cemmap

centre for microdata methods and practice

Nonparametric stochastic discount factor decomposition

Timothy Christensen

The Institute for Fiscal Studies
Department of Economics, UCL

cemmap working paper CWP24/15

An ESRC Research Centre

Institute for Fiscal Studies, 7 Ridgmount Street, London WC1E 7AE **Tel:** 44 (0)20 7291 4800 **Email:** info@cemmap.ac.uk **www.cemmap.ac.uk**

Nonparametric Stochastic Discount Factor Decomposition*

Timothy M. Christensen†

First version: May 2, 2013. Revised version: December 10, 2014.

Abstract

We introduce econometric methods to perform estimation and inference on the permanent and transitory components of the stochastic discount factor (SDF) in dynamic Markov environments. The approach is nonparametric in that it does not impose parametric restrictions on the law of motion of the state process. We propose sieve estimators of the eigenvalue-eigenfunction pair which are used to decompose the SDF into its permanent and transitory components, as well as estimators of the long-run yield and the entropy of the permanent component of the SDF, allowing for a wide variety of empirically relevant setups. Consistency and convergence rates are established. The estimators of the eigenvalue, yield and entropy are shown to be asymptotically normal and semiparametrically efficient when the SDF is observable. We also introduce nonparametric estimators of the continuation value under Epstein-Zin preferences, thereby extending the scope of our estimators to an important class of recursive preferences. The estimators are simple to implement, perform favorably in simulations, and may be used to numerically compute the eigenfunction and its eigenvalue in fully specified models when analytical solutions are not available.

Keywords: Nonparametric estimation, Sieve estimation, Stochastic discount factor, Permanent-transitory decomposition, Long run, Entropy.

JEL codes: C13, C14, C58.

---

*This paper is based on the first two chapter of my doctoral dissertation at Yale. I am very grateful to my advisors Xiaohong Chen and Peter Phillips and committee member Donald Andrews for their support and advice. I would also like to thank the Co-Editor and three referees for many helpful comments. I have benefited from feedback from participants of seminars at Chicago Booth, Columbia, Cornell, Duke, Indiana, Monash, Montreal, Northwestern, NYU, Penn, Princeton, Sydney, UCL, UNSW, Wisconsin, participants of the 2012 SoFiE/FGV conference on Asset Pricing and Portfolio Allocation in the Long Run, the 2013 European and North American Summer Meetings of the Econometric Society, the 2014 Toulouse Financial Econometrics Conference, and discussants Caio Almeida and Jean-Pierre Florens. A longer version of this paper, entitled "Estimating the Long-Run Implications of Dynamic Asset Pricing Models", was previously circulated as my job market paper. While writing this paper I was generously supported by a Carl Arvid Anderson Fellowship from the Cowles Foundation.

†Department of Economics, New York University, 19 W. 4th Street, 6th floor, New York, NY 10012, USA. E-mail address: `timothy.christensen@nyu.edu`

1 Introduction

Dynamic asset pricing models link the prices assets with sources of risk, the payoff horizon, and the preferences of economic agents. A large and growing literature in macroeconomics and asset pricing has shown how to extract information about the long-run valuation implications of a model by analyzing the permanent and transitory components of the stochastic discount factor (SDF).[1] The permanent-transitory decomposition of the SDF provides a powerful and robust means for investigating the connection between macroeconomic fundamentals, asset returns, and agents' preferences. Despite this recent activity, econometric methods for performing estimation and inference on the permanent and transitory components of the SDF have not yet been well developed.

We introduce econometric methods for performing estimation and inference on features of the permanent and transitory components of the SDF in discrete-time Markov environments. The estimators are nonparametric in that we do not impose no parametric restrictions on the law of motion, or "dynamics", of the Markov state process. This approach is in the spirit of generalized method of moments (GMM) (Hansen, 1982). One very attractive feature of GMM is that it allows important structural parameters of a model to be estimated without fully specifying the data generating process. Similar to GMM, our estimators may be used to extract information about the long-run pricing implications of a model without fully specifying the dynamics of the state process. To date, researchers have typically imposed simple parametric models in order to obtain analytical formulas for terms related to the permanent and transitory components. In contrast, economic theory is vague as to how the dynamics of the state process should be modeled. This nonparametric approach avoids potential distortion of the long-run implications which may arise due to misspecification of simple parametric models for the dynamics. This approach also permits researchers to use data to investigate the long-run valuation implications of different preferences without having to specify the dynamics and SDF in a way that makes analytical solution feasible.

Hansen and Scheinkman (2009) show that the permanent and transitory components of the SDF may be extracted by studying a positive eigenfunction and eigenvalue of a pricing operator. Their long-run pricing approximation shows that the positive eigenfunction characterizes the state dependence of the price of long-horizon assets and its eigenvalue encodes the yield on long-term bonds. The eigenvalue is also related to the entropy of the permanent component of the SDF, which is a measure of the persistence and dispersion of the SDF. Alvarez and Jermann (2005) derive bounds for the entropy of the permanent component as a function of returns relative to long-term bonds and estimate the bounds from historical returns data. Dual to their approach, we show how to estimate the entropy of the permanent component of the SDF obtained under different, possibly counterfactual, preference specifications using historical data on the state. Our estimators may be used in conjunction with the Alvarez and Jermann (2005) bounds to establish whether different

[1] Prominent examples include Alvarez and Jermann (2005), Hansen and Scheinkman (2009), Hansen (2012), Bakshi and Chabi-Yo (2012), and Backus, Chernov, and Zin (2014).

preference specifications can generate a sufficiently large entropy of the permanent component of the SDF to rationalize historical return premia.

The central focus of this paper is nonparametric sieve estimation of the positive eigenfunction, its eigenvalue, the long-run yield, and the entropy of the permanent component. This approach is inspired by earlier work of Chen, Hansen, and Scheinkman (2000) on nonparametric estimation of diffusion processes. Sieve methods are useful in this context as they reduce an intractable, infinite-dimensional eigenfunction problem to a low-dimensional matrix eigenvector problem which is then easily estimated from time series data on the state.[2] The estimators are particularly easy to implement: there is no simulation, optimization, or numerical integration. The estimators may also be used to numerically compute the long-run implications of fully specified asset pricing models for which analytical solutions are unavailable.

The scope of this paper is confined to stationary, discrete-time environments so as to simplify the econometric analysis. We present identification conditions for the positive eigenfunction in stationary discrete-time environments to complement those Hansen and Scheinkman (2009) provide for possibly nonstationary, continuous-time environments. The identification conditions are weaker than other nonparametric identification conditions for positive eigenfunctions which have been derived recently using similar operator methods (see Chen, Chernozhukov, Lee, and Newey (2014) and references therein). Following earlier work by Hansen and Scheinkman (1995) on Markov processes, we study a "time-reversed" version of the pricing operator.[3] Existence of a positive eigenfunction of the time-reversed operator is one of the identification conditions. We also present a version of the long-run pricing approximation of Hansen and Scheinkman (2009) which is formulated in terms of the positive eigenfunctions of the forward- and reverse-time pricing operators.

We establish consistency and convergence rates of the estimators of the forward- and reverse-time positive eigenfunctions, the eigenvalue, the long-run yield, and the entropy of the permanent component of the SDF, allowing for a variety of modeling setups. When specialized to the case in which the SDF is observable, the estimators of the eigenvalue, long-run yield, and entropy of the permanent component are shown to be asymptotically normal and semiparametrically efficient.

Certain SDFs contain components that depend on forward-looking expectations and are therefore not directly observable when we model the dynamics nonparametrically. For example, the SDF obtained under Kreps and Porteus (1978), Epstein and Zin (1989), and Weil (1990) recursive preference specifications depends on the continuation value function of future consumption which is unobservable to the econometrician when the dynamics are modeled nonparametrically. To extend the ambit of our estimators to models with recursive preferences, we also introduce nonparametric

[2] I am grateful to a referee for suggesting an alternative kernel-based approach in which (i) the Markov transition density is estimated nonparametrically and plugged into the pricing operator, then (ii) the estimator is recast as a $n \times n$ matrix eigenvector problem using similar reasoning as in Darolles, Fan, Florens, and Renault (2011) (where n is the sample size). A comparison of the relative merits of the two approaches is beyond the scope of this paper.

[3] Note, however, that we do not require the state process to be time reversible.

sieve estimators of the continuation value of future consumption under Epstein-Zin preferences. The continuation-value estimators may be plugged in to the sieve eigenfunction/eigenvalue estimators to nonparametrically estimate the positive eigenfunction, eigenvalue, long-run yield and entropy of the permanent component under counterfactual preference parameters. We show, via simulations, that these quantities may be estimated to a high degree of accuracy under Epstein-Zin preferences despite the fact that a nonparametric estimate of the continuation value is first plugged in to the eigenfunction/eigenvalue estimators. The continuation value estimators may also be used as an alternative to discretization methods (e.g. Tauchen and Hussey (1991)) to numerically solve for the continuation value and SDF in fully specified models when a solution is not available analytically.

The estimators are applied to extract the permanent and transitory components consistent with historical data on consumption and corporate earnings under both constant relative risk aversion (CRRA) preferences and Epstein-Zin preferences with unit elasticity of intertemporal substitution. We find that the two preference specifications generate virtually indistinguishable permanent components. Coherently with the well-documented shortcomings of the C-CAPM, neither preference specification can explain the historically high return on equities relative to long-term bonds under reasonable values of the risk aversion and time preference parameters. The Epstein-Zin specification has some success at explaining the level of historical long-term yields, but cannot explain the volatility of returns on long-term bonds.

The theoretical contributions of this paper have broader application to nonparametric identification and estimation. First, other quantities of interest, such as a habit formation component in a semiparametric C-CAPM (Chen et al., 2014) and marginal utility in nonparametric Euler equations (Lewbel, Linton, and Srisuma, 2011; Escanciano and Hoderlein, 2012), may be written as positive eigenfunctions of nonselfadjoint operators. Our estimators provide a computationally simple means of estimating these other models in a representative agent setting. In contrast, implementation of the kernel-based estimator of Lewbel et al. (2011) requires nonparametrically estimating conditional densities, a numerical integration step, and then solving a high-dimensional eigenfunction problem. The estimators and large-sample theory presented herein may be extended to study sieve estimation of these other models in a heterogeneous agent setting using mirco-level data.

Second, the derivation of the large sample properties of the estimators is nonstandard as the eigenfunction-eigenvalue pair are defined implicitly by an unknown, nonselfadjoint operator. The literature on nonparametric eigenfunction estimation to date has focused almost exclusively on the selfadjoint case (see Chen et al. (2000) and Gobet, Hoffmann, and Reiß (2004) for sieve estimation and Darolles, Florens, and Renault (1998), Darolles, Florens, and Gourieroux (2004), and Carrasco, Florens, and Renault (2007) for a kernel approach).[4] A notable exception is the working paper Lew-

[4]Darolles et al. (1998), Darolles et al. (2004), and Carrasco et al. (2007) estimate the singular value decomposition (SVD) of a possibly nonselfadjoint conditional expectation operator. The SVD is obtained as the eigendecomposition of two composite operators formed as the product of the operator and its adjoint. The composite operators are selfadjoint even when the conditional expectation operator is nonselfadjoint.

bel et al. (2011) who establish asymptotic normality of kernel-based eigenfunction and eigenvalue estimators for nonparametric Euler equations. Our derivation of the large-sample theory is rather different from theirs because with sieves the dimension of the function space is expanding with the sample size.

The paper is organized as follows. Section 2 presents the setting and briefly reviews SDF decomposition. Section 3 presents identification and existence conditions and a long-run pricing approximation. Section 4 describes the estimators and derives their large-sample properties. Section 5 outlines nonparametric continuation value estimation under recursive preferences and presents the simulations and empirical application. Section 6 concludes. A Supplementary Appendix contains a discussion of the connection between our identification conditions and those in Hansen and Scheinkman (2009), verification of the identification conditions for common parametric models, further results on estimation, and all proofs.

2 SDF decomposition

This section reviews briefly the relationship between the positive eigenfunction, the long-term implications of asset pricing models as exposited by Hansen and Scheinkman (2009), and related work by Alvarez and Jermann (2005), Hansen (2012), and Backus et al. (2014).

Consider a discrete-time environment such that at each date $t \in \{0, 1, \ldots\}$ the random vector X_t of state variables summarizes all relevant information for assigning values to future state-contingent payoffs. We assume that the state process $\{X_t\}$ on $\mathcal{X} \subseteq \mathbb{R}^d$ is a strictly stationary and ergodic, time-homogeneous, first-order Markov process defined on a complete probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, \mathbb{P})$, where $\mathcal{F}_t = \sigma(X_t, X_{t-1}, \ldots)$ is the σ-algebra generated by the history $\{\ldots, X_{t-1}, X_t\}$. We follow Alvarez and Jermann (2005) and Backus et al. (2014) and assume a stationary environment whereas Hansen and Scheinkman (2009) allow for possibly non-stationary environments.

We further assume there exists a pricing kernel process $\{M_t\}$ such that M_t is adapted to $\mathcal{F}_t$ for each t, and for which $M_{t+n} = M_t M_n(\theta_t)$ where θ_t is the shift operator that moves the time subscript forward by t units (see Hansen and Scheinkman (2009)).[5] At each date t, the price assigned to a claim to Z_{t+n}, payable at the future date $t + n$, is given by

$$\mathbb{E}\left[\frac{M_{t+n}}{M_t} Z_{t+n} \middle| X_t \right] . \tag{1}$$

It is convenient to write

$$\frac{M_{t+1}}{M_t} = m(X_t, X_{t+1}) \tag{2}$$

for some time-homogeneous, non-negative function $m : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ which will be referred to as the

[5] We refer to M_t as the pricing kernel and M_{t+1}/M_t as the SDF following Alvarez and Jermann (2005).

SDF, and the random variable $m(X_t, X_{t+1})$ will be referred to as the date-t SDF.[6]

Hansen and Scheinkman (2009) show that, by restricting (1) to payoffs of the form $Z_{t+n} = \psi(X_{t+n})$ for $\psi : \mathcal{X} \to \mathbb{R}$, we may define a collection $\{\mathbb{M}_n : n \geq 1\}$ of linear operators of the form

$$\mathbb{M}_n \psi(x) = \mathbb{E}\left[\frac{M_{t+n}}{M_t} \psi(X_{t+n}) \middle| X_t = x \right]$$

on a space of payoff functions ψ. As a consequence of the multiplicative functional and Markov structure, the factorization $\mathbb{M}_n = \mathbb{M}^n$ holds for each $n \geq 1$, where the *pricing operator* $\mathbb{M}$ is

$$\mathbb{M}\psi(x) = \mathbb{E}[m(X_t, X_{t+1})\psi(X_{t+1})|X_t = x]$$

using the notation (2). Thus $\mathbb{M}_n \psi$ may be calculated by iteratively applying $\mathbb{M}$ to ψ for n times.

A function ϕ is an *eigenfunction* of $\{\mathbb{M}_n : n \geq 1\}$ with *eigenvalue* ρ if

$$\mathbb{M}_n \phi = \rho^n \phi \tag{3}$$

for each $n \geq 1$. If, in addition, the eigenfunction ϕ is positive, then ϕ is referred to as the *principal eigenfunction*, its eigenvalue ρ as the *principal eigenvalue*, and the pair (ρ, ϕ) as the *principal eigenpair*.[7] As a consequence of the factorization $\mathbb{M}_n = \mathbb{M}^n$, the pair (ρ, ϕ) are the principal eigenpair if and only if

$$\mathbb{M}\phi = \rho\phi \tag{4}$$

where ϕ is positive.

Alvarez and Jermann (2005) decompose the pricing kernel into its permanent and transitory components M_t^P and M_t^T, respectively. Their decomposition is $M_t = M_t^P M_t^T$ where $\mathbb{E}_t[M_{t+1}^P] = M_t^P$ and where $M_t^T / M_{t+1}^T = R_{t+1,\infty}$ is the gross return from t to $t+1$ on a risk-free bond with infinite maturity. Hansen and Scheinkman (2009) show that, in Markov environments, the permanent and transitory components of the pricing kernel are $M_t^P = \rho^{-t} M_t \phi(X_t)$ and $M_t^T = \rho^t \phi(X_t)^{-1}$ and the permanent and transitory components of the SDF are

$$\frac{M_{t+1}^P}{M_t^P} = \rho^{-1} \frac{M_{t+1}}{M_t} \frac{\phi(X_{t+1})}{\phi(X_t)} \tag{5}$$

$$\frac{M_{t+1}^T}{M_t^T} = \rho \frac{\phi(X_t)}{\phi(X_{t+1})} \tag{6}$$

for each date $t \in \{0, 1, \ldots\}$. The martingale property of the permanent component may be used to

[6] If the SDF depends on additional variables, then $m(X_t, X_{t+1})$ may be interpreted as the conditional expectation of the SDF with respect to (X_t, X_{t+1}) (Hansen and Scheinkman, 2013).

[7] Under the identification conditions presented below, ρ will be the largest eigenvalue of $\mathbb{M}$ and ϕ will be the unique positive eigenfunction of $\mathbb{M}$ (in an appropriately chosen parameter space).

define a distorted conditional expectation $\widetilde{\mathbb{E}}$, where

$$\widetilde{\mathbb{E}}[\psi(X_{t+n})|X_t = x] := \mathbb{E}\left[\frac{M^P_{t+n}}{M^P_t}\psi(X_{t+n}) \middle| X_t = x \right] .$$

Hansen and Scheinkman (2009) provide a set of stochastic stability conditions for continuous-time environments under which

$$\widetilde{\mathbb{E}}[\psi(X_{t+n})|X_t = x] \to \int_{\mathcal{X}} \frac{\psi(u)}{\phi(u)} \,\mathrm{d}\hat{\varsigma}(u) \tag{7}$$

for some probability measure $\hat{\varsigma}$.[8] The measure $\hat{\varsigma}$ will, in general, be different from the unconditional distribution of X_t. Further, Hansen and Scheinkman (2009) show that (7) implies

$$\rho^{-n}\mathbb{M}_n\psi(X_t) \to \left(\int_{\mathcal{X}} \frac{\psi(u)}{\phi(u)} \,\mathrm{d}\hat{\varsigma}(u) \right) \phi(X_t) \tag{8}$$

as $n \to \infty$. Equation (8) makes precise the sense in which ρ captures the yield on long-horizon assets and ϕ captures state dependence of the prices of long-horizon assets. In particular, $y := -\log \rho$ may be interpreted as the yield on a bond with infinite maturity. We derive a version of (8) for stationary, discrete time environments under different conditions from Hansen and Scheinkman (2009) (see Theorem 3.3). We show that $\hat{\varsigma}$ in stationary environments is characterized by ϕ and a positive eigenfunction ϕ^* of a time-reversed pricing operator.

Alvarez and Jermann (2005), Hansen (2012) and Backus et al. (2014) study the entropy of a SDF and its permanent and transitory components. Recall that the entropy of a positive random variable Z is defined as $L(Z) = \log \mathbb{E}[Z] - \mathbb{E}[\log Z]$. In stationary, discrete-time environments, the entropy of the permanent component of the SDF takes the convenient form

$$L\Big(\frac{M^P_{t+1}}{M^P_t}\Big) = \log \rho - \mathbb{E}\Big[\log \Big(\frac{M_{t+1}}{M_t} \Big) \Big] \tag{9}$$

whenever $\mathbb{E}[\log \phi(X_0)]$ is finite, which we assume hereafter.

Alvarez and Jermann (2005) derive the bound

$$L\Big(\frac{M^P_{t+1}}{M^P_t}\Big) \geq \mathbb{E}[\log R_{t+1}] - \mathbb{E}[\log R_{t+1,\infty}] \tag{10}$$

where R_{t+1} is the gross return from t to $t+1$ on a generic asset. The inequality (10) complements the Hansen and Jagannathan (1991) bound in that it restricts the size of the permanent component of the SDF as a function of observed asset prices. The entropy bound (10) has a short-term counterpart,

[8] The stochastic stability conditions in Hansen and Scheinkman (2009) are for continuous-time environments. See Appendix A for a version of their conditions for discrete-time environments. Appendix A also shows the relation between the conditions imposed in the present paper and the discrete-time version of Hansen and Scheinkman's (2009) stochastic stability conditions.

namely

$$L\Big(\frac{M_{t+1}}{M_t}\Big) \geq \mathbb{E}[\log R_{t+1}] - \mathbb{E}[\log R_{t+1,f}] \tag{11}$$

where $R_{t+1,f}$ denotes the gross return from t to $t+1$ on the risk-free asset. Alvarez and Jermann (2005) also derive bounds on the size of the transitory component and conditional versions of (10). Bakshi and Chabi-Yo (2012, 2014) refine the bounds of Alvarez and Jermann (2005) and derive revealing bounds for the entropy of the square of the permanent component of the SDF.

The permanent component of the SDF is invariant under a certain transformations of the pricing kernel. As argued by Bansal and Lehmann (1997), Hansen (2012) and Backus et al. (2014), models may generate different short-term asset pricing implications but behave very similarly over long horizons. Following Hansen (2012), we may construct alternative pricing kernel processes $\{M_t^*\}$ with

$$M_t^* = M_t \frac{f^*(X_t)}{f^*(X_0)} \tag{12}$$

for each $t \in \{0, 1, \ldots\}$ where f^* is a positive function. For example, M_t could be the C-CAPM pricing kernel and f^* may capture internal or external habit persistence. Crucially, this modification of $\{M_t\}$ does not alter its long-run pricing implications: both $\{M_t\}$ and $\{M_t^*\}$ have the same permanent component (and therefore the same entropy of the permanent component of the SDF), the same principal eigenvalue ρ, and the same long-run yield (see Hansen (2012) and Backus et al. (2014)). By analyzing the permanent component of the SDF for a single model we can, therefore, make inferences about a much broader class of models. In particular, the long-term pricing implications of many models with habit persistence and models with a limiting type of recursive preferences are identical to the long-term pricing implications the C-CAPM.

3 Identification, existence, and long-run pricing

3.1 Identification

Multiplicity of positive eigenfunctions is an issue without further restrictions on the parameter space for ϕ (i.e. the space of functions to which ϕ is assumed to belong). Hansen and Scheinkman (2009) do not restrict the parameter space ex ante. Instead, they apply Markov process theory to derive a set of stochastic stability conditions for possibly non-stationary, continuous-time environments. Their stability conditions imply that the positive eigenfunction which is germane to their long-run approximation is unique, even though there may exist multiple positive eigenfunctions. In contrast, we follow Alvarez and Jermann (2005) and Backus et al. (2014) and confine our analysis to stationary, discrete-time environments. Under stationarity there is a natural way to restrict the parameter space for ϕ. Restricting the parameter space in this manner allows alternative identification conditions to be derived using operator theory. Our identification result shows that there is at most one positive eigenfunction in the parameter space. A long-run approximation result (Theorem

3.3) shows that this positive eigenfunction is indeed germane to the long-run approximation. The link between our regularity conditions are summarized below and discussed in greater detail in Appendix A.

Following previous work on stationary continuous-time Markov processes,[9] a natural parameter space in which to consider identification is the space $L^p(\mathcal{X}, \mathscr{X}, Q)$ with $1 \leq p < \infty$ where $\mathcal{X} \subseteq \mathbb{R}^d$ is the support of the state process, $\mathscr{X}$ is the Borel σ-algebra on $\mathcal{X}$, and Q is the stationary (i.e. unconditional) distribution of the state process. Note that we use do not require that $\{X_t\}$ be embeddable by a continuous-time process. Let L^p denote the space $L^p(\mathcal{X}, \mathscr{X}, Q)$.

