Supplemental Appendix for "Nonparametric Two-Step Sieve M Estimation and Inference"

Jinyong Hahn[*] Zhipeng Liao[†] Geert Ridder[‡]

UCLA UCLA USC

This version: December 2017

This supplemental appendix provides some auxiliary materials for "Nonparametric Two-Step Sieve M Estimation and Inference"(cited as HLR in this appendix). Section 1 provides sufficient conditions for Assumptions 3.2 and 3.4 in HLR which are the key high-level conditions for asymptotic normality of the two-step sieve M estimator. Section 2 presents some lemmas which are used in proving Theorem 5.1 in HLR. Section 3 contains verification of the high-level assumptions for asymptotic normality in the nonparametric triangular simultaneous equation model. Section 4 contains some extra simulation results. Section 5 establishes general theory on the consistency and convergence rate of the nonparametric two-step sieve M estimator.

1 Sufficient Conditions for Assumptions 3.2 and 3.4 in HLR

In this section, we provide sufficient conditions for the high-level assumptions (Assumptions 3.2 and 3.4) of the asymptotic normality of the nonparametric two-step sieve M estimator. These sufficient conditions are verified in the nonparametric triangular simultaneous equation model in Section 3 of the Appendix. We assume that the data $\{Z_i\}_{i=1}^n$ is i.i.d. in this section.

Assumption 1.1 *(i) For any $z_2 \in \mathcal{Z}_2$, any $\alpha \in \mathcal{N}_\alpha$ and any $v_{g,1}, v_{g,2} \in \mathcal{V}_2$, the following directional derivatives exist*

$$\Delta_\psi(z_2,\alpha)[v_{g,1}] = \left.\frac{\partial \psi(z_2, g+\tau v_{g,1}, h)}{\partial \tau}\right|_{\tau=0} \text{ and } r_{\psi,g}(z_2,\alpha)[v_{g,1}, v_{g,2}] = \left.\frac{\partial \Delta_\psi(z_2, g+\tau v_{g,2}, h)[v_{g,1}]}{\partial \tau}\right|_{\tau=0};$$

[*]Department of Economics, UCLA, Los Angeles, CA 90095-1477 USA. Email: hahn@econ.ucla.edu

[†]Department of Economics, UCLA, Los Angeles, CA 90095-1477 USA. Email: zhipeng.liao@econ.ucla.edu

[‡]Department of Economics, University of Southern California, Los Angeles, CA 90089. Email: ridder@usc.edu.

(ii) there exists $\Lambda_{1,n}(z_2)$ with $\mathbb{E}[\Lambda_{1,n}(Z_2)] \leq C$ such that

$$\sup_{\alpha \in \mathcal{N}_n} \left| \psi(z_2, g^*, h) - \psi(z_2, \alpha) - \Delta_\psi(z_2, \alpha)[\pm \kappa_n u^*_{g_n}] - \kappa_n^2 r_{\psi,g}(z_2, \alpha)[u^*_{g_n}, u^*_{g_n}] \right| \leq \kappa_n^2 \Lambda_{1,n}(z_2);$$

(iii) there exists $\Lambda_{2,n}(z_2)$ with $\mathbb{E}[\Lambda_{2,n}(Z_2)] \leq C$ such that

$$\sup_{\alpha \in \mathcal{N}_n} \left| r_{\psi,g}(z_2, \alpha)[u^*_{g_n}, u^*_{g_n}] - r_{\psi,g}(z_2, \alpha_o)[u^*_{g_n}, u^*_{g_n}] \right| \leq \Lambda_{2,n}(z_2);$$

(iv) $\mathbb{E}\left[\left| r_{\psi,g}(Z_2, \alpha_o)[u^*_{g_n}, u^*_{g_n}] \right|\right] \leq C$; (v) $\mathbb{E}\left[r_{\psi,h}(Z_2, \alpha_o)[h_{o,n} - h_o, u^*_{g_n}] \right] = o(n^{-1/2})$.

Assumption 1.2 *(i) For any $z_2 \in \mathcal{Z}_2$, any $\alpha \in \mathcal{N}_\alpha$, any $v_h \in \mathcal{V}_1$ and any $v_g \in \mathcal{V}_2$, the following directional derivative exists*

$$\left. \frac{\partial \Delta_\psi(z_2, g, h + \tau v_h)[v_g]}{\partial \tau} \right|_{\tau=0} = r_{\psi,h}(z_2, \alpha)[v_g, v_h];$$

(ii) there exists $\Lambda_{3,n}(z_2, \alpha)$ such that for any $\alpha \in \mathcal{N}_n$,

$$\left| \Delta_\psi(z_2, g, h)[u^*_{g_n}] - \Delta_\psi(z_2, g_o, h)[u^*_{g_n}] - r_{\psi,g}(z_2, g_o, h)[g - g_o, u^*_{g_n}] \right| \leq \Lambda_{3,n}(z_2, \alpha);$$

(iii) there exists $\Lambda_{4,n}(z_2, \alpha)$ such that for any $\alpha \in \mathcal{N}_n$,

$$\left| \Delta_\psi(z_2, g_o, h)[u^*_{g_n}] - \Delta_\psi(z_2, g_o, h_o)[u^*_{g_n}] - r_{\psi,h}(z_2, g_o, h_o)[h - h_o, u^*_{g_n}] \right| \leq \Lambda_{4,n}(z_2, \alpha);$$

(iv) there exists $\Lambda_{5,n}(z_2, \alpha)$ such that for any $\alpha \in \mathcal{N}_n$,

$$\left| r_{\psi,g}(z_2, g_o, h)[g - g_o, u^*_{g_n}] - r_{\psi,g}(z_2, g_o, h_o)[g - g_o, u^*_{g_n}] \right| \leq \Lambda_{5,n}(z_2, \alpha);$$

(v) $\max_{j=3,4,5} \sup_{\alpha \in \mathcal{N}_n} n^{-1/2} \sum_{i=1}^n \Lambda_{j,n}(Z_{2,i}, \alpha) = o_p(1)$; (vi) $\max_{j=3,4,5} \sup_{\alpha \in \mathcal{N}_n} \mathbb{E}[\Lambda_{j,n}(Z_2, \alpha)] = o(n^{1/2})$.

By Assumption 1.1.(i) and the definition of $\|\cdot\|_\psi$, we have

$$\langle v_{g,1}, v_{g,2} \rangle_\psi = \mathbb{E}[r_{\psi,g}(Z_2, \alpha_o)[v_{g,1}, v_{g,2}]]$$

for any $v_{g,1}, v_{g,2} \in \mathcal{V}_2$. By Assumption 1.2.(i), we have

$$\Gamma(\alpha_o)[v_h, v_g] = \mathbb{E}[r_{\psi,h}(Z_2, \alpha_o)[v_h, v_g]]$$

for any $v_h \in \mathcal{V}_1$ and any $v_g \in \mathcal{V}_2$.

Suppose that $\mathcal{F}$ is a class of functions of Z. Let F denote an envelope of $\mathcal{F}$,

$$F(z) \geq \sup_{f \in \mathcal{F}} |f(z)| \text{ for any } z \in \mathcal{Z}$$

where $\mathcal{Z}$ denotes the support of Z. For a probability measure Q and a constant q, such that $\|F\|_{Q,q} > 0$ (where $\|\cdot\|_{Q,q}$ denotes the L_q-norm under Q), we use $N(\varepsilon \|F\|_{Q,q}, \mathcal{F}, \|\cdot\|_{Q,q})$ to denote the minimal number of $\|\cdot\|_{Q,q}$-balls of radius $\varepsilon \|F\|_{Q,q}$ needed to cover $\mathcal{F}$. The supremum of $N(\varepsilon \|F\|_{Q,q}, \mathcal{F}, \|\cdot\|_{Q,q})$ over all finitely-discrete probability measures Q, is a uniform entropy number of $\mathcal{F}$.

Define

$$\mathcal{F}^*_{1,n} = \left\{ z_2 \mapsto r_{\psi,h}(z_2, \alpha_o)[h - h_{o,n}, u^*_{g_n}] : h \in \mathcal{N}_{h,n} \right\},$$
$$\mathcal{F}^*_{2,n} = \left\{ z_2 \mapsto r_{\psi,g}(z_2, \alpha_o)[g - g_{o,n}, u^*_{g_n}] : g \in \mathcal{N}_{g,n} \right\},$$

where $h_{o,n} \in \mathcal{H}_n$ and $g_{o,n} \in \mathcal{G}_n$ are such that $\|h_{o,n} - h_o\|_{\mathcal{H}} = O(\delta^*_{1,n})$ and $\|g_{o,n} - g_o\|_{\mathcal{G}} = O(\delta^*_{2,n})$.

Assumption 1.3 *(i)* $\mathbb{E}\left[\left|r_{\psi,h}(Z_2, \alpha_o)[h_{o,n} - h_o, u^*_{g_n}]\right|\right] = o(n^{-1/2})$*; (ii) let* $F^*_{1,n}$ *denote an envelope of* $\mathcal{F}^*_{1,n}$*, then*

$$\sup_Q N(\varepsilon \left\|F^*_{1,n}\right\|_{Q,2}, \mathcal{F}^*_{1,n}, L_2(Q)) \leq (C/\varepsilon)^{CL} \textit{ for any } \varepsilon \in (0,1];$$

(iii) $\mathbb{E}\left[\left|r_{\psi,g}(Z_2, \alpha_o)[g_{o,n} - g_o, u^*_{g_n}]\right|\right] = o(n^{-1/2})$*; (iv) let* $F^*_{2,n}$ *denote an envelope of* $\mathcal{F}^*_{2,n}$*, then*

$$\sup_Q N(\varepsilon \left\|F^*_{2,n}\right\|_{Q,2}, \mathcal{F}^*_{2,n}, L_2(Q)) \leq (C/\varepsilon)^{CK} \textit{ for any } \varepsilon \in (0,1];$$

(v) $\max_{j=1,2,}(\sup_{f \in \mathcal{F}^*_{j,n}} \mathbb{E}\left[f^2\right] + (K+L) \sup_{z_2 \in \mathcal{Z}_2} |F^2_{j,n}(z_2)| \log(n) n^{-1})^{1/2}((K+L)\log(n))^{1/2} = o(1)$.

Lemma 1.1 *Under Assumptions 1.1-1.3, Assumption 3.2 in HLR holds.*

Proof of Lemma 1.1. By Assumptions 1.1.(i)-(ii), and the triangle inequality,

$$\sup_{\alpha \in \mathcal{N}_n} \left| n^{-1} \sum_{i=1}^n \left[\begin{array}{c} \psi(Z_{2,i}, g^*, h) - \psi(Z_{2,i}, g, h) \\ -\Delta_\psi(Z_{2,i}, g, h)[\pm \kappa_n u^*_{g_n}] - \kappa_n^2 r_{\psi,g}(Z_{2,i}, g, h)[u^*_{g_n}, u^*_{g_n}] \end{array} \right] \right| \leq C\kappa_n^2 n^{-1} \sum_{i=1}^n \Lambda_{1,n}(Z_{2,i}) \quad (1.1)$$

which together with $\mathbb{E}\left[\Lambda_{1,n}(Z_2)\right] \leq C$ and the Markov inequality implies that

$$\sup_{\alpha\in\mathcal{N}_n}\left|n^{-1}\sum_{i=1}^{n}\left[\begin{array}{c}\psi(Z_{2,i},g^*,h)-\psi(Z_{2,i},g,h)\\ -\Delta_{\psi}(Z_{2,i},g,h)[\pm\kappa_n u^*_{g_n}]-\kappa_n^2 r_{\psi,g}(Z_{2,i},g,h)[u^*_{g_n},u^*_{g_n}]\end{array}\right]\right| = O_p(\kappa_n^2). \tag{1.2}$$

Similarly, by Assumptions 1.1.(i)-(ii), and the triangle inequality,

$$\sup_{\alpha\in\mathcal{N}_n}\left|\mathbb{E}\left[\begin{array}{c}\psi(Z_2,g^*,h)-\psi(Z_2,g,h)\\ -\Delta_{\psi}(Z_2,g,h)[\pm\kappa_n u^*_{g_n}]-\kappa_n^2 r_{\psi,g}(Z_2,g,h)[u^*_{g_n},u^*_{g_n}]\end{array}\right]\right| = O(\kappa_n^2), \tag{1.3}$$

which together with (1.2) implies that

$$\sup_{\alpha\in\mathcal{N}_n}\left|\mu_n\left\{\begin{array}{c}\psi(Z_2,g^*,h)-\psi(Z_2,g,h)\\ -\Delta_{\psi}(Z_2,g,h)[\pm\kappa_n u^*_{g_n}]-\kappa_n^2 r_{\psi,g}(Z_2,g,h)[u^*_{g_n},u^*_{g_n}]\end{array}\right\}\right| = O_p(\kappa_n^2). \tag{1.4}$$

By Assumptions 1.1.(iii), the triangle inequality and the Markov inequality,

$$\sup_{\alpha\in\mathcal{N}_n}\left|\mu_n\left\{r_{\psi,g}(Z_2,\alpha)[u^*_{g_n},u^*_{g_n}]-r_{\psi,g}(Z_2,\alpha_o)[u^*_{g_n},u^*_{g_n}]\right\}\right| = O_p(1) \tag{1.5}$$

which together with Assumptions 1.1.(iv), the triangle inequality and the Markov inequality implies that

$$\sup_{\alpha\in\mathcal{N}_n}\left|\mu_n\left\{r_{\psi,g}(Z_2,\alpha)[u^*_{g_n},u^*_{g_n}]\right\}\right| = O_p(1). \tag{1.6}$$

Combining the results in (1.4) and (1.6), and then applying the triangle inequality, we prove condition (12) of Assumption 3.2.(i) in HLR.

By Assumptions 1.2.(ii), 1.2.(v)-(vi), the triangle inequality and the Markov inequality,

$$\sup_{\alpha\in\mathcal{N}_n}\left|\mu_n\left\{\Delta_{\psi}(Z_2,g,h)[u^*_{g_n}]-\Delta_{\psi}(Z_2,g_o,h)[u^*_{g_n}]-r_{\psi,g}(Z_2,g_o,h)[g-g_o,u^*_{g_n}]\right\}\right| = o_p(n^{-1/2}). \tag{1.7}$$

Similarly, by Assumptions 1.2.(iv)-(vi), the triangle inequality and the Markov inequality,

$$\sup_{\alpha\in\mathcal{N}_n}\left|\mu_n\left\{r_{\psi,g}(Z_2,g_o,h)[g-g_o,u^*_{g_n}]-r_{\psi,g}(Z_2,g_o,h_o)[g-g_o,u^*_{g_n}]\right\}\right| = o_p(n^{-1/2}). \tag{1.8}$$

By Assumption 1.3.(iii), the triangle inequality and the Markov inequality,

$$\left|\mu_n\left\{r_{\psi,g}(Z_2,\alpha_o)[g_{o,n}-g_o,u^*_{g_n}]\right\}\right| = o_p(n^{-1/2}). \tag{1.9}$$

By Assumptions 1.3.(iv)-(v), we can use Lemma 22 in Belloni, et. al (2016) to show that

$$\sup_{g\in\mathcal{N}_{g,n}}\left|\mu_n\left\{r_{\psi,g}(Z_2,\alpha_o)[g-g_{o,n},u^*_{g_n}]\right\}\right| = o_p(n^{-1/2}), \tag{1.10}$$

which together with (1.9) implies that

$$\sup_{g\in\mathcal{N}_{g,n}}\left|\mu_n\left\{r_{\psi,g}(Z_2,\alpha_o)[g-g_o,u^*_{g_n}]\right\}\right| = o_p(n^{-1/2}). \tag{1.11}$$

Collecting the results in (1.7), (1.8) and (1.11), we get

$$\sup_{\alpha\in\mathcal{N}_n}\left|\mu_n\left\{\Delta_\psi(Z_2,g,h)[u^*_{g_n}]-\Delta_\psi(Z_2,g_o,h)[u^*_{g_n}]\right\}\right| = o_p(n^{-1/2}). \tag{1.12}$$

By Assumptions 1.2.(iii), 1.2.(v)-(vi), the triangle inequality and the Markov inequality,

$$\sup_{\alpha\in\mathcal{N}_n}\left|\mu_n\left\{\Delta_\psi(Z_2,g_o,h)[u^*_{g_n}]-\Delta_\psi(Z_2,\alpha_o)[u^*_{g_n}]-r_{\psi,h}(Z_2,\alpha_o)[h-h_o,u^*_{g_n}]\right\}\right| = o_p(n^{-1/2}). \tag{1.13}$$

By Assumption 1.3.(i), the triangle inequality and the Markov inequality,

$$\left|\mu_n\left\{r_{\psi,h}(Z_2,\alpha_o)[h_{o,n}-h_o,u^*_{g_n}]\right\}\right| = o_p(n^{-1/2}). \tag{1.14}$$

By Assumptions 1.3.(ii) and 1.3.(v), we can use Lemma 22 in Belloni, et. al (2016) to show that

$$\sup_{h\in\mathcal{N}_{h,n}}\left|\mu_n\left\{r_{\psi,h}(Z_2,\alpha_o)[h-h_{o,n},u^*_{g_n}]\right\}\right| = o_p(n^{-1/2}), \tag{1.15}$$

which together with (1.14) implies that

$$\sup_{g\in\mathcal{N}_{g,n}}\left|\mu_n\left\{r_{\psi,h}(Z_2,\alpha_o)[h-h_o,u^*_{g_n}]\right\}\right| = o_p(n^{-1/2}). \tag{1.16}$$

Collecting the results in (1.13) and (1.16), we get

$$\sup_{\alpha\in\mathcal{N}_n}\left|\mu_n\left\{\Delta_\psi(Z_2,g_o,h)[u^*_{g_n}]-\Delta_\psi(Z_2,\alpha_o)[u^*_{g_n}]\right\}\right| = o_p(n^{-1/2}). \tag{1.17}$$

Combining the results in (1.12) and (1.17), and then applying the triangle inequality, we immediately prove condition (13) of Assumption 3.2.(i) in HLR.

By Assumptions 1.1.(ii)-(iv),

$$\mathbb{E}\left[\psi(Z_2, g^*, h) - \psi(Z_2, g, h)\right] = \pm\kappa_n \mathbb{E}\left[\Delta_\psi(Z_2, g, h)[u^*_{g_n}]\right] + O(\kappa_n^2), \tag{1.18}$$

uniformly over $\alpha \in \mathcal{N}_n$. As $\mathbb{E}\left[\Delta_\psi(Z_2, g_o, h_o)[u^*_{g_n}]\right] = 0$, by Assumptions 1.2.(ii)-(iv) and 1.2.(vi)

$$\begin{aligned}
&\mathbb{E}\left[\Delta_\psi(Z_2, g, h)[u^*_{g_n}]\right] \\
&= \mathbb{E}\left[\Delta_\psi(Z_2, g, h)[u^*_{g_n}] - \Delta_\psi(Z_2, g_o, h)[u^*_{g_n}] - r_{\psi,g}(Z_2, g_o, h)[g - g_o, u^*_{g_n}]\right] \\
&+ \mathbb{E}\left[\Delta_\psi(Z_2, g_o, h)[u^*_{g_n}] - \Delta_\psi(Z_2, g_o, h_o)[u^*_{g_n}] - r_{\psi,h}(Z_2, g_o, h_o)[h - h_o, u^*_{g_n}]\right] \\
&+ \mathbb{E}\left[r_{\psi,g}(Z_2, g_o, h)[g - g_o, u^*_{g_n}] - r_{\psi,g}(Z_2, g_o, h_o)[g - g_o, u^*_{g_n}]\right] \\
&+ \mathbb{E}\left[r_{\psi,g}(Z_2, g_o, h_o)[g - g_o, u^*_{g_n}]\right] + \mathbb{E}\left[r_{\psi,h}(Z_2, g_o, h_o)[h - h_o, u^*_{g_n}]\right] \\
&= \mathbb{E}\left[r_{\psi,g}(Z_2, \alpha_o)[g - g_o, u^*_{g_n}]\right] + \mathbb{E}\left[r_{\psi,h}(Z_2, \alpha_o)[h - h_o, u^*_{g_n}]\right] + o(n^{-1/2}) \\
&= \langle g - g_o, u^*_{g_n}\rangle_\psi + \Gamma(\alpha_o)\left[h - h_o, u^*_{g_n}\right] + o(n^{-1/2})
\end{aligned} \tag{1.19}$$

where the second equality is by the definition of the inner product $\langle\cdot,\cdot\rangle_\psi$ and the functional $\Gamma(\alpha_o)[\cdot,\cdot]$. By Assumption 1.1.(v), (1.18), (1.19) and the definition of $K_\psi(g, h)$, we have

$$K_\psi(g, h) - K_\psi(g^*, h) = \mp\kappa_n\left[\langle g - g_o, u^*_{g_n}\rangle_\psi + \Gamma(\alpha_o)\left[h - h_{o,n}, u^*_{g_n}\right]\right] + O(\kappa_n^2). \tag{1.20}$$

By the definition of $||\cdot||_\psi$ and Assumption 1.1.(iv),

$$\frac{||g^* - g_o||^2_\psi - ||g - g_o||^2_\psi}{2} = \langle g - g_o, \pm\kappa_n u^*_{g_n}\rangle_\psi + O(\kappa_n^2). \tag{1.21}$$

Collecting the results in (1.20) and (1.21), we immediately prove Assumption 3.2.(ii) in HLR. ■

We next provide sufficient conditions for Assumptions 3.2 and 3.4 in HLR when the criterion function in the second-step M estimation takes the following form

$$\psi(Z_2, g, h) = \tau(Z_1, h)\psi^*(Z_2, g, h). \tag{1.22}$$

We will assume that Assumptions 1.1.(i) and 1.2.(i) hold for $\psi^*(Z_2, g, h)$. Define

$$\Delta^*_\psi(z_2, \alpha)[v_{g,1}] = \left.\frac{\partial\psi^*(z_2, g + \tau v_{g,1}, h)}{\partial\tau}\right|_{\tau=0} \text{ and } r^*_{\psi,g}(z_2, \alpha)[v_{g,1}, v_{g,2}] = \left.\frac{\partial\Delta^*_\psi(z_2, g + \tau v_{g,2}, h)[v_{g,1}]}{\partial\tau}\right|_{\tau=0},$$

for any $z_2 \in \mathcal{Z}_2$, any $\alpha \in \mathcal{N}_\alpha$ and any $v_{g,1}, v_{g,2} \in \mathcal{V}_2$. Then we have

$$\Delta_\psi(z_2,\alpha)[v_{g,1}] = \tau(z_1,h)\Delta^*_\psi(z_2,\alpha)[v_{g,1}] \text{ and } r_{\psi,g}(z_2,\alpha)[v_{g,1},v_{g,2}] = \tau(z_1,h)r^*_{\psi,g}(z_2,\alpha)[v_{g,1},v_{g,2}]$$

for any $\alpha \in \mathcal{N}_\alpha$ and any $v_{g,1}, v_{g,2} \in \mathcal{V}_2$. Define

$$r_{\psi,h}(z_2,\alpha)[v_h,v_g] = \tau(z_1,h)r^*_{\psi,h}(z_2,\alpha)[v_h,v_g],$$

where

$$r^*_{\psi,h}(z_2,\alpha)[v_h,v_g] = \left.\frac{\partial\Delta^*_\psi(z_2,g,h+\tau v_h)[v_g]}{\partial\tau}\right|_{\tau=0}.$$

Let ξ_n denote a non-decreasing real positive sequence, and $\delta^*_{\tau,n}$ denote a real positive sequence.

Assumption 1.4 *(i) $sup_{z_1\in\mathcal{Z}_1,h\in\mathcal{N}_{h,n}}[|\tau(z_1,h)|+|\tau(z_1,h_o)|] \leq C$; (ii) Assumptions 1.1.(i)-(ii) and 1.1.(v) hold; (iii) equation (19) in HLR holds; (iv) $\Delta^*_\psi(z_2,\alpha)[v_g]$ satisfies Assumption 1.2.(i); (v) Assumptions 1.2.(ii) and 1.2.(v)-(vi) hold; (vi) $\sup_{z_1\in\mathcal{Z}_1}\mathbb{E}\left[(\Delta^*_\psi(Z_2,\alpha_o)[u^*_{g_n}])^2\middle| Z_1=z_1\right] \leq \xi_n^2$; (vii)*

$$\sup_{h\in\mathcal{N}_{h,n}} n^{-1}\sum_{i=1}^n(\tau(Z_{1,i},h)-\tau(Z_{1,i},h_o))^2 = O_p(\delta^*_{\tau,n})$$

where $\delta^*_{\tau,n}\xi_n^2 = o(1)$.

Assumption 1.5 *(i) there exists $\Lambda_{6,n}(z_2,\alpha)$ such that for any $\alpha\in\mathcal{N}_n$*

$$\left|\tau(z_1,h)\left(r^*_{\psi,g}(z_2,g_o,h)[g-g_o,u^*_{g_n}]-r^*_{\psi,g}(z_2,\alpha_o)[g-g_o,u^*_{g_n}]\right)\right| \leq \Lambda_{6,n}(z_2,\alpha);$$

(ii) there exists $\Lambda_{7,n}(z_2,\alpha)$ such that for any $\alpha\in\mathcal{N}_n$

$$\left|\tau(z_1,h)\left(\Delta^*_\psi(z_2,g_o,h)[u^*_{g_n}]-\Delta^*_\psi(z_2,\alpha_o)[u^*_{g_n}]-r^*_{\psi,h}(z_2,\alpha_o)[h-h_o,u^*_{g_n}]\right)\right| \leq \Lambda_{7,n}(z_2,\alpha);$$

(iii)

$$\sup_{h\in\mathcal{N}_{h,n}}\left|\mathbb{E}\left[(\tau(Z_1,h)-\tau(Z_1,h_o))r^*_{\psi,h}(Z_2,\alpha_o)[h-h_o,u^*_{g_n}]\right]\right| = o(n^{-1/2});$$

(iv)

$$\sup_{\alpha\in\mathcal{N}_n}\left|\mathbb{E}\left[(\tau(Z_1,h)-\tau(Z_1,h_o))r^*_{\psi,g}(Z_2,\alpha_o)[g-g_o,u^*_{g_n}]\right]\right| = o(n^{-1/2});$$

(v) there exists $\Lambda_{8,n}(z_2)$ *with* $\mathbb{E}\left[\Lambda_{8,n}(Z_2)\right] \leq C$

$$\sup_{\alpha \in \mathcal{N}_n} \left| \tau(z_1, h)(r^*_{\psi,g}(z_2, \alpha)[u^*_{g_n}, u^*_{g_n}] - r^*_{\psi,g}(z_2, \alpha_o)[u^*_{g_n}, u^*_{g_n}]) \right| \leq \Lambda_{8,n}(z_2);$$

(vi) $\mathbb{E}\left[\left| r^*_{\psi,g}(Z_2, \alpha_o)[u^*_{g_n}, u^*_{g_n}] \right|\right] \leq C$*; (vii)* $\max_{j=6,7} \sup_{\alpha \in \mathcal{N}_n} n^{-1} \sum_{i=1}^{n} \Lambda_{j,n}(Z_{2,i}, \alpha) = o_p(n^{-1/2})$*; (viii)* $\max_{j=6,7} \sup_{\alpha \in \mathcal{N}_n} \mathbb{E}\left[\Lambda_{j,n}(Z_2, \alpha)\right] = o(n^{-1/2})$.

Define

$$\mathcal{F}^*_{3,n} = \left\{ z_2 \mapsto \tau(z_1, h) r^*_{\psi,h}(z_2, \alpha_o)[h - h_{o,n}, u^*_{g_n}] : h \in \mathcal{N}_{h,n} \right\},$$
$$\mathcal{F}^*_{4,n} = \left\{ z_2 \mapsto \tau(z_1, h) r^*_{\psi,g}(z_2, \alpha_o)[g - g_{o,n}, u^*_{g_n}] : h \in \mathcal{N}_{h,n}, g \in \mathcal{N}_{g,n} \right\}.$$

Assumption 1.6 *(i)* $\mathbb{E}[|r^*_{\psi,h}(Z_2, \alpha_o)[h_{o,n} - h_o, u^*_{g_n}]|] = o(n^{-1/2})$*; (ii) let* $F^*_{3,n}$ *denote an envelope of* $\mathcal{F}^*_{3,n}$*, then*

$$\sup_{Q} N(\varepsilon \left\| F^*_{3,n} \right\|_{Q,2}, \mathcal{F}^*_{3,n}, L_2(Q)) \leq (C/\varepsilon)^{CL} \text{ for any } \varepsilon \in (0, 1];$$

(iii) $\mathbb{E}\left[\left| r^*_{\psi,g}(Z_2, \alpha_o)[g_{o,n} - g_o, u^*_{g_n}] \right|\right] = o(n^{-1/2})$*; (iv) let* $F^*_{4,n}$ *denote an envelope of* $\mathcal{F}^*_{4,n}$*, then*

$$\sup_{Q} N(\varepsilon \left\| F^*_{4,n} \right\|_{Q,2}, \mathcal{F}^*_{4,n}, L_2(Q)) \leq (C/\varepsilon)^{C(L+K)} \text{ for any } \varepsilon \in (0, 1];$$

(v) $\max_{j=3,4}(\sup_{f \in \mathcal{F}^*_{j,n}} \mathbb{E}\left[f^2\right] + (K+L) \sup_{z_2 \in \mathcal{Z}_2} |F^*_{j,n}(z_2)| \log(n) n^{-1})^{1/2} ((K+L)\log(n))^{1/2} = o(1)$.

By definition, we have $\langle v_{g,1}, v_{g,2} \rangle_\psi = \mathbb{E}\left[\tau(Z_1, h_o) r^*_{\psi,g}(z_2, \alpha_o)[v_{g,1}, v_{g,2}]\right]$ for any $v_{g,1}, v_{g,2} \in \mathcal{V}_2$. Moreover, by (19) in HLR,

$$\Gamma(\alpha_o)\left[v_h, v_g\right] = \mathbb{E}\left[\tau(Z_1, h_o) r^*_{\psi,h}(Z_2, \alpha_o)[v_h, v_g]\right]$$

for any $v_h \in \mathcal{V}_1$ and any $v_g \in \mathcal{V}_2$.

Lemma 1.2 *Under Assumptions 1.4-1.6, condition (13) of Assumption 3.2, Assumption 3.2.(ii) and Assumption 3.4 in HLR holds.*

Proof of Lemma 1.2. By Assumptions 1.1.(i)-(ii), we can use the same arguments in the proof of Lemma 1.1 to show that

$$\sup_{\alpha \in \mathcal{N}_n} \left| \mu_n \left\{ \begin{array}{c} \psi(Z_2, g^*, h) - \psi(Z_2, g, h) \\ -\Delta_\psi(Z_2, g, h)[\pm \kappa_n u^*_{g_n}] - \kappa_n^2 r_{\psi,g}(Z_2, g, h)[u^*_{g_n}, u^*_{g_n}] \end{array} \right\} \right| = O_p(\kappa_n^2). \quad (1.23)$$

By Assumptions 1.5.(v), 1.5.(vii)-(viii), the triangle inequality and the Markov inequality,

$$\sup_{\alpha\in\mathcal{N}_n}\left|\mu_n\left\{\tau(Z_1,h)(r^*_{\psi,g}(Z_2,\alpha)[u^*_{g_n},u^*_{g_n}]-r^*_{\psi,g}(Z_2,\alpha_o)[u^*_{g_n},u^*_{g_n}])\right\}\right|=O_p(1). \tag{1.24}$$

By Assumptions 1.4.(i) and 1.5.(vi), the triangle inequality and the Markov inequality, which together with (1.24) and the triangle inequality implies that

$$\sup_{\alpha\in\mathcal{N}_n}\left|\mu_n\left\{\tau(Z_1,h)r^*_{\psi,g}(Z_2,\alpha_o)[u^*_{g_n},u^*_{g_n}]\right\}\right|=O_p(1). \tag{1.25}$$

Combining the results in (1.23)-(1.25), and then applying the triangle inequality, we prove condition (13) of Assumption 3.2.(i) in HLR.

By Assumption 1.5.(v),

$$\sup_{\alpha\in\mathcal{N}_n}\left|\mathbb{E}\left[\tau(Z_1,h)(r^*_{\psi,g}(Z_2,\alpha)[u^*_{g_n},u^*_{g_n}]-r^*_{\psi,g}(Z_2,\alpha_o)[u^*_{g_n},u^*_{g_n}])\right]\right|=O(1). \tag{1.26}$$

By Assumptions 1.4.(i) and 1.5.(vi),

$$\sup_{h\in\mathcal{N}_{h,n}}\left|\mathbb{E}\left[(\tau(Z_1,h)-\tau(Z_1,h_o))r^*_{\psi,g}(Z_2,\alpha_o)[u^*_{g_n},u^*_{g_n}]\right]\right|=O(1) \tag{1.27}$$

and

$$\left|\mathbb{E}\left[\tau(Z_1,h_o)r^*_{\psi,g}(Z_2,\alpha_o)[u^*_{g_n},u^*_{g_n}]\right]\right|=O(1), \tag{1.28}$$

which together with (1.26) and the triangle inequality implies that

$$\sup_{\alpha\in\mathcal{N}_n}\left|\mathbb{E}\left[r_{\psi,g}(Z_2,\alpha)[u^*_{g_n},u^*_{g_n}]\right]\right|=O(1). \tag{1.29}$$

By Assumptions 1.1.(ii), (1.29) and the triangle inequality,

$$\mathbb{E}\left[\tau(Z_1,h)\left[\psi^*(Z_2,g^*,h)-\psi^*(Z_2,g,h)\right]\right]=\pm\kappa_n\mathbb{E}\left[\tau(Z_1,h)\Delta^*_{\psi}(Z_2,g,h)[u^*_{g_n}]\right]+O(\kappa_n^2), \tag{1.30}$$

uniformly over $\alpha \in \mathcal{N}_n$. By $\mathbb{E}[\tau(Z_1,h)\Delta^*_\psi(Z_2,\alpha)[u^*_{g_n}]] = 0$, Assumptions 1.4.(v), 1.5.(i)-(iv) and 1.5.(viii),

$$\begin{aligned}
&\mathbb{E}\left[\tau(Z_1,h)\Delta^*_\psi(Z_2,g,h)[u^*_{g_n}]\right] \\
&= \mathbb{E}\left[\tau(Z_1,h)\Delta^*_\psi(Z_2,g_o,h)[u^*_{g_n}]\right] \\
&\quad + \mathbb{E}\left[\tau(Z_1,h)r^*_{\psi,g}(Z_2,g_o,h)[g-g_o,u^*_{g_n}]\right] + o(n^{-1/2}) \\
&= \mathbb{E}\left[\tau(Z_1,h)r^*_{\psi,h}(z_2,\alpha_o)[h-h_o,u^*_{g_n}]\right] \\
&\quad + \mathbb{E}\left[\tau(Z_1,h)r^*_{\psi,g}(Z_2,\alpha_o)[g-g_o,u^*_{g_n}]\right] + o(n^{-1/2}) \\
&= \mathbb{E}\left[r_{\psi,h}(z_2,\alpha_o)[h-h_o,u^*_{g_n}]\right] + \mathbb{E}\left[r_{\psi,g}(Z_2,\alpha_o)[g-g_o,u^*_{g_n}]\right] + o(n^{-1/2}) \\
&= \Gamma(\alpha_o)\left[h-h_o,u^*_{g_n}\right] + \langle g-g_o,u^*_{g_n}\rangle_\psi + o(n^{-1/2}), \qquad (1.31)
\end{aligned}$$

where the last equality is by the definitions of the inner product $\langle\cdot,\cdot\rangle_\psi$ and the functional $\Gamma(\alpha_o)[\cdot,\cdot]$. By Assumption 1.1.(v), (1.30), (1.31) and the definition of $K_\psi(g,h)$, we have

$$K_\psi(g,h) - K_\psi(g^*,h) = \mp\kappa_n\left[\langle g-g_o,u^*_{g_n}\rangle_\psi + \Gamma(\alpha_o)\left[h-h_{o,n},u^*_{g_n}\right]\right] + O(\kappa_n^2). \qquad (1.32)$$

By the definition of $||\cdot||_\psi$, Assumptions 1.4.(i) and 1.5.(vi),

$$\frac{||g^*-g_o||^2_\psi - ||g-g_o||^2_\psi}{2} = \langle g-g_o,\pm\kappa_n u^*_{g_n}\rangle_\psi + O(\kappa_n^2). \qquad (1.33)$$

Collecting the results in (1.32) and (1.33), we immediately prove Assumption 3.2.(ii) in HLR.

