Online Supplementary Appendix to "Testing a Class of Semi- or Nonparametric Conditional Moment Restriction Models using Series Methods"

Jesper Riis-Vestergaard Sørensen
Department of Economics, University of Copenhagen

October 11, 2022

Abstract

This online appendix contains supplementary discussion, proof details, and supporting lemmas. Specifically, Section S.1 contains some discussion of distance measures other than the Cramér-von Mises-type (CM-type) used in the paper. A proof of Lemma A.2, key to establishing the asymptotic equivalence in Lemma 1, is provided in Section S.2. Section S.3 contains the proof of the bootstrap equivalence claimed in Lemma 3. Supporting lemmas are gathered in Section S.4.

S.1 Other Distance Measures

The CM-type statistic (2.9) arises from an equi-weighted sum of squares of (empirical) L^2-type norms. Other measures of distance are certainly possible. For example, the maximum of (empirical) L^∞-type norms leads to a statistic $T_n^{\mathrm{KS}} \coloneqq \sqrt{n}\max_{1\leqslant\ell\leqslant L}\max_{1\leqslant i\leqslant n}|\widehat{M}_\ell(X_{\ell i})|$ akin to the classical Kolmogorov-Smirnov (KS) statistic. One could also, as in Bravo (2012), base a test on the profile empirical (log-)likelihood process $\widehat{D}_{\mathrm{EL}} : \times_{\ell=1}^{L}\mathcal{X}_\ell \to \mathbf{R}_+$ defined by $\widehat{D}_{\mathrm{EL}}(t) \coloneqq -2\sum_{i=1}^n \ln\widehat{\pi}_i(t)$, where $\{\widehat{\pi}_i(t)\}_{i=1}^n$ solves

$$\max_{0\leqslant\pi_i\leqslant 1}\sum_{i=1}^n \ln\pi_i \quad \text{s.t.} \quad \sum_{i=1}^n \pi_i = 1 \quad \text{and} \quad \sum_{i=1}^n \pi_i\widehat{m}_i(t) = 0,$$

with $\widehat{m}_i(t) := (\widehat{m}_{i1}(t_1), \ldots, \widehat{m}_{iL}(t_L))'$ and $\widehat{m}_{i\ell}(t_\ell) := \rho_\ell(Z_i, \widehat{\beta}, \widehat{h}_\ell(W_{\ell i}))\omega_\ell(t_\ell, X_i)$. Let $\widehat{F}_X$ denote the empirical distribution of the union of distinct elements of the X_ℓ's. Then, under regularity conditions, the asymptotic null distribution of the resulting CM-type statistic $D_{\mathrm{EL},n}^{\mathrm{CM}} := \int_{\times_{\ell=1}^{L}\mathcal{X}_\ell} \widehat{D}_{\mathrm{EL}}(t)\,\mathrm{d}\widehat{F}_X(t)$ and that of the test statistic T_n studied here only differ in terms of the (implicit) weighting employed. This observation follows from the internal studentization property of (generalized) empirical likelihood (Bravo, 2012, Footnote 5). See also Bravo (2012, Eq. (2.8)), which should be compared to T_n in (2.9). Moreover, the same conclusion holds for KS-type statistics, including T_n^{KS}, and for the test based on generalized empirical likelihood processes (Bravo, 2012, Remark 2.1). The primary reason for the CM-type of statistic pursued in this paper is its computational convenience. Moreover, preliminary simulation experiments (not reported) suggest that the CM-based test has somewhat better power properties than its KS equivalent. Similar experiences are noted in Rothe and Wied (2013, p. 316).

S.2 Proof of Lemma A.2

Proof of Lemma A.2. The proof proceeds in a number of steps. Since the lemma is stated for a given ℓ, for notational convenience we drop the ℓ subscripts throughout, refer to the (ℓth) index set ($\mathcal{X}_\ell$) as $\mathcal{T}$ itself, and use d_t for its dimension.

Step 0 (Main)

Let $t \in \mathcal{T}$ be arbitrary. Assumption 1 implies that $\|\widehat{\beta} - \beta_0\| \lesssim_{\mathrm{P}} n^{-1/2} \to 0$, so letting $\mathcal{N}$ be any open neighborhood of β_0 (again provided by Assumption 1), $\widehat{\beta} \in \mathcal{N}$ wp $\to 1$. To simplify notation and ensure that objects are globally well defined, in what follows we will—without loss of generality—assume that $\widehat{\beta} \in \mathcal{N}$ with *probability one for all* n. Then by Assumption 3, for any z, v, we may conduct a mean value expansion of $\beta \mapsto \rho(z, \beta, v)$ at $\widehat{\beta}$ around β_0 to get

$$\widehat{M}(t) = \sqrt{n}\mathbb{E}_n[\omega(t, X_i)\,\rho(Z_i, \beta_0, \widehat{h}(W_i))] + \mathrm{I}_n(t)'\sqrt{n}(\widehat{\beta} - \beta_0),$$
$$\mathrm{I}_n(t) \coloneqq \mathbb{E}_n\left[\omega(t, X_i)\,(\partial/\partial\beta)\,\rho(Z_i, \overline{\beta}, \widehat{h}(W_i))\right],$$

where $\overline{\beta}$ lies on the line segment connecting $\widehat{\beta}$ and β_0, thus satisfying $\|\overline{\beta} - \beta_0\| \leqslant \|\widehat{\beta} - \beta_0\| \to_{\mathrm{P}} 0$. Recall the definition of $b(t)$ in (3.7), which is well defined on $\mathcal{T}$ since

β_0 is interior to $\mathcal{B}$ (Assumption 1). Step 1 below shows that $\sup_{t\in\mathcal{T}}\|\mathrm{I}_n(t)-b(t)\| \to_{\mathrm{P}} 0$, and that b is bounded on $\mathcal{T}$, so Assumption 1 and the previous display combine to yield

$$\sqrt{n}\widehat{M}(t) = \sqrt{n}\mathbb{E}_n[\omega(t,X_i)\,\rho(Z_i,\beta_0,\widehat{h}(W_i))] + b(t)'\sqrt{n}\mathbb{E}_n[s(Z_i)] + o_{\mathrm{P}}(1), \tag{S.1}$$

uniformly on $\mathcal{T}$.

The remainder of the proof is about adjusting for estimation of h^*. Given that β_0 is held fixed throughout this argument, we will suppress the β argument and write $\rho(z,v) \coloneqq \rho(z,\beta_0,v)$. For the purpose of the adjustment, denote the first term on the right-hand side of (S.1)

$$\sqrt{n}\widehat{M}^*(t) \coloneqq \sqrt{n}\mathbb{E}_n[\omega(t,X_i)\,\rho(Z_i,\widehat{h}(W_i))], \tag{S.2}$$

and conduct a MVE of $v \mapsto \rho(Z_i,v)$ at $\widehat{h}(W_i)$ around $h^*(W_i)$ to arrive at

$$\sqrt{n}\widehat{M}^*(t) = \sqrt{n}\mathbb{E}_n\left[\omega(t,X_i)\left\{\rho(Z_i,h^*(W_i)) + \frac{\partial}{\partial h'}\rho(Z_i,\overline{h}(W_i))[\widehat{h}(W_i)-h^*(W_i)]\right\}\right],$$

where $\overline{h}(W_i)$ lies on the line segment connecting $\widehat{h}(W_i)$ and $h^*(W_i)$. Such an expansion is justified by Assumption 3. Further decomposition of the right-hand side yields

$$\begin{aligned}
&\sqrt{n}\widehat{M}^*(t) \\
&= \sqrt{n}\mathbb{E}_n\left[\omega(t,X_i)\,\rho(Z_i,h^*(W_i)) + \delta(t,W_i)'\{Y_i - h^*(W_i)\}\right] \\
&\quad + \sqrt{n}\mathbb{E}_n\left[\omega(t,X_i)\left\{\frac{\partial}{\partial h'}\rho(Z_i,\overline{h}(W_{\ell i})) - \frac{\partial}{\partial h'}\rho(Z_i,h^*(W_i))\right\}\{\widehat{h}(W_i)-h^*(W_i)\}\right] \\
&\quad + \mathbb{G}_n\left[\omega(t,X_i)\,\frac{\partial}{\partial h'}\rho(Z_i,h^*(W_i))\right][\widehat{h}(W_i)-h^*(W_i)] \\
&\quad + \sqrt{n}\bigg(\mathrm{E}_Z\left[\omega(t,X)\,\frac{\partial}{\partial h'}\rho(Z,h^*(W))\,[\widehat{h}(W)-h^*(W)]\right] \\
&\qquad\quad - \mathbb{E}_n[\delta(t,W_i)'\{Y_i-h^*(W_i)\}]\bigg) \\
&=: \sqrt{n}\mathbb{E}_n\left[\omega(t,X_i)\,\rho(Z_i,h^*(W_i)) + \delta(t,W_i)'\{Y_i-h^*(W_i)\}\right] \\
&\quad + \mathrm{II}_n(t) + \mathrm{III}_n(t) + \mathrm{IV}_n(t),
\end{aligned} \tag{S.3}$$

where $\mathrm{E}_Z[\cdot]$ denotes integration with respect to the distribution of Z, and $\delta(t,Z)$ is defined as in (3.8). The $k \times k$ matrix $Q_k = \mathrm{E}[p^k(W)p^k(W)']$ is invertible by Assumption 5. Let h_k and $\delta_k(t,\cdot)$ denote the mean-square projections of h^* and $\delta(t,\cdot)$, respectively, onto the span of p^k, i.e.,

$$h_{m,k}(\cdot) \coloneqq p^k(\cdot)' Q_k^{-1}\mathrm{E}[p^k(W)h_m^*(W)] = p^k(\cdot)'\pi_{h_m,k}, \tag{S.4}$$

$$\delta_{m,k}(t,\cdot) \coloneqq p^k(\cdot)' Q_k^{-1}\mathrm{E}[p^k(W)\delta_m(t,W)] = p^k(\cdot)'\pi_{\delta_m,k}(t), \tag{S.5}$$

where $\pi_{h_m,k}$ and $\pi_{\delta_m,k}$ are defined in (3.9) and (3.10), respectively. Consequently,

$$\begin{gathered}\mathrm{E}[\{h_{m,k}(W) - h_m^*(W)\}^2] = r_{h_m,k}^2,\\ \mathrm{E}[\{\delta_{m,k}(t,W) - \delta_m(t,W)\}^2] = r_{\delta_m,k}^2(t),\\ \mathrm{E}\{\|\delta_{m,k}(\cdot,W) - \delta_m(\cdot,W)\|_{\mathcal{T}}^2\} = R_{\delta_m,k}^2,\end{gathered}$$

for $r_{h_m,k}^2$, $r_{\delta_m,k}^2$ and $R_{\delta_m,k}^2$ defined in (3.11), (3.12) and (3.13), respectively. Steps 2–4 below show that the three remainder terms in the decomposition (S.3) satisfy:

$$\begin{aligned}\|\mathrm{II}_n\|_{\mathcal{T}} &\lesssim_{\mathrm{P}} \mathrm{E}[R(Z)]\sqrt{n}\max_{1\leqslant m\leqslant d}\|\widehat{h}_m - h_m^*\|_{\mathcal{W}}^{1+\gamma},\\ \|\mathrm{III}_n\|_{\mathcal{T}} &\lesssim_{\mathrm{P}} \max_{1\leqslant m\leqslant d}\Big(\sum_{j=1}^{k_{m,n}}\|p_j\|_{\mathcal{W}}^2\Big)^{1/2}\Big(\sqrt{k_{m,n}/n} + k_{m,n}^{-\alpha_m}\Big), \quad \text{and}\\ \|\mathrm{IV}_n\|_{\mathcal{T}} &\lesssim_{\mathrm{P}} \max_{1\leqslant m\leqslant d}\Big\{\sqrt{n}r_{h_m,k_{m,n}}\sup_{t\in\mathcal{T}} r_{\delta_m,k_{m,n}}(t) + \sqrt{\zeta_{k_{m,n}}^2 k_{m,n}\ln(k_{m,n})/n}\\ &\qquad + R_{\delta_m,k_{m,n}}\sqrt{\ln\Big(k_{m,n}/R_{\delta_m,k_{m,n}}\Big)} + \zeta_{k_{m,n}} r_{h_m,k_{m,n}}\Big\}.\end{aligned}$$

Plug (S.3) into (S.1), apply T and use the definition of $\widehat{M}^*$ in (S.2) to get the claimed in-probability bound.

Step 1: I_n **and** b

In this step we show that I_n defined in (S.1) and b defined (3.7) satisfy

$$\text{(a)}\ \sup_{t\in\mathcal{T}}\|\mathrm{I}_n(t) - b(t)\| \xrightarrow{\mathrm{P}} 0 \quad \text{and} \quad \text{(b)}\ \sup_{t\in\mathcal{T}}\|b(t)\| < \infty.$$

Decompose I_n as

$$\begin{aligned}\mathrm{I}_n(t) &= \mathbb{E}_n\left[\omega(t, X_i)(\partial/\partial\beta)\rho(Z_i, \overline{\beta}, h^*(W))\right] \\ &\quad + \mathbb{E}_n\left[\omega(t, X_i)\left\{(\partial/\partial\beta)\rho(Z_i, \overline{\beta}, \widehat{h}_n(W_i)) - (\partial/\partial\beta)\rho(Z_i, \overline{\beta}_n, h^*(W_i))\right\}\right] \\ &=: \mathrm{I}_{a,n}(t) + \mathrm{I}_{b,n}(t).\end{aligned}$$

Since $\|\overline{\beta} - \beta_0\| \leqslant \|\widehat{\beta} - \beta_0\|$ and $\widehat{\beta} \in \mathcal{N}$, we must have $\overline{\beta} \in \mathcal{N}$ wp $\to 1$, so using T, Assumptions 2, 3 and 7 and Lemma S.7.4, we get

$$\begin{aligned}\sup_{t\in\mathcal{T}} \|\mathrm{I}_{b,n}(t)\| &\leqslant \mathbb{E}_n\left[a(Z_i)\|\widehat{h}(W_i) - h^*(W_i)\|^c\right] \\ &\leqslant \sqrt{d}\mathbb{E}_n[a(Z_i)] \max_{1\leqslant m\leqslant d} \|\widehat{h}_m - h^*_m\|^c_{\mathcal{W}} \\ &\lesssim_{\mathrm{P}} \max_{1\leqslant m\leqslant d} \|\widehat{h}_m - h^*_m\|^c_{\mathcal{W}} \xrightarrow{\mathrm{P}} 0,\end{aligned}$$

where we have used M to deduce $\mathbb{E}_n[a(Z_i)] \lesssim_{\mathrm{P}} 1$.

Given that $\beta_0 \in \mathcal{N}$ open, there is an $r > 0$ such that the open ball $B_r(\beta_0)$ in $\mathbf{R}^{d_\beta}$ centered at β_0 with radius r is contained in $\mathcal{N}$. Let $\overline{B} \coloneqq \overline{B}_{r/2}(\beta_0)$ denote the closed ball in $\mathbf{R}^{d_\beta}$ with the same center but half the radius. Given that $\overline{B}$ is a closed and bounded subset of a finite-dimensional Euclidean space, by the Heine–Borel theorem it is compact. Assumptions 2 and 3 imply that $(t, \beta) \mapsto \omega(t, x)(\partial/\partial\beta)\rho(z, \beta, h^*(w))$ is continuous on $\mathcal{T} \times \mathcal{N}$ for each $z \in \mathcal{Z}$, hence on the subset $\mathcal{T} \times \overline{B}$, and this function is dominated by an integrable function depending on z only. Moreover, via Tychonoff's theorem, $\mathcal{T}$ and $\overline{B}$ compact imply that is $\mathcal{T} \times \overline{B}$ compact. Combining these observations with the fact that the data are i.i.d., Newey and McFadden (1994, Lemma 2.4) tells us that

(i) $(t, \beta) \mapsto \mathrm{E}[\omega(t, X)(\partial/\partial\beta)\rho(Z, \beta, h^*(W))]$ is continuous on $\mathcal{T} \times \overline{B}$,

(ii) $\sup_{(t,\beta)\in\mathcal{T}\times\overline{B}} \|(\mathbb{E}_n - \mathrm{E})[\omega(t, X_i)(\partial/\partial\beta)\rho(Z_i, \beta, h^*(W_i))]\| \xrightarrow{\mathrm{P}} 0.$

Given (i) and $\mathcal{T} \times \overline{B}$ compact, we must have (cf. Rudin, 1976, Theorem 4.19) that

(iii) $(t, \beta) \mapsto \mathrm{E}[\omega(t, X)(\partial/\partial\beta)\rho(Z, \beta, h^*(W))]$ is uniformly continuous on $\mathcal{T} \times \overline{B}$.

Let $\widetilde{\beta}$ be an arbitrary consistent estimator of β_0. Then $\widetilde{\beta} \in \overline{B}$ wp $\to 1$, and, on this event,

$$
\begin{aligned}
&\sup_{t\in\mathcal{T}}\left\|\mathbb{E}_n\left[\omega\left(t,X_i\right)\left(\partial/\partial\beta\right)\rho(Z_i,\widetilde{\beta},h^*\left(W_i\right))\right]-b\left(t\right)\right\| \\
&\leqslant \sup_{t\in\mathcal{T}}\left\|\left(\mathbb{E}_n-\mathrm{E}_Z\right)\left[\omega\left(t,X_i\right)\left(\partial/\partial\beta\right)\rho(Z_i,\widetilde{\beta},h^*\left(W_i\right))\right]\right\| \\
&\quad+\sup_{t\in\mathcal{T}}\left\|\mathrm{E}_Z\left[\omega\left(t,X\right)\left(\partial/\partial\beta\right)\rho(Z,\widetilde{\beta},h^*\left(W\right))\right]-b\left(t\right)\right\| \\
&\leqslant \sup_{(t,\beta)\in\mathcal{T}\times\overline{B}}\left\|\left(\mathbb{E}_n-\mathrm{E}\right)\left[\omega\left(t,X_i\right)\left(\partial/\partial\beta\right)\rho(Z_i,\beta,h^*\left(W_i\right))\right]\right\| \\
&\quad+\sup_{t\in\mathcal{T}}\left\|\mathrm{E}_Z\left[\omega\left(t,X\right)\left(\partial/\partial\beta\right)\rho(Z,\widetilde{\beta},h^*\left(W\right))\right]-b\left(t\right)\right\|\overset{\mathrm{P}}{\to}0,
\end{aligned}
$$

where the first inequality is due to T, the second uses $\{\widetilde{\beta}\in\overline{B}\}$, and we have used (ii) uniform convergence and (iii) uniform continuity. Invoking the conclusion of the previous display for the mean value $\widetilde{\beta}=\overline{\beta}$ we see that $\sup_{t\in\mathcal{T}}\|\mathrm{I}_{a,n}\left(t\right)-b\left(t\right)\|\to_{\mathrm{P}}0$, which combined with $\sup_{t\in\mathcal{T}}\|\mathrm{I}_{b,n}\left(t\right)\|\to_{\mathrm{P}}0$ and T establishes Part (a).

Continuity and $\mathcal{T}\times\overline{B}$ compact also imply $(t,\beta)\mapsto\mathrm{E}[\omega\left(t,X\right)\left(\partial/\partial\beta\right)\rho(Z,\beta,h^*\left(W\right))]$ is bounded on $\mathcal{T}\times\overline{B}$ (cf. Rudin, 1976, Theorem 4.15). Part (b) then follows from $\beta_0\in\overline{B}$.