Assumption 3.1 $\mathbb{M} : L^p \to L^p$ *is a bounded linear operator of the form*

$$\mathbb{M}\psi(x) = \int_{\mathcal{X}} \mathcal{K}_{\mathbb{M}}(x, y)\psi(y)\,\mathrm{d}Q(y)$$

for some measurable $\mathcal{K}_{\mathbb{M}} : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$.

When the joint density $f_{0,1}(x_t, x_{t+1})$ of (X_t, X_{t+1}) exists and the unconditional density $f(x_t)$ of X_t exists and is positive, then

$$\mathcal{K}_{\mathbb{M}}(x, y) = \frac{f_{0,1}(x, y)}{f(x)f(y)} m(x, y)\,.$$

Boundedness of $\mathbb{M}$ may be verified using the Schur test.

Assumption 3.2 *(a)* $\mathcal{K}_{\mathbb{M}}(x, y) \geq 0$ *a.e.-*$[Q \otimes Q]$*; and*

(b) for all $S \in \mathscr{X}$ *with* $0 < Q(S) < 1$ *we have*

$$\int_{\mathcal{X}\setminus S} \left(\int_S \mathcal{K}_{\mathbb{M}}(x, y)\,\mathrm{d}Q(x) \right) \mathrm{d}Q(y) > 0\,.$$

Let q be such that $p^{-1} + q^{-1} = 1$ and let L^q denote the space $L^q(\mathcal{X}, \mathscr{X}, Q)$. Following earlier work by Hansen and Scheinkman (1995) on continuous-time Markov processes, define the *time-reversed pricing operator* $\mathbb{M}^* : L^q \to L^q$ as

$$\mathbb{M}^*\psi(x) = \mathbb{E}[m(X_0, X_1)\psi(X_0)|X_1 = x]$$

for $\psi \in L^q$.

[9] See Hansen and Scheinkman (1995), Florens, Renault, and Touzi (1998), and Chen, Hansen, and Scheinkman (2009) who study continuous-time Markov processes. In contrast, we deal with general discrete-time Markov processes.

Assumption 3.3 *There exists $\phi \in L^p$, $\phi^* \in L^q$ and a constant ρ such that*

$$\begin{aligned} \mathbb{M}\phi &= \rho\phi \\ \mathbb{M}^*\phi^* &= \rho\phi^* \end{aligned}$$

with $\phi, \phi^* > 0$ a.e.-$[Q]$.

Assumption 3.1 is a mild condition which places some basic structure on $\mathbb{M}$ and ensures that $\mathbb{M}^*$ is well defined. Assumption 3.2(a) is included mainly for completeness, and is trivially satisfied when $f_{0,1}(x_t, x_{t+1})$ and $f(x_t)$ exist and are positive. Assumption 3.2(b) is a direct extension of the irreducibility criterion in the Perron-Frobenius theorem for matrices. The stronger "strict positivity" condition $K(x, y) > 0$ a.e.-$[Q \otimes Q]$ implies Assumption 3.2 but is stronger than needed for identification and existence. Assumption 3.3 merely requires that both $\mathbb{M}$ and its time-reversed counterpart $\mathbb{M}^*$ have positive eigenfunctions corresponding to the same eigenvalue ρ. This condition is verified directly for three parametric models in Appendix B, namely exponentially affine and exponentially quadratic SDFs, and a model with recursive preferences with stochastic volatility.

Theorem 3.1 *Let Assumptions 3.1, 3.2 and 3.3 hold. Then:*

(a) if $\zeta \in L^p$ is a positive eigenfunction of $\mathbb{M}$ then $\frac{\zeta(x)}{\phi(x)}$ is constant a.e.-$[Q]$

(b) if $\zeta^* \in L^q$ is a positive eigenfunction of $\mathbb{M}^*$ then $\frac{\zeta^*(x)}{\phi^*(x)}$ is constant a.e.-$[Q]$

(c) ρ is the unique eigenvalue of $\mathbb{M}$ with a positive eigenfunction.

To summarize the connection between Assumptions 3.1–3.3 and Hansen and Scheinkman (2009), Assumption 3.2 is analogous to their positivity and irreducibility conditions (their Assumptions 7.1 and 7.3). Assumption 3.3 and the maintained assumption of stationarity are sufficient for existence of an invariant distribution for the distorted conditional expectations (their Assumption 7.2). Assumption 3.3 is altogether different from their Harris recurrence condition (their Assumption 7.4). Harris recurrence, together with their other assumptions, implies convergence of the distorted conditional expectations which they use to achieve identification. We instead assume existence of ϕ^* which does not imply convergence of the distorted conditional expectations (cf. Theorem 3.3). Appendix A provides a more detailed comparison of the two sets of identification conditions.

A similar result to part (a) of Theorem 3.1 is reported in a preliminary draft of Hansen and Scheinkman (2009) from 2005.[10] In their earlier draft, the parameter space is a L^p space defined by an arbitrary measure and the environment is again formulated in continuous time. They impose

[10]See `http://www.cirano.qc.ca/realisations/grandes_conferences/methodes_econometriques/hansen.pdf`, dated November 24, 2005.

different positivity and irreducibility conditions on the (continuous-time) semigroup of pricing operators and assume that the semigroup and dual semigroup both have positive eigenfunctions ϕ and ϕ^* corresponding to the same eigenvalue.[11] Their identification result establishes uniqueness of ϕ but does not establish uniqueness of ϕ^*.

Escanciano and Hoderlein (2012), Chen et al. (2014), and Christensen (2014) have used related function-analytic methods to derive identification conditions of positive eigenfunctions of various operators. Each of these papers impose various positivity or irreducibility conditions on the operator and assume either that the operator is compact (Escanciano and Hoderlein, 2012; Chen et al., 2014) or power compact (Christensen, 2014). Theorem 3.1 does not require compactness or power compactness of $\mathbb{M}$ to nonparametrically identify ϕ and ϕ^*. Compactness or power compactness of $\mathbb{M}$ is a stronger sufficient condition for Assumption 3.3 (see Theorem 3.2 below). Further, Assumption 3.3 is simple to verify for common parametric models (see Appendix B).

In certain cases $\mathbb{M}$ may not be represented as an integral operator with measurable kernel. This includes when X_t is formed by stacking a rth-order Markov process into a first-order process and certain in habit formation models when X_t includes r lags of consumption growth. Nevertheless, $\mathbb{M}_n$ might be represented as an integral operator with measurable kernel for some $n \geq r$. Theorem 3.1 can then be applied with $\mathbb{M}_n$ in place of $\mathbb{M}$ to achieve identification of ϕ and ϕ^*.[12]

3.2 Existence

We now use Perron-Frobenius theory to establish existence of positive eigenfunctions $\phi \in L^p$ and $\phi^* \in L^q$ of the forward and reverse-time pricing operators $\mathbb{M}$ and $\mathbb{M}^*$. In contrast, Hansen and Scheinkman (2009) use Markov process theory to provide a constructive proof for the existence of ϕ without restricting the space ex ante. Appendix A discusses the connection between their existence conditions and the existence conditions presented here.

Assumption 3.4 $\mathbb{M}_n$ *is compact for some* $n \geq 1$.

Assumption 3.4 is weaker than requiring $\mathbb{M}$ to be compact (which corresponds to taking $n = 1$). There exist different sufficient conditions for Assumption 3.4 for different choices of p. Consider L^2, which will be the space of interest for estimation. Iterated kernels may be calculated via the recurrence relation

$$\mathcal{K}_{\mathbb{M}}^{(n)}(x,y) = \int_{\mathcal{X}} \mathcal{K}_{\mathbb{M}}^{(n-1)}(x,u)\mathcal{K}_{\mathbb{M}}(u,y)\,\mathrm{d}Q(u)$$

[11] Note that ϕ^* has the interpretation of an eigenfunction of the time-reversed operator only when $\{X_t\}$ is stationary and its stationary distribution Q is used to define the L^p space.

[12] If $\mathbb{M}$ has multiple positive eigenfunctions then $\mathbb{M}_n$ must also have multiple positive eigenfunctions. Thus, uniqueness of positive eigenfunctions of $\mathbb{M}_n$ implies uniqueness of positive eigenfunctions of $\mathbb{M}$.

with $\mathcal{K}_{\mathbb{M}}^{(1)}(x,y) = \mathcal{K}_{\mathbb{M}}(x,y)$. If there exists $n \geq 1$ for which

$$\int_{\mathcal{X}} \int_{\mathcal{X}} \mathcal{K}_{\mathbb{M}}^{(n)}(x,y)^2 \, \mathrm{d}Q(x) \, \mathrm{d}Q(y) < \infty \tag{13}$$

then $\mathbb{M}_n$ is Hilbert-Schmidt and therefore compact (see Carrasco et al. (2007)).[13] Assumption 3.4 is satisfied for L^1 if there exists $n \geq 1$ such that $\mathbb{M}_n$ maps L^1 into L^r for some $r > 1$; moreover, Assumption 3.4 is satisfied for L^p with $1 < p < \infty$ if there exists $n \geq 1$ such that $\mathbb{M}_n$ maps L^p into L^∞ (Schaefer, 1974, p. 337). In what follows, we say that the eigenvalue ρ is simple if it has a unique eigenfunction (up to scale) and we say that ρ is isolated if there exists a neighborhood of ρ such that ρ is the unique element of the spectrum of $\mathbb{M}$ belonging to the neighborhood.

Theorem 3.2 *Let Assumptions 3.1, 3.2 and 3.4 hold. Then:*

(a) there exist positive eigenfunctions ϕ and ϕ^* satisfying Assumption 3.3

(b) ρ is simple, isolated, and the largest real eigenvalue of $\mathbb{M}$.

It follows from Theorems 3.1 and 3.2 that ϕ and ϕ^* exist and are identified under Assumptions 3.1, 3.2 and 3.4. A similar existence result to part (a) was presented in a 2005 preliminary version of Hansen and Scheinkman (2009). There, they assumed that the spectral radius of $\mathbb{M}$ was positive and that their (continuous-time) semigroup of operators had an element which was compact. Their latter compactness condition is a continuous-time counterpart to Assumption 3.3. The further properties of ρ that we establish in part (b) of Theorem 3.2 are essential to our derivation of the large-sample theory. A similar proposition was derived under different positivity and irreducibility conditions in Christensen (2014).

3.3 Long-run pricing

We now derive a version of the long-run pricing approximation of Hansen and Scheinkman (2009) under a slight strengthening of the identification and existence conditions. To do so, we replace Assumption 3.2 with the following condition.

Assumption 3.5 $\mathcal{K}_{\mathbb{M}}(x,y) > 0$ *a.e.*-$[Q \otimes Q]$.

We impose the normalizations $\mathbb{E}[\phi(X_0)^p] = 1$ and $\mathbb{E}[\phi(X_0)\phi^*(X_0)] = 1$ and define the operator $(\phi \otimes \phi^*) : L^p \to L^p$ by

$$(\phi \otimes \phi^*)\psi(x) = \phi(x) \int_{\mathcal{X}} \phi^*(u)\psi(u) \, \mathrm{d}Q(u) \, .$$

[13] Similar Hilbert-Schmidt conditions have been used recently by Darolles et al. (2011), Escanciano and Hoderlein (2012), Connor, Hagmann, and Linton (2012), and Chen et al. (2014) to study nonparametric identification and/or estimation in other contexts. In these other applications the Hilbert-Schmidt condition is used to establish compactness of the operator whereas here we use it to establish power compactness.

Theorem 3.3 *Let Assumptions 3.1, 3.4, and 3.5 hold. Then there exists $c > 0$ such that*

$$\sup_{\psi \in L^p : \mathbb{E}[|\psi(X_0)|^p] \leq 1} \int_{\mathcal{X}} \left| \rho^{-n} \mathbb{M}_n \psi(x) - (\phi \otimes \phi^*) \psi(x) \right|^p \mathrm{d}Q(x) = O(e^{-cn})$$

as $n \to \infty$.

Theorem 3.3 establishes convergence of $\rho^{-n}\mathbb{M}_n$ uniformly in L^p norm, with the approximation error vanishing exponentially in the payoff horizon n. A similar proposition (without the exponential rate of convergence) was reported in a 2005 preliminary draft of Hansen and Scheinkman (2009). There, they assumed directly that the distorted conditional expectations converged do an unconditional expectation characterized by ϕ, ϕ^*, and the arbitrary measure used to define the L^p space. Here, we instead show that convergence obtains under a very slight strengthening of the conditions of Theorem 3.2.

Theorem 3.3 is similar to the various long-run approximation results presented in Section 7 of Hansen and Scheinkman (2009) which imply

$$\rho^{-n}\mathbb{M}_n\psi(x) \to \left(\int_{\mathcal{X}} \frac{\psi(u)}{\phi(u)} \, \mathrm{d}\hat{\varsigma}(u) \right) \phi(x) \tag{14}$$

for some probability measure $\hat{\varsigma}$. In general, possibly non-stationary environments it not clear how to calculate the measure $\hat{\varsigma}$. In stationary environments, however, Theorem 3.3 shows that

$$\rho^{-n}\mathbb{M}_n\psi(x) \to \left(\int_{\mathcal{X}} \frac{\psi(u)}{\phi(u)} \phi(u)\phi^*(u) \, \mathrm{d}Q(u) \right) \phi(x) \, . \tag{15}$$

Comparing (14) and (15), we see that the Radon-Nikodym derivative of $\hat{\varsigma}$ with respect to Q is

$$\frac{\mathrm{d}\hat{\varsigma}(x)}{\mathrm{d}Q(x)} = \phi(x)\phi^*(x) \, .$$

Therefore, in stationary, discrete-time environments the measure $\hat{\varsigma}$ is characterized by the distribution Q and the normalized positive eigenfunctions of $\mathbb{M}$ and $\mathbb{M}^*$.

Remark 3.1 *The preceding analysis could equally be applied to study valuation with a stochastic growth component as in Hansen, Heaton, and Li (2008), Hansen and Scheinkman (2009), Hansen (2012), Lettau and Wachter (2007, 2011), and others. Assuming the reference growth process from time t to $t+1$ is $G(X_t, X_{t+1})$ for some measurable $G : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ and existence of $f_{0,1}$ and f, the stochastic growth operator $\mathbb{S}$ and the valuation-with-stochastic-growth operator $\mathbb{T}$ have kernels*

$$\mathcal{K}_{\mathbb{S}}(x,y) = \frac{f_{0,1}(x,y)}{f(x)f(y)} G(x,y) \quad \textit{and} \quad \mathcal{K}_{\mathbb{T}}(x,y) = \frac{f_{0,1}(x,y)}{f(x)f(y)} G(x,y) m(x,y) \, ,$$

respectively.

4 Estimators and large-sample theory

This section introduces the nonparametric estimators of ρ, ϕ, ϕ^*, the long-run yield, and the entropy of the permanent component of the SDF and describes the large-sample properties of the estimators.

Sieve methods work by projecting the infinite-dimensional eigenfunction problem (4) onto a subspace spanned by finitely many known basis functions. This finite-dimensional approximation means that, at the cost of introducing some bias, the eigenfunction problem (4) can be rewritten as a low-dimensional matrix eigenvector problem. The matrices are estimated from a time series of data on $\{X_t\}$,[14] from which the estimators of ρ, ϕ, and related quantities are easily calculated. The estimators may also be used to numerically compute for ρ, ϕ, and related quantities fully specified models for which analytical solutions are unavailable.

The estimators introduced below build on previous work by Chen et al. (2000) and Gobet et al. (2004) who applied sieve methods to nonparametrically estimate eigenfunctions of the selfadjoint conditional expectation operator associated with a stationary, scalar diffusion process. However, in our context the operator $\mathbb{M}$ will typically be nonselfadjoint. This introduces several additional technicalities. First, if $\mathbb{M}$ is selfadjoint then the problem of estimating the time-reversed positive eigenfunction ϕ^* disappears because, in that circumstance, $\phi = \phi^*$. Thus our results on estimating time-reversed eigenfunctions are new. Second, if $\mathbb{M}$ is selfadjoint then ρ and ϕ solve an infinite-dimensional maximization problem (by the Courant-Fischer minimax theorem); this is not so for the nonselfadjoint case. Therefore, we apply perturbation methods, rather than extremum estimator asymptotics, to derive the large-sample theory. Our convergence rates for estimators of ρ and ϕ are obtained by modifying some arguments in Gobet et al. (2004); the convergence rates for the estimator of ϕ^* and the derivation of the asymptotic distribution and efficiency bounds of the parametric estimators are all new.

For the remainder of this section we use L^2 as the parameter space for ϕ and ϕ^* because it is endowed with the inner product $\langle \psi_1, \psi_2 \rangle = \mathbb{E}[\psi_1(X_0)\psi_2(X_0)]$.

4.1 Sieve approximation

Let $b_{K1}, \ldots, b_{KK} \in L^2$ be a dictionary of linearly independent basis functions (polynomials, splines, wavelets, Fourier series, etc) and let B_K denote the closed linear span of $b_{K1}, \ldots, b_{KK}$. We now construct an approximation of (4) in the space B_K. Any function $\psi \in B_K$ may be written as

$$\psi(x) = b^K(x)' c_K(\psi)$$

[14] As discussed below, the convergence rate calculations may be applied to study models with latent state variables despite our maintained assumption of an observable state vector.

where $b^K(x) = (b_{K1}(x), \ldots, b_{KK}(x))'$ is a vector of basis functions and $c_K(\psi) \in \mathbb{R}^K$ is a vector of coefficients. Define the Gram matrix

$$\mathbf{G}_K = \mathbb{E}[b^K(X_0)b^K(X_0)'] \tag{16}$$

which is $K \times K$, symmetric, and positive definite. The sieve space B_K is isometrically isomorphic to $\mathbb{R}^K$ endowed with the inner product $(u, v) \mapsto u'\mathbf{G}_K v$ under the isometry $\psi \mapsto c_K(\psi)$ because

$$\mathbb{E}[\psi_1(X_0)\psi_2(X_0)] = c_K(\psi_1)'\mathbf{G}_K c_K(\psi_2)$$

for all $\psi_1, \psi_2 \in B_K$.

To describe the finite-dimensional approximation of (4), let $\Pi_K : L^2 \to B_K$ denote the orthogonal projection onto B_K. Consider the eigenfunction problem

$$(\Pi_K \mathbb{M})\phi_K = \rho_K \phi_K \tag{17}$$

where ρ_K is the largest real eigenvalue of $\Pi_K\mathbb{M}$. This problem will be well defined for all K sufficiently large under Assumptions 4.1 and 4.2(a) below: ρ_K will be positive and simple and will therefore have a unique eigenfunction ϕ_K. Since $\phi_K \in B_K$ we may write $\phi_K = b^K(x)'c_K$ where $c_K = c(\phi_K)$ is a vector of coefficients. Also define the $K \times K$ matrix

$$\mathbf{M}_K = \mathbb{E}[b^K(X_0)m(X_0, X_1)b^K(X_1)'] \,. \tag{18}$$

To simplify notation, let $\mathbf{M} = \mathbf{M}_K$ and $\mathbf{G} = \mathbf{G}_K$ hereafter. The approximate eigenvalue problem (17) may then be rewritten as

$$b^K(x)'\mathbf{G}^{-1}\mathbf{M}c_K = \rho_K b^K(x)c_K$$

or, equivalently,

$$\mathbf{G}^{-1}\mathbf{M}c_K = \rho_K c_K$$

where ρ_K is the largest real eigenvalue of $\mathbf{G}_K^{-1}\mathbf{M}_K$ and c_K is its eigenvector. Similar logic leads us to approximate $\phi^*(x)$ by $\phi_K^+(x) = b^K(x)'c_K^*$ where

$$\mathbf{G}^{-1}\mathbf{M}'c_K^* = \rho_K c_K^*$$

(here we use the superscript "+" to denote that $b^K(x)c_K^*$ is the eigenfunction of the adjoint of $\Pi_K\mathbb{M}$ with respect to B_K rather than L^2; see Appendix C). Moreover, ρ_K is the largest real generalized eigenvalue of the pair $(\mathbf{M}, \mathbf{G})$ and c_K and c_K^* are its right- and left eigenvectors:

$$\begin{aligned} \mathbf{M}c_K &= \rho_K \mathbf{G}c_K \\ c_K^{*\prime}\mathbf{M} &= \rho_K c_K^{*\prime}\mathbf{G}\,. \end{aligned} \tag{19}$$

Display (19) provides the basis for numerical computation of ρ, ϕ, and ϕ^* in models for which analytical solutions are unavailable. For such models, the matrices $\mathbf{M}$ and $\mathbf{G}$ may be computed directly, via simulation or numerical integration, from which the approximate solutions ρ_K, ϕ_K and ϕ_K^+ for ρ, ϕ and ϕ^* can be recovered by solving (19). Lemma C.1 in Appendix C provides bounds on the approximation errors $\rho_K - \rho$, $\phi_K - \phi$, and $\phi_K^+ - \phi^*$.

4.2 Estimators

To estimate ρ, ϕ and ϕ^* we replace $\mathbf{M}$ and $\mathbf{G}$ in (19) by estimators $\widehat{\mathbf{M}}$ and $\widehat{\mathbf{G}}$, where $\widehat{\mathbf{G}}$ is positive definite and symmetric. We then solve

$$\begin{aligned} \widehat{\mathbf{G}}^{-1}\widehat{\mathbf{M}}\,\widehat{c} &= \widehat{\rho}\widehat{c} \\ \widehat{\mathbf{G}}^{-1}\widehat{\mathbf{M}}'\widehat{c}^* &= \widehat{\rho}\widehat{c}^* \end{aligned} \qquad (20)$$

where $\widehat{\rho}$ is the maximum real eigenvalue of $\widehat{\mathbf{G}}^{-1}\widehat{\mathbf{M}}$. The estimator of ϕ is $\widehat{\phi}(x) = b^K(x)'\widehat{c}$ and the estimator of ϕ^* is $\widehat{\phi}^*(x) = b^K(x)'\widehat{c}^*$. Under Assumptions 4.1, 4.2, and 4.3(a) below, with probability approaching one, $\widehat{\rho}$ will be positive and simple and so its eigenvectors $\widehat{c}$ and $\widehat{c}^*$ will be unique.[15]

Given a time series of data $\{X_0, X_1, \ldots, X_n\}$, we estimate $\mathbf{G}$ using

$$\widehat{\mathbf{G}} = \frac{1}{n}\sum_{t=0}^{n-1} b^K(X_t)b^K(X_t)'\,. \qquad (21)$$

We consider three possibilities for estimating $\mathbf{M}$.