We next verify Assumption 3.4 in HLR. Assumptions 3.4.(i)-(ii) are assumed directly. By definition,

$$\begin{aligned}
&\Delta_\psi(z_2,g,h)[u^*_{g_n}] - \Delta_\psi(z_2,g_o,h)[u^*_{g_n}] \\
&= \tau(z_1,h)r^*_{\psi,g}(z_2,\alpha_o)[g-g_{o,n},u^*_{g_n}] \\
&+ \tau(z_1,h)r^*_{\psi,g}(z_2,\alpha_o)[g_{o,n}-g_o,u^*_{g_n}] \\
&+ \left[\Delta_\psi(z_2,g,h)[u^*_{g_n}] - \Delta_\psi(z_2,g_o,h)[u^*_{g_n}] - r_{\psi,g}(z_2,g_o,h)[g-g_o,u^*_{g_n}]\right] \\
&+ \tau(z_1,h)\left(r^*_{\psi,g}(z_2,g_o,h)[g-g_o,u^*_{g_n}] - r^*_{\psi,g}(z_2,\alpha_o)[g-g_o,u^*_{g_n}]\right). \qquad (1.34)
\end{aligned}$$

By Assumptions 1.4.(v), 1.2.(v)-(vi), 1.5.(i) and 1.5.(vii)-(viii), and the Markov inequality,

$$\sup_{\alpha\in\mathcal{N}_n}\left|\mu_n\left\{\begin{array}{c}\Delta_\psi(Z_2,g,h)[u^*_{g_n}] - \Delta_\psi(Z_2,g_o,h)[u^*_{g_n}] \\ -r_{\psi,g}(Z_2,g_o,h)[g-g_o,u^*_{g_n}]\end{array}\right\}\right| = o_p(n^{-1/2}), \qquad (1.35)$$

and

$$\sup_{\alpha\in\mathcal{N}_n}\left|\mu_n\left\{\tau(Z_1,h)\left(\begin{array}{c} r^*_{\psi,g}(Z_2,g_o,h)[g-g_o,u^*_{g_n}] \\ -r^*_{\psi,g}(Z_2,g_o,h_o)[g-g_o,u^*_{g_n}] \end{array}\right)\right\}\right| = o_p(n^{-1/2}). \tag{1.36}$$

By Assumptions 1.4.(i), 1.6.(iii) and the Markov inequality,

$$\sup_{\alpha\in\mathcal{N}_n}\left|\mu_n\{\tau(Z_1,h)r^*_{\psi,g}(Z_2,\alpha_o)[g_{o,n}-g_o,u^*_{g_n}]\}\right| = o_p(n^{-1/2}). \tag{1.37}$$

By Assumptions 1.6.(iv)-(v), we can use Lemma 22 in Belloni, et. al (2016) to show that

$$\sup_{h\in\mathcal{N}_{h,n}}\left|\mu_n\left\{\tau(Z_1,h)r^*_{\psi,g}(Z_2,\alpha_o)[g-g_{o,n},u^*_{g_n}]\right\}\right| = o_p(n^{-1/2}). \tag{1.38}$$

Collecting the results in (1.34)-(1.38), and then applying the triangle inequality, we get

$$\sup_{h\in\mathcal{N}_{h,n}}\left|\mu_n\left\{\Delta_\psi(Z_2,g,h)[u^*_{g_n}]-\Delta_\psi(Z_2,g_o,h)[u^*_{g_n}]\right\}\right| = o_p(n^{-1/2}), \tag{1.39}$$

which proves condition (20) in Assumption 3.4.(iii). By definition,

$$\begin{aligned}
&\tau(z_1,h)(\Delta^*_\psi(z_2,g_o,h)[u^*_{g_n}]-\Delta^*_\psi(z_2,g_o,h_o)[u^*_{g_n}]) \\
&=\tau(z_1,h)r^*_{\psi,h}(z_2,\alpha_o)[h-h_{o,n},u^*_{g_n}] \\
&+\tau(z_1,h)r^*_{\psi,h}(z_2,\alpha_o)[h_{o,n}-h_o,u^*_{g_n}] \\
&+\tau(z_1,h)(\Delta^*_\psi(z_2,g_o,h)[u^*_{g_n}]-\Delta^*_\psi(z_2,\alpha_o)[u^*_{g_n}]-r^*_{\psi,h}(z_2,\alpha_o)[h-h_o,u^*_{g_n}]).
\end{aligned} \tag{1.40}$$

By Assumptions 1.5.(ii), 1.5.(vii)-(viii), the Markov inequality and the triangle inequality,

$$\sup_{h\in\mathcal{N}_{h,n}}\left|\mu_n\left\{\begin{array}{c}\tau(Z_1,h)(\Delta^*_\psi(Z_2,g_o,h)[u^*_{g_n}]-\Delta^*_\psi(Z_2,\alpha_o)[u^*_{g_n}] \\ -r^*_{\psi,h}(Z_2,\alpha_o)[h-h_o,u^*_{g_n}])\end{array}\right\}\right| = o_p(n^{-1/2}). \tag{1.41}$$

By Assumptions 1.4.(i), 1.6.(i), the Markov inequality and the triangle inequality,

$$\sup_{\alpha\in\mathcal{N}_n}\left|\mu_n\{\tau(Z_1,h)r^*_{\psi,h}(Z_2,\alpha_o)[h_{o,n}-h_o,u^*_{g_n}]\}\right| = o_p(n^{-1/2}). \tag{1.42}$$

By Assumptions 1.6.(ii)-(iii), we can use Lemma 22 in Belloni, et. al (2016) to show that

$$\sup_{h\in\mathcal{N}_{h,n}}\left|\mu_n\{\tau(Z_1,h)r^*_{\psi,h}(Z_2,\alpha_o)[h-h_{o,n},u^*_{g_n}]\}\right| = o_p(n^{-1/2}). \tag{1.43}$$

Collecting the results in (1.40)-(1.43), and then applying the triangle inequality, we get

$$\sup_{h\in\mathcal{N}_{h,n}} \left|\mu_n\left\{\tau(Z_1,h)(\Delta_\psi^*(Z_2,g_o,h)[u_{g_n}^*]-\Delta_\psi^*(Z_2,g_o,h_o)[u_{g_n}^*])\right\}\right| = o_p(n^{-1/2}), \tag{1.44}$$

which proves condition (21) in Assumption 3.4.(iii). Finally, Assumptions 3.4.(iv) in HLR follows by Assumptions 1.4.(vi)-(vii). ■

2 Some Auxiliary Lemmas for Theorem 5.1 of HLR

For the completeness of this section, we list the sufficient conditions of Theorem 5.1 in HLR. To facilitate the presentation, we first review some notations introduced in Section 5 and Appendix D of HLR. Recall that the basis functions used in the first-step and second-step M estimations are $L\times 1$ vector $R(x)$ and $K\times 1$ vector $P(\varepsilon)$ respectively. For $j=1,2$, we define $\upsilon_{j,K}=\sup_{\varepsilon\in\mathcal{E}_\eta}\left\|\partial^j P(\varepsilon)'\beta_{o,K}\right\|$, where $\mathcal{E}_\eta=[a-\eta,b+\eta]$ for some $a<b$ and small $\eta>0$, and $\beta_{o,K}\in\mathbb{R}^K$ is defined in Assumption 2.2.(iii) below. Let $\mathcal{N}_{g,n}=\{g\in\mathcal{G}_n:\|g-g_o\|_2\leq\delta_{2,n}^*\log(\log(n))\}$ denote the local neighborhood of g_o, where $\mathcal{G}_n$ denotes the sieve space of estimating g_o, $\delta_{2,n}^*=K^{1/2}n^{-1/2}+K^{-\rho_g}+\upsilon_{1,K}\delta_{h,n}^*$ and $\delta_{h,n}^*=L^{1/2}n^{-1/2}+L^{-\rho_h}$. For any column vector a, let $\|a\|$ denote its ℓ_2-norm; for any square matrix A, the operator norm is denoted by $||A||$; $\omega_{\max}(A)$ and $\omega_{\min}(A)$ denote the largest and smallest eigenvalues of a square matrix A, respectively. We use C to denote some generic finite positive constant larger than 1. For d a nonnegative integer, let $|g|_d=\max_{|\tau|\leq d}\sup_{\varepsilon\in\mathcal{E}}|\partial^\tau g(\varepsilon)|$ for any $g\in\mathcal{G}$ where $\mathcal{G}$ is the function space containing g_o. Let $\|\cdot\|_\infty$ denote the uniform norm. For any function f, $\mu_n(f)=n^{-1}\sum_{i=1}^n[f(Z_i)-\mathbb{E}[f(Z_i)]]$ denotes the empirical process indexed by f.

Assumption 2.1 *(i) The data $\{y_i,x_i,s_i\}_{i=1}^n$ is i.i.d.; (ii) $\mathbb{E}\left[\varepsilon_i^4\middle|x_i\right]<C$ and $\mathbb{E}[\varepsilon_i^2|x_i]>C^{-1}$; (iii) there exist $\rho_h>0$ and $\gamma_{o,L}\in\mathbb{R}^L$ such that*

$$\|h_{o,L}-h_o\|_\infty=O(L^{-\rho_h})$$

where $h_{o,L}(\cdot)\equiv R(\cdot)'\gamma_{o,L}$; (iv) the eigenvalues of Q_L are between C^{-1} and C for all L; (v) there exists a nondecreasing sequence ζ_L such that $\sup_{x\in\mathcal{X}}\|R(x)\|\leq\zeta_L$.

Assumption 2.2 *(i) $\mathbb{E}[u_i^4|\varepsilon_i]<C$ and $\mathbb{E}[u_i^2|\varepsilon_i]>C^{-1}$; (ii) $g_o(\varepsilon)$ is twice continuously differentiable;*

(iii) there exist $\rho_g > 0$ and $\beta_{o,K} \in \mathbb{R}^K$ such that

$$|g_{o,K} - g_o|_d = O(K^{-\rho_g})$$

where $g_{o,K}(\cdot) = P(\cdot)' \beta_{o,K}$ and $d = 1$; (iv) the eigenvalues of Q_K are between C^{-1} and C for all K; (v) for $j = 0, 1, 2$, there exists a nondecreasing sequence $\xi_{j,K}$ such that $\sup_{\varepsilon \in \mathcal{E}_\eta} \left\| \partial^j P(\varepsilon) \right\| \leq \xi_{j,K}$.

Assumption 2.3 *(i) $||v^*_{g_n}||_2 \geq C$ for all n; (ii) the functional $\rho(\cdot)$ satisfies*

$$\sup_{g \in \mathcal{N}_{g,n}} \left| \frac{\rho(g) - \rho(g_o) - \partial\rho(g_o)[g - g_o]}{\|v^*_n\|_{sd}} \right| = o(n^{-1/2});$$

(iii) $\left| \|v^*_n\|^{-1}_{sd}\, \partial\rho(g_o)[g_{o,n} - g_o] \right| = o(n^{-1/2})$; (iv) $\sup_{g \in \mathcal{N}_{g,n}} \|\partial\rho(g)[P] - \partial\rho(g_o)[P]\| = o(1)$.

Assumption 2.4 *The following conditions hold:*

(i) $n^{-1/2}(K + L)^{1/2}(\xi_{0,K} + \zeta_L)(\log(n))^{1/2} = o(1)$;

(ii) $n^{-1}(L\xi^2_{1,K} \log(n) + \zeta_L \xi_{1,K}) = o(1)$;

(iii) $n^{-1/2}\zeta_L(L\xi_{2,K} + L^{1/2}\xi_{1,K})(n^{-1/2}K^{1/2} + K^{-\rho_g} + v_{1,K}n^{-1/2}L^{1/2}) \log(n) = o(1)$;

(iv) $n^{-1/2}\zeta_L(L + L^{1/2}v_{1,K} + Lv_{2,K}) \log(n) = o(1)$;

(v) $nL^{1-2\rho_h} + K^{-\rho_g} = o(1)$.

Assumption 2.5 *The following conditions hold:*

(i) $||v^*_{g_n}||_2 \leq C$ for all n.

(ii) $(n^{-1}K\xi^2_{1,K} + (\zeta^2_L + \xi^2_{0,K} + \xi^2_{1,K})K^{-2\rho_g}) \log(n) = o(1)$;

(iii) $n^{-1}(\zeta^2_L + \xi^2_{0,K} + \xi^2_{1,K})v^2_{1,K} L \log(n) = o(1)$.

Lemma 2.1 *Under Assumptions 2.1, 2.2.(iv)-(v), 2.4.(i) and 2.4.(v), we have*

$$\left\| \widehat{Q}_{n,K} - Q_K \right\| = O_p(\xi^2_{1,K}\delta^{*2}_{h,n} + \xi_{1,K}\delta^*_{h,n} + n^{-1/2}\xi_{0,K}(\log K)^{1/2}),$$

where $\delta^*_{h,n} = L^{1/2}n^{-1/2} + L^{-\rho_h}$.

Proof of Lemma 2.1. Let $B_K = \{\lambda_K \in \mathbb{R}^K : \lambda'_K \lambda_K = 1\}$. Under Assumptions 2.1.(i), 2.2.(iv)-(v) and 2.4.(i), we can invoke Lemma 6.2 of Belloni, et al. (2015) to get

$$\sup_{\lambda_K \in B_K} \left| n^{-1} \sum_{i=1}^{n} \left[\left| \lambda'_K P(\varepsilon_i) \right|^2 \right] - \mathbb{E}\left[\left| \lambda'_K P(\varepsilon_i) \right|^2 \right] \right| = O_p(n^{-1/2}\xi_{0,K}(\log K)^{1/2}), \tag{2.1}$$

which (together with Assumption 2.4.(i)) further implies that

$$\|Q_{n,K} - Q_K\| = o_p(1) \tag{2.2}$$

Under Assumptions 2.1 and 2.4.(i), arguments in the proof of Theorem 4.1 in Belloni et al. (2015) show that

$$\|\widehat{\gamma}_n - \gamma_{o,L}\| = O_p(\delta^*_{h,n}), \tag{2.3}$$

which together with Assumptions 2.1.(iii)-(iv), and (2.52) below (which is proved under Assumptions 2.1 and 2.4.(i)) implies that

$$\begin{aligned} n^{-1}\sum_{i=1}^{n}\left[\left|\widehat{h}_n(x_i) - h_o(x_i)\right|^2\right] &\le 2n^{-1}\sum_{i=1}^{n}\left[\left|\widehat{h}_n(x_i) - h_{o,L}(x_i)\right|^2\right] + 2n^{-1}\sum_{i=1}^{n}\left[\left|h_{o,L}(x_i) - h_o(x_i)\right|^2\right] \\ &= 2(\widehat{\gamma}_n - \gamma_{o,L})' Q_{n,L}(\widehat{\gamma}_n - \gamma_{o,L}) + O(L^{-2\rho_h}) \\ &\le \omega_{\min}(Q_{n,L})\left\|\widehat{\gamma}_n - \gamma_{o,L}\right\|^2 + O(L^{-2\rho_h}) = O_p(\delta^{*2}_{h,n}). \end{aligned} \tag{2.4}$$

Then by (2.4), and the definition of $\widehat{\varepsilon}_i$,

$$n^{-1}\sum_{i=1}^{n}\left[\left|\widehat{\varepsilon}_i - \varepsilon_i\right|^2\right] = n^{-1}\sum_{i=1}^{n}\left[\left|\widehat{h}_n(x_i) - h_o(x_i)\right|^2\right] = O_p(\delta^{*2}_{h,n}). \tag{2.5}$$

Using (2.3), Assumptions 2.1.(iii), (v) and 2.4.(i), 2.4.(v), we have

$$\begin{aligned} \left\|\widehat{h}_n - h_o\right\|_\infty &\le \left\|\widehat{h}_n - h_{o,K}\right\|_\infty + \|h_{o,K} - h_o\|_\infty \\ &= \left\|R(x)'(\widehat{\gamma}_n - \gamma_{o,L})\right\|_\infty + O(K^{-\rho_h}) \\ &\le \zeta_L\left\|\widehat{\gamma}_n - \gamma_{o,L}\right\| + O(K^{-\rho_h}) = O_p(\zeta_L\delta^*_{h,n}), \end{aligned} \tag{2.6}$$

which implies that

$$\max_{i\le n}\left|\widehat{\varepsilon}_i - \varepsilon_i\right| = \max_{i\le n}\left|\widehat{h}_n(x_i) - h_o(x_i)\right| \le \left\|\widehat{h}_n - h_o\right\|_\infty = o_p(1). \tag{2.7}$$

For any $\lambda_K \in B_K$, by the mean value expansion, the triangle inequality and the Cauchy-Schwarz inequality,

$$\begin{aligned}
\left| \left|\lambda_K' P(\widehat{\varepsilon}_i)\right|^2 - \left|\lambda_K' P(\varepsilon_i)\right|^2 \right| &\leq \left|\lambda_K'(P(\widehat{\varepsilon}_i) - P(\varepsilon_i))\right|^2 + 2\left|\lambda_K'(P(\widehat{\varepsilon}_i) - P(\varepsilon_i))\lambda_K' P(\varepsilon_i)\right| \\
&= \left|\lambda_K' \partial P(\widetilde{\varepsilon}_i)(\widehat{\varepsilon}_i - \varepsilon_i)\right|^2 + 2\left|\lambda_K' \partial P(\widetilde{\varepsilon}_i)\lambda_K' P(\varepsilon_i)(\widehat{\varepsilon}_i - \varepsilon_i)\right| \\
&\leq \left\|\partial P(\widetilde{\varepsilon}_i)\right\|^2 \left|\widehat{\varepsilon}_i - \varepsilon_i\right|^2 + 2\left\|\partial P(\widetilde{\varepsilon}_i)\right\| \left|\lambda_K' P(\varepsilon_i)(\widehat{\varepsilon}_i - \varepsilon_i)\right| \qquad (2.8)
\end{aligned}$$

where $\widetilde{\varepsilon}_i$ is between $\widehat{\varepsilon}_i$ and ε_i for each $\lambda_K \in R^K$. By (2.5), Assumption 2.2.(v) and $\widetilde{\varepsilon}_i \in \mathcal{E}_\eta$ for all $i \leq n$ wpa1 (which is implied by (2.7)),

$$\frac{\max_{i\leq n}\left\|\partial P(\widetilde{\varepsilon}_i)\right\|^2}{n}\sum_{i=1}^{n}\left|\widehat{\varepsilon}_i - \varepsilon_i\right|^2 = O_p(\xi_{1,K}^2 \delta_{h,n}^{*2}). \qquad (2.9)$$

By the Cauchy-Schwarz inequality,

$$\begin{aligned}
&\sup_{\lambda_K \in B_K} \frac{\max_{i\leq n}\left\|\partial P(\widetilde{\varepsilon}_i)\right\|}{n}\sum_{i=1}^{n}\left|\lambda_K' P(\varepsilon_i)(\widehat{\varepsilon}_i - \varepsilon_i)\right| \\
&\leq \sup_{\lambda_K \in B_K}\max_{i\leq n}\left\|\partial P(\widetilde{\varepsilon}_i)\right\| \left(n^{-1}\sum_{i=1}^{n}\left|\widehat{\varepsilon}_i - \varepsilon_i\right|^2\right)^{1/2}\left(n^{-1}\sum_{i=1}^{n}\left|\lambda_K' P(\varepsilon_i)\right|^2\right)^{1/2} \\
&= O_p(\xi_{1,K}\delta_{h,n}^{*}), \qquad (2.10)
\end{aligned}$$

where the equality is by (2.9) and $\sup_{\lambda_K \in B_K} n^{-1}\sum_{i=1}^{n}\left|\lambda_K' P(\varepsilon_i)\right|^2 = O_p(1)$ which is implied by (2.1), $\xi_{0,K}(\log K)^{1/2}n^{-1/2} = o(1)$ and $\sup_{\lambda_K \in B_K}\mathbb{E}\left[\left|\lambda_K' P(\varepsilon)\right|^2\right] \leq \omega_{\max}(Q_K) \leq C$. By (2.8), (2.9) and (2.10),

$$\sup_{\lambda_K \in B_K}\left| n^{-1}\sum_{i=1}^{n}\left[\left|\lambda_K' P(\widehat{\varepsilon}_i)\right|^2\right] - n^{-1}\sum_{i=1}^{n}\left[\left|\lambda_K' P(\varepsilon_i)\right|^2\right]\right| = O_p(\xi_{1,K}^2\delta_{h,n}^{*2} + \xi_{1,K}\delta_{h,n}^{*}) \qquad (2.11)$$

which together with (2.1) proves the claim of the Lemma. ■

Lemma 2.2 *Suppose that Assumptions 2.1, 2.2, 2.4.(i)-(ii) and 2.4.(v) hold. Then we have*

$$\left\|\widehat{\beta}_n - \beta_{o,K}\right\| = O_p(K^{1/2}n^{-1/2} + K^{-\rho_g} + \upsilon_{1,K}\delta_{h,n}^{*}),$$

where $\upsilon_{1,K} = \sup_{\varepsilon \in \mathcal{E}_\eta}\left|\partial P(\varepsilon)'\beta_{o,K}\right|$.

Proof of Lemma 2.2. Let $G_n = [g_o(\varepsilon_1), \ldots, g_o(\varepsilon_n)]'$, $\widehat{G}_{K,n} = [g_{o,K}(\widehat{\varepsilon}_1), \ldots, g_{o,K}(\widehat{\varepsilon}_n)]'$ and $U_n = [u_1, \ldots, u_n]'$. By definition,

$$\widehat{\beta}_n = n^{-1}\widehat{Q}_{n,K}^{-1}\widehat{P}_n'(G_n + U_n) = \beta_{o,K} + n^{-1}\widehat{Q}_{n,K}^{-1}\widehat{P}_n'\left[(G_n - G_{n,K}) + (G_{n,K} - \widehat{G}_{n,K}) + U_n\right], \tag{2.12}$$

where $\widehat{Q}_{n,K} = n^{-1}\widehat{P}_n'\widehat{P}_n$ and $G_{n,K} = [g_{o,K}(\varepsilon_1), \ldots, g_{o,K}(\varepsilon_n)]'$. By Assumptions 2.4.(ii) and 2.4.(v), $\xi_{1,K}\delta_{h,n}^* = o(1)$ which together with Assumption 2.4.(i) and Lemma 2.1 implies that

$$(2C)^{-1} < \omega_{\min}(\widehat{Q}_{n,K}) \leq \omega_{\max}(\widehat{Q}_{n,K}) < 2C \text{ wpa1}. \tag{2.13}$$

By (2.13) and Assumption 2.2.(iii),

$$\begin{aligned}
&n^{-2}(G_n - G_{K,n})'\widehat{P}_n\widehat{Q}_{n,K}^{-2}\widehat{P}_n'(G_n - G_{K,n}) \\
&\leq \omega_{\min}^{-1}(\widehat{Q}_{n,K})n^{-2}(G_n - G_{K,n})'\widehat{P}_n\widehat{Q}_{n,K}^{-1}\widehat{P}_n'(G_n - G_{K,n}) \\
&= \omega_{\min}^{-1}(\widehat{Q}_{n,K})n^{-1}(G_n - G_{K,n})'\widehat{P}_n(\widehat{P}_n'\widehat{P}_n)^{-1}\widehat{P}_n'(G_n - G_{K,n}) \\
&\leq O_p(1)n^{-1}\sum_{i=1}^{n}\left[|g_o(\varepsilon_i) - g_{o,K}(\varepsilon_i)|^2\right] = O_p(K^{-2\rho_g}),
\end{aligned} \tag{2.14}$$

where the first equality is by the definition of $\widehat{Q}_{n,K}$, the second inequality is by the fact that $\widehat{P}_n(\widehat{P}_n'\widehat{P}_n)^{-1}\widehat{P}_n'$ is an idempotent matrix. Similarly

$$\begin{aligned}
&n^{-2}(G_{K,n} - \widehat{G}_{K,n})'\widehat{P}_n\widehat{Q}_{n,K}^{-2}\widehat{P}_n'(G_{K,n} - \widehat{G}_{K,n}) \\
&\leq O_p(1)n^{-1}(G_{K,n} - \widehat{G}_{K,n})'\widehat{P}_n(\widehat{P}_n'\widehat{P}_n)^{-1}\widehat{P}_n'(G_{K,n} - \widehat{G}_{K,n}) \\
&\leq O_p(1)n^{-1}\sum_{i=1}^{n}\left[|g_{o,K}(\varepsilon_i) - g_{o,K}(\widehat{\varepsilon}_i)|^2\right].
\end{aligned} \tag{2.15}$$

By the mean value expansion and the Cauchy-Schwarz inequality,

$$|g_{o,K}(\varepsilon_i) - g_{o,K}(\widehat{\varepsilon}_i)| = \left|\partial P(\widetilde{\varepsilon}_i)'\beta_{o,K}(\widehat{\varepsilon}_i - \varepsilon_i)\right| \leq \max_{i\leq n}\left|\partial P(\widetilde{\varepsilon}_i)'\beta_{o,K}\right| |\widehat{\varepsilon}_i - \varepsilon_i|, \tag{2.16}$$

where $\widetilde{\varepsilon}_i$ is between ε_i and $\widehat{\varepsilon}_i$. Using (2.16), we get

$$n^{-1}\sum_{i=1}^{n}\left[|g_{o,K}(\varepsilon_i) - g_{o,K}(\widehat{\varepsilon}_i)|^2\right] \leq \max_{i\leq n}\left|\partial P(\widetilde{\varepsilon}_i)'\beta_{o,K}\right|^2 n^{-1}\sum_{i=1}^{n}\left[|\widehat{\varepsilon}_i - \varepsilon_i|^2\right] = O_p(\upsilon_{1,K}^2\delta_{h,n}^{*2}), \tag{2.17}$$

where the equality is by (2.5) and $\max_{i\leq n}|\partial P(\widetilde{\varepsilon}_i)'\beta_{o,K}|^2 = O_p(\upsilon_{1,K}^2)$ which is implied by the definition of $\upsilon_{1,K}$ and $\widetilde{\varepsilon}_i \in \mathcal{E}_\eta$ for all $i \leq n$ wpa1 (which is implied by (2.7)). Combining the results in (2.15) and (2.17), we get

$$n^{-2}(G_{K,n} - \widehat{G}_{K,n})'\widehat{P}_n\widehat{Q}_{n,K}^{-2}\widehat{P}_n'(G_{K,n} - \widehat{G}_{K,n}) = O_p(\upsilon_{1,K}^2\delta_{h,n}^{*2}). \tag{2.18}$$

By Assumptions 2.1.(i) and 2.2.(i)

$$\begin{aligned}
&\mathbb{E}\left[n^{-2}U_n'\widehat{P}_n\widehat{Q}_{n,K}^{-1}\widehat{P}_n'U_n\right|\{x_i,s_i\}_{i=1}^n\Big] \\
&= tr\left(n^{-2}\widehat{P}_n\widehat{Q}_{n,K}^{-1}\widehat{P}_n'\mathbb{E}\left[U_nU_n'\right|x_i,s_i\}_{i=1}^n\Big]\right) \\
&\leq \frac{C}{n}tr\left(\widehat{Q}_{n,K}^{-1}\widehat{P}_n'\widehat{P}_n/n\right) = O(Kn^{-1})
\end{aligned} \tag{2.19}$$

which together with (2.13) and the Markov inequality implies that

$$n^{-2}U_n'\widehat{P}_n\widehat{Q}_{n,K}^{-2}\widehat{P}_n'U_n = O_p(Kn^{-1}). \tag{2.20}$$

Collecting the results in (2.12), (2.14), (2.18) and (2.20), we prove the claim of the Lemma. ■

Lemma 2.3 *Suppose that Assumptions 2.1, 2.2, 2.3 and 2.4 hold. Then Assumption 3.1 in HLR holds.*

Proof of Lemma 2.3. The the definition of $\|v_n^*\|_{sd}^2$, Assumptions 2.2.(i) and 2.3.(i),

$$\|v_n^*\|_{sd}^2 = \left\|v_{\Gamma_n}^*(x)\varepsilon\right\|_2^2 + \left\|v_{g_n}^*(\varepsilon)u\right\|_2^2 \geq \left\|v_{g_n}^*(\varepsilon)u\right\|_2^2 \geq C^{-1}\left\|v_{g_n}^*\right\|_2^2 \geq C^{-1} \tag{2.21}$$

for all n, which verifies Assumption 3.1.(i) in HLR. Assumption 3.1.(ii) in HLR is directly assumed in Assumption 2.3.(ii). By Lemma 2.2, we know that $\delta_{2,n}^* = n^{-1/2}K^{1/2} + K^{-\rho_g} + \upsilon_{1,K}\delta_{h,n}^*$, where $\upsilon_{1,K} = \sup_{\varepsilon\in\mathcal{E}_\eta}|\partial P(\varepsilon)'\beta_{o,K}|$. Let $g_n = g_{o,K}$, then by Assumption 2.2.(iii), we have $\|g_n - g_o\|_2 = O(K^{-\rho_g}) = O(\delta_{2,n}^*)$. By the definitions of $\|\cdot\|_\varphi$ and $\|\cdot\|_\psi$, we can set $c_\varphi = 1$ and $c_\psi = 1$ such that $\|v_h\|_\varphi \leq c_\varphi\|v_h\|_{\mathcal{H}}$ and $\|v_g\|_\psi \leq c_\psi\|v_g\|_{\mathcal{G}}$ for any $v_h \in \mathcal{V}_1$ and $v_g \in \mathcal{V}_2$. This verifies Assumption 3.1.(iii) in HLR. Assumption 3.1.(iv) in HLR is assumed in Assumptions 2.3.(iii). ■

Lemma 2.4 *Suppose that Assumptions 2.1, 2.2, 2.3 and 2.4 hold. Then Assumption 3.2 in HLR holds.*

Proof of Lemma 2.4. For ease of notation, we define $\varepsilon_h = s - h(x)$. By definition,

$$
\begin{aligned}
&\psi(Z_2, g^*, h) - \psi(Z_2, g, h) - \Delta_\psi(Z_2, g, h)[\pm\kappa_n u^*_{g_n}] \\
&= -\frac{1}{2}\left[\left|y - g(\varepsilon_h) \mp \kappa_n u^*_{g_n}(\varepsilon)\right|^2\right] + \frac{1}{2}\left[|y - g(\varepsilon_h)|^2\right] - [y - g(\varepsilon_h)]\,(\pm\kappa_n u^*_{g_n}) \\
&= -\frac{1}{2}\kappa_n^2 (u^*_{g_n}(\varepsilon))^2. \qquad (2.22)
\end{aligned}
$$

By Assumption 2.2.(i),

$$
\mathbb{E}\left[(u^*_{g_n}(\varepsilon))^2\right] = \frac{\mathbb{E}\left[|v^*_{g,n}(\varepsilon)|^2\right]}{\left\|v^*_{\Gamma_n}(x)\varepsilon\right\|_2^2 + \left\|v^*_{g_n}(\varepsilon)u\right\|_2^2} \leq \frac{\mathbb{E}\left[|v^*_{g,n}(\varepsilon)|^2\right]}{\left\|v^*_{\Gamma_n}(x)\varepsilon\right\|_2^2 + C^{-1}\left\|v^*_{g_n}(\varepsilon)\right\|_2^2} \leq C \qquad (2.23)
$$

which together with the Markov inequality, Assumption 2.1.(i) and (2.22) verifies Assumption 3.2.(i) in HLR.

By definition,

$$
\Delta_\psi(Z_2, g, h)[u^*_{g_n}] - \Delta_\psi(Z_2, g_o, h_o)[u^*_{g_n}] = \left(g_o(\varepsilon) - g(\varepsilon_h)\right) u^*_{g_n}. \qquad (2.24)
$$

Recall that $\mathcal{N}_{h,n} = \{h \in \mathcal{H}_n : \|h - h_o\|_2 \leq \delta_{1,n}\}$, where $\delta_{1,n} = \delta^*_{h,n}\log(\log(n))$. It is clear that for any $h(\cdot) = R(\cdot)'\gamma_L \in \mathcal{N}_{h,n}$, we have

$$
\begin{aligned}
\|h - h_o\|_\infty &\leq \|h - h_{o,L}\|_\infty + \|h_{o,L} - h_o\|_\infty \\
&\leq \left\|R(x)'(\gamma_L - \gamma_{o,L})\right\|_\infty + CL^{-\rho_h} \\
&\leq \zeta_L \|\gamma_L - \gamma_{o,L}\| + CL^{-\rho_h} \\
&\leq \zeta_L \omega_{\min}^{-1/2}(Q_L)((\gamma_L - \gamma_{o,L})'Q_L(\gamma_L - \gamma_{o,L}))^{1/2} + CL^{-\rho_h} \\
&= \zeta_L \omega_{\min}^{-1/2}(Q_L)\,\|h - h_{o,K}\|_2 + CL^{-\rho_h} \\
&\leq \zeta_L \omega_{\min}^{-1/2}(Q_L)\left[\|h - h_o\|_2 + \|h_{o,K} - h_o\|_2\right] + CL^{-\rho_h} \leq C\zeta_L\delta_{1,n} \qquad (2.25)
\end{aligned}
$$

where the last inequality is by Assumption 2.1.(iii)-(iv) and the definition of $\delta_{1,n}$. Define

$$
\mathcal{F}_n = \left\{f(s,x,h,g) : f(s,x,h,g) = \left(g_o(\varepsilon) - g(\varepsilon_h)\right) u^*_{g_n}(\varepsilon),\ g \in \mathcal{N}_{g,n},\ h \in \mathcal{N}_{h,n}\right\},
$$

where $\mathcal{N}_{g,n} = \{g \in \mathcal{G}_n : \|g - g_o\|_2 \leq \delta_{2,n}\}$ and $\delta_{2,n} = \delta^*_{2,n}\log(\log(n))$. By Assumptions 2.4.(i) and 2.4.(v), $\zeta_L\delta_{1,n} = o(1)$. Hence by (2.25) we can let n sufficiently large such that $\zeta_L\delta_{1,n} < \eta/2$ and $\varepsilon_h \in \mathcal{E}_\eta$ for any $h \in \mathcal{N}_{h,n}$. By the mean value expansion, $g(\varepsilon_h) - g(\varepsilon) = \partial P(\widetilde{\varepsilon}_h)'\beta(\varepsilon_h - \varepsilon)$ where $\widetilde{\varepsilon}_h$ is between ε_h and ε.

