SUPPLEMENT TO "NONPARAMETRIC STOCHASTIC DISCOUNT FACTOR DECOMPOSITION"

(*Econometrica*, Vol. 85, No. 5, September 2017, 1501–1536)

TIMOTHY M. CHRISTENSEN
Department of Economics, New York University

THIS SUPPLEMENT CONTAINS sufficient conditions for several assumptions in Sections 3 and 4 and proofs of all results in the main text.

APPENDIX C: SOME SUFFICIENT CONDITIONS

This appendix presents sufficient conditions for Assumptions 3.3, 3.4(b), and 4.3 and bounds for the terms $\eta_{n,k}$ and $\eta^*_{n,k}$ in display (23) and $\nu_{n,k}$ in display (37). Proofs of results in this appendix are contained in the Online Appendix.

C.1. *Sufficient Conditions for Assumptions 3.3 and 3.4(b)*

We assume that the state process $X = \{X_t : t \in T\}$ is either beta-mixing or rho-mixing. The beta-mixing coefficient between two σ-algebras $\mathcal{A}$ and $\mathcal{B}$ is

$$2\beta(\mathcal{A}, \mathcal{B}) = \sup \sum_{(i,j)\in I\times J} \left|\mathbb{P}(A_i \cap B_j) - \mathbb{P}(A_i)\mathbb{P}(B_j)\right|$$

with the supremum taken over all $\mathcal{A}$-measurable finite partitions $\{A_i\}_{i\in I}$ and $\mathcal{B}$-measurable finite partitions $\{B_j\}_{j\in J}$. The beta-mixing coefficients of X are defined as

$$\beta_q = \sup_t \beta\big(\sigma(\dots, X_{t-1}, X_t), \sigma(X_{t+q}, X_{t+q+1}, \dots)\big).$$

We say that X is *exponentially beta-mixing* if $\beta_q \le Ce^{-cq}$ for some $C, c > 0$. The rho-mixing coefficients of X are defined as

$$\rho_q = \sup_{\psi\in L^2:\mathbb{E}[\psi]=0, \|\psi\|=1} \mathbb{E}\big[\mathbb{E}\big[\psi(X_{t+q})|X_t\big]^2\big]^{1/2}.$$

We say that X is *exponentially rho-mixing* if $\rho_q \le e^{-cq}$ for some $c > 0$.

We use the sequence $\xi_k = \sup_x \|\mathbf{G}^{-1/2}b^k(x)\|$ to bound convergence rates. When X has bounded rectangular support and Q has a density that is bounded away from 0 and ∞, ξ_k is known to be $O(\sqrt{k})$ for (tensor-product) spline, cosine, and certain wavelet bases and $O(k)$ for (tensor-product) polynomial series (see, e.g., Newey (1997), Chen and Christensen (2015)). It is also possible to derive alternative sufficient conditions in terms of higher moments of $\|\mathbf{G}^{-1/2}b^k(X_t)\|$ (instead of $\sup_x \|\mathbf{G}^{-1/2}b^k(x)\|$) by extending arguments in Hansen (2015) to accommodate weakly dependent data and asymmetric matrices.

Timothy M. Christensen: timothy.christensen@nyu.edu

© 2017 The Econometric Society DOI: 10.3982/ECTA11600

C.1.1. *Sufficient Conditions in Case 1*

The first result below uses an exponential inequality for weakly dependent random matrices from Chen and Christensen (2015). The second extends arguments from Gobet, Hoffmann, and Reiß (2004).

LEMMA C.1: *Let the following hold*:
(a) *X is exponentially beta-mixing*,
(b) $\mathbb{E}[m(X_t, X_{t+1})^r] < \infty$ *for some* $r > 2$,
(c) $\xi_k^{2+4/r}(\log n)^2/n = o(1)$.
Then:
(1) *Assumption* 3.3 *holds*.
(2) *We may take* $\eta_{n,k} = \eta^*_{n,k} = \xi_k^{1+2/r}(\log n)/\sqrt{n}$ *in display* (23).
(3) *If, in addition*, $\xi_k^{4+8/r}(\log n)^4/n = o(1)$, *then Assumption* 3.4(b) *holds*.

LEMMA C.2: *Let the following hold*:
(a) *X is exponentially rho-mixing*,
(b) $\mathbb{E}[m(X_t, X_{t+1})^r] < \infty$ *for some* $r > 2$,
(c) $\xi_k^{2+4/r}k/n = o(1)$.
Then:
(1) *Assumption* 3.3 *holds*.
(2) *We may take* $\eta_{n,k} = \eta^*_{n,k} = \xi_k^{1+2/r}/\sqrt{n}$ *in display* (23).
(3) *If, in addition*, $\xi_k^{4+8/r}k^2/n = o(1)$, *then Assumption* 3.4(b) *also holds*.

C.1.2. *Sufficient Conditions in Case 2 With Parametric First Stage*

The following lemma presents one set of sufficient conditions for Assumptions 3.3 and 3.4(b) when $\alpha_0 \in \mathcal{A} \subseteq \mathbb{R}^{d_\alpha}$ is a finite-dimensional parameter.

LEMMA C.3: *Let the conditions of Lemma* C.1 *hold for* $m(x_0, x_1) = m(x_0, x_1; \alpha_0)$, *and let*:
(a) $\|\hat{\alpha} - \alpha_0\| = O_p(n^{-1/2})$,
(b) *$m(x_0, x_1; \alpha)$ be continuously differentiable in α on a neighborhood N of α_0 for all $(x_0, x_1) \in \mathcal{X}^2$ and let there exist a function $\bar{m} : \mathcal{X}^2 \to \mathbb{R}$ with $\mathbb{E}[\bar{m}(X_t, X_{t+1})^2] < \infty$ such that*

$$\sup_{\alpha \in N}\left\| \frac{\partial m(x_0, x_1; \alpha)}{\partial \alpha} \right\| \le \bar{m}(x_0, x_1) \quad \textit{for all } (x_0, x_1) \in \mathcal{X}^2.$$

Then:
(1) *Assumption* 3.3 *holds*.
(2) *We may take* $\eta_{n,k} = \eta^*_{n,k} = \xi_k^{1+2/r}(\log n)/\sqrt{n}$ *in display* (23).
(3) *If, in addition*, $\xi_k^{4+8/r}(\log n)^4/n = o(1)$, *then Assumption* 3.4(b) *holds*.

The conditions on k and bounds for $\eta_{n,k}$ and $\eta^*_{n,k}$ are the same as Lemma C.1. Therefore, here first-stage estimation of α does not reduce the convergence rates of $\widehat{\mathbf{G}}$ and $\widehat{\mathbf{M}}$ relative to Case 1.

C.1.3. *Sufficient Conditions in Case 2 With Semi/Nonparametric First Stage*

We now present one set of sufficient conditions for Assumptions 3.3 and 3.4(b) when $\alpha_0 \in \mathcal{A} \subseteq \mathbb{A}$ is an infinite-dimensional parameter and the parameter space is $\mathcal{A} \subseteq \mathbb{A}$ (a Banach space) equipped with a norm $\|\cdot\|_{\mathcal{A}}$. This includes the case in which α is a function, that is, $\alpha = h$ with $\mathbb{A} = \mathbb{H}$ a function space, and the case in which α consists of both finite-dimensional and function parts, that is, $\alpha = (\theta, h)$ with $\mathbb{A} = \Theta \times \mathbb{H}$ where $\Theta \subseteq \mathbb{R}^{\dim(\theta)}$.

For each $\alpha \in \mathcal{A}$, we define $\mathbb{M}^{(\alpha)}$ as the operator $\mathbb{M}^{(\alpha)}\psi(x) = \mathbb{E}[m(X_t, X_{t+1}; \alpha)\psi(X_{t+1})| X_t = x]$ with the understanding that $\mathbb{M}^{(\alpha_0)} = \mathbb{M}$. Let $\mathcal{M} = \{m(x_0, x_1; \alpha) - m(x_0, x_1; \alpha_0) : \alpha \in \mathcal{A}\}$. We say $\mathcal{M}$ has an envelope function E if there exists some measurable $E : \mathcal{X}^2 \to [1, \infty)$ such that $|m(x_0, x_1)| \le E(x_0, x_1)$ for every $(x_0, x_1) \in \mathcal{X}$ and $m \in \mathcal{M}$. Let $\mathcal{M}^* = \{m/E : m \in \mathcal{M}\}$. The functions in $\mathcal{M}^*$ are clearly bounded by ± 1. Let $N_{[]}(u, \mathcal{M}^*, \|\cdot\|_p)$ denote the entropy with bracketing of $\mathcal{M}^*$ with respect to the L^p norm $\|\cdot\|_p$. Finally, let $\ell^*(\alpha) = \|\mathbb{M}^{(\alpha)} - \mathbb{M}\|$ and observe that $\ell^*(\alpha_0) = 0$.

LEMMA C.4: *Let the conditions of Lemma* C.1 *hold for* $m(x_0, x_1) = m(x_0, x_1; \alpha_0)$, *and let*:

(a) *$\mathcal{M}$ have envelope function E with $\|E\|_{4s} < \infty$ for some $s > 1$*,

(b) $\log N_{[]}(u, \mathcal{M}^*, \|\cdot\|_{\frac{4sv}{2s-v}}) \le C_{[]}u^{-2\zeta}$ *for some constants* $C_{[]} > 0$, $\zeta \in (0, 1)$ *and* $v \in (1, 2s)$,

(c) *$\ell^*(\alpha)$ is pathwise differentiable at α_0 with* $|\ell^*(\alpha) - \ell^*(\alpha_0) - \dot{\ell}^*_{\alpha_0}[\alpha - \alpha_0]| = O(\|\alpha - \alpha_0\|^2_{\mathcal{A}})$, $\|\hat{\alpha} - \alpha_0\|_{\mathcal{A}} = o_p(n^{-1/4})$, *and* $\sqrt{n}\dot{\ell}^*_{\alpha_0}[\hat{\alpha} - \alpha_0] = O_p(1)$,

(d) $\xi_k^{4-\frac{2s-v}{sv}}(k \log k)/n = o(1)$, $\xi_k^{\zeta\frac{2s-v}{2sv}} = O(\sqrt{k \log k})$, *and* $(\log n) = O(\xi_k^{1/3})$.

Then:

(1) *Assumption* 3.3 *holds*.

(2) *We may take* $\eta_{n,k} = \eta^*_{n,k} = \xi_k^{1+2/r}(\log n)/\sqrt{n} + \xi_k^{2-\frac{2s-v}{2sv}}\sqrt{(k \log k)/n}$ *in display* (23).

(3) *If, in addition*, $[\xi_k^{4+8/r}(\log n)^4 + \xi_k^{8-\frac{4s-2v}{sv}}(k \log k)^2]/n = o(1)$, *then Assumption* 3.4(b) *holds*.

Note that the condition $\xi_k^{\zeta\frac{2s-v}{2sv}} = O(\sqrt{k \log k})$ is trivially satisfied when $\xi_k = O(\sqrt{k})$.

C.2. *Sufficient Conditions for Assumption 4.3*

The following is one set of sufficient conditions for Assumption 4.3 assuming beta-mixing. Recall that $\xi_k = \sup_x \|\mathbf{G}^{-1/2}b^k(x)\|$.

LEMMA C.5: *Let the following hold*:

(a) *X is exponentially beta-mixing*,

(b) $\mathbb{E}[(G_{t+1}^{1-\gamma})^{2s}] < \infty$ *for some* $s > 1$,

(c) $[\xi_k^2(\log n)^2 + \xi_k^{2+2\beta}k]/n = o(1)$ *and* $(\log n)^{\frac{2s-1}{s-1}}k/n = o(1)$.

Then:

(1) *Assumption* 4.3 *holds*.

(2) *We may take* $\nu_{n,k} = \xi_k^{1+\beta}\sqrt{k/n} + \xi_k(\log n)/\sqrt{n}$ *in display* (37).

APPENDIX D: PROOFS OF RESULTS IN THE MAIN TEXT

Notation: For $v \in \mathbb{R}^k$, define

$$\|v\|^2_{\mathbf{G}} = v'\mathbf{G}_k v,$$

or equivalently, $\|v\|_{\mathbf{G}} = \|\mathbf{G}_k^{1/2} v\|$. For any matrix $\mathbf{A} \in \mathbb{R}^{k\times k}$, we define

$$\|\mathbf{A}\|_{\mathbf{G}} = \sup\{\|\mathbf{A}v\|_{\mathbf{G}} : v \in \mathbb{R}^k, \|v\|_{\mathbf{G}} = 1\}.$$

We also define the inner product weighted by $\mathbf{G}_k$, namely, $\langle u, v\rangle_{\mathbf{G}} = u'\mathbf{G}_k v$. The inner product $\langle\cdot,\cdot\rangle_{\mathbf{G}}$ and its norm $\|\cdot\|_{\mathbf{G}}$ are germane for studying convergence of the matrix estimators, as $(\mathbb{R}^k, \langle\cdot,\cdot\rangle_{\mathbf{G}})$ is isometrically isomorphic to $(B_k, \langle\cdot,\cdot\rangle)$. The notation $a_n \lesssim b_n$ for two positive sequences a_n and b_n means that there exists a finite positive constant C such that $a_n \le C b_n$ for all n sufficiently large; $a_n \asymp b_n$ means $a_n \lesssim b_n$ and $b_n \lesssim a_n$.