Case 1: SDF is observable First, consider the case in which the one-period SDF $m(X_t, X_{t+1})$ is known. This is the case for the C-CAPM under CRRA preferences in which $m(X_t, X_{t+1}) = m(X_t, X_{t+1}; \beta, \gamma) = \beta G_{t+1}^{-\gamma}$ with fixed β and γ, provided $G_{t+1} = G(X_t, X_{t+1})$ for some measurable function G. Other examples include the SDFs obtained under the external habit preference specifications of Abel (1990) and Galí (1994). As the SDF is observable, we may estimate $\mathbf{M}$ using

$$\widehat{\mathbf{M}} = \frac{1}{n}\sum_{t=0}^{n-1} b^K(X_t)m(X_t, X_{t+1})b^K(X_{t+1})'\,. \qquad (22)$$

Case 2: SDF has unobservable components There exist several popular asset pricing models in which components of the SDF depend implicitly on the law of motion of $\{X_t\}$. For these models the functional form of m is unknown when we model the dynamics nonparametrically. Nevertheless, our estimators can still be applied provided an estimate of the unobservable component is first

[15] When $\widehat{\mathbf{G}}^{-1}\widehat{\mathbf{M}}$ has no real positive eigenvalues or when its maximum real eigenvalue is not positive and simple, then we can simply take $\rho = 1$ and set $\widehat{\phi}(x) = \widehat{\phi}^*(x) = 1$ for all x without altering the convergence rates or limiting distribution of the estimators.

"plugged in" to the SDF. For example, under Epstein-Zin preferences the SDF depends implicitly on the continuation value of future consumption which is unobservable when the dynamics of the state are modeled nonparametrically. In Section 5 we introduce a procedure to estimate the continuation value from a time series $\{X_0, X_1, \dots, X_n\}$ using sieve methods. By plugging in the estimated continuation value $\widehat{v}$, we can form a time series $\widehat{m}(X_t, X_{t+1}) = m(X_t, X_{t+1}; \widehat{v})$ for $t = 0, 1, \dots, n-1$. Our estimator of $\mathbf{M}$ in this case is

$$\widehat{\mathbf{M}} = \frac{1}{n} \sum_{t=0}^{n-1} b^K(X_t) \widehat{m}(X_t, X_{t+1}) b^K(X_{t+1})' . \tag{23}$$

Models with internal habit formation also have an unobservable forward-looking expectation in the marginal utility of consumption. It is not clear how to nonparametrically estimate this forward-looking component from data on the state alone. In the absence of such an estimator, the procedure introduced in this paper cannot be used to study internal habit formation models under counterfactual preference parameters.

Case 3: SDF is estimated The third case we consider is that in which the econometrician wishes to extract ρ, ϕ, and related quantities from a SDF that has been estimated from data on both X_t and asset returns over the period $t = 0, \dots, n$. This is different from Case 2 in two respects. First, the data used for estimation includes both data on the state and returns, whereas Case 2 only uses data on the state. Second, here the parameters in the estimated SDF are those that are implied by the returns data. In contrast, the approach taken in Case 2 allows us to estimate components of the SDF that are consistent with given, possibly counterfactual, preference parameters.

The method by which $\widehat{m}$ is extracted is not important for our purposes. All that we require is that the researcher may evaluate $\widehat{m}(X_t, X_{t+1})$ for each date $t = 0, \dots, n-1$. If so, then $\mathbf{M}$ may be estimated as in (23). For example, conventional moment-based methods such as GMM, minimum distance (such as the estimator for internal habit formation models introduced in Chen and Ludvigson (2009)), or empirical likelihood may be used to estimate the SDF from moment restrictions based on the Euler equation. Procedures based on options data, such the extended method of moments (Gagliardini, Gourieroux, and Renault, 2011) or nonparametric state-price density estimators (Aït-Sahalia and Lo, 1998), could also be used. This approach allows researchers to use sieve methods to recover the permanent component of the SDF implicit in option prices without discretizing the state space (cf. Ross (2014)).

It remains to introduce estimators for the long-run yield y and entropy of the permanent component $L(M_{t+1}^P / M_t^P)$ which, for simplicity, we denote by L. Coherently with the formulae $y = -\log \rho$ and $L = \log \rho - \mathbb{E}[\log m(X_0, X_1)]$, we estimate y using

$$\widehat{y} = -\log \widehat{\rho} \tag{24}$$

and we estimate L using

$$\widehat{L} = \log \widehat{\rho} - \frac{1}{n} \sum_{t=0}^{n-1} \log m(X_t, X_{t+1}) \tag{25}$$

in Case 1, and

$$\widehat{L} = \log \widehat{\rho} - \frac{1}{n} \sum_{t=0}^{n-1} \log \widehat{m}(X_t, X_{t+1}) \tag{26}$$

in Cases 2 and 3.

4.3 Consistency and convergence rates

We first state three basic assumptions which are used to develop the large-sample theory. The first is an identifying assumption, the second relates to the approximation properties of the sieve, and the third relates to the convergence properties of the matrix estimators.

In what follows, we let $\|\cdot\|$ denote the L^2 norm when applied to functions and the L^2 operator norm, given by $\|A\| = \sup\{\|A\psi\| : \psi \in L^2, \|\psi\| = 1\}$, when applied to linear operators $A : L^2 \to L^2$. We also let $\|\cdot\|_{\mathbf{G}}$ denote the vector and matrix norms on $\mathbb{R}^K$ induced by $\mathbf{G}$, i.e., $\|v\|_{\mathbf{G}}^2 = v'\mathbf{G}v$ for vectors and $\|\mathbf{A}\|_{\mathbf{G}} = \sup\{\|\mathbf{A}v\|_{\mathbf{G}} : v \in \mathbb{R}^K, \|v\|_{\mathbf{G}} = 1\}$ for matrices.

Assumption 4.1 *Assumptions 3.1, 3.2, and 3.4 hold for the space L^2.*

Assumption 4.2 *(a) $\|\Pi_K \mathbb{M} - \mathbb{M}\| = o(1)$;*

(b) $\|\Pi_K \phi - \phi\| = O(\delta_K)$; and

(c) $\|\Pi_K \phi^* - \phi^*\| = O(\delta_K^*)$.

Assumption 4.3 *(a) $\|\widehat{\mathbf{G}}^{-1}\widehat{\mathbf{M}} - \mathbf{G}^{-1}\mathbf{M}\|_{\mathbf{G}} = o_p(1)$;*

(b) $\|(\widehat{\mathbf{G}}^{-1}\widehat{\mathbf{M}} - \mathbf{G}^{-1}\mathbf{M})c_K\|_{\mathbf{G}} = O_p(\eta_{n,K})$; and

(c) $\|(\widehat{\mathbf{G}}^{-1}\widehat{\mathbf{M}}' - \mathbf{G}^{-1}\mathbf{M}')c_K^* / \|c_K^*\|_{\mathbf{G}}\|_{\mathbf{G}} = O_p(\eta_{n,K}^*)$.

Assumption 4.1 guarantees existence and identification of the positive eigenfunctions ϕ and ϕ^* in the space L^2. Assumption 4.1 also guarantees that ρ is isolated and simple. These two properties are used extensively in the derivation of the large sample theory. Assumption 4.1 can be replaced with the higher-level assumption that ϕ and ϕ^* exist and correspond to an isolated, simple eigenvalue ρ where ρ is the maximum eigenvalue of $\mathbb{M}$.

Assumption 4.2(a) is a condition on how well the range of $\mathbb{M}$ can be approximated over the sieve space B_K. This assumption necessarily requires that $\mathbb{M}$ is compact,[16] as has been assumed previously in the literature on sieve estimation of eigenfunctions (see, e.g., Gobet et al. (2004)). If $\mathbb{M}$ is not compact but $\mathbb{M}_\ell$ is compact for some $\ell \geq 2$, then one can apply the estimators to $\mathbb{M}_\ell$ in place of $\mathbb{M}$; consistency and convergence rates of ρ^ℓ, ϕ and ϕ^* would then follow directly from Theorem 4.1. Assumption 4.2(b)(c) are conditions on how well ϕ and ϕ^* may be approximated by elements of B_K. Values for δ_K and δ_K^* are known for common choices of sieve under standard smoothness assumptions (see Chen (2007)).

Assumption 4.3(a) requires that the estimator $\widehat{\mathbf{G}}^{-1}\widehat{\mathbf{M}}$ converges to $\mathbf{G}^{-1}\mathbf{M}$ in the matrix norm induced by $\mathbf{G}$. This condition imposes a restriction on the maximum rate at which K can grow with n, which will be determined by both the choice of sieve and the weak dependence properties of the data. When m is a function of other estimators (of, say, an unobservable component as in Case 2 or unknown parameters as in Case 3) then Assumption 4.3(a) also places restrictions on the rate at which the first-stage estimators must converge. Sufficient conditions for Assumption 4.3 are presented in Appendix C. Finally, Assumptions 4.2(a) and 4.3(a) imply $\delta_K, \delta_K^* = o(1)$ and $\eta_{n,K}, \eta_{n,K}^* = o(1)$, respectively. The purpose of introducing the terms δ_K, δ_K^*, $\eta_{n,K}$, and $\eta_{n,K}^*$ is to obtain more refined convergence rates for ϕ and ϕ^*.

The following Theorem, which is the main result of this section, establishes consistency and convergence rates of the estimators. Recall that for Theorem 3.3 we have normalized ϕ and ϕ^* so that $\mathbb{E}[\phi(X_0)^2] = 1$ and $\mathbb{E}[\phi(X_0)\phi^*(X_0)] = 1$. As eigenfunctions are only normalized up to scale, we also impose the normalizations $\mathbb{E}[\widehat{\phi}(X_0)^2] = 1$ and $\mathbb{E}[\widehat{\phi}(X_0)\widehat{\phi}^*(X_0)] = 1$.

Theorem 4.1 *Let Assumptions 4.1, 4.2 and 4.3 hold. Then:*

(a) $|\widehat{\rho} - \rho| = O_p(\delta_K + \eta_{n,K})$

(b) $\|\widehat{\phi} - \phi\| = O_p(\delta_K + \eta_{n,K})$

(c) $\|\widehat{\phi}^*/\|\widehat{\phi}^*\| - \phi^*/\|\phi^*\|\| = O_p(\delta_K^* + \eta_{n,K}^*)$.

Remark 4.1 *Theorem 4.1 does note require that* $\widehat{\mathbf{G}}$ *and* $\widehat{\mathbf{M}}$ *be estimated as in (21), (22), and (23). In fact, Theorem 4.1 holds for estimators* $\widehat{\rho}$, $\widehat{\phi}$ *and* $\widehat{\phi}^*$ *calculated from any estimators* $\widehat{\mathbf{G}}$ *and* $\widehat{\mathbf{M}}$ *of* $\mathbf{G}$ *and* $\mathbf{M}$ *provided* $\widehat{\mathbf{G}}$ *is positive definite and symmetric and* $\widehat{\mathbf{G}}$ *and* $\widehat{\mathbf{M}}$ *satisfy Assumption 4.3.*

Remark 4.2 *Theorem 4.1 may be used to estimate eigenvalues and eigenfunctions of the stochastic growth and valuation-with-stochastic growth operators* $\mathbb{S}$ *and* $\mathbb{T}$ *(see Remark 3.1) by replacing* $\mathbf{M}$ *with* $\mathbf{S} = E[b^K(X_0)G(X_0, X_1)b^K(X_1)']$ *and* $\mathbf{T} = [b^K(X_0)G(X_0, X_1)m(X_0, X_1)b^K(X_1)']$*, respectively.*

[16] An operator is compact if and only if it is the limit (in operator norm) of a sequence of operators with finite-dimensional range (Carrasco et al., 2007, Theorem 2.29). Each $\Pi_K\mathbb{M}$ has range B_K where $\dim(B_K) = K < \infty$.

Theorem 4.1 is a direct consequence of Lemmas C.1 and C.2 which derive separately the convergence rates of the bias and variance terms. The bias calculations in Lemma C.1 may be used to bound the finite-dimensional approximation error when the estimators are used to numerically compute ρ, ϕ, and ϕ^* in fully-specified models. Theorem 4.1 exhibits the usual bias-variance tradeoff encountered in nonparametric estimation. The terms δ_K and δ_K^*, which represent the bias terms, will typically be decreasing in K because ϕ and ϕ^* will be approximated over increasingly rich subspaces as K increases. On the other hand, increasing K means that more parameters in $\widehat{\mathbf{G}}$ and $\widehat{\mathbf{M}}$ need to be estimated, which introduces additional sampling error. Therefore, the variance terms $\eta_{n,K}$ and $\eta_{n,K}^*$ will typically be increasing in K and decreasing in n. Note that Conclusions (a) and (b) of Theorem 4.1 hold under Assumptions 4.1, 4.2(a)(b), and 4.3(a)(b).

Remark 4.1 suggests that preceding Theorem might be applied to models with latent state variables. Fully nonparametric model with latent variables are not well identified. Yet certain latent processes possess enough structure that a filter or similar device may be used to estimate the latent time series from a related, observable time series. For such processes it might be possible to construct $\widehat{\mathbf{G}}$ and $\widehat{\mathbf{M}}$ from the estimate of the latent data. Consistency and convergence rates would then follow from Theorem 4.1 so long as Assumption 4.3 could be verified.

Corollary 4.1 *Let Assumptions 4.1, 4.2(a)(b) and 4.3(a)(b) hold. Then:*

(a) $|\widehat{y} - y| = O_p(\delta_K + \eta_{n,K})$.

If, in addition,

$$n^{-1}\sum_{t=0}^{n-1} \log m(X_t, X_{t+1}) - \mathbb{E}[\log m(X_0, X_1)] = O_p(\eta_{n,K}^L)$$

in Case 1, or

$$n^{-1}\sum_{t=0}^{n-1} \log \widehat{m}(X_t, X_{t+1}) - \mathbb{E}[\log m(X_0, X_1)] = O_p(\eta_{n,K}^L)$$

in Cases 2 or 3, then:

(b) $|\widehat{L} - L| = O_p(\delta_K + \eta_{n,K} + \eta_{n,K}^L)$.

To further investigate the theoretical properties of the estimators, we now derive the convergence rate of $\widehat{\phi}$ in Case 1, where $\widehat{\mathbf{G}}$ and $\widehat{\mathbf{M}}$ are as in (21) and (22), under standard conditions from the statistics literature on optimal convergence rates. Although the following conditions are not particularly appropriate in an asset pricing context, the result is informative about the convergence properties of $\widehat{\phi}$ relative to conventional nonparametric estimators.

Corollary 4.2 *Let Assumptions 3.1, 3.2 and the following conditions hold: (i)* $\mathcal{X} \subset \mathbb{R}^d$ *is compact, rectangular and has nonempty interior; (ii)* Q *has density* f *which is continuous and strictly positive* *on $\mathcal{X}$; (iii) $\mathbb{M}$ is a bounded operator from L^2 into a Hölder space $\Lambda^{p_0}(\mathcal{X})$ of smoothness $p_0 > 0$ (see Section 2.3.1 of Chen (2007)); (iv) $\phi \in \Lambda^p(\mathcal{X})$ with $p \geq p_0$; (v) $\mathbb{E}[|m(X_0, X_1)|^r]^{1/r} < \infty$ for some $2 \leq r \leq \infty$; (vi) B_K is a spanned by (a tensor product of) polynomial splines of degree $\nu \geq p$ with uniformly bounded mesh ratio (see Chapter 12 of Schumaker (2007)); and (vii) $\{X_t\}$ is exponentially rho-mixing. Then:*

(a) Assumptions 4.1 and 4.2(a)(b) hold with $\delta_K = O(K^{-p/d})$, and Assumption 4.3(a)(b) holds with $\eta_{n,K} = O(K^{(r+2)/2r}/\sqrt{n})$ provided $K^{(2r+2)/r}/n = o(1)$

(b) $\|\widehat{\phi} - \phi\| = O_p(n^{\frac{-rp}{2rp+(2+r)d}})$ when $K \asymp n^{\frac{rd}{2rp+(2+r)d}}$ and $p > d/2$.

The convergence rate obtained in Corollary 4.2 when $r = \infty$ (i.e. m is bounded) is $n^{-p/(2p+d)}$. This rate is the same as the optimal L^2 convergence rate for nonparametric regression estimators with i.i.d. data when the unknown regression function belongs to $\Lambda^p(\mathcal{X})$ and conditions (i) and (ii) of Corollary 4.2 hold (see, e.g., Stone (1982)).

4.4 Asymptotic normality

We now establish to asymptotic normality of $\widehat{\rho}$, $\widehat{y}$ and $\widehat{L}$ in Case 1. The limit theory is derived via a novel sieve perturbation expansion because the usual derivation for extremum estimators cannot be applied. For the sake of brevity, we focus on Case 1 (observable SDF) with $\widehat{\mathbf{G}}$ and $\widehat{\mathbf{M}}$ as in (21) and (22). In Cases 2 and 3 the limiting distribution will depend on how the unobservable components or unknown parameters in m are estimated in the first stage, as is typical of two-step plug-in estimators. Appendix C presents a general expansion for $\widehat{\rho}$, from which the asymptotic distribution of the estimators may be derived on a case-by-case basis.

For Case 1, we derive the representation

$$\sqrt{n}(\widehat{\rho} - \rho) = \frac{1}{\sqrt{n}} \sum_{t=0}^{n-1} \psi_\rho(X_t, X_{t+1}) + o_p(1) \tag{27}$$

where $\psi_\rho(x, y) := \phi^*(x)m(x, y)\phi(y) - \rho\phi^*(x)\phi(x)$. Expression (27) shows that $\widehat{\rho}$ behaves asymptotically like a sample average even though $\widehat{\rho}$ is a highly nonlinear function of $\widehat{\mathbf{G}}$ and $\widehat{\mathbf{M}}$. Conveniently, $\{\psi_\rho(X_t, X_{t+1}), \mathcal{F}_t\}$ is a martingale difference sequence. Thus, the asymptotic distribution of $\widehat{\rho}$ follows from (27) by a martingale central limit theorem. Rather than assuming a specific type of weak dependence in order to derive (27), we instead impose a high-level assumption regarding the rate at which the estimation and approximation errors vanish. This assumption can be verified using the results in Appendix C for different weak dependence conditions.

Let $\mathbf{G}^{-1/2}$ denote the inverse of the positive definite square root of $\mathbf{G}$ and let $\widetilde{b}^K(x) = \mathbf{G}^{-1/2} b^K(x)$ denote the orthogonalized vector of basis functions. Coherently with (21) and (22), define the orthogonalized estimators

$$\begin{aligned}\widehat{\mathbf{G}}^o &= \frac{1}{n}\sum_{t=0}^{n-1}\widetilde{b}^K(X_t)\widetilde{b}^K(X_t)'\\ \widehat{\mathbf{M}}^o &= \frac{1}{n}\sum_{t=0}^{n-1}\widetilde{b}^K(X_t)m(X_t,X_{t+1})\widetilde{b}^K(X_{t+1})'\,.\end{aligned}$$

Note that $\mathbb{E}[\widehat{\mathbf{G}}^o] = I$ (the $K \times K$ identity matrix) and define $\mathbb{M}^o = \mathbb{E}[\widehat{\mathbf{M}}^o]$. Let $\bar{\eta}_{n,K,1}$ and $\bar{\eta}_{n,K,2}$ be such that $\|\widehat{\mathbf{G}}^o - I\| = O_p(\bar{\eta}_{n,K,1})$ and $\|\widehat{\mathbf{M}}^o - \mathbf{M}^o\| = O_p(\bar{\eta}_{n,K,2})$ where, without confusion, we let $\|\cdot\|$ denote the matrix spectral norm (largest singular value). Define

$$\tau_{n,K} = \frac{1}{n}\sum_{t=0}^{n-1}\Big\{\phi_K^+(X_t)m(X_t,X_{t+1})\phi_K(X_{t+1}) - \rho_K\phi_K^+(X_t)\phi_K(X_t) - \psi_\rho(X_t,X_{t+1})\Big\}$$

where ϕ_K and ϕ_K^+ are normalized so that $\mathbb{E}[\phi_K(X_0)^2] = 1$ and $\mathbb{E}[\phi_K(X_0)\phi_K^+(X_0)] = 1$. Let $V_\rho = \mathbb{E}[\psi_\rho(X_0,X_1)^2]$ and $V_y = \rho^{-2}V_\rho$, which will be the asymptotic variances of $\widehat{\rho}$ and $\widehat{y}$, respectively. Also define $\psi_L(x,y) = \rho^{-1}\psi_\rho(x,y) - \log m(x,y) + \mathbb{E}[\log m(X_0,X_1)]$ and let V_L denote the long-run variance of $\{\psi_L(X_t,X_{t+1})\}$. Finally, let $a \vee b := \max\{a,b\}$.

Assumption 4.4 *(a)* $\delta_K = o(n^{-1/2})$*;*

(b) $(\eta_{n,K} \vee \bar{\eta}_{n,K,1}) \times (\bar{\eta}_{n,K,1} \vee \bar{\eta}_{n,K,2}) = o(n^{-1/2})$*;*

(c) $\tau_{n,K} = o_p(n^{-1/2})$*;*

(d) V_ρ *is positive and finite; and*

(e) $n^{-1/2}\sum_{t=0}^{n-1}\psi_L(X_t,X_{t+1}) \to_d N(0,V_L)$ *where* V_L *is positive and finite.*

Assumption 4.4(a) is an undersmoothing condition which ensures that the approximation bias $\rho - \rho_K$ vanishes sufficiently quickly that it does not distort the limiting distribution. Assumption 4.4(b)(c) ensures the higher-order terms in (27) are $o_p(1)$; sufficient conditions for Assumption 4.4(b) under different weak dependence assumptions are presented in Appendix C. Note that the summands in $\tau_{n,K}$ have expectation zero, and that ϕ_K, ϕ_K^+ and ρ_K are converging to ϕ, ϕ^*, and ρ by Lemma C.1. Assumption 4.4(d) ensures the estimators have a non-degenerate limiting distribution and finite asymptotic variance.[17] Parts (a)–(d) of Assumption 4.4, together with the earlier assumptions, are sufficient to derive the asymptotic distribution of $\widehat{\rho}$ and $\widehat{y}$. An analogous expansion to (27) holds for $\widehat{L}$ with $\psi_L(X_t,X_{t+1})$ in place of $\psi_\rho(X_t,X_{t+1})$. However, $\{\psi_L(X_t,X_{t+1}),\mathcal{F}_t\}$ is not necessarily a martingale difference sequence. Therefore we make the high-level Assumption 4.4(e) in order to derive the asymptotic distribution of $\widehat{L}$.

[17] This is made to rule out certain pathological cases. For instance, if $m(x,y) = \bar{m}$ for all x,y then then $\rho = \bar{m}$, $\phi = 1$, and $\phi^* = 1$ which yields $\psi_\rho(x,y) = 0$ for all x,y.

Theorem 4.2 *Let Assumptions 4.1, 4.2, 4.3, and 4.4 hold. Then (27) holds and:*

(a) $\sqrt{n}(\widehat{\rho} - \rho) \to_d N(0, V_\rho)$

(b) $\sqrt{n}(\widehat{y} - y) \to_d N(0, V_y)$

(c) $\sqrt{n}(\widehat{L} - L) \to_d N(0, V_L)$.

Variance estimators of V_ρ, V_y and V_L are presented in Appendix C, together with the asymptotic distributions of t-statistics for $\widehat{\rho}$, $\widehat{y}$ and $\widehat{L}$.

We conclude this section by deriving the semiparametric efficiency bounds for Case 1. To derive the efficiency bound we require a further technical condition (Assumption D.1), which is deferred to the Appendix.

