Supplemental Material to

OPTIMAL SUP-NORM RATES AND UNIFORM INFERENCE ON NONLINEAR FUNCTIONALS OF NONPARAMETRIC IV REGRESSION

By

Xiaohong Chen and Timothy M. Christensen

August 2013
Updated February 2017

COWLES FOUNDATION DISCUSSION PAPER NO. 1923R2

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281

http://cowles.yale.edu/

Main Online Appendix to

Optimal Sup-norm Rates and Uniform Inference on Nonlinear Functionals of Nonparametric IV Regression

Xiaohong Chen Timothy M. Christensen

First version: August 2013. Revised: January 2017.

This main online supplementary appendix contains material to support our paper "Optimal Sup-norm Rates and Uniform Inference on Nonlinear Functionals of Nonparametric IV Regression ". Appendix D presents pointwise normality of sieve t statistics for nonlinear functionals of NPIV under low-level sufficient conditions. Appendix E contains background material on B-spline and wavelet bases and the equivalence between Besov and wavelet sequence norms. Appendix F contains material on useful matrix inequalities and convergence results for random matrices. The secondary online supplementary appendix contains additional technical lemmas and all of the proofs (Appendix G).

D Pointwise asymptotic normality of sieve t-statistics

In this section we derive the pointwise asymptotic normality of sieve t-statistics for nonlinear functionals of a NPIV function under low-level sufficient conditions. Previously under some high-level conditions, Chen and Pouzo (2015) established the pointwise asymptotic normality of sieve t statistics for (possibly) nonlinear functionals of h_0 satisfying general semi/nonparametric conditional moment restrictions including NPIV and nonparametric quantile IV models as special cases. As the sieve NPIV estimator $\widehat{h}$ has a closed-form expression and for the sake of easy reference, we derive the limit theory directly rather than appealing to the general theory in Chen and Pouzo (2015). Our low-level sufficient conditions are tailored to the case in which the functional $f(\cdot)$ is *irregular* in h_0 (i.e. slower than root-n estimable), so that they are directly comparable to the sufficient conditions for the uniform inference theory in Section 4.

We consider a functional $f : \mathcal{H} \subset L^\infty(X) \to \mathbb{R}$ for which $Df(h)[v] = \lim_{\delta \to 0^+}[\delta^{-1} f(h + \delta v)]$ exists for all $v \in \mathcal{H} - \{h_0\}$ for all h in a small neighborhood of h_0. Recall that the sieve 2SLS Riesz representer of $Df(h_0)$ is

$$v_n(f)(x) = \psi^J(x)'[S'G_b^{-1}S]^{-1}Df(h_0)[\psi^J] \, ,$$

and let

$$[s_n(f)]^2 = \|\Pi_K T v_n(f)\|^2_{L^2(W)} = (Df(h_0)[\psi^J])'[S'G_b^{-1}S]^{-1}Df(h_0)[\psi^J]$$

denote its weak norm. Chen and Pouzo (2015) called that the functional $f(\cdot)$ is an irregular (i.e. slower than $\sqrt{n}$-estimable) functional of h_0 if $s_n(f) \nearrow \infty$ and a regular (i.e. $\sqrt{n}$-estimable) functional of h_0 if $\lim_n s_n(f) < \infty$. Denote

$$\widehat{v}_n(f)(x) = \psi^J(x)'[S'G_b^{-1}S]^{-1}Df(\widehat{h})[\psi^J] \ .$$

It is clear that $v_n(f) = \widehat{v}_n(f)$ whenever $f(\cdot)$ is linear.

Recall that $\Omega = E[u_i^2 b^K(W_i)b^K(W_i)']$, and the "2SLS covariance matrix" for $\widehat{c}$ (given in equation (2)) is

$$\mho = [S'G_b^{-1}S]^{-1}S'G_b^{-1}\Omega G_b^{-1}S[S'G_b^{-1}S]^{-1} \ ,$$

and the sieve variance for $f(\widehat{h})$ is

$$[\sigma_n(f)]^2 = \big(Df(h_0)[\psi^J]\big)'\mho\big(Df(h_0)[\psi^J]\big) \ .$$

Under Assumption 2(i)(iii) we have that $[\sigma_n(f)]^2 \asymp [s_n(f)]^2$. Therefore $f()$ is an irregular functional of h_0 iff $\sigma_n(f) \nearrow +\infty$ as $n \to \infty$. Recall the sieve variance estimator is

$$[\widehat{\sigma}(f)]^2 = \big(Df(\widehat{h})[\psi^J]\big)'\widehat{\mho}\big(Df(\widehat{h})[\psi^J]\big) \ .$$

where $\widehat{\mho}$ is defined in equation (6).

Assumption 2 (continued)**.** *(iv') $\sup_w E[u_i^2\{|u_i| > \ell(n)\}|W_i = w] = o(1)$ for any positive sequence with $\ell(n) \nearrow \infty$.*

Assumption 2(iv') is a mild condition which is trivially satisfied if $E[|u_i|^{2+\epsilon}|W_i = w]$ is uniformly bounded for some $\epsilon > 0$.

Assumption 5' *Assumption 5 holds with $f_t = f$ and $\mathcal{T}$ a singleton.*

Assumption 5'(a) and 5'(b)(i)(ii) is similar to Assumption 3.5 of Chen and Pouzo (2015). Assumption 5'(b)(iii) controls any additional error arising in the estimation of $\sigma_n(f)$ due to nonlinearity of $f(\cdot)$ and is automatically satisfied when $f(\cdot)$ is a linear functional.

Remark D.1. *Remark 4.1 presents sufficient conditions for Assumption 5' as a special case, with $f_t = f$, $\underline{\sigma}_n = \sigma_n(f)$, and $\mathcal{T}$ a singleton.*

Again these sufficient conditions are formulated to take advantage of the sup-norm rate results in Section 3. Denote

$$\widehat{\mathbb{Z}}_n \equiv \frac{(Df(h_0)[\psi^J])'[S'G_b^{-1}S]^{-1}S'G_b^{-1}}{\sigma_n(f)}\frac{1}{\sqrt{n}}\sum_{i=1}^n b^K(W_i)u_i \, ,$$

and $\delta_{V,n} \equiv \big[\zeta_{b,K}^{(2+\delta)/\delta}\sqrt{(\log K)/n}\big]^{\delta/(1+\delta)} + \tau_J\zeta\sqrt{(\log J)/n} + \delta_{h,n}$, where $\delta_{h,n} = o_p(1)$ is a positive finite sequence such that $\|\widehat{h} - h_0\|_\infty = O_p(\delta_{h,n})$.

Theorem D.1. *(1) Let Assumptions 1(iii), 2(i)(iii)(iv'), 4(i), and either 5'(a) or 5'(b)(i)(ii) hold, and let $\tau_J \zeta \sqrt{(J \log J)/n} = o(1)$. Then:*

$$\sqrt{n}\frac{(f(\widehat{h}) - f(h_0))}{\sigma_n(f)} = \widehat{\mathbb{Z}}_n + o_p(1) \to_d N(0,1)\,.$$

(2) If $\|\widehat{h} - h_0\|_\infty = o_p(1)$ and Assumptions 2(ii) and 3(iii) hold (and 5'(b)(iii) also holds if f is nonlinear), then:

$$\left|\frac{\widehat{\sigma}(f)}{\sigma_n(f)} - 1\right| \;=\; = O_p(\delta_{V,n} + \eta_n') = o_p(1)\ ,$$

and

$$\sqrt{n}\frac{(f(\widehat{h}) - f(h_0))}{\widehat{\sigma}(f)} = \widehat{\mathbb{Z}}_n + o_p(1) \to_d N(0,1)\,.$$

By exploiting the closed form expression of the sieve NPIV estimator and by applying exponential inequalities for random matrices, Theorem D.1 derives the pointwise limit theory under lower-level sufficient conditions than those in Chen and Pouzo (2015) for irregular nonlinear functionals. In particular, when specialized to the exogenous case of $X_i = W_i$, $h_0(x) = E[Y_i|W_i = x]$, $K = J$ and $b^K = \psi^J$ with $\tau_J = 1$, the regularity conditions for Theorem D.1 become about the same mild conditions for Theorem 3.2 in Chen and Christensen (2015) on asymptotic normality of sieve t statistics for nonlinear functionals of series LS estimators. It is now obvious that one could also derive the asymptotic normality of sieve t-statistics for regular (i.e., root-n estimable) nonlinear functionals of a NPIV function under lower-level sufficient conditions by using our sup-norm rates results to verify Assumption 3.5(ii) and Remark 3.1 in Chen and Pouzo (2015).

E Spline and wavelet bases

In this section we bound the terms $\xi_{\psi,J}$, $e_J = \lambda_{\min}(G_{\psi,J})$ and $\kappa_\psi(J)$ for B-spline and CDV wavelet bases. Although we state the results for the space Ψ_J, they may equally be applied to B_K when B_K is constructed using B-spline or CDV wavelet bases.

E.1 Spline bases

We construct a univariate B-spline basis of order $r \geq 1$ (or degree $r-1 \geq 0$) with $m \geq 0$ interior knots and support $[0,1]$ in the following way. Let $0 = t_{-(r-1)} = \ldots = t_0 \leq t_1 \leq \ldots \leq t_m \leq t_{m+1} = \ldots = t_{m+r} = 1$ denote the extended knot sequence and let $I_1 = [t_0, t_1), \ldots, I_m = [t_m, t_{m+1}]$. A basis of order 1 is constructed by setting

$$N_{j,1}(x) = \begin{cases} 1 & \text{if } x \in I_j \\ 0 & \text{otherwise} \end{cases}$$

for $j = 0, \ldots m$. Bases of order $r > 1$ are generated recursively according to

$$N_{j,r}(x) = \frac{x - t_j}{t_{j+r-1} - t_j} N_{j,r-1}(x) + \frac{t_{j+r} - x}{t_{j+r} - t_{j+1}} N_{j+1,r-1}(x)$$

for $j = -(r-1), \ldots, m$ where we adopt the convention $\frac{1}{0} := 0$ (see Section 5 of DeVore and Lorentz (1993)). This results in a total of $m + r$ splines of order r, namely $N_{-(r-1),r}, \ldots, N_{m,r}$. Each spline is a polynomial of degree $r - 1$ on each interior interval $I_1, \ldots, I_m$ and is $(r-2)$-times continuously differentiable on $[0, 1]$ whenever $r \geq 2$. The mesh ratio is defined as

$$\text{mesh}(m) = \frac{\max_{0 \leq j \leq m}(t_{j+1} - t_j)}{\min_{0 \leq j \leq m}(t_{j+1} - t_j)} .$$

Clearly $\text{mesh}(m) = 1$ whenever the knots are placed evenly (i.e. $t_i = \frac{i}{m+1}$ for $i = 1, \ldots, m$ and $m \geq 1$) and we say that the mesh ratio is *uniformly bounded* if $\text{mesh}(m) \lesssim 1$ as $m \to \infty$. Each of has continuous derivatives of orders $\leq r - 2$ on $(0, 1)$. We let the space $\text{BSpl}(r, m, [0, 1])$ be the closed linear span of the $m + r$ splines $N_{-(r-1),r}, \ldots, N_{m,r}$.

We construct B-spline bases for $[0, 1]^d$ by taking tensor products of univariate bases. First generate d univariate bases $N_{-(r-1),r,i}, \ldots, N_{m,r,i}$ for each of the d components x_i of x as described above. Then form the vector of basis functions ψ^J by taking the tensor product of the vectors of univariate basis functions, namely:

$$\psi^J(x_1, \ldots, x_d) = \bigotimes_{i=1}^{d} \begin{pmatrix} N_{-(r-1),r,i}(x_i) \\ \vdots \\ N_{m,r,i}(x_i) \end{pmatrix} .$$

The resulting vector ψ^J has dimension $J = (r + m)^d$. Let $\psi_{J1}, \ldots, \psi_{JJ}$ denote its J elements.

Stability properties: The following two Lemmas bound $\xi_{\psi,J}$, and the minimum eigenvalue and condition number of $G_\psi = G_{\psi,J} = E[\psi^J(X_i)\psi^J(X_i)']$ when $\psi_{J1}, \ldots, \psi_{JJ}$ is constructed using univariate and tensor-products of B-spline bases with uniformly bounded mesh ratio.

Lemma E.1. *Let X have support $[0, 1]$ and let $\psi_{J1} = N_{-(r-1),r}, \ldots, \psi_{JJ} = N_{m,r}$ be a univariate B-spline basis of order $r \geq 1$ with $m = J - r \geq 0$ interior knots and uniformly bounded mesh ratio. Then: (a) $\xi_{\psi,J} = 1$ for all $J \geq r$; (b) If the density of X is uniformly bounded away from 0 and ∞ on $[0, 1]$, then there exists finite positive constants c_ψ and C_ψ such that $c_\psi J \leq \lambda_{\max}(G_\psi)^{-1} \leq \lambda_{\min}(G_\psi)^{-1} \leq C_\psi J$ for all $J \geq r$; (c) $\lambda_{\max}(G_\psi)/\lambda_{\min}(G_\psi) \leq C_\psi/c_\psi$ for all $J \geq r$.*

Lemma E.2. *Let X have support $[0, 1]^d$ and let $\psi_{J1}, \ldots, \psi_{JJ}$ be a B-spline basis formed as the tensor product of d univariate bases of order $r \geq 1$ with $m = J^{1/d} - r \geq 0$ interior knots and uniformly bounded mesh ratio. Then: (a) $\xi_{\psi,J} = 1$ for all $J \geq r^d$; (b) If the density of X is uniformly bounded* *away from* 0 *and* ∞ *on* $[0,1]^d$*, then there exists finite positive constants* c_ψ *and* C_ψ *such that* $c_\psi J \leq \lambda_{\max}(G_\psi)^{-1} \leq \lambda_{\min}(G_\psi)^{-1} \leq C_\psi J$ *for all* $J \geq r^d$*; (c)* $\lambda_{\max}(G_\psi)/\lambda_{\min}(G_\psi) \leq C_\psi/c_\psi$ *for all* $J \geq r^d$*.*

E.2 Wavelet bases

We construct a univariate wavelet basis with support $[0,1]$ following Cohen, Daubechies, and Vial (1993) (CDV hereafter). Let (φ,ψ) be a Daubechies pair such that φ has support $[-N+1,N]$. Given j such that $2^j-2N>0$, the orthonormal (with respect to the $L^2([0,1])$ inner product) basis for the space V_j includes 2^j-2N interior scaling functions of the form $\varphi_{j,k}(x)=2^{j/2}\varphi(2^jx-k)$, each of which has support $[2^{-j}(-N+1+k),2^{-j}(N+k)]$ for $k=N,\ldots,2^j-N-1$. These are augmented with N left scaling functions of the form $\varphi^0_{j,k}(x)=2^{j/2}\varphi^l_k(2^jx)$ for $k=0,\ldots,N-1$ (where $\varphi^l_0,\ldots,\varphi^l_{N-1}$ are fixed independent of j), each of which has support $[0,2^{-j}(N+k)]$, and N right scaling functions of the form $\varphi_{j,2^j-k}(x)=2^{j/2}\varphi^r_{-k}(2^j(x-1))$ for $k=1,\ldots,N$ (where $\varphi^r_{-1},\ldots,\varphi^r_{-N}$ are fixed independent of j), each of which has support $[1-2^{-j}(1-N-k),1]$. The resulting 2^j functions $\varphi^0_{j,0},\ldots,\varphi^0_{j,N-1},\varphi_{j,N},\ldots,\varphi_{j,2^j-N-1},\varphi^1_{j,2^j-N},\ldots,\varphi^1_{j,2^j-1}$ form an orthonormal basis (with respect to the $L^2([0,1])$ inner product) for their closed linear span V_j.

An orthonormal wavelet basis for the space W_j, defined as the orthogonal complement of V_j in V_{j+1}, is similarly constructed form the mother wavelet. This results in an orthonormal basis of 2^j functions, denoted $\psi^0_{j,0},\ldots,\psi^0_{j,N-1},\psi_{j,N},\ldots,\psi_{j,2^j-N-1},\psi^1_{j,2^j-N},\ldots,\psi^1_{j,2^j-1}$ (we use this conventional notation without confusion with the ψ_{Jj} basis functions spanning Ψ_J) where the "interior" wavelets $\psi_{j,N},\ldots,\psi_{j,2^j-N-1}$ are of the form $\psi_{j,k}(x)=2^{j/2}\psi(2^jx-k)$. To simplify notation we ignore the 0 and 1 superscripts on the left and right wavelets and scaling functions henceforth. Let L_0 and L be integers such that $2N<2^{L_0}\leq 2^L$. A wavelet space at resolution level L is the 2^{L+1}-dimensional set of functions given by

$$\mathrm{Wav}(L,[0,1])=\left\{\sum_{k=0}^{2^{L_0}-1}a_{L_0,k}\varphi_{L_0,k}+\sum_{j=L_0}^{L}\sum_{k=0}^{2^j-1}b_{j,k}\psi_{j,k}:a_{L_0,k},b_{j,k}\in\mathbb{R}\right\}.$$

We say that $\mathrm{Wav}(L,[0,1])$ has *regularity* γ if $\psi\in C^\gamma$ (which can be achieved by choosing N sufficiently large) and write $\mathrm{Wav}(L,[0,1],\gamma)$ for a wavelet space of regularity γ with continuously differentiable basis functions.

We construct wavelet bases for $[0,1]^d$ by taking tensor products of univariate bases. We again take L_0 and L to be integers such that $2N<2^{L_0}\leq 2^L$. Let $\widetilde{\psi}_{j,k,G}(x)$ denote an orthonormal tensor-product wavelet for $L^2([0,1]^d)$ at resolution level j where $k=(k_1,\ldots,k_d)\in\{0,\ldots,2^j-1\}^d$ and where $G\in G_{j,L}\subseteq\{w_\varphi,w_\psi\}^d$ denotes which elements of the tensor product are ψ_{j,k_i} (indices corresponding to w_ψ) and which are φ_{j,k_i} (indices corresponding to w_φ). For example, $\widetilde{\psi}_{j,k,w^d_\psi}=\prod_{i=1}^d\psi_{j,k_i}(x_i)$. Note that each $G\in G_{j,L}$ with $j>L$ has an element that is w_ψ (see Triebel (2006) for details). We have

$\#(G_{L_0,L_0}) = 2^d$, $\#(G_{j,L_0}) = 2^d - 1$ for $j > L_0$. Let $\mathrm{Wav}(L, [0,1]^d, \gamma)$ denote the space

$$\mathrm{Wav}(L, [0,1]^d, \gamma) = \left\{ \sum_{j=L_0}^{L} \sum_{G \in G_{j,L_0}} \sum_{k \in \{0,\dots,2^j-1\}^d} a_{j,k,G} \widetilde{\psi}_{j,k,G} : a_{j,k,G} \in \mathbb{R} \right\} \tag{25}$$

where each univariate basis has regularity γ. This definition clearly reduces to the above definition for $\mathrm{Wav}(L, [0,1], \gamma)$ in the univariate case.

Stability properties: The following two Lemmas bound $\xi_{\psi,J}$, as well as the minimum eigenvalue and condition number of $G_\psi = G_{\psi,J} = E[\psi^J(X_i)\psi^J(X_i)']$ when $\psi_{J1}, \dots, \psi_{JJ}$ is constructed using univariate and tensor-products of CDV wavelet bases.

Lemma E.3. *Let X have support $[0,1]$ and let be a univariate CDV wavelet basis of resolution level $L = \log_2(J) - 1$. Then: (a) $\xi_{\psi,J} = O(\sqrt{J})$ for each sieve dimension $J = 2^{L+1}$; (b) If the density of X is uniformly bounded away from 0 and ∞ on $[0,1]$, then there exists finite positive constants c_ψ and C_ψ such that $c_\psi \le \lambda_{\max}(G_\psi)^{-1} \le \lambda_{\min}(G_\psi)^{-1} \le C_\psi$ for each J; (c) $\lambda_{\max}(G_\psi)/\lambda_{\min}(G_\psi) \le C_\psi/c_\psi$ for each J.*

Lemma E.4. *Let X have support $[0,1]^d$ and let $\psi_{J1}, \dots, \psi_{JJ}$ be a wavelet basis formed as the tensor product of d univariate bases of resolution level L. Then: (a) $\xi_{\psi,J} = O(\sqrt{J})$ each J; (b) If the density of X is uniformly bounded away from 0 and ∞ on $[0,1]^d$, then there exists finite positive constants c_ψ and C_ψ such that $c_\psi \le \lambda_{\max}(G_\psi)^{-1} \le \lambda_{\min}(G_\psi)^{-1} \le C_\psi$ for each J; (c) $\lambda_{\max}(G_\psi)/\lambda_{\min}(G_\psi) \le C_\psi/c_\psi$ for each J.*

Wavelet characterization of Besov norms: When the wavelet basis just described is of regularity $\gamma > 0$, the norms $\|\cdot\|_{B^p_{\infty,\infty}}$ for $p < \gamma$ can be restated in terms of the wavelet coefficients. We briefly explain the multivariate case as it nests the univariate case. Any $f \in L^2([0,1]^d)$ may be represented as

$$f = \sum_{j,G,k} a_{j,k,G}(f) \widetilde{\psi}_{j,k,G}$$

with the sum is understood to be taken over the same indices as in display (25). If $f \in B^p_{\infty,\infty}([0,1]^d)$ then

$$\|f\|_{B^p_{\infty,\infty}} \asymp \|f\|_{b^p_{\infty,\infty}} := \sup_{j,k,G} 2^{j(p+d/2)} |a_{j,k,G}(f)| \, .$$

and if $f \in B^p_{2,2}([0,1])$ then

$$\|f\|^2_{B^p_{2,2}} \asymp \|f\|^2_{b^p_{2,2}} := \sum_{j,k,G} 2^{jp} a_{j,k,G}(f)^2$$

See Johnstone (2013) and Triebel (2006) for more thorough discussions.

F Useful results on random matrices

Notation: For a $r \times c$ matrix A with $r \leq c$ and full row rank r we let A_l^- denote its left pseudoinverse, namely $(A'A)^- A'$ where $'$ denotes transpose and $^-$ denotes generalized inverse. We let $s_{\min}(A)$ denote the minimum singular value of a rectangular matrix A. For a positive-definite symmetric matrix A we let $\lambda_{\min}(A)$ and $\lambda_{\max}(A)$ denote its minimum and maximum eigenvalue, respectively.

F.1 Some matrix inequalities

The following Lemmas are used throughout the proofs in this paper and are stated here for convenience.

Lemma F.1 (Weyl's inequality)**.** *Let $A, B \in \mathbb{R}^{r \times c}$ and let $s_i(A)$, $s_i(B)$ denote the ith (ordered) singular value of A and B respectively, for $1 \leq i \leq (r \wedge c)$. Then: $|s_i(A) - s_i(B)| \leq \|A - B\|_{\ell^2}$ for all $1 \leq i \leq (r \wedge c)$. In particular, $|s_{\min}(A) - s_{\min}(B)| \leq \|A - B\|_{\ell^2}$.*

Lemma F.2. *Let $A \in \mathbb{R}^{r \times r}$ be nonsingular. Then: $\|A^{-1} - I_r\|_{\ell^2} \leq \|A^{-1}\|_{\ell^2} \|A - I_r\|_{\ell^2}$.*

Lemma F.3 (Schmitt (1992))**.** *Let $A, B \in \mathbb{R}^{r \times r}$ be positive definite. Then:*

$$\|A^{1/2} - B^{1/2}\|_{\ell^2} \leq \frac{1}{\sqrt{\lambda_{\min}(B)} + \sqrt{\lambda_{\min}(A)}} \|A - B\|_{\ell^2} .$$

Lemma F.4. *Let $A, B \in \mathbb{R}^{r \times c}$ with $r \leq c$ and let A and B have full row rank r. Then:*

$$\|B_l^- - A_l^-\|_{\ell^2} \leq \frac{1 + \sqrt{5}}{2} (s_{\min}(A)^{-2} \vee s_{\min}(B)^{-2}) \|A - B\|_{\ell^2} .$$

If, in addition, $\|A - B\|_{\ell^2} \leq \frac{1}{2} s_{\min}(A)$ then

$$\|B_l^- - A_l^-\|_{\ell^2} \leq 2(1 + \sqrt{5}) s_{\min}(A)^{-2} \|A - B\|_{\ell^2} .$$

Lemma F.5. *Let $A \in \mathbb{R}^{r \times c}$ with $r \leq c$ have full row rank r. Then: $\|A_l^-\|_{\ell^2} \leq s_{\min}(A)^{-1}$.*

Lemma F.6. *Let $A, B \in \mathbb{R}^{r \times c}$ with $r \leq c$ and let A and B have full row rank r. Then:*

$$\|A'(AA')^{-1}A - B'(BB')^{-1}B\|_{\ell^2} \leq (s_{\min}(A)^{-1} \vee s_{\min}(B)^{-1}) \|A - B\|_{\ell^2} .$$

F.2 Convergence of the matrix estimators

Before presenting the following lemmas, we define the *orthonormalized* matrix estimators

$$\begin{aligned}
\widehat{G}_b^o &= G_b^{-1/2} \widehat{G}_b G_b^{-1/2} \\
\widehat{G}_\psi^o &= G_\psi^{-1/2} \widehat{G}_\psi G_\psi^{-1/2} \\
\widehat{S}^o &= G_b^{-1/2} \widehat{S} G_\psi^{-1/2}
\end{aligned}$$

and let $G_b^o = I_K$, $G_\psi^o = I_J$ and S^o denote their respective expected values.

Lemma F.7. *The orthonormalized matrix estimators satisfy the exponential inequalities:*

$$\begin{aligned}
\mathbb{P}\left(\|\widehat{G}_\psi^o - G_\psi^o\|_{\ell^2} > t\right) &\leq 2\exp\left\{\log J - \frac{t^2/2}{\zeta_{\psi,J}^2(1+2t/3)/n}\right\} \\
\mathbb{P}\left(\|\widehat{G}_b^o - G_b^o\|_{\ell^2} > t\right) &\leq 2\exp\left\{\log K - \frac{t^2/2}{\zeta_{b,K}^2(1+2t/3)/n}\right\} \\
\mathbb{P}\left(\|\widehat{S}^o - S^o\|_{\ell^2} > t\right) &\leq 2\exp\left\{\log K - \frac{t^2/2}{(\zeta_{b,K}^2 \vee \zeta_{\psi,J}^2)/n + 2\zeta_{b,K}\zeta_{\psi,J}t/(3n)}\right\}
\end{aligned}$$

and therefore

$$\begin{aligned}
\|\widehat{G}_\psi^o - G_\psi^o\|_{\ell^2} &= O_p(\zeta_{\psi,J}\sqrt{(\log J)/n}) \\
\|\widehat{G}_b^o - G_b^o\|_{\ell^2} &= O_p(\zeta_{b,K}\sqrt{(\log K)/n}) \\
\|\widehat{S}^o - S^o\|_{\ell^2} &= O_p((\zeta_{b,K} \vee \zeta_{\psi,J})\sqrt{(\log K)/n})\,.
\end{aligned}$$

as $n, J, K \to \infty$ provided $(\zeta_{b,K} \vee \zeta_{\psi,J})\sqrt{(\log K)/n} = o(1)$.

Lemma F.8 (Newey (1997), p. 162)**.** *Let Assumption 2(i) hold. Then: $\|G_b^{-1/2}B'u/n\|_{\ell^2} = O_p(\sqrt{K/n})$.*

Lemma F.9. *Let $h_J(x) = \psi^J(x)'c_J$ for any deterministic $c_J \in \mathbb{R}^J$ and $H_J = (h_J(X_1), \ldots, h_J(X_n))' = \Psi c_J$. Then:*

$$\begin{aligned}
&\|G_b^{-1/2}(B'(H_0 - \Psi c_J)/n - E[b^K(W_i)(h_0(X_i) - h_J(X_i))])\|_{\ell^2} \\
&= O_p\left(\left(\sqrt{K/n} \times \|h_0 - h_J\|_\infty\right) \wedge \left(\zeta_{b,K}/\sqrt{n} \times \|h_0 - h_J\|_{L^2(X)}\right)\right)\,.
\end{aligned}$$

Lemma F.10. *Let $s_{JK}^{-1}\zeta\sqrt{(\log J)/n} = o(1)$ and let $J \leq K = O(J)$. Then:*

(a) $\|(\widehat{G}_b^{-1/2}\widehat{S})_l^- \widehat{G}_b^{-1/2} G_b^{1/2} - (G_b^{-1/2}S)_l^-\|_{\ell^2} = O_p\left(s_{JK}^{-2}\zeta\sqrt{(\log J)/(ne_J)}\right)$

(b) $\|G_\psi^{1/2}\{(\widehat{G}_b^{-1/2}\widehat{S})_l^- \widehat{G}_b^{-1/2} G_b^{1/2} - (G_b^{-1/2}S)_l^-\}\|_{\ell^2} = O_p\left(s_{JK}^{-2}\zeta\sqrt{(\log J)/n)}\right)$

(c) $\|G_b^{-1/2}S\{(\widehat{G}_b^{-1/2}\widehat{S})_l^- \widehat{G}_b^{-1/2} G_b^{1/2} - (G_b^{-1/2}S)_l^-\}\|_{\ell^2} = O_p\left(s_{JK}^{-1}\zeta\sqrt{(\log J)/n}\right)$.

