Empirical Likelihood Covariate Adjustment for Regression Discontinuity Designs

Jun Ma[*] Zhengfei Yu[†]

Proof of Lemma 1. (a) follows from LIE and change of variables. (b) is a straightforward extension of Bickel and Doksum (2015, Proposition 11.3.1), which follows from LIE and $(p+1)$-th order Taylor expansion. For (c), denote $\bar{q}(V, X \mid h) := h^{-1/2} W_{p;\mathrm{s}}^{k} V$ and $\bar{\mathfrak{Q}} := \{\bar{q}(\cdot \mid h) : h \in \mathbb{H}\}$. Denote $\mathbb{P}_n^V f := n^{-1} \sum_i f(V_i, X_i)$, $\mathbb{P}^V f := \mathrm{E}[f(V, X)]$ and $\mathbb{G}_n^V := \sqrt{n}\left(\mathbb{P}_n^V - \mathbb{P}^V\right)$. Then we have

$$\left\|\mathbb{G}_n^V\right\|_{\bar{\mathfrak{Q}}} = \sup_{h \in \mathbb{H}} \left| \frac{1}{\sqrt{nh}} \sum_i \left(W_{p;\mathrm{s},i}^{k} V_i - \mathrm{E}\left[W_{p;\mathrm{s}}^{k} V\right]\right) \right|.$$

Let $\sigma_{\bar{\mathfrak{Q}}}^2 := \sup_{f \in \bar{\mathfrak{Q}}} \mathbb{P}^V f^2$. It follows from LIE and change of variables that $\sigma_{\bar{\mathfrak{Q}}}^2 = \sup_{h \in \mathbb{H}} \mathrm{E}\left[h^{-1} W_{p;\mathrm{s}}^{2k} g_{V^2}(X)\right] = O(1)$. Assume $\mathrm{s} = +$ without loss of generality. By definition and the assumption that K is supported on $[-1, 1]$, $\bar{q}(v, x \mid h) = \mathcal{K}_{p;-}^{k}(x/h)\, h^{-1/2} 1(0 < x < h)\, v$. Since Assumption 3 also implies that $\mathcal{K}_{p;-}^{k}$ has bounded variation $\forall k \in \mathbb{N}$. By Giné and Nickl (2015, Proposition 3.6.12), $\left\{x \mapsto \mathcal{K}_{p;-}^{k}(x/h) : h \in \mathbb{H}\right\}$ is VC-type with respect to a constant envelope and its VC characteristics are independent of n. By Kosorok (2007, Lemma 9.6), $\left\{(x, v) \mapsto h^{-1/2} 1(0 < x < h)\, v : h \in \mathbb{H}\right\}$ is VC-subgraph with an envelope $(x, v) \mapsto \underline{h}^{-1/2} 1\left(0 < x < \overline{h}\right) |v|$ and VC index being at most 3. By Kosorok (2007, Theorem 9.3) and Chernozhukov et al. (2014, Corollary A.1), $\bar{\mathfrak{Q}}$ is VC-type with respect to an envelope $F_{\bar{\mathfrak{Q}}}(v, x) \propto \underline{h}^{-1/2} 1\left(0 < x < \overline{h}\right) |v|$. By Chen and Kato (2020, Corollary 5.5), $\mathrm{E}\left[\left\|\mathbb{G}_n^V\right\|_{\bar{\mathfrak{Q}}}\right] \lesssim \sigma_{\bar{\mathfrak{Q}}} \sqrt{\log(n)} + \log(n) \left(\mathbb{P}^V \left|F_{\bar{\mathfrak{Q}}}\right|^r\right)^{1/r} n^{1/r} / \sqrt{n}$, where $\mathbb{P}^V \left|F_{\bar{\mathfrak{Q}}}\right|^r = O\left(\overline{h} / \underline{h}^{r/2}\right)$. (c) follows from Markov's inequality. ■

Proof of Lemma 2. Let $\mathcal{L}_\sharp := \left\{\lambda \in \mathbb{R}^{2d_u} : \|\lambda\| \leq \log(n) / \sqrt{nh}\right\}$. By

$$\max_i \|\mathcal{U}_i\| / \sqrt{nh} \lesssim \max_i 1\left(|X_i| \leq \overline{h}\right) \|U_i\| / \sqrt{\underline{n}} \leq \left(\sum_i 1\left(|X_i| \leq \overline{h}\right) \|U_i\|^{12}\right)^{1/12} / \sqrt{\underline{n}}$$

This version: April 26, 2022

[*]School of Economics, Renmin University of China
[†]Faculty of Humanities and Social Sciences, University of Tsukuba

and Markov's inequality, we have $\max_i \|\mathcal{U}_i\| / \sqrt{nh} = O_p\left(\overline{n}^{1/12}/\underline{n}^{1/2}\right)$. It follows that $\max_i \sup_{\lambda \in \mathcal{L}_\sharp} \left|\lambda^\top \mathcal{U}_i\right| = O_p\left(\log(n)\left(\overline{n}^{1/12}/\underline{n}^{1/2}\right)\right)$ and $\max_i \sup_{\lambda \in \mathcal{L}_\sharp} \left|\lambda^\top \mathcal{U}_i\right| < 1/2$ $\forall h \in \mathbb{H}$ wpa1. Therefore, $\mathcal{L}_\sharp \subseteq \mathcal{L}(\vartheta)$, $\forall h \in \mathbb{H}$ wpa1. Since $S(\cdot, \vartheta)$ is continuous and $\mathcal{L}_\sharp$ is compact, $\lambda_\sharp \coloneqq \operatorname{argmax}_{\lambda \in \mathcal{L}_\sharp} S(\lambda, \vartheta)$ exists $\forall h \in \mathbb{H}$ wpa1. By the definition of $\lambda_\sharp$ and second-order Taylor expansion,

$$
\begin{aligned}
0 = S\left(0_{2d_u}, \vartheta\right) \leq S\left(\lambda_\sharp, \vartheta\right) &= 2\left(\sqrt{nh}\lambda_\sharp\right)^\top \overline{\mathcal{U}} - \left(\sqrt{nh}\lambda_\sharp\right)^\top \left(\frac{1}{nh}\sum_i \frac{\mathcal{U}_i \mathcal{U}_i^\top}{\left(1 + \dot{\lambda}_\sharp^\top \mathcal{U}_i\right)^2}\right)\left(\sqrt{nh}\lambda_\sharp\right) \\
&\leq 2\left\|\sqrt{nh}\lambda_\sharp\right\| \left\|\overline{\mathcal{U}}\right\| - \left(\sqrt{nh}\lambda_\sharp\right)^\top \left(\frac{1}{nh}\sum_i \frac{\mathcal{U}_i \mathcal{U}_i^\top}{\left(1 + \max_i \sup_{\lambda \in \mathcal{L}_\sharp} \left|\lambda^\top \mathcal{U}_i\right|\right)^2}\right)\left(\sqrt{nh}\lambda_\sharp\right), \quad \text{(S1)}
\end{aligned}
$$

where $\dot{\lambda}_\sharp$ is the mean value that lies on the line joining 0_{2d_u} and $\lambda_\sharp$. Since $\max_i \sup_{\lambda \in \mathcal{L}_\sharp} \left|\lambda^\top \mathcal{U}_i\right| < 1/2$ $\forall h \in \mathbb{H}$ wpa1, by (S1),

$$
0 \leq S\left(\lambda_\sharp, \vartheta\right) \leq 2\left\|\sqrt{nh}\lambda_\sharp\right\| \left\|\overline{\mathcal{U}}\right\| - \frac{4}{9}\left(\sqrt{nh}\lambda_\sharp\right)^\top \left(\overline{\Delta}_{\mathcal{U}\mathcal{U}^\top} - \Delta_{\mathcal{U}\mathcal{U}^\top}\right)\left(\sqrt{nh}\lambda_\sharp\right) - \frac{4}{9}\left(\sqrt{nh}\lambda_\sharp\right)^\top \Delta_{\mathcal{U}\mathcal{U}^\top}\left(\sqrt{nh}\lambda_\sharp\right),
$$

$\forall h \in \mathbb{H}$ wpa1 and therefore,

$$
\varrho_{\min}\left(\Delta_{\mathcal{U}\mathcal{U}^\top}\right)\left\|\sqrt{nh}\lambda_\sharp\right\|^2 \leq \frac{9}{2}\left\|\sqrt{nh}\lambda_\sharp\right\| \left\|\overline{\mathcal{U}}\right\| + \left\|\overline{\Delta}_{\mathcal{U}\mathcal{U}^\top} - \Delta_{\mathcal{U}\mathcal{U}^\top}\right\| \left\|\sqrt{nh}\lambda_\sharp\right\|^2, \quad \text{(S2)}
$$

$\forall h \in \mathbb{H}$ wpa1. Since $\overline{\mathcal{U}} = (nh)^{-1/2}\sum_{i=1}^n \left(\mathcal{U}_i - \mathrm{E}[\mathcal{U}]\right) + \sqrt{nh}\Delta_{\mathcal{U}}$, it follows from Lemma 1 that $\left\|\overline{\mathcal{U}}\right\| = O_p\left(\sqrt{\log(n)}\right)$. It also follows from Lemma 1 that $\overline{\Delta}_{\mathcal{U}\mathcal{U}^\top} - \Delta_{\mathcal{U}\mathcal{U}^\top} = O_p\left(\sqrt{\log(n)/\underline{n}} + \log(n)\left(\overline{n}^{1/6}/\underline{n}\right)\right)$ and $\Delta_{\mathcal{U}\mathcal{U}^\top} = \operatorname{diag}\left(\psi_{UU^\top,+}, \psi_{UU^\top,-}\right) + O\left(\overline{h}\right)$. Since $\operatorname{diag}\left(\psi_{UU^\top,+}, \psi_{UU^\top,-}\right)$ is positive definite, $\varrho_{\min}\left(\Delta_{\mathcal{U}\mathcal{U}^\top}\right)$ is bounded away from zero when n is sufficiently large. By assumption, $\left\|\sqrt{nh}\lambda_\sharp\right\| \leq \log(n)$. It follows from these results and (S2) that $\sqrt{nh}\lambda_\sharp = O_p\left(\sqrt{\log(n)}\right)$. By this result, $\Pr\left[\sqrt{nh}\lambda_\sharp \leq \log(n)/2, \forall h \in \mathbb{H}\right] \to 1$ and therefore, wpa1, $\forall h \in \mathbb{H}$, $\lambda_\sharp$ is in the interior of $\mathcal{L}_\sharp$ and the first-order condition is satisfied: $\partial S(\lambda, \vartheta)/\partial\lambda|_{\lambda=\lambda_\sharp} = 0_{2d_u}$. Since $S(\cdot, \vartheta)$ is concave, $\lambda_\sharp$ attains $\sup_{\lambda \in \mathcal{L}(\vartheta)} S(\lambda, \vartheta)$ $\forall h \in \mathbb{H}$ wpa1 and therefore, $\sup_{\lambda \in \mathcal{L}(\vartheta)} S(\lambda, \vartheta) = S\left(\lambda_\sharp, \vartheta\right) \leq 2\left\|\sqrt{nh}\lambda_\sharp\right\| \left\|\overline{\mathcal{U}}\right\| = O_p(\log(n))$. Denote $\lambda_\natural \coloneqq \sqrt{\log(n)/(nh)}\widehat{\mathcal{U}}/\left\|\widehat{\mathcal{U}}\right\|$. It can be shown by using similar arguments, boundedness of Θ and $\widehat{\mathcal{U}}_i = \mathcal{U}_i - \mathcal{G}_i\widehat{\eta}_p$ that $\max_i \left\|\widehat{\mathcal{U}}_i\right\| / \sqrt{nh} = O_p\left(\overline{n}^{1/12}/\underline{n}^{1/2}\right)$. By second-order Taylor expansion,

$$
S\left(\lambda_\natural, \widehat{\vartheta}_p\right) = 2\left(\sqrt{nh}\lambda_\natural\right)^\top \widehat{\mathcal{U}} - \left(\sqrt{nh}\lambda_\natural\right)^\top \left(\frac{1}{nh}\sum_i \frac{\widehat{\mathcal{U}}_i \widehat{\mathcal{U}}_i^\top}{\left(1 + \dot{\lambda}_\natural^\top \widehat{\mathcal{U}}_i\right)^2}\right)\left(\sqrt{nh}\lambda_\natural\right)
$$

$$\geq 2\left(\sqrt{nh}\lambda_\natural\right)^\top \widehat{\mathcal{U}} - \left(\sqrt{nh}\lambda_\natural\right)^\top \left(\frac{1}{nh}\sum_i \frac{\widehat{\mathcal{U}}_i\widehat{\mathcal{U}}_i^\top}{\left(1-\sqrt{\log(n)/(nh)}\left(\max_i\left\|\widehat{\mathcal{U}}_i\right\|\right)\right)^2}\right)\left(\sqrt{nh}\lambda_\natural\right), \tag{S3}$$

where $\dot{\lambda}_\natural$ is the mean value that lies on the line joining 0_{2d_u} and $\lambda_\natural$. Then, $\sqrt{\log(n)}\left\|\widehat{\mathcal{U}}\right\| \leq S\left(\lambda_\natural, \widehat{\vartheta}_p\right) + 2\left((nh)^{-1}\sum_i\left\|\widehat{\mathcal{U}}_i\right\|^2\right)\log(n)$, $\forall h\in\mathbb{H}$, wpa1. By $\widehat{\mathcal{U}}_i = \mathcal{U}_i - \mathcal{G}_i\widehat{\eta}_p$, Lemma 1 and boundedness of Θ, we have $(nh)^{-1}\sum_i\left\|\widehat{\mathcal{U}}_i\right\|^2 = O_p(1)$. By the definition of $\widehat{\vartheta}_p$, $S\left(\lambda_\natural,\widehat{\vartheta}_p\right) \leq \sup_{\lambda\in\mathcal{L}(\widehat{\vartheta}_p)} S\left(\lambda,\widehat{\vartheta}_p\right) \leq \sup_{\lambda\in\mathcal{L}(\vartheta)} S(\lambda,\vartheta)$. Since $\sup_{\lambda\in\mathcal{L}(\vartheta)} S(\lambda,\vartheta) = O_p(\log(n))$, it follows that $\widehat{\mathcal{U}} = O_p\left(\sqrt{\log(n)}\right)$. Since $\widehat{\mathcal{U}} = \overline{\mathcal{U}} - \overline{\Delta}_{\mathcal{G}}\sqrt{nh}\widehat{\eta}_p$, then,

$$\left\|\sqrt{nh}\widehat{\eta}_p\right\|\sqrt{\varrho_{\mathtt{min}}\left(\overline{\Delta}_{\mathcal{G}}^\top\overline{\Delta}_{\mathcal{G}}\right)} \leq \left\|\overline{\Delta}_{\mathcal{G}}\sqrt{nh}\widehat{\eta}_p\right\| \leq \left\|\widehat{\mathcal{U}}\right\| + \left\|\overline{\mathcal{U}}\right\|. \tag{S4}$$

