

SUPPLEMENT TO “ESTIMATION OF NONPARAMETRIC
 CONDITIONAL MOMENT MODELS WITH POSSIBLY
 NONSMOOTH GENERALIZED RESIDUALS”
 (*Econometrica*, Vol. 80, No. 1, January 2012, 277–321)

BY XIAOHONG CHEN AND DEMIAN POUZO

In this supplemental document, we first provide a brief summary of commonly used function spaces and sieve spaces. We then provide mathematical proofs of all the theorems, corollaries, propositions, and lemmas that appear in the main text and the Appendix.

A. BRIEF SUMMARY OF FUNCTION SPACES AND SIEVES

HERE, WE BRIEFLY SUMMARIZE some definitions and properties of function spaces that are used in the main text; see Edmunds and Triebel (1996) for details. Let $\mathcal{S}(\mathcal{R}^d)$ be the Schwartz space of all complex-valued, rapidly decreasing, infinitely differentiable functions on \mathcal{R}^d . Let $\mathcal{S}^*(\mathcal{R}^d)$ be the space of all tempered distributions on \mathcal{R}^d , which is the topological dual of $\mathcal{S}(\mathcal{R}^d)$. For $h \in \mathcal{S}(\mathcal{R}^d)$, we let \widehat{h} denote the Fourier transform of h (i.e., $\widehat{h}(\xi) = (2\pi)^{-d/2} \int_{\mathcal{R}^d} \exp\{-iy'\xi\}h(y) dy$) and let $(g)^\vee$ denote the inverse Fourier transform of g (i.e., $(g)^\vee(y) = (2\pi)^{-d/2} \int_{\mathcal{R}^d} \exp\{iy'\xi\}g(\xi) d\xi$). Let $\varphi_0 \in \mathcal{S}(\mathcal{R}^d)$ be such that $\varphi_0(x) = 1$ if $|x| \leq 1$ and $\varphi_0(x) = 0$ if $|x| \geq 3/2$. Let $\varphi_1(x) = \varphi_0(x/2) - \varphi_0(x)$ and $\varphi_k(x) = \varphi_1(2^{-k+1}x)$ for all integer $k \geq 1$. Then the sequence $\{\varphi_k : k \geq 0\}$ forms a dyadic resolution of unity (i.e., $1 = \sum_{k=0}^\infty \varphi_k(x)$ for all $x \in \mathcal{R}^d$). Let $\nu \in \mathcal{R}$ and $p, q \in (0, \infty]$. The Besov space $\mathcal{B}_{p,q}^\nu(\mathcal{R}^d)$ is the collection of all functions $h \in \mathcal{S}^*(\mathcal{R}^d)$ such that $\|h\|_{\mathcal{B}_{p,q}^\nu}$ is finite:

$$\|h\|_{\mathcal{B}_{p,q}^\nu} \equiv \left(\sum_{j=0}^\infty \{2^{j\nu} \|(\varphi_j \widehat{h})^\vee\|_{L^p(\text{leb})}\}^q \right)^{1/q} < \infty$$

(with the usual modification if $q = \infty$). Let $\nu \in \mathcal{R}$, $p \in (0, \infty)$, and $q \in (0, \infty]$. The F space $\mathcal{F}_{p,q}^\nu(\mathcal{R}^d)$ is the collection of all functions $h \in \mathcal{S}^*(\mathcal{R}^d)$ such that $\|h\|_{\mathcal{F}_{p,q}^\nu}$ is finite:

$$\|h\|_{\mathcal{F}_{p,q}^\nu} \equiv \left\| \left(\sum_{j=0}^\infty \{2^{j\nu} |(\varphi_j \widehat{h})^\vee(\cdot)|\}^q \right)^{1/q} \right\|_{L^p(\text{leb})} < \infty$$

(with the usual modification if $q = \infty$). For $\nu > 0$ and $p, q \geq 1$, it is known that $\mathcal{F}_{p',q'}^{-\nu}(\mathcal{R}^d)$ ($\mathcal{B}_{p',q'}^{-\nu}(\mathcal{R}^d)$) is the dual space of $\mathcal{F}_{p,q}^\nu(\mathcal{R}^d)$ ($\mathcal{B}_{p,q}^\nu(\mathcal{R}^d)$) with $1/p' + 1/p = 1$ and $1/q' + 1/q = 1$.

Let $\mathcal{T}_{p,q}^\nu(\mathcal{R}^d)$ denote either $\mathcal{B}_{p,q}^\nu(\mathcal{R}^d)$ or $\mathcal{F}_{p,q}^\nu(\mathcal{R}^d)$. Then $\mathcal{T}_{p,q}^\nu(\mathcal{R}^d)$ gets larger with increasing q (i.e., $\mathcal{T}_{p,q_1}^\nu(\mathcal{R}^d) \subseteq \mathcal{T}_{p,q_2}^\nu(\mathcal{R}^d)$ for $q_1 \leq q_2$), gets larger with decreasing p (i.e., $\mathcal{T}_{p_1,q}^\nu(\mathcal{R}^d) \subseteq \mathcal{T}_{p_2,q}^\nu(\mathcal{R}^d)$ for $p_1 \geq p_2$), and gets larger with decreasing ν (i.e., $\mathcal{T}_{p,q}^{\nu_1}(\mathcal{R}^d) \subseteq \mathcal{T}_{p,q}^{\nu_2}(\mathcal{R}^d)$ for $\nu_1 \geq \nu_2$). Also, $\mathcal{T}_{p,q}^\nu(\mathcal{R}^d)$ becomes a *Banach* space when $p, q \geq 1$. The spaces $\mathcal{T}_{p,q}^\nu(\mathcal{R}^d)$ include many well known function spaces as special cases. For example, $L^p(\mathcal{R}^d, \text{leb}) = \mathcal{F}_{p,2}^0(\mathcal{R}^d)$ for $p \in (1, \infty)$, the *Hölder* space $\Lambda^r(\mathcal{R}^d) = \mathcal{B}_{\infty,\infty}^r(\mathcal{R}^d)$ for any real-valued $r > 0$, the *Hilbert–Sobolev* space $W_2^k(\mathcal{R}^d) = \mathcal{B}_{2,2}^k(\mathcal{R}^d)$ for integer $k > 0$, and the (*fractional*) *Sobolev* space $W_p^\nu(\mathcal{R}^d) = \mathcal{F}_{p,2}^\nu(\mathcal{R}^d)$ for any $\nu \in \mathcal{R}$ and $p \in (1, \infty)$, which has the equivalent norm $\|h\|_{W_p^\nu} \equiv \|((1 + |\cdot|^2)^{\nu/2} \widehat{h}(\cdot))^\vee\|_{L^p(\text{leb})} < \infty$ (note that for $\nu > 0$, the norm $\|h\|_{W_p^{-\nu}}$ is a shrinkage in the Fourier domain).

Let $\mathcal{T}_{p,q}^\nu(\Omega)$ be the corresponding space on an (arbitrary) bounded domain Ω in \mathcal{R}^d . Then the embedding of $\mathcal{T}_{p_1,q_1}^{\nu_1}(\Omega)$ into $\mathcal{T}_{p_2,q_2}^{\nu_2}(\Omega)$ is compact if $\nu_1 - \nu_2 > d \max\{p_1^{-1} - p_2^{-1}, 0\}$, and $-\infty < \nu_2 < \nu_1 < \infty$, $0 < q_1, q_2 \leq \infty$, and $0 < p_1, p_2 \leq \infty$ ($0 < p_1, p_2 < \infty$ for $\mathcal{F}_{p,q}^\nu(\Omega)$).

We define “weighted” versions of the space $\mathcal{T}_{p,q}^\nu(\mathcal{R}^d)$ as follows. Let $w(\cdot) = (1 + |\cdot|^2)^{\zeta/2}$, $\zeta \in \mathcal{R}$, be a weight function and define $\|h\|_{\mathcal{T}_{p,q}^\nu(\mathcal{R}^d, w)} = \|wh\|_{\mathcal{T}_{p,q}^\nu(\mathcal{R}^d)}$, that is, $\mathcal{T}_{p,q}^\nu(\mathcal{R}^d, w) = \{h : \|wh\|_{\mathcal{T}_{p,q}^\nu(\mathcal{R}^d)} < \infty\}$. Then the embedding of $\mathcal{T}_{p_1,q_1}^{\nu_1}(\mathcal{R}^d, w_1)$ into $\mathcal{T}_{p_2,q_2}^{\nu_2}(\mathcal{R}^d, w_2)$ is compact if and only if $\nu_1 - \nu_2 > d(p_1^{-1} - p_2^{-1})$, $w_2(x)/w_1(x) \rightarrow 0$ as $|x| \rightarrow \infty$, and $-\infty < \nu_2 < \nu_1 < \infty$, $0 < q_1, q_2 \leq \infty$, and $0 < p_1 \leq p_2 \leq \infty$ ($0 < p_1 \leq p_2 < \infty$ for $\mathcal{F}_{p,q}^\nu(\Omega)$).

If $\mathcal{H} \subseteq \mathbf{H}$ is a Besov space, then a *wavelet* basis $\{\psi_j\}$ is a natural choice of $\{q_j\}_j$ to satisfy Assumption 5.1 in Section 5. A real-valued function ψ is called a *mother wavelet* of degree γ if it satisfies (a) $\int_{\mathcal{R}} y^k \psi(y) dy = 0$ for $0 \leq k \leq \gamma$, (b) ψ and all its derivatives up to order γ decrease rapidly as $|y| \rightarrow \infty$ and (c) $\{2^{k/2} \psi(2^k y - j) : k, j \in \mathbb{Z}\}$ forms a Riesz basis of $L^2(\text{leb})$, that is, the linear span of $\{2^{k/2} \psi(2^k y - j) : k, j \in \mathbb{Z}\}$ is dense in $L^2(\text{leb})$ and

$$\left\| \sum_{k=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} a_{kj} 2^{k/2} \psi(2^k y - j) \right\|_{L^2(\mathcal{R})}^2 \asymp \sum_{k=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} |a_{kj}|^2$$

for all doubly biinfinite square-summable sequences $\{a_{kj} : k, j \in \mathbb{Z}\}$. A scaling function φ is called a *father wavelet* of degree γ if it satisfies (a') $\int_{\mathcal{R}} \varphi(y) dy = 1$, (b') φ and all its derivatives up to order γ decrease rapidly as $|y| \rightarrow \infty$, and (c') $\{\varphi(y - j) : j \in \mathbb{Z}\}$ forms a Riesz basis for a closed subspace of $L^2(\text{leb})$.

Some examples of sieves follow:

Orthogonal Wavelets: Given an integer $\gamma > 0$, there exist a father wavelet φ of degree γ and a mother wavelet ψ of degree γ , both compactly supported,

such that for any integer $k_0 \geq 0$, any function h in $L^2(\text{leb})$ has the wavelet γ -regular multiresolution expansion

$$h(y) = \sum_{j=-\infty}^{\infty} a_{k_0 j} \varphi_{k_0 j}(y) + \sum_{k=k_0}^{\infty} \sum_{j=-\infty}^{\infty} b_{kj} \psi_{kj}(y), \quad y \in \mathcal{R},$$

where $\{\varphi_{k_0 j}, j \in \mathbb{Z}; \psi_{kj}, k \geq k_0, j \in \mathbb{Z}\}$ is an orthonormal basis of $L^2(\text{leb})$; see Meyer (1992, Theorem 3.3). For an integer $K_n > k_0$, we consider the finite-dimensional linear space spanned by this wavelet basis of order γ :

$$h_n(y) = \psi^{k_n}(y)' \Pi = \sum_{j=0}^{2^{K_n-1}} \pi_{K_n, j} \varphi_{K_n, j}(y), \quad k(n) = 2^{K_n}.$$

Cardinal B-Spline Wavelets of Order γ :

$$(SM.1) \quad h_n(y) = \psi^{k_n}(y)' \Pi = \sum_{k=0}^{K_n} \sum_{j \in \mathcal{K}_n} \pi_{kj} 2^{k/2} B_\gamma(2^k y - j), \quad k(n) = 2^{K_n} + 1,$$

where $B_\gamma(\cdot)$ is the cardinal B-spline of order γ :

$$B_\gamma(y) = \frac{1}{(\gamma-1)!} \sum_{i=0}^{\gamma} (-1)^i \binom{\gamma}{i} [\max(0, y-i)]^{\gamma-1}.$$

Polynomial Splines of Order q_n :

$$(SM.2) \quad h_n(y) = \psi^{k_n}(y)' \Pi \\ = \sum_{j=0}^{q_n} \pi_j (y)^j + \sum_{k=1}^{r_n} \pi_{q_n+k} (y - \nu_k)_+^{q_n}, \quad k(n) = q_n + r_n + 1,$$

where $(y - \nu)_+^q = \max\{(y - \nu)^q, 0\}$ and $\{\nu_k\}_{k=1, \dots, r_n}$ are the knots. In the empirical application, for any given number of knots value r_n , the knots $\{\nu_k\}_{k=1, \dots, r_n}$ are simply chosen as the empirical quantiles of the data.