Theorem 4.3 *Let Assumptions 4.1, 4.2, 4.3, 4.4, and D.1 hold. Then:*

(a) the semiparametric efficiency bounds for ρ, y and L are V_ρ, V_y and V_L, respectively

(b) $\widehat{\rho}$, $\widehat{y}$ and $\widehat{L}$ are semiparametrically efficient.

Theorem 4.3 provides further theoretical justification for using sieve methods to nonparametrically estimate ρ, ϕ, and related quantities.

5 Simulation and empirical application

The empirical performance of the proposed estimators is explored first in a simulation and then in an empirical application. In both illustrations we assume the SDF is determined by a representative agent and consider two specifications of the agent's preferences over future consumption, namely time-separable CRRA preferences and a recursive preference specification following Kreps and Porteus (1978), Epstein and Zin (1989) and Weil (1990). To implement the estimators with recursive preferences, we first introduce a new approach for nonparametrically estimating the continuation value (CV) of future consumption from a time series of data on $\{X_t\}$. By estimating the CV directly, rather than using a proxy for the return on the aggregate wealth portfolio, we avoid any potential issues related to imperfect proxies which may arise if, for example, human capital and other intangible/non-tradable assets are significant components of aggregate wealth. The simulation results below show that the estimated CV can be plugged in to the SDF in order to estimate ρ, ϕ, and related quantities with a high degree of accuracy.

5.1 Nonparametric continuation value estimation under recursive preferences

In this subsection we briefly describe an approach to nonparametrically estimate the continuation value in models with Epstein-Zin recursive preferences when we do not place parametric restrictions on the law of motion of $\{X_t\}$. This procedure may also be used to numerically solve for the continuation value and SDF in models for which analytical solutions are not available, as an alternative to discretization-based methods (e.g. Tauchen and Hussey (1991)).

Under Epstein-Zin preferences, the quantity V_t, which denotes the date-t utility of the representative agent, is defined in terms of V_{t+1} and current consumption C_t via the recursion

$$V_t = \left\{(1-\beta)C_t^{1-\theta} + \beta \mathbb{E}[V_{t+1}^{1-\gamma}|\mathcal{F}_t]^{\frac{1-\theta}{1-\gamma}}\right\}^{\frac{1}{1-\theta}}$$

where $1/\theta$ is the elasticity of intertemporal substitution (EIS), β is the time discount parameter, and γ is the risk aversion parameter. Assume hereafter that consumption growth $G_{t+1} := C_{t+1}/C_t$ is a measurable function of X_{t+1} where $\{X_t\}$ is a strictly stationary first-order Markov process. Hansen and Scheinkman (2012) show that, in this environment, the scaled continuation value V_t/C_t is of the form $V_t/C_t =: V(X_t)$ where V solves the fixed point equation

$$V(X_t) = \left\{(1-\beta) + \beta \mathbb{E}\left[\left(V(X_{t+1})\frac{C_{t+1}}{C_t}\right)^{1-\gamma}\middle| X_t\right]^{\frac{1-\theta}{1-\gamma}}\right\}^{\frac{1}{1-\theta}} \tag{28}$$

with $V \geq 0$. We now show how (28) may be used to estimate the continuation value function nonparametrically for given (θ, β, γ) from a time series of data on $\{X_t\}$.

In the remainder of this section we follow Tallarini (2000) and Hansen et al. (2008) focus on the case $\theta = 1$ so that (28) may be solved analytically for V. This allows us to evaluate the performance of the estimators in the simulation exercise.[18] Nevertheless, the following sieve methods may certainly be used to estimate the continuation value when $\theta \neq 1$ by appropriately modifying the conditional moment restriction (29) below.

When $\theta = 1$, the fixed point equation (28) becomes

$$v(X_t) = \frac{\beta}{1-\gamma} \log \mathbb{E}[e^{(1-\gamma)(v(X_{t+1})+g_{t+1})}|X_t]$$

where $v(x) = \log V(x)$ and $g_{t+1} = \log(C_{t+1}/C_t)$ (Hansen et al., 2008, Section III). This expression

[18] When $\theta \neq 1$ the CV is not known analytically so it would be difficult to evaluate the accuracy of our estimators in simulations. One possible approach for doing so when $\theta \neq 1$ might be to consider a log-quadratic approximation to the CV and SDF obtained under small perturbations of θ from 1, as in Hansen et al. (2008) and Backus et al. (2014).

for v may be rearranged to obtain the conditional moment restriction

$$\mathbb{E}\left[e^{(1-\gamma)(v(X_{t+1})+g_{t+1})-\frac{1-\gamma}{\beta}v(X_t)}-1\middle| X_t\right]=0 \tag{29}$$

upon which our estimator of v is based.

We now describe a sieve minimum distance (SMD) approach to nonparametrically estimate v given β, γ, and a time series of data on $\{X_t\}$ (see Ai and Chen (2003) and Chen and Pouzo (2012) for background material on SMD estimation with i.i.d. data). Chen, Favilukis, and Ludvigson (2013) have recently applied SMD methods to estimate models with recursive preferences.[19] In the SMD procedure, the function v is first approximated by a linear combination $\sum_{k=1}^{K_1} c_{1k} p_{K_1 k}$ of K_1 basis functions $p_{K_1 1}, \ldots, p_{K_1 K_1}$. The moment restriction (29) is used to form a criterion function which is minimized with respect to the coefficients $c_{11}, \ldots, c_{1K_1}$. To form the criterion function, the conditional expectation in (29) is estimated by series regression on a second basis $\psi_{K_2 1}, \ldots, \psi_{K_2 K_2}$ of dimension K_2 with $K_1 \leq K_2$. Let $p^{K_1}(x) = (p_{K_1 1}(x), \ldots, p_{K_1 K_1}(x))'$, let $\psi^{K_2}(x) = (\psi_{K_2 1}(x), \ldots, \psi_{K_2 K_2}(x))'$, and let $\Psi = (\psi^{K_2}(X_0), \ldots, \psi^{K_2}(X_{n-1}))'$. The estimator of the conditional moment restriction (29) evaluated at $X_t = x$ and $v(x) = c_1' p^{K_1}(x)$ is

$$\widehat{u}(x, c_1) = \psi^{K_2}(x)'(\Psi'\Psi/n)^{-1}\left(\frac{1}{n}\sum_{t=0}^{n-1}\psi^{K_2}(X_t)\left(e^{(1-\gamma)(c_1'p^{K_1}(X_{t+1})+g_{t+1})-\frac{1-\gamma}{\beta}c_1'p^{K_1}(X_t)}-1\right)\right).$$

Our estimator of v is $\widehat{v}(x) = \widehat{c}_1' p^{K_1}(x)$ where

$$\widehat{c}_1 = \arg\min_{c_1 \in \mathbb{R}^{K_1}} \frac{1}{n}\sum_{t=0}^{n-1}\widehat{u}(X_t, c_1)^2 .$$

Approximating v and the conditional moment over the finite dimensional subspaces spanned by $p_{K_1 1}, \ldots, p_{K_1 K_1}$ and $\psi_{K_2 1}, \ldots, \psi_{K_2 K_2}$ introduces approximation bias. Increasing K_1 and K_2 will typically reduce the bias but will introduce additional sampling error as there will be more parameters to be estimated. Therefore, nonparametric estimation of c will be subject to a similar bias-variance tradeoff to that which is encountered in nonparametric estimation of ρ and ϕ. The theoretical properties $\widehat{v}$ could be derived by a time-series extension of Ai and Chen (2003) or Chen and Pouzo (2012). However, the literature on SMD estimation has almost exclusively focused on i.i.d. data to date so defer such an endeavor to future research.

The SDF obtained under Epstein-Zin preferences with $\theta = 1$ is of the form $m(X_t, X_{t+1}) = \beta \exp\{-\gamma g_{t+1} + (1-\gamma)v(X_{t+1}) - \frac{1-\gamma}{\beta}v(X_t)\}$. Our estimator $\widehat{v}$ may be plugged in to this func-

[19] There are several important differences between the method introduced here and that in Chen et al. (2013). There, they focus on the general case in which $\theta \neq 1$ and use time series data consumption and returns to estimate $(V, \beta, \gamma, \theta)$ nonparametrically using the conditional Euler equation. They also assume consumption growth is a function of a latent univariate Markov state variable. Here we estimate v from data on $\{X_t\}$ for fixed, possibly counterfactual, values of (β, γ, θ) using the fixed-point equation (28) instead of the Euler equation. We require an observable state vector but allow it to be of an arbitrary dimension.

tional form to obtain

$$\widehat{m}(X_t, X_{t+1}) = \beta \exp\Big\{ -\gamma g_{t+1} + (1-\gamma)\widehat{v}(X_{t+1}) - \frac{1-\gamma}{\beta}\widehat{v}(X_t)\Big\}. \tag{30}$$

The matrix $\mathbf{M}$ can then be estimated as described in (23).

5.2 Simulation

The following Monte Carlo (MC) experiment investigates the performance of our estimators when applied to a consumption-based asset pricing model under CRRA and Epstein-Zin preferences. The state variable is simply taken as $X_t = g_t$, where g_t denotes log consumption growth, which is assumed to evolve as a Gaussian AR(1) process:

$$g_{t+1} - \mu = \kappa(g_t - \mu) + \sigma e_{t+1}$$

where the e_t are i.i.d. $N(0,1)$ random variables. The parameters for the simulation are $\mu = 0.008$, $\kappa = 0.6$, and $\sigma = 0.01$. The data are constructed to be somewhat representative of quarterly U.S. real per capita growth in consumption of nondurables and services (for which $\kappa \approx 0.3$ and $\sigma \approx 0.005$) but we make $\{g_t\}$ are twice as persistent ($\kappa = 0.60$) to produce greater nonlinearity in the eigenfunctions and twice as volatile ($\sigma = 0.01$) to produce a more challenging estimation problem. The parameters in the utility function are set to $\beta = 0.994$ and $\gamma = 10$. For each design we generate 1000 samples of length 400, 800, 1600, and 3200: the smallest sample size is roughly the sample size with aggregate monthly or quarterly consumption data, whereas the larger sizes are used to illustrate the convergence properties of the estimators.

To implement the estimators $\widehat{\rho}$, $\widehat{\phi}$, and $\widehat{\phi}^*$, we use $\widehat{\mathbf{G}}$ in (21) for both preference specifications and use $\widehat{\mathbf{M}}$ in (22) for the CRRA design and $\widehat{\mathbf{M}}$ in (23) for the Epstein-Zin design with $\widehat{m}$ from (30). We also compute $\widehat{L}$ using the estimators in (25) and (26) for CRRA and Epstein-Zin preferences, respectively. A basis of dimension $K = 8$ is used for b^K to approximate ϕ under both utility specifications; for the recursive preference specification a sieve of dimension $K_1 = 6$ is used for p^{K_1} to approximate v and a sieve of dimension $K_2 = 12$ for ψ^{K_2} to estimate the conditional moment in the SMD procedure. The simulations are performed for using Hermite polynomial sieves for b^K, p^{K_1}, and ψ^{K_2}, and again with B-Spline sieves for b^K, p^{K_1}, and ψ^{K_2}. As is standard practice, the Hermite bases were centered and scaled by the sample mean and sample standard deviation of g, and the knots of the cubic B-spline sieve were placed at the empirical quantiles of the data. The MC results were reasonably insensitive both to the choice of sieve and to the dimension of the sieve space. Only the results for the Hermite polynomial sieve are presented below; the results for B-spline sieves and the parameterization $\kappa = 0.3$, $\sigma = 0.005$ are presented in Appendix E.

For each simulation configuration we estimate ϕ, ϕ^*, ρ, y, L; we also estimate v for the Epstein-Zin design. To calculate the root mean square error (RMSE) for $\widehat{\phi}$, $\widehat{\phi}^*$, and $\widehat{v}$, for each replication we

	n	CRRA $\widehat{\phi}$	CRRA $\widehat{\phi}^*$	Epstein-Zin $\widehat{\phi}$	Epstein-Zin $\widehat{\phi}^*$	Epstein-Zin $\widehat{v}$
Bias	400	0.0097	0.0110	0.0196	0.0317	0.1101
	800	0.0035	0.0045	0.0098	0.0154	0.0443
	1600	0.0034	0.0042	0.0056	0.0081	0.0264
	3200	0.0009	0.0011	0.0024	0.0045	0.0068
RMSE	400	0.0675	0.0799	0.0415	0.1325	0.2825
	800	0.0397	0.0492	0.0193	0.0879	0.1893
	1600	0.0283	0.0342	0.0113	0.0645	0.1357
	3200	0.0176	0.0212	0.0059	0.0416	0.0928

Table 1: Bias and RMSE of $\widehat{\phi}$ and $\widehat{\phi}^*$ under both preference specifications, and bias and RMSE of $\widehat{v}$ under Epstein-Zin preferences. Results are obtained from 1000 replications of the MC design using the sample size shown and Hermite polynomial bases for b^K, p^{K_1} and ψ^{K_2} with $K = 8$, $K_1 = 6$ and $K_2 = 12$.

	n	CRRA $\widehat{\rho}$	CRRA $\widehat{y}$	CRRA $\widehat{L}$	Epstein-Zin $\widehat{\rho}$	Epstein-Zin $\widehat{y}$	Epstein-Zin $\widehat{L}$
Bias	400	0.0012	-0.0007	0.0011	0.0030	-0.0029	0.0007
	800	0.0006	-0.0005	0.0004	0.0015	-0.0015	0.0004
	1600	0.0004	-0.0003	0.0005	0.0010	-0.0010	0.0005
	3200	0.0003	-0.0003	0.0001	0.0004	-0.0004	0.0001
RMSE	400	0.0385	0.0288	0.0251	0.0151	0.0131	0.0141
	800	0.0103	0.0105	0.0061	0.0040	0.0039	0.0059
	1600	0.0086	0.0086	0.0058	0.0049	0.0047	0.0059
	3200	0.0050	0.0051	0.0025	0.0009	0.0009	0.0025

Table 2: Bias and RMSE of $\widehat{\rho}$, $\widehat{y}$ and $\widehat{L}$ under both preference specifications. Results are obtained from 1000 replications of the MC design using the sample size shown and Hermite polynomial bases for b^K, p^{K_1} and ψ^{K_2} with $K = 8$, $K_1 = 6$ and $K_2 = 12$.

calculate the L^2 distance between the estimators and their population counterparts, then take the average over the MC replications. To calculate the bias we take the average of the estimators across the MC replications, then compute the L^2 distance between the average of the estimates across MC replications and their population counterparts.[20] Similar calculations are performed for $\widehat{\rho}$, $\widehat{y}$, and $\widehat{L}$.

Results of the MC exercise are presented in Tables 1 and 2. Table 1 shows that ϕ and ϕ^* may be estimated with small bias using a reasonably low-dimensional sieve, and that the sampling error vanishes as the sample size increases. It is slightly surprising that the RMSEs for $\widehat{\phi}$ under recursive preferences are about one half of the RMSEs for $\widehat{\phi}$ under CRRA preferences, even though with recursive preferences the continuation value must be first estimated nonparametrically. In contrast,

[20]The use of the "bias" here is not to be confused with the bias term in the convergence rate calculations. There "bias" measures how close ϕ_K and ρ_K are to ϕ and ρ. Here "bias" of an estimator refers to the distance between the parameter and the average of its estimates across the MC replications.

the bias for $\widehat{\phi}$ and $\widehat{\phi}^*$ and the RMSE for $\widehat{\phi}^*$ is larger under recursive preferences than CRRA preferences. The results in Table 1 show that $\widehat{v}$ is more difficult to estimate than $\widehat{\phi}$ and $\widehat{\phi}^*$, but may be estimated with a reasonably small degree of bias in moderate samples. Appendix E presents additional MC results under the parameterization $\kappa = 0.3$ and $\sigma = 0.005$ which yields bias and RMSEs about one third of the values presented in Table 1.

Table 2 presents similar results for $\widehat{\rho}$, $\widehat{y}$ and $\widehat{L}$. The bias and RMSEs of the estimators are all reasonably small, and are decreasing in the sample size n. As with $\widehat{\phi}$ and $\widehat{\phi}^*$, the RMSEs of $\widehat{\rho}$, $\widehat{y}$ and $\widehat{L}$ under recursive preferences are smaller than under CRRA preferences even though the continuation value is first estimated nonparametrically.

5.3 Empirical application

We now apply the methods developed in this paper to investigate the time-series properties and asset pricing implications of the permanent and transitory components of the SDF under Epstein-Zin preferences (with unit elasticity of intertemporal substitution) and CRRA preferences. As is well known, the permanent component of the SDF under many external and internal habit formation specifications is the same as the permanent component under CRRA preferences (Hansen, 2012; Backus et al., 2014). Our analysis of the permanent component obtained under CRRA preferences therefore extends to a much broader class of preferences. Two specifications of the state process are used, namely $X_t = g_t$ and $X_t = (g_t, g_{e,t})$ where g_t denotes the logarithm of real per capita consumption growth and $g_{e,t}$ denotes the logarithm of real per capita corporate earnings growth.

Data on consumption, corporate earnings, and population were sourced from the National Income and Product Accounts (NIPA) tables and span the period 1947:Q1 to 2012:Q4 (263 observations). The consumption and earnings growth series are formed by taking seasonally adjusted consumption of nondurables and services data (NIPA Table 2.3.5) and after tax corporate earnings (NIPA Table 1.12), deflating by the implicit price deflator for personal consumption expenditures (PCE; NIPA Table 2.3.4), and then calculating per capita growth rates using the deflated series and population data (NIPA Table 2.1). For data on the risk-free rate and market return, we take the 90-day T-bill rate and value-weighted return on the combined NYSE/AMEX/NASDAQ index including dividends (both from CRSP) and convert these series to real rates using the PCE deflator data. We proxy the holding period return on a bond of infinite maturity by the quarterly return on the 30 year U.S. Treasury index (from CRSP) which we deflate using the PCE data. Finally, for GDP data we use quarterly real seasonally adjusted data from the Federal Reserve.

The estimators $\widehat{\rho}$, $\widehat{\phi}$, $\widehat{\phi}^*$, $\widehat{v}$, $\widehat{y}$, and $\widehat{L}$ are implemented as described in the simulation exercise. We use Hermite polynomial sieves of dimension $K = 8$, $K_1 = 6$, and $K_2 = 12$ for b^K, p^{K_1} and ψ^{K_2} in the univariate case and a tensor-product sieves of dimension $K = 16$, $K_1 = 16$, and $K_2 = 25$ for b^K, p^{K_1} and ψ^{K_2} in the bivariate case. The following results were reasonably insensitive to the choice of sieve dimension and basis.

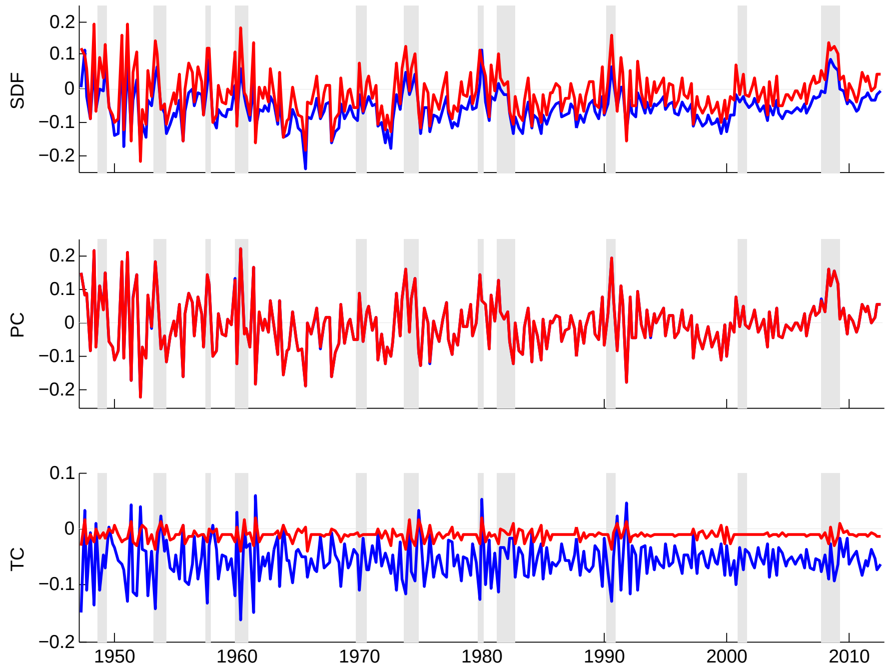

Figure 1: Time series of estimates of the logarithm of the SDF, its permanent component (PC), and its transitory component (TC). Blue lines are the estimates obtained under CRRA preferences and red lines are estimates obtained under Epstein-Zin preferences with EIS $= 1$. Shaded bars denote NBER recession indicators. The state variable is $X_t = g_t$ and SDF parameters are $\beta = 0.994$ and $\gamma = 10$.

For both preference and state process specifications, we construct time series of the SDF and its permanent and transitory components by substituting the estimated $\widehat{\rho}$ and $\widehat{\phi}$ (and $\widehat{m}$ with Epstein-Zin preferences) into the formulae for M^P_{t+1}/M^P_t and M^T_{t+1}/M^T_t in (5) and (6). Time series of the logarithm of M_{t+1}/M_t, M^P_{t+1}/M^P_t and M^T_{t+1}/M^T_t have been plotted in Figure 1 (for the case $X_t = g_t$) and Figure 2 (for the case $X_t = (g_t, g_{e,t})'$).Both figures show that the permanent components obtained under the two preference specifications are almost indistinguishable. The Epstein-Zin SDF with EIS $= 1$ and $\beta \approx 1$ is of approximately the same form as (12), because

$$\frac{M_{t+1}}{M_t} = \beta G_{t+1}^{-\gamma} \frac{V(X_{t+1})^{1-\gamma}}{V(X_t)^{(1-\gamma)/\beta}} \approx \beta G_{t+1}^{-\gamma} \frac{V(X_{t+1})^{1-\gamma}}{V(X_t)^{1-\gamma}}$$

when $\beta \approx 1$. Reasoning as in Hansen (2012) and Backus et al. (2014) would then suggest that the two SDFs should have similar permanent components. Nevertheless, it is perhaps surprising just how indistinguishable the two permanent components are. In contrast, the time series of the SDF under Epstein-Zin preferences is typically "rougher" than under CRRA preferences, and vice versa for the trajectories of the transitory components.

Figures 1 and 2 also show that the both the SDF and its permanent component are countercyclical whereas the transitory component is acyclical. The correlation between the log permanent component series and log GDP growth is around -0.35 for both preference and state specifications.

