Supplement to “Adaptive, Rate-Optimal Hypothesis Testing in Nonparametric IV Models”

Christoph Breunig   Xiaohong Chen

First version: August 2018; Latest revision: July 30, 2024

This supplementary appendix contains materials to support our main paper. Appendix C presents additional simulation results. Appendix D provides proofs of our results on confidence sets in Subsection 4.3. Appendix E presents additional technical lemmas and all the proofs.

C. Additional Simulations

This section provides additional simulation results. All the simulation results are based on 5000 Monte Carlo replications for every experiment and are at the nominal level $\alpha = 0.05$.

C.1. Adaptive Testing for Monotonicity: Simulation Design II

We generate the dependent variable Y according to the NPIV model (2.1), where

$$h(x) = c_0(x/5 + x^2) + c_A \sin(2\pi x) \ , \tag{C.1}$$

$c_0 \in \{0, 1\}$, $c_A \in [0, 0.6]$, and $W = \Phi(W^*)$, $X = \Phi\big(\xi W^* + \sqrt{1-\xi^2}\epsilon\big)$, $U = (0.3\epsilon + \sqrt{1-(0.3)^2}\nu)/2$, where (W^*, ϵ, ν) follows a multivariate standard normal distribution. This design with $(c_0, c_A) = (1, 0)$ and $\xi \in \{0.3, 0.5\}$ is the one in Chetverikov and Wilhelm (2017). The null hypothesis is that the NPIV function $h(\cdot)$ is weakly increasing on the support of X. The null is satisfied when $c_A \in [0, 0.184)$, and is violated when $c_A \geq 0.184$. We note that $c_0 = 0$, $c_A = 0.0$ corresponds to the boundary of the null hypothesis. Note that the degree of nonlinearity/complexity of h given in (C.1) becomes larger as $c_A > 0$ increases.

We implement our adaptive test $\widehat{\mathtt{T}}_n$ given in (2.12) in the main paper, and the Fang and Seo (2021) test for monotonicity of a NPIV function, denoted as FS. The FS test is computed using R language translation of their Matlab program code, with their deterministically chosen $J = 3$, $K \geq 3$ and other tuning parameter choices detailed in their 2019 arXiv version (also see the description in our main paper).

Table A reports the empirical size of our adaptive test $\widehat{\mathtt{T}}_n$, with $K(J) \in \{2J,\ 4J,\ 8J\}$, and using quadratic B-spline basis functions with varying number of knots for the unrestricted NPIV h. We also report the empirical size of the FS test, using $J = 3$ and

n	c_0	c_A	ξ	$\widehat{\mathtt{T}}_n$ $K(J)=2J$	$\widehat{J}$ $K(J)=2J$	$\widehat{\mathtt{T}}_n$ $K(J)=4J$	$\widehat{J}$ $K(J)=4J$	$\widehat{\mathtt{T}}_n$ $K(J)=8J$	$\widehat{J}$ $K(J)=8J$	FS $K=5$	FS $K=12$	FS $K=24$
500	0	0.0	0.3	0.004	3.01	0.012	3.03	0.011	3.21	0.005	0.013	0.018
			0.5	0.016	3.32	0.018	3.38	0.021	3.40	0.035	0.035	0.036
			0.7	0.025	3.57	0.030	3.58	0.026	3.49	0.050	0.049	0.042
	1	0.0	0.3	0.002	3.01	0.005	3.03	0.003	3.12	0.000	0.000	0.000
			0.5	0.004	3.38	0.004	3.36	0.004	3.25	0.000	0.000	0.000
			0.7	0.004	3.71	0.004	3.65	0.004	3.38	0.000	0.000	0.000
	1	0.1	0.3	0.002	3.01	0.006	3.03	0.005	3.12	0.000	0.000	0.000
			0.5	0.007	3.37	0.007	3.35	0.007	3.25	0.001	0.001	0.001
			0.7	0.009	3.64	0.008	3.59	0.008	3.34	0.000	0.000	0.000
1000	0	0.0	0.3	0.009	3.01	0.016	3.07	0.015	3.27	0.011	0.021	0.026
			0.5	0.023	3.50	0.025	3.47	0.028	3.45	0.051	0.046	0.044
			0.7	0.034	3.87	0.034	3.97	0.034	3.52	0.059	0.055	0.047
	1	0.0	0.3	0.003	3.02	0.005	3.06	0.004	3.15	0.000	0.000	0.000
			0.5	0.006	3.63	0.005	3.46	0.006	3.28	0.000	0.000	0.000
			0.7	0.003	4.23	0.003	4.22	0.003	3.46	0.000	0.000	0.000
	1	0.1	0.3	0.004	3.02	0.008	3.06	0.005	3.15	0.000	0.001	0.001
			0.5	0.009	3.59	0.009	3.44	0.010	3.29	0.001	0.001	0.001
			0.7	0.011	4.09	0.010	4.10	0.009	3.38	0.000	0.000	0.000
5000	0	0.0	0.3	0.020	3.38	0.019	3.42	0.026	3.39	0.040	0.040	0.044
			0.5	0.038	3.56	0.036	3.62	0.035	3.49	0.056	0.057	0.055
			0.7	0.045	4.14	0.042	4.12	0.035	3.75	0.056	0.059	0.058
	1	0.0	0.3	0.005	3.44	0.006	3.35	0.006	3.23	0.000	0.001	0.000
			0.5	0.004	3.81	0.003	3.80	0.003	3.47	0.000	0.000	0.000
			0.7	0.002	4.74	0.002	4.69	0.002	3.98	0.000	0.000	0.000
	1	0.1	0.3	0.009	3.42	0.008	3.35	0.009	3.24	0.001	0.002	0.001
			0.5	0.013	3.70	0.013	3.69	0.011	3.40	0.000	0.000	0.000
			0.7	0.008	4.52	0.006	4.46	0.006	3.75	0.000	0.000	0.000

Table A: Testing Monotonicity - Empirical Size of our adaptive test $\widehat{\mathtt{T}}_n$ and of the FS test (with $J=3$). Monte Carlo average value $\widehat{J}$. Nominal level $\alpha = 0.05$. Design from Appendix C.1 with NPIV function (C.1). Instrument strength increases in ξ.

$K \in \{5,\ 12,\ 24\}$ as comparison to our adaptive test's $K(J) \in \{2J,\ 4J,\ 8J\}$. From Table A we observe that our adaptive test $\widehat{\mathtt{T}}_n$ is slightly under sized across different sample sizes, different instrument strength, different $K(J)$ and different design specifications. The FS test is mostly undersized, but is slightly oversized at the boundary ($c_0 = 0,\ c_A = 0.0$) for sample sizes $n = 1000,\ 5000$ and strong instrument strength $\xi = 0.7$ even when $J = 3,\ K = 5$ (the most powerful choice in the 2019 arXiv version of Fang and Seo (2021)).

Figure A provides empirical rejection probabilities of our adaptive test $\widehat{\mathtt{T}}_n$ (blue and red solid lines) with $K(J) \in \{4J,\ 8J\}$ and of the FS test (with $J = 3,\ K = 5$; green dashed lines). The power curves of all tests improve as the instrument strength ξ increases. Our adaptive test with $K(J) = 8J$ has better empirical power in finite samples when instrument is weak, but the choice of $K(J)$ is less significant as the sample size or the instrument strength increases. For instrument strength $\xi = 0.3$, the FS test has almost trivial power for $c_A \in [0.2, 0.5]$ even for large sample size $n = 5000$, while our adaptive

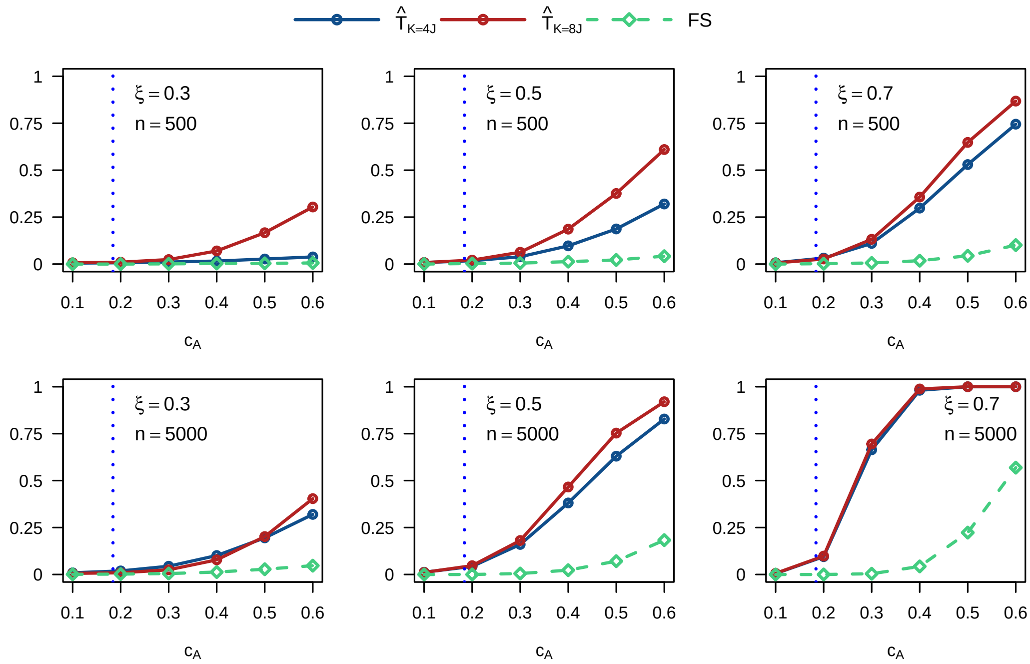

Figure A: Testing Monotonicity – Empirical power of our adaptive test $\widehat{\mathtt{T}}_n$ with $K(J) = 4J$ (blue solid lines) and $K(J) = 8J$ (red solid lines) and the FS test (with $J = 3$, $K = 5$, green dashed lines). Design from Appendix C.1 model (C.1) with $c_0 = 1$. The vertical dotted line indicates when the null hypothesis is violated (when $c_A \geq 0.184$). Instrument strength increases in ξ.

test $\widehat{\mathtt{T}}_n$ has non-trivial power for all $c_A \geq 0.3$. Moreover, the finite sample power of our adaptive test $\widehat{\mathtt{T}}_n$ increases much faster than the FS test as $c_A > 0.2$ becomes larger. Figure A shows the substantial finite sample power gains through adaptation even in small sample size $n = 500$.

Remark C.1. *When testing for inequality restrictions (IR) $\mathcal{H}_0 = \{h \in \mathcal{H} : \partial^l h \geq 0\}$, such as monotonicity and convexity, we could also compute our adaptive test $\widehat{\mathtt{T}}_n$ using modified critical values in Step 2 as follows: The estimator in (2.6) can be written as $\widehat{h}^{\mathrm{R}}_J(\cdot) = \psi^J(\cdot)'\widehat{\beta}^{\mathrm{R}}$. By construction of the estimator we have $\partial^l \widehat{h}^{\mathrm{R}}_J(X_i) \geq 0$, for all $1 \leq i \leq n$, or equivalently $\partial^l \Psi \widehat{\beta}^{\mathrm{R}} \geq 0$, where the application of the derivative operator is understood elementwise and rank$(\partial^l \Psi) \leq J$. Let Ψ_{act} be a submatrix of Ψ such that $\partial^l \Psi_{act} \widehat{\beta}^{\mathrm{R}} = 0$. Set $\widehat{\gamma}_J = \max\big(1, \text{rank}(\partial^l \Psi_{act})\big)$ and compute for a given nominal level $\alpha \in (0,1)$:*

$$\widehat{\eta}_J(\alpha) = \frac{q\big(\alpha/\#(\widehat{\mathcal{I}}_n), \widehat{\gamma}_J\big) - \widehat{\gamma}_J}{\sqrt{\widehat{\gamma}_J}}, \tag{C.2}$$

where $q(a, \gamma)$ denotes the $100(1-a)\%$-quantile of the chi-square distribution with γ degrees of freedom. Assuming that $J^c \leq \widehat{\gamma}_J$, $J \in \mathcal{I}_n$, for some constant $0 < c \leq 1$ with probability approaching one uniformly for $h \in \mathcal{H}$, Breunig and Chen (2021) establishes size control of the test statistic using the modified critical values given in (C.2). See Breunig and

Chen (2021) also for simulations and real data application of testing for monotonicity and convexity using this modified critical values. The simulations and empirical findings reported in Breunig and Chen (2021) are virtually the same, in terms of empirical size and power, as the ones reported in this revised version for testing inequalities.

C.2. Simulations for Multivariate Instruments

This section presents additional simulations for testing parametric hypotheses in the presence of multivariate conditioning variable $W = (W_1, W_2)$. We set $X_i = \Phi(X_i^*)$, $W_{1i} = \Phi(W_{1i}^*)$, and $W_{2i} = \Phi(W_{2i}^*)$, where

$$\begin{pmatrix} X_i^* \\ W_{1i}^* \\ W_{2i}^* \\ U_i \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \xi & 0.4 & 0.3 \\ \xi & 1 & 0 & 0 \\ 0.4 & 0 & 1 & 0 \\ 0.3 & 0 & 0 & 1 \end{pmatrix}\right). \tag{C.3}$$

We generate the dependent variable Y according to the NPIV model (2.1) where $h(x) = -x/5 + c_A x^2$. We test the null hypothesis of linearity, i.e., whether $c_A = 0$.

