Supplementary Material

Supplement to "Optimal sup-norm rates and uniform inference on nonlinear functionals of nonparametric IV regression"

(*Quantitative Economics*, Vol. 9, No. 1, March 2018, 39–84)

Xiaohong Chen
Cowles Foundation for Research in Economics, Yale University

Timothy M. Christensen
Department of Economics, New York University

This supplementary appendix contains material to support our paper. 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 supplementary appendix contains additional technical lemmas and all of the proofs (Appendix G).

Appendix 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 $Df(h)[v] = \lim_{\delta \to 0^+}[\delta^{-1} \times \{f(h + \delta v) - f(h)\}]$ 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

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

Xiaohong Chen: xiaohong.chen@yale.edu
Timothy M. Christensen: timothy.christensen@nyu.edu

https://doi.org/10.3982/QE722

and let

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

denote its weak norm. Chen and Pouzo (2015) called the functional $f(\cdot)$ 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

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

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

Recall that $\Omega = E[u_i^2 b^K(W_i)b^K(W_i)']$, that 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 that the sieve variance for $f(\widehat{h})$ is

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

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

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

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

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

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

Assumption 5′. *Let* η_n *and* η'_n *be sequences of nonnegative numbers such that* $\eta_n = o(1)$ *and* $\eta'_n = o(1)$. *Let* $\sigma_n(f) \nearrow +\infty$ *as* $n \to \infty$. *Either* (a) *or* (b) *of the following options holds*:

(a) *The functional* f is a linear functional and $\sqrt{n}(\sigma_n(f))^{-1}|f(\widetilde{h}) - f(h_0)| = O_p(\eta_n)$.

(b) (i) *The functional* $v \mapsto Df(h_0)[v]$ *is a linear functional*; (ii)

$$\left|\sqrt{n}\frac{f(\widehat{h}) - f(h_0)}{\sigma_n(f)} - \sqrt{n}\frac{Df(h_0)[\widehat{h} - \widetilde{h}]}{\sigma_n(f)}\right| = O_p(\eta_n);$$

(iii) $\frac{\|\Pi_K T(\widehat{v}_n(f) - v_n(f))\|_{L^2(W)}}{\sigma_n(f)} = O_p(\eta'_n)$.

Assumption 5′(a) and 5′(b)(i) and (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

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

and $\delta_{V,n} \equiv [\zeta_{b,K}^{(2+\delta)/\delta}\sqrt{(\log K)/n}]^{\delta/(1+\delta)} + \tau_J\zeta\sqrt{(\log J)/n} + \delta_{h,n}$, 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. (i) *Let Assumptions* 1(iii), 2(i), (iii), *and* (iv′), 4(i), *and either* 5′(a) *or* 5′(b)(i) *and* (ii) *hold, and let* $\tau_J\zeta\sqrt{(J\log J)/n} = o(1)$. *Then*

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

(ii) *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*

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

and

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

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).

APPENDIX 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 $0 = t_{-(r-1)} = \cdots = t_0 < t_1 < \cdots < t_m < t_{m+1} = \cdots = t_{m+r} = 1$ denote the extended knot sequence and let $I_0 = [t_0, t_1), \ldots, I_m = [t_m, t_{m+1}]$. 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

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

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 $N_{j,r}$ 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 exist finite positive constants* c_ψ *and* C_ψ *such that* $c_\psi J \le \lambda_{\max}(G_\psi)^{-1} \le \lambda_{\min}(G_\psi)^{-1} \le C_\psi J$ *for all* $J \ge r$; (c) $\lambda_{\max}(G_\psi)/\lambda_{\min}(G_\psi) \le C_\psi/c_\psi$ *for all* $J \ge r$.

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

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 $\varphi_{j,k}(x) = 2^{j/2}\varphi(2^j x - k)$, each of which has support $[2^{-j}(-N+1+k), 2^{-j}(N+k)]$ for $k = N,\dots,2^j - N - 1$. These are augmented with N left scaling functions of the form $\varphi^0_{j,k}(x) = 2^{j/2}\varphi^l_k(2^j x)$ for $k = 0,\dots,N-1$ (where $\varphi^l_0,\dots,\varphi^l_{N-1}$ 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 $k = 1,\dots,N$ (where $\varphi^r_{-1},\dots,\varphi^r_{-N}$ are fixed independent of j), each of which has support $[1-2^{-j}(1-N-k),1]$. The resulting 2^j functions $\varphi^0_{j,0},\dots,\varphi^0_{j,N-1},\varphi_{j,N},\dots,\varphi_{j,2^j-N-1},\varphi^1_{j,2^j-N},\dots,\varphi^1_{j,2^j-1}$ 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 from the mother wavelet. This results in an orthonormal basis of 2^j functions, denoted $\psi^0_{j,0},\dots,\psi^0_{j,N-1},\psi_{j,N},\dots,\psi_{j,2^j-N-1},\psi^1_{j,2^j-N}$, $\dots,\psi^1_{j,2^j-1}$ (we use this conventional notation without confusion with the ψ_{Jj} basis functions spanning Ψ_J), where the "interior" wavelets $\psi_{j,N},\dots,\psi_{j,2^j-N-1}$ are of the form $\psi_{j,k}(x) = 2^{j/2}\psi(2^j x - k)$. To simplify notation, we ignore the 0 and 1 superscripts on the left and right wavelets and the scaling functions henceforth. Let L_0 and L be integers such that $2N < 2^{L_0} \le 2^L$. A wavelet space at resolution level L is the 2^{L+1}-dimensional set of functions given by

