartheta_n^2 \overset{\text{def}}{=} \vartheta_n^{[1]} - \vartheta_n^{[2]} = \left[\frac{1}{2n} \sum_{i=1}^{n} (Y_i - g^{-1}(\tilde{Z}_i)) \rho_1^*(\tilde{Z}_i) Q_i^{\otimes 2} \right] - \left[\frac{1}{2n} \sum_{i=1}^{n} \rho_2(\tilde{Z}_i) Q_i^{\otimes 2} \right].
\]

We now consider $\vartheta_n^{[1]}$. Taylor expansion entails that
\[
\vartheta_n^{[1]} = \frac{1}{2n} \sum_{i=1}^{n} (Y_i - g^{-1}(\tilde{Z}_i)) \rho_1^*(\tilde{Z}_i) Q_i^{\otimes 2} + \frac{1}{2n} \sum_{i=1}^{n} (Y_i - g^{-1}(\tilde{Z}_i)) \rho_1^*(\tilde{Z}_i)(\tilde{Z}_i - Z_i) Q_i^{\otimes 2}
\]
\[
= \frac{1}{2n} \sum_{i=1}^{n} \frac{dg^{-1}(t)}{dt} \bigg|_{t=Z_i} \rho_1^*(\tilde{Z}_i)(\tilde{Z}_i - Z_i) Q_i^{\otimes 2} + \frac{1}{2n} \sum_{i=1}^{n} \frac{dg^{-1}(t)}{dt} \bigg|_{t=Z_i} \rho_1^*(\tilde{Z}_i)(\tilde{Z}_i - Z_i) Q_i^{\otimes 2}.
\]

Note that $E(Y_i - g^{-1}(\tilde{Z}_i)|U_i, \xi_i, W_i, X_i) = 0$. Then the first term above is $o_P(1)$ by the law of large numbers. Furthermore, as $\tilde{Z}_i - Z_i = (\hat{\alpha}_s(U_i) - \alpha(U_i))^T X_i$, by using Lemma A.2 and Condition (C7), we know that $\max_{i} |\hat{\alpha}_s(U_i) - \alpha(U_i)| = O_P((nh_o)^{-1/2}) = o_P(n^{-1/4})$ for $j = 1, \ldots, p$. This convergence together with Lemma A.1 in Li and Liang (2008) indicate that $\frac{1}{2n} \sum_{i=1}^{n} (Y_i - g^{-1}(\tilde{Z}_i)) \rho_1^*(\tilde{Z}_i)(\tilde{Z}_i - Z_i) Q_i^{\otimes 2} = o_P(n^{-1/2}) = o_P(1)$. Using Lemma A.2 again, we know that each elements of $\frac{1}{2n} \sum_{i=1}^{n} \frac{dg^{-1}(t)}{dt} \bigg|_{t=Z_i} \rho_1^*(\tilde{Z}_i)(\tilde{Z}_i - Z_i) Q_i^{\otimes 2}$ is $o_P((nh_o)^{-1/2}) = o_P(1)$. Similarly, we have $\frac{1}{2n} \sum_{i=1}^{n} \frac{dg^{-1}(t)}{dt} \bigg|_{t=Z_i} \rho_1^*(\tilde{Z}_i)(\tilde{Z}_i - Z_i) Q_i^{\otimes 2} = o_P(1)$. As a consequence, we obtain that $\vartheta_n^{[1]} = o_P(1)$. For the argument $\vartheta_n^{[2]}$. Taylor expansion entails that $\vartheta_n^{[2]} = \frac{1}{2n} \sum_{i=1}^{n} \rho_2(\tilde{Z}_i) Q_i^{\otimes 2} + \frac{1}{2n} \sum_{i=1}^{n} \rho_2(\tilde{Z}_i) Q_i^{\otimes 2} = E\left[\rho_2(Z) Q^{\otimes 2}\right] = o_P(1)$. Thus, we obtain that
\[
\vartheta_n^2 = E\left[\rho_2(Z) Q^{\otimes 2}\right] + o_P(1) \overset{\text{def}}{=} \Sigma + o_P(1).
\]

**Step 2.2.** We analyze the asymptotic expansion of $\vartheta_n^1$. 

\[\vartheta_n^1 = \vartheta_n^{[1]} - \vartheta_n^{[2]} = \left[\frac{1}{\sqrt{n}} \sum_{i=1}^{n} (Y_i - g^{-1}(\tilde{Z}_i)) \rho_1(\tilde{Z}_i) + \rho_1(\tilde{Z}_i) \omega_i \right] Q_i - \left[\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \rho_2(\tilde{Z}_i) Q_i \omega_i \right].\]

**Step 2.2.1.** Note that
\[
\vartheta_n^{[1]} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (Y_i - g^{-1}(\tilde{Z}_i)) \rho_1(\tilde{Z}_i) \omega_i - \left[\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \rho_2(\tilde{Z}_i) Q_i \omega_i \right] + R_n.
\]
Similar to \( \vartheta_{n1}^{[1]} \) in Step 2.1, as \( nh^2 \rightarrow \infty \), Lemma A.2 entails that \( \max_j |\hat{\alpha}_j(U_i) - \alpha_j(U_i)| = o_P(n^{-1/4}) \) for \( j = 1, \ldots, p \). Using Lemma A.1 in Li and Liang (2008), we can show that \( \sum_{i=1}^n (Y_i - g^{-1}(Z_i)) \left\{ \rho'(Z_i)(\tilde{Z}_i - Z_i) + \rho''(Z_i)(\tilde{Z}_i - Z_i) \hat{\omega}_i \right\} Q_i = o_P(n^{1/2}) \) and \( \sum_{i=1}^n Q_i \frac{d}{dt} \left| \begin{array}{c} 1 \\ Z_i = Z_i \\ \tilde{Z}_i = \tilde{Z}_i \\ Z_i = Z_i \\ \tilde{Z}_i = \tilde{Z}_i \\ \end{array} \right\} \rho'(Z_i)(\tilde{Z}_i - Z_i) + \rho''(Z_i) \hat{\omega}_i (\tilde{Z}_i - Z_i) \right\} = o_P(n^{1/2}) \). As a result, \( R_n = o_P(1) \).