References

Chen, X. and T. M. Christensen (2015). Optimal uniform convergence rates and asymptotic normality for series estimators under weak dependence and weak conditions. *Journal of Econometrics 188*(2), 447 – 465.

Chen, X. and D. Pouzo (2015). Sieve Wald and QLR inferences on semi/nonparametric conditional moment models. *Econometrica 83*(3), 1013–1079.

Cohen, A., I. Daubechies, and P. Vial (1993). Wavelets on the interval and fast wavelet transforms. *Applied and Computational Harmonic Analysis 1*, 54–81.

DeVore, R. A. and G. G. Lorentz (1993). *Constructive Approximation*. Grundlehren der mathematischen Wissenschaften 303. Springer-Verlag, Berlin.

Johnstone, I. M. (2013). Gaussian estimation: Sequence and wavelet models. Manuscript.

Newey, W. K. (1997). Convergence rates and asymptotic normality for series estimators. *Journal of Econometrics 79*(1), 147–168.

Schmitt, B. A. (1992). Perturbation bounds for matrix square roots and pythagorean sums. *Linear Algebra and its Applications 174*(0), 215–227.

Triebel, H. (2006). *Theory of Function Spaces III*. Birkhäuser, Basel.

Secondary Online Appendix to

Optimal Sup-norm Rates and Uniform Inference on Nonlinear Functionals of Nonparametric IV Regression

Xiaohong Chen Timothy M. Christensen

First version: August 2013. Revised: January 2017.

G Supplementary Lemmas and Proofs

All the notation follow from the main text and the main online appendix. For a $r \times c$ matrix A with $r \leq c$ and full row rank r we let A_l^- denote its left pseudoinverse, namely $(A'A)^- A'$ where $'$ denotes transpose and $^-$ denotes generalized inverse. We let $s_{\min}(A)$ denote the minimum singular value of a rectangular matrix A.

Let $s_{JK} = s_{\min}(G_b^{-1/2} S G_\psi^{-1/2})$. Throughout the proofs in the appendix we use the identity

$$\begin{aligned} \psi^J(x)'(G_b^{-1/2}S)_l^- &= \psi^J(x)'(S'G_b^{-1}S)^{-1}S'G_b^{-1/2} \\ &= \psi^J(x)'G_\psi^{-1/2}(G_\psi^{-1/2}S'G_b^{-1}SG_\psi^{-1/2})^{-1}G_\psi^{-1/2}S'G_b^{-1/2} \\ &= \psi^J(x)'G_\psi^{-1/2}(G_b^{-1/2}SG_\psi^{-1/2})_l^- \end{aligned}$$

which implies that

$$\begin{aligned} \|\psi^J(x)'(G_b^{-1/2}S)_l^-\|_{\ell^2} &\leq \|\psi^J(x)'G_\psi^{-1/2}\|_{\ell^2}\|(G_b^{-1/2}SG_\psi^{-1/2})_l^-\|_{\ell^2} \\ &\leq \zeta_{\psi,J}\|(G_b^{-1/2}SG_\psi^{-1/2})_l^-\|_{\ell^2} \\ &\leq \zeta_{\psi,J}s_{JK}^{-1} \end{aligned} \tag{26}$$

by definition of $\zeta_{\psi,J}$ and the fact that $\|A_l^-\|_{\ell^2} \leq s_{\min}(A)^{-1}$ (see Lemma F.5).

G.1 Proofs for Appendix A and Section 3.1

Since the proofs of results in Section 3.1 built upon those for results in Appendix A, we shall present the proofs for Appendix A first.

G.1.1 Proofs for Appendix A

Proof of Lemma A.1. First note that $\tau_J > 0$ for all J by compactness and injectivity of T. Then:

$$s_{JK} = \inf_{h \in \Psi_J : \|h\|_{L^2(X)}=1} \|\Pi_K T h\|_{L^2(W)} \leq \inf_{h \in \Psi_J : \|h\|_{L^2(X)}=1} \|Th\|_{L^2(W)} = \tau_J^{-1}$$

holds uniformly in J because Π_K is a contraction, whence $\tau_J \le s_{JK}^{-1}$. To derive a lower bound on τ_J, the triangle inequality and Assumption 4(i) yield:

$$
\begin{aligned}
s_{JK} &= \inf_{h\in\Psi_{J,1}} \|\Pi_K T h\|_{L^2(W)} \\
&\ge \inf_{h\in\Psi_{J,1}} \|Th\|_{L^2(W)} - \sup_{h\in\Psi_{J,1}} \|(\Pi_K T - T)h\|_{L^2(W)} \\
&= (1-o(1))\tau_J^{-1}\,.
\end{aligned}
$$

Therefore, $s_{JK}^{-1} \le (1-o(1))^{-1}\tau_J$. $\square$

It is clear that Lemma A.2 is implied by the following lemma.

Lemma G.1. *Let Assumptions 1(iii) and 4(ii) hold. Then:*
(1) (a) $\|h_0 - \pi_J h_0\|_{L^2(X)} \asymp \|h_0 - \Pi_J h_0\|_{L^2(X)}$*; and*
(b) $\tau_J \times \|T(h_0 - \pi_J h_0)\|_{L^2(W)} \le \text{const} \times \|h_0 - \pi_J h_0\|_{L^2(X)}$.
(2) If Assumption 4(i) also holds, then: (a) $\|Q_J h_0 - \pi_J h_0\|_{L^2(X)} \le o(1) \times \|h_0 - \pi_J h_0\|_{L^2(X)}$*; and*
(b) $\|h_0 - \Pi_J h_0\|_{L^2(X)} \asymp \|h_0 - Q_J h_0\|_{L^2(X)}$.
(3) If Assumption 4(iii') also holds, then: $\|Q_J h_0 - \pi_J h_0\|_\infty \le O(1) \times \|h_0 - \pi_J h_0\|_{L^2(X)}$.
(4) Further, if Condition (24) also holds, then Assumption 4(iii) is satisfied.

Proof of Lemma G.1. In what follows, "const" denotes a generic positive constant that may be different from line to line. Assumption 1(iii) guarantees τ_J and $\pi_J h_0$ are well defined. For part (1.a), we have:

$$
\begin{aligned}
\|h_0 - \Pi_J h_0\|_{L^2(X)} &\le \|h_0 - \pi_J h_0\|_{L^2(X)} \\
&\le \|h_0 - \Pi_J h_0\|_{L^2(X)} + \|\Pi_J h_0 - \pi_J h_0\|_{L^2(X)} \\
&\le \|h_0 - \Pi_J h_0\|_{L^2(X)} + \tau_J \|T(\pi_J h_0 - \Pi_J h_0)\|_{L^2(W)} \\
&= \|h_0 - \Pi_J h_0\|_{L^2(X)} + \tau_J \|T \pi_J(h_0 - \Pi_J h_0)\|_{L^2(W)} \\
&\le \|h_0 - \Pi_J h_0\|_{L^2(X)} + \tau_J \|T(h_0 - \Pi_J h_0)\|_{L^2(W)} \\
&= (1+\text{const}) \times \|h_0 - \Pi_J h_0\|_{L^2(X)}
\end{aligned}
$$

where the third line is by definition of τ_J, the fourth is because $\pi_J h = h$ for all $h \in \Psi_J$, the final line is by Assumption 4(ii), and the fifth is because π_J is a weak contraction under the norm $h \mapsto \|Th\|_{L^2(W)}$. More precisely by the definition of $\pi_J h_0$ we have:

$$
\langle Th, T(h_0 - \pi_J h_0)\rangle_W = 0 \tag{27}
$$

for all $h \in \Psi_J$, where $\langle \cdot,\cdot\rangle_W$ denotes the $L^2(W)$ inner product. With $h = \pi_J h_0 - \Pi_J h_0 \in \Psi_J$ this implies

$$
\langle T(\pi_J h_0 - \Pi_J h_0), T(h_0 - \pi_J h_0)\rangle_W = 0.
$$

$$
\langle T(\pi_J h_0 - \Pi_J h_0), T(h_0 - \Pi_J h_0)\rangle_W = \langle T(\pi_J h_0 - \Pi_J h_0), T(\pi_J h_0 - \Pi_J h_0)\rangle_W .
$$

Thus $\|T(\pi_J h_0 - \Pi_J h_0)\|_{L^2(W)} \le \|T(h_0 - \Pi_J h_0)\|_{L^2(W)}$.

For part (1.b):

$$\begin{aligned}\tau_J\|T(h_0-\pi_J h_0)\|_{L^2(W)} &\leq \tau_J\|T(h_0-\Pi_J h_0)\|_{L^2(W)}\\ &\leq \text{const}\times\|h_0-\Pi_J h_0\|_{L^2(X)}\\ &\leq \text{const}\times\|h_0-\pi_J h_0\|_{L^2(X)}\end{aligned}$$

where the first and final inequalities are by definition of $\pi_J h_0$ and $\Pi_J h_0$ and the second inequality is by Assumption 4(ii).

For part (2.a), Lemma A.1 guarantees that $Q_J h_0$ is well defined and that $s_{JK}^{-1} \leq 2\tau_J$ for all J sufficiently large. By definition of $Q_J h_0$ we have:

$$\langle \Pi_K Th, T(h_0-Q_J h_0)\rangle_W = 0 \tag{28}$$

for all $h \in \Psi_J$, where we use the fact that $\langle \Pi_K f, g\rangle_W = \langle \Pi_K f, \Pi_K g\rangle_W$ holds for any $f, g \in L^2(W)$ since Π_K is a projection). Substituting $h = Q_J h_0 - \pi_J h_0 \in \Psi_J$ into the two equations (27) and (28) yields:

$$\langle (T-\Pi_K T)(Q_J h_0-\pi_J h_0), T(h_0-\pi_J h_0)\rangle_W + \langle \Pi_K T(Q_J h_0-\pi_J h_0), T(h_0-\pi_J h_0)\rangle_W = 0 \tag{29}$$

$$\langle \Pi_K T(Q_J h_0-\pi_J h_0), T(h_0-Q_J h_0)\rangle_W = 0\,. \tag{30}$$

By subtracting (30) from (29) we obtain

$$\langle (T-\Pi_K T)(Q_J h_0-\pi_J h_0), T(h_0-\pi_J h_0)\rangle_W + \|\Pi_K T(Q_J h_0-\pi_J h_0)\|^2_{L^2(W)} = 0$$

We have therefore proved

$$\|\Pi_K T(Q_J h_0-\pi_J h_0)\|^2_{L^2(W)} = |\langle (T-\Pi_K T)(Q_J h_0-\pi_J h_0), T(h_0-\pi_J h_0)\rangle_W|\,. \tag{31}$$

It follows from (31), the Cauchy-Schwarz inequality, and Assumption 4(i) that:

$$\begin{aligned}s_{JK}^2\|Q_J h_0-\pi_J h_0\|^2_{L^2(X)} &\leq \|\Pi_K T(Q_J h_0-\pi_J h_0)\|^2_{L^2(W)}\\ &\leq \|(T-\Pi_K T)(Q_J h_0-\pi_J h_0)\|_{L^2(W)}\|T(h_0-\pi_J h_0)\|_{L^2(W)} &(32)\\ &\leq o(\tau_J^{-1})\|Q_J h_0-\pi_J h_0\|_{L^2(X)}\|T(h_0-\pi_J h_0)\|_{L^2(W)}\,. &(33)\end{aligned}$$

It follows by (33) and the relation $s_{JK}^{-1} \leq 2\tau_J$ for all J large that:

$$\begin{aligned}\|Q_J h_0-\pi_J h_0\|_{L^2(X)} &\leq o(1)\times\tau_J\|T(h_0-\pi_J h_0)\|_{L^2(W)}\\ &\leq o(1)\times\text{const}\times\|h_0-\pi_J h_0\|_{L^2(X)}\end{aligned}$$

where the final line is by part (1.b). For part (2.b), by definition of Q_J, Π_J and results in part (1.a) and part (2.a), we have:

$$\begin{aligned}\|h_0-\Pi_J h_0\|_{L^2(X)} &\leq \|h_0-Q_J h_0\|_{L^2(X)}\\ &\leq \|h_0-\pi_J h_0\|_{L^2(X)} + \|\pi_J h_0-Q_J h_0\|_{L^2(X)}\\ &\leq \|h_0-\pi_J h_0\|_{L^2(X)} + o(1)\times\|h_0-\pi_J h_0\|_{L^2(X)}\\ &= (1+\text{const})\times\|h_0-\Pi_J h_0\|_{L^2(X)}.\end{aligned}$$

This proves part (2.b).

For part (3), it follows from (32) and Assumption 4(iii') that

$$s_{JK}^2 \|Q_J h_0 - \pi_J h_0\|_{L^2(X)} \le \text{const} \times (\zeta_{\psi,J}\tau_J)^{-1} \|T(h_0 - \pi_J h_0)\|_{L^2(W)} \,.$$

and hence

$$\begin{aligned} \|Q_J h_0 - \pi_J h_0\|_{L^2(X)} &\le \text{const} \times \zeta_{\psi,J}^{-1} \times \tau_J \|T(h_0 - \pi_J h_0)\|_{L^2(W)} \\ &\le \zeta_{\psi,J}^{-1} \times \text{const} \times \|h_0 - \pi_J h_0\|_{L^2(X)} \end{aligned} \tag{34}$$

by Part (1.b) and the fact that $s_{JK}^{-1} \le 2\tau_J$ for all J large. Therefore,

$$\begin{aligned} \|Q_J h_0 - \pi_J h_0\|_\infty &\le \zeta_\psi \|Q_J h_0 - \pi_J h_0\|_{L^2(X)} \\ &\le \text{const} \times \|h_0 - \pi_J h_0\|_{L^2(X)} \end{aligned}$$

where the last inequality is due to (34).

For part (4), by the triangle inequality, the results in part (1.a) and (3) and Condition (24) we have:

$$\begin{aligned} \|Q_J(h_0 - \Pi_J h_0)\|_\infty &\le \|Q_J h_0 - \pi_J h_0\|_\infty + \|\pi_J h_0 - \Pi_J h_0\|_\infty \\ &\le \text{const} \times \|h_0 - \Pi_J h_0\|_{L^2(X)} + \|h_0 - \Pi_J h_0\|_\infty \\ &\le O(1) \times \|h_0 - \Pi_J h_0\|_\infty \end{aligned}$$

which completes the proof. □

Note that we may write $\Pi_J h_0(x) = \psi^J(x)' c_J$ for some c_J in $\mathbb{R}^J$. We use this notation hereafter.

Proof of Lemma A.3. We first prove **Result (1)**. We begin by writing

$$\begin{aligned} \widetilde{h}(x) - \Pi_J h_0(x) &= Q_J(h_0 - \Pi_J h_0)(x) \\ &\quad + \psi^J(x)'(G_b^{-1/2}S)_l^- \{G_b^{-1/2}(B'(H_0 - \Psi c_J)/n - E[b^K(W_i)(h_0(X_i) - \Pi_J h_0(X_i))])\} \\ &\quad + \psi^J(x)'\{(\widehat{G}_b^{-1/2}\widehat{S})_l^- \widehat{G}_b^{-1/2} G_b^{1/2} - (G_b^{-1/2}S)_l^-\} G_b^{-1/2} B'(H_0 - \Psi c_J)/n \\ &=: T_1 + T_2 + T_3 \end{aligned}$$

where $Q_J : L^2(X) \to \Psi_J$ is the sieve 2SLS projection operator given by

$$Q_J h(x) = \psi^J(x)'[S' G_b^{-1} S]^{-1} S' G_b^{-1} E[b^K(W_i) h(X_i)] \,.$$

Note that $Q_J h = h$ for all $h \in \Psi_J$.

Control of $\|T_1\|_\infty$: $\|T_1\|_\infty = O(1) \times \|h_0 - \Pi_J h_0\|_\infty$ by Assumption 4(iii).

Control of $\|T_2\|_\infty$: Using equation (26), the Cauchy-Schwarz inequality, and Lemma F.9, we obtain:

$$\begin{aligned} \|T_2\|_\infty &\le \sup_x \|\psi^J(x)(G_b^{-1/2}S)_l^-\|_{\ell^2} \|G_b^{-1/2}(B'(H_0 - \Psi c_J)/n - E[b^K(W_i)(h_0(X_i) - \Pi_J h_0(X_i))])\|_{\ell^2} \\ &\le \zeta_{\psi,J} s_{JK}^{-1} \|G_b^{-1/2}(B'(H_0 - \Psi c_J)/n - E[b^K(W_i)(h_0(X_i) - \Pi_J h_0(X_i))])\|_{\ell^2} \\ &= \zeta_{\psi,J} s_{JK}^{-1} \times O_p(\sqrt{K/n}) \times \|h_0 - \Pi_J h_0\|_\infty \,. \end{aligned}$$

It then follows by the relations $s_{JK}^{-1} \asymp \tau_J$ (Lemma A.1) and $\zeta_{\psi,J} \geq \sqrt{J} \asymp \sqrt{K}$ and Assumption 3(ii) that:

$$\|T_2\|_\infty = O_p(\tau_J \zeta_{\psi,J} \sqrt{J/n}) \times \|h_0 - \Pi_J h_0\|_\infty = O_p(1) \times \|h_0 - \Pi_J h_0\|_\infty .$$

Control of $\|T_3\|_\infty$: Similar to T_2 in the proof of Lemma 3.1, we may use Lemmas F.10(b) and A.1 to obtain:

$$\begin{aligned} \|T_3\|_\infty &\leq \zeta_{\psi,J} \|G_\psi^{1/2} \{(\widehat{G}_b^{-1/2} \widehat{S})_l^- \widehat{G}_b^{-1/2} G_b^{1/2} - (G_b^{-1/2} S)_l^- \}\|_{\ell^2} \|G_b^{-1/2} B'(H_0 - \Psi c_J)/n\|_{\ell^2} \\ &= \zeta_{\psi,J} \times O_p(\tau_J^2 \zeta \sqrt{(\log J)/n)}) \times \|G_b^{-1/2} B'(H_0 - \Psi c_J)/n\|_{\ell^2} . \end{aligned} \tag{35}$$

Then by Lemma F.9 and the triangle inequality, we have:

$$\begin{aligned} \|G_b^{-1/2} B'(H_0 - \Psi c_J)/n\|_{\ell^2} &\leq O_p(\sqrt{K/n}) \times \|h_0 - \Pi_J h_0\|_\infty + \|\Pi_K T(h_0 - \Pi_J h_0)\|_{L^2(W)} \\ &\leq O_p(\sqrt{K/n}) \times \|h_0 - \Pi_J h_0\|_\infty + \|T(h_0 - \Pi_J h_0)\|_{L^2(W)} . \end{aligned} \tag{36}$$

Substituting (36) into (35) and using Assumptions 3(ii) and 4(ii):

$$\begin{aligned} \|T_3\|_\infty &\leq O_p(\tau_J \zeta^2/\sqrt{n}) \times \Big(O_p(\tau_J \sqrt{K(\log J)/n}) \times \|h_0 - \Pi_J h_0\|_\infty + \tau_J \|T(h_0 - \Pi_J h_0)\|_{L^2(W)}\Big) \\ &= O_p(1) \times \big(o_p(1) \times \|h_0 - \Pi_J h_0\|_\infty + O_p(1) \times \|h_0 - \Pi_J h_0\|_{L^2(X)}\big) \\ &\leq O_p(1) \times \|h_0 - \Pi_J h_0\|_\infty \end{aligned}$$

where the final line is by the relation between the $L^2(X)$ and sup norms.

Result (2) then follows because

$$\begin{aligned} \|\widetilde{h} - h_0\|_\infty &\leq \|\widetilde{h} - \Pi_J h_0\|_\infty + \|\Pi_J h_0 - h_0\|_\infty \\ &\leq (1 + O_p(1)) \|\Pi_J h_0 - h_0\|_\infty \\ &\leq (1 + O_p(1))(1 + \|\Pi_J\|_\infty) \|h_0 - h_{0,J}\|_\infty . \end{aligned}$$

where the second inequality is by Result (1) and the final line is by Lebesgue's lemma. □

G.1.2 Proofs for Section 3.1

Proof of Lemma 3.1. Let $u = (u_1, \ldots, u_n)'$. Let M_n be a sequence of positive constants diverging to $+\infty$, and decompose $u_i = u_{1,i} + u_{2,i}$ where

$$\begin{aligned} u_{1,i} &= u_i\{|u_i| \leq M_n\} - E[u_i\{|u_i| \leq M_n\}|W_i] \\ u_{2,i} &= u_i\{|u_i| > M_n\} - E[u_i\{|u_i| > M_n\}|W_i] \\ u_1 &= (u_{1,1}, \ldots, u_{1,n})' \\ u_2 &= (u_{2,1}, \ldots, u_{2,n})' . \end{aligned}$$

For **Result (1)**, recall that $\xi_{\psi,J} = \sup_x \|\psi^J(x)\|_{\ell^1}$. By Hölder's inequality we have

$$\|\widehat{h} - \widetilde{h}\|_\infty = \sup_x |\psi^J(x)'(\widehat{c} - \widetilde{c})| \leq \xi_{\psi,J} \|(\widehat{c} - \widetilde{c})\|_{\ell^\infty} .$$

To derive the sup-norm convergence rate of the standard deviation term $\widehat{h}-\widetilde{h}$, it suffices to bound the ℓ^∞ norm of the $J\times 1$ random vector $(\widehat{c}-\widetilde{c})$. Although this appears like a crude bound, $\xi_{\psi,J}$ grows slowly in J for certain sieves whose basis functions have local support. For such bases the above bound, in conjunction with the following result

$$\|\widehat{c}-\widetilde{c}\|_{\ell^\infty} = O_p\left(s_{JK}^{-1}\sqrt{(\log J)/(ne_J)}\right) \tag{37}$$

leads to a tight bound on the convergence rate of $\|\widehat{h}-\widetilde{h}\|_\infty$.

To prove (37), we begin by writing

$$\begin{aligned}
\widehat{c}-\widetilde{c} &= (\widehat{G}_b^{-1/2}\widehat{S})_l^-\widehat{G}_b^{-1/2}B'u/n \\
&= (G_b^{-1/2}S)_l^- G_b^{-1/2}B'u/n + \{(\widehat{G}_b^{-1/2}\widehat{S})_l^-\widehat{G}_b^{-1/2} - (G_b^{-1/2}S)_l^- G_b^{-1/2}\}B'u/n \\
&=: T_1+T_2\,.
\end{aligned}$$

We will show that $\|T_1\|_{\ell^\infty} = O_p\left(s_{JK}^{-1}\sqrt{(\log J)/(ne_J)}\right)$ and $\|T_2\|_{\ell^\infty} = O_p\left(s_{JK}^{-1}\sqrt{(\log J)/(ne_J)}\right)$.

Control of $\|T_1\|_{\ell^\infty}$. Note that $T_1 = (G_b^{-1/2}S)_l^- G_b^{-1/2}B'u_1/n + (G_b^{-1/2}S)_l^- G_b^{-1/2}B'u_2/n$.

Let $(a)_j$ denote the jth element of a vector a. By the definition of $\|\cdot\|_{\ell^\infty}$ and the union bound,

$$\begin{aligned}
\mathbb{P}\left(\|(G_b^{-1/2}S)_l^- G_b^{-1/2}B'u_1/n\|_{\ell^\infty} > t\right) &\le \mathbb{P}\left(\bigcup_{j=1}^J |((G_b^{-1/2}S)_l^- G_b^{-1/2}B'u_1/n)_j| > t\right) \\
&\le \sum_{j=1}^J \mathbb{P}\left(|((G_b^{-1/2}S)_l^- G_b^{-1/2}B'u_1/n)_j| > t\right) \\
&= \sum_{j=1}^J \mathbb{P}\left(\left|\sum_{i=1}^n q_{j,JK}(W_i)u_{1,i}/n\right| > t\right)
\end{aligned} \tag{38}$$

where $q_{j,JK}(W_i) = ((G_b^{-1/2}S)_l^- G_b^{-1/2}b^K(W_i))_j$. The summands may be bounded by noting that

$$\begin{aligned}
|q_{j,JK}(W_i)| &\le \|(G_b^{-1/2}S)_l^-\|_{\ell^2}\|G_b^{-1/2}b^K(W_i)\|_{\ell^2} \\
&= \|G_\psi^{-1/2}[G_\psi^{-1/2}S'G_b^{-1}SG_\psi^{-1/2}]^{-1}G_\psi^{-1/2}S'G_b^{-1/2}\|_{\ell^2}\|G_b^{-1/2}b^K(W_i)\|_{\ell^2} \\
&\le \|G_\psi^{-1/2}\|_{\ell^2}\|[G_\psi^{-1/2}S'G_b^{-1}SG_\psi^{-1/2}]^{-1}G_\psi^{-1/2}S'G_b^{-1/2}\|_{\ell^2}\|G_b^{-1/2}b^K(W_i)\|_{\ell^2} \\
&\le \frac{\zeta_{b,K}}{s_{JK}\sqrt{e_J}}
\end{aligned} \tag{39}$$

uniformly in i and j. Therefore,

$$|q_{j,JK}(W_i)u_{1,i}/n| \le \frac{2M_n\zeta_{b,K}}{ns_{JK}\sqrt{e_J}} \tag{40}$$

uniformly in i and j.

Let $(A)_{j|}$ denote the jth row of the matrix A and let $(A)_{jj}$ denote its jth diagonal element. The second moments of the summands may be bounded by observing that

$$
\begin{aligned}
E[q_{j,JK}(W_i)^2] &= E[((G_b^{-1/2}S)_l^-)_{j|}G_b^{-1/2}b^K(W_i))^2] \\
&= E[((G_b^{-1/2}S)_l^-)_{j|}G_b^{-1/2}b^K(W_i)b^K(W_i)'G_b^{-1/2}((G_b^{-1/2}S)_l^-)'_{j|}] \\
&= ((G_b^{-1/2}S)_l^-((G_b^{-1/2}S)_l^-)')_{jj} \\
&= ((S'G_b^{-1}S)^{-1})_{jj} \\
&\leq \|(S'G_b^{-1}S)^{-1}\|_{\ell^2} \\
&= \|G_\psi^{-1/2}[G_\psi^{-1/2}S'G_b^{-1}SG_\psi^{-1/2}]^{-1}G_\psi^{-1/2}\|_{\ell^2} \\
&\leq \frac{1}{s_{JK}^2 e_J} \qquad (41)
\end{aligned}
$$

and so

$$
E[(q_{j,JK}(W_i)u_{1,i}/n)^2] \leq \frac{\overline{\sigma}^2}{n^2 s_{JK}^2 e_J} \qquad (42)
$$

by Assumption 2(i) and the law of iterated expectations. Bernstein's inequality and expressions (38), (40) and (42) yield

$$
\begin{aligned}
&\mathbb{P}\left(\|(G_b^{-1/2}S)_l^- G^{-1/2}B'u_1/n\|_{\ell^\infty} > Cs_{JK}^{-1}\sqrt{(\log J)/(ne_J)}\right) \\
&\leq \quad 2\exp\left\{\log J - \frac{C^2(\log J)/(ns_{JK}^2 e_J)}{c_1/(ns_{JK}^2e_J) + c_2 C M_n \zeta_{b,K}\sqrt{\log J}/(n^{3/2}s_{JK}^2 e_J)}\right\} \\
&= \quad 2\exp\left\{\log J - \frac{C^2(\log J)/(ns_{JK}^2 e_J)}{1/(ns_{JK}^2 e_J)[c_1 + c_2 C M_n \zeta_{b,K}\sqrt{(\log J)/n}]}\right\} \qquad (43)
\end{aligned}
$$

for finite positive constants c_1 and c_2. Then (43) is $o(1)$ for all large C provided $M_n\zeta_{b,K}\sqrt{(\log J)/n} = o(1)$.

