Online Appendix for Nonparametric Stochastic Discount Factor Decomposition

Timothy M. Christensen

May 19, 2017

This Online Appendix contains material to support the paper "Nonparametric Stochastic Discount Factor Decomposition". Appendix E presents additional simulation evidence. Appendix F provides further details on the relation between the identification and existence conditions in Section 2.3 and the identification and existence conditions in Hansen and Scheinkman (2009) and Borovička et al. (2016). Appendix G presents proofs of results in Appendix C of the supplementary material and this online appendix.

E Additional Monte Carlo evidence

This section presents additional simulation results using a cubic B-spline basis of dimension $k = 8$ for the Monte Carlo design described in Section 5 of the main text. The knots of the B-splines were placed evenly at the empirical quantiles of the data. As with the results obtained using Hermite polynomials, the simulation results were reasonably insensitive to the dimension of the sieve space.

Tables 4 and 5 present bias and RMSE of the estimators across simulations. Figures 5a–5e present (pointwise) confidence intervals for ϕ, ϕ^* and χ computed across simulations of different sample sizes.

F Additional results on identification

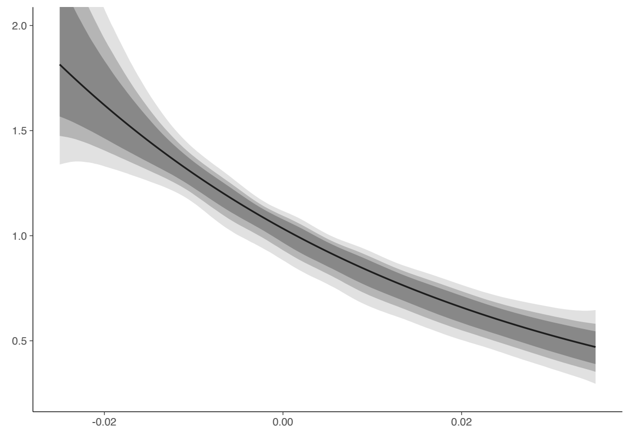
In this appendix we discuss separately existence and identification, and compare the conditions in the present paper with the stochastic stability conditions in Hansen and Scheinkman (2009) (HS hereafter) and Borovička et al. (2016) (BHS hereafter).

	n	Power Utility $\hat{\phi}$	$\hat{\phi}^*$	Recursive Preferences $\hat{\phi}$	$\hat{\phi}^*$	$\hat{\chi}$
Bias	400	0.0144	0.0141	0.0009	0.0241	0.0116
	800	0.0113	0.0132	0.0011	0.0190	0.0086
	1600	0.0078	0.0101	0.0010	0.0145	0.0057
	3200	0.0049	0.0068	0.0009	0.0128	0.0034
RMSE	400	0.1106	0.1334	0.0283	0.3479	0.0988
	800	0.0851	0.1043	0.0270	0.3151	0.0734
	1600	0.0650	0.0814	0.0235	0.2747	0.0547
	3200	0.0500	0.0627	0.0222	0.1702	0.0414

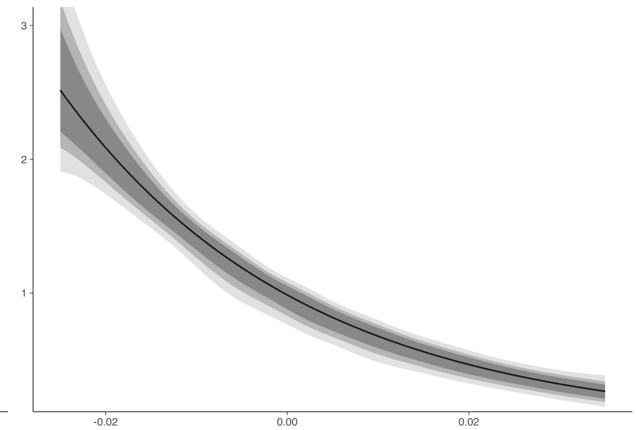
Table 4: Simulation results for $\hat{\phi}$, $\hat{\phi}^*$ and $\hat{\chi}$ with a cubic B-spline sieve of dimension $k = 8$.

	n	Power Utility $\hat{\rho}$	$\hat{y}$	$\hat{L}$	Recursive Preferences $\hat{\rho}$	$\hat{y}$	$\hat{L}$	$\hat{\lambda}$
Bias	400	0.0036	-0.0030	0.0030	0.0010	-0.0009	0.0031	0.0028
	800	0.0027	-0.0024	0.0024	0.0011	-0.0011	0.0027	0.0019
	1600	0.0019	-0.0017	0.0017	0.0010	-0.0009	0.0020	0.0013
	3200	0.0012	-0.0011	0.0011	0.0007	-0.0006	0.0013	0.0008
RMSE	400	0.0345	0.0330	0.0272	0.0154	0.0130	0.0305	0.0348
	800	0.0254	0.0244	0.0206	0.0155	0.0133	0.0244	0.0209
	1600	0.0190	0.0182	0.0157	0.0163	0.0136	0.0208	0.0153
	3200	0.0142	0.0135	0.0118	0.0148	0.0123	0.0165	0.0110

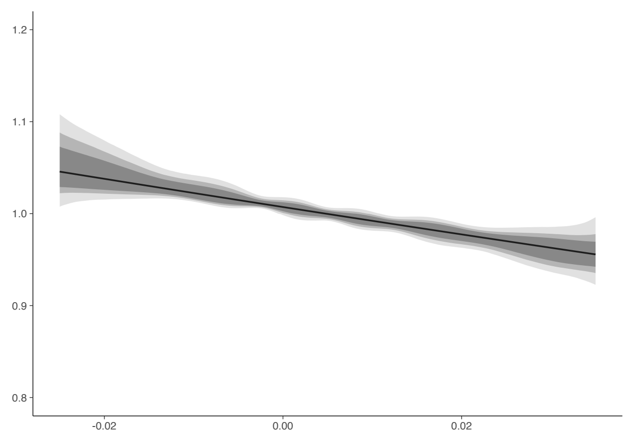
Table 5: Simulation results for $\hat{\rho}$, $\hat{y}$, $\hat{L}$ and $\hat{\lambda}$ with a cubic B-spline sieve of dimension $k = 8$.



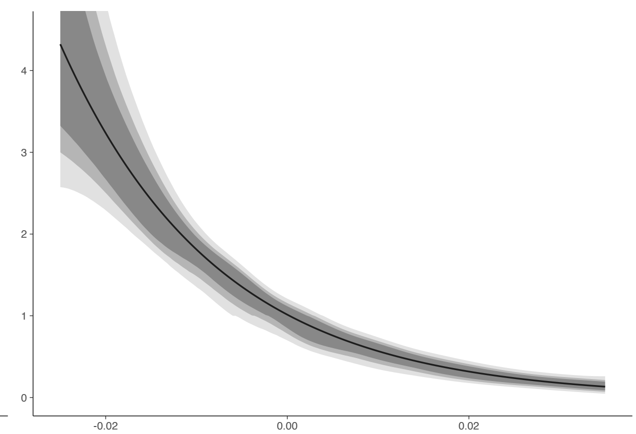
(a) $\hat{\phi}(x)$ for power utility



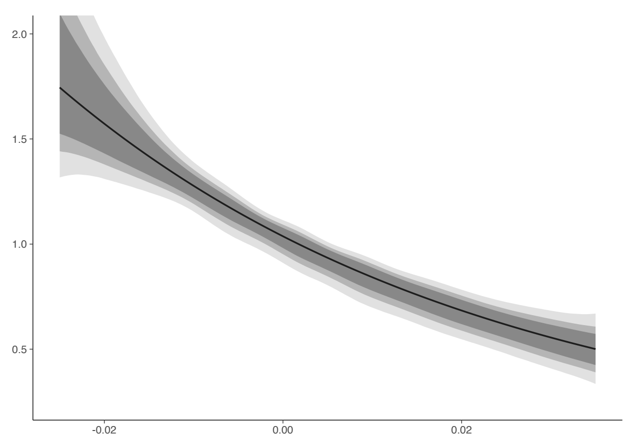
(b) $\hat{\phi}^*(x)$ for power utility



(c) $\hat{\phi}(x)$ for recursive preferences



(d) $\hat{\phi}^*(x)$ for recursive preferences



(e) $\hat{\chi}(x)$ for recursive preferences

Figure 5: Simulation results for a cubic B-spline basis with $k = 8$. Panels (a)–(d) display pointwise 90% confidence intervals for ϕ and ϕ^* across simulations (light, medium and dark correspond to $n = 400$, 800, and 1600 respectively; the true ϕ and ϕ^* plotted as solid lines). Panel (e) displays results for the positive eigenfunction χ of the continuation value operator

F.1 Identification

Assumption F.1 *Let the following hold:*

(a) $\mathbb{M}$ *is bounded*

(b) There exists positive functions $\phi, \phi^* \in L^2$ *and a positive scalar* ρ *such that* (ρ, ϕ) *solves (6) and* (ρ, ϕ^*) *solves (7)*

(c) $\mathbb{M}\psi$ *is positive for each non-negative* $\psi \in L^2$ *that is not identically zero.*

Note that no compactness or power-compactness condition appears in Assumption F.1.