Next, we consider the second term in \( \vartheta_{n1}^{[1]} \). By using (2.13), we have

\[
\frac{1}{\sqrt{n}} \sum_{i=1}^n (Y_i - g^{-1}(Z_i)) Q_i \rho'_1(Z_i) \hat{\omega}_i = \left[ \frac{b^2 \mu\lambda_2}{2\sqrt{n}} \sum_{i=1}^n (Y_i - g^{-1}(Z_i)) Q_i \rho'_1(Z_i) \xi^{(2)}(V_i) \right] \beta + \left[ \frac{1}{\sqrt{n^3}} \sum_{i=1}^n \sum_{j=1}^n Q_i \rho'_1(Z_i) \overline{f}(V_i) (Y_i - g^{-1}(Z_i)) \left\{ L_k(V_j - V_i) e_j \right\} \right] \beta \stackrel{\text{def}}{=} \Upsilon_{n1} + \Upsilon_{n2}. \tag{A.11}
\]

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 \( b \rightarrow 0 \). That is \( \Upsilon_{n1} = o_P(1) \). For the argument \( \Upsilon_{n2} \), 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 \( \chi^2_1 \) 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 \( \pi = (I_p, 0, 0, 0)_{p \times (2p+d)} \). We have \( (nh)^{-1/2} \pi \hat{\delta}_s = \hat{\alpha}_s(u) - \alpha(u) \). By using Lemma A.2, we have

\[
\frac{1}{\sqrt{n}} \sum_{i=1}^n Q_i \times (\tilde{Z}_i - Z_i) \rho_2(Z_i) = \frac{1}{\sqrt{n}} \sum_{i=1}^n Q_i X_i^T (\hat{\alpha}_s(U_i) - \alpha(U_i)) \rho_2(Z_i) \tag{A.12}
\]

Write \( \varpi(U_i) = E\{ \rho_2(Z)S_{\mu}(\xi, W, X)|U = U_i \} \) and let \( \kappa_k(U_i) \) be the \( k \)-th element of \( E\{ \rho_2(Z) \}

\[
N^{(2)}|U = U_i \}^{-1} N_i \right\} \). Using the asymptotic expression of \( h_{n1}^{[3]} \) obtained in Step 1.3 and the projection of \( U \)-statistics (Serfling; 1980, Section 5.5.1), as \( nh^4 \rightarrow 0 \), we have

\[
\frac{1}{n \sqrt{h}} \sum_{i=1}^n Q_i \rho_2(Z_i) X_i^T \pi A_{\pi}^{-1} h_{n1}^{[1]} \tag{A.13}
\]

Furthermore, let \( \varpi(u, v) = E\{ \rho_2(Z)(N^T, 0)^T|U = u, V = v \} \), \( \kappa_k(u, v) \) be the \( k \)-th element of \( \varpi(u) \), and \( \kappa_k(u, v) = E\{ Q X_k \rho_2(Z) u_k(U, v)f_{v}(U, v)f_{v}(V) \}. \) Using Lemma A.2 and the projection
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

$$
\varphi^{[1]}_{n1} = -\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \sum_{k=1}^{p} \left( E\{Q X_k \rho_2(Z) \mid U = U_i\} \kappa_k(U_i) q_1(Z_i, Y_i) - \kappa_k(V_i) e_i^T \beta \right) \\
+ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} Q_i q_1(Z_i, Y_i) + o_P(1). 
$$

(A.15)

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

$$
\varphi^{[2]}_{n1} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{\rho_2(Z_i) Q_i}{f_V(V_i)} \left\{ L_b(V_j - V_i) e_j \right\}^T \beta + o_P(1) \\
= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} E\{\rho_2(Z) Q \mid V = V_i\} e_i^T \beta + o_P(1). 
$$

(A.16)

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

$$
\varphi_{n1} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left\{ Q_i - \sum_{k=1}^{p} E\{Q X_k \rho_2(Z) \mid U = U_i\} \kappa_k(U_i) \right\} q_1(Z_i, Y_i) \\
+ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left\{ \sum_{k=1}^{p} \kappa_k(V_i) - E\{\rho_2(Z) Q \mid V = V_i\} \right\} e_i^T \beta + o_P(1). 
$$

(A.17)

Recalling the definition of $\Upsilon_n(\sigma_s)$, 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 $\Upsilon_n(\sigma_s)$ with respect to $\sigma_s$ has the asymptotic expression $\hat{\sigma} = \Sigma^{-1} \varphi_{n1} + o_P(1)$, where $\Sigma$ 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 $\hat{\beta}_{\lambda_1}, \hat{\beta}_{\lambda_2}$. Let $q_n = n^{-1/2} + a_n + b_n$, $S_1 = (s_1, \ldots, s_{1d})^T$, $S_2 = (s_2, \ldots, s_{2r})^T$ and $\|S_1\| = \|S_2\| = C$ for some constant $C$. Furthermore, define $\beta(n) = \beta + q_n S_1$, $\theta(n) = \theta + q_n S_2$, and $\sigma(n) = (\beta^T(n), \theta^T(n))^T$, $\sigma = (\beta^T, \theta^T)^T$. Let $S = \{(S_1, S_2) : \|S_1\| = C, \|S_2\| = C\}$. It suffices to show that, for any given $\varphi > 0$, there exists a large constant $C$ such that

$$
P\left\{ \sup_{S} \mathcal{L}_P(\sigma(n)) < \mathcal{L}_P(\sigma) \right\} \geq 1 - \varphi. 
$$