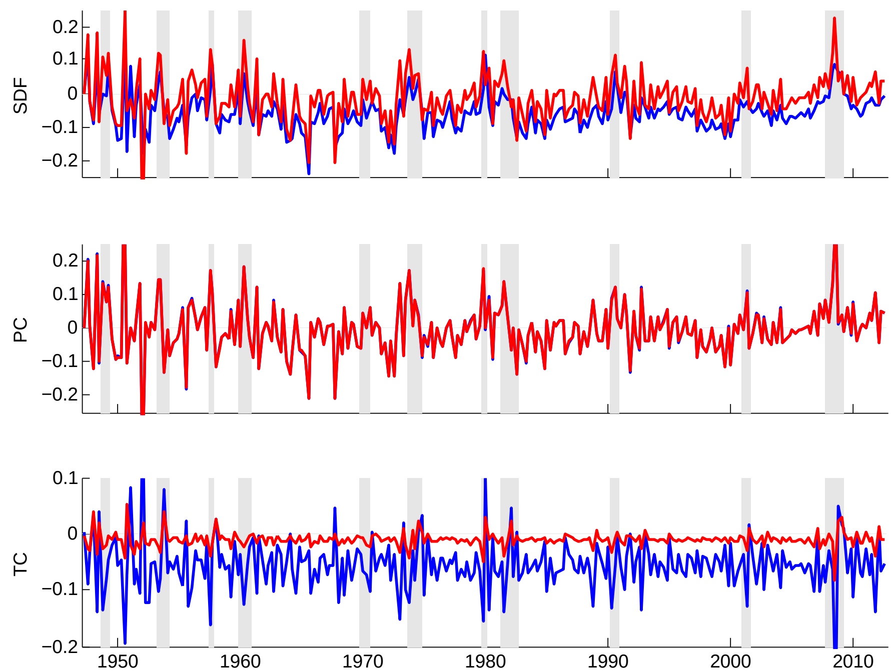

Figure 2: Time series of estimates of the logarithm of the SDF, its permanent component (PC), and its transitory component (TC). Blue lines are the estimates obtained under CRRA preferences and red lines are estimates obtained under Epstein-Zin preferences with EIS $= 1$. Shaded bars denote NBER recession indicators. The state variable is $X_t = (g_t, g_{e,t})'$ and SDF parameters are $\beta = 0.994$ and $\gamma = 10$.

In contrast, the correlation between the log transitory component series and log GDP growth is approximately 0.01 with $X_t = G_t$ and 0.06 when $X_t = (G_t, G_{e,t})'$. Further, the permanent and transitory components are strongly negatively correlated (around -0.75 with CRRA preferences and -0.6 with Epstein-Zin preferences).

We now turn to investigating whether the SDF and its permanent and transitory components are compatible with historical returns data. The bounds (11) and (10) show that the entropy of the SDF and the entropy of its permanent component must be at least as large as the return on assets relative to short- and long-term (zero-coupon) bonds, respectively. The quarterly premium on the combined market index relative to the 90-day T-bill rate and 30-year Treasury index were 1.83% and 1.54%, respectively, over the sample period. We take these historical premia as benchmark entropy bounds, though these bounds may be tightened further by including additional asset returns or returns on growth-optimal portfolios. Table 3 shows that none of the preference or state specifications can rationalize either benchmark premium despite the fact that $\gamma = 10$ might be regarded as reasonably large. Estimates of the entropy of the permanent component of the SDF are around 0.0030 for each preference and state specification, which is roughly one fifth of the level required to explain the premium of 1.54%. Table 3 also reports estimates of the entropy of the SDF, which we estimate by

X_t	$L(\frac{M_{t+1}}{M_t})$	$L(\frac{M^P_{t+1}}{M^P_t})$	$\mathrm{Var}(\frac{M^T_{t+1}}{M^T_t})$	$-\log\rho$
		CRRA		
g_t	0.0015 (0.0010,0.0020)	0.0032 (0.0015,0.0050)	0.0012 (0.0001,0.0023)	0.0550 (0.0469,0.0631)
$(g_t, g_{e,t})'$	0.0015 (0.0010,0.0020)	0.0026 (0.0009,0.0042)	0.0018 (0.0003,0.0034)	0.0557 (0.0475,0.0639)
		Epstein-Zin, EIS= 1		
g_t	0.0025 (0.0014,0.0036)	0.0030 (0.0015,0.0044)	0.0001 (−0.0004,0.0006)	0.0105 (0.0093,0.0118)
$(g_t, g_{e,t})'$	0.0024 (0.0013,0.0034)	0.0028 (0.0014,0.0043)	0.0002 (−0.0005,0.0008)	0.0107 (0.0092,0.0122)

Table 3: Nonparametric estimates of the entropy of the SDF, the entropy of the permanent component of the SDF, the variance of the transitory component of the SDF, and the long-run yield, under CRRA preferences and Epstein-Zin preferences (with EIS = 1 and continuation value estimated nonparametrically) using $\beta = 0.994$ and $\gamma = 10$. 90% confidence intervals (CIs) are reported in parentheses. CIs for $L(M_{t+1}/M_t)$, $L(M^P_{t+1}/M^P_t)$, and long-run yield with CRRA preferences are asymptotic CIs: for the long-run yield are computed as described in Appendix C, CIs for $L(M^P_{t+1}/M^P_t)$ are computed as described in Appendix C using an OSLRV estimator with a cosine basis of dimension 10; CIs for $L(M_{t+1}/M_t)$ are also formed using an OSLRV estimator with a cosine basis of dimension 10. Remaining CIs are computed using the bootstrap percentile method from 5000 replications of the stationary bootstrap with expected block size 6.

its sample analogue

$$\log\left(n^{-1}\sum_{t=0}^{n-1} m(X_t, X_{t+1})\right) - n^{-1}\sum_{t=0}^{n-1}\log m(X_t, X_{t+1})$$

under CRRA preferences; under Epstein-Zin preferences we replace m in the above display by $\widehat{m}$ from (30). Estimated entropies of the SDFs are even smaller, at around 0.0015 (CRRA) and 0.0025 (Epstein-Zin). So although the Epstein-Zin specification can generate a larger SDF than the CRRA specification, and therefore account for a larger (though still too small) premium relative to the short-term risk-free rate, the permanent components of the SDFs under the two specifications are of almost equal size. Both models are therefore unable to account for the historical return on equities relative to long-term bonds under the parameterization $\gamma = 10$ and $\beta = 0.994$. Repeating the exercise with $\gamma = 20$ yields estimates of $L(M^P_{t+1}/M^P_t)$ around 0.012 under both preference and state specifications, which is still somewhat short of the benchmark. Moreover, this shortcoming of the C-CAPM cannot be alleviated by adding internal or external habit formation in a way that results in transitory modifications to the pricing kernel.

The estimated long-run yield reported in Table 3 for Epstein-Zin preferences is around 1.05%

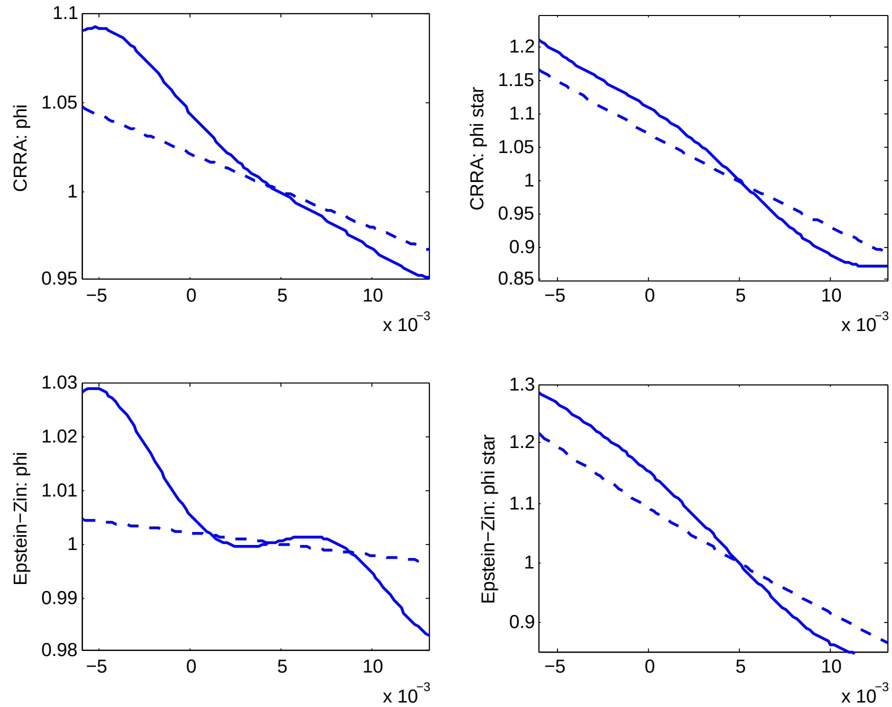

Figure 3: Nonparametric estimates of ϕ and ϕ^* (solid lines) obtained under CRRA preferences and Epstein-Zin preferences (with EIS = 1 and continuation value estimated nonparametrically) using $\beta = 0.994$ and $\gamma = 10$ and $X_t = G_t$. Dashed lines are parametric estimates obtained assuming $\log G_t$ is a Gaussian AR(1) process.

per quarter which compares favorably with historical long-term yields. For instance, the historical average real quarterly yield on the longest maturity (either 30 or 20 year) Treasury Constant Maturity index over the period April 1953 to December 2012 is 0.78%.[21] As expected, the estimated long-run yield with CRRA preferences, at around 5% per quarter, is much higher than historical long-term yields. Despite the relative success of the Epstein-Zin specification in matching the level of long-term yields, it cannot match the volatility observed in historical yield data. The bound

$$\mathrm{Var}\Big(\frac{M_{t+1}^T}{M_t^T}\Big) \geq \frac{(1 - \mathbb{E}[R_{t+1,\infty}]\mathbb{E}[R_{t+1,\infty}^{-1}])^2}{\mathrm{Var}(R_{t+1,\infty})} \tag{31}$$

is a consequence of the identity $M_t^T/M_{t+1}^T = R_{t+1,\infty}$.[22] Using the quarterly return on the 30 year U.S. Treasury index as a proxy for $R_{t+1,\infty}$, our estimate of the right-hand side of (31) is 0.0031. Table 3 presents the sample variance of the estimated transitory component, from which it is apparent that the volatility of the transitory component under Epstein-Zin preferences is roughly an order of magnitude too small (at least to the extent that our proxy for long-term yields is representative of the historical return on a bond of infinite maturity).

[21]Nominal Treasury Constant Maturity yields were taken from the Federal Reserve H-15 release and converted to real yields using the PCE deflator.

[22]The bound (31) obtains by substituting $M_t^T/M_{t+1}^T = R_{t+1,\infty}$ into the bound for $\mathrm{Var}(M_{t+1}^T/M_t^T)$ reported in Proposition 2 of Bakshi and Chabi-Yo (2012).

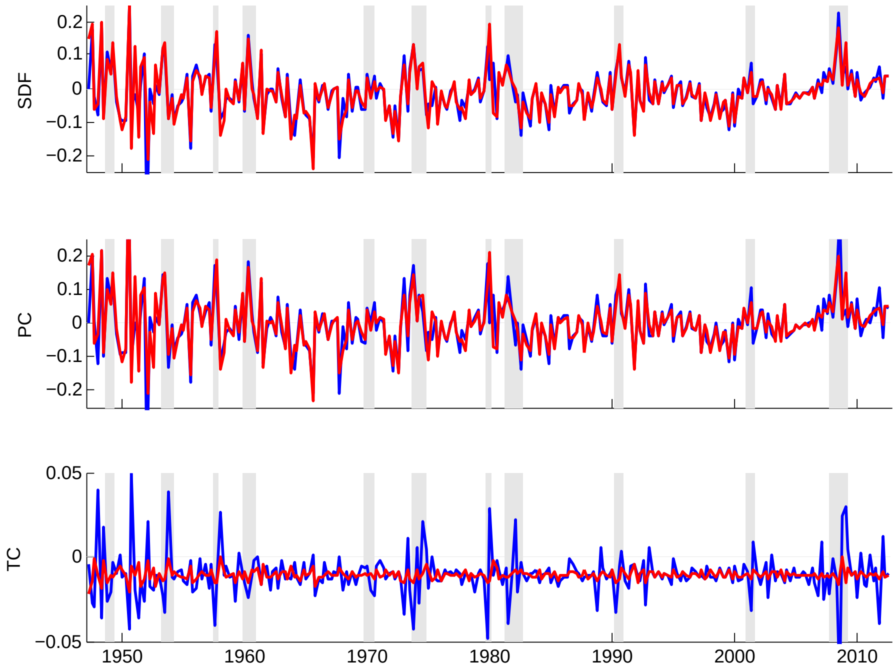

Figure 4: Time series of estimates of the logarithm of the SDF, its permanent component (PC), and its transitory component (TC) obtained under Epstein-Zin preferences with EIS = 1, $\beta = 0.994$, $\gamma = 10$ and $X_t = (g_t, g_{e,t})'$. Blue lines are SDF, PC, and TC extracted nonparametrically. Red lines are corresponding SDF, PC, TC with continuation value, ρ, and ϕ calculated assuming X_t is a Gaussian VAR(1) process. Shaded bars denote NBER recession indicators.

To examine how the nonparametric estimates compare with a parametric model for X_t, ρ, ϕ, ϕ^*, the Epstein-Zin continuation value, and the quantities $L(M_{t+1}/M_t)$, $L(M^P_{t+1}/M^P_t)$, $\text{Var}(M^T_{t+1}/M^T_t)$, and $-\log \rho$ were estimated assuming X_t is a Gaussian VAR(1) process. Figure 4 displays nonparametric estimates of ϕ and ϕ^* for both preference specifications together with parametric estimates assuming g_t is a Gaussian AR(1) process. To obtain these estimates, for each preference specification we calculate analytical formulae for the relevant quantities and then evaluate the formulae at the quasi maximum likelihood estimates (QMLEs) of the VAR(1) parameters. Figure 3 shows that the nonparametric estimates of ϕ and ϕ^* are steeper and more nonlinear than parametric estimates for both CRRA and Epstein-Zin preferences. The parametric estimates of the entropy of the permanent component (0.0027 with CRRA and EZ), the entropy of the SDF (0.0016 with CRRA and 0.0029 with EZ) and the long-run yield (0.055 with CRRA and 0.0107 with EZ) are similar to those obtained parametrically.

It is difficult to judge from Figure 4 just how the differences between the nonparametric and parametric estimates of ϕ translate to differences in the permanent and transitory component. Therefore, Figure 4 plots time series of the logarithm of the SDF and its permanent and transitory components obtained under the Epstein-Zin preference specification both nonparametrically and also assuming the state evolves as a Gaussian VAR(1) process (with the continuation value, ρ and ϕ calculated analytically and evaluated at the QMLEs). The overall trajectories of the SDF and PC are similar when obtained parametrically and nonparametrically, though the parametric trajectories appear somewhat less rough. The transitory components are quite different when extracted parametrically and nonparametrically, with the transitory component obtained parametrically appearing much too smooth. Similar results are obtained with $X_t = g_t$. With CRRA preferences the permanent components extracted parametrically and nonparametrically are similar, and the parametric transitory component is again more smooth than that which is obtained nonparametrically.

6 Conclusion

This paper introduces econometric methods for performing estimation and inference on the long-term valuation implications of dynamic asset pricing models. We introduce nonparametric sieve estimators of the positive eigenfunction and its eigenvalue, the long-run yield, and the entropy of the permanent component of the SDF. We establish consistency and convergence rates of the estimators allowing for a wide variety of empirically relevant setups, and establish asymptotic normality and efficiency of the estimators for the case in which the SDF is observed. To extend the ambit of our estimators to an important class of recursive preferences, we introduce new nonparametric estimators of the continuation value function in Markov environments. A simulation exercise shows that the principal eigenpair and related quantities can be estimated with a high degree of accuracy by plugging the nonparametric estimate of the continuation value function into the eigenvalue/eigenfunction estimators. When applied to aggregate U.S. consumption and corporate earnings data, our estimators reveal that the permanent components of the SDF obtained under Epstein-Zin preferences with unit EIS and under CRRA preferences are remarkably similar. Neither preference specification is able to account for historical returns on equities relative to long-term bounds under reasonable parameterizations. We also present identification conditions and a long-run pricing approximation for stationary, discrete-time environments which complements the analysis of Hansen and Scheinkman (2009) for general, continuous-time environments.

The present paper may be extended along several dimensions. First, one natural extension is to models with latent state variables. Second, the identification conditions and estimators may be applied to study identification and estimation of other semi/nonparametric models. Third, results of the simulation and empirical application under Epstein-Zin preferences were obtained assuming unit EIS. Further work is required to investigate the performance of the estimators and empirical findings under alternate elasticities of intertemporal substitution and state process specifications. Finally, the estimators may also be applied to study valuation with a baseline stochastic growth component.

References

Abel, A. B. (1990): "Asset Prices under Habit Formation and Catching up with the Joneses," *American Economic Review*, 80, 38–42.

Ai, C. and X. Chen (2003): "Efficient Estimation of Models with Conditional Moment Restrictions Containing Unknown Functions," *Econometrica*, 71, 1795–1843.

Aït-Sahalia, Y. and A. W. Lo (1998): "Nonparametric Estimation of State-Price Densities Implicit in Financial Asset Prices," *The Journal of Finance*, 53, 499–547.

Alvarez, F. and U. J. Jermann (2005): "Using Asset Prices to Measure the Persistence of the Marginal Utility of Wealth," *Econometrica*, 73, 1977–2016.

Backus, D., M. Chernov, and S. Zin (2014): "Sources of Entropy in Representative Agent Models," *Journal of Finance*, 69, 51–99.

Bakshi, G. and F. Chabi-Yo (2012): "Variance Bounds on the Permanent and Transitory Components of Stochastic Discount Factors," *Journal of Financial Economics*, 105, 191–208.

——— (2014): "New Entropy Restrictions and the Quest for Better Specified Asset Pricing Models," Fisher College of Business WP 2014-03-07, Ohio State University.

Bansal, R. and B. N. Lehmann (1997): "Growth-Optimal Portfolio Restrictions on Asset Pricing Models," *Macroeconomic Dynamics*, 1, 333–354.

Carrasco, M., J.-P. Florens, and E. Renault (2007): "Linear Inverse Problems in Structural Econometrics Estimation Based on Spectral Decomposition and Regularization," in *Handbook of Econometrics*, ed. by J. J. Heckman and E. E. Leamer, Elsevier, vol. 6, Part B, chap. 77, 5633–5751.

Chen, X. (2007): "Large Sample Sieve Estimation of Semi-Nonparametric Models," in *Handbook of Econometrics*, ed. by J. J. Heckman and E. E. Leamer, Elsevier, vol. 6, Part B, chap. 76, 5549–5632.

Chen, X., V. Chernozhukov, S. Lee, and W. K. Newey (2014): "Local Identification of Nonparametric and Semiparametric Models," *Econometrica*, 82, 785–809.

Chen, X., J. Favilukis, and S. C. Ludvigson (2013): "An Estimation of Economic Models with Recursive Preferences," *Quantitative Economics*, 4, 39–83.

Chen, X., L. P. Hansen, and J. A. Scheinkman (2000): "Shape-Preserving Estimation of Diffusions," Working paper, University of Chicago.

——— (2009): "Nonlinear Principal Components and Long-Run Implications of Multivariate Diffusions," *Annals of Statistics*, 37, 4279–4312.

Chen, X. and S. C. Ludvigson (2009): "Land of Addicts? An Empirical Investigation of Habit-Based Asset Pricing Models," *Journal of Applied Econometrics*, 24, 1057–1093.

Chen, X. and D. Pouzo (2012): "Estimation of Nonparametric Conditional Moment Models With Possibly Nonsmooth Generalized Residuals," *Econometrica*, 80, 277–321.

Christensen, T. M. (2014): "Nonparametric Identification of Positive Eigenfunctions," *Econometric Theory*, forthcoming.

Connor, G., M. Hagmann, and O. Linton (2012): "Efficient Semiparametric Estimation of the Fama-French Model and Extensions," *Econometrica*, 80, 713–754.

Darolles, S., Y. Fan, J.-P. Florens, and E. Renault (2011): "Nonparametric Instrumental Regression," *Econometrica*, 79, 1541–1565.

Darolles, S., J.-P. Florens, and C. Gourieroux (2004): "Kernel-Based Nonlinear Canonical Analysis and Time Reversibility," *Journal of Econometrics*, 119, 323–353.

Darolles, S., J.-P. Florens, and E. Renault (1998): "Nonlinear principal components and inference on a conditional expectation operator with applications to Markov processes," Working paper, Paris-Berlin Confernece, Garchy.

Epstein, L. G. and S. E. Zin (1989): "Substitution, Risk Aversion, and the Temporal Behavior of Consumption and Asset Returns: A Theoretical Framework," *Econometrica*, 57, 937–969.

Escanciano, J. C. and S. Hoderlein (2012): "Nonparametric Identification of Euler Equations," Working paper, Indiana University.

Florens, J.-P., E. Renault, and N. Touzi (1998): "Testing for Embeddability by Stationary Reversible Continuous-Time Markov Processes," *Econometric Theory*, 14, 744–769.

Gagliardini, P., C. Gourieroux, and E. Renault (2011): "Efficient Derivative Pricing by the Extended Method of Moments," *Econometrica*, 79, 1181–1232.

Galí, J. (1994): "Keeping up with the Joneses: Consumption Externalities, Portfolio Choice, and Asset Prices," *Journal of Money, Credit and Banking*, 26, 1–8.

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. (1982): "Large sample properties of generalized method of moments estimators," *Econometrica*, 50, 1029–1054.

——— (2012): "Dynamic Valuation Decomposition Within Stochastic Economies," *Econometrica*, 80, 911–967.

Hansen, L. P., J. C. Heaton, and N. Li (2008): "Consumption Strikes Back? Measuring Long-Run Risk," *Journal of Political Economy*, 116, 260–302.

Hansen, L. P. and R. Jagannathan (1991): "Implications of Security Market Data for Models of Dynamic Economies," *Journal of Political Economy*, 99, 225–262.

Hansen, L. P. and J. A. Scheinkman (1995): "Back to the Future: Generating Moment Implications for Continuous-Time Markov Processes," *Econometrica*, 63, 767–804.

——— (2009): "Long-Term Risk: An Operator Approach," *Econometrica*, 77, 177–234.

——— (2012): "Recursive Utility in a Markov Environment with Stochastic Growth," *Proceedings of the National Academy of Sciences*, 109, 11967–11972.

——— (2013): "Stochastic Compounding and Uncertain Valuation," Working paper, University of Chicago.

Kreps, D. M. and E. L. Porteus (1978): "Temporal Resolution of Uncertainty and Dynamic Choice Theory," *Econometrica*, 46, 185–200.

Lettau, M. and J. A. Wachter (2007): "Why Is Long-Horizon Equity Less Risky? A Duration-Based Explanation of the Value Premium," *The Journal of Finance*, 62, 55–92.

——— (2011): "The Term Structures of Equity and Interest Rates," *Journal of Financial Economics*, 101, 90–113.

Lewbel, A., O. B. Linton, and S. Srisuma (2011): "Nonparametric Euler Equation Identification and Estimation," Working paper, Boston College and London School of Economics.

Ross, S. A. (2014): "The Recovery Theorem," *Journal of Finance*, forthcoming.

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

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

Stone, C. J. (1982): "Optimal Global Rates of Convergence for Nonparametric Regression," *Annals of Statistics*, 10, 1040–1053.

Tallarini, T. D. (2000): "Risk-Sensitive Real Business Cycles," *Journal of Monetary Economics*, 45, 507–532.

Tauchen, G. and R. Hussey (1991): "Quadrature-Based Methods for Obtaining Approximate Solutions to Nonlinear Asset Pricing Models," *Econometrica*, 59, pp. 371–396.

Weil, P. (1990): "Nonexpected Utility in Macroeconomics," *The Quarterly Journal of Economics*, 105, 29–42.