Proposition F.1 *Let Assumption F.1 hold. Then: the functions* ϕ *and* ϕ^* *are the unique solutions (in* L^2*) to (6) and (7), respectively.*

We now compare the identification results with those in HS and BHS. Some of HS's conditions related to the generator of the semigroup of conditional expectation operators $\widetilde{\mathbb{E}}[\cdot|X_t = x]$ under the change of conditional probability induced by M_t^P, namely:

$$\widetilde{\mathbb{E}}[\psi(X_{t+\tau})|X_t = x] := \mathbb{E}\left[\frac{M_{t+\tau}^P}{M_t^P}\psi(X_{t+\tau})\middle|X_t = x\right]. \tag{OA.1}$$

In discrete-time environments, both multiplicative functionals and semigroups are indexed by non-negative integers. Therefore, the "generator" in discrete-time is just the single-period distorted conditional expectation operator $\psi \mapsto \widetilde{\mathbb{E}}[\psi(X_{t+1})|X_t = \cdot\,]$.

The following are discrete-time versions of Assumptions 6.1, 7.1, 7.2, 7.3, and 7.4 in HS.

Condition F.1 *(a)* $\{M_t^P : t \in T\}$ *is a positive multiplicative functional*

(b) There exists a probability measure $\hat{\varsigma}$ *such that*

$$\int \widetilde{\mathbb{E}}[\psi(X_{t+1})|X_t = x]\,\mathrm{d}\hat{\varsigma}(x) = \int \psi(x)\,\mathrm{d}\hat{\varsigma}(x)$$

for all bounded measurable $\psi : \mathcal{X} \to \mathbb{R}$

(c) For any $\Lambda \in \mathscr{X}$ *with* $\hat{\varsigma}(\Lambda) > 0$,

$$\widetilde{\mathbb{E}}\left[\sum_{t=1}^{\infty} 1\!\!1\{X_t \in \Lambda\}\middle| X_0 = x\right] > 0$$

for all $x \in \mathcal{X}$

(d) *For any* $\Lambda \in \mathscr{X}$ *with* $\hat{\varsigma}(\Lambda) > 0$,

$$\widetilde{\mathbb{P}}\left(\sum_{t=1}^{\infty} 1\!\!1\{X_t \in \Lambda\} = \infty \middle| X_0 = x \right) = 1$$

for all $x \in \mathcal{X}$, *where*

$$\widetilde{\mathbb{P}}(\{X_s\}_{s=0}^t \in A | X_0 = x) = \int \mathbb{E}[(M_t^P/M_0^P) 1\!\!1\{\{X_s\}_{s=0}^t \in A\} | X_0 = x] \, \mathrm{d}\hat{\varsigma}(x)$$

for each $A \in \mathcal{F}_t$.

Condition F.1(a) is satisfied by construction of M^P in (8). For Condition F.1(b), let ϕ and ϕ^* be as in Assumption F.1(b) and normalize ϕ^* such that $\mathbb{E}[\phi(X_t)\phi^*(X_t)] = 1$. Under this normalization we can define a probability measure $\hat{\varsigma}$ by $\hat{\varsigma}(A) = \mathbb{E}[\phi(X_t)\phi^*(X_t) 1\!\!1\{X_t \in A\}]$ for all $A \in \mathscr{X}$. Proposition F.3 below shows that this probability measure is precisely the measure used to define the unconditional expectation $\widetilde{\mathbb{E}}$ in the long-run approximation (9). Recall that Q is the stationary distribution of X. We then have:

$$\begin{aligned}
&\int \widetilde{\mathbb{E}}[\psi(X_{t+1}) | X_t = x] \, \mathrm{d}\hat{\varsigma}(x) \\
&= \int \mathbb{E}\left[\rho^{-1} m(X_t, X_{t+1}) \frac{\phi(X_{t+1})}{\phi(X_t)} \psi(X_{t+1}) \middle| X_t = x \right] \phi(x)\phi^*(x) \, \mathrm{d}Q(x) \\
&= \rho^{-1} \mathbb{E}\left[\phi^*(X_t)(\mathbb{M}(\phi\psi)(X_t)) \right] \\
&= \rho^{-1} \mathbb{E}\left[((\mathbb{M}^*\phi^*)(X_{t+1}))\phi(X_{t+1})\psi(X_{t+1}) \right] \\
&= \mathbb{E}[\phi^*(X_{t+1})\phi(X_{t+1})\psi(X_{t+1})] = \int \psi(x) \, \mathrm{d}\hat{\varsigma}(x) \, .
\end{aligned}$$

Therefore, Condition F.1(b) is satisfied. A similar derivation is reported for continuous-time semigroups in an preliminary 2005 draft of HS with Q replaced by an arbitrary measure.

For Condition F.1(c), note that $\hat{\varsigma}(\Lambda) > 0$ implies $Q(\Lambda) > 0$ under our construction of $\hat{\varsigma}$. Therefore, $\hat{\varsigma}(\Lambda) > 0$ implies $\phi(x) 1\!\!1\{x \in \Lambda\}$ is positive on a set of positive Q measure. Moreover, by definition of $\widetilde{\mathbb{E}}$ we have:

$$\begin{aligned}
\widetilde{\mathbb{E}}\left[\sum_{t=1}^{\infty} 1\!\!1\{X_t \in \Lambda\} \middle| X_0 = x \right] &= \frac{1}{\phi(x)} \sum_{t=1}^{\infty} \rho^{-t} \mathbb{M}_t(\phi(\cdot) 1\!\!1\{\cdot \in \Lambda\})(x) \\
&\geq \frac{1}{\phi(x)} \sum_{t=1}^{\infty} \lambda^{-t} \mathbb{M}_t(\phi(\cdot) 1\!\!1\{\cdot \in \Lambda\})(x)
\end{aligned}$$

for any $\lambda \geq r(\mathbb{M})$ where $r(\mathbb{M})$ denotes the spectral radius of $\mathbb{M}$. Assumption F.1(c) implies $\mathbb{M}$ is irreducible and, by definition of irreducibility, $\sum_{t=1}^{\infty} \lambda^{-t} \mathbb{M}_t(\phi(\cdot) 1\!\!1\{\cdot \in \Lambda\})(x) > 0$ (almost everywhere) holds for $\lambda > r(\mathbb{M})$. Therefore, Assumption F.1(c) implies Condition F.1(c), up to the "almost everywhere" qualification.

Part (d) is a Harris recurrence condition which does not translate clearly in terms of the operator $\mathbb{M}$. When combined with existence of an invariant measure and irreducibility (Condition F.1(b) and (c), respectively), it ensures both uniqueness of $\hat{\varsigma}$ as the invariant measure for the distorted expectations as well as ϕ-ergodicity, i.e.,

$$\lim_{\tau \to \infty} \sup_{0 \leq \psi \leq \phi} \left| \widetilde{\mathbb{E}} \left[\frac{\psi(X_{t+\tau})}{\phi(X_{t+\tau})} \middle| X_t = x \right] - \int \frac{\psi(x)}{\phi(x)} \, \mathrm{d}\hat{\varsigma}(x) \right| = 0 \tag{OA.2}$$

(almost everywhere) where the supremum is taken over all measurable ψ such that $0 \leq \psi \leq \phi$ (Meyn and Tweedie, 2009, Proposition 14.0.1). Result (OA.2) is a discrete-time version of Proposition 7.1 in HS, which they use to establish identification of ϕ. Assumption F.1 alone is not enough to obtain a convergence result like (OA.2). On the other hand, the conditions in the present paper assume existence of ϕ^* whereas no positive eigenfunction of the adjoint of $\mathbb{M}$ is guaranteed under the conditions in HS. Indeed, for non-stationary environments it is not even clear how to restrict the class of functions appropriately to define an adjoint (for instance, HS do not appear to restrict ϕ to belong to a Banach space). This suggests the Harris recurrence condition is of a very different nature from Assumption F.1.

BHS assume that X is ergodic under the $\widetilde{\mathbb{P}}$ probability measure, for which Conditions F.1(b)–(d) are sufficient. Also notice that Condition F.1(a) is satisfied by construction in BHS.

The identification results in HS and the proof of proposition 3.3 in BHS shows that uniqueness is established in the space of functions ψ for which $\widetilde{\mathbb{E}}[\psi(X_t)/\phi(X_t)]$ is finite, where $\widetilde{\mathbb{E}}$ denotes expectation under the stationary distribution corresponding to (OA.1). Under Assumption F.1, their result establishes identification in the space of functions ψ for which

$$\widetilde{\mathbb{E}}[\psi(X_t)/\phi(X_t)] = \mathbb{E}[\psi(X_t)\phi^*(X_t)]$$

is finite. The right-hand side is finite for all $\psi \in L^2$ (by Cauchy-Schwarz). So in this sense the identification result in HS and BHS applies to a larger class of functions than our result.

F.2 Existence

We obtain the following existence result by replacing Assumption F.1(b)(c) by the slightly stronger quasi-compactness and positivity conditions in Assumption 2.1. The following result is essentially Theorems 6 and 7 of Sasser (1964).[15] Say that $\mathbb{M}$ is *quasi-compact* if $\mathbb{M}$ is bounded and there exists $\tau \in T$ and a bounded linear operator $\mathbb{V}$ such that $\mathbb{M}_\tau - \mathbb{V}$ is compact and $r(\mathbb{V}) < r(\mathbb{M})^\tau$. Quasi-compactness of $\mathbb{M}$ is implied by Assumption 2.1.

Proposition F.2 *Let Assumption 2.1(a) hold and let $\mathbb{M}$ be quasi-compact. Then:*

(a) There exists positive functions $\phi, \phi^* \in L^2$ and a positive scalar ρ such that (ρ, ϕ) solves (6) and (ρ, ϕ^*) solves (7).