Hermite Polynomials of Order $k(n) - 1$:

$$(SM.3) \quad h_n(y) = \psi^{k_n}(y)' \Pi = \sum_{j=0}^{k_n-1} \pi_j (y - \nu_1)^j \exp\left\{-\frac{(y - \nu_1)^2}{2\nu_2^2}\right\},$$

where ν_1 and ν_2^2 can be chosen as the sample mean and variance of the data.

B. CONSISTENCY: PROOFS OF THEOREMS

PROOF OF THEOREM 3.1: Under the assumption that $E[m(X, h)'W(X) \times m(X, h)]$ is lower semicontinuous on finite-dimensional closed and bounded sieve spaces \mathcal{H}_k , we have that for all $\varepsilon > 0$ and each fixed $k \geq 1$,

$$\begin{aligned} g(k, \varepsilon) &\equiv \inf_{h \in \mathcal{H}_k^{M_0} : \|h - h_0\|_s \geq \varepsilon} E[\|m(X, h)\|_W^2] \\ &\geq \min_{h \in \mathcal{H}_k : \|h - h_0\|_s \geq \varepsilon} E[\|m(X, h)\|_W^2] \end{aligned}$$

exists and is strictly positive (under Assumption 3.1(i) and (ii)). Moreover, for fixed k , $g(k, \varepsilon)$ increases as ε increases. For any fixed $\varepsilon > 0$, $g(k, \varepsilon)$ decreases as k increases, and could go to zero as k goes to infinity. Following the proof of Lemma A.4(i) with $\mathcal{T} = \|\cdot\|_s$ topology, $\mathcal{H}_{k(n)}^{M_0} \subseteq \mathcal{H}_{k(n)}$, $\lambda_n P(h) \geq 0$, and $\eta_n = O(\eta_{0,n})$, we have, for all $\varepsilon > 0$ and n sufficiently large,

$$\begin{aligned} &\Pr(\|\widehat{h}_n - h_0\|_s \geq \varepsilon) \\ &\leq \Pr(\|\widehat{h}_n - h_0\|_s \geq \varepsilon, \widehat{h}_n \in \mathcal{H}_{k(n)}^{M_0}) + \varepsilon \\ &\leq \Pr\left(\inf_{h \in \mathcal{H}_{k(n)}^{M_0} : \|h - h_0\|_s \geq \varepsilon} \{cE[\|m(X, h)\|_W^2] + \lambda_n P(h)\}\right) \\ &\leq c'E[\|m(X, \Pi_n h_0)\|_W^2] \\ &\quad + O_p(\eta_{0,n}) + O_p(\bar{\delta}_{m,n}^2) + \lambda_n P(h_0) + O_p(\lambda_n) + \varepsilon \\ &\leq \Pr\left(\inf_{h \in \mathcal{H}_{k(n)}^{M_0} : \|h - h_0\|_s \geq \varepsilon} \{cE[\|m(X, h)\|_W^2]\}\right) \leq c'E[\|m(X, \Pi_n h_0)\|_W^2] \\ &\quad + O_p(\eta_{0,n}) + O_p(\bar{\delta}_{m,n}^2) + \lambda_n P(h_0) + O_p(\lambda_n) + \varepsilon \\ &\leq \Pr(g(k(n), \varepsilon)) \\ &\leq O_p(\max\{\bar{\delta}_{m,n}^2, \eta_{0,n}, E(\|m(X, \Pi_n h_0)\|_W^2), \lambda_n\}) + \varepsilon, \end{aligned}$$

which goes to zero under $\max\{\bar{\delta}_{m,n}^2, \eta_{0,n}, E(\|m(X, \Pi_n h_0)\|_W^2), \lambda_n\} = o(g(k(n), \varepsilon))$. Thus $\|\widehat{h}_n - h_0\|_s = o_p(1)$. *Q.E.D.*

PROOF OF THEOREM 3.2: Under the assumptions that $E[m(X, h)'W(X) \times m(X, h)]$ is lower semicontinuous and $P(h)$ is lower semicompact on $(\mathcal{H}, \|\cdot\|_s)$, we have that for all $\varepsilon > 0$,

$$g(\varepsilon) \equiv \min_{h \in \mathcal{H}^M : \|h - h_0\|_s \geq \varepsilon} E[m(X, h)'W(X)m(X, h)]$$

exists (by Theorem 38.B in Zeidler (1985)) and is strictly positive (under Assumption 3.1(i) and (ii)) for $\mathcal{H}^M = \{h \in \mathcal{H} : P(h) \leq M\}$ with some large but finite $M \geq M_0$. By Lemma A.4(i) with $\mathcal{T} = \|\cdot\|_s$ topology, $\mathcal{H}_{k(n)}^{M_0} \subseteq \mathcal{H}^M$, $\lambda_n > 0$, $P(h) \geq 0$, $\eta_n = O(\eta_{0,n})$, and $\max\{\eta_{0,n}, E[\|m(X, \Pi_n h_0)\|_W^2]\} = O(\lambda_n)$, we have, for all $\varepsilon > 0$ and n sufficiently large,

$$\begin{aligned}
& \Pr(\|\widehat{h}_n - h_0\|_s \geq \varepsilon) \\
& \leq \Pr(\|\widehat{h}_n - h_0\|_s \geq \varepsilon, \widehat{h}_n \in \mathcal{H}_{k(n)}^{M_0}) + \varepsilon \\
& \leq \Pr\left(\inf_{h \in \mathcal{H}_{k(n)}^{M_0} : \|h - h_0\|_s \geq \varepsilon} \{cE[\|m(X, h)\|_W^2] + \lambda_n P(h)\}\right) \\
& \leq O_p(\bar{\delta}_{m,n}^2) + \lambda_n P(h_0) + O_p(\lambda_n) + \varepsilon \\
& \leq \Pr\left(\inf_{h \in \mathcal{H}^M : \|h - h_0\|_s \geq \varepsilon} E[\|m(X, h)\|_W^2]\right) \\
& \leq O_p(\bar{\delta}_{m,n}^2) + \lambda_n P(h_0) + O_p(\lambda_n) + \varepsilon \\
& \leq \Pr(g(\varepsilon) \leq O_p(\max\{\bar{\delta}_{m,n}^2, \lambda_n\})) + \varepsilon,
\end{aligned}$$

which goes to zero under $\max\{\bar{\delta}_{m,n}^2, \lambda_n\} = o(1)$. Thus $\|\widehat{h}_n - h_0\|_s = o_p(1)$.
Q.E.D.

PROOF OF THEOREM 3.3: We divide the proof in two steps: first we show consistency under the weak topology; second we establish consistency under the strong norm.

Step 1. We can establish consistency in the weak topology by applying Lemma A.1, either verifying its conditions or following its proof directly. Under stated conditions, $\widehat{h}_n \in \mathcal{H}_{k(n)}$ with probability approaching 1. By Lemma A.3(ii) with $\max\{\eta_{0,n}, E[\|m(X, \Pi_n h_0)\|_W^2]\} = o(\lambda_n)$ and $\eta_n = O(\eta_{0,n})$, we have $P(\widehat{h}_n) - P(h_0) \leq o_p(1)$; thus we can focus on the set $\{h \in \mathcal{H}_{k(n)} : P(h) \leq M_0\} = \mathcal{H}_{k(n)}^{M_0}$ for all n large enough. Let $\mathcal{B}_w(h_0)$ denote any open neighborhood (in the weak topology) around h_0 , and let $\mathcal{B}_w^c(h_0)$ denote its complement (under the weak topology) in \mathcal{H} . By Lemma A.4(ii) with $\mathcal{B}_T(h_0) = \mathcal{B}_w(h_0)$, $\lambda_n P(h) \geq 0$, $\mathcal{H}_{k(n)}^{M_0} \subseteq \mathcal{H}$, $\eta_n = O(\eta_{0,n})$, and $\max\{\eta_{0,n}, E[\|m(X, \Pi_n h_0)\|_W^2]\} = o(\lambda_n)$, we have, for all nonempty open balls $\mathcal{B}_w(h_0)$, all $\varepsilon > 0$, and n sufficiently large,

$$\begin{aligned}
& \Pr(\widehat{h}_n \notin \mathcal{B}_w(h_0)) \\
& \leq \Pr(\widehat{h}_n \notin \mathcal{B}_w(h_0), \widehat{h}_n \in \mathcal{H}_{k(n)}^{M_0}) + \varepsilon \\
& \leq \Pr\left(\inf_{\mathcal{H}_{k(n)}^{M_0} : h \notin \mathcal{B}_w(h_0)} \{cE[\|m(X, h)\|_W^2] + \lambda_n P(h)\}\right)
\end{aligned}$$

$$\begin{aligned} &\leq O_p(\bar{\delta}_{m,n}^2) + \lambda_n P(h_0) + o(\lambda_n) + \varepsilon \\ &\leq \Pr\left(\inf_{\mathcal{H}: h \notin \mathcal{B}_w(h_0)} E[\|m(X, h)\|_W^2] \leq O_p(\max\{\bar{\delta}_{m,n}^2, \lambda_n\})\right) + \varepsilon. \end{aligned}$$

Let $E[\|m(X, h)\|_W^2]$ be weak sequentially lower semicontinuous on \mathcal{H} . Since $\mathcal{H} \cap \mathcal{B}_w^c(h_0)$ is weakly compact (weakly closed and bounded), by Assumption 3.4(ii) and Theorem 38.A in Zeidler (1985), there exists $h^*(\mathcal{B}) \in \mathcal{H} \cap \mathcal{B}_w^c(h_0)$ such that $\inf_{\mathcal{H}: h \notin \mathcal{B}_w(h_0)} E[\|m(X, h)\|_W^2] = E[\|m(X, h^*(\mathcal{B}))\|_W^2]$. It must hold that $g(\mathcal{B}) \equiv E[\|m(X, h^*(\mathcal{B}))\|_W^2] > 0$; otherwise, by Assumption 3.1(i) and (ii), $\|h^*(\mathcal{B}) - h_0\|_s = 0$. But if this is the case, then for any $t \in \mathbf{H}^*$ we have $|\langle t, h^*(\mathcal{B}) - h_0 \rangle_{\mathbf{H}^*, \mathbf{H}}| \leq \text{const.} \times \|h^*(\mathcal{B}) - h_0\|_s = 0$, a contradiction to the fact that $h^*(\mathcal{B}) \notin \mathcal{B}_w(h_0)$. Thus

$$\begin{aligned} &\Pr(\widehat{h}_n \notin \mathcal{B}_w(h_0), \widehat{h}_n \in \mathcal{H}_{k(n)}^{M_0}) \\ &\leq \Pr(E[\|m(X, h^*(\mathcal{B}))\|_W^2] \leq O_p(\max\{\bar{\delta}_{m,n}^2, \lambda_n\})), \end{aligned}$$

which goes to zero since $\max\{\bar{\delta}_{m,n}^2, \lambda_n\} = o(1)$. Hence $\Pr(\widehat{h}_n \notin \mathcal{B}_w(h_0)) \rightarrow 0$.