By Lemma 1, $\overline{\Delta}_{\mathcal{G}} = \begin{bmatrix}\mu_{G,+}^\top & \mu_{G,-}^\top\end{bmatrix}^\top + O_p\left(\sqrt{\log(n)/\underline{n}} + \overline{h}\right)$. $\begin{bmatrix}\mu_{G,+}^\top & \mu_{G,-}^\top\end{bmatrix}^\top$ has full column rank, if $\mu_{D,+}\neq\mu_{D,-}$. By using the fact that $|\varrho_{\mathtt{min}}(\mathrm{A}) - \varrho_{\mathtt{min}}(\mathrm{B})| \leq \|\mathrm{A}-\mathrm{B}\|$, $\varrho_{\mathtt{min}}\left(\overline{\Delta}_{\mathcal{G}}^\top\overline{\Delta}_{\mathcal{G}}\right)$ is bounded away from zero $\forall h\in\mathbb{H}$, wpa1. (a) follows easily from this result, (S4) and the fact that $\left\|\widehat{\mathcal{U}}\right\|$ and $\left\|\overline{\mathcal{U}}\right\|$ are both $O_p\left(\sqrt{\log(n)}\right)$. By $\max_i\left\|\widehat{\mathcal{U}}_i\right\|/\sqrt{nh} = O_p\left(\overline{n}^{1/12}/\underline{n}^{1/2}\right)$ and the definition of $\mathcal{L}_\sharp$, $\max_i\sup_{\lambda\in\mathcal{L}_\sharp}\left|\lambda^\top\widehat{\mathcal{U}}_i\right| = O_p\left(\log(n)\left(\overline{n}^{1/12}/\underline{n}^{1/2}\right)\right)$ and therefore $\max_i\sup_{\lambda\in\mathcal{L}_\sharp}\left|\lambda^\top\widehat{\mathcal{U}}_i\right| < 1/2$ $\forall h\in\mathbb{H}$ wpa1. Therefore, $\mathcal{L}_\sharp\subseteq\mathcal{L}\left(\widehat{\vartheta}_p\right)$, $\forall h\in\mathbb{H}$ wpa1. Since $S\left(\cdot,\widehat{\vartheta}_p\right)$ is continuous and $\mathcal{L}_\sharp$ is compact, $\widehat{\lambda}_\sharp \coloneqq \mathrm{argmax}_{\lambda\in\mathcal{L}_\sharp} S\left(\lambda,\widehat{\vartheta}_p\right)$ exists $\forall h\in\mathbb{H}$ wpa1. By the definition of $\widehat{\lambda}_\sharp$ and similar arguments used to show (S2), we have $\varrho_{\min}\left(\Delta_{\mathcal{U}\mathcal{U}^\top}\right)\left\|\sqrt{nh}\widehat{\lambda}_\sharp\right\| \lesssim \left\|\widehat{\mathcal{U}}\right\| + \left\|\widehat{\Delta}_{\mathcal{U}\mathcal{U}^\top} - \Delta_{\mathcal{U}\mathcal{U}^\top}\right\|\left\|\sqrt{nh}\widehat{\lambda}_\sharp\right\|$. Since $\widehat{\Delta}_{\mathcal{U}\mathcal{U}^\top} - \overline{\Delta}_{\mathcal{U}\mathcal{U}^\top} = (nh)^{-1}\sum_i\left\{\mathcal{G}_i\widehat{\eta}_p\mathcal{U}_i^\top + \mathcal{U}_i\widehat{\eta}_p^\top\mathcal{G}_i^\top + \mathcal{G}_i\widehat{\eta}_p\widehat{\eta}_p^\top\mathcal{G}_i^\top\right\}$, it follows from Lemma 1 and (a) that $\widehat{\Delta}_{\mathcal{U}\mathcal{U}^\top} - \overline{\Delta}_{\mathcal{U}\mathcal{U}^\top} = O_p\left(\sqrt{\log(n)/\underline{n}}\right)$ and therefore, $\widehat{\Delta}_{\mathcal{U}\mathcal{U}^\top} - \Delta_{\mathcal{U}\mathcal{U}^\top} = O_p\left(\sqrt{\log(n)/\underline{n}} + \log(n)\left(\overline{n}^{1/6}/\underline{n}\right)\right)$. Since $\left\|\sqrt{nh}\widehat{\lambda}_\sharp\right\| \leq \log(n)$ by construction, it follows from these results that $\sqrt{nh}\widehat{\lambda}_\sharp = O_p\left(\sqrt{\log(n)}\right)$. Wpa1, $\forall h\in\mathbb{H}$, $\widehat{\lambda}_\sharp$ is in the interior of $\mathcal{L}_\sharp$ and the first-order condition is satisfied: $\partial S\left(\lambda,\widehat{\vartheta}_p\right)/\partial\lambda\Big|_{\lambda=\widehat{\lambda}_\sharp} = 0_{2d_u}$. It follows from the concavity of $S\left(\cdot,\widehat{\vartheta}_p\right)$ that $\widehat{\lambda}_\sharp$ also attains $\sup_{\lambda\in\mathcal{L}(\widehat{\vartheta}_p)} S\left(\lambda,\widehat{\vartheta}_p\right)$ $\forall h\in\mathbb{H}$ wpa1. Then (b) follows from setting $\widehat{\lambda}_p = \widehat{\lambda}_\sharp$. (c) and (d) follow from similar arguments. ■

Proof of Lemma 3. It is shown in the proof of Lemma 2 that $\widehat{\lambda}_p$ satisfies the first-order condition which can be written as $\sum_i\widehat{\mathcal{U}}_i/\left(1+\widehat{\lambda}_p^\top\widehat{\mathcal{U}}_i\right) = 0_{2d_u}$ $\forall h\in\mathbb{H}$ wpa1. We also showed that $\max_i\left\|\widehat{\mathcal{U}}_i\right\|/\sqrt{nh} = O_p\left(\overline{n}^{1/12}/\underline{n}^{1/2}\right)$ and $\sqrt{nh}\widehat{\lambda}_p = O_p\left(\sqrt{\log(n)}\right)$. Therefore, we have $\max_i\left|\widehat{\lambda}_p^\top\widehat{\mathcal{U}}_i\right| = o_p(1)$. By $\widehat{\mathcal{U}}_i = \mathcal{U}_i - \mathcal{G}_i\widehat{\eta}_p$, Lemma 1 and boundedness of Θ, we have $(nh)^{-1}\sum_i\left\|\widehat{\mathcal{U}}_i\right\|^3 = O_p\left(1+\log(n)\left(\overline{n}^{1/4}/\underline{n}\right)\right)$ and $(nh)^{-1}\sum_i\left\|\widehat{\mathcal{U}}_i\right\|^4 = O_p\left(1+\log(n)\left(\overline{n}^{1/3}/\underline{n}\right)\right)$. By these results, Lemma 2 and simple algebra, $(nh)^{-1}\sum_i\widehat{\mathcal{U}}_i\widehat{\mathcal{U}}_i^\top/\left(1+\widehat{\lambda}_p^\top\widehat{\mathcal{U}}_i\right)^2 = \widehat{\Delta}_{\mathcal{U}\mathcal{U}^\top} + o_p(1)$. It is shown in the proof of Lemma 2 that $\widehat{\Delta}_{\mathcal{U}\mathcal{U}^\top} = \mathrm{diag}\left(\psi_{UU^\top,+},\psi_{UU^\top,-}\right) + o_p(1)$. Therefore, $\varrho_{\mathtt{min}}\left((nh)^{-1}\sum_i\widehat{\mathcal{U}}_i\widehat{\mathcal{U}}_i^\top/\left(1+\widehat{\lambda}_p^\top\widehat{\mathcal{U}}_i\right)^2\right)$ is bounded away from zero $\forall h\in\mathbb{H}$, wpa1. By the implicit function theorem, wpa1 $\forall h \in \mathbb{H}$, there exists a continuously differentiable function $\lambda(\cdot)$ defined on some open neighborhood $\mathbb{B}\left(\widehat{\vartheta}_p\right)$ of $\widehat{\vartheta}_p$ such that $\widehat{\lambda}_p = \lambda\left(\widehat{\vartheta}_p\right)$ and $(nh)^{-1}\sum_i \mathcal{U}_i(\theta)/\left(1+\lambda(\theta)^\top \mathcal{U}_i(\theta)\right) = 0_{2d_u}$ $\forall \theta \in \mathbb{B}\left(\widehat{\vartheta}_p\right)$. Since $S(\cdot,\theta)$ is concave, $S(\lambda(\theta),\theta) = \sup_{\lambda\in\mathcal{L}(\theta)} S(\lambda,\theta)$ and $\widehat{\vartheta}_p = \operatorname{argmin}_{\theta\in\mathbb{B}(\widehat{\vartheta}_p)} S(\lambda(\theta),\theta)$. By the chain rule and $\sum_i \widehat{\mathcal{U}}_i/\left(1+\widehat{\lambda}_p^\top\widehat{\mathcal{U}}_i\right) = 0_{2d_u}$, the first-order condition for $\widehat{\vartheta}_p$ can be written as $\sum_i \mathcal{G}_i^\top \widehat{\lambda}_p/\left(1+\widehat{\lambda}_p^\top\widehat{\mathcal{U}}_i\right) = 0_{2d_u}$, which holds $\forall h \in \mathbb{H}$ wpa1. By simple algebra we have

$$0_{2d_u} = \sum_i \left\{\widehat{\mathcal{U}}_i - \widehat{\mathcal{U}}_i\widehat{\mathcal{U}}_i^\top\widehat{\lambda}_p + \frac{\widehat{\mathcal{U}}_i\left(\widehat{\mathcal{U}}_i^\top\widehat{\lambda}_p\right)^2}{1+\widehat{\lambda}_p^\top\widehat{\mathcal{U}}_i}\right\} \text{ and } 0_{d_\vartheta} = \sum_i \left\{\mathcal{G}_i^\top\widehat{\lambda}_p - \frac{\mathcal{G}_i^\top\widehat{\lambda}_p\left(\widehat{\lambda}_p^\top\widehat{\mathcal{U}}_i\right)}{1+\widehat{\lambda}_p^\top\widehat{\mathcal{U}}_i}\right\}. \tag{S5}$$

By $\max_i\left|\widehat{\lambda}_p^\top\widehat{\mathcal{U}}_i\right| = o_p(1)$, $(nh)^{-1}\sum_i\left\|\widehat{\mathcal{U}}_i\right\| = O_p(1)$, $(nh)^{-1}\sum_i\left\|\widehat{\mathcal{U}}_i\right\|^3 = O_p\left(1+\log(n)\left(\overline{n}^{1/4}/\underline{n}\right)\right)$ and Lemma 2, $(nh)^{-1/2}\sum_i\left\{\widehat{\mathcal{U}}_i\left(\widehat{\mathcal{U}}_i^\top\widehat{\lambda}_p\right)^2\right\}/\left(1+\widehat{\lambda}_p^\top\widehat{U}_i\right) = O_p\left(\upsilon_n^\ddagger\right)$ and $(nh)^{-1/2}\sum_i\left\{\mathcal{G}_i^\top\widehat{\lambda}_p\left(\widehat{\lambda}_p^\top\widehat{\mathcal{U}}_i\right)\right\}/\left(1+\widehat{\lambda}_p^\top\widehat{\mathcal{U}}_i\right) = O_p\left(\upsilon_n^\ddagger\right)$, where $\upsilon_n^\ddagger := \log(n)/\sqrt{\underline{n}} + \log(n)^2\left(\overline{n}^{1/4}/\underline{n}^{3/2}\right)$. By these results and $\widehat{\mathcal{U}}_i = \mathcal{U}_i - \mathcal{G}_i\widehat{\eta}_p$, (S5) can be written as $\widehat{\Delta}_{\mathcal{U}\mathcal{U}^\top}\sqrt{nh}\widehat{\lambda}_p + \overline{\Delta}_{\mathcal{G}}\sqrt{nh}\widehat{\eta}_p = \overline{\mathcal{U}} + O_p\left(\upsilon_n^\ddagger\right)$ and $\overline{\Delta}_{\mathcal{G}}^\top\sqrt{nh}\widehat{\lambda}_p = O_p\left(\log(n)/\sqrt{\underline{n}}\right)$. By $\widehat{\Delta}_{\mathcal{U}\mathcal{U}^\top} - \Delta_{\mathcal{U}\mathcal{U}^\top} = O_p\left(\sqrt{\log(n)/\underline{n}} + \log(n)\left(\overline{n}^{1/6}/\underline{n}\right)\right)$, $\overline{\Delta}_{\mathcal{G}} - \Delta_{\mathcal{G}} = O_p\left(\sqrt{\log(n)/\underline{n}}\right)$ and Lemma 2, we have

$$\Delta_{\mathcal{U}\mathcal{U}^\top}\sqrt{nh}\widehat{\lambda}_p + \Delta_{\mathcal{G}}\sqrt{nh}\widehat{\eta}_p = \overline{\mathcal{U}} + O_p\left(\upsilon_n^\dagger\right) \text{ and } \Delta_{\mathcal{G}}^\top\sqrt{nh}\widehat{\lambda}_p = O_p\left(\upsilon_n^\dagger\right). \tag{S6}$$