Supplementary Appendix to

Nonparametric Stochastic Discount Factor Decomposition

Timothy M. Christensen

December 10, 2014

This appendix contains material to support the paper "Nonparametric Stochastic Discount Factor Decomposition". Appendix A presents further details on the relation between the identification and existence conditions in Section 3 and the identification and existence conditions in Hansen and Scheinkman (2009). Appendix C contains further results on nonparametric estimation of positive eigenfunctions. Appendix B presents formulae for ρ, ϕ and ϕ^* for three parametric specifications in the literature, thereby verifying Assumption 3.3 for these models. The proofs of all results in the main text and this supplement are presented in Appendix D. Finally, Appendix E presents additional Monte Carlo results.

A Further discussion of identification and existence conditions

A.1 Identification

To establish identification of ϕ, Hansen and Scheinkman (2009) impose a set of stochastic stability conditions under which there is at most one positive eigenfunction that is germane to their long-run approximation. We now present a version of their stochastic stability conditions that are tailored to discrete-time environments and discuss the connection between their conditions and the identification conditions in the present paper. Some of the identification conditions in Hansen and Scheinkman (2009) pertain 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. 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 $\psi \mapsto \widetilde{\mathbb{E}}[\psi(X_1)|X_0 = x]$.

The following are Assumptions 6.1, 7.1, 7.2, 7.3, and 7.4 of Hansen and Scheinkman (2009) tailored to discrete-time environments (so here the time index t takes values in the set $\{0, 1, \ldots\}$).

Assumption A.1 *(a) $\{M_t^P : t \geq 0\}$ is a multiplicative martingale;*

(b) $\{M_t : t \geq 0\}$ is a strictly positive process with probability 1;

(c) there exists a probability measure $\hat{\varsigma}$ such that

$$\int \widetilde{\mathbb{E}}[\psi(X_1)|X_0 = x] \, \mathrm{d}\hat{\varsigma}(x) = \int \psi(x) \, \mathrm{d}\hat{\varsigma}(x)$$

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

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

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

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

(e) for any $\Lambda \in \mathscr{X}$ with $\hat{\varsigma}(\Lambda) > 0$,

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

for all $x \in \mathcal{X}$, where the probability $\widetilde{\mathbb{P}}(\cdot|X_0 = x)$ is absolutely continuous with respect to $\widetilde{\mathbb{P}}$ (where $\widetilde{\mathbb{P}}$ is given by $\widetilde{\mathbb{P}}(A) = \int \mathbb{E}[(M_t^P/M_0^P)\chi_A|X_0 = x] \, \mathrm{d}\hat{\varsigma}(x)$ for each $A \in \mathcal{F}_t$) conditioned on $X_0 = x$ when restricted to $\mathcal{F}_t$ for each $t \geq 0$.

We now discuss the relation between Assumption A.1 (i.e. Hansen and Scheinkman's (2009) assumptions in a discrete-time setting) and Assumptions 3.1, 3.2 and 3.3 in the present paper.

Part (a) is satisfied by our construction of the permanent component, and part (b) is analogous to the condition $\mathcal{K}_{\mathbb{M}}(x, y) \geq 0$ a.e.-$[Q \otimes Q]$ in Assumption 3.2.

For part (c), let ϕ and ϕ^* be as in Assumption 3.3 and normalize ϕ^* such that $\mathbb{E}[\phi(X_0)\phi^*(X_0)] = 1$. Under this normalization we can define a probability measure $\hat{\varsigma}$ by $\hat{\varsigma}(A) = \mathbb{E}[\phi(X_0)\phi^*(X_0)\chi_A(X_0)]$ for all $A \in \mathscr{X}$. We then have:

$$\begin{aligned}
\int \widetilde{\mathbb{E}}[\psi(X_1)|X_0 = x] \, \mathrm{d}\hat{\varsigma}(x) &= \int \mathbb{E}\left[\rho^{-1} m(X_0, X_1) \frac{\phi(X_1)}{\phi(X_0)} \psi(X_1) \middle| X_0 = x \right] \phi(x)\phi^*(x) \, \mathrm{d}Q(x) \\
&= \rho^{-1} \mathbb{E}\left[\phi^*(X_0)(\mathbb{M}(\phi\psi)(X_0))\right] \\
&= \rho^{-1} \mathbb{E}\left[((\mathbb{M}^*\phi^*)(X_1))\phi(X_1)\psi(X_1)\right] \\
&= \mathbb{E}[\phi^*(X_1)\phi(X_1)\psi(X_1)] \\
&= \int \psi(x) \, \mathrm{d}\hat{\varsigma}(x) \, .
\end{aligned}$$

Therefore, Assumption A.1(c) is satisfied under Assumption 3.3 and our maintained assumption of stationarity. A similar derivation is reported for continuous-time semigroups in an preliminary 2005 draft of Hansen and Scheinkman (2009), but there the stationary distribution Q is replaced by an arbitrary measure.

For part (d), note that $\hat{\varsigma}(\Lambda) > 0$ implies $Q(\Lambda) > 0$ under our construction of $\hat{\varsigma}$. Therefore, $\hat{\varsigma}(\Lambda) > 0$ implies $\phi\chi_\Lambda$ is positive on a set of positive Q measure. Moreover, by definition of $\widetilde{\mathbb{E}}$ we have:

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

for any $\lambda \geq \rho$. Assumption 3.2(b) is necessary and sufficient $\mathbb{M}$ to be irreducible and, by definition of irreducibility of $\mathbb{M}$, $\sum_{t=1}^{\infty}\lambda^{-t}\mathbb{M}_t(\phi\chi_\Lambda)(x) > 0$ a.e.-$[Q]$ holds for $\lambda > \rho$ (Schaefer, 1999, p. 317). Therefore, Assumption 3.2(b) implies Assumption A.1(d), up to the qualification "a.e.-$[Q]$".

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

$$\lim_{n\to\infty}\sup_{0\leq\psi\leq\phi}\left|\widetilde{\mathbb{E}}\left[\frac{\psi(X_n)}{\phi(X_n)}\middle| X_0 = x\right] - \int\frac{\psi(x)}{\phi(x)}\,\mathrm{d}\hat{\varsigma}(x)\right| = 0 \quad \text{a.e.-}[\hat{\varsigma}] \tag{32}$$

as $n \to \infty$, where the supremum is taken over all measurable ψ such that $0 \leq \psi \leq \phi$ (Meyn and Tweedie, 2009, Proposition 14.3.1).

The result (32) is a discrete-time version of Proposition 7.1 in Hansen and Scheinkman (2009), which they use to establish identification of ϕ. Our identification conditions alone are not enough to obtain a convergence result like (32) (cf. Theorem 3.3). 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 Hansen and Scheinkman (2009). Indeed, for non-stationary environments it is not even clear how to restrict the class of functions appropriately to define an adjoint (for instance, Hansen and Scheinkman (2009) 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 3.3.

A.2 Existence

Hansen and Scheinkman (2009) 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 environment 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.

Assumption A.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 $x \in \mathcal{X}$*

(b) there exists a measure ν on $(\mathcal{X}, \mathscr{X})$ such that $\mathbb{F}\chi_\Lambda(x) > 0$ for any $\Lambda \in \mathscr{X}$ with $\nu(\Lambda) > 0$, where $\mathbb{F}$ is given by

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

for $a > \underline{a}$

(c) the operator $\mathbb{G}$ given by

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

is bounded on the space of bounded functions, where $s : \mathcal{X} \to \mathbb{R}$ is a non-negative function such that $\int s(x)\,\mathrm{d}\nu(x) > 0$ such that $\mathbb{F}\psi(x) \geq s(x) \int \psi(u)\,\mathrm{d}\nu(u)$ for all $\psi \geq 0$ (s exists by part (b)), $(s \otimes \nu)\psi(x) := s(x) \int \psi(u)\,\mathrm{d}\nu(u)$, and λ belongs to the spectrum of $\mathbb{F}$.

Hansen and Scheinkman (2009) show that $\phi := V\mathbb{G}s$ is a positive eigenfunction of $\mathbb{M}$. Let us now consider how these existence conditions compare with the existence conditions in the present paper.

Part (b) is satisfied under Assumption 3.2 with $\nu = Q$ whenever $a > r(\mathbb{M})$ where $r(\mathbb{M})$ denotes the spectral radius of $\mathbb{M}$. By positivity and irreducibility of $\mathbb{M}$ (cf. Assumption 3.2(a) and (b), respectively), if $\Lambda \in \mathscr{X}$ with $Q(\Lambda) > 0$ then

$$\sum_{t=1}^{\infty} a^{-t}\mathbb{M}_t(V\chi_\Lambda)(x) \geq \sum_{t=1}^{\infty} a^{-t}\mathbb{M}_t\chi_\Lambda(x) > 0 \quad \text{a.e.-}[Q]$$

where the first inequality is by positivity and the second is by irreducibility. It follows that $\mathbb{F}\chi_\Lambda(x) > 0$ a.e.-$[Q]$. This verifies part (b) (up to the qualification "a.e.-$[Q]$").

On the other hand, parts (a) and (c) of Assumption A.2 seem quite different from the conditions of Theorem 3.1. For instance, the conditions of Theorem 3.2 do not presume existence of the function V but impose a power compactness condition. Hansen and Scheinkman (2009) 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 $\mathbb{G}$ be bounded (or the sufficient conditions for this provided in Hansen and Scheinkman (2009)) do not seem to translate clearly in terms of the operator $\mathbb{M}$ or the kernel $\mathcal{K}_\mathbb{M}$.

B Verification of the identification conditions in some parametric models

B.1 Exponentially affine SDF

Let X_t denote the $N \times 1$ vector of state variables each period. Let ϵ_{t+1} denote a $N \times 1$ vector of $N(0, I)$ shocks that are independent of X_t. Assume X_t evolves according to

$$X_{t+1} = AX_t + \sigma\epsilon_{t+1} \tag{33}$$

where all the eigenvalues of the $N \times N$ matrix A lie within the unit circle, and $\Sigma := \sigma\sigma'$ is positive definite. We also assume the SDF is of the form

$$m(X_t, X_{t+1}) = \alpha_0 \exp\left\{\alpha_1' X_t + \alpha_2' \sigma\epsilon_{t+1}\right\} \tag{34}$$

where $\alpha_0 \in \mathbb{R}$ and $\alpha_1, \alpha_2 \in \mathbb{R}^N$. If log consumption growth is an affine function of X_t then this SDF specification is obtained, e.g., under Epstein-Zin preferences with unit EIS (Hansen et al., 2008) and also under CRRA preferences.

Lemma B.1 *Let $\{X_t\}$ follow (33) and let the SDF be as in (34). Then: Assumption 3.1 holds for all L^p spaces with $1 < p < \infty$; Assumption 3.3 holds with*

$$\begin{aligned}
\phi(x) &= \beta_0 \exp\left\{\alpha_1'(I-A)^{-1}x\right\} \\
\phi^*(x) &= \beta_0^* \exp\left\{[\alpha_1' VA' + \alpha_2'\Sigma](I-A')^{-1}V^{-1}x\right\} \\
\rho &= \alpha_0 \exp\{\tfrac{1}{2}(\alpha_1'(I-A)^{-1} + \alpha_2')\Sigma((I-A')^{-1}\alpha_1 + \alpha_2)\}\,.
\end{aligned}$$

for positive constants β_0 and β_0^*, where $V = \sum_{j\geq 0} A^j \Sigma A'^j$; and Assumption 3.5 holds (and therefore Assumption 3.2 also holds). Note that $\phi \in L^p$ and $\phi^* \in L^q$ for all $1 \leq p, q < \infty$.

B.2 Exponentially quadratic SDF

Here we maintain the dynamic specification (33) but now assume that the stochastic discount factor is of the form

$$m(X_t, X_{t+1}) = \alpha_0 \exp\left\{\alpha_1' X_t + \alpha_2'\sigma\epsilon_{t+1} + X_t'\Gamma_1 X_t + X_t'\Gamma_2\sigma\epsilon_{t+1}\right\} \tag{35}$$

where α_0, α_1, and α_2 are as in the exponentially affine case, $\Gamma_1 \in \mathbb{R}^{N\times N}$ is symmetric with $-\Gamma_1$ non-nonnegative definite, and $\Gamma_2 \in \mathbb{R}^{N\times N}$. This specification has been used as a reduced-form SDF to examine the term structure of equity (see, e.g., Lettau and Wachter (2007, 2011)) and in the extensive literature on affine term structure models. In what follows, we impose the "essentially affine" restriction $\Gamma_1 + \frac{1}{2}\Gamma_2\Sigma\Gamma_2' = 0$ (Duffee, 2002) because of the extensive use of this restriction in the literature.[23]

Lemma B.2 *Let $\{X_t\}$ follow (33), let the SDF be as in (35) with $\Gamma_1 + \frac{1}{2}\Gamma_2\Sigma\Gamma_2' = 0$, and let all eigenvalues of $A' + \Gamma_2\Sigma$ lie inside the unit circle. Then: Assumption 3.1 holds for all L^p spaces with $1 < p < \infty$; Assumption 3.3 holds with*

$$\begin{aligned}
\phi(x) &= \beta_0 \exp\left\{[\alpha_1' + \alpha_2'\Sigma\Gamma_2'][I - A - \Sigma\Gamma_2']^{-1}x\right\} \\
\phi^*(x) &= \beta_0^* \exp\left\{\beta_1^* x - x'B^*x\right\} \\
\rho &= \alpha_0 \exp\{\tfrac{1}{2}(\alpha_1' + \alpha_2'(I - A))(I - A - \Sigma\Gamma_2')^{-1}\Sigma(I - A' - \Gamma_2\Sigma)^{-1}(\alpha_1 + (I - A')\alpha_2)\}
\end{aligned}$$

for positive constants β_0 and β_0^*, where

$$\begin{aligned}
B^* &= \frac{1}{2}\left(\Big(\sum_{j\geq 0}(A + \Sigma\Gamma_2')^j\Sigma(A' + \Gamma_2\Sigma)^j\Big)^{-1} - V^{-1}\right) \\
\beta_1^* &= (I - S'D^{-1})^{-1}(\alpha_2 + SD^{-1}(\alpha_1 - A'\alpha_2))
\end{aligned}$$

with $D = S\Sigma S' + V^{-1} + 2B^*$, $S = A'\Sigma^{-1} + \Gamma_2$, and $V = \sum_{j\geq 0} A^j\Sigma A'^j$; and Assumption 3.5 holds (and therefore Assumption 3.2 also holds). Note that $\phi \in L^p$ for all $1 \leq p < \infty$ and $\phi^* \in L^q$ for all $q \in (1, \bar{q})$ where $\bar{q} = \sup\{q > 1 : qB^* + \frac{1}{2}V^{-1}$ is positive definite$\}$.

The "stability condition" requiring all eigenvalues of $A' + \Gamma_2\Sigma$ to lie inside the unit circle implies that (i) B^* is the unique solution of the discrete-time algebraic Riccati equation

$$2B^* = \Sigma^{-1} - V^{-1} - S'(2B^* + V^{-1} + S\Sigma S'))^{-1}S \tag{36}$$

and (ii) $\phi^* \in L^q$ for some $q > 1$. If the stability condition fails then there may be multiple B^* solving (36) and therefore multiple ϕ^* solving $\mathbb{M}^*\phi^* = \rho\phi^*$. However, none of these ϕ^* would belong to L^q for any $q > 1$ because the stability condition is also a necessary condition for positive definiteness of $qB^* + \frac{1}{2}V^{-1}$ for any $q > 1$ (see the proof of Lemma B.2).

B.3 Recursive preferences and stochastic volatility

The previous two examples assumed the state process was linear. We now show Assumption 3.3 is easily verified for a nonlinear example. We assume that the log consumption growth process for $g_t = \log(C_t/C_{t-1})$ is:

$$\begin{aligned}
g_{t+1} &= (1 - \kappa)\bar{g} + \kappa g_t + \sigma v_t^{1/2}\epsilon_{t+1} \\
v_{t+1} &= \text{ARG}(c_v, \varphi_v, \delta_v)
\end{aligned} \tag{37}$$

[23]The essentially affine condition can be relaxed but the analysis becomes substantially more complicated.

where ϵ_{t+1} is an independent $N(0,1)$ random variable and $\text{ARG}(c_v, \varphi_v, \delta_v)$ is an autoregressive gamma (ARG) process of order 1 which is parameterized by $(c_v, \varphi_v, \delta_v)$ (see Gourieroux and Jasiak (2006) for details). The ARG process is a discrete-time version of the familiar continuous-time square root process. The state vector is $X_t = (g_t, v_t)'$.

We assume that $m(X_t, X_{t+1})$ is of the form

$$m(X_t, X_{t+1}) = \alpha_0 \exp\left\{\alpha_1 g_{t+1} + \alpha_2 v_t + \alpha_3 v_{t+1} + \alpha_4 \sigma v_t^{1/2} \epsilon_{t+1}\right\} \tag{38}$$

where $\alpha_1 = -1$, and where $\alpha_2 = \alpha_3 = 0$ when $\sigma = 0$. This SDF is obtained under Epstein-Zin preferences with EIS = 1 by following Appendix H of Backus et al. (2014).

To solve for ρ, ϕ, and ϕ^* we conjecture solutions that are exponentially affine in $X_t = (g_t, v_t)'$ because the characteristic function of the ARG process is exponentially affine. Further, to solve for ϕ^* we use the fact that $\{v_t\}_{t=-\infty}^{\infty}$ is time reversible, and that $\{g_t\}_{t=-\infty}^{\infty}$ conditioned on $\{v_t\}_{t=-\infty}^{\infty}$ is time reversible because it a scalar linear process with Gaussian innovations.

Lemma B.3 *Let $\{X_t\}$ follow (37) and let the SDF be as in (38). Then:*

$$\begin{aligned}
\phi(g,v) &= \beta_0 \exp\left\{\frac{\alpha_1 \kappa}{1-\kappa} g + \beta_2 v\right\} \\
\phi^*(g,v) &= \beta_0^* \exp\left\{\left[\frac{\alpha_1}{1-\kappa} + \alpha_4(1+\kappa)\right] g + \beta_2^* v\right\} \\
\rho &= \alpha_0 \exp\left\{\alpha_1 \bar{g} - \delta_v \log(1 - (\alpha_3 + \beta_2) c_v)\right\}
\end{aligned}$$

solve $\mathbb{M}\phi = \rho\phi$ and $\mathbb{M}^*\phi^* = \rho\phi^*$, where

$$\begin{aligned}
\beta_2 &= \frac{(1 + (\xi - \alpha_3)c_v - \varphi_v) - \sqrt{(1 + (\xi - \alpha_3)c_v - \varphi_v)^2 - 4c_v(\xi - \alpha_3\xi + \varphi_v\alpha_3)}}{2c_v} \\
\beta_2^* &= \beta_2 + \alpha_3 - \xi \\
\xi &= \alpha_2 + \frac{1}{2}\sigma^2 \left(\frac{\alpha_1}{1-\kappa} + \alpha_4\right)^2
\end{aligned}$$

and β_0 and β_0^* are positive constants.

C Additional results on estimation

In this section we derive convergence rates of the bias and variance terms, supplementary results useful for verification of Assumption 4.3, and some additional results related to inference.

C.1 Bias and variance calculations

Before presenting the results, it is worth emphasizing the distinction between ϕ_K^+ and ϕ_K^*. Recall that $\phi_K^+(x) = b^K(x)'c_K^*$ is the eigenfunction corresponding to ρ_K of the adjoint of $\Pi_K\mathbb{M}$ with respect to the subspace B_K. Let ϕ_K^* denote the eigenfunction corresponding to ρ_K of the adjoint of $\Pi_K\mathbb{M}$ with respect to the space L^2 (these quantities are uniquely defined for all K sufficiently large under Assumptions 4.1 and 4.2(a)). That is,

$$\begin{aligned}\mathbb{E}[\phi_K^*(X_0)\Pi_K\mathbb{M}\psi(X_0)] &= \rho_K\mathbb{E}[\phi_K^*(X_0)\psi(X_0)] \\ \mathbb{E}[\phi_K^+(X_0)\Pi_K\mathbb{M}\psi_K(X_0)] &= \rho_K\mathbb{E}[\phi_K^+(X_0)\psi_K(X_0)]\end{aligned}$$

for all $\psi \in L^2$ and $\psi_K \in B_K$. It follows that $\Pi_K\phi_K^* = \phi_K^+$. Lemma C.1 below shows that ϕ_K^* and ϕ_K^+ converge to ϕ^* at the same rate.

As eigenfunctions are only identified up to sign and scale, for the remainder of this section we impose the scale normalizations $\|\phi\| = 1$, $\|\phi_K\| = 1$, $\|\widehat{\phi}\| = 1$, $\langle\phi,\phi^*\rangle = 1$, $\langle\phi_K,\phi_K^*\rangle = 1$, and $\langle\widehat{\phi},\widehat{\phi}^*\rangle = 1$, and the sign normalizations $\langle\phi,\phi_K\rangle \geq 0$, $\langle\phi^*,\phi_K^*\rangle \geq 0$, $\langle\phi^*,\phi_K^+\rangle \geq 0$, and $\langle\widehat{\phi}^*,\phi_K^+\rangle \geq 0$.

The following Lemma provides convergence rates for the bias terms. These rates bound the approximation error introduced when the infinite-dimensional eigenfunction problem (4) is approximated by the matrix eigenvector problem (19).

Lemma C.1 *Let Assumptions 4.1 and 4.2(a) hold. Then there exists $\bar{K} \in \mathbb{N}$ such that for all $K \geq \bar{K}$:*

(a) ρ_K *is real and simple*

(b) ϕ_K *is the unique eigenfunction of* $\Pi_K\mathbb{M}$ *corresponding to the eigenvalue* ρ_K

(c) $(\Pi_K\mathbb{M})^*$ *has a unique eigenfunction* ϕ_K^* *corresponding to the eigenvalue* ρ_K.

If, in addition, Assumption 4.2(b) holds:

(d) $|\rho_K - \rho| = O(\delta_K)$

(e) $\|\phi_K - \phi\| = O(\delta_K)$.

Further, if Assumption 4.2(c) also holds:

(f) $\|\phi_K^*/\|\phi_K^*\| - \phi^*/\|\phi^*\|\| = O(\delta_K^*)$

(h) $\|\phi_K^+/\|\phi_K^+\| - \phi^*/\|\phi^*\|\| = O(\delta_K^*)$.

With Lemma C.1 in hand it remains to derive convergence rates for the variance terms.

Lemma C.2 *Let Assumptions 4.1, 4.2(a), and 4.3(a) hold. Then with probability approaching one:*

(a) $\widehat{\rho}$ *is real and simple*

(b) $\widehat{c} \in \mathbb{R}^K$ *is the unique eigenvector of* $\widehat{\mathbf{G}}^{-1}\widehat{\mathbf{M}}$ *corresponding to the eigenvalue* $\widehat{\rho}$

(c) $\widehat{c}^* \in \mathbb{R}^K$ *is the unique eigenvector of* $\widehat{\mathbf{G}}^{-1}\widehat{\mathbf{M}}'$ *corresponding to the eigenvalue* $\widehat{\rho}$.

If, in addition, Assumption 4.3(b) holds:

(d) $|\widehat{\rho} - \rho_K| = O_p(\eta_{n,K})$

(e) $\|\widehat{\phi} - \phi_K\| = O_p(\eta_{n,K})$.