(b) The functions ϕ and ϕ^* are the unique solutions (in L^2) to (6) and (7), respectively.

(c) The eigenvalue ρ is simple and isolated and it is the largest eigenvalue of $\mathbb{M}$.

A similar existence result to part (a) was presented in a 2005 preliminary version of HS. For that result, HS assumed that $r(\mathbb{M})$ was positive and that the (continuous-time) semigroup of operators had an element which was compact. The further properties of ρ that we establish in part (c) of Proposition F.2 are essential to our derivation of the large-sample theory. A similar proposition was derived under different conditions in Christensen (2015).

HS establish existence of ϕ in possibly non-stationary, continuous-time environments by appealing to the theory of ergodic Markov processes. Equivalent conditions for discrete-time environments are now presented and compared with our identification conditions. As with the identification conditions, we use analogues of generators and resolvents for discrete-time semigroups where appropriate.

Condition F.2 *(a) There exists a function $V : \mathcal{X} \to \mathbb{R}$ with $V \geq 1$ and a finite constant $\underline{a} > 0$ such that $\mathbb{M}V(x) \leq \underline{a}V(x)$ for all $x \in \mathcal{X}$*

[15] I thank an anonymous referee for bringing Theorems 6 and 7 of Sasser (1964) to my attention. Theorems 6 and 7 of Sasser (1964) replace Assumption 2.1(a) in Proposition F.2 by the condition that $\mathbb{M}$ is *quasi-positive*, i.e. for each non-negative ψ and ψ^* in L^2 that are not identically zero there exists $\tau \in T$ such that $\langle \psi^*, \mathbb{M}_\tau \psi \rangle > 0$. Notice that quasi-compactness also requires that $r(\mathbb{M}) > 0$. Assumption 2.1(a) is sufficient for these two conditions (i.e. quasi-positivity and $r(\mathbb{M}) > 0$). The condition $r(\mathbb{M}) > 0$ together with power-compactness of $\mathbb{M}$ (Assumption 2.1(b)) is sufficient for quasi-compactness.

(b) *There exists a measure ν on $(\mathcal{X}, \mathscr{X})$ such that $\mathbb{J}1\!\!1\{\cdot \in \Lambda\}(x) > 0$ for any $\Lambda \in \mathscr{X}$ with $\nu(\Lambda) > 0$, where $\mathbb{J}$ is given by*

$$\mathbb{J}\psi(x) = \sum_{t=0}^{\infty} a^{-(t+1)} \frac{\mathbb{M}_t(V\psi)(x)}{V(x)}$$

for $a > \underline{a}$

(c) *The operators $\mathbb{J}$ and $\mathbb{K}$ are bounded, where $\mathbb{K}$ is given by*

$$\mathbb{K}\psi(x) = \sum_{t=0}^{\infty} \lambda^{-t}((\mathbb{J} - s \otimes \nu)^t \psi)(x)$$

where $s : \mathcal{X} \to \mathbb{R}_+$ is such that $\int s \, \mathrm{d}\nu > 0$ and $\mathbb{J}\psi(x) \geq s(x) \int \psi \, \mathrm{d}\nu$ for all $\psi \geq 0$ (s exists by part (b)), $(s \otimes \nu)\psi(x) := s(x) \int \psi \, \mathrm{d}\nu$, and $\lambda \in \sigma(\mathbb{J})$.

HS show that $\mathbb{K}s$ is a positive eigenfunction of $\mathbb{M}$ under the preceding conditions (see their Lemma D.3). Condition F.2(b) is satisfied under Assumption 2.1 with $\nu = Q$ whenever $a > r(\mathbb{M})$. To see this, take $\Lambda \in \mathscr{X}$ with $Q(\Lambda) > 0$ and observe that:

$$\sum_{t=1}^{\infty} a^{-t} \mathbb{M}_t(V(\cdot)1\!\!1\{\cdot \in \Lambda\})(x) \geq \sum_{t=1}^{\infty} a^{-t} \mathbb{M}_t 1\!\!1\{\cdot \in \Lambda\} > 0$$

(almost everywhere) where the first inequality is by positivity and the second is by irreducibility. It follows that $\mathbb{J}1\!\!1\{\cdot \in \Lambda\}(x) > 0$ (almost everywhere). This verifies part (b), up to the "almost everywhere" qualification.

On the other hand, Conditions F.2(a)(c) seem quite different from the conditions of Proposition F.2. For instance, Assumption 2.1 does not presume existence of the function V but imposes a quasi-compactness condition. HS do not restrict the function space for $\mathbb{M}$ ex ante so there is no notion of a bounded or power-compact operator on the space to which ϕ belongs. The requirement that $\mathbb{K}$ be bounded (or the sufficient conditions for this provided in HS) do not seem to translate clearly in terms of the operator $\mathbb{M}$.

F.3 Long-run pricing

We now present a version of the long-run pricing approximation of HS that holds under our existence and identification conditions. We impose the normalization $\mathbb{E}[\phi(X_t)\phi^*(X_t)] = 1$

and define the operator $(\phi \otimes \phi^*) : L^2 \to L^2$ by:

$$(\phi \otimes \phi^*)\psi(x) = \phi(x) \int \phi^* \psi \, \mathrm{d}Q \, .$$

Proposition F.3 *Let Assumption 2.1 hold. Then: there exists $c > 0$ such that:*

$$\|\rho^{-\tau}\mathbb{M}_\tau - (\phi \otimes \phi^*)\| = O(e^{-c\tau})$$

as $\tau \to \infty$.

Proposition F.3 is similar to Proposition 7.4 in HS. Proposition F.3 establishes convergence of $\rho^{-\tau}\mathbb{M}_\tau$ to $(\phi\otimes\phi^*)$, with the approximation error vanishing exponentially in the payoff horizon n. A similar proposition (without the rate of convergence) was reported in a 2005 draft of HS. There, HS assumed directly that the distorted conditional expectations converged to an unconditional expectation characterized by ϕ, ϕ^*, and an arbitrary measure. Proposition F.3 shows that in stationary environments the unconditional expectation $\widetilde{\mathbb{E}}[\psi(X_t)/\phi(X_t)]$ appearing in the long-run approximation (9) is characterized by ϕ, ϕ^* and Q, namely:

$$\widetilde{\mathbb{E}}\left[\frac{\psi(X_t)}{\phi(X_t)}\right] = \mathbb{E}[\psi(X_t)\phi^*(X_t)] \, .$$

G Proofs of results in Appendices C and F

G.1 Proofs for Appendix C.1

Proof of Lemma C.1. Lemma 2.2 of Chen and Christensen (2015) gives the bound $\|\widehat{\mathbf{G}}^o - \mathbf{I}\| = O_p(\xi_k(\log n)/\sqrt{n})$. We first prove that $\|\widehat{\mathbf{M}}^o - \mathbf{M}^o\| = O_p(\xi_k^{1+2/r}(\log n)/\sqrt{n})$.

Let $\{T_n : n \geq 1\}$ be a sequence of positive constants to be defined below. Let $\tilde{b}^k = \mathbf{G}^{-1/2}b^k$ be the orthogonalized basis functions and let $\Xi_{t,n} = n^{-1}\tilde{b}^k(X_t)m(X_t, X_{t+1})\tilde{b}^k(X_{t+1})'$. Write:

$$\begin{aligned}
\widehat{\mathbf{M}}^o - \mathbf{M}^o &= \sum_{t=0}^{n-1} \Xi_{t,n}^{trunc} + \sum_{t=0}^{n-1} \Xi_{t,n}^{tail} \quad \text{where} \\
\Xi_{t,n}^{trunc} &= \Xi_{t,n}1\!\!1\{\|\Xi_{t,n}\| \leq T_n/n\} - \mathbb{E}[\Xi_{t,n}1\!\!1\{\|\Xi_{t,n}\| \leq T_n/n\}] \\
\Xi_{t,n}^{tail} &= \Xi_{t,n}1\!\!1\{\|\Xi_{t,n}\| > T_n/n\} - \mathbb{E}[\Xi_{t,n}1\!\!1\{\|\Xi_{t,n}\| > T_n/n\}] \, .
\end{aligned}$$

Note $\mathbb{E}[\Xi_{t,n}^{trunc}] = 0$ and $\|\Xi_{t,n}^{trunc}\| \leq 2n^{-1}T_n$ by construction. Let $S^{k-1} = \{u \in \mathbb{R}^k : \|u\| = 1\}$. For any $u, v \in S^{k-1}$ and any $0 \leq t, s \leq n-1$, we have:

$$\begin{aligned}
|u'\mathbb{E}[\Xi_{t,n}^{trunc}(\Xi_{s,n}^{trunc})']v| &\lesssim \frac{\xi_k^2}{n^2}\mathbb{E}[|u'\tilde{b}^k(X_t)m(X_t, X_{t+1})m(X_s, X_{s+1})\tilde{b}^k(X_s)'v|] \\
&\leq \frac{\xi_k^2}{n^2}\mathbb{E}[|m(X_t, X_{t+1})|^r]^{2/r} \times \mathbb{E}[|(u'\tilde{b}^k(X_t))|^q]^{1/q} \times \mathbb{E}[|(v'\tilde{b}^k(X_s))|^q]^{1/q} \\
&\lesssim \frac{\xi_k^2}{n^2}\mathbb{E}[|(u'\tilde{b}^k(X_t))|^q]^{1/q} \times \mathbb{E}[|(v'\tilde{b}^k(X_s))|^q]^{1/q}
\end{aligned}$$

where the second line is by Hölder's inequality choosing q such that $1 = \frac{2}{r} + \frac{2}{q}$ and the third is because $\mathbb{E}[|m(X_t, X_{t+1})|^r] < \infty$. Since $E[(\tilde{b}^k(X_0)'u)^2] = \|u\|^2 = 1$ for any $u \in S^{k-1}$, we have:

$$\mathbb{E}[|(u'\tilde{b}^k(X_t))|^q]^{1/q} \leq (\xi_k^{q-2}\mathbb{E}[(u'\tilde{b}^k(X_t))^2])^{1/q} = \xi_k^{1-2/q}$$

and so:

$$\|\mathbb{E}[\Xi_{t,n}^{trunc}(\Xi_{s,n}^{trunc})']\| \lesssim \sup_{u,v \in S^{k-1}} |u'\mathbb{E}[\Xi_{t,n}^{trunc}(\Xi_{s,n}^{trunc})']v| = O(\xi_k^{2+4/r}/n^2)\,.$$