Step 2: $\|\mathbf{II}_n\|_{\mathcal{T}}$

In this step we show that II_n defined in (S.3) satisfies

$$
\|\mathrm{II}_n\|_{\mathcal{T}}\lesssim_{\mathrm{P}}\mathrm{E}\left[R\left(Z\right)\right]\sqrt{n}\max_{1\leqslant m\leqslant d}\|\widehat{h}_m-h_m^*\|_{\mathcal{W}}^{1+\gamma}
$$

for R and γ given by Assumption 3. Using T and CS, Assumptions 2 and 3 imply that

$$
\begin{aligned}
\|\mathrm{II}_n\|_T&\leqslant\|\omega\|_{\mathcal{T}\times\mathcal{T}}\sqrt{n}\mathbb{E}_n\left[\left\|\frac{\partial}{\partial h}\rho(Z_i,\overline{h}\left(W_i\right))-\frac{\partial}{\partial h}\rho(Z_i,h^*\left(W_i\right))\right\|\left\|\widehat{h}\left(W_i\right)-h\left(W_i\right)\right\|\right] \\
&\lesssim\sqrt{n}\mathbb{E}_n[R\left(Z_i\right)\|\overline{h}\left(W_i\right)-h^*\left(W_i\right)\|^{\gamma}\|\widehat{h}\left(W_i\right)-h^*\left(W_i\right)\|] \\
&\leqslant\sqrt{n}\mathbb{E}_n[R\left(Z_i\right)\|\widehat{h}\left(W_i\right)-h^*\left(W_i\right)\|^{1+\gamma}] \\
&\leqslant d^{(1+\gamma)/2}\mathbb{E}_n\left[R\left(Z_i\right)\right]\sqrt{n}\max_{1\leqslant m\leqslant d}\|\widehat{h}_m-h_m^*\|_{\mathcal{W}}^{1+\gamma} \\
&\lesssim_{\mathrm{P}}\mathrm{E}\left[R\left(Z\right)\right]\sqrt{n}\max_{1\leqslant m\leqslant d}\|\widehat{h}_m-h_m^*\|_{\mathcal{W}}^{1+\gamma},
\end{aligned}
$$

where $\overline{h}(W_i)$ is on the line segment connecting $\widehat{h}(W_i)$ and $h(W_i)$, thus satisfying $\|\overline{h}(W_i) - h^*(W_i)\| \leqslant \|\widehat{h}(W_i) - h^*(W_i)\|$, and $\mathbb{E}_n[R(Z_i)] \lesssim_{\mathrm{P}} \mathrm{E}[R(Z)]$ follows from M.

Step 3: $\|\mathbf{III}_n\|_{\mathcal{T}}$

In this step we show that III_n defined in (S.3) satisfies

$$\|\mathrm{III}_n\|_{\mathcal{T}} \lesssim_{\mathrm{P}} \max_{1\leqslant m\leqslant d}\left(\sum_{j=1}^{k_{m,n}}\|p_j\|_{\mathcal{W}}^2\right)^{1/2}\left(\sqrt{k_{m,n}/n}+k_{m,n}^{-\alpha_m}\right)$$

for α given by Assumption 6. For $h:\mathcal{W}\to\mathbf{R}^d$ composed by maps $\{h_m\}_1^d$ in $L^2(W)$, define the map D

$$D(t,z,h) \coloneqq \omega(t,x)\,(\partial/\partial h')\rho(z,h^*(w))\,h(w) \tag{S.6}$$

such that $h\mapsto D(t,z,h)$ is a linear functional for given $(t,z)\in\mathcal{T}\times\mathcal{Z}$. Let Δ denote the centered version of D, i.e.,

$$\begin{aligned}\Delta(t,z,h) \coloneqq{}& \omega(t,x)\,(\partial/\partial h')\rho(z,h^*(w))\,h(w)\\ &- \mathrm{E}_Z[\omega(t,X)\,(\partial/\partial h')\rho(Z,h^*(W))\,h(W)]\end{aligned} \tag{S.7}$$

which is also linear in h. Letting $\widetilde{h}_m = p^{k\prime}\widetilde{\pi}_m$ be as in Assumption 6, by linearity we may write

$$\begin{aligned}\mathrm{III}_n(t) &= \sqrt{n}\mathbb{E}_n\left[\Delta(t,Z_i,\widehat{h}-h^*)\right]\\ &= \sqrt{n}\mathbb{E}_n\left[\Delta(t,Z_i,\widehat{h}-\widetilde{h})\right]+\sqrt{n}\mathbb{E}_n\left[\Delta(t,Z_i,\widetilde{h}-h^*)\right]\\ &\eqqcolon \mathrm{III}_{a,n}(t)+\mathrm{III}_{b,n}(t)\,.\end{aligned} \tag{S.8}$$

Given that $\zeta_k = \sup_{w\in\mathcal{W}}[\sum_{j=1}^k p_j(w)^2]^{1/2}$ and $\sqrt{k}\lesssim\zeta_k$ (implied by Assumption 5), $\zeta_{k_n}\to\infty$ and thus $\sum_{j=1}^{k_n}\|p_j\|_{\mathcal{W}}^2\to\infty$. In particular, $\sum_{j=1}^{k_n}\|p_j\|_{\mathcal{W}}^2$ is bounded away from zero as $n \to \infty$. By T, the desired conclusion will therefore follow from showing

$$\|\mathrm{III}_{a,n}\|_{\mathcal{T}} \lesssim_{\mathrm{P}} \max_{1 \leqslant m \leqslant d} \left(\sum_{j=1}^{k_{m,n}} \|p_j\|_{\mathcal{W}}^2 \right)^{1/2} \left(\sqrt{k_{m,n}/n} + k_{m,n}^{-\alpha_m} \right),$$

$$\|\mathrm{III}_{b,n}\|_{\mathcal{T}} \lesssim_{\mathrm{P}} \max_{1 \leqslant m \leqslant d} k_{m,n}^{-\alpha_m}.$$

Step 3a: $\|\mathbf{III}_{a,n}\|_{\mathcal{T}}$ In this step we show that $\mathrm{III}_{a,n}$ defined in (S.8) satisfies

$$\|\mathrm{III}_{a,n}\|_{\mathcal{T}} \lesssim_{\mathrm{P}} \max_{1 \leqslant m \leqslant d} \left(\sum_{j=1}^{k_{m,n}} \|p_j\|_{\mathcal{W}}^2 \right)^{1/2} \left(\sqrt{k_{m,n}/n} + k_{m,n}^{-\alpha_m} \right)$$

for α_m given by Assumption 6. Given that

$$\mathrm{III}_{a,n}(t) = \sqrt{n}\mathbb{E}_n \left[\Delta(t, Z_i, \widehat{h} - \widetilde{h}) \right] = \sum_{m=1}^{d} \sqrt{n}\mathbb{E}_n \left[\Delta_m(t, Z_i, \widehat{h}_m - \widetilde{h}_m) \right],$$

$$\begin{aligned}\Delta_m(t, Z_i, h_m) &\coloneqq \omega(t,x)\,(\partial/\partial h_m)\rho(z, h^*(w))\,h_m(w) \\ &\quad - \mathrm{E}\left[\omega(t,X)\,(\partial/\partial h_m)\rho(Z, h^*(W))\,h_m(W)\right],\end{aligned}$$

by T, we may focus on bounding a single $\sup_{t\in\mathcal{T}}|\sqrt{n}\mathbb{E}_n[\Delta_m(t, Z_i, \widehat{h}_m - \widetilde{h}_m)]|$ in probability. For the remainder of this section we therefore drop the m subscript and write $(\partial/\partial h)\,\rho(Z, h^*(Z))$ for the scalar $(\partial/\partial h_m)\,\rho(Z, h^*(Z))$. Let

$$\Delta_i^k(t) \coloneqq (\Delta(t, Z_i, p_1), \ldots, \Delta(t, Z_i, p_k))'.$$

Then CS implies

$$\begin{aligned}\|\mathrm{III}_{a,n}\|_{\mathcal{T}} &= \sup_{t\in\mathcal{T}} \left| \sqrt{n}\mathbb{E}_n \left[\Delta(t, Z_i, p^{k_n\prime}(\widehat{\pi} - \widetilde{\pi})) \right] \right| = \sup_{t\in\mathcal{T}} \left| \sqrt{n} \left\{ \mathbb{E}_n[\Delta_i^{k_n}(t)] \right\}' (\widehat{\pi} - \widetilde{\pi}) \right| \\ &\leqslant \|\widehat{\pi} - \widetilde{\pi}\| \sup_{t\in\mathcal{T}} \left\| \sqrt{n}\mathbb{E}_n[\Delta_i^{k_n}(t)] \right\|.\end{aligned}$$

Lemma S.7 tells us that $\|\widehat{\pi} - \widetilde{\pi}\| \lesssim_{\mathrm{P}} \sqrt{k_n/n} + k_n^{-\alpha}$, so it remains to show that

$$\sup_{t\in\mathcal{T}} \left\| \sqrt{n}\mathbb{E}_n[\Delta_i^{k_n}(t)] \right\| \lesssim_{\mathrm{P}} \left(\sum_{j=1}^{k_n} \|p_j\|_{\mathcal{W}}^2 \right)^{1/2}.$$

By M it suffices to show the finite-sample moment bound, for any $k \in \mathbf{N}$,

$$\mathrm{E}\left[\sup_{t\in\mathcal{T}}\left\|\sqrt{n}\mathbb{E}_n[\Delta_i^k(t)]\right\|^2\right] \lesssim \sum_{j=1}^k \|p_j\|_{\mathcal{W}}^2.$$

Given that

$$\mathrm{E}\left[\sup_{t\in\mathcal{T}}\left\|\sqrt{n}\mathbb{E}_n[\Delta_i^k(t)]\right\|^2\right] \leqslant \sum_{j=1}^k \mathrm{E}\left[\sup_{t\in\mathcal{T}}\left|\sqrt{n}\mathbb{E}_n[\Delta(t, Z_i, p_j)\right|^2\right],$$

it suffices to show that

$$\mathrm{E}\left[\sup_{t\in\mathcal{T}}\left|\sqrt{n}\mathbb{E}_n[\Delta(t, Z_i, p_j)\right|^2\right] \lesssim \|p_j\|_{\mathcal{W}}^2, \quad j \in \{1, \dots, k\}.$$

To this end, fix $j \in \{1, \dots, k\}$, and consider the function class $\mathcal{F}_j \coloneqq \mathcal{F}_j(\mathcal{T}) := \{f : z \mapsto \Delta(t, z, p_j); t \in \mathcal{T}\}$. For $f_1 \coloneqq f(\cdot; t_1), f_2 \coloneqq f(\cdot; t_2) \in \mathcal{F}_j$ arbitrary, by T, J and Assumptions 2 and 3,

$$\begin{aligned}
&|f_1(z) - f_2(z)| \\
&= \Bigg| [\omega(t_1, x) - \omega(t_2, x)] \frac{\partial}{\partial h}\rho(z, h^*(w)) p_j(w) \\
&\quad - \mathrm{E}\left[\{\omega(t_1, X) - \omega(t_2, X)\} \frac{\partial}{\partial h}\rho(Z, h^*(W)) p_j(W)\right]\Bigg| \\
&\leqslant |\omega(t_1, x) - \omega(t_2, x)| \left|\frac{\partial}{\partial h}\rho(z, h^*(w))\right| |p_j(w)| \\
&\quad + \mathrm{E}\left[|\omega(t_1, X) - \omega(t_2, X)| \left|\frac{\partial}{\partial h}\rho(Z, h^*(W))\right| |p_j(W)|\right] \\
&\lesssim \left(\left|\frac{\partial}{\partial h}\rho(z, h^*(w))\right| |p_j(w)| + \mathrm{E}\left[\left|\frac{\partial}{\partial h}\rho(Z, h^*(W))\right| |p_j(W)|\right]\right) \|t_1 - t_2\| \\
&\leqslant \left(\left|\frac{\partial}{\partial h}\rho(z, h^*(w))\right| + \mathrm{E}\left[\left|\frac{\partial}{\partial h}\rho(Z, h^*(W))\right|\right]\right) \|p_j\|_{\mathcal{W}} \|t_1 - t_2\| \\
&= L_1(z) \|p_j\|_{\mathcal{W}} \|t_1 - t_2\|,
\end{aligned}$$

such that we may write

$$|f_1(z) - f_2(z)| \leqslant F_{1j}(z) \|t_1 - t_2\|, \quad F_{1j}(z) \coloneqq C_1 L_1(z) \|p_j\|_{\mathcal{W}},$$

for some constant $C_1 \in (0,\infty)$. Similarly, for $f \coloneqq f(\cdot;t) \in \mathcal{F}_j$ arbitrary, by T, J and Assumptions 2 and 3,

$$|f(z)| = \left|\omega(t,x)\frac{\partial}{\partial h}\rho(z,h^*(w))\,p_j(w) - \mathrm{E}_Z\left[\omega(t,X)\frac{\partial}{\partial h}\rho(Z,h^*(W))\,p_j(W)\right]\right| \\ \lesssim L_1(z)\,\|p_j\|_{\mathcal{W}},$$

such that we may write

$$|f(z)| \leqslant F_{2j}(z), \quad F_{2j}(z) \coloneqq C_2 L_1(z)\,\|p_j\|_{\mathcal{W}},$$

for some constant $C_2 \in (0,\infty)$. Let $C_3 \coloneqq C_1 \vee C_2$ and

$$F_j(z) \coloneqq C_3 L_1(z)\,\|p_j\|_{\mathcal{W}}.$$

Then $\|F_j\|_{P,2} \lesssim \|p_j\|_{\mathcal{W}}$, so F_j is an square-integrable envelope for $\mathcal{F}_j$ satisfying

$$|f_1(z) - f_2(z)| \leqslant F_j(z)\,\|t_1 - t_2\|.$$

Given that $\mathcal{T}$ is compact (Assumption 2), we must have $\mathrm{diam}(\mathcal{T}) < \infty$. Pollard (1990, Lemma 4.1) and the fact that covering numbers are bounded by packing numbers (cf. van der Vaart and Wellner, 1996, p. 98) therefore combine to yield $N(\varepsilon,\mathcal{T},\|\cdot\|) \leqslant (3\mathrm{diam}(\mathcal{T})/\varepsilon)^{d_t}$ for $\varepsilon \in (0,\mathrm{diam}(\mathcal{T})]$. Hence, by van der Vaart and Wellner (1996, Theorem 2.7.11) and the previous display,

$$N_{[\,]}(\varepsilon\|F_j\|_{P,2},\mathcal{F}_j,L^2(P)) \leqslant N(\varepsilon/2,\mathcal{T},\|\cdot\|) \leqslant (6\mathrm{diam}(\mathcal{T})/\varepsilon)^d \leqslant (C/\varepsilon)^d$$

for $\varepsilon \in (0,\mathrm{diam}(\mathcal{T})]$ (and $= 1$ otherwise). The bracketing integral of $\mathcal{F}_j$ therefore satisfies the bound

$$J_{[\,]}\left(\delta,\mathcal{F}_j,L^2(P)\right) \leqslant \int_0^\delta \sqrt{1 + C\ln(1/\varepsilon)}\,\mathrm{d}\varepsilon.$$

Note that the right-hand side depends on neither j nor k. In particular, the integral $J_{[\,]}(1,\mathcal{F}_j,L^2(P))$ is bounded uniformly in $j \in \{1,\ldots,k\}, k \in \mathbf{N}$. By construction, $\mathrm{E}[f(Z)] = \mathrm{E}[\Delta(t,Z,p_j)] = 0$ for any $f \in \mathcal{F}_j$, so we may view the stochastic process $\{\sqrt{n}\mathbb{E}_n[\Delta(t,Z_i,p_j)]; t \in \mathcal{T}\}$ as an empirical process $\{\mathbb{G}_n(f); f \in \mathcal{F}_j\}$. van der Vaart and Wellner (1996, Theorem 2.14.2) therefore implies the finite-sample bound

$$\mathrm{E}\left[\|\mathbb{G}_n\|_{\mathcal{F}_j}\right] \lesssim J_{[\,]}\left(1, \mathcal{F}_j, L^2(P)\right) \|F_j\|_{P,2} \lesssim \|F_j\|_{P,2} \lesssim \|p_j\|_{\mathcal{W}}.$$

van der Vaart and Wellner (1996, Theorem 2.14.5) now shows

$$\left(\mathrm{E}[\|\mathbb{G}_n\|^2_{\mathcal{F}_j}]\right)^{1/2} \lesssim \mathrm{E}\left[\|\mathbb{G}_n\|_{\mathcal{F}_j}\right] + \|F_j\|_{P,2} \lesssim \|p_j\|_{\mathcal{W}},$$

which is the desired bound.

Step 3b: $\|\mathbf{III}_{b,n}\|_{\mathcal{T}}$ In this step we show that $\mathrm{III}_{b,n}$ defined in (S.8) satisfies

$$\|\mathrm{III}_{b,n}\|_{\mathcal{T}} \lesssim_{\mathrm{P}} \max_{1\leqslant m\leqslant d} k_{m,n}^{-\alpha_m},$$

for α_m given by Assumption 6. Given that

$$\mathrm{III}_{b,n}(t) = \sqrt{n}\mathbb{E}_n\left[\Delta(t, Z_i, \widetilde{h} - h^*)\right] = \sum_{m=1}^{d} \sqrt{n}\mathbb{E}_n\left[\Delta_m(t, Z_i, \widetilde{h}_m - h^*_m)\right],$$

as was the case for $\|\mathrm{III}_{a,n}\|_{\mathcal{T}}$, by T we may focus on bounding each right-hand side term in probability and therefore drop the m subscript. For this purpose, fix $k \in \mathbf{N}$ and consider the function class $\mathcal{F}_k \coloneqq \mathcal{F}_k(\mathcal{T}) \coloneqq \{f\colon z \mapsto \Delta(t, z, \widetilde{h} - h^*); t \in \mathcal{T}\}$. For $f \coloneqq f(\cdot, t), f_1 \coloneqq f(\cdot, t_1), f_2 \coloneqq f(\cdot, t_2) \in \mathcal{F}_k$ arbitrary, arguments analogous to the ones applied to handle $\|\mathrm{III}_{a,n}\|_{\mathcal{T}}$ establish that

$$|f_1(z) - f_2(z)| \leqslant C_1 L_1(z) \|\widetilde{h} - h^*\|_{\mathcal{W}} \|t_1 - t_2\|,$$
$$|f(z)| \leqslant C_2 L_1(z) \|\widetilde{h} - h^*\|_{\mathcal{W}}.$$

Define $C_3 \coloneqq C_1 \vee C_2$ and $F_k(z) \coloneqq C_3 L_1(z) \|\widetilde{h} - h^*\|_{\mathcal{W}}$. Then $\|F_k\|_{P,2} = C_4\|\widetilde{h} - h^*\|_{\mathcal{W}} \lesssim k^{-\alpha}$ by Assumption 6. Hence F_k is an square-integrable envelope for $\mathcal{F}_k$, and arguments analogous to the ones used for $\|\mathrm{III}_{a,n}\|_{\mathcal{T}}$ show that the resulting bracketing integral $J_{[\,]}(\delta, \mathcal{F}_k, L^2(P))$ is bounded by a constant independent of k. van der Vaart and Wellner (1996, Theorem 2.14.2) therefore implies

$$\mathrm{E}\left[\|\mathbb{G}_n\|_{\mathcal{F}_k}\right] \lesssim J_{[\,]}\left(1, \mathcal{F}_k, L^2(P)\right) \|F_k\|_{P,2} \lesssim \|F_k\|_{P,2} \lesssim k^{-\alpha},$$

and the claim follows from M.