(A.18)
Write

\[ F_{n1} = \sum_{i=1}^{n} \left\{ Q\{g^{-1}(\hat{\alpha}_s(U_i)^T X_i + \xi_i^T \beta(n) + W_i^T \theta(n)), Y_i) \right\} \\
- Q\{g^{-1}(\hat{\alpha}_s(U_i)^T X_i + \xi_i^T \beta + W_i^T \theta), Y_i) \right\}, \]

\[ F_{n2} = -n \sum_{j=1}^{d_1} \{ p_{\lambda_{ij}}(|\beta_j|) + p_{\lambda_{1j}}(|\beta_j|) \} - n \sum_{i=1}^{r_1} \{ p_{\lambda_{ij}}(|\theta_i| + q_n s_{2i}) - p_{\lambda_{ij}}(|\theta_i|) \}, \]

where \( d_1, r_1 \) are the numbers of elements of \( \beta_1, \theta_1 \), 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 \( \hat{Z}_i = \hat{\alpha}_s(U_i)^T X_i + \xi_i^T \beta + W_i^T \theta \). Taylor expansion entails that

\[ F_{n1} = \sum_{i=1}^{n} q_1(\hat{Z}_i, Y_i) q_n \cdot (\xi_i^T S_1 + W_i^T S_2) + \frac{1}{2} \sum_{i=1}^{n} q_2(\hat{Z}_i + \hat{\xi}_i^*, Y_i) q_n^2 \cdot (\xi_i^T S_1 + W_i^T S_2)^2 \]

\[ = F_{n1}^{[1]} - F_{n1}^{[2]}, \]

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

\[ F_{n1}^{[1]} = q_n \sum_{i=1}^{n} q_1(Z_i, Y_i) (\xi_i^T S_1 + W_i^T S_2) + q_n \sum_{i=1}^{n} q_2(Z_i, Y_i) (\hat{Z}_i - Z_i) (\xi_i^T S_1 + W_i^T S_2) \]

\[ + \frac{q_n}{2} \sum_{i=1}^{n} q_3(Z_i + \hat{\xi}_i^*, Y_i) (\hat{Z}_i - Z_i)^2 (\xi_i^T S_1 + W_i^T S_2) \]

\[ = \Delta_{n1}^{[1]} + \Delta_{n1}^{[2]} + \Delta_{n1}^{[3]}, \]

where \(|\hat{\xi}_i^*|\) is between 0 and \(|Z_i - \hat{Z}_i|\). 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 \( \Xi_i \)'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 \( \Delta_{n1}^{[3]} = O_P(1/nh + b^4 + \log b^{-1}/nb) \). Thus, under Condition (C7), we have \( \Delta_{n1}^{[3]} = o_P(1)C \).

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

\[ F_{n1}^{[2]} = \frac{q_n^2}{2} \sum_{i=1}^{n} (Y_i - g^{-1}(Z_i)) \rho_1'(Z_i) (\xi_i^T S_1 + W_i^T S_2)^2 - \frac{q_n^2}{2} \sum_{i=1}^{n} \rho_2(Z_i) (\xi_i^T S_1 + W_i^T S_2)^2 \]

\[ + \frac{q_n^2}{2} \sum_{i=1}^{n} \left\{ \frac{dg^{-1}(t)}{dt} \bigg|_{t=Z_i} \rho_1'(Z_i) + (Y_i - g^{-1}(Z_i)) \rho_1''(Z_i) - \rho_2(Z_i) \right\} (\xi_i^T S_1 + W_i^T S_2)^2 \]

\[ \times \left[ (Z_i - \hat{Z}_i) + \hat{\xi}_i^* \right] \]

\[ = \Omega_{n1}^{[1]} - \Omega_{n1}^{[2]} + \Omega_{n1}^{[3]}, \]

Note that \( q_n^2 = O(n^{-1}) \). Using (2.1.3) and \( E[(Y - g^{-1}(Z)) \rho'(Z) U, \xi, \xi, W, X] = 0 \), we can obtain that \( \Omega_{n1}^{[1]} = \frac{q_n^2}{2} \sum_{i=1}^{n} (Y_i - g^{-1}(Z_i)) \rho_1'(Z_i) (\xi_i^T S_1 + W_i^T S_2)^2 + o_P(1)C^2 = o_P(1)C^2 \). Moreover, \(|\hat{\xi}_i^*|\) is between 0 and \(|Z_i - \hat{Z}_i|\). Then (2.1.3) and Lemma A.2 guarantee \(|\hat{\xi}_i^*| = O_P(1/\sqrt{nh} + \)
\[ b^2 + \sqrt{\log b^{-1}/nb} \]. Hence, we have \( \Omega_{n1}^{(3)} = o_P(1)C^2 \). For the argument \( \Omega_{n1}^{(2)} \), we have \( \Omega_{n1}^{(2)} = \frac{q_n^2}{T} \sum_{i=1}^n \rho_{2}(Z_i) (\xi_i^T S_1 + W_w^T S_2)^2 + o_P(1)C^2 = (S_1, S_2) \Sigma(S_1^T, S_2^T)^T + o_P(1)C^2 \). As a result, these arguments for \( F_{n1}^{(1)} \) and \( F_{n1}^{(2)} \) mean

\[ F_{n1} = q_n \sum_{i=1}^n \{ q_i(Z_i, Y_i) Q_i^1 + \Xi_i^T \} (S_1^T, S_2^T)^T - 0.5(S_1, S_2) \Sigma(S_1^T, S_2^T)^T + o_P(1)C^2. \] (A.19)