Horowitz (2006) assumes $d_x = d_w$ and hence we cannot compare our adaptive test with his for Design (C.3). Instead we will compare our adaptive test $\widehat{\mathtt{T}}_n$ against an adaptive image-space test (IT), which is our proposed adaptive version of Bierens (1990)'s type test for semi-nonparametric conditional moment restrictions.[1] Specifically, our image-space test (IT) is based on a leave-one-out sieve estimator of the quadratic functional $\mathrm{E}[\mathrm{E}[Y - h^{\mathrm{R}}(X)|W]^2]$, given by

$$\widehat{D}_K = \frac{2}{n(n-1)} \sum_{1 \le i < i' \le n} \big(Y_i - \widehat{h}^{\mathrm{R}}(X_i)\big)\big(Y_{i'} - \widehat{h}^{\mathrm{R}}(X_{i'})\big) b^K(W_i)'(B'B/n)^{-} b^K(W_{i'}),$$

where $\widehat{h}^{\mathrm{R}}$ is a null restricted parametric estimator for the null parametric function h^{R}. The data-driven IT statistic is:

$$\widehat{\mathtt{IT}}_n = \mathbb{1}\Big\{\text{there exists } K \in \widehat{\mathcal{I}}_n \text{ such that } n\widehat{D}_K/\widehat{V}_K > \big(q\big(\alpha/\#(\widehat{\mathcal{I}}_n), K\big) - K\big)/\sqrt{K}\Big\}$$

with the estimator $\widehat{V}_K = \big\|(B'B)^{-1/2} \sum_{i=1}^n (Y_i - \widehat{h}^{\mathrm{R}}(X_i))^2 b^K(W_i) b^K(W_i)'(B'B)^{-1/2}\big\|_F$, and the adjusted index set $\widehat{\mathcal{I}}_n = \{K \le \widehat{K}_{\max} : K = \underline{K}2^k \text{ where } k = 0, 1, \ldots, k_{\max}\}$, where $\underline{K} := \lfloor\sqrt{\log\log n}\rfloor$, $k_{\max} := \lceil \log_2(n^{1/3}/\underline{K})\rceil$, and the empirical upper bound $\widehat{K}_{\max} = \min\big\{K > \underline{K} : 10\,\zeta^2(K)\sqrt{(\log K)/n} \ge s_{\min}\big((B'B/n)^{-1/2}\big)\big\}$. Finally $q\big(a, K\big)$ is the $100(1-a)$%-quantile of the chi-square distribution with K degrees of freedom. In this simulation, it is convenient to additionally weight the basis functions by $(B'B/n)^{-1/2}$ to improve the finite

[1] We refer readers to Breunig and Chen (2020) for the theoretical properties of the adaptive image-space test.

sample performance of the IT statistic. Table B compares the empirical size of the adaptive

n	Design	ξ	$\widehat{\mathtt{T}}_n,\ K(J)=4J$	$\widehat{J}$	$\widehat{\mathtt{IT}}_n$	$\widehat{K}$
500	(5.1)	0.3	0.023	3.03	0.046	3.38
	$d_x = d_w$	0.5	0.028	3.40	0.046	3.37
		0.7	0.035	3.56	0.046	3.37
	(C.3)	0.3	0.034	3.45	0.034	6.00
	$d_x < d_w$	0.5	0.035	3.49	0.035	6.00
		0.7	0.038	3.55	0.040	6.00
1000	(5.1)	0.3	0.022	3.07	0.053	3.40
		0.5	0.027	3.48	0.051	3.39
		0.7	0.037	3.58	0.049	3.39
	(C.3)	0.3	0.039	3.47	0.032	6.93
		0.5	0.040	3.50	0.038	6.92
		0.7	0.043	3.58	0.037	6.90
5000	(5.1)	0.3	0.032	3.43	0.049	3.38
		0.5	0.043	3.55	0.045	3.39
		0.7	0.049	3.63	0.042	3.38
	(C.3)	0.3	0.049	3.51	0.048	10.28
		0.5	0.048	3.57	0.046	10.27
		0.7	0.050	3.80	0.051	10.25

Table B: Testing Parametric Form - Empirical size of our adaptive tests $\widehat{\mathtt{T}}_n$ and of $\widehat{\mathtt{IT}}_n$. Nominal level $\alpha = 0.05$. Monte Carlo average value $\widehat{J}$. Design from Appendix C.2. Instrument strength increases in ξ.

image space test $\widehat{\mathtt{IT}}_n$ with our adaptive structural space test $\widehat{\mathtt{T}}_n$, at the 5% nominal level. We see that both tests provide accurate size control. We also report the average choices of sieve dimension parameters, as described in Section 5. The multivariate design (C.3) leads to larger sieve dimension choices $\widehat{K}$ in adaptive image-space tests $\widehat{\mathtt{IT}}_n$ while the sieve dimension choices $\widehat{J}$ of our adaptive structural-space test $\widehat{\mathtt{T}}_n$ is not sensitive to the dimensionality (d_w) of the conditional instruments.

Figure B compares the empirical power of $\widehat{\mathtt{IT}}_n$ and of $\widehat{\mathtt{T}}_n$, at the 5% nominal level, using the sample sizes $n = 500$ (1st and 2nd rows) and $n = 1000$ (3rd and 4th rows). The finite sample empirical power curves of both tests increase with ξ and sample size n. For the scalar conditional instrument case, while our adaptive structural space test $\widehat{\mathtt{T}}_n$ is more powerful when $\xi \in \{0.3, 0.5\}$ (weaker strength of instruments), the finite sample power curves of both tests are similar when $\xi = 0.7$. For the multivariate conditional instruments case, while the power of our adaptive structural space test $\widehat{\mathtt{T}}_n$ increases with larger dimension d_w, the adaptive image space test $\widehat{\mathtt{IT}}_n$ suffers from larger d_w and has lower power. The same patterns are also present when we compare the two tests using size-adjusted empirical power curves (see our arXiv:2006.09587v3 version, Appendix C.3).

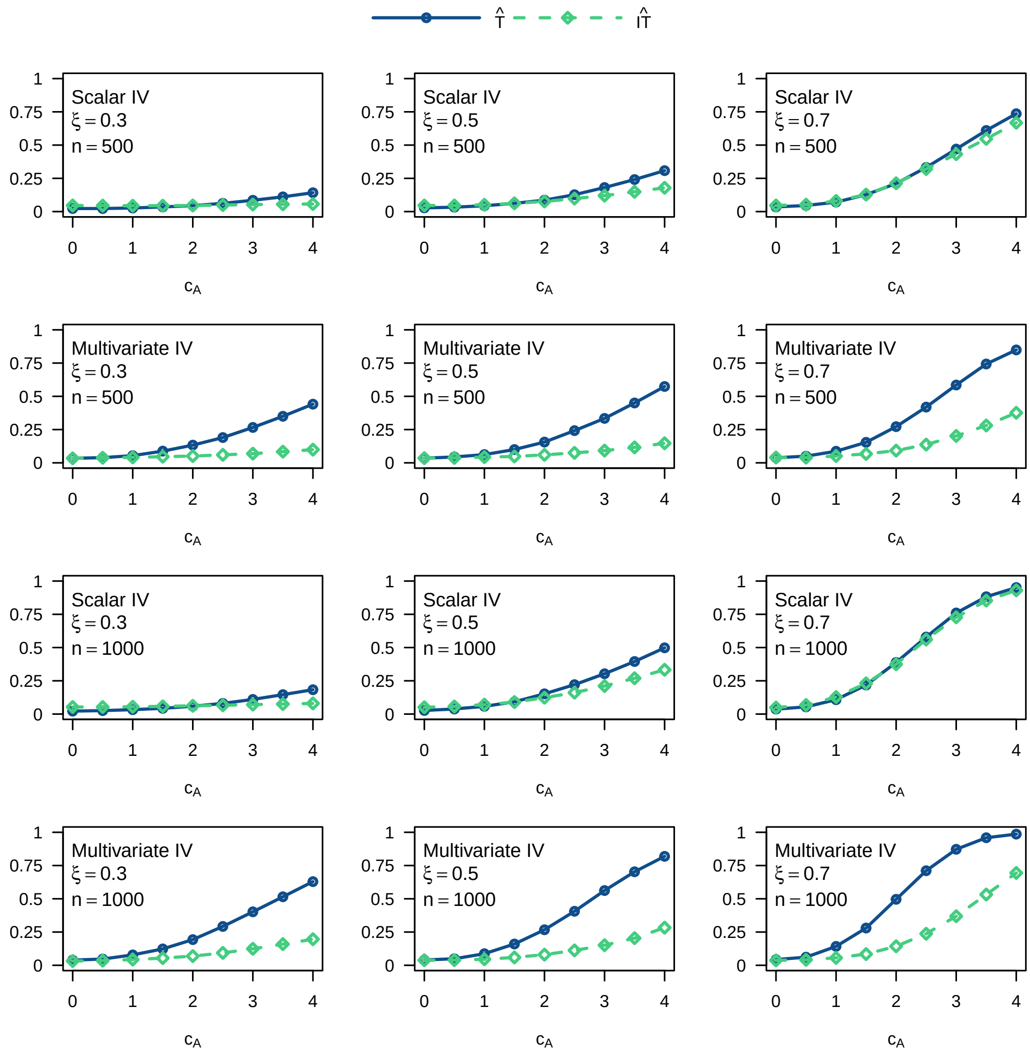

Figure B: Testing Parametric Form - Empirical power of our adaptive tests $\widehat{\mathtt{T}}_n$ (blue solid lines) and of $\widehat{\mathtt{IT}}_n$ (green dashed lines). 1st and 3rd rows: power comparisons in scalar IV case ($d_w = 1$); 2nd and 4th rows: power comparisons in multivariate IV case ($d_w > 1$). Design from Appendix C.2. Instrument strength increases in ξ.

D. Proofs of Inference Results in Subsection 4.3

Proof of Corollary 4.1. Proof of (4.8). We observe

$$\limsup_{n\to\infty}\sup_{h\in\mathcal{H}_0} \mathrm{P}_h\left(h \notin \mathcal{C}_n(\alpha)\right) = \limsup_{n\to\infty}\sup_{h\in\mathcal{H}_0} \mathrm{P}_h\left(\max_{J\in\widehat{\mathcal{I}}_n} \frac{n\widehat{D}_J(h)}{\widehat{\eta}_J(\alpha)\,\widehat{V}_J} > 1\right) \leq \alpha,$$

where the last inequality is due to step 1 of the proof of Theorem 4.1 and step 3 of the proof of Theorem 4.2.

Proof of (4.9). Let J^* be as be as in step 2 of the proof of Theorem 4.1. We observe uniformly for $h \in \mathcal{H}_1(\delta^\circ \mathsf{r}_n)$ that

$$\mathrm{P}_h\left(h \notin \mathcal{C}_n(\alpha)\right) = \mathrm{P}_h\left(\max_{J \in \widehat{\mathcal{I}}_n} \frac{n\widehat{D}_J(h)}{\widehat{\eta}_J(\alpha)\widehat{V}_J} > 1\right) = 1 - \mathrm{P}_h\left(\max_{J \in \widehat{\mathcal{I}}_n} \frac{n\widehat{D}_J(h)}{\widehat{\eta}_J(\alpha)\,\widehat{V}_J} \leq 1\right) = 1 - o(1),$$

where the last equation is due to step 2 of the proof of Theorem 4.1 and step 3 of the proof of Theorem 4.2. □

Proof of Corollary 4.2. For any $h \in \mathcal{H}_0$, we analyze the diameter of the confidence set $\mathcal{C}_n(\alpha)$ under P_h. Lemma B.8 implies $\sup_{h \in \mathcal{H}_0} \mathrm{P}_h\big(\widehat{J}_{\max} > \overline{J}\big) = o(1)$ and hence, it is sufficient to consider the deterministic index set $\mathcal{I}_n$ given in (4.2). For all $h_1 \in \mathcal{C}_n(\alpha) \subset \mathcal{H}_0$ it holds for all $J \in \mathcal{I}_n$ by using the definition of the projection Q_J given in (B.1):

$$\begin{aligned} \|h - h_1\|_{L^2(X)} &\leq \|Q_J \Pi_J (h - h_1)\|_{L^2(X)} + \|\Pi_J h - h\|_{L^2(X)} + \|\Pi_J h_1 - h_1\|_{L^2(X)} \\ &\leq \|Q_J(h - h_1)\|_{L^2(X)} + O(J^{-p/d_x}), \end{aligned} \tag{D.1}$$

due to the triangular inequality and the sieve approximation bound from the smoothness restrictions imposed on $\mathcal{H}$. By Theorem B.1 we have

$$\Big| \|Q_J(h - h_1)\|^2_{L^2(X)} - \widehat{D}_J(h_1) \Big| \lesssim n^{-1} s_J^{-2} \sqrt{J} + n^{-1/2} s_J^{-1} \big(\|h - h_1\|_{L^2(X)} + J^{-p/d_x} \big)$$

wpa1 uniformly for $h \in \mathcal{H}_0$. Consequently, the definition of the confidence set $\mathcal{C}_n(\alpha)$ with $h_1 \in \mathcal{C}_n(\alpha)$ gives for all $J \in \mathcal{I}_n$:

$$\begin{aligned} \|Q_J(h - h_1)\|^2_{L^2(X)} &\lesssim \; n^{-1} \widehat{\eta}_J(\alpha)\, \widehat{V}_J + n^{-1/2} s_J^{-1} \big(\|h - h_1\|_{L^2(X)} + J^{-p/d_x} \big) + n^{-1} s_J^{-2} \sqrt{J} \\ &\lesssim n^{-1} \sqrt{\log \log n}\, s_J^{-2} \sqrt{J} + n^{-1/2} s_J^{-1} \big(\|h - h_1\|_{L^2(X)} + J^{-p/d_x} \big) \end{aligned}$$

wpa1 uniformly for $h \in \mathcal{H}_0$ by using Lemmas B.2, B.5 and B.4(ii). Consequently, inequality (D.1) yields for all $J \in \mathcal{I}_n$:

$$\|h - h_1\|^2_{L^2(X)} \lesssim \frac{n^{-1} \sqrt{\log \log n}\, s_J^{-2} \sqrt{J} + J^{-2p/d_x}}{1 - C_B n^{-1/2} s_J^{-1}}$$

wpa1 uniformly for $h \in \mathcal{H}_0$. Now using that $n^{-1/2} s_J^{-1} = o(1)$ for all $J \in \mathcal{I}_n$ by Assumption 4(i) we obtain $\|h - h_1\|_{L^2(X)} \lesssim n^{-1/2} (\log \log n)^{1/4}\, s_J^{-1} J^{1/4} + J^{-p/d_x}$ with probability approaching one uniformly for $h \in \mathcal{H}_0$. We may choose $J = cJ^\circ \in \mathcal{I}_n$ for some constant $c > 0$ and n sufficiently large and hence, the result follows. □

E. Technical Results

Below, $\lambda_{\max}(\cdot)$ denotes the maximal eigenvalue of a matrix.

Lemma E.1. *Let Assumptions 1(ii)-(iii) and 2 hold. Then, wpa1 uniformly for $h \in \mathcal{H}$:*

$$\frac{1}{n(n-1)}\sum_{i\neq i'}\big(Y_i-\Pi_{\mathcal{H}_0}h(X_i)\big)\big(Y_{i'}-\Pi_{\mathcal{H}_0}h(X_{i'})\big)b^K(W_i)'\Big(A'A-\widehat{A}'\widehat{A}\Big)b^K(W_{i'})$$
$$\lesssim n^{-1}V_J+n^{-1/2}s_J^{-1}\big(\|h-\Pi_{\mathcal{H}_0}h\|_{L^2(X)}+J^{-p/d_x}\big).$$

Proof. Let $\Pi_{\mathcal{H}_0}^{\perp}:=\mathrm{id}-\Pi_{\mathcal{H}_0}$. We establish an upper bound of

$$\begin{aligned}
&\frac{1}{n^2}\sum_{i,i'}\big(Y_i-\Pi_{\mathcal{H}_0}h(X_i)\big)\big(Y_{i'}-\Pi_{\mathcal{H}_0}h(X_{i'})\big)b^K(W_i)'\Big(A'A-\widehat{A}'\widehat{A}\Big)b^K(W_{i'})\\
&=\mathrm{E}[\Pi_{\mathcal{H}_0}^{\perp}h(X)b^K(W)]'\Big(A'A-\widehat{A}'\widehat{A}\Big)\,\mathrm{E}[\Pi_{\mathcal{H}_0}^{\perp}h(X)b^K(W)]\\
&+2\Big(\frac{1}{n}\sum_i\big(Y_i-\Pi_{\mathcal{H}_0}h(X_i)\big)b^K(W_i)-\mathrm{E}[\Pi_{\mathcal{H}_0}^{\perp}h(X)b^K(W)]\Big)'\Big(A'A-\widehat{A}'\widehat{A}\Big)\,\mathrm{E}[\Pi_{\mathcal{H}_0}^{\perp}h(X)b^K(W)]\\
&+\Big(\frac{1}{n}\sum_i\big(Y_i-\Pi_{\mathcal{H}_0}h(X_i)\big)b^K(W_i)'-\mathrm{E}[\Pi_{\mathcal{H}_0}^{\perp}h(X)b^K(W)]'\Big)\Big(A'A-\widehat{A}'\widehat{A}\Big)\\
&\qquad\times\Big(\frac{1}{n}\sum_i\big(Y_i-\Pi_{\mathcal{H}_0}h(X_i)\big)b^K(W_i)'-\mathrm{E}[\Pi_{\mathcal{H}_0}^{\perp}h(X)b^K(W)]'\Big)
\end{aligned}$$

uniformly for $h \in \mathcal{H}$. It is sufficient to bound the first summand on the right hand side. We make use of the decomposition

$$\begin{aligned}
&\mathrm{E}[\Pi_{\mathcal{H}_0}^{\perp}h(X)b^K(W)]'\Big(A'A-\widehat{A}'\widehat{A}\Big)\,\mathrm{E}[\Pi_{\mathcal{H}_0}^{\perp}h(X)b^K(W)]\\
&\qquad=2\,\mathrm{E}[\Pi_{\mathcal{H}_0}^{\perp}h(X)b^K(W)]'A'(A-\widehat{A})\,\mathrm{E}[\Pi_{\mathcal{H}_0}^{\perp}h(X)b^K(W)]\\
&\qquad\quad-\mathrm{E}[\Pi_{\mathcal{H}_0}^{\perp}h(X)b^K(W)]'(A-\widehat{A})'(A-\widehat{A})\,\mathrm{E}[\Pi_{\mathcal{H}_0}^{\perp}h(X)b^K(W)]=:2T_1-T_2.
\end{aligned}$$