D.1. *Proofs of Results in Sections 2, 3, and 4*

PROOF OF PROPOSITION 2.1: Theorem V.6.6 of Schaefer (1974) implies, in view of Assumption 2.1, that $\rho := r(\mathbb{M}) > 0$ and that $\mathbb{M}$ has a unique positive eigenfunction $\phi \in L^2$ corresponding to ρ. Applying the result to $\mathbb{M}^*$ in place of $\mathbb{M}$ guarantees existence of $\phi^* \in L^2$. This proves part (a). Theorem V.6.6 of Schaefer (1974) also implies that ρ is isolated and the largest eigenvalue of $\mathbb{M}$. Theorem V.5.2(iii) of Schaefer (1974), in turn, implies that ρ is simple, completing the proof of part (c). Theorem V.5.2(iv) of Schaefer (1974) implies that ϕ is the unique positive solution to (6). The same result applied to $\mathbb{M}^*$ guarantees uniqueness of ϕ^*, proving part (b). Part (d) follows from Proposition F.3 in the Online Appendix. *Q.E.D.*

PROOF OF THEOREM 3.1: Immediate from Lemmas A.2 and A.4. *Q.E.D.*

PROOF OF COROLLARY 3.1: We first verify Assumption 3.2. By Theorem 12.8 of Schumaker (2007) and (ii)–(iv), for each $\psi \in L^2$ there exists a $h_k(\mathbb{M}\psi) \in B_k$ such that

$$\|\mathbb{M}\psi - h_k(\mathbb{M}\psi)\| \lesssim k^{-\bar{p}/d}\|\mathbb{M}\psi\|_{W^{\bar{p}}} \lesssim k^{-\bar{p}/d}\|\psi\|.$$

Therefore,

$$\begin{aligned}\|\mathbb{M}\psi - \Pi_k\mathbb{M}\psi\| &= \|\mathbb{M}\psi - h_k(\mathbb{M}\psi) + \Pi_k(h_k(\mathbb{M}\psi) - \mathbb{M}\psi)\| \\ &\le 2\|\mathbb{M}\psi - h_k(\mathbb{M}\psi)\| \lesssim k^{-\bar{p}/d}\|\psi\|,\end{aligned}$$

and so $\|\mathbb{M} - \Pi_k\mathbb{M}\| = O(k^{-\bar{p}/d}) = o(1)$ as required.

Similar arguments yield $\delta_k = O(k^{-p/d})$ and $\delta_k^* = O(k^{-p/d})$.

By Lemma C.2, conditions (iv)–(vii) are sufficient for Assumption 3.3 and we may take $\eta_{n,k} = \eta_{n,k}^* = k^{(r+2)/(2r)}/\sqrt{n}$. Choosing $k \asymp n^{\frac{rd}{2rp+(2+r)d}}$ balances bias and variance terms and we obtain the convergence rates as stated. *Q.E.D.*

PROOF OF REMARK 3.1: First observe that $\mathbb{M}\phi = \sum_{n=1}^\infty \mu_n\langle\phi, \varphi_n\rangle g_n$. Taking the inner product of both sides of $\mathbb{M}\phi = \rho\phi$ with g_n, we obtain $\mu_n\langle\phi,\varphi_n\rangle = \rho\langle\phi, g_n\rangle$ for each $n \in \mathbb{N}$. By Parseval's identity, $\|\phi\|^2 = \sum_{n\in\mathbb{N}}\langle\phi,\varphi_n\rangle^2 \ge \rho^2\sum_{n\in\mathbb{N}:\mu_n>0}\mu_n^{-2}\langle\phi, g_n\rangle^2$. Similarly, $\|\phi^*\|^2 \ge \rho^2\sum_{n\in\mathbb{N}:\mu_n>0}\mu_n^{-2}\langle\phi^*,\varphi_n\rangle^2$. Note that $\langle\phi, g_n\rangle = 0$ and $\langle\phi^*,\varphi_n\rangle = 0$ if $\mu_n = 0$.

As B_k spans the linear subspace in L^2 generated by $\{g_n\}_{n=1}^k$, we have $\phi_k := \sum_{n=1}^k \langle \phi, g_n \rangle g_n \in B_k$. Therefore, assuming $\mu_{k+1} > 0$ (else the result is trivially true):

$$\|\phi - \phi_k\|^2 = \sum_{n \geq k+1} \langle \phi, g_n \rangle^2 = \mu_{k+1}^2 \sum_{n \geq k+1} \frac{\langle \phi, g_n \rangle^2}{\mu_{k+1}^2} \leq \mu_{k+1}^2 \sum_{n \geq k+1 : \mu_n > 0} \frac{\langle \phi, g_n \rangle^2}{\mu_n^2} \leq \mu_{k+1}^2 \frac{\|\phi\|^2}{\rho^2}.$$

It follows that

$$\delta_k = \|\phi - \Pi_k \phi\| = \big\|\phi - \phi_k + \Pi_k(\phi_k - \phi)\big\| \leq 2\|\phi - \phi_k\| = O(\mu_{k+1}).$$

A similar argument gives $\delta_k^* = O(\mu_{k+1})$. *Q.E.D.*

Before proving Theorem 3.2, we first present a lemma that controls higher-order bias terms involving ϕ_k and ϕ_k^*. Define

$$\psi_{k,\rho}(x_0, x_1) = \phi_k^*(x_0) m(x_0, x_1) \phi_k(x_1) - \rho_k \phi_k^*(x_0) \phi_k(x_0)$$

with ϕ_k and ϕ_k^* normalized so that $\|\phi_k\| = 1$ and $\langle \phi_k, \phi_k^* \rangle = 1$, and

$$\Delta_{\psi,n,k} = \frac{1}{n} \sum_{t=0}^{n-1} \big(\psi_{\rho,k}(X_t, X_{t+1}) - \psi_\rho(X_t, X_{t+1})\big),$$

where ψ_ρ is from display (25).

To simplify notation, let $\phi_t = \phi(X_t)$, $\phi_t^* = \phi^*(X_t)$, $\phi_{k,t} = \phi_k(X_t)$, and $\phi_{k,t}^* = \phi_k^*(X_t)$.

LEMMA D.1: *Assumptions* 3.1 *and* 3.2 *hold*. *Then* $\Delta_{\psi,n,k} = O_p(\delta_k + \delta_k^*)$.

PROOF OF LEMMA D.1: First write

$$\begin{aligned}
\Delta_{\psi,n,k} &= \frac{1}{n} \sum_{t=0}^{n-1} \big(\phi_{k,t}^* - \phi_t^*\big) m(X_t, X_{t+1}) \phi_{k,t+1} + \frac{1}{n} \sum_{t=0}^{n-1} \phi_t^* m(X_t, X_{t+1}) (\phi_{k,t+1} - \phi_{t+1}) \\
&\quad - (\rho_k - \rho) \frac{1}{n} \sum_{t=0}^{n-1} \phi_{k,t}^* \phi_{k,t} - \rho \frac{1}{n} \sum_{t=0}^{n-1} \big(\phi_{k,t}^* \phi_{k,t} - \phi_t^* \phi_t\big) \\
&=: \widehat{T}_1 + \widehat{T}_2 + \widehat{T}_3 + \widehat{T}_4.
\end{aligned}$$

By iterated expectations,

$$\begin{aligned}
\mathbb{E}\big[\big|\big(\phi_{k,t}^* - \phi_t^*\big) m(X_t, X_{t+1}) \phi_{k,t+1}\big|\big] &= \big\langle \big|\phi_k^* - \phi^*\big|, \mathbb{M}\big(|\phi_k|\big)\big\rangle \\
&\leq \big\|\phi_k^* - \phi^*\big\| \|\mathbb{M}\| \|\phi_k\| = O\big(\delta_k^*\big)
\end{aligned}$$

using Cauchy–Schwarz, boundedness of $\mathbb{M}$ (Assumption 3.1), and Lemma A.2 (note that the normalizations $\langle \phi_k^*, \phi_k \rangle = 1$ and $\langle \phi, \phi^* \rangle = 1$ instead of $\|\phi_k^*\| = 1$ and $\|\phi^*\| = 1$ do not affect the conclusions of Lemma A.2). Markov's inequality then implies $\widehat{T}_1 = O_p(\delta_k^*)$.

Similarly,

$$\mathbb{E}\big[\big|\phi_t^* m(X_t, X_{t+1})(\phi_{k,t+1} - \phi_{t+1})\big|\big] = \big\langle \phi^*, \mathbb{M}\big(|\phi_k - \phi|\big)\big\rangle \le \|\phi^*\| \|\mathbb{M}\| \|\phi_k - \phi\| = O(\delta_k)$$

and so $\widehat{T}_2 = O_p(\delta_k)$.

Since $\rho_k - \rho = O(\delta_k)$ by Lemma A.2(a) and $\frac{1}{n}\sum_{t=0}^{n-1} \phi_{k,t}^* \phi_{k,t} = O_p(1)$ follows from Lemma A.2(b),(c), we obtain $\widehat{T}_3 = O_p(\delta_k)$. Finally,

$$\mathbb{E}\big[\big|\phi_{k,t}^* \phi_{k,t} - \phi_t^* \phi_t\big|\big] \le \|\phi_k^* - \phi^*\| \|\phi_k\| + \|\phi^*\| \|\phi_k - \phi\| = O\big(\delta_k + \delta_k^*\big)$$

again by Cauchy–Schwarz and Lemma A.2. Therefore, $\widehat{T}_4 = O_p(\delta_k + \delta_k^*)$. *Q.E.D.*

PROOF OF THEOREM 3.2: First note that

$$\begin{aligned} \sqrt{n}(\hat{\rho} - \rho) &= \sqrt{n}(\hat{\rho} - \rho_k) + \sqrt{n}(\rho_k - \rho) \\ &= \sqrt{n}(\hat{\rho} - \rho_k) + o(1) \\ &= \sqrt{n} c_k^{*\prime}(\widehat{\mathbf{M}} - \rho_k \widehat{\mathbf{G}}) c_k + o_p(1), \end{aligned} \tag{S.1}$$

where the second line is by Assumption 3.4(a) and the third line is by Lemma B.1 and Assumption 3.4(b) (under the normalizations $\|\mathbf{G} c_k\| = 1$ and $c_k^{*\prime} \mathbf{G} c_k = 1$). By identity, we may write the first term on the right-hand side of display (S.1) as

$$\begin{aligned} \sqrt{n} c_k^{*\prime}(\widehat{\mathbf{M}} - \rho_k \widehat{\mathbf{G}}) c_k &= \frac{1}{\sqrt{n}} \sum_{t=0}^{n-1} \psi_\rho(X_t, X_{t+1}) + \sqrt{n} \times \Delta_{\psi,n,k} \\ &= \frac{1}{\sqrt{n}} \sum_{t=0}^{n-1} \psi_\rho(X_t, X_{t+1}) + o_p(1), \end{aligned} \tag{S.2}$$

where the second line is by Lemma D.1 and Assumption 3.4(a). The result follows by substituting (S.2) into (S.1) and applying a CLT for stationary and ergodic martingale differences (e.g., Billingsley (1961)), which is valid in view of Assumption 3.4(c). *Q.E.D.*

PROOF OF THEOREM 3.3: This is a consequence of Theorem B.1 in Appendix B. *Q.E.D.*

PROOF OF THEOREM 3.4: Let $m_t(\alpha) = m(X_t, X_{t+1}; \alpha)$. By Assumption 3.4(a), Lemma B.1, and Assumption 3.4(b),

$$\begin{aligned} \sqrt{n}(\hat{\rho} - \rho) &= \frac{1}{\sqrt{n}} \sum_{t=0}^{n-1} \big(\phi_{k,t}^* \phi_{k,t+1} m(X_t, X_{t+1}, \hat{\alpha}) - \rho_k \phi_{k,t}^* \phi_{k,t}\big) + o_p(1) \\ &= \frac{1}{\sqrt{n}} \sum_{t=0}^{n-1} \psi_\rho(X_t, X_{t+1}) + \frac{1}{\sqrt{n}} \sum_{t=0}^{n-1} \phi_{k,t}^* \phi_{k,t+1}\big(m_t(\hat{\alpha}) - m_t(\alpha_0)\big) + o_p(1), \end{aligned} \tag{S.3}$$

where the second equality is by Lemma D.1.