Step 2. Consistency under the weak topology implies that $\langle t_0, \widehat{h}_n - h_0 \rangle_{\mathbf{H}^*, \mathbf{H}} = o_p(1)$. By Assumption 3.4(i), $P(\widehat{h}_n) - P(h_0) \geq \langle t_0, \widehat{h}_n - h_0 \rangle_{\mathbf{H}^*, \mathbf{H}} + g(\|\widehat{h}_n - h_0\|_s)$. Lemma A.3(ii) implies that $P(\widehat{h}_n) - P(h_0) \leq o_p(1)$ under $\max\{\eta_{0,n}, E[\|m(X, \Pi_n h_0)\|_W^2]\} = o(\lambda_n)$, $\eta_n = O(\eta_{0,n})$. Thus $g(\|\widehat{h}_n - h_0\|_s) = o_p(1)$ and $\|\widehat{h}_n - h_0\|_s = o_p(1)$ by our assumption over $g(\cdot)$. This, $\langle t_0, \widehat{h}_n - h_0 \rangle_{\mathbf{H}^*, \mathbf{H}} = o_p(1)$, and Assumption 3.4(i) imply that $P(\widehat{h}_n) - P(h_0) \geq o_p(1)$. But $P(\widehat{h}_n) \leq P(h_0) + o_p(1)$ by Lemma A.3(ii). Thus $P(\widehat{h}_n) - P(h_0) = o_p(1)$. *Q.E.D.*

VERIFICATION OF REMARK 3.2: Claim (i) follows from Proposition 38.7 of Zeidler (1985). Claim (ii) follows from Corollary 41.9 of Zeidler (1985). For claim (iii), the fact that $\sqrt{W(\cdot)}m(\cdot, h) : \mathcal{H} \rightarrow L^2(f_X)$ is compact and Frechet differentiable implies that its Frechet derivative is also a compact operator; see Zeidler (1985, Proposition 7.33). This and the chain rule imply that the functional $E[\|m(X, \cdot)\|_W^2] : \mathcal{H} \rightarrow [0, \infty)$ is Frechet differentiable and its Frechet derivative is compact on \mathcal{H} . Hence $E[\|m(X, h)\|_W^2]$ has a compact Gateaux derivative on \mathcal{H} and, by claim (ii), is weak sequentially lower semicontinuous on \mathcal{H} . *Q.E.D.*

PROOF OF THEOREM A.1: For result (i), we first show that the set of minimum penalization solutions, \mathcal{M}_0^P , is not empty. Since $E[\|m(X, h)\|_W^2]$ is convex and lower semicontinuous in $h \in \mathcal{H}$ and \mathcal{H} is a convex, closed, and bounded subset of a reflexive Banach space (Assumption 3.4(ii)), by Proposition 38.15 of Zeidler (1985), \mathcal{M}_0 is convex, closed, and bounded (and nonempty). Since $P(\cdot)$ is convex and lower semicontinuous on \mathcal{M}_0 , by applying Proposition 38.15 of Zeidler (1985), we have that the set \mathcal{M}_0^P is nonempty, convex, closed, and a

bounded subset of \mathcal{M}_0 . Next, we show uniqueness of the minimum penalization solution. Suppose that there exist $h_1, h_0 \in \mathcal{M}_0^p$ such that $\|h_1 - h_0\|_s > 0$. Since \mathcal{M}_0^p is a subset of \mathcal{M}_0 and since \mathcal{M}_0 is convex, $h' = \lambda h_1 + (1 - \lambda)h_0 \in \mathcal{M}_0$. Since $P(\cdot)$ is strictly convex on \mathcal{M}_0 (in $\|\cdot\|_s$), thus $P(h') < P(h_0)$, but this is a contradiction since h_0 is a minimum penalization solution. Thus we have established result (i).

For result (ii), first, as already shown earlier, $\widehat{h}_n \in \mathcal{H}_{k(n)}$ with probability approaching 1. We now show its consistency under the weak topology. To establish this, we adapt Step 1 in the [proof](#) of Theorem 3.3 to the case where Assumption 3.1(ii) (identification) may not hold, but h_0 is the minimum penalization solution. Let $\mathcal{B}_w(h_0)$ denote any open neighborhood (in the weak topology) around h_0 , and let $\mathcal{B}_w^c(h_0)$ denote its complement (under the weak topology) in \mathcal{H} . By Lemma A.3(ii), $P(\widehat{h}_n) = O_p(1)$. By Lemma A.4(ii) with $\mathcal{B}_T(h_0) = \mathcal{B}_w(h_0)$, $\mathcal{H}_{k(n)}^{M_0} \subseteq \mathcal{H}_{k(n)}$, $\eta_n = O(\eta_{0,n})$, and $\max\{\bar{\delta}_{m,n}^2, \eta_{0,n}, E[\|m(X, \Pi_n h_0)\|_W^2]\} = o(\lambda_n)$, we have, for all nonempty open balls $\mathcal{B}_w(h_0)$,

$$\begin{aligned} & \Pr(\widehat{h}_n \notin \mathcal{B}_w(h_0), \widehat{h}_n \in \mathcal{H}_{k(n)}^{M_0}) \\ & \leq \Pr\left(\inf_{\mathcal{H}_{k(n)}^{M_0}: h \notin \mathcal{B}_w(h_0)} \{cE[\|m(X, h)\|_W^2] + \lambda_n P(h)\}\right) \\ & \leq \lambda_n P(h_0) + o_p(\lambda_n) \\ & \leq \Pr\left(\inf_{\mathcal{H}_{k(n)}: h \notin \mathcal{B}_w(h_0)} \{cE[\|m(X, h)\|_W^2] + \lambda_n P(h)\}\right) \\ & \leq \lambda_n P(h_0) + o_p(\lambda_n). \end{aligned}$$

By Assumptions 3.1(iii) and 3.4(ii), $\mathcal{H}_{k(n)}$ is weakly sequentially compact. Since $\mathcal{B}_w^c(h_0)$ is closed under the weak topology, the set $\mathcal{H}_{k(n)} \cap \mathcal{B}_w^c(h_0)$ is weakly sequentially compact. By Assumption 3.4(ii) and the assumption that $E[\|m(X, h)\|_W^2]$ is convex and lower semicontinuous on \mathcal{H} , $cE[\|m(X, h)\|_W^2] + \lambda_n P(h)$ is weakly sequentially lower semicontinuous on $\mathcal{H}_{k(n)}$. Thus $g(k(n), \varepsilon, \lambda_n) \equiv \inf_{\mathcal{H}_{k(n)} \cap \mathcal{B}_w^c(h_0)} \{cE[\|m(X, h)\|_W^2] + \lambda_n P(h)\} \geq 0$ exists, and we denote its minimizer as $h_n(\varepsilon) \in \mathcal{H}_{k(n)} \cap \mathcal{B}_w^c(h_0)$. Hence, with $\max\{\bar{\delta}_{m,n}^2, \eta_{0,n}, E[\|m(X, \Pi_n h_0)\|_W^2]\} = o(\lambda_n)$ and $\lambda_n > 0$, we have

$$\begin{aligned} & \Pr(\widehat{h}_n \notin \mathcal{B}_w(h_0), \widehat{h}_n \in \mathcal{H}_{k(n)}^{M_0}) \\ & \leq \Pr\left(cE[\|m(X, h_n(\varepsilon))\|_W^2] + \lambda_n P(h_n(\varepsilon)) \leq \lambda_n P(h_0) + o_p(\lambda_n)\right) \\ & = \Pr\left(\frac{g(k(n), \varepsilon, \lambda_n) - \lambda_n P(h_0)}{\lambda_n} \leq o_p(1)\right). \end{aligned}$$

If $\liminf_n E[\|m(X, h_n(\varepsilon))\|_W^2] = \text{const.} > 0$, then $\Pr(\widehat{h}_n \notin \mathcal{B}_w(h_0), \widehat{h}_n \in \mathcal{H}_{k(n)}^{M_0}) \rightarrow 0$ trivially. So we assume $\liminf_n E[\|m(X, h_n(\varepsilon))\|_W^2] = \text{const.} = 0$. Since $\mathcal{H} \cap \mathcal{B}_w^c(h_0)$ is weakly compact, there exists a subsequence $\{h_{n_k}(\varepsilon)\}_k$ that converges (weakly) to $h_\infty(\varepsilon) \in \mathcal{H} \cap \mathcal{B}_w^c(h_0)$. By weakly lower semicontinuity of $E[\|m(X, h)\|_W^2]$ on \mathcal{H} , $h_\infty(\varepsilon) \in \mathcal{M}_0$. By definition of h_0 and the assumption that $P(h)$ is strictly convex in $h \in \mathcal{M}_0$, it must be that $P(h_\infty(\varepsilon)) - P(h_0) \geq \text{const.} > 0$ by result (i). Note that this is true for any convergent subsequence. Therefore, we have established that

$$\liminf_n \frac{g(k(n), \varepsilon, \lambda_n) - \lambda_n P(h_0)}{\lambda_n} \geq \text{const.} > 0;$$

thus $\Pr(\widehat{h}_n \notin \mathcal{B}_w(h_0), \widehat{h}_n \in \mathcal{H}_{k(n)}^{M_0}) \rightarrow 0$. Hence, by similar calculations to those in Lemma A.4(ii), for any $\varepsilon > 0$ and sufficiently large n , $\Pr(\widehat{h}_n \notin \mathcal{B}_w(h_0)) \leq \Pr(\widehat{h}_n \notin \mathcal{B}_w(h_0), \widehat{h}_n \in \mathcal{H}_{k(n)}^{M_0}) + \varepsilon \leq 2\varepsilon$.

Given the consistency under the weak topology, Assumption 3.4(i) and Lemma A.3(ii), we obtain $\|\widehat{h}_n - h_0\|_s = o_p(1)$ and $P(\widehat{h}_n) - P(h_0) = o_p(1)$ by following Step 2 in the [proof](#) of Theorem 3.3. *Q.E.D.*

C. CONSISTENCY: PROOFS OF LEMMAS

PROOF OF LEMMA A.1: By definition of the infimum, $\widehat{\alpha}_n$ always exists, and $\widehat{\alpha}_n \in \mathcal{A}_{k(n)}$ with outer probability approaching 1 ($\widehat{\alpha}_n$ may not be measurable). It follows that for all $\mathcal{B}_{\mathcal{T}}(\alpha_0)$,

$$\begin{aligned} & \Pr^*(\widehat{\alpha}_n \in \mathcal{A}_{k(n)}, \widehat{\alpha}_n \notin \mathcal{B}_{\mathcal{T}}(\alpha_0)) \\ & \leq \Pr^*\left(\inf_{\alpha \in \mathcal{A}_{k(n)}: \alpha \notin \mathcal{B}_{\mathcal{T}}(\alpha_0)} \widehat{Q}_n(\alpha) \leq \widehat{Q}_n(\Pi_n \alpha_0) + O_{p^*}(\eta_n)\right) \\ & \leq \Pr^*\left(\inf_{\alpha \in \mathcal{A}_{k(n)}: \alpha \notin \mathcal{B}_{\mathcal{T}}(\alpha_0)} \{K \overline{Q}_n(\alpha) - O_{p^*}(c_n)\}\right) \\ & \leq K_0 \overline{Q}_n(\Pi_n \alpha_0) + O_{p^*}(c_{0,n}) + O_{p^*}(\eta_n) \\ & \leq \Pr^*\left(\inf_{\alpha \in \mathcal{A}_{k(n)}: \alpha \notin \mathcal{B}_{\mathcal{T}}(\alpha_0)} \overline{Q}_n(\alpha) \leq O_{p^*}(\max\{c_n, c_{0,n}, \overline{Q}_n(\Pi_n \alpha_0), \eta_n\})\right) \\ & \leq \Pr^*(g_0(n, k(n), \mathcal{B}) \leq O_{p^*}(\max\{c_n, c_{0,n}, \overline{Q}_n(\Pi_n \alpha_0), \eta_n\})) \\ & \text{by condition a(ii) in Lemma A.1,} \end{aligned}$$

which goes to 0 by condition d(iii) in Lemma A.1.

Q.E.D.

PROOF OF LEMMA A.2: Under condition c(ii) of Lemma A.2, $\widehat{\alpha}_n$ is well defined and measurable. It follows that for any $\varepsilon > 0$,

$$\begin{aligned}
& \Pr(\|\widehat{\alpha}_n - \alpha_0\|_s > \varepsilon) \\
& \leq \Pr\left(\inf_{\alpha \in \mathcal{A}_{k(n)}: \|\alpha - \alpha_0\|_s \geq \varepsilon} \widehat{Q}_n(\alpha) \leq \widehat{Q}_n(\Pi_n \alpha_0) + O_p(\eta_n)\right) \\
& \leq \Pr\left(\inf_{\alpha \in \mathcal{A}_{k(n)}: \|\alpha - \alpha_0\|_s \geq \varepsilon} \{|\overline{Q}_n(\alpha) - |\widehat{Q}_n(\alpha) - \overline{Q}_n(\alpha)|\}\right. \\
& \quad \left. \leq \overline{Q}_n(\Pi_n \alpha_0) + |\widehat{Q}_n(\Pi_n \alpha_0) - \overline{Q}_n(\Pi_n \alpha_0)| + O_p(\eta_n)\right) \\
& \leq \Pr\left(\inf_{\alpha \in \mathcal{A}_{k(n)}: \|\alpha - \alpha_0\|_s \geq \varepsilon} \overline{Q}_n(\alpha) \leq 2\widehat{c}_n + \overline{Q}_n(\Pi_n \alpha_0) + O_p(\eta_n)\right) \\
& = \Pr\left(\inf_{\alpha \in \mathcal{A}_{k(n)}: \|\alpha - \alpha_0\|_s \geq \varepsilon} \overline{Q}_n(\alpha) - \overline{Q}_n(\alpha_0)\right. \\
& \quad \left. \leq 2\widehat{c}_n + \overline{Q}_n(\Pi_n \alpha_0) - \overline{Q}_n(\alpha_0) + O_p(\eta_n)\right) \\
& \leq \Pr(g_0(n, k(n), \varepsilon) \leq 2\widehat{c}_n + |\overline{Q}_n(\Pi_n \alpha_0) - \overline{Q}_n(\alpha_0)| + O_p(\eta_n))
\end{aligned}$$

which goes to 0 by condition d of Lemma A.2.