We first consider the term T_1 as follows:

$$\begin{aligned}
T_1&=\mathrm{E}[\Pi_{\mathcal{H}_0}^{\perp}h(X)b^K(W)]'A'(\widehat{A}-A)\,\mathrm{E}[\Pi_J\Pi_{\mathcal{H}_0}^{\perp}h(X)b^K(W)]\\
&+\mathrm{E}[\Pi_{\mathcal{H}_0}^{\perp}h(X)b^K(W)]'A'(\widehat{A}-A)\,\mathrm{E}[(\Pi_{\mathcal{H}_0}^{\perp}h-\Pi_J\Pi_{\mathcal{H}_0}^{\perp}h)(X)b^K(W)]:=A_1+A_2.
\end{aligned}\tag{E.1}$$

We now consider the term A_1. Recall that $Q_J\Pi_J h=\Pi_J h$ and $\widehat{S}G^{-1}\langle h,\psi^J\rangle_{L^2(X)}=$

$n^{-1}\sum_i \Pi_J h(X_i)b^K(W_i)$. We have:

$$
\begin{aligned}
&\big((G_b^{-1/2}S)_l^- \operatorname{E}[\Pi_{\mathcal{H}_0}^\perp h(X)\widetilde{b}^K(W)]\big)' G\big((G_b^{-1/2}S)_l^- - (\widehat{G}_b^{-1/2}\widehat{S})_l^-\widehat{G}_b^{-1/2}G_b^{1/2}\big)\operatorname{E}[\Pi_J\Pi_{\mathcal{H}_0}^\perp h(X)\widetilde{b}^K(W)] \\
&= \big\langle Q_J\Pi_{\mathcal{H}_0}^\perp h, \Pi_J\Pi_{\mathcal{H}_0}^\perp h - (\psi^J)'(\widehat{G}_b^{-1/2}\widehat{S})_l^-\widehat{G}_b^{-1/2}\operatorname{E}[\Pi_{\mathcal{H}_0}^\perp h(X)b^K(W)]\big\rangle_{L^2(X)} \\
&= \langle Q_J\Pi_{\mathcal{H}_0}^\perp h, \psi^J\rangle_{L^2(X)}'(\widehat{G}_b^{-1/2}\widehat{S})_l^-\widehat{G}_b^{-1/2}\Big(\frac{1}{n}\sum_i \Pi_J\Pi_{\mathcal{H}_0}^\perp h(X_i)b^K(W_i) - \operatorname{E}[\Pi_{\mathcal{H}_0}^\perp h(X)b^K(W)]\Big) \\
&= \langle Q_J\Pi_{\mathcal{H}_0}^\perp h, \psi^J\rangle_{L^2(X)}'(G_b^{-1/2}S)_l^-\Big(\frac{1}{n}\sum_i \Pi_J\Pi_{\mathcal{H}_0}^\perp h(X_i)\widetilde{b}^K(W_i) - \operatorname{E}[\Pi_J\Pi_{\mathcal{H}_0}^\perp h(X)\widetilde{b}^K(W)]\Big) \\
&\quad + \langle Q_J\Pi_{\mathcal{H}_0}^\perp h, \psi^J\rangle_{L^2(X)}'(G_b^{-1/2}S)_l^- G_b^{-1/2}S'\Big((\widehat{G}_b^{-1/2}\widehat{S})_l^-\widehat{G}_b^{-1/2}G_b^{1/2} - (G_b^{-1/2}S)_l^-\Big) \\
&\qquad \times\Big(\frac{1}{n}\sum_i \Pi_J\Pi_{\mathcal{H}_0}^\perp h(X_i)\widetilde{b}^K(W_i) - \operatorname{E}[\Pi_J\Pi_{\mathcal{H}_0}^\perp h(X)\widetilde{b}^K(W)]\Big) =: A_{11} + A_{12},
\end{aligned}
$$

where we used the notation $\widetilde{b}^K(\cdot) = G_b^{-1/2}b^K(\cdot)$. Consider A_{11} we have:

$$
\begin{aligned}
\operatorname{E}|A_{11}|^2 &\le n^{-1}\operatorname{E}\Big|\langle Q_J\Pi_{\mathcal{H}_0}^\perp h, \psi^J\rangle_{L^2(X)}'(G_b^{-1/2}S)_l^-\Pi_J\Pi_{\mathcal{H}_0}^\perp h(X)\widetilde{b}^K(W)\Big|^2 \\
&\le 2n^{-1}\big\|\langle Q_J\Pi_{\mathcal{H}_0}^\perp h, \psi^J\rangle_{L^2(X)}'(G_b^{-1/2}S)_l^-\big\|^2\|\Pi_K T\Pi_{\mathcal{H}_0}^\perp h\|_{L^2(W)}^2 \\
&\quad + 2n^{-1}\big\|\langle Q_J\Pi_{\mathcal{H}_0}^\perp h, \psi^J\rangle_{L^2(X)}'(G_b^{-1/2}S)_l^-\big\|^2\|\Pi_K T(\Pi_{\mathcal{H}_0}^\perp h - \Pi_J\Pi_{\mathcal{H}_0}^\perp h)\|_{L^2(W)}^2 \\
&\lesssim n^{-1}\big\|\langle Q_J\Pi_{\mathcal{H}_0}^\perp h, \psi^J\rangle_{L^2(X)}'(G_b^{-1/2}S)_l^-\big\|^2,
\end{aligned}
$$

where the second bound is due to the Cauchy-Schwarz inequality and the third bound is due to Assumption 2(iv). Consider A_{12}, we infer from Chen and Christensen (2018, Lemma F.10(c)) and Assumption 2(ii) that

$$
\begin{aligned}
|A_{12}|^2 \le& \big\|\langle Q_J\Pi_{\mathcal{H}_0}^\perp h, \psi^J\rangle_{L^2(X)}'(G_b^{-1/2}S)_l^-\big\|^2\Big\|G_b^{-1/2}S'\Big((\widehat{G}_b^{-1/2}\widehat{S})_l^-\widehat{G}_b^{-1/2}G_b^{1/2} - (G_b^{-1/2}S)_l^-\Big)\Big\|^2 \\
&\times\Big\|\frac{1}{n}\sum_i \Pi_J\Pi_{\mathcal{H}_0}^\perp h(X_i)b^K(W_i) - \operatorname{E}[\Pi_J\Pi_{\mathcal{H}_0}^\perp h(X)b^K(W)]\Big\|^2 \\
&\lesssim \big\|\langle Q_J\Pi_{\mathcal{H}_0}^\perp h, \psi^J\rangle_{L^2(X)}'(G_b^{-1/2}S)_l^-\big\|^2 \times n^{-1}s_J^{-2}\zeta_J^2(\log J)\times n^{-1}\zeta_J^2 \\
&\lesssim n^{-1}\big\|\langle Q_J\Pi_{\mathcal{H}_0}^\perp h, \psi^J\rangle_{L^2(X)}'(G_b^{-1/2}S)_l^-\big\|^2
\end{aligned}
$$

wpa1 uniformly for $h \in \mathcal{H}$. Next we consider the term A_2 of (E.1). Following the upper bound of A_{12} we obtain wpa1 uniformly for $h \in \mathcal{H}$:

$$
\begin{aligned}
&\Big|\operatorname{E}[\Pi_{\mathcal{H}_0}^\perp h(X)b^K(W)]'A'G(\widehat{A}-A)\operatorname{E}[(h - \Pi_{\mathcal{H}_0}h - \Pi_J\Pi_{\mathcal{H}_0}^\perp h)(X)b^K(W)]\Big|^2 \\
&\le \big\|\langle Q_J\Pi_{\mathcal{H}_0}^\perp h, \psi^J\rangle_{L^2(X)}'(G_b^{-1/2}S)_l^-\big\|^2\Big\|G_b^{-1/2}S\Big((\widehat{G}_b^{-1/2}\widehat{S})_l^-\widehat{G}_b^{-1/2}G_b^{1/2} - (G_b^{-1/2}S)_l^-\Big)\Big\|^2 \\
&\qquad \times\big\|\langle T(\Pi_{\mathcal{H}_0}^\perp h - \Pi_J\Pi_{\mathcal{H}_0}^\perp h), \widetilde{b}^K\rangle_{L^2(W)}\big\|^2 \\
&\lesssim \big\|\langle Q_J\Pi_{\mathcal{H}_0}^\perp h, \psi^J\rangle_{L^2(X)}'(G_b^{-1/2}S)_l^-\big\|^2\|\Pi_K T(\Pi_{\mathcal{H}_0}^\perp h - \Pi_J\Pi_{\mathcal{H}_0}^\perp h)\|_{L^2(W)}^2 \times n^{-1}s_J^{-2}\zeta_J^2(\log J) \\
&\lesssim n^{-1}\big\|\langle Q_J\Pi_{\mathcal{H}_0}^\perp h, \psi^J\rangle_{L^2(X)}'(G_b^{-1/2}S)_l^-\big\|^2
\end{aligned}
$$

using that $s_J^{-2}\|\Pi_K T(\Pi_{\mathcal{H}_0}^{\perp} h - \Pi_J \Pi_{\mathcal{H}_0}^{\perp} h)\|^2_{L^2(W)} \lesssim \|\Pi_{\mathcal{H}_0}^{\perp} h - \Pi_J \Pi_{\mathcal{H}_0}^{\perp} h\|^2_{L^2(X)}$ by Assumption 2(iv) and $\zeta_J^2(\log J)\|h - \Pi_J h\|^2_{L^2(X)} = O(1)$ by Assumption 2(iii). Finally, we obtain $|T_1| \leq |A_1| + |A_2| \lesssim n^{-1/2}\|\langle Q_J \Pi_{\mathcal{H}_0}^{\perp} h, \psi^J\rangle'_{L^2(X)}(G_b^{-1/2}S)_l^-\|$ wpa1 uniformly for $h \in \mathcal{H}$.

We next consider the term T_2 using the decomposition:

$$T_2 \leq 2\,\mathrm{E}[\Pi_J \Pi_{\mathcal{H}_0}^{\perp} h(X) b^K(W)]'(\widehat{A} - A)' G (\widehat{A} - A)\,\mathrm{E}[\Pi_J \Pi_{\mathcal{H}_0}^{\perp} h(X) b^K(W)] \\ + 2\,\mathrm{E}[\Pi_J^{\perp} \Pi_{\mathcal{H}_0}^{\perp} h(X) b^K(W)]'(\widehat{A} - A)' G (\widehat{A} - A)\,\mathrm{E}[\Pi_J^{\perp} \Pi_{\mathcal{H}_0}^{\perp} h(X) b^K(W)] =: 2T_{21} + 2T_{22}$$

where $\Pi_J^{\perp} = \mathrm{id} - \Pi_J$ is the projection. We first bound T_{21} using Assumption 2(ii):

$$\begin{aligned} T_{21} &\leq \Big| \langle \Pi_J \Pi_{\mathcal{H}_0}^{\perp} h, \psi^J \rangle'_{L^2(X)} \Big((\widehat{G}_b^{-1/2}\widehat{S})_l^- \widehat{G}_b^{-1/2} S - I_J \Big)' (\widehat{G}_b^{-1/2}\widehat{S})_l^- \widehat{G}_b^{-1/2} \\ &\qquad \times \Big(\frac{1}{n}\sum_i \Pi_J \Pi_{\mathcal{H}_0}^{\perp} h(X_i) \widetilde{b}^K(W_i) - \mathrm{E}[\Pi_J \Pi_{\mathcal{H}_0}^{\perp} h(X) \widetilde{b}^K(W)] \Big) \Big| \\ &\leq \big\| \langle \Pi_J \Pi_{\mathcal{H}_0}^{\perp} h, \psi^J \rangle_{L^2(X)} \big\| \|S - \widehat{S}\| \Big\| (\widehat{G}_b^{-1/2}\widehat{S})_l^- \widehat{G}_b^{-1/2} \Big\|^2 \\ &\qquad \times \Big\| \frac{1}{n}\sum_i \Pi_J \Pi_{\mathcal{H}_0}^{\perp} h(X_i) \widetilde{b}^K(W_i) - \mathrm{E}[\Pi_J \Pi_{\mathcal{H}_0}^{\perp} h(X) \widetilde{b}^K(W)] \Big\| \\ &\lesssim \|\Pi_J(h - \Pi_{\mathcal{H}_0} h)\|_{L^2(X)} n^{-1/2} s_J^{-2} \zeta_J \sqrt{\log J} \times n^{-1/2} s_J^{-1} \zeta_J \lesssim n^{-1/2} s_J^{-1} \|\Pi_J(h - \Pi_{\mathcal{H}_0} h)\|_{L^2(X)} \end{aligned}$$

wpa1 uniformly for $h \in \mathcal{H}$. For T_{22}, we note that uniformly in $h \in \mathcal{H}$, $\|\,\mathrm{E}[\Pi_J^{\perp} \Pi_{\mathcal{H}_0}^{\perp} h(X) \widetilde{b}^K(W)]\| = \|\Pi_K T(\Pi_J \Pi_{\mathcal{H}_0}^{\perp} h - \Pi_{\mathcal{H}_0}^{\perp} h)\|_{L^2(W)} \lesssim s_J J^{-p/d_x}$ by Assumption 2(iv). Thus, following the upper bound derivations of T_{21}, we obtain $T_{22} \lesssim n^{-1/2} s_J^{-1} J^{-p/d_x}$ wpa1 uniformly for $h \in \mathcal{H}$. □

Lemma E.2. *Under Assumptions 2(i) it holds for $\widetilde{h} \in \{h, \Pi_{\mathcal{H}_0} h\}$ that*

$$\sup_{J \in \mathcal{I}_n} \sup_{h \in \mathcal{H}} \lambda_{\max}\big(\mathrm{E}_h \big[(Y - \widetilde{h}(X))^2 \widetilde{b}^{K(J)}(W) \widetilde{b}^{K(J)}(W)' \big] \big) \leq \overline{\sigma}^2 < \infty.$$

Proof. We have for any $\gamma \in \mathbb{R}^K$ where $K = K(J)$ that

$$\begin{aligned} \gamma'\,\mathrm{E}_h \big[(Y - \widetilde{h}(X))^2 \widetilde{b}^K(W) \widetilde{b}^K(W)' \big] \gamma &\leq \mathrm{E}\Big[\mathrm{E}_h[(Y - \widetilde{h}(X))^2 | W] \big(\gamma' \widetilde{b}^K(W) \big)^2 \Big] \\ &\leq \overline{\sigma}^2\, \mathrm{E}\Big[\big(\gamma' \widetilde{b}^K(W) \big)^2 \Big] = \overline{\sigma}^2 \gamma' G_b^{-1/2}\,\mathrm{E}\big[b^K(W) b^K(W)' \big] G_b^{-1/2} \gamma = \overline{\sigma}^2 \|\gamma\|^2 \end{aligned}$$

uniformly for $h \in \mathcal{H}$ and $J \in \mathcal{I}_n$, where the second inequality is due to Assumption 2(i). □