As $\varepsilon_h \in \mathcal{E}_\eta$ for any $h \in \mathcal{N}_{h,n}$, we have $\widetilde{\varepsilon}_h \in \mathcal{E}_\eta$. Hence for any $g(\cdot) = P(\cdot)'\beta$ with $g(\cdot) \in \mathcal{N}_{g,n}$ and any $h \in \mathcal{N}_{h,n}$, we have

$$\begin{aligned}
|g(\varepsilon_h) - g(\varepsilon)| &\leq \left|\partial P(\widetilde{\varepsilon}_h)'(\beta - \beta_{o,K})(\varepsilon_h - \varepsilon)\right| + \left|\partial P(\widetilde{\varepsilon}_h)'\beta_{o,K}(\varepsilon_h - \varepsilon)\right| \\
&= \left|\partial P(\widetilde{\varepsilon}_h)'(\beta - \beta_{o,K})(h(x) - h_o(x))\right| + \left|\partial P(\widetilde{\varepsilon}_h)'\beta_{o,K}(h(x) - h_o(x))\right| \\
&\leq \left[\|\partial P(\widetilde{\varepsilon}_h)\| \|\beta - \beta_{o,K}\| + \left|\partial P(\widetilde{\varepsilon}_h)'\beta_{o,K}\right|\right] \|h - h_o\|_\infty \\
&\leq (\xi_{1,K}\delta_{2,n} + \upsilon_{1,K})\zeta_L\delta_{1,n} \leq (\xi_{1,K}\delta_{1,n} + 1)\zeta_L\delta_{2,n} \leq C\zeta_L\delta_{2,n} \qquad (2.26)
\end{aligned}$$

where the first inequality is by the mean value expansion and the triangle inequality, the equality is by the definitions of ε_h and ε, the second inequality is by the Cauchy-Schwarz inequality, the third inequality is by Assumption 2.2.(v), (2.25), the definitions of $\upsilon_{1,K}$ and $\mathcal{N}_{h,n}$, and

$$\|\beta - \beta_{o,K}\| \leq \omega_{\min}^{-1}(Q_K)\left(\|g - g_o\|_2 + \|g_o - g_{o,K}\|_2\right) \leq C\delta_{2,n} \qquad (2.27)$$

which is implied by Assumption 2.2.(iii) and the definition of $\mathcal{N}_{g,n}$, the fourth inequality is because $\upsilon_{1,K}\delta_{1,n} \leq \delta_{2,n}$ by definition, the last inequality in (2.26) is by $\xi_{1,K}\delta_{1,n} = O(1)$ which is implied by Assumptions 2.4.(ii) and 2.4.(v). By the triangle inequality and the Cauchy-Schwarz inequality,

$$|g(\varepsilon) - g_o(\varepsilon)| \leq \|\beta - \beta_{o,K}\|\,\xi_{0,K} + \|g_o - g_{o,K}\|_\infty \leq C\xi_{0,K}\delta_{2,n} \qquad (2.28)$$

where the last inequality is by Assumption 2.2.(iii) and (2.27). By the definition of $u^*_{g_n}$, Assumptions 2.2.(iv)-(v) and (2.21),

$$\sup_{\varepsilon\in\mathcal{E}}\left|u^*_{g_n}(\varepsilon)\right|^2 \leq \frac{\xi_{0,K}^2\partial\rho(g_o)\left[P\right]' Q_K^{-2}\partial\rho(g_o)\left[P\right]}{C^{-1}\left\|v^*_{g_n}\right\|_2^2} = \frac{C\xi_{0,K}^2\partial\rho(g_o)\left[P\right]' Q_K^{-2}\partial\rho(g_o)\left[P\right]}{\partial\rho(g_o)[P]'Q_K^{-1}\partial\rho(g_o)\left[P\right]} \leq C\xi_{0,K}^2. \qquad (2.29)$$

Combining the results in (2.26), (2.28) and (2.29), we get

$$\begin{aligned}
\sup_{f\in\mathcal{F}_n}\|f\|_\infty &\leq \sup_{g\in\mathcal{N}_{g,n},\ h\in\mathcal{N}_{h,n},\ \varepsilon\in\mathcal{E}}\left[|g(\varepsilon_h) - g(\varepsilon)| + |g(\varepsilon) - g_o(\varepsilon)|\right]\sup_{\varepsilon\in\mathcal{E}}\left|u^*_{g_n}(\varepsilon)\right| \\
&\leq C(\zeta_L + \xi_{0,K})\xi_{0,K}\delta_{2,n} \equiv M_n. \qquad (2.30)
\end{aligned}$$

For any $f \in \mathcal{F}_n$, by (2.26) and (2.28),

$$
\begin{aligned}
\mathbb{E}\left[f^2\right] &\leq 2E\left[\left(g(\varepsilon_h) - g(\varepsilon)\right)^2 \left(u^*_{g_n}(\varepsilon)\right)^2\right] + 2E\left[\left(g(\varepsilon) - g_o(\varepsilon)\right)^2 \left(u^*_{g_n}(\varepsilon)\right)^2\right] \\
&\leq C(\zeta_L^2 + \xi_{0,K}^2)\delta_{2,n}^2 \mathbb{E}\left[\left(u^*_{g_n}(\varepsilon)\right)^2\right] \leq C(\zeta_L^2 + \xi_{0,K}^2)\delta_{2,n}^2 \equiv d_n^2 \qquad (2.31)
\end{aligned}
$$

where the last inequality is by (2.23). For any $f_1 = f(\cdot, h_1, g_1)$ and any $f_2 = f(\cdot, h_2, g_2)$ where $h_1, h_2 \in \mathcal{N}_{h,n}$ and $g_1, g_2 \in \mathcal{N}_{g,n}$, by the triangle inequality,

$$
\begin{aligned}
|f_1 - f_2| &\leq \left|\left(g_1(\varepsilon_{h_1}) - g_1(\varepsilon_{h_2})\right) u^*_{g_n}(\varepsilon)\right| + \left|\left(g_1(\varepsilon_{h_2}) - g_2(\varepsilon_{h_2})\right) u^*_{g_n}(\varepsilon)\right| \\
&\leq \left|\partial P(\widetilde{\varepsilon}_h)'\beta_1(\varepsilon_{h_1} - \varepsilon_{h_2}) u^*_{g_n}(\varepsilon)\right| + \xi_{0,K}\left|u^*_{g_n}(\varepsilon)\right| \|\beta_1 - \beta_2\| \\
&= \left|\partial P(\widetilde{\varepsilon}_h)'\left[(\beta_1 - \beta_{o,K}) + \beta_{o,K}\right](h_1(x) - h_2(x)) u^*_{g_n}(\varepsilon)\right| + \xi_{0,K}\left|u^*_{g_n}(\varepsilon)\right| \|\beta_1 - \beta_2\| \\
&\leq \left[\|\partial P(\widetilde{\varepsilon}_h)\| \|\beta_1 - \beta_{o,K}\| + \left|\partial P(\widetilde{\varepsilon}_h)'\beta_{o,K}\right|\right] \left|R(x)'(\gamma_1 - \gamma_2)\right| \left|u^*_{g_n}(\varepsilon)\right| + \xi_{0,K}\left|u^*_{g_n}(\varepsilon)\right| \|\beta_1 - \beta_2\| \\
&\leq \left[\xi_{1,K}\delta_{2,n} + \upsilon_{1,K}\right]\zeta_L \left|u^*_{g_n}(\varepsilon)\right| \|\gamma_1 - \gamma_2\| + \xi_{0,K}\left|u^*_{g_n}(\varepsilon)\right| \|\beta_1 - \beta_2\| \\
&\leq F_n(\varepsilon)(\|\beta_1 - \beta_2\| + \|\gamma_1 - \gamma_2\|), \qquad (2.32)
\end{aligned}
$$

where $F_n(\varepsilon) = C(\xi_{1,K}\zeta_L\delta_{2,n} + \upsilon_{1,K}\zeta_L + \xi_{0,K})\left|u^*_{g_n}(\varepsilon)\right|$, the equality is by the definitions of ε_{h_1} and ε_{h_2}, the fourth inequality is by $\|\partial P(\widetilde{\varepsilon}_h)\| \leq \xi_{1,K}$ and $\|\beta_1 - \beta_{o,K}\| \leq \delta_{2,n}$ for any $h_1, h_2 \in \mathcal{N}_{h,n}$, and

$$
\left|R(x)'(\gamma_1 - \gamma_2)\right| \leq \|R(x)\| \|\gamma_1 - \gamma_2\| \leq \zeta_L \|\gamma_1 - \gamma_2\| \qquad (2.33)
$$

which is implied by the triangle inequality and the definition of ζ_L. By (2.23), $\|F_n\|_2 \leq C(\xi_{1,K}\zeta_L\delta_{2,n} + \upsilon_{1,K}\zeta_L + \xi_{0,K}) \equiv \xi_{F_n}$. Let $H_{[]}\left(u, \mathcal{F}_n, \|\cdot\|_2\right)$ denote the u-bracketing entropy number of the function space $\mathcal{F}_n$ under the L_2-norm. By Example 19.7 in Van der Vaart (1998), $H_{[]}\left(u\|F_n\|_2, \mathcal{F}_n, \|\cdot\|_2\right) \leq (Cu^{-1})^{L+K}$ for all $u \in (0,1)$. Hence

$$
J_{[]}\left(d_n, \mathcal{F}_n, \|\cdot\|_2\right) = \int_0^{d_n} \left(\log H_{[]}\left(u, \mathcal{F}_n, \|\cdot\|_2\right)\right)^{1/2} du \leq C(K+L)^{1/2}(\log(n))^{1/2} d_n \qquad (2.34)
$$

where the inequality is by $d_n^{-1} \leq Cn$ and $\xi_{F_n} \leq Cn$ which are implied by Assumption 2.4. By (2.30), (2.31), (2.34) and Lemma 19.36 in Van der Vaart (1998),

$$\begin{aligned}
&\mathbb{E}\left[\sup_{\alpha\in\mathcal{N}_n}\left|\mu_n\left\{\Delta_\psi(Z_2,g,h)[u_{g_n}^*]-\Delta_\psi(Z_2,g_o,h_o)[u_{g_n}^*]\right\}\right|\right] \\
&\leq \frac{J_{[]}\left(d_n,\mathcal{F}_n,\|\cdot\|_2\right)}{n^{1/2}}\left(1+\frac{J_{[]}\left(d_n,\mathcal{F}_n,\|\cdot\|_2\right)}{d_n^2 n^{1/2}}M_n\right) \\
&\leq C\frac{(K+L)^{1/2}(\log(n))^{1/2}}{n^{1/2}}d_n\left(1+\frac{(K+L)^{1/2}(\log(n))^{1/2}}{d_n n^{1/2}}M_n\right) \\
&\leq C\frac{(K+L)^{1/2}(\log(n))^{1/2}d_n}{n^{1/2}}\left(1+\frac{(K+L)^{1/2}\xi_{0,K}(\log(n))^{1/2}}{n^{1/2}}\right)=o_p(1) \qquad (2.35)
\end{aligned}$$

where the equality is by Assumptions 2.4.(i), and 2.4.(v). Using (2.35) and the Markov inequality, we get

$$\sup_{\alpha\in\mathcal{N}_n}\left|\mu_n\left\{\Delta_\psi(Z_2,g,h)[u_{g_n}^*]-\Delta_\psi(Z_2,g_o,h_o)[u_{g_n}^*]\right\}\right|=o_p(n^{-1/2}), \qquad (2.36)$$

which verifies Assumption 3.2.(ii) in HLR.

By Assumption 2.2.(i), (2.23) and $\mathbb{E}\left[u|\,\varepsilon\right]=0$,

$$\begin{aligned}
K_\psi(g,h)-K_\psi(g^*,h) &= \mathbb{E}\left[-\frac{1}{2}\left|y-g(\varepsilon_h)\right|^2\right]-\mathbb{E}\left[-\frac{1}{2}\left|y-g^*(\varepsilon_h)\right|^2\right] \\
&= \mathbb{E}\left[-\frac{1}{2}\left|y-g(\varepsilon_h)\right|^2\right]-\mathbb{E}\left[-\frac{1}{2}\left|y-g(\varepsilon_h)\mp\kappa_n u_{g_n}^*(\varepsilon)\right|^2\right] \\
&= \mathbb{E}\left[\kappa_n^2\frac{(u_{g_n}^*(\varepsilon))^2}{2}\mp\kappa_n u_{g_n}^*(\varepsilon)u\pm\kappa_n u_{g_n}^*(\varepsilon)(g(\varepsilon_h)-g_o(\varepsilon))\right] \\
&= \pm\kappa_n\mathbb{E}\left[u_{g_n}^*(\varepsilon)(g(\varepsilon_h)-g_o(\varepsilon))\right]+O(\kappa_n^2). \qquad (2.37)
\end{aligned}$$

By the second order expansion, $g(\varepsilon_h)-g(\varepsilon)=\partial g(\varepsilon)(\varepsilon_h-\varepsilon)+\partial^2 g(\widetilde{\varepsilon}_h)(\varepsilon_h-\varepsilon)^2$, where $\widetilde{\varepsilon}_h\in\mathcal{E}_\eta$ for any $h\in\mathcal{N}_{h,n}$. For any $g\left(\cdot\right)=P\left(\cdot\right)'\beta\in\mathcal{N}_g$, we have

$$\|\beta\|\leq\|\beta-\beta_{o,K}\|+\|\beta_{o,K}\|\leq C\delta_{2,n}+\|\beta_{o,K}\|\leq C \qquad (2.38)$$

where the second inequality is (2.27), Assumptions 2.4.(i)-(ii) and 2.4.(v), the third inequality is by

$$\|\beta_{o,K}\|\leq\omega_{\min}^{-1}(Q_K)\left\|g_{o,K}\right\|_2\leq\omega_{\min}^{-1}(Q_K)\left[\left\|g_{o,K}-g_o\right\|_2+\|g_o\|\right]\leq C \qquad (2.39)$$

where the third inequality is by Assumptions 2.2.(ii)-(iv). Note that for any $g(\cdot) = P(\cdot)'\beta_K \in N_{g,n}$, we have $\|g - g_o\|_2 \leq \delta_{2,n}$, which together with Assumption 2.2.(iii) and the definition of $\delta_{2,n}$ implies that

$$\|g - g_{o,K}\|_2 \leq \|g - g_o\|_2 + \|g_{o,K} - g_o\|_2 \leq 2\delta_{2,n}. \tag{2.40}$$

By (2.40) and Assumption 2.2.(iv),

$$\|\beta - \beta_{o,K}\|^2 \leq \omega_{\min}^{-1}(Q_K)(\beta - \beta_{o,K})'Q_K(\beta - \beta_{o,K}) = \omega_{\min}^{-1}(Q_K)\,\|g - g_{o,K}\|_2^2 \leq C\delta_{2,n}. \tag{2.41}$$

By (2.38), $\left|\partial^2 g(\widetilde{\varepsilon}_h)\right| \leq C\xi_{2,K}$, which together with (2.23), (2.4), (2.6) and (2.41) implies that

$$\begin{aligned}
\mathbb{E}\left[\left|\partial^2 g(\widetilde{\varepsilon}_h)(\varepsilon_h - \varepsilon)^2 u_{g_n}^*(\varepsilon)\right|\right] &\leq \mathbb{E}\left[\left|\partial^2 P(\widetilde{\varepsilon}_h)'(\beta - \beta_{o,K})(\varepsilon_h - \varepsilon)^2 u_{g_n}^*(\varepsilon)\right|\right] \\
&\quad + \mathbb{E}\left[\left|\partial^2 P(\widetilde{\varepsilon}_h)'\beta_{o,K}(\varepsilon_h - \varepsilon)^2 u_{g_n}^*(\varepsilon)\right|\right] \\
&\leq \left(\xi_{2,K}\,\|\beta - \beta_{o,K}\| + \upsilon_{2,K}\right)\mathbb{E}\left[\left|(\varepsilon_h - \varepsilon)^2 u_{g_n}^*(\varepsilon)\right|\right] \\
&\leq \left(\xi_{2,K}\delta_{2,n} + \upsilon_{2,K}\right)\zeta_L\delta_{1,n}\mathbb{E}\left[\left|(\varepsilon_h - \varepsilon)u_{g_n}^*(\varepsilon)\right|\right] \\
&\leq \left(\xi_{2,K}\delta_{2,n} + \upsilon_{2,K}\right)\zeta_L\delta_{1,n}^2 = o(n^{-1/2})
\end{aligned} \tag{2.42}$$

for any $g \in \mathcal{N}_{g,n}$ and $h \in \mathcal{N}_{h,n}$, where the equality is by Assumptions 2.4.(iii)-(v). By (2.42),

$$\mathbb{E}\left[u_{g_n}^*(\varepsilon)(g(\varepsilon_h) - g(\varepsilon))\right] = \pm\kappa_n\mathbb{E}\left[u_{g_n}^*(\varepsilon)\partial g(\varepsilon)(\varepsilon_h - \varepsilon)\right] + o(n^{-1/2}). \tag{2.43}$$

By Jensen's inequality, the Holder inequality, (2.23), Assumptions 2.1.(iii), 2.2.(ii), 2.4.(v) and the definition of $h_{o,n}$,

$$\begin{aligned}
\left|\mathbb{E}\left[u_{g_n}^*(\varepsilon)\partial g(\varepsilon)(\varepsilon_{h_{o,n}} - \varepsilon)\right]\right| &= \left|\mathbb{E}\left[u_{g_n}^*(\varepsilon)\partial g(\varepsilon)(h_o - h_{o,n})\right]\right| \\
&\leq C(\mathbb{E}\left[(u_{g_n}^*(\varepsilon))^2\right]\mathbb{E}\left[(h_o - h_{o,n})^2\right])^{1/2} \\
&\leq C(\mathbb{E}\left[(h_o - h_{o,n})^2\right])^{1/2} = o(n^{-1/2}).
\end{aligned} \tag{2.44}$$

Combining the results in (2.37), (2.43) and (2.44), we get

$$K_\psi(g,h) - K_\psi(g^*,h) = \mp\kappa_n\Gamma(\alpha_o)\left[h - h_{o,n}, u_{g_n}^*\right] \pm \kappa_n\mathbb{E}\left[u_{g_n}^*(\varepsilon)(g(\varepsilon) - g_o(\varepsilon))\right] + o(n^{-1}). \tag{2.45}$$

By the definition of $\|\cdot\|_{\psi}$ and (2.23),

$$\frac{||g^* - g_o||_{\psi}^2 - ||g - g_o||_{\psi}^2}{2} = \pm\kappa_n \mathbb{E}\left[u_{g_n}^*(\varepsilon)(g(\varepsilon) - g_o(\varepsilon))\right] + o_p(n^{-1}) \tag{2.46}$$

which together with (2.45) verifies Assumption 3.2.(iii) in HLR. ■

Lemma 2.5 *Suppose that Assumptions 2.1, 2.2, 2.3 and 2.4 hold. Then Assumption 3.3 in HLR holds.*

Proof of Lemma 2.5. As the functional value $\rho(g_o)$ only depends on g_o, we know that $u_{h_n}^* = 0$. By Assumption 2.1.(i),

$$\mathbb{E}\left[(u_{\Gamma_n}^*(x))^2\right] \leq \frac{\left\|v_{\Gamma_n}^*(x)\right\|_2^2}{\left\|v_{\Gamma_n}^*(x)\right\|_2^2 C^{-1} + \left\|v_{g_n}^*(\varepsilon)u\right\|_2^2} \leq C, \tag{2.47}$$

which together with the Hölder inequality and Assumption 2.1.(iii) implies that

$$\left|\langle h_{o,L} - h_o, u_{\Gamma_n}^* \rangle_{\varphi}\right| \leq \|h_{o,L} - h_o\|_2 \left\|u_{\Gamma_n}^*\right\|_2 = O(L^{-\rho_h}). \tag{2.48}$$

By the definition of $\widehat{h}_n$,

$$\langle \widehat{h}_n - h_{o,L}, u_{\Gamma_n}^* \rangle_{\varphi} = \mathbb{E}\left[u_{\Gamma_n}^*(x)R(x)'\right]\left(R_n R_n'\right)^{-1} R_n(S_n - H_{n,L}), \tag{2.49}$$

where $H_{n,L} = [h_{o,L}(x_1), \ldots, h_{o,L}(x_n)]'$. By the Cauchy-Schwarz inequality and the Hölder inequality, we have

$$\left\|\mathbb{E}\left[u_{\Gamma_n}^*(x)R(x)\right]\right\|^2 \leq \mathbb{E}\left[(u_{\Gamma_n}^*(x))^2\right]\mathbb{E}\left[R(x)'R(x)\right] \leq CL \tag{2.50}$$

where the second inequality is by (2.47) and Assumption 2.1.(iv). Under Assumptions 2.1 and 2.4.(i), we can invoke Lemma 6.2 of Belloni, et al. (2015) to get

$$\|Q_L - Q_{n,L}\| = O_p(\zeta_L(\log L)^{1/2} n^{-1/2}), \tag{2.51}$$

where $Q_{n,L} = n^{-1}R_n R_n'$, which together with Assumption 2.4.(i) implies that

$$(2C)^{-1} < \omega_{\min}(Q_{n,L}) \leq \omega_{\max}(Q_{n,L}) < 2C \text{ wpa1}. \tag{2.52}$$

By the Cauchy-Schwarz inequality, (2.50), (2.52) and Assumption 2.1.(iii)

$$
\begin{aligned}
&\left|\mathbb{E}\left[u^*_{\Gamma_n}(x)R(x)'\right]\left(R_nR_n'\right)^{-1}R_n(H_n-H_{n,L})\right|^2 \\
&\leq \left\|\mathbb{E}\left[u^*_{\Gamma_n}(x)R(x)\right]\right\|^2(H_n-H_{n,L})'R_n'\left(R_nR_n'\right)^{-2}R_n(H_n-H_{n,L}) \\
&\leq O_p(Ln^{-1})(H_n-H_{n,L})'(H_n-H_{n,L})=O_p(L^{1-2\rho_h}),
\end{aligned} \tag{2.53}
$$

which together with $S_n - H_{n,L} = (H_n - H_{n,L}) + e_n$ (where $e_n = [\varepsilon_1, \ldots, \varepsilon_n]'$), (2.49) and Assumption 2.4.(v) implies that

$$
\langle \widehat{h}_n - h_{o,L}, u^*_{\Gamma_n}\rangle_{\varphi} = \mathbb{E}\left[u^*_{\Gamma_n}(x)R(x)'\right]\left(R_nR_n'\right)^{-1}R_ne_n + o_p(n^{-1/2}). \tag{2.54}
$$

By Assumptions 2.1.(i)-(ii) and 2.1.(iv), and (2.52),

$$
\begin{aligned}
\mathbb{E}\left[\left\|n^{-1}Q_L^{-1}R_ne_n\right\|^2\middle|\{x_i\}_{i=1}^n\right] &= \mathbb{E}\left[n^{-2}e_n'R_n'Q_L^{-2}R_ne_n\middle|\{x_i\}_{i=1}^n\right] \\
&\leq n^{-2}\omega_{\min}^{-2}(Q_L)tr\left(R_n'\mathbb{E}\left[e_ne_n'\middle|\{x_i\}_{i=1}^n\right]R_n\right) \\
&\leq Cn^{-2}\omega_{\min}^{-2}(Q_L)tr\left(R_nR_n'\right) \\
&\leq Cn^{-1}\omega_{\min}^{-2}(Q_L)tr\left(Q_{n,L}\right) = O_p(Ln^{-1})
\end{aligned} \tag{2.55}
$$

which together with the Markov inequality implies that

$$
\left\|n^{-1}Q_L^{-1}R_ne_n\right\| = O_p(L^{1/2}n^{-1/2}). \tag{2.56}
$$

By the Cauchy-Schwarz inequality,

$$
\begin{aligned}
&\left|\mathbb{E}\left[u^*_{\Gamma_n}(x)R(x)'\right]Q_{L,n}^{-1}\frac{R_ne_n}{n} - \mathbb{E}\left[u^*_{\Gamma_n}(x)R(x)'\right]n^{-1}Q_L^{-1}R_ne_n\right|^2 \\
&= \left|\mathbb{E}\left[u^*_{\Gamma_n}(x)R(x)'\right]Q_{L,n}^{-1}\left(Q_{L,n}-Q_L\right)n^{-1}Q_L^{-1}R_ne_n\right| \\
&\leq \left\|\mathbb{E}\left[u^*_{\Gamma_n}(x)R(x)'\right]Q_{L,n}^{-1}\right\|\left\|Q_{L,n}-Q_L\right\|\left\|n^{-1}Q_L^{-1}R_ne_n\right\| \\
&= O_p(\zeta_L(\log L)^{1/2}Ln^{-1}) = o_p(n^{-1/2})
\end{aligned} \tag{2.57}
$$

where the second equality is by (2.50), (2.51), (2.52) and (2.55), the last equality is by Assumption 2.4.(iv). Collecting the results in (2.54) and (2.57), we get

$$\langle \widehat{h}_n - h_{o,L}, u^*_{\Gamma_n} \rangle_\varphi = \mathbb{E}\left[u^*_{\Gamma_n}(x)R(x)'\right] Q_L^{-1} \frac{R_n e_n}{n} + o_p(n^{-1/2}). \tag{2.58}$$

By the definition of $u^*_{\Gamma_n}(x)$,

$$\mathbb{E}\left[u^*_{\Gamma_n}(x)R(x)'\right] Q_L^{-1} \frac{R_n e_n}{n} = \mathbb{E}\left[\partial g_o(\varepsilon) v^*_{g_n}(\varepsilon) R(x)'\right] Q_L^{-1} \frac{R_n e_n}{n \left\| v^*_n \right\|_{sd}}, \tag{2.59}$$

and moreover

$$\Delta_\varphi(Z_{1,i}, h_o)[u^*_{\Gamma_n}] = \mathbb{E}\left[\partial g_o(\varepsilon) v^*_{g_n}(\varepsilon) R(x)'\right] Q_L^{-1} \frac{R(x_i)\varepsilon_i}{\left\| v^*_n \right\|_{sd}}. \tag{2.60}$$

Hence we have

$$\mu_n \left\{ \Delta_\varphi(Z_1, h_o)[u^*_{\Gamma_n}] \right\} = \mathbb{E}\left[u^*_{\Gamma_n}(x)R(x)'\right] Q_L^{-1} \frac{R_n e_n}{n} \tag{2.61}$$

which together with (2.48), (2.58) and Assumption 2.4.(v) verifies Assumption 3.3.(i) in HLR.

By definition,

$$\Delta_\varphi(Z_1, h_o)[u^*_{\Gamma_n}] + \Delta_\psi(Z_2, g_o, h_o)[u^*_{g_n}] = \frac{v^*_{\Gamma_n}(x)\varepsilon + v^*_{g_n}(\varepsilon)u}{\left\| v^*_n \right\|_{sd}}. \tag{2.62}$$

By the Cauchy-Schwarz inequality, Assumptions 2.1.(iv)-(v), 2.2.(ii) and (2.21),

$$\begin{aligned}
\frac{\sup_{x\in\mathcal{X}} \left| v^*_{\Gamma_n}(x) \right|^2}{\left\| v^*_n \right\|_{sd}^2} &= \frac{\zeta_L^2}{\omega^2_{\min}(Q_L)} \frac{\left\| \mathbb{E}\left[\partial g_o(\varepsilon) v^*_{g_n}(\varepsilon) R(x)\right] \right\|^2}{\left\| v^*_n \right\|_{sd}^2} \\
&\leq \frac{C\zeta_L^2}{\omega^2_{\min}(Q_L)} \frac{\mathbb{E}\left[(\partial g_o(\varepsilon) v^*_{g_n}(\varepsilon))^2\right] \mathbb{E}\left[R(x)'R(x)\right]}{\left\| v^*_{g_n} \right\|_2^2} \\
&\leq \frac{C\zeta_L^2 \sup_{\varepsilon\in\mathcal{E}} (\partial g_o(\varepsilon))^2}{\omega^2_{\min}(Q_L)} \frac{\mathbb{E}\left[(v^*_{g_n}(\varepsilon))^2\right] \mathbb{E}\left[R(x)'R(x)\right]}{\left\| v^*_{g_n} \right\|_2^2} = O(L\zeta_L^2).
\end{aligned} \tag{2.63}$$

By Assumptions 2.1.(ii), 2.2.(i), 2.4.(i) and 2.4.(iv), and the results in (2.23), (2.29), (2.47) and (2.63),

$$
\begin{aligned}
\frac{\mathbb{E}\left[\left|v_{\Gamma_n}^*(x)\varepsilon+v_{g_n}^*(\varepsilon)u\right|^4\right]}{n\left\|v_n^*\right\|_{sd}^4} &\leq 8\frac{\mathbb{E}\left[\left|v_{\Gamma_n}^*(x)\varepsilon\right|^4\right]+\mathbb{E}\left[\left|v_{g_n}^*(\varepsilon)u\right|^4\right]}{n\left\|v_n^*\right\|_{sd}^4} \\
&\leq C\frac{\mathbb{E}\left[\left|v_{\Gamma_n}^*(x)\right|^4\right]+\mathbb{E}\left[\left|v_{g_n}^*(\varepsilon)\right|^4\right]}{n\left\|v_n^*\right\|_{sd}^4} \\
&\leq Cn^{-1}(\xi_{0,K}^2+L\zeta_L^2)\left(\mathbb{E}\left[\left|u_{\Gamma_n}^*(x)\right|^2\right]+\mathbb{E}\left[\left|u_{g_n}^*(\varepsilon)\right|^2\right]\right) \\
&= O(\xi_{0,K}^2 n^{-1}+L\zeta_L^2 n^{-1})=o(1), \qquad (2.64)
\end{aligned}
$$

which together with Assumption 2.1.(i) and the Linderberge CLT verifies Assumption 3.3.(ii) in HLR. The condition $\varepsilon_{2,n}=O(\kappa_n)$ and $\kappa_n\delta_{2,n}^{*-1}=o(1)$ in Assumption 3.3.(iii) of HLR hold by $\varepsilon_{2,n}=0$ and by $n^{-1/2}\delta_{2,n}^{*-1}=O(1)$ respectively. Moreover $||u_{g_n}^*||_\psi^2\leq C$ by the definition of $\|\cdot\|_\psi$ and (2.23). This verifies Assumption 3.3.(iii) in HLR. ■

Recall that $\mathcal{N}_{h,n}=\{h\in\mathcal{H}_n:\|h-h_o\|_2\leq\delta_{h,n}^*\log(\log(n))\}$ and $\mathcal{N}_n=\mathcal{N}_{h,n}\times\mathcal{N}_{g,n}$. In Section 4 of HLR, we define $\mathcal{W}_{1,n}=\{h\in\mathcal{V}_{1,n}:\|h\|_2\leq 1\}$ and $\mathcal{W}_{2,n}=\{g\in\mathcal{V}_{2,n}:\|g\|_2\leq 1\}$.

Lemma 2.6 *Suppose that Assumptions 2.1, 2.2, 2.3, 2.4 and 2.5 hold. Then Assumptions 4.1 and 4.2 in HLR hold.*

Proof of Lemma 2.6. Assumptions 4.1.(i) and 4.1.(ii) in HLR hold by the definition of $\langle\cdot,\cdot\rangle_\psi$. By the Cauchy-Schwarz inequality,

$$
\begin{aligned}
&\sup_{\alpha\in\mathcal{N}_n}\sup_{v_{g_1},v_{g_2}\in\mathcal{W}_{2,n}}\left|n^{-1}\sum_{i=1}^n r_\psi(Z_{2,i},\alpha)[v_{g_1},v_{g_2}]-\mathbb{E}\left[r_\psi(Z_2,\alpha_o)[v_{g_1},v_{g_2}]\right]\right| \\
&\quad=\sup_{v_{g_1},v_{g_2}\in\mathcal{W}_{2,n}}\left|n^{-1}\sum_{i=1}^n v_{g_1}(\varepsilon_i)v_{g_2}(\varepsilon_i)-\mathbb{E}\left[v_{g_1}(\varepsilon)v_{g_2}(\varepsilon)\right]\right| \\
&\quad\leq\|Q_{n,K}-Q_K\|=O_p(\xi_{0,K}(\log K)^{1/2}n^{-1/2})=o_p(1) \qquad (2.65)
\end{aligned}
$$

where the second equality is by (2.1), the third equality is by Assumption 2.4.(i). This means that Assumption 4.1.(iii) in HLR holds. Assumption 4.1.(iv) in HLR is assumed in Assumption 2.3.(iv). This verifies Assumption 4.1 in HLR.

Assumptions 4.2.(i) and 4.2.(ii) in HLR hold by the definition of $\langle\cdot,\cdot\rangle_{\varphi}$. By the Cauchy-Schwarz inequality,

$$
\begin{aligned}
&\sup_{h\in\mathcal{N}_{h,n}}\sup_{v_{h_1},v_{h_2}\in\mathcal{W}_{1,n}}\left|n^{-1}\sum_{i=1}^{n}r_{\varphi}(Z_{1,i},h)[v_{h_1},v_{h_2}]-\mathbb{E}\left[r_{\varphi}(Z_1,h_o)[v_{h_1},v_{h_2}]\right]\right|\\
&\quad=\sup_{v_{h_1},v_{h_2}\in\mathcal{W}_{1,n}}\left|n^{-1}\sum_{i=1}^{n}v_{h_1}(x_i)v_{h_2}(x_i)-\mathbb{E}\left[v_{h_1}(x)v_{h_2}(x)\right]\right|\\
&\quad\leq\|Q_{n,L}-Q_L\|=O_p(\zeta_L(\log L)^{1/2}n^{-1/2})=o_p(1) \qquad (2.66)
\end{aligned}
$$

where the second equality is by (2.51), and the third equality is by Assumption 2.4.(i). This means that Assumption 4.2.(iii) in HLR holds. As $\partial\rho(\alpha)[v_h]=0$ for any α in this example, Assumption 4.2.(iv) in HLR holds.

Under Assumptions 2.4.(v) and 2.5,

$$
\xi_{1,K}\delta_{2,n}\leq\xi_{1,K}(K^{1/2}n^{-1/2}+K^{-\rho_g}+\upsilon_{1,K}L^{1/2}n^{-1/2})\log(\log(n))=o(1). \qquad (2.67)
$$

By definition, for any $\alpha\in\mathcal{N}_n$, we have

$$
\begin{aligned}
&\Gamma_n(\alpha)\left[v_h,v_g\right]-\Gamma(\alpha_o)\left[v_h,v_g\right]\\
&=n^{-1}\sum_{i=1}^{n}\left[\partial g(\varepsilon_{h,i})-\partial g(\varepsilon_i)\right]v_h(x_i)v_g(\varepsilon_i)\\
&+n^{-1}\sum_{i=1}^{n}\left[\partial g(\varepsilon_i)-\partial g_o(\varepsilon_i)\right]v_h(x_i)v_g(\varepsilon_i)\\
&+n^{-1}\sum_{i=1}^{n}\partial g_o(\varepsilon_i)v_h(x_i)v_g(\varepsilon_i)-\mathbb{E}\left[\partial g_o(\varepsilon)v_h(x)v_g(\varepsilon)\right]. \qquad (2.68)
\end{aligned}
$$

By the Cauchy-Schwarz inequality,

$$
\begin{aligned}
&\sup_{v_h\in\mathcal{W}_{1,n},v_g\in\mathcal{W}_{2,n}}\left|n^{-1}\sum_{i=1}^{n}\left|v_h(x_i)v_g(\varepsilon_i)\right|\right|^2\\
&\leq\sup_{v_h\in\mathcal{W}_{1,n},v_g\in\mathcal{W}_{2,n}}\left[n^{-1}\sum_{i=1}^{n}\left|v_h(x_i)\right|^2\times n^{-1}\sum_{i=1}^{n}\left|v_g(\varepsilon_i)\right|^2\right]\\
&\leq\|Q_{L,n}\|\left\|Q_{K,n}\right\|=O_p(1) \qquad (2.69)
\end{aligned}
$$

where the equality is by Assumptions 2.1.(iv), 2.2.(iv), and results in (2.1) and (2.51). Recall that $\mathcal{B}^*_{2,n} \equiv \{v \in \mathcal{V}_{2,n} : \|v - v^*_{g_n}\|_\psi \|v^*_{g_n}\|_\psi^{-1} \leq \delta_{v_g,n}\}$, where $\delta_{v_g,n} = o(1)$ is some positive sequence such that $\widehat{v}^*_{g_n} \in \mathcal{B}^*_{2,n}$ wpa1. For any $v_g \in \mathcal{B}^*_{2,n}$, we have

$$\left| \|v_g\|_2 \|v^*_{g_n}\|_2^{-1} - 1 \right| \leq \|v_g - v^*_{g_n}\|_2 \|v^*_{g_n}\|_2^{-1} = o(1) \tag{2.70}$$

which implies that

$$\sup_{v_g \in \mathcal{B}^*_{2,n}} \|v_g\|_2 \|v^*_{g_n}\|_2^{-1} \leq 2 \tag{2.71}$$

for all large n. By (2.71), the mean value expansion, the triangle inequality and the Cauchy-Schwarz inequality,

$$\begin{aligned}
&\sup_{v_h \in \mathcal{W}_{1,n}, v_g \in \mathcal{B}^*_{2,n}} \left| n^{-1} \sum_{i=1}^n \left[\partial g(\varepsilon_{h,i}) - \partial g(\varepsilon_i)\right] v_h(x_i) v_g(\varepsilon_i) \right| \\
&\leq 2 \|v^*_{g_n}\|_2 \sup_{v_h \in \mathcal{W}_{1,n}, v_g \in \mathcal{W}_{2,n}} \left| n^{-1} \sum_{i=1}^n \partial P(\widetilde{\varepsilon}_{h,i})'(\beta - \beta_{o,K})(\varepsilon_{h,i} - \varepsilon_i) v_h(x_i) v_g(\varepsilon_i) \right| \\
&+ 2 \|v^*_{g_n}\|_2 \sup_{v_h \in \mathcal{W}_{1,n}, v_g \in \mathcal{W}_{2,n}} \left| n^{-1} \sum_{i=1}^n \partial P(\widetilde{\varepsilon}_{h,i})'\beta_{o,K}(\varepsilon_{h,i} - \varepsilon_i) v_h(x_i) v_g(\varepsilon_i) \right| \\
&\leq C \|v^*_{g_n}\|_2 \left[\xi_{1,K} \|\beta - \beta_{o,K}\| + \upsilon_{1,K}\right] \zeta_L \delta_{1,n} \left(\sup_{v_h \in \mathcal{W}_{1,n}, v_g \in \mathcal{W}_{2,n}} n^{-1} \sum_{i=1}^n |v_h(x_i) v_g(\varepsilon_i)| \right) \\
&\leq C \|v^*_{g_n}\|_2 \left[\xi_{1,K} \delta_{2,n} + \upsilon_{1,K}\right] \zeta_L \delta_{1,n} \left(\sup_{v_h \in \mathcal{W}_{1,n}, v_g \in \mathcal{W}_{2,n}} n^{-1} \sum_{i=1}^n |v_h(x_i) v_g(\varepsilon_i)| \right),
\end{aligned} \tag{2.72}$$

for any $h \in N_{h,n}$ and any $g \in N_{g,n}$, where the second inequality is by (2.25), the third inequality is by (2.41). Equation (2.72) together with Assumptions 2.5.(iii)-(v), (2.67), and (2.69) implies that

$$\begin{aligned}
&\sup_{\alpha \in \mathcal{N}_n} \sup_{v_h \in \mathcal{W}_{1,n}, v_g \in \mathcal{B}^*_{2,n}} \left| n^{-1} \sum_{i=1}^n \left[\partial g(\varepsilon_{h,i}) - \partial g(\varepsilon_i)\right] v_h(x_i) v_g(\varepsilon_i) \right| \\
&= O_p((\xi_{1,K}\delta_{2,n} + \upsilon_{1,K})\zeta_L \delta_{1,n}) \\
&= O_p(n^{-1/2} L^{1/2} \zeta_L \xi_{1,K} (n^{-1/2} K^{1/2} + K^{-\rho_g} + \upsilon_{1,K} n^{-1/2} L^{1/2}) + n^{-1/2} L^{1/2} \zeta_L \upsilon_{1,K}) = o_p(1).
\end{aligned} \tag{2.73}$$