Q.E.D.

PROOF OF LEMMA A.3: We first show that $\widehat{h}_n \in \mathcal{H}_n$ w.p.a.1. The infimum $\inf_{\gamma_n} \widehat{Q}_n(h)$ exists w.p.a.1 and hence, for any $\varepsilon > 0$, there is a sequence, $(h_{j,n}(\varepsilon))_j \subseteq \mathcal{H}_n$ such that $\widehat{Q}_n(h_{j,n}(\varepsilon)) \leq \inf_{\gamma_n} \widehat{Q}_n(h) + \varepsilon$ w.p.a.1. Let $\widehat{h}_n \equiv h_{n,n}(\eta_n)$. Then such a choice satisfies $\widehat{h}_n \in \mathcal{H}_n$ w.p.a.1.

Next, by definition of \widehat{h}_n , we have for any $\lambda_n > 0$,

$$\begin{aligned}
\lambda_n \widehat{P}_n(\widehat{h}_n) & \leq \frac{1}{n} \sum_{i=1}^n \|\widehat{m}(X_i, \widehat{h}_n)\|_{\widehat{W}}^2 + \lambda_n \widehat{P}_n(\widehat{h}_n) \\
& \leq \frac{1}{n} \sum_{i=1}^n \|\widehat{m}(X_i, \Pi_n h_0)\|_{\widehat{W}}^2 + \lambda_n \widehat{P}_n(\Pi_n h_0) + O_p(\eta_n)
\end{aligned}$$

and

$$\begin{aligned}
& \lambda_n \{P(\widehat{h}_n) - P(h_0)\} + \lambda_n \{\widehat{P}_n(\widehat{h}_n) - P(\widehat{h}_n)\} \\
& \leq \frac{1}{n} \sum_{i=1}^n \|\widehat{m}(X_i, \Pi_n h_0)\|_{\widehat{W}}^2 + \lambda_n \{\widehat{P}_n(\Pi_n h_0) - P(\Pi_n h_0)\} \\
& \quad + \lambda_n \{P(\Pi_n h_0) - P(h_0)\} + O_p(\eta_n).
\end{aligned}$$

Thus

$$\begin{aligned}
& \lambda_n \{P(\widehat{h}_n) - P(h_0)\} \\
& \leq \frac{1}{n} \sum_{i=1}^n \|\widehat{m}(X_i, \Pi_n h_0)\|_{\widehat{W}}^2 + 2\lambda_n \sup_{h \in \mathcal{H}_n} |\widehat{P}_n(h) - P(h)| \\
& \quad + \lambda_n |P(\Pi_n h_0) - P(h_0)| + O_p(\eta_n) \\
& \leq O_p(\eta_{0,n} + E[\|m(X, \Pi_n h_0)\|_{\widehat{W}}^2]) + 2\lambda_n \sup_{h \in \mathcal{H}_n} |\widehat{P}_n(h) - P(h)| \\
& \quad + \lambda_n |P(\Pi_n h_0) - P(h_0)|,
\end{aligned}$$

where the last inequality is due to Assumption 3.3(i) and $\eta_n = O(\eta_{0,n})$. Therefore, for all $M > 0$,

$$\begin{aligned}
& \Pr(P(\widehat{h}_n) - P(h_0) > M) \\
& = \Pr(\lambda_n \{P(\widehat{h}_n) - P(h_0)\} > \lambda_n M) \\
& \leq \Pr\left(O_p(\eta_{0,n} + E[\|m(X, \Pi_n h_0)\|_{\widehat{W}}^2]) + 2\lambda_n \sup_{h \in \mathcal{H}_n} |\widehat{P}_n(h) - P(h)| \right. \\
& \quad \left. + \lambda_n |P(\Pi_n h_0) - P(h_0)| > \lambda_n M\right).
\end{aligned}$$

(i) Under Assumption 3.2(b), $\lambda_n \sup_{h \in \mathcal{H}_n} |\widehat{P}_n(h) - P(h)| + \lambda_n |P(\Pi_n h_0) - P(h_0)| = O_p(\lambda_n)$, we have

$$\begin{aligned}
& \Pr(P(\widehat{h}_n) - P(h_0) > M) \\
& \leq \Pr\left(O_p(\max\{\eta_{0,n} + E[\|m(X, \Pi_n h_0)\|_{\widehat{W}}^2], \lambda_n\}) > \lambda_n M\right) \\
& \leq \Pr\left(O_p\left(\frac{\eta_{0,n} + E[\|m(X, \Pi_n h_0)\|_{\widehat{W}}^2]}{\lambda_n}\right) + O_p(1) > M\right),
\end{aligned}$$

which, under $\max\{\eta_{0,n}, E[\|m(X, \Pi_n h_0)\|_{\widehat{W}}^2]\} = O(\lambda_n)$, goes to zero as $M \rightarrow \infty$. Thus $P(\widehat{h}_n) - P(h_0) = O_p(1)$. Since $0 \leq P(h_0) < \infty$, we have $P(\widehat{h}_n) = O_p(1)$.

(ii) Under Assumption 3.2(c), $\lambda_n \sup_{h \in \mathcal{H}_n} |\widehat{P}_n(h) - P(h)| + \lambda_n |P(\Pi_n h_0) - P(h_0)| = o_p(\lambda_n)$, we have

$$\begin{aligned}
& \Pr(P(\widehat{h}_n) - P(h_0) > M) \\
& \leq \Pr\left(O_p\left(\frac{\eta_{0,n} + E[\|m(X, \Pi_n h_0)\|_{\widehat{W}}^2]}{\lambda_n}\right) + o_p(1) > M\right),
\end{aligned}$$

which, under $\max\{\eta_{0,n}, E[\|m(X, \Pi_n h_0)\|_W^2]\} = o(\lambda_n)$, goes to zero for all $M > 0$. Thus $P(\hat{h}_n) - P(h_0) \leq o_p(1)$. *Q.E.D.*

PROOF OF LEMMA A.4: It suffices to consider $\lambda_n P(\cdot) > 0$ only. By the fact that $\Pr(A) \leq \Pr(A \cap B) + \Pr(B^c)$ for any measurable sets A and B , we have

$$\Pr(\hat{h}_n \notin \mathcal{B}_T(h_0)) \leq \Pr(\hat{h}_n \notin \mathcal{B}_T(h_0), P(\hat{h}_n) \leq M_0) + \Pr(P(\hat{h}_n) > M_0).$$

For any $\varepsilon > 0$, choose $M_0 \equiv M_0(\varepsilon)$ such that $\Pr(P(\hat{h}_n) > M_0) < \varepsilon$ for sufficiently large n . Note that such a M_0 always exists by Lemma A.3. Thus, we can focus on the set $\mathcal{H}_{k(n)}^{M_0} \equiv \{h \in \mathcal{H}_{k(n)} : \lambda_n P(h) \leq \lambda_n M_0\}$ and bound $\Pr(\hat{h}_n \notin \mathcal{B}_T(h_0), P(\hat{h}_n) \leq M_0)$.

By definition of \hat{h}_n and $\Pi_n h_0$, Assumptions 3.3 and 3.1(iii), and $\eta_n = O(\eta_{0,n})$, we have, for all $\mathcal{B}_T(h_0)$,

$$\begin{aligned} & \Pr(\hat{h}_n \notin \mathcal{B}_T(h_0), \hat{h}_n \in \mathcal{H}_{k(n)}^{M_0}) \\ & \leq \Pr\left(\inf_{h \in \mathcal{H}_{k(n)}^{M_0}, h \notin \mathcal{B}_T(h_0)} \left\{ \frac{1}{n} \sum_{i=1}^n \|\hat{m}(X_i, h)\|_W^2 + \lambda_n \hat{P}(h) \right\}\right) \\ & \leq \frac{1}{n} \sum_{i=1}^n \|\hat{m}(X_i, \Pi_n h_0)\|_W^2 + \lambda_n \hat{P}(\Pi_n h_0) + O_p(\eta_n) \\ & \leq \Pr\left(\inf_{h \in \mathcal{H}_{k(n)}^{M_0}, h \notin \mathcal{B}_T(h_0)} \{cE[\|m(X, h)\|_W^2] + \lambda_n \hat{P}(h)\}\right) \\ & \leq c'E[\|m(X, \Pi_n h_0)\|_W^2] + O_p(\bar{\delta}_{m,n}^2) + \lambda_n \hat{P}(\Pi_n h_0) + O_p(\eta_{0,n}). \end{aligned}$$

By Assumption 3.2(b), we have $\lambda_n \sup_{h \in \mathcal{H}_n} |\hat{P}(h) - P(h)| = O_p(\lambda_n)$ and $\lambda_n |P(\Pi_n h_0) - P(h_0)| = O(\lambda_n)$. Thus, with $\max\{\eta_{0,n}, E[\|m(X, \Pi_n h_0)\|_W^2]\} = O(\lambda_n)$, for all $\mathcal{B}_T(h_0)$,

$$\begin{aligned} & \Pr(\hat{h}_n \notin \mathcal{B}_T(h_0), \hat{h}_n \in \mathcal{H}_{k(n)}^{M_0}) \\ & \leq \Pr\left(\inf_{h \in \mathcal{H}_{k(n)}^{M_0}, h \notin \mathcal{B}_T(h_0)} \{cE[\|m(X, h)\|_W^2] + \lambda_n P(h)\}\right) \\ & \leq O_p(\bar{\delta}_{m,n}^2) + \lambda_n P(h_0) + O_p(\lambda_n). \end{aligned}$$

By Assumption 3.2(c), we have $\lambda_n \sup_{h \in \mathcal{H}_n} |\hat{P}(h) - P(h)| = o_p(\lambda_n)$ and $\lambda_n |P(\Pi_n h_0) - P(h_0)| = o(\lambda_n)$ for $\lambda_n > 0$. Thus, with $\max\{\eta_{0,n}, E[\|m(X,$

$\Pi_n h_0\|_W^2\}} = o(\lambda_n)$, for all $\mathcal{B}_T(h_0)$,

$$\begin{aligned} & \Pr(\widehat{h}_n \notin \mathcal{B}_T(h_0), \widehat{h}_n \in \mathcal{H}_{k(n)}^{M_0}) \\ & \leq \Pr\left(\inf_{h \in \mathcal{H}_{k(n)}^{M_0}: h \notin \mathcal{B}_T(h_0)} \{cE[\|m(X, h)\|_W^2] + \lambda_n P(h)\}\right) \\ & \leq O_p(\bar{\delta}_{m,n}^2) + \lambda_n P(h_0) + o_p(\lambda_n). \end{aligned}$$

Hence we obtain results (i) and (ii).

Q.E.D.

D. CONVERGENCE RATE: PROOFS OF THEOREMS

The proof of Theorem 4.1 directly follows from Lemma B.1 and the definition of $\omega_n(\delta, \mathcal{H}_{osn})$. The proof of Corollary 5.1 directly follows from Theorem 4.1 and Lemma B.2. The proof of Corollary 5.2 directly follows from Theorem 4.1 and Lemmas B.2 and B.3.

PROOF OF COROLLARY 5.3: By Theorem 4.1, Lemmas B.2 and B.3(ii), results of Corollary 5.2 are obviously true. We now specialize Corollary 5.2 to the PSMD estimator using a series LS estimator $\widehat{m}(X, h)$. For this case, we have $\delta_{m,n}^{*2} = \frac{J_n^*}{n} \asymp b_{m,J_n^*}^2$.