Step 4: $\|\mathbf{IV}_n\|_{\mathcal{T}}$

In this step we show that IV_n defined in (S.3) satisfies

$$\|\mathrm{IV}_n\|_{\mathcal{T}} \lesssim_{\mathrm{P}} \max_{1\leqslant m\leqslant d}\Big\{\sqrt{n}r_{h_m,k_{m,n}}\sup_{t\in\mathcal{T}} r_{\delta_m,k_{m,n}}(t)+\sqrt{\zeta^2_{k_{m,n}}k_{m,n}\ln(k_{m,n})/n} \\ + R_{\delta_m,k_{m,n}}\sqrt{\ln\left(k_{m,n}/R_{\delta_m,k_{m,n}}\right)}+\zeta_{k_{m,n}}r_{h_m,k_{m,n}}\Big\},$$

where $\zeta_k, r_{h_m,k}, r_{\delta_m,k}$ and $R_{\delta_m,k}$ are defined in (3.5), (3.11), (3.12) and (3.13), respectively. Given the decomposition

$$\begin{aligned}\mathrm{IV}_n(t) &= \sqrt{n}\Big(\mathrm{E}_Z\Big[\omega(t,X)(\partial/\partial h')\rho(Z,h^*(W))\{\widehat{h}(W)-h^*(W)\}\Big] \\ &\quad - \mathbb{E}_n[\delta(t,W_i)'\{Y_i-h^*(W_i)\}]\Big) \\ &= \sum_{m=1}^{d}\sqrt{n}\Big(\mathrm{E}_Z\Big[\omega(t,X)(\partial/\partial h_m)\rho(Z,h^*(W))\{\widehat{h}_m(W)-h^*_m(W)\}\Big] \\ &\quad - \mathbb{E}_n[\delta_m(t,W_i)\{Y_{mi}-h^*_m(W_i)\}]\Big),\end{aligned}$$

by T we may drop the m subscript and focus on bounding a single summand uniformly over $\mathcal{T}$ in probability. For this purpose, recall that h_k and $\delta_k(t,\cdot)$ are the mean-square projections of $h^*$and $\delta(t,\cdot)$, respectively, onto the linear span of p^k and $r^2_{h,k}$ and $r^2_{\delta,k}(t)$ are the mean-square errors resulting from these projections. Define

$$\psi_k(t) \coloneqq \mathrm{E}\left[\delta(t,W)p^k(W)\right]. \tag{S.9}$$

By Assumption 5, the population least-square coefficients $\pi_k = Q_k^{-1}\mathrm{E}[p^k(W)Y]$ are well defined for all $k\in\mathbf{N}$. Applying Lemma S.3, we see that the inverse of $\widehat{Q}_{k_n} \coloneqq \mathbb{E}_n[p^{k_n}(W_i)p^{k_n}(W_i)']$ exists wp $\to 1$. As a consequence, the sample least-squares coefficients take the form $\widehat{\pi} = \widehat{Q}_{k_n}^{-1}\mathbb{E}_n[p^{k_n}(W_i)Y_i]$ wp $\to 1$. Assuming—without loss of generality—that $\widehat{Q}_{k_n}^{-1}$ exists with probability one for all n,

$$
\begin{aligned}
\sqrt{n}\mathrm{E}_W\{\delta(t,W)[\widehat{h}(W)-h_{k_n}(W)]\} &= \sqrt{n}\mathrm{E}_W\{\delta(t,W)\,p^{k_n}(W)'(\widehat{\pi}-\pi_{k_n})\} \\
&= \psi_{k_n}(t)'\sqrt{n}(\widehat{\pi}-\pi_{k_n}) \\
&= \psi_{k_n}(t)'\sqrt{n}\left(\widehat{Q}_{k_n}^{-1}\mathbb{E}_n\left[p^{k_n}(W_i)Y_i\right]-\pi_{k_n}\right) \\
&= \psi_{k_n}(t)'\widehat{Q}_{k_n}^{-1}\sqrt{n}\left(\mathbb{E}_n\left[p^{k_n}(W_i)Y_i\right]-\widehat{Q}_{k_n}\pi_{k_n}\right) \\
&= \psi_{k_n}(t)'\widehat{Q}_{k_n}^{-1}\sqrt{n}\mathbb{E}_n\left[p^{k_n}(W_i)\{Y_i-h_{k_n}(W_i)\}\right],
\end{aligned}
$$

where $\mathrm{E}_W[\cdot]$ denotes integration with respect to the distribution of W. By definition of $\delta(t,W)$ [see (3.8)] and iterated expectations, for a nonrandom function h of W alone,

$$
\mathrm{E}[\omega(t,X)(\partial/\partial h)\rho(Z,h^*(W))\,h(W)] = \mathrm{E}[\delta(t,W)\,h(W)].
$$

Using the previous two displays and adding and subtracting

$$
\begin{aligned}
&\sqrt{n}\mathbb{E}_n[\delta_{k_n}(t,W_i)\{Y_i-h_{k_n}(W_i)\}] \\
&= \sqrt{n}\mathbb{E}_n\left[p^{k_n}(W_i)'Q_{k_n}^{-1}\mathrm{E}[p^{k_n}(W)\,\delta(t,W)]\{Y_i-h_{k_n}(W_i)\}\right] \\
&= \psi_{k_n}(t)'Q_{k_n}^{-1}\sqrt{n}\mathbb{E}_n\left[p^{k_n}(W_i)\{Y_i-h_{k_n}(W_i)\}\right],
\end{aligned}
$$

we may decompose $\mathrm{IV}_n(t)$ as

$$
\begin{aligned}
\mathrm{IV}_n(t) &= \sqrt{n}\mathrm{E}_W[\delta(t,W)\{\widehat{h}(W)-h^*(W)\}] - \sqrt{n}\mathbb{E}_n[\delta(t,W_i)\{Y_i-h^*(W_i)\}] \\
&= \sqrt{n}\mathrm{E}_W\{\delta(t,W)[h_{k_n}(W)-h^*(W)]\} + \sqrt{n}\mathrm{E}_W[\delta(t,W)\{\widehat{h}(W)-h_{k_n}(W)\}] \\
&\quad + \sqrt{n}\mathbb{E}_n[\delta(t,W_i)\{Y_i-h^*(W_i)\}] \\
&= \sqrt{n}\mathrm{E}_W[\delta(t,W)\{h_{k_n}(W)-h^*(W)\}] \\
&\quad + \psi_{k_n}(t)'(\widehat{Q}_{k_n}^{-1}-Q_{k_n}^{-1})\sqrt{n}\mathbb{E}_n\left[p^{k_n}(W_i)\{Y_i-h_{k_n}(W_i)\}\right] \\
&\quad + \sqrt{n}\mathbb{E}_n[\delta_{k_n}(t,W_i)\{Y_i-h_{k_n}(W_i)\}-\delta(t,W_i)\{Y_i-h^*(W_i)\}] \\
&=: \mathrm{IV}_{a,n}(t)+\mathrm{IV}_{b,n}(t)+\mathrm{IV}_{c,n}(t).
\end{aligned}
$$

By T it therefore suffices to show that

$$\begin{aligned}
\|\mathrm{IV}_{a,n}\|_{\mathcal{T}} &\leqslant \sqrt{n} r_{h,k_n} \sup_{t\in\mathcal{T}} r_{\delta,k_n}(t), \\
\|\mathrm{IV}_{b,n}\|_{\mathcal{T}} &\lesssim_{\mathrm{P}} \sqrt{\zeta_{k_n}^2 k_n \ln(k_n)/n}, \quad \text{and} \\
\|\mathrm{IV}_{c,n}\|_{\mathcal{T}} &\lesssim_{\mathrm{P}} R_{\delta,k_n}\sqrt{\ln(k_n/R_{\delta,k_n})} + \zeta_{k_n} r_{h,k_n}.
\end{aligned}$$

Step 4a: $\|\mathbf{IV}_{a,n}\|_{\mathcal{T}}$ In order to establish the inequality

$$\|\mathrm{IV}_{a,n}\|_{\mathcal{T}} \leqslant \sqrt{n} r_{h,k_n} \sup_{t\in\mathcal{T}} r_{\delta,k_n}(t),$$

recall that h_k defined in (S.4) is the mean-square projection of h^* onto the span of p^k, so by orthogonality of projections we have $\mathrm{E}\left[\delta_k(t,W)\{h_k(W) - h^*(W)\}\right] = 0$ for each $t \in \mathcal{T}$. Now J followed by CS yield

$$\begin{aligned}
\|\mathrm{IV}_{a,n}\|_{\mathcal{T}} &= \sqrt{n} \sup_{t\in\mathcal{T}} \left|\mathrm{E}\left[\delta(t,W)\{h_{k_n}(W) - h^*(W)\}\right]\right| \\
&= \sqrt{n} \sup_{t\in\mathcal{T}} \left|\mathrm{E}\left[\{\delta_{k_n}(t,W) - \delta(t,W)\}\{h_{k_n}(W) - h^*(W)\}\right]\right| \\
&\leqslant \sqrt{n}\, \|h_{k_n} - h^*\|_{P,2} \sup_{t\in\mathcal{T}} \|\delta_{k_n}(t,\cdot) - \delta(t,\cdot)\|_{P,2} = \sqrt{n} r_{h,k_n} \sup_{t\in\mathcal{T}} r_{\delta,k_n}(t).
\end{aligned}$$

Step 4b: $\|\mathbf{IV}_{b,n}\|_{\mathcal{T}}$ In this step we show that

$$\|\mathrm{IV}_{b,n}\|_{\mathcal{T}} \lesssim_{\mathrm{P}} \sqrt{\zeta_{k_n}^2 k_n \ln(k_n)/n}.$$

Using the fact that mean-square projections and conditional expectations are $L^2(P)$-contractions followed by Assumptions 2 and 3, we see that

$$\begin{aligned}
\psi_k(t)' Q_k^{-1} \psi_k(t) &= \{Q_k^{-1}\mathrm{E}[p^k(W)\,\delta(t,W)]\}' Q_k \{Q_k^{-1}\mathrm{E}[p^k(W)\,\delta(t,W)]\} \\
&= \mathrm{E}[\delta_k(t,W)^2] \leqslant \mathrm{E}[\delta(t,W)^2] \leqslant \mathrm{E}[\omega(t,X)^2 \,(\partial/\partial h)\,\rho(Z,h^*(W))^2] \\
&\lesssim \mathrm{E}[(\partial/\partial h)\,\rho(Z,h^*(W))^2] < \infty,
\end{aligned}$$

with an upper bound that depends on neither t nor k. By the Min-Max Theorem, Assumption 5, and the previous display, it follows that

$$\begin{aligned}\|\psi_k(t)\,Q_k^{-1}\|^2 &= [\psi_k(t)\,Q_k^{-1/2}]'Q_k^{-1}[Q_k^{-1/2}\psi_K(t)] \lesssim \|\psi_k(t)\,Q_k^{-1/2}\|^2\\ &\leqslant \sup_{k\in\mathbf{N},t\in\mathcal{T}}|\psi_k(t)'\,Q_k^{-1}\psi_k(t)| < \infty,\end{aligned}$$

thus implying $\sup_{k\in\mathbf{N},t\in\mathcal{T}}\|\psi_k(t)\,Q_k^{-1}\| < \infty$. By Lemma S.6 we have $\|\widehat{Q}_{k_n} - Q_{k_n}\|_{\text{op}} \lesssim_{\text{P}} [\zeta_{k_n}^2 \ln(k_n)/n]^{1/2} \to 0$ under Assumption 7. Moreover, Lemma S.3 shows that $\|\widehat{Q}_{k_n}^{-1}\|_{\text{op}} \lesssim_{\text{P}} 1$. Using these observations and the previous display,

$$\begin{aligned}\sup_{t\in\mathcal{T}}\|\psi_{k_n}(t)'\,\widehat{Q}_{k_n}^{-1} - \psi_{k_n}(t)'\,Q_k^{-1}\| &= \sup_{t\in\mathcal{T}}\|\psi_{k_n}(t)'\,Q_{k_n}^{-1}(Q_{k_n} - \widehat{Q}_{k_n})\widehat{Q}_{k_n}^{-1}\|\\ &\leqslant \|(Q_{k_n} - \widehat{Q}_{k_n})\widehat{Q}_{k_n}^{-1}\|_{\text{op}}\sup_{t\in\mathcal{T}}\|\psi_{k_n}(t)'\,Q_{k_n}^{-1}\|\\ &\leqslant \|\widehat{Q}_{k_n} - Q_{k_n}\|_{\text{op}}\|\widehat{Q}_{k_n}^{-1}\|_{\text{op}}\sup_{t\in\mathcal{T}}\|\psi_{k_n}(t)'\,Q_{k_n}^{-1}\|\\ &\lesssim_{\text{P}} \sqrt{\zeta_{k_n}^2\ln(k_n)/n} \to 0.\end{aligned}$$

From the previous display and $\sup_{k\in\mathbf{N},t\in\mathcal{T}}\|\psi_k(t)'\,Q_k^{-1}\| < \infty$ it follows that

$$\sup_{t\in\mathcal{T}}\|\psi_{k_n}(t)'\,\widehat{Q}_{k_n}^{-1}\| \lesssim_{\text{P}} 1.$$

Observe also that, by the Assumption 5, the Min-Max theorem, and the fact that $\text{E}[p^k(W)\{Y - h_k(W)\}] = 0$ (which follows from h_k being the mean-square projection of h^*),

$$\begin{aligned}&\text{E}\Big[\Big\|Q_k^{-1}\sqrt{n}\mathbb{E}_n\left[p^k(W_i)\{Y_i - h_k(W_i)\}\right]\Big\|^2\Big]\\ &\lesssim \text{E}\Big[\Big\|Q_k^{-1/2}\sqrt{n}\mathbb{E}_n\left[p^k(W_i)\{Y_i - h_k(W_i)\}\right]\Big\|^2\Big]\\ &= \text{E}\left[p^k(W)'\,Q_k^{-1}p^k(W)\{Y - h_k(W)\}^2\right]\\ &= \text{E}\Big[U^2p^k(W)'\,Q_k^{-1}p^k(W)\Big] + \text{E}\left[p^k(W)'\,Q_k^{-1}p^k(W)\{h_k(W) - h^*(W)\}^2\right],\end{aligned}$$

where we have used $U = Y - h^*(W)$. By Assumption 4, $\text{E}[U^2|W]$ is bounded, so

$$\begin{aligned}\text{E}[U^2p^k(W)'\,Q_k^{-1}p^k(W)] &= \text{E}[\text{E}\left[U^2|W\right]p^k(W)'\,Q_k^{-1}p^k(W)]\\ &\lesssim \text{E}[p^k(W)'\,Q_k^{-1}p^k(W)] = k.\end{aligned}$$

Moreover, by Assumption 5,

$$\begin{aligned}&\mathrm{E}\left[p^{k}\left(W\right)^{\prime}Q_{k}^{-1}p^{k}\left(W\right)\left\{h_{k}\left(W\right)-h^{*}\left(W\right)\right\}^{2}\right]\\&\lesssim\mathrm{E}\left[\|p^{k}\left(W\right)\|^{2}\{h_{k}\left(W\right)-h^{*}\left(W\right)\}^{2}\right]\leqslant\zeta_{k}^{2}r_{h,k}^{2}.\end{aligned}$$

Given Assumption 7, $\zeta_k^2 r_{h,k}^2 = (\zeta_k r_{h,k})^2 \to 0$ as $k \to \infty$, so

$$\mathrm{E}\Big[\Big\|Q_{k}^{-1}\sqrt{n}\mathbb{E}_{n}\left[p^{k}\left(W_{i}\right)\left\{Y_{i}-h_{k}\left(W_{i}\right)\right\}\right]\Big\|^{2}\Big]\lesssim k.$$

M now implies

$$\left\|Q_{k}^{-1}\sqrt{n}\mathbb{E}_{n}\left[p^{k_{n}}\left(W_{i}\right)\left\{Y_{i}-h_{k_{n}}\left(W_{i}\right)\right\}\right]\right\|\lesssim_{\mathrm{P}}\sqrt{k_{n}}.$$

Using CS we therefore arrive at

$$\begin{aligned}\|\mathrm{IV}_{b,n}\|_{\mathcal{T}}&=\sup_{t\in\mathcal{T}}\left|\psi_{k_{n}}\left(t\right)^{\prime}\widehat{Q}_{k_{n}}^{-1}(Q_{k_{n}}-\widehat{Q}_{k_{n}})Q_{k_{n}}^{-1}\sqrt{n}\mathbb{E}_{n}\left[p^{k_{n}}\left(W_{i}\right)\left\{Y_{i}-h_{k_{n}}\left(W_{i}\right)\right\}\right]\right|\\&\leqslant\left\|Q_{k_{n}}^{-1}\sqrt{n}\mathbb{E}_{n}\left[p^{k_{n}}\left(W_{i}\right)\left\{Y_{i}-h_{k_{n}}\left(W_{i}\right)\right\}\right]\right\|\sup_{t\in\mathcal{T}}\|\psi_{k_{n}}\left(t\right)^{\prime}\widehat{Q}_{k_{n}}^{-1}(Q_{k_{n}}-\widehat{Q}_{k_{n}})\|\\&\leqslant\left\|Q_{k_{n}}^{-1}\sqrt{n}\mathbb{E}_{n}\left[p^{k_{n}}\left(W_{i}\right)\left\{Y_{i}-h_{k_{n}}\left(W_{i}\right)\right\}\right]\right\|\|\widehat{Q}_{k_{n}}-Q_{k_{n}}\|_{\mathrm{op}}\sup_{t\in\mathcal{T}}\|\psi_{k_{n}}\left(t\right)^{\prime}\widehat{Q}_{k_{n}}^{-1}\|\\&\lesssim_{\mathrm{P}}\sqrt{k_{n}}\sqrt{\zeta_{k_{n}}^{2}\ln\left(k_{n}\right)/n}.\end{aligned}$$

Step 4c: $\|\mathbf{IV}_{c,n}\|_{\mathcal{T}}$ In this section we show that

$$\|\mathrm{IV}_{c,n}\|_{\mathcal{T}}\lesssim_{\mathrm{P}}R_{\delta,k_{n}}\sqrt{\ln\left(k_{n}/R_{\delta,k_{n}}\right)}+\zeta_{k_{n}}r_{h,k_{n}}.$$

Letting $U_i \coloneqq Y_i - h^*(W_i)$, we may decompose $\mathrm{IV}_{c,n}(t)$ as

$$\begin{aligned}&\mathrm{IV}_{c,n}\left(t\right)\\&=\sqrt{n}\mathbb{E}_{n}\left[U_{i}\left\{\delta_{k_{n}}\left(t,W_{i}\right)-\delta\left(t,W_{i}\right)\right\}\right]-\sqrt{n}\mathbb{E}_{n}\left[\delta_{k_{n}}\left(t,W_{i}\right)\left\{h_{k_{n}}\left(W_{i}\right)-h^{*}\left(W_{i}\right)\right\}\right]\\&=:\mathrm{IV}_{d,n}\left(t\right)+\mathrm{IV}_{e,n}\left(t\right).\end{aligned}$$

By T it therefore suffices to show that

$$\|\mathrm{IV}_{d,n}\|_{\mathcal{T}}\lesssim_{\mathrm{P}}R_{\delta,k_{n}}\sqrt{\ln\left(k_{n}/R_{\delta,k_{n}}\right)}\quad\text{and}\quad\|\mathrm{IV}_{e,n}\|_{T}\lesssim_{\mathrm{P}}\zeta_{k_{n}}r_{h,k_{n}}.$$

For the purpose of bounding $\|\mathrm{IV}_{d,n}\|_{\mathcal{T}}$, consider the function class $\mathcal{F}_k \coloneqq \mathcal{F}_k(\mathcal{T}) \coloneqq \{f : z \mapsto \{y - h^*(w)\}\{\delta_k(t,w) - \delta(t,w)\}; t \in \mathcal{T}\}$. Note that $\mathrm{E}[f(Z)] = 0$ for any $f \in \mathcal{F}_k$, so we may view the stochastic process $\{\mathrm{IV}_{d,n}(t); t \in \mathcal{T}\}$ as an empirical process $\{\mathbb{G}_n(f); f \in \mathcal{F}_k\}$. For any $t_1, t_2 \in \mathcal{T}$, by conditional J we have

$$\begin{aligned}
|\delta(t_1,w) - \delta(t_2,w)| &= |\mathrm{E}[\{\omega(t_1,X) - \omega(t_2,X)\}(\partial/\partial h)\rho(Z,h^*(W))|\, W = w]| \\
&\lesssim \mathrm{E}[|(\partial/\partial h)\rho(Z,h^*(W))|\,|W = w]\,\|t_1 - t_2\|.
\end{aligned}$$

Consequently, using Assumption 3 and the fact that conditional expectations are $L^2(P)$ contractions,

$$\begin{aligned}
\mathrm{E}\left[\{\delta(t_1,W) - \delta(t_2,W)\}^2\right] &\lesssim \mathrm{E}\left[\left\{\mathrm{E}\left[\left|\frac{\partial}{\partial h}\rho(Z,h^*(W))\right| |W\right]\right\}^2\right]\|t_1 - t_2\|^2 \\
&\leqslant \mathrm{E}\left[\left\{\frac{\partial}{\partial h}\rho(Z,h^*(W))\right\}^2\right]\|t_1 - t_2\|^2 \lesssim \|t_1 - t_2\|^2.
\end{aligned}$$