Since it follows from Lemma 1 that $\Delta_{\mathcal{U}\mathcal{U}^\top} = \operatorname{diag}\left(\psi_{UU^\top,+},\psi_{UU^\top,-}\right) + O\left(\overline{h}\right)$ and $\Delta_{\mathcal{G}} = \begin{bmatrix} \mu_{G,+}^\top & \mu_{G,-}^\top \end{bmatrix}^\top + O\left(\overline{h}^{p+1}\right)$, $\Delta_{\mathcal{U}\mathcal{U}^\top}$ and $\Delta_{\mathcal{G}}^\top\Delta_{\mathcal{U}\mathcal{U}^\top}^{-1}\Delta_{\mathcal{G}}$ are invertible $\forall h \in \mathbb{H}$, when n is sufficiently large. (a) follows from

$$\begin{bmatrix} \Delta_{\mathcal{U}\mathcal{U}^\top} & \Delta_{\mathcal{G}} \\ \Delta_{\mathcal{G}}^\top & 0_{d_\vartheta\times d_\vartheta} \end{bmatrix}^{-1} = \begin{bmatrix} \mathrm{Q} & \mathrm{N} \\ \mathrm{N}^\top & -\mathrm{O} \end{bmatrix}$$

and (S6). (b) follows from similar arguments. ■

Proof of Lemma 4. By Taylor expansion, $S\left(\widehat{\lambda}_p,\widehat{\vartheta}_p\right)$ is equal to the sum of $2\widehat{\lambda}_p^\top\left(\sum_i\widehat{\mathcal{U}}_i\right) - \sum_i\left(\widehat{\lambda}_p^\top\widehat{\mathcal{U}}_i\right)^2$ and a remainder term that is bounded up to a constant by $\sum_i\left|\widehat{\lambda}_p^\top\widehat{\mathcal{U}}_i\right|^3/\left(1-\left|\widehat{\lambda}_p^\top\widehat{\mathcal{U}}_i\right|\right)^3$. By using $(nh)^{-1}\sum_i\left\|\widehat{\mathcal{U}}_i\right\|^3 = O_p\left(1+\log(n)\left(\overline{n}^{1/4}/\underline{n}\right)\right)$ and Lemma 2, $\sum_i\left|\widehat{\lambda}_p^\top\widehat{\mathcal{U}}_i\right|^3 = O_p\left(\sqrt{\log(n)}\upsilon_n^\ddagger\right)$. By these results and $\max_i\left|\widehat{\lambda}_p^\top\widehat{\mathcal{U}}_i\right| = o_p(1)$, $S\left(\widehat{\lambda}_p,\widehat{\vartheta}_p\right) = 2\widehat{\lambda}_p^\top\left(\sum_i\widehat{\mathcal{U}}_i\right) - \sum_i\left(\widehat{\lambda}_p^\top\widehat{\mathcal{U}}_i\right)^2 + O_p\left(\sqrt{\log(n)}\upsilon_n^\ddagger\right)$. It was shown in the proof of Lemma 3 that $\widehat{\mathcal{U}} = \widehat{\Delta}_{\mathcal{U}\mathcal{U}^\top}\left(\sqrt{nh}\widehat{\lambda}_p\right) + O_p\left(\upsilon_n^\ddagger\right)$. It follows from these results, Lemma 2 and $\widehat{\Delta}_{\mathcal{U}\mathcal{U}^\top} - \Delta_{\mathcal{U}\mathcal{U}^\top} = O_p\left(\sqrt{\log(n)/\underline{n}} + \log(n)\left(\overline{n}^{1/6}/\underline{n}\right)\right)$ that $S\left(\widehat{\lambda}_p,\widehat{\vartheta}_p\right) = \left(\sqrt{nh}\widehat{\lambda}_p\right)^\top\Delta_{\mathcal{U}\mathcal{U}^\top}\left(\sqrt{nh}\widehat{\lambda}_p\right) + O_p\left(\sqrt{\log(n)}\upsilon_n^\dagger\right)$. By Lemma 3 and $\overline{\mathcal{U}} = O_p\left(\sqrt{\log(n)}\right)$, $S\left(\widehat{\lambda}_p,\widehat{\vartheta}_p\right) = \overline{\mathcal{U}}^\top\mathrm{Q}\overline{\mathcal{U}} + O_p\left(\sqrt{\log(n)}\upsilon_n^\dagger\right)$. Similarly, we have $S\left(\widetilde{\lambda}_p,\vartheta_0,\widetilde{\vartheta}_p\right) = \overline{\mathcal{U}}^\top\mathrm{Q}_\dagger\overline{\mathcal{U}} + O_p\left(\sqrt{\log(n)}\upsilon_n^\dagger\right)$. By definition, $LR_p(\vartheta_0 \mid h) = S\left(\widetilde{\lambda}_p,\vartheta_0,\widetilde{\vartheta}_p\right) - S\left(\widehat{\lambda}_p,\widehat{\vartheta}_p\right)$. Therefore, $LR_p(\vartheta_0 \mid h) =$

$\overline{\mathcal{U}}^\top (\mathrm{Q}_\dagger - \mathrm{Q})\overline{\mathcal{U}} + O_p\left(\sqrt{\log(n)}\upsilon_n^\dagger\right)$. Then, by straightforward algebraic calculations,

$$\begin{aligned}
\mathrm{Q}_\dagger - \mathrm{Q} &= \Delta_{\mathcal{U}\mathcal{U}^\top}^{-1}\left\{\Delta_{\mathcal{G}}\left(\Delta_{\mathcal{G}}^\top \Delta_{\mathcal{U}\mathcal{U}^\top}^{-1}\Delta_{\mathcal{G}}\right)^{-1}\Delta_{\mathcal{G}}^\top - \Delta_{\mathcal{G}_\dagger}\left(\Delta_{\mathcal{G}_\dagger}^\top \Delta_{\mathcal{U}\mathcal{U}^\top}^{-1}\Delta_{\mathcal{G}_\dagger}\right)^{-1}\Delta_{\mathcal{G}_\dagger}^\top\right\}\Delta_{\mathcal{U}\mathcal{U}^\top}^{-1} \\
&= \Delta_{\mathcal{U}\mathcal{U}^\top}^{-1}\left(\Delta_{\mathcal{G}_0} - \Delta_{\mathcal{G}_\dagger}\Phi_{\dagger\dagger}^{-1}\Phi_{\dagger 0}\right)\left(\Phi_{00} - \Phi_{0\dagger}\Phi_{\dagger\dagger}^{-1}\Phi_{\dagger 0}\right)^{-1}\left(\Delta_{\mathcal{G}_0}^\top - \Phi_{0\dagger}\Phi_{\dagger\dagger}^{-1}\Delta_{\mathcal{G}_\dagger}^\top\right)\Delta_{\mathcal{U}\mathcal{U}^\top}^{-1}. \qquad \text{(S7)}
\end{aligned}$$

Then by this result, (12), (14) and $\left(\Phi_{00} - \Phi_{0\dagger}\Phi_{\dagger\dagger}^{-1}\Phi_{\dagger 0}\right)^{-1} = \Sigma_\Delta / \left(\Delta_{\mathcal{D}_+}/\Delta_+ - \Delta_{\mathcal{D}_-}/\Delta_-\right)^2$

$$\begin{aligned}
\overline{\mathcal{U}}^\top (\mathrm{Q}_\dagger - \mathrm{Q})\overline{\mathcal{U}} &= \left\{\mathrm{e}_{d_u,1}^\top \Phi_\pm^{-1}\left(\overline{\mathcal{U}}_+/\Delta_+ - \overline{\mathcal{U}}_-/\Delta_-\right)\right\}^2 \Sigma_\Delta \\
&= \left\{\left(\overline{\mathcal{M}}_+/\Delta_+ - \overline{\mathcal{M}}_-/\Delta_-\right) - \left(\overline{\mathcal{Z}}_+/\Delta_+ - \overline{\mathcal{Z}}_-/\Delta_-\right)^\top \gamma_\Delta\right\}^2 / \Sigma_\Delta. \qquad \text{(S8)}
\end{aligned}$$

By using $\gamma_\Delta = \gamma_{\mathsf{adj}} + O\left(\overline{h}\right)$ and (11), $\Sigma_\Delta = \Delta_{\mathcal{E}^2}/\varphi^2 + O\left(\overline{h}\right)$. By $\left\|\overline{\mathcal{U}}\right\| = O_p\left(\sqrt{\log(n)}\right)$ and $\gamma_\Delta = \gamma_{\mathsf{adj}} + O\left(\overline{h}\right)$, the numerator on the right hand side of the second equality in (S8) is $\left\{(nh)^{-1/2}\sum_i \mathcal{E}_i\right\}^2 + O_p\left(\log(n)\overline{h}\right)$. Let $\tilde{q}(T_i, X_i \mid h) \coloneqq h^{-1/2}\mathcal{E}_i/\sqrt{\Delta_{\mathcal{E}^2}}$ and $\tilde{\mathfrak{Q}} \coloneqq \{\tilde{q}(\cdot \mid h) : h \in \mathbb{H}\}$. Then it is clear that $\left\{(nh)^{-1/2}\sum_i \mathcal{E}_i\right\}^2 / \Delta_{\mathcal{E}^2} = \left\{\mathbb{G}_n^T \tilde{q}(\cdot \mid h)\right\}^2$ and therefore, $LR_p(\vartheta_0 \mid h) = \left\{\mathbb{G}_n^T \tilde{q}(\cdot \mid h)\right\}^2 + O_p\left(\log(n)\overline{h} + \sqrt{\log(n)}\upsilon_n^\dagger\right)$. Also denote $\mathfrak{Q} \coloneqq \{q(\cdot \mid h) : h \in \mathbb{H}\}$ and $\mathfrak{D} \coloneqq \{q(\cdot \mid h) - \tilde{q}(\cdot \mid h) : h \in \mathbb{H}\}$. By similar arguments as in the proof of Lemma 1, $\tilde{\mathfrak{Q}}$ and $\mathfrak{Q}$ are both VC-type with respect to the envelopes $\left(F_{\tilde{\mathfrak{Q}}}, F_{\mathfrak{Q}}\right)$ satisfying $F_{\tilde{\mathfrak{Q}}}(T_i, X_i) \propto \underline{h}^{-1/2}1\left(|X_i| \leq \overline{h}\right)|\epsilon_i - \mu_\epsilon| / \sqrt{\inf_{h \in \mathbb{H}}\Delta_{\mathcal{E}^2}}$ and $F_{\mathfrak{Q}}(T_i, X_i) \propto \underline{h}^{-1/2}1\left(|X_i| \leq \overline{h}\right)|\epsilon_i - \mu_\epsilon| / \sqrt{\xi(|X_i|) f_{|X|}(|X_i|)}$, respectively. By change of variables, $\mathbb{P}^T F_{\tilde{\mathfrak{Q}}}^{12} \asymp \mathbb{P}^T F_{\mathfrak{Q}}^{12} = O\left(\overline{h}/\underline{h}^6\right)$. By Chernozhukov et al. (2014, Lemma A.6), $\mathfrak{D}$ is VC-type with respect to the envelope $F_{\mathfrak{D}} = F_{\tilde{\mathfrak{Q}}} + F_{\mathfrak{Q}}$. Let

$$\sigma_{\mathfrak{D}}^2 \coloneqq \sup_{f \in \mathfrak{D}} \mathbb{P}^T f^2 = \sup_{h \in \mathbb{H}} \mathrm{E}\left[\left(q(T, X \mid h) - \tilde{q}(T, X \mid h)\right)^2\right].$$

By LIE and the fact that $\left(W_{p;+} + W_{p;-}\right)^2 = \mathcal{K}_{p;+}(|X|/h)$,

$$\begin{aligned}
\mathrm{E}\left[\left(q(T, X \mid h) - \tilde{q}(T, X \mid h)\right)^2\right] &= \mathrm{E}\left[\frac{1}{h}\left(W_{p;+} + W_{p;-}\right)^2 (\epsilon - \mu_\epsilon)^2 \left(\frac{1}{\sqrt{\Delta_{\mathcal{E}^2}}} - \frac{1}{\sqrt{\xi(|X|) f_{|X|}(|X|)\,\omega_{p;+}^{0,2}}}\right)^2\right] \\
&= \int_0^\infty \frac{1}{h}\mathcal{K}_{p;+}\left(\frac{z}{h}\right)^2 \left(\sqrt{\frac{\xi(z) f_{|X|}(z)}{\Delta_{\mathcal{E}^2}}} - \frac{1}{\sqrt{\omega_{p;+}^{0,2}}}\right)^2 \mathrm{d}z. \qquad \text{(S9)}
\end{aligned}$$

Note that $\Delta_{\mathcal{E}^2} = \int_0^\infty h^{-1}\mathcal{K}_{p;+}(z/h)^2 \xi(z) f_{|X|}(z)\,\mathrm{d}z$ and therefore, it follows from mean value expansion and (S9) that $\sigma_{\mathfrak{D}}^2 = O\left(\overline{h}^2\right)$. By Chen and Kato (2020, Corollary 5.5), $\mathrm{E}\left[\left\|\mathbb{G}_n^T\right\|_{\mathfrak{D}}\right] \lesssim \sigma_{\mathfrak{D}}\sqrt{\log(n)} + \log(n)\left\|F_{\mathfrak{D}}\right\|_{\mathbb{P}^T,12} n^{1/12}/\sqrt{n}$ and therefore, $\mathrm{E}\left[\left\|\mathbb{G}_n^T\right\|_{\mathfrak{D}}\right] = O\left(\sqrt{\log(n)} \cdot \overline{h} + \log(n)\left(\overline{n}^{1/12}/\underline{n}^{1/2}\right)\right)$. Let $\sigma_{\tilde{\mathfrak{Q}}}^2 \coloneqq \sup_{f \in \tilde{\mathfrak{Q}}} \mathbb{P}^T f^2$ and $\sigma_{\mathfrak{Q}}^2 \coloneqq \sup_{f \in \mathfrak{Q}} \mathbb{P}^T f^2$. It is easy to see that $\mathbb{P}^T f^2 = 1$, if $f \in \mathfrak{Q}$ or $f \in \tilde{\mathfrak{Q}}$ and therefore,

$\sigma^2_{\tilde{\mathfrak{Q}}} = \sigma^2_{\mathfrak{Q}} = 1$. Similarly, $\mathrm{E}\left[\left\|\mathbb{G}_n^T\right\|_{\tilde{\mathfrak{Q}}}\right] \lesssim \sigma_{\tilde{\mathfrak{Q}}}\sqrt{\log(n)} + \log(n)\left\|F_{\tilde{\mathfrak{Q}}}\right\|_{\mathbb{P}^T,12} n^{1/12}/\sqrt{n}$ and a similar inequality with $\tilde{\mathfrak{Q}}$ replaced by $\mathfrak{Q}$ holds. Therefore, $\mathrm{E}\left[\left\|\mathbb{G}_n^T\right\|_{\tilde{\mathfrak{Q}}}\right] \asymp \mathrm{E}\left[\left\|\mathbb{G}_n^T\right\|_{\mathfrak{Q}}\right] = O\left(\sqrt{\log(n)}\right)$. Then it follows from Markov's inequality that $\left\{\mathbb{G}_n^T \tilde{q}(\cdot \mid h)\right\}^2 - \left\{\mathbb{G}_n^T q(\cdot \mid h)\right\}^2 = O_p\left(\log(n)\,\overline{h} + \log(n)^{3/2}\left(\overline{n}^{1/12}/\underline{n}^{1/2}\right)\right)$. The conclusion follows from this result and $LR_p(\vartheta_0 \mid h) = \left\{\mathbb{G}_n^T \tilde{q}(\cdot \mid h)\right\}^2 + O_p\left(\log(n)\,\overline{h} + \sqrt{\log(n)}\upsilon_n^\dagger\right)$. ∎