Using Taylor expansion and Cauchy-Schwartz inequality, we conclude that \( F_{n2} \) is bounded by \( n\sqrt{\bar{I}_1} q_n a_n^\ast \rho_0 S_1 || + nq_n^2 a_n^{\ast\ast} || S_1 ||^2 + n\sqrt{\bar{I}_1} q_n b_n^\ast || S_2 || + nq_n^2 b_n^{\ast\ast} || S_2 ||^2 = O(\sqrt{\bar{I}_1} + \sqrt{\bar{I}_2})C + O(a_n^{\ast\ast}C^2 + b_n^{\ast\ast}C^2). \) As \( a_n^\ast \rightarrow 0 \) and \( b_n^\ast \rightarrow 0 \), 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 \( \hat{\beta}_{\lambda_1} = \beta + O_P(n^{-1/2}), \hat{\beta}_{\lambda_2} = \theta + O_P(n^{-1/2}). \)  

**A.5. Proof of Theorem 3**

From Lemma A.3, we know that \( \hat{\beta}_{\lambda_1(2)} = 0_{(d-d_1)x1} \) and \( \hat{\beta}_{\lambda_2(2)} = 0_{(r-r_1)x1} \). We next investigate asymptotic normality of \( \hat{\beta}_{\lambda_1(1)}, \hat{\beta}_{\lambda_2(1)}. \) Define \( \delta_{\lambda_1(1)} = \sqrt{\bar{I}_1} (\hat{\beta}_{\lambda_1(1)} - \beta_i(1))^T, (\hat{\beta}_{\lambda_2(1)} - \beta_1(1))^T \), \( \hat{Q}(1) = (\xi_{i1}^T, W_{i1}^T)^T, \hat{Z}_{i1} = \beta_i^T (\xi_{i1} + \theta_i^T W_{i1} + \alpha_i(U_i) X_i, Z_{i1} = \beta_{i1}^T \xi_{i1} + \theta_{i1}^T W_{i1} + \alpha(U_i) X_i, \hat{\omega}_{i1} = \xi_{i1}^T (\xi_{i1} - \xi_{i1}) + (\beta_{i1})^T (\xi_{i1} - \xi_{i1}). \) Then \( \delta_{\lambda_1(1)} \) maximizes

\[ \sum_{i=1}^n \left\{ Q g^{-1} \left( \hat{Z}_{i1}, \hat{\omega}_{i1}, n^{-1/2} \sigma_{1}^T \hat{Q}_{i1} \right) \right\} - n \sum_{j=1}^{d_1} p_{\lambda_{1j}}(|\beta_{j*}|) \]

\[ - n \sum_{s=1}^{r_1} p_{\lambda_{2s}}(|\theta_{s*}|) \triangleq \mathbb{E} \left( \sum_{i=1}^n \left\{ Q g^{-1} \left( \hat{Z}_{i1}, \hat{\omega}_{i1}, n^{-1/2} \sigma_{1}^T \hat{Q}_{i1} \right) \right\} - n \sum_{j=1}^{d_1} p_{\lambda_{1j}}(|\beta_{j*}|) - n \sum_{s=1}^{r_1} p_{\lambda_{2s}}(|\theta_{s*}|) \right), \] (A.20)

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

By using Taylor expansion, \( p_{\lambda_{1j}}(|\beta_{j*}|) = p_{\lambda_{1j}}(|\beta_{j}|) + p_{\lambda_{1j}}^\prime (|\beta_{j}|) \text{sign}(\beta_{j} - \beta_{j}) + 1/2 p_{\lambda_{1j}}^{''}(|\beta_{j}|)(\beta_{j} - \beta_{j})^2. \) Similar to the analysis for \( \mathbb{E} \left( \sum_{i=1}^n \left\{ Q g^{-1} \left( \hat{Z}_{i1}, \hat{\omega}_{i1}, n^{-1/2} \sigma_{1}^T \hat{Q}_{i1} \right) \right\} - n \sum_{j=1}^{d_1} p_{\lambda_{1j}}(|\beta_{j*}|) - n \sum_{s=1}^{r_1} p_{\lambda_{2s}}(|\theta_{s*}|) \right) \) 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

\[ \mathbb{E} \left( \sum_{i=1}^n \left\{ Q g^{-1} \left( \hat{Z}_{i1}, \hat{\omega}_{i1}, n^{-1/2} \sigma_{1}^T \hat{Q}_{i1} \right) \right\} - n \sum_{j=1}^{d_1} p_{\lambda_{1j}}(|\beta_{j*}|) - n \sum_{s=1}^{r_1} p_{\lambda_{2s}}(|\theta_{s*}|) \right) = \left[ n^{-1/2} \sum_{i=1}^n \{ q_1(\hat{Z}_{i1}, Y_i) Q_{i1} + q_2(\hat{Z}_{i1}, Y_i) Q_{i1} \}^T + \sqrt{n} R_{n, \lambda_1, \lambda_2} \right] \sigma_{1*} \]

\[ + \sigma_{1*}^T \left[ \frac{1}{n} \sum_{i=1}^n q_2(\hat{Z}_{i1}, Y_i) Q_{i1}^2 + \frac{1}{2} \sum_{n, \lambda_1, \lambda_2} \right] \sigma_{1*} + C_n, \] (A.21)

where \( C_n \) is irrelevant to \( \sigma_{1*}. \) Define \( \Sigma_{1*} = E \left( \rho_{2}(Z_{i1}) Q_{i1}^2 \right). \) Similar to the proof of Theorem 1,
Furthermore, it follows from the arguments similar to the proof of Lemma A.2 that
\[
\sqrt{n} (\Sigma_{(1)} + \Sigma_{n,\lambda_1,\lambda_2}) \left\{ \left( (\hat{\beta}_1(1) - \beta(1))^T, (\hat{\beta}_2(1) - \theta(1))^T \right)^T + (\Sigma_{(1)} + \Sigma_{n,\lambda_1,\lambda_2})^{-1} R_{n,\lambda_1,\lambda_2} \right\}
\]
\[
= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left\{ Q_i(1) - \sum_{k=1}^{p} E\{ Q_1 X_k \rho_2(Z(1)) | U = U_i \} \kappa_{1k} U_i \right\} q_1(Z_i(1), Y_i)
\]
\[
+ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left\{ \sum_{k=1}^{p} \kappa_{1k} (v_i) - E\{ \rho_2(Z(1)) Q(1) | V = V_i \} \right\} e_i(1) \beta(1) + o_P(1).
\]

We complete the proof.