By the triangle and Markov inequalities and (39), we have

$$
\begin{aligned}
\mathbb{P}\left(\|(G_b^{-1/2}S)_l^- G^{-1/2}B'u_2/n\|_{\ell^\infty} > t\right) &= \mathbb{P}\left(\max_{1\leq j\leq J}\left|\sum_{i=1}^n q_{j,JK}(W_i)u_{2,i}/n\right| > t\right) \\
&\leq \mathbb{P}\left(\frac{\zeta_{b,K}}{s_{JK}\sqrt{e_J}}\sum_{i=1}^n |u_{2,i}/n| > t\right) \\
&\leq \frac{2\zeta_{b,K}}{ts_{JK}\sqrt{e_J}}E[|u_i|\{|u_i| > M_n\}] \\
&\leq \frac{2\zeta_{b,K}}{ts_{JK}\sqrt{e_J}M_n^{1+\delta}}E[|u_i|^{2+\delta}\{|u_i| > M_n\}]
\end{aligned}
$$

which, by Assumption 2(ii), is $o(1)$ when $t = Cs_{JK}^{-1}\sqrt{(\log J)/(ne_J)}$ provided $\zeta_{b,K}\sqrt{n/(\log J)} = O(M_n^{1+\delta})$.

Choosing $M_n^{1+\delta} \asymp \zeta_{b,K}\sqrt{n/\log J}$ satisfies the condition $\zeta_{b,K}\sqrt{n/(\log J)} = O(M_n^{1+\delta})$ trivially, and satisfies the condition $M_n\zeta_{b,K}\sqrt{(\log J)/n} = o(1)$ provided $\zeta_{b,K}^{(2+\delta)/\delta}\sqrt{(\log J)/n} = o(1)$, which holds by Assumption 3(iii).

Control of $\|T_2\|_{\ell^\infty}$: Using the fact that $\|\cdot\|_{\ell^\infty} \leq \|\cdot\|_{\ell^2}$ on $\mathbb{R}^J$ and Lemmas F.10(a) and F.8, we have:

$$\begin{aligned}
\|T_2\|_{\ell^\infty} &= \|\{(\widehat{G}_b^{-1/2}\widehat{S})_l^- \widehat{G}_b^{-1/2} G_b^{1/2} - (G_b^{-1/2}S)_l^-\} G_b^{-1/2} B'u/n\|_{\ell^\infty} \\
&\leq \|(\widehat{G}_b^{-1/2}\widehat{S})_l^- \widehat{G}_b^{-1/2} G_b^{1/2} - (G_b^{-1/2}S)_l^-\|_{\ell^2} \|G_b^{-1/2} B'u/n\|_{\ell^2} \\
&= O_p\left(s_{JK}^{-2}\zeta\sqrt{(\log K)/(ne_J)}\right) \times O_p(\sqrt{K/n}) \\
&= O_p\left(s_{JK}^{-1}\sqrt{(\log J)/(ne_J)}\right) \times O_p(s_{JK}^{-1}\zeta\sqrt{K/n}) \\
&= O_p\left(s_{JK}^{-1}\sqrt{(\log J)/(ne_J)}\right)
\end{aligned}$$

where the last equality follows from Assumption 3(ii) and the facts that $\zeta \geq \sqrt{K}$ and $J \asymp K$.

For **Result (2)**, we begin by writing

$$\begin{aligned}
\widehat{h}(x) - \widetilde{h}(x) &= \psi^J(x)'(\widehat{G}_b^{-1/2}\widehat{S})_l^- \widehat{G}_b^{-1/2} B'u/n \\
&= \psi^J(x)'(G_b^{-1/2}S)_l^- G_b^{-1/2} B'u/n + \psi^J(x)'\{(\widehat{G}_b^{-1/2}\widehat{S})_l^- \widehat{G}_b^{-1/2} - (G_b^{-1/2}S)_l^- G_b^{-1/2}\} B'u/n \\
&=: T_1 + T_2\,.
\end{aligned}$$

We will show that $\|T_1\|_\infty = O_p\left(\tau_J \zeta_{\psi,J}\sqrt{(\log n)/n}\right)$ and $\|T_2\|_\infty = O_p\left(\tau_J \zeta_{\psi,J}\sqrt{(\log n)/n}\right)$.

Control of $\|T_1\|_\infty$. Note that $T_1 = \psi^J(x)'(G_b^{-1/2}S)_l^- G_b^{-1/2} B'u_1/n + \psi^J(x)'(G_b^{-1/2}S)_l^- G_b^{-1/2} B'u_2/n$.

Let $\mathcal{X}_n \subset \mathcal{X}$ be a grid of finitely many points such that for each $x \in \mathcal{X}$ there exits a $\bar{x}_n(x) \in \mathcal{X}_n$ such that $\|x - \bar{x}_n(x)\| \lesssim (\zeta_{\psi,J} J^{-(\omega+\frac{1}{2})})^{1/\omega'}$, where ω, ω' are as in Assumption 3(i). By compactness and convexity of the support $\mathcal{X}$ of X_i, we may choose $\mathcal{X}_n$ to have cardinality $\#(\mathcal{X}_n) \lesssim n^\beta$ for some $0 < \beta < \infty$. Therefore,

$$\begin{aligned}
&\sup_x \|\psi^J(x)'(G_b^{-1/2}S)_l^- G_b^{-1/2} B'u_1/n\|_\infty \\
&\leq \max_{x_n \in \mathcal{X}_n} |\psi^J(x_n)'(G_b^{-1/2}S)_l^- G_b^{-1/2} B'u_1/n| + \sup_x |\{\psi^J(x) - \psi^J(\bar{x}_n(x))\}'(G_b^{-1/2}S)_l^- G_b^{-1/2} B'u_1/n| \\
&\leq \max_{x_n \in \mathcal{X}_n} |\psi^J(x_n)'(G_b^{-1/2}S)_l^- G_b^{-1/2} B'u_1/n| + C_\omega J^\omega \zeta_{\psi,J} J^{-(\omega+\frac{1}{2})} s_{JK}^{-1} \|G_b^{-1/2} B'u_1/n\|_{\ell^2} \\
&= \max_{x_n \in \mathcal{X}_n} |\psi^J(x_n)'(G_b^{-1/2}S)_l^- G_b^{-1/2} B'u_1/n| + C_\omega J^\omega \zeta_{\psi,J} J^{-(\omega+\frac{1}{2})} s_{JK}^{-1} \times O_p(\sqrt{J/n}) \\
&= \max_{x_n \in \mathcal{X}_n} |\psi^J(x_n)'(G_b^{-1/2}S)_l^- G_b^{-1/2} B'u_1/n| + o_p(s_{JK}^{-1}\zeta_{\psi,J}\sqrt{(\log J)/n})
\end{aligned}$$

for some finite positive constant C_ω, where the first inequality is by the triangle inequality, the second is by Hölder continuity of the basis for Ψ_J and similar reasoning to that used in equation (26), the first equality is by Lemma F.8 and the fact that $J \asymp K$, and the final equality is because $(\log J)^{-1/2} = o(1)$. For each $x_n \in \mathcal{X}_n$ we may write

$$\begin{aligned}
\psi^J(x_n)'(G_b^{-1/2}S)_l^- G_b^{-1/2} B'u_1/n &= \frac{1}{n}\sum_{i=1}^n g_{n,i}(x_n) u_{1,i}, \text{ where} \\
g_{n,i}(x_n) &= \psi^J(x_n)'(G_b^{-1/2}S)_l^- G_b^{-1/2} b^K(W_i)\,.
\end{aligned}$$

It follows from equation (26) and the Cauchy-Schwarz inequality that the bounds

$$\begin{aligned}
|g_{n,i}(x_n)| &\leq s_{JK}^{-1}\zeta_{\psi,J}\zeta_{b,K} \\
E[g_{n,i}(x_n)^2] &= \psi^J(x_n)'(G_b^{-1/2}S)_l^- ((G_b^{-1/2}S)_l^-)'\psi^J(x_n) \quad \leq \quad s_{JK}^{-2}\zeta_{\psi,J}^2
\end{aligned}$$

hold uniformly for $x_n \in \mathcal{X}_n$. Therefore, by Assumption 2(i) and iterated expectations, the bounds

$$\begin{array}{rcl} |g_{n,i}(x_n)u_{1,i}| & \leq & 2s_{JK}^{-1}\zeta_{\psi,J}\zeta_{b,K}M_n \\ E[g_{n,i}(x_n)^2u_{1,i}^2] & \leq & \overline{\sigma}^2\psi^J(x_n)'(G_b^{-1/2}S)_l^-((G_b^{-1/2}S)_l^-)'\psi^J(x_n) \quad \leq \quad \overline{\sigma}^2 s_{JK}^{-2}\zeta_{\psi,J}^2 \end{array}$$

hold uniformly for $x_n \in \mathcal{X}_n$. It follows by the union bound and Bernstein's inequality that

$$\begin{array}{rl} & \mathbb{P}\left(\max\limits_{x_n\in\mathcal{X}_n} |\psi^J(x_n)'(G_b^{-1/2}S)_l^- G_b^{-1/2}B'u_1/n| > Cs_{JK}^{-1}\zeta_{\psi,J}\sqrt{(\log n)/n}\right) \\ \leq & \#(\mathcal{X}_n)\max\limits_{x_n\in\mathcal{X}_n}\mathbb{P}\left(\left|\dfrac{1}{n}\sum\limits_{i=1}^n g_{n,i}(x_n)u_{1,i}\right| > Cs_{JK}^{-1}\zeta_{\psi,J}\sqrt{(\log n)/n}\right) \\ \lesssim & \exp\left\{\beta\log n - \dfrac{C^2\zeta_{\psi,J}^2(\log n)/(ns_{JK}^2)}{c_1\zeta_{\psi,J}^2/(ns_{JK}^2)[1+(c_2/c_1)(CM_n\zeta_{b,K}\sqrt{(\log n)/n}]}\right\} \end{array} \tag{44}$$

for finite positive constants c_1 and c_2. Then (44) is $o(1)$ for all large C provided $M_n\zeta_{b,K}\sqrt{(\log n)/n} = o(1)$.

By the triangle and Markov inequalities and equation (26), we have

$$\begin{array}{rcl} \mathbb{P}\left(\|\psi^J(x)(G_b^{-1/2}S)_l^- G_b^{-1/2}B'u_2/n\|_\infty > t\right) & \leq & \mathbb{P}\left(s_{JK}^{-1}\zeta_{\psi,J}\|G_b^{-1/2}B'u_2/n\|_{\ell^2} > t\right) \\ & \leq & \mathbb{P}\left(s_{JK}^{-1}\zeta_{\psi,J}\zeta_{b,K}\sum\limits_{i=1}^n |u_{2,i}/n| > t\right) \\ & \leq & \dfrac{2\zeta_{\psi,J}\zeta_{b,K}}{ts_{JK}}E[|u_i|\{|u_i| > M_n\}] \\ & \leq & \dfrac{2\zeta_{\psi,J}\zeta_{b,K}}{ts_{JK}M_n^{1+\delta}}E[|u_i|^{2+\delta}\{|u_i| > M_n\}] \end{array}$$

which, by Assumption 2(ii), is $o(1)$ when $t = Cs_{JK}^{-1}\zeta_{\psi,J}\sqrt{(\log n)/n}$ provided $\zeta_{b,K}\sqrt{n/(\log n)} = O(M_n^{1+\delta})$.

Choosing $M_n^{1+\delta} \asymp \zeta_{b,K}\sqrt{n/\log n}$ satisfies the condition $\zeta_{b,K}\sqrt{n/(\log n)} = O(M_n^{1+\delta})$ trivially, and satisfies the condition $M_n\zeta_{b,K}\sqrt{(\log n)/n} = o(1)$ provided $\zeta_{b,K}^{(2+\delta)/\delta}\sqrt{(\log n)/n} = o(1)$, which holds by Assumption 3(iii). We have therefore proved that $\|T_1\|_\infty = O_p(s_{JK}^{-1}\zeta_{\psi,J}\sqrt{(\log n)/n})$. It follows by the relation $\tau_J \asymp s_{JK}^{-1}$ (Lemma A.1) that $\|T_1\|_\infty = O_p(\tau_J\zeta_{\psi,J}\sqrt{(\log n)/n})$.

Control of $\|T_2\|_\infty$: Using the fact that $\|h\|_\infty \leq \zeta_{\psi,J}\|h\|_{L^2(X)}$ on Ψ_J and Lemmas F.10(b) and F.8 and the relation $\tau_J \asymp s_{JK}^{-1}$, we have:

$$\begin{array}{rcl} \|T_2\|_\infty & \leq & \zeta_{\psi,J}\|G_\psi^{1/2}\{(\widehat{G}_b^{-1/2}\widehat{S})_l^-\widehat{G}_b^{-1/2}G_b^{1/2} - (G_b^{-1/2}S)_l^-\}G_b^{-1/2}B'u/n\|_{\ell^2} \\ & \leq & \zeta_{\psi,J}\|G_\psi^{1/2}\{(\widehat{G}_b^{-1/2}\widehat{S})_l^-\widehat{G}_b^{-1/2}G_b^{1/2} - (G_b^{-1/2}S)_l^-\}\|_{\ell^2}\|G_b^{-1/2}B'u/n\|_{\ell^2} \\ & = & \zeta_{\psi,J}O_p\left(\tau_J^2\zeta\sqrt{(\log J)/n}\right) \times O_p(\sqrt{K/n}) \\ & = & O_p\left(\tau_J\zeta_{\psi,J}\sqrt{(\log J)/n}\right) \times O_p(\tau_J\zeta\sqrt{K/n}) \\ & = & O_p\left(\tau_J\zeta_{\psi,J}\sqrt{(\log J)/n}\right) \times O_p(1) \end{array}$$

where the last equality follows from Assumption 3(ii) and the fact that $\zeta \geq \sqrt{J} \asymp \sqrt{K}$. ☐

Proof of Theorem 3.1. We decompose $\|\widehat{h}-h_0\|_\infty$ into three parts:

$$\|\widehat{h}-h_0\|_\infty \le \|\widehat{h}-\widetilde{h}\|_\infty + \|\widetilde{h}-\Pi_J h_0\|_\infty + \|\Pi_J h_0 - h_0\|_\infty .$$

where $\|\widehat{h}-\widetilde{h}\|_\infty = O_p(\tau_J \xi_{\psi,J}\sqrt{(\log J)/(ne_J)})$ by Lemma 3.1(1) and $\|\widetilde{h}-\Pi_J h_0\|_\infty = O_p(1)\times \|\Pi_J h_0 - h_0\|_\infty$ by Lemma A.3. □

Proof of Corollary 3.1. For **Result (1)**, note that Assumption 3(ii) is satisfied with $\zeta = O(J^{1/2})$ for Ψ_J and B_K being spline, or wavelet or cosine sieves. Next, by the lemmas in Appendix E, $\xi_{\psi,J}/\sqrt{e_J} = O(J^{1/2})$ for Ψ_J being spline or wavelet sieves. Also, $\|\Pi_J\|_\infty \lesssim 1$ for Ψ_J being a spline sieve (Huang (2003)) or a tensor product CDV wavelet sieve (Chen and Christensen (2015)). For $h_0 \in B_\infty(p,L)$ and Ψ_J being spline or wavelet sieves, Lemma A.3 implies that

$$\|\widetilde{h}-h_0\|_\infty = O_p(J^{-p/d}).$$

Note that Bernstein inequalities (or inverse estimates) from approximation theory imply that

$$\|\partial^\alpha h\|_\infty = O(J^{|\alpha|/d})\|h\|_\infty$$

for all $h \in \Psi_J$ (see Schumaker (2007) for splines and Cohen (2003) for wavelets on domains). Therefore,

$$\begin{aligned}
\|\partial^\alpha \widetilde{h} - \partial^\alpha h_0\|_\infty &\le \|\partial^\alpha \widetilde{h} - \partial^\alpha(\Pi_J h_0)\|_\infty + \|\partial^\alpha(\Pi_J h_0) - \partial^\alpha h_0\|_\infty \\
&\le O(J^{|\alpha|/d})\|\widetilde{h}-\Pi_J h_0\|_\infty + \|\partial^\alpha(\Pi_J h_0) - \partial^\alpha h_0\|_\infty \\
&\le O_p(J^{-(p-|\alpha|)/d}) + \|\partial^\alpha(\Pi_J h_0) - \partial^\alpha h_0\|_\infty
\end{aligned}$$

Let h_J be any element of Ψ_J. Since $\Pi_J h_J = h_J$, we have:

$$\begin{aligned}
\|\partial^\alpha(\Pi_J h_0) - \partial^\alpha h_0\|_\infty &= \|\partial^\alpha(\Pi_J(h_0-h_J)) + \partial^\alpha h_J - \partial^\alpha h_0\|_\infty \\
&\le O(J^{|\alpha|/d})\|\Pi_J(h_0-h_J))\|_\infty + \|\partial^\alpha h_J - \partial^\alpha h_0\|_\infty \\
&\le O(J^{|\alpha|/d}) \times \text{const} \times \|h_0 - h_J\|_\infty + \|\partial^\alpha h_J - \partial^\alpha h_0\|_\infty .
\end{aligned}$$

The above inequality holds uniformly in $h_J \in \Psi_J$. Choosing h_J such that $\|h_0 - h_J\|_\infty = O(J^{-p/d})$ and $\|\partial^\alpha h_J - \partial^\alpha h_0\|_\infty = O(J^{-(p-|\alpha|)/d})$ yields the desired result.

For **Result (2)**, Theorem 3.1 implies that

$$\|\widehat{h}-h_0\|_\infty = O_p(J^{-p/d} + \tau_J\sqrt{(J\log J)/n}).$$

By similar arguments to the above, we have:

$$\begin{aligned}
\|\partial^\alpha \widehat{h} - \partial^\alpha h_0\|_\infty &\le \|\partial^\alpha \widehat{h} - \partial^\alpha \widetilde{h}\|_\infty + \|\partial^\alpha \widetilde{h} - \partial^\alpha h_0\|_\infty \\
&\le O(J^{|\alpha|/d})\|\widehat{h}-\widetilde{h}\|_\infty + \|\partial^\alpha \widetilde{h} - \partial^\alpha h_0\|_\infty \\
&\le O_p\left(J^{|\alpha|/d}\left(\tau_J\sqrt{(J\log J)/n}\right)\right) + \|\partial^\alpha \widetilde{h} - \partial^\alpha h_0\|_\infty
\end{aligned}$$

and the result follows by Result (1).

For **Results (2.a) and (2.b)**, Assumption 3(ii)(iii) is satisfied if $\tau_J \times J/\sqrt{n} = O(1)$ and $J^{(2+\delta)/\delta}(\log n)/n = o(1)$. This is satisfied given the stated conditions with the optimal choice of J for mildly ill-posed case and severely ill-posed case respectively. $\square$

G.2 Proofs for Section 3.2

Proof of Theorem 3.2. Consider the Gaussian reduced-form NPIR model with known operator T:

$$\begin{aligned} Y_i &= Th_0(W_i) + u_i \\ u_i|W_i &\sim N(0, \sigma^2(W_i)) \end{aligned} \tag{45}$$

for $1 \leq i \leq n$, where W_i is continuously distributed over $\mathcal{W}$ with density uniformly bounded away from 0 and ∞. As in Chen and Reiss (2011), Theorem 3.2 is proved by (i) noting that the risk (in sup-norm loss) for the NPIV model is at least as large as the risk (in sup-norm loss) for the NPIR model, and (ii) calculating a lower bound (in sup-norm loss) for the NPIR model. Theorem 3.2 therefore follows from a sup-norm analogue of Lemma 1 of Chen and Reiss (2011) and Theorem G.1, which establishes a lower bound on minimax risk over Hölder classes under sup-norm loss for the NPIR model. $\square$

Theorem G.1. *Let Condition LB hold for the NPIR model (45) with a random sample $\{(W_i, Y_i)\}_{i=1}^n$. Then for any $0 \leq |\alpha| < p$:*

$$\liminf_{n\to\infty} \inf_{\widehat{g}_n} \sup_{h \in B_\infty(p,L)} \mathbb{P}_h \left(\|\widehat{g}_n - \partial^\alpha h\|_\infty \geq c(n/\log n)^{-(p-|\alpha|)/(2(p+\varsigma)+d)} \right) \geq c' > 0$$

in the mildly ill-posed case, and

$$\liminf_{n\to\infty} \inf_{\widehat{g}_n} \sup_{h \in B_\infty(p,L)} \mathbb{P}_h \left(\|\widehat{g}_n - \partial^\alpha h\|_\infty \geq c(\log n)^{-(p-|\alpha|)/\varsigma} \right) \geq c' > 0$$

in the severely ill-posed case, where $\inf_{\widehat{g}_n}$ denotes the infimum over all estimators of $\partial^\alpha h$ based on the sample of size n, $\sup_{h \in B_\infty(p,L)} \mathbb{P}_h$ denotes the sup over $h \in B_\infty(p, L)$ and distributions (W_i, u_i) which satisfy Condition LB with ν fixed, and the finite positive constants c, c' depend only on p, L, d, ς and σ_0.

Proof of Theorem G.1. We establish the lower bound by applying Theorem 2.5 of Tsybakov (2009) (see Theorem G.2 below). We first explain the scalar ($d = 1$) case in detail. Let $\{\phi_{j,k}, \psi_{j,k}\}_{j,k}$ be a wavelet basis of regularity $\gamma > p$ for $L^2([0,1])$ as described in Appendix E. Recall that this basis is generated by a Daubechies pair (φ, ψ) where φ has support $[-N+1, N]$. We will define a family of submodels in which we perturb h_0 by elements of the wavelet space W_j, where we choose j deterministically with n. For given j, recall that the wavelet space W_j consists of 2^j functions $\{\psi_{j,k}\}_{0 \leq k \leq 2^j - 1}$, such that $\{\psi_{j,k}\}_{r \leq k \leq 2^j - N - 1}$ are interior wavelets for which $\psi_{j,k}(\cdot) = 2^{j/2}\psi(2^j(\cdot) - k)$.

By construction, the support of each interior wavelet is an interval of length $2^{-j}(2r-1)$. Thus for all j sufficiently large (hence the lim inf in our statement of the Lemma) we may choose a set $M \subset \{r, \ldots, 2^j - N - 1\}$ of interior wavelets with $\#(M) \gtrsim 2^j$ such that $\text{support}(\psi_{j,m}) \cap \text{support}(\psi_{j,m'}) = \emptyset$ for all $m, m' \in M$ with $m \neq m'$. Note also that by construction we have $\#(M) \leq 2^j$ (since there are $2^j - 2N$ interior wavelets).

Recall the norms $\|\cdot\|_{b^p_{\infty,\infty}}$ defined in Appendix E. Let $h_0 \in B_\infty(p, L)$ be such that $\|h_0\|_{B^p_{\infty,\infty}} \leq L/2$, and for each $m \in M$ let

$$h_m = h_0 + c_0 2^{-j(p+1/2)} \psi_{j,m}$$

where c_0 is a positive constant to be defined subsequently. Noting that

$$\begin{aligned} c_0 2^{-j(p+1/2)} \|\psi_{j,m}\|_{B^p_{\infty,\infty}} &\lesssim c_0 2^{-j(p+1/2)} \|\psi_{j,m}\|_{b^p_{\infty,\infty}} \\ &\leq c_0 \end{aligned}$$

it follows by the triangle inequality that $\|h_m\|_{B^p_{\infty,\infty}} \leq L$ uniformly in m for all sufficiently small c_0. By Condition LB, let W_i be distributed such that X_i has uniform marginal distribution on $[0,1]$. For $m \in \{0\} \cup M$ let P_m be the joint distribution of $\{(W_i, Y_i)\}_{i=1}^n$ with $Y_i = Th_m(W_i) + u_i$ for the Gaussian NPIR model (45).

For any $m \in M$

$$\begin{aligned} \|\partial^\alpha h_0 - \partial^\alpha h_m\|_\infty &= c_0 2^{-j(p+1/2)} \|\partial^\alpha \psi_{j,m}\|_\infty \\ &= c_0 2^{-j(p-|\alpha|)} \|\psi^{(|\alpha|)}\|_\infty \end{aligned}$$

where $\psi^{(|\alpha|)}$ denotes the $|\alpha|$th derivative of ψ. Moreover, for any $m, m' \in M$ with $m \neq m'$

$$\begin{aligned} \|\partial^\alpha h_m - \partial^\alpha h_{m'}\|_\infty &= c_0 2^{-j(p+1/2)} \|\partial^\alpha \psi_{j,m} - \partial^\alpha \psi_{j,m'}\|_\infty \\ &= 2c_0 2^{-j(p-|\alpha|)} \|\psi^{(|\alpha|)}\|_\infty \end{aligned}$$

by virtue of the disjoint support of $\{\psi_{j,m}\}_{m \in M}$.

By Condition LB(iii),

$$\|T\psi_{j,m}(W_i)\|_{L^2(W)} \lesssim \nu(2^j)^2 \langle \psi_{j,m}, \psi_{j,m} \rangle_X^2 = \nu(2^j)^2$$

(because $c_0 2^{-j(p+1/2)} \psi_{j,m} \in \mathcal{H}_2(p, L)$ for sufficiently small c_0) where $\nu(2^j) = 2^{-j\varsigma}$ in the mildly ill-posed case and $\nu(2^j) = \exp(-2^{j\varsigma})$ in the severely ill-posed case. The KL distance $K(P_m, P_0)$ is

$$\begin{aligned} K(P_m, P_0) &\leq \frac{1}{2} \sum_{i=1}^n (c_0 2^{-j(p+1/2)})^2 E\left[\frac{(T\psi_{j,m}(W_i))^2}{\sigma^2(W_i)}\right] \\ &\leq \frac{1}{2} \sum_{i=1}^n (c_0 2^{-j(p+1/2)})^2 \frac{E\left[(T\psi_{j,m}(W_i))^2\right]}{\underline{\sigma}^2} \\ &\lesssim n(c_0 2^{-j(p+1/2)})^2 \nu(2^j)^2 \,. \end{aligned}$$

In the mildly ill-posed case ($\nu(2^j) = 2^{-j\varsigma}$) we choose $2^j \asymp (n/(\log n))^{1/(2(p+\varsigma)+1)}$. This yields:

$$\begin{aligned} K(P_m, P_0) &\lesssim c_0^2 \log n \quad \text{ uniformly in } m \\ \log(\#(M)) &\gtrsim \log n + \log\log n \,. \end{aligned}$$

since $\#(M) \asymp 2^j$.

In the severely ill-posed case ($\nu(2^j) = \exp(-\frac{1}{2} 2^{j\varsigma})$) we choose $2^j = (c_1 \log n)^{1/\varsigma}$ with $c_1 > 1$. This yields:

$$\begin{aligned} K(P_m, P_0) &\lesssim n^{-(c_1-1)} \quad \text{ uniformly in } m \\ \log(\#(M)) &\gtrsim \log\log n \,. \end{aligned}$$

In both the mildly and severely ill-posed cases, we may choose c_0 sufficiently small that both $\|h_m\|_{B^p_{\infty,\infty}} \leq L$ and $K(P_m, P_0) \leq \frac{1}{8} \log(\#(M))$ hold uniformly in m for all n sufficiently large. All conditions of Theorem 2.5 of Tsybakov (2009) are satisfied and hence we obtain the lower bound result.

In the multivariate case ($d > 1$) we let $\widetilde{\psi}_{j,k,G}(x)$ denote an orthonormal tensor-product wavelet for $L^2([0,1]^d)$ at resolution level j (see Appendix E). We construct a family of submodels analogously to the univariate case, setting $h_m = h_0 + c_0 2^{-j(p+d/2)}\widetilde{\psi}_{j,m,G}$ where $\widetilde{\psi}_{j,m,G}$ is now the tensor product of d interior univariate wavelets at resolution level j with $G = (w_\psi)^d$ and where $\#(M) \asymp 2^{jd}$. By condition LB we obtain

$$\|\partial^\alpha h_m - \partial^\alpha h_{m'}\|_\infty \gtrsim c_0 2^{-j(p-|\alpha|)}$$

for each $m, m' \in \{0\} \cup M$ with $m \neq m'$, and

$$K(P_m, P_0) \quad \lesssim \quad n(c_0 2^{-j(p+d/2)})^2 \nu(2^j)^2$$

for each $m \in M$, where $\nu(2^j) = 2^{-j\varsigma}$ in the mildly ill-posed case and $\nu(2^j) \asymp \exp(-2^{j\varsigma})$ in the severely ill-posed case. We choose $2^j \asymp (n/\log n)^{1/(2(p+\varsigma)+d)}$ in the mildly ill-posed case and $2^j = (c_1 \log n)^{1/\varsigma}$ in the severely ill-posed case. The result follows as in the univariate case. $\square$

The following theorem is a special case of Theorem 2.5 on p. 99 of Tsybakov (2009) which we use to prove the minimax lower bounds in sup- and L^2-norm loss for h_0 and its derivatives. We state the result here for convenience.