Proof of Lemma 5. Let $r_n := \sqrt{\bar{n}/\log(n)}$, $\overline{V}_i := V_i 1(V_i > r_n)$, $\underline{V}_i := V_i 1(V_i \leq r_n)$ and $\left(\underline{V}, \overline{V}\right)$ be defined similarly. Then write $\bar{n}^{-1/2}\sum_i \left(W^k_{p;\mathrm{s},i} V_i - \mathrm{E}\left[W^k_{p;\mathrm{s}} V\right]\right) = \overline{\mathcal{W}} + \underline{\mathcal{W}}$, where $\overline{\mathcal{W}} := \bar{n}^{-1/2}\sum_i \left(W^k_{p;\mathrm{s},i}\overline{V}_i - \mathrm{E}\left[W^k_{p;\mathrm{s}}\overline{V}\right]\right)$ and $\underline{\mathcal{W}} := \bar{n}^{-1/2}\sum_i \left(W^k_{p;\mathrm{s},i}\underline{V}_i - \mathrm{E}\left[W^k_{p;\mathrm{s}}\underline{V}\right]\right)$. Let $\sigma^2_{\underline{\mathcal{W}}} := \mathrm{Var}\left[h^{-1/2} W^k_{p;\mathrm{s}}\underline{V}\right]$. By $\sigma^2_{\underline{\mathcal{W}}} \leq \mathrm{E}\left[h^{-1} W^{2k}_{p;\mathrm{s}}\underline{V}^2\right]$, LIE and change of variables, $\sigma^2_{\underline{\mathcal{W}}} = O(1)$. $\left|W^k_{p;\mathrm{s},i}\underline{V}_i - \mathrm{E}\left[W^k_{p;\mathrm{s}}\underline{V}\right]\right|$ is bounded by an upper bound that is proportional to r_n. Let $c > 0$ denote an arbitrary positive constant. By Giné and Nickl (2015, Theorem 3.1.7 and Equation 3.24) with $u = \log(n^c)$, $\Pr\left[|\underline{\mathcal{W}}| \geq \left(\sqrt{2c\sigma^2_{\underline{\mathcal{W}}}} + c/3\right)\sqrt{\log(n)}\right] \leq 2n^{-c}$. By $\sigma^2_{\underline{\mathcal{W}}} = O(1)$ and taking c to be sufficiently large, $\underline{\mathcal{W}} = O_p^\star\left(\sqrt{\log(n)}\right)$. By Markov's inequality, the fact that $\overline{V}^2 \leq \overline{V}^2 |V/r_n|^3$ and change of variables, $\Pr\left[|\overline{\mathcal{W}}| \geq \sqrt{\log(n)}\right] \leq \mathrm{E}\left[h^{-1} W^{2k}_{p;\mathrm{s}}\overline{V}^2\right]/\log(n) \leq \mathrm{E}\left[h^{-1} W^{2k}_{p;\mathrm{s}}|V|^5\right]/\left(r_n^3 \cdot \log(n)\right) = O\left(\log(n)/\bar{n}^{3/2}\right)$ and therefore, $\overline{\mathcal{W}} = O_p^\star\left(\sqrt{\log(n)}\right)$. ∎

Proof of Lemma 6. By Markov's inequality, $\Pr\left[\bar{n}^{-1}\sum_i \|\mathcal{U}_i\|^5 > \Delta_{\|\mathcal{U}\|^5} + c\right]$ is bounded above by the fourth central moment of $\bar{n}^{-1}\sum_i \|\mathcal{U}_i\|^5$ divided by c^4, where $c > 0$ is an arbitrary positive constant. By straightforward calculation and change of variables, its fourth central moment is bounded above by $3n^{-2}\left(\mathrm{E}\left[h^{-2}\|\mathcal{U}\|^{10}\right]\right)^2 + n^{-3}\mathrm{E}\left[h^{-4}\|\mathcal{U}\|^{20}\right] = O\left(\bar{n}^{-2}\right)$. Therefore, $\bar{n}^{-1}\sum_i \|\mathcal{U}_i\|^5 = O_p^\star(1)$ and by $\max_i \|\mathcal{U}_i\| \leq \left(\sum_i \|\mathcal{U}_i\|^5\right)^{1/5}$, $\max_i \|\mathcal{U}_i\| = O_p^\star\left(\bar{n}^{1/5}\right)$. Then, by this result and the definition of $\mathcal{L}_\sharp$, $\Pr\left[\max_i \sup_{\lambda \in \mathcal{L}_\sharp}\left|\lambda^\top \mathcal{U}_i\right| \geq 1/2\right]$ is bounded above by $\Pr\left[\max_i \|\mathcal{U}_i\| \geq \left(\sqrt{\bar{n}}/\log(n)\right)/2\right] = O\left(\bar{n}^{-2}\right)$. Therefore, $\mathcal{L}_\sharp \subseteq \mathcal{L}(\vartheta)$ wp* and $\lambda_\sharp := \mathrm{argmax}_{\lambda \in \mathcal{L}_\sharp} S(\lambda, \vartheta)$ exists wp*. By using $\overline{\mathcal{U}} = O_p^\star\left(\sqrt{\log(n)}\right)$ and $\overline{\Delta}_{\mathcal{U}\mathcal{U}^\top} - \Delta_{\mathcal{U}\mathcal{U}^\top} = O_p^\star\left(\sqrt{\log(n)/\bar{n}}\right)$, which follow from Lemma 5, and repeating the steps in the proof of Lemma 2, $\sqrt{\bar{n}}\lambda_\sharp = O_p^\star\left(\sqrt{\log(n)}\right)$. Then, $\sqrt{\bar{n}}\lambda_\sharp \leq \log(n)/2$ wp* and $S(\lambda_\sharp, \vartheta) = \sup_{\lambda \in \mathcal{L}(\vartheta)} S(\lambda, \vartheta) = O_p^\star(\log(n))$. By similar arguments, boundedness of Θ and $\widehat{\mathcal{U}}_i = \mathcal{U}_i - \mathcal{G}_i \widehat{\eta}_p$, $\max_i \left\|\widehat{\mathcal{U}}_i\right\| = O_p^\star\left(\bar{n}^{1/5}\right)$ and $\bar{n}^{-1}\sum_i \left\|\widehat{\mathcal{U}}_i\right\|^2 = O_p^\star(1)$. By repeating the steps in the proof of Lemma 2, $\sqrt{\log(n)}\left\|\widehat{\mathcal{U}}\right\| \leq \sup_{\lambda \in \mathcal{L}(\vartheta)} S(\lambda, \vartheta) + 2\left(\bar{n}^{-1}\sum_i \left\|\widehat{\mathcal{U}}_i\right\|^2\right)\log(n) = O_p^\star(\log(n))$. (a) follows from (S4), $\overline{\mathcal{U}} = O_p^\star\left(\sqrt{\log(n)}\right)$, $\widehat{\mathcal{U}} = O_p^\star\left(\sqrt{\log(n)}\right)$ and the fact that $\varrho_{\min}\left(\overline{\Delta}_{\mathcal{G}}^\top \overline{\Delta}_{\mathcal{G}}\right)$ is bounded away from zero wp*, which follows from Lemmas 1 and 5. The proof of (b) parallels that of Lemma 2(b) and uses the fact $\widehat{\Delta}_{\mathcal{U}\mathcal{U}^\top} - \Delta_{\mathcal{U}\mathcal{U}^\top} = O_p^\star\left(\sqrt{\log(n)/\bar{n}}\right)$. (c) and (d) follow from similar arguments. ∎

Proof of Lemma 7. A decomposition $LR^\star = \bar{n}\left(\tilde{R}_1^2 + 2\tilde{R}_1\tilde{R}_2 + 2\tilde{R}_1\tilde{R}_3 + \tilde{R}_2^2\right)$ can be derived. $\tilde{R}_k$ is a homogeneous k-th order polynomial of $\left(A^{\mathsf{k}}, A^{\mathsf{kl}}, A^{\mathsf{klm}}, C^{\mathsf{k},n}, C^{\mathsf{k;l},n}\right)$ so that $\tilde{R}_1 = O_p^\star\left(\sqrt{\log(n)/\bar{n}}\right)$, $\tilde{R}_2 = O_p^\star(\log(n)/\bar{n})$ and $\tilde{R}_3 = O_p^\star\left((\log(n)/\bar{n})^{3/2}\right)$. $-$M is a projection matrix onto the orthogonal complement of the column space of $\Pi_\dagger$. Let ϖ_0 be a vector spanning the one-dimensional orthogonal complement of the column space of $\Pi_\dagger$ so that $-\mathrm{M} = \varpi_0 \left(\varpi_0^\top \varpi_0\right)^{-1} \varpi_0^\top$. Let $\varpi := \varpi_0 / \sqrt{\varpi_0^\top \varpi_0}$. Then, $\varpi^\top \varpi = 1$ and $-\mathrm{M} = \varpi \varpi^\top$. The expressions of $\left(\tilde{R}_1, \tilde{R}_2, \tilde{R}_3\right)$ can be readily obtained in a special case of Ma (2017). Algebraic calculations in Ma (2017) show that by setting $\tilde{R}_1 := \varpi^{(k)} A^k$,

$$\tilde{R}_2 := \frac{1}{2}\mathrm{M}^{(mk)}\varpi^{(n)}A^{mn}A^k - \varpi^{(n)}A^{n\,d_\vartheta+a}A^{d_\vartheta+a} + \left\{\frac{1}{3}\alpha^{vmn}\mathrm{M}^{(vl)}\mathrm{M}^{(mk)}\varpi^{(n)} - \gamma^{m;v,o}\Omega^{(on)}\mathrm{P}^{(nk)}\mathrm{M}^{(ml)}\varpi^{(v)}\right\}$$
$$\times A^l A^k + \left\{\left(\gamma^{d_\vartheta+a;v,m}[d_\vartheta+a,v]\right)\Omega^{(mo)}\mathrm{P}^{(ok)}\varpi^{(v)} - \alpha^{vm\,d_\vartheta+a}\mathrm{M}^{(vk)}\varpi^{(m)}\right\} A^k A^{d_\vartheta+a} - \Omega^{(ko)}\mathrm{P}^{(om)}\varpi^{(l)}$$
$$\times C^{l,k}A^m + \left\{\alpha^{v\,d_\vartheta+a\,d_\vartheta+b}\varpi^{(v)} - \gamma^{d_\vartheta+a;d_\vartheta+b,m}\Omega^{(mn)}\varpi^{(n)}\right\} A^{d_\vartheta+a}A^{d_\vartheta+b} + \Omega^{(km)}\varpi^{(m)}C^{d_\vartheta+a,k}A^{d_\vartheta+a}, \quad \text{(S10)}$$

where $\gamma^{d_\vartheta+a;v,m}[d_\vartheta+a,v]$ denotes $\gamma^{d_\vartheta+a;v,m} + \gamma^{v;d_\vartheta+a,m}$ and $\tilde{R}_3$ to be given by the formula provided in Ma (2017, Appendix D.3), we have $LR^\star = \bar{n}\left(\tilde{R}_1^2 + 2\tilde{R}_1\tilde{R}_2 + 2\tilde{R}_1\tilde{R}_3 + \tilde{R}_2^2\right)$. (S10) is formally the same as Ma (2017, (D.2)) with terms that depend on the second derivatives removed. The expression of $\tilde{R}_3$ is also essentially the same as that of R_3 in Ma (2017, Appendix D.3) with terms that depend on the higher-order derivatives removed and hence omitted for brevity.

Let $\alpha^{\mathsf{k}} := \Delta_{\mathcal{V}^{(\mathsf{k})}}$ and $\mathring{A}^{\mathsf{k}} := A^{\mathsf{k}} - \alpha^{\mathsf{k}}$. By replacing A^{k} with $\mathring{A}^{\mathsf{k}} + \alpha^{\mathsf{k}}$, we have $\tilde{R}_1 = \tilde{R}_{10} + \tilde{R}_{11}$, where $\tilde{R}_{10} := \varpi^{(k)}\alpha^k$ and $\tilde{R}_{11} := \varpi^{(k)}\mathring{A}^k$. Similarly, we replace A^{k} with $\mathring{A}^{\mathsf{k}} + \alpha^{\mathsf{k}}$ to decompose $\tilde{R}_2 = \tilde{R}_{22} + \tilde{R}_{21} + \tilde{R}_{20}$ so that $\tilde{R}_{2k}$ is a homogeneous $(2-k)$ −th order polynomial of $\alpha^1, ..., \alpha^{2d_u}$:

$$\tilde{R}_{21} := \frac{1}{2}\mathrm{M}^{(mk)}\varpi^{(n)}A^{mn}\alpha^k - \varpi^{(n)}A^{n\,d_\vartheta+a}\alpha^{d_\vartheta+a} + \frac{2}{3}\alpha^{vmn}\mathrm{M}^{(vl)}\mathrm{M}^{(mk)}\varpi^{(n)}\mathring{A}^l\alpha^k - \gamma^{m;v,o}\Omega^{(on)}\mathrm{P}^{(nk)}\mathrm{M}^{(ml)}\varpi^{(v)}$$
$$\times\left(\mathring{A}^l\alpha^k[l,k]\right) + \left\{\left(\gamma^{d_\vartheta+a;v,m}[d_\vartheta+a,v]\right)\Omega^{(mo)}\mathrm{P}^{(ok)}\varpi^{(v)} - \alpha^{vm\,d_\vartheta+a}\mathrm{M}^{(vk)}\varpi^{(m)}\right\}\left(\alpha^k\mathring{A}^{d_\vartheta+a}[k,d_\vartheta+a]\right)$$
$$- \Omega^{(ko)}\mathrm{P}^{(om)}\varpi^{(l)}C^{l,k}\alpha^m + \left\{\alpha^{v\,d_\vartheta+a\,d_\vartheta+b}\varpi^{(v)} - \gamma^{d_\vartheta+a;d_\vartheta+b,m}\Omega^{(mn)}\varpi^{(n)}\right\}\left(\alpha^{d_\vartheta+a}\mathring{A}^{d_\vartheta+b}[d_\vartheta+a,d_\vartheta+b]\right)$$
$$+ \Omega^{(km)}\varpi^{(m)}C^{d_\vartheta+a,k}\alpha^{d_\vartheta+a}, \quad \text{(S11)}$$

$\tilde{R}_{22}$ is defined by the right hand side of (S10) with A^{k} replaced by $\mathring{A}^{\mathsf{k}}$ and $\tilde{R}_{20} := \tilde{R}_2 - \tilde{R}_{22} - \tilde{R}_{21} = O\left(\|\Delta_{\mathcal{U}}\|^2\right)$. Let $R_0 := \tilde{R}_{10} + \tilde{R}_{20}$, $R_1 := \tilde{R}_{11} + \tilde{R}_{21}$ and $R_2 := \tilde{R}_{22}$. We decompose $\tilde{R}_3 = \tilde{R}_{33} + \tilde{R}_{32} + \tilde{R}_{31} + \tilde{R}_{30}$ in a similar manner and let $R_3 := \tilde{R}_{33}$. R_3 is given by the formula of $\tilde{R}_3$ with A^{k} replaced by $\mathring{A}^{\mathsf{k}}$. Then, let $R := R_1 + R_2 + R_3$. By Lemma 5, $\tilde{R}_1 + \tilde{R}_2 + \tilde{R}_3 = R_0 + R + O_p^\star\left(\|\Delta_{\mathcal{U}}\| \log(n)/\bar{n}\right)$ and therefore, $LR^\star = \bar{n}\left(R_0 + R\right)^2 + O_p^\star\left(\upsilon_n^\sharp\right)$.