Proof of Theorem B.1. From the definition of Q_J given in (B.1) we infer

$$\|Q_J(h - \Pi_{\mathcal{H}_0} h)\|^2_{L^2(X)} = \big\| A\,\mathrm{E}_h[(Y - \Pi_{\mathcal{H}_0} h(X)) b^K(W)] \big\|^2 = \|\,\mathrm{E}_h[U^J]\|^2$$

using the notation $U_i^J = (Y_i - \Pi_{\mathcal{H}_0} h(X_i)) A b^K(W_i)$. The definition of $\widehat{D}_J$ implies

$$\widehat{D}_J(\Pi_{\mathcal{H}_0} h) - \|Q_J(h - \Pi_{\mathcal{H}_0} h)\|^2_{L^2(X)} = \frac{1}{n(n-1)} \sum_{j=1}^{J} \sum_{i \neq i'} \big(U_{ij} U_{i'j} - \mathrm{E}_h[U_{1j}]^2\big) \tag{E.2}$$

$$+ \frac{1}{n(n-1)} \sum_{i \neq i'} \big(Y_i - \Pi_{\mathcal{H}_0} h(X_i)\big)\big(Y_{i'} - \Pi_{\mathcal{H}_0} h(X_{i'})\big) b^K(W_i)' \Big(A'A - \widehat{A}'\widehat{A}\Big) b^K(W_{i'}). \tag{E.3}$$

Consider the summand in (E.2), we observe

$$\Big| \sum_{j=1}^{J} \sum_{i \neq i'} \big(U_{ij} U_{i'j} - \mathrm{E}_h[U_{1j}]^2\big)\Big|^2 = \sum_{j,j'=1}^{J} \sum_{i \neq i'} \sum_{i'' \neq i'''} \big(U_{ij} U_{i'j} - \mathrm{E}_h[U_{1j}]^2\big)\big(U_{i''j'} U_{i'''j'} - \mathrm{E}_h[U_{1j'}]^2\big).$$

We distinguish three different cases. First: i, i', i'', i''' are all different, second: either $i = i''$ or $i' = i'''$, or third: $i = i'$ and $i' = i'''$. We thus calculate for each $j, j' \geq 1$ that

$$\begin{aligned}
&\sum_{i \neq i'} \sum_{i'' \neq i'''} \big(U_{ij} U_{i'j} - \mathrm{E}_h[U_{1j}]^2\big)\big(U_{i''j'} U_{i'''j'} - \mathrm{E}_h[U_{1j'}]^2\big) \\
&\qquad = \sum_{i,i',i'',i''' \text{all different}} \big(U_{ij} U_{i'j} - \mathrm{E}_h[U_{1j}]^2\big)\big(U_{i''j'} U_{i'''j'} - \mathrm{E}_h[U_{1j'}]^2\big) \\
&\qquad\quad + 2 \sum_{i \neq i' \neq i''} \big(U_{ij} U_{i'j} - \mathrm{E}_h[U_{1j}]^2\big)\big(U_{i''j'} U_{i'j'} - \mathrm{E}_h[U_{1j'}]^2\big) \\
&\qquad\quad + \sum_{i \neq i'} \big(U_{ij} U_{i'j} - \mathrm{E}_h[U_{1j}]^2\big)\big(U_{ij'} U_{i'j'} - \mathrm{E}_h[U_{1j'}]^2\big).
\end{aligned}$$

The expectation of the first term on the right hand side vanishes due to independent observations and thus, we have

$$\begin{aligned}
&\mathrm{E}_h \Big| \sum_{j=1}^{J} \sum_{i \neq i'} \big(U_{ij} U_{i'j} - \mathrm{E}_h[U_{1j}]^2\big)\Big|^2 \\
&= 2n(n-1)(n-2) \underbrace{\sum_{j,j'=1}^{J} \mathrm{E}_h \Big[\big(U_{1j} U_{2j} - \mathrm{E}_h[U_{1j}]^2\big)\big(U_{3j'} U_{2j'} - \mathrm{E}_h[U_{1j'}]^2\big)\Big]}_{I} \\
&+ n(n-1) \underbrace{\sum_{j,j'=1}^{J} \mathrm{E}_h \Big[\big(U_{1j} U_{2j} - \mathrm{E}_h[U_{1j}]^2\big)\big(U_{1j'} U_{2j'} - \mathrm{E}_h[U_{1j'}]^2\big)\Big]}_{II}.
\end{aligned}$$

Now using $\|(G_b^{-1/2} S G^{-1/2})_l^-\| = s_J^{-1}$ together with the notation $\widetilde{\psi}^J = G^{-1/2}\psi^J$ we obtain

$$\begin{aligned}
\big\|\langle Q_J(h - \Pi_{\mathcal{H}_0} h), \psi^J\rangle'_{L^2(X)} (G_b^{-1/2} S)_l^-\big\| &= \big\|\langle Q_J(h - \Pi_{\mathcal{H}_0} h), \widetilde{\psi}^J\rangle'_{L^2(X)} (G_b^{-1/2} S G^{-1/2})_l^-\big\| \\
&\leq s_J^{-1} \big\|\langle Q_J(h - \Pi_{\mathcal{H}_0} h), \widetilde{\psi}^J\rangle_{L^2(X)}\big\| \lesssim s_J^{-1}\big(\|h - \Pi_{\mathcal{H}_0} h\|_{L^2(X)} + J^{-p/d_x}\big),
\end{aligned} \tag{E.4}$$

where the last equation is due to Lemma B.1(i). Consequently, we bound the term I by

$$
\begin{aligned}
I &= \sum_{j,j'=1}^{J} \mathrm{E}_h[U_{1j}]\,\mathrm{E}_h[U_{1j'}]\,\mathbb{C}ov_h(U_{1j}, U_{1j'}) = \mathrm{E}_h[U_1^J]'\,\mathbb{C}ov_h(U_1^J, U_1^J)\,\mathrm{E}_h[U_1^J] \\
&\leq \lambda_{\max}\big(\mathbb{V}ar_h((Y - \Pi_{\mathcal{H}_0} h(X))\widetilde{b}^K(W))\big)\big\|(G_b^{-1/2} S G^{-1/2})_l^- \,\mathrm{E}_h[U_1^J]\big\|^2 \\
&\leq \overline{\sigma}^2 \Big\| \Big((G_b^{-1/2}S)_l^- \,\mathrm{E}_h[(Y - \Pi_{\mathcal{H}_0} h(X))\widetilde{b}^K(W)]\Big)' G (G_b^{-1/2}S)_l^- \Big\|^2 \\
&= \overline{\sigma}^2 \Big\| \langle Q_J(h - \Pi_{\mathcal{H}_0} h), \psi^J \rangle'_{L^2(X)} (G_b^{-1/2}S)_l^- \Big\|^2 \lesssim s_J^{-2}\big(\|h - \Pi_{\mathcal{H}_0} h\|^2_{L^2(X)} + J^{-2p/d_x}\big),
\end{aligned}
$$

using $U_i^J = (Y_i - \Pi_{\mathcal{H}_0} h(X_i))(G_b^{-1/2} S G^{-1/2})_l^- \widetilde{b}^K(W_i)$ and Lemma E.2. For term II we observe

$$
II = \sum_{j,j'=1}^{J} \mathrm{E}_h[U_{1j}U_{1j'}]^2 - \Big(\sum_{j=1}^{J} \mathrm{E}_h[U_{1j}]^2\Big)^2 \leq \sum_{j,j'=1}^{J} \mathrm{E}_h[U_{1j}U_{1j'}]^2 = V_J^2.
$$

Thus, the upper bounds derived for the terms I and II imply for all $n \geq 2$:

$$
\mathrm{E}_h \Big| \frac{1}{n(n-1)} \sum_{j=1}^{J} \sum_{i \neq i'} \big(U_{ij}U_{i'j} - \mathrm{E}_h[U_{1j}]^2\big)\Big|^2 \lesssim \frac{\|h - \Pi_{\mathcal{H}_0} h\|^2_{L^2(X)} + J^{-2p/d_x}}{n s_J^2} + \frac{V_J^2}{n^2}. \quad \text{(E.5)}
$$

Thus equality (E.3) implies the result by employing Lemma B.2 and Lemma E.1. □

Proof of Lemma A.1. By Lemma E.1 and the decomposition (E.2–E.3) we obtain

$$
\mathrm{P}_{h_0}\left(\frac{n\widehat{D}_J(h_0)}{V_J} > \eta_J(\alpha)\right) = \mathrm{P}_{h_0}\left(\frac{1}{V_J(n-1)} \sum_{j=1}^{J} \sum_{i \neq i'} U_{ij}U_{i'j} > \eta_J(\alpha)\right) + o(1).
$$

Using the martingale central limit theorem (see, e.g., Breunig (2020, Lemma A.3)) we obtain

$$
\mathrm{P}_{h_0}\left(\frac{1}{\sqrt{2}V_J(n-1)} \sum_{j=1}^{J} \sum_{i \neq i'} U_{ij}U_{i'j} > z_{1-\alpha}\right) = \alpha + o(1),
$$

where $z_{1-\alpha}$ denotes the $(1-\alpha)$-quantile of the standard normal distribution. Further, Lemma B.4(i) implies $V_J/\widehat{V}_J = 1$ wpa1 uniformly for $h \in \mathcal{H}$ and since $\eta_J(\alpha)/\sqrt{2} = \frac{q(\alpha,J)-J}{\sqrt{2J}}$ converges to $z_{1-\alpha}$ as J tends to infinity, the result follows. □

Proof of Lemma B.1. Proof of (i): Using the notation $\widetilde{b}^K(\cdot) := G_b^{-1/2} b^K(\cdot)$, we observe for all $h \in \mathcal{H}$ that

$$
\begin{aligned}
\|Q_J(h - \Pi_{\mathcal{H}_0} h)\|_{L^2(X)} &= \big\|(G_b^{-1/2} S G^{-1/2})_l^- \operatorname{E}[\widetilde{b}^K(W)(h - \Pi_{\mathcal{H}_0} h)(X)]\big\| \\
&\leq \big\|(G_b^{-1/2} S G^{-1/2})_l^- \operatorname{E}[\widetilde{b}^K(W)(\Pi_J h - \Pi_J \Pi_{\mathcal{H}_0} h)(X)]\big\| \\
&\quad + \big\|(G_b^{-1/2} S G^{-1/2})_l^- \operatorname{E}[\widetilde{b}^K(W)((h - \Pi_{\mathcal{H}_0} h)(X) - (\Pi_J h - \Pi_J \Pi_{\mathcal{H}_0} h)(X))]\big\| \\
&\leq \|\Pi_J h - \Pi_J \Pi_{\mathcal{H}_0} h\|_{L^2(X)} + s_J^{-1} \|\Pi_K T((h - \Pi_{\mathcal{H}_0} h) - (\Pi_J h - \Pi_J \Pi_{\mathcal{H}_0} h))\|_{L^2(W)} \\
&\leq \|\Pi_J h - \Pi_J \Pi_{\mathcal{H}_0} h\|_{L^2(X)} + O\big(J^{-p/d_x}\big)
\end{aligned}
$$

by Assumption 2(iv).

Proof of (ii): We observe $\|Q_J h - h\|_{L^2(X)} \leq \|Q_J(h - \Pi_J h)\|_{L^2(X)} + \|\Pi_J h - h\|_{L^2(X)}$. The result thus follows by replacing $\Pi_{\mathcal{H}_0} h$ with $\Pi_J h$ in the derivation of (i). $\square$

Proof of Lemma B.2. For any $J \times J$ matrix M it holds $\|M\|_F \leq \sqrt{J}\|M\|$ and hence

$$
\begin{aligned}
V_J^2 &= \left\| \left(G_b^{-1/2} S G^{-1/2}\right)_l^- \operatorname{E}_h\left[(Y - h(X))^2 \widetilde{b}^K(W) \widetilde{b}^K(W)'\right] \left(G_b^{-1/2} S G^{-1/2}\right)_l^- \right\|_F^2 \\
&\leq J \left\| \left(G_b^{-1/2} S G^{-1/2}\right)_l^- \right\|^4 \left\| \operatorname{E}_h\left[(Y - h(X))^2 \widetilde{b}^K(W) \widetilde{b}^K(W)'\right] \right\|^2.
\end{aligned}
$$

The result now follows from $\|(G_b^{-1/2} S G^{-1/2})_l^-\| = s_J^{-1}$ and Lemma E.2. $\square$

Proof of Lemma B.3. In the following, let e_j be the unit vector with 1 at the j–th position. Introduce a unitary matrix Q such that by Schur decomposition $Q' A G_b A' Q = \operatorname{diag}(s_1^{-2}, \ldots, s_J^{-2})$. We make use of the notation $\widetilde{U}_i^J = (Y_i - h(X_i)) Q' A b^K(W_i)$. Now since the Frobenius norm is invariant under unitary matrix multiplication we have

$$
V_J^2 = \sum_{j,j'=1}^J \operatorname{E}_h[\widetilde{U}_{1j} \widetilde{U}_{1j'}]^2 \geq \sum_{j=1}^J \operatorname{E}_h[\widetilde{U}_{1j}^2]^2 = \sum_{j=1}^J \big(\operatorname{E}_h |(Y - h(X)) e_j' Q' A b^K(W)|^2\big)^2.
$$

Consequently, using the lower bound $\inf_{w \in \mathcal{W}} \inf_{h \in \mathcal{H}} \operatorname{E}_h[(Y - h(X))^2 | W = w] \geq \underline{\sigma}^2$ by Assumption 1(i), we obtain uniformly for $h \in \mathcal{H}$:

$$
\begin{aligned}
V_J^2 &\geq \underline{\sigma}^4 \sum_{j=1}^J \big(\operatorname{E}[e_j' Q' A b^K(W) b^K(W)' A' Q e_j]\big)^2 = \underline{\sigma}^4 \sum_{j=1}^J \big(e_j' Q' A G_b A' Q e_j\big)^2 \\
&= \underline{\sigma}^4 \sum_{j=1}^J \big(e_j' \operatorname{diag}(s_1^{-2}, \ldots, s_J^{-2}) e_j\big)^2 \geq \underline{\sigma}^4 \sum_{j=1}^J s_j^{-4},
\end{aligned}
$$

which proves the result. $\square$

Recall the definition $\mathcal{C}_h = \max_{\mathsf{e} \in \mathcal{S}^{K\circ}} \int_0^1 \big(1 + \log N_{[]}\,\big(\epsilon \|F_{h,\mathsf{e}}\|_{L^2(Z)}, \mathcal{F}_{h,\mathsf{e}}, L^2(Z)\big)\big)^{1/2} d\epsilon$.

Lemma E.3. *Let Assumptions 1(ii)-(iii), 2(i), 4(i)(iii), and 5(ii) hold. Then, for $J = J^\circ$, we have wpa1 uniformly for $h \in \mathcal{H}_1(\delta^\circ \mathsf{r}_n)$:*

$$\left| \frac{1}{n(n-1)} \sum_{i \neq i'} U_i(\widehat{h}_J^{\mathrm{R}}) U_{i'}(\widehat{h}_J^{\mathrm{R}}) a_{J,ii'} - \mathrm{E}_h \left[U_i(\widehat{h}_J^{\mathrm{R}}) U_{i'}(\widehat{h}_J^{\mathrm{R}}) a_{J,ii'} \right] \right| \lesssim n^{-1/2} s_J^{-1} \mathcal{C}_h \big(\|h - \mathcal{H}_0\|_{L^2(X)} + J^{-p/d_x} \big) + n^{-1} s_J^{-2} \sqrt{J},$$

where $U_i(\phi) = Y_i - \phi(X_i)$ and $a_{J,ii'} = b^K(W_i)' A' A b^K(W_{i'})$.