Theorem G.2 (Tsybakov (2009))**.** *Assume that $\#(M) \geq 2$ and suppose that $(\mathcal{H}, \|\cdot\|_{\mathcal{H}})$ contains elements $\{h_m : m \in \{0\} \cup M\}$ such that:*
(i) $\|\partial^\alpha h_m - \partial^\alpha h_{m'}\|_{\mathcal{H}} \geq 2s > 0$ for each $m, m' \in M \cup \{0\}$ with $m \neq m'$;
(ii) $P_m \ll P_0$ for each $m \in M$ and

$$\frac{1}{\#(M)} \sum_{m \in M} K(P_m, P_0) \leq \mathsf{a} \log(\#(M))$$

with $0 < \mathsf{a} < \frac{1}{8}$ and where P_m denotes the distribution of the data when $h = h_m$ for each $m \in \{0\} \cup M$. Then:

$$\inf_{\widehat{g}} \sup_{h \in \mathcal{H}} \mathbb{P}_h(\|\widehat{g} - \partial^\alpha h\|_{\mathcal{H}} \geq s) \geq \frac{\sqrt{\#(M)}}{1 + \sqrt{\#(M)}} \left(1 - 2\mathsf{a} - \sqrt{\frac{2\mathsf{a}}{\log(\#(M))}} \right) > 0\,.$$

G.3 Proofs for Section 3.3

Proof of Lemma 3.2. We first prove **Result (1)**. Let $P_{J-1,z} = clsp\{\phi_{01,z}, \ldots, \phi_{0J-1,z}\}$ and let $P_{J-1,z}^\perp$ denote its orthogonal complement in $L^2(X_1|Z = z)$. Observe that by definition of the singular values, for each z we have:

$$\begin{aligned} \sup_{h_z \in P_{J-1,z}^\perp : \|h_z\|_{L^2(X_1|Z=z)}=1} \|T_z h_z\|^2_{L^2(W_1|Z=z)} &= \sup_{h_z \in P_{J-1,z}^\perp : \|h_z\|_{L^2(X_1|Z=z)}=1} \langle (T_z^* T_z) h_z, h_z \rangle_{X_1|Z=z} \\ &= \mu_{J,z}^2\,. \end{aligned} \tag{46}$$

Then let $P_{J-1}^{\perp} = \{h(x_1, z) \in L^2(X) : h(\cdot, z) \in P_{J-1,z}^{\perp} \text{ for each } z\}$. Note that $\phi_{0j} \in \{h \in P_{J-1}^{\perp} : \|h(\cdot, z)\|_{L^2(X_1|Z=z)} = 1\,\forall z\}$ for each $j \geq J$. Then:

$$\begin{aligned}
\tau_J^{-2} &= \inf_{h \in \Psi_J : \|h\|_{L^2(X)=1}} \|Th\|^2_{L^2(W)} \\
&\leq \inf_{h \in \Psi_J \cap P_{J-1}^{\perp} : \|h\|_{L^2(X_1|Z=z)=1}\ \forall z} \|Th\|^2_{L^2(W)} \\
&\leq \sup_{h \in \Psi_J \cap P_{J-1}^{\perp} : \|h\|_{L^2(X_1|Z=z)=1}\ \forall z} \|Th\|^2_{L^2(W)} \\
&\leq \sup_{h \in P_{J-1}^{\perp} : \|h\|_{L^2(X_1|Z=z)=1}\ \forall z} \|Th\|^2_{L^2(W)} \,.
\end{aligned} \tag{47}$$

Let F_Z denote the distribution of Z. For any $h \in P_{J-1}^{\perp}$ let $h_z(x_1) = h(x_1, z)$ and observe that $h_z \in P_{J-1,z}^{\perp}$. By iterated expectations and (46), for any $h \in P_{J-1}^{\perp}$ with $\|h_z\|_{L^2(X_1|Z=z)} = 1$ for each z, we have:

$$\begin{aligned}
\|Th\|^2_{L^2(W)} &= \int \|E[h(X_{1i}, z)|W_{1i}, Z_i = z]\|^2_{L^2(W_1|Z=z)} \,\mathrm{d}F_Z(z) \\
&= \int \|T_z h_z\|^2_{L^2(W_1|Z=z)} \,\mathrm{d}F_Z(z) \\
&\leq \int \mu_{J,z}^2 \|h_z\|^2_{L^2(X_1|Z=z)} \,\mathrm{d}F_Z(z) \\
&= \int \mu_{J,z}^2 \,\mathrm{d}F_Z(z) = E[\mu_{J,Z_i}^2] \,.
\end{aligned} \tag{48}$$

It follows by substituting (48) into (47) that $\tau_J \geq E[\mu_{J,Z_i}^2]^{-1/2}$.

To prove **Result (2)**, note that any $h \in \Psi_J$ with $h \neq 0$ can be written as $\sum_{j=1}^J a_j \phi_{0j}$ for constants $a_j = a_j(h)$ where

$$\|h\|^2_{L^2(X)} = E\left[E\left[\left(\sum_{j=1}^J a_j \phi_{0j}(X_{1i}, Z_i)\right)^2 \middle| Z_i\right]\right] = \sum_{j=1}^J a_j^2$$

since $E[\phi_{0j,z}(X_i)\phi_{0k,z}(X_i)|Z_i = z] = \delta_{jk}$ where δ_{jk} denotes the Kronecker delta. Moreover:

$$\begin{aligned}
\|Th\|^2_{L^2(W)} &= E\left[\left(E\left[\sum_{j=1}^J a_j \phi_{0j}(X_{1i}, Z_i) \middle| W_{1i}, Z_i\right]\right)^2\right] \\
&= E\left[\left(E\left[\sum_{j=1}^J a_j \phi_{0j,Z_i}(X_{1i}) \middle| W_{1i}, Z_i\right]\right)^2\right] \\
&= E\left[\left(\sum_{j=1}^J a_j \mu_{j,Z_i} \phi_{1j,Z_i}(W_{1i})\right)^2\right] \\
&= E\left[E\left[\left(\sum_{j=1}^J a_j \mu_{j,Z_i} \phi_{1j,Z_i}(W_{1i})\right)^2 \middle| Z_i\right]\right] = \sum_{j=1}^J a_j^2 E\left[\mu_{j,Z_i}^2\right] \geq \|h\|^2_{L^2(X)} E[\mu_{J,Z_i}^2]
\end{aligned}$$

since $E[\phi_{1j,z}(W_{1i})\phi_{1k,z}(W_{1i})|Z_i = z] = \delta_{jk}$. Therefore,

$$\tau_J = \sup_{h \in \Psi_J} \frac{\|h\|_{L^2(X)}}{\|Th\|_{L^2(W)}} \le \frac{1}{E[\mu_{J,Z_i}^2]^{1/2}}$$

as required. $\square$

G.4 Proofs for Appendix D and Section 4

Since the proofs for uniform inference theories (in Section 4) built upon that for the pointwise normality Theorem D.1 (in Appendix D), we shall present the proof of Theorem D.1 first.

G.4.1 Proofs for Appendix D

Proof of Theorem D.1. We first prove **Result (1)**. By Assumption 5'(a) or 5'(b)(i)(ii) we have:

$$\sqrt{n}\frac{(f(\widehat{h}) - f(h_0))}{\sigma_n(f)} = \sqrt{n}\frac{Df(h_0)[\widehat{h} - \widetilde{h}]}{\sigma_n(f)} + o_p(1)\,.$$

Define

$$Z_n(W_i) = \frac{(Df(h_0)[\psi^J])'[S'G_b^{-1}S]^{-1}S'G_b^{-1}b^K(W_i)}{\sigma_n(f)} = \Pi_K T u_n(f)(W_i)$$

where $u_n(f) = v_n(f)/\sigma_n(f)$ is the scaled sieve 2SLS Riesz representer. Note that $E[(Z_n(W_i)u_i)^2] = 1$. Then

$$\begin{aligned}
\sqrt{n}\frac{Df(h_0)[\widehat{h} - \widetilde{h}]}{\sigma_n(f)} &= \frac{1}{\sqrt{n}}\sum_{i=1}^n Z_n(W_i)u_i \\
&\quad + \frac{(Df(h_0)[\psi^J])'((\widehat{S}'\widehat{G}_b^-\widehat{S})^-\widehat{S}'\widehat{G}_b^- - (S'G_b^{-1}S)^{-1}S'G_b^{-1})(B'u/\sqrt{n})}{\sigma_n(f)} \\
&=: T_1 + T_2\,.
\end{aligned}$$

We first show $T_1 \to_d N(0,1)$ by the Lindeberg-Feller theorem. To verify the Lindeberg condition, note that

$$|Z_n(W_i)| \le \frac{\left\|(Df(h_0)[\psi^J])'(S'G_b^{-1}S)^{-1}S'G_b^{-1/2}\right\| \left\|G_b^{-1/2}b^K(W_i)\right\|}{(\inf_w E[u_i^2|W_i = w])^{1/2}\left\|(Df(h_0)[\psi^J])'(S'G_b^{-1}S)^{-1}S'G_b^{-1/2}\right\|} \le \underline{\sigma}^{-1}\zeta_b(K)$$

by the Cauchy-Schwarz inequality and Assumption 2(iii). Therefore,

$$E[u_i^2 Z_n(W_i)^2\{|Z_n(W_i)u_i| > \eta\sqrt{n}\}] \le \sup_w E[u_i^2\{|u_i| \gtrsim \eta(\sqrt{n}/\zeta_b(K))\}|W_i = w] = o(1)$$

by Assumption 2(iv') and the condition on J. Therefore, $T_1 \to_d N(0,1)$.

For T_2, observe that

$$
\begin{aligned}
|T_2| &= \left|\frac{(Df(h_0)[\psi^J])'((\widehat{G}_b^{-1/2}\widehat{S})_l^-\widehat{G}_b^{-1/2}G_b^{1/2} - (G_b^{-1/2}S)_l^-)(G_b^{-1/2}B'u/\sqrt{n})}{\sigma_n(f)}\right| \\
&= \left|\frac{[(Df(h_0)[\psi^J])'(G_b^{-1/2}S)_l^-]G_b^{-1/2}S\{(\widehat{G}_b^{-1/2}\widehat{S})_l^-\widehat{G}_b^{-1/2}G_b^{1/2} - (G_b^{-1/2}S)_l^-\}(G_b^{-1/2}B'u/\sqrt{n})}{\sigma_n(f)}\right| \\
&\leq \frac{\left\|(Df(h_0)[\psi^J])'(G_b^{-1/2}S)_l^-\right\| \left\|G_b^{-1/2}S\{(\widehat{G}_b^{-1/2}\widehat{S})_l^-\widehat{G}_b^{-1/2}G_b^{1/2} - (G_b^{-1/2}S)_l^-\}\right\| \left\|G_b^{-1/2}B'u/\sqrt{n}\right\|}{(\inf_w E[u_i^2|W_i = w])^{1/2}\left\|(Df(h_0)[\psi^J])'(G_b^{-1/2}S)_l^-\right\|} \\
&\leq \underline{\sigma}^{-1}\left\|G_b^{-1/2}S\{(\widehat{G}_b^{-1/2}\widehat{S})_l^-\widehat{G}_b^{-1/2}G_b^{1/2} - (G_b^{-1/2}S)_l^-\}\right\| \left\|G_b^{-1/2}B'u/\sqrt{n}\right\| \\
&= O_p(s_{JK}^{-1}\zeta\sqrt{(J\log J)/n})
\end{aligned}
$$

where the first inequality is by the Cauchy-Schwarz inequality, the second is by Assumption 2(iii), and the final line is by Lemmas F.10(c) and F.8. The result follows by the equivalence $\tau_J \asymp s_{JK}^{-1}$ (see Lemma A.1) and the condition $\tau_J\zeta\sqrt{(J\log n)/n} = o(1)$.

Result (2) follows directly from Result (1) and Lemma G.2. $\square$

Lemma G.2. *Let Assumptions 1(iii), 2(i)–(iii), 3(iii) and 4(i) hold, $\tau_J\zeta\sqrt{(\log n)/n} = o(1)$, and Assumption 5'(b)(iii) hold (with $\eta_n' = 0$ if $f(\cdot)$ is linear). Let $\|\widehat{h} - h_0\|_\infty = O_p(\delta_{h,n}) = o_p(1)$, and $\delta_{V,n} \equiv \big[\zeta_{b,K}^{(2+\delta)/\delta}\sqrt{(\log K)/n}\big]^{\delta/(1+\delta)} + \tau_J\zeta\sqrt{(\log J)/n} + \delta_{h,n}$. Then:*

$$
\left|\frac{\widehat{\sigma}_n(f)}{\sigma_n(f)} - 1\right| = O_p(\delta_{V,n} + \eta_n') = o_p(1)\,.
$$

Proof of Lemma G.2. First write

$$
\begin{aligned}
\frac{\widehat{\sigma}_n(f)^2}{\sigma_n(f)^2} - 1 &= \left(\frac{\widehat{\gamma}_n'\Omega^o\widehat{\gamma}_n}{\sigma_n(f)^2} - 1\right) + \frac{\widehat{\gamma}_n'(\widehat{\Omega}^o - \Omega^o)\widehat{\gamma}_n}{\sigma_n(f)^2} \\
&= \left(\frac{(\widehat{\gamma}_n - \gamma_n)'\Omega^o(\widehat{\gamma}_n + \gamma_n)}{\sigma_n(f)^2}\right) + \frac{\widehat{\gamma}_n'(\widehat{\Omega}^o - \Omega^o)\widehat{\gamma}_n}{\sigma_n(f)^2} \quad =: \quad T_1 + T_2
\end{aligned}
$$

where

$$
\begin{aligned}
\widehat{\Omega}^o &= G_b^{-1/2}\widehat{\Omega}G_b^{-1/2} & \qquad \widehat{\gamma}_n &= G_b^{1/2}\widehat{G}_b^{-1}\widehat{S}[\widehat{S}'\widehat{G}_b^{-1}\widehat{S}]^{-1}Df(\widehat{h})[\psi^J] \\
\Omega^o &= G_b^{-1/2}\Omega G_b^{-1/2} & \qquad \gamma_n &= G_b^{-1/2}S[S'G_b^{-1}S]^{-1}Df(h_0)[\psi^J]
\end{aligned}
$$

and observe that $\gamma_n'\Omega^o\gamma_n = \sigma_n(f)^2$ and $\widehat{\gamma}_n'\widehat{\Omega}^o\widehat{\gamma}_n = \widehat{\sigma}_n(f)^2$.

Control of T_1: We first show that

$$
\frac{\|\widehat{\gamma}_n - \gamma_n\|_{\ell^2}}{\sigma_n(f)} = O_p(\tau_J\zeta\sqrt{(\log J)/n} + \eta_n') = o_p(1)\,. \tag{49}
$$

To simplify notation, let

$$
\partial = \frac{Df(h_0)[\psi^J]}{s_n(f)} \quad \text{and} \quad \widehat{\partial} = \frac{Df(\widehat{h})[\psi^J]}{s_n(f)}
$$

and note that $\|\partial'(G_b^{-1/2}S)_l^-\|_{\ell^2} = s_n(f)/\sigma_n(f) \asymp 1$ under Assumptions 2(i)(iii) and that $\widehat{\partial} = \partial$ if $f(\cdot)$ is linear.

Then we have:

$$
\begin{aligned}
\frac{\|\widehat{\gamma}_n-\gamma_n\|_{\ell^2}}{\sigma_n(f)} &= \|\widehat{\partial}'(\widehat{G}_b^{-1/2}\widehat{S})_l^-\widehat{G}_b^{-1/2}G_b^{1/2}-\partial'(G_b^{-1/2}S)_l^-\|_{\ell^2} \\
&\leq \|\widehat{\partial}'(G_b^{-1/2}S)_l^-\|_{\ell^2}\|G_b^{-1/2}S\{(\widehat{G}_b^{-1/2}\widehat{S})_l^-\widehat{G}_b^{-1/2}G_b^{1/2}-(G_b^{-1/2}S)_l^-\}\|_{\ell^2}+\underline{\sigma}^{-1}\|(\widehat{\partial}'-\partial')(G_b^{-1/2}S)_l^-)\|_{\ell^2} \\
&= O_p(1)\times O_p(s_{JK}^{-1}\zeta\sqrt{(\log J)/n})+\underline{\sigma}^{-1}\frac{\|\Pi_K T(\widehat{v}_n(f)-v_n(f))\|_{L^2(W)}}{s_n(f)} \\
&= O_p(1)\times O_p(s_{JK}^{-1}\zeta\sqrt{(\log J)/n})+O_p(\eta_n')
\end{aligned}
$$

where the third line is Lemma F.10(c) and the final line is by Assumption 5'(b)(iii). Therefore, (49) holds by the equivalence $s_{JK}^{-1}\asymp\tau_J$ (Lemma A.1) and the condition $\tau_J\zeta\sqrt{(\log n)/n}$.

Finally, since all eigenvalues of Ω^o are bounded between $\underline{\sigma}^2$ and $\overline{\sigma}^2$ under Assumption 2(i)(iii), it follows from (49) and Cauchy-Schwarz that $|T_1|=o_p(1)$.

Control of T_2: Equation (49) implies that $\|\widehat{\gamma}_n\|/\sigma_n(f)=O_p(1)$. Therefore, $|T_2|\leq O_p(1)\times\|\widehat{\Omega}^o-\Omega^o\|_{\ell^2}=o_p(1)$ by Lemma G.3. $\square$

Lemma G.3. *Let Assumptions 2(i)(ii) hold, let $\zeta_{b,K}\sqrt{(\log K)/n}=o(1)$, and let $\|\widehat{h}-h_0\|_\infty=O_p(\delta_{h,n})$ with $\delta_{h,n}=o(1)$. Then:*

$$
\|\widehat{\Omega}^o-\Omega^o\|_{\ell^2}=O_p\Big(\big(\zeta_{b,K}^{(2+\delta)/\delta}\sqrt{(\log K)/n}\big)^{\delta/(1+\delta)}+\delta_{h,n}\Big)
$$

Proof of Lemma G.3. By the triangle inequality:

$$
\begin{aligned}
\|\widehat{\Omega}^o-\Omega^o\|_{\ell^2} &\leq \left\|G_b^{-1/2}\left(\frac{1}{n}\sum_{i=1}^n u_i^2 b^K(W_i)b^K(W_i)'\right)G_b^{-1/2}\right\|_{\ell^2} \\
&\quad+\left\|G_b^{-1/2}\left(\frac{1}{n}\sum_{i=1}^n 2u_i(\widehat{u}_i-u_i)b^K(W_i)b^K(W_i)'\right)G_b^{-1/2}\right\|_{\ell^2} \\
&\quad+\left\|G_b^{-1/2}\left(\frac{1}{n}\sum_{i=1}^n(\widehat{u}_i-u_i)^2 b^K(W_i)b^K(W_i)'\right)G_b^{-1/2}\right\|_{\ell^2} \\
&\leq O_p((\zeta_{b,K}^{(2+\delta)/\delta}\sqrt{(\log K)/n})^{\delta/(1+\delta)})+\|\widehat{h}-h_0\|_\infty\times O_p(1)+\|\widehat{h}-h_0\|_\infty^2\times O_p(1)
\end{aligned}
$$

where the first term may easily be deduced from the proof of Lemma 3.1 of Chen and Christensen (2015), the second then follows because $2u_i(\widehat{u}_i-u_i)\leq 2(1+u_i^2)\|\widehat{h}-h_0\|_\infty$, and the third follows similarly because $\|\widehat{G}_b^o\|_{\ell^2}=O_p(1)$ by Lemma F.7. $\square$

G.4.2 Proofs for Section 4

Proof of Lemma 4.1. Recall that

$$
\begin{aligned}
\widehat{\mathbb{Z}}_n(t) &= \frac{(Df_t(h_0)[\psi^J])'[S'G_b^{-1}S]^{-1}S'G_b^{-1/2}}{\sigma_n(f_t)}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n G_b^{-1/2}b^K(W_i)u_i\right), \\
\mathbb{Z}_n(t) &= \frac{(Df_t(h_0)[\psi^J])'[S'G_b^{-1}S]^{-1}S'G_b^{-1/2}}{\sigma_n(f_t)}\mathcal{Z}_n \quad \text{where } \mathcal{Z}_n\sim N(0,\Omega^o) \text{ with } \Omega^o=G_b^{-1/2}\Omega G_b^{-1/2}.
\end{aligned}
$$

Step 1: Uniform Bahadur representation. By Assumption 5(a) or (b)(i)(ii), we have

$$\begin{aligned}
\sup_{t\in\mathcal{T}}\left|\sqrt{n}\frac{f_t(\widehat{h})-f_t(h_0)}{\widehat{\sigma}_n(f_t)}-\widehat{\mathbb{Z}}_n(t)\right| &\leq \sup_{t\in\mathcal{T}}\left|\sqrt{n}\frac{Df_t(h_0)[\widehat{h}-\widetilde{h}]}{\sigma_n(f_t)}-\widehat{\mathbb{Z}}_n(t)\right|+O_p(\eta_n)\times\sup_{t\in\mathcal{T}}\left|\frac{\sigma_n(f_t)}{\widehat{\sigma}_n(f_t)}\right| \\
&\quad +\sup_{t\in\mathcal{T}}\left|\frac{\sigma_n(f_t)}{\widehat{\sigma}_n(f_t)}-1\right|\times\sup_{t\in\mathcal{T}}\left|\sqrt{n}\frac{Df_t(h_0)[\widehat{h}-\widetilde{h}]}{\sigma_n(f_t)}\right| \\
&=: T_1+T_2+T_3\,.
\end{aligned}$$

Control of T_1: As in the proof of Theorem D.1,

$$\begin{aligned}
T_1 &= \sup_{t\in\mathcal{T}}\left|\frac{(Df_t(h_0)[\psi^J])'((\widehat{G}_b^{-1/2}\widehat{S})_l^-\widehat{G}_b^{-1/2}G_b^{1/2}-(G_b^{-1/2}S)_l^-)(G_b^{-1/2}B'u/\sqrt{n})}{\sigma_n(f_t)}\right| \\
&= \sup_{t\in\mathcal{T}}\left|\frac{[(Df_t(h_0)[\psi^J])'(G_b^{-1/2}S)_l^-]G_b^{-1/2}S\{(\widehat{G}_b^{-1/2}\widehat{S})_l^-\widehat{G}_b^{-1/2}G_b^{1/2}-(G_b^{-1/2}S)_l^-\}(G_b^{-1/2}B'u/\sqrt{n})}{\sigma_n(f_t)}\right| \\
&\leq \sup_{t\in\mathcal{T}}\frac{\left\|(Df_t(h_0)[\psi^J])'(G_b^{-1/2}S)_l^-\right\|\left\|G_b^{-1/2}S\{(\widehat{G}_b^{-1/2}\widehat{S})_l^-\widehat{G}_b^{-1/2}G_b^{1/2}-(G_b^{-1/2}S)_l^-\}\right\|\left\|G_b^{-1/2}B'u/\sqrt{n}\right\|}{(\inf_w E[u_i^2|W_i=w])^{1/2}\left\|(Df_t(h_0)[\psi^J])'(G_b^{-1/2}S)_l^-\right\|} \\
&\leq \underline{\sigma}^{-1}\left\|G_b^{-1/2}S\{(\widehat{G}_b^{-1/2}\widehat{S})_l^-\widehat{G}_b^{-1/2}G_b^{1/2}-(G_b^{-1/2}S)_l^-\}\right\|\left\|G_b^{-1/2}B'u/\sqrt{n}\right\| \\
&= O_p(\tau_J\zeta\sqrt{(J\log J)/n})=o_p(r_n)
\end{aligned}$$

where the first inequality is by the Cauchy-Schwarz inequality, the second is by Assumption 2(iii), and the final line is by Lemmas F.10(c) and F.8 and the equivalence $\tau_J \asymp s_{JK}^{-1}$ (see Lemma A.1), and the last $o_p(r_n)$ is by Assumption 6(ii.2).

Control of T_2: Lemma G.4 below shows that

$$\sup_{t\in\mathcal{T}}\left|\frac{\widehat{\sigma}_n(f_t)}{\sigma_n(f_t)}-1\right|=O_p(\delta_{V,n}+\eta_n')=o_p(1)$$

from which it follows that $T_2=O_p(\eta_n)\times O_p(1)=O_p(\eta_n)$.

Control of T_3: By Lemma G.4 below and the bound for T_1, we have:

$$\begin{aligned}
T_3 &= O_p(\delta_{V,n}+\eta_n')\times\sup_{t\in\mathcal{T}}\left|\sqrt{n}\frac{Df_t(h_0)[\widehat{h}-\widetilde{h}]}{\sigma_n(f_t)}\right| \\
&= O_p(\delta_{V,n}+\eta_n')\times\big[\sup_{t\in\mathcal{T}}|\widehat{\mathbb{Z}}_n(t)|+o_p(r_n)\big] \\
&= O_p(\delta_{V,n}+\eta_n')\times\big[o_p(r_n)+\sup_{t\in\mathcal{T}}|\mathbb{Z}_n(t)|+o_p(r_n)\big] \\
&= O_p(\delta_{V,n}+\eta_n')\times\big[o_p(r_n)+O_p(c_n)\big]
\end{aligned}$$

where the second-last line is by display (54) step 2 below and the final line is by Lemma G.5 below. Therefore we have proved:

$$\begin{aligned}
\sup_{t\in\mathcal{T}}\left|\sqrt{n}\frac{f_t(\widehat{h})-f_t(h_0)}{\widehat{\sigma}_n(f_t)}-\widehat{\mathbb{Z}}_n(t)\right| &= O_p(\tau_J\zeta\sqrt{(J\log J)/n})+O_p(\eta_n)+O_p(\delta_{V,n}+\eta_n')\times\big[o_p(r_n)+O_p(c_n)\big] \\
&= o_p(r_n) \qquad (50)
\end{aligned}$$

where the final line is by Assumption 6(ii.2).