Further, if Assumption 4.3(c) also holds:

(f) $\|\widehat{\phi}^*/\|\widehat{\phi}^*\| - \phi_K^+/\|\phi_K^+\|\| = O_p(\eta_{n,K})$.

C.2 Convergence results for matrix estimators

We now provide some sufficient conditions for Assumption 4.3 in Cases 1, 2, and 3. We derive the results assuming $\{X_t\}$ is either beta-mixing or rho-mixing because many popular models for macroeconomic or financial time series imply this form of weak dependence. Examples include copula-based Markov models (Chen, Wu, and Yi, 2009; Beare, 2010) and discretely sampled Markov diffusion processes (Chen, Hansen, and Carrasco, 2010). The results presented below for beta-mixing data use an exponential inequality for sums of weakly-dependent random matrices derived by Chen and Christensen (2014); the results for rho-mixing data adapt arguments in Gobet et al. (2004).

We first present sufficient conditions for Assumption 4.3 in Case 1, for $\widehat{\mathbf{G}}$ and $\widehat{\mathbf{M}}$ in (21) and (22). Let $\widetilde{b}^K(x) = \mathbf{G}_K^{-1/2} b^K(x)$ denote the orthonormalized vector of basis functions, where $\mathbf{G}_K^{-1/2}$ denotes the inverse of the positive definite square root of $\mathbf{G}_K$. Let $\xi_K = \sup_x \|b^K(x)\|$ denote the usual measure of roughness of the sieve basis functions and let $\lambda_K = 1/\sqrt{\lambda_{\min}(\mathbf{G}_K)}$ denote the reciprocal of the square root of the minimum eigenvalue of $\mathbf{G}$.

Lemma C.3 *Let* $\{X_t\}$ *be strictly stationary and exponentially beta-mixing, let* $\mathbb{E}[|m(X_0, X_1)|^r] < \infty$ *for some* $2 \leq r \leq \infty$, *and let* $\xi_K \lambda_K (\log n)/\sqrt{n} = O(1)$. *Then:*

(a) $\|\widehat{\mathbf{G}}^{-1}\widehat{\mathbf{M}} - \mathbf{G}^{-1}\mathbf{M}\|_{\mathbf{G}} = O_p((\xi_K \lambda_K)^{1+2/r}(\log n)/\sqrt{n})$

(b) Assumption 4.3(b)(c) holds with $\eta_{n,K} = \eta^*_{n,K} = O((\xi_K \lambda_K)^{1+2/r}(\log n)/\sqrt{n})$.

Lemma C.4 *Let $\{X_t\}$ be strictly stationary and exponentially rho-mixing and let $\mathbb{E}[|m(X_0,X_1)|^r] < \infty$ for some $2 \leq r \leq \infty$. Then:*

(a) $\|\widehat{\mathbf{G}}^{-1}\widehat{\mathbf{M}} - \mathbf{G}^{-1}\mathbf{M}\|_{\mathbf{G}} = O_p((\xi_K\lambda_K)^{1+2/r}\sqrt{K}/\sqrt{n})$

(b) Assumption 4.3(b)(c) holds with $\eta_{n,K} = \eta^*_{n,K} = O((\xi_K\lambda_K)^{1+2/r}/\sqrt{n})$.

We now present sufficient conditions for Cases 2 and 3 using $\widehat{\mathbf{G}}$ and $\widehat{\mathbf{M}}$ in (21) and (23).

Lemma C.5 *Let $\{X_t\}$ be strictly stationary and exponentially beta-mixing, let $\mathbb{E}[|m(X_0,X_1)|^r] < \infty$ for some $2 \leq r \leq \infty$, let $(\frac{1}{n}\sum_{t=0}^{n-1}(\widehat{m}(X_t,X_{t+1}) - m(X_t,X_{t+1}))^2)^{1/2} = O_p(\nu_n)$ where $\nu_n = o(1)$, and let $\xi_K\lambda_K(\log n)/\sqrt{n} = o(1)$. Then:*

(a) $\|\widehat{\mathbf{G}}^{-1}\widehat{\mathbf{M}} - \mathbf{G}^{-1}\mathbf{M}\|_{\mathbf{G}} = O_p((\xi_K\lambda_K)^{1+2/r}(\log n)/\sqrt{n} + \xi_K\lambda_K\nu_n)$

(b) Assumption 4.3(b)(c) holds with $\eta_{n,K} = \eta^*_{n,K} = O((\xi_K\lambda_K)^{1+2/r}(\log n)/\sqrt{n} + \xi_K\lambda_K\nu_n)$.

Lemma C.6 *Let $\{X_t\}$ be strictly stationary and exponentially rho-mixing, let $\mathbb{E}[|m(X_0,X_1)|^r] < \infty$ for some $2 \leq r \leq \infty$, let $(\frac{1}{n}\sum_{t=0}^{n-1}(\widehat{m}(X_t,X_{t+1}) - m(X_t,X_{t+1}))^2)^{1/2} = O_p(\nu_n)$ where $\nu_n = o(1)$, and let $\xi_K\lambda_K\sqrt{K}/\sqrt{n} = o(1)$. Then:*

(a) $\|\widehat{\mathbf{G}}^{-1}\widehat{\mathbf{M}} - \mathbf{G}^{-1}\mathbf{M}\|_{\mathbf{G}} = O_p((\xi_K\lambda_K)^{1+2/r}\sqrt{K}/\sqrt{n} + \xi_K\lambda_K\nu_n)$

(b) Assumption 4.3(b)(c) holds with $\eta_{n,K} = \eta^*_{n,K} = O((\xi_K\lambda_K)^{1+2/r}/\sqrt{n} + \xi_K\lambda_K\nu_n)$.

Remark C.1 *Lemmas C.5 and C.6 show that the mean-square convergence rate of $\widehat{m}$ to m, namely ν_n, affects the convergence rates of the eigenfunction estimators. With all else being equal, if $\nu_n \to 0$ at a slower rate then $\eta_{n,K}$ and $\eta^*_{n,K}$ will be larger for given K (see parts (b) of Corollaries C.5 and C.6). Moreover, larger ν_n means that K will have to increase more slowly with with the sample size to verify Assumption C.1, which means δ_K and δ^*_K will vanish more slowly as n increases. Consequently, the convergence rates of the estimators $\widehat{\rho}$, $\widehat{\phi}$ and $\widehat{\phi}^*$ will be slower.*

C.3 Sieve perturbation expansion

The following result shows that, to first order, $\widehat{\rho} - \rho_K$ behaves as a linear functional of $\widehat{\mathbf{M}} - \rho_K\widehat{\mathbf{G}}$. This result is used to derive the limiting distribution of $\widehat{\rho}$, $\widehat{y}$ and $\widehat{L}$ in Theorem 4.2, and may be applied to derive the asymptotic distribution of $\widehat{\rho}$, $\widehat{y}$ and $\widehat{L}$ in Cases 2 and 3. To introduce the following result, let $\mathbf{M}^o$, $\widehat{\mathbf{G}}^o$, and $\widehat{\mathbf{M}}^o$ by obtained by pre- and post-multiplying $\mathbf{M}$, $\widehat{\mathbf{G}}$ and $\widehat{\mathbf{M}}$ by $\mathbf{G}^{-1/2}$ (where $\mathbf{G}^{-1/2}$ denotes the inverse of the positive definite square root of $\mathbf{G}$).

Lemma C.7 *Let $\|\widehat{\mathbf{G}}^o - I\| = O_p(\bar\eta_{n,K,1})$ and $\|\widehat{\mathbf{M}}^o - \mathbf{M}^o\| = O_p(\bar\eta_{n,K,2})$ where $\bar\eta_{n,K,1} = o(1)$, and let Assumptions 4.1, 4.2 and 4.3 hold. Then:*

$$\widehat\rho - \rho_K = c_K^{*\prime}(\widehat{\mathbf{M}} - \rho_K \widehat{\mathbf{G}})c_K + O_p((\eta_{n,K} \vee \bar\eta_{n,K,1})(\bar\eta_{n,K,1} \vee \bar\eta_{n,K,2}))\,.$$

C.4 Asymptotic inference in Case 1

We now turn to providing estimators of the asymptotic variances V_ρ, V_y, and V_L in Case 1. The rates of convergence in Theorem 4.1 are for estimators under the scale normalization $\mathbb{E}[\widehat\phi(X_0)^2] = 1$ and $\mathbb{E}[\widehat\phi^*(X_0)\widehat\phi(X_0)] = 1$. In practice the measure Q is unknown so these normalizations are infeasible. Therefore, we let $\widehat\phi^f$ and $\widehat\phi^{*f}$ denote $\widehat\phi$ and $\widehat\phi^*$ renormalized under the empirical measure, i.e. $n^{-1}\sum_{t=0}^{n-1}\widehat\phi^f(X_t)^2 = 1$ and $n^{-1}\sum_{t=0}^{n-1}\widehat\phi^f(X_t)\widehat\phi^{*f}(X_t) = 1$. Our estimator of V_ρ is

$$\widehat V_\rho = \frac{1}{n}\sum_{t=0}^{n-1}\Big(\widehat\phi^{*f}(X_t)m(X_t,X_{t+1})\widehat\phi^f(X_{t+1}) - \widehat\rho\widehat\phi^{*f}(X_t)\widehat\phi^f(X_t)\Big)^2$$

and our estimator of V_y is $\widehat V_y = \widehat\rho^{-2}\widehat V_\rho$. Estimating V_L involves estimating a long-run variance. We use an orthogonal series long-run variance (OSLRV) estimator of Phillips (2005) in conjunction with fixed-bandwidth asymptotics as in Chen, Liao, and Sun (2012).[24] Let $\{h_j : j \geq 0\}$ be a continuously differentiable orthonormal basis for $L_2[0,1]$ (the space of measurable functions on $[0,1]$ that are square-integrable with respect to Lebesgue measure), such as a cosine or Legendre polynomial basis. Let $h_0 = 1$ so that $\int_0^1 h_j(u)\,\mathrm{d}u = 0$ for each $j \geq 1$. For each $j = 1,\ldots,J$, define

$$\widehat\Lambda_j = \frac{1}{\sqrt n}\sum_{t=0}^{n-1} h_j\Big(\frac{t+1}{n}\Big)\widehat\psi_L(X_t,X_{t+1})$$

where

$$\widehat\psi_L(X_t,X_{t+1}) = \widehat\rho^{-1}\widehat\phi^{*f}(X_t)m(X_t,X_{t+1})\widehat\phi^f(X_{t+1}) - \widehat\phi^{*f}(X_t)\widehat\phi^f(X_t) - \big(\log m(X_t,X_{t+1}) - \overline{lm}_n\big)$$

and $\overline{lm}_n = n^{-1}\sum_{t=0}^{n-1}\log m(X_t,X_{t+1})$. Our estimator for V_L is

$$\widehat V_{L,J} = \frac{1}{J}\sum_{j=1}^{J}\widehat\Lambda_j^2\,. \tag{39}$$

In what follows we use "fixed-bandwidth asymptotics" in that we keep J fixed as $n \to \infty$. As a result, we will obtain $\sqrt n \widehat V_{L,J}^{-1/2}(\widehat L - L) \to_d t_J$ rather than the usual $N(0,1)$ limit obtained with consistent long-run variance estimation.

[24] A considerable literature has shown that asymptotic inference with consistent kernel-based truncated-lag estimators can suffer size and power distortions in finite samples, and has proposed fixed-bandwidth asymptotics as a remedy (see, e.g., Kiefer, Vogelsang, and Bunzel (2000); Jansson (2004); Müller (2007)).

To simplify notation, let $\phi_t := \phi(X_t)$, let $m_{t,t+1} := m(X_t, X_{t+1})$, and so on.

Assumption C.1 *(a) each of* $\mathbb{E}[\phi_0^{*2} m_{0,1} \phi_0 \phi_1]$, $\mathbb{E}[\phi_0^{*2} m_{0,1}^2 \phi_1^2]$, *and* $\mathbb{E}[\phi_0^{*2} \phi_0^2]$ *are finite;*

(b) $\xi_K \lambda_K (\|\widehat{\phi}^{*f} - \phi^*\| \vee \|\widehat{\phi}^f - \phi\|) = o_p(1)$;

(c) there exists a neighborhood N_K *of* (ϕ, ϕ^*) *such that* $(\widehat{\phi}^f, \widehat{\phi}^{*f}) \in N_K$ *wpa1 and for which*

$$\sup_{(f,f^*) \in N_K} \sum_{t=0}^{n-1} h_j\Big(\frac{t+1}{n}\Big)\{\phi_t^* \phi_t - f_t^* f_t - \mathbb{E}[\phi_0^* \phi_0 - f_0^* f_0]\} = o_p(n^{1/2})$$

and

$$\sup_{(f,f^*) \in N_K} \sum_{t=0}^{n-1} h_j\Big(\frac{t+1}{n}\Big)\{m_{t,t+1}(\phi_t^* \phi_{t+1} - f_t^* f_{t+1}) - \mathbb{E}[m_{0,1}(\phi_0^* \phi_1 - f_0^* f_1)]\} = o_p(n^{1/2})\,;$$

(d) $n^{-1} \sum_{t=0}^{n-1} h_j(\frac{t+1}{n})(\phi_t^* m_{t,t+1} \phi_{t+1} - \mathbb{E}[\phi_0^* m_{0,1} \phi_1]) = o_p(1)$; *and*

(e) for any $(v_0, v_1, \ldots, v_J)' \in \mathbb{R}^{J+1}$ *we have*

$$\Pr\Big(\bigcap_{j=0}^{J}\Big\{\frac{1}{\sqrt{n}}\sum_{t=0}^{n-1} h_j\Big(\frac{t+1}{n}\Big)\psi_L(X_t, X_{t+1}) \le v_j\Big\}\Big) = \Pr\Big(\bigcap_{j=0}^{J}\Big\{\frac{1}{\sqrt{n}}\sum_{t=0}^{n-1} h_j\Big(\frac{t+1}{n}\Big)\omega_t \le v_j\Big\}\Big) + o(1)$$

where $\omega_0, \omega_1, \ldots, \omega_{n-1}$ *are i.i.d.* $N(0, V_L)$ *random variables.*

Assumption C.1(a) ensures the individual terms in V_ρ are well defined. Assumptions C.1(c)(e) are versions of Assumption 5.2(i)(iv) of Chen et al. (2012); C.1(e) is weaker than assuming a functional central limit theorem applies. Assumption C.1(c)(d)(e) may be verified under specific weak dependence assumptions.

Theorem C.1 *Let Assumptions 4.1, 4.2, 4.3, 4.4 and C.1(a)(b) hold. Then:*

(a) $\widehat{V}_\rho \to_p V_\rho$ *and* $\sqrt{n}\widehat{V}_\rho^{-1/2}(\widehat{\rho} - \rho) \to_d N(0,1)$

(b) $\widehat{V}_y \to_p V_y$ *and* $\sqrt{n}\widehat{V}_y^{-1/2}(\widehat{y} - y) \to_d N(0,1)$.

If, in addition, Assumption C.1(c)(d)(e) holds, then:

(c) $J\widehat{V}_{L,J}/V_L \to_d \chi_J^2$ *and* $\sqrt{n}\widehat{V}_{L,J}^{-1/2}(\widehat{L} - L) \to_d t_J$.

D Proofs

D.1 Proof of results in the main text

If $\mathbb{T}$ is a bounded linear operator on a Banach space $\mathcal{B}$ we define its spectrum $\sigma(\mathbb{T})$ as the complement in $\mathbb{C}$ of the set of all $z \in \mathbb{C}$ for which the resolvent operator $\mathcal{R}(\mathbb{T}, z) := (\mathbb{T} - zI)^{-1}$ is a bounded linear operator on $\mathcal{B}$. We also let $r(\mathbb{T}) := \sup\{|z| : z \in \sigma(\mathbb{T})\}$ denote the spectral radius of $\mathbb{T}$.

Proof of Theorem 3.1. We first prove part (c). Let $\mathbb{M}\xi = \lambda\xi$ where $\xi \in L^p$ is positive. Then

$$\lambda\mathbb{E}[\xi(X_0)\phi^*(X_0)] = \mathbb{E}[((\mathbb{M}\xi)(X_0))\phi^*(X_0)] = \mathbb{E}[\xi(X_0)((\mathbb{M}^*\phi^*)(X_0))] = \rho\mathbb{E}[\xi(X_0)\phi^*(X_0)] \qquad (40)$$

by Assumption 3.3. Moreover, $\mathbb{E}[\xi(X_0)\phi^*(X_0)] > 0$ because ξ and ϕ^* are positive. Therefore $\rho = \lambda$, proving (c).

For part (a), let $F = \{\xi \in L^p : \mathbb{M}\xi = \rho\xi\}$. Clearly $F \neq \{0\}$ because $\phi \in F$ by Assumption 3.3. If $\xi \in F$ let $|\xi|$ denote the function given by $|\xi|(x) = |\xi(x)|$.

Claim 1: $\xi \in F$ implies $|\xi| \in F$.

Proof of Claim 1: Since $\mathbb{M}$ is a positive operator (Assumption 3.2(a)), for any $\xi \in F$ we have $\mathbb{M}|\xi| \geq |\mathbb{M}\xi| = |\rho\xi| = \rho|\xi|$ for any $\xi \in F$ which implies $\mathbb{M}|\xi| - \rho|\xi| \geq 0$ a.e.-$[Q]$. On the other hand,

$$\mathbb{E}[\phi^*(X_0)((\mathbb{M}|\xi|)(X_0) - \rho|\xi|(X_0))] = \mathbb{E}[((\mathbb{M}^*\phi^*)(X_0))|\xi|(X_0)] - \rho\mathbb{E}[\phi^*(X_0)|\xi|(X_0)] = 0\,.$$

But $\phi^*(x) > 0$ a.e.-$[Q]$ by Assumption 3.3. Therefore, $\mathbb{M}|\xi| = \rho|\xi|$ whence $|\xi| \in F$, proving Claim 1.

Claim 2: $\xi \in F$ implies $\xi = |\xi|$ a.e.-$[Q]$ or $-\xi = |\xi|$ a.e.-$[Q]$.

Proof of Claim 2: This is trivially true when $\xi = 0$, so consider $\xi \neq 0$. Let $\xi = |\xi|$ on a set of positive Q measure. We prove, by contradiction, that this implies $|\xi| = \xi$. Assume $|\xi| \neq \xi$ on a set of positive Q measure. Then $|\xi| - \xi \geq 0$ a.e.-$[Q]$ and $|\xi| - \xi \neq 0$. Note that $\mathbb{M}(|\xi| - \xi) = \rho(|\xi| - \xi)$ (by Claim 1) and that $\rho \leq r(\mathbb{M})$ (by definition of the spectral radius). Then for any $\lambda > r(\mathbb{M})$ we have

$$\frac{(\rho/\lambda)}{1 - (\rho/\lambda)}(|\xi| - \xi) = \sum_{n \geq 1} \left(\frac{\rho}{\lambda}\right)^n (|\xi| - \xi) = \sum_{n \geq 1} \lambda^{-n}\mathbb{M}^n(|\xi| - \xi) > 0 \text{ a.e.-}[Q]$$

by Assumption 3.2 (Schaefer, 1974, p. 337) whence $|\xi| > \xi$ a.e.-$[Q]$. This contradicts the fact that $\xi = |\xi|$ on a set of positive Q measure. A similar proof shows that if $-\xi = |\xi|$ holds on a set of positive Q measure then $-\xi = |\xi|$, proving Claim 2.

We know by part (c) that if $\zeta \in L^p$ is a positive eigenfunction of $\mathbb{M}$ then $\zeta \in F$. Define the sets $S_+ = \{s \in \mathbb{R} : \zeta \geq s\phi \text{ a.e.-}[Q]\}$ and $S_- = \{s \in \mathbb{R} : \zeta \leq s\phi \text{ a.e.-}[Q]\}$.

Claim 3: S_+, S_- are nonempty, closed, convex, and $\mathbb{R} = S_+ \cup S_-$.

Proof of Claim 3: To prove S_+ and S_- are nonempty, note that we must have $(-\infty, 0] \subseteq S_+$ because ζ is positive and ϕ is positive. Suppose S_- is empty. Then $\zeta > s\phi$ on a set of positive measure for all $s \in (0, \infty]$. But by Claim 2 this implies that $\zeta \geq s\phi$ a.e.-[Q] for all $s \in (0, \infty)$, which is clearly impossible because $\phi > 0$ a.e.-[Q]. Therefore S_- is nonempty.

It is straightforward to verify that S_+ and S_- are convex and closed.

It remains to show $\mathbb{R} = S_+ \cup S_-$. Take any $s \in \mathbb{R}$. Clearly $\zeta - s\phi \in F$. By Claim 2 we know that $\zeta - s\phi \geq 0$ a.e.-[Q] (implying $s \in S_+$) or $\zeta - s\phi \leq 0$ a.e.-[Q] (implying $s \in S_-$) holds. Therefore $\mathbb{R} = S_+ \cup S_-$. This completes the proof of Claim 3.

Claim 3 implies that $S_+ \cap S_-$ must be nonempty. Therefore $S_+ \cap S_- = \{s^*\}$ because $S_+ \cap S_-$ must be a singleton (otherwise $\zeta = s\phi$ and $\zeta = s'\phi$ with $s \neq s'$) and so $\zeta = s^*\phi$ a.e.-[Q], proving (a).

For part (b), a similar argument to (c) shows that ρ is the only eigenvalue of $\mathbb{M}^*$ with a non-negative eigenfunction. The result then follows by similar arguments to the proof of (a). ∎

The following theorem is originally due to Schaefer (1960). The version presented below is Theorem 3.2 on p. 318 of Schaefer (1999).

Theorem D.1 *Let E be an ordered real Banach space with positive cone C, and suppose that u is an irreducible positive endomorphism whose spectral radius r is a pole of the resolvent [of u]. Then:*

i. $r > 0$ and r is a pole of order 1

ii. there exist positive eigenvectors, pertaining to r, of u and u' [its adjoint]. Each positive eigenvector for r is quasi-interior to C, and each positive eigenvector for u' is a strictly positive linear form

iii. each of the following assumptions implies that the multiplicity of $d(r)$ of r is 1: (a) C has non-empty interior, (b) $d(r)$ is finite, (c) E is a Banach lattice.

Proof of Theorem 3.2. For part (a), first note that $\mathbb{M}^n$ is compact so $\mathbb{M}$ has discrete spectrum whose only limit point is zero and any nonzero eigenvalue of $\mathbb{M}$ is a pole of the resolvent of $\mathbb{M}$ (Dunford and Schwartz, 1958, Theorem 6, p. 579). Assumptions 3.1 and 3.2(a) imply $\mathbb{M}$ is a positive operator, so $r(\mathbb{M}) \in \sigma(\mathbb{M})$ (Schaefer, 1999, p. 312). Assumptions 3.1 and 3.2 also imply that $r(\mathbb{M}) > 0$ (Schaefer, 1974, p. 337). Therefore, $r(\mathbb{M})$ is a pole of the resolvent of $\mathbb{M}$. Further, Assumption 3.2(b) implies $\mathbb{M}$ is irreducible. Existence of ϕ and ϕ^* follows from Theorem D.1(ii). Note $\rho = r(\mathbb{M})$ because only one eigenvalue of $\mathbb{M}$ has a positive eigenfunction (Theorem 3.1(c)).