Proof. For simplicity of notation, we write J instead of J° throughout the proof. We observe for all $h \in \mathcal{H}_1(\delta^\circ \mathsf{r}_n)$ that

$$\begin{aligned}
&\frac{1}{n(n-1)} \sum_{i \neq i'} U_i(\widehat{h}_J^{\mathrm{R}}) U_{i'}(\widehat{h}_J^{\mathrm{R}}) a_{J,ii'} - \mathrm{E}_h \left[U_i(\widehat{h}_J^{\mathrm{R}}) U_{i'}(\widehat{h}_J^{\mathrm{R}}) a_{J,ii'} \right] \\
&= \frac{1}{n(n-1)} \sum_{i \neq i'} U_i(\Pi_{\mathcal{H}_0} h) U_{i'}(\Pi_{\mathcal{H}_0} h) a_{J,ii'} - \mathrm{E}_h \left[U_i(\Pi_{\mathcal{H}_0} h) U_{i'}(\Pi_{\mathcal{H}_0} h) a_{J,ii'} \right] \\
&+ \frac{2}{n(n-1)} \sum_{i \neq i'} U_i(\Pi_{\mathcal{H}_0} h)(\Pi_{\mathcal{H}_0} h - \widehat{h}_J^{\mathrm{R}})(X_{i'}) a_{J,ii'} - \mathrm{E}_h \left[U_i(\Pi_{\mathcal{H}_0} h)(\Pi_{\mathcal{H}_0} h - \widehat{h}_J^{\mathrm{R}})(X_{i'}) a_{J,ii'} \right] \\
&+ \frac{1}{n(n-1)} \sum_{i \neq i'} (\Pi_{\mathcal{H}_0} h - \widehat{h}_J^{\mathrm{R}})(X_i)(\Pi_{\mathcal{H}_0} h - \widehat{h}_J^{\mathrm{R}})(X_{i'}) a_{J,ii'} - \mathrm{E}_h \left[(\Pi_{\mathcal{H}_0} h - \widehat{h}_J^{\mathrm{R}})(X_i)(\Pi_{\mathcal{H}_0} h - \widehat{h}_J^{\mathrm{R}})(X_{i'}) a_{J,ii'} \right] \\
&=: T_1 + 2\, T_2 + T_3.
\end{aligned}$$

From the proof of Theorem B.1 we conclude $\sup_{h \in \mathcal{H}_1(\delta^\circ \mathsf{r}_n)} \mathrm{E}_h |T_1| \lesssim n^{-1} s_J^{-2} \sqrt{J}$. Consider T_2. Below, we let $a_i^J = A b^K(W_i) = (G_b^{-1/2} S G^{-1/2})_\ell^- \widetilde{b}^K(W_i)$. By Assumption 5(ii) $\sup_{h \in \mathcal{H}_1(\delta^\circ \mathsf{r}_n)} \mathrm{P}_h \big(\zeta_J \mathcal{C}_h \|\widehat{h}_J^{\mathrm{R}} - \Pi_{\mathcal{H}_0} h\|_{L^2(X)} > C \big) \to 0$ and consequently may assume that $\widehat{h}_J^{\mathrm{R}} \in \mathcal{H}_{0,J}(h) := \{ \|\phi - \Pi_{\mathcal{H}_0} h\|_{L^2(X)} \le [\zeta_J \mathcal{C}_h]^{-1} : \phi \in \mathcal{H}_{0,J} \}$. We have for all $h \in \mathcal{H}_1(\delta^\circ \mathsf{r}_n)$ that the absolute value of T_2 is bounded by

$$\begin{aligned}
&\sup_{\phi \in \mathcal{H}_{0,J}(h)} \left| \frac{1}{n(n-1)} \sum_{i \neq i'} \Big(U_i(\Pi_{\mathcal{H}_0} h) a_i^J - \mathrm{E}_h [U(\Pi_{\mathcal{H}_0} h) a^J] \Big)' \Big((\Pi_{\mathcal{H}_0} h - \phi)(X_{i'}) a_{i'}^J - \mathrm{E}\left[(\Pi_{\mathcal{H}_0} h - \phi)(X) a^J \right] \Big) \right| \\
&+ \left| \frac{1}{n} \sum_i \Big(U_i(\Pi_{\mathcal{H}_0} h) a_i^J - \mathrm{E}_h [U(\Pi_{\mathcal{H}_0} h) a^J] \Big)' \mathrm{E}\left[(\Pi_{\mathcal{H}_0} h - \widehat{h}_J^{\mathrm{R}})(X) a^J \right] \Big) \right| \\
&+ \sup_{\phi \in \mathcal{H}_{0,J}(h)} \left| \frac{1}{n} \sum_i \Big((\Pi_{\mathcal{H}_0} h - \phi)(X_i) a_i^J - \mathrm{E}\left[(\Pi_{\mathcal{H}_0} h - \phi)(X) a^J \right] \Big)' \mathrm{E}_h [U(\Pi_{\mathcal{H}_0} h) a^J] \right| =: T_{21} + T_{22} + T_{23}.
\end{aligned}$$

Below we let $a_{k,i} = b^K(W_i)' A' A G_b^{1/2} e_k$. Note that $\mathrm{E}\, \|a_{k,i}\|^2 \le \|(G_b^{-1/2} S G^{-1/2})_\ell^-\|^4 = s_J^{-4}$ for all $k = 1, \ldots, K$. We obtain uniformly for $h \in \mathcal{H}_1(\delta^\circ \mathsf{r}_n)$ by van der Vaart and Wellner

(2000, Theorem 2.14.2) that

$$
\begin{aligned}
\mathrm{E}_h T_{21} &\leq \sum_{k=1}^{K} \mathrm{E}_h \Big| \frac{1}{n} \sum_i U_i(\Pi_{\mathcal{H}_0} h) a_{k,i} - \mathrm{E}_h[U(\Pi_{\mathcal{H}_0} h) a_k] \Big| \\
&\qquad \times \mathrm{E}_h \sup_{\phi \in \mathcal{H}_{0,J}(h)} \Big| \frac{1}{n-1} \sum_{i'} (\Pi_{\mathcal{H}_0} h - \phi)(X_{i'}) \widetilde{b}_k(W_{i'}) - \mathrm{E}_h \big[(\Pi_{\mathcal{H}_0} h - \phi)(X) \widetilde{b}_k(W) \big] \Big| \\
&\lesssim \frac{\mathcal{C}_h}{n} \sqrt{\sum_{k=1}^{K} \mathrm{E}_h \Big[|U_i(\Pi_{\mathcal{H}_0} h)|^2 \| G_b^{-1/2} A' A b^K(W) \|^2 \Big]} \sqrt{\sum_{k=1}^{K} \mathrm{E}_h \sup_{\phi \in \mathcal{H}_{0,J}(h)} |(\Pi_{\mathcal{H}_0} h - \phi)(X) \widetilde{b}_k(W)|^2} \\
&\lesssim \frac{\mathcal{C}_h}{n} \overline{\sigma} s_J^{-2} \sqrt{J} \zeta_J \| \Pi_{\mathcal{H}_0} h - \Phi_J \|_{L^2(X)} \lesssim n^{-1} s_J^{-2} \sqrt{J}
\end{aligned}
$$

for some $\Phi_J \in \mathcal{H}_{0,J}(h)$ and using that $\mathrm{E}_h[|U(\Pi_{\mathcal{H}_0} h)|^2 | W] \leq \overline{\sigma}^2$ by Assumption 2(i). Further, we evaluate uniformly for $h \in \mathcal{H}_1(\delta^\circ \mathsf{r}_n)$:

$$
\begin{aligned}
\mathrm{E}_h T_{22} &= \overline{\sigma} n^{-1/2} \sqrt{\mathrm{E} \Big| (a^J)' \mathrm{E}_h \big[\big(\Pi_{\mathcal{H}_0} h - \widehat{h}_J^{\mathrm{R}} \big)(X) a^J \big] \Big|^2} \\
&\leq \overline{\sigma} n^{-1/2} s_J^{-2} \sup_{\phi \in \mathcal{H}_{0,J}(h)} \| \Pi_K T \big(\Pi_{\mathcal{H}_0} h - \phi \big) \|_{L^2(W)} \lesssim n^{-1/2} s_J^{-1} \big(\| h - \mathcal{H}_0 \|_{L^2(X)} + J^{-p/d_x} \big),
\end{aligned}
$$

where in the last equation, we used Assumption 2(iv) and $\| h - \Pi_{\mathcal{H}_0} h \|_{L^2(X)} = \| h - \mathcal{H}_0 \|_{L^2(X)}$. Consider T_{23}. Below, we make use of the relation $\mathrm{E}[U(\Pi_{\mathcal{H}_0} h) a^J]' a_i^J = \langle Q_J(h - \Pi_{\mathcal{H}_0} h), \psi^J \rangle'_{L^2(X)} (G_b^{-1/2} S)_\ell^- \widetilde{b}^K(W_i)$ and obtain uniformly for $h \in \mathcal{H}_1(\delta^\circ \mathsf{r}_n)$:

$$
\begin{aligned}
\mathrm{E}_h T_{23} &\leq \big\| \langle Q_J(h - \Pi_{\mathcal{H}_0} h), \psi^J \rangle'_{L^2(X)} (G_b^{-1/2} S)_\ell^- \big\| \\
&\quad \times \mathrm{E}_h \sup_{\mathsf{e} \in \mathcal{S}^{K^\circ - 1}} \sup_{\phi \in \mathcal{H}_{0,J}(h)} \Big| \frac{1}{n} \sum_i (\Pi_{\mathcal{H}_0} h - \phi)(X_i) \widetilde{b}^K(W_i)' \mathsf{e} - \mathrm{E} \big[(\Pi_{\mathcal{H}_0} h - \phi)(X) \widetilde{b}^K(W)' \mathsf{e} \big] \Big| \\
&\lesssim \big\| \langle Q_J(h - \Pi_{\mathcal{H}_0} h), \psi^J \rangle'_{L^2(X)} (G_b^{-1/2} S)_\ell^- \big\| \times \mathcal{C}_h n^{-1/2} \zeta_J \| \Pi_{\mathcal{H}_0} h - \Phi_J \|_{L^2(X)} \\
&\lesssim \mathcal{C}_h n^{-1/2} s_J^{-1} \big(\| h - \mathcal{H}_0 \|_{L^2(X)} + J^{-p/d_x} \big),
\end{aligned}
$$

where we used that $\sup_w |\widetilde{b}^K(w)' \mathsf{e}| \leq \zeta_J$ for all $\mathsf{e} \in \mathcal{S}^{K^\circ}$. Consider T_3. We have

$$
\begin{aligned}
|T_3| \leq & \sup_{\phi \in \mathcal{H}_{0,J}(h)} \Big| \frac{1}{n(n-1)} \sum_{i \neq i'} \Big((\Pi_{\mathcal{H}_0} h - \phi)(X_i) a_i^J - \mathrm{E} \big[(\Pi_{\mathcal{H}_0} h - \phi)(X) a^J \big] \Big)' \\
&\qquad \times \Big((\Pi_{\mathcal{H}_0} h - \phi)(X_{i'}) a_{i'}^J - \mathrm{E} \big[(\Pi_{\mathcal{H}_0} h - \phi)(X) a^J \big] \Big) \Big| \\
& + 2 \sup_{\phi \in \mathcal{H}_{0,J}(h)} \Big| \frac{1}{n} \sum_i \Big((\Pi_{\mathcal{H}_0} h - \phi)(X_i) a_i^J - \mathrm{E} \big[(\Pi_{\mathcal{H}_0} h - \phi)(X) a^J \big] \Big)' \mathrm{E} \big[(\Pi_{\mathcal{H}_0} h - \phi)(X) a^J \big] \Big| \\
& =: T_{31} + T_{32}.
\end{aligned}
$$

We evaluate for the first term on the right hand side that uniformly for $h \in \mathcal{H}_1(\delta^\circ \mathsf{r}_n)$:

$$
\begin{aligned}
\mathrm{E}\, T_{31} &\leq s_J^{-2} \sum_{k=1}^{K} \Big(\mathrm{E} \sup_{\phi \in \mathcal{H}_{0,J}(h)} \Big| \frac{1}{n} \sum_i (\Pi_{\mathcal{H}_0} h - \phi)(X_i) \widetilde{b}_k(W_i) - \mathrm{E}\left[(\Pi_{\mathcal{H}_0} h - \phi)(X) \widetilde{b}_k(W) \right] \Big| \Big)^2 \\
&\lesssim \frac{\mathcal{C}_h^2}{n s_J^2} \mathrm{E} \sup_{\phi \in \mathcal{H}_{0,J}(h)} \| (\Pi_{\mathcal{H}_0} h - \phi)(X) \widetilde{b}^K(W) \|^2 \lesssim \frac{\mathcal{C}_h^2}{n s_J^2} \zeta_J^2 \| \Pi_{\mathcal{H}_0} h - \Phi_J \|_{L^2(X)}^2 \lesssim \frac{\sqrt{J}}{n s_J^2},
\end{aligned}
$$

for some $\Phi_J \in \mathcal{H}_{0,J}(h)$ and using that $\mathcal{C}_h^2 \lesssim \sqrt{J}$. Further, we have $\mathrm{E}[(\Pi_{\mathcal{H}_0} h - \phi)(X) a^J]' a_i^J = \langle Q_J(\Pi_{\mathcal{H}_0} h - \phi), \psi^J \rangle'_{L^2(X)} (G_b^{-1/2} S)_\ell^- \widetilde{b}^K(W_i)$ and thus, following the derivation of the bound of T_{23}, we obtain

$$
\begin{aligned}
\mathrm{E}\, T_{32} &\leq \sup_{\phi \in \mathcal{H}_{0,J}(h)} \big\| \langle Q_J(\phi - \Pi_{\mathcal{H}_0} h), \psi^J \rangle'_{L^2(X)} (G_b^{-1/2} S)_\ell^- \big\| \\
&\quad \times \mathrm{E} \sup_{\mathsf{e} \in \mathcal{S}^{K^\circ}} \sup_{\phi \in \mathcal{H}_{0,J}(h)} \Big| \frac{1}{n} \sum_i (\phi - \Pi_{\mathcal{H}_0} h)(X_i) \widetilde{b}^K(W_i)' \mathsf{e} - \mathrm{E}\left[(\phi - \Pi_{\mathcal{H}_0} h)(X) \widetilde{b}^K(W)' \mathsf{e} \right] \Big| \\
&\lesssim n^{-1/2} s_J^{-1} \mathcal{C}_h \big(\| h - \Pi_{\mathcal{H}_0} h \|_{L^2(X)} + J^{-p/d_x} \big)
\end{aligned}
$$

uniformly for $h \in \mathcal{H}_1(\delta^\circ \mathsf{r}_n)$, where the last equation is due to Assumption 5(ii). Finally, the result follows from an application of Markov's inequality. $\square$

Lemma E.4. *Let Assumptions 1(ii)-(iii), 2(i), 4(i)(iii), and 5(ii) hold. Then, for $J = J^\circ$, we have wpa1 uniformly for $h \in \mathcal{H}_1(\delta^\circ \mathsf{r}_n)$:*

$$
\begin{aligned}
&\frac{1}{n(n-1)} \sum_{i \neq i'} \big(Y_i - \widehat{h}_J^{\mathrm{R}}(X_i) \big) \big(Y_{i'} - \widehat{h}_J^{\mathrm{R}}(X_{i'}) \big) b^K(W_i)' \Big(A'A - \widehat{A}'\widehat{A} \Big) b^K(W_{i'}) \\
&\qquad\qquad \lesssim n^{-1/2} s_J^{-1} \mathcal{C}_h \big(\| h - \mathcal{H}_0 \|_{L^2(X)} + J^{-p/d_x} \big) + n^{-1} s_J^{-2} \sqrt{J}.
\end{aligned}
$$