The same argument gives $\|\mathbb{E}[(\Xi_{t,n}^{trunc})'\Xi_{s,n}^{trunc}]\| = O(\xi_k^{2+4/r}/n^2)$. By Corollary 4.2 of Chen and Christensen (2015):

$$\left\| \sum_{t=0}^{n-1} \Xi_{t,n}^{trunc} \right\| = O_p(\xi_k^{1+2/r}(\log n)/\sqrt{n})$$

provided $T_n(\log n)/n = o(\xi_k^{1+2/r}/\sqrt{n})$.

Now consider the remaining term. If m is bounded we can set $\Xi_{t,n}^{tail} \equiv 0$ by taking $T_n = C\xi_k^2$ for sufficiently large C. Otherwise, by the triangle and Jensen inequalities:

$$\begin{aligned}
\mathbb{E}\left[\left\| \sum_{t=0}^{n-1} \Xi_{t,n}^{tail} \right\|\right] &\leq 2n\mathbb{E}[\|\Xi_{t,n}\| 1\!\!1\{\|\Xi_{t,n}\| > T_n/n\}] \\
&\leq \frac{2n^r}{T_n^{r-1}}\mathbb{E}[\|\Xi_{t,n}\|^r 1\!\!1\{\|\Xi_{t,n}\| > T_n/n\}] \leq \frac{2\xi_k^{2r}}{T_n^{r-1}}\mathbb{E}[|m(X_0, X_1)|^r]\,.
\end{aligned}$$

By Markov's inequality:

$$\left\| \sum_{t=0}^{n-1} \Xi_{t,n}^{tail} \right\| = O_p(\xi_k^{2r}/T_n^{r-1})\,.$$

choosing T_n so that $\xi_k^{2r}/T_n^{r-1} \asymp \xi_k^{1+2/r}(\log n)/\sqrt{n}$, we obtain:

$$\left\| \sum_{t=0}^{n-1} \Xi_{t,n}^{tail} \right\| = O_p(\xi_k^{1+2/r}(\log n)/\sqrt{n})\,.$$

The condition $T_n(\log n)/n = o(\xi_k^{1+2/r}/\sqrt{n})$ is, with this choice of T_n, equivalent to the condition $(\xi_k(\log n)/\sqrt{n})^{(r-2)/(r-1)} = o(1)$, which holds because $\xi_k(\log n)/\sqrt{n} = o(1)$ and $r > 2$. We have therefore shown that $\|\widehat{\mathbf{M}}^o - \mathbf{M}^o\| = O_p(\xi_k^{1+2/r}(\log n)/\sqrt{n})$.

Result (1) now follows from Lemma D.3(b), noting that

$$\|(\widehat{\mathbf{G}}^o)^{-1}\widehat{\mathbf{M}}^o - \mathbf{M}^o\| = O_p(\xi_k^{1+2/r}(\log n)/\sqrt{n})$$

which is $o_p(1)$ under the condition $\xi_k^{1+2/r}(\log n)/\sqrt{n} = o(1)$. Result (2) follows from Result (1) and definition of the operator norm. Result (3) is immediate from the fact that $\|\widehat{\mathbf{G}}^o - \mathbf{I}\| = O_p(\xi_k(\log n)/\sqrt{n})$ and $\|\widehat{\mathbf{M}}^o - \mathbf{M}^o\| = O_p(\xi_k^{1+2/r}(\log n)/\sqrt{n})$. ■

Proof of Lemma C.2. Similar arguments to the proof of Lemmas 4.8 and 4.12 of Gobet et al. (2004) give the bounds $\|\widehat{\mathbf{G}}^o - \mathbf{I}\| = O_p(\xi_k\sqrt{k/n})$, $\|(\widehat{\mathbf{G}}^o - \mathbf{I})\tilde{c}_k\| = O_p(\xi_k/\sqrt{n})$, and $\|\tilde{c}_k^{*\prime}(\widehat{\mathbf{G}}^o - I)\| = O_p(\xi_k/\sqrt{n})$. We first establish analogous bounds for $\widehat{\mathbf{M}}^o$.

Let $u_1, \ldots, u_k$ be an orthonormal basis for $\mathbb{R}^k$. Then:

$$\begin{aligned}
\mathbb{E}[\|\widehat{\mathbf{M}}^o - \mathbf{M}^o\|^2] &\leq \sum_{l=1}^{k} \mathbb{E}[\|(\widehat{\mathbf{M}}^o - \mathbf{M}^o)u_l\|^2] \\
&= \sum_{l=1}^{k}\sum_{j=1}^{k} \mathrm{Var}\left[\frac{1}{n}\sum_{t=1}^{n}(\tilde{b}_{kj}(X_t)^2 m(X_t, X_{t+1})^2(\tilde{b}^k(X_{t+1})'u_l)\right].
\end{aligned}$$

Now, by the covariance inequality for rho-mixing processes:

$$\begin{aligned}
\mathbb{E}[\|\widehat{\mathbf{M}}^o - \mathbf{M}^o\|^2] &\leq \frac{C}{n}\sum_{l=1}^{k}\sum_{j=1}^{k}\mathbb{E}\left[\tilde{b}_{kj}(X_t)^2 m(X_t, X_{t+1})^2(\tilde{b}^k(X_{t+1})'u_l)^2\right] \\
&\leq \frac{C\xi_k^2}{n}\sum_{l=1}^{k}\mathbb{E}\left[m(X_t, X_{t+1})^2(\tilde{b}^k(X_{t+1})'u_l)^2\right]
\end{aligned}$$

where the constant C depends only on the rho-mixing coefficients. By Hölder's inequality:

$$\begin{aligned}
\mathbb{E}[m(X_t, X_{t+1})^2(\tilde{b}^k(X_{t+1})'u_l)^2] &\leq \mathbb{E}[|m(X_0, X_1)|^r]^{2/r} \times \mathbb{E}[(\tilde{b}^k(X_0)'u_l)^{\frac{2r}{r-2}}]^{\frac{r-2}{r}} \\
&\leq \mathbb{E}[|m(X_0, X_1)|^r]^{2/r} \times \xi_k^{4/r} \times \mathbb{E}[(\tilde{b}^k(X_0)'u_l)^2]^{\frac{r-2}{r}} \lesssim \xi_k^{4/r}
\end{aligned}$$

since $\mathbb{E}[|m(X_0, X_1)|^r] < \infty$ and $\|u_l\| = 1$. Substituting into the above, we obtain

$$\mathbb{E}[\|\widehat{\mathbf{M}}^o - \mathbf{M}^o\|^2] \lesssim \xi_k^{2+4/r}k/n$$

which, by Markov's inequality, yields $\|\widehat{\mathbf{M}}^o - \mathbf{M}^o\| = O_p(\xi_k^{1+2/r}\sqrt{k/n})$. Similar arguments give $\|(\widehat{\mathbf{M}}^o - \mathbf{M}^o)\tilde{c}_k\| = O_p(\xi_k^{1+2/r}/\sqrt{n})$ and $\|\tilde{c}_k^{*\prime}(\widehat{\mathbf{M}}^o - \mathbf{M}^o)\| = O_p(\xi_k^{1+2/r}/\sqrt{n})$.

Result (1) now follows from Lemma D.3(b), noting that

$$\|(\widehat{\mathbf{G}}^o)^{-1}\widehat{\mathbf{M}}^o - \mathbf{M}^o\| = O_p(\xi_k^{1+2/r}\sqrt{k/n})$$

which is $o_p(1)$ under the condition $\xi_k^{1+2/r}\sqrt{k/n} = o(1)$.

For result (2), note that whenever $\|\widehat{\mathbf{G}}^o - \mathbf{I}\| \leq \frac{1}{2}$, we have $\|(\widehat{\mathbf{G}}^o)^{-1}\| \leq 2$ and hence:

$$\begin{aligned}
\|((\widehat{\mathbf{G}}^o)^{-1}\widehat{\mathbf{M}}^o - \mathbf{M}^o)\tilde{c}_k\| &\leq \|(\widehat{\mathbf{G}}^o)^{-1}(\widehat{\mathbf{M}}^o - \mathbf{M}^o)\tilde{c}_k\| + \|((\widehat{\mathbf{G}}^o)^{-1} - \mathbf{I})\mathbf{M}^o\tilde{c}_k\| \\
&\leq 2\|(\widehat{\mathbf{M}}^o - \mathbf{M}^o)\tilde{c}_k\| + 2\rho_k\|(\widehat{\mathbf{G}}^o - \mathbf{I})\tilde{c}_k\|\,.
\end{aligned}$$

The result for $\tilde{c}_k$ follows form the bounds $\|(\widehat{\mathbf{G}}^o - \mathbf{I})\tilde{c}_k\| = O_p(\xi_k/\sqrt{n})$ and $\|(\widehat{\mathbf{M}}^o - \mathbf{M}^o)\tilde{c}_k\| = O_p(\xi_k^{1+2/r}/\sqrt{n})$. The result for $\tilde{c}_k^*$ follows similarly.