A.6. Proof of Theorem 4

Write \( X_i^u = (X_i^T, h^{-1} X_i (U_i - u)^T) \), \( \eta(u) = (n h)^{1/2} ((\hat{\alpha}(u) - \alpha(u))^T, h (\hat{B}(u) - \alpha(u))^T) \), \( \hat{Z}_i(u) = \alpha(u)^T X_i + \alpha'(u) X_i (U_i - u) + \beta^T \xi_i + \hat{\theta} \xi_i \). We know that \( \eta(u) \) is the maximizer of (A.22) with respect to \( \eta_* \),
\[
\varphi_n(\eta_*) = h \sum_{i=1}^{n} \left\{ Q\{ g^{-1}(r_n^{-1} \eta_*^T X_i^u + (\hat{\xi}_i - \xi_i)^T \hat{\beta}_i + \hat{Z}_i(u)), Y_i \} \right. \\
\left. - Q\{ g^{-1}(X_i^u), Y_i \} \right\} K_h(U_i - u).
\]

Similar to handling \( \ell_n(\delta^*) \) in Lemma A.2, we express (A.22) as
\[
\varphi_n(\eta_*) = \left[ r_n h \sum_{i=1}^{n} \left\{ q_1(\hat{Z}_i(u), Y_i) + q_2(\hat{Z}_i(u), Y_i) (\hat{\xi}_i - \xi_i)^T \beta \right\} K_h(U_i - u) X_i^u \right] \eta_*
\]
\[
+ \eta_* \left[ \frac{n}{2} h \sum_{i=1}^{n} q_2(\hat{Z}_i(u), Y_i) (X_i^u)^{\otimes 2} K_h(U_i - u) \right] \eta_*
\]
\[
+ \left[ h \sum_{i=1}^{n} \left\{ q_1(\hat{Z}_i(u), Y_i) (\hat{\xi}_i - \xi_i)^T \beta + \frac{1}{2} q_2(\hat{Z}_i(u), Y_i) ((\hat{\xi}_i - \xi_i)^T \beta)^2 \right\} K_h(U_i - u) \right]
\]
\[
\overset{\text{def}}{=} h_{n1} \eta_* + h_{n2} \eta_* h_{n2} \eta_* + h_{n3}.
\]

Furthermore, it follows from the arguments similar to the proof of Lemma A.2 that
\[
h_{n2} = \frac{1}{2} f_{U}(u) E\{ \rho_2(Z) \text{diag} (X^{\otimes 2}, \mu K_2 X^{\otimes 2}) | U = u \} + o_P(1).
\]

We now consider \( h_{n1} \), which equals to
\[
r_n h \sum_{i=1}^{n} q_1(\hat{Z}_i(u), Y_i) K_h(U_i - u) X_i^u + r_n h \sum_{i=1}^{n} q_2(\hat{Z}_i(u), Y_i) (\hat{\xi}_i - \xi_i)^T \beta K_h(U_i - u) X_i^u
\]
\[
\overset{\text{def}}{=} h_{n1}^{[1]} + h_{n1}^{[2]}.
\]
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
\[
\begin{align*}
\hat{h}^{|1|}_{n1} &= \sqrt{\frac{h}{n}} \sum_{i=1}^{n} q_1(\tilde{Z}_i(u), Y_i) K_h(U_i - u) X_i^u \\
&\quad - \sqrt{\frac{h}{n}} \sum_{i=1}^{n} \rho_2(\tilde{Z}_i(u)) K_h(U_i - u) X_i^u \{ \xi_i^T (\hat{\beta} - \beta) + W_i^T (\hat{\theta} - \theta) \} \\
&\quad + \sqrt{\frac{h}{n}} \sum_{i=1}^{n} (Y_i - g^{-1}(\tilde{Z}_i(u))) \rho_1(\tilde{Z}_i(u)) K_h(U_i - u) X_i^u \{ \xi_i^T (\hat{\beta} - \beta) + W_i^T (\hat{\theta} - \theta) \} \\
&\quad + \sqrt{\frac{h}{n}} \sum_{i=1}^{n} \frac{dg^{-1}(t)}{dt} \bigg|_{t=\tilde{Z}_i(u)} \rho_1(\tilde{Z}_i(u)) K_h(U_i - u) X_i^u \{ \xi_i^T (\hat{\beta} - \beta) + W_i^T (\hat{\theta} - \theta) \}^2 \\
&\quad \overset{\text{def}}{=} \hat{h}^{|1|}_{n1}(1) + \hat{h}^{|1|}_{n1}(2) + \hat{h}^{|1|}_{n1}(3) + \hat{h}^{|1|}_{n1}(4).
\end{align*}
\]