Proof. For simplicity of notation, we write J instead of J° throughout the proof. Following the proof of Lemma E.1 it is sufficient to control

$$
\begin{aligned}
&\mathrm{E}_h[(h - \widehat{h}_J^{\mathrm{R}})(X) b^K(W)]' \Big(A'A - \widehat{A}'\widehat{A} \Big) \mathrm{E}_h[(h - \widehat{h}_J^{\mathrm{R}})(X) b^K(W)] \\
&\quad = 2\, \mathrm{E}_h[(h - \widehat{h}_J^{\mathrm{R}})(X) b^K(W)]' A'(A - \widehat{A})\, \mathrm{E}_h[(h - \widehat{h}_J^{\mathrm{R}})(X) b^K(W)] \\
&\qquad - \mathrm{E}_h[(h - \widehat{h}_J^{\mathrm{R}})(X) b^K(W)]' (A - \widehat{A})' (A - \widehat{A})\, \mathrm{E}_h[(h - \widehat{h}_J^{\mathrm{R}})(X) b^K(W)] =: 2T_1 - T_2,
\end{aligned}
$$

We first consider the term T_1 using the decomposition:

$$
\begin{aligned}
T_1 &= \mathrm{E}_h[(h - \widehat{h}_J^{\mathrm{R}})(X) b^K(W)]' A'(\widehat{A} - A)\, \mathrm{E}_h[\Pi_J(h - \widehat{h}_J^{\mathrm{R}})(X) b^K(W)] \\
&\quad + \mathrm{E}_h[(h - \widehat{h}_J^{\mathrm{R}})(X) b^K(W)]' A'(\widehat{A} - A)\, \mathrm{E}_h[(h - \widehat{h}_J^{\mathrm{R}} - \Pi_J(h - \widehat{h}_J^{\mathrm{R}}))(X) b^K(W)]. \quad \text{(E.6)}
\end{aligned}
$$

Consider the first summand on the right hand side of equation (E.6). By Assumption 5(ii) $\sup_{h \in \mathcal{H}_1(\delta^\circ \mathsf{r}_n)} \mathrm{P}_h\big(\zeta_J \mathcal{C}_h \| \widehat{h}_J^{\mathrm{R}} - \Pi_{\mathcal{H}_0} h \|_{L^2(X)} > C \big) \to 0$ and consequently may assume that

$\widehat{h}_J^{\text{R}} \in \mathcal{H}_{0,J}(h) := \{\phi \in \mathcal{H}_{0,J} : \|\phi - \Pi_{\mathcal{H}_0} h\|_{L^2(X)} \leq [\zeta_J \mathcal{C}_h]^{-1}\}$. We calculate

$$
\begin{aligned}
&\sup_{\phi\in\mathcal{H}_{0,J}(h)} \Big|\big((G_b^{-1/2}S)_l^- \operatorname{E}[(h-\phi)(X)\widetilde{b}^K(W)]\big)' G\big((G_b^{-1/2}S)_l^- - (\widehat{G}_b^{-1/2}\widehat{S})_l^- \widehat{G}_b^{-1/2}G_b^{1/2}\big) \operatorname{E}[(h-\phi)(X)\widetilde{b}^K(W)]\Big| \\
&= \sup_{\phi\in\mathcal{H}_{0,J}(h)} \Big|\langle Q_J(h-\phi), \psi^J\rangle'_{L^2(X)} (G_b^{-1/2}S)_l^- \Big(\frac{1}{n}\sum_i \Pi_J(h-\phi)(X_i)\widetilde{b}^K(W_i) - \operatorname{E}[\Pi_J(h-\phi)(X)\widetilde{b}^K(W)]\Big)\Big| \\
&\quad + \sup_{\phi\in\mathcal{H}_{0,J}(h)} \Big|\langle Q_J(h-\phi), \psi^J\rangle'_{L^2(X)} (G_b^{-1/2}S)_l^- G_b^{-1/2} S\Big((\widehat{G}_b^{-1/2}\widehat{S})_l^- \widehat{G}_b^{-1/2}G_b^{1/2} - (G_b^{-1/2}S)_l^-\Big) \\
&\qquad\qquad \times \Big(\frac{1}{n}\sum_i \Pi_J(h-\phi)(X_i)\widetilde{b}^K(W_i) - \operatorname{E}[\Pi_J(h-\phi)(X)\widetilde{b}^K(W)]\Big)\Big| =: T_{11} + T_{12}.
\end{aligned}
$$

Consider T_{11}, which coincides with the term T_{32} in the proof of Lemma E.3 and thus, we have $\operatorname{E}|T_{11}| \lesssim n^{-1/2} s_J^{-1} \mathcal{C}_h\big(\|h - \mathcal{H}_0\|_{L^2(X)} + J^{-p/d_x}\big)$. To establish an upper bound for T_{12} we infer from Chen and Christensen (2018, Lemma F.10(c)) that

$$
\begin{aligned}
|T_{12}|^2 &\leq \sup_{\phi\in\mathcal{H}_{0,J}(h)} \big\|\langle Q_J(h-\phi), \psi^J\rangle'_{L^2(X)} (G_b^{-1/2}S)_l^-\big\|^2 \\
&\quad \times \Big\|G_b^{-1/2}S\Big((\widehat{G}_b^{-1/2}\widehat{S})_l^- \widehat{G}_b^{-1/2}G_b^{1/2} - (G_b^{-1/2}S)_l^-\Big)\Big\|^2 \\
&\quad \times \sup_{\phi\in\mathcal{H}_{0,J}(h)} \Big\|\frac{1}{n}\sum_i \Pi_J(h-\phi)(X_i)b^K(W_i) - \operatorname{E}[\Pi_J(h-\phi)(X)b^K(W)]\Big\|^2 \\
&\lesssim \sup_{\phi\in\mathcal{H}_{0,J}(h)} \big\|\langle Q_J(h-\phi), \psi^J\rangle'_{L^2(X)} (G_b^{-1/2}S)_l^-\big\|^2 \times n^{-1}s_J^{-2}\zeta_J^2(\log J) \times n^{-1}\zeta_J^2\mathcal{C}_h^2 \\
&\lesssim n^{-1}s_J^{-2}\mathcal{C}_h^2\big(\|h - \Pi_{\mathcal{H}_0}h\|_{L^2(X)}^2 + J^{-2p/d_x}\big)
\end{aligned}
$$

wpa1 uniformly for $h \in \mathcal{H}_1(\delta^\circ \mathsf{r}_n)$, where the last equation is due to $s_J^{-1}\zeta_J^2\sqrt{(\log J)/n} = O(1)$ from Assumption 4(i). Consider the second summand on the right hand side of equation (E.6). Following the upper bound of T_{12} we obtain

$$
\begin{aligned}
&\sup_{\phi\in\mathcal{H}_{0,J}(h)} \Big|\operatorname{E}[(h-\phi)(X)b^K(W)]'A'G(\widehat{A} - A)\operatorname{E}[(h-\phi-\Pi_J(h-\phi))(X)b^K(W)]\Big|^2 \\
&\leq \sup_{\phi\in\mathcal{H}_{0,J}(h)} \big\|\langle Q_J(h-\phi), \psi^J\rangle'_{L^2(X)} (G_b^{-1/2}S)_l^-\big\|^2 \Big\|G_b^{-1/2}S'\Big((\widehat{G}_b^{-1/2}\widehat{S})_l^- \widehat{G}_b^{-1/2}G_b^{1/2} - (G_b^{-1/2}S)_l^-\Big)\Big\|^2 \\
&\quad \times \sup_{\phi\in\mathcal{H}_{0,J}(h)} \big\|\langle T(h-\phi-\Pi_J(h-\phi)), \widetilde{b}^K\rangle_{L^2(W)}\big\|^2 \\
&\lesssim \sup_{\phi\in\mathcal{H}_{0,J}(h)} \big\|\langle Q_J(h-\phi), \psi^J\rangle'_{L^2(X)} (G_b^{-1/2}S)_l^-\big\|^2 \sup_{\phi\in\mathcal{H}_{0,J}(h)} \|\Pi_K T(h-\phi-\Pi_J(h-\phi))\|_{L^2(W)}^2 \\
&\quad \times n^{-1}s_J^{-2}\zeta_J^2(\log J) \\
&\lesssim n^{-1}s_J^{-2}\big(\|h - \Pi_{\mathcal{H}_0}h\|_{L^2(X)}^2 + J^{-2p/d_x}\big)
\end{aligned}
$$

wpa1 uniformly for $h \in \mathcal{H}_1(\delta^\circ \mathsf{r}_n)$, using that $s_J^{-2}\|\Pi_K T(h - \Pi_{\mathcal{H}_0}h - \Pi_J(h - \Pi_{\mathcal{H}_0}h))\|_{L^2(W)}^2 \lesssim \|h - \Pi_{\mathcal{H}_0}h - \Pi_J(h - \Pi_{\mathcal{H}_0}h)\|_{L^2(X)}^2$ by Assumption 4(i) and $\zeta_J^2(\log J)\|h - \Pi_J h\|_{L^2(X)}^2 = O(1)$ by Assumption 4(iii).

We now consider the term T_2 using the decomposition

$$\begin{aligned}
T_2 &\leq 2 \sup_{\phi \in \mathcal{H}_{0,J}(h)} \Big| \mathrm{E}[\Pi_J(h-\phi)(X)b^K(W)]'(\widehat{A}-A)'G(\widehat{A}-A)\,\mathrm{E}[\Pi_J(h-\phi)(X)b^K(W)]\Big| \\
&\quad + 2 \sup_{\phi \in \mathcal{H}_{0,J}(h)} \Big| \mathrm{E}[\Pi_J^{\perp}(h-\phi)(X)b^K(W)]'(\widehat{A}-A)'G(\widehat{A}-A)\,\mathrm{E}[\Pi_J^{\perp}(h-\phi)(X)b^K(W)]\Big| \\
&=: 2T_{21} + 2T_{22},
\end{aligned}$$

where $\Pi_J^{\perp} = \mathrm{id} - \Pi_J$ is the projection. We bound T_{21} as follows:

$$\begin{aligned}
T_{21} &\leq \sup_{\phi \in \mathcal{H}_{0,J}(h)} \Big| \langle \Pi_J(h-\phi), \psi^J\rangle'_{L^2(X)} \Big((\widehat{G}_b^{-1/2}\widehat{S})_l^- \widehat{G}_b^{-1/2} S - I_J \Big)' (\widehat{G}_b^{-1/2}\widehat{S})_l^- \widehat{G}_b^{-1/2} \\
&\qquad\qquad \times \Big(\frac{1}{n}\sum_i \Pi_J(h-\phi)(X_i)\widetilde{b}^K(W_i) - \mathrm{E}[\Pi_J(h-\phi)(X)\widetilde{b}^K(W)] \Big) \Big| \\
&\leq \sup_{\phi \in \mathcal{H}_{0,J}(h)} \big\| \langle \Pi_J(h-\phi), \psi^J\rangle_{L^2(X)} \big\| \|S - \widehat{S}\| \Big\| (\widehat{G}_b^{-1/2}\widehat{S})_l^- \widehat{G}_b^{-1/2} \Big\|^2 \\
&\qquad\qquad \times \Big\| \frac{1}{n}\sum_i \Pi_J(h-\phi)(X_i)\widetilde{b}^K(W_i) - \mathrm{E}[\Pi_J(h-\phi)(X)\widetilde{b}^K(W)] \Big\| \\
&\lesssim \sup_{\phi \in \mathcal{H}_{0,J}(h)} \|\Pi_J(h-\phi)\|_{L^2(X)} \times n^{-1/2} s_J^{-2} \zeta_J \sqrt{\log J} \times n^{-1/2}\zeta_J \mathcal{C}_h \\
&\lesssim n^{-1/2} s_J^{-1} \mathcal{C}_h \big(\|h - \Pi_{\mathcal{H}_0} h\|_{L^2(X)} + J^{-p/d_x} \big).
\end{aligned}$$

wpa1 uniformly for $h \in \mathcal{H}_1(\delta^\circ \mathrm{r}_n)$. For T_{22}, we note that uniformly in $h \in \mathcal{H}$ and $\phi \in \mathcal{H}_{0,J}(h)$, $\|\mathrm{E}[\Pi_J^{\perp}(h-\phi)(X)\widetilde{b}^K(W)]\| = \|\Pi_K T \Pi_J^{\perp}(h-\phi)\|_{L^2(W)} \lesssim s_J J^{-p/d_x}$ by Assumption 2(iv). Thus, following the upper bound derivations of T_{21}, we obtain $T_{22} \lesssim n^{-1/2} s_J^{-1} J^{-p/d_x}$ wpa1 uniformly for $h \in \mathcal{H}_1(\delta^\circ \mathrm{r}_n)$. □

Lemma E.5. *Let Assumptions 1(i)-(iii), 2(i), and 4 be satisfied. Then, using the notation $S^o := G_b^{-1/2} S G^{-1/2}$, we have for some constant $C > 0$:*

$$(i)\ \mathrm{P}\left(\max_{J \in \mathcal{I}_n} \left\{ \frac{s_J^2 \sqrt{n}}{\zeta_J \sqrt{\log J}} \left\| (\widehat{G}_b^{-1/2} \widehat{S} \widehat{G}^{-1/2})_l^- \widehat{G}_b^{-1/2} G_b^{1/2} - (S^o)_l^- \right\| \right\} > C \right) = o(1),$$

$$(ii)\ \mathrm{P}\left(\max_{J \in \mathcal{I}_n} \left\{ \frac{s_J^2 \sqrt{n}}{\zeta_J \sqrt{\log J}} \left\| S^o \Big((\widehat{G}_b^{-1/2} \widehat{S} \widehat{G}^{-1/2})_l^- \widehat{G}_b^{-1/2} G_b^{1/2} - (S^o)_l^- \Big) \right\| \right\} > C \right) = o(1).$$

Proof. The results can be established by following the same proof from Chen, Christensen, and Kankanala (2024, Lemma C.4) with their $(\tau_J, \sqrt{J})$ replaced by our (s_J^{-1}, ζ_J). □

Lemma E.6. *Let Assumptions 1(i)-(iii), 2(i) and 4(i) hold. Then, we have*

$$\mathrm{P}_h\Big(\max_{J \in \mathcal{I}_n} \Big| \frac{(\log\log J)^{-1/2}}{(n-1)V_J} \sum_{i \neq i'} U_i(\Pi_{\mathcal{H}_0} h) U_{i'}(\Pi_{\mathcal{H}_0} h) b^K(W_i)' \big(A'A - \widehat{A}'\widehat{A} \big) b^K(W_{i'}) \Big| > \frac{1-c_0}{8} \Big) = o(1)$$

uniformly for $h \in \mathcal{H}_0$, where $U_i(\phi) = Y_i - \phi(X_i)$ and c_0 is as in the proof of Theorem 4.1.