We decompose the second term on the right-hand side of (S.3) as

$$\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{t=0}^{n-1}\phi_{k,t}^{*}\phi_{k,t+1}\big(m_t(\hat{\alpha})-m_t(\alpha_0)\big)\\
&\quad=\frac{1}{\sqrt{n}}\sum_{t=0}^{n-1}\phi_t^{*}\phi_{t+1}\frac{\partial m_t(\alpha_0)}{\partial\alpha'}(\hat{\alpha}-\alpha_0)\\
&\qquad+\frac{1}{\sqrt{n}}\sum_{t=0}^{n-1}\phi_t^{*}\phi_{t+1}\left(m_t(\hat{\alpha})-m_t(\alpha_0)-\frac{\partial m_t(\alpha_0)}{\partial\alpha'}(\hat{\alpha}-\alpha_0)\right)\qquad\text{(S.4)}\\
&\qquad+\frac{1}{\sqrt{n}}\sum_{t=0}^{n-1}\big(\phi_{k,t}^{*}\phi_{k,t+1}-\phi_t^{*}\phi_{t+1}\big)\big(m_t(\hat{\alpha})-m_t(\alpha_0)\big)\\
&\quad=:\frac{1}{\sqrt{n}}\sum_{t=0}^{n-1}\phi_t^{*}\phi_{t+1}\frac{\partial m_t(\alpha_0)}{\partial\alpha'}(\hat{\alpha}-\alpha_0)+\widehat{T}_1+\widehat{T}_2.
\end{aligned}$$

For term $\widehat{T}_1$, whenever $\hat{\alpha}\in N$ (which it is wpa1), we may take a mean value expansion to obtain

$$\widehat{T}_1=\frac{1}{n}\sum_{t=0}^{n-1}\phi_t^{*}\phi_{t+1}\left(\frac{\partial m_t(\tilde{\alpha})}{\partial\alpha'}-\frac{\partial m_t(\alpha_0)}{\partial\alpha'}\right)\times\sqrt{n}(\hat{\alpha}-\alpha_0),$$

where $\tilde{\alpha}$ is in the segment between $\hat{\alpha}$ and α_0. It follows by routine arguments (e.g., Lemma 4.3 of Newey and McFadden (1994), replacing the law of large numbers by the ergodic theorem) that

$$\frac{1}{n}\sum_{t=0}^{n-1}\phi_t^{*}\phi_{t+1}\left(\frac{\partial m_t(\tilde{\alpha})}{\partial\alpha}-\frac{\partial m_t(\alpha_0)}{\partial\alpha}\right)=o_p(1)\qquad\text{(S.5)}$$

holds under Assumption 3.5(c),(d). Moreover, $\sqrt{n}(\hat{\alpha}-\alpha_0)=O_p(1)$ by Assumption 3.5(a),(b). Therefore, $\widehat{T}_1=o_p(1)$.

For term $\widehat{T}_2$, observe that by Assumption 3.5(c), whenever $\hat{\alpha}\in N$ (which it is wpa1), we have

$$\big|m_t(\hat{\alpha})-m_t(\alpha_0)\big|\le\bar{m}(X_t,X_{t+1})\times\|\hat{\alpha}-\alpha_0\|,$$

where $\max_{0\le t\le n-1}|\bar{m}(X_t,X_{t+1})|=o_p(n^{1/s})$ because $E[\bar{m}(X_t,X_{t+1})^s]<\infty$. Therefore, wpa1, we have

$$\begin{aligned}
\widehat{T}_2&\le\sqrt{n}\times\frac{1}{n}\sum_{t=0}^{n-1}\big|\phi_{k,t}^{*}\phi_{k,t+1}-\phi_t^{*}\phi_{t+1}\big|\times\max_{0\le t\le n-1}\big|\bar{m}(X_t,X_{t+1})\big|\times\|\hat{\alpha}-\alpha_0\|\\
&=\frac{1}{n}\sum_{t=0}^{n-1}\big|\phi_{k,t}^{*}\phi_{k,t+1}-\phi_t^{*}\phi_{t+1}\big|\times o_p\big(n^{1/s}\big)=O_p\big(\delta_k+\delta_k^{*}\big)\times o_p\big(n^{1/s}\big)
\end{aligned}$$

by similar arguments to the proof of Lemma D.1. Finally, observe that $n^{1/s}(\delta_k+\delta_k^{*})=o(1)$ by Assumption 3.4(a) and the condition $s\ge 2$. Therefore, $\widehat{T}_2=o_p(1)$.

Since $\widehat{T}_1$ and $\widehat{T}_2$ in display (S.4) are both $o_p(1)$, we have

$$\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{t=0}^{n-1}\phi_{k,t}^*\phi_{k,t+1}\big(m_t(\hat\alpha)-m_t(\alpha_0)\big)\\
&\quad=\left(\frac{1}{n}\sum_{t=0}^{n-1}\phi_t^*\phi_{t+1}\frac{\partial m_t(\alpha_0)}{\partial\alpha'}\right)\sqrt{n}(\hat\alpha-\alpha_0)+o_p(1)\\
&\quad=\mathbb{E}\left[\phi^*(X_t)\phi(X_{t+1})\frac{\partial m(X_t,X_{t+1};\alpha_0)}{\partial\alpha'}\right]\sqrt{n}(\hat\alpha-\alpha_0)+o_p(1).
\end{aligned}$$

Substituting into (S.3) and using Assumption 3.5(a):

$$\sqrt{n}(\hat\rho-\rho)=\frac{1}{\sqrt{n}}\sum_{t=0}^{n-1}h'_{[2a]}\begin{pmatrix}\psi_{\rho,t}\\ \psi_{\alpha,t}\end{pmatrix}+o_p(1)$$

and the result follows by Assumption 3.5(b). *Q.E.D.*

PROOF OF THEOREM 3.5: We follow similar arguments to the proof of Theorem 3.4. Here, we can decompose the second term on the right-hand side of display (S.3) as

$$\begin{aligned}
\frac{1}{\sqrt{n}}\sum_{t=0}^{n-1}\phi_{k,t}^*\phi_{k,t+1}\big(m_t(\hat\alpha)-m_t(\alpha_0)\big)&=\sqrt{n}\big(\ell(\hat\alpha)-\ell(\alpha_0)\big)+\widehat{T}_1+\widehat{T}_2\\
&=\frac{1}{\sqrt{n}}\sum_{t=0}^{n-1}\psi_{\ell,t}+o_p(1)+\widehat{T}_1+\widehat{T}_2,
\end{aligned}$$

where the second line is by Assumption 3.6(b),(c), with

$$\begin{aligned}
\widehat{T}_1&=\frac{1}{\sqrt{n}}\sum_{t=0}^{n-1}\Big(\phi_t^*\phi_{t+1}\big(m_t(\hat\alpha)-m_t(\alpha_0)\big)-\big(\ell(\hat\alpha)-\ell(\alpha_0)\big)\Big),\\
\widehat{T}_2&=\frac{1}{\sqrt{n}}\sum_{t=0}^{n-1}\big(\phi_{k,t}^*\phi_{k,t+1}-\phi_t^*\phi_{t+1}\big)\big(m_t(\hat\alpha)-m_t(\alpha_0)\big).
\end{aligned}$$

The result will follow by Assumption 3.6(c),(d) provided $\widehat{T}_1$ and $\widehat{T}_2$ are both $o_p(1)$.

For term $\widehat{T}_1$, notice that $\widehat{T}_1=\mathcal{Z}_n(g_{\hat\alpha})$ where $\mathcal{Z}_n$ denotes the centered empirical process on $\mathcal{G}$. We have $\mathbb{K}(g_{\hat\alpha},g_{\hat\alpha})=o_p(1)$ by Assumption 3.6(c). Appropriately modifying the arguments of Lemma 19.24 in van der Vaart (1998) (i.e., replacing the L^2 norm by the norm induced by $\mathbb{K}$, which is the appropriate semimetric for the weakly dependent case) gives $\mathcal{Z}_n(g_{\hat\alpha})\to_p 0$.

For term $\widehat{T}_2$, observe that

$$\mathbb{E}\big[\big|\big(\phi_{k,t}^*\phi_{k,t+1}-\phi_t^*\phi_{t+1}\big)\big(m_t(\hat\alpha)-m_t(\alpha_0)\big)\big|\big]\lesssim\mathbb{E}\big[\big|\big(\phi_{k,t}^*\phi_{k,t+1}-\phi_t^*\phi_{t+1}\big)\big|^{s/(s-1)}\big]^{(s-1)/s}$$

by Assumption 3.6(e) and Hölder's inequality. We complete the proof assuming $\|\phi_k\|_{2s/(s-2)}=O(1)$ and $\|\phi^*\|_{2s/(s-2)}<\infty$; the proof under the alternative condition in

Assumption 3.6(e) is analogous. By the Minkowski and Hölder inequalities and Assumption 3.6(e), we have

$$\begin{aligned}
&\mathbb{E}\big[\big|\big(\phi^*_{k,t}\phi_{k,t+1}-\phi^*_t\phi_{t+1}\big)\big|^{s/(s-1)}\big]^{(s-1)/s} \\
&\quad\le \mathbb{E}\big[\big|\big(\phi^*_{k,t}-\phi^*_t\big)\phi_{k,t+1}\big|^{s/(s-1)}\big]^{(s-1)/s}+\mathbb{E}\big[\big|\phi^*_t(\phi_{k,t+1}-\phi_{t+1})\big|^{s/(s-1)}\big]^{(s-1)/s} \\
&\quad\le \big\|\phi^*_k-\phi^*\big\|\|\phi_k\|_{2s/(s-2)}+\|\phi_k-\phi\|\big\|\phi^*\big\|_{2s/(s-2)} \\
&\quad= O(1)\times O\big(\delta^*_k+\delta_k\big).
\end{aligned}$$

It follows by Assumption 3.4(a) and Markov's inequality that $\widehat{T}_2=o_p(1)$. *Q.E.D.*

The following lemma is based on Lemma 6.10 in Akian, Gaubert, and Nussbaum (2016).

LEMMA D.2: *Let the conditions of Proposition* 4.1 *hold*. *Then there exist finite positive constants* C, c, *and a neighborhood* N *of* h *such that*

$$\big\|\mathbb{T}^n\psi-h\big\|\le Ce^{-cn}$$

for all $\psi\in N$.

PROOF OF LEMMA D.2: Fix some constant $\bar{a}$ such that $r(\mathbb{D}_h)<\bar{a}<1$. By the Gelfand formula, there exists $m\in\mathbb{N}$ such that $\|\mathbb{D}^m_h\|<\bar{a}^m$. Fréchet differentiability of $\mathbb{T}$ at h together with the chain rule for Fréchet derivatives implies that

$$\big\|\mathbb{T}^m\psi-\mathbb{T}^m h-\mathbb{D}^m_h(\psi-h)\big\|=o\big(\|\psi-h\|\big)\quad \text{as } \|\psi-h\|\to 0,$$

hence

$$\big\|\mathbb{T}^m\psi-h\big\|\le\big\|\mathbb{D}^m_h\big\|\|\psi-h\|+o\big(\|\psi-h\|\big)<\big(\bar{a}^m+o(1)\big)\times\|\psi-h\|.$$

We may choose $\epsilon>0$ and $a\in(\bar{a},1)$ such that $\|\mathbb{T}^m\psi-h\|\le a^m\|\psi-h\|$ for all $\psi\in B_\epsilon(h):=\{\psi\in L^2:\|\psi-h\|<\epsilon\}$. ($B_\epsilon(h)$ is the neighborhood in the statement of the lemma.) Then for any $\psi\in B_\epsilon(h)$ and any $k\in\mathbb{N}$, we have

$$\big\|\mathbb{T}^{km}\psi-h\big\|\le a^{km}\|\psi-h\|. \tag{S.6}$$

It is straightforward to show via induction that boundedness of $\mathbb{G}$ and homogeneity of degree β of $\mathbb{T}$ together imply

$$\big\|\mathbb{T}^n\psi_1-\mathbb{T}^n\psi_2\big\|\le\big(1+\|\mathbb{G}\|\big)^{\frac{1}{1-\beta}}\|\psi_1-\psi_2\|^{\beta^n} \tag{S.7}$$

for any $\psi_1,\psi_2\in L^2$.

Take any $n\ge m$ and let $k=\lfloor n/m\rfloor$. By (S.6) and (S.7), we have

$$\begin{aligned}
\big\|\mathbb{T}^n\psi-h\big\|&=\big\|\mathbb{T}^{(n-km)}\mathbb{T}^{km}\psi-\mathbb{T}^{(n-km)}h\big\| \\
&\le\big(1+\|\mathbb{G}\|\big)^{\frac{1}{1-\beta}}\big\|\mathbb{T}^{km}\psi-h\big\|^{\beta^{(n-km)}} \\
&\le\big(1+\|\mathbb{G}\|\big)^{\frac{1}{1-\beta}}\epsilon^{\beta^{(n-km)}}\big(a^{km}\big)^{\beta^{(n-km)}}
\end{aligned}$$

for any $\psi\in B_\epsilon(h)$. The result follows for suitable choice of C and c. *Q.E.D.*

PROOF OF PROPOSITION 4.1: Take C and c from Lemma D.2 and $B_\epsilon(h)$ from the proof of Lemma D.2. Let $N = \{\psi \in L^2 : \|\psi - \chi\| < \epsilon/\|h\|\}$ and note that $\{\|h\|\psi : \psi \in N\} = B_\epsilon(h)$.