Let $\mathcal{F} := \left(W_p \otimes U, W_p, \left(W_{p;+}^2, W_{p;-}^2\right)^\top \otimes \left(U, U^2\right), \left(W_{p;+}^3, W_{p;-}^3\right)^\top \otimes U^3\right)$. $\mathcal{F}_i$ is defined analogously and let d_f denote the dimension of $\mathcal{F}$. It can be shown that $\sqrt{\bar{n}}R := h_n\left(\overline{\mathcal{F}}\right)$, where $\overline{\mathcal{F}} := \bar{n}^{-1/2}\sum_i\left(\mathcal{F}_i - \mathrm{E}\left[\mathcal{F}\right]\right)$ and h_n is a cubic polynomial. E.g., $\sqrt{\bar{n}}\varpi^{(k)}\mathring{A}^k = \tilde{\varpi}^\top\Delta_{\mathcal{U}\mathcal{U}^\top}^{-1/2}\left(\bar{n}^{-1/2}\sum_i\left(\mathcal{U}_i - \mathrm{E}\left[\mathcal{U}\right]\right)\right)$, where $\tilde{\varpi} := \mathrm{S}\begin{bmatrix} \varpi^\top & 0_{d_z}^\top \end{bmatrix}^\top$. It can be shown that other terms on the right hand side of (S11) can also be written as linear functions of $\overline{\mathcal{F}}$. Similarly, it can be shown by tedious algebra that $\sqrt{\bar{n}}R_2$ and $\sqrt{\bar{n}}R_3$ are homogenous quadratic and cubic polynomials of $\overline{\mathcal{F}}$. A more lucid proof of this fact uses the observation that $\ell_p\left(\widehat{\vartheta}_p \mid h\right) = \inf_{\theta_2}\sup_{\lambda_2} 2\sum_i \log\left(1+\lambda_2^\top \bar{\mathcal{U}}_i\left(\theta_2\right)\right)$. Let $\mathcal{M}_i\left(\theta_0,\theta_1\right) \coloneqq W_{p,i} \otimes \left(Y_i - \theta_0 D_i - \theta_1\right)$. By rearranging the moment conditions, $\ell_p\left(\widehat{\vartheta}_p \mid h\right) = \inf_{\theta_0,\theta_1,\theta_2}\sup_{\lambda_1,\lambda_2} 2\sum_i \log\left(1+\lambda_1^\top \mathcal{M}_i\left(\theta_0,\theta_1\right) + \lambda_2^\top \bar{\mathcal{U}}_i\left(\theta_2\right)\right)$. Let $\mathcal{W}_i\left(\theta\right) \coloneqq \left(\mathcal{M}_i\left(\theta_0,\theta_1\right), \bar{\mathcal{U}}_i\left(\theta_2\right)\right)$. $\widehat{\vartheta}_p$ and $\widehat{\lambda} \coloneqq \left(\widehat{\lambda}_1,\widehat{\lambda}_2\right)$ satisfy the first-order conditions wp*:

$$\sum_i \frac{\mathcal{M}_i\left(\widehat{\vartheta}_{p,0},\widehat{\vartheta}_{p,1}\right)}{1+\widehat{\lambda}^\top \mathcal{W}_i\left(\widehat{\vartheta}_p\right)} = 0_2,\ \sum_i \frac{\bar{\mathcal{U}}_i\left(\widehat{\vartheta}_{p,2}\right)}{1+\widehat{\lambda}^\top \mathcal{W}_i\left(\widehat{\vartheta}_p\right)} = 0_{2d_z},\ \sum_i \frac{\left(W_{p,i}\otimes\left(D_i,1\right)\right)^\top \widehat{\lambda}_1}{1+\widehat{\lambda}^\top \mathcal{W}_i\left(\widehat{\vartheta}_p\right)} = 0_2,\ \sum_i \frac{\bar{\mathcal{G}}_i^\top \widehat{\lambda}_2}{1+\widehat{\lambda}^\top \mathcal{W}_i\left(\widehat{\vartheta}_p\right)} = 0_{d_z}.$$

The third condition implies that $\widehat{\lambda}_1 = 0_2$ wp*. Therefore, $\ell_p\left(\widehat{\vartheta}_p \mid h\right) = 2\sum_i \log\left(1+\widehat{\lambda}_2^\top \bar{\mathcal{U}}_i\left(\widehat{\vartheta}_{p,2}\right)\right)$ and the second and fourth conditions are $\sum_i \bar{\mathcal{U}}_i\left(\widehat{\vartheta}_{p,2}\right) / \left(1+\widehat{\lambda}_2^\top \bar{\mathcal{U}}_i\left(\widehat{\vartheta}_{p,2}\right)\right) = 0_{2d_z}$ and $\sum_i \bar{\mathcal{G}}_i^\top \widehat{\lambda}_2 / \left(1+\widehat{\lambda}_2^\top \bar{\mathcal{U}}_i\left(\widehat{\vartheta}_{p,2}\right)\right) = 0_{d_z}$, which coincide with the first-order conditions of $\inf_{\theta_2}\sup_{\lambda_2}\sum_i \log\left(1+\lambda_2^\top \bar{\mathcal{U}}_i\left(\theta_2\right)\right)$. Therefore, we have $\ell_p\left(\widehat{\vartheta}_p \mid h\right) = \inf_{\theta_2}\sup_{\lambda_2} 2\sum_i \log\left(1+\lambda_2 \bar{\mathcal{U}}_i\left(\theta_2\right)\right)$. By expansion and Lemma 6, we get approximations for $\widehat{\lambda}_2$, $\widehat{\vartheta}_{p,2}$ and $\ell_p\left(\widehat{\vartheta}_p \mid h\right)$ which are similar to (27) and (28). Then it is clear that by replacing sample averages with sums of their centered versions and population counterparts we can get further approximations which are polynomials in $\bar{n}^{-1/2}\sum_i\left(\bar{\mathcal{F}}_i - \mathrm{E}\left[\bar{\mathcal{F}}\right]\right)$, where $\left(\bar{\mathcal{F}}_i,\bar{\mathcal{F}}\right)$ are defined by the formulae of $\left(\mathcal{F}_i,\mathcal{F}\right)$ with $\left(U_i,U\right)$ replaced by $\left(\bar{U}_i,\bar{U}\right)$. Similarly, the stochastic expansion of $\ell_p\left(\vartheta_0,\widetilde{\vartheta}_p \mid h\right)$ should involve only terms in $\overline{\mathcal{F}}$.

Let $\kappa_j\left(V\right)$ denote the j-th cumulant of a random variable V. We follow arguments in the proof of Calonico et al. (2018, Theorem S.1) and apply Skovgaard (1986, Theorem 3.4) with $s=4$ to $S_n \coloneqq \mathrm{B}^{-1/2}\overline{\mathcal{F}}$ where $\mathrm{B} \coloneqq \mathrm{Var}\left[\mathcal{F}\right]/h$. For any $t \in \mathbb{R}^{d_f}$ with $\|t\|=1$, by change of variables and calculation of the moments (see, e.g., DiCiccio et al., 1988, Page 12), $\kappa_3\left(t^\top S_n\right) = \mathrm{E}\left[\left(t^\top S_n\right)^3\right] = O\left(\bar{n}^{-1/2}\right)$, $\kappa_4\left(t^\top S_n\right) = \mathrm{E}\left[\left(t^\top S_n\right)^4\right] - 3\left(\mathrm{E}\left[\left(t^\top S_n\right)^2\right]\right)^2 = O\left(\bar{n}^{-1}\right)$ and $\rho_{s,n}\left(t\right) \coloneqq \max\left\{\left|\kappa_3\left(t^\top S_n\right)\right|/3!, \sqrt{\left|\kappa_4\left(t^\top S_n\right)\right|/4!}\right\} = O\left(\bar{n}^{-1/2}\right)$, uniformly in t. Condition I and II of Skovgaard (1986, Theorem 3.4) are satisfied by taking $a_n\left(t\right) \propto \sqrt{\bar{n}}$ and $\epsilon_n = \bar{n}^{-3/2}$. Let $\hat{\Psi}_V\left(t\right) \coloneqq \mathrm{E}\left[\exp\left(\mathrm{i}t^\top V\right)\right]$ denote the characteristic function of a random vector V, where $\mathrm{i} \coloneqq \sqrt{-1}$. Let $\mathcal{F}_{\mathrm{s}} \coloneqq \left(W_{p;\mathrm{s}}U, W_{p;\mathrm{s}}, W_{p;\mathrm{s}}^2\left(U,U^2\right), W_{p;\mathrm{s}}^3 U^3\right)$, $\mathrm{s} \in \{-,+\}$. Then, $\hat{\Psi}_{\mathcal{F}}\left(t\right) = \mathrm{E}\left[\exp\left(\mathrm{i}t_+^\top \mathcal{F}_+\right)1\left(X \geq 0\right)\right] + \mathrm{E}\left[\exp\left(\mathrm{i}t_-^\top \mathcal{F}_-\right)1\left(X < 0\right)\right]$, where $\left(t_-,t_+\right)$ denote corresponding coordinates of t. By change of variables, $\mathrm{E}\left[\exp\left(\mathrm{i}t_+^\top \mathcal{F}_+\right)1\left(X \geq 0\right)\right] = h\left(f_X\left(0\right)E_+\left(t_+\right) + O\left(h\right)\right) + \Pr\left[X > h\right]$, where E_+ is the characteristic function of $\mathcal{K}_{p;+}\left(V\right)\left(U,1\right)$, $\mathcal{K}_{p;+}\left(V\right)^2\left(U,U^2\right)$, $\mathcal{K}_{p;+}\left(V\right)^3 U^3$, where $\left(V,U\right)$ has the joint density given by $\left(v,u\right) \mapsto 1\left(0 \leq v \leq 1\right) f_{U|X}\left(u \mid 0\right)$. A similar result holds for $\mathrm{E}\left[\exp\left(\mathrm{i}t_-^\top \mathcal{F}_-\right)1\left(X < 0\right)\right]$ with $E_-\left(t_-\right)$ defined similarly. Therefore, $\hat{\Psi}_{\mathcal{F}}\left(t\right) = 1 - \Pr\left[-h < X \leq h\right] + hf_X\left(0\right)\left(E_+\left(t_+\right) + E_-\left(t_-\right)\right) + O\left(h^2\right)$. By Assumption 5, the vector-valued functions $\left(v,u\right) \mapsto \left(1, \left(\mathcal{K}_{p;+}\left(v\right), \mathcal{K}_{p;+}\left(v\right)^2, \mathcal{K}_{p;+}\left(v\right)^3\right) \otimes \left(1,u,u^2,u^3\right)\right)$ are linearly independent. By invoking the same arguments as in the proof of Calonico et al. (2018, Lemma S.9), $\forall\varepsilon > 0$, $\exists c_\varepsilon > 0$ such that $\sup_{\|t\|>\varepsilon}\left|E_+\left(t_+\right)\right| < 1 - c_\varepsilon$. A similar result holds for E_-. Then by these results, $\forall\varepsilon > 0$, $\exists c_\varepsilon > 0$ such that $\sup_{\|t\|>\varepsilon}\left|\hat{\Psi}_{\mathcal{F}}(t)\right|<1-c_{\varepsilon}h$, when n is sufficiently large. It follows from this result and arguments in the proof of Calonico et al. (2018, Theorem S.1) that $\forall\delta>0$, $\exists c_{\delta}>0$ such that $\sup_{\|t\|>\delta\sqrt{\bar{n}}}\left|\hat{\Psi}_{S_n}(t)\right|\leq(1-c_{\delta}h)^n$ when n is sufficiently large. It is also easy to see that $\forall\delta>0$, $(1-c_{\delta}h)^n\leq\epsilon_n^{d_f/2+2}$, when n is sufficiently large. Therefore, Condition III''_{α} of Skovgaard (1986, Theorem 3.4 and Remark 3.5) is satisfied with $\alpha=1$. Verification of Condition IV of Skovgaard (1986, Theorem 3.4) follows from essentially the same calculations and arguments in the proof of Calonico et al. (2018, Theorem S.1). Now all conditions for Skovgaard (1986, Theorem 3.4) are verified. It shows that S_n admits a valid Edgeworth expansion, i.e., conditions (3.1), (3.2) and (3.3) of Skovgaard (1981) are satisfied with $U_n=S_n$, $s=4$, $\beta_{s,n}=\bar{n}^{-1}$ and the Edgeworth expansion holds uniformly over the class of all convex sets in $\mathbb{R}^{d_f}$. Note that we can write $\sqrt{\bar{n}}R=h_n\left(\mathrm{B}^{1/2}S_n\right)$. Then we apply Skovgaard (1981) to show that the Edgeworth expansion is preserved by smooth transformations. Condition (3.4) of Skovgaard (1981) is satisfied with g_n taken to be $x\mapsto h_n\left(\mathrm{B}^{1/2}x\right)$ whose the gradient at zero $\nabla g_n(0)$ is given by $\nabla g_n(0)=\mathrm{B}^{1/2}\left(\tilde{\varpi}^{\top}\Delta_{\mathcal{U}\mathcal{U}^{\top}}^{-1/2},0_{d_f-2d_u}^{\top}\right)^{\top}+O\left(\|\Delta_{\mathcal{U}}\|\right)$ by the chain rule. Then we apply Skovgaard (1981, Theorem 3.2) to $f_n(S_n)\coloneqq B_n^{-1}g_n(S_n)$, where $B_n^2\coloneqq\nabla g_n(0)^{\top}\nabla g_n(0)$. Then, $B_n^2=\tilde{\varpi}^{\top}\Delta_{\mathcal{U}\mathcal{U}^{\top}}^{-1/2}\left(\mathrm{Var}\left[\mathcal{U}\right]/h\right)\Delta_{\mathcal{U}\mathcal{U}^{\top}}^{-1/2}\tilde{\varpi}+O\left(\|\Delta_{\mathcal{U}}\|\right)=1+O\left(\|\Delta_{\mathcal{U}}\|\right)$. Condition I of Skovgaard (1981, Assumption 3.1) is satisfied with $p=4$. Condition II of Skovgaard (1981, Assumption 3.1) is satisfied with $\lambda_n=O\left(\bar{n}^{-1/2}\right)$ so that $\lambda_n^{p-1}=o\left(\bar{n}^{-1}\right)$. Now all conditions for Skovgaard (1981, Theorem 3.2) are verified. It is left to compute the approximate cumulants.