By Theorem 1, we know that \( \hat{\beta} - \beta = O_P(n^{-1/2}), \hat{\theta} - \theta = O_P(n^{-1/2}) \). Furthermore, \( (nh)^{1/2} \left\{ \frac{1}{n} \sum_{i=1}^{n} \rho_2(\tilde{Z}_i(u)) K_h(U_i - u) X_i^u \xi_i^T \right\} \) 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 \( \sqrt{n} \sum_{i=1}^{n} \rho_2(\tilde{Z}_i(u)) K_h(U_i - u) X_i^u \xi_i^T (\hat{\beta} - \beta) = O_P(h^{1/2}) + O_P(1)(\hat{\beta} - \beta) = o_P(1) \)

and then \( \hat{h}^{|1|}_{n1}(2) = o_P(1) \). Similarly, we can show that \( \hat{h}^{|1|}_{n1}(3) = o_P(1), \hat{h}^{|1|}_{n1}(4) = o_P(1) \). For \( \hat{h}^{|1|}_{n1}(1) \), we have
\[
\begin{align*}
(nh)^{1/2} \left\{ \frac{1}{n} \sum_{i=1}^{n} \left\{ q_1(\tilde{Z}_i(u), Y_i) K_h(U_i - u) X_i^u - E \left[ q_1(\tilde{Z}_i(u), Y_i) K_h(U_i - u) X_i^u \right] \right\} \right\} \\
\overset{\mathcal{L}}{\longrightarrow} N(0, \Sigma_{s1}(u)),
\end{align*}
\]
where \( \Sigma_{s1}(u) = f_U(u) E\{ \rho_2(Z) \text{diag}(v_{K_0} X_1^\otimes, v_{K_2} X_2^\otimes) \big| U = u \} \).

Step 4.2. We deal with \( \hat{h}^{|2|}_{n1} \). In fact, we only need to consider
\[
\sqrt{n h} \left\{ \frac{1}{n} \sum_{i=1}^{n} \rho_2(\tilde{Z}_i(u)) \xi_i^{(2)}(V_i) K_h(U_i - u) X_i^u - E \left[ \rho_2(\tilde{Z}_i(u)) \xi_i^{(2)}(V_i) K_h(U_i - u) X_i^u \right] \right\}
\]
\[
\overset{\mathcal{L}}{\longrightarrow} N(0, A^*(u))
\]
for some matrix \( A^*(u) \). From (A.24), we achieve that \( 0.5 \mu_1 \beta^2 \sqrt{n} \sum_{i=1}^{n} \rho_2(\tilde{Z}_i(u)) K_h(U_i - u) \xi_i^{(2)}(V_i)^T \)
\( \beta X_i^u = O_P(\sqrt{nhb^2}) + O_P(b^2) = o_P(1) \). Moreover, similar to the analysis for (A.14), we have
\[
\begin{align*}
(nh)^{1/2} \sum_{i=1}^{n} \sum_{j=1}^{n} \rho_2(\tilde{Z}_i(u)) \left[ L_b(V_j - V_i) e_j \right]^T \beta K_h(U_i - u) X_i^u = O_P(h^{1/2}) + O_P(h^{1/2}(b^2 + h^2)) = o_P(1)
\end{align*}
\]
by using the projection of \( U \)-statistics (Serfling (1980), Section 5.3.1). As such, \( \hat{h}^{|2|}_{n1} = o_P(1) \).

As a consequence, \( \hat{h}^{|3|}_{n1} = o_P(1) \), together with Condition (C7), i.e., \( nh^4 \to 0, nb^4 \to 0 \), and (A.23), mean that \( \hat{h}_{n1} \overset{\mathcal{L}}{\longrightarrow} N(0, \Sigma_{s1}(u)) \). Similar to the proof of Lemma A.2, the maximizer \( \hat{\eta} \) of \( \psi_n(\eta) \) is asymptotically equivalent to \( \hat{\eta} = -2h_{n2}^{-1} \hat{h}_{n1} + o_P(1) \). Invoking the asymptotic result of \( \hat{h}_{n2} \) 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

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

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

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

\[
\mathcal{I}_{n4} = \sum_{i=1}^{n} \left[ Q \{ g^{-1}(\beta^T \xi_i + \theta^T W_i) \} - Q \{ g^{-1}(\beta^T \xi_i + \theta^T W_i) \} \right].
\]

By similar arguments used in the proof of Theorem 3 in Li and Liang (2008), we can show that \( 2\mathcal{I}_{n2} \overset{\mathcal{L}}{\rightarrow} \chi^2_2 \) and \( 2\mathcal{I}_{n4} \overset{\mathcal{L}}{\rightarrow} \chi^2_2 \). Next, we analyze \( \mathcal{I}_{n1} \) and \( \mathcal{I}_{n3} \).

**Step 5.1.** Taylor expansion entails that

\[
\mathcal{I}_{n1} \overset{\text{def}}{=} \mathcal{I}_{n1}^{[1]} + \mathcal{I}_{n1}^{[2]} = \sum_{i=1}^{n} q_1(\hat{\alpha}(U_i)^T X_i + \beta^T \xi_i + \theta^T W_i, Y_i)(\hat{\xi}_i - \xi)^T \hat{\beta}
\]

\[
+ \frac{1}{2} \sum_{i=1}^{n} q_2(\hat{\alpha}(U_i)^T X_i + \beta^T \xi_i + \theta^T W_i, Y_i)(\hat{\xi}_i - \xi)^T \hat{\beta}^2.
\]

**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

\[
\mathcal{I}_{n1}^{[1]} = \sum_{i=1}^{n} q_1(Z_i, Y_i) (\hat{\xi}_i - \xi)^T \hat{\beta} + \sum_{i=1}^{n} \left[ (Y_i - g^{-1}(Z_i)) \rho'(Z_i) + \rho_2(Z_i) \right] (\hat{\xi}_i - \xi)^T \hat{\beta} \Delta Z_i
\]

\[
+ \sum_{i=1}^{n} \left[ \left( \frac{dg^{-1}(t)}{dt} \bigg|_{t=Z_i} \right) \rho'(Z_i) \right] (\hat{\xi}_i - \xi)^T \hat{\beta} (\Delta Z_i)^2
\]

\[
\overset{\text{def}}{=} \mathcal{I}_{n1}^{[1]}(1) + \mathcal{I}_{n1}^{[1]}(2) + \mathcal{I}_{n1}^{[1]}(3).
\]