Take any $\psi \in \{af : f \in N, a \in \mathbb{R} \setminus \{0\}\}$. For any such ψ, we can write $\psi = (a/\|h\|)f^*$ where $f^* = \|h\|f \in B_\epsilon(h)$. By homogeneity of $\mathbb{T}$:

$$\chi_{n+1}(\psi) = \frac{\mathbb{T}^n(\chi_1(\psi))}{\|\mathbb{T}^n(\chi_1(\psi))\|} = \frac{\mathbb{T}^n(\chi_1(f^*))}{\|\mathbb{T}^n(\chi_1(f^*))\|} = \chi_{n+1}(f^*)$$

for each $n \geq 1$ (note positivity of $\mathbb{G}$ ensures that $\|\mathbb{T}^n f^*\| > 0$ for each n and each $f^* \in N$). It follows from Lemma D.2 that

$$\|\chi_{n+1}(\psi) - \chi\| = \|\chi_{n+1}(f^*) - \chi\| = \left\| \frac{\mathbb{T}^n(f^*)}{\|\mathbb{T}^n(f^*)\|} - \frac{h}{\|h\|} \right\| \leq \frac{2}{\|h\|}\|\mathbb{T}^n(f^*) - h\| \leq \frac{2}{\|h\|}Ce^{-cn}$$

as required. *Q.E.D.*

PROOF OF COROLLARY 4.1: The result for χ is stated in the text. For h, let C, c, and $B_\epsilon(h)$ be as in Lemma D.2 and its proof. Suppose h' is a fixed point of $\mathbb{T}$ belonging to $B_\epsilon(h)$. Then, by Lemma D.2,

$$\|h' - h\| = \|\mathbb{T}^n h' - h\| \leq Ce^{-cn} \to 0,$$

hence $h' = h$. *Q.E.D.*

PROOF OF THEOREM 4.1: Immediate from Lemmas A.6 and A.8. *Q.E.D.*

D.2. *Proofs for Appendix A.1*

PROOF OF LEMMA A.1: We first prove that there exists $K \in \mathbb{N}$ such that the maximum eigenvalue ρ_k of the operator $\Pi_k\mathbb{M} : L^2 \to L^2$ is real and simple whenever $k \geq K$.

Under Assumption 3.1, ρ is a simple isolated eigenvalue of $\mathbb{M}$. Therefore, there exists an $\epsilon > 0$ such that $|\lambda - \rho| > 2\epsilon$ for all $\lambda \in \sigma(\mathbb{M}) \setminus \{\rho\}$. Let Γ denote a positively oriented circle in $\mathbb{C}$ centered at ρ with radius ϵ. Let $\mathcal{R}(\mathbb{M}, z) = (\mathbb{M} - zI)^{-1}$ denote the resolvent of $\mathbb{M}$ evaluated at $z \in \mathbb{C} \setminus \sigma(\mathbb{M})$, where I is the identity operator. Note that

$$C_{\mathcal{R}} := \sup_{z\in\Gamma} \|\mathcal{R}(\mathbb{M}, z)\| < \infty \tag{S.8}$$

because $\mathcal{R}(\mathbb{M}, z)$ is a holomorphic function on Γ and Γ is compact.

By Assumption 3.2, there exists $K \in \mathbb{N}$ such that

$$C_{\mathcal{R}} \times \|\Pi_k\mathbb{M} - \mathbb{M}\| < 1 \tag{S.9}$$

holds for all $k \geq K$. It follows by Theorem IV.3.18 on p. 214 of Kato (1980) that whenever $k \geq K$: (i) the operator $\Pi_k\mathbb{M}$ has precisely one eigenvalue ρ_k inside Γ and ρ_k is simple; (ii) $\Gamma \subset (\mathbb{C} \setminus \sigma(\Pi_k\mathbb{M}))$; and (iii) $\sigma(\Pi_k\mathbb{M}) \setminus \{\rho_k\}$ lies on the exterior of Γ. Note that ρ_k must be real whenever $k \geq K$ because complex eigenvalues come in conjugate pairs. Thus, if ρ_k

were complex-valued, then its conjugate would also be inside Γ, which would contradict the fact that ρ_k is the unique eigenvalue of $\Pi_k\mathbb{M}$ on the interior of Γ.

Any nonzero eigenvalue of $\Pi_k\mathbb{M}$ is also an eigenvalue of $(\mathbf{M}, \mathbf{G})$ with the same multiplicity. Therefore, the largest eigenvalue ρ_k of $(\mathbf{M}, \mathbf{G})$ is positive and simple whenever $k \geq K$. *Q.E.D.*

Let $\Pi_k\mathbb{M}|_{B_k} : B_k \to B_k$ denote the restriction of $\Pi_k\mathbb{M}$ to B_k. Recall that $\phi_k^*(x) = b^k(x)'c_k^*$ where c_k^* solves the left-eigenvector problem in (15). Here, ϕ_k^* is the eigenfunction of the adjoint $(\Pi_k\mathbb{M}|_{B_k})^* : B_k \to B_k$ corresponding to ρ_k. That is, $\langle(\Pi_k\mathbb{M}|_{B_k})^*\phi_k^*, \psi\rangle = \rho_k\langle\phi_k^*, \psi\rangle$ for all $\psi \in B_k$.

Another adjoint is also relevant for the next proof, namely, $(\Pi_k\mathbb{M})^* : L^2 \to L^2$, which is the adjoint of $\Pi_k\mathbb{M}$ in the space L^2. It follows from Lemma A.1 that $(\Pi_k\mathbb{M})^*$ has an eigenfunction, say ϕ_k^+, corresponding to ρ_k whenever $k \geq K$. That is, $\langle(\Pi_k\mathbb{M})^*\phi_k^+, \psi\rangle = \rho_k\langle\phi_k^+, \psi\rangle$ for all $\psi \in L^2$. Notice that ϕ_k^+ does not necessarily belong to B_k, so we may have that $\phi_k^* \neq \phi_k^+$.

PROOF OF LEMMA A.2: Step 1: Proof of part (b). By Proposition 4.2 of Gobet, Hoffmann, and Reiß (2004) (taking $T = \mathbb{M}$, $T_\varepsilon = \Pi_k\mathbb{M}$, and $\Gamma =$ the boundary of $B(\kappa, \rho)$ in their notation), the inequality

$$\|\phi - \phi_k\| \leq \text{const} \times \big\|(\Pi_k\mathbb{M} - \mathbb{M})\phi\big\|$$

holds for all k sufficiently large, where the constant depends only on $C_{\mathcal{R}}$. The result follows by noticing that

$$\big\|(\Pi_k\mathbb{M} - \mathbb{M})\phi\big\| = \rho \times \|\Pi_k\phi - \phi\| = O(\delta_k). \tag{S.10}$$

Step 2: Proof of part (a). By Corollary 4.3 of Gobet, Hoffmann, and Reiß (2004), the inequality

$$|\rho - \rho_k| \leq \text{const} \times \big\|(\Pi_k\mathbb{M} - \mathbb{M})\phi\big\|$$

holds for all k sufficiently large, where the constant depends only on $C_{\mathcal{R}}$ and $\|\mathbb{M}\|$. The result follows by (S.10).

Step 3: Proof that $\|\phi_k^+ - \phi^*\| = O(\delta_k^*)$ under the normalizations $\|\phi^*\| = 1$ and $\|\phi_k^+\| = 1$. Let P_k^* denote the spectral projection on the eigenspace of $(\Pi_k\mathbb{M})^*$ corresponding to ρ_k. By the proof of Proposition 4.2 of Gobet, Hoffmann, and Reiß (2004) (taking $T = \mathbb{M}^*$, $T_\varepsilon = (\Pi_k\mathbb{M})^*$, and $\Gamma =$ the boundary of $B(\kappa, \rho)$ in their notation and noting that $\|\mathcal{R}(\mathbb{M}^*, z)\| = \|\mathcal{R}(\mathbb{M}, \bar{z})\|$ holds for all $z \in \Gamma$), the inequality

$$\big\|\phi^* - P_k^*\phi^*\big\| \leq \text{const} \times \big\|\big((\Pi_k\mathbb{M})^* - \mathbb{M}^*\big)\phi^*\big\|$$

for all k sufficiently large, where the constant depends only on $C_{\mathcal{R}}$. Moreover,

$$\begin{aligned}\big\|\big((\Pi_k\mathbb{M})^* - \mathbb{M}^*\big)\phi^*\big\| &= \big\|\big(\mathbb{M}^*\Pi_k - \mathbb{M}^*\big)\phi^*\big\| = \big\|\mathbb{M}^*\big(\Pi_k\phi^* - \phi^*\big)\big\| \\ &\leq \|\mathbb{M}\|\big\|\Pi_k\phi^* - \phi^*\big\| = O\big(\delta_k^*\big)\end{aligned}$$

by definition of δ_k^* (cf. display (22)) and boundedness of $\mathbb{M}$. Therefore,

$$\big\|\phi^* - P_k^*\phi^*\big\| = O\big(\delta_k^*\big). \tag{S.11}$$

Define $(\phi_k^+ \otimes \phi_k)\psi(x) = \langle \phi_k, \psi \rangle \times \phi_k^+(x)$ for any $\psi \in L^2$. We use the fact that

$$P_k^* = \frac{1}{\langle \phi_k, \phi_k^+ \rangle}(\phi_k^+ \otimes \phi_k)$$

under the normalizations $\|\phi_k\| = 1$ and $\|\phi_k^+\| = 1$ (Chatelin (1983, p. 113)). Then, under the sign normalization $\langle \phi^*, \phi_k^+ \rangle \geq 0$, we have

$$\|\phi^* - \phi_k^+\|^2 \leq 2\|\phi^* - (\phi_k^+ \otimes \phi_k^+)\phi^*\|^2$$

(see the proof of Proposition 4.2 of Gobet, Hoffmann, and Reiß (2004)). Moreover,

$$\|\phi^* - (\phi_k^+ \otimes \phi_k^+)\phi^*\|^2 \leq \left\| \phi^* - \left(\phi_k^+ \otimes \frac{\phi_k}{\langle \phi_k, \phi_k^+ \rangle}\right)\phi^* \right\|^2 \equiv \|\phi^* - P_k^* \phi^*\|^2.$$

It follows by (S.11) that $\|\phi^* - \phi_k^+\| = O(\delta_k^*)$.

Step 4: Proof that $\|\phi_k^* - \phi^*\| = O(\delta_k^*)$. To relate ϕ_k^+ to ϕ_k^*, observe that by definition of $(\Pi_k \mathbb{M})^*$ and $(\Pi_k \mathbb{M}|_{B_k})^*$, we must have

$$\begin{aligned} \mathbb{E}[\phi_k^+(X)\Pi_k \mathbb{M}\psi(X)] &= \rho_k \mathbb{E}[\phi_k^+(X)\psi(X)] \quad \text{for all } \psi \in L^2, \\ \mathbb{E}[\phi_k^*(X)\Pi_k \mathbb{M}\psi_k(X)] &= \rho_k \mathbb{E}[\phi_k^*(X)\psi_k(X)] \quad \text{for all } \psi_k \in B_k. \end{aligned}$$

It follows from taking $\psi = \psi_k$ in the first line of the above display that $\Pi_k \phi_k^+ = \phi_k^*$. Now by the triangle inequality and the fact that Π_k is a weak contraction, we have

$$\begin{aligned} \|\phi^* - \phi_k^*\| = \|\phi^* - \Pi_k \phi_k^+\| &\leq \|\phi^* - \Pi_k \phi^*\| + \|\Pi_k \phi^* - \Pi_k \phi_k^+\| \\ &\leq \|\phi^* - \Pi_k \phi^*\| + \|\phi^* - \phi_k^+\| = O(\delta_k^*) + O(\delta_k^*), \end{aligned}$$

where the final equality is by definition of δ_k^* (see display (22)) and Step 3. *Q.E.D.*

The following lemma collects some useful bounds on the orthogonalized estimators.