Step 2: Approximating $\widehat{\mathbb{Z}}_n(t)$ by a Gaussian process $\mathbb{Z}_n(t)$. We use Yurinskii's coupling (Pollard, 2002, Theorem 10, p. 244) to show that there exists a sequence of $N(0,\Omega^o)$ random vectors $\mathcal{Z}_n$ such that

$$\left\| \frac{1}{\sqrt{n}} \sum_{i=1}^n G_b^{-1/2} b^K(W_i) u_i - \mathcal{Z}_n \right\|_{\ell^2} = o_p(r_n)\,. \tag{51}$$

By Assumption 2(iv) we have

$$\sum_{i=1}^n E[\|n^{-1/2} G_b^{-1/2} b^K(W_i) u_i\|_{\ell^2}^3] \lesssim n^{-1/2} \zeta_{b,K} E[\|G_b^{-1/2} b^K(W_i)\|_{\ell^2}^2] = \frac{\zeta_{b,K} K}{\sqrt{n}} = O\left(\frac{\zeta_{b,K} J}{\sqrt{n}}\right).$$

Existence of $\mathcal{Z}_n$ follows under the condition (Assumption 6(ii.1))

$$\frac{\zeta_{b,K} J^2}{r_n^3 \sqrt{n}} = o(1)\,.$$

The process $\mathbb{Z}_n(t)$ is a centered Gaussian process with the covariance function

$$E[\mathbb{Z}_n(t_1)\mathbb{Z}_n(t_2)] = \frac{(Df_{t_1}(h_0)[\psi^J])'[S'G_b^{-1}S]^{-1}S'G_b^{-1}\Omega G_b^{-1}S[S'G_b^{-1}S]^{-1}Df_{t_2}(h_0)[\psi^J]}{\sigma_n(f_{t_1})\sigma_n(f_{t_2})}\,. \tag{52}$$

Now observe that

$$\sup_{t\in\mathcal{T}} \left\| \frac{(Df_t(h_0)[\psi^J])'[S'G_b^{-1}S]^{-1}S'G_b^{-1/2}}{\sigma_n(f_t)} \right\|_{\ell^2} = \sup_{t\in\mathcal{T}} \frac{s_n(f_t)}{\sigma_n(f_t)} \asymp 1 \tag{53}$$

by Assumption 2(i)(iii). Therefore,

$$\sup_{t\in\mathcal{T}} \left| \widehat{\mathbb{Z}}_n(t) - \mathbb{Z}_n(t) \right| = o_p(r_n) \tag{54}$$

by equations (51) and (53) and Cauchy-Schwarz. □

Lemma G.4. *Let Assumptions 1(iii), 2(i)–(iii), 3(ii)(iii) and 4(i) hold, $\tau_J \zeta \sqrt{(\log n)/n} = o(1)$, and Assumption 5(b)(iii) hold (with $\eta_n' = 0$ if $f_t(\cdot)$ is linear). Let $\|\widehat{h} - h_0\|_\infty = O_p(\delta_{h,n})$ with $\delta_{h,n} = o(1)$. Then:*

$$\sup_{t\in\mathcal{T}} \left| \frac{\sigma_n(f_t)}{\widehat{\sigma}_n(f_t)} - 1 \right| = O_p(\delta_{V,n} + \eta_n') = o_p(1).$$

Proof of Lemma G.4. The proof follows by identical arguments to the proof of Lemma G.2. □

Proof of Theorem 4.1. Recall that

$$\mathbb{Z}_n^*(t) = \frac{(Df_t(\widehat{h})[\psi^J])'[\widehat{S}'\widehat{G}_b^{-1}\widehat{S}]^{-1}\widehat{S}'\widehat{G}_b^{-1}}{\widehat{\sigma}(f_t)} \left(\frac{1}{\sqrt{n}} \sum_{i=1}^n b^K(W_i)\widehat{u}_i \varpi_i \right) \quad \text{for each } t \in \mathcal{T}\,.$$

Step 1: Approximating $\mathbb{Z}_n^*(t)$ by a Gaussian process $\widetilde{\mathbb{Z}}_n^*$. Each of the terms $n^{-1/2} G_b^{-1/2} b^K(W_i)\widehat{u}_i\varpi_i$ is centered under $\mathbb{P}^*$ because $E[\varpi_i | Z^n] = 0$ for $i = 1, \ldots, n$. Moreover,

$$\sum_{i=1}^n E[(n^{-1/2} G_b^{-1/2} b^K(W_i)\widehat{u}_i\varpi_i)(n^{-1/2} G_b^{-1/2} b^K(W_i)\widehat{u}_i\varpi_i)' | Z^n] = \widehat{\Omega}^o$$

where $G_b^{-1/2}\widehat{\Omega}G_b^{-1/2} = \widehat{\Omega}^o$, and

$$\sum_{i=1}^n E[\|n^{-1/2}G_b^{-1/2}b^K(W_i)\widehat{u}_i\varpi_i\|_{\ell^2}^3|Z^n] \lesssim n^{-3/2}\sum_{i=1}^n E[\|G_b^{-1/2}b^K(W_i)\|_{\ell^2}^2|\widehat{u}_i|^3]$$

because $E[|\varpi_i|^3|Z^n] < \infty$ uniformly in i, and where

$$n^{-3/2}\sum_{i=1}^n E[\|G_b^{-1/2}b^K(W_i)\|_{\ell^2}^2|\widehat{u}_i|^3] \lesssim \frac{\zeta_{b,K}K}{\sqrt{n}}$$

holds wpa1 (by Markov's inequality using $|\widehat{u}_i|^3 \lesssim |u_i|^3 + \|\widehat{h}-h_0\|_\infty^3$ and Assumption 2(iv)). A second application of Yurinskii's coupling conditional on the data Z^n then yields existence of a sequence of $N(0,\widehat{\Omega}^o)$ random vectors $\mathcal{Z}_n^*$ such that

$$\left\|\frac{1}{\sqrt{n}}\sum_{i=1}^n G_b^{-1/2}b^K(W_i)\widehat{u}_i\varpi_i - \mathcal{Z}_n^*\right\|_{\ell^2} = o_{p^*}(r_n)$$

wpa1. Therefore:

$$\sup_{t\in\mathcal{T}}\left|\mathbb{Z}_n^*(t) - \frac{(Df_t(\widehat{h})[\psi^J])'[\widehat{S}'\widehat{G}_b^{-1}\widehat{S}]^{-1}\widehat{S}'\widehat{G}_b^{-1/2}}{\widehat{\sigma}_n(f_t)}\mathcal{Z}_n^*\right| = o_{p^*}(r_n) \tag{55}$$

wpa1. Now observe that we can define a centered Gaussian process $\widetilde{\mathbb{Z}}_n^*$ under $\mathbb{P}^*$ by

$$\widetilde{\mathbb{Z}}_n^*(t) \equiv \frac{(Df_t(h_0)[\psi^J])'[S'G_b^{-1}S]^{-1}S'G_b^{-1/2}}{\sigma_n(f_t)}(\Omega^o)^{1/2}(\widehat{\Omega}^o)^{-1/2}\mathcal{Z}_n^*$$

which has the same covariance function as $\mathbb{Z}_n$ (see equation (52)) whenever $\widehat{\Omega}^o$ is invertible (which it is wpa1). Therefore, by Lemma G.6 below we have:

$$\sup_{t\in\mathcal{T}}\left|\frac{(Df_t(\widehat{h})[\psi^J])'[\widehat{S}'\widehat{G}_b^{-1}\widehat{S}]^{-1}\widehat{S}'\widehat{G}_b^{-1/2}}{\widehat{\sigma}_n(f_t)}\mathcal{Z}_n^* - \widetilde{\mathbb{Z}}_n^*(t)\right| = o_p^*(r_n) \tag{56}$$

wpa1. It follows from equations (55) and (56) and Assumption 6(ii) that

$$\sup_{t\in\mathcal{T}}\left|\mathbb{Z}_n^*(t) - \widetilde{\mathbb{Z}}_n^*(t)\right| \quad = \quad o_{p^*}(r_n) + o_p^*(r_n) \quad = \quad o_p^*(r_n) \tag{57}$$

wpa1.

Step 2: Consistency. By Lemma 4.1 and display (54), we have:

$$\sup_{t\in\mathcal{T}}\left|\frac{\sqrt{n}(f_t(\widehat{h}) - f_t(h_0))}{\widehat{\sigma}_n(f_t)} - \mathbb{Z}_n(t)\right| = o_p(r_n) + o_p(r_n) = o_p(r_n)\,.$$

Therefore, we may choose a sequence of positive constants ϵ_n with $\epsilon_n = o(1)$ such that

$$\sup_{t\in\mathcal{T}}\left|\frac{\sqrt{n}(f_t(\widehat{h}) - f_t(h_0))}{\widehat{\sigma}_n(f_t)} - \mathbb{Z}_n(t)\right| \leq \epsilon_n r_n \tag{58}$$

holds wpa1. By an anti-concentration inequality (Chernozhukov, Chetverikov, and Kato, 2014, Theorem 2.1) and Lemma G.5 below, we have:

$$\sup_{s\in\mathbb{R}} \mathbb{P}\left(\sup_{t\in\mathcal{T}} |\mathbb{Z}_n(t) - s| \leq \epsilon_n r_n\right) \lesssim \epsilon_n r_n E[\sup_{t\in\mathcal{T}} |\mathbb{Z}_n(t)|] \lesssim \epsilon_n r_n c_n = o(1)$$

due to $r_n c_n \lesssim 1$ (Assumption 6(ii.1)). This, together with (58), yields:

$$\sup_{s\in\mathbb{R}} \left| \mathbb{P}\left(\sup_{t\in\mathcal{T}} \left|\frac{\sqrt{n}(f_t(\widehat{h}) - f_t(h_0))}{\widehat{\sigma}_n(f_t)}\right| \leq s\right) - \mathbb{P}\left(\sup_{t\in\mathcal{T}} |\mathbb{Z}_n(t)| \leq s\right)\right| = o(1)\,. \tag{59}$$

Moreover, by (57) we may choose a sequence of positive constants ϵ_n' with $\epsilon_n' = o(1)$ such that

$$\sup_{t\in\mathcal{T}} \left|\mathbb{Z}_n^*(t) - \widetilde{\mathbb{Z}}_n^*(t)\right| \leq \epsilon_n' r_n$$

holds wpa1. Similar arguments then yield:

$$\sup_{s\in\mathbb{R}} \mathbb{P}^*\left(\sup_{t\in\mathcal{T}} |\widetilde{\mathbb{Z}}_n^*(t) - s| \leq \epsilon_n' r_n\right) \lesssim \epsilon_n' = o(1)$$

wpa1. This, together with equation (57), yields:

$$\sup_{s\in\mathbb{R}} \left|\mathbb{P}^*\left(\sup_{t\in\mathcal{T}} |\mathbb{Z}_n^*(t)| \leq s\right) - \mathbb{P}^*\left(\sup_{t\in\mathcal{T}} |\widetilde{\mathbb{Z}}_n^*(t)| \leq s\right)\right| = o_p(1)\,. \tag{60}$$

The result is immediate from equations (59) and (60) and the fact that

$$\mathbb{P}\left(\sup_{t\in\mathcal{T}} |\mathbb{Z}_n(t)| \leq s\right) = \mathbb{P}^*\left(\sup_{t\in\mathcal{T}} |\widetilde{\mathbb{Z}}_n^*(t)| \leq s\right) \quad \text{wpa1 in } \mathbb{P}$$

holds uniformly in s. □

Lemma G.5. *Let Assumption 6(i) hold. Then:* $E[\sup_{t\in\mathcal{T}} |\mathbb{Z}_n(t)|] \lesssim c_n$ *and* $\sup_{t\in\mathcal{T}} |\mathbb{Z}_n(t)| = O_p(c_n)$*.*

Proof of Lemma G.5. Observe that $d_n(t_1, t_2) := E[(\mathbb{Z}_n(t_1) - \mathbb{Z}_n(t_2))^2]^{1/2}$. By Corollary 2.2.8 of van der Vaart and Wellner (1996) and Assumption 6(i), there exists a universal constant C such that

$$E[\sup_t |\mathbb{Z}_n(t)|] \leq E[|\mathbb{Z}_n(\bar{t})|] + C \int_0^\infty \sqrt{\log N(\mathcal{T}, d_n, \epsilon)}\, \mathrm{d}\epsilon$$

for any $\bar{t} \in \mathcal{T}$, where $E[|\mathbb{Z}_n(\bar{t})|] = \sqrt{2/\pi}$ because $\mathbb{Z}_n(\bar{t}) \sim N(0,1)$. Therefore, $E[\sup_t |\mathbb{Z}_n(t)|] \lesssim c_n$. The second result follows by Markov's inequality. □

Lemma G.6. *Let Assumptions 1(iii), 2, 3(ii)(iii), 4(i) and 6 hold,* $\tau_J \zeta \sqrt{(\log n)/n} = o(1)$ *and* $\|\widehat{h} - h_0\|_\infty = O_p(\delta_{h,n})$ *with* $\delta_{h,n} = o(1)$*. Let Assumption 5(b)(iii) hold with* $\eta_n' \sqrt{J} = o(r_n)$ *for nonlinear* $f_t()$*. Let* $\mathcal{Z}_n^*$ *and* $\widetilde{\mathbb{Z}}_n^*(t)$ *be as in the proof of Theorem 4.1. Then:*

$$\sup_{t\in\mathcal{T}} \left|\frac{(Df_t(\widehat{h})[\psi^J])'[\widehat{S}'\widehat{G}_b^{-1}\widehat{S}]^{-1}\widehat{S}'\widehat{G}_b^{-1/2}}{\widehat{\sigma}_n(f_t)} \mathcal{Z}_n^* - \widetilde{\mathbb{Z}}_n^*(t)\right| = o_p^*(r_n) \quad \textit{wpa1 in } \mathbb{P}.$$

Proof of Lemma G.6. First note that because $\mathcal{Z}_n^* \sim N(0, \widehat{\Omega}^o)$ and the minimum and maximum eigenvalues of $\widehat{\Omega}^o$ are uniformly bounded away from 0 and ∞ wpa1 (by Lemma G.3 and Assumptions 2(i)(iii)), we have $\|\mathcal{Z}_n^*\| = O_{p^*}(\sqrt{K})$ wpa1 by Chebyshev's inequality.

Now, writing out term by term we have:

$$\begin{aligned}
&\sup_{t\in\mathcal{T}}\left|\left(\frac{(Df_t(\widehat{h})[\psi^J])'[\widehat{S}'\widehat{G}_b^{-1}\widehat{S}]^{-1}\widehat{S}'\widehat{G}_b^{-1/2}}{\widehat{\sigma}_n(f_t)} - \frac{(Df_t(h_0)[\psi^J])'[S'G_b^{-1}S]^{-1}S'G_b^{-1/2}(\Omega^o)^{1/2}(\widehat{\Omega}^o)^{-1/2}}{\sigma_n(f_t)}\right)\mathcal{Z}_n^*\right| \\
&\leq \sup_{t\in\mathcal{T}}\left|\frac{\left(Df_t(\widehat{h})[\psi^J] - Df_t(h_0)[\psi^J]\right)'(G_b^{-1/2}S)_l^- G_b^{-1/2}S[\widehat{S}'\widehat{G}_b^{-1}\widehat{S}]^{-1}\widehat{S}'\widehat{G}_b^{-1/2}}{\widehat{\sigma}_n(f_t)}\mathcal{Z}_n^*\right| \\
&+ \sup_{t\in\mathcal{T}}\left|\frac{(Df_t(h_0)[\psi^J])'\left([\widehat{S}'\widehat{G}_b^{-1}\widehat{S}]^{-1}\widehat{S}'\widehat{G}_b^{-1/2} - [S'G_b^{-1}S]^{-1}S'G_b^{-1/2}(\Omega^o)^{1/2}(\widehat{\Omega}^o)^{-1/2}\right)}{\sigma_n(f_t)}\mathcal{Z}_n^*\right| \times \sup_{t\in\mathcal{T}}\frac{\sigma_n(f_t)}{\widehat{\sigma}_n(f_t)} \\
&+ \sup_{t\in\mathcal{T}}\left|\frac{\sigma_n(f_t)}{\widehat{\sigma}_n(f_t)} - 1\right| \times \sup_{t\in\mathcal{T}}\left|\frac{(Df_t(h_0)[\psi^J])'[S'G_b^{-1}S]^{-1}S'G_b^{-1/2}(\Omega^o)^{1/2}(\widehat{\Omega}^o)^{-1/2}}{\sigma_n(f_t)}\mathcal{Z}_n^*\right| \quad =: T_1 + T_2 + T_3\,.
\end{aligned}$$

Control of T_1: By Cauchy-Schwarz, we have:

$$\begin{aligned}
T_1 &\leq \sup_{t\in\mathcal{T}}\frac{\|\Pi_K T(\widehat{v}_n(f_t) - v_n(f_t))\|_{L^2(W)}}{\sigma_n(f_t)} \times \sup_{t\in\mathcal{T}}\frac{\sigma_n(f_t)}{\widehat{\sigma}_n(f_t)} \times \left\|G_b^{-1/2}S[\widehat{S}'\widehat{G}_b^{-1}\widehat{S}]^{-1}\widehat{S}'\widehat{G}_b^{-1/2}\right\|_{\ell^2} \times \|(\Omega^o)^{1/2}(\widehat{\Omega}^o)^{-1/2}\|_{\ell^2} \times \|\mathcal{Z}_n^*\|_{\ell^2} \\
&= o_p(\eta_n') \times O_p(1) \times O_p(1) \times O_p(1) \times O_{p^*}(\sqrt{K}) = o_p^*(r_n)
\end{aligned}$$

where the first term is by Assumption 5(b)(iii) (or zero if the f_t are linear functionals), the second term is by Lemma G.4, the third is by Lemma F.10(c) (using the fact that $s_{JK} \asymp \tau_J$, see Lemma A.1) and the fact that $\|G_b^{-1/2}S[S'G_b^{-1}S]^{-1}S'G_b^{-1/2}\|_{\ell^2} = 1$, and the fourth term is by Lemma G.3. Therefore, $T_1 = O_{p^*}(\eta_n'\sqrt{J})$ wpa1 (since $K \asymp J$), and is therefore $= o_p^*(r_n)$ wpa1 by the condition stated in this Lemma.

Control of T_2: Let $\Delta\mathbb{Z}_n(t)$ denote the Gaussian process (under $\mathbb{P}^*$) defined by

$$\Delta\mathbb{Z}_n(t) = \frac{(Df_t(h_0)[\psi^J])'\left([\widehat{S}'\widehat{G}_b^{-1}\widehat{S}]^{-1}\widehat{S}'\widehat{G}_b^{-1/2} - [S'G_b^{-1}S]^{-1}S'G_b^{-1/2}(\Omega^o)^{1/2}(\widehat{\Omega}^o)^{-1/2}\right)}{\widehat{\sigma}_n(f_t)}\mathcal{Z}_n^*$$

for each $t \in \mathcal{T}$. The intrinsic semi-metric $\Delta d_n(t_1,t_2)$ of $\Delta\mathbb{Z}_n(t)$ is $\Delta d_n(t_1,t_2)^2 = E^*[(\Delta\mathbb{Z}_n(t_1) - \Delta\mathbb{Z}_n(t_2))^2]$ for each $t_1, t_2 \in \mathcal{T}$, where E^* denotes expectation under the measure $\mathbb{P}^*$. Observe that:

$$\begin{aligned}
\Delta d_n(t_1,t_2) =& \left\|\left(\frac{Df_{t_1}(h_0)[\psi^J]}{\sigma_n(f_{t_1})} - \frac{Df_{t_2}(h_0)[\psi^J]}{\sigma_n(f_{t_2})}\right)'[S'G_b^{-1}S]^{-1}S'G_b^{-1/2}(\Omega^o)^{1/2}\right. \\
&\left.\times (\Omega^o)^{-1/2}G_b^{-1/2}S\left\{[\widehat{S}'\widehat{G}_b^{-1}\widehat{S}]^{-1}\widehat{S}'\widehat{G}_b^{-1/2} - [S'G_b^{-1}S]^{-1}S'G_b^{-1/2}(\Omega^o)^{1/2}(\widehat{\Omega}^o)^{-1/2}\right\}(\widehat{\Omega}^o)^{1/2}\right\|_{\ell^2} \\
\lesssim& d_n(t_1,t_2) \times \left\|G_b^{-1/2}S'\left\{[\widehat{S}'\widehat{G}_b^{-1}\widehat{S}]^{-1}\widehat{S}'\widehat{G}_b^{-1/2} - [S'G_b^{-1}S]^{-1}S'G_b^{-1/2}(\Omega^o)^{1/2}(\widehat{\Omega}^o)^{-1/2}\right\}\right\|_{\ell^2}
\end{aligned}$$

wpa1, where the first line uses the fact that $\widehat{\Omega}^o$ is invertible wpa1 and the second line uses the fact that Ω^o and $\widehat{\Omega}^o$ have eigenvalue uniformly bounded away from 0 and ∞ wpa1. It follows by Lemma F.10(c) and Lemmas

G.3 and F.3 that

$$
\begin{aligned}
&\left\| G_b^{-1/2} S' \left\{ [\widehat{S}'\widehat{G}_b^{-1}\widehat{S}]^{-1}\widehat{S}'\widehat{G}_b^{-1/2} - [S'G_b^{-1}S]^{-1}S'G_b^{-1/2}(\Omega^o)^{1/2}(\widehat{\Omega}^o)^{-1/2} \right\} \right\|_{\ell^2} \\
&\leq \left\| G_b^{-1/2} S' \left\{ [\widehat{S}'\widehat{G}_b^{-1}\widehat{S}]^{-1}\widehat{S}'\widehat{G}_b^{-1/2} - [S'G_b^{-1}S]^{-1}S'G_b^{-1/2} \right\} \right\|_{\ell^2} + \left\| I - (\Omega^o)^{1/2}(\widehat{\Omega}^o)^{-1/2} \right\|_{\ell^2} \\
&= O_p(\tau_J \zeta \sqrt{(\log J)/n}) + O_p\Big(\big(\zeta_{b,K}^{(2+\delta)/\delta} \sqrt{(\log K)/n}\big)^{\delta/(1+\delta)} + \delta_{h,n} \Big) \\
&= O_p(\delta_{V,n})\,.
\end{aligned}
$$

Therefore,

$$
\Delta d_n(t_1,t_2) \leq O_p(\delta_{V,n}) \times d_n(t_1,t_2)
$$

wpa1. Moreover, by similar arguments we have

$$
\sup_{t \in \mathcal{T}} E^*[(\Delta \mathbb{Z}_n(t))^2]^{1/2} \lesssim O_p(\delta_{V,n})
$$

wpa1. Therefore, we can scale $\Delta \mathbb{Z}_n(t)$ by dividing through by a sequence of positive constants of order $\delta_{V,n}$ to obtain

$$
E^*[\sup_{t \in \mathcal{T}} |\Delta \mathbb{Z}_n(t)|] \lesssim O_p(\delta_{V,n}) \times c_n
$$

wpa1 by identical arguments to the proof of Lemma G.5. Therefore,

$$
T_2 \leq O_p(\delta_{V,n} \times c_n) \times \sup_{t \in \mathcal{T}} \frac{\sigma_n(f_t)}{\widehat{\sigma}_n(f_t)} = O_p(\delta_{V,n} \times c_n) \times O_p(1)
$$

wpa1 by Lemma G.4 and so $T_2 = o_{p^*}(r_n)$ under Assumption 6(ii.2).

Control of T_3: The second term in T_3 is the supremum of a Gaussian process with the same distribution (under $\mathbb{P}^*$) as $\mathbb{Z}_n(t)$ (under $\mathbb{P}$). Therefore, by Lemmas G.4 and G.5 we have:

$$
T_3 = O_p(\delta_{V,n} + \eta_n') \times O_{p^*}(c_n)
$$

and so $T_3 = o_p^*(r_n)$ wpa1 under Assumption 6(ii.2). □

Proof of Remark 4.2. For any $t_1, t_2 \in \mathcal{T}$ we have:

$$
\begin{aligned}
d_n(t_1,t_2) &\leq \frac{2}{\sigma_n(f_{t_1}) \vee \sigma_n(f_{t_2})} \left\| \Omega^{1/2} G_b^{-1} S [S'G_b^{-1}S]^{-1} \left(Df_{t_1}(h_0)[\psi^J] - Df_{t_2}(h_0)[\psi^J] \right) \right\|_{\ell^2} \\
&\leq \frac{2\overline{\sigma} s_{JK}^{-1}}{\underline{\sigma}_n} \left\| G_\psi^{-1/2} \left(Df_{t_1}(h_0)[\psi^J] - Df_{t_2}(h_0)[\psi^J] \right) \right\|_{\ell^2} \\
&\lesssim \frac{\tau_J \Gamma_n}{\underline{\sigma}_n} \|t_1 - t_2\|_{\ell^2}^{\gamma_n}
\end{aligned}
\tag{61}
$$

where the first inequality is because $\|x/\|x\| - y/\|y\|\| \leq 2\|x-y\|/(\|x\| \vee \|y\|)$ whenever $\|x\|, \|y\| \neq 0$ and the third is by the equivalence $s_{JK}^{-1} \asymp \tau_J$ (see Lemma A.1). By (61) and compactness of $\mathcal{T}$, we have $N(\mathcal{T}, d_n, \epsilon) \leq C(\tau_J \Gamma_n/(\epsilon \underline{\sigma}_n))^{d_T/\gamma_n} \vee 1$ for some finite constant C. □

Proof of Corollary 4.1. We verify the conditions of Lemma 4.1 (or Theorem 4.1). By assumption we may take $\underline{\sigma}_n \asymp \tau_J J^a$ with $a = \frac{1}{2} + \frac{|\alpha|}{d}$. Assumption 5 is therefore satisfied with $\eta_n = \sqrt{n}\tau_J^{-1} J^{-(p/d+1/2)}$ by Remark 4.1(a') and Lemma A.3.

The continuity condition in Remark 4.2 holds with $\Gamma_n = O(J^{a'})$ for some $a' > 0$ and $\gamma_n = 1$ since Ψ_J is spanned by a B-spline basis of order $\gamma > (p \vee 2 + |\alpha|)$ (DeVore and Lorentz, 1993, Section 5.3). Assumption 6(i) therefore holds with $c_n = O(\sqrt{\log J})$ by Remark 4.2 because $(\tau_J \Gamma_n/(\epsilon \underline{\sigma}_n)) \lesssim (J^{a'-a}\epsilon^{-1})$. We can therefore take $r_n = (\log J)^{-\kappa}$ for $\kappa \in [1/2, 1]$ in Assumption 6(ii). The first condition in Assumption 6(ii) then holds provided $J^5(\log J)^{6\kappa}/n = o(1)$. Since $\eta_n' = 0$, the second condition in Assumption 6(ii) holds provided

$$\tau_J J\sqrt{(\log J)/n} + \sqrt{n}\tau_J^{-1}J^{-(p/d+1/2)} + \left([J^{\frac{2+\delta}{2\delta}}\sqrt{(\log J)/n}]^{\frac{\delta}{1+\delta}} + J^{-p/d} + \tau_J\sqrt{J(\log J)/n}\right)\sqrt{\log J} = o((\log J)^{-\kappa})$$

(using Corollary 3.1 for $\delta_{h,n}$). In applying Corollary 3.1 we require that the conditions $\tau_J J/\sqrt{n} = O(1)$ and $J^{(2+\delta)/2\delta}\sqrt{(\log n)/n} = o(1)$ hold. Finally to apply Theorem 4.1 we also need $\tau_J J\sqrt{(\log J)/n} = o(1)$. Sufficient conditions for all these restrictions on J are provided in the statement of this corollary. In particular, we note that $[J^{\frac{2+\delta}{2\delta}}\sqrt{(\log n)/n}](\log J)^{\frac{1+\delta}{\delta}}$ decreases as $\delta > 0$ increases. Hence the condition $J^5(\log n)^{6\kappa}/n = o(1)$ (for $\kappa \in [1/2, 1]$) implies that $[J^{\frac{2+\delta}{2\delta}}\sqrt{(\log n)/n}]^{\frac{\delta}{1+\delta}}(\log J)^{\kappa+0.5} = o(1)$ holds for all $\delta \geq 1$. $\square$

G.5 Proofs for Section 5

Proof of Theorem 5.1. The result will follow from Theorem D.1. Assumption 2(i)–(iii)(iv') is satisfied under Assumption CS(iii). Assumption 3(i)(ii)(iii) is satisfied by Assumption CS (iv) and the second part of Assumption CS(v), noting that $\zeta_{\psi,J} = O(\sqrt{J})$ and $\zeta_{b,K} = O(\sqrt{K}) = O(\sqrt{J})$. Since the basis spanning Ψ_J is a Riesz basis for T Assumption 4 is satisfied with $\tau_J \asymp \mu_J^{-1}$.