Then we calculate the formal cumulants of $f_n(S_n)=B_n^{-1}\sqrt{\bar{n}}R$. In the calculations, we repeatedly use formulae for moments of products of sample averages (e.g., DiCiccio et al., 1988, Page 12) and Lemma 1. By definition, $\mathrm{E}\left[R_1\right]=0$. We calculate $\mathrm{E}\left[R_2\right]$, let the remainder term absorb the terms that involve $\alpha^1,\ldots,\alpha^{2d_u}$ and get $\mathrm{E}\left[R_2\right]=\bar{n}^{-1}\bar{\kappa}_1+O\left(\|\Delta_{\mathcal{U}}\|/n\right)$ where $\bar{\kappa}_1\coloneqq\alpha^{mnk}\mathrm{M}^{(mk)}\varpi^{(n)}/6-\Omega^{(ko)}\mathrm{P}^{(om)}\varpi^{(l)}\gamma^{m;l,k}$. By formulae for third moments and Lemma 1, $\mathrm{E}\left[R_3\right]=O\left(\bar{n}^{-2}\right)$. Therefore, $\kappa_1\left(\sqrt{\bar{n}}R\right)=\tilde{\kappa}_{1,n}+O\left(\bar{n}^{-1/2}\|\Delta_{\mathcal{U}}\|h+\bar{n}^{-3/2}\right)$ with $\tilde{\kappa}_{1,n}\coloneqq\bar{n}^{-1/2}\bar{\kappa}_1$. For the second cumulant, by definition, $\kappa_2(R)=\mathrm{E}\left[R^2\right]-\left(\mathrm{E}\left[R\right]\right)^2$ and by formulae for fifth and sixth moments and Lemma 1, $\mathrm{E}\left[R^2\right]=\mathrm{E}\left[R_1^2\right]+2\cdot\mathrm{E}\left[R_1R_2\right]+2\cdot\mathrm{E}\left[R_1R_3\right]+\mathrm{E}\left[R_2^2\right]+O\left(\bar{n}^{-3}\right)$. By $R_1=\tilde{R}_{11}+\tilde{R}_{21}$ and calculation, $\mathrm{E}\left[R_1^2\right]=\mathrm{E}\left[\tilde{R}_{11}^2\right]+2\cdot\mathrm{E}\left[\tilde{R}_{21}\tilde{R}_{11}\right]+O\left(\bar{n}^{-1}\|\Delta_{\mathcal{U}}\|^2\right)$, $\mathrm{E}\left[R_1R_2\right]+\mathrm{E}\left[R_1R_3\right]=\mathrm{E}\left[\tilde{R}_{11}R_2\right]+\mathrm{E}\left[\tilde{R}_{11}R_3\right]+O\left(\bar{n}^{-2}\|\Delta_{\mathcal{U}}\|\right)$. Then by calculation, $\mathrm{E}\left[\tilde{R}_{11}^2\right]=\bar{n}^{-1}+O\left(\|\Delta_{\mathcal{U}}\|^2/n\right)$ and $2\cdot\mathrm{E}\left[\tilde{R}_{21}\tilde{R}_{11}\right]=\bar{n}^{-1}\tilde{\kappa}_{21,n}+O\left(\|\Delta_{\mathcal{U}}\|^2/n\right)$, where $\tilde{\kappa}_{21,n}\coloneqq\alpha^{mno}\mathrm{M}^{(no)}\mathrm{M}^{(mk)}\alpha^k/3-2\gamma^{l;d_{\vartheta}+a,k}\Omega^{(km)}\mathrm{M}^{(ml)}\alpha^{d_{\vartheta}+a}$. Then, $\mathrm{E}\left[R_1^2\right]=\bar{n}^{-1}\left(1+\tilde{\kappa}_{21,n}\right)+O\left(\bar{n}^{-1}\|\Delta_{\mathcal{U}}\|^2\right)$. Calculation of $2\cdot\mathrm{E}\left[\tilde{R}_{11}R_2\right]+2\cdot\mathrm{E}\left[\tilde{R}_{11}R_3\right]+\mathrm{E}\left[R_2^2\right]$ follows from replication of calculations in Ma (2017) and we can directly use the results therein. By calculations in Ma (2017), we have

$$2\cdot\mathrm{E}\left[\tilde{R}_{11}R_2\right]+2\cdot\mathrm{E}\left[\tilde{R}_{11}R_3\right]+\mathrm{E}\left[R_2^2\right]=\bar{n}^{-2}\sum_{j=1}^{8}\bar{\kappa}_{2j}+O\left(\bar{n}^{-2}\|\Delta_{\mathcal{U}}\|h+\bar{n}^{-3}\right),$$

for some bounded constants $\bar{\kappa}_{21}, ..., \bar{\kappa}_{28}$, e.g., $\bar{\kappa}_{21} \coloneqq \alpha^{vmn}\mathrm{M}^{(vo)}\mathrm{M}^{(ml)}\mathrm{M}^{(nk)}\alpha^{klo}/3 - \alpha^{vm\,d_\vartheta+a}\mathrm{M}^{(vo)}\mathrm{M}^{(mn)}\alpha^{on\,d_\vartheta+a} + \alpha^{n\,d_\vartheta+a\,d_\vartheta+b}\mathrm{M}^{(nm)}\alpha^{m\,d_\vartheta+a\,d_\vartheta+b}$ and the $O\left(\bar{n}^{-2}\|\Delta_{\mathcal{U}}\| h + \bar{n}^{-3}\right)$ remainder collects terms that depend on $\alpha^1, ..., \alpha^{2d_u}$ and higher-order terms from the fourth moment calculation. The expressions of $\bar{\kappa}_{22}, ..., \bar{\kappa}_{28}$ are also easily obtained from Ma (2017) and hence omitted. Therefore, $\kappa_2\left(\sqrt{\bar{n}}R\right) = \tilde{\kappa}_{2,n} + O\left(\|\Delta_{\mathcal{U}}\|^2 + \bar{n}^{-1}\|\Delta_{\mathcal{U}}\| + \bar{n}^{-2}\right)$, where $\tilde{\kappa}_{2,n} \coloneqq 1 + \tilde{\kappa}_{21,n} + \tilde{\kappa}_{22,n}$ and $\tilde{\kappa}_{22,n} \coloneqq \bar{n}^{-1}\left(\sum_{j=1}^{8}\bar{\kappa}_{2j} - \bar{\kappa}_1^2\right)$. By definition, $\kappa_3(R) = \mathrm{E}\left[R^3\right] - 3\cdot\mathrm{E}[R]\,\mathrm{E}\left[R^2\right] + 2\left(\mathrm{E}[R]\right)^3$ and by $\mathrm{E}[R] = \mathrm{E}[R_2] + O\left(\bar{n}^{-2}\right)$, $\mathrm{E}[R_2] = O\left(\bar{n}^{-1}\right)$, $\mathrm{E}\left[R^2\right] = \mathrm{E}\left[R_1^2\right] + O\left(\bar{n}^{-2}\right)$ and $\mathrm{E}\left[R^3\right] = \mathrm{E}\left[R_1^3\right] + 3\cdot\mathrm{E}\left[R_2R_1^2\right] + O\left(\bar{n}^{-3}\right)$, which follows from formulae for higher moments, we have $\kappa_3(R) = \mathrm{E}\left[R_1^3\right] - 3\left(\mathrm{E}\left[R_2R_1^2\right] - \mathrm{E}[R_2]\,\mathrm{E}\left[R_1^2\right]\right) + O\left(\bar{n}^{-3}\right)$. It is easy to check that $\mathrm{E}\left[R_1^3\right] = \mathrm{E}\left[\tilde{R}_{11}^3\right] + O\left(\bar{n}^{-2}\|\Delta_{\mathcal{U}}\|\right)$, $\mathrm{E}\left[R_2R_1^2\right] = \mathrm{E}\left[R_2\tilde{R}_{11}^2\right] + O\left(\bar{n}^{-2}\|\Delta_{\mathcal{U}}\|\right)$. By these results and $\mathrm{E}\left[R_1^2\right] = \mathrm{E}\left[\tilde{R}_{11}^2\right] + O\left(\bar{n}^{-1}\|\Delta_{\mathcal{U}}\|\right)$, $\kappa_3(R) = \mathrm{E}\left[\tilde{R}_{11}^3\right] - 3\left(\mathrm{E}\left[R_2\tilde{R}_{11}^2\right] - \mathrm{E}[R_2]\,\mathrm{E}\left[\tilde{R}_{11}^2\right]\right) + O\left(\bar{n}^{-3} + \bar{n}^{-2}\|\Delta_{\mathcal{U}}\|\right)$. Calculation and expansion of $\mathrm{E}\left[\tilde{R}_{11}^3\right] - 3\left(\mathrm{E}\left[R_2\tilde{R}_{11}^2\right] - \mathrm{E}[R_2]\,\mathrm{E}\left[\tilde{R}_{11}^2\right]\right)$ follows from replication of calculations in Ma (2017). For example, by calculation using formulae for moments (DiCiccio et al., 1988),

$$\mathrm{E}\left[\tilde{R}_{11}^3\right] = n^{-2}\left(\mathrm{E}\left[\left(h^{-1}\varpi^{(k)}\mathcal{V}^{(k)} - \varpi^{(k)}\alpha^k\right)^3\right]\right) = n^{-2}\mathrm{E}\left[\left(h^{-1}\varpi^{(k)}\mathcal{V}^{(k)}\right)^3\right] + O\left(\bar{n}^{-2}\|\Delta_{\mathcal{U}}\| h\right),$$

and the $O\left(\bar{n}^{-2}\|\Delta_{\mathcal{U}}\| h\right)$ remainder collects all terms in the expansion of the third moment which depend on $\alpha^1, ..., \alpha^{2d_u}$. Note that we can write $\mathrm{E}\left[h^{-1}\left(\varpi^{(k)}\mathcal{V}^{(k)}\right)^3\right] = \varpi^{(k)}\varpi^{(l)}\varpi^{(m)}\alpha^{klm}$ in coordinate notations. Similarly, we calculate $\mathrm{E}\left[R_2\tilde{R}_{11}^2\right] - \mathrm{E}[R_2]\,\mathrm{E}\left[\tilde{R}_{11}^2\right]$. We note that coefficients of terms of order $\bar{n}^{-2}$ in $\mathrm{E}\left[\tilde{R}_{11}^3\right] - 3\left(\mathrm{E}\left[R_2\tilde{R}_{11}^2\right] - \mathrm{E}[R_2]\,\mathrm{E}\left[\tilde{R}_{11}^2\right]\right)$ are formally the same as those of the leading terms in the calculation of the formal third cumulant in Ma (2017). Calculations in Ma (2017) show that the sum of these coefficients are exactly zero and therefore, the leading term vanishes so that $\kappa_3\left(\sqrt{\bar{n}}R\right) = O\left(\|\Delta_{\mathcal{U}}\|/\sqrt{\bar{n}} + \bar{n}^{-3/2}\right)$. By this result, the fact that $\kappa_4(R) = \mathrm{E}\left[R^4\right] - 3\left(\mathrm{E}\left[R^2\right]\right)^2 - 4\cdot\mathrm{E}[R]\,\kappa_3(R) + 2\left(\mathrm{E}[R]\right)^4$, $\mathrm{E}[R] = O\left(\bar{n}^{-1}\right)$, $R = \tilde{R}_{11} + \tilde{R}_{21} + R_2 + R_3$ and standard calculations,

$$\begin{aligned}\kappa_4(R) = \mathrm{E}\left[R^4\right] - 3\left(\mathrm{E}\left[R^2\right]\right)^2 + O\left(\bar{n}^{-3}\|\Delta_{\mathcal{U}}\| + \bar{n}^{-4}\right) = \left\{\mathrm{E}\left[\tilde{R}_{11}^4\right] - 3\left(\mathrm{E}\left[\tilde{R}_{11}^2\right]\right)^2\right\} \\ + 4\left\{\mathrm{E}\left[R_2\tilde{R}_{11}^3\right] - 3\cdot\mathrm{E}\left[R_2\tilde{R}_{11}\right]\mathrm{E}\left[\tilde{R}_{11}^2\right]\right\} + 6\left\{\mathrm{E}\left[R_2^2\tilde{R}_{11}^2\right] - \mathrm{E}\left[R_2^2\right]\mathrm{E}\left[\tilde{R}_{11}^2\right]\right\} \\ + 4\left\{\mathrm{E}\left[R_2\tilde{R}_{11}^3\right] - 3\cdot\mathrm{E}\left[R_2\tilde{R}_{11}\right]\mathrm{E}\left[\tilde{R}_{11}^2\right]\right\} + O\left(\bar{n}^{-3}\|\Delta_{\mathcal{U}}\| + \bar{n}^{-4}\right). \quad \text{(S12)}\end{aligned}$$

And by standard calculations,

$$\begin{aligned}\mathrm{E}\left[\tilde{R}_{11}^4\right] - 3\left(\mathrm{E}\left[\tilde{R}_{11}^2\right]\right)^2 &= n^{-3}\left(\mathrm{E}\left[\left(h^{-1}\varpi^{(k)}\mathcal{V}^{(k)} - \varpi^{(k)}\alpha^k\right)^4\right] - 3\left(\mathrm{E}\left[\left(h^{-1}\varpi^{(k)}\mathcal{V}^{(k)} - \varpi^{(k)}\alpha^k\right)^2\right]\right)^2\right) \\ &= n^{-3}\left(\mathrm{E}\left[\left(h^{-1}\varpi^{(k)}\mathcal{V}^{(k)}\right)^4\right] - 3\left(\mathrm{E}\left[\left(h^{-1}\varpi^{(k)}\mathcal{V}^{(k)}\right)^2\right]\right)\right) + O\left(\bar{n}^{-3}\|\Delta_{\mathcal{U}}\| h\right),\end{aligned}$$

and the $O\left(\bar{n}^{-3}\left\|\Delta_{\mathcal{U}}\right\|h\right)$ remainder collects all terms that depend on $\alpha^1,\ldots,\alpha^{2d_u}$. Similarly, we also calculate $\mathrm{E}\left[R_2\tilde{R}_{11}^3\right]-3\cdot\mathrm{E}\left[R_2\tilde{R}_{11}\right]\mathrm{E}\left[\tilde{R}_{11}^2\right]$, $\mathrm{E}\left[R_2^2\tilde{R}_{11}^2\right]-\mathrm{E}\left[R_2^2\right]\mathrm{E}\left[\tilde{R}_{11}^2\right]$ and $\mathrm{E}\left[R_2\tilde{R}_{11}^3\right]-3\cdot\mathrm{E}\left[R_2\tilde{R}_{11}\right]\mathrm{E}\left[\tilde{R}_{11}^2\right]$ on the right hand side of the second equality in (S12), ignore small-order terms that depend on $\alpha^1,\ldots,\alpha^{2d_u}$ and take the sum of the leading terms. We do not need to rework on the calculations since they are formally the same as those done in Ma (2017). Calculations in Ma (2017) show that the sum of the leading terms on the right hand side of (S12) is exactly zero so that it follows from this result and (S12) that $\kappa_4\left(\sqrt{\bar{n}}R\right)=O\left(\bar{n}^{-1}\left\|\Delta_{\mathcal{U}}\right\|+\bar{n}^{-2}\right)$. By previous calculations and $B_n=1+O\left(\left\|\Delta_{\mathcal{U}}\right\|\right)$, we get the approximate cumulants for $f_n\left(S_n\right)$: $\kappa_1\left(f_n\left(S_n\right)\right)=B_n^{-1}\tilde{\kappa}_{1,n}+O\left(\bar{n}^{-1/2}\left\|\Delta_{\mathcal{U}}\right\|h+\bar{n}^{-3/2}\right)$, $\kappa_2\left(f_n\left(S_n\right)\right)=B_n^{-2}\tilde{\kappa}_{2,n}+O\left(\left\|\Delta_{\mathcal{U}}\right\|^2+\bar{n}^{-1}\left\|\Delta_{\mathcal{U}}\right\|+\bar{n}^{-2}\right)$, $\kappa_3\left(f_n\left(S_n\right)\right)=O\left(\left\|\Delta_{\mathcal{U}}\right\|/\sqrt{\bar{n}}+\bar{n}^{-3/2}\right)$ and $\kappa_4\left(f_n\left(S_n\right)\right)=O\left(\bar{n}^{-1}\left\|\Delta_{\mathcal{U}}\right\|+\bar{n}^{-2}\right)$.