LEMMA D.3:
(a) *If* $\widehat{\mathbf{G}}$ *is invertible, then*

$$(\widehat{\mathbf{G}}^o)^{-1}\widehat{\mathbf{M}}^o - \mathbf{M}^o = \widehat{\mathbf{M}}^o - \widehat{\mathbf{G}}^o\mathbf{M}^o + (\widehat{\mathbf{G}}^o)^{-1}\left((\widehat{\mathbf{G}}^o - \mathbf{I})^2\mathbf{M}^o + (\mathbf{I} - \widehat{\mathbf{G}}^o)(\widehat{\mathbf{M}}^o - \mathbf{M}^o)\right).$$

(b) *In particular, if* $\|\widehat{\mathbf{G}}^o - \mathbf{I}\| \leq \frac{1}{2}$, *we obtain*

$$\|(\widehat{\mathbf{G}}^o)^{-1}\widehat{\mathbf{M}}^o - \mathbf{M}^o\| \leq \|\widehat{\mathbf{M}}^o - \mathbf{M}^o\| + 2\|\widehat{\mathbf{G}}^o - \mathbf{I}\| \times (\|\mathbf{M}^o\| + \|\widehat{\mathbf{M}}^o - \mathbf{M}^o\|).$$

PROOF OF LEMMA D.3: If $\widehat{\mathbf{G}}$ is invertible, we have

$$\begin{aligned} (\widehat{\mathbf{G}}^o)^{-1}\widehat{\mathbf{M}}^o - \mathbf{M}^o &= (\mathbf{I} - (\widehat{\mathbf{G}}^o)^{-1}(\widehat{\mathbf{G}}^o - \mathbf{I}))\widehat{\mathbf{M}}^o - \mathbf{M}^o \\ &= \widehat{\mathbf{M}}^o - \mathbf{M}^o - (\widehat{\mathbf{G}}^o)^{-1}(\widehat{\mathbf{G}}^o - \mathbf{I})\mathbf{M}^o - (\widehat{\mathbf{G}}^o)^{-1}(\widehat{\mathbf{G}}^o - \mathbf{I})(\widehat{\mathbf{M}}^o - \mathbf{M}^o). \end{aligned}$$

Part (b) follows by the triangle inequality, noting that $\|(\widehat{\mathbf{G}}^o)^{-1}\| \le 2$ whenever $\|\widehat{\mathbf{G}}^o - \mathbf{I}\| \le \frac{1}{2}$. Substituting $(\widehat{\mathbf{G}}^o)^{-1} = (\mathbf{I} - (\widehat{\mathbf{G}}^o)^{-1}(\widehat{\mathbf{G}}^o - \mathbf{I}))$ into the preceding display yields

$$\begin{aligned}
&(\widehat{\mathbf{G}}^o)^{-1}\widehat{\mathbf{M}}^o - \mathbf{M}^o \\
&\quad = \widehat{\mathbf{M}}^o - \mathbf{M}^o - \big(\mathbf{I} - (\widehat{\mathbf{G}}^o)^{-1}(\widehat{\mathbf{G}}^o - \mathbf{I})\big)(\widehat{\mathbf{G}}^o - \mathbf{I})\mathbf{M}^o - (\widehat{\mathbf{G}}^o)^{-1}(\widehat{\mathbf{G}}^o - \mathbf{I})(\widehat{\mathbf{M}}^o - \mathbf{M}^o) \\
&\quad = \widehat{\mathbf{M}}^o - \widehat{\mathbf{G}}^o\mathbf{M}^o + (\widehat{\mathbf{G}}^o)^{-1}(\widehat{\mathbf{G}}^o - \mathbf{I})^2\mathbf{M}^o - (\widehat{\mathbf{G}}^o)^{-1}(\widehat{\mathbf{G}}^o - \mathbf{I})(\widehat{\mathbf{M}}^o - \mathbf{M}^o),
\end{aligned}$$

as required. *Q.E.D.*

PROOF OF LEMMA A.3: Step 1: We show that

$$\|\mathcal{R}(\Pi_k\mathbb{M}|_{B_k}, z)\| \le \|\mathcal{R}(\Pi_k\mathbb{M}, z)\|$$

holds for all $z \in \mathbb{C} \setminus (\sigma(\Pi_k\mathbb{M}) \cup \sigma(\Pi_k\mathbb{M}|_{B_k}))$. Fix any such z. For any $\psi_k \in B_k$, we have $\mathcal{R}(\Pi_k\mathbb{M}|_{B_k}, z)\psi_k = \zeta_k$ where $\zeta_k = \zeta_k(\psi_k) \in B_k$ is given by $\psi_k = (\Pi_k\mathbb{M} - zI)\zeta_k$. For any $\psi \in L^2$, we have $\mathcal{R}(\Pi_k\mathbb{M}, z)\psi = \zeta$ where $\zeta = \zeta(\psi) \in L^2$ is given by $\psi = (\Pi_k\mathbb{M} - zI)\zeta$. In particular, taking $\psi_k \in B_k$, we must have $\zeta_k(\psi_k) = \zeta(\psi_k)$. Therefore, $\mathcal{R}(\Pi_k\mathbb{M}|_{B_k}, z)\psi_k = \mathcal{R}(\Pi_k\mathbb{M}, z)\psi_k$ holds for all $\psi_k \in B_k$. We now have

$$\begin{aligned}
\|\mathcal{R}(\Pi_k\mathbb{M}|_{B_k}, z)\| &= \sup\{\|\mathcal{R}(\Pi_k\mathbb{M}|_{B_k}, z)\psi_k\| : \psi_k \in B_k, \|\psi_k\| = 1\} \\
&= \sup\{\|\mathcal{R}(\Pi_k\mathbb{M}, z)\psi_k\| : \psi_k \in B_k, \|\psi_k\| = 1\} \\
&\le \sup\{\|\mathcal{R}(\Pi_k\mathbb{M}, z)\psi\| : \psi \in L^2, \|\psi\| = 1\} = \|\mathcal{R}(\Pi_k\mathbb{M}, z)\|.
\end{aligned}$$

Step 2: We show that $(\widehat{\mathbf{M}}, \widehat{\mathbf{G}})$ has a unique eigenvalue $\hat{\rho}$ inside Γ wpa1, where Γ is from the proof of Lemma A.1.

As the nonzero eigenvalues of $\Pi_k\mathbb{M}$, $\Pi_k\mathbb{M}|_{B_k}$, and $\mathbf{G}^{-1}\mathbf{M}$ are the same, it follows from the proof of Lemma A.1 that for all $k \ge K$, the curve Γ encloses precisely one eigenvalue of $\mathbf{G}^{-1}\mathbf{M}$, namely, ρ_k, and that ρ_k is a simple eigenvalue of $\mathbf{G}^{-1}\mathbf{M}$.

Recall that $\mathbf{G}^{-1}\mathbf{M}$ is isomorphic to $\Pi_k\mathbb{M}|_{B_k}$ on $(\mathbb{R}^k, \langle\cdot,\cdot\rangle_{\mathbf{G}})$. Let $\mathcal{R}(\mathbf{G}^{-1}\mathbf{M}, z)$ denote the resolvent of $\mathbf{G}^{-1}\mathbf{M}$ on $(\mathbb{R}^k, \langle\cdot,\cdot\rangle_{\mathbf{G}})$. By step 1, we then have

$$\sup_{z\in\Gamma}\|\mathcal{R}(\mathbf{G}^{-1}\mathbf{M}, z)\|_{\mathbf{G}} = \sup_{z\in\Gamma}\|\mathcal{R}(\Pi_k\mathbb{M}|_{B_k}, z)\| \le \sup_{z\in\Gamma}\|\mathcal{R}(\Pi_k\mathbb{M}, z)\|. \tag{S.12}$$

The second resolvent identity gives $\mathcal{R}(\Pi_k\mathbb{M}, z) = \mathcal{R}(\mathbb{M}, z) + \mathcal{R}(\Pi_k\mathbb{M}, z)(\mathbb{M} - \Pi_k\mathbb{M})\mathcal{R}(\mathbb{M}, z)$. It follows that whenever (S.9) holds (which it does for all $k \ge K$),

$$\sup_{z\in\Gamma}\|\mathcal{R}(\Pi_k\mathbb{M}, z)\| \le \frac{C_{\mathcal{R}}}{1 - C_{\mathcal{R}}\|\Pi_k\mathbb{M} - \mathbb{M}\|} = C_{\mathcal{R}}(1 + o(1)) \tag{S.13}$$

by Assumption 3.2. Combining (S.12) and (S.13), we obtain

$$\sup_{z\in\Gamma}\|\mathcal{R}(\mathbf{G}^{-1}\mathbf{M}, z)\|_{\mathbf{G}} = O(1). \tag{S.14}$$

By Lemma D.3(b), Assumption 3.3, and boundedness of $\mathbb{M}$:

$$\|\widehat{\mathbf{G}}^{-1}\widehat{\mathbf{M}} - \mathbf{G}^{-1}\mathbf{M}\|_{\mathbf{G}} = \|(\widehat{\mathbf{G}}^o)^{-1}\widehat{\mathbf{M}}^o - \mathbf{M}^o\| = o_p(1).$$

It follows by (S.14) that the inequality

$$\big\|\widehat{\mathbf{G}}^{-1}\widehat{\mathbf{M}} - \mathbf{G}^{-1}\mathbf{M}\big\|_{\mathbf{G}} \times \sup_{z\in\Gamma}\big\|\mathcal{R}\big(\mathbf{G}^{-1}\mathbf{M}, z\big)\big\|_{\mathbf{G}} < 1 \tag{S.15}$$

holds wpa1.

By Theorem IV.3.18 on p. 214 of Kato (1980), whenever (S.15) holds: $\widehat{\mathbf{G}}^{-1}\widehat{\mathbf{M}}$ has precisely one eigenvalue, say $\hat{\rho}$, inside Γ; $\hat{\rho}$ is simple; and, the remaining eigenvalues of $\widehat{\mathbf{G}}^{-1}\widehat{\mathbf{M}}$ are on the exterior of Γ. Note that $\hat{\rho}$ must necessarily be real whenever (S.15) holds (because complex eigenvalues come in conjugate pairs); hence the corresponding left- and right-eigenvectors $\hat{c}^*$ and $\hat{c}$ are also real and unique (up to scale). *Q.E.D.*

PROOF OF LEMMA A.4: Take $k \geq K$ from Lemma A.1 and work on the sequence of events upon which

$$\big\|\widehat{\mathbf{G}}^{-1}\widehat{\mathbf{M}} - \mathbf{G}^{-1}\mathbf{M}\big\|_{\mathbf{G}} \times \sup_{z\in\Gamma}\big\|\mathcal{R}\big(\mathbf{G}^{-1}\mathbf{M}, z\big)\big\|_{\mathbf{G}} < \frac{1}{2} \tag{S.16}$$

holds. By the proof of Lemma A.3, this inequality holds wpa1 and $\hat{\rho}$, $\hat{c}$, and $\hat{c}^*$ to (16) are unique on this sequence of events.

Step 1: Proof of part (b). Under the normalizations $\|\hat{c}\|_{\mathbf{G}} = 1$ and $\|\hat{c}^*\|_{\mathbf{G}} = 1$, whenever (S.16) holds (which it does wpa1), we have

$$\|\hat{\phi} - \phi_k\|^2 = \|\hat{c} - c_k\|^2_{\mathbf{G}} \leq \sqrt{8}\sup_{z\in\Gamma}\big\|\mathcal{R}\big(\mathbf{G}^{-1}\mathbf{M}, z\big)\big\|_{\mathbf{G}} \times \big\|\big(\widehat{\mathbf{G}}^{-1}\widehat{\mathbf{M}} - \mathbf{G}^{-1}\mathbf{M}\big)c_k\big\|_{\mathbf{G}}$$

by Proposition 4.2 of Gobet, Hoffmann, and Reiß (2004) (setting $\widehat{\mathbf{G}}^{-1}\widehat{\mathbf{M}} = T_\varepsilon$, $\mathbf{G}^{-1}\mathbf{M} = T$, and $\Gamma =$ the boundary of $B(\kappa, \rho)$ in their notation). The result now follows by (S.14) and the fact that

$$\big\|\big(\widehat{\mathbf{G}}^{-1}\widehat{\mathbf{M}} - \mathbf{G}^{-1}\mathbf{M}\big)c_k\big\|_{\mathbf{G}} = \big\|\big(\big(\widehat{\mathbf{G}}^o\big)^{-1}\widehat{\mathbf{M}}^o - \mathbf{M}^o\big)\tilde{c}_k\big\| = O_p(\eta_{n,k}) \tag{S.17}$$

(cf. display (23)).

Step 2: Proof of part (a). In view of (S.16), (S.14), and the fact that $\|\mathbf{G}^{-1}\mathbf{M}\|_{\mathbf{G}} = \|\Pi_k\mathbb{M}|_{B_k}\| \leq \|\mathbb{M}\| < \infty$, by Corollary 4.3 of Gobet, Hoffmann, and Reiß (2004), we have

$$|\hat{\rho} - \rho_k| \leq O(1) \times \big\|\big(\widehat{\mathbf{G}}^{-1}\widehat{\mathbf{M}} - \mathbf{G}^{-1}\mathbf{M}\big)c_k\big\|_{\mathbf{G}}.$$

The result follows by (S.17).

Step 3: Proof of part (c). Identical arguments to the proof of part (b) yield

$$\big\|\hat{\phi}^* - \phi_k^*\big\| = \big\|\hat{c}^* - c_k^*\big\|_{\mathbf{G}} \leq \sqrt{8}\sup_{z\in\Gamma}\big\|\mathcal{R}\big(\mathbf{G}^{-1}\mathbf{M}', z\big)\big\|_{\mathbf{G}} \times \big\|\big(\widehat{\mathbf{G}}^{-1}\widehat{\mathbf{M}}' - \mathbf{G}^{-1}\mathbf{M}'\big)c_k^*\big\|_{\mathbf{G}}$$

under the normalization $\|\hat{c}^*\|_{\mathbf{G}} = \|c_k^*\|_{\mathbf{G}} = 1$. The result now follows by (S.14), noting that $\sup_{z\in\Gamma}\|\mathcal{R}(\mathbf{G}^{-1}\mathbf{M}', z)\|_{\mathbf{G}} = \sup_{z\in\Gamma}\|\mathcal{R}(\mathbf{G}^{-1}\mathbf{M}, z)\|_{\mathbf{G}}$, and the fact that

$$\big\|\big(\widehat{\mathbf{G}}^{-1}\widehat{\mathbf{M}}' - \mathbf{G}^{-1}\mathbf{M}'\big)c_k^*\big\|_{\mathbf{G}} = \big\|\big(\big(\widehat{\mathbf{G}}^o\big)^{-1}\widehat{\mathbf{M}}^{o\prime} - \mathbf{M}^{o\prime}\big)\tilde{c}_k^*\big\| = O_p\big(\eta_{n,k}^*\big)$$

(cf. display (23)). *Q.E.D.*

D.3. *Proofs for Appendix A.2*

Some of the proofs in this subsection make use of properties of fixed-point indices. We refer the reader to Section 19.5 of Krasnosel'skii, Vainikko, Zabreiko, Rutitskii, and Stetsenko (1972) for details.