By the triangle inequality, the Cauchy-Schwarz inequality, Assumption 2.2.(iii) and 2.2.(v)

$$\begin{aligned}\sup_{\varepsilon\in\mathcal{E}}\left|\partial g(\varepsilon)-\partial g_o(\varepsilon)\right| &\le \sup_{\varepsilon\in\mathcal{E}}\left|\partial g(\varepsilon)-\partial g_{o,K}(\varepsilon)\right|+\sup_{\varepsilon\in\mathcal{E}}\left|\partial g_{o,K}(\varepsilon)-\partial g_o(\varepsilon)\right| \\ &\le \xi_{1,K}\left\|\beta-\beta_{o,K}\right\|+K^{-\rho_g}, \end{aligned} \tag{2.74}$$

which together with the definition of $\mathcal{N}_n$ and (2.27) implies that

$$\sup_{g\in\mathcal{N}_{g,n}}\sup_{\varepsilon\in\mathcal{E}}\left|\partial g(\varepsilon)-\partial g_o(\varepsilon)\right|\le C\xi_{1,K}\delta_{2,n}+K^{-\rho_g}=o(1) \tag{2.75}$$

where the equality is by Assumption 2.4.(v) and (2.67). Using (2.75) and the triangle inequality

$$\begin{aligned}&\sup_{g\in\mathcal{N}_{g,n}}\sup_{v_h\in\mathcal{W}_{1,n},v_g\in\mathcal{B}^*_{2,n}}\left|n^{-1}\sum_{i=1}^{n}\left[\partial g(\varepsilon_i)-\partial g_o(\varepsilon_i)\right]v_h(x_i)v_g(\varepsilon_i)\right| \\ &\le \sup_{g\in\mathcal{N}_{g,n}}\sup_{\varepsilon\in\mathcal{E}}\left|\partial g(\varepsilon)-\partial g_o(\varepsilon)\right|\times\sup_{v_h\in\mathcal{W}_{1,n},v_g\in\mathcal{B}^*_{2,n}}n^{-1}\sum_{i=1}^{n}\left|v_h(x_i)v_g(\varepsilon_i)\right| \\ &\le C\left\|v^*_{g_n}\right\|_2\left(\xi_{1,K}\delta_{2,n}+K^{-\rho_g}\right)\left(\sup_{v_h\in\mathcal{W}_{1,n},v_g\in\mathcal{W}_{2,n}}n^{-1}\sum_{i=1}^{n}\left|v_h(x_i)v_g(\varepsilon_i)\right|\right)=o_p(1)\end{aligned} \tag{2.76}$$

where the equality is by Assumption 2.5.(i) and (2.69). By Assumptions 2.1.(i), 2.1.(v) 2.2.(ii), 2.2.(iv) and the Cauchy-Schwarz inequality,

$$\mathbb{E}\left[\left\|\mu_n\left\{\partial g_o(\varepsilon)R(x)P(\varepsilon)'\right\}\right\|^2\right]\le n^{-1}\mathbb{E}\left[\left|\partial g_o(\varepsilon)\right|^2\left|P(\varepsilon)'R(x)\right|^2\right]\le CK\zeta_L n^{-1}=o(1) \tag{2.77}$$

where the equality is by Assumption 2.4.(i). By the Cauchy-Schwarz inequality,

$$\begin{aligned}&\sup_{v_h\in\mathcal{W}_{1,n},v_g\in\mathcal{B}^*_{2,n}}\left|\mu_n\left\{\partial g_o(\varepsilon)v_h(x)v_g(\varepsilon)\right\}\right| \\ &\le 2\left\|v^*_{g_n}\right\|_2\sup_{v_h\in\mathcal{W}_{1,n},v_g\in\mathcal{W}_{2,n}}\left|\mu_n\left\{\partial g_o(\varepsilon)v_h(x)v_g(\varepsilon)\right\}\right| \\ &\le 2\left\|v^*_{g_n}\right\|_2\left\|\mu_n\left\{\partial g_o(\varepsilon)R(x)P(\varepsilon)'\right\}\right\|=o_p(1)\end{aligned} \tag{2.78}$$

where the equality is by Assumption 2.5.(i), (2.77) and the Markov inequality. Collecting the results in (2.68), (2.73), (2.76) and (2.78), we get

$$\sup_{\alpha\in\mathcal{N}_n}\sup_{v_h\in\mathcal{W}_{1,n},v_g\in\mathcal{B}^*_{2,n}}\left|\Gamma_n(\alpha)\left[v_h,v_g\right]-\Gamma(\alpha_o)\left[v_h,v_g\right]\right|=o_p(1). \tag{2.79}$$

By the Hölder inequality and Assumption 2.2.(ii)

$$\begin{aligned}
&\sup_{v_h \in \mathcal{W}_{1,n}, v_g \in \mathcal{B}^*_{2,n}} \left| \Gamma(\alpha_o) \left[v_h, v_g - v^*_{g_n} \right] \right| \\
&\leq C \left\| v^*_{g_n} \right\|_2 \sup_{v_h \in \mathcal{W}_{1,n}, v_g \in \mathcal{B}^*_{2,n}} \left[\left\| v_h \right\|_2 \left\| v_g - v^*_{g_n} \right\|_2 \left\| v^*_{g_n} \right\|_2^{-1} \right] \\
&\leq C \left\| v^*_{g_n} \right\|_2 \left\| Q_L \right\| \sup_{v_g \in \mathcal{B}^*_{2,n}} \left\| v_g - v^*_{g_n} \right\|_2 \left\| v^*_{g_n} \right\|_2^{-1} = o(1) \qquad (2.80)
\end{aligned}$$

where the equality is by Assumption 2.5.(i), Assumption 2.1.(iv) and the definition of $\mathcal{B}^*_{2,n}$. Combining the results in (2.79) and (2.80), we verify Assumption 4.2.(v) in HLR. ■

Lemma 2.7 *Suppose that Assumptions 2.1, 2.2, 2.3, 2.4 and 2.5 hold. Then Lemma C.3 in HLR holds..*

Proof of Lemma 2.7. By definition for any $h \in \mathcal{N}_{h,n}$,

$$\Delta^2_\varphi(Z_1, h)[v_h] = \varepsilon^2 v_h^2(x) + (h(x) - h_o(x))^2 v_h^2(x) - 2\varepsilon v_h^2(x)(h(x) - h_o(x)). \qquad (2.81)$$

By the definitions of $\mathcal{W}_{1,n}$ and the operator norm,

$$\sup_{v_h \in \mathcal{W}_{1,n}} \left| \mu_n \left\{ \varepsilon^2 v_h^2(x) \right\} \right| \leq \left\| \mu_n \left\{ \varepsilon^2 R(x) R(x)' \right\} \right\|. \qquad (2.82)$$

By Assumptions 2.1 and the Cauchy-Schwarz inequality,

$$\mathbb{E}\left[\left\| \mu_n \left\{ \varepsilon^2 R(x) R(x)' \right\} \right\|^2 \right] \leq n^{-1} \mathbb{E}\left[\varepsilon^4 \left| R(x)' R(x) \right|^2 \right] \leq L \zeta_L^2 n^{-1} \qquad (2.83)$$

which together with (2.82), the Markov inequality and Assumption 2.4.(i) implies that

$$\sup_{v_h \in \mathcal{W}_{1,n}} \left| \mu_n \left\{ \varepsilon^2 v_h^2(x) \right\} \right| = o_p(1). \qquad (2.84)$$

By the definition of $\mathcal{N}_{h,n}$,

$$\begin{aligned}
&\sup_{h \in \mathcal{N}_{h,n}} \sup_{v_h \in \mathcal{W}_{1,n}} n^{-1} \sum_{i=1}^n (h(x_i) - h_o(x_i))^2 v_h^2(x_i) \\
&\leq \left(\sup_{h \in \mathcal{N}_{h,n}} \left\| h - h_o \right\|_\infty^2 \right) \left(\sup_{v_h \in \mathcal{W}_{1,n}} n^{-1} \sum_{i=1}^n v_h^2(x_i) \right) = O_p(\zeta_L^2 \delta_{h,n}^2) = o_p(1) \qquad (2.85)
\end{aligned}$$

where $\delta^{*2}_{h,n} = Ln^{-1} + L^{-2\rho_h}$, the first equality is by (2.25) and

$$\sup_{v_h \in \mathcal{W}_{1,n}} n^{-1} \sum_{i=1}^{n} v_h^2(x_i) = O_p(1), \tag{2.86}$$

which follows by arguments in showing (2.69), the second equality is by Assumptions 2.4.(i) and 2.4.(v). By the Cauchy-Schwarz inequality,

$$\begin{aligned} &\sup_{h \in \mathcal{N}_{h,n}} \sup_{v_h \in \mathcal{W}_{1,n}} \left| n^{-1} \sum_{i=1}^{n} \varepsilon_i v_h^2(x_i)(h(x_i) - h_o(x_i)) \right|^2 \\ &\leq \left(\sup_{h \in \mathcal{N}_{h,n}} \|h - h_o\|_\infty^2 \right) \left(\sup_{v_h \in \mathcal{W}_{1,n}} n^{-1} \sum_{i=1}^{n} \varepsilon_i^2 v_h^2(x_i) \right) \left(\sup_{v_h \in \mathcal{W}_{1,n}} n^{-1} \sum_{i=1}^{n} v_h^2(x_i) \right). \end{aligned} \tag{2.87}$$

By Assumptions 2.1.(ii) and 2.1.(iv),

$$\left| \sup_{v_h \in \mathcal{W}_{1,n}} \mathbb{E}\left[\varepsilon_i^2 v_h^2(x_i)\right] \right| \leq \left\| \mathbb{E}\left[\varepsilon^2 R(x) R(x)'\right] \right\| \leq C \tag{2.88}$$

which together with (2.84) implies that

$$\sup_{v_h \in \mathcal{W}_{1,n}} n^{-1} \sum_{i=1}^{n} \varepsilon_i^2 v_h^2(x_i) = O_p(1). \tag{2.89}$$

By (2.86), (2.87), (2.89) and the definition of $\mathcal{N}_{h,n}$,

$$\sup_{h \in \mathcal{N}_{h,n}} \sup_{v_h \in \mathcal{W}_{1,n}} \left| n^{-1} \sum_{i=1}^{n} \varepsilon_i v_h^2(x_i)(h(x_i) - h_o(x_i)) \right|^2 = O_p(\zeta_L^2 \delta_{h,n}^2) = o_p(1) \tag{2.90}$$

where $\delta^{*2}_{h,n} = Ln^{-1} + L^{-2\rho_h}$, the second equality is by Assumptions 2.4.(i) and 2.4.(v). Collecting the results in (2.81), (2.84), (2.85) and (2.90), we show that Lemma C.3.(i) in HLR holds.

By definition

$$\begin{aligned} &n^{-1} \sum_{i=1}^{n} \Delta_\psi^2(Z_{2,i}, \alpha)[v_g] - \mathbb{E}\left[\Delta_\psi^2(Z_2, \alpha_o)[v_g]\right] \\ &= \mu_n \left\{ u^2 v_g^2(\varepsilon) \right\} + n^{-1} \sum_{i=1}^{n} (g(\varepsilon_{h,i}) - g_o(\varepsilon_i))^2 v_g^2(\varepsilon_i) \\ &- 2n^{-1} \sum_{i=1}^{n} u_i (g(\varepsilon_{h,i}) - g_o(\varepsilon_i)) v_g^2(\varepsilon_i). \end{aligned} \tag{2.91}$$

Using similar arguments in showing (2.84), we can show that

$$\sup_{v_g \in \mathcal{W}_{2,n}} \mu_n \left\{ u^2 v_g^2(\varepsilon) \right\} = O_p(K\xi_{0,K} n^{-1}) = o_p(1), \tag{2.92}$$

where the equality is by Assumption 2.4.(i). By (2.26) and (2.28),

$$\begin{aligned}
&\sup_{\alpha \in \mathcal{N}_n} \sup_{v_g \in \mathcal{W}_{2,n}} n^{-1} \sum_{i=1}^n (g(\varepsilon_{h,i}) - g_o(\varepsilon_i))^2 v_g^2(\varepsilon_i) \\
&\leq C(\zeta_L^2 + \xi_{0,K}^2)\delta_{2,n}^2 \sup_{v_g \in \mathcal{W}_{2,n}} n^{-1} \sum_{i=1}^n v_g^2(\varepsilon_i) \\
&= o(1) \sup_{v_g \in \mathcal{W}_{2,n}} n^{-1} \sum_{i=1}^n v_g^2(\varepsilon_i) = o_p(1) \qquad (2.93)
\end{aligned}$$

where the first equality is by

$$(\zeta_L^2 + \xi_{0,K}^2)\delta_{2,n}^2 = o(1), \tag{2.94}$$

which is implied by Assumption 2.4.(i), 2.4.(v) and $(\zeta_L^2 + \xi_{0,K}^2)v_{1,K}^2 L n^{-1} = o(1)$ (which is implied by Assumption 2.5), the second equality in (2.93) is by

$$\sup_{v_g \in \mathcal{W}_{2,n}} n^{-1} \sum_{i=1}^n v_g^2(\varepsilon_i) = O_p(1) \tag{2.95}$$

which follows by arguments in showing (2.69). Similarly by (2.26) and (2.28),

$$\begin{aligned}
&\sup_{\alpha \in \mathcal{N}_n} \sup_{v_g \in \mathcal{W}_{2,n}} \left| n^{-1} \sum_{i=1}^n u_i (g(\varepsilon_{h,i}) - g_o(\varepsilon_i)) v_g^2(\varepsilon_i) \right|^2 \\
&\leq C(\zeta_L^2 + \xi_{0,K}^2)\delta_{2,n}^2 \sup_{v_g \in \mathcal{W}_{2,n}} n^{-1} \sum_{i=1}^n u_i^2 v_g^2(\varepsilon_i) \sup_{v_g \in \mathcal{W}_{2,n}} n^{-1} \sum_{i=1}^n v_g^2(\varepsilon_i) \\
&= o_p(1) \sup_{v_g \in \mathcal{W}_{2,n}} n^{-1} \sum_{i=1}^n u_i^2 v_g^2(\varepsilon_i) = o_p(1) \qquad (2.96)
\end{aligned}$$

where the first equality is by (2.94) and (2.95), the second equality is by

$$\sup_{v_g \in \mathcal{W}_{2,n}} n^{-1} \sum_{i=1}^n u_i^2 v_g^2(\varepsilon_i) = O_p(1) \tag{2.97}$$

which follows by similar arguments in showing (2.89). Collecting the results in (2.91), (2.92), (2.93) and (2.96), we show that Lemma C.3.(ii) in HLR holds.

By definition

$$\begin{aligned}
&\Delta_\varphi(Z_1,h)[v_h]\Delta_\psi(Z_2,\alpha)[v_g]-\mathbb{E}_Z\left[\Delta_\varphi(Z_1,h_o)[v_h]\Delta_\psi(Z_2,\alpha_o)[v_g]\right]\\
&=u\varepsilon v_g(\varepsilon)v_h(x)-\mathbb{E}\left[u\varepsilon v_g(\varepsilon)v_h(x)\right]\\
&+(h(x)-h_o(x))uv_h(x)v_g(\varepsilon)+(g(\varepsilon_h)-g_o(\varepsilon))\varepsilon v_h(x)v_g(\varepsilon)\\
&+(g(\varepsilon_h)-g_o(\varepsilon))(h(x)-h_o(x))v_h(x)v_g(\varepsilon), \qquad (2.98)
\end{aligned}$$

for any $\alpha\in\mathcal{N}_n$. By the Cauchy-Schwarz inequality and Assumptions 2.1.(i)-(ii), 2.1.(v), 2.2.(i) and 2.2.(v),

$$\begin{aligned}
\mathbb{E}\left[\left\|\mu_n\left\{u\varepsilon R(x)P(\varepsilon)'\right\}\right\|^2\right]&=n^{-1}\mathbb{E}\left[u^2\varepsilon^2P(\varepsilon)'P(\varepsilon)R(x)'R(x)\right]\\
&\le n^{-1}\sqrt{\mathbb{E}\left[(u^2P(\varepsilon)'P(\varepsilon))^2\right]}\sqrt{\mathbb{E}\left[(\varepsilon^2R(x)'R(x))^2\right]}\\
&\le n^{-1}\sqrt{\xi_{0,K}^2\mathbb{E}\left[P(\varepsilon)'P(\varepsilon)\right]}\sqrt{\zeta_L^2\mathbb{E}\left[R(x)'R(x)\right]}\\
&\le Cn^{-1}\zeta_L\xi_{0,K}L^{1/2}K^{1/2}\le Cn^{-1}(L+K)(\zeta_L^2+K^2)=o(1), \qquad (2.99)
\end{aligned}$$

where the third inequality is by $\mathbb{E}\left[P(\varepsilon)'P(\varepsilon)\right]\le tr(Q_K)=O(K)$ and $\mathbb{E}\left[R(x)'R(x)\right]\le tr(Q_L)=O(L)$, and the last equality is by Assumption 2.4.(i). By the Cauchy-Schwarz inequality, the Markov inequality and (2.99), we have

$$\sup_{v_h\in\mathcal{W}_{1,n},v_g\in\mathcal{W}_{2,n}}\mu_n\left\{u\varepsilon v_g(\varepsilon)v_h(x)\right\}=o_p(1). \qquad (2.100)$$

By (2.26), (2.28) and the Cauchy-Schwarz inequality,

$$\begin{aligned}
&\sup_{\alpha\in\mathcal{N}_n}\sup_{v_h\in\mathcal{W}_{1,n},v_g\in\mathcal{W}_{2,n}}\left|n^{-1}\sum_{i=1}^n(g(\varepsilon_{h,i})-g_o(\varepsilon_i))\varepsilon_iv_h(x_i)v_g(\varepsilon_i)\right|^2\\
&\le C(\zeta_L^2+\xi_{0,K}^2)\delta_{2,n}^2\left(\sup_{v_h\in\mathcal{W}_{1,n}}n^{-1}\sum_{i=1}^n\varepsilon_i^2v_h^2(x_i)\right)\left(\sup_{v_g\in\mathcal{W}_{2,n}}n^{-1}\sum_{i=1}^nv_g^2(\varepsilon_i)\right)=o_p(1) \qquad (2.101)
\end{aligned}$$

where the equality is by (2.94), (2.89) and (2.95). By the Cauchy-Schwarz inequality,

$$\begin{aligned}
&\sup_{h\in\mathcal{N}_{h,n}}\sup_{v_h\in\mathcal{W}_{1,n},v_g\in\mathcal{W}_{2,n}}\left|n^{-1}\sum_{i=1}^n(h(x_i)-h_o(x_i))u_iv_h(x_i)v_g(\varepsilon_i)\right|^2\\
&\le\sup_{h\in\mathcal{N}_{h,n}}\|h-h_o\|_\infty^2\left(\sup_{v_h\in\mathcal{W}_{1,n}}n^{-1}\sum_{i=1}^nv_h^2(x_i)\right)\left(\sup_{v_g\in\mathcal{W}_{2,n}}n^{-1}\sum_{i=1}^nu_iv_g^2(\varepsilon_i)\right)=o_p(1) \qquad (2.102)
\end{aligned}$$

where the equality is by (2.86), (2.97) and $\sup_{h\in\mathcal{N}_{h,n}}\|h-h_o\|_\infty^2=\zeta_L^2\delta_{1,n}^2=o(1)$ which is implied by Assumption 2.4.(i). Similarly,

$$\begin{aligned}
&\sup_{\alpha\in\mathcal{N}_n}\sup_{v_h\in\mathcal{W}_{1,n},v_g\in\mathcal{W}_{2,n}}\left|n^{-1}\sum_{i=1}^{n}(h(x_i)-h_o(x_i))(g(\varepsilon_{h,i})-g_o(\varepsilon_i))v_h(x_i)v_g(\varepsilon_i)\right|^2\\
&\leq C(\zeta_L^2+\xi_{0,K}^2)\delta_{2,n}^2\sup_{h\in\mathcal{N}_{h,n}}\|h-h_o\|_\infty^2\\
&\times\left(\sup_{v_h\in\mathcal{W}_{1,n}}n^{-1}\sum_{i=1}^{n}v_h^2(x_i)\right)\left(\sup_{v_g\in\mathcal{W}_{2,n}}n^{-1}\sum_{i=1}^{n}v_g^2(\varepsilon_i)\right)\\
&=O_p((\zeta_L^2+\xi_{0,K}^2)\zeta_L^2\delta_{1,n}^2\delta_{2,n}^2)=o_p(1)
\end{aligned}\tag{2.103}$$

where the first equality is by (2.86), (2.95) and $\sup_{h\in\mathcal{N}_{h,n}}\|h-h_o\|_\infty^2=\zeta_L^2\delta_{1,n}^2$, the second equality is by (2.94), and $\zeta_L^2\delta_{1,n}^2=o(1)$ which is implied by Assumption 2.4.(i). Collecting the results in (2.98), (2.100), (2.101), (2.102) and (2.103), we show that Lemma C.3.(iii) in HLR holds. ■

Lemma 2.8 *Suppose that Assumptions 2.1, 2.2, 2.3 and 2.4 hold. Then Assumption 4.3.(iv) in HLR holds.*

Proof of Lemma 2.8. By the definition of $\Delta_\varphi(Z_1,h_o)[v_h]$, we have

$$\sup_{v_h\in\mathcal{W}_{1,n}}\mathbb{E}\left[|\Delta_\varphi(Z_1,h_o)[v_h]|^2\right]\leq\left\|\mathbb{E}\left[\varepsilon^2R(x)R(x)'\right]\right\|\leq C\|Q_L\|\leq C\tag{2.104}$$

where the second inequality is by Assumption 2.1.(ii), the third inequality is by Assumption 2.1.(iv). Similarly,

$$\sup_{v_g\in\mathcal{W}_{2,n}}\mathbb{E}\left[|\Delta_\psi(Z_2,\alpha_o)[v_g]|^2\right]\leq\left\|\mathbb{E}\left[u^2P(\varepsilon)P(\varepsilon)'\right]\right\|\leq C\|Q_K\|\leq C\tag{2.105}$$

where the second inequality is by Assumption 2.2.(i), the third inequality is by Assumption 2.2.(iv). By $v_{h_n}^*=0$, (2.21), (2.23) and (2.47),

$$(\left\|v_{h_n}^*\right\|_\varphi+\left\|v_{\Gamma_n}^*\right\|_\varphi+\left\|v_{g_n}^*\right\|_\psi)\left\|v_n^*\right\|_{sd}^{-1}=\left\|u_{\Gamma_n}^*\right\|_2+\left\|u_{g_n}^*\right\|_2\leq C,\tag{2.106}$$

which verifies (4.10) in HLR. By (2.104), (2.105) and (2.106), Assumption 4.3.(iv) in HLR is verified. ■

3 Verification of Assumptions 3.2 and 3.4 in Example 2.1

In this section, we use the nonparametric triangular simultaneous equation model in Newey, Powell and Vella (1999) to illustrate the high-level sufficient conditions for the asymptotic normality of the two-step sieve estimator.

The first step nonparametric estimation takes the following form:

$$\widehat{h}_n = \arg\max_{h\in\mathcal{H}_n} -\frac{1}{2n}\sum_{i=1}^{n}\left(x_i - h\left(w_{1,i}\right)\right)^2 \tag{3.1}$$

where $\mathcal{H}_n = \{h : h(\cdot) = R(\cdot)'\gamma,\, \gamma \in \mathbb{R}^{L(n)}\}$. Let $R(w_{1,i}) = \left[r_1(w_{1,i}), \ldots, r_{L(n)}(w_{1,i})\right]'$ for $i = 1, \ldots, n$, and $R_n = [R(w_{1,1}), \ldots, R(w_{1,n})]$. The first step M estimator $\widehat{h}_n$ has a closed form expression

$$\widehat{h}_n(\cdot) = R(\cdot)'\left(R_n R_n'\right)^{-1} R_n X_n = R(\cdot)'\widehat{\gamma}_n \tag{3.2}$$

where $X_n = [x_1, \ldots, x_n]'$. To define the second step M estimation, let $P(w) = \left[p_1(w), \ldots, p_{K(n)}(w)\right]'$ be a vector of approximating functions of $w = (x, w_2', u)'$ such that each $p_k(w)$ depends on (x, w_2) or on u, but not both. From the first step estimator, we calculate $\widehat{u}_i = x_i - \widehat{h}_n(w_{1,i})$ for $i = 1, \ldots, n$. Let $P(\widehat{w}_i) = [p_1(\widehat{w}_i), \ldots, p_{K(n)}(\widehat{w}_i)]'$ and $\widehat{P}_n = [\widehat{\tau}_1 P(\widehat{w}_1), \ldots, \widehat{\tau}_n P(\widehat{w}_n)]'$, where $\widehat{w}_i = (x_i, w_{2,i}', \widehat{u}_i)'$ and $\widehat{\tau}_i = \prod_{j=1}^{d_{w_2}+2} I\{a_j \leq \widehat{w}_{j,i} \leq b_j\}$ for $i = 1, \ldots, n$, where d_{w_2} denotes the dimension of w_2 and $\widehat{w}_{j,i}$ is the j-th component of $\widehat{w}_i$ for $j = 1, \ldots, d_{w_2}+2$. Let $g_o(w) = m_o(x, w_2) + \lambda_o(u)$ and $\eta = y - m_o(x, w_2) - \lambda_o(u)$. By the definition of $\lambda_o(u)$, and the conditional moment restrictions in (3) of HLR, we have

$$\mathbb{E}[\eta | x, w_1] = 0. \tag{3.3}$$

Let $\mathcal{T}_w = \{w : \tau(w) = 1\}$ where $\tau(w) = \prod_{j=1}^{d_{w_2}+2} I\{a_j \leq w_{j,i} \leq b_j\}$. The second step M estimator (of g_o) is

$$\widehat{g}_n = \arg\max_{g\in\mathcal{G}_n} -n^{-1}\sum_{i=1}^{n}\widehat{\tau}_i(y_i - g(\widehat{w}_i))^2 \tag{3.4}$$

where $\mathcal{G}_n = \{g(\cdot) : g(\cdot) = \tau(\cdot)P(\cdot)'\beta,\, \beta \in \mathbb{R}^{K(n)}\}$. The second step M estimator $\widehat{g}_n$ also has a closed form expression

$$\widehat{g}_n(w) = P(w)'(\widehat{P}_n'\widehat{P}_n)^{-1}\widehat{P}_n' Y_n = P(w)'\widehat{\beta}_n \tag{3.5}$$

for any $w \in \mathcal{T}_w$, where $Y_n = [y_1, \ldots, y_n]'$. The plug-in estimator of $\rho(g_o)$ is $\rho(\widehat{g}_n)$, where $\rho(\cdot)$ is a linear functional of g.

We next list the low level sufficient conditions for the asymptotic normality of $\rho(\widehat{g}_n)$. These assumptions are from Newey, Powell and Vella (1999).

Assumption 3.1 *$\{(y_i, x_i, w_{1,i})\}_{i=1}^n$ is i.i.d., $var(x|w_1)$ and $var(y|x, w_1)$ are bounded.*

Assumption 3.2 *w_1 is continuously distributed with density that is bounded away from zero on its support, and the support of w_1 is a cartesian product of compact, connected intervals. Also w is continuously distributed and its density is bounded away from zero on $\mathcal{T}_w$, and $\mathcal{T}_w$ is contained in the interior of the support of w.*

Assumption 3.3 *$h_o(w_1)$ is continuously differentiable of order s_1 on the support of w_1 and $m_o(x, w_2)$ and $\lambda_o(u)$ are Lipschitz and continuous differentiable of order s on $\mathcal{T}_w$.*

In the rest of the section, we write L and K for $L(n)$ and $K(n)$ respectively for notational simplicity. Following Newey, Powell and Vella (1999), we consider two types of approximating functions for $R(w_1)$ and $P(w)$: the power series and splines.

Assumption 3.4 *Either (a) for power series, $(K^3 + K^2L)(L^{1/2}n^{-1/2} + L^{-s_1/d_{w_1}}) = o(1)$; or (b) for splines, $(K^2 + KL)(L^{1/2}n^{-1/2} + L^{-s_1/d_{w_1}}) = o(1)$.*

By Assumption 3.3, there exists $\gamma_{o,L} \in \mathbb{R}^L$ such that

$$\sup_{w_1 \in \mathcal{W}_1} |h_{o,L}(w_1) - h_o(w_1)| \leq CL^{-s_1/d_{w_1}}, \tag{3.6}$$

where $h_{o,L}(w_1) = R(w_1)'\gamma_{o,L}$, $\mathcal{W}_1$ denotes the support of w_1 and d_{w_1} denotes the dimension of w_1, and there exists $\beta_{o,K} \in \mathbb{R}^K$ such that

$$\sup_{w \in \mathcal{T}_w} |g_{o,K}(w) - g_o(w)| \leq CK^{-s/d} \tag{3.7}$$

where $g_{o,K}(w) = P(w)'\beta_{o,K}$ and d denotes the dimension of $(x, w_2')'$.

We next calculate the Riesz representors $v^*_{g_n}$ and $v^*_{\Gamma_n}$. Let $Z_{1,i} = (x_i, w'_{1,i})'$ and $\varphi(Z_{1,i}, h) = -(x_i - h(w_{1,i}))^2/2$. By definition, $\langle v_{h_1}, v_{h_2}\rangle_\varphi = \mathbb{E}[v_{h_1}(w_1)v_{h_2}(w_1)]$ for any $v_{h_1}, v_{h_2} \in \mathcal{V}_1$. Let $Z_{2,i} = (y_i, x_i, w'_{1,i})'$, $u_{h,i} = x_i - h(w_{1,i})$ and $w_{h,i} = (x_i, w'_{2,i}, u_{h,i})'$. The criterion function of the second step estimation is

$$\psi(Z_{2,i}, g, h) = -\tau(w_{h,i})(y_i - m(x_i, w_{2,i}) - \lambda(x_i - h(w_{1,i})))^2/2.$$

By definition, $\langle v_{g_1}, v_{g_2}\rangle_\psi = \mathbb{E}\left[\tau(w)v_{g_1}(w)v_{g_2}(w)\right]$ for any $v_{g_1}, v_{g_2} \in \mathcal{V}_2$. By some simple calculation, we get

$$v^*_{g_n}(\cdot) = \tau(\cdot)P(\cdot)'Q_K^{-1}\rho(P_K),$$

where $Q_K = \mathbb{E}\left[\tau(w)P(w)P(w)'\right]$ and $\rho(P_K) = [\rho(p_1), \ldots, \rho(p_K)]'$. Moreover, by the conditional moment condition (3.3), we have

$$\Gamma(\alpha_o)\left[v_h, v_g\right] = \mathbb{E}\left[\tau(w)\partial_u g_o(w)v_h(w_1)v_g(w)\right]$$

where $\partial_u g_o(w) = \partial g_o(w)/\partial u$, which implies that

$$v^*_{\Gamma_n}(\cdot) = R(\cdot)'Q_L^{-1}\mathbb{E}\left[\tau(w)\partial_u g_o(w)v^*_{g_n}(w)\right] = R(\cdot)'Q_L^{-1}HQ_K^{-1}\rho(P_K),$$

where $H = \mathbb{E}\left[\tau(w)\partial_u g_o(w)R(w_1)P(w)'\right]$ and $Q_L = \mathbb{E}\left[R(w_1)R(w_1)'\right]$. Using the sieve Riesz representors $v^*_{g_n}$ and $v^*_{\Gamma_n}$, and the i.i.d. assumption, we have

$$\begin{aligned}\|v^*_n\|^2_{sd} &= \mathrm{Var}\left[n^{-\frac{1}{2}}\sum\nolimits_{i=1}^n \left(u_i v^*_{\Gamma_n}(w_{1,i}) + \eta_i\tau(w_i)v^*_{g_n}(w_i)\right)\right] \\ &= \mathbb{E}\left[u^2(v^*_{\Gamma_n}(w_1))^2\right] + \mathbb{E}\left[\eta^2\tau(w)(v^*_{g_n}(w))^2\right]\end{aligned} \tag{3.8}$$

where the second equality is by (3.3). Let $\Sigma_K = \mathbb{E}\left[\eta^2\tau(w)P(w)P(w)'\right]$ and $\Sigma_L = \mathbb{E}\left[u^2R(w_1)R(w_1)'\right]$. By the explicit expressions of $v^*_{g_n}$ and $v^*_{\Gamma_n}$,

$$\begin{aligned}\|v^*_n\|^2_{sd} &= \rho(P_K)'Q_K^{-1}H'Q_L^{-1}\mathbb{E}\left[u^2R(w_1)R(w_1)'\right]Q_L^{-1}HQ_K^{-1}\rho(P_K) \\ &\quad + \rho(P_K)'Q_K^{-1}\mathbb{E}\left[\eta^2\tau(w)P(w)P(w)'\right]Q_K^{-1}\rho(P_K) \\ &= \rho(P_K)'Q_K^{-1}\left[\Sigma_K + H'Q_L^{-1}\Sigma_L Q_L^{-1}H\right]Q_K^{-1}\rho(P_K)\end{aligned}$$

which is the same as the variance-covariance matrix V of the two-step estimator defined on page 596 of Newey, Powell and Vella (1999).

Assumption 3.5 *$\sigma^2(x, w_1) = var(y|x, w_1)$ is bounded away from zero, $\mathbb{E}[\eta^4|x, w_1]$ is bounded, and $\mathbb{E}[u^4|x, w_1]$ is bounded. Also $g_o(w)$ is twice continuously differentiable in u with bounded first and second derivatives.*

Assumption 3.6 *There exists $v^*_g(w)$ and $\beta_{v,K}$ such that $\mathbb{E}[\tau(w)\left|v^*_g(w)\right|^2] < \infty$, $\rho(g_o) = \mathbb{E}[\tau(w)v^*_g(w)g_o(w)]$, $\rho(p_k) = \mathbb{E}[\tau(w)v^*_g(w)p_k(w)]$ and $\mathbb{E}[\tau(w)\left|v^*_g(w) - P(w)'\beta_{v,K}\right|^2] \to 0$ as $K \to \infty$.*

For any $d_w \times 1$ vector a of nonnegative integers, let $|a| = \sum_{j=1}^{d_w} a_j$, $\partial^a g(w) = \partial^{|a|} g(w)/\partial w_1 \cdots \partial w_{d_w}$. Let $\xi_{\delta,K}$ ($\delta = 0,1$) and ζ_L be nondecreasing sequences such that $\max_{|a|\leq\delta} \sup_{w\in\mathcal{T}_w} \|\partial^a P(w)\| \leq \xi_{\delta,K}$ and $\sup_{w_1\in\mathcal{W}_1} \|R(w_1)\| \leq \zeta_L$ respectively. The following assumption is on the numbers of generic approximating functions in the first step and second step estimations.

Assumption 3.7 *$n^{1/2}K^{-s/d} = o(1)$ and $n^{1/2}L^{-s_1/d_{w_1}} = o(1)$, and*

$$\xi_{0,K}^2(L^2+K^2)\log(n)n^{-1} + \xi_{0,K}^2\zeta_L^2 L(\zeta_L^2 L + \xi_{0,K}^2 K)n^{-1} + \xi_{1,K}^2 LKn^{-1} = o(1). \tag{3.9}$$

When the power series are used in the two-step estimation, we have $\zeta_L \leq CL$ and $\xi_{\delta,K} \leq CK^{1+2\delta}$ ($\delta = 0,1$). Under the conditions that $n^{1/2}K^{-s/d} = o(1)$ and $n^{1/2}L^{-s_1/d_{w_1}} = o(1)$, the sufficient condition for (3.9) becomes

$$(K^7L + K^5L^3 + K^2L^6)n^{-1} = o(1)$$

which is implied by Assumption 8 in Newey, Powell and Vella (1999). When the splines are used in the two-step estimation, we have $\zeta_L \leq CL^{1/2}$ and $\xi_{\delta,K} \leq CK^{1/2+\delta}$ ($\delta = 0,1$). Under the conditions that $n^{1/2}K^{-s/d} = o(1)$ and $n^{1/2}L^{-s_1/d_{w_1}} = o(1)$, the sufficient condition for (3.9) becomes

$$(K^4L + K^3L^2 + KL^4)n^{-1} = o(1)$$

which is also implied by Assumption 8 in Newey, Powell and Vella (1999).

Theorem 3.1 *Under Assumptions 3.1-3.7, we have*

$$\frac{\sqrt{n}\left[\rho(\widehat{g}_n) - \rho(g_o)\right]}{\|v_n^*\|_{sd}} \to_d N(0,1). \tag{3.10}$$

Proof of Theorem 3.1. Define $\delta_{h,n} = \delta_{h,n}^*\varrho_n$ and $\delta_{g,n} = \delta_{g,n}^*\varrho_n$ where $\delta_{h,n}^* = L^{1/2}n^{-1/2} + L^{-s_1/d_{w_1}}$, $\delta_{g,n}^* = K^{1/2}n^{-1/2} + K^{-s/d} + \delta_{h,n}^*$ and $\{\varrho_n\}_n$ is a slowly divergent real positive sequence. Let $\mathcal{N}_{\gamma,n} = \{\gamma \in \mathbb{R}^L\colon ||\gamma - \gamma_{o,L}|| \leq \delta_{h,n}\}$ where $\delta_{h,n} = \delta_{h,n}^*\varrho_n$ and $\{\varrho_n\}_n$ is a slowly divergent real positive sequence. Similarly, define $\mathcal{N}_{\beta,n} = \{\beta \in \mathbb{R}^K\colon ||\beta - \beta_{o,K}|| \leq \delta_{g,n}\}$ where $\delta_{g,n} = \delta_{g,n}^*\varrho_n$. By Lemma 3.2.(b) and Lemma 3.2.(d), we have $\widehat{\gamma}_n \in \mathcal{N}_{\gamma,n}$ and $\widehat{\beta}_n \in \mathcal{N}_{\beta,n}$ wpa1. Define $\mathcal{N}_{h,n} = \{h(\cdot) = R(\cdot)'\gamma\colon \gamma \in \mathcal{N}_{\gamma,n}\}$ and $\mathcal{N}_{g,n} = \{g(\cdot) = P(\cdot)'\beta\colon \beta \in \mathcal{N}_{\gamma,n}\}$.[1] By Lemma 3.2.(b) and Lemma 3.2.(d), we have $\widehat{h}_n \in \mathcal{N}_{h,n}$ and

[1]Let $\|h\|_2 = (\mathbb{E}\left[h(w_1)^2\right])^{1/2}$ denote the L_2-norm and $\|g\|_{2,\tau} = (\mathbb{E}\left[\tau(w)g(w)^2\right])^{1/2}$ denote the restricted L_2-norm. One may also define the local neighborhoods of h_o and g_o as: $\mathcal{N}_{h,n}' = \{h(\cdot) = R(\cdot)'\gamma\colon \|h - h_o\|_2 \leq \delta_{h,L}\varrho_n'\}$ and $\mathcal{N}_{g,n}' = \{g(\cdot) = P(\cdot)'\beta\colon \|g - g_o\|_{2,\tau} \leq \delta_{g,L}\varrho_n'\}$ respectively, where $\{\varrho_n'\}_n$ is a slowly divergent real sequence. For any $h = R(\cdot)'\gamma_h \in \mathcal{N}_{h,n}'$,

$\widehat{g}_n \in \mathcal{N}_{g,n}$ wpa1. The proof of the theorem is divided into three steps.