Let $\phi\left(\cdot\mid\mu,\sigma^2\right)$ denote the PDF of $\mathrm{N}\left(\mu,\sigma^2\right)$. By applying Skovgaard (1981, Theorem 3.2) to $f_n\left(S_n\right)=B_n^{-1}\sqrt{\bar{n}}R$,

$$\Pr\left[\bar{n}\left(R_0+R\right)^2\leq x\right]=\int_{\left|t+\left(\sqrt{\bar{n}}R_0\right)/B_n\right|\leq\sqrt{x}/B_n}\phi\left(t\mid B_n^{-1}\tilde{\kappa}_{1,n},B_n^{-2}\tilde{\kappa}_{2,n}\right)\mathrm{d}t+O\left(\left\|\Delta_{\mathcal{U}}\right\|/\sqrt{\bar{n}}+\bar{n}^{-3/2}\right),\quad\text{(S13)}$$

uniformly in $x>0$. By using the recurrence properties of non-central χ^2 (Cohen, 1988) and mean value expansion, we have $\partial F\left(x\mid\lambda\right)/\partial\lambda|_{\lambda=\overline{\lambda}}=-xf_{\chi_1^2}\left(x\right)+O\left(\overline{\lambda}\right)$. By this result, $B_n^2=1+O\left(\left\|\Delta_{\mathcal{U}}\right\|\right)$, change of variables and mean value expansion,

$$\begin{aligned}&\int_{\left|t+\left(\sqrt{\bar{n}}R_0\right)/B_n\right|\leq\sqrt{x}/B_n}\phi\left(t\mid B_n^{-1}\tilde{\kappa}_{1,n},B_n^{-2}\tilde{\kappa}_{2,n}\right)\mathrm{d}t=\int_{|t|\leq\sqrt{x/\tilde{\kappa}_{2,n}}}\phi\left(t\mid\left(\sqrt{\bar{n}}R_0+\tilde{\kappa}_{1,n}\right)/\sqrt{\tilde{\kappa}_{2,n}},1\right)\mathrm{d}t\\&=F\left(\frac{x}{\tilde{\kappa}_{2,n}}\mid\frac{\left(\sqrt{\bar{n}}R_0+\tilde{\kappa}_{1,n}\right)^2}{\tilde{\kappa}_{2,n}}\right)=F_{\chi_1^2}\left(x\right)-xf_{\chi_1^2}\left(x\right)\left(\left(\sqrt{\bar{n}}\tilde{R}_{10}+\tilde{\kappa}_{1,n}\right)^2+\tilde{\kappa}_{21,n}+\tilde{\kappa}_{22,n}\right)+O\left(\nu_n^{\sharp}\right).\quad\text{(S14)}\end{aligned}$$

By (S13) and (S14),

$$\Pr\left[\bar{n}\left(R_0+R\right)^2\leq x\right]=F_{\chi_1^2}\left(x\right)-\tilde{\mathscr{C}}_p^{\mathsf{pre}}\left(n,h\right)xf_{\chi_1^2}\left(x\right)+O\left(\nu_n^{\sharp}\right),\quad\text{(S15)}$$

where $\tilde{\mathscr{C}}_p^{\mathsf{pre}}\left(n,h\right):=\bar{n}\tilde{R}_{10}^2+2\sqrt{\bar{n}}\tilde{R}_{10}\tilde{\kappa}_{1,n}+\tilde{\kappa}_{21,n}+\bar{n}^{-1}\sum_{j=1}^{8}\bar{\kappa}_{2j}$. By tedious and lengthy algebra, we can directly show that $\tilde{R}_{10}^2=\mathscr{B}_p^{\dagger}-\mathscr{B}_p^{\ddagger}$ and $\sum_{j=1}^{8}\bar{\kappa}_{2j}=\sum_{j=1}^{4}\left(\mathscr{V}_{p,j}^{\dagger}-\mathscr{V}_{p,j}^{\ddagger}\right)+O\left(h\right)$ and $2\sqrt{\bar{n}}\tilde{R}_{10}\tilde{\kappa}_{1,n}+\tilde{\kappa}_{21,n}=O\left(h\left\|\Delta_{\mathcal{U}}\right\|\right)$. By calculating $\mathrm{E}\left[LR^{\star}\right]$ with arguments used repeatedly in previous proofs, we find that $\tilde{\mathscr{C}}_p^{\mathsf{pre}}\left(n,h\right)$ is just the leading term in the expansion $\mathrm{E}\left[LR^{\star}\right]-1=\tilde{\mathscr{C}}_p^{\mathsf{pre}}\left(n,h\right)+o\left(\upsilon_n^{\natural}\right)$, where $\upsilon_n^{\natural}:=\left\|\Delta_{\mathcal{U}}\right\|+\bar{n}\left\|\Delta_{\mathcal{U}}\right\|^2+\bar{n}^{-1}$. We use the fact that $\ell_p\left(\widehat{\vartheta}_p\mid h\right)=\inf_{\theta_2}\sup_{\lambda_2}2\sum_i\log\left(1+\lambda_2^{\top}\bar{\mathcal{U}}_i\left(\theta_2\right)\right)$ and an alternative expression for $LR^{\star}=\bar{n}\left(\widetilde{\ell}^{\star}-\widehat{\ell}^{\star}\right)$ to get a more lucid proof.

We consider the singular value decomposition of $\Delta_{\bar{\mathcal{U}}\bar{\mathcal{U}}^{\top}}^{-1/2}\left(-\Delta_{\bar{\mathcal{G}}}\right)$ such that $\bar{\mathrm{S}}^{\top}\Delta_{\bar{\mathcal{U}}\bar{\mathcal{U}}^{\top}}^{-1/2}\left(-\Delta_{\bar{\mathcal{G}}}\right)\bar{\mathrm{T}}=\left[\begin{array}{cc}\bar{\Lambda}&0_{d_z\times d_z}\end{array}\right]^{\top}$

where $\bar{\mathrm{S}}^\top\bar{\mathrm{S}} = \mathrm{I}_{2d_z}$, $\bar{\mathrm{T}}^\top\bar{\mathrm{T}} = \mathrm{I}_{d_z}$ and $\bar{\Lambda}$ is a d_z-dimensional diagonal matrix. We apply the rotation by $\bar{\mathcal{V}}_i(\theta_2) := \bar{\Gamma}\bar{\mathcal{U}}_i(\theta_2)$ where $\bar{\Gamma} := \bar{\mathrm{S}}^\top\Delta_{\bar{\mathcal{U}}\bar{\mathcal{U}}^\top}^{-1/2}$ so that $\ell_p\left(\widehat{\vartheta}_p \mid h\right) = \inf_{\theta_2}\sup_{\lambda_2} 2\sum_i \log\left(1+\lambda_2^\top\bar{\mathcal{V}}_i(\theta_2)\right)$ and calculations from Matsushita and Otsu (2013) can be applied. Also denote $\bar{\mathcal{V}}_i := \bar{\Gamma}\bar{\mathcal{U}}_i$, $\bar{\mathcal{H}}_i := \bar{\Gamma}\left(-\bar{\mathcal{G}}_i\right)$ ($\bar{\mathcal{V}}$ and $\bar{\mathcal{H}}$ defined similarly) and $\bar{\Omega} := \left(\bar{\Lambda}\bar{\mathrm{T}}^\top\right)^{-1}$. Then it follows that $\Delta_{\bar{\mathcal{V}}\bar{\mathcal{V}}^\top} = \mathrm{I}_{2d_z\times 2d_z}$ and

$$\begin{bmatrix} -\Delta_{\bar{\mathcal{V}}\bar{\mathcal{V}}^\top} & \Delta_{\bar{\mathcal{H}}} \\ \Delta_{\bar{\mathcal{H}}}^\top & 0_{d_z\times d_z}\end{bmatrix}^{-1} = \begin{bmatrix} 0_{d_z\times d_z} & 0_{d_z\times d_z} & \bar{\Omega}^\top \\ 0_{d_z\times d_z} & -\mathrm{I}_{d_z} & 0_{d_z\times d_z} \\ \bar{\Omega} & 0_{d_z\times d_z} & \bar{\Omega}\bar{\Omega}^\top \end{bmatrix} = \begin{bmatrix} -\left(\bar{\Gamma}^\top\right)^{-1}\bar{\mathrm{Q}}\bar{\Gamma}^{-1} & -\left(\bar{\Gamma}^\top\right)^{-1}\bar{\mathrm{N}} \\ -\bar{\mathrm{N}}^\top\bar{\Gamma}^{-1} & \bar{\mathrm{O}} \end{bmatrix}. \tag{S16}$$

Let $\left(A_\ddagger^{\mathsf{a}}, A_\ddagger^{\mathsf{ab}}, A_\ddagger^{\mathsf{abc}}, C_\ddagger^{\mathsf{a},s}, C_\ddagger^{\mathsf{a};\mathsf{b},s}\right)$, $\left(\alpha_\ddagger^{\mathsf{a}}, \alpha_\ddagger^{\mathsf{ab}}, \alpha_\ddagger^{\mathsf{abc}}, \alpha_\ddagger^{\mathsf{abcd}}\right)$ and $\left(\gamma_\ddagger^{\mathsf{a},s}, \gamma_\ddagger^{\mathsf{a};\mathsf{b},s}, \gamma_\ddagger^{\mathsf{a},s;\mathsf{b},t}, \gamma_\ddagger^{\mathsf{a};\mathsf{b};\mathsf{c},s}\right)$ be defined by the same formulae as those of $\left(A^{\mathsf{k}}, A^{\mathsf{kl}}, A^{\mathsf{klm}}, C^{\mathsf{k},n}, C^{\mathsf{k};\mathsf{l},n}\right)$, $\left(\alpha^{\mathsf{k}}, \alpha^{\mathsf{kl}}, \alpha^{\mathsf{klm}}, \alpha^{\mathsf{klmn}}\right)$ and $\left(\gamma^{\mathsf{k},n}, \gamma^{\mathsf{k};\mathsf{l},n}, \gamma^{\mathsf{k},n;\mathsf{l},o}, \gamma^{\mathsf{k};\mathsf{l};\mathsf{m},n}\right)$, with $(\mathcal{V}, \mathcal{H}, \mathcal{V}_i, \mathcal{H}_i)$ replaced by $\left(\bar{\mathcal{V}}, \bar{\mathcal{H}}, \bar{\mathcal{V}}_i, \bar{\mathcal{H}}_i\right)$. The leading terms in the stochastic expansion of $\bar{n}^{-1}\ell_p\left(\widehat{\vartheta}_p \mid h\right)$ is given by $\bar{n}^{-1}\widehat{\ell}^\star = \tilde{R}_{\ddagger 1}^{d_z+a}\tilde{R}_{\ddagger 1}^{d_z+a} + 2\tilde{R}_{\ddagger 1}^{d_z+a}\tilde{R}_{\ddagger 2}^{d_z+a} + 2\tilde{R}_{\ddagger 1}^{d_z+a}\tilde{R}_{\ddagger 3}^{d_z+a} + \tilde{R}_{\ddagger 2}^{d_z+a}\tilde{R}_{\ddagger 2}^{d_z+a}$, where the expressions of $\left(\tilde{R}_{\ddagger 1}^{d_z+a}, \tilde{R}_{\ddagger 2}^{d_z+a}, \tilde{R}_{\ddagger 3}^{d_z+a}\right)$ are readily obtained in a special case of Matsushita and Otsu (2013) when the moment conditions are linear in parameters. E.g., $\tilde{R}_{\ddagger 1}^{d_z+a} := A_\ddagger^{d_z+a}$,

$$\tilde{R}_{\ddagger 2}^{d_z+a} := -\frac{1}{2}A_\ddagger^{d_z+b}A_\ddagger^{d_z+a\,d_z+b} + \frac{1}{3}\alpha_\ddagger^{d_z+a\,d_z+b\,d_z+c}A_\ddagger^{d_z+b}A_\ddagger^{d_z+c} - \bar{\Omega}^{(st)}C_\ddagger^{d_z+a,s}A_\ddagger^t + \bar{\Omega}^{(st)}\gamma_\ddagger^{d_z+a;d_z+b,s}A_\ddagger^{d_z+b}A_\ddagger^t$$