Given that mean-square projections are also $L^2(P)$ contractions,

$$\begin{aligned}
&\left\|Q_k^{-1/2}\mathrm{E}\left[p^k(W)\{\delta(t_1,W) - \delta(t_2,W)\}\right]\right\|^2 \\
&= \mathrm{E}\left[\left(p^k(W)' Q_k^{-1}\mathrm{E}\left[p^k(W)\{\delta(t_1,W) - \delta(t_2,W)\}\right]\right)^2\right] \\
&\leqslant \mathrm{E}\left[\{\delta(t_1,W) - \delta(t_2,W)\}^2\right]
\end{aligned}$$

so by CS and the previous two displays,

$$\begin{aligned}
|\delta_k(t_1,w) - \delta_k(t_2,w)| &= \left|p^k(w)' Q_k^{-1}\mathrm{E}\left[p^k(W)\{\delta(t_1,W) - \delta(t_2,W)\}\right]\right| \\
&\leqslant \left\|p^k(w)' Q_k^{-1/2}\right\| \left\|Q_k^{-1/2}\mathrm{E}\left[p^k(W)\{\delta(t_1,W) - \delta(t_2,W)\}\right]\right\| \\
&\lesssim \left\|p^k(w)' Q_k^{-1/2}\right\| \|t_1 - t_2\|. \qquad \text{(S.10)}
\end{aligned}$$

Thus, for any $f_1 \coloneqq f(\cdot, t_1), f_2 \coloneqq f(\cdot, t_2) \in \mathcal{F}_k$, by T,

$$
\begin{aligned}
&|f_1(z) - f_2(z)| \\
&\leqslant |y - h^*(w)| \left(|\delta_k(t_1, w) - \delta_k(t_2, w)| + |\delta(t_1, w) - \delta(t_2, w)|\right) \\
&\leqslant C\,|y - h^*(w)| \left\{ \left\| p^k(w)^\top Q_k^{-1/2} \right\| + \mathrm{E}\left[|(\partial/\partial h)\,\rho(Z, h^*(W))| \,|W = w \right] \right\} \|t_1 - t_2\| \\
&=: F_{1k}(z)\,\|t_1 - t_2\|\,.
\end{aligned}
$$

Moreover, for any $f \coloneqq f(\cdot, t) \in \mathcal{F}_k$,

$$
\begin{aligned}
|f(z)| &= |y - h^*(w)|\,|\delta_k(t, w) - \delta(t, w)| \\
&\leqslant |y - h^*(w)|\,\|\delta_k(\cdot, w) - \delta(\cdot, w)\|_{\mathcal{T}} =: F_{2k}(z)\,.
\end{aligned}
$$

Using Assumptions 3 and 4, the inequality $(a+b)^2 \leqslant 2a^2 + 2b^2$, and the fact that conditional expectations are $L^2(P)$ contractions, we see that

$$
\begin{aligned}
\mathrm{E}[F_{1k}(Z)^2] &\lesssim \mathrm{E}\left[U^2 \left\{ \left\| p^k(W)' Q_k^{-1/2} \right\| + \mathrm{E}\left[\left| \frac{\partial}{\partial h} \rho(Z, h^*(W)) \right| |W \right] \right\}^2 \right] \\
&\lesssim \mathrm{E}\left[\left\| p^k(W)' Q_k^{-1/2} \right\|^2 \right] + \mathrm{E}\left[\left\{ \mathrm{E}\left[\left| \frac{\partial}{\partial h} \rho(Z, h^*(W)) \right| |W \right] \right\}^2 \right] \\
&\leqslant k + \mathrm{E}\left[\left(\partial/\partial h \right) \rho(Z, h^*(W))^2 \right] \lesssim k.
\end{aligned}
$$

Given Assumptions 4 and 7, we get

$$
\mathrm{E}[F_{2k}(Z)^2] = \mathrm{E}\left[U^2 \,\|\delta_k(\cdot, W) - \delta(\cdot, W)\|_{\mathcal{T}}^2 \right] \lesssim \mathrm{E}\left[\|\delta_k(\cdot, W) - \delta(\cdot, W)\|_{\mathcal{T}}^2 \right] = R_{\delta,k}^2 \to 0
$$

as $k \to \infty$. Thus, defining $F_k \coloneqq F_{1,k} + F_{2,k}$ we must have

$$
\mathrm{E}[F_k(Z)^2] \lesssim k + R_{\delta,k}^2 \lesssim k \quad \text{as } k \to \infty,
$$

and it follows that F_k is a square-integrable envelope for $\mathcal{F}_k$ satisfying

$$
|f_1(z) - f_2(z)| \leqslant F_k(z)\,\|t_1 - t_2\| \quad \text{and} \quad \|F_k\|_{P,2} \lesssim k^{1/2} \ \text{ as } k \to \infty.
$$

Using $\mathcal{T}$ compact and the previous display, van der Vaart and Wellner (1996, Theorem 2.7.11) implies that

$$N_{[\,]}(\varepsilon\|F_k\|_{P,2},\mathcal{F}_k,L^2(P))\leqslant(C/\varepsilon)^{d_t}\,,\quad\varepsilon\in(0,1],$$

and thus

$$J_{[\,]}\left(\delta,\mathcal{F}_k,L^2(P)\right)\leqslant\int_0^\delta\sqrt{1+d_t\ln(C/\varepsilon)}\mathrm{d}\varepsilon,\quad\delta\in(0,1],$$

where the right-hand side does not depend on k. In particular, $J_{[\,]}(1,\mathcal{F}_{k_n},L^2(P))\lesssim 1$.

Defining

$$\sigma_n^2\coloneqq\sup_{f\in\mathcal{F}_{k_n}}\mathbb{E}_n\left[f(Z_i)^2\right]$$

we see that

$$\sigma_n^2=\sup_{t\in\mathcal{T}}\mathbb{E}_n\left[U_i^2\left\{\delta_{k_n}(t,W_i)-\delta(t,W_i)\right\}^2\right]\leqslant\mathbb{E}_n\left[U_i^2\|\delta_{k_n}(\cdot,W_i)-\delta(\cdot,W_i)\|_{\mathcal{T}}^2\right],$$

such that

$$\mathrm{E}\left[\sigma_n^2\right]\leqslant\mathrm{E}\left[U^2\|\delta_{k_n}(\cdot,W)-\delta(\cdot,W)\|_{\mathcal{T}}^2\right]\lesssim\mathrm{E}\left[\|\delta_{k_n}(\cdot,W)-\delta(\cdot,W)\|_{\mathcal{T}}^2\right]=R_{\delta,k_n}^2.$$

There are two cases: (1) $R_{\delta,k_n}/\|F_{k_n}\|_{P,2}\to 0$ and (2) $R_{\delta,k_n}/\|F_{k_n}\|_{P,2}\nrightarrow 0$.

Case 1: $R_{\delta,k_n}/\|F_{k_n}\|_{P,2}\to 0$. Given that $\sqrt{\mathrm{E}[\sigma_n^2]}\leqslant C_1R_{\delta,k_n}$, by the change of variables $\varepsilon':=\varepsilon/C_1$ we have

$$\begin{aligned}J_{[\,]}\left(\sqrt{\mathrm{E}[\sigma_n^2]}/\|F_{k_n}\|_{P,2},\mathcal{F}_{k_n},L^2(P)\right)&\leqslant J_{[\,]}\left(C_1R_{\delta,k_n}/\|F_{k_n}\|_{P,2},\mathcal{F}_{k_n},L^2(P)\right)\\&=C_1\int_0^{R_{\delta,k_n}/\|F_{k_n}\|_{P,2}}\sqrt{1+d_t\ln(C_3/\varepsilon')}\mathrm{d}\varepsilon'\\&=:C_1\overline{J}_{[\,]}\left(R_{\delta,k_n}/\|F_{k_n}\|_{P,2}\right).\end{aligned}\tag{S.11}$$

van der Vaart and Wellner (2011, p. 196) establishes the maximal inequality

$$\mathrm{E}\left[\|\mathbb{G}_n\|_{\mathcal{F}_{k_n}}\right]\lesssim J_{[\,]}\left(\sqrt{\mathrm{E}[\sigma_n^2]}/\|F_{k_n}\|_{P,2},\mathcal{F}_{k_n},L^2(P)\right)\|F_{k_n}\|_{P,2}.$$

The previous two displays show that

$$\mathrm{E}\left[\|\mathbb{G}_n\|_{\mathcal{F}_{k_n}}\right]) \lesssim \overline{J}_{[\,]}\left(R_{\delta,k_n}/\|F_{k_n}\|_{P,2}\right)\|F_{k_n}\|_{P,2}$$

and from van der Vaart and Wellner (1996, p. 239) we know that an integral of the form $\int_0^\delta [1+\ln(1/u)]^{1/2}\mathrm{d}u$—as in (S.11)—satisfies $\int_0^\delta [1+\ln(1/u)]^{1/2}\mathrm{d}u \lesssim \delta\sqrt{\ln(1/\delta)}$ as $\delta \to 0_+$. Since $R_{\delta,k_n}/\|F_{\delta,k_n}\|_{P,2} \to 0$ holds by hypothesis, the previous display combined with $\|F_{k_n}\|_{P,2} \lesssim \sqrt{k_n}$ and M yields

$$\begin{aligned}\|\mathbb{G}_n\|_{\mathcal{F}_{k_n}} &\lesssim_{\mathrm{P}} \left(R_{\delta,k_n}/\|F_{k_n}\|_{\mathrm{P},2}\right)\sqrt{\ln\left(\|F_{k_n}\|_{P,2}/R_{\delta,k_n}\right)}\|F_{k_n}\|_{P,2}\\ &= R_{\delta,k_n}\sqrt{\ln\left(\|F_{k_n}\|_{P,2}/R_{\delta,k_n}\right)} \lesssim R_{\delta,k_n}\sqrt{\ln\left(k_n/R_{\delta,k_n}\right)}.\end{aligned}$$

Case 2. $R_{\delta,k_n}/\|F_{k_n}\|_{P,2} \nrightarrow 0$. Given that $R_{\delta,k_n} \to 0$ (Assumption 7), we must have $\|F_{k_n}\|_{P,2} \lesssim R_{\delta,k}$. van der Vaart and Wellner (1996, Theorem 2.14.2) and $J_{[\,]}\left(1,\mathcal{F}_{k_n},L^2(P)\right) \lesssim 1$ yield

$$\mathrm{E}\left[\|\mathbb{G}_n\|_{\mathcal{F}_{k_n}}\right] \lesssim J_{[\,]}\left(1,\mathcal{F}_{k_n},L^2(P)\right)\|F_{k_n}\|_{P,2} \lesssim \|F_{k_n}\|_{P,2} \lesssim R_{\delta,k_n} \lesssim R_{\delta,k_n}\sqrt{\ln\left(\frac{k_n}{R_{\delta,k_n}}\right)}.$$

M now yields the same rate as in Case 1. In either case, we observe that $\|\mathrm{IV}_{d,n}\|_{\mathcal{T}} \lesssim_{\mathrm{P}} R_{\delta,k_n}\sqrt{\ln\left(k_n/R_{\delta,k_n}\right)}$.

For the purpose of bounding $\|\mathrm{IV}_{e,n}\|_{\mathcal{T}}$, consider the function class $\mathcal{F}_k \coloneqq \{f : z \mapsto \delta_k(t,w)\{h_k(w)-h^*(w)\}\,;t\in\mathcal{T}\}$. Note that, by orthogonality of mean–square projections we have $\mathrm{E}[f(Z)] = 0$ for any $f \in \mathcal{F}_k$, so we may view the stochastic process $\{\mathrm{IV}_{e,n}(t)\,;t\in\mathcal{T}\}$ as an empirical process $\{\mathbb{G}_n(f); f \in \mathcal{F}_{k_n}\}$. For any $t_1,t_2 \in \mathcal{T}$, using the bound in (S.10) we have that $f_1 \coloneqq f(\cdot;t_1), f_2 \coloneqq f(\cdot;t_2) \in \mathcal{F}_k$, satisfy

$$\begin{aligned}|f_1(z)-f_2(z)| &= |\delta_k(t_1,w)-\delta_k(t_2,w)|\,|h_k(w)-h^*(w)|\\ &\lesssim \left\|p^k(w)' Q_k^{-1/2}\right\||h_k(w)-h^*(w)|\,\|t_1-t_2\|\\ &\lesssim \zeta_k |h_k(w)-h^*(w)|\,\|t_1-t_2\|.\end{aligned}$$

The previous display implies

$$|f_1(z)-f_2(z)| \leqslant F_{1,k}(z)\,\|t_1-t_2\|,$$

for $F_{1k}(z) \coloneqq C_1\zeta_k |h_k(w) - h^*(w)|$ and some $C_1 \in (0,\infty)$. Since conditional expectations are $L^2(P)$ contractions, by Assumptions 2 and 3,

$$\begin{aligned}\mathrm{E}[\delta(t,W)^2] &= \mathrm{E}\left[\left\{\mathrm{E}\Big[\omega(t,X)\Big|\frac{\partial}{\partial h}\rho(Z,h^*(W))\Big|\Big|W\Big]\right\}^2\right] \\ &\leqslant \mathrm{E}\left[\omega(t,X)^2\frac{\partial}{\partial h}\rho(Z,h^*(W))^2\right] \lesssim \mathrm{E}\Big[\frac{\partial}{\partial h}\rho(Z,h^*(W))^2\Big] < \infty,\end{aligned}$$

thus implying $\sup_{t\in\mathcal{T}}\mathrm{E}[\delta(t,W)^2] < \infty$. By CS and using that mean–square projections are $L^2(P)$ contractions as well, we get

$$\begin{aligned}|\delta_k(t,w)| &= |p^k(w)' Q_k^{-1}\mathrm{E}[p^k(W)\delta(t,W)]| \\ &\leqslant \left\|p^k(w)' Q_k^{-1/2}\right\| \left\|Q_k^{-1/2}\mathrm{E}[p^k(W)\delta(t,W)]\right\| \\ &\lesssim \left\|p^k(w)\right\| \mathrm{E}[\delta(t,W)^2] \lesssim \zeta_k,\end{aligned}$$

which implies that for any $f \coloneqq f(\cdot;t) \in \mathcal{F}_k$,

$$|f(z)| = |\delta_k(t,w)|\,|h_k(w) - h^*(w)| \lesssim \zeta_k |h_k(w) - h^*(w)|.$$

The previous diplay shows that $|f(z)| \leqslant F_{2k}(z)$ for $F_{2k}(z) \coloneqq C_2\zeta_k |h_k(w) - h^*(w)|$ and some $C_2 \in (0,\infty)$. Let $C_3 \coloneqq C_1 \vee C_2$, and define $F_k(z) \coloneqq C_3\zeta_k |h_k(w) - h^*(w)|$. Then by Assumption 7,

$$\|F_k\|_{P,2} = C_3\zeta_k\|h_k - h^*\|_{P,2} = C_3\zeta_k r_{h,k} \to 0 \text{ as } k \to \infty.$$

In particular, $\|F_k\|_{P,2} \lesssim 1$. Now, F_k is a square-integrable envelope for $\mathcal{F}_k$ satisfying

$$|f_1(z) - f_2(z)| \leqslant F_k(z)\,\|t_1 - t_2\|.$$

Using $\mathcal{T}$ compact and the previous display, by van der Vaart and Wellner (1996, Theorem 2.7.11) we see that

$$N_{[\,]}(\varepsilon\|F_k\|_{P,2}, \mathcal{F}_k, L^2(\mathrm{P})) \leqslant (C/\varepsilon)^{d_t}, \quad \varepsilon \in (0,1],$$

and thus

$$J_{[\,]}\left(\delta,\mathcal{F}_k,L^2\left(P\right)\right)\leqslant\int_0^\delta\sqrt{1+d_t\ln\left(C/\varepsilon\right)}\mathrm{d}\varepsilon,\quad\delta\in(0,1],$$

where the right-hand side does not depend on k. In particular, $J_{[\,]}\left(1,\mathcal{F}_k,L^2\left(P\right)\right)\lesssim 1$. Using van der Vaart and Wellner (1996, Theorem 2.14.2) $J_{[\,]}\left(1,\mathcal{F}_{k_n},L^2\left(P\right)\right)\lesssim 1$, we arrive at

$$\mathrm{E}\left[\|\mathbb{G}_n\|_{\mathcal{F}_{k_n}}\right]\lesssim J_{[\,]}\left(1,\mathcal{F}_{k_n},L^2\left(P\right)\right)\|F_{k_n}\|_{P,2}\lesssim\|F_{k_n}\|_{P,2}\lesssim\zeta_{k_n}r_{h,k_n},$$

so $\|\mathrm{IV}_{e,n}\|_{\mathcal{T}}\lesssim_{\mathrm{P}}\zeta_{k_n}r_{h,k_n}$ by M. □

S.3 Proofs for Section 3.2

Define the stochastic processes $\widehat{G}^u$ and G_n^{*u} by

$$\widehat{G}^u\left(t\right)\coloneqq\frac{1}{\sqrt{n}}\sum_{i=1}^n\xi_i\widehat{g}\left(t,Z_i\right)\quad\text{and}\quad G_n^{*u}\left(t\right)\coloneqq\frac{1}{\sqrt{n}}\sum_{i=1}^n\xi_i g\left(t,Z_i\right).$$

which are the '*u*ncentered' versions of $\widehat{G}$ and G_n^* defined in (3.18) and (3.15), respectively, i.e., the displayed processes are not centered at the sample mean. The following lemma shows that the uncentered processes are asymptotically equivalent.

Lemma S.1. *If Assumptions 1–8 hold, then* $\max_{1\leqslant\ell\leqslant L}\|\widehat{G}_\ell^u-G_{\ell n}^{*u}\|_{\mathcal{X}_\ell}\to_{\mathrm{P}}0$.

Proof of Lemma S.1. The proof proceeds in a number of steps parallelling the proof of Lemma A.2. It suffices to establish the claimed convergence for given ℓ. We therefore drop the ℓ subscripts throughout, refer to the (ℓth) index set ($\mathcal{X}_\ell$) as $\mathcal{T}$ itself, and use d_t for its dimension.

Step 0 (Main)

For fixed $t \in \mathcal{T}$ a decomposition yields

$$\begin{aligned}
\widehat{G}^u(t) - G_n^{*u}(t) &= \sqrt{n}\mathbb{E}_n\left[\xi_i\{\widehat{g}(t, Z_i) - g(t, Z_i)\}\right] \\
&= \sqrt{n}\mathbb{E}_n\left[\xi_i\omega(t, X_i)\{\rho(Z_i, \widehat{\beta}, \widehat{h}(W_i)) - \rho(Z_i, \beta_0, h^*(W_i))\}\right] \\
&\quad - [\widehat{b}(t) - b(t)]'\sqrt{n}\mathbb{E}_n[\xi_i s(Z_i)] \\
&\quad - \widehat{b}(t)'\sqrt{n}\mathbb{E}_n[\xi_i\{\widehat{s}(Z_i) - s(Z_i)\}] \\
&\quad + \sqrt{n}\mathbb{E}_n\left[\xi_i(\widehat{\delta}(t, W_i)'\{Y_i - \widehat{h}(W_i)\} - \delta(t, W_i)' U_i)\right], \\
&=: \mathrm{I}_n(t) + \mathrm{II}_n(t) + \mathrm{III}_n(t) + \mathrm{IV}_n(t). \qquad \text{(S.12)}
\end{aligned}$$

where $U_i = Y_i - h^*(W_i)$. The following steps show that the four remainder terms $\to_{\mathrm{P}} 0$ uniformly over $\mathcal{T}$. The claim therefore follows from T.