PROOF OF LEMMA A.5: By Assumption 4.1 and Corollary 4.1, we may choose $\varepsilon > 0$ such that $\overline{N} = \{\psi \in L^2 : \|\psi - h\| \le \varepsilon\}$ contains only one fixed point of $\mathbb{T}$, namely, h. We verify the conditions of Theorem 19.4 in Krasnosel'skii et al. (1972) where, in our notation, $\Omega = \overline{N}$, $E_n = B_k$, $P_n = \Pi_k$, $T = \mathbb{T}$, and $T_n = \Pi_k \mathbb{T}|_{B_k}$ (i.e., the restriction of $\Pi_k\mathbb{T}$ to B_k). The compactness condition is satisfied by Assumption 4.1(b) (recall that compactness of $\mathbb{G}$ implies compactness of $\mathbb{T}$). The fixed point h has nonzero index by Assumption 4.1(c); see result (5) on p. 300 of Krasnosel'skii et al. (1972). Finally, condition (19.28) in Krasnosel'skii et al. (1972) holds by Assumption 4.2(b) and their condition (19.29) is trivially satisfied. *Q.E.D.*

PROOF OF REMARK A.1: This follows by the proof of result (19.31) in Theorem 19.3 in Krasnosel'skii et al. (1972). *Q.E.D.*

PROOF OF REMARK A.2: This follows by Theorem 19.7 in Krasnosel'skii et al. (1972). *Q.E.D.*

PROOF OF LEMMA A.6: Part (c) follows by the proof of display (19.50) on p. 310 in Krasnosel'skii et al. (1972) where, in our notation, $x_0 = h$, $x_n = h_k$, $P_n = \Pi_k$, $P^{(n)} = I - \Pi_k$, $T = \mathbb{T}$, and $T'(x_0) = \mathbb{D}_h$. Note that Assumption 4.2(a) implies their condition $\|T'(x_0) - P_n T'(x_0)\| \to 0$ as $n \to \infty$. Part (b) then follows from the inequality

$$\left\| \frac{h}{\|h\|} - \frac{h_k}{\|h_k\|} \right\| \le \frac{2}{\|h\|} \|h - h_k\|.$$

Finally, part (a) follows from the fact that $|\|h\| - \|h_k\|| = O(\tau_k)$ and continuous differentiability of $x \mapsto x^{1-\beta}$ at each $x > 0$. *Q.E.D.*

The next lemma presents some bounds on the estimators which are used in the proof of Lemmas A.7 and A.8.

LEMMA D.4:
(a) *Let Assumptions* 4.1(b) *and* 4.3 *hold. Then*

$$\sup_{v \in \mathbb{R}^k : \|v\|_{\mathbf{G}} \le c} \left\| \widehat{\mathbf{G}}^{-1} \widehat{\mathbf{T}} v - \mathbf{G}^{-1} \mathbf{T} v \right\|_{\mathbf{G}} = o_p(1).$$

(b) *Moreover,*

$$\sup_{v \in \mathbb{R}^k : \|v' b^k - h\| \le \varepsilon} \left\| \widehat{\mathbf{G}}^{-1} \widehat{\mathbf{T}} v - \mathbf{G}^{-1} \mathbf{T} v \right\|_{\mathbf{G}} = O_p(\nu_{n,k}),$$

where $\nu_{n,k}$ is from display (37).

PROOF OF LEMMA D.4: By definition of $\widehat{\mathbf{G}}^o$, $\widehat{\mathbf{T}}^o$, and $\mathbf{T}^o$, we have

$$\sup_{v \in \mathbb{R}^k : \|v\|_{\mathbf{G}} \le c} \left\| \widehat{\mathbf{G}}^{-1} \widehat{\mathbf{T}} v - \mathbf{G}^{-1} \mathbf{T} v \right\|_{\mathbf{G}} = \sup_{v \in \mathbb{R}^k : \|v\| \le c} \left\| \left(\widehat{\mathbf{G}}^o\right)^{-1} \widehat{\mathbf{T}}^o v - \mathbf{T}^o v \right\|.$$

Whenever $\|\mathbf{I} - \widehat{\mathbf{G}}^o\| < 1$ (which it is wpa1 by Assumption 4.3), for any $v \in \mathbb{R}^k$, we have

$$\begin{aligned}&(\widehat{\mathbf{G}}^o)^{-1}\widehat{\mathbf{T}}^o v - \mathbf{T}^o v \\ &\quad = \widehat{\mathbf{T}}^o v - \mathbf{T}^o v - (\widehat{\mathbf{G}}^o)^{-1}(\widehat{\mathbf{G}}^o - \mathbf{I})\mathbf{T}^o v - (\widehat{\mathbf{G}}^o)^{-1}(\widehat{\mathbf{G}}^o - \mathbf{I})(\widehat{\mathbf{T}}^o v - \mathbf{T}^o v).\end{aligned} \tag{S.18}$$

Part (a) follows by the triangle inequality and Assumption 4.3, noting that $\sup_{v\in\mathbb{R}^k:\|v\|\le c}\|\mathbf{T}^o v\| \le \sup_{\psi:\|\psi\|\le c}\|\mathbb{T}\psi\| < \infty$ holds for each c by Assumption 4.1(b).

Part (b) follows by definition of $\widehat{\mathbf{G}}^o$, $\widehat{\mathbf{T}}^o$, $\mathbf{T}^o$, and $\nu_{n,k}$ in display (37). *Q.E.D.*

PROOF OF LEMMA A.7: Let ε, K, and N_k be as in Lemma A.5. Also define the sets $N = \{\psi \in L^2 : \|\psi - h\| < \varepsilon\}$, $\Gamma = \{\psi \in L^2 : \|\psi - h\| = \varepsilon\}$, $\Gamma_k = \{\psi \in B_k : \|\psi - h\| = \varepsilon\}$, $\mathbf{N}_k = \{v \in \mathbb{R}^k : v'b^k(x) \in N_k\}$, and $\boldsymbol{\Gamma}_k = \{v \in \mathbb{R}^k : v'b^k(x) \in \Gamma_k\}$.

Let $\gamma(I - \mathbb{T}; \Gamma)$ denote the rotation of the field $(I - \mathbb{T})\psi$ on Γ. Assumption 4.1 implies that $|\gamma(I - \mathbb{T}; \Gamma)| = 1$; see result (5) on p. 300 of Krasnosel'skii et al. (1972). Also notice that

$$\sup_{\psi\in\Gamma}\|\mathbb{T}\psi - \Pi_k\mathbb{T}\psi\| < \inf_{\psi\in\Gamma}\|\psi - \mathbb{T}\psi\| \tag{S.19}$$

holds for all k sufficiently large by Assumption 4.2(b) (note that $\inf_{\psi\in\Gamma}\|\psi - \mathbb{T}\psi\| > 0$, otherwise $\mathbb{T}$ would have a fixed point on Γ, contradicting the definition of $\overline{N}$ in the proof of Lemma A.5). Result (2) on p. 299 of Krasnosel'skii et al. (1972) then implies that whenever (S.19) holds, we have $|\gamma(I - \Pi_k\mathbb{T}; \Gamma)| = |\gamma(I - \mathbb{T}; \Gamma)| = 1$. Result (3) on p. 299 of Krasnosel'skii et al. (1972) then implies that $|\gamma(I - \Pi_k\mathbb{T}|_{B_k}; \Gamma_k)| = 1$ whenever (S.19) holds. Finally, by isomorphism, we have that $|\gamma(\mathbf{I} - \mathbf{G}^{-1}\mathbf{T}; \boldsymbol{\Gamma}_k)| = 1$ whenever (S.19) holds.

We now show that the inequality

$$\sup_{v\in\boldsymbol{\Gamma}_k}\big\|(\widehat{\mathbf{G}}^{-1}\widehat{\mathbf{T}} - \mathbf{G}^{-1}\mathbf{T})v\big\|_{\mathbf{G}} < \inf_{\psi\in\Gamma_k}\|\psi - \Pi_k\mathbb{T}\psi\| \tag{S.20}$$

holds wpa1. The left-hand side is $o_p(1)$ by Lemma D.4(a). For the right-hand side, we claim that $\liminf_{k\to\infty}\inf_{\psi\in\Gamma_k}\|\psi - \Pi_k\mathbb{T}\psi\| > 0$. Suppose the claim is false. Then there exists a subsequence $\{\psi_{k_l} : l \ge 1\}$ with $\psi_{k_l} \in \Gamma_{k_l}$ such that $\psi_{k_l} - \Pi_{k_l}\mathbb{T}\psi_{k_l} \to 0$. Since $\mathbb{T}$ is compact, there exists a convergent subsequence $\{\mathbb{T}\psi_{k_{l_j}} : j \ge 1\}$. Let $\psi^* = \lim_{j\to\infty}\mathbb{T}\psi_{k_{l_j}}$. Then

$$\big\|\psi_{k_{l_j}} - \psi^*\big\| \le \|\psi_{k_{l_j}} - \Pi_{k_{l_j}}\mathbb{T}\psi_{k_{l_j}}\| + \big\|\Pi_{k_{l_j}}\mathbb{T}\psi_{k_{l_j}} - \Pi_{k_{l_j}}\psi^*\big\| + \big\|\Pi_{k_{l_j}}\psi^* - \psi^*\big\| \to 0$$

as $j \to \infty$, where the first term vanishes by definition of ψ_{k_l}, the second vanishes by definition of ψ^*, and the third vanishes by Assumption 4.2(b). Therefore, $\psi^* \in \Gamma$. Moreover, by continuity of $\mathbb{T}$ and definition of ψ^*,

$$\big\|\mathbb{T}\psi^* - \psi^*\big\| \le \big\|\mathbb{T}\psi^* - \mathbb{T}\psi_{k_{l_j}}\big\| + \big\|\mathbb{T}\psi_{k_{l_j}} - \psi^*\big\| \to 0$$

as $j \to \infty$, hence $\psi^* \in \Gamma$ is a fixed point of $\mathbb{T}$. But this contradicts the fact that h is the unique fixed point of $\mathbb{T}$ in $\overline{N} = N \cup \Gamma$ (cf. the proof of Lemma A.5). This proves the claim.

Result (2) on p. 299 of Krasnosel'skii et al. (1972) then implies that whenever (S.19) and (S.20) hold (which they do wpa1), we have $\gamma(\mathbf{I} - \widehat{\mathbf{G}}^{-1}\widehat{\mathbf{T}}; \boldsymbol{\Gamma}_k) = \gamma(\mathbf{I} - \mathbf{G}^{-1}\mathbf{T}; \boldsymbol{\Gamma}_k)$. Therefore, $|\gamma(\mathbf{I} - \widehat{\mathbf{G}}^{-1}\widehat{\mathbf{T}}; \boldsymbol{\Gamma}_k)| = 1$ also holds wpa1 and hence, by result (1) on p. 299 of Krasnosel'skii et al. (1972), $\widehat{\mathbf{G}}^{-1}\widehat{\mathbf{T}}$ has at least one fixed point $\hat{v} \in \mathbf{N}_k$. We have therefore shown that

$\hat{h}(x) = b^k(x)'\hat{v}$ is well defined wpa1 and $\|\hat{h} - h\| < \varepsilon$ wpa1. Consistency of $\hat{h}$ follows by repeating the preceding argument with any positive $\varepsilon' < \varepsilon$. *Q.E.D.*

PROOF OF REMARK A.3: Fix any positive $\varepsilon' < \varepsilon$ and let $A = \{\psi \in L^2 : \varepsilon' \le \|\psi - h\| \le \varepsilon\}$, $A_k = \{\psi \in B_k : \varepsilon' \le \|\psi - h\| \le \varepsilon\}$, and $\mathbf{A}_k = \{v \in \mathbb{R}^k : v'b^k(x) \in A_k\}$. Clearly, $\mathbb{T}$ has no fixed point in A. Moreover, similar arguments to the proof of result (19.31) in Theorem 19.3 in Krasnosel'skii et al. (1972) imply that A_k contains no fixed points of $\Pi_k\mathbb{T}$ for all k sufficiently large. By similar arguments to the proof of Lemma A.7, we may deduce that $\liminf_{k\to\infty} \inf_{\psi\in A_k} \|\psi - \Pi_k\mathbb{T}\psi\| =: c^* > 0$. Then for any $v \in \mathbf{A}_k$, we have $\|v - \widehat{\mathbf{G}}^{-1}\widehat{\mathbf{T}}v\| \ge c^* - o_p(1)$ where the $o_p(1)$ term holds uniformly over $\mathbf{A}_k$ by Lemma D.4(a). Therefore, $\|v - \widehat{\mathbf{G}}^{-1}\widehat{\mathbf{T}}v\| \ge c^*/2$ holds for all $v \in \mathbf{A}_k$ wpa1. On the other hand, any fixed point $\hat{v}$ of $\widehat{\mathbf{G}}^{-1}\widehat{\mathbf{T}}$ with $b^k(x)'\hat{v} \in N_k$ necessarily has $\|\hat{v} - \widehat{\mathbf{G}}^{-1}\widehat{\mathbf{T}}\hat{v}\| = 0$. Therefore, no such fixed point $\hat{v}$ belongs to $\mathbf{A}_k$ wpa1. *Q.E.D.*