Proof. Let I_{s_J} denote the J dimensional identity matrix multiplied by the vector $C_0(s_1, \dots, s_J)'$ for some sufficiently large constant C_0 and where s_j^{-1}, $1 \le j \le J$, are the nondecreasing singular values of $AG_b^{1/2} = (G_b^{-1/2} S G^{-1/2})_l^-$. There exists a unitary matrix Q such that

$$\begin{aligned}
&\sum_{i \neq i'} U_i(\Pi_{\mathcal{H}_0} h) U_{i'}(\Pi_{\mathcal{H}_0} h) b^K(W_i)' \big(A'A - \widehat{A}'\widehat{A}\big) b^K(W_{i'}) \\
&\le \Big\| \sum_i U_i(\Pi_{\mathcal{H}_0} h) \widetilde{b}^K(W_i)' Q I_{s_J}^{-1} \Big\|^2 \big\| I_{s_J} Q' G_b^{1/2} (A'A - \widehat{A}'\widehat{A}) G_b^{1/2} Q I_{s_J} \big\| \\
&= \sum_{i \neq i'} U_i(\Pi_{\mathcal{H}_0} h) U_{i'}(\Pi_{\mathcal{H}_0} h) \widetilde{b}^K(W_i)' Q I_{s_J}^{-2} Q' \widetilde{b}^K(W_{i'}) \big\| I_{s_J} Q' G_b^{1/2} (A'A - \widehat{A}'\widehat{A}) G_b^{1/2} Q I_{s_J} \big\| \\
&\quad + \sum_i \Big\| U_i(\Pi_{\mathcal{H}_0} h) \widetilde{b}^K(W_i) Q I_{s_J}^{-1} \Big\|^2 \big\| I_{s_J} Q' G_b^{1/2} (A'A - \widehat{A}'\widehat{A}) G_b^{1/2} Q I_{s_J} \big\|.
\end{aligned}$$

The fourth moment condition imposed in Assumption 2(i) implies uniformly for $h \in \mathcal{H}_0$:

$$\begin{aligned}
&\mathrm{E}_h \max_{J \in \mathcal{I}_n} \Big| \frac{1}{nV_J} \sum_i \Big(\| U_i(\Pi_{\mathcal{H}_0} h) \widetilde{b}^K(W_i) Q I_{s_J}^{-1} \|^2 - \mathrm{E}_h \| U(\Pi_{\mathcal{H}_0} h) \widetilde{b}^K(W) Q I_{s_J}^{-1} \|^2 \Big) \Big|^2 \\
&\qquad \lesssim n^{-1} \zeta_J^2 \sum_{J \in \mathcal{I}_n} V_J^{-2} s_J^{-4} \lesssim n^{-1} \zeta_J^2 \sum_{J \in \mathcal{I}_n} \Big(\sum_{j=1}^J s_J^4 s_j^{-4} \Big)^{-1} \lesssim n^{-1} \zeta_J^2 \sum_{J \in \mathcal{I}_n} J^{-1} = o(1),
\end{aligned}$$

due to Lemma B.3 and the definition of the index set $\mathcal{I}_n$. Consequently, from the second moment condition imposed in Assumption 2(i) we obtain uniformly for $J \in \mathcal{I}_n$:

$$n^{-1} \sum_i \| (Y_i - \Pi_{\mathcal{H}_0} h(X_i)) \widetilde{b}^K(W_i) Q I_{s_J}^{-1} \|^2 \le \overline{\sigma}^2 c_0^{-1} \zeta_J \Big(\sum_{j=1}^J s_j^{-4} \Big)^{1/2} \le \overline{\sigma}^2 \underline{\sigma}^{-2} c_0^{-1} \zeta_J V_J$$

with probability approaching one (under $h \in \mathcal{H}_0$), by making use of Lemma B.3. Further, we obtain uniformly for $h \in \mathcal{H}_0$:

$$\begin{aligned}
&\mathrm{P}_h\Big(\max_{J \in \mathcal{I}_n} \Big| \frac{(\log\log J)^{-1/2}}{(n-1)V_J} \sum_{i,i'} U_i(\Pi_{\mathcal{H}_0} h) U_{i'}(\Pi_{\mathcal{H}_0} h) b^K(W_i)' \Big(A'A - \widehat{A}'\widehat{A} \Big) b^K(W_{i'}) \Big| > \frac{1-c_0}{8} \Big) \\
&\le \mathrm{P}_h\Big(\max_{J \in \mathcal{I}_n} \Big| \frac{(\log\log J)^{-1/2}}{(n-1)V_J} \sum_{i \neq i'} U_i(\Pi_{\mathcal{H}_0} h) U_{i'}(\Pi_{\mathcal{H}_0} h) \widetilde{b}^K(W_i)' Q I_{s_J}^{-2} Q' \widetilde{b}^K(W_{i'}) \Big| > \frac{1-c_0}{8} \Big) \\
&\quad + \mathrm{P}_h\Big(\max_{J \in \mathcal{I}_n} \big(\| I_{s_J} Q G_b^{1/2} (A'A - \widehat{A}'\widehat{A}) G_b^{1/2} Q I_{s_J} \| \big) > \frac{1-c_0}{16} \Big) \\
&\quad + \mathrm{P}_h\Big(\max_{J \in \mathcal{I}_n} \big(\overline{\sigma}^2 \underline{\sigma}^{-2} c_0^{-1} \zeta_J (\log\log J)^{-1/2} \| I_{s_J} Q G_b^{1/2} (A'A - \widehat{A}'\widehat{A}) G_b^{1/2} Q' I_{s_J} \| \big) > \frac{1-c_0}{16} \Big) + o(1) \\
&=: T_1 + T_2 + T_3 + o(1).
\end{aligned}$$

Note that T_1 is arbitrarily small for C_0 sufficiently large by following step 1 in the proof of

Theorem 4.1. Consider T_2. We make use of the inequality

$$\big\|I_{s_J}QG_b^{1/2}(\widehat{A}'\widehat{A}-A'A)G_b^{1/2}Q'I_{s_J}\big\| \leq 2\big\|I_{s_J}QG_b^{1/2}\big(\widehat{A}-A\big)'AG_b^{1/2}Q'I_{s_J}\big\|+\big\|(\widehat{A}-A)G_b^{1/2}QI_{s_J}\|^2.$$

It is sufficient to consider the first summand on the right hand side. Note that $\|AG_b^{1/2}Q'I_{s_J}\| \leq C_0^{-1}$. Consequently, from Lemma E.5(ii) we infer

$$\mathrm{P}\left(\max_{J\in\mathcal{I}_n}\left\{\frac{s_J^2\sqrt{n}}{\zeta_J\sqrt{\log J}}\big\|I_{s_J}QG_b^{1/2}\big(\widehat{A}-A\big)'AG_b^{1/2}Q'I_{s_J}\big\|\right\} > C\right) = o(1).$$

Assumption 4(i), i.e., $s_J^{-1}\zeta_J^2\sqrt{(\log J)/n} = O(1)$ uniformly for $J \in \mathcal{I}_n$ thus implies $T_3 = o(1)$. □

Proof of Lemma B.4. It is sufficient to prove (ii). Let $\Sigma = \mathrm{E}_h[(Y-h(X))^2b^{K(J)}(W)b^{K(J)}(W)']$ and $\widehat{\Sigma} = n^{-1}\sum_i\big(Y_i-\widehat{h}_J(X_i)\big)^2b^{K(J)}(W_i)b^{K(J)}(W_i)'$. Then $V_J = \big\|A\Sigma A'\big\|_F$ and $\widehat{V}_J = \big\|\widehat{A}\widehat{\Sigma}\widehat{A}'\big\|_F$. For all $J\in\mathcal{I}_n$ the triangular inequality implies

$$|\widehat{V}_J-V_J| \leq \big\|\widehat{A}\widehat{\Sigma}\widehat{A}'-A\Sigma A'\big\|_F \leq 2\big\|\big(\widehat{A}-A\big)\widehat{\Sigma}\,A'\big\|_F+\big\|\big(\widehat{A}-A\big)\widehat{\Sigma}^{1/2}\big\|_F^2+\big\|A\big(\widehat{\Sigma}-\Sigma\big)A'\big\|_F.$$

In the remainder of this proof, it is sufficient to consider $\|(\widehat{A}-A)\Sigma\,A'\|_F+\|A(\widehat{\Sigma}-\Sigma)\,A'\|_F =: T_1+T_2$. Consider T_1. By Lemma E.2 we have the upper bound $\|G_b^{-1/2}\Sigma G_b^{-1/2}\| \leq \overline{\sigma}$. Below we make use of the inequality $\|m_1m_2\|_F \leq \|m_1\|\|m_2\|_F$ for matrices m_1 and m_2. Since the Frobenius norm is invariant under rotation, we calculate uniformly for $J\in\mathcal{I}_n$ that

$$\begin{aligned}T_1 &= \|(G_b^{1/2}SG^{1/2})(\widehat{A}-A)\Sigma\,A'AG_b^{1/2}\|\\ &\leq \|(G_b^{1/2}SG^{1/2})(\widehat{A}-A)G_b^{1/2}\|\|G_b^{-1/2}\Sigma G_b^{-1/2}\|\|(G_b^{1/2}SG^{1/2})_l^{-2}\|_F \lesssim \frac{\zeta_J}{s_J}\Big(\frac{\log(J)}{n}\sum_{j=1}^{J}s_j^{-4}\Big)^{1/2}\end{aligned}$$

wpa1 uniformly or $h\in\mathcal{H}$, by making use of Lemma E.5(i) and the Schur decomposition as in the proof of Lemma B.3. From Assumption 4(i), i.e., $s_J^{-1}\zeta_J^2\sqrt{(\log J)/n} = O(1)$, uniformly for $J\in\mathcal{I}_n$ we infer $T_1/V_J = J^{-1/2}(\sum_{j=1}^J s_j^{-4})^{1/2}/V_J \to 0$ wpa1 uniformly or $h\in\mathcal{H}$, where the last equation is due to Lemma B.3. Consider T_2. Again using Lemma B.3 we obtain $T_2 \leq \underline{\sigma}^{-2}\|G_b^{-1/2}(\widehat{\Sigma}-\Sigma)G_b^{-1/2}\|$ by using the upper bound as derived for T_1. Further, evaluate

$$\begin{aligned}\|G_b^{-1/2}(\widehat{\Sigma}-\Sigma)G_b^{-1/2}\| &= \Big\|\frac{1}{n}\sum_i\big((Y_i-\widehat{h}_J(X_i))^2-(Y_i-h(X_i))^2\big)\widetilde{b}^K(W_i)\widetilde{b}^K(W_i)'\Big\|\\ &\leq \Big\|\frac{1}{n}\sum_i\big(\widehat{h}_J(X_i)-h(X_i)\big)^2\widetilde{b}^K(W_i)\widetilde{b}^K(W_i)'\Big\|\\ &\quad+2\Big\|\frac{1}{n}\sum_i\big(\widehat{h}_J(X_i)-h(X_i)\big)(Y_i-h(X_i))\widetilde{b}^K(W_i)\widetilde{b}^K(W_i)'\Big\| =: T_{21}+T_{22}.\end{aligned}$$

Consider T_{21}. The definition of the unrestricted sieve NPIV estimator in (2.5) implies uniformly for $J \in \mathcal{I}_n$

$$
\begin{aligned}
T_{21} &\leq \Big\| \frac{1}{n} \sum_i \big(\widehat{h}_J(X_i) - Q_J h(X_i)\big)^2 \widetilde{b}^K(W_i) \widetilde{b}^K(W_i)' \Big\| + \Big\| \frac{1}{n} \sum_i \big(Q_J h(X_i) - h(X_i)\big)^2 \widetilde{b}^K(W_i) \widetilde{b}^K(W_i)' \Big\| \\
&\leq \zeta_J^2 \Big\| \widehat{A} \frac{1}{n} \sum_i Y_i b^K(W_i) - A \,\mathrm{E}_h[Y b^K(W)] \Big\|^2 \times \Big\| \frac{1}{n} \sum_i \psi^J(X_i) \psi^J(X_i)' \Big\| \\
&\quad + \zeta_J^2 \Big\| \frac{1}{n} \sum_i \big(Q_J h(X_i) - h(X_i)\big)^2 \Big\| \lesssim \zeta_J^4 s_J^{-2} n^{-1} + \max_{J \in \mathcal{I}_n} \big\{ \zeta_J^2 \| Q_J h - h \|_{L^2(X)} \big\}
\end{aligned}
$$

wpa1 uniformly for $h \in \mathcal{H}$, where the right hand side tends to zero. This follows by the rate condition imposed in Assumption 4(i) and that $\|Q_J h - h\|_{L^2(X)} = O(J^{-p/d_x})$ uniformly for $J \in \mathcal{I}_n$ and $h \in \mathcal{H}$ by Lemma B.1(ii). Analogously, we obtain that $\max_{J \in \mathcal{I}_n} T_{22}$ vanishes wpa1 uniformly for $h \in \mathcal{H}$. □

Proof of Lemma B.5. We first prove the lower bound. By the definition of the RES index set $\widehat{\mathcal{I}}_n$ we have that any element $J \in \widehat{\mathcal{I}}_n$, tends slowly to infinity as $n \to \infty$. Let $\widehat{j}_{\max} \leq j_{\max}$ be the largest integer such that $\underline{J} 2^{\widehat{j}_{\max}} \leq \widehat{J}_{\max}$. Consequently, the definition of the RES index set implies for all $J \in \widehat{\mathcal{I}}_n$ that

$$
\log(J) \leq \log(\underline{J} 2^{\widehat{j}_{\max}}) = \widehat{j}_{\max} \log(2) + \log(\underline{J}) \leq \widehat{j}_{\max} + 1 = \#(\widehat{\mathcal{I}}_n)
$$

for n sufficiently large. From the lower bounds for quantiles of the chi-squared distribution established in Inglot (2010, Theorem 5.2) we deduce for all $J \in \widehat{\mathcal{I}}_n$ and n sufficiently large:

$$
\begin{aligned}
\widehat{\eta}_J(\alpha) &= \frac{q\big(\alpha/\#(\widehat{\mathcal{I}}_n), J\big) - J}{\sqrt{J}} \geq \frac{q\big(\alpha/(\log J), J\big) - J}{\sqrt{J}} \\
&\geq \frac{\sqrt{\log\big((\log J)/\alpha\big)}}{4} + \frac{2 \log\big((\log J)/\alpha\big)}{\sqrt{J}} \geq \frac{\sqrt{\log\log(J) - \log(\alpha)}}{4}
\end{aligned}
$$

using the lower bounds for quantiles of the chi-squared distribution established in Inglot (2010, Theorem 5.2). We now consider the upper bound. From the definition of $\#(\widehat{\mathcal{I}}_n)$ we infer $\#(\widehat{\mathcal{I}}_n) = \widehat{j}_{\max} + 1 \leq \lceil \log_2(n^{1/3}/\underline{J}) \rceil + 1 \leq \log(n^{1/3}/\underline{J}) + 1$ and thus $\#(\widehat{\mathcal{I}}_n) \leq \log(n)$. Consequently, we calculate for all $J \in \widehat{\mathcal{I}}_n$ and n sufficiently large:

$$
\begin{aligned}
\widehat{\eta}_J(\alpha) &\leq \frac{q\big(\alpha/(\log n), J\big) - J}{\sqrt{J}} \leq 2\sqrt{\log\big((\log n)/\alpha\big)} + \frac{2 \log\big((\log n)/\alpha\big)}{\sqrt{J}} \\
&\leq 2\sqrt{\log\big((\log n)/\alpha\big)}(1 + o(1)) \leq 4\sqrt{\log\log(n) - \log(\alpha)},
\end{aligned}
$$

where the second inequality is due to Laurent and Massart (2000, Lemma 1). □

Proof of Lemma B.6. Result B.6(i) directly follows from Houdré and Reynaud-Bouret

(2003, Theorem 3.4); see also Gine and Nickl (2016, Theorem 3.4.8). We next prove the bounds on $\Lambda_1,\ \Lambda_2,\ \Lambda_3,\ \Lambda_4$ for Result B.6(ii).