Result (3) is immediate from the fact that $\|\widehat{\mathbf{G}}^o - \mathbf{I}\| = O_p(\xi_k\sqrt{k/n})$ and $\|\widehat{\mathbf{M}}^o - \mathbf{M}^o\| = O_p(\xi_k^{1+2/r}\sqrt{k/n})$. ■

Proof of Lemma C.3. The proof will follow by the same arguments as the proof of results (1)–(3) in Lemma C.1, provided we show that $\|\widehat{\mathbf{M}}^o - \mathbf{M}^o\| = O_p(\xi_k^{1+2/r}(\log n)/\sqrt{n})$ also holds in this case. First write:

$$\begin{aligned}
\widehat{\mathbf{M}}^o - \mathbf{M}^o &= \left(\frac{1}{n}\sum_{t=0}^{n-1}\tilde{b}^k(X_t)\Big(m(X_t,X_{t+1};\hat{\alpha}) - m(X_t,X_{t+1};\alpha_0)\Big)\tilde{b}^k(X_{t+1})\right) \\
&\quad + \left(\frac{1}{n}\sum_{t=0}^{n-1}\tilde{b}^k(X_t)m(X_t,X_{t+1};\alpha_0)\tilde{b}^k(X_{t+1}) - \mathbf{M}^o\right) =: \widehat{\Delta}_{1,k} + \widehat{\Delta}_{2,k}
\end{aligned}$$

where $\|\widehat{\Delta}_{2,k}\| = O_p(\xi_k^{1+2/r}(\log n)/\sqrt{n})$ by the proof of Lemma C.1. For $\widehat{\Delta}_{1,k}$, condition (a) implies that $\hat{\alpha} \in N$ wpa1. Whenever $\hat{\alpha} \in N$ we may take a mean value expansion (valid by condition (b)) to obtain:

$$\|\widehat{\Delta}_{1,k}\| = \left\|\frac{1}{n}\sum_{t=0}^{n-1}\tilde{b}^k(X_t)\tilde{b}^k(X_{t+1})'\left(\frac{\partial m(X_t,X_{t+1};\tilde{\alpha})}{\partial\alpha'}(\hat{\alpha}-\alpha_0)\right)\right\| \quad \text{wpa1}$$

for $\tilde{\alpha}$ in the segment between $\hat{\alpha}$ and α_0. Therefore, wpa1 we have:

$$
\begin{aligned}
\|\widehat{\Delta}_{1,k}\| &= \sup_{u,v \in S^{k-1}} \left| \frac{1}{n} \sum_{t=0}^{n-1} (u'\tilde{b}^k(X_t))(v'\tilde{b}^k(X_{t+1})) \left(\frac{\partial m(X_t, X_{t+1}; \tilde{\alpha})}{\partial \alpha'} (\hat{\alpha} - \alpha_0) \right) \right| \\
&\le \xi_k \times \left(\sup_{u \in S^{k-1}} \frac{1}{n} \sum_{t=0}^{n-1} |u'\tilde{b}^k(X_t)| \times \bar{m}(X_t, X_{t+1}) \right) \times \|\hat{\alpha} - \alpha_0\| \\
&\le \xi_k \times \left(\sup_{u \in S^{k-1}} u'\widehat{\mathbf{G}}^o u \right)^{1/2} \times \left(\frac{1}{n} \sum_{t=0}^{n-1} \bar{m}(X_t, X_{t+1})^2 \right)^{1/2} \times \|\hat{\alpha} - \alpha_0\|
\end{aligned}
$$

where the first line is because $\|\mathbf{A}\| = \sup_{u,v \in S^{k-1}} |u'\mathbf{A}v|$ and the second and third lines are by condition (b) and the Hölder and Cauchy-Schwarz inequalities. Finally, notice that $\sup_{u \in S^{k-1}} u'\widehat{\mathbf{G}}^o u = \|\widehat{\mathbf{G}}^o\| = 1 + o_p(1)$ by the proof of Lemma C.1, and $\frac{1}{n}\sum_{t=0}^{n-1} \bar{m}(X_t, X_{t+1})^2 = O_p(1)$ by the ergodic theorem and condition (b). Therefore $\|\widehat{\Delta}_{1,k}\| = O_p(\xi_k/\sqrt{n})$ and so $\|\widehat{\mathbf{M}}^o - \mathbf{M}^o\| = O_p(\xi_k^{1+2/r}(\log n)/\sqrt{n})$, as required. ■

Proof of Lemma C.4. The proof will follow by the same arguments as the proof of results (1)–(3) in Lemma C.1, provided we show that

$$
\|\widehat{\mathbf{M}}^o - \mathbf{M}^o\| = O_p\left(\frac{\xi_k^{1+2/r}(\log n)}{\sqrt{n}} + \frac{\xi_k^{2 - \frac{2s-v}{2sv}} \sqrt{k \log k}}{\sqrt{n}} \right).
$$

As in the proof of Lemma C.3, it suffices to bound:

$$
\widehat{\Delta}_{1,k} := \frac{1}{n} \sum_{t=0}^{n-1} \tilde{b}^k(X_t) \Big(m(X_t, X_{t+1}; \hat{\alpha}) - m(X_t, X_{t+1}; \alpha_0) \Big) \tilde{b}^k(X_{t+1}) .
$$

Let $h_\alpha(x_0, x_1) = m(x_0, x_1; \alpha) - m(x_0, x_1; \alpha_0)$ and let:

$$
\begin{aligned}
h_\alpha^{trunc}(x_0, x_1) &= h_\alpha(x_0, x_1) 1\!\!1\{\|\tilde{b}^k(x_0)\| \|\tilde{b}^k(x_1)\| E(x_0, x_1) \le T_n\} \\
h_\alpha^{tail}(x_0, x_1) &= h_\alpha(x_0, x_1) 1\!\!1\{\|\tilde{b}^k(x_0)\| \|\tilde{b}^k(x_1)\| E(x_0, x_1) > T_n\}
\end{aligned}
$$

where $\{T_n : n \geq 1\}$ be a sequence of positive constants to be defined below. Then:

$$\begin{aligned}
\|\widehat{\Delta}_{1,k}\| &\leq \sup_{\alpha\in\mathcal{A}} \left\| \frac{1}{n}\sum_{t=0}^{n-1} \tilde{b}^k(X_t) h_\alpha^{trunc}(X_t, X_{t+1}) \tilde{b}^k(X_{t+1}) - \mathbb{E}[\tilde{b}^k(X_t) h_\alpha^{trunc}(X_t, X_{t+1}) \tilde{b}^k(X_{t+1})] \right\| \\
&\quad + \sup_{\alpha\in\mathcal{A}} \left\| \frac{1}{n}\sum_{t=0}^{n-1} \tilde{b}^k(X_t) h_\alpha^{tail}(X_t, X_{t+1}) \tilde{b}^k(X_{t+1}) \right\| \\
&\quad + \sup_{\alpha\in\mathcal{A}} \left\| \mathbb{E}[\tilde{b}^k(X_t) h_\alpha^{tail}(X_t, X_{t+1}) \tilde{b}^k(X_{t+1})] \right\| + \left\| \mathbb{E}[\tilde{b}^k(X_t) h_{\hat{\alpha}}(X_t, X_{t+1}) \tilde{b}^k(X_{t+1})] \right\| \\
&=: \widehat{\Delta}_{1,k,1} + \widehat{\Delta}_{1,k,2} + \widehat{\Delta}_{1,k,3} + \widehat{\Delta}_{1,k,4} \,.
\end{aligned}$$

Let $\mathcal{H}_{n,k} = \{(c_0'\tilde{b}^k(x_0))(c_1'\tilde{b}^k(x_1))h_\alpha^{trunc}(x_0,x_1) : c_0, c_1 \in S^{k-1}, \alpha \in \mathcal{A}\}$ where S^{k-1} is the unit sphere in $\mathbb{R}^k$. Then:

$$\widehat{\Delta}_{1,k,1} \leq n^{-1/2} \times \sup_{h\in\mathcal{H}_{n,k}} |\mathcal{Z}_n(h)|$$

by definition of the operator norm, where $\mathcal{Z}_n$ is the centered empirical process on $\mathcal{H}_{n,k}$. By Theorem 2 of Doukhan et al. (1995):

$$\mathbb{E}[\sup_{h\in\mathcal{H}_{n,k}} |\mathcal{Z}_n(h)|] = O\left(\varphi(\sigma_{n,k}) + \frac{T_n q \varphi^2(\sigma_{n,k})}{\sigma_{n,k}^2 \sqrt{n}} + \sqrt{n} T_n \beta_q \right) \tag{OA.3}$$

where $q \in \{1, 2, \ldots\}$, $\sigma_{n,k} \geq \sup_{h\in\mathcal{H}_{n,k}} \|h\|_{2,\beta}$ for the norm $\|\cdot\|_{2,\beta}$ defined on p. 400 of Doukhan et al. (1995), and $\varphi(\sigma)$ is the bracketing entropy integral:

$$\varphi(\sigma) = \int_0^\sigma \sqrt{\log N_{[\,]}(u, \mathcal{H}_{n,k}, \|\cdot\|_{2,\beta})}\, \mathrm{d}u \,.$$

Exponential β-mixing and Lemma 2 of Doukhan et al. (1995) (with $\phi(x) = x^v$) imply:

$$\|\cdot\|_{2,\beta} \leq C\|\cdot\|_{2v} \quad \text{on } L^{2v} \tag{OA.4}$$

for any $v > 1$, where the constant $C < \infty$ depends only on v and the β-mixing coefficients. Taking $1 < v < 2s$, by Hölder's inequality and condition (a) we have:

$$\sup_{h\in\mathcal{H}_{n,k}} \|h\|_{2,\beta} \leq C \sup_{h\in\mathcal{H}_{n,k}} \|h\|_{2v} \leq C\xi_k^{2-\frac{2s-v}{2sv}} \|E\|_{4s} \,.$$

We therefore take $\sigma_{n,k} = C\xi_k^{2-\frac{2s-v}{2sv}}\|E\|_{4s}$.