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 \( \mathcal{I}_{n1}^{[1]}(3) = O_P(n(b^2 + (\log b^{-1})^{1/2})/nh) \). 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

\[
\sum_{i=1}^{n} \left[ (Y_i - g^{-1}(Z_i)) \rho'(Z_i) + \rho_2(Z_i) \right] (\hat{\xi}_i - \xi)^T \hat{\beta} \times \left( \hat{\alpha}(U_i)^T X_i + (\hat{\beta} - \beta)^T \xi_i + (\hat{\theta} - \theta)^T W_i \right) = O_P(n h^{-1} b^2) + O_P(b^2) = o_P(1).
\]

Furthermore, using the projection of \( U \)-statistics (Serfling; 1980, Section 5.3.1), we have

\[
\frac{1}{n} \sum_{i=1}^{n} \left[ (Y_i - g^{-1}(Z_i)) \rho'(Z_i) + \rho_2(Z_i) \right] \left[ L_b(V_i - V_j) \right] e_i^T \hat{\beta}
\]

\[
(\hat{\alpha}(U_i)^T X_i + (\hat{\beta} - \beta)^T \xi_i + (\hat{\theta} - \theta)^T W_i) = O_P(1) = o_P(h^{-1/2}).
\]

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
\[
I_{n1}^{[2]} \overset{\text{def}}{=} I_{n1}^{[2]}(1) - I_{n1}^{[2]}(2) = 0.5 \sum_{i=1}^{n} (Y_i - g^{-1}(Z_i + \Delta Z_i))\rho'(Z_i + \Delta Z_i)((\hat{\xi}_i - \xi)^T \hat{\beta})^2 \\
- 0.5 \sum_{i=1}^{n} \rho_2(Z_i + \Delta Z_i)((\hat{\xi}_i - \xi)^T \hat{\beta})^2.
\]

Taylor expansion of $g^{-1}(Z_i + \Delta Z_i)$ and $\rho'(Z_i + \Delta Z_i)$ with respect to $Z_i$ yields that
\[
I_{n1}^{[2]}(1) = 0.5 \sum_{i=1}^{n} (Y_i - g^{-1}(Z_i)) \rho'(Z_i + \Delta Z_i)((\hat{\xi}_i - \xi)^T \hat{\beta})^2 \\
+ 0.5 \sum_{i=1}^{n} \left. \frac{dg^{-1}(t)}{dt} \right|_{t=Z_i} (\rho'(Z_i) + \Delta Z_i)((\hat{\xi}_i - \xi)^T \hat{\beta})^2 \Delta Z_i.
\]

Together with (2.1.3) and Lemma A.1, we know that $\{((\hat{\xi}_i - \xi)^T \hat{\beta})^2 = O_P(b^4 + \log b^{-1}/(nb))$. Thus, we have $\sqrt{nh}\{((\hat{\xi}_i - \xi)^T \hat{\beta})^2 = O_P(n^{-1/4})$ since $nh^6 \to 0$ and $nh^2/\log h^{-1} \to \infty$. So Lemma A.1 in Li and Liang (2008) entails that $\frac{1}{\sqrt{nh}} \sum_{i=1}^{n} (Y_i - g^{-1}(Z_i)) \rho'(Z_i + \Delta Z_i) \{\sqrt{nh}((\hat{\xi}_i - \xi)^T \hat{\beta})^2\} = O_P(h^{-1/2})$. By Theorems 1 and 4, we know that, under $H_0$, $\Delta Z_i = O_P(1/\sqrt{nh})$. Again together with the result $\{((\hat{\xi}_i - \xi)^T \hat{\beta})^2 = O_P(b^4 + \log b^{-1}/(nb))$, we have
\[
\frac{1}{2} \sum_{i=1}^{n} \left. \frac{dg^{-1}(t)}{dt} \right|_{t=Z_i} (\rho'(Z_i) + \Delta Z_i)((\hat{\xi}_i - \xi)^T \hat{\beta})^2 \Delta Z_i = nO_P(1/\sqrt{nh})O_P(b^4 + \log b^{-1}/(nb)) = o_P(h^{-1/2}), \quad (A.26)
\]

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

For the argument $I_{n1}^{[2]}(2)$, Taylor expansion $\rho_2(Z_i + \Delta Z_i) = \rho_2(Z_i) + \rho_2'(Z_i) \Delta Z_i$ and the fact that $\sum_{i=1}^{n} \rho_2(Z_i)((\hat{\xi}_i - \xi)^T \hat{\beta})^2 \Delta Z_i = o_P(h^{-1/2})$ yield
\[
I_{n1}^{[2]}(2) = \frac{b^4}{2} \sum_{i=1}^{n} \rho_2(Z_i)\{\xi^{(2)}(V_i)^T \hat{\beta}\}^2 + \frac{1}{2} \sum_{i=1}^{n} \rho_2(Z_i)\left. \frac{1}{nf(V_i)} \sum_{j=1}^{n} [L_0(V_j - V_i)e_j]^T \hat{\beta} \right|^2 \\
+ \frac{b^2}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{\rho_2(Z_i)}{f_i(V_j)} \{\xi^{(2)}(V_i)^T \hat{\beta}\} \{L_0(V_j - V_i)e_j\}^T \hat{\beta} + o_P(h^{-1/2}). \quad (A.27)
\]