Step 1. We verify Assumption 3.1 in HLR. By Assumptions 3.5 and 3.6, Lemma 3.1 implies that

$$\|v_n^*\|_{sd} \to \mathbb{E}\left[\eta^2\tau(w)(v_g^*(w))^2\right] + \mathbb{E}\left[u^2(v_\Gamma^*(w_1))^2\right] \tag{3.11}$$

as $K \to \infty$ and $L \to \infty$, where $v_\Gamma^*(w_1) = \mathbb{E}\left[\tau(w)v_g^*(w)\partial_u g_o(w)|w_1\right]$. The above limit is the same as the asymptotic variance defined in (5.7) of Newey, Powell and Vella (1999). By Assumption 3.5, $\mathbb{E}\left[\eta^2|x,w_1\right] > C_\eta$ where C_η is a finite positive constant. This means that

$$\mathbb{E}\left[\eta^2\tau(w)(v_g^*(w))^2\right] \geq C_\eta\mathbb{E}\left[\tau(w)(v_g^*(w))^2\right] > 0 \tag{3.12}$$

where the last inequality is by the fact that $\rho(g_o)$ is an unknown value. If $\mathbb{E}\left[\tau(w)(v_g^*(w))^2\right] = 0$, we have $\tau(w)(v_g^*(w))^2 = 0$ almost surely which together with (5.6) in Newey, Powell and Vella (1999) implies that $\rho(g) = 0$ for any $g \in \mathcal{G}$, where $\mathcal{G}$ includes all additive functions satisfying Assumptions 3.3 and 3.5. In such a case, $\rho(g_o)$ will be a known (to zero) value. Combining the results in (3.11) and (3.12) we have $\liminf_n \|v_n^*\|_{sd} > 0$, which verifies Assumption 3.1.(i). Because $\rho(\cdot)$ is a linear functional and $\|v_n^*\|_{sd}$ is bounded away from zero, Assumption 3.1.(ii) holds trivially. The strong norms $\|\cdot\|_{\mathcal{H}}$ and $\|\cdot\|_{\mathcal{G}}$ used to establish the convergence rate of $\widehat{h}_n$ and $\widehat{g}_n$ respectively are the L_2-norm $\|h\|_2 = (\mathbb{E}\left[(h(w_1))^2\right])^{1/2}$ and the restricted L_2-norm $\|g\|_{2,\tau} = (\mathbb{E}\left[\tau(w)(g(w))^2\right])^{1/2}$ respectively (see footnote 1 for details). By the definitions of $\|\cdot\|_\varphi$ and $\|\cdot\|_\psi$, we can set $c_\varphi = 1$ and $c_\psi = 1$ such that $\|v_h\|_\varphi \leq c_\varphi \|v_h\|_{\mathcal{H}}$ and $\|v_g\|_\psi \leq c_\psi \|v_g\|_{\mathcal{G}}$ for any $v_h \in \mathcal{V}_1$ and $v_g \in \mathcal{V}_2$. Under Assumptions 3.1-3.4, we can use Lemma 4.1 of Newey, Powell and Vella (1999) to get

$$\|\widehat{g}_n(w) - g_o\|_{\mathcal{G}} = \delta_{2,n}^* \tag{3.13}$$

where $\delta_{2,n}^* = K^{1/2}n^{-1/2} + K^{-s/d} + L^{1/2}n^{-1/2} + L^{-s_1/d_{w_1}}$. Let $g_n(\cdot) = g_{o,K}(\cdot)$ where $g_{o,K}$ is defined in (3.7). Then by (3.7) we have

$$\|g_n - g_o\|_{\mathcal{G}} = \|g_{o,K} - g_o\|_{\mathcal{G}} \leq \sup_{w\in\mathcal{T}_w} |g_{o,K}(w) - g_o(w)| = O(\delta_{2,n}^*), \tag{3.14}$$

by the triangle inequality,

$$\|h - h_{o,n}\| \leq \|h - h_o\| + \|h_{o,n} - h_o\| \leq 2\delta_{h,L}\varrho_n'$$

which implies that $||\gamma_h - \gamma_{o,L}|| \leq 2\omega_{\min}^{-1}(Q_L)\delta_{h,L}\varrho_n'$, where $\omega_{\min}(Q_L)$ denotes the smallest eigenvalue of Q_L which is bounded away from zero by Assumption 3.2. Hence if we let $\varrho_n = 2\omega_{\min}^{-1}(Q_L)\varrho_n'$, then $\gamma_h \in \mathcal{N}_{\gamma,n}$ which implies that $h \in \mathcal{N}_{h,n}$ and hence $\mathcal{N}_{h,n}' \subset \mathcal{N}_{h,n}$. Similarly, we can appropriately choose ϱ_n such that $\mathcal{N}_{g,n}' \subset \mathcal{N}_{g,n}$. This means the high-level sufficient conditions verified under $\mathcal{N}_{h,n}$ and/or $\mathcal{N}_{g,n}$ holds for their counterparts under $\mathcal{N}_{h,n}'$ and/or $\mathcal{N}_{g,n}'$.

which finishes verification of Assumption 3.1.(iii). For Assumption 3.1.(iv), as $\rho(g)$ is linear and it only depends on g, it is sufficient to show that

$$\frac{1}{\|v_n^*\|_{sd}} |\rho(g_{o,n} - g_o)| = o(n^{-\frac{1}{2}}) \tag{3.15}$$

where $g_{o,n}$ denotes the projection of g_o on the finite dimensional sieve space with respect to the restricted L_2-norm $\|\cdot\|_{2,\tau}$. By (5.6) in Newey, Powell and Vella (1999),

$$\begin{aligned}|\rho(g_{o,n} - g_o)|^2 &= \left|\mathbb{E}\left[\tau(w)v_g^*(w)(g_{o,n}(w) - g_o(w))\right]\right|^2 \\ &\leq \mathbb{E}\left[\tau(w)(v_g^*(w))^2\right]\mathbb{E}\left[\tau(w)(g_{o,n}(w) - g_o(w))^2\right] \\ &\leq \mathbb{E}\left[\tau(w)(v_g^*(w))^2\right]\mathbb{E}\left[\tau(w)(g_{o,K}(w) - g_o(w))^2\right] = O(K^{-2s/d}) \end{aligned} \tag{3.16}$$

where the first inequality is by Hölder's inequality, the second inequality is by the definition of $g_{o,n}$, the last equality is by (3.7) and Assumption 3.6. By Assumption 3.1.(i) (which has already been verified), (3.16) and Assumption 3.7, we prove (3.15) and hence Assumption 3.1.(iv).

Step 2. We verify Assumption 3.2 of HLR. Let $u_h = x - h(w_1)$ and $w_h = (x, w_2', u_h)'$. By definition

$$\begin{aligned}&\psi(Z_2, g^*, h) - \psi(Z_2, g, h) - \Delta_\psi(Z_2, g, h)[\pm\kappa_n u_{g_n}^*] \\ &= -\frac{\tau(w_h)(y - g(w_h) \mp \kappa_n u_{g_n}^*(w))^2}{2} \\ &+ \frac{\tau(w_h)(y - g(w_h))^2}{2} - \tau(w_h)(y - g(w_h))(\pm\kappa_n u_{g_n}^*) \\ &= -\frac{\kappa_n^2}{2}\tau(w_h)(u_{g_n}^*(w))^2, \end{aligned} \tag{3.17}$$

where $u_{g_n}^*(w) = v_{g_n}^*(w)/\|v_n^*\|_{sd}$ and $\|v_n^*\|_{sd}$ is defined in (3.8). By the triangle inequality, Lemma 3.3.(e)-(f) and (3.17),

$$\begin{aligned}&\sup_{h\in\mathcal{N}_{h,n}, g\in\mathcal{N}_{g,n}} \left|\mu_n\left\{\psi(Z_2, g^*, h) - \psi(Z_2, g, h) - \Delta_\psi(Z_2, g, h)[\pm\kappa_n u_{g_n}^*]\right\}\right| \\ &\leq \frac{\kappa_n^2}{2} n^{-1}\sum_{i=1}^n (u_{g_n}^*(w_i)^2 + \mathbb{E}[u_{g_n}^*(w)^2]) = O_p(\kappa_n^2) \end{aligned} \tag{3.18}$$

which verifies the first condition (12) of Assumption 3.2.(i) in HLR. Instead of verifying (13) of Assumption 3.2.(i) in HLR, we show that Assumption 3.4 holds. Assumption 3.4.(i) is implied by Assumption 3.1.

Let $\tau(Z_1,h)=\tau(w_h)$ and $\Delta^*_\psi(Z_2,g,h)[u^*_{g_n}]=(y-g(w_h))u^*_{g_n}(w)$. By definition,

$$\Delta_\psi(Z_2,g,h)[u^*_{g_n}]=\tau(w_h)(y-g(w_h))u^*_{g_n}(w)=\tau(Z_1,h)\Delta^*_\psi(Z_2,g,h)[u^*_{g_n}]. \tag{3.19}$$

Therefore equation (18) of HLR holds. By definition, $\tau(w_h)$ and $u^*_{g_n}(w)$ only depend on (x,w_1). By (3.3),

$$\mathbb{E}\left[\Delta^*_\psi(Z_2,g_o,h_o)[u^*_{g_n}]\middle|\,Z_1\right]=\mathbb{E}\left[(y-g_o(w))u^*_{g_n}(w)\middle|\,x,w_1\right]=u^*_{g_n}(w)\mathbb{E}\left[\eta|\,x,w_1\right]=0 \tag{3.20}$$

which verifies (19) of HLR. By (3.19) and (3.20) we show that Assumption 3.4.(ii) of HLR holds. By definition,

$$\Delta_\psi(Z_2,g,h)[u^*_{g_n}]-\Delta_\psi(Z_2,g_o,h)[u^*_{g_n}]=\tau(w_h)(g_o(w_h)-g(w_h))u^*_{g_n}(w), \tag{3.21}$$

and

$$\tau(Z_2,h)(\Delta^*_\psi(Z_2,g_o,h)[u^*_{g_n}]-\Delta_\psi(Z_2,g_o,h_o)[u^*_{g_n}])=\tau(w_h)(g_o(w)-g_o(w_h))u^*_{g_n}(w). \tag{3.22}$$

Hence Assumption 3.4.(iii) follows by Lemmas 3.6 and 3.7. By Assumption 3.5 and Lemma 3.3.(f) we have for any h

$$\begin{aligned}&(\tau(Z_1,h)-\tau(Z_1,h_o))^2\mathbb{E}\left[(\Delta^*_\psi(Z_2,g_o,h_o)[u^*_{g_n}])^2\middle|\,Z_1\right]\\&=(\tau(w_h)-\tau(w))^2(u^*_{g_n}(w))\mathbb{E}\left[\eta^2\middle|\,Z_1\right]\le C\xi^2_{0,K}(\tau(w_h)-\tau(w))^2,\end{aligned} \tag{3.23}$$

which together with Lemma 3.3.(d) implies that

$$\begin{aligned}&\sup_{h\in\mathcal{N}_{h,n}}n^{-1}\sum_{i=1}^n(\tau(Z_{1,i},h)-\tau(Z_{1,i},h_o))^2\mathbb{E}\left[(\Delta^*_\psi(Z_{2,i},g_o,h_o)[u^*_{g_n}])^2\middle|\,Z_{1,i}\right]\\&\le C\xi^2_{0,K}\sup_{h\in\mathcal{N}_{h,n}}n^{-1}\sum_{i=1}^n(\tau(Z_{1,i},h)-\tau(Z_{1,i},h_o))^2=O_p(\xi^2_{0,K}\zeta_L\delta_{h,n})\end{aligned} \tag{3.24}$$

where the $\xi^2_{0,K}\zeta_L\delta_{h,n}=o(1)$ by Assumption 3.7. This proves Assumption 3.4.(iv) and hence finishes verification of Assumption 3.4.

We next verify Assumption 3.2.(ii) of HLR. By definition,

$$\psi(Z_2,g,h)-\psi(Z_2,g^*,h)=\tau(w_h)(y-g(w_h))(\mp\kappa_n u^*_{g_n})+\frac{\kappa_n^2}{2}\tau(w_h)(u^*_{g_n}(w))^2, \tag{3.25}$$

which together with Lemma 3.3.(e) and the definition of $K_\psi(g,h)$ implies that

$$K_\psi(g,h) - K_\psi(g^*,h) = \mp\kappa_n \mathbb{E}\left[\tau(w_h)(y - g(w_h))u_{g_n}^*(w)\right] + O(\kappa_n^2). \tag{3.26}$$

By (3.3),

$$\mathbb{E}\left[(\tau(w_h) - \tau(w))(y - g_o(w))u_{g_n}^*(w)\right] = 0 \tag{3.27}$$

which implies that

$$\begin{aligned} &\mathbb{E}\left[\tau(w_h)(y - g(w_h))u_{g_n}^*(w)\right] \\ &= \mathbb{E}\left[\tau(w_h)(g_o(w) - g_o(w_h))u_{g_n}^*(w)\right] \\ &+ \mathbb{E}\left[\tau(w_h)(g_o(w_h) - g(w_h))u_{g_n}^*(w)\right]. \end{aligned} \tag{3.28}$$

Using the second order expansion in (3.111),

$$\begin{aligned} &\sup_{h\in\mathcal{N}_{h,n}} \left|\mathbb{E}\left[\tau(w_h)(g_o(w) - g_o(w_h) - \partial_u g_o(w)(h(w_1) - h_o(w_1)))u_{g_n}^*(w)\right]\right| \\ &\leq C\sup_w |u_{g_n}^*(w)| \sup_{h\in\mathcal{N}_{h,n}} \mathbb{E}\left[(h(w) - h_o(w))^2\right] \leq C\xi_{0,K}\delta_{h,n}^2 = o_p(n^{-1/2}), \end{aligned} \tag{3.29}$$

where the second inequality is by Lemma 3.3.(b) and 3.3.(f), the equality is by Assumption 3.7. By Assumption 3.5, (3.6), Lemma 3.2.(c) and 3.3.(g) and the definition of $\mathcal{N}_{h,n}$,

$$\begin{aligned} &\sup_{h\in\mathcal{N}_{h,n}} \left|\mathbb{E}\left[(\tau(w_h) - \tau(w))\partial_u g_o(w)(h(w_1) - h_o(w_1))u_{g_n}^*(w)\right]\right| \\ &\leq C\sup_w \left|u_{g_n}^*(w)\right| \left(\sup_{h\in\mathcal{N}_{h,n}} \sup_{w_1} |h(w_1) - h_o(w_1)|\right) \sup_{h\in\mathcal{N}_{h,n}} \mathbb{E}\left[|\tau(w_h) - \tau(w)|\right] \\ &\leq C\xi_{0,K}\zeta_L\delta_{h,n}\left(\zeta_L \sup_{\gamma\in\mathcal{N}_{\gamma,n}} \|\gamma - \gamma_{o,L}\| + CL^{-s_1/d_{w_1}}\right) \\ &\leq C\xi_{0,K}\zeta_L^2\delta_{h,n}^2 = o_p(n^{-1/2}), \end{aligned} \tag{3.30}$$

where the equality is by Assumption 3.7. By (3.29), (3.30) and the triangle inequality,

$$\mathbb{E}\left[\tau(w_h)(g_o(w) - g_o(w_h))u_{g_n}^*(w)\right] = \mathbb{E}\left[\tau(w)\partial_u g_o(w)(h(w_1) - h_o(w_1))u_{g_n}^*(w)\right] + o_p(n^{-1/2}), \tag{3.31}$$

uniformly over $(h, g) \in \mathcal{N}_n$. By (3.120) in the proof of Lemma 3.7,

$$
\begin{aligned}
&\sup_{h\in\mathcal{N}_{h,n}, g\in\mathcal{N}_{g,n}} \mathbb{E}\left[\left|\tau(w_h)(g_o(w_h) - g(w_h) - g_o(w) + g(w))u^*_{g_n}(w)\right|\right] \\
&\leq \xi_{1,K} \sup_{\beta\in\mathcal{N}_{\beta,n}} \|\beta - \beta_{o,K}\| \sup_{h\in\mathcal{N}_{h,n}} \mathbb{E}\left[\left|u^*_{g_n}(w)(h(w_1) - h_o(w_1))\right|\right] \\
&\leq \xi_{1,K}\delta_{g,n} \left\|u^*_{g_n}\right\|_2 \sup_{h\in\mathcal{N}_{h,n}} \|h - h_o\|_2 \\
&\leq \xi_{1,K}\delta_{g,n}\delta_{h,n} = o_p(n^{-1/2}), \qquad (3.32)
\end{aligned}
$$

where the second inequality is by Hölder's inequality and the definition of $\mathcal{N}_{\beta,n}$, the third inequality is by Lemma 3.3.(e) and the definition of $\mathcal{N}_{h,n}$, the equality is by Assumption 3.7. Similarly by (3.7), Lemma 3.3.(b), 3.3.(c) and 3.3.(g),

$$
\begin{aligned}
&\sup_{h\in\mathcal{N}_{h,n}, g\in\mathcal{N}_{g,n}} \mathbb{E}\left[\left|(\tau(w_h) - \tau(w))(g_o(w) - g(w))u^*_{g_n}(w)\right|\right] \\
&\leq \sup_w \left|u^*_{g_n}(w)\right| \sup_{g\in\mathcal{N}_{g,n}} \sup_w |g(w) - g_o(w)| \sup_{h\in\mathcal{N}_{h,n}} \mathbb{E}\left[|\tau(w_h) - \tau(w)|\right] \\
&\leq C\xi_{0,K}\zeta_L\delta_{h,n} \sup_{\beta\in\mathcal{N}_{\beta,n}} \left[\xi_{0,K}\|\beta - \beta_{o,K}\| + CK^{-s/d}\right] \\
&\leq C\xi^2_{0,K}\zeta_L\delta_{g,n}\delta_{h,n} = o_p(n^{-1/2}), \qquad (3.33)
\end{aligned}
$$

where the equality is by Assumption 3.7. By (3.32), (3.33) and the triangle inequality,

$$
\mathbb{E}\left[\tau(w_h)(g_o(w_h) - g(w_h))u^*_{g_n}(w)\right] = \mathbb{E}\left[\tau(w)(g_o(w) - g(w))u^*_{g_n}(w)\right] + o_p(n^{-1/2}), \qquad (3.34)
$$

uniformly over $(h, g) \in \mathcal{N}_n$. Collecting the results in (3.26), (3.28), (3.31) and (3.34), we deduce that

$$
\begin{aligned}
K_\psi(g, h) - K_\psi(g^*, h) &= \mathbb{E}\left[\tau(w_h)(y - g(w_h))(\mp\kappa_n u^*_{g_n}(w))\right] \\
&= \mathbb{E}\left[\tau(w)\partial_u g_o(w)(h(w_1) - h_o(w_1))(\mp\kappa_n u^*_{g_n}(w))\right] \\
&\quad + \mathbb{E}\left[\tau(w)(g_o(w) - g(w))(\mp\kappa_n u^*_{g_n}(w))\right] + o_p(n^{-1/2}) \qquad (3.35)
\end{aligned}
$$

uniformly over $(h, g) \in \mathcal{N}_n$. By definition,

$$
\Gamma(\alpha_o)\left[h - h_o, u^*_{g_n}\right] = \mathbb{E}\left[\tau(w)\partial_u g_o(w)(h(w_1) - h_o(w_1))u^*_{g_n}(w)\right], \qquad (3.36)
$$

for any $h \in \mathcal{N}_{h,n}$. By Jensen's inequality, (3.6), Assumptions 3.5, 3.7 and the definition of $h_{o,n}$,

$$\begin{aligned}\left|\Gamma(\alpha_o)\left[h_{o,n}-h_o,u_{g_n}^*\right]\right| &\leq \mathbb{E}\left[\left|\tau(w)\partial_u g_o(w)(h(w_1)-h_o(w_1))u_{g_n}^*(w)\right|\right]\\ &\leq C(\mathbb{E}[|h(w_1)-h_o(w_1)|^2])^{1/2}=o(n^{-1/2}).\end{aligned} \tag{3.37}$$

Moreover,

$$\begin{aligned}&\frac{||g^*-g_o||_\psi^2-||g-g_o||_\psi^2}{2}\\ &=\mathbb{E}\left[(g(w)-g_o(w))(\mp\kappa_n u_{g_n}^*(w))\right]+\frac{\kappa_n^2}{2}\mathbb{E}\left[u_{g_n}^*(w)^2\right]\\ &=\mathbb{E}\left[(g(w)-g_o(w))(\mp\kappa_n u_{g_n}^*(w))\right]+O_p(\kappa_n^2)\end{aligned} \tag{3.38}$$

uniformly over $g \in \mathcal{N}_{g,n}$, where the second equality is by Lemma 3.3.(e). Collecting the results in (3.35), (3.36), (3.37) and (3.38) proves Assumption 3.2(ii).

Step 3. We verify Assumption 3.3 of HLR. As $\rho(g)$ does not depend on h, we only need to show that

$$\left|\langle\widehat{h}_n-h_o,u_{\Gamma_n}^*\rangle_\varphi-\mu_n\left\{\Delta_\varphi(Z_1,h_o)[u_{\Gamma_n}^*]\right\}\right|=O_p(\kappa_n). \tag{3.39}$$

By definition

$$\langle\widehat{h}_n-h_o,u_{\Gamma_n}^*\rangle_\varphi=\langle\widehat{h}_n-h_{o,L},u_{\Gamma_n}^*\rangle_\varphi+\langle h_{o,L}-h_o,u_{\Gamma_n}^*\rangle_\varphi. \tag{3.40}$$

By Hölder's inequality, (3.6), Lemma 3.3.(h) and Assumption 3.7,

$$\left|\langle h_{o,L}-h_o,u_{\Gamma_n}^*\rangle_\varphi\right|\leq\left\|u_{\Gamma_n}^*\right\|\left\|h_{o,L}-h_o\right\|=o(n^{-1/2}). \tag{3.41}$$

By definition,

$$\widehat{h}_n(w_1)-h_{o,L}(w_1)=R(w_1)'\left(R_nR_n'\right)^{-1}R_n(H_n-H_{L,n})+R(w_1)'\left(R_nR_n'\right)^{-1}R_nU_n \tag{3.42}$$

where $H_n=(h_o(w_{1,1}),\ldots,h_o(w_{1,n}))'$, $H_{L,n}=(h_{o,L}(w_{1,1}),\ldots,h_{o,L}(w_{1,n}))'$ and $U_n=(u_1,\ldots,u_n)'$. By definition

$$\begin{aligned}\langle\widehat{h}_n-h_{o,L},u_{\Gamma_n}^*\rangle_\varphi&=\|v_n^*\|_{sd}^{-1}\,\rho(P_K)'Q_K^{-1}H'(R_nR_n')^{-1}R_nU_n\\ &\quad+\|v_n^*\|_{sd}^{-1}\,\rho(P_K)'Q_K^{-1}H'(R_nR_n')^{-1}R_n(H_n-H_{L,n}).\end{aligned} \tag{3.43}$$

By the Cauchy-Schwarz inequality,

$$
\begin{aligned}
&\left| \|v_n^*\|_{sd}^{-1} \rho(P_K)' Q_K^{-1} H' (R_n R_n')^{-1} R_n (H_n - H_{L,n}) \right|^2 \\
&= \left| n^{-1} \|v_n^*\|_{sd}^{-1} \rho(P_K)' Q_K^{-1} H' (\widehat{Q}_{n,L})^{-1} R_n (H_n - H_{L,n}) \right|^2 \\
&\leq \frac{\rho(P_K)' Q_K^{-1} H' (\widehat{Q}_{n,L})^{-1} H Q_K^{-1} \rho(P_K)}{\|v_n^*\|_{sd}} \\
&\times \frac{(H_n - H_{L,n})' R_n' (R_n R_n')^{-1} R_n (H_n - H_{L,n})}{n} \\
&\leq \omega_{\min}^{-1}(\widehat{Q}_{n,L}) \omega_{\max}(Q_L) \sup_{w_1} |h_o(w_1) - h_{o,L}(w_1)| \frac{\rho(P_K)' Q_K^{-1} H' Q_L^{-1} H Q_K^{-1} \rho(P_K)}{\|v_n^*\|_{sd}} \\
&\leq C \omega_{\min}^{-1}(\widehat{Q}_{n,L}) \omega_{\max}(Q_L) L^{-2s_1/d_{w_1}} \frac{\mathbb{E}\left[\left| v_{\Gamma_n}^*(w) \right|^2 \right]}{\|v_n^*\|_{sd}} = o_p(n^{-1})
\end{aligned}
\tag{3.44}
$$

where $\widehat{Q}_{n,L} = n^{-1} R_n R_n'$, the second inequality is by the fact that $R_n'(R_n R_n')^{-1} R_n$ is an idempotent matrix, the third inequality is by (3.6) and the definition of $v_{\Gamma_n}^*$, the last equality is by Lemma 3.2.(a), 3.3.(h) and Assumption 3.7. Hence we have

$$
\|v_n^*\|_{sd}^{-1} \rho(P_K)' Q_K^{-1} H' (R_n R_n')^{-1} R_n (H_n - H_{L,n}) = o_p(n^{-1/2}). \tag{3.45}
$$

By the i.i.d. assumption, Assumption 3.5 and Lemma 3.2.(a),

$$
\mathbb{E}\left[\left\| n^{-1} Q_L^{-1} R_n U_n \right\|^2 \middle| \{w_{1,i}\}_{i=1}^n \right] \leq \mathbb{E}\left[u^2 \middle| w_1 \right] n^{-1} tr(Q_L^{-1} \widehat{Q}_{n,L}) = O_p(n^{-1}), \tag{3.46}
$$

which together with the Markov inequality implies that

$$
\left\| n^{-1} Q_L^{-1} R_n U_n \right\| = O_p(n^{-1/2}). \tag{3.47}
$$

By the definition of $v_{\Gamma_n}^*$, Assumption 3.5, Lemma 3.2.(a) and 3.3.(h),

$$
\begin{aligned}
&\|v_n^*\|_{sd}^{-2} \rho(P_K)' Q_K^{-1} H' (\widehat{Q}_{n,L})^{-2} H Q_K^{-1} \rho(P_K) \\
&\leq \|v_n^*\|_{sd}^{-2} \omega_{\min}^{-2}(\widehat{Q}_{n,L}) \omega_{\max}(Q_L) \rho(P_K)' Q_K^{-1} H' Q_L^{-1} H Q_K^{-1} \rho(P_K) \\
&\leq \|v_n^*\|_{sd}^{-2} \omega_{\min}^{-2}(\widehat{Q}_{n,L}) \omega_{\max}(Q_L) \mathbb{E}\left[(v_{\Gamma_n}^*(w_1))^2 \right] = O_p(1).
\end{aligned}
\tag{3.48}
$$

By Lemma 3.2.(a), (3.47), (3.48) and the Cauchy-Schwarz inequality

$$
\begin{aligned}
&\left|(n\left\|v_n^*\right\|_{sd})^{-1}\rho(P_K)'Q_K^{-1}H'((\widehat{Q}_{n,L})^{-1}-Q_L^{-1})R_nU_n\right| \\
&=\left|(n\left\|v_n^*\right\|_{sd})^{-1}\rho(P_K)'Q_K^{-1}H'(\widehat{Q}_{n,L})^{-1}(\widehat{Q}_{n,L}-Q_L)Q_L^{-1}R_nU_n\right| \\
&\leq\left\|\left\|v_n^*\right\|_{sd}^{-1}\rho(P_K)'Q_K^{-1}H'(\widehat{Q}_{n,L})^{-1}\right\|\left\|\widehat{Q}_{n,L}-Q_L\right\|\left\|n^{-1}Q_L^{-1}R_nU_n\right\| \\
&=O_p(\zeta_L L^{1/2}n^{-1})=o_p(n^{-1/2})
\end{aligned}
\tag{3.49}
$$

where the last equality is by Assumption 3.7. By (3.40), (3.41), (3.43), (3.44) and (3.49),

$$
\begin{aligned}
\langle\widehat{h}_n-h_o,u_{\Gamma_n}^*\rangle_\varphi &=(n\left\|v_n^*\right\|_{sd})^{-1}\rho(P_K)'Q_K^{-1}H'Q_KQ_L^{-1}R_nU_n+o_p(n^{-1/2}) \\
&=(n\left\|v_n^*\right\|_{sd})^{-1}\sum_{i=1}^n v_{\Gamma_n}^*(w_{1,i})u_i+o_p(n^{-1/2}) \\
&=\mu_n\left\{\Delta_\varphi(Z_1,h_o)[u_{\Gamma_n}^*]\right\}+o_p(n^{-1/2})
\end{aligned}
\tag{3.50}
$$

where the second equality is by the definition of $v_{\Gamma_n}^*$, and the third equality is by the definition of $\Delta_\varphi(Z_1,h_o)[u_{\Gamma_n}^*]$. This verifies Assumption 3.3.(i) in HLR. To verify Assumption 3.3.(ii) in HLR, we notice that

$$
\begin{aligned}
&n^{-\frac{1}{2}}\sum_{i=1}^n\left\{\Delta_\varphi(Z_{1,i},h_o)[u_{\Gamma_n}^*]+\Delta_\psi(Z_{2,i},g_o,h_o)[u_{g_n}^*]\right\} \\
&=n^{-\frac{1}{2}}\sum_{i=1}^n\left\{u_{\Gamma_n}^*(w_{1,i})u_i+u_{g_n}^*(w_i)\eta_i\right\}.
\end{aligned}
\tag{3.51}
$$

To show the asymptotic normality of the above partial sum, we apply the Lindbergh-Feller CLT. By the Cauchy-Schwarz inequality, Assumption 3.5, Lemma 3.3.(h)

$$
\begin{aligned}
\frac{\sup_{w_1\in\mathcal{W}_1}\left|v_{\Gamma_n}^*(w_1)\right|^2}{\left\|v_n^*\right\|_{sd}^2} &\leq\zeta_L^2\left\|v_n^*\right\|_{sd}^{-2}\left\|Q_L^{-1}HQ_K^{-1}\rho(P_K)\right\|^2 \\
&\leq\omega_{\min}^{-1}(Q_L)\zeta_L^2\left\|v_n^*\right\|_{sd}^{-2}\rho(P_K)'Q_K^{-1}H'Q_L^{-1}HQ_K^{-1}\rho(P_K) \\
&=\frac{C\zeta_L^2}{\omega_{\min}(Q_L)}\frac{\mathbb{E}\left[(v_{\Gamma_n}^*(w_1))^2\right]}{\left\|v_{g_n}^*\right\|_2^2}=O(\zeta_L^2),
\end{aligned}
\tag{3.52}
$$

where the first equality is by the definition of $v^*_{\Gamma_n}$. By Assumption 3.5, (3.52), Lemma 3.3.(f)-3.3.(h),

$$\begin{aligned}
&\frac{\mathbb{E}\left[(v^*_{\Gamma_n}(w_1)u + v^*_{g_n}(w)\eta)^4\right]}{n\left\|v^*_n\right\|^4_{sd}} \\
&\leq 8\frac{\mathbb{E}\left[(v^*_{\Gamma_n}(w_1)u)^4\right] + \mathbb{E}\left[(v^*_{g_n}(w)\eta)^4\right]}{n\left\|v^*_n\right\|^4_{sd}} \\
&\leq 8C\frac{\mathbb{E}\left[(v^*_{\Gamma_n}(w_1))^4\right] + \mathbb{E}\left[(v^*_{g_n}(w))^4\right]}{n\left\|v^*_n\right\|^4_{sd}} \\
&\leq 8C\frac{\sup_{w_1}(v^*_{\Gamma_n}(w_1))^2 + \sup_{w}(v^*_{g_n}(w))^2}{n\left\|v^*_n\right\|^2_{sd}}\frac{\mathbb{E}\left[(v^*_{\Gamma_n}(w_1))^2\right] + \mathbb{E}\left[(v^*_{g_n}(w))^2\right]}{\left\|v^*_n\right\|^2_{sd}} \\
&= O((\zeta_L^2 + \xi_{0,K}^2)n^{-1}) = o(1) \qquad (3.53)
\end{aligned}$$

where the equality is by Assumption 3.7. This verifies the Lindbergh's condition. Hence Assumption 3.3.(ii) in HLR follows by the i.i.d. assumption and the Lindbergh-Feller CLT. Finally, we verify Assumption 3.3.(iii) in HLR. First, we have $\varepsilon_{2,n} = 0$ because the estimators in both the first step and the second step have closed form expressions. By definition, $\delta^*_{2,n} = K^{1/2}n^{-1/2} + K^{-s/d} + L^{1/2}n^{-1/2} + L^{-s_1/d_{w_1}}$ which together with $K \to \infty$ and $L \to \infty$ implies that $n^{1/2}(\delta^*_{2,n})^{-1} = o(1)$. Moreover by Lemma 3.3.(e), $||u^*_{g_n}||_\psi = (\mathbb{E}\left[(v^*_{g_n}(w))^2\right])^{1/2}\left\|v^*_n\right\|^{-1}_{sd} = O(1)$ which finishes verification of Assumption 3.3.(iii) in HLR. ■

Corollary 3.2 *Under Assumptions 3.1-3.7, Assumptions 1.4-1.6 hold.*

Proof of Lemma 3.1. We first verify Assumption 1.4. By definition,

$$_{z_1\in\mathcal{Z}_1, h\in\mathcal{N}_{h,n}}\left[|\tau(z_1,h)| + |\tau(z_1,h_o)|\right] \leq 2 \qquad (3.54)$$

which shows that Assumption 1.4.(i) holds. By definition, $\psi^*(z_2,\alpha) = -\frac{1}{2}(y - g(w_h))^2$, which implies that

$$\Delta^*_\psi(z_2,\alpha)[v_{g,1}] = (y - g(w_h))v_{g,1} \qquad (3.55)$$

and

$$r^*_{\psi,g}(z_2,\alpha)[v_{g,1}, v_{g,2}] = -v_{g,1}v_{g,2} \qquad (3.56)$$

for any $v_{g,1}$, $v_{g,2} \in \mathcal{V}_2$, which implies that Assumption 1.1.(i) holds. Moreover

$$\begin{aligned}
&\psi(z_2, g^*, h) - \psi(z_2, \alpha) - \Delta_\psi(z_2, \alpha)[\pm\kappa_n u^*_{g_n}] - \kappa_n^2 r_{\psi,g}(z_2, \alpha)[u^*_{g_n}, u^*_{g_n}] \\
&= \tau(z_1, h) \left[\begin{array}{c} \psi^*(z_2, g^*, h) - \psi^*(z_2, \alpha) \\ -\Delta^*_\psi(z_2, \alpha)[\pm\kappa_n u^*_{g_n}] - \kappa_n^2 r^*_{\psi,g}(z_2, \alpha)[u^*_{g_n}, u^*_{g_n}] \end{array} \right] = 0, \quad (3.57)
\end{aligned}$$

for any $\alpha \in \mathcal{N}_n$ and any $z_2 \in \mathcal{Z}_2$. This means that Assumption 1.1.(ii) holds for $\psi(z_2, \alpha)$ with $\Lambda_{1,n}(z_2) = 0$. By definition,

$$\begin{aligned}
\Gamma(\alpha_o)\left[h - h_o, u^*_{g_n}\right] &= \mathbb{E}\left[\tau(w)\partial_u g_o(w)(h(w_1) - h_o(w_1))u^*_{g_n}(w)\right] \\
&= \mathbb{E}\left[\tau(w) r^*_{\psi,h}(Z_2, \alpha_o)[h_{o,n} - h_o, u^*_{g_n}]\right]. \quad (3.58)
\end{aligned}$$

Therefore, Assumption 1.1.(v) has been verified in (3.37) above. By (3.3) and (3.55),

$$\mathbb{E}\left[\Delta^*_\psi(Z_2, g_o, h_o)[u^*_{g_n}]\middle| Z_1\right] = \mathbb{E}\left[\eta u^*_{g_n}(w)\middle| x, w_1\right] = u^*_{g_n}(w)\mathbb{E}\left[\eta | x, w_1\right] = 0 \quad (3.59)$$

which verifies Assumption 1.4.(iii). By definition,

$$r^*_{\psi,h}(z_2, \alpha)[v_g, v_h] = \partial_u g(w_h) v_g v_h \quad (3.60)$$

for any $z_2 \in \mathcal{Z}_2$, any $\alpha \in \mathcal{N}_\alpha$, any $v_h \in \mathcal{V}_1$ and any $v_g \in \mathcal{V}_2$, which implies that Assumption 1.4.(iv) holds. By the triangle inequality, (3.7) and (3.120) in the proof of Lemma 3.7, for any $\alpha \in \mathcal{N}_n$,

$$\begin{aligned}
&\left|\Delta_\psi(z_2, g, h)[u^*_{g_n}] - \Delta_\psi(z_2, g_o, h)[u^*_{g_n}] - r_{\psi,g}(z_2, g_o, h)[g - g_o, u^*_{g_n}]\right| \\
&= \left|\tau(w_h)u^*_{g_n}(w)\left((y - g(w_h)) - (y - g_o(w_h)) + (g(w) - g_o(w))\right)\right| \\
&= \left|\tau(w_h)u^*_{g_n}(w)\left((g_o(w_h) - g(w_h)) + (g(w) - g_o(w))\right)\right| \\
&\leq \left|\tau(w_h)u^*_{g_n}(w)\left((g_{o,K}(w_h) - g(w_h)) - (g_{o,K}(w) - g(w))\right)\right| \\
&+ \left|\tau(w_h)u^*_{g_n}(w)\left((g_{o,K}(w_h) - g_o(w_h)) - (g_{o,K}(w) - g_o(w))\right)\right| \\
&\leq \xi_{1,K}\left\|\beta - \beta_{o,K}\right\|\left|(h(w_1) - h_o(w_1))u^*_{g_n}\right| + CK^{-s/d}|u^*_{g_n}| \\
&\leq \xi_{1,K}\delta_{g,n}\left|(h(w_1) - h_o(w_1))u^*_{g_n}\right| + CK^{-s/d}|u^*_{g_n}|. \quad (3.61)
\end{aligned}$$

Let $\Lambda_{3,n}(z_2, \alpha) = \xi_{1,K}\delta_{g,n}\left|(h(w_1) - h_o(w_1))u^*_{g_n}\right| + CK^{-s/d}|u^*_{g_n}|$. By Lemma 3.3.(a), 3.3.(f) and Assumption 3.7,

$$\sup_{h\in\mathcal{N}_{h,n}}\sum_{i=1}^{n}\Lambda_{3,n}(Z_{2,i},\alpha)=o_p(n^{-1/2}). \tag{3.62}$$

Similarly, by Lemma 3.3.(b), 3.3.(e) and Assumption 3.7,

$$\sup_{h\in\mathcal{N}_{h,n}}\mathbb{E}\left[\Lambda_{3,n}(Z_2,\alpha)\right]=o(n^{-1/2}). \tag{3.63}$$

This verifies Assumption 1.4.(v). By definition,

$$\mathbb{E}\left[(\Delta^*_\psi(Z_2,\alpha_o)[u^*_{g_n}])^2\middle| Z_1=z_1\right]=(u^*_{g_n}(w))^2\mathbb{E}\left[\eta^2\middle| Z_1=z_1\right]\le C\xi^2_{0,K} \tag{3.64}$$

where the inequality is by Assumption 3.5 and Lemma 3.3.(g). By Lemma 3.3.(d), Assumption 1.4.(vii) holds with $\delta^*_{\tau,n}=\zeta_L\delta_{h,n}$, and $\delta^*_{\tau,n}\xi^2_{0,K}=o(1)$ follows by Assumption 3.7.