It remains to verify Assumption 5'(b). By the Riesz basis property and Assumption 2(i)–(iii) we have $[\sigma_n(f_{CS})]^2 \asymp \sum_{j=1}^J (a_j/\mu_j)^2 \lesssim J\mu_J^{-2}$ (see Section 6 of Chen and Pouzo (2015)). For f_{CS} we have

$$Df_{CS}(h_0)[h - h_0] = \int_0^1 \left(\{h(\mathsf{p}(t), \mathsf{y} - S_{\mathsf{y}}(t)) - h_0(t, \mathsf{y} - S_{\mathsf{y}}(t))\}e^{-\int_0^t \partial_2 h_0(\mathsf{p}(v), \mathsf{y} - S_{\mathsf{y}}(v))\mathsf{p}'(v)\,\mathrm{d}v}\mathsf{p}'(t)\right)\mathrm{d}t$$

(Hausman and Newey, 1995, p. 1471) which is clearly a linear functional (Assumption 5'(b)(i)). Note that $\sigma_n(f_{CS}) \lesssim \sqrt{J}\mu_J^{-1}$ and Assumption CS(v) together imply $\mu_J^{-1}J^{3/2}\sqrt{(\log J)/n} = o(1)$. This, $p > 2$ and Corollary 3.1 together imply that $\|\widehat{h} - h_0\|_{B^2_{\infty,\infty}} = o_p(1)$ and $\|\widetilde{h} - h_0\|_{B^2_{\infty,\infty}} = o_p(1)$, and

$$\|\widehat{h} - h_0\|_\infty = O_p\left(J^{-p/2} + \mu_J^{-1}\sqrt{(J\log J)/n}\right), \quad \|\widehat{h} - h_0\|_{B^1_{\infty,\infty}} = O_p\left(\sqrt{J}\left(J^{-p/2} + \mu_J^{-1}\sqrt{(J\log J)/n}\right)\right) \tag{62}$$

$$\|\widetilde{h} - h_0\|_\infty = O_p\left(J^{-p/2}\right), \quad \|\widetilde{h} - h_0\|_{B^1_{\infty,\infty}} = O_p\left(\sqrt{J}\left(J^{-p/2}\right)\right). \tag{63}$$

Applying Lemma A1 of Hausman and Newey (1995), we obtain

$$\left|f_{CS}(\widehat{h}) - f_{CS}(h_0) - Df_{CS}(h_0)[\widehat{h} - h_0]\right| = O_p\left(\sqrt{J}\left(J^{-p} + \mu_J^{-2}\frac{J\log J}{n}\right)\right)$$
$$\left|Df_{CS}(h_0)[\widetilde{h} - h_0]\right| = O_p(J^{-p/2}).$$

Since $p > 2$, Assumption CS(v) guarantees that Assumption 5'(b)(ii) holds with

$$\eta_n = \frac{\sqrt{n}}{\sigma_n(f_{CS})} \times \left(J^{-p/2} + \mu_J^{-2}\frac{J^{3/2}\log J}{n}\right) = o(1).$$

Finally, for Assumption 5'(b)(iii), we have

$$\frac{\|\Pi_K T(\widehat{v}_n(f_{CS}) - v_n(f_{CS}))\|_{L^2(W)}}{\sigma_n(f_{CS})} \lesssim \frac{\tau_J \sqrt{\sum_{j=1}^J \left(Df_{CS}(\widehat{h})[(G_\psi^{-1/2}\psi^J)_j] - Df_{CS}(h_0)[(G_\psi^{-1/2}\psi^J)_j] \right)^2}}{\sigma_n(f_{CS})}$$

where $\tau_J \asymp \mu_J^{-1}$. Moreover,

$$\left| Df_{CS}(\widehat{h})[(G_\psi^{-1/2}\psi^J)_j] - Df_{CS}(h_0)[(G_\psi^{-1/2}\psi^J)_j] \right| \lesssim \sqrt{J} \times O_p\left(\sqrt{J}\left(J^{-p/2} + \mu_J^{-1}\sqrt{(J\log J)/n}\right)\right)$$

(uniformly in $j = 1, \ldots J$) by Lemma A1 of Hausman and Newey (1995). Therefore

$$\frac{\|\Pi_K T(\widehat{v}_n(f_{CS}) - v_n(f_{CS}))\|_{L^2(W)}}{\sigma_n(f_{CS})} \lesssim \frac{J^{3/2}\mu_J^{-1} O_p\left(J^{-p/2} + \mu_J^{-1}\sqrt{(J\log J)/n}\right)}{\left(\sum_{j=1}^J (a_j/\mu_j)^2\right)^{1/2}} = O_p(\eta_n')$$

which is $o_p(1)$ by Assumption CS(v). Finally we note that the condition $\mu_J^{-1} J\sqrt{(\log J)/n} = o(1)$ of Theorem D.1 is trivially implied by $\mu_J^{-1}J^{3/2}\sqrt{(\log J)/n} = o(1)$ (which is in turn implied by Assumption CS(v)). This proves the result. □

Proof of Corollary 5.1. For **Result (1)**, since $\sigma_n(f_{CS}) \asymp J^{(a+\varsigma+1)/2}$, the first part of Assumption CS(v) is satisfied provided

$$\frac{\sqrt{n}}{J^{(a+\varsigma+1)/2}}\left(J^{-p/2} + J^{\varsigma+2}\sqrt{(\log J)}/n\right) = o(1)$$

for which a sufficient condition is $nJ^{-(p+a+\varsigma+1)} = o(1)$ and $J^{3+\varsigma-a}(\log n)/n = o(1)$. Moreover, the condition $\mu_J^{-1}J^{3/2}\sqrt{(\log J)/n} = o(1)$ is implied by $J^{3+\varsigma-(a\wedge 0)}(\log n)/n = o(1)$. The condition $J^{3+\varsigma-(a\wedge 0)}(\log n)/n = o(1)$ also implies that $J^{(2+\delta)/(2\delta)}\sqrt{(\log n)/n} = o(1)$ holds whenever $\delta \geq 2/(2+\varsigma-(a\wedge 0))$.

For **Result (2)**, we have $\sigma_n(f_{CS})^2 \gtrsim a_J^2/\mu_J^2 \asymp \exp(J^{\varsigma/2} + a\log J)$. Take $J = (\log(n/(\log n)^\varrho))^{2/\varsigma}$. Then

$$\begin{aligned}
\sigma_n(f_{CS})^2 &\gtrsim \exp\left(\log(n/(\log n)^\varrho) + \log[(\log(n/(\log n)^\varrho))^{2a/\varsigma}]\right) \\
&= \exp\left(\log[n/(\log n)^\varrho \times (\log(n/(\log n)^\varrho))^{2a/\varsigma}]\right) \\
&= n/(\log n)^\varrho \times (\log(n/(\log n)^\varrho))^{2a/\varsigma}
\end{aligned}$$

and so

$$\sigma_n(f_{CS}) \gtrsim \frac{\sqrt{n}}{(\log n)^{\varrho/2}} \times (\log(n/(\log n)^\varrho))^{a/\varsigma}.$$

The first part of Assumption CS(v) is then satisfied provided

$$\frac{(\log n)^{\varrho/2}}{(\log(n/(\log n)^\varrho))^{a/\varsigma}}\left((\log(n/(\log n)^\varrho))^{-p/\varsigma} + (\log n)^{-\varrho} \times (\log(n/(\log n)^\varrho))^{4/\varsigma}\log\log n\right) = o(1)$$

which holds provided $2p > \varrho\varsigma - 2a$ and $\varrho\varsigma > 8 - 2a$. The condition $J^{(2+\delta)/(2\delta)}\sqrt{(\log n)/n} = o(1)$ holds for any $\delta > 0$. The remaining condition $\mu_J^{-1}J^{3/2}\sqrt{(\log J)/n} = o(1)$ is implied by

$$\frac{\sqrt{n}}{(\log n)^{\varrho/2}}(\log(n/(\log n)^\varrho))^{3/\varsigma}\sqrt{(\log\log n)/n} = o(1)$$

for which a sufficient condition is $\varrho\varsigma > 6$. Now, we may always choose $\varrho > 0$ so that $\varrho\varsigma > 6 \vee (8-2a)$. The remaining condition then holds provided $2p > 6 \vee (8-2a) - 2a$. $\square$

Proof of Theorem 5.2. The proof follows by identical arguments to those of Theorem 5.1, noting that

$$f_{DL}(h) - f_{CS}(h) = (\mathsf{p}^1 - \mathsf{p}^0)h(\mathsf{p}^1, \mathsf{y})$$

and so

$$\begin{aligned} f_{DL}(\widehat{h}) - f_{DL}(h_0) &= f_{CS}(\widehat{h}) - f_{CS}(h_0) + (\mathsf{p}^1 - \mathsf{p}^0)\left(\widehat{h}(\mathsf{p}^1, \mathsf{y}) - h_0(\mathsf{p}^1, \mathsf{y})\right) \\ Df_{DL}(\widehat{h})[h - h_0] &= Df_{CS}(h_0)[h - h_0] + (\mathsf{p}^1 - \mathsf{p}^0)\left(\widehat{h}(\mathsf{p}^1, \mathsf{y}) - h_0(\mathsf{p}^1, \mathsf{y})\right) \\ Df_{DL}(\widehat{h})[v] - Df_{DL}(h_0)[v] &= Df_{CS}(\widehat{h})[v] - Df_{CS}(h_0)[v] \end{aligned}$$

where clearly $|(\mathsf{p}^1 - \mathsf{p}^0)(\widehat{h}(\mathsf{p}^1, \mathsf{y}) - h_0(\mathsf{p}^1, \mathsf{y}))| \leq \text{const} \times \|\widehat{h} - h_0\|_\infty$. Since $\sigma_n(f_{DL}) \asymp \mu_J^{-1}\sqrt{J}$, the stated conditions on J in this theorem imply that Assumption CS(v) holds. $\square$

Proof of Theorem 5.3. The result will follow from Theorem D.1, and is very similar to that of Theorem 5.1. Assumptions 2(i)–(iii)(iv'), 3(i)(ii)(iii), and 4 are verified as in the proof of Theorem 5.1. It remains to verify Assumption 5'(b). As in the proof of Theorem 5.1 we have $\tau_J \asymp \mu_J^{-1}$ and $[\sigma_n(f_A)]^2 \asymp \sum_{j=1}^J (a_j/\mu_j)^2$ (see Section 6 of Chen and Pouzo (2015)). Simple expansion of f_A yields

$$Df_A(h_0)[h - h_0] = \int w(\mathsf{p}) e^{h_0(\log \mathsf{p}, \log \mathsf{y})}(h(\log \mathsf{p}, \log \mathsf{y}) - h_0(\log \mathsf{p}, \log \mathsf{y}))\, \mathrm{d}\mathsf{p}$$

which is clearly a linear functional (Assumption 5'(b)(i)), and

$$\begin{aligned} &\left| f_A(\widehat{h}) - f_A(h_0) - Df_A(h_0)[\widehat{h} - h_0] \right| \\ &= \int w(\mathsf{p}) \Big(e^{\widehat{h}(\log \mathsf{p}, \log \mathsf{y}) - h_0(\log \mathsf{p}, \log \mathsf{y})} - 1 - \big(\widehat{h}(\log \mathsf{p}, \log \mathsf{y}) - h_0(\log \mathsf{p}, \log \mathsf{y})\big)\Big) e^{h_0(\log \mathsf{p}, \log \mathsf{y})}\, \mathrm{d}\mathsf{p}\,. \end{aligned}$$

Therefore, by Corollary 3.1 we have

$$\begin{aligned} \left| f_A(\widehat{h}) - f_A(h_0) - Df_A(h_0)[\widehat{h} - h_0] \right| &= O_p\left(J^{-p} + \mu_J^{-2} \frac{J \log J}{n} \right) \\ \left| Df_A(h_0)[\widetilde{h} - h_0] \right| &= O_p(J^{-p/2})\,. \end{aligned}$$

Since $p > 0$, the stated conditions on J in this theorem guarantees that Assumption 5'(b)(ii) holds with

$$\eta_n = \frac{\sqrt{n}}{\sigma_n(f_A)} \times \left(J^{-p/2} + \mu_J^{-2} \frac{J \log J}{n} \right) = o(1)\,.$$

Finally, for Assumption 5'(b)(iii), we have

$$\frac{\|\Pi_K T(\widehat{v}_n(f_A) - v_n(f_A))\|_{L^2(W)}}{\sigma_n(f_A)} \lesssim \frac{\tau_J}{\sigma_n(f_A)} \sqrt{\left(\sum_{j=1}^J \left(Df_A(\widehat{h})[(G_\psi^{-1/2}\psi^J)_j] - Df_A(h_0)[(G_\psi^{-1/2}\psi^J)_j] \right)^2 \right)}$$

where $\tau_J \asymp \mu_J^{-1}$ and where a first-order Taylor expansion of Df_A yields

$$\begin{aligned}\left|Df_A(\widehat{h})[(G_\psi^{-1/2}\psi^J)_j] - Df_A(h_0)[(G_\psi^{-1/2}\psi^J)_j]\right| &\lesssim \|(G_\psi^{-1/2}\psi^J)_j\|_\infty \times \|\widehat{h}-h_0\|_\infty \\ &= O_p\left(\sqrt{J}\left(J^{-p/2} + \mu_J^{-1}\sqrt{(J\log J)/n}\right)\right).\end{aligned}$$

It follows that

$$\frac{\|\Pi_K T(\widehat{v}_n(f_A) - v_n(f_A))\|_{L^2(W)}}{[\sigma_n(f_A)]} \lesssim \frac{\mu_J^{-1} J \times O_p\left(J^{-p/2} + \mu_J^{-1}\sqrt{(J\log J)/n}\right)}{\left(\sum_{j=1}^J (a_j/\mu_j)^2\right)^{1/2}} = O_p(\eta_n')$$

which is $o_p(1)$ by the displayed condition on J in this theorem. Finally we note that the condition $\mu_J^{-1} J\sqrt{(\log J)/n} = o(1)$ of Theorem D.1 is implied by the displayed condition on J in this theorem and the fact that $\sigma_n(f_A) \lesssim \sqrt{J}\mu_J^{-1}$. This proves the result. □

Proof of Theorem 5.4. We verify the conditions of Lemma 4.1 and Theorem 4.1. Assumptions 1 and 2 are satisfied by Assumption CS(i)(ii)(iv) and U-CS(i). Assumption 3(iii) is satisfied by Assumption U-CS(iii). Assumption 4(i) is satisfied by the Riesz basis condition. For Assumption 5(b), we check the conditions of Remark 4.1(b'). It is clear that $Df_{CS,t}[h-h_0]$ (see display (14)) is a linear functional of $h-h_0$ for each $t \in \mathcal{T}$. As in the proof of Theorem 5.1 we have $[\sigma_n(f_{CS,t})]^2 \asymp \sum_{j=1}^J (a_{j,t}/\mu_j)^2$ uniformly in t. Thus, $\underline{\sigma}_n \lesssim \sqrt{J}\mu_J^{-1}$. This and Assumption U-CS(iv.1) together imply $\mu_J^{-1} J^{3/2}\sqrt{(\log J)/n} = o(1)$. Also we note that the first part of Assumption U-CS(iii) and $\delta \geq 1$ imply that $J^{(2+\delta)/(2\delta)}\sqrt{(\log n)/n} = o(1)$ holds. These results and $p > 2$ and Corollary 3.1 together imply that $\|\widehat{h} - h_0\|_{B^2_{\infty,\infty}} = o_p(1)$ and $\|\widetilde{h} - h_0\|_{B^2_{\infty,\infty}} = o_p(1)$, and equations (62) and (63) hold. Therefore, $\widehat{h}$ and $\widetilde{h}$ are within an ϵ neighborhood (in Hölder norm of smoothness 2) of h_0 wpa1. As $\mathcal{T} = [\underline{\mathsf{p}}^0, \overline{\mathsf{p}}^0] \times [\underline{\mathsf{p}}^1, \overline{\mathsf{p}}^1] \times [\underline{\mathsf{y}}, \overline{\mathsf{y}}]$ where the intervals $[\underline{\mathsf{p}}^0, \overline{\mathsf{p}}^0]$ and $[\underline{\mathsf{p}}^1, \overline{\mathsf{p}}^1]$ are in the interior of the support of P_i and $[\underline{\mathsf{y}}, \overline{\mathsf{y}}]$ is in the interior of the support of Y_i and $h_0 \in B_\infty(p, L)$ with $p > 2$ and $0 < L < \infty$, it is straightforward to extend Lemma A1 of Hausman and Newey (1995) to show

$$\begin{aligned}\sup_{t\in\mathcal{T}}\left|f_{CS,t}(\widehat{h}) - f_{CS,t}(h_0) - Df_{CS,t}(h_0)[\widehat{h}-h_0]\right| &= O_p\left(\sqrt{J}\left(J^{-p} + \mu_J^{-2}\frac{J\log J}{n}\right)\right), \\ \sup_{t\in\mathcal{T}}\left|Df_{CS,t}(h_0)[\widetilde{h}-h_0]\right| &= O_p(J^{-p/2})\end{aligned}$$

by Corollary 3.1. Since $\underline{\sigma}_n \lesssim \sqrt{J}\mu_J^{-1}$, Assumption U-CS(iii) guarantees that Assumption 5(b)(ii) holds with

$$\eta_n = \frac{\sqrt{n}}{\underline{\sigma}_n} \times \left(J^{-p/2} + \mu_J^{-2}\frac{J^{3/2}\log J}{n}\right).$$

For Assumption 5(b)(iii), we have

$$\sup_{t\in\mathcal{T}}\frac{\|\Pi_K T(\widehat{v}_n(f_{CS,t}) - v_n(f_{CS,t}))\|_{L^2(W)}}{[\sigma_n(f_{CS,t})]} \lesssim \sup_{t\in\mathcal{T}} \tau_J \frac{\sqrt{\sum_{j=1}^J\left(Df_{CS,t}(\widehat{h})[(G_\psi^{-1/2}\psi^J)_j] - Df_{CS,t}(h_0)[(G_\psi^{-1/2}\psi^J)_j]\right)^2}}{[\sigma_n(f_{CS,t})]}$$

where $\tau_J \asymp \mu_J^{-1}$. By straightforward extension of Lemma A1 of Hausman and Newey (1995) and (62) and (63):

$$\sup_{t\in\mathcal{T}}\left|Df_{CS,t}(\widehat{h})[(G_\psi^{-1/2}\psi^J)_j] - Df_{CS,t}(h_0)[(G_\psi^{-1/2}\psi^J)_j]\right| \lesssim \sqrt{J} \times O_p\left(\sqrt{J}\left(J^{-p/2} + \mu_J^{-1}\sqrt{(J\log J)/n}\right)\right)$$

whence Assumption 5(b)(iii) holds with

$$\eta_n' = \frac{J^{3/2}\mu_J^{-1}}{\underline{\sigma}_n} \times \left(J^{-p/2} + \mu_J^{-1}\sqrt{J(\log J)/n}\right).$$

which is $o(1)$ by Assumption U-CS(iv). This verifies Assumption 5(b).

Finally, Assumption 6(i) holds with $c_n = O(\sqrt{\log J})$ by Assumption U-CS(ii) and Remark 4.2. For Assumption 6(ii) we take $r_n = [\log J]^{-1/2}$. Assumption 6(ii.1) then holds provided $J^5(\log J)^3/n = o(1)$. Assumption 6(ii.2) holds provided

$$\tau_J J\sqrt{(\log J)/n} + \eta_n + \left([J^{\frac{2+\delta}{2\delta}}\sqrt{(\log J)/n}]^{\frac{\delta}{1+\delta}} + J^{-p/d} + \tau_J\sqrt{J(\log J)/n} + \eta_n'\right)\sqrt{\log J} = o((\log J)^{-1/2})$$

(using Corollary 3.1 for $\delta_{h,n}$), which is satisfied provided

$$\tau_J J\sqrt{(\log J)/n} + \eta_n + \eta_n'\sqrt{\log J} = o((\log J)^{-1/2})$$

which is in turn implied by Assumption U-CS(iii) and U-CS(iv.1) and the property $\underline{\sigma}_n \lesssim \sqrt{J}\mu_J^{-1}$. Thus Lemma 4.1 applies to $f_t = f_{CS,t}$ with a rate $r_n = [\log J]^{-1/2}$.

Next we note that the condition $\eta_n'\sqrt{J} = o((\log J)^{-1/2})$ needed for Theorem 4.1 is directly implied by Assumption U-CS(iv.2). □

Proof of Theorem 5.5. Follows by similar arguments to the proofs of Theorems 5.2 and 5.4, noting that

$$|Df_{DL,t_1}(h_0)[h] - Df_{DL,t_2}(h_0)[h]| \le |Df_{CS,t_1}(h_0)[h] - Df_{CS,t_2}(h_0)[h]| + |(\mathsf{p}_1^1 - \mathsf{p}_1^0)h(\mathsf{p}_1^1, \mathsf{y}_1) - (\mathsf{p}_2^1 - \mathsf{p}_2^0)h(\mathsf{p}_2^1, \mathsf{y}_2)|$$

and so $c_n = O(\sqrt{\log J})$ by Assumption U-CS(ii) and Remark 4.2 (see the proof of Corollary 4.1). We can then take $r_n = [\log J]^{-1/2}$. □

G.6 Proofs for Appendix B

Proof of Theorem B.1. As with the proof of Theorem 3.1, we first decompose the error into three parts:

$$\begin{aligned} \|\widehat{h} - h_0\|_{L^2(X)} &\le \|\widehat{h} - \widetilde{h}\|_{L^2(X)} + \|\widetilde{h} - \Pi_J h_0\|_{L^2(X)} + \|\Pi_J h_0 - h_0\|_{L^2(X)} \\ &=: T_1 + T_2 + \|h_0 - \Pi_J h_0\|_{L^2(X)}. \end{aligned}$$

To prove **Result (1)** it is enough to show that $T_2 \le O_p(1) \times \|h_0 - \Pi_J h_0\|_{L^2(X)}$. To do this, bound

$$\begin{aligned} T_2 &\le \|G_\psi^{1/2}(S'G_b^{-1/2})_l^- G_b^{-1/2}B'(H_0 - \Psi c_J)/n\|_{\ell^2} \\ &\quad + \|G_\psi^{1/2}\{(\widehat{G}_b^{-1/2}\widehat{S})_l^- \widehat{G}_b^{-1/2}G_b^{1/2} - (G_b^{-1/2}S)_l^-\}G_b^{-1/2}B'(H_0 - \Psi c_J)/n\|_{\ell^2} \quad =: \quad T_{21} + T_{22}. \end{aligned}$$

For T_{21},

$$\begin{aligned} T_{21} &\leq s_{JK}^{-1}\|G_b^{-1/2}B'(H_0-\Psi c_J)/n\|_{\ell^2} \\ &\leq O_p(\tau_J\zeta_{b,K}/\sqrt{n})\times\|h_0-\Pi_J h_0\|_{L^2(X)}+\tau_J\|\Pi_K T(h_0-\Pi_J h_0)\|_{L^2(W)} \\ &\leq O_p(\tau_J\zeta_{b,K}/\sqrt{n})\times\|h_0-\Pi_J h_0\|_{L^2(X)}+\tau_J\|T(h_0-\Pi_J h_0)\|_{L^2(W)} \\ &= O_p(1)\times\|h_0-\Pi_J h_0\|_{L^2(X)} \end{aligned}$$

where the second line is by Lemma F.9 and the relations $J\asymp K$ and $\tau_J\asymp s_{JK}^{-1}$, and the final line is by Assumption 4(ii) and the condition $\tau_J\zeta\sqrt{(\log J)/n}=o(1)$. Similarly,

$$\begin{aligned} T_{22} &\leq \|G_\psi^{1/2}\{(\widehat{G}_b^{-1/2}\widehat{S})_l^-\widehat{G}_b^{-1/2}G_b^{1/2}-(G_b^{-1/2}S)_l^-\}\|_{\ell^2}\|G_b^{-1/2}B'(H_0-\Psi c_J)/n\|_{\ell^2} \\ &\leq O_p(s_{JK}^{-2}\zeta\sqrt{(\log J)/n})\times\big(O_p(\zeta_{b,K}/\sqrt{n})\times\|h_0-\Pi_J h_0\|_{L^2(X)}+\|T(h_0-\Pi_J h_0)\|_{L^2(W)}\big) \\ &= o_p(\tau_J\zeta\sqrt{(\log J)/n})^2\times\|h_0-\Pi_J h_0\|_{L^2(X)}+O_p(1)\times\tau_J\|T(h_0-\Pi_J h_0)\|_{L^2(W)} \\ &= O_p(1)\times\|h_0-\Pi_J h_0\|_{L^2(X)} \end{aligned}$$

where the second line is by Lemmas F.9 and F.10(b) and the relations $J\asymp K$ and $\tau_J\asymp s_{JK}^{-1}$, and the final line is by the condition $\tau_J\zeta\sqrt{(\log J)/n}=o(1)$ and Assumption 4(ii). This proves Result (1).

To prove **Result (2)** it remains to control T_1. To do this, bound

$$\begin{aligned} T_1 &\leq \|G_\psi^{1/2}(S'G_b^{-1/2})_l^-G_b^{-1/2}B'u/n\|_{\ell^2}+\|G_\psi^{1/2}\{(\widehat{G}_b^{-1/2}\widehat{S})_l^-\widehat{G}_b^{-1/2}G_b^{1/2}-(G_b^{-1/2}S)_l^-\}G_b^{-1/2}B'u/n\|_{\ell^2} \\ &=: T_{11}+T_{12}\,. \end{aligned}$$

For T_{11}, by definition of s_{JK} and Lemma F.8 we have:

$$T_{11}\leq s_{JK}^{-1}\|G_b^{-1/2}B'u/n\|_{\ell^2}=O_p(s_{JK}^{-1}\sqrt{K/n})=O_p(\tau_J\sqrt{J/n})$$

where the final line is because $J\asymp K$ and $\tau_J\asymp s_{JK}^{-1}$ (Lemma A.1). Similarly,

$$\begin{aligned} T_{12} &\leq \|G_\psi^{1/2}\{(\widehat{G}_b^{-1/2}\widehat{S})_l^-\widehat{G}_b^{-1/2}G_b^{1/2}-(G_b^{-1/2}S)_l^-\}\|_{\ell^2}\|G_b^{-1/2}B'u/n\|_{\ell^2} \\ &= O_p(\tau_J^2\zeta\sqrt{(\log J)/n})\times O_p(\sqrt{J/n}) \\ &= o_p(\tau_J\sqrt{J/n}) \end{aligned}$$

where the second line is by Lemmas F.8 and F.10(b) and the relations $J\asymp K$ and $\tau_J\asymp s_{JK}^{-1}$, and the final line is by the condition $\tau_J\zeta\sqrt{(\log J)/n}=o(1)$. □

Proof of Corollary B.1. Analogous to the proof of Corollary 3.1. □

Proof of Theorem B.2. As in the proof of Theorem 3.2, it suffices to prove a lower bound for the Gaussian reduced-form NPIR model (45). Theorem G.3 below does just this. □

Theorem G.3. *Let Condition LB hold with $B_2(p,L)$ in place of $B_\infty(p,L)$ hold for the NPIR model (45) with a random sample $\{(W_i,Y_i)\}_{i=1}^n$. Then for any $0\leq|\alpha|<p$:*

$$\liminf_{n\to\infty}\inf_{\widehat{g}_n}\sup_{h\in B_2(p,L)}\mathbb{P}_h\left(\|\widehat{g}_n-\partial^\alpha h\|_{L^2(X)}\geq cn^{-(p-|\alpha|)/(2(p+\varsigma)+d)}\right)\geq c'>0$$

in the mildly ill-posed case, and

$$\liminf_{n\to\infty}\inf_{\widehat{g}_n}\sup_{h\in B_2(p,L)}\mathbb{P}_h\left(\|\widehat{g}_n-\partial^\alpha h\|_{L^2(X)}\geq c(\log n)^{-(p-|\alpha|)/\varsigma}\right)\geq c'>0$$

in the severely ill-posed case, in the severely ill-posed case, where $\inf_{\widehat{g}_n}$ *denotes the infimum over all estimators of* $\partial^\alpha h$ *based on the sample of size* n, $\sup_{h\in B_2(p,L)}\mathbb{P}_h$ *denotes the sup over* $h\in B_2(p,L)$ *and distributions* (W_i,u_i) *which satisfy Condition LB with* ν *fixed, and the finite positive constants* c,c' *depend only on* p,L,d,ς *and* σ_0.

Proof of Theorem G.3. We use similar arguments to the proof of Theorem G.1, using Theorem 2.5 of Tsybakov (2009) (see Theorem G.2). Again, we first explain the scalar ($d=1$) case in detail. Let $\{\phi_{j,k},\psi_{j,k}\}_{j,k}$ be a wavelet basis of regularity $\gamma>p$ for $L^2([0,1])$ as described in Appendix E.

By construction, the support of each interior wavelet is an interval of length $2^{-j}(2r-1)$. Thus for all j sufficiently large (hence the lim inf in our statement of the Lemma) we may choose a set $M\subset\{r,\dots,2^j-r-1\}$ of interior wavelets with cardinality $\#(M)\asymp 2^j$ such that $\mathrm{support}(\psi_{j,m})\cap\mathrm{support}(\psi_{j,m'})=\emptyset$ for all $m,m'\in M$ with $m\neq m'$.

Take $g_0\in B(p,L/2)$ and for each $m\in M$ define $\theta=\{\theta_m\}_{m\in M}$ where each $\theta_m\in\{0,1\}$ and define

$$h_\theta=g_0+c_0 2^{-j(p+1/2)}\sum_{m\in M}\theta_m\psi_{j,m}$$

for each θ, where c_0 is a positive constant to be defined subsequently. Note that this gives $2^{(\#(M))}$ such choices of h_θ. By the equivalence $\|\cdot\|_{B^p_{2,2}}\asymp\|\cdot\|_{b^p_{2,2}}$, for each θ we have:

$$\begin{aligned}\|h_\theta\|_{B^p_{2,2}} &\leq L/2+\left\|c_0 2^{-j(p+1/2)}\sum_{m\in M}\theta_m\psi_{j,m}\right\|_{B^p_{2,2}}\\ &\leq L/2+\mathrm{const}\times\left\|c_0 2^{-j(p+1/2)}\sum_{m\in M}\theta_m\psi_{j,m}\right\|_{b^p_{2,2}}\\ &= L/2+\mathrm{const}\times c_0 2^{-j(p+1/2)}\left(\sum_{m\in M}\theta_m^2 2^{2jp}\right)^{1/2}\\ &\leq L/2+\mathrm{const}\times c_0\,.\end{aligned}$$

Therefore, we can choose c_0 sufficiently small that $h_\theta\in B_2(p,L)$ for each θ.