For the bound on Λ_1, we recall the notation $U_i^J = U_i A b^K(W_i)$ with $U_i = Y_i - h(X_i)$ for $h \in \mathcal{H}_0$. Then, under $\mathcal{H}_0$ we have:

$$\begin{aligned}
\mathrm{E}_h[R_1^2(Z_1, Z_2)] \le \mathrm{E}_h \left|U_1 b^K(W_1)' A' A b^K(W_2) U_2\right|^2 &= \mathrm{E}_h\left[(U^J)' \,\mathrm{E}_h\left[U^J (U^J)'\right] U^J\right]\\
&= \sum_{j,j'=1}^{J} \mathrm{E}_h[U_{1j} U_{1j'}]^2 = V_J^2.
\end{aligned}$$

For the bound on Λ_2, for any function ν and κ with $\|\nu\|_{L^2(Z)} \le 1$ and $\|\kappa\|_{L^2(Z)} \le 1$, respectively, we obtain

$$\begin{aligned}
|\,\mathrm{E}_h[R_1(Z_1, Z_2)\nu(Z_1)\kappa(Z_2)]| \le& \left|\,\mathrm{E}_h[U \mathbb{1}_M b^K(W)'\nu(Z)] A' A \,\mathrm{E}_h[U \mathbb{1}_M b^K(W)\kappa(Z)]\right|\\
\le& \|A \,\mathrm{E}_h[U \mathbb{1}_M b^K(W)\kappa(Z)]\| \, \|A \,\mathrm{E}_h[U \mathbb{1}_M b^K(W)\nu(Z)]\|\\
\le& \|A G_b^{1/2}\|^2 \sqrt{\mathrm{E}\left[|\,\mathrm{E}_h[U \mathbb{1}_M \kappa(Z)|W]|^2\right]} \times \sqrt{\mathrm{E}\left[|\,\mathrm{E}_h[U \mathbb{1}_M \nu(Z)|W]|^2\right]}
\end{aligned}$$

Now observe $\mathrm{E}\left[|\,\mathrm{E}_h[U\mathbb{1}_M \kappa(Z)|W]|^2\right] \le \mathrm{E}\left[\,\mathrm{E}_h[U^2|W]\kappa^2(Z)\right] \le \overline{\sigma}^2$ by Assumption 2(i) and using that $\|\kappa\|_{L^2(Z)} \le 1$, which yields the upper bound by using $\|A G_b^{1/2}\| = s_J^{-1}$.

For the bound on Λ_3, observe that for any $z = (u, w)$

$$\begin{aligned}
\left|\,\mathrm{E}_h[R_1^2(Z_1, z)]\right| \le& \,\mathrm{E}_h \left|U \mathbb{1}\{|U| \le M_n\} b^K(W)' A' A b^K(w) u \mathbb{1}\{|u| \le M_n\}\right|^2\\
&\le \|A b^K(w) u \mathbb{1}\{|u| \le M_n\}\|^2 \,\mathrm{E}_h \|A b^K(W) U\|^2 \le \overline{\sigma}^2 M_n^2\, \zeta_{b,K}^2 \,\|A G_b^{1/2}\|^4,
\end{aligned}$$

again by using Assumption 2(i) and hence the upper bound on Λ_3 follows.

For the bound on Λ_4, observe that for any $z_1 = (u_1, w_1)$ and $z_2 = (u_2, w_2)$ we get

$$\begin{aligned}
\left|R_1(z_1, z_2)\right| \le& \left|u_1 \mathbb{1}\{|u_1| \le M_n\} b^K(w_1)' A' A b^K(w_2) u_2 \mathbb{1}\{|u_2| \le M_n\}\right|\\
&\le \sup_{u,w} \left\|A b^K(w) u \mathbb{1}\{|u| \le M_n\}\right\|^2 \le M_n^2 \zeta_{b,K}^2 \|A G_b^{1/2}\|^2,
\end{aligned}$$

which completes the proof. □

Proof of Lemma B.7. It suffices to prove (ii) for a simple null $\mathcal{H}_0 = \{h_0\}$. For any $h \in \mathcal{H}_1(\delta^\circ \mathsf{r}_n)$, we denote $B_J = (\|\,\mathrm{E}_h[U^J]\| - \|h - h_0\|_{L^2(X)})^2$. Applying $\|\,\mathrm{E}_h[U^{J^*}]\|^2 = \|Q_{J^*}(h - h_0)\|_{L^2(X)}^2$ and Lemma B.1(i) we obtain: $B_{J^*} = \left(\|Q_{J^*}(h-h_0)\|_{L^2(X)} - \|h - h_0\|_{L^2(X)}\right)^2 \le C_B\, \mathsf{r}_n^2$ for some constant C_B. By the inequality $\|\,\mathrm{E}_h[U^{J^*}]\|^2 \ge \|h - h_0\|_{L^2(X)}^2/2 - B_{J^*}$ we have uniformly for $h \in \mathcal{H}_1(\delta^\circ r_{n,J})$:

$$\mathrm{P}_h\left(n\,\widehat{D}_{J^*}(h_0) \leq 2c_1\sqrt{\log\log n}\,U_{J^*}\right) = \mathrm{P}_h\left(\|\,\mathrm{E}_h[U^{J^*}]\|^2 - \widehat{D}_{J^*}(h_0) > \|\,\mathrm{E}_h[U^{J^*}]\|^2 - \frac{2c_1\sqrt{\log\log n}\,U_{J^*}}{n}\right)$$
$$\leq \mathrm{P}_h\Big(\Big|\frac{4}{n(n-1)}\sum_{j=1}^{J^*}\sum_{i<i'}\big(U_{ij}U_{i'j} - \mathrm{E}_h[U_{1j}]^2\big)\Big| > \rho_h\Big)$$
$$+\,\mathrm{P}_h\Big(\Big|\frac{4}{n(n-1)}\sum_{i<i'}(Y_i - h_0(X_i))(Y_{i'} - h_0(X_{i'})b^{K^*}(W_i)'\big(A'A - \widehat{A}'\widehat{A}\big)b^{K^*}(W_{i'})\Big| > \rho_h\Big) = T_1 + T_2,$$

where $\rho_h = \|h - h_0\|^2_{L^2(X)}/2 - 2c_1 n^{-1}\sqrt{\log\log n}V_{J^*} - B_{J^*}$. To bound term T_1, we apply inequality (E.5) and Markov's inequality:

$$T_1 \lesssim n^{-1}s_{J^*}^{-2}\rho_h^{-2}\big(\|h - h_0\|^2_{L^2(X)} + (J^*)^{-2p/d_x}\big) + n^{-2}V_{J^*}^2\rho_h^{-2}. \tag{E.7}$$

In the following, we distinguish between two cases. First, consider the case where $n^{-2}V_{J^*}^2\rho_h^{-2}$ dominates the right hand side. For any $h \in \mathcal{H}_1(\delta^\circ \mathsf{r}_n)$ we have $\|h - h_0\|_{L^2(X)} \geq \delta^\circ \mathsf{r}_n$ and hence, we obtain the lower bound

$$\rho_h = \|h - h_0\|^2_{L^2(X)}/2 - 2c_1 n^{-1}\sqrt{\log\log n}V_{J^*} - B_{J^*} \geq \kappa_0 \mathsf{r}_n^2 \tag{E.8}$$

where $\kappa_0 := (\delta^\circ)^2/2 - C - C_B$ for some constant $C > 0$ and $\kappa_0 > 0$ whenever $\delta^\circ > \sqrt{2(C + C_B)}$. From inequality (E.7) we infer $T_1 \lesssim n^{-2}V_{J^*}^2(J^*)^{4p/d_x} = o(1)$. Second, consider the case where $n^{-1}s_{J^*}^{-2}\rho_h^{-2}\big(\|h-h_0\|^2_{L^2(X)}+(J^*)^{-2p/d_x}\big)$ dominates. For any $h \in \mathcal{H}_1(\delta^\circ \mathsf{r}_n)$ we have $\|h - h_0\|^2_{L^2(X)} \geq (\delta^\circ)^2\mathsf{r}_n^2 \geq 5c_1 n^{-1}V_{J^*}\sqrt{\log\log n}$ for δ° sufficiently large and hence, we obtain $\rho_h \geq \kappa_1\,\|h - h_0\|^2_{L^2(X)}$ for some constant $\kappa_1 := 1/5 - C_B/(\delta^\circ)^2$ which is positive for any $\delta^\circ > \sqrt{5C_B}$. Hence, inequality (E.7) yields uniformly for $h \in \mathcal{H}_1(\delta^\circ \mathsf{r}_n)$ that

$$T_1 \lesssim n^{-1}s_{J^*}^{-2}\left(\|h - h_0\|^{-2}_{L^2(X)} + \|h - h_0\|^{-4}_{L^2(X)}(J^*)^{-2p/d}\right) \lesssim n^{-1}s_{J^*}^{-2}\mathsf{r}_n^{-2} = o(1).$$

Finally, $T_2 = o(1)$ uniformly for $h \in \mathcal{H}_1(\delta^\circ \mathsf{r}_n)$ by making use of Lemma E.1. □

Proof of Lemma B.8. Recall the definition of $\overline{J} = \sup\{J : \zeta^2(J)\sqrt{(\log J)/n} \leq \bar{c}\,s_J\}$. Following the proof of Chen, Christensen, and Kankanala (2024, Lemma C.6), using Weyl's inequality (see e.g. Chen and Christensen (2018, Lemma F.1)) together with Chen and Christensen (2018, Lemma F.7) we obtain that $|\widehat{s}_J - s_J| \leq c_0 s_J$ uniformly in $J \in \mathcal{I}_n$ for some $0 < c_0 < 1$ with probability approaching one uniformly for $h \in \mathcal{H}$.
Proof of (i). By making use of the definition of $\widehat{J}_{\max}$ given in (2.11), we obtain uniformly for $h \in \mathcal{H}$:

$$\mathrm{P}_h\left(\widehat{J}_{\max} > \overline{J}\right) \leq \mathrm{P}_h\left(\zeta^2(\overline{J})\sqrt{\log(\overline{J})/n} < \frac{3}{2}\widehat{s}_{\overline{J}}\right) \leq \mathrm{P}_h\left(\zeta^2(\overline{J})\sqrt{\log(\overline{J})/n} < \frac{3}{2}(1 + c_0)s_{\overline{J}}\right) + o(1)$$

The upper bound imposed on the growth of $\overline{J}$ is determined by a sufficiently large constant $\overline{c} > 0$ and hence, there exists a constant $\underline{c} \geq 3(1+c_0)/2$ such that $s_{\overline{J}}^{-1}\zeta^2(\overline{J})\sqrt{\log(\overline{J})/n} \geq \underline{c}$. Consequently, we obtain

$$\mathrm{P}_h\left(\widehat{J}_{\max} > \overline{J}\right) \leq \mathrm{P}_h\left(s_{\overline{J}}^{-1}\zeta^2(\overline{J})\sqrt{\log(\overline{J})/n} < \frac{3}{2}(1+c_0)\right) + o(1) = o(1).$$

Proof of (ii). From the definition of J° given in (4.3) we infer as above for some constant $0 < c_0 < 1$ and uniformly for $h \in \mathcal{H}$:

$$\mathrm{P}_h\left(J^\circ > \widehat{J}_{\max}\right) \leq \mathrm{P}_h\left((1-c_0)n^{-1}\sqrt{\log\log n}\,\widehat{J}_{\max}^{2p/d_x+1/2} \leq \widehat{s}^2_{\widehat{J}_{\max}}\right) + o(1).$$

Consider the case $\zeta(J) = \sqrt{J}$. The definition of $\widehat{J}_{\max}$ in (2.11) yields uniformly for $h \in \mathcal{H}$:

$$\begin{aligned}
\mathrm{P}_h\left(J^\circ > \widehat{J}_{\max}\right) &\leq \mathrm{P}_h\left((1-c_0)\sqrt{\log\log n}\,\widehat{J}_{\max}^{2p/d_x-3/2} \leq (\log \overline{J})\right) + o(1)\\
&\leq \mathrm{P}_h\left((1-c_0)\widehat{s}_{\widehat{J}_{\max}}\sqrt{n} \leq \frac{2}{3}\sqrt{\log\overline{J}}\Big(\frac{\log\overline{J}}{\sqrt{\log\log n}}\Big)^{1/(2p/d_x-3/2)}\right) + o(1)\\
&\leq \mathrm{P}_h\left((1-c_0)^2 s_{\overline{J}}\sqrt{n} \leq \frac{2}{3}\sqrt{\log\overline{J}}\Big(\frac{\log\overline{J}}{\sqrt{\log\log n}}\Big)^{1/(2p/d_x-3/2)}\right) + o(1)\\
&\leq \mathrm{P}_h\left(\frac{(1-c_0)^2}{\overline{c}}\overline{J} \leq \frac{2}{3}\Big(\frac{\log\overline{J}}{\sqrt{\log\log n}}\Big)^{1/(2p/d_x-3/2)}\right) + o(1),
\end{aligned}$$

where the last inequality follows from the definition of $\overline{J}$, i.e., $s_{\overline{J}} \geq \overline{c}^{-1}\overline{J}\sqrt{\log(\overline{J})/n}$. From Assumption 4(iii), i.e., $p \geq 3d_x/4$, we infer $\mathrm{P}_h(J^\circ > \widehat{J}_{\max}) = o(1)$ and, in particular, $\mathrm{P}_h(2J^\circ > \widehat{J}_{\max}) = o(1)$ uniformly for $h \in \mathcal{H}$. The proof of $\zeta(J) = J$ follows analogously using the condition $p \geq 7d_x/4$. □

References

Bierens, H. J. (1990): "A consistent conditional moment test of functional form," *Econometrica*, 58(6), 1443–1458.

Breunig, C. (2020): "Specification testing in nonparametric instrumental quantile regression," *Econometric Theory*, 36(4), 583–625.

Breunig, C., and X. Chen (2020): "Adaptive, Rate-Optimal Testing in Instrumental Variables Models," *arXiv preprint arXiv:2006.09587v1.*

——— (2021): "Adaptive, Rate-Optimal Hypothesis Testing in Nonparametric IV Models," *arXiv preprint arXiv:2006.09587v2.*

Chen, X., T. Christensen, and S. Kankanala (2024): "Adaptive Estimation and Uniform Confidence Bands for Nonparametric Structural Functions and Elasticities," *The Review of Economic Studies, forthcoming.*

Chen, X., and T. M. Christensen (2018): "Optimal sup-norm rates and uniform inference on nonlinear functionals of nonparametric IV regression," *Quantitative Economics*, 9(1), 39–84.

Chetverikov, D., and D. Wilhelm (2017): "Nonparametric instrumental variable estimation under monotonicity," *Econometrica*, 85, 1303–1320.

Fang, Z., and J. Seo (2021): "A Projection Framework for Testing Shape Restrictions That Form Convex Cones," *Econometrica*, 89(5), 2439–2458.

Gine, E., and R. Nickl (2016): *Mathematical Foundations of Infinite-Dimensional Statistical Models*. Cambridge University Press.

Horowitz, J. L. (2006): "Testing a Parametric Model Against a Nonparametric Alternative with Identification Through Instrumental Variables," *Econometrica*, 74(2), 521–538.

Houdré, C., and P. Reynaud-Bouret (2003): "Exponential inequalities, with constants, for U-statistics of order two," in *Stochastic inequalities and applications*, pp. 55–69. Springer.

Inglot, T. (2010): "Inequalities for quantiles of the chi-square distribution," *Probability and Mathematical Statistics*, 30(2), 339–351.

Laurent, B., and P. Massart (2000): "Adaptive estimation of a quadratic functional by model selection," *The Annals of Statistics*, pp. 1302–1338.

van der Vaart, A., and J. Wellner (2000): *Weak Convergence and Empirical Processes: With Applications to Statistics (Springer Series in Statistics)*. Springer.