By Assumption 5.4(ii) and the condition that either $P(h) \geq \sum_{j=1}^{\infty} \nu_j^{2\alpha} |\langle h, q_j \rangle_s|^2$ for all $h \in \mathcal{H}_{os}$ or $\mathcal{H}_{os} \subseteq \mathcal{H}_{\text{ellipsoid}}$, we have, for all $h \in \mathcal{H}_{os}$,

$$\begin{aligned} c_2 E[m(X, h)' W(X) m(X, h)] & \leq \|h - h_0\|^2 \\ & \leq \text{const.} \sum_{j=1}^{\infty} \{\varphi(\nu_j^{-2})\} \langle h - h_0, q_j \rangle_s^2. \end{aligned}$$

On the other hand, the choice of penalty and the definition of \mathcal{H}_{os} imply that $\sum_j \nu_j^{2\alpha} \langle h - h_0, q_j \rangle_s^2 \leq \text{const.}$ for all $h \in \mathcal{H}_{os}$. Denote $\eta_j = \{\varphi(\nu_j^{-2})\} \langle h - h_0, q_j \rangle_s^2$. Then $\sum_j \nu_j^{2\alpha} \{\varphi(\nu_j^{-2})\}^{-1} \eta_j \leq M$. Therefore, the class $\{g \in L^2(\mathcal{X}, \|\cdot\|_{L^2(f_X)}) : g(\cdot) = \sqrt{W(\cdot)} m(\cdot, h), h \in \mathcal{H}_{os}\}$ is embedded in the ellipsoid $\{g \in L^2(\mathcal{X}, \|\cdot\|_{L^2(f_X)}) : \|g\|_{L^2(f_X)}^2 = \sum_j \eta_j, \text{ and } \sum_j \nu_j^{2\alpha} \{\varphi(\nu_j^{-2})\}^{-1} \eta_j \leq M'\}$ for some finite constant M' . By invoking the results of Yang and Barron (1999), it follows that the J_n th approximation error rate of this ellipsoid satisfies $b_{m,J_n}^2 \leq \text{const.} \nu_{J_n}^{-2\alpha} \{\varphi(\nu_{J_n}^{-2})\}$. Hence $\delta_{m,n}^{*2} = \frac{J_n^*}{n} \asymp b_{m,J_n^*}^2 \leq \text{const.} \nu_{J_n^*}^{-2\alpha} \{\varphi(\nu_{J_n^*}^{-2})\}$ and $\|\widehat{h} - h_0\|_s = O_p(\nu_{J_n^*}^{-\alpha}) = O_p(\sqrt{\frac{J_n^*}{n} \{\varphi(\nu_{J_n^*}^{-2})\}^{-1}})$. *Q.E.D.*

E. CONVERGENCE RATE: PROOFS OF LEMMAS

PROOF OF LEMMA B.1: Let $r_n^2 = \max\{\delta_{m,n}^2, \lambda_n \delta_{p,n}, \|\Pi_n h_0 - h_0\|^2, \lambda_n |P(\Pi_n h_0) - P(\widehat{h}_n)|\} = o_p(1)$. Since $\widehat{h}_n \in \mathcal{H}_{osn}$ with probability approaching 1, we have, for all $M > 1$,

$$\begin{aligned} & \Pr\left(\frac{\|\widehat{h}_n - h_0\|}{r_n} \geq M\right) \\ & \leq \Pr\left(\inf_{\{h \in \mathcal{H}_{osn} : \|h - h_0\| \geq Mr_n\}} \left\{ \frac{1}{n} \sum_{i=1}^n \|\widehat{m}(X_i, h)\|_{\widehat{W}}^2 + \lambda_n \widehat{P}_n(h) \right\}\right) \\ & \leq \frac{1}{n} \sum_{i=1}^n \|\widehat{m}(X_i, \Pi_n h_0)\|_{\widehat{W}}^2 + \lambda_n \widehat{P}_n(\Pi_n h_0) + O_p(\eta_n) \\ & \leq \Pr\left(\inf_{\{h \in \mathcal{H}_{osn} : \|h - h_0\| \geq Mr_n\}} \left\{ \frac{1}{n} \sum_{i=1}^n \|\widehat{m}(X_i, h)\|_{\widehat{W}}^2 + \lambda_n P(h) \right\}\right) \\ & \leq \frac{1}{n} \sum_{i=1}^n \|\widehat{m}(X_i, \Pi_n h_0)\|_{\widehat{W}}^2 + \lambda_n P(\Pi_n h_0) + 2\lambda_n \delta_{p,n} + O_p(\eta_n), \end{aligned}$$

where the last inequality is due to $\sup_{h \in \mathcal{H}_{osn}} |\widehat{P}_n(h) - P(h)| = O_p(\delta_{p,n}) = O_p(1)$. By Assumption 3.3 with $\eta_{0,n} = O(\delta_{m,n}^2)$ and $\eta_n = O(\eta_{0,n})$, and definitions of \mathcal{H}_{osn} and $\delta_{m,n}^2$, there are two finite constants $c, c_0 > 0$ such that

$$\begin{aligned} \text{(SM.4)} \quad & cE(\|m(X, \widehat{h}_n)\|_{\widehat{W}}^2) + \lambda_n P(\widehat{h}_n) \\ & \leq O_p(\delta_{m,n}^2 + \lambda_n \delta_{p,n}) + c_0 E(\|m(X, \Pi_n h_0)\|_{\widehat{W}}^2) + \lambda_n P(\Pi_n h_0), \end{aligned}$$

which implies

$$\begin{aligned} cE(\|m(X, \widehat{h}_n)\|_{\widehat{W}}^2) & \leq O_p(\delta_{m,n}^2 + \lambda_n \delta_{p,n}) + c_0 E(\|m(X, \Pi_n h_0)\|_{\widehat{W}}^2) \\ & \quad + \lambda_n |P(\Pi_n h_0) - P(\widehat{h}_n)|. \end{aligned}$$

This, $\|\widehat{h}_n - h_0\|_s = o_p(1)$, and Assumption 4.1 imply that

$$\begin{aligned} & \Pr\left(\frac{\|\widehat{h}_n - h_0\|}{r_n} \geq M\right) \\ & \leq \Pr\left(M^2 r_n^2 \right. \\ & \quad \left. \leq O_p(\max\{\delta_{m,n}^2, \lambda_n \delta_{p,n}, \|\Pi_n h_0 - h_0\|^2, \lambda_n |P(\Pi_n h_0) - P(\widehat{h}_n)|\})\right), \end{aligned}$$

which, given our choice of r_n , goes to zero as $M \rightarrow \infty$; hence $\|\widehat{h}_n - h_0\| = O_p(r_n)$.

By definition of \mathcal{H}_{osn} (or under Assumption 3.2(b)), $\lambda_n|P(\Pi_n h_0) - P(\widehat{h}_n)| = O_p(\lambda_n)$ and $\delta_{p,n} = O_p(1)$; hence result (i) follows.

Under Assumption 3.2(c), $\lambda_n|P(\Pi_n h_0) - P(\widehat{h}_n)| = o_p(\lambda_n)$ and $\delta_{p,n} = o_p(1)$; hence result (ii) follows.

For result (iii), using the same argument as that for results (i) and (ii), inequality (SM.4) still holds. By condition (iii) of Theorem 4.1, $\lambda_n(P(\widehat{h}_n) - P(\Pi_n h_0)) \geq \lambda_n \langle t_0, \widehat{h}_n - \Pi_n h_0 \rangle_{\mathbb{H}^*, \mathbb{H}}$. Thus

$$\begin{aligned} & cE(\|m(X, \widehat{h}_n)\|_W^2) + \lambda_n \langle t_0, \widehat{h}_n - \Pi_n h_0 \rangle_{\mathbb{H}^*, \mathbb{H}} \\ & \leq O_p(\delta_{m,n}^2 + \lambda_n \delta_{p,n}) + c_0 E(\|m(X, \Pi_n h_0)\|_W^2); \end{aligned}$$

hence

$$\begin{aligned} cE(\|m(X, \widehat{h}_n)\|_W^2) & \leq O_p(\delta_{m,n}^2 + \lambda_n \delta_{p,n}) + c_0 E(\|m(X, \Pi_n h_0)\|_W^2) \\ & \quad + \text{const.} \lambda_n \|\widehat{h}_n - \Pi_n h_0\|_s. \end{aligned}$$

By Assumption 4.1, Lemma B.1(iii) follows by choosing $r_n^2 = \max\{\delta_{m,n}^2, \lambda_n \delta_{p,n}, \|\Pi_n h_0 - h_0\|^2, \lambda_n \|\widehat{h}_n - \Pi_n h_0\|_s\} = o_p(1)$. *Q.E.D.*

PROOF OF LEMMA B.2: To simplify notation, we denote $b_j = \varphi(v_j^{-2})$. Result (i) follows directly from the definition of $\omega_n(\delta, \mathcal{H}_{osn})$ as well as the fact that $\{q_j\}_{j=1}^\infty$ is a Riesz basis, and hence for any $h \in \mathcal{H}_{osn}$, there is a finite constant $c_1 > 0$ such that

$$\begin{aligned} c_1 \|h\|_s^2 & \leq \sum_{j \leq k(n)} |\langle h, q_j \rangle_s|^2 \\ & \leq \left(\max_{j \leq k(n)} b_j^{-1} \right) \sum_{j \leq k(n)} b_j |\langle h, q_j \rangle_s|^2 \leq \frac{1}{c b_{k(n)}} \|h\|^2, \end{aligned}$$

where the last inequality is due to Assumption 5.2(i) and $\{b_j\}$ nonincreasing. Similarly, Assumption 5.2(ii) implies result (ii) since

$$\begin{aligned} c_2 \|h_0 - \Pi_n h_0\|_s^2 & \geq \sum_{j > k(n)} |\langle h_0 - \Pi_n h_0, q_j \rangle_s|^2 \\ & \geq c \left(\min_{j > k(n)} b_j^{-1} \right) \sum_{j > k(n)} b_j |\langle h_0 - \Pi_n h_0, q_j \rangle_s|^2 \\ & \geq \frac{c'}{b_{k(n)}} \|h_0 - \Pi_n h_0\|^2 \end{aligned}$$

for some finite positive constants c_2 , c , and c' . Result (iii) directly follows from results (i) and (ii). *Q.E.D.*

PROOF OF LEMMA B.3: Denote $b_j = \varphi(\nu_j^{-2})$. For any $h \in \mathcal{H}_{os}$ with $\|h\|^2 \leq O(\delta^2)$ and for any $k \geq 1$, Assumptions 5.3 and 5.4(i) imply that there are finite positive constants c_1 and c such that

$$\begin{aligned} c_1 \|h\|_s^2 &\leq \sum_{j \leq k} \langle h, q_j \rangle_s^2 + \sum_{j > k} \langle h, q_j \rangle_s^2 \\ &\leq \left(\max_{j \leq k} b_j^{-1} \right) \sum_j b_j \langle h, q_j \rangle_s^2 + M^2(\nu_{k+1})^{-2\alpha} \\ &\leq \frac{1}{c} b_k^{-1} \delta^2 + M^2(\nu_{k+1})^{-2\alpha}. \end{aligned}$$

Given that $M > 0$ is a fixed finite number and δ is small, we can assume $M^2(\nu_2)^{-2\alpha} > \frac{1}{c} \delta^2 / b_1$. Since $\{b_j\}$ is nonincreasing and $\{\nu_j\}_{j=1}^\infty$ is strictly increasing in $j \geq 1$, we have that there is a $k^* \equiv k^*(\delta) \in (1, \infty)$ such that

$$\frac{\delta^2}{b_{k^*-1}} < cM^2(\nu_{k^*})^{-2\alpha} \quad \text{and} \quad \frac{\delta^2}{b_{k^*}} \geq cM^2(\nu_{k^*})^{-2\alpha} \geq cM^2(\nu_{k^*+1})^{-2\alpha}$$

and

$$\omega(\delta, \mathcal{H}_{os}) \equiv \sup_{h \in \mathcal{H}_{os}: \|h - h_0\| \leq \delta} \|h - h_0\|_s \leq \text{const.} \frac{\delta}{\sqrt{b_{k^*}}};$$

thus result (i) holds. Result (ii) follows from Lemma B.2 and result (i). *Q.E.D.*

F. PROOFS OF LEMMAS FOR SERIES LS ESTIMATION OF $m(\cdot)$

Denote $\tilde{m}(X, h) \equiv P^{J_n}(X)'(P'P)^{-1}P'm(h)$ and $m(h) = (m(X_1, h), \dots, m(X_n, h))'$.