To bound the bracketing entropy, define $\mathcal{H}_{n,k}^* = \{b_0(x_0)b_1(x_1)h(x_0,x_1) : b_0, b_1 \in \mathcal{B}_k^*, h \in \mathcal{H}_n^*\}$ where $\mathcal{B}_k^* = \{(c'\tilde{b}^k)/\xi_k : c \in S^{k-1}\}$ and $\mathcal{H}_n^* = \{h_\alpha^{trunc}/E : \alpha \in \mathcal{A}\}$. For $\mathcal{B}_k^*$, note that

$|c_0'\tilde{b}^k(x)/\xi_k - c_1'\tilde{b}^k(x)/\xi_k| \leq (\xi_k^{-1}\|\tilde{b}^k(x)\|) \times \|c_0 - c_1\|$ where $\|(\|\tilde{b}^k(x)\|/\xi_k)\|_p \leq (k/\xi_k^2)^{1/p}$ for any $p > 2$. By Theorem 2.7.11 of van der Vaart and Wellner (1996) and Lemma 2.5 of van de Geer (2000):

$$N_{[\,]}(u, \mathcal{B}_k^*, \|\cdot\|_p) \leq N\left(\frac{u}{2(k/\xi_k^2)^{1/p}}, S^{k-1}, \|\cdot\|\right) \leq \left(\frac{8(k/\xi_k^2)^{1/p}}{u} + 1\right)^k.$$

It follows by Lemma 9.25(ii) in Kosorok (2008) that:

$$N_{[\,]}(3u, \mathcal{H}_{n,k}^*, \|\cdot\|_p) \leq \left(\frac{8(k/\xi_k^2)^{1/p}}{u} + 1\right)^{2k} N_{[\,]}(u, \mathcal{H}_n^*, \|\cdot\|_p)\,. \tag{OA.5}$$

Let $[f_l, f_u]$ be a ε-bracket for $\mathcal{H}_{n,k}^*$ under the $L^{\frac{4sv}{2s-v}}$ norm. Then $[\xi_k^2 E f_l, \xi_k^2 E f_u]$ is a $\xi_k^2\|E\|_{4s}\varepsilon$-bracket for $\mathcal{H}_{n,k}$ under the L^{2v} norm, because $\|\xi_k^2 E(f_u - f_l)\|_{2v} \leq \xi_k^2\|E\|_{4s}\|f_u - f_l\|_{\frac{4sv}{2s-v}}$. Taking $p = \frac{4sv}{2s-v}$ in display (OA.5) and using the fact that truncation of $\mathcal{M}^*$ doesn't increase its bracketing entropy, we obtain:

$$\begin{aligned}
N_{[\,]}(u, \mathcal{H}_{n,k}, \|\cdot\|_{2v}) &\leq N_{[\,]}\Big(\frac{u}{\xi_k^2\|E\|_{4s}}, \mathcal{H}_{n,k}^*, \|\cdot\|_{\frac{4sv}{2s-v}}\Big) \\
&\leq \Big(\frac{24\|E\|_{4s}\xi_k^{2-\frac{2s-v}{2sv}}k^{\frac{2s-v}{4sv}}}{u} + 1\Big)^{2k} N_{[\,]}\Big(\frac{u}{3\xi_k^2\|E\|_{4s}}, \mathcal{M}^*, \|\cdot\|_{\frac{4vs}{2s-v}}\Big)\,.
\end{aligned} \tag{OA.6}$$

Now, by displays (OA.4) and (OA.6) and condition (b):

$$\begin{aligned}
\varphi(\sigma) &= \int_0^\sigma \sqrt{\log N_{[\,]}(u, \mathcal{H}_{n,k}, \|\cdot\|_{2,\beta})}\,\mathrm{d}u \\
&\leq \int_0^\sigma \sqrt{\log N_{[\,]}(u/C, \mathcal{H}_{n,k}, \|\cdot\|_{2v})}\,\mathrm{d}u \\
&\lesssim k^{1/2}\int_0^\sigma \sqrt{\log\left(1 + 24C\|E\|_{4s}\xi_k^{2-\frac{2s-v}{2sv}}k^{\frac{2s-v}{4sv}}/u\right)}\,\mathrm{d}u + (\xi_k^2\|E\|_{4s})^\zeta\frac{\sigma^{1-\zeta}}{1-\zeta} \\
&\lesssim \|E\|_{4s}\xi_k^{2-\frac{2s-v}{2sv}}k^{\frac{1}{2}+\frac{2s-v}{4sv}}\int_0^{\sigma/(24C\|E\|_{4s}\xi_k^{2-\frac{2s-v}{2sv}}k^{\frac{2s-v}{4sv}})}\sqrt{\log(1+1/u)}\,\mathrm{d}u + (\xi_k^2\|E\|_{4s})^\zeta\frac{\sigma^{1-\zeta}}{1-\zeta}\,.
\end{aligned}$$

Since $\sigma_{n,k} = C\xi_k^{2-\frac{2s-v}{2sv}}\|E\|_{4s}$, we obtain:

$$\begin{aligned}
\varphi(\sigma_{n,k}) &\lesssim \|E\|_{4s}\xi_k^{2-\frac{2s-v}{2sv}}k^{\frac{1}{2}+\frac{2s-v}{4sv}}\int_0^{\frac{1}{24}k^{-\frac{2s-v}{4sv}}}\sqrt{\log(1+1/u)}\,\mathrm{d}u + (\xi_k^2\|E\|_{4s})^\zeta(\xi_k^{2-\frac{2s-v}{2sv}}\|E\|_{4s})^{1-\zeta} \\
&\lesssim \|E\|_{4s}\xi_k^{2-\frac{2s-v}{2sv}}\sqrt{k\log k} + \|E\|_{4s}\xi_k^{2-\frac{2s-v}{2sv}+\zeta\frac{2s-v}{2sv}}
\end{aligned}$$

since $\int_0^\delta \sqrt{\log(1+1/u)}\,\mathrm{d}u = O(\delta\sqrt{-\log\delta})$ as $\delta \to 0^+$. If $\xi_k^{\zeta\frac{2s-v}{2sv}} \lesssim \sqrt{k\log k}$ then the first term dominates and we obtain $\varphi(\sigma_{n,k}) = O(\xi_k^{2-\frac{2s-v}{2sv}}\sqrt{k\log k})$. It follows by display (OA.3) that:

$$\widehat{\Delta}_{1,k,1} = O_p\left(\frac{\xi_k^{2-\frac{2s-v}{2sv}}\sqrt{k\log k}}{\sqrt{n}} + \frac{T_n q k\log k}{n} + T_n\beta_q\right).$$

By Markov's inequality we may deduce $\widehat{\Delta}_{1,k,2} = O_p(\xi_k^{8s}/T_n^{4s-1})$ and $\widehat{\Delta}_{1,k,3} = O(\xi_k^{8s}/T_n^{4s-1})$. Choosing T_n so that:

$$\frac{\xi_k^{8s}}{T_n^{4s-1}} \asymp \frac{\xi_k^{2-\frac{2s-v}{2sv}}\sqrt{k\log k}}{\sqrt{n}}$$

and $q = C_0\log n$ for sufficiently large C_0 ensures, in view of the condition $\log n = O(\xi_k^{1/3})$, that $\widehat{\Delta}_{1,k,1}$, $\widehat{\Delta}_{1,k,2}$, and $\widehat{\Delta}_{1,k,3}$ are all $O_p(\xi^{2-\frac{2s-v}{2sv}}\sqrt{(k\log k)/n})$. For the remaining term, by condition (c) we have:

$$\widehat{\Delta}_{1,k,4} = \|\Pi_k(\mathbb{M}^{(\hat\alpha)} - \mathbb{M})|_{B_k}\| \le \ell^*(\hat\alpha) = \frac{1}{\sqrt{n}}\sqrt{n}\dot\ell^*_{\alpha_0}[\hat\alpha - \alpha_0] + O(\|\hat\alpha-\alpha_0\|^2_{\mathcal{A}}) = O_p(n^{-1/2})$$

which is of smaller order. ∎

G.2 Proofs for Appendix C.2

Proof of Lemma C.5. Lemma 2.2 of Chen and Christensen (2015) gives the bound $\|\widehat{\mathbf{G}}^o - \mathbf{I}\| = O_p(\xi_k(\log n)/\sqrt{n})$. Let $\{T_n : n \ge 1\}$ be a sequence of positive constants to be defined and let:

$$\begin{aligned} G_{t+1}^{trunc} &= G_{t+1}^{1-\gamma}1\!\!1\{\|\tilde b^k(x_t)\|\|\tilde b^k(x_{t+1})\|^\beta |G_{t+1}^{1-\gamma}| \le T_n\} \\ G_{t+1}^{tail} &= G_{t+1}^{1-\gamma}1\!\!1\{\|\tilde b^k(x_t)\|\|\tilde b^k(x_{t+1})\|^\beta |G_{t+1}^{1-\gamma}| > T_n\}\,. \end{aligned}$$