and the expression of $\tilde{R}_{\ddagger 3}^{d_z+a}$ is omitted for brevity (see Matsushita and Otsu, 2013, A.1). Let $\mathring{A}_\ddagger^{\mathsf{a}} := A_\ddagger^{\mathsf{a}} - \alpha_\ddagger^{\mathsf{a}}$. We again replace $A_\ddagger^{\mathsf{a}}$ by $\mathring{A}_\ddagger^{\mathsf{a}} + \alpha_\ddagger^{\mathsf{a}}$ to obtain $\tilde{R}_{\ddagger 1}^{d_z+a} = \tilde{R}_{\ddagger 11}^{d_z+a} + \tilde{R}_{\ddagger 10}^{d_z+a}$, $\tilde{R}_{\ddagger 2}^{d_z+a} = \tilde{R}_{\ddagger 22}^{d_z+a} + \tilde{R}_{\ddagger 21}^{d_z+a} + \tilde{R}_{\ddagger 20}^{d_z+a}$ and $\tilde{R}_{\ddagger 3}^{d_z+a} = \tilde{R}_{\ddagger 33}^{d_z+a} + \tilde{R}_{\ddagger 32}^{d_z+a} + \tilde{R}_{\ddagger 31}^{d_z+a} + \tilde{R}_{\ddagger 30}^{d_z+a}$. Then by standard calculations, $\mathrm{E}\left[\bar{n}^{-1}\widehat{\ell}^\star\right]$ is equal to the sum of $\tilde{R}_{\ddagger 10}^{d_z+a}\tilde{R}_{\ddagger 10}^{d_z+a}$, $\tilde{R}_{\ddagger 10}^{d_z+a}\mathrm{E}\left[\tilde{R}_{\ddagger 22}^{d_z+a}\right]$, $\mathrm{E}\left[\tilde{R}_{\ddagger 11}^{d_z+a}\tilde{R}_{\ddagger 21}^{d_z+a}\right]$, $\mathrm{E}\left[\tilde{R}_{\ddagger 11}^{d_z+a}\tilde{R}_{\ddagger 11}^{d_z+a}\right]$ and $2\cdot\mathrm{E}\left[\tilde{R}_{\ddagger 11}^{d_z+a}\tilde{R}_{\ddagger 22}^{d_z+a}\right] + 2\cdot\mathrm{E}\left[\tilde{R}_{\ddagger 11}^{d_z+a}\tilde{R}_{\ddagger 33}^{d_z+a}\right] + \mathrm{E}\left[\tilde{R}_{\ddagger 22}^{d_z+a}\tilde{R}_{\ddagger 22}^{d_z+a}\right]$ with an $o\left(\upsilon_n^\natural\right)$ remainder term. By inverting using the second equality of (S16), $\tilde{R}_{\ddagger 10}^{d_z+a}\tilde{R}_{\ddagger 10}^{d_z+a} = \alpha_\ddagger^{d_z+a}\alpha_\ddagger^{d_z+a} = \bar{\mathrm{Q}}^{(\mathsf{ab})}\bar{\Upsilon}^{\mathsf{a}}\bar{\Upsilon}^{\mathsf{b}}$. By calculation and $\Delta_{\bar{\mathcal{V}}\bar{\mathcal{V}}^\top} = \mathrm{I}_{2d_z\times 2d_z}$, $\mathrm{E}\left[\tilde{R}_{\ddagger 11}^{d_z+a}\tilde{R}_{\ddagger 11}^{d_z+a}\right] = \bar{n}^{-1}d_z + O\left(\|\Delta_{\mathcal{U}}\|^2/n\right)$. It is easy to calculate that $\mathrm{E}\left[\tilde{R}_{\ddagger 22}^{d_z+a}\right] = -\bar{n}^{-1}\alpha_\ddagger^{d_z+a\,d_z+b\,d_z+b}/6 - \bar{\Omega}^{(st)}\gamma_\ddagger^{t;d_z+a,s} + O\left(\|\Delta_{\mathcal{U}}\|/n\right)$. Then by (S16),

$$\tilde{R}_{\ddagger 10}^{d_z+a}\mathrm{E}\left[\tilde{R}_{\ddagger 22}^{d_z+a}\right] = -\bar{n}^{-1}\left(\frac{1}{6}\bar{\Upsilon}^{\mathsf{abc}}\bar{\mathrm{Q}}^{(\mathsf{ab})}\bar{\mathrm{Q}}^{(\mathsf{cd})}\bar{\Upsilon}^{\mathsf{d}} + \bar{\Gamma}^{\mathsf{a};\mathsf{b},s}\bar{\mathrm{N}}^{(\mathsf{a}s)}\bar{\mathrm{Q}}^{(\mathsf{bc})}\bar{\Upsilon}^{\mathsf{c}}\right) + o\left(\upsilon_n^\natural/\bar{n}\right).$$

By calculation and using (S16), $\mathrm{E}\left[\tilde{R}_{\ddagger 11}^{d_z+a}\tilde{R}_{\ddagger 21}^{d_z+a}\right] = \bar{n}^{-1}\bar{\Upsilon}^{\mathsf{abc}}\bar{\mathrm{Q}}^{(\mathsf{ab})}\bar{\mathrm{Q}}^{(\mathsf{cd})}\bar{\Upsilon}^{\mathsf{d}}/6 + o\left(\upsilon_n^\natural/\bar{n}\right)$. By calculation in Matsushita and Otsu (2013, A.4),

$$2\cdot\mathrm{E}\left[\tilde{R}_{\ddagger 11}^{d_z+a}\tilde{R}_{\ddagger 22}^{d_z+a}\right] + 2\cdot\mathrm{E}\left[\tilde{R}_{\ddagger 11}^{d_z+a}\tilde{R}_{\ddagger 33}^{d_z+a}\right] + \mathrm{E}\left[\tilde{R}_{\ddagger 22}^{d_z+a}\tilde{R}_{\ddagger 22}^{d_z+a}\right] = \bar{n}^{-2}\sum_{j=1}^{8}\bar{\kappa}_{\ddagger 2j} + o\left(\upsilon_n^\natural/\bar{n}\right),$$

where the constants are defined by

$$\left(\bar{\kappa}_{\dagger 21},\bar{\kappa}_{\ddagger 22},\bar{\kappa}_{\ddagger 23},\bar{\kappa}_{\ddagger 24},\bar{\kappa}_{\ddagger 25},\bar{\kappa}_{\ddagger 26},\bar{\kappa}_{\ddagger 27},\bar{\kappa}_{\ddagger 28}\right) \coloneqq \Big(\frac{1}{2}\bar{\Upsilon}^{\mathsf{abcd}}\bar{\mathrm{Q}}^{(\mathsf{ab})}\bar{\mathrm{Q}}^{(\mathsf{cd})}, -\frac{1}{3}\bar{\Upsilon}^{\mathsf{abc}}\bar{\mathrm{Q}}^{(\mathsf{ad})}\bar{\mathrm{Q}}^{(\mathsf{be})}\bar{\mathrm{Q}}^{(\mathsf{cf})}\bar{\Upsilon}^{\mathsf{def}},$$

$$2\bar{\Gamma}^{\mathsf{a;b;c},s}\bar{\mathrm{N}}^{(\mathsf{a}s)}\bar{\mathrm{Q}}^{(\mathsf{bc})}, -\bar{\Gamma}^{\mathsf{a;b},s}\bar{\mathrm{Q}}^{(\mathsf{ac})}\bar{\mathrm{Q}}^{(\mathsf{bd})}\bar{\mathrm{N}}^{(\mathsf{e}s)}\bar{\Upsilon}^{\mathsf{cde}}, -\bar{\Gamma}^{\mathsf{a},s;\mathsf{b},t}\bar{\mathrm{Q}}^{(\mathsf{ab})}\bar{\mathrm{O}}^{(st)}$$

$$\bar{\Gamma}^{\mathsf{a;c},s}\bar{\mathrm{Q}}^{(\mathsf{ab})}\bar{\mathrm{Q}}^{(\mathsf{cd})}\bar{\mathrm{O}}^{(st)}\bar{\Gamma}^{\mathsf{b;d},t}, -\bar{\Gamma}^{\mathsf{a;c},s}\bar{\mathrm{N}}^{(\mathsf{a}t)}\bar{\mathrm{Q}}^{(\mathsf{cd})}\bar{\mathrm{N}}^{(\mathsf{b}s)}\bar{\Gamma}^{\mathsf{b;d},t}, \bar{\Gamma}^{\mathsf{a;c},s}\bar{\mathrm{N}}^{(\mathsf{a}s)}\bar{\mathrm{Q}}^{(\mathsf{cd})}\bar{\mathrm{N}}^{(\mathsf{b}t)}\bar{\Gamma}^{\mathsf{b;d},t}\Big).$$

Note that $(\bar{\kappa}_{\dagger 21},\bar{\kappa}_{\ddagger 22},\bar{\kappa}_{\ddagger 23},\bar{\kappa}_{\ddagger 25}) = \left(\mathscr{V}^{\ddagger}_{p,1},\mathscr{V}^{\ddagger}_{p,2},\mathscr{V}^{\ddagger}_{p,3},\mathscr{V}^{\ddagger}_{p,4}\right)$. Therefore,

$$\mathrm{E}\left[\bar{n}\widehat{\ell}^{\star}\right] = d_z + \bar{n}\mathscr{B}^{\ddagger}_p - \bar{\Gamma}^{\mathsf{a;b},s}\bar{\mathrm{N}}^{(\mathsf{a}s)}\bar{\mathrm{Q}}^{(\mathsf{bc})}\bar{\Upsilon}^{\mathsf{c}} + \bar{n}^{-1}\sum_{j=1}^{8}\bar{\kappa}_{\ddagger 2j} + o\left(\upsilon^{\natural}_n\right).$$

Let $\bar{\kappa}_{\dagger 2j}$ be defined by the formula of $\bar{\kappa}_{\ddagger 2j}$ with $\left(\bar{\Upsilon},\bar{\mathrm{Q}},\bar{\mathrm{N}},\bar{\mathrm{O}},\bar{\Gamma}\right)$ replaced by $\left(\Upsilon,\mathrm{Q}_{\dagger},\mathrm{N}_{\dagger},\mathrm{O}_{\dagger},\Gamma_{\dagger}\right)$ and also $(\bar{\kappa}_{\dagger 21},\bar{\kappa}_{\dagger 22},\bar{\kappa}_{\dagger 23},\bar{\kappa}_{\dagger 25}) = \left(\mathscr{V}^{\dagger}_{p,1},\mathscr{V}^{\dagger}_{p,2},\mathscr{V}^{\dagger}_{p,3},\mathscr{V}^{\dagger}_{p,4}\right)$. By following the same steps, we get a similar expansion for $\mathrm{E}\left[\bar{n}\widetilde{\ell}^{\star}\right]$. And, then we have $\mathrm{E}\left[\bar{n}\left(\widetilde{\ell}^{\star}-\widehat{\ell}^{\star}\right)\right]-1 = \tilde{\mathscr{C}}^{\mathsf{pre}}_p(n,h) + o\left(\upsilon^{\natural}_n\right)$

$$\tilde{\mathscr{C}}^{\mathsf{pre}}_p(n,h) = \bar{n}\left(\mathscr{B}^{\dagger}_p - \mathscr{B}^{\ddagger}_p\right) - \Gamma^{\mathsf{k;l},u}_{\dagger}\mathrm{N}^{(\mathsf{k}u)}_{\dagger}\mathrm{Q}^{(\mathsf{lm})}_{\dagger}\Upsilon^{\mathsf{m}} + \bar{\Gamma}^{\mathsf{a;b},s}\bar{\mathrm{N}}^{(\mathsf{a}s)}\bar{\mathrm{Q}}^{(\mathsf{bc})}\bar{\Upsilon}^{\mathsf{c}} + \bar{n}^{-1}\sum_{j=1}^{8}\left(\bar{\kappa}_{\dagger 2j}-\bar{\kappa}_{\ddagger 2j}\right).$$

It is easy to see that by Lemma 1, $\Gamma^{\mathsf{k;l},u}_{\dagger} \asymp \bar{\Gamma}^{\mathsf{a;b},s} = O(h)$. Therefore, $\Gamma^{\mathsf{k;l},u}_{\dagger}\mathrm{N}^{(\mathsf{k}u)}_{\dagger}\mathrm{Q}^{(\mathsf{lm})}_{\dagger}\Upsilon^{\mathsf{m}} \asymp \bar{\Gamma}^{\mathsf{a;b},s}\bar{\mathrm{N}}^{(\mathsf{a}s)}\bar{\mathrm{Q}}^{(\mathsf{bc})}\bar{\Upsilon}^{\mathsf{c}} = O\left(\|\Delta_{\mathcal{U}}\|h\right)$, $\bar{\kappa}_{\dagger 24} \asymp \bar{\kappa}_{\ddagger 24} = O(h)$ and $\bar{\kappa}_{\dagger 26} \asymp \bar{\kappa}_{\dagger 27} \asymp \bar{\kappa}_{\dagger 28} \asymp \bar{\kappa}_{\ddagger 26} \asymp \bar{\kappa}_{\ddagger 27} \asymp \bar{\kappa}_{\ddagger 28} = O\left(h^2\right)$. It follows from these results that $\tilde{\mathscr{C}}^{\mathsf{pre}}_p(n,h) = \mathscr{C}^{\mathsf{pre}}_p(n,h) + O\left(\|\Delta_{\mathcal{U}}\|h + n^{-1}\right)$.

It is easily seen that the result (S13) with the weak inequality replaced by a strict inequality still holds (see Skovgaard, 1981, Theorem 3.2). By $LR^{\star} = \bar{n}\left(R_0+R\right)^2 + O^{\star}_p\left(\upsilon^{\sharp}_n\right)$ and the fact (21),

$$\left|\Pr\left[LR^{\star}\le x\right] - \Pr\left[\bar{n}\left(R_0+R\right)^2 \le x\right]\right| \le \Pr\left[\left|\bar{n}\left(R_0+R\right)^2 - x\right| \le c_1\upsilon^{\sharp}_n\right] + c_2\left(\log(n)/\bar{n}^{3/2}\right) = O\left(\upsilon^{\sharp}_n\right), \tag{S17}$$

where the equality follows from (S13) and boundedness of $\phi\left(\cdot \mid \tilde{\kappa}_{1,n},\tilde{\kappa}_{2,n}\right)$. The conclusion follows from (S15), (S17) and $\tilde{\mathscr{C}}^{\mathsf{pre}}_p(n,h) = \mathscr{C}^{\mathsf{pre}}_p(n,h) + O\left(\|\Delta_{\mathcal{U}}\|h + n^{-1}\right)$. ■

References

Bickel, P. J. and K. A. Doksum (2015). *Mathematical statistics: basic ideas and selected topics*, Volume 2. CRC Press.

Calonico, S., M. D. Cattaneo, and M. H. Farrell (2018). Coverage error optimal confidence intervals for local polynomial regression. *arXiv preprint arXiv:1808.01398*.

Chen, X. and K. Kato (2020). Jackknife multiplier bootstrap: finite sample approximations to the u-process supremum with applications. *Probability Theory and Related Fields 176*(3-4), 1–67.

Chernozhukov, V., D. Chetverikov, and K. Kato (2014). Gaussian approximation of suprema of empirical processes. *Annals of Statistics 42*(4), 1564–1597.

Cohen, J. D. (1988). Noncentral chi-square: Some observations on recurrence. *The American Statistician 42*(2), 120–122.

DiCiccio, T., P. Hall, and J. Romano (1988). Bartlett adjustments for empirical likelihood. Technical report No. 298, Department of Statistics, Stanford University.

Giné, E. and R. Nickl (2015). *Mathematical foundations of infinite-dimensional statistical models*, Volume 40. Cambridge University Press.

Kosorok, M. R. (2007). *Introduction to empirical processes and semiparametric inference.* Springer Science & Business Media.

Ma, J. (2017). Second-order refinement of empirical likelihood ratio tests of nonlinear restrictions. *The Econometrics Journal 20*(1), 139–148.

Matsushita, Y. and T. Otsu (2013). Second-order refinement of empirical likelihood for testing overidentifying restrictions. *Econometric Theory 29*(02), 324–353.

Skovgaard, I. M. (1981). Transformation of an edgeworth expansion by a sequence of smooth functions. *Scandinavian Journal of Statistics*, 207–217.

Skovgaard, I. M. (1986). On multivariate edgeworth expansions. *International Statistical Review/Revue Internationale de Statistique*, 169–186.