We next verify Assumption 1.5. By (3.56), Assumption 1.5.(i) holds with $\Lambda_{6,n}(z_2,\alpha)=0$ for any $z_2\in\mathcal{Z}_2$ and any $\alpha\in\mathcal{N}_n$. This also means that Assumptions 1.5.(vii)-(viii) also hold for $\Lambda_{6,n}(z_2,\alpha)$. By Assumption 3.3, (3.55) and (3.60),

$$\begin{aligned}
&\left|\Delta^*_\psi(z_2,g_o,h)[u^*_{g_n}]-\Delta^*_\psi(z_2,\alpha_o)[u^*_{g_n}]-r^*_{\psi,h}(z_2,\alpha_o)[h-h_o,u^*_{g_n}]\right|\\
&=\left|u^*_{g_n}(w)\left((y-g_o(w_h))-(y-g_o(w))+\partial_u g_o(w)(h(w)-h_o(w))\right)\right|\\
&=\left|u^*_{g_n}(w)\left((g_o(w)-g_o(w_h))-\partial_u g_o(w)(h_o(w)-h(w))\right)\right|\\
&\le C\left|u^*_{g_n}(w)(h(w)-h_o(w))^2\right|.
\end{aligned} \tag{3.65}$$

Let $\Lambda_{7,n}(z_2,\alpha)=C\left|u^*_{g_n}(w)(h(w)-h_o(w))^2\right|$. Then by Lemma 3.3.(a) and 3.3.(g), and Assumption 3.7,

$$\sup_{h\in\mathcal{N}_{h,n}}\sum_{i=1}^{n}\Lambda_{7,n}(Z_{2,i},\alpha)=o_p(n^{-1/2}). \tag{3.66}$$

Similarly, by Lemma 3.3.(b) and 3.3.(g), and Assumption 3.7,

$$\sup_{h\in\mathcal{N}_{h,n}}\mathbb{E}\left[\Lambda_{7,n}(Z_2,\alpha)\right]=o_p(n^{-1/2}). \tag{3.67}$$

This shows that Assumptions 1.5.(ii) and 1.5.(vii)-(viii) hold. For any $h \in \mathcal{N}_{h,n}$,

$$
\begin{aligned}
&\left|(\tau(Z_1,h)-\tau(Z_1,h_o))r^*_{\psi,h}(Z_2,\alpha_o)[h-h_o,u^*_{g_n}]\right| \\
&= \left|(\tau(w_h)-\tau(w))\partial_u g_o(w)(h(w_1)-h_o(w_1))u^*_{g_n}\right| \\
&\leq C\zeta_L\xi_{0,K}\delta_{h,n}\left|\tau(w_h)-\tau(w)\right| \qquad (3.68)
\end{aligned}
$$

where the inequality is by Assumption 3.5, (3.91) in the proof of Lemma 3.3 and Lemma 3.3.(g). By (3.68), Lemma 3.3.(c) and Assumption 3.7,

$$
\sup_{h\in\mathcal{N}_{h,n}}\left|\mathbb{E}\left[(\tau(Z_1,h)-\tau(Z_1,h_o))r^*_{\psi,h}(Z_2,\alpha_o)[h-h_o,u^*_{g_n}]\right]\right| \leq C\zeta_L^2\xi_{0,K}\delta^2_{h,n}=o(n^{-1/2}), \qquad (3.69)
$$

which verifies Assumption 1.5.(iii). By (3.7) and Lemma 3.2.(d), for any $g \in \mathcal{N}_{g,n}$

$$
\begin{aligned}
&\sup_w\left|(g(w)-g_o(w))u^*_{g_n}(w)\right| \\
&=\sup_w\left|\tau(w)(g(w)-g_o(w))u^*_{g_n}(w)\right| \\
&\leq \sup_{w\in\mathcal{T}_w}\left|\tau(w)(g(w)-g_{o,K}(w))u^*_{g_n}(w)\right| + \sup_{w\in\mathcal{T}_w}\left|\tau(w)(g_{o,K}(w)-g_o(w))u^*_{g_n}(w)\right| \\
&\leq \sup_w\left|u^*_{g_n}(w)\right|\sup_{w\in\mathcal{T}_w}\left|g(w)-g_{o,K}(w)\right| \\
&+\sup_w\left|u^*_{g_n}(w)\right|\sup_{w\in\mathcal{T}_w}\left|g(w)-g_{o,K}(w)\right| \\
&\leq C\xi_{0,K}(\|\beta-\beta_{o,K}\|\,\xi_{0,K}+K^{-s/d})\leq C\xi^2_{0,K}\delta_{g,n} \qquad (3.70)
\end{aligned}
$$

where the first equality is by $\tau(w)^2=\tau(w)$, the first inequality is by the triangle inequality, the third inequality is by (3.7) and Lemma 3.3.(g), and the last inequality is by the definition of $\mathcal{N}_{g,n}$. For any $\alpha \in \mathcal{N}_n$,

$$
\begin{aligned}
&\left|(\tau(Z_1,h)-\tau(Z_1,h_o))r^*_{\psi,g}(Z_2,\alpha_o)[g-g_o,u^*_{g_n}]\right| \\
&= \left|(\tau(w_h)-\tau(w))(g(w)-g_o(w))u^*_{g_n}\right| \\
&\leq C\xi^2_{0,K}\delta_{g,n}\left|\tau(Z_1,h)-\tau(Z_1,h_o)\right| \qquad (3.71)
\end{aligned}
$$

where the inequality is by (3.70) and the definition of $\mathcal{N}_n$. By (3.68), Lemma 3.3.(c) and Assumption 3.7,

$$\sup_{h\in\mathcal{N}_{h,n}} \left|\mathbb{E}\left[(\tau(Z_1,h)-\tau(Z_1,h_o))r^*_{\psi,g}(Z_2,\alpha_o)[g-g_o,u^*_{g_n}]\right]\right| \leq C\zeta_L\xi^2_{0,K}\delta_{h,n}\delta_{g,n} = o(n^{-1/2}), \tag{3.72}$$

which verifies Assumption 1.5.(iv). By (3.56), Assumption 1.5.(v) holds with $\Lambda_{8,n}(z_2,\alpha)=0$ for any $z_2\in\mathcal{Z}_2$ and any $\alpha\in\mathcal{N}_n$. By (3.56) and Lemma 3.3.(e), Assumption 1.5.(vi) also holds. Assumptions 1.5.(vii) and 1.5.(viii) have been verified together with Assumptions 1.5.(i) and 1.5.(ii).

Finally, we verify Assumption 1.6. Let $h_{o,n}=h_{o,L}$. By (3.60), 3.6), Assumption 3.7 and Lemma 3.3.(e)

$$\begin{aligned}
&\mathbb{E}\left[\left|r^*_{\psi,h}(Z_2,\alpha_o)[h_{o,L}-h_o,u^*_{g_n}]\right|\right]\\
&=\mathbb{E}\left[\left|\partial_u g_o(w)(h_{o,L}-h_o)u^*_{g_n}\right|\right]\\
&\leq C(\mathbb{E}\left[\left|(h_{o,L}-h_o)^2\right|\right])^{1/2}(\mathbb{E}\left[\left|(u^*_{g_n})^2\right|\right])^{1/2}\\
&\leq CL^{-s_1/d_{w_1}} = o(n^{-1/2})
\end{aligned}\tag{3.73}$$

which verifies Assumption 1.6.(i). Assumption 1.6.(ii) can be verified using the same arguments of the proof of Lemma 3.4. Let $g_{o,n}=g_{o,K}$. By (3.56), (3.7), Assumption 3.7 and Lemma 3.3.(e)

$$\begin{aligned}
&\mathbb{E}\left[\left|r^*_{\psi,g}(Z_2,\alpha_o)[g_{o,K}-g_o,u^*_{g_n}]\right|\right]\\
&=\mathbb{E}\left[\left|(g_{o,K}-g_o)u^*_{g_n}\right|\right]\\
&\leq C(\mathbb{E}\left[\left|(g_{o,K}-g_o)^2\right|\right])^{1/2}(\mathbb{E}\left[\left|(u^*_{g_n})^2\right|\right])^{1/2}\\
&\leq CL^{-s/d} = o(n^{-1/2})
\end{aligned}\tag{3.74}$$

which verifies Assumption 1.6.(iii). For any $h\in\mathcal{N}_{h,n}$,

$$\begin{aligned}
&\mathbb{E}\left[\tau(Z_1,h)^2(r^*_{\psi,h}(z_2,\alpha_o)[h-h_{o,L},u^*_{g_n}])^2\right]\\
&=\mathbb{E}\left[\tau(w_h)^2(\partial_u g_o(w)(h-h_{o,L})u^*_{g_n})^2\right]\\
&\leq C\sup_{w_1\in\mathcal{W}_1}\left|(h(w_1)-h_{o,L}(w_1))^2\right|\mathbb{E}\left[(u^*_{g_n})^2\right]\leq C\zeta_L^2\delta^2_{h,n}
\end{aligned}\tag{3.75}$$

where the first inequality is by $\tau(w_h)^2<1$ for any $h\in\mathcal{N}_{h,n}$ and Assumption 3.5, the last inequality is by (3.91) and Lemma 3.3.(e). Moreover, for any $f \in \mathcal{F}_{3,n}^*$,

$$\sup_{z_2 \in \mathcal{Z}_2} |f(z_2)| \leq \left(\sup_{h \in \mathcal{N}_{h,n}} \sup_{w_1 \in \mathcal{W}_1} \left| (h(w_1) - h_{o,L}(w_1))^2 \right| \right) \left(\sup_{w \in \mathcal{W}} \left| (u_{g_n}^*(w))^2 \right| \right) \leq C\xi_{0,K}^2 \zeta_L^2 \delta_{h,n}^2, \tag{3.76}$$

which together with (3.75) and Assumption 3.7 implies that

$$\begin{aligned} &(\sup_{f \in \mathcal{F}_{3,n}^*} \mathbb{E}\left[f^2\right] + (K+L) \sup_{z_2 \in \mathcal{Z}_2} |F_{3,n}^*(z_2)| \log(n) n^{-1})(K+L)\log(n) \\ &\leq C(\zeta_L^2 \delta_{h,n}^2 (K+L) + (K+L)^2 \xi_{0,K}^2 \zeta_L^2 \delta_{h,n}^2 \log(n) n^{-1}) \log(n) = o(1). \end{aligned} \tag{3.77}$$

This verifies Assumption 1.6.(v) for $\mathcal{F}_{3,n}^*$. For any $h \in \mathcal{N}_{h,n}$ and $g \in \mathcal{N}_{g,n}$,

$$\tau(z_1, h) r_{\psi,g}^*(z_2, \alpha_o)[g - g_{o,K}, u_{g_n}^*] = \tau(w_h) u_{g_n}^*(w) P(w)'(\beta - \beta_{o,K}). \tag{3.78}$$

Hence Assumption 1.6.(iv) can be verified using the same arguments of Lemma 3.5. For any $h \in \mathcal{N}_{h,n}$ and $g \in \mathcal{N}_{g,n}$,

$$\begin{aligned} &\mathbb{E}\left[\tau(Z_1, h)^2 (r_{\psi,g}^*(z_2, \alpha_o)[g - g_{o,K}, u_{g_n}^*])^2\right] \\ &= \mathbb{E}\left[\tau(w_h)^2 ((g - g_{o,K}) u_{g_n}^*)^2\right] \\ &\leq C \sup_{w_1 \in \mathcal{W}_1} \left| \tau(w)(g(w) - g_{o,K}(w))^2 \right| \mathbb{E}\left[(u_{g_n}^*)^2\right] \leq C\xi_{0,K}^2 \delta_{g,n}^2 \end{aligned} \tag{3.79}$$

where the first inequality is by $\tau(w)^2 = \tau(w)$ and $\tau(w_h)^2 < 1$ for any w and any $h \in \mathcal{N}_{h,n}$, the second inequality is by the definition of $\mathcal{N}_{g,n}$ and Lemma 3.3.(e). Moreover, for any $f \in \mathcal{F}_{4,n}^*$,

$$\sup_{z_2 \in \mathcal{Z}_2} |f(z_2)| \leq \left(\sup_{g \in \mathcal{N}_{h,n}} \sup_{w \in \mathcal{T}_w} \left| (g(w) - g_{o,K}(w))^2 \right| \right) \left(\sup_{w \in \mathcal{W}} \left| (u_{g_n}^*(w))^2 \right| \right) \leq C\xi_{0,K}^4 \delta_{g,n}^2, \tag{3.80}$$

which together with (3.79) and Assumption 3.7 implies that

$$\begin{aligned} &(\sup_{f \in \mathcal{F}_{4,n}^*} \mathbb{E}\left[f^2\right] + (K+L) \sup_{z_2 \in \mathcal{Z}_2} |F_{4,n}^*(z_2)| \log(n) n^{-1})(K+L)\log(n) \\ &\leq C(\xi_{0,K}^2 \delta_{g,n}^2 (K+L) + (K+L)^2 \xi_{0,K}^4 \delta_{g,n}^2 \log(n) n^{-1}) \log(n) = o(1). \end{aligned} \tag{3.81}$$

This verifies Assumption 1.6.(v) for $\mathcal{F}_{3,n}^*$. ■

Lemma 3.1 *Let $v_\Gamma^*(w_1) = \mathbb{E}\left[\tau(w)v_g^*(w)\partial_u g_o(w)|w_1\right]$. Under Assumptions 3.5 and 3.6, we have*

(a) $\mathbb{E}[\tau(w)|v_{g_n}^*(w) - v_g^*(w)|^2] \to 0$ as $K \to \infty$;

(b) $\mathbb{E}\left[\eta^2\tau(w)(v_{g_n}^*(w))^2\right] \to \mathbb{E}\left[\eta^2\tau(w)(v_g^*(w))^2\right]$ as $K \to \infty$;

(c) $\mathbb{E}\left[|v_{\Gamma_n}^*(w_1) - v_\Gamma^*(w_1)|^2\right] \to 0$ as $K \to \infty$ and $L \to \infty$;

(d) $\mathbb{E}\left[u^2(v_{\Gamma_n}^*(w_1))^2\right] \to \mathbb{E}\left[u^2(v_\Gamma^*(w_1))^2\right]$ as $K \to \infty$ and $L \to \infty$.

Proof of Lemma 3.1. (a) By the definition of $v_{g_n}^*$ and Assumption 3.6,

$$\mathbb{E}[\tau(w)P(w)(v_{g_n}^*(w) - v_g^*(w))] = \mathbf{0}_{K\times 1} \tag{3.82}$$

which immediately implies that

$$\begin{aligned}
&\mathbb{E}[\tau(w)|P(w)'\beta_{g,K} - v_g^*(w)|^2] \\
&= \mathbb{E}[\tau(w)|P(w)'\beta_{g,K} - v_{g_n}^*(w)|^2] + \mathbb{E}[\tau(w)|v_{g_n}^*(w) - v_g^*(w)|^2] \\
&\geq \mathbb{E}[\tau(w)|v_{g_n}^*(w) - v_g^*(w)|^2]
\end{aligned} \tag{3.83}$$

for any $\beta_{g,K} \in \mathbb{R}^K$. Hence as $K \to \infty$,

$$\mathbb{E}[\tau(w)|v_{g_n}^*(w) - v_g^*(w)|^2] \leq \mathbb{E}[\tau(w)|P(w)'\beta_{v,K} - v_g^*(w)|^2] \to 0, \tag{3.84}$$

where $\beta_{v,K}$ is defined in Assumption 3.6.

(b) By Assumption 3.5, Jensen's inequality and Hölder's inequality,

$$\begin{aligned}
&\left|\mathbb{E}\left[\eta^2\tau(w)(v_{g_n}^*(w) - v_g^*(w))v_g^*(w)\right]\right| \\
&\leq C\mathbb{E}\left[\tau(w)\left|(v_{g_n}^*(w) - v_g^*(w))v_g^*(w)\right|\right] \\
&\leq C(\mathbb{E}\left[\tau(w)\left|(v_{g_n}^*(w) - v_g^*(w))^2\right|\right]\mathbb{E}\left[\tau(w)(v_g^*(w))^2\right])^{1/2}
\end{aligned} \tag{3.85}$$

which together with Assumption 3.6 and the result proved in (a) implies that

$$\left|\mathbb{E}\left[\eta^2\tau(w)(v_{g_n}^*(w) - v_g^*(w))v_g^*(w)\right]\right| \to 0 \text{ as } K \to \infty. \tag{3.86}$$

By the triangle inequality,

$$
\begin{aligned}
&\left|\mathbb{E}\left[\eta^2\tau(w)(v^*_{g_n}(w))^2\right] - \mathbb{E}\left[\eta^2\tau(w)(v^*_g(w))^2\right]\right| \\
&\leq \mathbb{E}\left[\eta^2\tau(w)(v^*_{g_n}(w) - v^*_g(w))^2\right] \\
&+ 2\left|\mathbb{E}\left[\eta^2\tau(w)(v^*_{g_n}(w) - v^*_g(w))v^*_g(w)\right]\right|, \qquad (3.87)
\end{aligned}
$$

which combined with the results in (3.84), (3.85) and (3.86) proves the claim (b).

(c) Let $v^*_{\Gamma,L}(w_1) = R(\cdot)'Q_L^{-1}\mathbb{E}\left[R(w_1)\tau(w)\partial_u g_o(w)v^*_g(w)\right]$. Then

$$
v^*_{\Gamma_n}(w_1) - v^*_{\Gamma,L}(w_1) = R(\cdot)'Q_L^{-1}\mathbb{E}\left[R(w_1)\tau(w)\partial_u g_o(w)(v^*_{g_n}(w) - v^*_g(w))\right]. \qquad (3.88)
$$

By the (matrix) Cauchy-Schwarz inequality, Assumption 3.5 and the result proved in (a),

$$
\begin{aligned}
\mathbb{E}\left[|v^*_{\Gamma_n}(w_1) - v^*_{\Gamma,L}(w_1)|^2\right] &\leq \mathbb{E}\left[\tau(w)(\partial_u g_o(w))^2(v^*_{g_n}(w) - v^*_g(w))^2\right] \\
&\leq C\mathbb{E}\left[\tau(w)(v^*_{g_n}(w) - v^*_g(w))^2\right] \to 0 \qquad (3.89)
\end{aligned}
$$

as $K \to \infty$. Using the same arguments after display (A.9) of Newey, Powell and Vella (1999) (their $b_L(z)$ and $\rho(z)$ are $v^*_{\Gamma,L}(w_1)$ and $v^*_\Gamma(w_1)$ here respectively), we can show that

$$
\mathbb{E}\left[|v^*_{\Gamma,L}(w_1) - v^*_\Gamma(w_1)|^2\right] \to 0 \text{ as } L \to \infty. \qquad (3.90)
$$

Combining the results in (3.89) and (3.90), we immediately prove the claim in (c).

(d) The proof follows similar arguments in the proof of claim (b) and hence is omitted. ■

Let $\widehat{Q}_{n,L} = n^{-1}R_nR_n'$ and $\widehat{Q}_{n,K} = n^{-1}\widehat{P}_n'\widehat{P}_n$, which are the estimators of $Q_L = \mathbb{E}\left[R(w_1)R(w_1)'\right]$ and $Q_K = \mathbb{E}\left[\tau(w)P(w)P(w)'\right]$ respectively. The following Lemma is useful to verify the high-level conditions for the asymptotic normality. The proof of the results in Lemmas 3.2.(a) and 3.2.(b) are in Newey (1997) and the proof of the remaining results are in Newey, Powell and Vella (1999).

Lemma 3.2 *Let $\delta^*_{h,n} = L^{1/2}n^{-1/2} + L^{-s_1/d_{w_1}}$ and $\delta^*_{g,n} = K^{1/2}n^{-1/2} + K^{-s/d} + \delta^*_{h,n}$. Under Assumptions 3.1-3.4, we have*

(a) $||\widehat{Q}_{n,L} - Q_L|| = O_p(\zeta_L L^{1/2}n^{-1/2})$;

(b) $||\widehat{\gamma}_n - \gamma_{o,L}|| = O_p(\delta^*_{h,n})$;

(c) $||\widehat{Q}_{n,K} - Q_K|| = O_p(\xi^2_{1,K}(\delta^*_{h,n})^2 + K^{1/2}\xi_{1,K}\delta^*_{h,n} + \xi^2_{0,K}\zeta_L\delta^*_{h,n})$;

(d) $||\widehat{\beta}_n - \beta_{o,K}|| = O_p(\delta^*_{g,n})$;

(e) $n^{-1}\sum_{i=1}^n |\widehat{\tau}_i - \tau_i| = O_p(\zeta_L \delta^*_{h,n})$.

Recall that $\mathcal{N}_{\gamma,n} = \{\gamma \in \mathbb{R}^L\colon\ ||\gamma - \gamma_{o,L}|| \leq \delta_{h,n}\}$ and $\mathcal{N}_{\beta,n} = \{\beta \in \mathbb{R}^K\colon\ ||\beta - \beta_{o,K}|| \leq \delta_{g,n}\}$ where $\delta_{h,n} = \delta^*_{h,n}\varrho_n$, $\delta_{g,n} = \delta^*_{g,n}\varrho_n$ and $\{\varrho_n\}_n$ is a slowly divergent real positive sequence. By Lemma 3.2.(b) and Lemma 3.2.(d), we have $\widehat{\gamma}_n \in \mathcal{N}_{\gamma,n}$ and $\widehat{\beta}_n \in \mathcal{N}_{\beta,n}$ wpa1. Define $\mathcal{N}_{h,n} = \{h(\cdot) = R(\cdot)'\gamma\colon\ \gamma \in \mathcal{N}_{\gamma,n}\}$ and $\mathcal{N}_{g,n} = \{g(\cdot) = P(\cdot)'\beta\colon\ \beta \in \mathcal{N}_{\gamma,n}\}$. The following Lemma is useful to verify the high-level conditions.

Lemma 3.3 *Under Assumptions 3.1-3.6, we have*

(a) $\sup_{\gamma\in\mathcal{N}_{\gamma,n}} n^{-1}\sum_{i=1}^n \left[|R(w_1)'\gamma - h_o(w_1)|^2\right] = O_p(\delta^2_{h,n})$;

(b) $\sup_{\gamma\in\mathcal{N}_{\gamma,n}} \mathbb{E}\left[|R(w_{1,i})'\gamma - h_o(w_{1,i})|^2\right] = O(\delta^2_{h,n})$;

(c) $\sup_{h\in\mathcal{N}_{h,n}} \mathbb{E}\left[|\tau(w_h) - \tau(w)|\right] \leq C\zeta_L\delta_{h,n}$;

(d) $\sup_{h\in\mathcal{N}_{h,n}} n^{-1}\sum_{i=1}^n \left[|\tau(w_{h,i}) - \tau(w_i)|\right] = O_p(\zeta_L\delta_{h,n})$;

(e) $\mathbb{E}\left[(v^*_{g_n}(w))^2\right] \leq C\,\|v^*_n\|^2_{sd}$;

(f) $n^{-1}\sum_{i=1}^n (v^*_{g_n}(w_i))^2\,\|v^*_n\|^{-2}_{sd} = O_p(1)$;

(g) $\sup_w \left|v^*_{g_n}(w)\,\|v^*_n\|^{-1}_{sd}\right| \leq C\xi_{0,K}$;

(h) $\mathbb{E}\left[\left|v^*_{\Gamma_n}(w_1)\right|^2\right] \leq C\,\|v^*_n\|^2_{sd}$.

Proof of Lemma 3.3. Following Newey (1997) we assume without loss of generality that $Q_L = I_L$ and $Q_K = I_K$. Such an assumption can be verified under Assumption 3.2 for the power series and splines using the arguments in the proof of Theorem 4 and Theorem 7 of Newey (1997) respectively.

(a) By Assumption 3.4, Lemma 3.2.(a), $Q_L = I_L$, the Cauchy-Schwarz inequality, the definition of $\mathcal{N}_{\gamma,n}$ and (3.6),

$$
\begin{aligned}
&\sup_{\gamma\in\mathcal{N}_{\gamma,n}} n^{-1}\sum_{i=1}^n \left[\left|R(w_{1,i})'\gamma - h_o(w_{1,i})\right|^2\right] \\
&\leq 2\sup_{\gamma\in\mathcal{N}_{\gamma,n}} n^{-1}\sum_{i=1}^n \left[\left|R(w_{1,i})'\gamma - h_{o,L}(w_{1,i})\right|^2\right] + 2\sup_{\gamma\in\mathcal{N}_{\gamma,n}} n^{-1}\sum_{i=1}^n \left[\left|h_{o,L}(w_{1,i}) - h_o(w_{1,i})\right|^2\right] \\
&\leq 2\sup_{\gamma\in\mathcal{N}_{\gamma,n}} (\gamma - \gamma_{o,L})'\widehat{Q}_{n,L}(\gamma - \gamma_{o,L}) + 2Ck^{-s_1/d_{w_1}} \\
&\leq 2\omega_{\max}(\widehat{Q}_{n,L})\sup_{\gamma\in\mathcal{N}_{\gamma,n}} \|\gamma - \gamma_{o,L}\|^2 + 2Ck^{-s_1/d_{w_1}} = O_p(\delta^2_{h,n}),
\end{aligned}
$$

which proves the claim in (a).

(b) The proof follows similar arguments to those in the proof of (a) and is omitted.

(c) For any $h(\cdot) = P(\cdot)'\gamma \in \mathcal{N}_{h,n}$,

$$
\begin{aligned}
|h(w_1) - h_o(w_1)| &\leq \left|P(w_1)'\gamma - h_{o,L}(w_1)\right| + |h_{o,L}(w_1) - h_o(w_1)| \\
&\leq \zeta_L \|\gamma - \gamma_{o,L}\| + CL^{-s_1/d_{w_1}} \leq C\zeta_L \delta_{h,n} \qquad (3.91)
\end{aligned}
$$

which implies that

$$
\begin{aligned}
|\tau(w_h) - \tau(w)| &\leq \left|I\left\{u \leq b + P(w_1)'\gamma - h_o(w_1)\right\} - I\{u \leq b\}\right| \\
&\quad + \left|I\left\{u \geq a + P(w_1)'\gamma - h_o(w_1)\right\} - I\{u \geq a\}\right| \\
&\leq I\left\{|u - b| \leq \left|P(w_1)'\gamma - h_o(w_1)\right|\right\} \\
&\quad + I\left\{|u - a| \leq \left|P(w_1)'\gamma - h_o(w_1)\right|\right\} \\
&\leq I\left\{|u - b| \leq C\zeta_L\delta_{h,n}\right\} + I\left\{|u - a| \leq C\zeta_L\delta_{h,n}\right\}, \qquad (3.92)
\end{aligned}
$$

where $\zeta_L\delta_{h,n} = o(1)$ by Assumption 3.7. As the density of u is bounded in the local neighborhoods of a and b (which is assumed in Lemma A3 of Newey, Powell and Vella (1999)), by (3.92) we get

$$
\mathbb{E}\left[\sup_{h \in \mathcal{N}_{h,n}} |\tau(w_h) - \tau(w)|\right] \leq C\zeta_L\delta_{h,n} \qquad (3.93)
$$

which finishes the proof.

(d) By (3.93) and the Markov inequality we immediately get the asserted result.

(e) By the definition of η and Assumption 3.5, $\mathbb{E}\left[\eta^2|x, w_1\right] \geq C_\eta$ where C_η is a finite positive constant. Thus

$$
\begin{aligned}
\mathbb{E}[(v^*_{g_n}(w))^2 \|v^*_n\|_{sd}^{-2}] &= \frac{\mathbb{E}[(v^*_{g_n}(w))^2]}{\mathbb{E}\left[u^2(v^*_{\Gamma_n}(w_1))^2\right] + \mathbb{E}\left[\eta^2\tau(w)(v^*_{g_n}(w))^2\right]} \\
&\leq \frac{\mathbb{E}[(v^*_{g_n}(w))^2]}{\mathbb{E}\left[u^2(v^*_{\Gamma_n}(w_1))^2\right] + C_\eta\mathbb{E}\left[\tau(w)(v^*_{g_n}(w))^2\right]} \\
&= \frac{\mathbb{E}[\tau(w)(v^*_{g_n}(w))^2]}{\mathbb{E}\left[u^2(v^*_{\Gamma_n}(w_1))^2\right] + C_\eta\mathbb{E}\left[\tau(w)(v^*_{g_n}(w))^2\right]} \leq C_\eta^{-1} \qquad (3.94)
\end{aligned}
$$

where the second equality is by the definition of $v^*_{g_n}$ and $\tau(w)^2 = \tau(w)$.

(f) The asserted result follows by (e) and the Markov inequality.

(g) By the Cauchy-Schwarz inequality and Assumption 3.5,

$$
\begin{aligned}
\left|v_{g_n}^*(w)\right|^2 \|v_n^*\|_{sd}^{-2} &= \frac{\left|\tau(w)P(w)'Q_K^{-1}\rho(P_K)\right|^2}{\mathbb{E}\left[u^2(v_{\Gamma_n}^*(w_1))^2\right]+\mathbb{E}\left[\eta^2\tau(w)(v_{g_n}^*(w))^2\right]} \\
&\leq \frac{\rho(P_K)'Q_K^{-2}\rho(P_K)\left\|\tau(w)P(w)\right\|^2}{\mathbb{E}\left[u^2(v_{\Gamma_n}^*(w_1))^2\right]+C_\eta\mathbb{E}\left[\tau(w)(v_{g_n}^*(w))^2\right]} \\
&\leq \frac{\xi_{0,K}^2\rho(P_K)'Q_K^{-2}\rho(P_K)}{\mathbb{E}\left[u^2(v_{\Gamma_n}^*(w_1))^2\right]+C_\eta\mathbb{E}\left[\tau(w)(v_{g_n}^*(w))^2\right]} \\
&\leq \frac{\xi_{0,K}^2\omega_{\min}^{-1}(Q_K)\mathbb{E}\left[\tau(w)(v_{g_n}^*(w))^2\right]}{\mathbb{E}\left[u^2(v_{\Gamma_n}^*(w_1))^2\right]+C_\eta\mathbb{E}\left[\tau(w)(v_{g_n}^*(w))^2\right]} \\
&\leq \xi_{0,K}^2\omega_{\min}^{-1}(Q_K)C_\eta^{-1}
\end{aligned}
\tag{3.95}
$$

for any w. This combined with $Q_K = I_K$ immediately proves the claim.

(h) By Lemmas 3.1.(b) and 3.1.(d),

$$
\mathbb{E}[\left|v_{\Gamma_n}^*(w_1)\right|^2 \|v_n^*\|_{sd}^{-2}] \to \frac{\mathbb{E}\left[(v_\Gamma^*(w_1))^2\right]}{\mathbb{E}\left[\eta^2\tau(w)(v_g^*(w))^2\right]+\mathbb{E}\left[u^2(v_\Gamma^*(w_1))^2\right]} \tag{3.96}
$$

as $K \to \infty$ and $L \to \infty$, where $v_\Gamma^*(w_1) = \mathbb{E}\left[\tau(w)v_g^*(w)\partial_u g_o(w)|w_1\right]$. By Assumption 3.5 and Jensen's inequality,

$$
\mathbb{E}\left[(v_\Gamma^*(w_1))^2\right] \leq C\mathbb{E}\left[(\mathbb{E}\left[\tau(w)v_g^*(w)|w_1\right])^2\right] \leq C\mathbb{E}\left[\tau(w)(v_g^*(w))^2\right]. \tag{3.97}
$$

By Assumption 3.5, $\mathbb{E}\left[\eta^2|x,w_1\right] \geq C_\eta$ where C_η is a finite positive constant, which together with (3.97) implies that

$$
\begin{aligned}
&\frac{\mathbb{E}\left[\tau(w)(v_g^*(w))^2\right]}{\mathbb{E}\left[\eta^2\tau(w)(v_g^*(w))^2\right]+\mathbb{E}\left[u^2(v_\Gamma^*(w_1))^2\right]} \\
&\leq \frac{\mathbb{E}\left[\tau(w)(v_g^*(w))^2\right]}{C_\eta\mathbb{E}\left[\tau(w)(v_g^*(w))^2\right]+\mathbb{E}\left[u^2(v_\Gamma^*(w_1))^2\right]} \leq C_\eta^{-1}.
\end{aligned}
\tag{3.98}
$$

The asserted claim follows from (3.96) and (3.98). ■

Lemma 3.4 *Define $\mathcal{F}_{1,n} = \{(x, w_1) \mapsto \partial_u g_o(w)\tau(w_h)(h(w_1) - h_o(w_1))u_{g_n}^*(w) : h \in \mathcal{N}_{h,n}\}$. Then the uniform entropy numbers of $\mathcal{F}_{1,n}$ satisfies*

$$
\sup_Q N(\varepsilon \left\|F_{1,n}\right\|_{Q,2}, \mathcal{F}_{1,n}, L_2(Q)) \leq (C/\varepsilon)^{CL} \text{ for any } \varepsilon \in (0,1], \tag{3.99}
$$

where C is a finite fixed constant, Q ranges over all finitely-discrete probabilities measures and $F_{1,n}$ denotes the envelope of $\mathcal{F}_{1,n}$.