Since $\psi_{j,m}\in C^\gamma$ with $\gamma>|\alpha|$ is compactly supported and X_i has density bounded away from 0 and ∞, we have $\|2^{j/2}\psi^{(|\alpha|)}(2^jx-m)\|_{L^2(X)}\asymp 1$ (uniformly in m). By this and the disjoint support of the $\psi_{j,m}$, for each θ,θ' we have:

$$\begin{aligned}\|\partial^\alpha h_\theta-\partial^\alpha h_{\theta'}\|_{L^2(X)} &= c_0 2^{-j(p-|\alpha|+1/2)}\left(\sum_{m\in M}(\theta_m-\theta'_m)^2\|2^{j/2}\psi^{(|\alpha|)}(2^j\,\cdot\,-m)\|^2_{L^2(X)}\right)^{1/2}\\ &\gtrsim c_0 2^{-j(p-|\alpha|+1/2)}\sqrt{\rho(\theta,\theta')}\end{aligned}$$

where $\rho(\theta,\theta')$ is the Hamming distance between θ and θ'. Take j large enough that $\#(M)\geq 8$. By the Varshamov-Gilbert bound (Tsybakov, 2009, Lemma 2.9) we may choose a subset $\theta^{(0)},\theta^{(1)},\dots,\theta^{(M^*)}$ such that $\theta^0=(0,\dots,0)$, $\rho(\theta^{(a)},\theta^{(b)})\geq\#(M)/8\gtrsim 2^j$ for all $0\leq a<b\leq M^*$ and $M^*\geq 2^{\#(M)/8}$ (recall that there were

$2^{(\#(M))}$ distinct vectors θ and $\#(M) \asymp 2^j$). For each $m \in \{0, 1, \ldots, m^*\}$ let $h_m := h_{\theta^{(m)}}$. Then we have

$$\|\partial^\alpha h_m - \partial^\alpha h_{m'}\|_{L^2(X)} \gtrsim c_0 2^{-j(p-|\alpha|)}$$

for each $0 \leq m < m' \leq M^*$.

For each $0 \leq m \leq M^*$, let P_m denote the joint distribution of $\{(W_i, Y_i)\}_{i=1}^n$ with $Y_i = Th_m(W_i) + u_i$ for the Gaussian NPIR model (45). It follows from Condition LB(ii)(iii) that for each $1 \leq m \leq M^*$ the KL distance $K(P_m, P_0)$ is

$$\begin{aligned} K(P_m, P_0) &\leq \frac{1}{2}\sum_{i=1}^n (c_0 2^{-j(p+\frac{1}{2})})^2 E\left[\frac{(T\sum_{k\in M}\theta_k^{(m)}(\psi_{j,k}(W_i)))^2}{\sigma^2(W_i)}\right] \\ &\leq \frac{n}{2\underline{\sigma}^2}(c_0 2^{-j(p+\frac{1}{2})})^2 \nu(2^j)^2 \sum_{k\in M}(\theta_k^{(m)})^2 \|\psi_{j,k}\|^2_{L^2(X)} \\ &\lesssim n c_0^2 2^{-2jp} \nu(2^j)^2 \end{aligned}$$

where the final line is because X_i has density bounded away from 0 and ∞ and $\sum_{k\in M}(\theta_k^{(m)})^2 \leq \#(M) \asymp 2^j$ for each $1 \leq m \leq M^*$.

In the mildly ill-posed case ($\nu(2^j) = 2^{-j\varsigma}$) we choose $2^j \asymp n^{1/(2(p+\varsigma)+1)}$. This yields:

$$\begin{aligned} K(P_m, P_0) &\lesssim c_0^2 n^{\frac{1}{2(p+\varsigma)+1}} \quad \text{uniformly in } m \\ \log(M^*) &\gtrsim 2^j \asymp n^{1/(2(p+\varsigma)+1)}. \end{aligned}$$

since $M^* \geq 2^{\#(M)/8}$ and $\#(M) \asymp 2^j$.

In the severely ill-posed case ($\nu(2^j) = \exp(-\frac{1}{2}2^{j\varsigma})$) we choose $2^j = (c_1 \log n)^{1/\varsigma}$ with $c_1 > 1$. This yields:

$$\begin{aligned} K(P_m, P_0) &\lesssim c_0^2 n^{-(c_1-1)} \quad \text{uniformly in } m \\ \log(M^*) &\gtrsim 2^j \asymp (\log n)^{1/\varsigma}. \end{aligned}$$

In both the mildly and severely ill-posed cases, the result follows by choosing c_0 sufficiently small that both $\|h_m\|_{B^p_{2,2}} \leq L$ and $K(P_m, P_0) < \frac{1}{8}\log(M^*)$ hold uniformly in m for all n sufficiently large. All conditions of Theorem 2.5 of Tsybakov (2009) are satisfied and hence we obtain the lower bound result.

In the multivariate case ($d > 1$) we let $\widetilde{\psi}_{j,k,G}(x)$ denote an orthonormal tensor-product wavelet for $L^2([0,1]^d)$ at resolution level j. We construct a family of submodels analogously to the univariate case, setting $h_\theta = g_0 + c_0 2^{-j(p+d/2)}\sum_{m\in M}\theta_m \widetilde{\psi}_{j,m,G}$ where $\psi_{j,m}$ is now the product of d interior univariate wavelets at resolution level j with $G = (w_\psi)^d$ (see Appendix E) and where $\#(M) \asymp 2^{jd}$. We then use the Varshamov-Gilbert bound to reduce this to a family of models h_m with $0 \leq m \leq M^*$ and $M^* \asymp 2^{jd}$. We then have:

$$\|\partial^\alpha h_m - \partial^\alpha h_{m'}\|_\infty \gtrsim c_0 2^{-j(p-|\alpha|)}$$

for each $0 \leq m < m' \leq M^*$, and

$$K(P_m, P_0) \lesssim n(c_0 2^{-j(p+d/2)})^2 \nu(2^j)^2$$

for each $1 \leq m \leq M^*$, where $\nu(2^j) = 2^{-j\varsigma}$ in the mildly ill-posed case and $\nu(2^j) \asymp \exp(-2^{j\varsigma})$ in the severely ill-posed case. We choose $2^j \asymp n^{1/(2(p+\varsigma)+d)}$ in the mildly ill-posed case and $2^j = (c_1 \log n)^{1/\varsigma}$ in the severely ill-posed case. The result follows as in the univariate case. $\square$

G.7 Proofs for Appendix C

Proof of Theorem C.1. As in the proof of Theorem 3.2, this follows from the lower bound for NPIR in Theorem G.4. $\square$

The following is a slightly stronger "in probability" version of Lemma 1 in Yu (1997), which is used to prove Theorem G.4. Let $\mathcal{P}$ be a family of probability measures, let $\theta(P)$ be a parameter with values in a pseudo-metric space $(\mathcal{D}, d)$ for some distribution $P \in \mathcal{P}$, and let $\hat{\theta}(P)$ be an estimator of $\theta(P)$ taking values in $(\mathcal{D}, d)$. If $\theta \in \mathcal{D}$ and $D \subset \mathcal{D}$, we let $d(\theta, D) = \inf_{\theta' \in D} d(\theta, \theta')$. Let $co(\mathcal{P})$ denote the convex hull of a set of measures $\mathcal{P}$. Finally, if $\mathbb{P}, \mathbb{Q} \in \mathcal{P}$ we let $\|\mathbb{P} - \mathbb{Q}\|_{TV}$ denote the total variation distance and $\mathrm{aff}(\mathbb{P}, \mathbb{Q}) = 1 - \|\mathbb{P} - \mathbb{Q}\|_{TV}$ denote the affinity between $\mathbb{P}$ and $\mathbb{Q}$.

Lemma G.7. *Suppose there are subsets $D_1, D_2 \subset \mathcal{D}$ that are 2δ separated for some $\delta > 0$ (i.e. $d(s_1, s_2) \geq 2\delta$ for all $s_1 \in D_1$ and $s_2 \in D_2$) and subsets $\mathbb{P}_1, \mathbb{P}_2 \subset \mathcal{P}$ for which $\theta(\mathbb{P}) \in D_1$ for all $\mathbb{P} \in \mathcal{P}_1$ and $\theta(\mathbb{P}) \in D_2$ for all $\mathbb{P} \in \mathcal{P}_2$. Then:*

$$2 \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{P}(d(\widehat{\theta}, \theta(\mathbb{P})) \geq \delta) \geq \sup_{\mathbb{P}_1 \in co(\mathcal{P}_1), \mathbb{P}_2 \in co(\mathcal{P}_2)} \mathrm{aff}(\mathbb{P}_1, \mathbb{P}_2)\,.$$

Proof of Lemma G.7. We proceed as in the proof of Lemma 1 in Yu (1997). Let $P_1 \in \mathcal{P}_1$ and $P_2 \in \mathcal{P}_2$. Then:

$$\begin{aligned} 2 \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{P}(d(\widehat{\theta}, \theta(\mathbb{P})) \geq \delta) &\geq \mathbb{P}_1(d(\widehat{\theta}, \theta(\mathbb{P}_1)) \geq \delta) + \mathbb{P}_2(d(\widehat{\theta}, \theta(\mathbb{P}_2)) \geq \delta) \\ &\geq \mathbb{P}_1(d(\widehat{\theta}, D_1) \geq \delta) + \mathbb{P}_2(d(\widehat{\theta}, D_2) \geq \delta)\,. \end{aligned}$$

Since the inequality $2 \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{P}(d(\widehat{\theta}, \theta(\mathbb{P})) \geq \delta) \geq \mathbb{P}_1(d(\widehat{\theta}, D_1) \geq \delta) + \mathbb{P}_2(d(\widehat{\theta}, D_2) \geq \delta)$ holds for any fixed $\mathbb{P}_1 \in \mathcal{P}_1$ and $\mathbb{P}_2 \in \mathcal{P}_2$, it must also hold for any $\mathbb{P}_1 \in co(\mathcal{P}_1)$ and $\mathbb{P}_2 \in co(\mathcal{P}_2)$. Also note that

$$\begin{aligned} 1\!\!1\{d(\widehat{\theta}, D_1) \geq \delta\} + 1\!\!1\{d(\widehat{\theta}, D_2) \geq \delta\} &\geq 1\!\!1\{d(\widehat{\theta}, D_1) + d(\widehat{\theta}, D_2) \geq 2\delta\} \\ &\geq 1\!\!1\{d(D_1, D_2) \geq 2\delta\} \quad = \quad 1 \end{aligned}$$

because $d(D_1, D_2) \geq 2\delta$. Now by definition of $\alpha(\cdot, \cdot)$, for any $\mathbb{P}_1 \in co(\mathcal{P}_1)$ and $\mathbb{P}_2 \in co(\mathcal{P}_2)$ we have:

$$\begin{aligned} 2 \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{P}(d(\widehat{\theta}, \theta(\mathbb{P})) \geq \delta) &\geq \mathbb{P}_1(d(\widehat{\theta}, D_1) \geq \delta) + \mathbb{P}_2(d(\widehat{\theta}, D_2) \geq \delta) \\ &\geq \inf\{\mathbb{P}_1 f + \mathbb{P}_2 g : f, g \text{ non negative and measurable with } f + g \geq 1\} \\ &= \mathrm{aff}(\mathbb{P}_1, \mathbb{P}_2)\,. \end{aligned}$$

The result follows by taking the supremum of the right-hand side over $\mathbb{P}_1$ and $\mathbb{P}_2$. $\square$

Theorem G.4. *Let Condition LB hold with $B_2(p, L)$ in place of $B_\infty(p, L)$ for the NPIR model (45) with a random sample $\{(W_i, Y_i)\}_{i=1}^n$. Then for any $0 \leq |\alpha| < p$:*

$$\liminf_{n \to \infty} \inf_{\widehat{g}_n} \sup_{h \in B_2(p, L)} \mathbb{P}_h \left(|\widehat{g}_n - f(h)| > c r_n \right) \geq c' > 0$$

where

$$r_n = \begin{cases} n^{-1/2} & \textit{in the mildly ill-posed case when } p \geq \varsigma + 2|\alpha| + d/4 \\ n^{-4(p-|\alpha|)/(4(p+\varsigma)+d)} & \textit{in the mildly ill-posed case when } \varsigma < p < \varsigma + 2|\alpha| + d/4 \\ (\log n)^{-2(p-|\alpha|)/\varsigma} & \textit{in the severely ill-posed case,} \end{cases}$$

$\inf_{\widehat{g}_n}$ *denotes the infimum over all estimators of* $f(h_0)$ *based on the sample of size* n*,* $\sup_{h \in B_2(p,L)} \mathbb{P}_h$ *denotes the sup over* $h \in B_2(p, L)$ *and distributions* (W_i, u_i) *which satisfy Condition LB with* ν *fixed, and the finite positive constants* c, c' *do not depend on* n*.*

Proof of Theorem G.4. We first prove the result for the scalar ($d = 1$) case, then describe the modifications required in the multivariate case.

Let $\{\phi_{j,k}, \psi_{j,k}\}_{j,k}$ be a CDV wavelet basis of regularity $\gamma > p$ for $L^2([0, 1])$, as described in Appendix E. As in the proof of Theorem G.3, we choose a set $M \subset \{r, \ldots, 2^j - r - 1\}$ of interior wavelets with cardinality $\mathsf{m} := \#(M) \asymp 2^j$ such that $\mathrm{support}(\psi_{j,m}) \cap \mathrm{support}(\psi_{j,m'}) = \emptyset$ for all $m, m' \in M$ with $m \neq m'$. Let $\theta = \{\theta_m\}_{m \in M}$ where each $\theta_m \in \{-1, 1\}$ and for each $\theta \in \{-1, 1\}^{\mathsf{m}}$ define:

$$h_\theta = \sum_{m \in M} \frac{\theta_m c_0 2^{-jp}}{\sqrt{\mathsf{m}}} \psi_{j,m} \,.$$

and let $h_0 = 0$. By the equivalence $\|\cdot\|_{b^p_{2,2}} \asymp \|\cdot\|_{B^p_{2,2}}$, we have:

$$\|h_\theta\|_{B^p_{2,2}} \quad \lesssim \quad \|h_\theta\|_{b^p_{2,2}} \quad = \quad \left(2^{2jp} \sum_{m \in M} \frac{\theta_m^2 c_0^2 2^{-2jp}}{\mathsf{m}} \right)^{1/2} \quad = \quad c_0 \,.$$

Therefore, we may choose c_0 sufficiently small that $h_\theta \in B_2(p, L)$ for all $\theta \in \{-1, 1\}^{\mathsf{m}}$.

Let $\psi^{(|\alpha|)}$ denote the $|\alpha|$th derivative of ψ. By disjoint support of the $\psi_{j,m}(x) = 2^{j/2}\psi(2^j x - m)$, $\mu(x) \geq \underline{\mu} > 0$, and a change of variables, we have:

$$\begin{aligned} |f(h_\theta) - f(h_0)| &= \mathsf{m}^{-1} \sum_{m \in M} \int (c_0 \theta_m 2^{-jp} \psi^{(|\alpha|)}_{j,m}(x))^2 \mu(x) \, \mathrm{d}x \\ &\gtrsim \mathsf{m}^{-1} \sum_{m \in M} \int (c_0 \theta_m 2^{-jp} \psi^{(|\alpha|)}_{j,m}(x))^2 \, \mathrm{d}x \\ &= c_0^2 2^{-2jp} \mathsf{m}^{-1} \sum_{m \in M} \int 2^{(2|\alpha|+1)j} \psi^{(|\alpha|)}(2^j x - m)^2 \, \mathrm{d}x \\ &= c_0^2 2^{-2j(p-|\alpha|)} \int \psi^{(|\alpha|)}(u)^2 \, \mathrm{d}u \quad \gtrsim \quad c_0^2 2^{-2j(p-|\alpha|)} \,. \end{aligned}$$

Therefore, there exists a constant $c_* > 0$ such that

$$|f(h_\theta) - f(h_0)| > 2c_* 2^{-2j(p-|\alpha|)} \tag{64}$$

holds for all for each $\theta \in \{-1, 1\}^{\mathsf{m}}$ whenever j is sufficiently large.

Let P_0 (respectively P_θ) denote the joint distribution of $\{(W_i, Y_i)\}_{i=1}^n$ with $Y_i = Th_0(W_i) + u_i$ (respectively $Y_i = Th_\theta(W_i) + u_i$) for the Gaussian NPIR model (45) where, under Condition LB, we may assume that X_i

and W_i have uniform marginals and that the joint density $f_{XW}(x,w)$ of (X_i, W_i) has wavelet expansion

$$f_{XW}(x,w) = \sum_{k=0}^{2^{r_0}-1} \lambda_{r_0}\varphi_{r_0,k}(x)\varphi_{r_0,k}(w) + \sum_{j=r_0}^{\infty}\sum_{k=0}^{2^j-1} \lambda_j \psi_{j,k}(x)\psi_{j,k}(w)\,.$$

Observe that

$$T\psi_{j,k}(w) = \int \psi_{j,k}(x) f_{XW}(x,w)\,\mathrm{d}x = \lambda_j \psi_{j,k}(w)$$

for each $0 \le k \le 2^j - 1$ and each $j \ge r_0$ (r_0 is fixed) and that $|\lambda_j| \asymp \nu(2^j)$ by Condition LB(iii). Let P^* denote the mixture distribution obtained by assigning weight $2^{-\mathsf{m}}$ to P_θ for each of the 2^{m} realizations of θ. Lemma G.8 yields

$$\|P^* - P_0\|_{TV}^2 \lesssim \frac{n^2 2^{-4jp}\nu(2^j)^4}{\mathsf{m}}\,. \tag{65}$$

In the mildly ill-posed case ($\nu(2^j) = 2^{j\varsigma}$) we have

$$\|P^* - P_0\|_{TV}^2 \lesssim n^2 2^{-j(4(p+\varsigma)+1)}$$

because $\mathsf{m} \asymp 2^j$. Choose $2^j \asymp c n^{2/(4(p+\varsigma)+1)}$ with c sufficiently small so $\|P^* - P_0\|_{TV} \le 1 - \epsilon$ for some $1 > \epsilon > 0$ and all n large enough, whence:

$$\mathrm{aff}(P^*, P_0) \quad = \quad 1 - \|P^* - P_0\|_{TV} \quad \ge \quad \epsilon \tag{66}$$

for all n sufficiently large. It now follows by Lemma G.7 and equations (64) and (66) that for all n sufficiently large, any estimator $\widehat{g}_n$ of $f(h)$ obeys the bound

$$\sup_{h \in B_2(p,L)} \mathbb{P}_h\left(|\widehat{g}_n - f(h)| > c_* 2^{-2j(p-|\alpha|)}\right) \quad \ge \quad \epsilon/2 \tag{67}$$

where $2^{-2j(p-|\alpha|)} \asymp n^{-4(p-|\alpha|)/(4(p+\varsigma)+1)}$. This is slower than $n^{-1/2}$ whenever $p \le \varsigma + 2|\alpha| + 1/4$.

In the severely ill-posed case ($\nu(2^j) = \exp(-\frac{1}{2}2^{\varsigma j})$) we choose $2^j = (c\log n)^{1/\varsigma}$ for some $c \in (0,1)$. This yields $\|P^* - P_0\|_{TV} = o(1)$ by (65) and hence there exists $\epsilon > 0$ such that $\mathrm{aff}(P^*, P_0) \ge \epsilon$ for all n sufficiently large. Then by Lemma G.7 and equation (64), for all n sufficiently large, any estimator $\widetilde{f}_n$ of $f(h)$ obeys the same bound (67) with $2^{-2j(p-|\alpha|)} \asymp (\log n)^{-2(p-|\alpha|)/\varsigma}$.

In the multivariate case ($d > 1$) we let $\widetilde{\psi}_{j,k,G}(x)$ denote an orthonormal tensor-product wavelet for $L^2([0,1]^d)$ at resolution level j, as described in Appendix E. We may choose a subset M of $\{0,\ldots,2^j-1\}^d$ with $\mathsf{m} := \#(M) \asymp 2^{dj}$ for which each $m \in M$ indexes a tensor-product of interior wavelets of the form $2^{j/2}\psi(2^j x_l - m_i)$, which we denote by $\widetilde{\psi}_{j,m}(x)$, such that $\widetilde{\psi}_{j,m}$ and $\widetilde{\psi}_{j,m'}$ have disjoint support for each $m, m' \in M$ with $m \ne m'$. For each $\theta \in \{-1,1\}^{\mathsf{m}}$ we define

$$h_\theta = \sum_{m \in M} \frac{\theta_m c_0 2^{-jp}}{\sqrt{\mathsf{m}}} \widetilde{\psi}_{j,m}(x)$$

with c_0 sufficiently small such that $h_\theta \in B_2(p,L)$ for each θ. Let $h_0(x) = 0$ for all $x \in [0,1]^d$. By disjoint support of the $\widetilde{\psi}_{j,m}$ and a change of variables, we have:

$$\begin{aligned} |f(h_\theta) - f(h_0)| &= c_0^2 2^{-2jp} \mathsf{m}^{-1} \sum_{m \in M} \int \left(\prod_{i=1}^d 2^{(2\alpha_i+1)j} \psi^{(\alpha_i)}(2^j x_i - m_i)^2 \right) \mu(x)\,\mathrm{d}x \\ &\gtrsim c_0^2 2^{-2jp} \mathsf{m}^{-1} \sum_{m \in M} \int \left(\prod_{i=1}^d 2^{(2\alpha_i+1)j} \psi^{(\alpha_i)}(2^j x_i - m_i)^2 \right) \mathrm{d}x \quad \gtrsim \quad c_0^2 2^{-2j(p-|\alpha|)} . \end{aligned}$$

Letting P_0, P_θ, and P^* be defined analogously to in the univariate case, we let X_i and W_i have uniform marginals on $[0,1]^d$ and their joint density $f_{XW}(x,w)$ has wavelet expansion

$$f_{XW}(x,w) = \sum_{j=r_0}^\infty \sum_{G \in G_{j,r_0}} \sum_k \lambda_j \widetilde{\psi}_{j,k,G}(x) \widetilde{\psi}_{j,k,G}(w) \tag{68}$$

with $|\lambda_j| \asymp \nu(2^j)$. Lemma G.8 again yields

$$\|P^* - P_0\|_{TV}^2 \lesssim \frac{n^2 2^{-4jp} \nu(2^j)^4}{\mathsf{m}} .$$

The result follows by choosing $2^j \asymp cn^{2/(4(p+\varsigma)+d)}$ with sufficiently small c in the mildly ill-posed case and $2^j = (c \log n)^{1/\varsigma}$ for some $c \in (0,1)$ in the severely ill-posed case. □

Lemma G.8. *Let the Condition LB hold with $B_2(p,L)$ in place of $B_\infty(p,L)$ for the NPIR model (45), let P^* and P_0 be as described in the proof of Theorem G.4, and let $2^{-jp}\nu(2^j) = o(1)$. Then:*

$$\|P^* - P_0\|_{TV}^2 \lesssim \frac{n^2 2^{-4jp} \nu(2^j)^4}{\mathsf{m}} .$$

Proof of Lemma G.8. We prove the result for the multivariate case. For each $\theta \in \{-1,1\}^{\mathsf{m}}$, the density of P_θ with respect to P_0 is

$$\begin{aligned} \frac{\mathrm{d}P_\theta}{\mathrm{d}P_0} &= \prod_{i=1}^n \exp\left\{ \frac{-1}{2\sigma_0^2} \left([T(h_\theta - h_0)(W_i)]^2 - 2u_i [T(h_\theta - h_0)(W_i)] \right) \right\} \\ &= \prod_{i=1}^n \exp\left\{ \frac{-1}{2\sigma_0^2} \left(\left[\sum_{m \in M} \frac{\theta_m c_0 2^{-jp} \lambda_j}{\sqrt{\mathsf{m}}} \widetilde{\psi}_{j,m}(W_i) \right]^2 - 2u_i \left[\sum_{m \in M} \frac{\theta_m c_0 2^{-jp} \lambda_j}{\sqrt{\mathsf{m}}} \widetilde{\psi}_{j,m}(W_i) \right] \right) \right\} . \end{aligned}$$

Since the $\widetilde{\psi}_{j,m}$ have disjoint support, we have

$$\begin{aligned} \frac{\mathrm{d}P_\theta}{\mathrm{d}P_0} &= \prod_{i=1}^n \exp\left\{ -\frac{1}{2} \sum_{m \in M} \frac{c_0^2 2^{-2jp} \lambda_j^2}{\sigma_0^2 \mathsf{m}} \widetilde{\psi}_{j,m}(W_i)^2 + \frac{u_i}{\sigma_0} \sum_{m \in M} \theta_m \frac{c_0 2^{-jp} \lambda_j}{\sigma_0 \sqrt{\mathsf{m}}} \widetilde{\psi}_{j,m}(W_i) \right\} \\ &= \prod_{i=1}^n \exp\left\{ \sum_{m \in M} \left(-\frac{1}{2} \Delta_{i,j,m}^2 + \theta_m \frac{u_i}{\sigma_0} \Delta_{i,j,m} \right) \right\} \end{aligned}$$

where $u_i \sim N(0, \sigma_0^2)$ under P_0 and

$$\Delta_{i,j,m} = \frac{c_0 2^{-jp} \lambda_j}{\sigma_0 \sqrt{\mathsf{m}}} \widetilde{\psi}_{j,m}(W_i) .$$

Therefore:

$$\frac{\mathrm{d}P_\theta}{\mathrm{d}P_0} = \prod_{i=1}^n (1 + A_{i,j}(\theta))$$

where

$$A_{i,j}(\theta) = \exp\left\{\sum_{m\in M}\left(-\frac{1}{2}\Delta^2_{i,j,m} + \theta_m\frac{u_i}{\sigma_0}\Delta_{i,j,m}\right)\right\} - 1$$
$$= \left\{\sum_{m\in M}\exp\left(-\frac{1}{2}\Delta^2_{i,j,m} + \theta_m\frac{u_i}{\sigma_0}\Delta_{i,j,m}\right)\right\} - 1$$

and the second line is again by disjoint support of the $\widetilde{\psi}_{j,m}$ (which implies $\Delta_{i,j,m}$ is nonzero for at most one m for each i).

Let E_0 be expectation under the measure P_0 and observe that $E_0[A_{i,j}(\theta)] = 0$ for each $\theta \in \{-1,1\}^{\mathsf{m}}$. For each θ, θ' we define the vector $\kappa_{\theta,\theta'} \in \mathbb{R}^n$ whose ith element is:

$$\begin{aligned}
\kappa_{\theta,\theta'}(i) &= E_0[A_{i,j}(\theta)A_{i,j}(\theta')] \\
&= \sum_{m\in M}\sum_{m'\in M} E_0\left[\exp\left\{-\frac{1}{2}\Delta^2_{i,j,m} - \frac{1}{2}\Delta^2_{i,j,m'} + \frac{u_i}{\sigma_0}(\theta_m\Delta_{i,j,m} + \theta'_{m'}\Delta_{i,j,m'})\right\} - 1\right] \\
&= \sum_{m\in M}\sum_{m'\in M} E_0\left[e^{\theta_m\theta'_{m'}\Delta_{i,j,m}\Delta_{i,j,m'}} - 1\right] \\
&= \sum_{m\in M} E_0\left[e^{\theta_m\theta'_m\Delta^2_{i,j,m}} - 1\right].
\end{aligned}$$

where the final line is again by disjoint support of the $\widetilde{\psi}_{j,m}$. Using $|\widetilde{\psi}_{j,k}| \lesssim 2^{dj/2}$, $E_0[\widetilde{\psi}_{j,k}(X_i)^2] = 1$, and $\mathsf{m} \asymp 2^{dj}$, it is straightforward to derive the bounds:

$$|\Delta_{i,j,m}| \lesssim \frac{c_0 2^{-jp}\lambda_j}{\sigma_0} \tag{69}$$
$$E_0[\Delta^2_{i,j,m}] = \frac{c_0^2 2^{-2jp}\lambda_j^2}{\sigma_0^2\mathsf{m}}. \tag{70}$$

By Taylor's theorem:

$$\begin{aligned}
\kappa_{\theta,\theta'}(i) &= \sum_{m\in M}\theta_m\theta'_m E_0[\Delta^2_{i,j,m}] + \sum_{m\in M}\frac{1}{2}E_0[\Delta^4_{i,j,m}] + r_3(\theta,\theta') \\
&= \sum_{m\in M}\theta_m\theta'_m E_0[\Delta^2_{i,j,m}] + r_2(\theta,\theta')
\end{aligned}$$

where r_2 and r_3 are remainder terms. Using the Lagrange remainder formula and (69) and (70), we may deduce that

$$r_2(\theta,\theta') \leq \frac{1}{2}\sum_m E_0\left[e^{\Delta^2_{i,j,m}}\Delta^4_{i,j,m}\right] \lesssim \exp\left\{\frac{Cc_0^2 2^{-2jp}\lambda_j^2}{\sigma_0^2}\right\}\frac{c_0^4 2^{-4jp}\lambda_j^4}{\sigma_0^4\mathsf{m}}.$$

and

$$r_3(\theta,\theta') \leq \frac{1}{6}\sum_m E_0\left[e^{\Delta^2_{i,j,m}}\Delta^6_{i,j,m}\right] \lesssim \exp\left\{\frac{Cc_0^2 2^{-2jp}\lambda_j^2}{\sigma_0^2}\right\}\frac{c_0^6 2^{-6jp}\lambda_j^6}{\sigma_0^6\mathsf{m}}.$$

where C is a finite positive constant and we again used the fact that $\Delta_{i,j,m}$ is nonzero for at most one m for each i.