LEMMA SM.1: *Let Assumptions C.1 and C.2(i) hold. Then there are finite constants $c, c' > 0$ such that, w.p.a.1,*

$$\begin{aligned} cE_X[\|\tilde{m}(X, h)\|_{\tilde{W}}^2] &\leq \frac{1}{n} \sum_{i=1}^n \|\tilde{m}(X_i, h)\|_{\tilde{W}}^2 \\ &\leq c'E_X[\|\tilde{m}(X, h)\|_{\tilde{W}}^2] \quad \text{uniformly in } h \in \mathcal{H}_{k(n)}^{M_0}. \end{aligned}$$

PROOF: Denote $\langle g, \bar{g} \rangle_{n,X} \equiv \frac{1}{n} \sum_{i=1}^n g(X_i) \bar{g}(X_i)$ and $\langle g, \bar{g} \rangle_X \equiv E_X[g(X) \times \bar{g}(X)]$, where $g(X)$ and $\bar{g}(X)$ are square integrable functions of X . We want to show that for all $t > 0$,

$$(SM.5) \quad \lim_{n \rightarrow \infty} \Pr \left(\sup_{h \in \mathcal{H}_{k(n)}^P} \left| \frac{\langle \tilde{m}(\cdot, h), \tilde{m}(\cdot, h) \rangle_{n,X} - \langle \tilde{m}(\cdot, h), \tilde{m}(\cdot, h) \rangle_X}{\langle \tilde{m}(\cdot, h), \tilde{m}(\cdot, h) \rangle_X} \right| > t \right) = 0.$$

Let $G_n \equiv \{g : g(x) = \sum_{k=1}^{J_n} \pi_k p_k(x); \pi_k \in \mathcal{R}, \sup_x |g(x)| < \infty\}$. By construction $\tilde{m}(X, h) = \arg \min_{g \in G_n} n^{-1} \sum_{i=1}^n \|m(X_i, h) - g(X_i)\|_I^2$, so $\tilde{m}(X, h) \in G_n$ and

$$\begin{aligned} & \sup_{h \in \mathcal{H}_{k(n)}^P} \left| \frac{\langle \tilde{m}(\cdot, h), \tilde{m}(\cdot, h) \rangle_{n,X} - \langle \tilde{m}(\cdot, h), \tilde{m}(\cdot, h) \rangle_X}{\langle \tilde{m}(\cdot, h), \tilde{m}(\cdot, h) \rangle_X} \right| \\ & \leq \sup_{g \in G_n : \|g\|_X = 1} |\langle g, g \rangle_{n,X} - \langle g, g \rangle_X|. \end{aligned}$$

Define $A_n \equiv \sup_{g \in G_n} \frac{\sup_x |g(x)|}{\sqrt{E[(g(X))^2]}}$. Then, under Assumption C.1(i)–(iii) and the definition of G_n , we have $A_n \asymp \xi_n$. Thus, by Assumption C.1(iv), Lemma 4 of Huang (1998) for general linear sieves $\{p_k\}_{k=1}^{J_n}$, and Corollary 3 of Huang (2003) for polynomial spline sieves, equation (SM.5) holds. So with $t = 0.5$, we obtain that uniformly over $h \in \mathcal{H}_{k(n)}^{M_0}$,

$$0.5E_X[\|\tilde{m}(X, h)\|_I^2] \leq \frac{1}{n} \sum_{i=1}^n \|\tilde{m}(X_i, h)\|_I^2 \leq 2E_X[\|\tilde{m}(X, h)\|_I^2]$$

except for an event w.p.a.0. By Assumption C.1(v), there are finite constants $K, K' > 0$ such that $KI \leq W(X) \leq K'I$ for almost all X . Thus, $K\|\tilde{m}(X, h)\|_I^2 \leq \|\tilde{m}(X, h)\|_W^2 \leq K'\|\tilde{m}(X, h)\|_I^2$ for almost all X . Also by Assumption C.1(v), uniformly over $h \in \mathcal{H}_{k(n)}$,

$$\begin{aligned} \|\tilde{m}(X, h)\|_W^2 &= \tilde{m}(X, h)' \{\widehat{W}(X) - W(X) + W(X)\} \tilde{m}(X, h) \\ &\leq \sup_{x \in \mathcal{X}} |\widehat{W}(x) - W(x)| \times \|\tilde{m}(X, h)\|_I^2 + \|\tilde{m}(X, h)\|_W^2 \\ &\leq (K' + o_p(1)) \|\tilde{m}(X, h)\|_I^2. \end{aligned}$$

Similarly,

$$\|\tilde{m}(X, h)\|_W^2 \geq (K - o_p(1)) \|\tilde{m}(X, h)\|_I^2.$$

Note that for n large, $\min\{K', K\} \pm o_p(1) > 0$. Therefore, uniformly over $h \in \mathcal{H}_{k(n)}^{M_0}$,

$$\begin{aligned} \text{const.} \times E_X[\|\tilde{m}(X, h)\|_W^2] &\leq \frac{1}{n} \sum_{i=1}^n \|\tilde{m}(X_i, h)\|_W^2 \\ &\leq \text{const.}' \times E_X[\|\tilde{m}(X, h)\|_W^2] \end{aligned}$$

except for a set w.p.a.0.

Q.E.D.

PROOF OF LEMMA C.1: By Assumption C.1(i) and (v) it suffices to establish the results for $W = I$. Result (i) directly follows from our Assumption C.1 and Lemma A.1 Part (C) of Ai and Chen (2003).

Result (iii) can be established in the same way as that of result (ii). For result (ii), let $\varepsilon(Z, h) \equiv \rho(Z, h) - m(X, h)$ and $\varepsilon(h) \equiv (\varepsilon(Z_1, h), \dots, \varepsilon(Z_n, h))'$. For any symmetric and positive matrix Ω ($d \times d$), we have the spectral decomposition $\Omega = U\Lambda U'$, where $\Lambda = \text{diag}\{\lambda_1, \dots, \lambda_d\}$ with $\lambda_i > 0$ and $UU' = I_d$. Denote $\lambda_{\min}(\Omega)$ as the smallest eigenvalue of the matrix Ω . By definition, we have

$$\begin{aligned} &\sup_{h \in \mathcal{H}_{k(n)}^{M_0}} \frac{1}{n} \sum_{i=1}^n \|\hat{m}(X_i, h) - \tilde{m}(X_i, h)\|_I^2 \\ &= \sup_{h \in \mathcal{H}_{k(n)}^{M_0}} \frac{1}{n} \sum_{i=1}^n \text{Tr}\{p^{J_n}(X_i)'(P'P)^{-1}P'\varepsilon(h)\varepsilon(h)'P(P'P)^{-1}p^{J_n}(X_i)\} \\ &= \sup_{h \in \mathcal{H}_{k(n)}^{M_0}} \frac{1}{n} \sum_{i=1}^n \text{Tr}\{\varepsilon(h)'P(P'P)^{-1}p^{J_n}(X_i)p^{J_n}(X_i)'(P'P)^{-1}P'\varepsilon(h)\} \\ &= \sup_{h \in \mathcal{H}_{k(n)}^{M_0}} \frac{1}{n} \text{Tr}\left\{\varepsilon(h)'P(P'P)^{-1} \sum_{i=1}^n \{p^{J_n}(X_i)p^{J_n}(X_i)'\}(P'P)^{-1}P'\varepsilon(h)\right\} \\ &= \sup_{h \in \mathcal{H}_{k(n)}^{M_0}} \frac{1}{n} \text{Tr}\{\varepsilon(h)'P(P'P)^{-1}P'\varepsilon(h)\} \\ &= \sup_{h \in \mathcal{H}_{k(n)}^{M_0}} \frac{1}{n^2} \text{Tr}\{\varepsilon(h)'P(P'P/n)^{-1}P'\varepsilon(h)\} \\ &\leq (\lambda_{\min}(P'P/n))^{-1} \times \sup_{\mathcal{H}_{k(n)}^{M_0}} \frac{1}{n^2} \text{Tr}\{\varepsilon(h)'PP'\varepsilon(h)\}. \end{aligned}$$

Note that

$$\varepsilon(h)'PP'\varepsilon(h) = \sum_{j=1}^{J_n} \left(\left| \sum_{i=1}^n p_j(X_i) \varepsilon(Z_i, h) \right| \right)^2.$$

Let $r_n = \frac{J_n}{n} C_n$. We have, for all $M \geq 1$,

$$\begin{aligned} & \Pr \left(\sup_{h \in \mathcal{H}_{k(n)}^{M_0}} \frac{1}{n} \sum_{i=1}^n \|\widehat{m}(X_i, h) - \widetilde{m}(X_i, h)\|_T^2 > Mr_n \right) \\ & \leq \Pr \left((\lambda_{\min}(P'P/n))^{-1} \right. \\ & \quad \times \left. \sup_{h \in \mathcal{H}_{k(n)}^{M_0}} \sum_{j=1}^{J_n} \left(\left| \frac{1}{n} \sum_{i=1}^n p_j(X_i) \varepsilon(Z_i, h) \right| \right)^2 > Mr_n \right) \\ & \leq \Pr \left((\lambda_{\min}(P'P/n))^{-1} \right. \\ & \quad \times \left. \sum_{j=1}^{J_n} \left(\sup_{h \in \mathcal{H}_{k(n)}^{M_0}} \left| \frac{1}{n} \sum_{i=1}^n p_j(X_i) \varepsilon(Z_i, h) \right| \right)^2 > Mr_n \right). \end{aligned}$$

Following Newey (1997, p. 162) and under Assumption C.1(i)–(iv), we have: $(\lambda_{\min}(P'P/n))^{-1} = O_p(1)$. Thus, to bound $\Pr(\sup_{h \in \mathcal{H}_{k(n)}^{M_0}} n^{-1} \sum_{i=1}^n \|\widehat{m}(X_i, h) - \widetilde{m}(X_i, h)\|_T^2 > Mr_n)$, it suffices to bound the probability

$$\begin{aligned} & \Pr \left(\sum_{j=1}^{J_n} \left(\sup_{h \in \mathcal{H}_{k(n)}^{M_0}} \left| \frac{1}{n} \sum_{i=1}^n p_j(X_i) \varepsilon(Z_i, h) \right| \right)^2 > Mr_n \right) \\ & \leq \frac{1}{Mr_n} E_{Z^n} \left[\sum_{j=1}^{J_n} \left(\sup_{h \in \mathcal{H}_{k(n)}^{M_0}} \left| \frac{1}{n} \sum_{i=1}^n p_j(X_i) \varepsilon(Z_i, h) \right| \right)^2 \right] \\ & \leq \frac{J_n}{nr_n M} \max_{1 \leq j \leq J_n} E_{Z^n} \left[\left(\sup_{h \in \mathcal{H}_{k(n)}^{M_0}} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^n p_j(X_i) \varepsilon(Z_i, h) \right| \right)^2 \right], \end{aligned}$$

where the first inequality is by Markov inequality and $E_{Z^n}(\cdot)$ denotes the expectation with respect to $Z^n \equiv (Z_1, \dots, Z_n)$. By Theorem 2.14.5 in Van der Vaart

and Wellner (1996) (VdV-W; also see Pollard (1990)), we have

$$\begin{aligned} & \max_{1 \leq j \leq J_n} E_{Z^n} \left[\left(\sup_{h \in \mathcal{H}_{k(n)}^{M_0}} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^n p_j(X_i) \varepsilon(Z_i, h) \right| \right)^2 \right] \\ & \leq \max_{1 \leq j \leq J_n} \left(E_{Z^n} \left[\sup_{h \in \mathcal{H}_{k(n)}^{M_0}} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^n p_j(X_i) \varepsilon(Z_i, h) \right| \right] \right. \\ & \quad \left. + \sqrt{E[|p_j(X) \bar{\rho}_n(Z)|^2]} \right)^2. \end{aligned}$$

By Assumption C.2(i) and $\max_{1 \leq j \leq J_n} E[|p_j(X)|^2] \leq \text{const.}$, we have

$$\max_{1 \leq j \leq J_n} E[|p_j(X) \bar{\rho}_n(Z)|^2] \leq \text{const.} < \infty.$$

By Theorem 2.14.2 in VdV-W, we have (up to some constant)

$$\begin{aligned} & \max_{1 \leq j \leq J_n} E_{Z^n} \left[\sup_{h \in \mathcal{H}_{k(n)}^{M_0}} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^n p_j(X_i) \varepsilon(Z_i, h) \right| \right] \\ & \leq \max_{1 \leq j \leq J_n} \left\{ \sqrt{E[|p_j(X) \bar{\rho}_n(Z)|^2]} \right. \\ & \quad \left. \times \int_0^1 \sqrt{1 + \log N_{[]}(\omega K, \mathcal{E}_{jn}, \|\cdot\|_{L^2(f_Z)})} d\omega \right\} \\ & \leq K \max_{1 \leq j \leq J_n} \int_0^1 \sqrt{1 + \log N_{[]}(\omega K, \mathcal{E}_{jn}, \|\cdot\|_{L^2(f_Z)})} d\omega, \end{aligned}$$

where $\mathcal{E}_{jn} \equiv \{p_j(\cdot) \varepsilon(\cdot, h) : h \in \mathcal{H}_{k(n)}^{M_0}\}$. Note that for any $h, h' \in \mathcal{H}_{k(n)}^{M_0}$, we have

$$\begin{aligned} |p_j(X)(\varepsilon(Z, h) - \varepsilon(Z, h'))| & \leq |p_j(X)| \{|\rho(Z, h) - \rho(Z, h')| \\ & \quad + |E[\rho(Z, h)|X] - E[\rho(Z, h')|X]|\} \end{aligned}$$

and

$$\begin{aligned} & |p_j(X)| |E[\rho(Z, h)|X] - E[\rho(Z, h')|X]| \\ & \leq |p_j(X)| E[|\rho(Z, h) - \rho(Z, h')||X]. \end{aligned}$$