Proof of Lemma 3.4. Let $\tau(x, w_2) = \prod_{j=1}^{d_{w_2}+1} I\{a_j \leq w_j \leq b_j\}$, $a = a_{d_{w_2}+2}$ and $b = b_{d_{w_2}+2}$. Then by definition,

$$\tau(w_h) = \tau(x, w_2) I\{a \leq x - h(w_1) \leq b\}. \tag{3.100}$$

Define

$$\mathcal{F}_{11,n} = \{(x, w_1) \mapsto I\{x \leq b + R(w_1)'\gamma : \gamma \in \mathcal{N}_{\gamma,n}\}; \tag{3.101}$$

$$\mathcal{F}_{12,n} = \{(x, w_1) \mapsto I\{x \geq a + R(w_1)'\gamma : \gamma \in \mathcal{N}_{\gamma,n}\}; \tag{3.102}$$

$$\mathcal{F}_{13,n} = \{(x, w_1) \mapsto \tau(x, w_2)\partial_u g_o(w)(R(w_1)'\gamma - h_o(w_1))u^*_{g_n}(w) : \gamma \in \mathcal{N}_{\gamma,n}\}. \tag{3.103}$$

Then by Lemmas 2.6.15 and 2.6.18 in van der Vaart and Wellner (1996), the VC-dimentions of $\mathcal{F}_{11,n}$, $\mathcal{F}_{12,n}$ and $\mathcal{F}_{13,n}$ are of order L. By Theorem 2.6.7 in van der Vaart and Wellner (1996), the uniform entropy number of $\mathcal{F}_{1j,n}$ satisfies

$$\sup_Q N(\varepsilon \left\|F_{1j,n}\right\|_{Q,2}, \mathcal{F}_{1j,n}, L_2(Q)) \leq (C/\varepsilon)^{CL} \text{ for any } \varepsilon \in (0, 1], \tag{3.104}$$

where C is a universal constant and $F_{1j,n}$ denotes the envelope of $\mathcal{F}_{1j,n}$ for $j = 1, 2, 3$. Because

$$\mathcal{F}_{1,n} \subset \{f_1 f_2 f_3 : f_1 \in \mathcal{F}_{11,n}, f_2 \in \mathcal{F}_{12,n}, f_3 \in \mathcal{F}_{13,n}\}, \tag{3.105}$$

by (A.6) and (A.7) in Andrews (1994),

$$\begin{aligned} &\sup_Q N(\varepsilon \left\|F_{11,n}F_{12,n}F_{13,n}\right\|_{Q,2}, \mathcal{F}_{1,n}, L_2(Q)) \\ &\leq \prod_{j=1}^{3} \sup_Q N(\varepsilon \left\|F_{1j,n}\right\|_{Q,2}/3, \mathcal{F}_{1j,n}, L_2(Q)) \leq (C/\varepsilon)^{CL} \end{aligned} \tag{3.106}$$

where the second inequality is by (3.104). This proves (3.99) with $F_{1,n} = F_{11,n}F_{12,n}F_{13,n}$. ∎

Lemma 3.5 *Define $\mathcal{F}_{2,n} = \{(x, w_1) \mapsto \tau(w_h)u^*_{g_n}(w)P(w)'\alpha : h \in \mathcal{N}_{h,n},\ \alpha \in \mathbb{S}^{K-1}\}$, where $\mathbb{S}^{K-1} = \{\alpha \in \mathbb{R}^K : \alpha'\alpha = 1\}$. Then the uniform entropy numbers of $\mathcal{F}_{2,n}$ satisfies*

$$\sup_Q N(\varepsilon \left\|F_{2,n}\right\|_{Q,2}, \mathcal{F}_{2,n}, L_2(Q)) \leq (C/\varepsilon)^{C(L+K)} \textit{ for any } \varepsilon \in (0, 1], \tag{3.107}$$

where C is a finite fixed constant, Q ranges over all finitely-discrete probabilities measures and $F_{2,n}$ denotes the envelope of $\mathcal{F}_{2,n}$.

Proof of Lemma 3.5. Define

$$\mathcal{F}_{21,n} = \{(x, w_1) \mapsto \tau(x, w_2) u^*_{g_n}(w) P(w)'\alpha : \alpha \in \mathbb{S}^{K-1}\}, \tag{3.108}$$

where $\tau(x, w_2)$ is defined in the proof of Lemma 3.4. Then by Lemmas 2.6.15 and 2.6.18 in van der Vaart and Wellner (1996), the VC-dimension of $\mathcal{F}_{21,n}$ is of order K. By Theorem 2.6.7 in van der Vaart and Wellner (1996), the uniform entropy number of $\mathcal{F}_{21,n}$ satisfies

$$\sup_Q N(\varepsilon \|F_{21,n}\|_{Q,2}, \mathcal{F}_{21,n}, L_2(Q)) \le (C/\varepsilon)^{CK} \text{ for any } \varepsilon \in (0, 1], \tag{3.109}$$

where C is a universal constant and $F_{21,n}$ denotes the envelope of $\mathcal{F}_{21,n}$. The rest of the proof is the same as Lemma 3.4, because

$$\mathcal{F}_{2,n} \subset \{f_1 f_2 f_3 : f_1 \in \mathcal{F}_{11,n}, f_2 \in \mathcal{F}_{12,n}, f_3 \in \mathcal{F}_{21,n}\}, \tag{3.110}$$

where $\mathcal{F}_{11,n}$ and $\mathcal{F}_{12,n}$ are defined in (3.101) and (3.102) respectively. Hence (3.107) holds with $F_{2,n} = F_{11,n} F_{12,n} F_{21,n}$. ∎

Lemma 3.6 *Under Assumptions 3.1-3.7,*

$$\sup_{h \in \mathcal{N}_{h,n}} \left| \mu_n \left\{ \tau(w_h)(g_o(w) - g_o(w_h)) u^*_{g_n}(w) \right\} \right| = o_p(n^{-1/2}).$$

Proof of Lemma 3.6. Let $u_h = x - h(w_1)$. As $u = x - h_o(w_1)$, we have $u - u_h = h(w_1) - h_o(w_1)$ by definition. By Assumption 3.5,

$$|g_o(w) - g_o(w_h) - \partial_u g_o(w)(h(w_1) - h_o(w_1))| \le C |h(w_1) - h_o(w_1)|^2 \tag{3.111}$$

which together with the triangle inequality, Lemmas 3.3.(a), 3.3.(b) and 3.3.(g) implies that

$$\begin{aligned}
&\sup_{h\in\mathcal{N}_{h,n}} \left|\mu_n \left\{\tau(w_h)(g_o(w) - g_o(w_h) - \partial_u g_o(w)(h(w_1) - h_o(w_1)))u^*_{g_n}(w)\right\}\right| \\
&\leq C \sup_{h\in\mathcal{N}_{h,n}} n^{-1}\sum_{i=1}^{n}\left[\left|h(w_{1,i}) - h_o(w_{1,i})\right|^2 \left|u^*_{g_n}(w_i)\right|\right] \\
&+ C \sup_{h\in\mathcal{N}_{h,n}} \mathbb{E}\left[\left|h(w_1) - h_o(w_1)\right|^2 \left|u^*_{g_n}(w)\right|\right] \\
&\leq C \sup_{w}\left|u^*_{g_n}(w)\right| \sup_{h\in\mathcal{N}_{h,n}} n^{-1}\sum_{i=1}^{n}\left[\left|h(w_{1,i}) - h_o(w_{1,i})\right|^2\right] \\
&+ C \sup_{w}\left|u^*_{g_n}(w)\right| \sup_{h\in\mathcal{N}_{h,n}} \mathbb{E}\left[\left|h(w_{1,i}) - h_o(w_{1,i})\right|^2\right] = O_p(\xi_{0,K}\delta^2_{h,n}).
\end{aligned} \tag{3.112}$$

By Assumption 3.7, $\xi_{0,K}\delta^2_{h,n} = o(n^{-1/2})$. Hence by (3.112) we have

$$\sup_{h\in\mathcal{N}_{h,n}} \left|\mu_n \left\{\tau(w_h)(g_o(w) - g_o(w_h) - \partial_u g_o(w)(h(w_1) - h_o(w_1)))u^*_{g_n}(w)\right\}\right| = o_p(n^{-1/2}). \tag{3.113}$$

We next show that

$$\sup_{h\in\mathcal{N}_{h,n}} \left|\mu_n \left\{\tau(w_h)\partial_u g_o(w)(h(w) - h_o(w))u^*_{g_n}(w)\right\}\right| = o_p(n^{-1/2}). \tag{3.114}$$

Let $\mathcal{F}_{1,n} = \{(x, w_1) \mapsto \partial_u g_o(w)\tau(w_h)(h(w_1) - h_o(w_1))u^*_{g_n}(w) : h \in \mathcal{N}_{h,n}\}$. By Assumption 3.5, Lemmas 3.3.(b) and 3.3.(g),

$$\begin{aligned}
\sup_{f\in\mathcal{F}_{1,n}} \mathbb{E}\left[f^2\right] &= \sup_{h\in\mathcal{N}_{h,n}} \mathbb{E}\left[(\partial_u g_o(w)\tau(w_h)(h(w_1) - h_o(w_1))u^*_{g_n}(w))^2\right] \\
&\leq C \sup_{h\in\mathcal{N}_{h,n}} \mathbb{E}\left[((h(w_1) - h_o(w_1))u^*_{g_n}(w))^2\right] \\
&\leq C \sup_{w}\left|u^*_{g_n}(w)\right|^2 \sup_{h\in\mathcal{N}_{h,n}} \mathbb{E}\left[(h(w_1) - h_o(w_1))^2\right] \leq C\xi^2_{0,K}\delta^2_{h,n}.
\end{aligned} \tag{3.115}$$

Moreover, by the definition of $\mathcal{N}_{\gamma,n}$, (3.6), Assumption 3.5, Lemmas 3.3.(b) and 3.3.(g),

$$
\begin{aligned}
&\sup_{h\in\mathcal{N}_{h,n}} \left|\partial_u g_o(w)\tau(w_h)(h(w_1)-h_o(w_1))u^*_{g_n}(w)\right| \\
&\leq C\sup_{h\in\mathcal{N}_{h,n}} \left|(h(w_1)-h_o(w_1))u^*_{g_n}(w)\right| \\
&\leq C\sup_{w}\left|u^*_{g_n}(w)\right| \sup_{h\in\mathcal{N}_{h,n}} \left[|h(w_1)-h_{o,L}(w_1)|+|h_{o,L}(w_1)-h_o(w_1)|\right] \\
&\leq C\sup_{w}\left|u^*_{g_n}(w)\right| \sup_{\gamma\in\mathcal{N}_{\gamma,n}} \left[\xi_{0,K}\left\|\gamma-\gamma_{o,L}\right\|+CL^{-s_1/d_{w_1}}\right] \leq C\xi^2_{0,K}\delta_{h,n}. \qquad (3.116)
\end{aligned}
$$

By Assumption 3.7,

$$
L\xi^2_{0,K}\delta^2_{h,n}\log(n)+\xi^2_{0,K}\delta_{h,n}L^2(\log(n))^2n^{-1}=o(1). \qquad (3.117)
$$

Collecting the results in Lemma 3.4, (3.115), (3.116) and (3.117), we can use Lemma 22 of Belloni et. al (2016) to show that

$$
\sup_{h\in\mathcal{N}_{h,n}} \left|\mu_n\left\{\partial_u g_o(w)\tau(w_h)(h(w_1)-h_o(w_1))u^*_{g_n}(w)\right\}\right| = o_p(n^{-1/2}). \qquad (3.118)
$$

The asserted result follows by (3.114), (3.118) and the triangle inequality. ■

Lemma 3.7 *Under Assumptions 3.1-3.7,*

$$
\sup_{h\in\mathcal{N}_{h,n},g\in\mathcal{N}_{g,n}} \left|\mu_n\left\{\tau(w_h)(g_o(w_h)-g(w_h))u^*_{g_n}\right\}\right| = o_p(n^{-1/2}).
$$

Proof of Lemma 3.7. By the triangle inequality, (3.7), Lemmas 3.3.(e)-(f)

$$
\begin{aligned}
&\sup_{h\in\mathcal{N}_{h,n},g\in\mathcal{N}_{g,n}} \left|\mu_n\left\{\tau(w_h)(g_o(w_h)-g_{o,K}(w_h))u^*_{g_n}(w)\right\}\right| \\
&\leq CK^{-s/d}n^{-1}\sum_{i=1}^{n}\left[\left|u^*_{g_n}(w_i)\right|+\mathbb{E}\left[\left|u^*_{g_n}(w_i)\right|\right]\right] = o_p(n^{-1/2}), \qquad (3.119)
\end{aligned}
$$

where the equality is by Assumption 3.7. By the first order expansion and the Cauchy-Schwarz inequality, for any $g \in \mathcal{N}_{g,n}$,

$$\begin{aligned}
&\left|\tau(w_h)(g_{o,K}(w_h) - g(w_h) - g_{o,K}(w) + g(w))u^*_{g_n}(w)\right| \\
&= \left|\tau(w_h)(\beta - \beta_{o,K})'(P(w_h) - P(w))u^*_{g_n}(w)\right| \\
&\leq \xi_{1,K} \left\|\beta - \beta_{o,K}\right\| \left|u^*_{g_n}(w)(h(w_1) - h_o(w_1))\right| \qquad (3.120)
\end{aligned}$$

which together with the definition of $\mathcal{N}_{h,n}$, the triangle inequality and Lemmas 3.3.(a) and 3.3.(f) implies that

$$\begin{aligned}
&\sup_{h\in\mathcal{N}_{h,n},g\in\mathcal{N}_{g,n}} n^{-1}\sum_{i=1}^{n}\left|\tau(w_{i,h})(g_{o,K}(w_{i,h}) - g(w_{i,h}) - g_{o,K}(w_i) + g(w_i))u^*_{g_n}(w_i)\right| \\
&\leq \xi_{1,K}\delta_{g,n} \sup_{h\in\mathcal{N}_{h,n}} n^{-1}\sum_{i=1}^{n}\left|u^*_{g_n}(w_i)(h(w_{1,i}) - h_o(w_{1,i}))\right| \\
&\leq \xi_{1,K}\delta_{g,n} \sup_{h\in\mathcal{N}_{h,n}} \left(n^{-1}\sum_{i=1}^{n}\left|u^*_{g_n}(w_i)\right|^2 n^{-1}\sum_{i=1}^{n}\left|h(w_{1,i}) - h_o(w_{1,i})\right|^2\right)^{1/2} \\
&= O_p(\xi_{1,K}\delta_{g,n}\delta_{h,n}) = o_p(n^{-1/2}) \qquad (3.121)
\end{aligned}$$

where the equality is by Assumption 3.7. Similarly, we can show that

$$\sup_{h\in\mathcal{N}_{h,n},g\in\mathcal{N}_{g,n}} \mathbb{E}\left[\left|\tau(w_{i,h})(g_{o,K}(w_{i,h}) - g(w_{i,h}) - g_{o,K}(w_i) + g(w_i))u^*_{g_n}(w_i)\right|\right] = o(n^{-1/2}), \qquad (3.122)$$

which together with (3.121) implies that

$$\sup_{h\in\mathcal{N}_{h,n},g\in\mathcal{N}_{g,n}} \left|\mu_n\left\{\tau(w_h)(g_{o,K}(w_h) - g(w_h) - g_{o,K}(w) + g(w))u^*_{g_n}(w)\right\}\right| = o_p(n^{-1/2}). \qquad (3.123)$$

Recall that $\mathcal{F}_{2,n} = \{(x, w_1) \mapsto \tau(w_h)u^*_{g_n}(w)P(w)'\alpha : h \in \mathcal{N}_{h,n},\ \alpha \in \mathbb{S}^{K-1}\}$, where $\mathbb{S}^{K-1} = \{\alpha \in \mathbb{R}^K : \alpha'\alpha = 1\}$. By Lemma 3.3.(g) and $\tau(w)^2 = \tau(w)$,

$$\begin{aligned}
\sup_{f\in\mathcal{F}_{2,n}} \mathbb{E}\left[f^2\right] &= \sup_{h\in\mathcal{N}_{h,n},\alpha\in\mathbb{S}^{K-1}} \mathbb{E}\left[(\tau(w_h)u^*_{g_n}(w)P(w)'\alpha)^2\right] \\
&\leq \sup_{w}(u^*_{g_n}(w))^2 \sup_{\alpha\in\mathbb{S}^{K-1}} \mathbb{E}\left[(\tau(w)P(w)'\alpha)^2\right] \leq C\xi^2_{0,K}. \qquad (3.124)
\end{aligned}$$

Similarly,

$$\sup_{h\in\mathcal{N}_{h,n},\alpha\in\mathbb{S}^{K-1}} \left|\tau(w_h)u^*_{g_n}(w)P(w)'\alpha\right| \leq \sup_{\alpha\in\mathbb{S}^{K-1}} \left|u^*_{g_n}(w)P(w)'\alpha\right| \leq C\xi^2_{0,K}. \qquad (3.125)$$

Collecting the results in Lemma 3.5, (3.124) and (3.125), we can use Lemma 22 of Belloni et. al (2016) to show that

$$\sup_{h\in\mathcal{N}_{h,n},\alpha\in\mathbb{S}^{K-1}} \left|\mu_n\left\{\tau(w_h)u^*_{g_n}(w)P(w)'\alpha\right\}\right| = O_p((L+K)^{1/2}\xi_{0,K}(\log(n))^{1/2}n^{-1/2}). \tag{3.126}$$

By the definition of $\mathcal{N}_{g,n}$ and (3.126),

$$\begin{aligned}
&\sup_{h\in\mathcal{N}_{h,n},g\in\mathcal{N}_{g,n}} \left|\mu_n\left\{\tau(w_h)(g_{o,K}(w)-g(w))u^*_{g_n}(w)\right\}\right| \\
&\leq \sup_{h\in\mathcal{N}_{h,n},\alpha\in\mathbb{S}^{K-1}} \left|\mu_n\left\{\tau(w_h)u^*_{g_n}(w)P(w)'\alpha\right\}\right| \sup_{\beta\in\mathcal{N}_{\beta,n}} \|\beta-\beta_{o,K}\| \\
&= O_p(\delta_{g,n}(L+K)^{1/2}\xi_{0,K}(\log(n))^{1/2}n^{-1/2}) = o_p(n^{-1/2})
\end{aligned} \tag{3.127}$$

where the second equality is by Assumption 3.7. Collecting the results in (3.119), (3.123) and (3.127), and applying the triangle inequality, we immediately prove the asserted result. ■

4 Extra Simulation Results

In this section, we study the finite sample performance of the two-step nonparametric M estimator and the proposed inference method when the nonparametric regressor may have unbounded support. The simulated data is from the following model

$$y_i = w_{1,i}\theta_o + m_o(h_o(x_i)) + u_i, \tag{4.1}$$

$$s_i = h_o(x_i) + \varepsilon_i, \tag{4.2}$$

where $\theta_o = 1$; $h_o(x) = 2\cos(\pi x)$, $m_o(w_2) = \sin(\pi w_2)$ and $w_2 = h_o(x)$. For $i = 1, \ldots, n$, we independently draw $(w_{1,i}, x_{*,i}, u_i, \varepsilon_i)'$ from $N(0, I_4)$ and then calculate

$$x_i = 2^{-1/2}(w_{1,i} + x_{*,i}). \tag{4.3}$$

The data $\{y_i, s_i, w_{1,i}, x_i\}_{i=1}^n$ are generated using the equations in (4.1) and (4.2).

The first-step and second-step nonparametric estimators and the consistent variance estimator take the same forms as their counterparts in Section 7 of HLR and hence are omitted here. We consider sample sizes n = 100, 250 and 500 in this simulation study. For each sample size, we generate 10000 simulated samples to evaluate the performances of the two-step sieve estimator and the proposed inference procedure. For each simulated sample, we calculate the sieve estimator of (θ_o, m_o), and the 0.90 confidence interval of θ_o for each combination of (L, K) where $L = 2, \ldots, 16$ and $K = 2, \ldots, 21$. The simulation results are reported in Figures 4.1 and 4.2.

The properties of the two-step sieve M estimator and the proposed confidence interval are similar to what we found in the other DGP employed in HLR. We list some important differences. First, when the unknown function estimated in the first-step has unbounded support, the optimal L which produces a two-step M estimator with the smallest MSE is much larger. Second, the ratio between the MSE of the cross-validated estimator of m_o and the optimal MSE does not seem to converge to 1 in all the sample sizes we considered. However, the MSE of the cross-validated estimator of θ_o does approach the optimal value quickly as the sample size increases. Third, when L is small (e.g., $L = 4$), the proposed confidence interval over-covers the unknown parameter θ_o and its length diverges with increasing K. Fourth, the coverage probability of the confidence interval based on the cross-validated sieve estimator is almost identical to the nominal level even when the sample size is small (e.g., $n = 100$).

5 Consistency and Convergence Rate

In this appendix, we first derive the consistency of the second-step sieve M estimator $\widehat{g}_n$ under the metric $\|\cdot\|_{\mathcal{G}}$ defined on $\mathcal{G}$. Given the consistency, we then focus on a local neighborhood of g_o to calculate the convergence rate of $\widehat{g}_n$. Under mild conditions, the first-step sieve M estimator $\widehat{h}_n$ is consistent (see, e.g., Theorem 3.1 of Chen, 2007), and also has rate of convergence under a pseudo-metric $\|\cdot\|_{\mathcal{H}}$.[2] Let $\delta^*_{h,n} = O(\varepsilon_{1,n})$ be a small positive number that goes to zero as $n \to \infty$. Without loss of generality we denote $||\widehat{h}_n - h_o||_{\mathcal{H}} = O_p(\delta^*_{h,n})$ as the convergence rate. Hence we can assume that $\widehat{h}_n$ belongs to a shrinking neighborhood $\mathcal{N}_{h,n} = \{h \in \mathcal{H}_n : \|h - h_o\|_{\mathcal{H}} \leq \delta_{h,n}\}$ of h_o wpa1, where $\delta_{h,n} = \delta^*_{h,n} \log(\log(n)) = o(1)$.

5.1 Consistency of the second step sieve M estimation

The following conditions are sufficient for the consistency of $\widehat{g}_n$ under $\|\cdot\|_{\mathcal{G}}$.

Assumption 5.1 *(i)* $\mathbb{E}\left[\psi\left(Z_2, g_o, h_o\right)\right] > -\infty$ *and if* $\mathbb{E}\left[\psi\left(Z_2, g_o, h_o\right)\right] = \infty$*, then* $\mathbb{E}\left[\psi\left(Z_2, g, h_o\right)\right] < \infty$ *for all* $g \in \mathcal{G}_n \backslash \{g_o\}$ *and for all* $n \geq 1$*; (ii) for all* $\varepsilon > 0$*, there exists some non-increasing positive sequence*

[2] See, e.g., Shen and Wong (1994) and Chen and Shen (1998) for the convergence rate of the one-step (approximate) sieve M estimator for i.i.d. data and weakly dependent data respectively.

Figure 4.1. The Mean Squared Errors of the Two-step Sieve M Estimators of m_o and θ_o (DGP2)

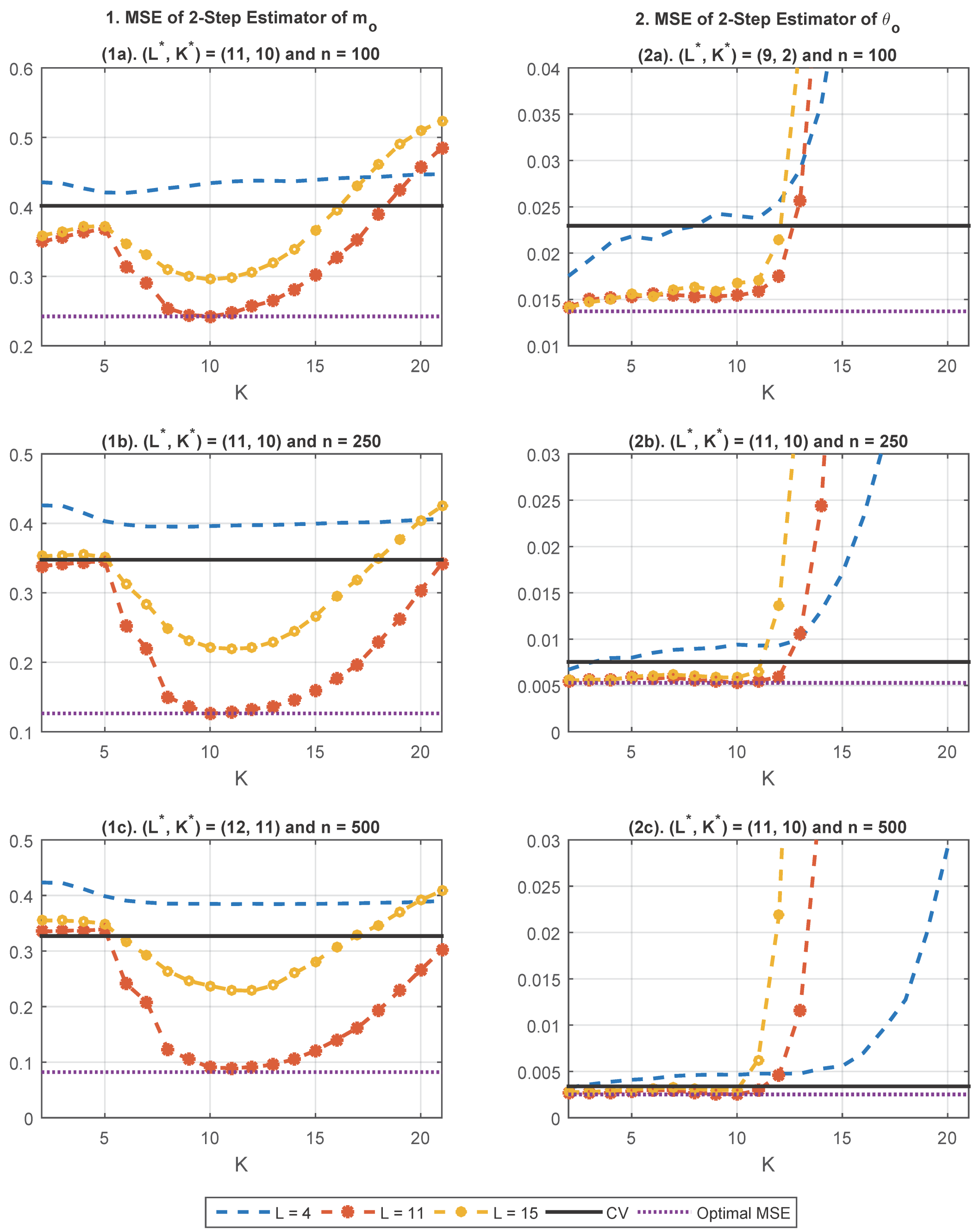

Figure 4.1: 1. The left panel represents the MSEs of the two-step sieve estimator of m_o for sample sizes n=100, 250 and 500 respectively; 2. the right panel represents the MSEs of the two-step sieve estimator of θ_o for sample sizes n=100, 250 and 500 respectively; 3. L^* and K^* denote the numbers of the series terms which produce sieve estimator of m_o with the smallest finite sample MSE (in the left panel) or sieve estimator of θ_o with the smallest finite sample MSE (in the left panel); 4. the dotted line represents the MSE of the two-step sieve M estimator with $L = L^*$ and $K = K^*$; 5. the solid line represents the MSE of the two-step sieve M estimator with L and K selected by 5-fold cross-validation.

Figure 4.2. The Convergence Probability and the Average Length of the Confidence Interval of θ_o (DGP2)

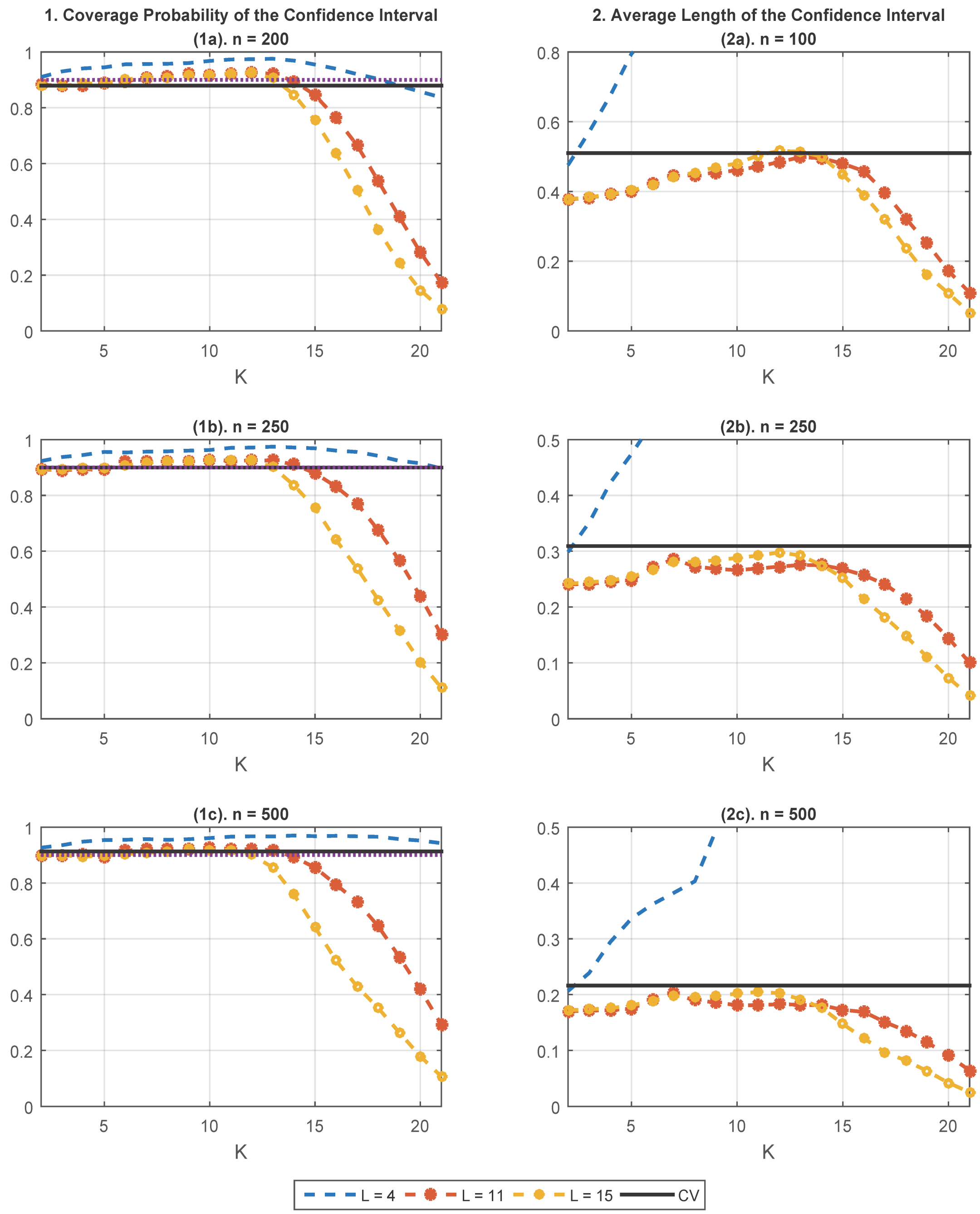

Figure 4.2: 1. The left panel presents the coverage probability of the confidence interval of θ_o for sample sizes n=100, 250 and 500 respectively; 2. the right panel presents the average length of the confidence interval of θ_o for sample sizes n=100, 250 and 500 respectively; 3. the dotted line in the left panel is the 0.90 line which represents the nominal coverage of the confidence interval; 4. the solid line represents the coverage probability of the confidence interval based on the two-step sieve estimator with K and L selected by 5-fold cross-validation.

$c_n(\varepsilon)$ such that for all $n \geq 1$

$$\mathbb{E}\left[\psi\left(Z_2, g_o, h_o\right)\right] - \sup_{\{g \in \mathcal{G}_n:\ \|g - g_o\|_{\mathcal{G}} \geq \varepsilon\}} \mathbb{E}\left[\psi\left(Z_2, g, h_o\right)\right] \geq c_n(\varepsilon) \tag{5.1}$$

and $\liminf_n c_n(\varepsilon) > 0$ for all $\varepsilon > 0$.

Assumption 5.1 is the identification uniqueness condition for g_o. For sieve M estimation a similar condition can be found in White and Wooldridge (1991). This assumption is stronger than Condition 3.1 of Theorem 3.1 in Chen (2007) and Condition a of Lemma A.2 in Chen and Pouzo (2012), because it requires $c_n(\varepsilon)$ to be bounded away from zero for all large n. It essentially requires that the second step sieve M estimation is well-posed under the strong metric $\|\cdot\|_{\mathcal{G}}$.

Assumption 5.2 *(i) $g_o \in \mathcal{G}$ and $\|\cdot\|_{\mathcal{G}}$ is a metric defined on $\mathcal{G}$ or some metric space containing $\mathcal{G}$; (ii) $\mathcal{G}_n \subset \mathcal{G}_{n+1} \subset \mathcal{G}$ for all $n \geq 1$ and there exists some $g_n \in \mathcal{G}_n$ such that*

$$\left|\mathbb{E}\left[\psi(Z_2, g_n, h_o) - \psi(Z_2, g_o, h_o)\right]\right| = O(\eta_{2,n}) \tag{5.2}$$

where $\eta_{2,n}$ is some finite positive non-increasing sequence.

Assumption 5.2 imposes conditions on the sieve spaces. It is essentially Condition b of Lemma A.2 in Chen and Pouzo (2012). It is also implied by Conditions 3.2 and 3.3 of Theorem 3.1 in Chen (2007). The condition in (5.2) is clearly implied by the convergence rate of the sieve approximation error of $\|g_n - g_o\|_{s,2}$ and the continuity of the criterion function $\mathbb{E}\left[\psi\left(Z_2, g, h_o\right)\right]$ for all $g \in \mathcal{G}_n$ in the local neighborhood of g_o. In the following we denote $\mu_n\left[\psi\left(Z_2, g, h\right)\right] \equiv \frac{1}{n}\sum_{i=1}^{n}\left\{\psi\left(Z_{2,i}, g, h\right) - \mathbb{E}\left[\psi\left(Z_2, g, h\right)\right]\right\}$.

Assumption 5.3 *(i) $\sup_{g \in \mathcal{G}_n, h \in \mathcal{N}_{h,n}} \left|\mu_n\left[\psi(Z_2, g, h)\right]\right| = O_p(\eta_{0,n})$ where $\{\eta_{0,n}\}$ is some finite positive non-increasing sequence going to zero; (ii) there is a finite positive non-increasing sequence $\{\eta_{1,n}\}$ going to zero such that*

$$\sup_{g \in \mathcal{G}_n, h \in \mathcal{N}_{h,n}} \left|\mathbb{E}\left[\psi(Z_2, g, h) - \psi\left(Z_2, g, h_o\right)\right]\right| = O(\eta_{1,n}).$$

Assumption 5.3 is similar to Condition 3.5 of Theorem 3.1 in Chen (2007) and the first part of Condition d of Lemma A.2 in Chen and Pouzo (2012). Assumption 5.3.(i) can be verified by applying a standard empirical process result. Assumption 5.3.(ii) can be verified by the convergence rate of the first-step sieve M estimator $\widehat{h}_n$ and the continuity of the criterion function $\mathbb{E}\left[\psi\left(Z_2, g, h\right)\right]$ in $h \in \mathcal{N}_{h,n}$ uniformly over $g \in \mathcal{G}_n$.

Theorem 5.1 *Let Assumptions 5.1, 5.2 and 5.3 hold. If*

$$\max\left\{\eta_{0,n}, \eta_{1,n}, \eta_{2,n}, \varepsilon_{2,n}^2\right\} = o(1) \tag{5.3}$$

then the second-step sieve M estimator is consistent under $\|\cdot\|_{\mathcal{G}}$*, i.e.* $\|\widehat{g}_n - g_o\|_{\mathcal{G}} = o_p(1)$.