We then have:

$$\begin{aligned} \sup_{v:\|v\|\le c}\|\widehat{\mathbf{T}}^o v - \mathbf{T}^o v\| \le{}& \sup_{v:\|v\|\le c}\left\|\frac{1}{n}\sum_{t=0}^{n-1}\tilde b^k(X_t)G_{t+1}^{trunc}|\tilde b^k(X_{t+1})'v|^\beta - \mathbb{E}[\tilde b^k(X_t)G_{t+1}^{trunc}|\tilde b^k(X_{t+1})'v|^\beta]\right\| \\ &+ \sup_{v:\|v\|\le c}\left\|\frac{1}{n}\sum_{t=0}^{n-1}\tilde b^k(X_t)G_{t+1}^{tail}|\tilde b^k(X_{t+1})'v|^\beta\right\| \\ &+ \sup_{v:\|v\|\le c}\left\|\mathbb{E}[\tilde b^k(X_t)G_{t+1}^{tail}|\tilde b^k(X_{t+1})'v|^\beta]\right\| \quad =: \quad \widehat{T}_1 + \widehat{T}_2 + \widehat{T}_3\,. \end{aligned}$$

Let $\mathcal{H}_{n,k} = \{w'\tilde{b}^k(x_0)G_1^{trunc}|\tilde{b}^k(x_1)'v|^\beta : v \in \mathbb{R}^k, \|v\| \leq c, w \in S^{k-1}\}$. Then:

$$\widehat{T}_1 \leq n^{-1/2} \times \sup_{h \in \mathcal{H}_{n,k}} |\mathcal{Z}_n(h)|$$

where $\mathcal{Z}_n$ is the centered empirical process on $\mathcal{H}_{n,k}$. Each $h \in \mathcal{H}_{n,k}$ is uniformly bounded by $c^\beta T_n$. Therefore, by Condition (a) and Theorem 2 of Doukhan et al. (1995):

$$\mathbb{E}[\sup_{h \in \mathcal{H}_{n,k}} |\mathcal{Z}_n(h)|] = O\left(\varphi(\sigma_{n,k}) + \frac{c^\beta T_n q \varphi^2(\sigma_{n,k})}{\sigma_{n,k}^2 \sqrt{n}} + \sqrt{n} c^\beta T_n \beta_q\right) \tag{OA.7}$$

where $q \in \{1, 2, \ldots\}$, $\sigma_{n,k} \geq \sup_{h \in \mathcal{H}_{n,k}} \|h\|_{2,\beta}$ for the norm $\|\cdot\|_{2,\beta}$ defined on p. 400 of Doukhan et al. (1995), and $\varphi(\sigma)$ is the bracketing entropy integral:

$$\varphi(\sigma) = \int_0^\sigma \sqrt{\log N_{[\,]}(u, \mathcal{H}_{n,k}, \|\cdot\|_{2,\beta})} \, \mathrm{d}u \, .$$

To calculate $\sigma_{n,k}$, by (OA.4) and Hölder's inequality we have:

$$\sup_{h \in \mathcal{H}_{n,k}} \|h\|_{2,\beta} \leq C \sup_{h \in \mathcal{H}_{n,k}} \|h\|_{2s} \leq C c^\beta \|G^{1-\gamma}\|_{2s} \xi_k^{1+\beta}$$

where $\|G^{1-\gamma}\|_{2s}$ is finite by condition (b). Set $\sigma_{n,k} = C c^\beta \|G^{1-\gamma}\|_{2s} \xi_k^{1+\beta}$.

To bound the bracketing entropy, first fix $q > 2$ and let $w_1, \ldots, w_{N_1}$ be a ε-cover for S^{k-1} and $v_1, \ldots, v_{N_2}$ be a $\varepsilon^{1/\beta}$-cover for $\{v \in \mathbb{R}^k : \|v\| \leq c\}$. For any $w \in S^{k-1}$ and $v \in \{v \in \mathbb{R}^k : \|v\| \leq c\}$ there exist $v_i \in \{v_1, \ldots, v_{N_1}\}$ and $w_j \in \{w_1, \ldots, w_{N_2}\}$ such that:

$$\begin{aligned}
&w_j'\tilde{b}^k(x_0)G_1^{trunc}|\tilde{b}^k(x_1)'v_i|^\beta - \varepsilon\Big((1 + c^\beta)\|\tilde{b}^k(x_0)\|\|\tilde{b}^k(x_1)\|^\beta|G_1^{trunc}|\Big) \\
&\leq w'\tilde{b}^k(x_0)G_1^{trunc}|\tilde{b}^k(x_1)'v|^\beta \\
&\leq w_j'\tilde{b}^k(x_0)G_1^{trunc}|\tilde{b}^k(x_1)'v_i|^\beta + \varepsilon\Big((1 + c^\beta)\|\tilde{b}^k(x_0)\|\|\tilde{b}^k(x_1)\|^\beta|G_1^{trunc}|\Big)
\end{aligned}$$

where:

$$\Big\|2\varepsilon\Big((1 + c^\beta)\|\tilde{b}^k(x_0)\|\|\tilde{b}^k(x_1)\|^\beta|G_1^{trunc}|\Big)\Big\|_{2s} \leq 2\varepsilon(1 + c^\beta)\|G^{1-\gamma}\|_{2s}\xi_k^{1+\beta} = \varepsilon C_0 \xi_k^{1+\beta} \, .$$

where $C_0 = 2(1 + c^\beta)\|G^{1-\gamma}\|_{2s}$. Therefore, given a ε-cover of S^{k-1} and a $\varepsilon^{1/\beta}$-cover for $\{v \in \mathbb{R}^k : \|v\| \leq c\}$ we can construct $\varepsilon C_0 \xi_k^{1+\beta}$-brackets for $\mathcal{H}_{n,k}$ under the L^{2s} norm, and so by Lemma 2.5 of van de Geer (2000):

$$N_{[\,]}\big(u, \mathcal{H}_{n,k}, \|\cdot\|_{2s}\big) \leq \Big(\frac{4C_0\xi_k^{1+\beta}}{u} + 1\Big)^k \Big(\frac{4c(C_0\xi_k^{1+\beta})^{1/\beta}}{u^{1/\beta}} + 1\Big)^k .$$

By (OA.4) and the above display:

$$\begin{aligned}
\varphi(\sigma) &= \int_0^\sigma \sqrt{\log N_{[\,]}(u, \mathcal{H}_{n,k}, \|\cdot\|_{2,\beta})}\,\mathrm{d}u \\
&\leq \int_0^\sigma \sqrt{\log N_{[\,]}(u/C, \mathcal{H}_{n,k}, \|\cdot\|_{2s})}\,\mathrm{d}u \\
&\leq k^{1/2}\left(\int_0^\sigma \sqrt{\log\big(1 + 4CC_0\xi_k^{1+\beta}/u\big)}\,\mathrm{d}u + \int_0^\sigma \sqrt{\log\big(1 + 4c(CC_0\xi_k^{1+\beta}/u)^{1/\beta}\big)}\,\mathrm{d}u\right) .
\end{aligned}$$

Since $\sigma_{n,k} = Cc^\beta \|G^{1-\gamma}\|_{2s}\xi_k^{1+\beta}$, by a change of variables we obtain $\varphi(\sigma_{n,k}) = O(\xi_k^{1+\beta}\sqrt{k})$. Substituting into (OA.7):

$$\widehat{T}_1 = O_p\bigg(\frac{\xi_k^{1+\beta}\sqrt{k}}{\sqrt{n}} + \frac{T_n q k}{n} + T_n\beta_q\bigg) .$$

By Markov's inequality we may deduce $\widehat{T}_2 = O_p(\xi_k^{(1+\beta)2s}/T_n^{2s-1})$ and $\widehat{T}_3 = O(\xi_k^{(1+\beta)2s}/T_n^{2s-1})$. Choosing T_n so that $\xi_k^{(1+\beta)2s}/T_n^{2s-1} \asymp \xi_k^{1+\beta}\sqrt{k/n}$ and $q = C_0 \log n$ for large enough C_0 ensures, in view of the condition $(\log n)^{(2s-1)/(s-1)}k/n = o(1)$, that $\widehat{T}_1$, $\widehat{T}_2$, and $\widehat{T}_3$ are all $O_p(\xi_k^{1+\beta}\sqrt{k/n})$.