Step 1: $\|\mathrm{I}_n\|_{\mathcal{T}} \to_{\mathrm{P}} 0$

Assumption 1 and M implies that $\widehat{\beta} \to_{\mathrm{P}} \beta_0$ interior (also under H_1), so letting $\mathcal{N}$ be the open neighborhood of β_0 provided by Assumption 3, $\widehat{\beta} \in \mathcal{N}$ wp $\to 1$. To simplify notation and ensure that objects are globally well defined, in what follows we will—without loss of generality—assume that $\widehat{\beta} \in \mathcal{N}$ with probability one for all n. Assumption 3, a MVE, J and CS followed by T then imply

$$\begin{aligned}
\|\mathrm{I}_n\|_{\mathcal{T}} &\leqslant \|\widehat{\beta} - \beta_0\| \sup_{t \in \mathcal{T}} \left\| \sqrt{n}\mathbb{E}_n\left[\xi_i\omega(t, X_i)\frac{\partial}{\partial\beta}\rho(Z_i, \beta_0, h^*(W_i))\right]\right\| \\
&\quad + \sqrt{n}\|\widehat{\beta} - \beta_0\| \sup_{t \in \mathcal{T}} \left\|\mathbb{E}_n\left[\xi_i\omega(t, X_i)\left\{\frac{\partial}{\partial\beta}\rho(Z_i, \overline{\beta}, \widehat{h}(W_i)) - \frac{\partial}{\partial\beta}\rho(Z_i, \beta_0, h^*(W_i))\right\}\right]\right\| \\
&=: \mathrm{I}_{a,n} + \mathrm{I}_{b,n},
\end{aligned}$$

where $\overline{\beta}$ satisfies $\|\overline{\beta}-\beta_0\| \leqslant \|\widehat{\beta}-\beta_0\|$ such that also $\overline{\beta} \in \mathcal{N}$. Per Assumptions 2 and 8.2, J and M,

$$\begin{aligned}
\mathrm{I}_{b,n} &\leqslant \sqrt{n}\|\widehat{\beta}-\beta_0\|\mathbb{E}_n\left[|\xi_i| \sup_{t\in\mathcal{T}}|\omega(t,X_i)|\left\|\frac{\partial}{\partial\beta}\rho(Z_i,\overline{\beta},\widehat{h}(W_i))-\frac{\partial}{\partial\beta}\rho(Z_i,\beta_0,h^*(W_i))\right\|\right] \\
&\lesssim \sqrt{n}\|\widehat{\beta}-\beta_0\|\mathbb{E}_n\left[|\xi_i|\, a'(Z_i)\left(\|\overline{\beta}-\beta_0\|+\|\widehat{h}(W_i)-h^*(W_i)\|\right)\right] \\
&\leqslant \sqrt{n}\|\widehat{\beta}-\beta_0\|\mathbb{E}_n\left[|\xi_i|\, a'(Z_i)\left(\|\widehat{\beta}-\beta_0\|+\sqrt{d}\max_{1\leqslant m\leqslant d}\|\widehat{h}_m-h_m^*\|_{\mathcal{W}}\right)\right] \\
&\lesssim_{\mathrm{P}} \mathrm{E}\left[a'(Z)\right]\sqrt{n}\|\widehat{\beta}-\beta_0\|\left(\|\widehat{\beta}-\beta_0\| \vee \max_{1\leqslant m\leqslant d}\|\widehat{h}_m-h_m^*\|_{\mathcal{W}}\right),
\end{aligned}$$

which $\to_{\mathrm{P}} 0$ also by Assumption 8.2. By now familiar arguments (see the proof of Lemma 1), each function class

$$\mathcal{F}_j' := \left\{(v,z)\mapsto v\omega(t,x)\frac{\partial}{\partial\beta_j}\rho(z,\beta_0,h^*(w))\,; t\in\mathcal{T}\right\}$$

may be proven Donsker. Weak convergence of $\sqrt{n}\mathbb{E}_n[\xi_i\omega(\cdot,X_i)(\partial/\partial\beta_j)\rho(Z_i,\beta_0,h^*(W_i))]$ in $L^\infty(\mathcal{T})$ follows and, therefore,

$$\begin{aligned}
&\sup_{t\in\mathcal{T}}\left\|\sqrt{n}\mathbb{E}_n\left[\xi_i\omega(t,X_i)\frac{\partial}{\partial\beta}\rho(Z_i,\beta_0,h^*(W_i))\right]\right\| \\
&\lesssim \max_{1\leqslant j\leqslant d_\beta}\sup_{t\in\mathcal{T}}\left|\sqrt{n}\mathbb{E}_n\left[\xi_i\omega(t,X_i)\frac{\partial}{\partial\beta_j}\rho(Z_i,\beta_0,h^*(W_i))\right]\right| \lesssim_{\mathrm{P}} 1.
\end{aligned}$$

It then follows from $\widehat{\beta}\to_{\mathrm{P}}\beta_0$ that also $\mathrm{I}_{a,n}\to_{\mathrm{P}} 0$, and, thus, $\|\mathrm{I}_n\|_{\mathcal{T}}\to_{\mathrm{P}} 0$.

Step 2: $\|\mathrm{II}_n\|_{\mathcal{T}}\to_{\mathrm{P}} 0$

By CS, II_n defined in (S.12) satisfies

$$\|\mathrm{II}_n\|_{\mathcal{T}} \leqslant \left\|\sqrt{n}\mathbb{E}_n\left[\xi_i s(Z_i)\right]\right\|\sup_{t\in\mathcal{T}}\left\|\widehat{b}(t)-b(t)\right\|,$$

To show $\|\mathrm{II}_n\|_{\mathcal{T}}\to_{\mathrm{P}} 0$, it therefore suffices to show $\|\sqrt{n}\mathbb{E}_n\left[\xi_i s(Z_i)\right]\| \lesssim_{\mathrm{P}} 1$ and $\sup_{t\in\mathcal{T}}\|\widehat{b}(t)-b(t)\|\to_{\mathrm{P}} 0$.

Step 2a: $\|\sqrt{n}\mathbb{E}_n[\xi_i s(Z_i)]\| \lesssim_{\mathrm{P}} 1$**.** Given that the ξ_i's are i.i.d., zero-mean, unit variance and independent of the data we have

$$\mathrm{E}\left[\left\|\sqrt{n}\mathbb{E}_n[\xi_i s(Z_i)]\right\|^2 \middle| \{Z_i\}_1^n\right] = \mathbb{E}_n\left[\|s(Z_i)\|^2\right].$$

The desired $\|\sqrt{n}\mathbb{E}_n[\xi_i s(Z_i)]\| \lesssim_{\mathrm{P}} 1$ now follows from iterated expectations, integrability of $\|s(Z)\|^2$ (Assumption 1) and M.

Step 2b: Behavior of $\widehat{b}$. In this step we show that

$$\text{(a)} \sup_{t\in\mathcal{T}}\|\widehat{b}(t) - b(t)\| \xrightarrow{\mathrm{P}} 0 \quad \text{and} \quad \text{(b)} \sup_{t\in\mathcal{T}}\|\widehat{b}(t)\| \lesssim_{\mathrm{P}} 1,$$

with b and $\widehat{b}$ defined in (3.7) and (3.21), respectively. To show (a), note that the argument used in Step 1 of the proof of Lemma A.2 shows that

$$\begin{aligned}
&(t,\beta) \mapsto \mathrm{E}[\omega(t,X)(\partial/\partial\beta)\rho(Z,\beta,h^*(W))] \text{ is uniformly continuous on } \mathcal{T}\times\overline{B},\\
&\text{and } \sup_{\mathcal{T}\times\overline{B}}\|(\mathbb{E}_n - \mathrm{E})\omega(t,X_i)(\partial/\partial\beta)\rho(Z_i,\beta,h^*(W_i))\| \xrightarrow{\mathrm{P}} 0,
\end{aligned}$$

where $\overline{B}\subset\mathcal{N}$ is any closed set containing β_0 in its interior (Assumption 1). By T we have

$$\begin{aligned}
\sup_{t\in\mathcal{T}}\|\widehat{b}(t) - b(t)\| \leqslant{}& \sup_{t\in\mathcal{T}}\left\|\mathbb{E}_n\left[\omega(t,X_i)\left\{\frac{\partial}{\partial\beta}\rho(Z_i,\widehat{\beta},\widehat{h}(W_i)) - \frac{\partial}{\partial\beta}\rho(Z_i,\widehat{\beta},h^*(W_i))\right\}\right]\right\|\\
&+ \sup_{t\in\mathcal{T}}\left\|(\mathbb{E}_n - \mathrm{E}_Z)\left[\omega(t,X_i)\frac{\partial}{\partial\beta}\rho(Z_i,\widehat{\beta},h^*(W_i))\right]\right\|\\
&+ \sup_{t\in\mathcal{T}}\left\|\mathrm{E}_Z\left[\omega(t,X)\frac{\partial}{\partial\beta}\rho(Z,\widehat{\beta},h^*(W))\right] - b(t)\right\|.
\end{aligned}$$

Given that $\widehat{\beta}\in\overline{B}$ wp $\to 1$, the second and third term on the right $\to_{\mathrm{P}} 0$ due to uniform convergence and uniform continuity, respectively. By T and Assumptions 2 and 3, the first term is bounded by a constant multiple of

$$\begin{aligned}
\mathbb{E}_n[a(Z_i)\|\widehat{h}(Z_i) - h^*(Z_i)\|^c] &\leqslant d^{c/2}\mathbb{E}_n[a(Z_i)]\max_{1\leqslant m\leqslant d}\|\widehat{h}_m - h_m^*\|_{\mathcal{W}}^c\\
&\lesssim_{\mathrm{P}} \max_{1\leqslant m\leqslant d}\|\widehat{h}_m - h_m^*\|_{\mathcal{W}}^c \xrightarrow{\mathrm{P}} 0,
\end{aligned}$$

where the $\lesssim_{\mathrm{P}}$ follows from M and the $\to_{\mathrm{P}} 0$ from Lemma S.7. The previous display finishes the proof of (a) and therefore the proof of Step 2 ($\|\mathrm{II}_n\|_{\mathcal{T}} \to_{\mathrm{P}} 0$).

To show (b), note that the argument used in Step 1 of the proof of Lemma A.2 also shows that $\sup_{t\in\mathcal{T}}\|b(t)\| \lesssim 1$. Two applications of T yield

$$\left|\sup_{t\in\mathcal{T}}\|\widehat{b}(t)\| - \sup_{t\in\mathcal{T}}\|b(t)\|\right| \leqslant \sup_{t\in\mathcal{T}}\left|\|\widehat{b}(t)\| - \|b(t)\|\right| \leqslant \sup_{t\in\mathcal{T}}\|\widehat{b}(t) - b(t)\| \xrightarrow{\mathrm{P}} 0,$$

which combined with $\sup_{t\in\mathcal{T}}\|b(t)\| \lesssim 1$ implies $\sup_{t\in\mathcal{T}}\|\widehat{b}(t)\| \lesssim_{\mathrm{P}} 1$.

Step 3: $\|\mathrm{III}_n\|_{\mathcal{T}} \to_{\mathrm{P}} 0$

By CS, III_n defined in (S.12) satisfies

$$\|\mathrm{III}_n\|_{\mathcal{T}} \leqslant \left\|\sqrt{n}\mathbb{E}_n\left[\xi_i\{\widehat{s}(Z_i) - s(Z_i)\}\right]\right\| \sup_{t\in\mathcal{T}}\left\|\widehat{b}(t)\right\|. \tag{S.13}$$

By Step 2b, $\sup_{t\in\mathcal{T}}\|\widehat{b}(t)\| \lesssim_{\mathrm{P}} 1$, so to show $\|\mathrm{III}_n\|_{\mathcal{T}} \to_{\mathrm{P}} 0$, it suffices to show that

$$\left\|\sqrt{n}\mathbb{E}_n\left[\xi_i\{\widehat{s}(Z_i) - s(Z_i)\}\right]\right\| \xrightarrow{\mathrm{P}} 0.$$

To this end, note that by the ξ_i's being i.i.d., zero-mean, unit variance and independent of the data, and Assumption 8, we have

$$\mathrm{E}\left[\left\|\sqrt{n}\mathbb{E}_n\left[\xi_i\{\widehat{s}(Z_i) - s(Z_i)\}\right]\right\|^2 \middle| \{Z_i\}_1^n\right] = \mathbb{E}_n\left[\|\widehat{s}(Z_i) - s(Z_i)\|^2\right] \xrightarrow{\mathrm{P}} 0,$$

so $\|\sqrt{n}\mathbb{E}_n\left[\xi_i\{\widehat{s}(Z_i) - s(Z_i)\}\right]\| \to_{\mathrm{P}} 0$ follows from Lemma S.9.

Step 4: $\|\mathrm{IV}_n\|_{\mathcal{T}} \to_{\mathrm{P}} 0$

Given that IV_n defined in (S.12) may be written as the sum

$$\mathrm{IV}_n(t) = \sum_{m=1}^{d} \sqrt{n}\mathbb{E}_n\left[\xi_i\{\widehat{\delta}_m(t, W_i)\{Y_{mi} - \widehat{h}_m(W_i)\} - \delta_m(t, W_i)U_{mi}\}\right],$$

it suffices to bound each summand uniformly over $\mathcal{T}$ in probability. We therefore omit also the m subscript for the remainder of Step 4 and interpret $(\partial/\partial h)\rho$ as a scalar derivative. By T, the (mth) summand satisfies the uniform bound

$$\begin{aligned}
&\sup_{t\in\mathcal{T}}\left|\sqrt{n}\mathbb{E}_n\left[\xi_i\{\widehat{\delta}(t,W_i)\{Y_i-\widehat{h}(W_i)\}-\delta(t,W_i)U_i\}\right]\right|\\
&\leqslant\sup_{t\in\mathcal{T}}\left|\sqrt{n}\mathbb{E}_n\left[\xi_iU_i\{\widehat{\delta}(t,W_i)-\delta(t,W_i)\}\right]\right|\\
&\quad+\sup_{t\in\mathcal{T}}\left|\sqrt{n}\mathbb{E}_n\left[\xi_i\widehat{\delta}(t,W_i)\{\widehat{h}(W_i)-h^*(W_i)\}\right]\right|=:\|\mathrm{IV}_{a,n}\|_{\mathcal{T}}+\|\mathrm{IV}_{b,n}\|_{\mathcal{T}}. \qquad (S.14)
\end{aligned}$$

We consider each term on the right-hand side in turn.

Step 4a: $\|\mathrm{IV}_{a,n}\|_{\mathcal{T}}\to_{\mathrm{P}}0$**.** Recalling the definitions of δ_k and ψ_k in (S.5) and (S.9), respectively, we may write $\delta_k(t,w)=p^k(w)'Q_k^{-1}\psi_k(t)$, such that by T,

$$\begin{aligned}
\|\mathrm{IV}_{a,n}\|_{\mathcal{T}}&=\sup_{t\in\mathcal{T}}\left|\sqrt{n}\mathbb{E}_n\left[\xi_iU_ip^{k_n}(W_i)'\widehat{Q}_{k_n}^{-1}\widehat{\psi}_{k_n}(t)\right]\right.\\
&\leqslant\sup_{t\in\mathcal{T}}\left|\widehat{\psi}_{k_n}(t)'(\widehat{Q}_{k_n}^{-1}-Q_{k_n}^{-1})\sqrt{n}\mathbb{E}_n\left[p^{k_n}(W_i)\xi_iU_i\right]\right|\\
&\quad+\sup_{t\in\mathcal{T}}\left|\{\widehat{\psi}_{k_n}(t)-\psi_{k_n}(t)\}'Q_{k_n}^{-1}\sqrt{n}\mathbb{E}_n\left[p^{k_n}(W_i)\xi_iU_i\right]\right|\\
&\quad+\sup_{t\in\mathcal{T}}\left|\sqrt{n}\mathbb{E}_n\left[\xi_iU_i\{\delta_{k_n}(t,W_i)-\delta(t,W_i)\}\right]\right|\\
&=:\|\mathrm{IV}_{a,1,n}\|_{\mathcal{T}}+\|\mathrm{IV}_{a,2,n}\|_{\mathcal{T}}+\|\mathrm{IV}_{a,3,n}\|_{\mathcal{T}}, \qquad (S.15)
\end{aligned}$$

where we employ the convenient shorthand

$$\widehat{\psi}_k(t)=\mathbb{E}_n\left[p^k(W_i)\,\omega(t,X_i)\frac{\partial}{\partial\beta}\rho(Z_i,\widehat{\beta},\widehat{h}(W_i))\right]$$

as defined in the (auxilliary) Step 4c below.

Step 4a(1): $\|\mathrm{IV}_{a,1,n}\|_{\mathcal{T}}\to_{\mathrm{P}}0$**.** Using Assumptions 4 and 5 and the ξ_i's being i.i.d., zero-mean, unit variance and independent of the data,

$$\begin{aligned}
\mathrm{E}\left[\left\|Q_{k_n}^{-1/2}\sqrt{n}\mathbb{E}_n\left[p^{k_n}(W_i)\xi_iU_i\right]\right\|^2\right]&=\mathrm{E}[\xi^2U^2p^{k_n}(W)'Q_{k_n}^{-1}p^{k_n}(W)]\\
&=\mathrm{E}[U^2p^{k_n}(W)'Q_{k_n}^{-1}p^{k_n}(W)]\\
&\lesssim\mathrm{E}[p^{k_n}(W)'Q_{k_n}^{-1}p^{k_n}(W)]=k_n,
\end{aligned}$$

so by M we have

$$\left\| Q_{k_n}^{-1/2} \sqrt{n} \mathbb{E}_n \left[p^{k_n} (W_i) \xi_i U_i \right] \right\| \lesssim_{\mathrm{P}} \sqrt{k_n}. \tag{S.16}$$

Step 4c shows that $\sup_{t \in \mathcal{T}} \| \widehat{\psi}_{k_n} (t)' \widehat{Q}_{k_n}^{-1} \| \lesssim_{\mathrm{P}} 1$, so by CS, Assumption 5, Lemma S.6, and the previous display,

$$\begin{aligned}
\| \mathrm{IV}_{a,1,n} \|_{\mathcal{T}} &= \sup_{t \in \mathcal{T}} \left| \widehat{\psi}_{k_n} (t)' \widehat{Q}_{k_n}^{-1} (Q_{k_n} - \widehat{Q}_{k_n}) Q_{k_n}^{-1} \sqrt{n} \mathbb{E}_n \left[p^{k_n} (W_i) \xi_i U_i \right] \right| \\
&\leqslant \left\| Q_{k_n}^{-1} \sqrt{n} \mathbb{E}_n \left[p^{k_n} (W_i) \xi_i U_i \right] \right\| \sup_{t \in \mathcal{T}} \left\| \widehat{\psi}_{k_n} (t)' \widehat{Q}_{k_n}^{-1} (Q_{k_n} - \widehat{Q}_{k_n}) \right\| \\
&\leqslant \left\| Q_{k_n}^{-1} \sqrt{n} \mathbb{E}_n \left[p^{k_n} (W_i) \xi_i U_i \right] \right\| \| \widehat{Q}_{k_n} - Q_{k_n} \|_{\mathrm{op}} \sup_{t \in \mathcal{T}} \left\| \widehat{\psi}_{k_n} (t)' \widehat{Q}_{k_n}^{-1} \right\| \\
&\lesssim \left\| Q_{k_n}^{-1/2} \sqrt{n} \mathbb{E}_n \left[p^{k_n} (W_i) \xi_i U_i \right] \right\| \| \widehat{Q}_{k_n} - Q_{k_n} \|_{\mathrm{op}} \sup_{t \in \mathcal{T}} \left\| \widehat{\psi}_{k_n} (t)' \widehat{Q}_{k_n}^{-1} \right\| \\
&\lesssim_{\mathrm{P}} \sqrt{k_n} [\zeta_{k_n}^2 \ln (k_n) / n]^{1/2} = [\zeta_{k_n}^2 k_n \ln (k_n) / n]^{1/2},
\end{aligned}$$

which $\to 0$ by Assumption 7.