PROOF OF LEMMA A.8: We first prove part (c). The Fréchet derivative of $\Pi_k\mathbb{T}|_{B_k}$ at h is $\Pi_k\mathbb{D}_h|_{B_k}$. This may be represented on $(\mathbb{R}^k, \langle\cdot,\cdot\rangle_{\mathbf{G}})$ by the matrix $\mathbf{G}^{-1}\mathbf{D}_h$ where $\mathbf{D}_h = \mathbb{E}[b^k(X_t)\beta G_{t+1}^{1-\gamma}h(X_t)^{\beta-1}b^k(X_{t+1})']$. By Lemma A.7, $\hat{v}$ (equivalently, $\hat{h}$) is well defined wpa1. Therefore, wpa1, we have

$$\big(\mathbf{I} - \mathbf{G}^{-1}\mathbf{D}_h\big)(v_k - \hat{v}) = \mathbf{G}^{-1}\mathbf{T}\hat{v} - \widehat{\mathbf{G}}^{-1}\widehat{\mathbf{T}}\hat{v} - \big(\mathbf{G}^{-1}\mathbf{T}\hat{v} - \mathbf{G}^{-1}\mathbf{T}v_k - \mathbf{G}^{-1}\mathbf{D}_h(\hat{v} - v_k)\big).$$

Note that $\|\mathbf{G}^{-1}\mathbf{T}\hat{v} - \widehat{\mathbf{G}}^{-1}\widehat{\mathbf{T}}\hat{v}\|_{\mathbf{G}} = O_p(\nu_{n,k})$ by Lemma D.4(b) and consistency of $\hat{h}$. Therefore,

$$\big\|\big(\mathbf{I} - \mathbf{G}^{-1}\mathbf{D}_h\big)(v_k - \hat{v})\big\|_{\mathbf{G}} \le O_p(\nu_{n,k}) + \big\|\mathbf{G}^{-1}\mathbf{T}\hat{v} - \mathbf{G}^{-1}\mathbf{T}v_k - \mathbf{G}^{-1}\mathbf{D}_h(\hat{v} - v_k)\big\|_{\mathbf{G}}. \tag{S.21}$$

By isomorphism, we have $\|(\mathbf{I} - \mathbf{G}^{-1}\mathbf{D}_h)(v_k - \hat{v})\|_{\mathbf{G}} = \|(I - \Pi_k\mathbb{D}_h)(h_k - \hat{h})\|$. Assumptions 4.1(c) and 4.2(a) together imply that $(I - \Pi_k\mathbb{D}_h)^{-1}$ exists for all k sufficiently large and the norms $\|(I - \Pi_k\mathbb{D}_h)^{-1}\|$ are uniformly bounded (for all k sufficiently large). Therefore,

$$\big\|\big(\mathbf{I} - \mathbf{G}^{-1}\mathbf{D}_h\big)(v_k - \hat{v})\big\|_{\mathbf{G}} \ge \text{const} \times \|h_k - \hat{h}\| \tag{S.22}$$

holds for all k sufficiently large. Also notice that

$$\begin{aligned}
&\big\|\mathbf{G}^{-1}\mathbf{T}\hat{v} - \mathbf{G}^{-1}\mathbf{T}v_k - \mathbf{G}^{-1}\mathbf{D}_h(\hat{v} - v_k)\big\|_{\mathbf{G}} \\
&\quad = \big\|\Pi_k\mathbb{T}\hat{h} - \Pi_k\mathbb{T}h_k - \Pi_k\mathbb{D}_h(\hat{h} - h_k)\big\| \\
&\quad \le \big\|\mathbb{T}\hat{h} - \mathbb{T}h - \mathbb{D}_h(\hat{h} - h) - \big(\mathbb{T}h_k - \mathbb{T}h - \mathbb{D}_h(h_k - h)\big)\big\| \\
&\quad \le \big\|\mathbb{T}\hat{h} - \mathbb{T}h - \mathbb{D}_h(\hat{h} - h)\big\| + \big\|\mathbb{T}h_k - \mathbb{T}h - \mathbb{D}_h(h_k - h)\big\| \\
&\quad = o(1) \times \big(\|\hat{h} - h_k\| + \|h_k - h\|\big) + o(1) \times \|h - h_k\|,
\end{aligned} \tag{S.23}$$

where the first inequality is because Π_k is a (weak) contraction on L^2 and the final line is by Assumption 4.1(c). Substituting (S.22) and (S.23) into (S.21) and rearranging, we obtain

$$\big(1 - o(1)\big) \times \|h_k - \hat{h}\| \le O_p(\nu_{n,k}) + o_p(\tau_k).$$

Parts (a) and (b) follow by similar arguments to the proof of Lemma A.6. *Q.E.D.*

D.4. *Proofs for Appendix B*

PROOF OF PROPOSITION B.1: First note that

$$\sqrt{n}(\hat{L}-L) = \sqrt{n}\left(\log\hat{\rho} - \log\rho - \frac{1}{n}\sum_{t=0}^{n-1}\log m(X_t, X_{t+1}) + \mathbb{E}\big[\log m(X_t, X_{t+1})\big]\right)$$
$$= \frac{1}{\sqrt{n}}\sum_{t=0}^{n-1}\big(\rho^{-1}\psi_{\rho,t} - \psi_{lm,t}\big) + o_p(1),$$

where the second line is by display (24) and a delta-method type argument. The result now follows from the joint convergence in the statement of the proposition. *Q.E.D.*

PROOF OF PROPOSITION B.2: Similar arguments to the proof of Proposition B.1 yield

$$\sqrt{n}(\hat{L}-L) = \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\big(\rho^{-1}\psi_{\rho,t} + \rho^{-1}\phi^*_{k,t}\phi_{k,t+1}\big(m_t(\hat{\alpha}) - m_t(\alpha_0)\big)$$
$$- \big(\log m_t(\hat{\alpha}) - \log m_t(\alpha_0)\big) - \psi_{lm,t}\big) + o_p(1).$$

By similar arguments to the proof of Theorem 3.4, we may deduce

$$\frac{1}{\sqrt{n}}\sum_{t=0}^{n-1}\big(\rho^{-1}\phi^*_{k,t}\phi_{k,t+1}\big(m_t(\hat{\alpha}) - m_t(\alpha_0)\big) - \big(\log m_t(\hat{\alpha}) - \log m_t(\alpha_0)\big)\big)$$
$$= D_{\alpha,lm}\sqrt{n}(\hat{\alpha}-\alpha_0) + o_p(1),$$

where

$$D_{\alpha,lm} = \mathbb{E}\left[\left(\frac{\phi^*(X_t)\phi(X_{t+1})}{\rho} - \frac{1}{m(X_t, X_{t+1}, \alpha)}\right)\frac{\partial m(X_t, X_{t+1}, \alpha)}{\partial\alpha'}\right].$$

Substituting into the expansion for $\hat{L}$ and using Assumption 3.5(a) yields

$$\sqrt{n}(\hat{L}-L) = \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\big(\rho^{-1}\psi_{\rho,t} + D_{\alpha,lm}\psi_{\alpha,t} - \psi_{lm,t}\big) + o_p(1).$$

The result follows by the joint CLT assumed in the statement of the proposition. *Q.E.D.*

PROOF OF THEOREM B.1: We prove part (1) first. We first characterize the tangent space as in pp. 878–880 of Bickel and Kwon (2001) (their arguments trivially extend to $\mathbb{R}^d$-valued Markov processes). Let Q_2 denote the stationary distribution of (X_t, X_{t+1}). Consider the tangent space $\mathcal{H}_0 = \{h(X_t, X_{t+1}) : \mathbb{E}[h(X_t, X_{t+1})^2] < \infty$ and $\mathbb{E}[h(X_t, X_{t+1})|X_t = x] = 0$ almost surely$\}$ endowed with the $L^2(Q_2)$ norm. Take any bounded $h \in \mathcal{H}_0$ and consider the one-dimensional parametric model which we identify with the collection of transition probabilities $\{P_1^{\tau,h} : |\tau| \le 1\}$, where each transition probability $P_1^{\tau,h}$ is dominated by P_1 (the true transition probability) and is given by

$$\frac{\mathrm{d}P_1^{\tau,h}(x_{t+1}|x_t)}{\mathrm{d}P_1(x_{t+1}|x_t)} = e^{\tau h(x_t,x_{t+1}) - A(\tau,x_t)},$$

where

$$A(\tau, x_t) = \log\biggl(\int e^{\tau h(x_t, x_{t+1})} P_1(\mathrm{d}x_{t+1}|x_t)\biggr).$$

For each τ, we define the linear operator $\mathbb{M}^{(\tau,h)}$ on L^2 by

$$\mathbb{M}^{(\tau,h)}\psi(x_t) = \int m(x_t, x_{t+1})\psi(x_{t+1}) P_1^{\tau,h}(\mathrm{d}x_{t+1}|x_t).$$

Observe that

$$\bigl(\mathbb{M}^{(\tau,h)} - \mathbb{M}\bigr)\psi(x_t) = \int m(x_t, x_{t+1})\psi(x_{t+1})\bigl(e^{\tau h(x_t,x_{t+1}) - A(\tau,x_t)} - 1\bigr) P_1(\mathrm{d}x_{t+1}|x_t) \tag{S.24}$$

is a bounded linear operator on L^2 (since $\|\mathbb{M}\| < \infty$ and h is bounded). By Taylor's theorem,

$$e^{\tau h(x_t,x_{t+1}) - A(\tau,x_t)} - 1 = \tau h(x_t, x_{t+1}) + O\bigl(\tau^2\bigr), \tag{S.25}$$

where the $O(\tau^2)$ term is uniform in (x_t, x_{t+1}). It now follows by boundedness of h that $\|\mathbb{M}^{(\tau,h)} - \mathbb{M}\| = O(\tau)$. Similar arguments to the proof of Lemma A.1 imply that there exist $\epsilon > 0$ and $\bar{\tau} > 0$ such that the largest eigenvalue $\rho_{(\tau,h)}$ of $\mathbb{M}^{(\tau,h)}$ is simple and lies in the interval $(\rho - \epsilon, \rho + \epsilon)$ for each $\tau < \bar{\tau}$. Taking a perturbation expansion of $\rho_{(\tau,h)}$ about $\tau = 0$ (see, e.g., equation (3.6) on p. 89 of Kato (1980) which also applies in the infinite-dimensional case, as made clear in Section VII.1.5 of Kato (1980)):

$$\begin{aligned} \rho_{(\tau,h)} - \rho &= \bigl\langle\bigl(\mathbb{M}^{(\tau,h)} - \mathbb{M}\bigr)\phi, \phi^*\bigr\rangle + O\bigl(\tau^2\bigr) \\ &= \tau\mathbb{E}\bigl[m(X_t, X_{t+1})h(X_t, X_{t+1})\phi(X_{t+1})\phi^*(X_t)\bigr] + O\bigl(\tau^2\bigr) \\ &= \tau\int m(x_t, x_{t+1})\phi(x_{t+1})\phi^*(x_t)h(x_t, x_{t+1})\,\mathrm{d}Q_2(x_t, x_{t+1}) + O\bigl(\tau^2\bigr) \end{aligned} \tag{S.26}$$

under the normalization $\langle\phi, \phi^*\rangle = 1$, where the second line is by (S.24) and (S.25). Expression (S.26) shows that the derivative of $\rho_{(\tau,h)}$ at $\tau = 0$ is $\tilde{\psi}_\rho = m(x_t, x_{t+1}) \times \phi(x_{t+1})\phi^*(x_t)$.

As bounded functions are dense in $\mathcal{H}_0$, we have shown that ρ is differentiable relative to $\mathcal{H}_0$ with derivative $\tilde{\psi}_\rho$. The efficient influence function for ρ is the projection of $\tilde{\psi}_\rho$ onto $\mathcal{H}_0$, namely,

$$\tilde{\psi}_\rho(x_t, x_{t+1}) - \mathbb{E}\bigl[\tilde{\psi}_\rho(X_t, X_{t+1})|X_t = x_t\bigr] = \psi_\rho(x_t, x_{t+1}),$$

because $\mathbb{E}[\tilde{\psi}_\rho(X_t, X_{t+1})|X_t = x_t] = \phi^*(x_t)\mathbb{M}\phi(x_t) = \rho\phi(x_t)\phi^*(x_t)$. It follows that $V_\rho = \mathbb{E}[\psi_\rho(X_t, X_{t+1})^2]$ is the efficiency bound for ρ. A similar argument shows that $h'(\rho)\psi_\rho$ is the efficient influence function for $h(\rho)$.

We now prove part (2). The efficient influence function for L is

$$\psi_L = \rho^{-1}\psi_\rho - \psi_{\log m},$$

where $\psi_{\log m}$ is the efficient influence function for $\mathbb{E}[\log m(X_t, X_{t+1})]$. It is well known that

$$\psi_{\log m}(x_0, x_1) = l(x_0, x_1) + \sum_{t=0}^{\infty}\bigl(\mathbb{E}\bigl[l(X_{t+1}, X_{t+2})\big|X_1 = x_1\bigr] - \mathbb{E}\bigl[l(X_t, X_{t+1})\big|X_0 = x_0\bigr]\bigr),$$

where $l(x_t, x_{t+1}) = \log m(x_t, x_{t+1})$ (see, e.g., Greenwood and Wefelmeyer (1995)). It may be verified using the telescoping property of the above sum that $V_L = \mathbb{E}[\psi_L(X_0, X_1)^2]$. *Q.E.D.*

PROOF OF LEMMA B.1: Take $k \geq K$ from Lemma A.1 and work on the sequence of events upon which (S.16) holds, so that $\hat{\rho}$, $\hat{c}$, and $\hat{c}^*$ are uniquely defined by Lemma A.3.