The expression for $\nu_{n,k}$ now follows from display (S.18) and the rates for $\widehat{\mathbf{G}}^o$ and $\widehat{\mathbf{T}}^o$. ∎

G.3 Proofs for Appendix F

Proof of Proposition F.1. We first show that any positive eigenfunction of $\mathbb{M}$ must have eigenvalue ρ. Suppose that there is some positive $\psi \in L^2$ and scalar λ such that $\mathbb{M}\psi(x) = \lambda\psi(x)$. Then we obtain:

$$\lambda\langle \phi^*, \psi\rangle = \langle \phi^*, \mathbb{M}\psi\rangle = \langle \mathbb{M}^*\phi^*, \psi\rangle = \rho\langle \phi^*, \psi\rangle$$

with $\langle \phi^*, \psi\rangle > 0$ because ϕ^* and ψ are positive, hence $\lambda = \rho$. A similar argument shows that any positive eigenfunction of $\mathbb{M}^*$ must correspond to the eigenvalue ρ.

It remains to show that ϕ and ϕ^* are the unique eigenfunctions (in L^2) of $\mathbb{M}$ and $\mathbb{M}^*$ with eigenvalue ρ. We do this in the following three steps. Let $F = \{\psi \in L^2 : \mathbb{M}\psi = \rho\psi\}$. We first show that if $\psi \in F$ then the function $|\psi|$ given by $|\psi|(x) = |\psi(x)|$ also is in F. In the second step we show that $\psi \in F$ implies $\psi = |\psi|$ or $\psi = -|\psi|$. Finally, in the third step we show that $F = \{s\phi : s \in \mathbb{R}\}$.

For the first step, first observe that $F \neq \{0\}$ because $\phi \in F$ by Assumption F.1(b). Then by Assumption F.1(c), for any $\psi \in F$ we have $\mathbb{M}|\psi| \geq |\mathbb{M}\psi| = \rho|\psi|$ and so $\mathbb{M}|\psi| - \rho|\psi| \geq 0$ (almost everywhere). On the other hand,

$$\langle \phi^*, \mathbb{M}|\psi| - \rho|\psi| \rangle = \langle \mathbb{M}^*\phi^*, |\psi| \rangle - \rho\langle \phi^*, |\psi| \rangle = 0$$

which implies that $\mathbb{M}|\psi| = \rho|\psi|$ and hence $|\psi| \in F$.

For the second step, take any $\psi \in F$ that is not identically zero. Suppose that $\psi = |\psi|$ on a set of positive Q measure (otherwise we can take $-\psi$ in place of ψ). We will prove by contradiction that this implies $|\psi| = \psi$. Assume not, i.e. $|\psi| \neq \psi$ on a set of positive Q measure. Then $|\psi| - \psi \geq 0$ (almost everywhere) and $|\psi| - \psi \neq 0$. But by step 1 we also have that $\mathbb{M}(|\psi| - \psi) = \rho(|\psi| - \psi)$. Then for any $\lambda > r(\mathbb{M})$ we have

$$\frac{(\rho/\lambda)}{1 - (\rho/\lambda)}(|\xi| - \xi) = \sum_{n \geq 1} \left(\frac{\rho}{\lambda}\right)^n (|\xi| - \xi) = \sum_{n \geq 1} \lambda^{-n}\mathbb{M}^n(|\xi| - \xi) > 0$$

(almost everywhere) by Assumption F.1(c). Therefore, $|\psi| > \psi$ (almost everywhere). This contradicts the fact that $\psi = |\psi|$ on a set of positive Q measure. A similar proof shows that if $-\psi = |\psi|$ holds on a set of positive Q measure then $-\psi = |\psi|$.

For the third step we use an argument based on the Archimedean axiom (see, e.g., p. 66 of Schaefer (1974)). Take any positive $\psi \in F$ and define the sets $S_+ = \{s \in \mathbb{R} : \psi \geq s\phi\}$ and $S_- = \{s \in \mathbb{R} : \zeta \leq s\phi\}$ (where the inequalities are understood to hold almost everywhere). It is easy to see that S_+ and S_- are convex and closed. We also have $(-\infty, 0] \subseteq S_+$ so S_+ is nonempty. Suppose S_- is empty. Then $\psi > s\phi$ on a set of positive measure for all $s \in (0, \infty)$. By step 2 we therefore have $\psi > s\phi$ (almost everywhere). But then because L^2 is a lattice we must have $\|\psi\| \geq s\|\phi\|$ for all $s \in (0, \infty)$ which is impossible because $\psi \in L^2$. Therefore S_- is nonempty. Finally, we show that $\mathbb{R} = S_+ \cup S_-$. Take any $s \in \mathbb{R}$. Clearly $\psi - s\phi \in F$. By Claim 2 we know that either: $\psi - s\phi \geq 0$ (almost everywhere) which implies $s \in S_+$ or $\psi - s\phi \leq 0$ (almost everywhere) which implies $s \in S_-$. Therefore $\mathbb{R} = S_+ \cup S_-$. The Archimedean axiom implies that the intersection $S_+ \cap S_-$ must be nonempty. Therefore $S_+ \cap S_- = \{s^*\}$ (the intersection must be a singleton else $\psi = s\phi$ and $\psi = s'\phi$ with $s \neq s'$) and so $\psi = s^*\phi$ (almost everywhere). This completes the proof of the third step.

A similar argument implies that ϕ^* is the unique positive eigenfunction of $\mathbb{M}^*$. ■

Proof of Proposition F.2. Assumption 2.1(a) implies that $r(\mathbb{M}) > 0$ (see Proposition IV.9.8 and Theorem V.6.5 of Schaefer (1974)). The result now follows by Theorems 6 and 7 of Sasser (1964) with $\rho = r(\mathbb{M})$. That ρ is isolated follows from the discussion on p. 1030 of Sasser (1964). ■

Proof of Proposition F.3. Consider the operator $\overline{\mathbb{M}} = \rho^{-1}\mathbb{M}$ with $\rho = r(\mathbb{M})$. Proposition F.2 implies that $\{1\} = \{\lambda \in \sigma(\overline{\mathbb{M}}) : |\lambda| = 1\}$. Further, since $\mathbb{M}$ is power compact it has discrete spectrum (Dunford and Schwartz, 1958, Theorem 6, p. 579). We therefore have $\sup\{|\lambda| : \lambda \in \sigma(\overline{\mathbb{M}}), \lambda \neq 1\} < 1$ and hence $\overline{\mathbb{M}} = (\phi \otimes \phi^*) + \mathbb{V}$ where $r(\mathbb{V}) < 1$ and $\overline{\mathbb{M}}$, $(\phi \otimes \phi^*)$ and $\mathbb{V}$ commute (see, e.g., p. 331 of Schaefer (1974) or pp. 1034-1035 of Sasser (1964)). Since these operators commute, a simple inductive argument yields:

$$\mathbb{V}^\tau = (\overline{\mathbb{M}} - (\phi \otimes \phi^*))^\tau = \overline{\mathbb{M}}^\tau - (\phi \otimes \phi^*) = \rho^{-\tau}\mathbb{M}_\tau - (\phi \otimes \phi^*)$$

for each $\tau \in T$. By the Gelfand formula, there exists $\epsilon > 0$ such that:

$$\lim_{\tau \to \infty} \|\mathbb{V}^\tau\|^{1/\tau} = r(\mathbb{V}) \leq 1 - \epsilon \tag{OA.8}$$

Let $\{\tau_k : k \geq 1\} \subseteq T$ be the maximal subset of T for which $\|\mathbb{V}^{\tau_k}\| > 0$. If this subsequence is finite then the proof is complete. If this subsequence is infinite, then by expression (OA.8),

$$\limsup_{\tau_k \to \infty} \frac{\log \|\mathbb{V}^{\tau_k}\|}{\tau_k} < 0\,.$$

Therefore, there exists a finite positive constant c such that for all τ_k large enough, we have:

$$\log \|\mathbb{V}^{\tau_k}\| \leq -c\tau_k$$

and hence:

$$\|\rho^{-\tau_k}\mathbb{M}_{\tau_k} - (\phi \otimes \phi^*)\| \leq e^{-c\tau_k}$$

as required. ■

References

Borovička, J., L. P. Hansen, and J. A. Scheinkman (2016). Misspecified recovery. *Journal of Finance 71*(6), 2493–2544.

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.

Christensen, T. M. (2015). Nonparametric identification of positive eigenfunctions. *Econometric Theory 31*(6), 1310–1330.

Doukhan, P., P. Massart, and E. Rio (1995). Invariance principles for absolutely regular empirical processes. *Annales de l'Institut Henri Poincaré (B) Probabilités et Statistiques 31*, 393–427.

Dunford, N. and J. T. Schwartz (1958). *Linear Operators, Part I: General Theory.* Interscience Publishers, New York.

Gobet, E., M. Hoffmann, and M. Reiß (2004). Nonparametric estimation of scalar diffusions based on low frequency data. *Annals of Statistics 32*, 2223–2253.

Hansen, L. P. and J. A. Scheinkman (2009). Long-term risk: An operator approach. *Econometrica 77*(1), 177–234.

Kosorok, M. R. (2008). *Introduction to Empirical Processes and Semiparametric Inference.* Springer.

Meyn, S. and R. L. Tweedie (2009). *Markov Chains and Stochastic Stability.* Cambridge University Press.

Sasser, D. W. (1964). Quasi-positive operators. *Pacific Journal of Mathematics 14*, 1029–1037.

Schaefer, H. H. (1974). *Banach Lattices and Positive Operators.* Springer-Verlag, Berlin.

van de Geer, S. (2000). *Empirical Processes in M-Estimation.* Cambridge University Press.

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