Step 4a(2): $\| \mathrm{IV}_{a,2,n} \|_{\mathcal{T}} \to_{\mathrm{P}} 0$. By CS, Assumption 5, (S.16) and Step 4c,

$$\begin{aligned}
\| \mathrm{IV}_{a,2,n} \|_{\mathcal{T}} &\leqslant \left\| Q_{k_n}^{-1} \sqrt{n} \mathbb{E}_n \left[p^{k_n} (W_i) \xi_i U_i \right] \right\| \sup_{t \in \mathcal{T}} \left\| \widehat{\psi}_{k_n} (t) - \psi_{k_n} (t) \right\| \\
&\lesssim \left\| Q_{k_n}^{-1/2} \sqrt{n} \mathbb{E}_n \left[p^{k_n} (W_i) \xi_i U_i \right] \right\| \sup_{t \in \mathcal{T}} \left\| \widehat{\psi}_{k_n} (t) - \psi_{k_n} (t) \right\| \\
&\lesssim_{\mathrm{P}} \sqrt{k_n} \Bigg[\left\{ \mathrm{E} \left[R' (Z_i)^2 \right] \right\}^{1/2} \zeta_{k_n} \left(\| \widehat{\beta} - \beta_0 \| \vee \max_{1 \leqslant m' \leqslant d} \| \widehat{h}_{m'} - h_{m'}^* \|_{n,2} \right) \\
&\quad + \left(\sum_{j=1}^{k_n} \| p_j \|_{\mathcal{W}}^2 \right)^{1/2} / \sqrt{n} \Bigg] \\
&= \left\{ \mathrm{E} \left[R' (Z_i)^2 \right] \right\}^{1/2} \zeta_{k_n} \sqrt{k_n} \left(\| \widehat{\beta} - \beta_0 \| \vee \max_{1 \leqslant m' \leqslant d} \| \widehat{h}_{m'} - h_{m'}^* \|_{n,2} \right) \\
&\quad + \left(\sum_{j=1}^{k_n} \| p_j \|_{\mathcal{W}}^2 \right)^{1/2} \sqrt{\frac{k_n}{n}},
\end{aligned}$$

which $\to_{\mathrm{P}} 0$ by Assumptions 7 and 8.4.

Step 4a(3): $\|\mathrm{IV}_{a,3,n}\|_{\mathcal{T}} \to_{\mathrm{P}} 0$. Fix k and let

$$\mathcal{F}_k' \coloneqq \{(v,z) \mapsto v\{y - h^*(w)\}\{\delta_k(t,w) - \delta(t,w)\}; t \in \mathcal{T}\}.$$

Given that each $\mathrm{E}[f(\xi, Z)] = 0$ for each $f \in \mathcal{F}_k'$, the stochastic process IV_n may be viewed as an empirical process $\mathbb{G}_n$ indexed by the changing classes $\mathcal{F}_{k_n}'$. For $f = f_t, f_1 = f_{t_1}, f_2 = f_{t_2} \in \mathcal{F}_{k_n}'$ arbitrary, by arguments parallel to those used in Step 4c in the proof of Lemma A.2, there exists a function $z \mapsto F_k(z)$ such that

$$|f(v,z)| \leqslant |v| F_k(z),$$
$$|f_1(v,z) - f_2(v,z)| \leqslant |v| F_k(z) \|t_1 - t_2\|,$$

and $\|F_k\|_{P,2} \lesssim \sqrt{k}$. The ξ_i's being zero mean, unit variance and independent of the data implies that $F_k' : (v,z) \mapsto |v| F_k(z)$ is an envelope for $\mathcal{F}_k'$ with $(\mathrm{E}[F_k'(\xi,Z)^2])^{1/2} = \|F_k\|_{P,2} \lesssim \sqrt{k}$ as $k \to \infty$, satisfying

$$|f_1(s,z) - f_2(s,z)| \leqslant F_k'(s,z) \|t_1 - t_2\|.$$

Using $\mathcal{T}$ compact and the previous display, by van der Vaart and Wellner (1996, Theorem 2.7.11) we see that

$$N_{[\,]}(\varepsilon(\mathrm{E}[F_k'(\xi,Z)^2])^{1/2}, \mathcal{F}_k', L^2(\xi,Z)) \leqslant (C/\varepsilon)^{d_t}, \quad \varepsilon \in (0,1].$$

and thus

$$J_{[\,]}\left(\delta, \mathcal{F}_k', L^2(\xi,Z)\right) \leqslant \int_0^\delta \sqrt{1 + d_t \ln(C/\varepsilon)} \mathrm{d}\varepsilon, \quad \delta \in (0,1].$$

where the right-hand side does not depend on k. In particular, $J_{[\,]}\left(1, \mathcal{F}_{k_n}', L^2(\xi,Z)\right) \lesssim 1$. Defining

$$\sigma_n^2 \coloneqq \sup_{f \in \mathcal{F}_{k_n}'} \mathbb{E}_n[f(\xi_i, Z_i)^2]$$

we see that

$$\sigma_n^2 = \sup_{t \in \mathcal{T}} \mathbb{E}_n\left[\xi_i^2 U_i^2 \{\delta_{k_n}(t, W_i) - \delta(t, W_i)\}^2\right] \leqslant \mathbb{E}_n\left[\xi_i^2 U_i^2 \|\delta_{k_n}(\cdot, W_i) - \delta(\cdot, W_i)\|_{\mathcal{T}}^2\right],$$

thus implying

$$\mathrm{E}\left[\sigma_n^2\right] \leqslant \mathrm{E}\left[\xi^2 U^2 \|\delta_{k_n}(\cdot, W) - \delta(\cdot, W)\|_{\mathcal{T}}^2\right] \leqslant C\mathrm{E}\left[\|\delta_{k_n}(\cdot, W) - \delta(\cdot, W)\|_{\mathcal{T}}^2\right] = CR_{\delta,k_n}^2,$$

where the second inequality follows from the ξ_i's being zero mean, unit variance, and independent of the data and Assumption 4, and the last equality follows from the definitions of δ_k and $R_{\delta,k}$ [the latter in (3.13)].

It suffices to consider the two cases (1) $R_{\delta,k_n}/\|F_{k_n}\|_{P,2} \to 0$ and (2) $R_{\delta,k_n}/\|F_{k_n}\|_{P,2} \nrightarrow 0$ in turn. *Case 1*: $R_{\delta,k_n}/\|F_{k_n}\|_{P,2} \to 0$. Given that $\sqrt{\mathrm{E}\left[\sigma_n^2\right]} \leqslant CR_{\delta,k_n}$, by the change of variables $\varepsilon' \coloneqq \varepsilon/C$ we have

$$\begin{aligned} J_{[\,]}\left(\sqrt{\mathrm{E}\left[\sigma_n^2\right]}/\|F_{k_n}\|_{P,2}, \mathcal{F}_k', L^2(\xi, Z)\right) &\leqslant J_{[\,]}\left(CR_{\delta,k_n}/\|F_{k_n}\|_{P,2}, \mathcal{F}_k', L^2(\xi, Z)\right) \\ &= C\int_0^{R_{\delta,k_n}/\|F_{k_n}\|_{P,2}} \sqrt{1 + d_t \ln(C'/\varepsilon')}\mathrm{d}\varepsilon' \\ &=: C\overline{J}_{[\,]}\left(R_{\delta,k_n}/\|F_{k_n}\|_{P,2}\right). \end{aligned} \tag{S.17}$$

By van der Vaart and Wellner (2011, p. 196) we have the maximal inequality

$$\begin{aligned} \mathrm{E}[\|\mathbb{G}_n\|_{\mathcal{F}_{k_n}'}] &\lesssim J_{[\,]}\left(\frac{\sqrt{\mathrm{E}\left[\sigma_n^2\right]}}{\|F_{k_n}\|_{P,2}}, \mathcal{F}_{k_n}', L^2(\xi, Z)\right)\|F_{k_n}\|_{P,2} \\ &\lesssim \overline{J}_{[\,]}\left(\frac{R_{\delta,k_n}}{\|F_{k_n}\|_{P,2}}\right)\|F_{k_n}\|_{P,2}, \end{aligned}$$

and from van der Vaart and Wellner (1996, p. 239) we know that an entropy integral (bound) of the form (S.17) satisfies $\overline{J}_{[\,]}(\delta) \lesssim \delta\sqrt{\ln(1/\delta)}$ as $\delta \to 0_+$. Since $R_{\delta,k_n}/\|F_{k_n}\|_{P,2} \to 0$ holds by hypothesis, the previous display combined with $\|F_{k_n}\|_{P,2} \lesssim \sqrt{k_n}$ yields

$$\begin{aligned} \mathrm{E}[\|\mathbb{G}_n\|_{\mathcal{F}_{k_n}'}] &\lesssim \frac{R_{\delta,k_n}}{\|F_{k_n}\|_{P,2}}\sqrt{\ln\left(\frac{\|F_{k_n}\|_{P,2}}{R_{\delta,k_n}}\right)}\|F_{k_n}\|_{P,2} = R_{\delta,k_n}\sqrt{\ln\left(\frac{\|F_{k_n}\|_{P,2}}{R_{\delta,k_n}}\right)} \\ &\lesssim R_{\delta,k_n}\sqrt{\ln(k_n/R_{\delta,k_n})}. \end{aligned}$$

Case 2. Suppose that $R_{\delta,k_n}/\|F_{k_n}\|_{P,2} \nrightarrow 0$. Given that $R_{\delta,k_n} \to 0$ (Assumption 7), we must have $\|F_{k_n}\|_{P,2} \lesssim R_{\delta,k_n}$. van der Vaart and Wellner (1996, Theorem 2.14.2) and $J_{[\,]}\left(1, \mathcal{F}'_{k_n}, L^2(\xi, Z)\right) \lesssim 1$ yield

$$\begin{aligned}\mathrm{E}[\|\mathbb{G}_n\|_{\mathcal{F}'_{k_n}}] &\lesssim J_{[\,]}\left(1, \mathcal{F}_{k'_n}, L^2(\xi, Z)\right) \|F_{k_n}\|_{P,2} \\ &\lesssim \|F_{k_n}\|_{P,2} \lesssim R_{\delta,k_n} \lesssim R_{\delta,k_n}\sqrt{\ln(k_n/R_{\delta,k_n})}\end{aligned}$$

as in Case 1. The claim $\|\mathrm{IV}_{a,3,n}\|_{\mathcal{T}}$ as defined in (S.15) $\to_{\mathrm{P}} 0$ now follows from M and $R_{\delta,k_n}\sqrt{\ln(k_n/R_{\delta,k_n})} \to 0$ (Assumption 7). Via (S.15), this $\to_{\mathrm{P}} 0$ in turn shows that $\|\mathrm{IV}_{a,n}\|_{\mathcal{T}}$ as defined in (S.14) $\to_{\mathrm{P}} 0$.

Step 4b: $\|\mathrm{IV}_{b,n}\|_{\mathcal{T}} \to_{\mathrm{P}} 0$. Employ the shorthand $\widehat{\psi}_k(t)$ defined in (S.18) below, such that $\widehat{\delta}_k(t, W_i) = p^k(W_i)'\widehat{Q}_k^{-1}\widehat{\psi}_k(t)$. Step 4c shows that $\sup_{\mathcal{T}}\|\widehat{\psi}_{k_n}(t)'\widehat{Q}_{k_n}^{-1}\| \lesssim_{\mathrm{P}} 1$, so by CS it follows that

$$\begin{aligned}\|\mathrm{IV}_{b,n}\|_{\mathcal{T}} &= \sup_{t\in\mathcal{T}}\left|\widehat{\psi}_{k_n}(t)'\widehat{Q}_{k_n}^{-1}\sqrt{n}\mathbb{E}_n\left[p^{k_n}(W_i)\xi_i\{\widehat{h}(W_i) - h^*(W_i)\}\right]\right| \\ &\leqslant \left\|\sqrt{n}\mathbb{E}_n\left[p^{k_n}(W_i)\xi_i\{\widehat{h}(W_i) - h^*(W_i)\}\right]\right\| \sup_{t\in\mathcal{T}}\left\|\widehat{\psi}_{k_n}(t)'\widehat{Q}_{k_n}^{-1}\right\| \\ &\lesssim_{\mathrm{P}} \left\|\sqrt{n}\mathbb{E}_n\left[p^{k_n}(W_i)\xi_i\{\widehat{h}(W_i) - h^*(W_i)\}\right]\right\|.\end{aligned}$$

To show that the right-hand side $\to_{\mathrm{P}} 0$, note that by the ξ_i's being i.i.d., zero-mean, unit variance and independent of $\{Z_i\}_1^n$, and $\widehat{h}$ being $\{Z_i\}_1^n$-measurable,

$$\begin{aligned}&\mathrm{E}\left[\left\|\sqrt{n}\mathbb{E}_n\left[p^{k_n}(W_i)\xi_i\{\widehat{h}(W_i) - h^*(W_i)\}\right]\right\|^2 \middle| \{Z_i\}_1^n\right] \\ &= \mathbb{E}_n\left[\|p^{k_n}(W_i)\|^2|\widehat{h}(W_i) - h^*(W_i)|^2\right] \leqslant \left(\sum_{j=1}^{k_n}\|p_j\|_{\mathcal{W}}^2\right)\|\widehat{h} - h^*\|_{n,2}^2 \\ &\lesssim_{\mathrm{P}} \left[\left(\sum_{j=1}^{k_n}\|p_j\|_{\mathcal{W}}^2\right)^{1/2}\left(\sqrt{k_n/n} + k_n^{-\alpha}\right)\right]^2 \to 0,\end{aligned}$$

where the $\lesssim_{\mathrm{P}}$ follows from Lemma S.7 and the $\to 0$ from Assumption 7. Lemma S.9 then implies $\|\sqrt{n}\mathbb{E}_n[p^{k_n}(W_i)\xi_i\{\widehat{h}(W_i) - h^*(W_i)\}]\| \to_{\mathrm{P}} 0$. This $\to_{\mathrm{P}} 0$ finishes the proof of the claim that $\|\mathrm{IV}_{b,n}\|_{\mathcal{T}}$ as defined in (S.14) $\to_{\mathrm{P}} 0$, which in turn shows that $\|\mathrm{IV}_n\|_{\mathcal{T}}$ as defined in (S.12) $\to_{\mathrm{P}} 0$.

Step 4c (auxilliary): Behavior of $\widehat{\psi}_{k_n}$ and $\widehat{Q}^{-}_{k_n}$. Motivated by the LOIE, we estimate $\psi_k(t) = \mathrm{E}[p^k(W)\,\delta(t,W)]$ as defined in (S.9) by

$$\widehat{\psi}_k(t) \coloneqq \mathbb{E}_n\left[p^k(W_i)\,\omega(t,X_i)\,\frac{\partial}{\partial h}\rho(Z_i,\widehat{\beta},\widehat{h}(W_i))\right]. \tag{S.18}$$

Note that this definition allows us to write $\widehat{\delta}$ defined in (3.22) as

$$(t,w) \mapsto \widehat{\delta}(t,w) = p^{k_n}(w)'\,\widehat{Q}^{-}_{k_n}\widehat{\psi}_{k_n}(t).$$

This section shows that

$$\begin{aligned}
&\text{(a)}\ \sup_{t\in\mathcal{T}}\left\|\widehat{\psi}_{k_n}(t) - \psi_{k_n}(t)\right\| \\
&\quad\lesssim_{\mathrm{P}} \left\{\mathrm{E}\left[R'(Z_i)^2\right]\right\}^{1/2}\zeta_{k_n}\left(\|\widehat{\beta}-\beta_0\| \vee \max_{1\leqslant m'\leqslant d}\|\widehat{h}_{m'} - h^*_{m'}\|_{n,2}\right) \\
&\qquad + \frac{1}{\sqrt{n}}\left(\sum_{j=1}^{k_n}\|p_j\|^2_{\mathcal{W}}\right)^{1/2} \to 0, \\
&\text{(b)}\ \sup_{t\in\mathcal{T}}\left\|\widehat{\psi}_{k_n}(t)'\,\widehat{Q}^{-}_{k_n} - \psi_{k_n}(t)'\,Q^{-1}_{k_n}\right\| \xrightarrow{\mathrm{P}} 0, \\
\text{and}\quad &\text{(c)}\ \sup_{t\in\mathcal{T}}\left\|\widehat{\psi}_{k_n}(t)'\,\widehat{Q}^{-}_{k_n}\right\| \lesssim_{\mathrm{P}} 1.
\end{aligned}$$

To show (a), recall $\Delta(t,z,h)$ from (S.7) and define

$$\Delta^k_i(t) \coloneqq (\Delta(t,Z_i,p_1),\ldots,\Delta(t,Z_i,p_k))'.$$

Then by T we have

$$\begin{aligned}
&\sup_{t\in\mathcal{T}}\left\|\widehat{\psi}_{k_n}(t) - \psi_{k_n}(t)\right\| \\
&\leqslant \sup_{t\in\mathcal{T}}\left\|\mathbb{E}_n\left[\omega(t,X_i)\left\{\frac{\partial}{\partial h}\rho(Z_i,\widehat{\beta},\widehat{h}(W_i)) - \frac{\partial}{\partial h}\rho(Z_i,\beta_0,h^*(W_i))\right\}p^{k_n}(W_i)\right]\right\| \\
&\quad + \sup_{t\in\mathcal{T}}\|(\mathbb{E}_n - \mathrm{E})\,\Delta^{k_n}_i(t)\|.
\end{aligned}$$

By Assumptions 1, 2 and 8 and T followed by CS and M,

$$
\begin{aligned}
&\sup_{t\in\mathcal{T}}\left\|\mathbb{E}_n\left[\omega\left(t,X_i\right)\left\{\frac{\partial}{\partial h}\rho(Z_i,\widehat{\beta},\widehat{h}\left(W_i\right))-\frac{\partial}{\partial h}\rho(Z_i,\beta_0,h^*\left(W_i\right))\right\}p^{k_n}\left(W_i\right)\right]\right\|\\
&\lesssim \mathbb{E}_n\left[\left\|p^{k_n}\left(W_i\right)\right\|R'\left(Z_i\right)\left(\|\widehat{\beta}-\beta_0\|+\|\widehat{h}\left(W_i\right)-h^*\left(W_i\right)\|\right)\right]\\
&\lesssim_{\mathrm{P}}\left\{\mathrm{E}\left[R'\left(Z_i\right)^2\right]\right\}^{1/2}\zeta_{k_n}\left(\|\widehat{\beta}-\beta_0\|\vee\max_{1\leqslant m'\leqslant d}\|\widehat{h}_{m'}-h^*_{m'}\|_{n,2}\right),
\end{aligned}
$$

which $\to_{\mathrm{P}} 0$ by Assumption 8.

Moreover, the argument used in Step 3a of the proof of Lemma A.2 shows that

$$\sup_{t\in\mathcal{T}}\left\|\mathbb{E}_n\left[\Delta_i^{k_n}\left(t\right)\right]\right\|\lesssim_{\mathrm{P}}\left(\sum_{j=1}^{k_n}\|p_j\|^2_{\mathcal{W}}\right)^{1/2}/\sqrt{n}.$$

Lemmas S.3 and S.6 and Assumptions 5 and 7 show that $\widehat{Q}_{k_n}$ is invertible wp $\to 1$ and $\lambda_{\min}(\widehat{Q}_{k_n})^{-1}\lesssim_{\mathrm{P}} 1$. To ease notation we will (without loss of generality) assume that $\widehat{Q}_{k_n}^{-1}$ exists with probability one for all n, such that $\widehat{Q}_{k_n}^{-}=\widehat{Q}_{k_n}^{-1}$. The argument used in Step 4b of the proof of Lemma A.2 shows that $\sup_{\mathcal{T}}\|\psi_{k_n}\left(t\right)^{\top}Q_{k_n}^{-1}\|\lesssim 1$, so by (a) and T,

$$
\begin{aligned}
&\sup_{t\in\mathcal{T}}\left\|\widehat{\psi}_{k_n}\left(t\right)'\widehat{Q}_{k_n}^{-1}-\psi_{k_n}\left(t\right)'Q_{k_n}^{-1}\right\|\\
&\leqslant\sup_{t\in\mathcal{T}}\left\|\{\widehat{\psi}_{k_n}\left(t\right)-\psi_{k_n}\left(t\right)\}'\widehat{Q}_{k_n}^{-1}\right\|+\sup_{t\in\mathcal{T}}\left\|\psi_{k_n}\left(t\right)'\left(\widehat{Q}_{k_n}^{-1}-Q_{k_n}^{-1}\right)\right\|\\
&\leqslant\|\widehat{Q}_{k_n}^{-1}\|_{\mathrm{op}}\sup_{t\in\mathcal{T}}\left\|\widehat{\psi}_{k_n}\left(t\right)-\psi_{k_n}\left(t\right)\right\|+\sup_{t\in\mathcal{T}}\left\|\psi_{k_n}\left(t\right)'Q_{k_n}^{-1}(\widehat{Q}_{k_n}-Q_{k_n})\widehat{Q}_{k_n}^{-1}\right\|\\
&\leqslant\|\widehat{Q}_{k_n}^{-1}\|_{\mathrm{op}}\left(\sup_{t\in\mathcal{T}}\left\|\widehat{\psi}_{k_n}\left(t\right)-\psi_{k_n}\left(t\right)\right\|+\|\widehat{Q}_{k_n}-Q_{k_n}\|_{\mathrm{op}}\sup_{t\in\mathcal{T}}\left\|\psi_{k_n}\left(t\right)'Q_{k_n}^{-1}\right\|\right)\xrightarrow{\mathrm{P}}0,
\end{aligned}
$$

which shows (b). Part (c) follows from (b) and $\sup_{t\in\mathcal{T}}\|\psi_{k_n}\left(t\right)'Q_{k_n}^{-1}\|\lesssim 1$. This concludes the proof of the claim that $\|\mathrm{IV}_n\|_{\mathcal{T}}$ as defined in (S.12) $\to_{\mathrm{P}} 0$ and hence the proof of Lemma S.1. □

Lemma S.2. *If Assumptions 1–8 hold, then*

$$\max_{1\leqslant\ell\leqslant L}\|\mathbb{E}_n[\widehat{g}_\ell\left(\cdot,Z_i\right)-g_\ell\left(\cdot,Z_i\right)]\|_{\mathcal{X}_\ell}\xrightarrow{\mathrm{P}}0.$$

Proof of Lemma S.2. The proof proceeds in a number of steps. Since the lemma is stated for a given ℓ, for notational convenience we drop the ℓ subscripts throughout and refer to the (ℓth) index set $\mathcal{X}_\ell$ as $\mathcal{T}$ itself.