Recall that $\mathcal{O}_{j_n} \equiv \{p_j(\cdot)\rho(\cdot, h) : h \in \mathcal{H}_{k(n)}^{M_0}\}$ and that

$$\max_{1 \leq j \leq J_n} \int_0^1 \sqrt{1 + \log N_{[]}(\omega K, \mathcal{O}_{j_n}, \|\cdot\|_{L^2(f_Z)})} d\omega \leq \sqrt{C_n} < \infty$$

by Assumption C.2(iii). We have:

$$\max_{1 \leq j \leq J_n} \int_0^1 \sqrt{1 + \log N_{[]}(\omega K, \mathcal{E}_{j_n}, \|\cdot\|_{L^2(f_Z)})} d\omega \leq \text{const.} \times \sqrt{C_n} < \infty$$

and hence

$$\max_{1 \leq j \leq J_n} E_{Z^n} \left[\sup_{h \in \mathcal{H}_{k(n)}^{M_0}} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^n p_j(X_i) \varepsilon(Z_i, h) \right| \right] \leq \text{const.} \times \sqrt{C_n}.$$

It then follows that

$$\begin{aligned} & \frac{J_n}{nr_n M} \max_{1 \leq j \leq J_n} E_{Z^n} \left[\left(\sup_{h \in \mathcal{H}_{k(n)}^{M_0}} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^n p_j(X_i) \varepsilon(Z_i, h) \right| \right)^2 \right] \\ & \leq \text{const.} \times \frac{J_n C_n}{nr_n M}, \end{aligned}$$

so $r_n = \frac{J_n}{n} C_n$ and letting $M \rightarrow \infty$, the desired result follows. *Q.E.D.*

PROOF OF LEMMA C.2: The proofs of results (i) and (iii) are the same as that of result (ii). For result (ii), by the fact $(a - b)^2 + b^2 \geq \frac{1}{2}a^2$, we have that uniformly over $h \in \mathcal{H}_{k(n)}$,

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n \|\widehat{m}(X_i, h)\|_{\widehat{w}}^2 & \geq \frac{1}{2} \frac{1}{n} \sum_{i=1}^n \|\widetilde{m}(X_i, h)\|_{\widehat{w}}^2 \\ & \quad - \frac{1}{n} \sum_{i=1}^n \|\widehat{m}(X_i, h) - \widetilde{m}(X_i, h)\|_{\widehat{w}}^2. \end{aligned}$$

By Lemma SM.1, there is a finite constant $c > 0$ such that, w.p.a.1 and uniformly over $h \in \mathcal{H}_{k(n)}^{M_0}$,

$$\begin{aligned} & \frac{1}{n} \sum_{i=1}^n \|\widehat{m}(X_i, h)\|_{\widehat{w}}^2 \\ & \geq \frac{c}{2} E_X [\|\widetilde{m}(X, h)\|_{\widehat{w}}^2] - \frac{1}{n} \sum_{i=1}^n \|\widehat{m}(X_i, h) - \widetilde{m}(X_i, h)\|_{\widehat{w}}^2 \end{aligned}$$

$$\begin{aligned}
&\geq \frac{c}{4} E_X [\|m(X, h)\|_W^2] - \left(\frac{c}{2} E_X [\|m(X, h) - \tilde{m}(X, h)\|_W^2] \right. \\
&\quad \left. + \frac{1}{n} \sum_{i=1}^n \|\widehat{m}(X_i, h) - \tilde{m}(X_i, h)\|_{\widehat{W}}^2 \right) \\
&\geq K E_X [\|m(X, h)\|_W^2] - O_p \left(b_{m, J_n}^2 + \frac{J_n}{n} C_n \right),
\end{aligned}$$

where the second inequality is due to the fact that $(a - b)^2 + b^2 \geq \frac{1}{2}a^2$ and the last inequality is due to Lemma C.1, Assumption C.2(ii), and $\frac{c}{4} \equiv \bar{K} > 0$.

Similarly, by the fact $(a + b)^2 \leq 2a^2 + 2b^2$, we have that uniformly over $h \in \mathcal{H}_{k(n)}$,

$$\begin{aligned}
\frac{1}{n} \sum_{i=1}^n \|\widehat{m}(X_i, h)\|_{\widehat{W}}^2 &\leq 2 \frac{1}{n} \sum_{i=1}^n \|\tilde{m}(X_i, h)\|_{\widehat{W}}^2 \\
&\quad + 2 \frac{1}{n} \sum_{i=1}^n \|\widehat{m}(X_i, h) - \tilde{m}(X_i, h)\|_{\widehat{W}}^2.
\end{aligned}$$

By Lemma SM.1, there is a finite constant $c' > 0$ such that, w.p.a.1 and uniformly over $h \in \mathcal{H}_{k(n)}^{M_0}$,

$$\begin{aligned}
&\frac{1}{n} \sum_{i=1}^n \|\widehat{m}(X_i, h)\|_{\widehat{W}}^2 \\
&\leq 2c' E_X [\|\tilde{m}(X, h)\|_W^2] + 2 \frac{1}{n} \sum_{i=1}^n \|\widehat{m}(X_i, h) - \tilde{m}(X_i, h)\|_{\widehat{W}}^2 \\
&\leq 4c' E_X [\|m(X, h)\|_W^2] + \left(4c' E_X [\|\tilde{m}(X, h) - m(X, h)\|_W^2] \right. \\
&\quad \left. + \frac{2}{n} \sum_{i=1}^n \|\widehat{m}(X_i, h) - \tilde{m}(X_i, h)\|_{\widehat{W}}^2 \right) \\
&\leq K' E_X [\|m(X, h)\|_W^2] + O_p \left(b_{m, J_n}^2 + \frac{J_n}{n} C_n \right),
\end{aligned}$$

where the second inequality is again due to the fact $(a + b)^2 \leq 2a^2 + 2b^2$, and the last inequality is due to Lemma C.1, Assumption C.2(ii), and $4c' \equiv K' < \infty$. *Q.E.D.*

PROOF OF LEMMA C.3: By Assumption C.1(i) and (v), it suffices to establish the results for $W = I$. Using the same notation and following the steps as in the

proof of Lemma C.1, we obtain

$$\begin{aligned}
& \sup_{h \in \mathcal{N}_{osn}} \frac{1}{n} \sum_{i=1}^n \|\widehat{m}(X_i, h) - \widehat{m}(X_i, h_0) - \widetilde{m}(X_i, h)\|_I^2 \\
&= \sup_{h \in \mathcal{N}_{osn}} \frac{1}{n^2} \text{Tr}\{[\varepsilon(h) - \varepsilon(h_0)]' P(P'P/n)^{-1} P'[\varepsilon(h) - \varepsilon(h_0)]\} \\
&\leq (\lambda_{\min}(P'P/n))^{-1} \\
&\quad \times \sup_{h \in \mathcal{N}_{osn}} \frac{1}{n^2} \text{Tr}\{[\varepsilon(h) - \varepsilon(h_0)]' P P' [\varepsilon(h) - \varepsilon(h_0)]\} \\
&= (\lambda_{\min}(P'P/n))^{-1} \\
&\quad \times \sup_{h \in \mathcal{N}_{osn}} \frac{1}{n^2} \sum_{j=1}^{J_n} \left(\left| \sum_{i=1}^n p_j(X_i) [\varepsilon(Z_i, h) - \varepsilon(Z_i, h_0)] \right| \right)^2.
\end{aligned}$$

Let $r_n = \frac{J_n}{n} (\delta_{s,n})^{2\kappa}$. For all $M \geq 1$, to bound

$$\Pr \left(\sup_{h \in \mathcal{N}_{osn}} \frac{1}{n} \sum_{i=1}^n \|\widehat{m}(X_i, h) - \widehat{m}(X_i, h_0) - \widetilde{m}(X_i, h)\|_I^2 > M r_n \right),$$

it suffices to bound the probability

$$\begin{aligned}
& \Pr \left(\sum_{j=1}^{J_n} \left(\sup_{h \in \mathcal{N}_{osn}} \left| \frac{1}{n} \sum_{i=1}^n p_j(X_i) [\varepsilon(Z_i, h) - \varepsilon(Z_i, h_0)] \right| \right)^2 > M r_n \right) \\
&\leq \frac{J_n}{n r_n M} \\
&\quad \times \max_{1 \leq j \leq J_n} E_{Z^n} \left[\left(\sup_{h \in \mathcal{N}_{osn}} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^n p_j(X_i) [\varepsilon(Z_i, h) - \varepsilon(Z_i, h_0)] \right| \right)^2 \right].
\end{aligned}$$

Let $\Delta \varepsilon(Z_i, h) \equiv \varepsilon(Z_i, h) - \varepsilon(Z_i, h_0)$. By Theorem 2.14.5 in VdV-W, we have

$$\begin{aligned}
& \max_{1 \leq j \leq J_n} E_{Z^n} \left[\left(\sup_{h \in \mathcal{N}_{osn}} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^n p_j(X_i) \Delta \varepsilon(Z_i, h) \right| \right)^2 \right] \\
&\leq \max_{1 \leq j \leq J_n} \left(E_{Z^n} \left[\sup_{h \in \mathcal{N}_{osn}} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^n p_j(X_i) \Delta \varepsilon(Z_i, h) \right| \right] \right)^2
\end{aligned}$$

$$+ \sqrt{E \left[\sup_{h \in \mathcal{N}_{osn}} |p_j(X) \Delta \varepsilon(Z, h)|^2 \right]} \Big)^2.$$

By Jensen's inequality,

$$\begin{aligned} & E \left[\sup_{h \in \mathcal{N}_{osn}} |p_j(X) \{m(X, h) - m(X, h_0)\}|^2 \right] \\ & \leq E \left[\sup_{h \in \mathcal{N}_{osn}} |p_j(X) \{\rho(Z, h) - \rho(Z, h_0)\}|^2 \right]. \end{aligned}$$

Hence

$$\begin{aligned} & \max_{1 \leq j \leq J_n} \sqrt{E \left[\sup_{h \in \mathcal{N}_{osn}} |p_j(X) \Delta \varepsilon(Z, h)|^2 \right]} \\ & \leq \max_{1 \leq j \leq J_n} \sqrt{2E \left[\sup_{h \in \mathcal{N}_{osn}} |p_j(X) \{\rho(Z, h) - \rho(Z, h_0)\}|^2 \right]} \\ & \leq \text{const.} \times (\delta_{s,n})^\kappa \end{aligned}$$

by condition (i) in Lemma C.3.

By Theorem 2.14.2 in VdV-W, Remark C.1, and conditions (i) and (ii) of Lemma C.3, we have (up to some constant)

$$\begin{aligned} & \max_{1 \leq j \leq J_n} E_{Z^n} \left[\sup_{h \in \mathcal{N}_{osn}} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^n p_j(X_i) \Delta \varepsilon(Z_i, h) \right| \right] \\ & \leq \max_{1 \leq j \leq J_n} \left\{ (\delta_{s,n})^\kappa \int_0^1 (1 + \log N_{[]} (w(\delta_{s,n})^\kappa, \right. \\ & \quad \left. \{p_j(\cdot) \Delta \varepsilon(\cdot, h) : h \in \mathcal{N}_{osn}\}, \|\cdot\|_{L^2(f_Z)}) \right)^{1/2} dw \Big\} \\ & \leq (\delta_{s,n})^\kappa \int_0^1 \sqrt{1 + \log N(w^{1/\kappa}, \mathcal{N}_{osn}, \|\cdot\|_s)} dw \leq \text{const.} \times (\delta_{s,n})^\kappa. \end{aligned}$$

Hence

$$\max_{1 \leq j \leq J_n} E_{Z^n} \left[\left(\sup_{h \in \mathcal{N}_{osn}} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^n p_j(X_i) \Delta \varepsilon(Z_i, h) \right| \right)^2 \right] = O((\delta_{s,n})^{2\kappa}).$$

The desired result follows.

Q.E.D.

G. APPLICATION: PROOFS OF PROPOSITIONS

PROOF OF PROPOSITION 6.1: We obtain the result by verifying that all the assumptions of Theorem 3.2 (lower semicompact penalty) are satisfied with $\widehat{W} = W = I$.