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

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 $2N < 2^{L_0} \le 2^L$. 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, $\widetilde{\psi}_{j,k,w_\psi^d} = \prod_{i=1}^d \psi_{j,k_i}(x_i)$. 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 $\#(G_{L_0,L_0}) = 2^d$ and $\#(G_{j,L_0}) = 2^d - 1$ for $j > L_0$. Let $\text{Wav}(L, [0,1]^d, \gamma)$ denote the space

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

where each univariate basis has regularity γ. This definition clearly reduces to the above definition for $\text{Wav}(L, [0,1], \gamma)$ 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 $\psi_{J1}, \ldots, \psi_{JJ}$ 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, and* (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 $\psi_{J1}, \ldots, \psi_{JJ}$ 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, and* (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

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

where 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

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

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.

Appendix 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 the prime ($'$) denotes transpose and the bar ($^-$) 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 $s_i(A)$ and $s_i(B)$ 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*

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

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*

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

F.2 *Convergence of the matrix estimators*

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

$$\widehat{G}_b^o = G_b^{-1/2}\widehat{G}_b G_b^{-1/2},$$
$$\widehat{G}_\psi^o = G_\psi^{-1/2}\widehat{G}_\psi G_\psi^{-1/2},$$
$$\widehat{S}^o = G_b^{-1/2}\widehat{S}G_\psi^{-1/2},$$

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*

$$\mathbb{P}\big(\|\widehat{G}^o_\psi - G^o_\psi\|_{\ell^2} > t\big) \le 2\exp\left\{\log J - \frac{t^2/2}{\zeta^2_{\psi,J}(1+2t/3)/n}\right\},$$

$$\mathbb{P}\big(\|\widehat{G}^o_b - G^o_b\|_{\ell^2} > t\big) \le 2\exp\left\{\log K - \frac{t^2/2}{\zeta^2_{b,K}(1+2t/3)/n}\right\},$$

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

and, therefore,

$$\|\widehat{G}^o_\psi - G^o_\psi\|_{\ell^2} = O_p\big(\zeta_{\psi,J}\sqrt{(\log J)/n}\big),$$
$$\|\widehat{G}^o_b - G^o_b\|_{\ell^2} = O_p\big(\zeta_{b,K}\sqrt{(\log K)/n}\big),$$
$$\|\widehat{S}^o - S^o\|_{\ell^2} = O_p\big((\zeta_{b,K} \vee \zeta_{\psi,J})\sqrt{(\log K)/n}\big)$$

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 let* $H_J = (h_J(X_1), \dots, h_J(X_n))' = \Psi c_J$. *Then*

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

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

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

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

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

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. [3]

Chen, X. and D. Pouzo (2015), "Sieve Wald and QLR inferences on semi/nonparametric conditional moment models." *Econometrica*, 83 (3), 1013–1079. [1, 2, 3]

Cohen, A., I. Daubechies, and P. Vial (1993), "Wavelets on the interval and fast wavelet transforms." *Applied and Computational Harmonic Analysis*, 1, 54–81. [5]

DeVore, R. A. and G. G. Lorentz (1993), *Constructive Approximation*. Grundlehren der mathematischen Wissenschaften, Vol. 303. Springer-Verlag, Berlin. [4]

Johnstone, I. M. (2013), "Gaussian estimation: Sequence and wavelet models." Manuscript. [6]

Newey, W. K. (1997), "Convergence rates and asymptotic normality for series estimators." *Journal of Econometrics*, 79 (1), 147–168. [8]

Schmitt, B. A. (1992), "Perturbation bounds for matrix square roots and Pythagorean sums." *Linear Algebra and its Applications*, 174 (0), 215–227. [7]

Triebel, H. (2006), *Theory of Function Spaces III*. Birkhäuser, Basel. [6]

Co-editor Rosa L. Matzkin handled this manuscript.

Manuscript received 25 May, 2016; final version accepted 22 March, 2017; available online 26 April, 2017.