Step 0 (Main)

For fixed $t \in \mathcal{T}$ we may write

$$\begin{aligned}\mathbb{E}_n[\widehat{g}(t, Z_i) - g(t, Z_i)] &= \mathbb{E}_n\left[\omega(t, X_i)\{\rho(Z_i, \widehat{\beta}, \widehat{h}(W_i)) - \rho(Z_i, \beta_0, h^*(W_i))\}\right] \\ &\quad - \{\widehat{b}(t) - b(t)\}'\mathbb{E}_n[s(Z_i)] - \widehat{b}(t)'\mathbb{E}_n[\widehat{s}(Z_i) - s(Z_i)] \\ &\quad + \mathbb{E}_n[\widehat{\delta}(t, W_i)'\{Y_i - \widehat{h}(W_i)\} - \delta(t, W_i)'U_i] \\ &=: \mathrm{I}_n(t) + \mathrm{II}_n(t) + \mathrm{III}_n(t) + \mathrm{IV}_n(t).\end{aligned}$$

The following steps show that the four remainder terms $\to_{\mathrm{P}} 0$ uniformly over $\mathcal{T}$. The claim therefore follows from T.

Step 1: $\|\mathrm{I}_n\|_{\mathcal{T}} \to_{\mathrm{P}} 0$

Assumption 1 implies that $\widehat{\beta} \to_{\mathrm{P}} \beta_0$ interior (also under H_1), so letting $\mathcal{N}$ be an open neighborhood of β_0, $\widehat{\beta} \in \mathcal{N}$ wp $\to 1$. To simplify notation and ensure that objects are globally well defined, in what follows we will—without loss of generality—assume that $\widehat{\beta} \in \mathcal{N}$ with probability equal to one for all n. A MVE of $\beta \mapsto \rho(Z_i, \beta, \widehat{h}(W_i))$ at $\widehat{\beta}$ around β_0 followed by CS show that

$$\begin{aligned}\|\mathrm{I}_n\|_{\mathcal{T}} &\leqslant \sup_{t\in\mathcal{T}}|\mathbb{E}_n[\omega(t, X_i)\{\rho(Z_i, \beta_0, \widehat{h}(W_i)) - \rho(Z_i, \beta_0, h^*(W_i))\}]| \\ &\quad + \|\widehat{\beta} - \beta_0\| \sup_{t\in\mathcal{T}}\|\mathbb{E}_n[\omega(t, X_i)(\partial/\partial\beta)\rho(Z_i, \overline{\beta}, \widehat{h}(W_i))]\| \\ &=: \|\mathrm{I}_{a,n}\|_{\mathcal{T}} + \|\widehat{\beta} - \beta_0\|\|\mathrm{I}_{b,n}\|_{\mathcal{T}},\end{aligned}$$

where $\overline{\beta}$ satisfies $\|\overline{\beta} - \beta_0\| \leqslant \|\widehat{\beta} - \beta_0\|$ such that $\overline{\beta} \in \mathcal{N}$. Since $\|\widehat{\beta} - \beta_0\| \to_{\mathrm{P}} 0$ it suffices to show that $\|\mathrm{I}_{a,n}\|_{\mathcal{T}} \to_{\mathrm{P}} 0$ and $\|\mathrm{I}_{b,n}\|_{\mathcal{T}} \lesssim_{\mathrm{P}} 1$. Step 1 in the proof of Lemma A.2 shows that

$$\sup_{t\in\mathcal{T}}\|\mathrm{I}_{b,n}(t) - b(t)\| \xrightarrow{\mathrm{P}} 0 \quad \text{and} \quad \sup_{t\in\mathcal{T}}\|b(t)\| < \infty,$$

which combine to yield $\|\mathrm{I}_{b,n}\|_{\mathcal{T}} \lesssim_{\mathrm{P}} 1$.

Step 1a: $\|\mathrm{I}_{a,n}\|_{\mathcal{T}} \to_{\mathrm{P}} 0$**.** Abbreviate $(z, v) \mapsto \rho(z, \beta_0, v)$ by ρ. By a MVE of $v \mapsto \rho(Z_i, v)$ at $\widehat{h}(W_i)$ around $h^*(W_i)$ and T we may bound $\|\mathrm{I}_{a,n}\|_{\mathcal{T}}$ by

$$\begin{aligned}
&\sup_{t\in\mathcal{T}}\left|\mathbb{E}_n\left[\omega(t, X_i)\left\{\frac{\partial}{\partial h'}\rho(Z_i, \overline{h}(W_i)) - \frac{\partial}{\partial h'}\rho(Z_i, h^*(W_i))\right\}\{\widehat{h}(W_i) - h^*(W_i)\}\right]\right| \\
&\quad + \sup_{t\in\mathcal{T}}\left|\mathbb{E}_n\left[\omega(t, X_i)\frac{\partial}{\partial h'}\rho(Z_i, h^*(W_i))\{\widehat{h}(W_i) - h^*(W_i)\}\right]\right| \\
&=: \|\mathrm{I}_{a,1,n}\|_{\mathcal{T}} + \|\mathrm{I}_{a,2,n}\|_{\mathcal{T}},
\end{aligned}$$

where $\|\overline{h}(W_i) - h^*(W_i)\| \leqslant \|\widehat{h}(W_i) - h^*(W_i)\|$. By T, CS, Assumptions 2 and 8 and M,

$$\|\mathrm{I}_{a,1,n}\|_{\mathcal{T}} \lesssim \mathbb{E}_n[R'(Z_i)\|\widehat{h}(W_i) - h^*(W_i)\|^2] \lesssim_{\mathrm{P}} \mathrm{E}[R'(Z_i)]\max_{1\leqslant m\leqslant d}\|\widehat{h}_m - h^*_m\|^2_{\mathcal{W}} \xrightarrow{\mathrm{P}} 0.$$

Similarly, by T, CS and Assumptions 2 and 3,

$$\begin{aligned}
\|\mathrm{I}_{a,2,n}\|_{\mathcal{T}} &\lesssim \mathbb{E}_n\left[\left|\frac{\partial}{\partial h'}\rho(Z_i, h^*(W_i))\{\widehat{h}(W_i) - h^*(W_i)\}\right|\right] \\
&\lesssim \mathbb{E}_n\left[\left\|\frac{\partial}{\partial h}\rho(Z_i, h^*(W_i))\right\|\right]\max_{1\leqslant m\leqslant d}\|\widehat{h}_m - h^*_m\|_{\mathcal{W}} \\
&\lesssim_{\mathrm{P}} \max_{1\leqslant m\leqslant d}\|\widehat{h}_m - h^*_m\|_{\mathcal{W}} \xrightarrow{\mathrm{P}} 0,
\end{aligned}$$

where the $\lesssim_{\mathrm{P}}$ follows from M and $\|(\partial/\partial h)\,\rho(Z, h^*(W))\|^2$ being integrable (Assumption 3) and the $\to_{\mathrm{P}} 0$ from Lemma S.7.

Step 2: $\|\mathrm{II}_n\|_{\mathcal{T}} \to_{\mathrm{P}} 0$

Step 2b in the proof of Lemma S.1 shows that $\sup_{t\in\mathcal{T}}\|\widehat{b}(t) - b(t)\| \to_{\mathrm{P}} 0$, so by T, CS, Assumption 1, and M

$$\|\mathrm{II}_n\|_{\mathcal{T}} \leqslant \|\mathbb{E}_n[s(Z_i)]\|\sup_{t\in\mathcal{T}}\|\widehat{b}(t) - b(t)\| \lesssim_{\mathrm{P}} \sup_{t\in\mathcal{T}}\|\widehat{b}(t) - b(t)\| \xrightarrow{\mathrm{P}} 0.$$

Step 3: $\|\mathrm{III}_n\|_{\mathcal{T}} \to_{\mathrm{P}} 0$

Step 2b in the proof of Lemma S.1 also shows $\sup_{t\in\mathcal{T}}\|\widehat{b}\,(t)\| \lesssim_{\mathrm{P}} 1$, so by T, CS and Assumption 8,

$$\|\mathrm{III}_n\|_{\mathcal{T}} \leqslant \|\mathbb{E}_n[\widehat{s}\,(Z_i) - s\,(Z_i)]\| \sup_{t\in\mathcal{T}}\|\widehat{b}\,(t)\| \lesssim_{\mathrm{P}} \left\{\mathbb{E}_n\left[\|\widehat{s}\,(Z_i) - s\,(Z_i)\|^2\right]\right\}^{1/2} \xrightarrow{\mathrm{P}} 0.$$

Step 4: $\|\mathrm{IV}_n\|_{\mathcal{T}} \to_{\mathrm{P}} 0$

Given that

$$\begin{aligned}&\mathbb{E}_n[\widehat{\delta}\,(t,W_i)'\,\{Y_i - \widehat{h}\,(W_i)\} - \delta\,(t,W_i)'\,U_i]\\&= \sum_{m=1}^{d}\mathbb{E}_n[\widehat{\delta}_m\,(t,W_i)\,\{Y_{mi} - \widehat{h}_m\,(W_i)\} - \delta_m\,(t,W_i)\,U_{mi}],\end{aligned}$$

by T, it suffices to bound each right-hand side summand uniformly over $\mathcal{T}$ in probability. We therefore drop also the m subscript for the remainder of this step. Now, for fixed $t \in \mathcal{T}$, adding and subtracting $p^{k_n}\,(W_i)'\,Q_{k_n}^{-1}\widehat{\psi}_{k_n}\,(t)\,U_i$ [with $\widehat{\psi}_k$ defined in (S.18)], recalling that $\delta_k\,(t,w) = p^k\,(w)'\,Q_k^{-1}\psi_k\,(t)$ we may decompose (the mth summand) as follows:

$$\begin{aligned}&\mathbb{E}_n[U_i\{\widehat{\delta}\,(t,W_i) - \delta\,(t,W_i)\}] - \mathbb{E}_n[\widehat{\delta}\,(t,W_i)\,\{\widehat{h}\,(W_i) - h^*\,(W_i)\}]\\&= \widehat{\psi}_{k_n}\,(t)'\,(\widehat{Q}_{k_n}^{-1} - Q_{k_n}^{-1})\mathbb{E}_n[p^{k_n}\,(W_i)\,U_i] + [\widehat{\psi}_{k_n}\,(t) - \psi_{k_n}\,(t)]'Q_{k_n}^{-1}\mathbb{E}_n[p^{k_n}\,(W_i)\,U_i]\\&\quad + \mathbb{E}_n[U_i\{\delta_{k_n}\,(t,W_i) - \delta\,(t,W_i)\}] - \widehat{\psi}_{k_n}\,(t)'\,\widehat{Q}_{k_n}^{-1}\mathbb{E}_n[p^{k_n}\,(W_i)\,\{\widehat{h}\,(W_i) - h^*\,(W_i)\}]\\&=: \mathrm{IV}_{a,n}\,(t) + \mathrm{IV}_{b,n}\,(t) + \mathrm{IV}_{c,n}\,(t) + \mathrm{IV}_{d,n}\,(t)\,.\end{aligned}$$

The desired $\|\mathrm{IV}_n\|_{\mathcal{T}} \to_{\mathrm{P}} 0$ will follow by T if we can show that the four remainder terms $\to_{\mathrm{P}} 0$. To this end, note first that by Assumptions 4 and 5,

$$\begin{aligned}\mathrm{E}\left[\|Q_k^{-1}\mathbb{E}_n[p^k\,(W_i)\,U_i]\|^2\right] \lesssim \mathrm{E}\left[\|Q_k^{-1/2}\mathbb{E}_n[p^k\,(W_i)\,U_i]\|^2\right] &= \mathrm{E}[U^2p^k\,(W)'\,Q_k^{-1}p^k\,(W)]/n\\&\lesssim \mathrm{E}[p^k\,(W)'\,Q_k^{-1}p^k\,(W)]/n = k/n,\end{aligned}$$

so by M and Assumption 7,

$$\|Q_{k_n}^{-1}\mathbb{E}_n[p^{k_n}\,(W_i)\,U_i]\| \lesssim_{\mathrm{P}} \sqrt{k_n/n} \to 0.$$

Step 4c in the proof of Lemma S.1 shows that $\sup_{t\in\mathcal{T}}\|\widehat{\psi}_{k_n}(t)'\,\widehat{Q}_{k_n}^{-1}\| \lesssim_{\mathrm{P}} 1$. Moreover, Lemma S.6 show that $\|\widehat{Q}_{k_n} - Q_{k_n}\|_{\mathrm{op}} \lesssim_{\mathrm{P}} [\zeta_{k_n}^2 \ln(k_n)/n]^{1/2} \to 0$, so by the previous display, CS and Assumption 7,

$$
\begin{aligned}
\|\mathrm{IV}_{a,n}\|_{\mathcal{T}} &= \left\|\widehat{\psi}_{k_n}(t)'\,\widehat{Q}_{k_n}^{-1}(Q_{k_n} - \widehat{Q}_{k_n})Q_{k_n}^{-1}\mathbb{E}_n\left[p^{k_n}(W_i)\,U_i\right]\right\|_{\mathcal{T}} \\
&\leqslant \left\|Q_{k_n}^{-1}\mathbb{E}_n\left[p^{k_n}(W_i)\,U_i\right]\right\| \|\widehat{Q}_{k_n} - Q_{k_n}\|_{\mathrm{op}} \sup_{t\in\mathcal{T}}\left\|\widehat{\psi}_{k_n}(t)'\,\widehat{Q}_{k_n}^{-1}\right\| \\
&\lesssim_{\mathrm{P}} (k_n/n)^{1/2}\left\{\zeta_{k_n}^2 \ln(k_n)/n\right\}^{1/2} \to 0.
\end{aligned}
$$

Step 4c in the proof of Lemma S.1 also shows that $\sup_{t\in\mathcal{T}}\|\widehat{\psi}_{k_n}(t) - \psi_{k_n}(t)\| \to_{\mathrm{P}} 0$, so by CS,

$$
\|\mathrm{IV}_{b,n}\|_{\mathcal{T}} \leqslant \left\|Q_{k_n}^{-1}\mathbb{E}_n\left[p^{k_n}(W_i)\,U_i\right]\right\| \sup_{\mathcal{T}}\left\|\widehat{\psi}_{k_n}(t) - \psi_{k_n}(t)\right\| \xrightarrow{\mathrm{P}} 0.
$$

Step 4c in the proof of Lemma A.2 shows that

$$
\|\mathrm{IV}_{c,n}\|_{t\in\mathcal{T}} = \sup_{t\in\mathcal{T}}|\mathbb{E}_n[U_i\{\delta_{k_n}(t, W_i) - \delta(t, W_i)\}]| \lesssim_{\mathrm{P}} R_{\delta,k_n}\sqrt{\ln(k_n/R_{\delta,k_n})} \to 0.
$$

Lastly, by CS, Lemma S.7 and $\sup_{t\in\mathcal{T}}\|\widehat{\psi}_{k_n}(t)'\,\widehat{Q}_{k_n}^{-1}\| \lesssim_{\mathrm{P}} 1$ we get

$$
\begin{aligned}
\|\mathrm{IV}_{d,n}\|_{\mathcal{T}} &\leqslant \left\|\mathbb{E}_n\left[p^{k_n}(W_i)\,\{\widehat{h}(W_i) - h^*(W_i)\}\right]\right\| \sup_{t\in\mathcal{T}}\left\|\widehat{\psi}_{k_n}(t)'\,\widehat{Q}_{k_n}^{-1}\right\| \\
&\lesssim \left(\sum_{j=1}^{k_n}\|p_j\|_{\mathcal{W}}^2\right)^{1/2} \max_{1\leqslant m'\leqslant d}\|\widehat{h}_{m'} - h_{m'}^*\|_{n,2} \sup_{t\in\mathcal{T}}\left\|\widehat{\psi}_{k_n}(t)'\,\widehat{Q}_{k_n}^{-1}\right\| \\
&\lesssim_{\mathrm{P}} \left(\sum_{j=1}^{k_n}\|p_j\|_{\mathcal{W}}^2\right)^{1/2} \max_{1\leqslant m'\leqslant d}\left(\sqrt{k_{m',n}/n} + k_{m',n}^{-\alpha_{m'}}\right),
\end{aligned}
$$

which $\to 0$ by Assumption 8.4. This finishes the proof of $\|\mathbb{E}_n[\widehat{g}(\cdot, Z_i) - g(\cdot, Z_i)]\|_{\mathcal{T}} \to_{\mathrm{P}} 0$. □

Proof of Lemma 3. Since the lemma is stated for a given ℓ, we drop the ℓ subscripts throughout and refer to the (ℓth) index set $\mathcal{X}_\ell$ as $\mathcal{T}$ itself. Then by T,

$$\begin{aligned}
\|\widehat{G} - G_n^*\|_{\mathcal{T}} &= \left\|\sqrt{n}\mathbb{E}_n\left[\left(\xi_i - \overline{\xi}\right)\widehat{g}\left(\cdot, Z_i\right)\right] - \sqrt{n}\mathbb{E}_n\left[\left(\xi_i - \overline{\xi}\right)g\left(\cdot, Z_i\right)\right]\right\|_{\mathcal{T}} \\
&= \left\|\sqrt{n}\mathbb{E}_n\left[\xi_i\widehat{g}\left(\cdot, Z_i\right)\right] - \sqrt{n}\mathbb{E}_n\left[\xi_i g\left(\cdot, Z_i\right)\right] - \sqrt{n}\cdot\overline{\xi}\cdot\mathbb{E}_n[\widehat{g}\left(\cdot, Z_i\right) - g\left(\cdot, Z_i\right)]\right\|_{\mathcal{T}} \\
&= \left\|\widehat{G}^u - G_n^{*u} - \sqrt{n}\cdot\overline{\xi}\cdot\mathbb{E}_n[\widehat{g}\left(\cdot, Z_i\right) - g\left(\cdot, Z_i\right)]\right\|_{\mathcal{T}} \\
&\leqslant \|\widehat{G}^u - G_n^{*u}\|_{\mathcal{T}} + |\sqrt{n}\cdot\overline{\xi}|\,\|\mathbb{E}_n[\widehat{g}\left(\cdot, Z_i\right)] - \mathbb{E}_n\left[g\left(\cdot, Z_i\right)\right]\|_{\mathcal{T}}.
\end{aligned}$$

The first term on the right $\to_{\mathrm{P}} 0$ by Lemma S.1. Given that $\sqrt{n}\cdot\overline{\xi} \sim \mathrm{N}(0,1)$, certainly $|\sqrt{n}\cdot\overline{\xi}| \lesssim_{\mathrm{P}} 1$. The second term therefore $\to_{\mathrm{P}} 0$ by Lemma S.2. □

S.4 Supporting Lemmas

For now, let Q and $\widehat{Q}$ be symmetric but otherwise arbitrary random matrices of possibly growing dimension. Also, denote the smallest and largest eigenvalue of a matrix A by $\lambda_{\min}(A)$ and $\lambda_{\max}(A)$, respectively.