First, Assumption 3.1(i) is trivially satisfied with $W = I$. For any $h \in \mathcal{H}$, we denote $h(y_1, y_2) = h_1(y_1) + h_2(y_2)$, $\Delta h(y_1, y_2) = h(y_1, y_2) - h_0(y_1, y_2) = \Delta h_1(y_1) + \Delta h_2(y_2)$, and $\Delta h_l(y_l) = h_l(y_l) - h_{0l}(y_l)$ for $l = 1, 2$. By the mean value theorem, Condition 6.1(iv), and the definitions of $K_{l,h}[\Delta h_l](X)$, we have

$$\begin{aligned}
 (\text{SM.6}) \quad m(X, h) - m(X, h_0) &= E[F_{Y_3|Y_1, Y_2, X}(h_1(Y_1) + h_2(Y_2)) \\
 &\quad - F_{Y_3|Y_1, Y_2, X}(h_{01}(Y_1) + h_{02}(Y_2)) | X] \\
 &= E\left[\left\{\int_0^1 f_{Y_3|Y_1, Y_2, X}(h_0(Y_1, Y_2) + t\Delta h(Y_1, Y_2)) dt\right\}\right. \\
 &\quad \left.\times [\Delta h_1(Y_1) + \Delta h_2(Y_2)] \Big| X\right] \\
 &= K_{1,h}[\Delta h_1](X) + K_{2,h}[\Delta h_2](X).
 \end{aligned}$$

Therefore, for any $h \in \mathcal{H}$ such that $m(X, h) - m(X, h_0) = 0$ almost surely X , under Condition 6.2(ii), we have $K_{1,h}[\Delta h_1](X) = 0$, $K_{2,h}[\Delta h_2](X) = 0$ almost surely X , which implies $\Delta h_l = 0$ almost surely Y_l for $l = 1, 2$ (by Condition 6.2(ii)). Thus, the identification Assumption 3.1(ii) holds. Given our choices of \mathcal{H} and \mathcal{H}_n (Condition 6.2(i) and (iii)), and $\|h\|_s = \|h\|_{\text{sup}} = \sup_{y_1} |h_1(y_1)| + \sup_{y_2} |h_2(y_2)|$, the sieve space \mathcal{H}_n is closed and we have, for $h_0 \in \mathcal{H}$, that there is $\Pi_n h_0 \in \mathcal{H}_n$ such that

$$\begin{aligned}
 \|h_0 - \Pi_n h_0\|_s &= \|h_0 - \Pi_n h_0\|_{\text{sup}} \\
 &\leq c\{k_1(n)\}^{-r_1} + c'\{k_2(n)\}^{-r_2} \\
 &= o(1), \quad \text{with } r_l = \alpha_l/d;
 \end{aligned}$$

thus Assumption 3.1(iii) holds. For any $h \in \mathcal{H}$ with $\Delta h(y_1, y_2) = \Delta h_1(y_1) + \Delta h_2(y_2)$ and $\Delta h_l(y_l) = h_l(y_l) - h_{0l}(y_l)$, $l = 1, 2$, equation (SM.6) implies that

$$\begin{aligned}
 |m(X, h) - m(X, h_0)| &\leq E\left[\sup_{t \in [0, 1]} f_{Y_3|Y_1, Y_2, X}(h_0(Y_1, Y_2) + t\Delta h(Y_1, Y_2)) \Big| X\right] \\
 &\quad \times \left[\sup_{y_1} |\Delta h_1(y_1)| + \sup_{y_2} |\Delta h_2(y_2)|\right].
 \end{aligned}$$

Since $m(X, h_0) = 0$ and by Condition 6.1(iv), we have

$$\begin{aligned} E[|m(X, h)|^2] &= E[|m(X, h) - m(X, h_0)|^2] \\ &\leq E\left[\left(\sup_{t \in [0, 1]} f_{Y_3|Y_1, Y_2, X}(h_0(Y_1, Y_2) + t\Delta h(Y_1, Y_2))\right)|X\right]^2 \\ &\quad \times (\|h - h_0\|_s)^2 \\ &\leq \text{const.} \times [\|h - h_0\|_s]^2. \end{aligned}$$

This and $\|\Pi_n h_0 - h_0\|_s = o(1)$ imply

$$\begin{aligned} E[|m(X, \Pi_n h_0)|^2] &\leq \text{const.} \|\Pi_n h_0 - h_0\|_s^2 \\ &\leq c\{k_1(n)\}^{-2r_1} + c'\{k_2(n)\}^{-2r_2} = o(1); \end{aligned}$$

hence Assumption 3.1(iv) holds. Assumption 3.2(b) directly follows from our choice of $\widehat{P}(\cdot) = P(\cdot)$.

Next, Condition 6.1(i) and (ii) and $\widehat{W} = W = I$ imply that Assumption C.1 holds. Assumption C.2(i) follows trivially with $\bar{\rho}_n(Z) \equiv 1$ since $\sup_{h \in \mathcal{H}} |\rho(Z, h)| \leq 1$. Condition 6.1(ii) and (iii) implies that Assumption C.2(ii) holds with $b_{m, J_n}^2 = J_n^{-2r_m}$. Thus Lemma C.2 result (i) is applicable and Assumption 3.3(i) is satisfied with $\eta_{0, n} = \frac{J_n}{n} + J_n^{-2r_m}$. This, $E([m(X, \Pi_n h_0)]^2) = O(\max[\{k_1(n)\}^{-2r_1}, \{k_2(n)\}^{-2r_2}])$, and $\max[\{k_1(n)\}^{-2r_1}, \{k_2(n)\}^{-2r_2}, \frac{J_n}{n} + J_n^{-2r_m}] = O(\lambda_n)$ together imply that Lemma A.3(i) holds. Moreover, it follows by our choice of penalty that $P(\Pi_n h_0) = O(1)$ and $P(\widehat{h}_n) = O_p(1)$. By our choice of sieves space it follows that $\log N(w^{1/2}, \mathcal{H}_{k(n)}^{M_0}, \|\cdot\|_{L^\infty}) \leq \min\{\frac{1}{2}k(n) \log(1/w), \text{const.}(1/w)^{d/2\alpha}\}$, where $\alpha \equiv \min\{\alpha_1, \alpha_2\} > 0$ (and const. can depend on M_0 , but not n); see, for example, Chen (2007) and Chen, Linton, and van Keilegom (2003). Following the verifications of Examples 1 and 2 in Chen, Linton, and van Keilegom (2003), we have that condition (18) in Remark C.1 holds with $\kappa = 1/2$. Hence, by Remark C.1 (with $\kappa = 1/2$), we have that Assumption C.2(iii) is satisfied with either $C_n \leq \text{const.} \times k(n)$ if $\alpha \leq d$ or $C_n = \text{const.} < \infty$ if $\alpha > d$. By Lemma C.2 result (ii) and the fact that $\frac{J_n k(n)}{n} = o(1)$, it follows that $\bar{\delta}_{m, n}^2 = o(1)$ and hence Assumption 3.3(ii) holds.

By the mean value theorem and Condition 6.1(iv), we have, for all $h, h' \in \mathcal{H}$, $\|h - h'\|_s = \|h - h'\|_{\text{sup}} = \sup_{y_1} |h_1(y_1) - h'_1(y_1)| + \sup_{y_2} |h_2(y_2) - h'_2(y_2)|$. Then

$$\begin{aligned} &|m(X, h) - m(X, h')| \\ &\leq E\left[\sup_{t \in [0, 1]} f_{Y_3|Y_1, Y_2, X}(h'(Y_1, Y_2) + t\{h - h'\}(Y_1, Y_2))\right] \|h - h'\|_s. \end{aligned}$$

This, Condition 6.1(iv), and $\sup_{x \in \mathcal{X}, h \in \mathcal{H}} |m(x, h)| \leq 1$ imply that

$$\begin{aligned} E[|m(X, h)|^2] - E[|m(X, h')|^2] &\leq 2E[|m(X, h) - m(X, h')|] \\ &\leq \text{const.} \times \|h - h'\|_s. \end{aligned}$$

Thus $E[|m(X, h)|^2]$ is continuous on $(\mathcal{H}, \|\cdot\|_s)$. We have that for any $M < \infty$, the embedding of the set $\{h \in \mathcal{H} : P(h) = \|h_1\|_{L^{\alpha_1}} + \|h_2\|_{L^{\alpha_2}} \leq M\}$ into \mathbf{H} is compact under the norm $\|\cdot\|_s = \|\cdot\|_{\text{sup}}$; hence $P(\cdot)$ is lower semicompact.

The condition $\max[\{k_1(n)\}^{-2r_1}, \{k_2(n)\}^{-2r_2}, \frac{J_n}{n} + J_n^{-2r_m}] = O(\lambda_n)$ and Theorem 3.2 now imply the desired consistency results. *Q.E.D.*

PROOF OF PROPOSITION 6.2: We obtain the results by verifying that all the assumptions of Corollary 5.1 are satisfied.

We first show that $\delta_{m,n} = \frac{J_n}{n} + b_{m,J_n}^2$. Similar to the [proof](#) of Proposition 6.1, $\log N(w^{1/2}, \mathcal{H}_{osn}, \|\cdot\|_{L^\infty}) \leq \min\{\frac{1}{2}k(n) \log(1/w), \text{const.}(1/w)^{d/2\alpha}\}$, where $\alpha = \min\{\alpha_1, \alpha_2\} > d$. By Remark C.1 (with $\kappa = 1/2$), we have that Assumption C.2(iv) is satisfied with $C < \infty$. Thus Lemma C.2 result (iii) is applicable and yields $\delta_{m,n}^2 = \frac{J_n}{n} + J_n^{-2r_m} = o(1)$.

Assumptions 3.1, 3.2, and 3.3 (with $\eta_{0,n} = \delta_{m,n}^2 = \frac{J_n}{n} + J_n^{-2r_m}$) are already verified in the [proof](#) of Proposition 6.1. Given the choice of the norm $\|h\|_s$, Assumption 5.1 is satisfied with $\|h_0 - \Pi_n h_0\|_s = O(\{k(n)\}^{-r})$ with $r = \alpha/d$. Condition 6.3(ii) implies Assumption 5.2. It remains to verify Assumption 4.1. By Condition 6.1(iv), we have

$$\begin{aligned} \frac{dm(X, h_0)}{dh}[h - h_0] &= T_{h_0}[h - h_0] \\ &= E[f_{Y_3|Y_1, Y_2, X}(h_{01}(Y_1) + h_{02}(Y_2)) \\ &\quad \times [h_1(Y_1) - h_{01}(Y_1) + h_2(Y_2) - h_{02}(Y_2)]|X], \\ \|h - h_0\|^2 &= E\left[\frac{dm(X, h_0)}{dh}[h - h_0]\right]^2 \leq \text{const.} \|h - h_0\|_s^2; \end{aligned}$$

hence Assumption 4.1(i) holds. Since

$$m(X, h) - m(X, h_0) = K_{1,h}[h_1 - h_{01}](X) + K_{2,h}[h_2 - h_{02}](X),$$

Condition 6.3(i) implies Assumption 4.1(ii). The results now follow from Corollary 5.1. *Q.E.D.*

REFERENCES

AI, C., AND X. CHEN (2003): "Efficient Estimation of Models With Conditional Moment Restrictions Containing Unknown Functions," *Econometrica*, 71, 1795–1844. [[17](#)]

- CHEN, X., O. LINTON, AND I. VAN KEILEGOM (2003): "Estimation of Semiparametric Models When the Criterion Function Is Not Smooth," *Econometrica*, 71, 1591–1608. [25]
- EDMUNDS, D., AND H. TRIEBEL (1996): *Function Spaces, Entropy Numbers, Differential Operators*. Cambridge: Cambridge University Press. [1]
- HUANG, J. (1998): "Projection Estimation in Multiple Regression With Application to Functional ANOVA Models," *The Annals of Statistics*, 26, 242–272. [16]
- (2003): "Local Asymptotics for Polynomial Spline Regression," *The Annals of Statistics*, 31, 1600–1635. [16]
- MEYER, Y. (1992): *Wavelets and Operators*. Cambridge, U.K.: Cambridge University Press. [3]
- NEWBY, W. K. (1997): "Convergence Rates and Asymptotic Normality for Series Estimators," *Journal of Econometrics*, 79, 147–168. [18]
- POLLARD, D. (1990): *Empirical Processes: Theory and Applications*. NSF–CBMS Regional Conference Series in Probability and Statistics, Vol. 2. Hayward, CA: Institute of Mathematical Statistics. [19]
- VAN DER VAART, A., AND J. WELLNER (1996): *Weak Convergence and Empirical Processes: With Applications to Statistics*. New York: Springer-Verlag. [19]
- YANG, Y., AND A. BARRON (1999): "Information-Theoretic Determination of Minimax Rates of Convergence," *The Annals of Statistics*, 27, 1564–1599. [12]
- ZEIDLER, E. (1985): *Nonlinear Functional Analysis and Its Applications III: Variational Methods and Optimization*. New York: Springer-Verlag. [5,6]

*Cowles Foundation for Research in Economics, Yale University, 30 Hillhouse,
Box 208281, New Haven, CT 06520, U.S.A.; xiaohong.chen@yale.edu*
and

*Department of Economics, UC at Berkeley, 508-1 Evans Hall 3880, Berkeley,
CA 94704-3880, U.S.A.; dpouzo@econ.berkeley.edu.*

Manuscript received April, 2008; final revision received December, 2010.