Normalize $\hat{c}$, $\hat{c}^*$, c_k, and c_k^* so that $\|\hat{c}\|_{\mathbf{G}} = 1$, $\|c_k\|_{\mathbf{G}} = 1$, $\hat{c}'\mathbf{G}\hat{c}^* = 1$, and $c_k'\mathbf{G}c_k^* = 1$. Let $\mathbf{P} = c_k c_k^{*\prime}\mathbf{G}$ and $\widehat{\mathbf{P}} = \hat{c}\hat{c}^{*\prime}\mathbf{G}$. We then have $\operatorname{trace}(\widehat{\mathbf{P}}) = 1$, $\operatorname{trace}(\mathbf{P}) = 1$, $\hat{\rho} = \operatorname{trace}(\widehat{\mathbf{P}}\widehat{\mathbf{G}}^{-1}\widehat{\mathbf{M}})$, $\rho_k = \operatorname{trace}(\mathbf{P}\mathbf{G}^{-1}\mathbf{M})$, $\widehat{\mathbf{G}}^{-1}\widehat{\mathbf{M}}\widehat{\mathbf{P}} = \hat{\rho}\widehat{\mathbf{P}}$, and $\mathbf{G}^{-1}\mathbf{M}\mathbf{P} = \mathbf{P}\mathbf{G}^{-1}\mathbf{M} = \rho_k\mathbf{P}$. Now observe that

$$\begin{aligned}\hat{\rho} - \rho_k &= \operatorname{trace}\big(\widehat{\mathbf{P}}\widehat{\mathbf{G}}^{-1}\widehat{\mathbf{M}}\big) - \operatorname{trace}\big(\mathbf{P}\mathbf{G}^{-1}\mathbf{M}\big) \\ &= \operatorname{trace}\big((\widehat{\mathbf{P}} - \mathbf{P})\widehat{\mathbf{G}}^{-1}\widehat{\mathbf{M}}\big) + \operatorname{trace}\big(\mathbf{P}\big(\widehat{\mathbf{G}}^{-1}\widehat{\mathbf{M}} - \mathbf{G}^{-1}\mathbf{M}\big)\big).\end{aligned}$$

By addition and subtraction of terms, we have

$$\begin{aligned}&\operatorname{trace}\big((\widehat{\mathbf{P}} - \mathbf{P})\widehat{\mathbf{G}}^{-1}\widehat{\mathbf{M}}\big) \\ &\quad= \hat{\rho} - \hat{\rho}\operatorname{trace}(\mathbf{P}\widehat{\mathbf{P}}) + \operatorname{trace}\big(\mathbf{P}\widehat{\mathbf{G}}^{-1}\widehat{\mathbf{M}}(\widehat{\mathbf{P}} - \mathbf{I})\big) \\ &\quad= \hat{\rho}\operatorname{trace}\big(\mathbf{P}(\mathbf{I} - \widehat{\mathbf{P}})\big) + \operatorname{trace}\big(\mathbf{P}\widehat{\mathbf{G}}^{-1}\widehat{\mathbf{M}}(\widehat{\mathbf{P}} - \mathbf{I})\big) \\ &\quad= (\hat{\rho} - \rho_k)\operatorname{trace}\big(\mathbf{P}(\mathbf{I} - \widehat{\mathbf{P}})\big) + \rho_k\operatorname{trace}\big(\mathbf{P}(\mathbf{I} - \widehat{\mathbf{P}})\big) + \operatorname{trace}\big(\mathbf{P}\widehat{\mathbf{G}}^{-1}\widehat{\mathbf{M}}(\widehat{\mathbf{P}} - \mathbf{I})\big) \\ &\quad= (\hat{\rho} - \rho_k)\operatorname{trace}\big(\mathbf{P}(\mathbf{I} - \widehat{\mathbf{P}})\big) + \operatorname{trace}\big(\mathbf{P}\mathbf{G}^{-1}\mathbf{M}(\mathbf{I} - \widehat{\mathbf{P}})\big) + \operatorname{trace}\big(\mathbf{P}\widehat{\mathbf{G}}^{-1}\widehat{\mathbf{M}}(\widehat{\mathbf{P}} - \mathbf{I})\big) \\ &\quad= (\hat{\rho} - \rho_k)\operatorname{trace}\big(\mathbf{P}(\mathbf{I} - \widehat{\mathbf{P}})\big) + \operatorname{trace}\big(\mathbf{P}\big(\widehat{\mathbf{G}}^{-1}\widehat{\mathbf{M}} - \mathbf{G}^{-1}\mathbf{M}\big)(\widehat{\mathbf{P}} - \mathbf{I})\big),\end{aligned} \tag{S.27}$$

where

$$\big|\operatorname{trace}\big(\mathbf{P}(\mathbf{I} - \widehat{\mathbf{P}})\big)\big| = \big|c_k^{*\prime}\mathbf{G}(c_k - \widehat{\mathbf{P}}c_k)\big| \leq \big\|c_k^*\big\|_{\mathbf{G}}\|c_k - \widehat{\mathbf{P}}c_k\|_{\mathbf{G}}. \tag{S.28}$$

By the proof of Proposition 4.2 of Gobet, Hoffmann, and Reiß (2004) (setting $\widehat{\mathbf{P}} = P_\varepsilon$, $\widehat{\mathbf{G}}^{-1}\widehat{\mathbf{M}} = T_\varepsilon$, $\mathbf{G}^{-1}\mathbf{M} = T$, and Γ from the proof of Lemma A.1 as the boundary of $B(\kappa, \rho)$ in their notation) and similar arguments to the proof of Lemma A.4:

$$\|c_k - \widehat{\mathbf{P}}c_k\|_{\mathbf{G}} \lesssim \big\|\big(\widehat{\mathbf{G}}^{-1}\widehat{\mathbf{M}} - \mathbf{G}^{-1}\mathbf{M}\big)c_k\big\|_{\mathbf{G}} = O_p(\eta_{n,k}). \tag{S.29}$$

Moreover,

$$\big\|c_k^*\big\|_{\mathbf{G}} = \|\mathbf{P}\|_{\mathbf{G}} \leq \left\|\frac{1}{2\pi\mathrm{i}}\int_\Gamma \mathcal{R}\big(\mathbf{G}^{-1}\mathbf{M}, z\big)\,\mathrm{d}z\right\|_{\mathbf{G}}$$

(see Kato (1980, expression (6.19), p. 178)) which is $O(1)$ by display (S.14). By displays (S.28) and (S.29) and the fact that $\hat{\rho} - \rho_k = O_p(\eta_{n,k})$ (by Lemma A.4), we obtain

$$(\hat{\rho} - \rho_k)\operatorname{trace}\big(\mathbf{P}(\mathbf{I} - \widehat{\mathbf{P}})\big) = O_p\big(\eta_{n,k}^2\big). \tag{S.30}$$

Moreover,

$$\begin{aligned}\big|\operatorname{trace}\big(\mathbf{P}\big(\widehat{\mathbf{G}}^{-1}\widehat{\mathbf{M}} - \mathbf{G}^{-1}\mathbf{M}\big)(\widehat{\mathbf{P}} - \mathbf{I})\big)\big| &= \big|c_k^{*\prime}\mathbf{G}\big(\widehat{\mathbf{G}}^{-1}\widehat{\mathbf{M}} - \mathbf{G}^{-1}\mathbf{M}\big)(\widehat{\mathbf{P}} - \mathbf{I})c_k\big| \\ &\leq \big\|c_k^*\big\|_{\mathbf{G}}\big\|\widehat{\mathbf{G}}^{-1}\widehat{\mathbf{M}} - \mathbf{G}^{-1}\mathbf{M}\big\|_{\mathbf{G}}\|c_k - \widehat{\mathbf{P}}c_k\|_{\mathbf{G}} \\ &= O_p(\eta_{n,k,1} + \eta_{n,k,2})O_p(\eta_{n,k})\end{aligned} \tag{S.31}$$

by Lemma D.3(b) and display (S.29). It follows by (S.27), (S.30), and (S.31) that

$$\hat{\rho} - \rho_k = \operatorname{trace}\big(\mathbf{P}\big(\widehat{\mathbf{G}}^{-1}\widehat{\mathbf{M}} - \mathbf{G}^{-1}\mathbf{M}\big)\big) + O_p(\eta_{n,k,1} + \eta_{n,k,2}) \times O_p(\eta_{n,k}) + O_p\big(\eta_{n,k}^2\big).$$

Finally,

$$\begin{aligned}\operatorname{trace}\big(\mathbf{P}\big(\widehat{\mathbf{G}}^{-1}\widehat{\mathbf{M}} - \mathbf{G}^{-1}\mathbf{M}\big)\big) &= c_k^{*\prime}\mathbf{G}\big(\widehat{\mathbf{G}}^{-1}\widehat{\mathbf{M}} - \mathbf{G}^{-1}\mathbf{M}\big)c_k \\ &= \tilde{c}_k^{*\prime}\big(\big(\widehat{\mathbf{G}}^o\big)^{-1}\widehat{\mathbf{M}}^o - \mathbf{M}^o\big)\tilde{c}_k \\ &= \tilde{c}_k^{*\prime}\big(\widehat{\mathbf{M}}^o - \widehat{\mathbf{G}}^o\mathbf{M}^o\big)\tilde{c}_k + O_p\big(\eta_{n,k,1} \times (\eta_{n,k,1} + \eta_{n,k,2})\big)\end{aligned}$$

by Lemma D.3(a) and the fact that $\|\tilde{c}_k^*\| = \|c_k^*\|_{\mathbf{G}} = O(1)$. The result follows by noting that

$$\tilde{c}_k^{*\prime}\big(\widehat{\mathbf{M}}^o - \widehat{\mathbf{G}}^o\mathbf{M}^o\big)\tilde{c}_k = c_k^{*\prime}(\widehat{\mathbf{M}} - \rho_k\widehat{\mathbf{G}})c_k$$

and that $\eta_{n,k}$ is of at least as small order as $\eta_{n,k,1}$ and $\eta_{n,k,2}$ (cf. Lemma D.3(a)). *Q.E.D.*

REFERENCES

AKIAN, M., S. GAUBERT, AND R. NUSSBAUM (2016): "Uniqueness of the Fixed Point of Nonexpansive Semidifferentiable Maps," *Transactions of the American Mathematical Society*, 368 (2), 1271–1320. [9]

BICKEL, P. J., AND J. KWON (2001): "Inference for Semiparametric Models: Some Questions and an Answer" (with discussion), *Statistica Sinica*, 11, 863–960. [18]

BILLINGSLEY, P. (1961): "The Lindeberg–Lévy Theorem for Martingales," *Proceedings of the American Mathematical Society*, 12 (1), 788–792. [6]

CHATELIN, F. (1983): *Spectral Approximation of Linear Operators*. New York: Academic Press. [12]

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. [1,2]

GOBET, E., M. HOFFMANN, AND M. REISS (2004): "Nonparametric Estimation of Scalar Diffusions Based on Low Frequency Data," *The Annals of Statistics*, 32, 2223–2253. [2,11,12,14,20]

GREENWOOD, P. E., AND W. WEFELMEYER (1995): "Efficiency of Empirical Estimators for Markov Chains," *The Annals of Statistics*, 32 (1), 132–143. [20]

HANSEN, B. E. (2015): "A Unified Asymptotic Distribution Theory for Parametric and Non-Parametric Least Squares," Working Paper, University of Wisconsin. [1]

KATO, T. (1980): *Perturbation Theory for Linear Operators*. Berlin: Springer-Verlag. [10,14,19,20]

KRASNOSEL'SKII, M. A., G. M. VAINIKKO, P. P. ZABREIKO, Y. B. RUTITSKII, AND V. Y. STETSENKO (1972): *Approximate Solution of Operator Equations*. Groningen: Wolters-Noordhoff. [15-17]

NEWEY, W. K. (1997): "Convergence Rates and Asymptotic Normality for Series Estimators," *Journal of Econometrics*, 79 (1), 147–168. [1]

NEWEY, W. K., AND D. MCFADDEN (1994): "Large Sample Estimation and Hypothesis Testing," in *Handbook of Econometrics*, Vol. 4. Elsevier, 2111–2245, Chapter 36. [7]

SCHAEFER, H. H. (1974): *Banach Lattices and Positive Operators*. Berlin: Springer-Verlag. [4]

SCHUMAKER, L. L. (2007): *Spline Functions: Basic Theory*. Cambridge: Cambridge University Press. [4]

VAN DER VAART, A. W. (1998): *Asymptotic Statistics*. Cambridge University Press. [8]

Co-editor Lars Peter Hansen handled this manuscript.

Manuscript received 5 May, 2013; final version accepted 9 May, 2017; available online 22 May, 2017.