Lemma S.3. *If* $\lambda_{\min}(Q) \geqslant c$ wp $\to 1$ *for some constant* $c \in (0,\infty)$ *and* $\|\widehat{Q} - Q\|_{\mathrm{op}} \to_{\mathrm{P}} 0$*, then* $\widehat{Q}$ *is invertible* wp $\to 1$ *and* $\lambda_{\min}(\widehat{Q})^{-1} \lesssim_{\mathrm{P}} 1$.[1]

Proof. Given that the eigenvalues of a symmetric (hence square) matrix A are bounded in absolute value by the operator norm, for conformable vectors v,

$$\begin{aligned}
\lambda_{\min}(\widehat{Q}) &= \min_{\|v\|=1}\left\{v'Qv + v'(\widehat{Q} - Q)v\right\} \\
&\geqslant \lambda_{\min}(Q) - \lambda_{\max}(Q - \widehat{Q}) \\
&\geqslant \lambda_{\min}(Q) - \|\widehat{Q} - Q\|_{\mathrm{op}}.
\end{aligned}$$

It follows that

$$\begin{aligned}
\mathrm{P}\left(\lambda_{\min}(\widehat{Q}) < c/2\right) &\leqslant \mathrm{P}\left(\lambda_{\min}(Q) - \|\widehat{Q} - Q\|_{\mathrm{op}} < c/2\right) \\
&\leqslant \mathrm{P}\left(\|\widehat{Q} - Q\|_{\mathrm{op}} \geqslant c/2\right) + \mathrm{P}\left(\lambda_{\min}(Q) < c\right) \to 0,
\end{aligned}$$

so $\mathrm{P}(\lambda_{\min}(\widehat{Q}) \geqslant c/2) \to 1$. Hence, $\widehat{Q}$ is invertible wp $\to 1$, and its smallest eigenvalue is bounded away from zero in probability. □

[1] This is Newey (1995, Lemma A.4) except that we state convergence in terms of the (weaker) operator matrix norm instead of the (stronger) Frobenius norm.

For now, let $Y, H \in \mathbf{R}^n, P \in \mathbf{R}^{n\times k}$ be arbitrary and of possibly growing dimensions n and k and abbreviate $U \coloneqq Y - H$, $\widehat{\pi} \coloneqq (P'P)^- P'Y$ and $\widehat{H} \coloneqq P\widehat{\pi}$.

Lemma S.4. *For any* $\pi \in \mathbf{R}^k$,

$$\|\widehat{H} - H\|^2 \leqslant U'P\left(P'P\right)^- P'U + \|P\pi - H\|^2,$$
$$\|\widehat{H} - P\pi\|^2 \leqslant 2U'P\left(P'P\right)^- P'U + 2\|P\pi - H\|^2.$$

Proof. Generalized inversion preserves symmetry, so $\mathcal{P}_P \coloneqq P\left(P'P\right)^- P'$ and $\mathcal{M}_P \coloneqq I_{n\times n} - \mathcal{P}_P$ are symmetric idempotent. Given that also $\mathcal{P}_P P = P$ [see, e.g., Rao (1973, 1b.5(vi)(a))], for any fixed $\pi \in \mathbf{R}^k$, we must have

$$\begin{aligned}\|\widehat{H} - H\|^2 &= \|\mathcal{P}_P Y - H\|^2 = \|\mathcal{P}_P U - \mathcal{M}_P H\|^2 = U'\mathcal{P}_P U + H'\mathcal{M}_P H \\ &= U'\mathcal{P}_P U + (H - P\pi)'\,\mathcal{M}_P\,(H - P\pi) \leqslant U'\mathcal{P}_P U + \|P\pi - H\|^2,\end{aligned}$$

where the inequality follows from the Min-Max theorem and an idempotent matrix having only zero or one eigenvalues. Similarly, abbreviating $H_\pi \coloneqq P\pi$,

$$\begin{aligned}\|\widehat{H} - P\pi\|^2 &= \|\mathcal{P}_P Y - H_\pi\|^2 = \|\mathcal{P}_P(U + H - H_\pi)\|^2 \\ &= (U + H - H_\pi)'\mathcal{P}_P(U + H - H_\pi) \\ &\leqslant 2U'\mathcal{P}_P U + 2(H - H_\pi)'\mathcal{P}_P(H - H_\pi) \\ &\leqslant 2U'\mathcal{P}_P U + 2\|H_\pi - H\|^2,\end{aligned}$$

where the first inequality follows from $(v + w)'A(v + w) \leqslant 2v'Av + 2w'Aw$ for A positive semi-definite (p.s.d.), and the second from idempotency of $\mathcal{P}_P$. □

Next, interpret $\{(Y_i, W_i)\}_1^n$ as i.i.d. $\mathbf{R}^{1+d}$-valued random variables with $d \in \mathbf{N}$ (fixed), $\mathrm{E}\left[Y^2\right] < \infty, \mathcal{W} \coloneqq \mathrm{supp}\,(W)$, and let $p^k : \mathbf{R}^d \to \mathbf{R}^k$ be a nonrandom vector function of possibly growing length satisfying $\zeta_k \coloneqq \sup_{w\in\mathcal{W}} \|p^k(w)\| < \infty$ for all $k \in \mathbf{N}$. Also, define $h(w) \coloneqq \mathrm{E}[Y|W = w]$, $\sigma^2(w) \coloneqq \mathrm{var}(Y|W = w), w \in \mathcal{W}$, and $U_i \coloneqq Y_i - h(W_i)$ and let $\mathbf{U}$ and P be the $n \times 1$ vector and $n \times k$ matrix of U_i's and $p^k(W_i)$'s, respectively.

Lemma S.5. $\mathrm{E}[\mathbf{U}'P\left(P'P\right)^- P'\mathbf{U}] \leqslant \|\sigma^2\|_{\mathcal{W}}\,(n \wedge k)\,.$

Proof. By the i.i.d. assumption, the positive semidefinite (p.s.d.) matrix

$$\mathrm{E}\left[\mathbf{U}\mathbf{U}' \middle| \{W_i\}_1^n\right] = \operatorname{diag}\left\{\sigma^2(W_i)\right\}_1^n.$$

Given that $\mathcal{P}_P \coloneqq P(P'P)^- P'$ is also p.s.d., using $\operatorname{tr}(AB) \leqslant \lambda_{\max}(A)\operatorname{tr}(B)$ for A, B p.s.d., we get

$$\begin{aligned}\mathrm{E}\left[\mathbf{U}'P(P'P)^- P'\mathbf{U}\middle| \{W_i\}_1^n\right] &= \operatorname{tr}\left(\mathrm{E}\left[\mathbf{U}\mathbf{U}' \middle| \{W_i\}_1^n\right]\mathcal{P}_P\right)\\ &\leqslant \max_{1\leqslant i\leqslant n}\sigma^2(W_i)\operatorname{tr}(\mathcal{P}_P) \leqslant \|\sigma^2\|_{\mathcal{W}}\operatorname{tr}(\mathcal{P}_P).\end{aligned}$$

Now, $\operatorname{tr}(\mathcal{P}_P) = \operatorname{tr}((P'P)^- P'P) = \operatorname{rank}(P'P)$, the latter equality following from Rao (1973, 1b(ii)(a)). The claim now follows from $\operatorname{rank}(P'P) = \operatorname{rank}(P) \leqslant n \wedge k$ upon taking expectation over the W_i's. □

Lemma S.6. *If the eigenvalues of $Q_k \coloneqq \mathrm{E}[p^k(W)p^k(W)']$ are bounded from above uniformly in k, then $\widehat{Q}_{k_n} \coloneqq \mathbb{E}_n[p^{k_n}(W_i)p^{k_n}(W_i)']$ satisfies*

$$\mathrm{E}\left[\left\|\widehat{Q}_{k_n} - Q_{k_n}\right\|_{op}\right] \lesssim \frac{\zeta_{k_n}^2 \ln k_n}{n} + \sqrt{\frac{\zeta_{k_n}^2 \ln k_n}{n}}.$$

Proof. The matrix $\widehat{Q}_k$ is the average of the n independent p.s.d. $k\times k$-matrix valued random variables $p^k(W_i)p^k(W_i)'$ with the matrix Q_k as their common mean. Given that

$$\begin{aligned}\left\|p^k(W_i)p^k(W_i)'\right\|_{\mathrm{op}} &\leqslant \left\|p^k(W_i)p^k(W_i)'\right\|_F\\ &= [\operatorname{tr}(p^k(W_i)p^k(W_i)'p^k(W_i)p^k(W_i)')]^{1/2}\\ &= \left\|p^k(W_i)\right\|^2 \leqslant \zeta_k^2,\end{aligned}$$

these n random matrices are bounded in operator norm by ζ_k^2. By hypothesis, $\|Q_k\|_{\mathrm{op}} = [\lambda_{\max}(Q_k'Q_k)]^{1/2} = \lambda_{\max}(Q_k) \lesssim 1$ as $k\to\infty$. The claim now follows from Belloni et al. (2015, Lemma 6.2), which builds on a fundamental result obtained by Rudelson (1999). □

Lemma S.7. *Let σ^2 be bounded on $\mathcal{W}$, the eigenvalues of $Q_k \coloneqq \mathrm{E}[p^k(W)p^k(W)']$ bounded from above and below uniformly in k, let $\widetilde{\pi} \in \mathbf{R}^k$ satisfy $\|p^{k\prime}\widetilde{\pi} - h\|_{\mathcal{W}} \lesssim k^{-\alpha}$ for some $\alpha \in (0,1)$, and define $\widetilde{h} \coloneqq p^{k\prime}\widetilde{\pi}$ and $\widehat{h} \coloneqq p^{k\prime}\widehat{\pi}$, where $\widehat{\pi} \coloneqq \widehat{Q}_{k_n}^- \mathbb{E}_n[p^{k_n}(W_i)Y_i]$.*

Then, provided $k_n/n \to 0$ and $\zeta_{k_n}^2 \ln(k_n)/n \to 0$, we have

1. $\|\widehat{h} - h\|_{n,2} \lesssim_{\mathrm{P}} \sqrt{k_n/n} + k_n^{-\alpha}$,
2. $\|\widehat{h} - \widetilde{h}\|_{n,2} \lesssim_{\mathrm{P}} \sqrt{k_n/n} + k_n^{-\alpha}$,
3. $\|\widehat{\pi} - \widetilde{\pi}\| \lesssim_{\mathrm{P}} \sqrt{k_n/n} + k_n^{-\alpha}$, *and*
4. $\|\widehat{h} - h\|_{\mathcal{W}} \lesssim_{\mathrm{P}} \zeta_{k_n}\left(\sqrt{k_n/n} + k_n^{-\alpha}\right)$.

Proof. By Lemma S.4,

$$\|\widehat{h} - h\|_{n,2}^2 \leqslant \mathbf{U}'P\left(P'P\right)^{-} P'\mathbf{U}/n + \|\widetilde{h} - h\|_{n,2}^2,$$
$$\|\widehat{h} - \widetilde{h}\|_{n,2}^2 \leqslant 2\mathbf{U}'P\left(P'P\right)^{-} P'\mathbf{U}/n + 2\|\widetilde{h} - h\|_{n,2}^2.$$

By hypothesis $\|\widetilde{h} - h\|_{n,2} \leqslant \|\widetilde{h} - h\|_{\mathcal{W}} \lesssim k^{-\alpha}$. Moreover, via M, Lemma S.5 and $\|\sigma^2\|_{\mathcal{W}} < \infty$ imply $\mathbf{U}'P\left(P'P\right)^{-} P'\mathbf{U} \lesssim_{\mathrm{P}} k_n$. The first two claims now follow from the previous display.

Via M, given that $\lambda_{\max}\left(Q_k\right) \lesssim 1$, Lemma S.6 and $\zeta_{k_n}^2 \ln(k_n)/n \to 0$ imply $\|\widehat{Q}_{k_n} - Q_{k_n}\|_{\mathrm{op}} \to_{\mathrm{P}} 0$. Given that also $\lambda_{\min}\left(Q_k\right)^{-1} \lesssim 1$, Lemma S.3 then implies that $\widehat{Q}_{k_n}$ is invertible wp $\to 1$ and $\lambda_{\min}(\widehat{Q}_{k_n})^{-1} \lesssim_{\mathrm{P}} 1$. Hence wp $\to 1$, by the Min-Max theorem

$$\|\widehat{\pi} - \widetilde{\pi}\|^2 \leqslant \lambda_{\min}(\widehat{Q}_{k_n})^{-1}\|P(\widehat{\pi} - \widetilde{\pi})\|^2/n = \lambda_{\min}(\widehat{Q}_{k_n})^{-1}\|\widehat{h} - \widetilde{h}\|_{n,2}^2 \lesssim_{\mathrm{P}} \|\widehat{h} - \widetilde{h}\|_{n,2}^2.$$

The third claim now follows from the second. Given that $\|\widehat{h}-\widetilde{h}\|_{\mathcal{W}} = \sup_{w\in\mathcal{W}} |p^k(w)'(\widehat{\pi}-\widetilde{\pi})| \leqslant \zeta_k\|\widehat{\pi} - \widetilde{\pi}\|$ and $\|\widetilde{h} - h\|_{\mathcal{W}} \lesssim k^{-\alpha}$, by T and the third claim,

$$\|\widehat{h} - h\|_{\mathcal{W}} \lesssim_{\mathrm{P}} \zeta_{k_n}\|\widehat{\pi} - \widetilde{\pi}\| + k_n^{-\alpha} \lesssim_{\mathrm{P}} \zeta_{k_n}\left(\sqrt{k_n/n} + k_n^{-\alpha}\right).$$

□

Lemma S.8. *Let X_n and Y_n be sequences of stochastic processes defined on a common probability space $(\Omega, \mathcal{F}, P)$ and taking values in a separable metric space $(\mathbb{D}, d)$, and let $\mathcal{F}_n$ be a sequence of sub-σ-algebras. If $X_n \rightsquigarrow_{\mathrm{P},\mathcal{F}} X$ in $\mathbb{D}$ and $d(X_n, Y_n) \to_{\mathrm{P}} 0$, then $Y_n \rightsquigarrow_{\mathrm{P},\mathcal{F}} X$ in $\mathbb{D}$.*

Proof. By T,

$$\begin{aligned}
&\sup_{h\in \mathrm{BL}_1(\mathbb{D})} \left|\mathrm{E}\left[h\left(Y_n\right)\right|\mathcal{F}_n\right] - \mathrm{E}\left[h\left(X\right)\right]| \\
&\leqslant \sup_{h\in \mathrm{BL}_1(\mathbb{D})} \left|\mathrm{E}\left[h\left(Y_n\right) - h\left(X_n\right)\right|\mathcal{F}_n\right]| + \sup_{h\in \mathrm{BL}_1(\mathbb{D})} \left|\mathrm{E}\left[h\left(X_n\right)\right|\mathcal{F}_n\right] - \mathrm{E}\left[h\left(X\right)\right]| \\
&\leqslant d\left(X_n, Y_n\right) \wedge 2 + o_{\mathrm{P}}\left(1\right) = o_{\mathrm{P}}\left(1\right).
\end{aligned}$$

□

Lemma S.9. *If X_n is a sequence of nonnegative random variables defined on a common probability space $(\Omega, \mathcal{F}, \mathrm{P})$, $\mathcal{F}_n$ is a sequence of sub-σ-algebras, and $\mathrm{E}\left[X_n\middle|\mathcal{F}_n\right] \to_{\mathrm{P}} 0$, then $X_n \to_{\mathrm{P}} 0$.*

Proof. Fix $n \in \mathbf{N}$, let $Y_n \coloneqq \mathrm{E}\left[X_n\middle|\mathcal{F}_n\right]$ and let $A_n \coloneqq \{Y_n = 0\}$. Then $X_n = 0$ almost everywhere on A_n. Indeed, if X_n is not zero almost everywhere on A_n, then there would exist a $C \in (0,\infty)$ such that $B_{n,C} \coloneqq \{\omega \in A_n; X_n(\omega) > 1/C\}$ satisfies $\mathrm{P}\left(B_{n,C}\right) > 0$. By definition of (a version of) the conditional expectation of X_n given $\mathcal{F}_n$, we must have $\int_A X_n \mathrm{d}P = \int_A Y_n \mathrm{d}P$ for every $A \in \mathcal{F}_n$ and, in particular, for A_n. Since $Y_n = 0$ on A_n and $B_{n,C}$ is a subset of A_n, it follows that

$$0 = \int_{A_n} Y_n \mathrm{d}P = \int_{A_n} X_n \mathrm{d}P \geqslant \int_{B_{n,C}} X_n \mathrm{d}P \geqslant \mathrm{P}\left(B_{n,C}\right)/C,$$

which contradicts $\mathrm{P}\left(B_{n,C}\right) > 0$. Since $n \in \mathbf{N}$ was arbitrary, we have shown that $X_n = 0$ on A_n for each $n \in \mathbf{N}$. Now, fix $\varepsilon, \delta > 0$. Then $\mathrm{P}\left(X_n > \varepsilon \cap Y_n = 0\right) = 0$ by the previous claim, and it follows that

$$\begin{aligned}
\mathrm{P}\left(X_n > \varepsilon\right) &= \mathrm{P}\left(X_n > \varepsilon \cap Y_n = 0\right) + \mathrm{P}\left(X_n > \varepsilon \cap 0 < Y_n \leqslant \delta\varepsilon\right) + \mathrm{P}\left(X_n > \varepsilon \cap Y_n > \delta\varepsilon\right) \\
&\leqslant \mathrm{P}\left(X_n > \delta^{-1} Y_n > 0\right) + \mathrm{P}\left(Y_n > \delta\varepsilon\right).
\end{aligned}$$

Given that Y_n is $\mathcal{F}_n$ measurable, by conditional M we have

$$\begin{aligned}
\mathrm{P}\left(X_n > \delta^{-1} Y_n > 0\right) &= \mathrm{E}\left[\mathbf{1}_{Y_n>0} \mathrm{P}\left(X_n > \delta^{-1} Y_n \middle| \mathcal{F}_n\right)\right] \leqslant \mathrm{E}\left[\mathbf{1}_{Y_n>0} \delta \mathrm{E}\left[X_n\middle|\mathcal{F}_n\right]/Y_n\right] \\
&= \delta \mathrm{P}\left(Y_n > 0\right) \leqslant \delta.
\end{aligned}$$

By $Y_n \to_{\mathrm{P}} 0$ and the previous two displays we see that for any $\varepsilon, \delta > 0, \limsup_{n\to\infty} \mathrm{P}(X_n > \varepsilon) \leqslant \delta$, so the claim follows from letting $\delta \to 0_+$. □

References

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.

Bravo, F. (2012): "Generalized empirical likelihood testing in semiparametric conditional moment restrictions models," *The Econometrics Journal*, 15, 1–31.

Newey, W. K. (1995): "Convergence rates for series estimators," in *Advances in Econometrics and Quantitative Economics*, ed. by G. Maddala, P. C. B. Phillips, and T. Srinivasan, Blackwell.

Newey, W. K. and D. McFadden (1994): "Large sample estimation and hypothesis testing," *Handbook of econometrics*, 4, 2111–2245.

Pollard, D. (1990): "Empirical processes: theory and applications," in *NSF-CBMS regional conference series in probability and statistics*, JSTOR, i–86.

Rao, C. R. (1973): *Linear statistical inference and its applications*, vol. 2, Wiley New York.

Rothe, C. and D. Wied (2013): "Misspecification testing in a class of conditional distributional models," *Journal of the American Statistical Association*, 108, 314–324.

Rudelson, M. (1999): "Random vectors in the isotropic position," *Journal of Functional Analysis*, 164, 60–72.

Rudin, W. (1976): *Principles of Mathematical Analysis*, International Series in Pure & Applied Mathematics, McGraw-Hill New York, third ed.

van der Vaart, A. and J. A. Wellner (2011): "A local maximal inequality under uniform entropy," *Electronic Journal of Statistics*, 5, 192.

van der Vaart, A. W. and J. A. Wellner (1996): *Weak Convergence and Empirical Processes*, Springer.