Proof of Theorem 5.1. Let $Q_n(g,h) \equiv \frac{1}{n}\sum_{i=1}^n \psi(Z_{2,i}, g, h)$ and $Q(g,h) \equiv \mathbb{E}[\psi(Z_2, g, h)]$. Let $I_n(\varepsilon) \equiv \Pr\left(\|\widehat{g}_n - g_o\|_{\mathcal{G}} > \varepsilon\right)$. For any $\varepsilon > 0$, by the definition of $\widehat{g}_n$, we have

$$I_n(\varepsilon) \leq \Pr\left(\sup_{\{g\in\mathcal{G}_n:\ ||g-g_o||_{\mathcal{G}}\geq\varepsilon\}} Q_n(g,\widehat{h}_n) \geq Q_n(g_n,\widehat{h}_n) - O_p\left(\varepsilon_{2,n}^2\right)\right). \tag{5.4}$$

Rewrite the inequality inside the parentheses on the RHS as

$$-\left[Q_n(g_n,\widehat{h}_n) - Q(g_o,h_o)\right] + O_p\left(\varepsilon_{2,n}^2\right) \geq Q(g_o,h_o) - \sup_{\{g\in\mathcal{G}_n:\ ||g-g_o||_{\mathcal{G}}\geq\varepsilon\}} Q_n(g,\widehat{h}_n). \tag{5.5}$$

Note that the first two terms on the LHS of the above inequality can be rewritten as

$$\begin{aligned}
&-\left[Q_n(g_n,\widehat{h}_n) - Q(g_o,h_o)\right]\\
&= -\mu_n\left[\psi(Z_2,g_n,\widehat{h}_n)\right] - \left[Q(g_n,\widehat{h}_n) - Q(g_n,h_o)\right] - \left[Q(g_n,h_o) - Q(g_o,h_o)\right]
\end{aligned}$$

which implies that if $\widehat{h}_n \in \mathcal{N}_{h,n}$ with probability approaching 1 (wpa1), then

$$-\left[Q_n(g_n,\widehat{h}_n) - Q(g_o,h_o)\right] \leq I_{1,n} + I_{2,n} + I_{3,n}, \tag{5.6}$$

where

$$\begin{aligned}
I_{1,n} &\equiv \sup_{g\in\mathcal{G}_n, h\in\mathcal{N}_{h,n}} \left|\mu_n\left[\psi(Z_2,g,h)\right]\right|,\\
I_{2,n} &\equiv \sup_{g\in\mathcal{G}_n, h\in\mathcal{N}_{h,n}} \left|Q(g,h) - Q(g,h_o)\right|,\\
I_{3,n} &\equiv \left|Q(g_n,h_o) - Q(g_o,h_o)\right|.
\end{aligned}$$

Similarly if $\widehat{h}_n \in \mathcal{N}_{h,n}$ wpa1, then for any $g \in \mathcal{G}_n$,

$$\begin{aligned} Q_n(g, \widehat{h}_n) &= \mu_n \left[\psi(Z_2, g, \widehat{h}_n)\right] + \left[Q(g, \widehat{h}_n) - Q(g, h_o)\right] + Q(g, h_o) \\ &\leq \sup_{g \in \mathcal{G}_n, h \in \mathcal{N}_{h,n}} |\mu_n [\psi(Z_2, g, h)]| + \sup_{g \in \mathcal{G}_n, h \in \mathcal{N}_{h,n}} |Q(g,h) - Q(g, h_o)| + Q(g, h_o) \\ &= I_{1,n} + I_{2,n} + Q(g, h_o). \end{aligned} \tag{5.7}$$

Therefore when $\widehat{h}_n \in \mathcal{N}_{h,n}$ wpa1, we may note that the term on the RHS of (5.5) is such that

$$\begin{aligned} &Q(g_o, h_o) - \sup_{\{g \in \mathcal{G}_n \colon ||g - g_o||_{\mathcal{G}} \geq \varepsilon\}} Q_n(g, \widehat{h}_n) \\ &\geq -I_{1,n} - I_{2,n} + Q(g_o, h_o) - \sup_{\{g \in \mathcal{G}_n \colon ||g - g_o||_{\mathcal{G}} \geq \varepsilon\}} Q(g, h_o). \end{aligned} \tag{5.8}$$

From (5.4), (5.5), (5.6) and (5.8), we get

$$I_n(\varepsilon) \leq \Pr\left(2\sum_{j=1}^{3} I_{j,n} + O_p(\varepsilon_{2,n}^2) \geq Q(g_o, h_o) - \sup_{\{g \in \mathcal{G}_n \colon ||g - g_o||_{\mathcal{G}} \geq \varepsilon\}} Q(g, h_o)\right) + \Pr\left(\widehat{h}_n \notin \mathcal{N}_{h,n}\right). \tag{5.9}$$

If $Q(g_o, h_o) = \infty$, then using Assumption 5.1.(i), we have

$$Q(g_o, h_o) - \sup_{\{g \in \mathcal{G}_n \colon ||g - g_o||_{\mathcal{G}} \geq \varepsilon\}} Q(g, h_o) = \infty. \tag{5.10}$$

However, from Assumption 5.2.(ii) and 5.3, we get $\max\{I_{1,n}, I_{2,n}, I_{3,n}\} = O_p(1)$, which together with (5.9), (5.10), $\varepsilon_{2,n} = o(1)$ and the definition of $\mathcal{N}_{h,n}$ implies that

$$I_n(\varepsilon) \leq \Pr\left(\widehat{h}_n \notin \mathcal{N}_{h,n}\right) \to 0 \text{ as } n \to \infty.$$

On the other hand, if $Q(g_o, h_o) < \infty$, then using (5.9) and Assumption 5.1.(ii), we get

$$I_n(\varepsilon) \leq \Pr\left(\frac{2I_{1,n} + 2I_{2,n} + 2I_{3,n} + O_p(\varepsilon_{2,n}^2)}{c_n(\varepsilon)} \geq 1\right) + \Pr\left(\widehat{h}_n \notin \mathcal{N}_{h,n}\right). \tag{5.11}$$

Assumption 5.1.(ii), Assumption 5.2.(ii), Assumption 5.3 and the condition (5.3) imply that

$$\frac{2I_{1,n} + 2I_{2,n} + 2I_{3,n} + O_p(\varepsilon_{2,n}^2)}{c_n(\varepsilon)} = o_p(1)$$

for any $\varepsilon > 0$. Combining this result with (5.11) and the definition of $\mathcal{N}_{h,n}$, we conclude that $I_n(\varepsilon) \to 0$ as n goes to infinity. This finishes the proof. ∎

5.2 Rate of convergence of the second step sieve M estimation

After the consistency of the second-step sieve M estimator $\widehat{g}_n$ is established, we can focus on the local neighborhood of g_o to compute the convergence rate of $\widehat{g}_n$ under $\|\cdot\|_{\mathcal{G}}$. Let K_2 be a generic finite and positive constant and define

$$\mathcal{N}_{2,K_2} \equiv \{g \in \mathcal{G}_n : ||g - g_o||_{\mathcal{G}} \leq K_2\},$$

then by the consistency of $\widehat{g}_n$, we have $\widehat{g}_n \in \mathcal{N}_{2,K_2}$ wpa1. Moreover, given the convergence rate $\delta^*_{1,n}$ of the first-step sieve M estimator $\widehat{h}_n$, we can define

$$\mathcal{N}_{1,K_1} \equiv \left\{h \in \mathcal{H}_n : ||h - h_o||_{\mathcal{H}}/\delta^*_{h,n} \leq K_1\right\}$$

such that for any small constant $\omega > 0$, there is a finite constant $K_\omega > 0$ such that

$$\Pr(\widehat{h}_n \notin \mathcal{N}_{1,K_\omega}) \leq \omega \text{ for all } n. \tag{5.12}$$

The following general conditions are sufficient for deriving the convergence rate of $\widehat{g}_n$.

Assumption 5.4 *There are some finite, positive and non-increasing sequences $\delta_{1,n}$, $\delta_{2,n}$ and δ_n that go to zero as $n \to \infty$ such that the following hold for any fixed finite constants $K_1 > 0$, $K_2 > 0$: (i)*

$$\sup_{h \in \mathcal{N}_{1,K_1}} |\mathbb{E}[\psi(Z_2, g_n, h) - \psi(Z_2, g_o, h)]| = O(\delta^2_{2,n}); \tag{5.13}$$

(ii) for any small constant $\delta, \widetilde{\delta} > 0$ and for any $g \in \mathcal{N}_{2,K_2}$ with $0 < \widetilde{\delta} < \|g - g_o\|_{\mathcal{G}} < \delta$

$$\sup_{h \in \mathcal{N}_{1,K_1}} \mathbb{E}[\psi(Z_2, g, h) - \psi(Z_2, g_o, h)] \leq c_{K_1,1}\delta_{1,n}\delta - c_{K_1,2}\delta^2, \tag{5.14}$$

where $c_{K_1,1}$ and $c_{K_1,2} > 0$ are finite constants only depending on K_1; (iii)

$$\sup_{g \in \mathcal{N}_{2,K_2}, h \in \mathcal{N}_{1,K_1}} |\mu_n[\psi(Z, g, h) - \psi(Z, g, h_o)]| = O_p(\delta^2_n); \tag{5.15}$$

(iv) for all n large enough and for any sufficiently small δ,

$$\mathbb{E}\left[\sup_{\left\{g\in\mathcal{N}_{2,K_2}:\ \|g-g_o\|_{\mathcal{G}}\leq\delta\right\}}\left|\mu_n\left[\psi(Z,g,h_o)-\psi(Z,g_o,h_o)\right]\right|\right]\leq\frac{c_1\phi_n(\delta)}{\sqrt{n}} \tag{5.16}$$

where $c_1 > 0$ is some finite constant and $\phi_n(\cdot)$ is some function such that $\delta^{-\gamma}\phi_n(\delta)$ is a decreasing function for some $\gamma \in (0, 2)$.

Assumption 5.4.(i) imposes local smoothness condition on the function $\mathbb{E}\left[\psi\left(Z_2,\cdot,h\right)\right]$ uniformly over h in some shrinking neighborhood. The rate $\delta_{2,n}$ is determined by the convergence rates of the sieve approximation error of g_o and the first step sieve estimator $\widehat{h}_n$. Assumption 5.4.(ii) is a local identification condition. The term $\delta_{1,n}$ on the right side of the inequality (5.14) represents the effect of first-step estimation on the second-step sieve estimate $\widehat{g}_n$. In Assumption 5.4.(i), (ii) and (iii), the uniform convergence is imposed over local neighborhoods $\mathcal{N}_{1,K_1}$ and/or $\mathcal{N}_{2,K_2}$. That is particularly useful for establishing the convergence rate of $\widehat{g}_n$, because by the consistency of $\widehat{g}_n$ and the convergence rate of $\widehat{h}_n$, we can bound the probabilities of the events $\{\widehat{g}_n \notin \mathcal{N}_{2,K_2}\}$ and $\{\widehat{h}_n \notin \mathcal{N}_{1,K_1}\}$ in finite samples by choosing sufficiently large K_1 and K_2. Assumption 5.4.(iv) is a stochastic equicontinuity condition which is similar to the one in Theorem 3.4.1 of Van der Vaart and Wellner (1996).

Theorem 5.2 *Suppose that the conditions in Theorem 5.1 and Assumption 5.4 are satisfied. Furthermore, if $\|g_n - g_o\|_{\mathcal{G}} = O(\delta^*_{2,n})$ where $\delta^*_{2,n}$ is defined below and there is a finite, positive and non-increasing sequence $\delta_{g,n}$ such that*

$$\left(\delta_{g,n}\right)^{-2}\phi_n(\delta_{g,n})\leq c_2\sqrt{n}, \tag{5.17}$$

then we have $\|\widehat{g}_n - g_o\|_{\mathcal{G}} = O_p\left(\delta^*_{2,n}\right)$, where $\delta^*_{2,n} \equiv \max\left\{\delta_{1,n}, \delta_{2,n}, \delta_n, \delta_{g,n}, \varepsilon_{2,n}\right\}$.

Proof of Theorem 5.2. Let $\omega > 0$ be some arbitrarily small constant. Because $\widehat{g}_n$ is consistent, we can choose a sufficiently large constant $K_M > 0$ such that

$$\Pr\left(||\widehat{g}_n-g_o||_{\mathcal{G}}>K_M\right)\leq\omega. \tag{5.18}$$

By $\|g_n - g_o\|_{\mathcal{G}} = o(1)$, we deduce that there is some sufficiently large K_{g_o} such that $||g_n - g_o||_{\mathcal{G}} \leq K_{g_o}$. Let $K^*_M = \max\{K_M, K_{g_o}\}$,

$$\mathcal{G}_n(M)\equiv\left\{g\in\mathcal{G}_n:2^M\delta^*_{2,n}<||g-g_o||_{\mathcal{G}}\leq K^*_M\right\}$$

and $I_{M,n}(\omega) \equiv \Pr\left(||\widehat{g}_n - g_o||_{\mathcal{G}} > 2^M \delta^*_{2,n}\right)$. Note that by (5.18), we have

$$I_{M,n}(\omega) = \Pr\left(\widehat{g}_n \in \mathcal{G}_n(M)\right) + \Pr\left(||\widehat{g}_n - g_o||_{\mathcal{G}} > K^*_M\right) \leq \Pr\left(\widehat{g}_n \in \mathcal{G}_n(M)\right) + \omega. \tag{5.19}$$

We will prove that

$$I_{M,n}(\omega) \leq \sum_{j \geq M, 2^{j-1}\delta^*_{2,n} \leq K^*_M} \frac{c_1 c_2 \left[(2^{j+1})^\gamma + K^\gamma_\varepsilon\right]}{|c_{K_1,2} 2^{2j} - K - c_{K_1,1} 2^j|} + 5\omega \tag{5.20}$$

where c_1 and c_2 are defined in Assumption 5.4.(iv) and (5.17), $c_{K_1,1}$, $c_{K_1,2}$, K_ε and K are some fixed finite constants which may depend on ω, and $\gamma \in (0,2)$ is defined in Assumption 5.4.(iv). As $\gamma < 2$, we can choose M sufficiently large such that

$$\sum_{j \geq M, 2^{j-1}\delta^*_{2,n} \leq K^*_M} \frac{c_1 c_2 \left[(2^{j+1})^\gamma + K^\gamma_\varepsilon\right]}{|c_{K_1,2} 2^{2j} - K - c_{K_1,1} 2^j|} < \omega,$$

which together with (5.20) implies that $I_{M,n}(\omega) \leq 6\omega$. As we can let ω arbitrarily small, this would establish that $||\widehat{g}_n - g_o||_{\mathcal{G}} = O_p\left(\delta^*_{2,n}\right)$. Equation (5.20) is established by combining (5.21), (5.31) and (5.33) below, which are proved in several steps.

<u>Step 1</u>: We prove that

$$I_{M,n}(\omega) \leq \Pr\left(\sup_{g \in \mathcal{G}_n(M), h \in \mathcal{N}_{1,K_1}} \left[I_{1,n}(g, h_o) + I_{2,n}(g,h)\right] + K\delta^{*2}_{2,n} \geq 0\right) + 5\omega \tag{5.21}$$

where K_1 is a fixed constant such that $\Pr\left(\widehat{h}_n \notin \mathcal{N}_{1,K_1}\right) \leq \omega$ for all n, K is some fixed constant defined below,

$$I_{1,n}(g, h_o) \equiv \mu_n\left[\psi(Z_2, g, h_o) - \psi(Z_2, g_n, h_o)\right],$$
$$\text{and } I_{2,n}(g,h) \equiv Q(g,h) - Q(g_o,h).$$

For this purpose, we first note that by the definition of $\widehat{g}_n$, we can choose some sufficiently large constant $K_1 > 0$ such that

$$\Pr\left(Q_n(\widehat{g}_n, \widehat{h}_n) - Q_n(g_n, \widehat{h}_n) + K_1\varepsilon^2_{2,n} < 0\right) \leq \omega. \tag{5.22}$$

Combining (5.19) and (5.22), we have

$$I_{M,n}(\omega) \leq \Pr\left(\sup_{g \in \mathcal{G}_n(M)} Q_n(g, \widehat{h}_n) - Q_n(g_n, \widehat{h}_n) + K_1 \varepsilon_{2,n}^2 \geq 0\right) + 2\omega. \tag{5.23}$$

It is clear that the term inside the parentheses on the RHS of (5.23) is such that

$$\begin{aligned}
&Q_n(g, \widehat{h}_n) - Q_n(g_n, \widehat{h}_n) \\
&= \mu_n\left[\psi(Z_2, g, \widehat{h}_n) - \psi(Z_2, g_n, \widehat{h}_n)\right] + Q(g, \widehat{h}_n) - Q(g_n, \widehat{h}_n) \\
&= \mu_n\left[\psi(Z_2, g, \widehat{h}_n) - \psi(Z_2, g, h_o)\right] + \mu_n\left[\psi(Z_2, g_n, h_o) - \psi(Z_2, g_n, \widehat{h}_n)\right] \\
&+ \mu_n\left[\psi(Z_2, g, h_o) - \psi(Z_2, g_n, h_o)\right] + Q(g, \widehat{h}_n) - Q(g_o, \widehat{h}_n) \\
&+ Q(g_o, \widehat{h}_n) - Q(g_n, \widehat{h}_n),
\end{aligned}$$

and therefore,

$$\begin{aligned}
&Q_n(g, \widehat{h}_n) - Q_n(g_n, \widehat{h}_n) \\
&= \mu_n\left[\psi(Z_2, g, \widehat{h}_n) - \psi(Z_2, g, h_o)\right] + \mu_n\left[\psi(Z_2, g_n, h_o) - \psi(Z_2, g_n, \widehat{h}_n)\right] \\
&+ Q(g_o, \widehat{h}_n) - Q(g_n, \widehat{h}_n) + I_{1,n}(g, h_o) + I_{2,n}(g, \widehat{h}_n).
\end{aligned} \tag{5.24}$$

From Assumption 5.4.(iii), we can choose some constant K_2 sufficiently large such that

$$\begin{aligned}
&\Pr\left(\sup_{g \in \mathcal{G}_n(M)} \mu_n\left[\psi(Z_2, g, \widehat{h}_n) - \psi(Z_2, g, h_o)\right] \geq K_2 \delta_n^2, \widehat{h}_n \in \mathcal{N}_{1,K_1}\right) \\
&\leq \Pr\left(\sup_{g \in \mathcal{N}_{2,K_M^*}, h \in \mathcal{N}_{1,K_1}} \left|\mu_n\left[\psi(Z_2, g, h) - \psi(Z_2, g, h_o)\right]\right| \geq K_2 \delta_n^2\right) \leq \omega.
\end{aligned} \tag{5.25}$$

Combining (5.23), (5.24), and (5.25), we obtain

$$I_{M,n}(\omega) \leq \Pr\left[\left(\begin{array}{c} \mu_n\left[\psi(Z_2, g_n, h_o) - \psi(Z_2, g_n, \widehat{h}_n)\right] \\ +Q(g_o, \widehat{h}_n) - Q(g_n, \widehat{h}_n) \\ +\sup_{g \in \mathcal{G}_n(M)}\left[I_{1,n}(g, h_o) + I_{2,n}(g, \widehat{h}_n)\right] \\ +K_1 \varepsilon_{2,n}^2 + K_2 \delta_n^2 \end{array}\right) \geq 0, \widehat{h}_n \in \mathcal{N}_{1,K_1}\right] + 4\omega. \tag{5.26}$$

By the definition of $\mathcal{N}_{2,K_M^*}$, we have $g_n \in \mathcal{N}_{2,K_M^*}$, which together with Assumption 5.4.(iii) implies that

$$\begin{aligned}
&\Pr\left(\mu_n\left[\psi(Z_2,g_n,h_o)-\psi(Z_2,g_n,\widehat{h}_n)\right]\geq K_2\delta_n^2,\widehat{h}_n\in\mathcal{N}_{1,K_1}\right)\\
&\leq \Pr\left(\sup_{g\in\mathcal{N}_{2,K_M^*},h\in\mathcal{N}_{1,K_1}}\left|\mu_n\left[\psi(Z_2,g,h_o)-\psi(Z_2,g,h)\right]\right|\geq K_2\delta_n^2\right)\leq\omega. \qquad (5.27)
\end{aligned}$$

By the same argument that led to (5.26), we obtain

$$I_{M,n}(\omega)\leq \Pr\left[\left(\begin{array}{c} Q(g_o,\widehat{h}_n)-Q(g_n,\widehat{h}_n)\\ +\sup_{g\in\mathcal{G}_n(M)}\left[I_{1,n}(g,h_o)+I_{2,n}(g,\widehat{h}_n)\right]\\ +K_1\varepsilon_{2,n}^2+2K_2\delta_n^2\end{array}\right)\geq 0,\widehat{h}_n\in\mathcal{N}_{1,K_1}\right]+5\omega. \qquad (5.28)$$

From Assumption 5.4.(i), we can choose some constant K_3 sufficiently large such that

$$\sup_{h\in\mathcal{N}_{1,K_1}}\left|\mathbb{E}\left[\psi\left(Z_2,g_o,h\right)-\psi\left(Z_2,g_n,h\right)\right]\right|<K_3\delta_{2,n}^2$$

which implies that

$$\begin{aligned}
&\Pr\left(Q(g_o,\widehat{h}_n)-Q(g_n,\widehat{h}_n)\geq K_3\delta_{2,n}^2,\widehat{h}_n\in\mathcal{N}_{1,K_1}\right)\\
&\leq \Pr\left(\sup_{h\in\mathcal{N}_{1,K_1}}\left|\mathbb{E}\left[\psi\left(Z_2,g_o,h\right)-\psi\left(Z_2,g_n,h\right)\right]\right|\geq K_3\delta_{2,n}^2\right)=0. \qquad (5.29)
\end{aligned}$$

By the same argument that led to (5.28), we obtain

$$I_{M,n}(\omega)\leq \Pr\left[\left(\begin{array}{c} \sup_{g\in\mathcal{G}_n(M)}\left[I_{1,n}(g,h_o)+I_{2,n}(g,\widehat{h}_n)\right]\\ +K_1\varepsilon_{2,n}^2+2K_2\delta_n^2+K_3\delta_{2,n}^2\end{array}\right)\geq 0,\widehat{h}_n\in\mathcal{N}_{1,K_1}\right]+5\omega. \qquad (5.30)$$

Recalling $\delta_{2,n}^*\equiv\max\{\delta_{1,n},\delta_n,\delta_{2,n},\delta_{g,n},\varepsilon_{2,n}\}$, we obtain

$$I_{M,n}(\omega)\leq \Pr\left(\sup_{g\in\mathcal{G}_n(M),h\in\mathcal{N}_{1,K_1}}\left[I_{1,n}(g,h_o)+I_{2,n}(g,h)\right]+K\delta_{2,n}^{*2}\geq 0\right)+5\omega,$$

where $K=4\times\max\{K_1,K_2,K_3\}$, which proves (5.21).

Step 2: Now we prove that

$$\Pr\left(\sup_{g\in\mathcal{G}_n(M),h\in\mathcal{N}_{1,K_1}}\left[I_{1,n}(g,h_o)+I_{2,n}(g,h)\right]+K\delta_{2,n}^{*2}\geq 0\right)$$
$$\leq\sum_{j\geq M,2^{j-1}\delta_{2,n}^{*}\leq K_M^{*}}\Pr\left(\sup_{g\in\mathcal{G}_{n,j}}I_{1,n}(g,h_o)\geq\left(c_{K_1,2}2^{2j}-K-c_{K_1,1}2^{j}\right)\delta_{2,n}^{*2}\right) \tag{5.31}$$

where $c_{K_1,1}$ and $c_{K_1,2}$ are some fixed constants defined below,

$$\mathcal{G}_{n,j}\equiv\left\{g:2^{j}\delta_{2,n}^{*}<||g-g_o||_{\mathcal{G}}\leq 2^{j+1}\delta_{2,n}^{*}\right\}.$$

We start by noting that we can divide $\mathcal{G}_n(M)$ into infinite (but countably) many disjoint pieces, i.e. $\mathcal{G}_n(M)=\cup_{j=M}^{\infty}\mathcal{G}_{n,j}$, where it is clear that $\mathcal{G}_{n,j}\cap\mathcal{G}_{n,j'}=\varnothing$ for any $j,j'\geq M$ and $j\neq j'$. By the sub-additivity of the probability measure,

$$\Pr\left(\sup_{g\in\mathcal{G}_n(M),h\in\mathcal{N}_{1,K_1}}\left[I_{1,n}(g,h_o)+I_{2,n}(g,h)\right]+K\delta_{2,n}^{*2}\geq 0\right)$$
$$\leq\sum_{j\geq M,2^{j-1}\delta_{2,n}^{*}\leq K_M^{*}}\Pr\left(\sup_{g\in\mathcal{G}_{n,j},h\in\mathcal{N}_{1,K_1}}\left[I_{1,n}(g,h_o)+I_{2,n}(g,h)\right]+K\delta_{2,n}^{*2}\geq 0\right). \tag{5.32}$$

By Assumption 5.4.(ii) and the definition of $\delta_{2,n}^{*}$, we have

$$\begin{aligned}\sup_{g\in\mathcal{G}_{n,j},h\in\mathcal{N}_{1,K_1}}I_{2,n}(g,h)&=\sup_{g\in\mathcal{G}_{n,j},h\in\mathcal{N}_{1,K_1}}\left[Q(g,h)-Q(g_o,h)\right]\\&\leq c_{K_1,1}\delta_{1,n}\left(2^{j}\delta_{2,n}^{*}\right)-c_{K_1,2}\left(2^{j}\delta_{2,n}^{*}\right)^{2}\\&\leq\left(c_{K_1,1}2^{j}-c_{K_1,2}2^{2j}\right)\delta_{2,n}^{*2},\end{aligned}$$

which together with (5.32) implies (5.31).

Step 3: We now prove that

$$\Pr\left(\sup_{g\in\mathcal{G}_{n,j}}I_{1,n}(g,h_o)\geq\left(c_{K_1,2}2^{2j}-K-c_{K_1,1}2^{j}\right)\delta_{2,n}^{*2}\right)\leq\frac{c_1c_2\left[(2^{j+1})^{\gamma}+K_{\varepsilon}^{\gamma}\right]}{|c_{K_1,2}2^{2j}-K-c_{K_1,1}2^{j}|} \tag{5.33}$$

where c denotes the generic constant defined in (5.16) or (5.17), and K_ε is some fixed constant defined below. For this purpose, we start by using Markov inequality and the triangle inequality to deduce that

$$\begin{aligned}
&\Pr\left(\sup_{g\in\mathcal{G}_{n,j}} I_{1,n}(g,h_o) \geq \left(c_{K_1,2}2^{2j} - K - c_{K_1,1}2^j\right)\delta_{2,n}^{*2}\right) \\
&\leq \frac{\mathbb{E}\left[\sup_{g\in\mathcal{G}_{n,j}} \left|\mu_n\left[\psi(Z_2,g,h_o) - \psi(Z_2,g_n,h_o)\right]\right|\right]}{\left|\left(c_{K_1,2}2^{2j} - K - c_{K_1,1}2^j\right)\delta_{2,n}^{*2}\right|} \\
&\leq \frac{\mathbb{E}\left[\sup_{g\in\mathcal{G}_{n,j}} \left|\mu_n\left[\psi(Z_2,g,h_o) - \psi(Z_2,g_o,h_o)\right]\right|\right]}{\left|c_{K_1,2}2^{2j} - K - c_{K_1,1}2^j\right|\delta_{2,n}^{*2}} \\
&+ \frac{\mathbb{E}\left[\left|\mu_n\left[\psi(Z_2,g_n,h_o) - \psi(Z_2,g_o,h_o)\right]\right|\right]}{\left|c_{K_1,2}2^{2j} - K - c_{K_1,1}2^j\right|\delta_{2,n}^{*2}}.
\end{aligned} \tag{5.34}$$

Using Assumption 5.4.(iv), we deduce that

$$\begin{aligned}
&\frac{\mathbb{E}\left[\sup_{g\in\mathcal{G}_{n,j}} \left|\mu_n\left[\psi(Z_2,g,h_o) - \psi(Z_2,g_o,h_o)\right]\right|\right]}{\left|c_{K_1,2}2^{2j} - K - c_{K_1,1}2^j\right|\delta_{2,n}^{*2}} \\
&\leq \frac{c_1\phi_n(2^{j+1}\delta_{2,n}^*)}{\sqrt{n}\left|c_{K_1,2}2^{2j} - K - c_{K_1,1}2^j\right|\delta_{2,n}^{*2}} \\
&= \frac{c_1(2^{j+1}\delta_{2,n}^*)^\gamma}{\sqrt{n}\left|c_{K_1,2}2^{2j} - K - c_{K_1,1}2^j\right|\delta_{2,n}^{*2}} \frac{\phi_n(2^{j+1}\delta_{2,n}^*)}{(2^{j+1}\delta_{2,n}^*)^\gamma} \\
&\leq \frac{c_1(2^{j+1})^\gamma}{\left|c_{K_1,2}2^{2j} - K - c_{K_1,1}2^j\right|} \frac{\phi_n(\delta_{2,n}^*)}{\sqrt{n}\delta_{2,n}^{*2}} \leq \frac{c_1c_2(2^{j+1})^\gamma}{\left|c_{K_1,2}2^{2j} - K - c_{K_1,1}2^j\right|},
\end{aligned} \tag{5.35}$$

where the last inequality uses the fact that $\phi_n(\delta)/\delta^\gamma$ is a decreasing function such that

$$\begin{aligned}
\frac{\phi_n(\delta_{2,n}^*)}{\sqrt{n}\delta_{2,n}^{*2}} &= \frac{1}{\sqrt{n}} \frac{\phi_n(\delta_{2,n}^*)}{\delta_{2,n}^{*\gamma}} \frac{1}{\delta_{2,n}^{*(2-\gamma)}} \\
&\leq \frac{1}{\sqrt{n}} \frac{\phi_n(\delta_{g,n})}{\delta_{g,n}^{\gamma}} \frac{1}{\delta_{g,n}^{(2-\gamma)}} = \frac{\phi_n(\delta_{g,n})}{\sqrt{n}\delta_{g,n}^2} \leq c_2.
\end{aligned}$$

From $||g_n - g_o||_{\mathcal{G}} = O(\delta^*_{2,n})$, we can choose $K_\varepsilon > 1$ large enough such that $||g_n - g_o||_{\mathcal{G}} \leq K_\varepsilon \delta^*_{2,n}$. Using Assumption 5.4.(iv) and similar arguments in showing (5.35), we deduce that

$$\begin{aligned}
&\frac{\mathbb{E}\left[\left|\mu_n\left[\psi(Z_2, g_n, h_o) - \psi(Z_2, g_o, h_o)\right]\right|\right]}{\left|c_{K_1,2}2^{2j} - K - c_{K_1,1}2^j\right| \delta^{*2}_{2,n}} \\
&\leq \frac{\mathbb{E}\left[\left|\sup_{\left\{g\in\mathcal{G}_n:||g-g_o||_{\mathcal{G}}\leq K_\varepsilon\delta^*_{2,n}\right\}} \mu_n\left[\psi(Z_2, g, h_o) - \psi(Z_2, g_o, h_o)\right]\right|\right]}{\left|c_{K_1,2}2^{2j} - K - c_{K_1,1}2^j\right| \delta^{*2}_{2,n}} \\
&\leq \frac{c_1(K_\varepsilon\delta^*_{2,n})^\gamma}{\sqrt{n}\left|c_{K_1,2}2^{2j} - K - c_{K_1,1}2^j\right| \delta^{*2}_{2,n}} \frac{\phi_n(K_\varepsilon\delta^*_{2,n})}{(K_\varepsilon\delta^*_{2,n})^\gamma} \\
&\leq \frac{c_1 K_\varepsilon^\gamma}{\left|c_{K_1,2}2^{2j} - K - c_{K_1,1}2^j\right|} \frac{\phi_n(\delta^*_{2,n})}{\sqrt{n}\delta^{*2}_{2,n}} = \frac{c_1 c_2 K_\varepsilon^\gamma}{\left|c_{K_1,2}2^{2j} - K - c_{K_1,1}2^j\right|}. \qquad (5.36)
\end{aligned}$$

From (5.34), (5.35) and (5.36), we get (5.33). ■

Theorem 5.2 indicates that the convergence rate of the second-step sieve M estimator is determined by the convergence rate $\max\{\delta_{1,n}, \delta_n\}$ of the estimation error introduced by the first-step sieve estimation, the rate $\delta_{2,n}$ of the sieve approximation error of g_o, the convergence rate $\varepsilon_{2,n}$ of the optimization error and the measure $\delta_{g,n}$ of the complexity of the sieve space $\mathcal{G}_n$.

Let $\Psi_{n,\delta} \equiv \left\{\psi(Z_2, g, h_o) - \psi(Z_2, g_o, h_o) : \ \|g - g_o\|_{\mathcal{G}} \leq \delta, g \in \mathcal{N}_{2,K}\right\}$ and let $H_{[]}\left(u, \Psi_{n,\delta}, \|\cdot\|_2\right)$ denote the bracket entropy of the function class $\Psi_{n,\delta}$ with respect to the $L_2(dF_Z)$-norm $\|\cdot\|_2$. Define

$$J_{[]}\left(\delta, \Psi_{n,\delta}, \|\cdot\|_2\right) = \int_0^\delta H_{[]}\left(u, \Psi_{n,\delta}, \|\cdot\|_2\right) du.$$

Assumption 5.4.(iii) and (iv) can be replaced by the following low level conditions.

Assumption 5.5 *(i) The data are i.i.d.; (ii)*

$$\sup_{\left\{g\in\mathcal{N}_{2,K}:\ \|g-g_o\|_{\mathcal{G}}\leq\delta\right\}} \mathbb{E}\left[\left|\psi(Z, g, h_o) - \psi(Z, g_o, h_o)\right|^2\right] \leq c\delta^2;$$

(iii) for any small $\delta > 0$, there exists a constant $s_1 \in (0, 2)$ such that

$$\sup_{\left\{g\in\mathcal{N}_{2,K}:\ \|g-g_o\|_{\mathcal{G}}\leq\delta\right\}} \left|\psi(Z, g, h_o) - \psi(Z, g_o, h_o)\right| \leq \delta^{s_1} U(Z)$$

where $\mathbb{E}\left[|U(Z)|^{s_2}\right] \leq c$ *for some* $s_2 \geq 2$*; (iv) there is a sequence of positive numbers* $\delta_{g,n}$ *such that*

$$\delta_{g,n} = \inf\left\{\delta \in (0,1) : \frac{J_{[]}\left(\delta, \Psi_{n,\delta}, \|\cdot\|_2\right)}{\sqrt{n}\delta^2} \leq c\right\},$$

where $\delta^{-\gamma} J_{[]}\left(\delta, \Psi_{n,\delta}, \|\cdot\|_2\right)$ *is a decreasing function for some* $\gamma \in (0,2)$*.*

Assumption 5.5.(i), (ii) and (iii) are directly from the sufficient conditions of Theorem 3.2 in Chen (2007) which establishes the convergence rate of one-step sieve M estimation with *i.i.d.* or m-dependent data. The low level conditions in Assumption 5.5 are easy to verify in practice. However, the advantage of the high level assumption (5.16) is that it integrates the data structure and the metric entropy restriction into one simple stochastic equicontinuity condition. As a result, the convergence rate of the second-step sieve M estimator derived in this paper applies to the general scenario with time series observation.

Corollary 5.3 *Suppose that the conditions in Theorem 5.1, Assumption 5.4.(i), (ii) and 5.5 are satisfied. Furthermore, if* $\|g_n - g_o\|_{\mathcal{G}} = O(\delta^*_{2,n})$*, then we have* $\|\widehat{g}_n - g_o\|_{\mathcal{G}} = O_p(\delta^*_{2,n})$*, where* $\delta^*_{2,n}$ *is defined in Theorem 5.2.*

Proof of Corollary 5.3. By Assumption 5.5.(iii), we know that for any small number $\omega > 0$, there exists a sufficiently large constant M_n such that

$$\begin{aligned}\Pr(|U(Z_i)| > M_n \text{ for all } i \leq n) &\leq \sum_{i=1}^{n} \Pr(|U(Z_i)| > M_n) \\ &\leq \sum_{i=1}^{n} \frac{\mathbb{E}\left[|U(Z)|^{s_2}\right]}{M_n^{s_2}} \leq cnM_n^{-s_2} \leq \omega,\end{aligned}$$

where the first inequality is by the Bonferroni inequality, and the second inequality is by the Markov inequality.

Now, conditioning on the event $\{|U(Z_i)| \leq M_n \text{ for all } i \leq n\}$ and using Assumption 5.5.(iii), we have

$$|\psi(Z_i, g, h_o) - \psi(Z_i, g_o, h_o)| \leq \delta^{s_1} M_n$$

for all $i \leq n$ and for any $\psi(Z, g, h_o) - \psi(Z, g_o, h_o) \in \Psi_{n,\delta}$, which together with Assumption 5.5.(i) and (ii), enables us to invoke Lemma 19.36 in Van der Vaart (1998) to get

$$\begin{aligned}
&\mathbb{E}\left[\sup_{\left\{g \in \mathcal{G}_n:\ \|g-g_o\|_{\mathcal{G}} \leq \delta\right\}} \left|\mu_n\left[\psi(Z, g, h_o) - \psi(Z, g_o, h_o)\right]\right|\right] \\
&\leq \frac{cJ_{[]}\left(\delta, \Psi_{n,\delta}, \|\cdot\|_2\right)}{\sqrt{n}} \left(1 + \frac{J_{[]}\left(\delta, \Psi_{n,\delta}, \|\cdot\|_2\right)}{\sqrt{n}\delta^2} M_n\right) \equiv \frac{\phi_n(\delta)}{\sqrt{n}}.
\end{aligned}$$

By Assumption 5.5.(iv), we know that the above function $\phi_n(\delta)$ satisfies the requirement (5.17) in Theorem 5.2. The rest of the proof is the same as that of Theorem 5.2 and hence is omitted. ■

References

[1] Belloni, A., V. Chernozhukov, D. Chetverikov, and K. Kato (2015): “Some New Asymptotic Theory for Least Squares Series: Pointwise and Uniform Results,” Journal of Econometrics, 186, 345–366.

[2] Belloni, A., Chernozhukov, D., Chernozhukov, V., and Fernández-val, I. (2016): ”Conditional Quantile Process Based on Series or Many Regressors,” Working Paper, Department of Economics, UCLA.

[3] Chen, X. (2007): “Large Sample Sieve Estimation of Semi-Nonparametric Models,” In: James J. Heckman and Edward E. Leamer, Editor(s), *Handbook of Econometrics*, 6B, Pages 5549-5632.

[4] Chen, X. and X. Shen (1998): “Sieve Extremum Estimates for Weakly Dependent Data,” *Econometrica*, 66, 289-314.

[5] Chen, X., and D. Pouzo (2012): “Estimation of Nonparametric Conditional Moment Models with Possibly Nonsmooth Generalized Residuals,” *Econometrica* 80, 277-321.

[6] Shen, X. and W.H. Wong (1994): “Convergence Rate of Sieve Estimates,” *Annals of Statistics*, 22(2) 580–615.

[7] Van der Vaart, A. and J. Wellner (1996): *Weak Convergence and Empirical Processes: with Applications to Statistics*, New York: Springer-Verlag

[8] Van der Vaart, A. (1998): *Asymptotic Statistics*, Cambridge: Cambridge University Press.

[9] White, H. and J. Wooldridge (1991): "Some Results on Sieve Estimation with Dependent Observations", in Barnett, W.A., J. Powell and G. Tauchen (eds.), *Non-parametric and Semi-parametric Methods in Econometrics and Statistics*, 459-493, Cambridge: Cambridge University Press.

[10] Wooldridge, J.M. (2002): *Econometric Analysis of Cross Section and Panel Data*, Cambridge: MIT Press.