Lemma 22 of Pollard (2000) provides the bound:

$$\|P^* - P_0\|_{TV}^2 \leq 2^{-2\mathsf{m}} \sum_{\theta \in \{-1,1\}^{\mathsf{m}}} \sum_{\theta' \in \{-1,1\}^{\mathsf{m}}} \Upsilon(\kappa_{\theta,\theta'})$$

where for any vector $c = (c_1, \ldots, c_n)' \in \mathbb{R}^n$ the function $\Upsilon(c)$ is defined as

$$\Upsilon(c) = -1 + \prod_{i=1}^{n}(1 + c_i) = \sum_{i=1}^{n} c_i + \sum_{i_1=1}^{n} \sum_{i_2=i_1+1}^{n} c_{i_1} c_{i_2} + \text{higher-order terms}$$

where the higher-order terms are sums over triples, quadruples, etc, with all distinct indices, up to $c_1 c_2 \ldots c_n$. Therefore:

$$\begin{aligned}
\|P^* - P_0\|_{TV}^2 \leq 2^{-2\mathsf{m}} \sum_{\theta,\theta'} \Bigg[& \sum_{i=1}^{n} \left(\sum_{m \in M} \theta_m \theta'_m E_0[\Delta_{i,j,m}^2] + \sum_{m \in M} \frac{1}{2} E_0[\Delta_{i,j,m}^4] + r_3(\theta,\theta') \right) \\
& + \sum_{i_1=1}^{n} \sum_{i_2=i_1+1}^{n} \left\{ \left(\sum_{m \in M} \theta_m \theta'_m E_0[\Delta_{i_1,j,m}^2] + r_2(\theta,\theta') \right) \left(\sum_{m' \in M} \theta_{m'} \theta'_{m'} E_0[\Delta_{i_2,j,m'}^2] + r_2(\theta,\theta') \right) \right\} \\
& + \text{higher order terms} \Bigg]. \qquad (71)
\end{aligned}$$

Since $\sum_{\theta,\theta'} \theta_m \theta'_{m'} = 0$ for all $m, m' \in M$ and $\sum_{\theta,\theta'} 1 = 2^{2\mathsf{m}}$, the first-order sum in (71) is:

$$\begin{aligned}
& 2^{-2\mathsf{m}} \sum_{i=1}^{n} \sum_{\theta,\theta'} \left(\sum_{m} \theta_m \theta'_{m'} E_0[\Delta_{i,j,m}^2] + \frac{1}{2} \sum_{m} E_0[\Delta_{i,j,m}^4] + r_3(\theta,\theta') \right) \\
& = \frac{n}{2} \sum_{m} E_0 \left[\Delta_{i,j,m}^4\right] + n 2^{-2\mathsf{m}} \sum_{\theta,\theta'} r_3(\theta,\theta') \\
& \lesssim \frac{n c_0^4 2^{-4jp} \lambda_j^4}{\sigma_0^4 \mathsf{m}} \left(1 + \exp\left\{ \frac{c_0^2 2^{-2jp} \lambda_j^2}{\sigma_0^2} \right\} \frac{c_0^2 2^{-2jp} \lambda_j^2}{\sigma_0^2 \mathsf{m}} \right). \qquad (72)
\end{aligned}$$

Also observe that

$$\sum_{\theta} \sum_{\theta'} (\theta_m \theta'_m)(\theta_{m'} \theta'_{m'}) = \begin{cases} 0 & \text{if } m \neq m' \\ 2^{2\mathsf{m}} & \text{if } m = m'. \end{cases}$$

The second-order sum in (71) is therefore:

$$\begin{aligned}
& \frac{n(n-1)}{2} \left(\mathsf{m} E_0[\Delta_{i,j,m}^2]^2 + O(\mathsf{m} \max_{m \in M} E[\Delta_{i,j,m}^2] \times \max_{\theta,\theta'} r_2(\theta,\theta')) + O(\max_{\theta,\theta'} r_2(\theta,\theta')^2) \right) \\
& \asymp \frac{n^2 c_0^4 2^{-4jp} \lambda_j^4}{\sigma_0^4 \mathsf{m}} \left(1 + O\left(\exp\left\{ \frac{2C c_0^2 2^{-2jp} \lambda_j^2}{\sigma_0^2} \right\} \left(2^{-2jp} \lambda_j^2 + \frac{2^{-4jp} \lambda_j^4}{\mathsf{m}} \right) \right) \right). \qquad (73)
\end{aligned}$$

The higher-order terms in (71) will be of asymptotically smaller order because $2^{-jp}\lambda_j \asymp 2^{-jp}\nu(2^j) = o(1)$. Substituting (72) and (73) into (71) yields:

$$\|P^* - P_0\|_{TV}^2 \quad \lesssim \quad \frac{n^2 2^{-4jp} \lambda_j^4}{\mathsf{m}} (1 + o(1)) \quad \asymp \quad \frac{n^2 2^{-4jp} \nu(2^j)^4}{\mathsf{m}}$$

as required. □

G.8 Proofs for Appendix E

Proof of Lemma E.1. Part (a) is equation (3.4) on p. 141 of DeVore and Lorentz (1993). For part (b), let $v \in \mathbb{R}^J$, let $f_X(x)$ denote the density of X_i and let $\underline{f}_X = \inf_x f_X(x)$ and $\overline{f}_X = \sup_x f_X(x)$. Then for any $v \in \mathbb{R}^J$:

$$\begin{aligned} v'E[\psi^J(X_i)\psi^J(X_i)']v &\geq \underline{f}_X \int_0^1 (\psi^J(x)'v)^2 \,\mathrm{d}x \\ &\geq \underline{f}_X c_1^2 \frac{\min_{-r+1\leq j\leq m}(t_{j+r}-t_r)}{r}\|v\|_{\ell^2}^2 \\ &\geq \underline{f}_X c_1^2 \frac{c_2 J^{-1}}{r}\|v\|_{\ell^2}^2 \end{aligned}$$

for some finite positive constant c_1, where the first inequality is by Assumption 1(i), the second is by Theorem 4.2 (p. 145) of DeVore and Lorentz (1993) with $p = 2$, and the third is by uniform boundedness of the mesh ratio. By the variational characterization of eigenvalues of selfadjoint matrices, we have:

$$\lambda_{\min}(G_\psi) = \min_{v\in\mathbb{R}^J, v\neq 0} \frac{v'E[\psi^J(X_i)\psi^J(X_i)']v}{\|v\|_{\ell^2}^2} \geq \underline{f}_X c_1^2 \frac{c_2 J^{-1}}{r}$$

This establishes the upper bound on $\lambda_{\min}(G_\psi)^{-1}$. The proof of the lower bound for $\lambda_{\max}(G_\psi)^{-1}$ follows analogously by Theorem 4.2 (p. 145) of DeVore and Lorentz (1993) with $p = 2$. Part (c) then follows directly from part (b). □

Proof of Lemma E.2. The ℓ^1 norm of the tensor product of vectors equals the product of the ℓ^1 norms of the factors, whence part (a) follows from Lemma E.1. As $\psi^J(x)$ is formed as the tensor-product of univariate B-splines, each element of $\psi^J(x)$ is of the form $\prod_{l=1}^d \psi_{Ji_l}(x_l)$ where $\psi_{Ji_l}(x_l)$ denotes the i_lth element of the vector of univariate B-splines. Let $v \in \mathbb{R}^J$. We may index the elements of v by the multi-indices $i_1,\ldots,i_d \in \{1,\ldots,m+r\}^d$. By boundedness of f_X away from zero and Fubini's theorem, we have:

$$\begin{aligned} v'E[\psi^J(X_i)\psi^J(X_i)']v &\geq \underline{f}_X \int_0^1 \cdots \int_0^1 \Big(\sum_{i_1,\ldots,i_l} v_{i_1\cdots i_d} \prod_{l=1}^d \psi_{Ji_l}(x_l)\Big)^2 \mathrm{d}x_1\cdots\mathrm{d}x_d \\ &= \underline{f}_X \int_0^1 \cdots \int_0^1 \sum_{i_2,\ldots,i_l}\sum_{j_2,\ldots,j_l} \left(\prod_{l=2}^d \psi_{Ji_l}(x_l)\right)\left(\prod_{l=2}^d \psi_{Jj_l}(x_l)\right) \\ &\qquad \left\{ \int_0^1 \sum_{i_1}\sum_{j_1} v_{i_1\cdots i_d} v_{j_1\cdots j_d} \psi_{Ji_1}(x_1)\psi_{Jj_1}(x_1)\mathrm{d}x_1 \right\} \mathrm{d}x_2\cdots\mathrm{d}x_d\,. \end{aligned}$$

Applying Theorem 4.2 (p. 145) of DeVore and Lorentz (1993) to the term in braces, and repeating for $x_2,\ldots,x_d$, we have:

$$\begin{aligned} v'E[\psi^J(X_i)\psi^J(X_i)']v &\geq \underline{f}_X \left(c_1^2 \frac{\min_{-r+1\leq j\leq m}(t_{j+r}-t_r)}{r}\right)^d \|v\|_{\ell^2}^2 \\ &\geq \underline{f}_X c_1^{2d} \frac{c_2^d J^{-1}}{r^d}\|v\|_{\ell^2}^2 \end{aligned}$$

where the second inequality is by uniform boundedness of the mesh ratio. The rest of the proof follows by identical arguments to Lemma E.1. □

Proof of Lemma E.3. Each of the interior $\varphi_{j,k}$ and $\psi_{j,k}$ have support $[2^{-j}(-N+1+k), 2^{-j}(N+k)]$, therefore $\varphi_{j,k}(x) \neq 0$ (respectively $\psi_{j,k}(x) \neq 0$) for less than or equal to $2N$ interior $\varphi_{j,k}$ (resp. $\psi_{j,k}$) and for any $x \in [0,1]$. Further, there are only N left and right $\varphi_{j,k}$ and $\psi_{j,k}$. Therefore, $\varphi_{j,k}(x) \neq 0$ (respectively $\psi_{j,k}(x) \neq 0$) for less than or equal to $3N$ of the $\varphi_{j,k}$ (resp. $\psi_{j,k}$) at resolution level j for each $x \in [0,1]$. By construction of the basis, each of φ, $\varphi^l_{j,k}$, $\psi^l_{j,k}$ for $k = 0, \ldots, N-1$ and $\varphi^r_{j,-k}$, $\psi^r_{j,-k}$ for $k = 1, \ldots, N$ are continuous and therefore attain a finite maximum on $[0,1]$. Therefore, each of the $\varphi_{j,k}$ and $\psi_{j,k}$ are uniformly bounded by some multiple of $2^{j/2}$ and so:

$$\xi_{\psi,J} \lesssim 3N \times (\underbrace{2^{L_0/2}}_{\text{for the } \varphi_{L_0,k}} + \underbrace{2^{L_0/2}}_{\text{for the } \psi_{L_0,k}} + \ldots + \underbrace{2^{L/2}}_{\text{for the } \psi_{L,k}}) \lesssim 2^{L/2}$$

The result then follows because $J = 2^{L+1}$. For part (b), because f_X is uniformly bounded away from 0 and ∞ and the wavelet basis is orthonormal for $L^2[0,1]$, we have

$$v' E[\psi^J(X_i)\psi^J(X_i)']v \asymp v' \left(\int_0^1 \psi^J(x)\psi^J(x)' \, \mathrm{d}x \right) v = \|v\|^2_{\ell^2}$$

and so all eigenvalues of G_ψ are uniformly (in J) bounded away from 0 and ∞. Part (c) follows directly. □

Proof of Lemma E.4. Lemma E.3 implies that each of the factor vectors in the tensor product at level j has ℓ^1 norm of order $O(2^{dj/2})$ uniformly for $x = (x_1, \ldots, x_d)' \in [0,1]^d$ and in j. There are at most 2^d such tensor products at each resolution level. Therefore, $\xi_{\psi,J} = O(2^{dL/2}) = O(\sqrt{J})$ since $J = O(2^{dL})$. Parts (b) and (c) follow by the same arguments of the proof of Lemma E.3 since the tensor-product basis is orthonormal for $L^2([0,1]^d)$. □

G.9 Proofs for Appendix F

Proof of Lemma F.2. $\|A^{-1} - I_r\|_{\ell^2} = \|A^{-1}(A - I_r)\|_{\ell^2} \leq \|A^{-1}\|_{\ell^2}\|A - I_r\|_{\ell^2}$. □

Proof of Lemma F.4. The first assertion is immediate by Theorem 3.3 of Stewart (1977) and definition of A^-_l and B^-_l. For the second part, Weyl's inequality implies that $s_{\min}(B) \geq \frac{1}{2} s_{\min}(A)$ whenever $\|A - B\|_{\ell^2} \leq \frac{1}{2} s_{\min}(A)$. □

Proof of Lemma F.5. $\|A^-_l\|^2_{\ell^2} = \lambda_{\max}(A^-_l (A^-_l)') = \lambda_{\max}((A'A)^{-1}) = 1/\lambda_{\min}(A'A) = s_{\min}(A)^{-2}$. □

Proof of Lemma F.6. The result follows from Li, Li, and Cui (2013) (see also Stewart (1977)). □

Proof of Lemma F.7. We prove the results for $\widehat{S}^o$; convergence of $\widehat{G}^o_\psi$ and $\widehat{G}^o_b$ is proved in Lemma 2.1 of Chen and Christensen (2015). Note that

$$\widehat{S}^o - S^o = \sum_{i=1}^n n^{-1} G_b^{-1/2} \{ b^K(W_i)\psi^J(X_i)' - E[b^K(W_i)\psi^J(X_i)'] \} G_\psi^{-1/2} =: \sum_{i=1}^n \Xi^o_i$$

where $\|\Xi_i\|_{\ell^2} \leq 2n^{-1}\zeta_{b,K}\zeta_{\psi,J}$. Also,

$$\begin{aligned}
\left\|\sum_{i=1}^n E[\Xi_i^o \Xi_i^{o\prime}]\right\|_{\ell^2} &\leq n^{-1}\|E[G_b^{-1/2}b^K(W_i)\psi^J(X_i)'G_\psi^{-1}\psi^J(X_i)b^K(W_i)'G_b^{-1/2}]\|_{\ell^2} \\
&\leq n^{-1}\zeta_{\psi,J}^2\|E[G_b^{-1/2}b^K(W_i)b^K(W_i)'G_b^{-1/2}]\|_{\ell^2} \\
&= n^{-1}\zeta_{\psi,J}^2\|I\|_{\ell^2} \\
&= n^{-1}\zeta_{\psi,J}^2
\end{aligned}$$

by the fact that $\|I\|_{\ell^2} = 1$. An identical argument yields the bound $\|\sum_{i=1}^n E[\Xi_i^{o\prime}\Xi_i^o]\|_{\ell^2} \leq n^{-1}\zeta_{b,K}^2$. Applying a Bernstein inequality for random matrices (Tropp, 2012, Theorem 1.6) yields

$$\mathbb{P}\left(\|\widehat{S}^o - S^o\|_{\ell^2} > t\right) \leq 2\exp\left\{\log K - \frac{-t^2/2}{(\zeta_{b,K}^2 \vee \zeta_{\psi,J}^2)/n + 2\zeta_{b,K}\zeta_{\psi,J}t/(3n)}\right\}.$$

The convergence rate $\|\widehat{S}^o - S^o\|_{\ell^2}$ from this inequality under appropriate choice of t. $\square$

Proof of Lemma F.9. Let $\widetilde{b}^K(x) = G_b^{-1/2}b^K(x)$ and denote $\widetilde{b}^K(x)' = (\widetilde{b}_{K1}(x), \ldots, \widetilde{b}_{KK}(x))$. As the summands have expectation zero, we have

$$\begin{aligned}
E\left[\|G_b^{-1/2}\{B'(H_0 - H_J)/n - E[b^K(W_i)(h_0(X_i) - h_J(X_i))]\}\|_{\ell^2}^2\right] &\leq \frac{1}{n}E\left[\sum_{k=1}^K (\widetilde{b}_{Kk}(W_i))^2(h_0(X_i) - h_J(X_i))^2\right] \\
&\leq \frac{K}{n}\|h_0 - h_J\|_\infty^2 \wedge \frac{\zeta_{b,K}^2}{n}\|h_0 - h_J\|_{L^2(X)}^2.
\end{aligned}$$

The result follows by Chebyshev's inequality. $\square$

Proof of Lemma F.10. We begin by rewriting the target in terms of the orthonormalized matrices

$$\begin{aligned}
(\widehat{G}_b^{-1/2}\widehat{S})_l^- \widehat{G}_b^{-1/2}G_b^{1/2} - (G_b^{-1/2}S)_l^- &= G_\psi^{-1/2}\{(\widehat{S}^{o\prime}\widehat{G}_b^{o-1}\widehat{S}^o)^{-1}\widehat{S}^{o\prime}\widehat{G}_b^{o-1} - (S^{o\prime}S^o)^{-1}S^{o\prime}\} \\
&= G_\psi^{-1/2}\{((\widehat{G}_b^o)^{-1/2}\widehat{S}^o)_l^-(\widehat{G}_b^o)^{-1/2} - (S^o)_l^-\}. \qquad (74)
\end{aligned}$$

We first bound the term in braces. By the triangle inequality,

$$\begin{aligned}
&\|((\widehat{G}_b^o)^{-1/2}\widehat{S}^o)_l^-(\widehat{G}_b^o)^{-1/2} - (S^o)_l^-\|_{\ell^2} \\
&\leq \|((\widehat{G}_b^o)^{-1/2}\widehat{S}^o)_l^- - (S^o)_l^-\|_{\ell^2}\|(\widehat{G}_b^o)^{-1/2}\|_{\ell^2} + \|(\widehat{G}_b^o)^{-1/2} - I\|_{\ell^2}\|(S^o)_l^-\|_{\ell^2}. \qquad (75)
\end{aligned}$$

Lemma F.7 provides that

$$\begin{aligned}
\|\widehat{G}_b^o - I_K\|_{\ell^2} &= O_p(\zeta_{b,K}\sqrt{(\log K)/n}) \qquad (76) \\
\|\widehat{S}^o - S^o\|_{\ell^2} &= O_p((\zeta_{b,K} \vee \zeta_{\psi,J})\sqrt{(\log K)/n}). \qquad (77)
\end{aligned}$$

Let $\mathcal{A}_n$ denote the event upon which $\|\widehat{G}_b^o - I_K\|_{\ell^2} \leq \frac{1}{2}$ and note that $\mathbb{P}(\mathcal{A}_n^c) = o(1)$ because $\|\widehat{G}_b^o - I_K\|_{\ell^2} = o_p(1)$. Then by Lemmas F.2 and F.3 we have

$$\begin{aligned}
\|(\widehat{G}_b^o)^{-1/2} - I_K\|_{\ell^2} &\leq \sqrt{2}\|(\widehat{G}_b^o)^{1/2} - I_K\|_{\ell^2} \\
&\leq \frac{2}{1+\sqrt{2}}\|\widehat{G}_b^o - I_K\|_{\ell^2}
\end{aligned}$$

on $\mathcal{A}_n$. It follows by expression (76) and the fact that $\mathbb{P}(\mathcal{A}_n^c) = o(1)$ that

$$\|(\widehat{G}_b^o)^{-1/2} - I\|_{\ell^2} = O_p(\zeta_{b,K}\sqrt{(\log K)/n}) \tag{78}$$

which in turn implies that $\|(\widehat{G}_b^o)^{-1/2}\| = 1 + o_p(1)$.

To bound $\|((\widehat{G}_b^o)^{-1/2}\widehat{S}^o)_l^- - (S^o)_l^-\|_{\ell^2}$, it follows by equations (77) and (78) and the fact that $\|S^o\|_{\ell^2} \le 1$ that:

$$\begin{aligned}
\|(\widehat{G}_b^o)^{-1/2}\widehat{S}^o - S^o\|_{\ell^2} &\le \|(\widehat{G}_b^o)^{-1/2} - I_K\|_{\ell^2}\|\widehat{S}^o\|_{\ell^2} + \|\widehat{S}^o - S^o\|_{\ell^2} \\
&= O_p((\zeta_{b,K} \vee \zeta_{\psi,J})\sqrt{(\log K)/n}) \,.
\end{aligned} \tag{79}$$

Let $\mathcal{A}_{n,1} \subseteq \mathcal{A}_n$ denote the event on which $\|(\widehat{G}_b^o)^{-1/2}\widehat{S}^o - S^o\|_{\ell^2} \le \frac{1}{2}s_{JK}$ and note that $\mathbb{P}(\mathcal{A}_{n,1}^c) = o(1)$ by virtue of the condition $s_{JK}^{-1}(\zeta_{b,K} \vee \zeta_{\psi,J})\sqrt{(\log K)/n} = o(1)$. Lemma F.4 provides that

$$\|((\widehat{G}_b^o)^{-1/2}\widehat{S}^o)_l^- - (S^o)_l^-\|_{\ell^2} \le 2(1+\sqrt{5})s_{JK}^{-2}\|(\widehat{G}_b^o)^{-1/2}\widehat{S}^o - S^o\|_{\ell^2} \tag{80}$$

on $\mathcal{A}_{n,1}$, and so

$$\|((\widehat{G}_b^o)^{-1/2}\widehat{S}^o)_l^- - (S^o)_l^-\|_{\ell^2} = O_p(s_{JK}^{-2}(\zeta_{b,K} \vee \zeta_{\psi,J})\sqrt{(\log K)/n}) \tag{81}$$

by (79) and (80). It follows from equations (81) and (74) that:

$$\|(\widehat{G}_b^{-1/2}\widehat{S})_l^-\widehat{G}_b^{-1/2}G_b^{1/2} - (G_b^{-1/2}S)_l^-\|_{\ell^2} = O_p\Big(s_{JK}^{-2}(\zeta_{b,K} \vee \zeta_{\psi,J})\sqrt{(\log K)/(ne_J)}\Big)$$

which, together with the condition $J \le K = O(J)$, proves part (a). Part (b) follows similarly.

For part (c), we pre and post multiply terms in the product by $G_b^{-1/2}$ and $G_\psi^{-1/2}$ to obtain:

$$\begin{aligned}
&\|G_b^{-1/2}S\{(\widehat{G}_b^{-1/2}\widehat{S})_l^-\widehat{G}_b^{-1/2}G_b^{1/2} - (G_b^{-1/2}S)_l^-\}\|_{\ell^2} \\
&\quad= \|S^o[\widehat{S}^{o\prime}(\widehat{G}_b^o)^-\widehat{S}^o]^-\widehat{S}^{o\prime}(\widehat{G}_b^o)^- - S^o[S^{o\prime}S^o]^{-1}S^{o\prime}\|_{\ell^2} \\
&\quad\le \|S^o[\widehat{S}^{o\prime}(\widehat{G}_b^o)^-\widehat{S}^o]^-\widehat{S}^{o\prime}(\widehat{G}_b^o)^{-1/2}((\widehat{G}_b^o)^{-1/2} - I_K)\|_{\ell^2} \\
&\qquad+\|(S^o - (\widehat{G}_b^o)^{-1/2}\widehat{S}^o)[\widehat{S}^{o\prime}(\widehat{G}_b^o)^-\widehat{S}^o]^-\widehat{S}^{o\prime}(\widehat{G}_b^o)^{-1/2}\|_{\ell^2} \\
&\qquad+\|(\widehat{G}_b^o)^{-1/2}\widehat{S}^o[\widehat{S}^{o\prime}(\widehat{G}_b^o)^-\widehat{S}^o]^-\widehat{S}^{o\prime}(\widehat{G}_b^o)^{-1/2} - S^o[S^{o\prime}S^o]^{-1}S^{o\prime}\|_{\ell^2} \,.
\end{aligned} \tag{82}$$

Note that $\|((\widehat{G}_b^o)^{-1/2}\widehat{S}^o)_l^-\|_{\ell^2} \le 2s_{JK}^{-1}$ on $\mathcal{A}_{n,1}$ by Lemma F.5, so

$$\|((\widehat{G}_b^o)^{-1/2}\widehat{S}^o)_l^-\|_{\ell^2} = O_p(s_{JK}^{-1}) \,. \tag{83}$$

It follows by substituting (78), (79), and (83) into (82) that

$$\begin{aligned}
&\|G_b^{-1/2}S\{(\widehat{G}_b^{-1/2}\widehat{S})_l^-\widehat{G}_b^{-1/2}G_b^{1/2} - (G_b^{-1/2}S)_l^-\}\|_{\ell^2} \\
&\quad\le O_p(s_{JK}^{-1}(\zeta_{b,K} \vee \zeta_{\psi,J})\sqrt{(\log K)/n}) \\
&\qquad+\|(\widehat{G}_b^o)^{-1/2}\widehat{S}^o[\widehat{S}^{o\prime}(\widehat{G}_b^o)^-\widehat{S}^o]^-\widehat{S}^{o\prime}(\widehat{G}_b^o)^{-1/2} - S^o[S^{o\prime}S^o]^{-1}S^{o\prime}\|_{\ell^2} \,.
\end{aligned} \tag{84}$$

The remaining term on the right-hand side of (84) is the ℓ^2 norm of the difference between the orthogonal projection matrices associated with S^o and $(\widehat{G}_b^o)^{-1/2}\widehat{S}^o$. Applying Lemma F.6, we obtain:

$$\|(\widehat{G}_b^o)^{-1/2}\widehat{S}^o[\widehat{S}^{o\prime}(\widehat{G}_b^o)^-\widehat{S}^o]^-\widehat{S}^{o\prime}(\widehat{G}_b^o)^{-1/2} - S^o[S^{o\prime}S^o]^{-1}S^{o\prime}\|_{\ell^2} \le 2s_{JK}^{-1}\|(\widehat{G}_b^o)^{-1/2}\widehat{S}^o - S^o\|_{\ell^2}$$

on $\mathcal{A}_{n,1}$. Result (c) then follows by (79) and (84). □

References

Chen, X. and T. M. Christensen (2015). Optimal uniform convergence rates and asymptotic normality for series estimators under weak dependence and weak conditions. *Journal of Econometrics 188*(2), 447 – 465.

Chen, X. and D. Pouzo (2015). Sieve Wald and QLR inferences on semi/nonparametric conditional moment models. *Econometrica 83*(3), 1013–1079.

Chen, X. and M. Reiss (2011). On rate optimality for ill-posed inverse problems in econometrics. *Econometric Theory 27*(3), 497–521.

Chernozhukov, V., D. Chetverikov, and K. Kato (2014). Anti-concentration and honest, adaptive confidence bands. *The Annals of Statistics 42*(5), 1787–1818.

Cohen, A. (2003). *Numerical Analysis of Wavelet Methods*. Elsevier.

DeVore, R. A. and G. G. Lorentz (1993). *Constructive Approximation*. Grundlehren der mathematischen Wissenschaften 303. Springer-Verlag, Berlin.

Hausman, J. A. and W. K. Newey (1995). Nonparametric estimation of exact consumers surplus and deadweight loss. *Econometrica 63*(6), pp. 1445–1476.

Huang, J. Z. (2003). Local asymptotics for polynomial spline regression. *The Annals of Statistics 31*(5), 1600–1635.

Li, B., W. Li, and L. Cui (2013). New bounds for perturbation of the orthogonal projection. *Calcolo 50*(1), 69–78.

Pollard, D. (2000). *Asymptopia*, Chapter 3: Distances and affinities between measures. Unpublished Manuscript.

Pollard, D. (2002). *A User's Guide to Measure Theoretic Probability*. Cambridge University Press, Cambridge.

Schumaker, L. L. (2007). *Spline Functions: Basic Theory*. Cambridge University Press, Cambridge.

Stewart, G. W. (1977). On the perturbation of pseudo-inverses, projections and linear least squares problems. *SIAM Review 19*(4), 634–662.

Tropp, J. A. (2012). User-friendly tail bounds for sums of random matrices. *Foundations of Computational Mathematics 12*, 389–434.

Tsybakov, A. B. (2009). *Introduction to Nonparametric Estimation*. Springer, New York.

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

Yu, B. (1997). Assouad, Fano, and Le Cam. In D. Pollard, E. Torgersen, and G. L. Yang (Eds.), *Festschrift for Lucien Le Cam*. Springer-Verlag New York.