As $nb^4 \to 0$, $b^4 \sum_{i=1}^{n} \rho_2(Z_i)\{\xi^{(2)}(V_i)^T \hat{\beta}\}^2 = o_P(1)$. 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
\[
\frac{1}{2} \sum_{i=1}^{n} \rho_2(Z_i)\left. \frac{1}{nf(V_i)} \sum_{j=1}^{n} [L_0(V_j - V_i)e_j]^T \hat{\beta} \right|^2 \\
= \frac{1}{2n^2b^2} \sum_{i=1}^{n} \rho_2(Z_i) f_i(V_i)^2 L(0)^2 \{e_i^T \hat{\beta}\}^2 + \frac{1}{2n^2} \sum_{i=1}^{n} \sum_{j=1, j \neq i}^{n} \rho_2(Z_i) f_i(V_i) \{[L_0(V_j - V_i)e_j]^T \hat{\beta}\}^2 \\
+ \frac{1}{2n^2} \sum_{i=1}^{n} \sum_{s=1}^{n} \sum_{j=1, j \neq s}^{n} \rho_2(Z_i) f_i(V_i) \{[L_0(V_s - V_i)e_j]^T \hat{\beta}\} \{[L_0(V_j - V_i)e_j]^T \hat{\beta}\}. \quad (A.28)
\]
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

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

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

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

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 \(\frac{L(0)}{n^2b} \sum_{i=1}^{n} \sum_{j=1, j \neq i}^{n} \frac{\rho_2(Z_i)}{f^2(V_i)} e^T e \left\{ L_b(V_j - V_i)e_j \right\}^T \beta = O_P(1/nb)\). We directly adopt the arguments used in (Fan et al.; 2001, Pages 182-186) and know that the second term of (A.30) equals

\[
\frac{b^{-1/2}}{2} \left\{ \frac{\sqrt{n} \sum_{j \neq i} \{ E(\rho_2(Z_i)|V_i) \}}{f_V(V_i)} \right\} E \left\{ \left( \frac{E(\rho_2(Z)|V)}{f_V(V)} \right)^2 \right\} + o_P(b^{-1/2})
\]

\[
= \frac{b^{-1/2}}{2} W_b(n) + o_P(b^{-1/2}). \tag{A.31}
\]

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

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

A combination of (A.27)-(A.31), together with that \(T_{n1}^{(1)} = o_P(h_o^{-1/2}), T_{n1}^{(2)}(1) = o_P(h_o^{-1/2}),\) indicates that

\[
T_{n1} = -\frac{b^{-1/2} h_o^{-1/2}}{\Sigma_e \beta} E \left\{ \frac{E(\rho_2(Z_i)|V_i)}{f(V_i)} \right\} - \frac{c_b^{-1/2} h_o^{-1/2}}{2} W_b(n) + o_P(h_o^{-1/2}). \tag{A.32}
\]

**Step 5.2.** Taylor expansion entails that \(T_{n3} = \sum_{i=1}^{n} q_1(Z_i, Y_i) \hat{\alpha}(U_i)^T X_i + \sum_{i=1}^{n} q_2(Z_i, Y_i) \hat{\alpha}(U_i)^T X_i)^2.\) Recalling the proof of Theorem 4, we have, under \(H_0, \sqrt{nh_o \hat{\alpha}}(u) = -2h_{12}^{-1} h_{11}^{[1]} - 2h_{12}^{-1} h_{11}^{[2]}\). By the results obtained in Step 4.2, we know that \(-2h_{12}^{-1} h_{11}^{[2]} = O_P(\sqrt{nh_o h^2} + h^{1/2}) = o_P(1), as nh_o \to 0.\) As such, we only need to consider the main term \(-2h_{12}^{-1} h_{11}^{[1]}\). It follows from a direct use of Theorem 10 of Fan et al. (2001) that

\[
T_{n3} = -\frac{p[\mu]}{c_b h_o} \left[ K(0) - 0.5v_K \right] - \frac{c_b^{-1/2} h_o^{-1/2}}{2} W_b(n) + o_P(h_o^{-1/2}), \tag{A.33}
\]
where \( v_{K_0} = \int K^2(u)du \) and \( W_h(n) \xrightarrow{c} N(0, \sigma^2_K) \) with \( \sigma^2_K = 2p \left\{ \int [2K(s) - K * K(s)]^2 ds \right\} \frac{2}{|U|} \). Note that \( h_{n1}^{[i]} \) 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 \( 2I_{n2} \xrightarrow{c} \chi^2_p = o_P(h_o^{-1/2}) \) and \( 2I_{n4} \xrightarrow{c} \chi^2_p = o_P(h_o^{-1/2}) \). Thus, together with (A.32) and (A.33), we have that

\[
\mathcal{H}_1 - \mathcal{H}_0 = \frac{v_{L_0} \beta^T \Sigma e \beta}{c_b h_o} E \left[ \frac{E(p_2(Z_i)|V_i)}{f_V(V_i)} \right] - \frac{p |U|}{c_b h_o} [K(0) - 0.5v_{K_0}] \tag{A.34}
\]

\[
- \frac{h_o^{-1/2}}{2} \left\{ c_b^{-1/2} W_b(n) + c_h^{-1/2} W_h(n) \right\} + o_P(h_o^{-1/2}).
\]

Note that \( c_b^{-1/2} W_b(n) + c_h^{-1/2} W_h(n) \xrightarrow{c} N(0, c_b^{-1} \sigma^2_K + c_h^{-1} \sigma^2_K) \). By (A.34), we know that \( r_{LK} \{\mathcal{H}_0 - \mathcal{H}_1 - \chi^2_{df_n}\} \xrightarrow{c} 0 \). The proof is complete.

**References**