For part (b) we know that $\rho = r(\mathbb{M})$ is a simple eigenvalue by Theorem D.1(iii). Further, ρ is isolated because the only limit point of $\sigma(\mathbb{M})$ is zero. ∎

Lemma D.1 *Let the Assumptions of Theorem 3.3 hold. Then there exists* $\epsilon > 0$ *such that* $|\lambda| \leq (1-\epsilon)\rho$ *for all* $\lambda \in \sigma(\mathbb{M}) \setminus \{\rho\}$.

Proof of Lemma D.1. It follows from Theorem 3.2(b) that there exists $\epsilon > 0$ such that $|\lambda| < (1-\epsilon)\rho$ for all $\lambda \in \{z \in \sigma(\mathbb{M}) : |z| \neq \rho\}$. It remains to show that ρ is the unique element of $\sigma(\mathbb{M})$ with modulus ρ. Since $\mathbb{M}$ is irreducible (by Assumption 3.5) it follows by Theorem V.5.4 of Schaefer (1974) that $\mathcal{S} := \{z \in \sigma(\mathbb{M}) : |z| = \rho\}$ consists of first-order poles of the resolvent of $\mathbb{M}$. As $\mathbb{M}$ has a discrete spectrum whose only limit point is zero there must be only finitely many elements of $\mathcal{S}$. Moreover, by power compactness of $\mathbb{M}$, each of the first-order poles of the resolvent of $\mathbb{M}$ are eigenvalues of $\mathbb{M}$ (Dunford and Schwartz, 1958, Theorem 6, p. 579). However, Assumption 3.5 implies that every eigenvalue λ of $\mathbb{M}$ with $\lambda \neq \rho$ has modulus $|\lambda| < \rho$ (Schaefer, 1974, Theorem 6.6, p. 337). Therefore $\mathcal{S} = \{\rho\}$. ■

Proof of Theorem 3.3. Let $\overline{\mathbb{M}} = \rho^{-1}\mathbb{M}$, whence $\sigma(\overline{\mathbb{M}}) = \rho^{-1}\sigma(\mathbb{M})$ by the spectral mapping theorem. In particular, $r(\overline{\mathbb{M}}) = 1$. Note that, by construction, $(\phi \otimes \phi^*)$ is the spectral projection of $\overline{\mathbb{M}}$ corresponding to the eigenvalue 1.

Consider the operator $\mathbb{V} = \overline{\mathbb{M}} - (\phi \otimes \phi^*)$. Lemma D.1 implies that $\mathbb{V}$ has spectral radius $r(\mathbb{V}) \leq 1 - \epsilon$ for some $\epsilon > 0$. Also note that

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

where the second equality is because $\overline{\mathbb{M}}$ and $(\phi \otimes \phi^*)$ commute and $(\phi \otimes \phi^*)$ is a projection.

Let $\|\cdot\|$ denote the L^p operator norm given by $\|A\|^p = \sup\{\mathbb{E}[|A\psi(X_0)|^p : \mathbb{E}[|\psi(X_0)|^p] \leq 1\}$. By the Gelfand formula (Dunford and Schwartz, 1958, p. 567),

$$\lim_{n\to\infty} \|\mathbb{V}^n\|^{1/n} = r(\mathbb{V}) \leq 1 - \epsilon \,. \tag{41}$$

Let $\{n_k : k \geq 1\} \subseteq \mathbb{N}$ be the maximal subset of $\mathbb{N}$ for which $\|\mathbb{V}^{n_k}\| > 0$. If this subsequence is finite then the proof is complete. If this subsequence is infinite, then by expression (41),

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

Therefore, there exists a finite positive constant c such that for all n_k large enough,

$$\log \|\mathbb{V}^{n_k}\| \leq -cn_k$$

and the result follows. ■

Proof of Theorem 4.1. For parts (a) and (b), we first bound

$$
\begin{aligned}
|\widehat{\rho} - \rho| &\leq |\widehat{\rho} - \rho_K| + |\rho_K - \rho| \\
\|\widehat{\phi} - \phi\| &\leq \|\widehat{\phi} - \phi_K\| + \|\phi_K - \phi\| .
\end{aligned}
$$

The desired results now follow by parts (d) and (e) of Lemmas C.1 and C.2.

For part (c), we have

$$
\left\| \frac{\widehat{\phi}^*}{\|\widehat{\phi}^*\|} - \frac{\phi^*}{\|\phi^*\|} \right\| \leq \left\| \frac{\widehat{\phi}^*}{\|\widehat{\phi}^*\|} - \frac{\phi_K^+}{\|\phi_K^+\|} \right\| + \left\| \frac{\phi_K^+}{\|\phi_K^+\|} - \frac{\phi^*}{\|\phi^*\|} \right\|
$$

and the result follows by Lemmas C.1(h) and C.2(f). ■

Proof of Corollary 4.1. Part (a) follows immediately by continuity of $\log(\cdot)$ on a neighborhood of $\rho > 0$. Part (b) then follows from part (a) by the triangle inequality, since:

$$
|\widehat{L} - L| \leq |y - y| + \left| n^{-1} \sum_{t=0}^{n-1} \log m(X_t, X_{t+1}) - \mathbb{E}[\log m(X_0, X_1)] \right|
$$

with the obvious modification in Cases 2 or 3. ■

Proof of Corollary 4.2. Condition (iii) implies $\mathbb{M}$ is compact because $\mathbb{M}$ maps L^2 into $\Lambda^{p_0}(\mathcal{X})$ and $\Lambda^{p_0}(\mathcal{X})$ is compactly embedded in $L^2(\mathcal{X}, \mathscr{X}, Leb)$ (Triebel, 2006, Proposition 4.6, p. 197) which, by condition (ii), is equivalent to $L^2(\mathcal{X}, \mathscr{X}, Q)$. Therefore Assumption 4.1 holds.

Let $\|\cdot\|_{p_0}$ denote the norm on $\Lambda^{p_0}(\mathcal{X})$ (see Section 2.3.1 of Chen (2007)). We may bound $\|\Pi_K \mathbb{M} - \mathbb{M}\|$ using the factorization

$$
\begin{aligned}
\|\Pi_K \mathbb{M} - \mathbb{M}\| &\leq \sup_{\psi \in \Lambda^{p_0}(\mathcal{X}) : \|\psi\|_{p_0} \neq 0} \frac{\|\Pi_K \psi - \psi\|}{\|\psi\|_{p_0}} \sup_{\psi \in L^2 : \|\psi\| \neq 0} \frac{\|\mathbb{M}\psi\|_{p_0}}{\|\psi\|} \\
&= O(K^{-p_0/d}) \times \text{const}
\end{aligned}
$$

where the $O(K^{-p_0/d})$ term is by Theorem 12.8 of Schumaker (2007) (under conditions (i)(ii)(vi)) and the constant term is by condition (iii). This verifies Assumption 4.2(a). Assumption 4.2(b) is satisfied with $\delta_K = O(K^{-p/d})$ by Theorem 12.8 of Schumaker (2007) (under conditions (i)(ii)(iv)(vi)).

For Assumption 4.3, note that the minimum eigenvalue of $\mathbf{G}$ is uniformly bounded away from zero and $\xi_K = O(\sqrt{K})$ under conditions (i)(ii)(vi) (see, e.g., Newey (1997)). As the data are exponentially rho-mixing (condition (vii)), Lemma C.4 implies $\|\widehat{\mathbf{G}}^{-1}\widehat{\mathbf{M}} - \mathbf{G}^{-1}\mathbf{M}\|_{\mathbf{G}} = O_p(K^{(r+1)/r}/\sqrt{n})$ and $\eta_{n,K} = O(K^{(r+2)/(2r)}/\sqrt{n})$. Choosing $K \asymp n^{\frac{rd}{2rp+(2+r)d}}$ sets $\delta_K \asymp \eta_{n,K}$. Further, $K^{(r+2)/(2r)}/\sqrt{n} = o(1)$ holds for this choice of K provided $p > d/2$. This verifies Assumption 4.3(a)(b). ■

Proof of Theorem 4.2. First consider the limiting distribution for $\widehat{\rho}$. By Lemmas C.7 and C.1(d),

$$\begin{aligned}\widehat{\rho}-\rho &= \frac{1}{n}\sum_{t=0}^{n-1}\{\phi_K^+(X_t)m(X_t,X_{t+1})\phi_K(X_{t+1})-\rho_K\phi_K^+(X_t)\phi_K(X_t)\}\\ &\quad +\rho_K-\rho+O_p((\eta_{n,K}\vee\bar{\eta}_{n,K,1})\times(\bar{\eta}_{n,K,1}\vee\bar{\eta}_{n,K,2}))\\ &= \frac{1}{n}\sum_{t=0}^{n-1}\{\phi^*(X_t)m(X_t,X_{t+1})\phi(X_{t+1})-\rho\phi^*(X_t)\phi(X_t)\}+o_p(n^{-1/2})\end{aligned}$$

where the $o_p(n^{-1/2})$ term is by Assumption 4.4(a)(b)(c), proving (27). Part (a) is then immediate by a central limit theorem for stationary and ergodic martingale differences (Billingsley, 1961) under Assumption 4.4(d). Part (b) follows directly from part (a) via the delta method. For part (c), by continuity of log on a neighborhood of $\rho>0$ and (27) we have

$$\log\widehat{\rho}-\log\rho=\frac{1}{n}\sum_{t=0}^{n-1}\rho^{-1}\psi_\rho(X_t,X_{t+1})+o_p(n^{-1/2})$$

and so

$$\widehat{L}-L=\frac{1}{n}\sum_{t=0}^{n-1}\psi_L(X_t,X_{t+1})+o_p(n^{-1/2})$$

and the result follows by Assumption 4.4(e). ■

We first state a further assumption required to prove Theorem 4.3. Let $P^n(\cdot|x)$ denote the conditional measure of X_n given $X_0=x$.

Assumption D.1 *(a) $P^1(\cdot|x)$ has density $f(\cdot|x)$ with respect to Lebesgue measure*

(b) there exists a probability measure Θ on $(\mathcal{X},\mathscr{X})$ and $n\in\mathbb{N}$ such that $P^n(A|x)\geq\Theta(A)$ for all $x\in\mathcal{X}$ and $A\in\mathscr{X}$

(c) $\mathbb{E}[\log m(X_0,X_1)^2]<\infty$.

Assumption D.1(a) is useful for characterizing the tangent space. Assumption D.1(b) implies that $\{X_t\}$ is uniformly ergodic and phi-mixing (Doukhan, 1994, Theorem 1, p. 88). Finally, Assumption D.1(c) just ensures that a component of V_L is well defined.

Proof of Theorem 4.3. We follow arguments in Bickel and Kwon (2001) and Greenwood, Schick, and Wefelmeyer (2001). Let $\mathcal{B}$ denote the space of all bounded measurable $f:\mathcal{X}\times\mathcal{X}\to\mathbb{R}$, and let $\mathcal{T}=\{h\in\mathcal{B}:\mathbb{E}[h(X_0,X_1)|X_0=x]=0$ for all $x\in\mathcal{X}\}$. Let $f(\cdot|x)$ denote the conditional density of X_1 given $X_0=x$. For any $h\in\mathcal{T}$ there is $N_h\in\mathbb{N}$ such that $\sup_{(x,y)\in\mathcal{X}^2}|h(x,y)|\leq n^{1/2}$ for all $n\geq N_h$. By Assumption D.1(a) we can define $f_{n,h}(y|x):=f(y|x)\{1+n^{-1/2}h(y,x)\}$ which is non-negative for all $n \geq N_h$ and

$$\int f_{n,h}(y|x) = \int_{\mathcal{X}} (1 + n^{-1/2} h(y,x)) f(y|x) \, \mathrm{d}y = 1 + n^{-1/2} \mathbb{E}[h(X_0, X_1)|X_0 = x] = 1 \, .$$

Therefore, for every $h \in \mathcal{T}$, $f_{n,h}$ is a conditional density for all $n \geq N_h$.

Let $\mathbb{P}_{n,h}$ denote the distribution of $\{X_0, X_1, \ldots, X_n\}$ when the conditional distribution of X_{t+1} given X_t is $f_{n,h}$, and let $\mathbb{P}_{n,0}$ denote the distribution of $\{X_0, X_1, \ldots, X_n\}$ under the true conditional density $f(y|x)$. A version of local asymptotic normality is known to obtain, i.e.

$$\log \frac{\mathrm{d}\mathbb{P}_{n,h}}{\mathrm{d}\mathbb{P}_{n,0}} = \frac{1}{\sqrt{n}} \sum_{t=0}^{n-1} h(X_t, X_{t+1}) - \frac{1}{2} \mathbb{E}[h(X_0, X_1)^2] + o_{\mathbb{P}_{n,0}}(1)$$

(see Greenwood et al. (2001)) where $n^{-1/2} \sum_{t=0}^{n-1} h(X_t, X_{t+1}) \to_d N(0, \mathbb{E}[h(X_0, X_1)^2])$ by a CLT stationary and ergodic martingale differences (Billingsley, 1961).

For any $h \in \mathcal{T}$ define $\mathbb{M}_{n,h} : L^2 \to L^2$ by

$$\mathbb{M}_{n,h} \psi(x) = \int_{\mathcal{X}} \mathcal{K}_{\mathbb{M}}(x,y)(1 + n^{-1/2} h(x,y)) \psi(y) \, \mathrm{d}Q(y) \, .$$

Therefore we may write:

$$(\mathbb{M}_{n,h} - \mathbb{M}) \psi(x) = n^{-1/2} \int_{\mathcal{X}} \mathcal{K}_{\mathbb{M}}(x,y) h(x,y) \psi(y) \, \mathrm{d}Q(y)$$

where $\psi(x) \mapsto \int_{\mathcal{X}} \mathcal{K}_{\mathbb{M}}(x,y) h(x,y) \psi(y) \, \mathrm{d}Q(y)$ is a bounded operator on L^2 because $\mathbb{M}$ is a bounded operator on L^2 and $h \in \mathcal{B}$. Thus $\|\mathbb{M}_{n,h} - \mathbb{M}\| = O(n^{-1/2})$ for each $h \in \mathcal{T}$. By the proof of Lemma C.2, for all n sufficiently large the operator $\mathbb{M}_{n,h}$ has one simple eigenvalue, say $\rho_{n,h}$, in the interval $[\rho - \epsilon, \rho + \epsilon]$. Further, by similar arguments to the proof of Lemma D.10,

$$\begin{aligned} \rho_{n,h} - \rho &= \mathbb{E}[\phi^*(X_0)(\mathbb{M}_{n,h} - \mathbb{M})\phi(X_0)] + o(\|\mathbb{M}_{n,h} - \mathbb{M}\|) \\ \Rightarrow \sqrt{n}(\rho_{n,h} - \rho) &= \mathbb{E}[\phi^*(X_0) m(X_0, X_1) \phi(X_0) h(X_0, X_1)] + o(1) \end{aligned}$$

because $\|\mathbb{M}_{n,h} - \mathbb{M}\| = O(n^{-1/2})$. The gradient of ρ is therefore $\phi^*(x) m(x,y) \phi(y)$ and its projection onto the closure of $\mathcal{T}$ (under the seminorm $h \mapsto \mathbb{E}[h(X_0, X_1)^2]^{1/2}$) is

$$\phi^*(x) m(x,y) \phi(y) - \mathbb{E}[\phi^*(X_0) m(X_0, X_1) \phi(X_1) | X_0 = x] = \psi_\rho(x,y) \, .$$

Therefore ψ_ρ is the efficient influence function and $V_\rho = \mathbb{E}[\psi_\rho(X_0, X_1)^2]$ is the efficiency bound for ρ, which is attained by $\widehat{\rho}$. The result for y follows by continuity.

As shown in Example 1 of Greenwood et al. (2001),

$$\begin{aligned}\psi_m(x,y) &= \log m(x,y) - \mathbb{E}[\log m(X_0,X_1)|X_0=x] \\ &\quad + \sum_{t=1}^{\infty}\left(\mathbb{E}[\log m(X_t,X_{t+1})|X_1=y] - \mathbb{E}[\log m(X_t,X_{t+1})|X_0=x]\right) \qquad (42)\end{aligned}$$

is the efficient influence function for $\mathbb{E}[\log m(X_0,X_1)]$. The efficient influence function for L is therefore $\widetilde{\psi}_L(x,y) = \rho^{-1}\psi_\rho(x,y) - \psi_m(x,y)$ and $\mathbb{E}[\widetilde{\psi}_L(X_0,X_1)^2]$ is the efficiency bound for L. It may be verified using the telescoping property of the sum in (42) that $V_L = \mathbb{E}[\widetilde{\psi}_L(X_0,X_1)^2]$. Therefore, $\widehat{L}$ is semiparametrically efficient. ■

D.2 Proof of results in Appendix B

In the following proofs we use the Gaussian integration formula

$$\int_{\mathbb{R}^N} \exp\left\{b'x + -\frac{1}{2}x'Ax\right\}\mathrm{d}x = \sqrt{\frac{(2\pi)^N}{|A|}}\exp\left\{\frac{1}{2}b'A^{-1}b\right\}$$

where $A \in \mathbb{R}^{N\times N}$ is positive definite and $b \in \mathbb{R}^N$.

Proof of Lemma B.1. First consider ϕ. Conjecture a solution of the form $\phi(x) = \beta_0\exp\{\beta_1'x\}$. Substituting into $\mathbb{M}\phi = \rho\phi$ yields

$$\begin{aligned}\rho\beta_0\exp\{\beta_1'x_t\} &= \alpha_0\beta_0\mathbb{E}[\exp\{\alpha_1'x_t + \alpha_2'\sigma\epsilon_{t+1} + \beta_1'(Ax_t + \sigma\epsilon_{t+1})\}|X_t = x_t] \\ &= \alpha_0\beta_0\exp\left\{\frac{1}{2}(\alpha_2'+\beta_1')\Sigma(\alpha_2+\beta_1)\right\}\exp\{(\alpha_1'+\beta_1'A)x_t\}\,.\end{aligned}$$

Equating coefficients of x_t on the left- and right-hand sides yields $\beta_1' = \alpha_1'(I-A)^{-1}$.

Using the same ansatz for ϕ^*, we obtain

$$\begin{aligned}\rho\exp\{\beta_1^{*\prime}x_{t+1}\} &= \alpha_0\mathbb{E}[\exp\{\alpha_1'X_t + \alpha_2'(x_{t+1} - AX_t) + \beta_1^{*\prime}X_t\}|X_{t+1} = x_{t+1}] \\ &= \alpha_0\exp\{\alpha_2'x_{t+1}\}\mathbb{E}[\exp\{\Gamma'X_t\}|X_{t+1} = x_{t+1}]\end{aligned}$$

where $\Gamma = \alpha_1 - A'\alpha_2 + \beta_1^*$. The conditional density for X_t given $X_{t+1} = x_{t+1}$ is

$$\begin{aligned}&f(X_t|X_{t+1} = x_{t+1}) \\ &= \frac{1}{\sqrt{(2\pi)^N|\Sigma|}}\exp\left\{-\frac{1}{2}\left(x_{t+1}'(\Sigma^{-1} - V^{-1})x_{t+1} + X_t'(A'\Sigma^{-1}A + V^{-1})X_t\right) + X_t'A'\Sigma^{-1}x_{t+1}\right\}\,.\end{aligned}$$

It follows that

$$\begin{aligned}
&\mathbb{E}[\exp\{\Gamma' X_t\}|X_{t+1} = x_{t+1}] \\
&= \frac{\exp\left\{\frac{1}{2}\Gamma'(A'\Sigma^{-1}A + V^{-1})^{-1}\Gamma\right\}}{\sqrt{|\Sigma||A'\Sigma^{-1}A + V^{-1}|}} \times \exp\left\{\Gamma'(A'\Sigma^{-1}A + V^{-1})^{-1}A'\Sigma^{-1}x_{t+1}\right\} \\
&\quad \times \exp\left\{-\frac{1}{2}x'_{t+1}(\Sigma^{-1} - V^{-1} - \Sigma^{-1}A(A'\Sigma^{-1}A + V^{-1})^{-1}A'\Sigma^{-1})x_{t+1}\right\}. \qquad (43)
\end{aligned}$$

Cconsider the quadratic in x_{t+1} in (43), which must be zero if the ansatz for ϕ^* is correct. Note that $\sigma' V^{-1}\sigma = (I + \widetilde{A}V\widetilde{A}')^{-1} = I + \sum_{j\geq 1}(-1)^j(\widetilde{A}V\widetilde{A}')^j$ where $\widetilde{A} = \sigma^{-1}A$. Therefore,

$$\begin{aligned}
&I - \sigma' V^{-1}\sigma - \widetilde{A}(\widetilde{A}'\widetilde{A} + V^{-1})^{-1}\widetilde{A}' \\
&= -\sum_{j=1}^{\infty}(-1)^j(\widetilde{A}V\widetilde{A}')^j - \widetilde{A}V^{-1/2}(V^{1/2}\widetilde{A}'\widetilde{A}V^{1/2} + I)^{-1}V^{-1/2}\widetilde{A}' \\
&= -\sum_{j=1}^{\infty}(-1)^j(\widetilde{A}V\widetilde{A}')^j - \widetilde{A}V^{1/2}\left(\sum_{j=0}^{\infty}(-1)^j(V^{1/2}\widetilde{A}'\widetilde{A}V^{1/2})^j\right)V^{1/2}\widetilde{A}' \quad = \quad 0
\end{aligned}$$

as required. Moreover, by the Woodbury formula for determinants we obtain:

$$|\Sigma||A'\Sigma^{-1}A + V^{-1}| \quad = \quad |\Sigma||V^{-1}||\Sigma^{-1}||\Sigma + AVA'| \quad = \quad 1$$

because $V = \Sigma + AVA'$. Substituting (43) into $\mathbb{M}^*\phi^* = \rho\phi^*$ yields

$$\rho\exp\{\beta_1^{*\prime}x_{t+1}\} = \alpha_0\exp\left\{\alpha_2' x_{t+1} + \frac{1}{2}\Gamma'(A'\Sigma^{-1}A + V^{-1})^{-1}\Gamma + \Gamma'(A'\Sigma^{-1}A + V^{-1})^{-1}A'\Sigma^{-1}x_{t+1}\right\}.$$

Using the relation $(A'\Sigma^{-1}A + V^{-1})^{-1} = V - VA'V^{-1}AV$ and equating the coefficients of x_{t+1}:

$$\begin{aligned}
\beta_1^{*\prime} &= \alpha_2' + [\alpha_1' - \alpha_2'A + \beta_1^{*\prime}](A'\Sigma^{-1}A + V^{-1})^{-1}A'\Sigma^{-1} \\
&= \alpha_2' + [\alpha_1' - \alpha_2'A + \beta_1^{*\prime}](VA'V^{-1}) \\
&= [\alpha_1'VA' + \alpha_2'\Sigma](I - A')^{-1}V^{-1}.
\end{aligned}$$

It remains to check that ϕ^* corresponds to the eigenvalue ρ. By (43) and the expression for ρ it is enough to show $\Gamma'(A'\Sigma^{-1}A + V^{-1})^{-1}\Gamma = (\alpha_2' + \alpha_1'(I - A)^{-1})\Sigma(\alpha_2 + (I - A')^{-1}\alpha_1)$ where

$$\begin{aligned}
\Gamma' &= \alpha_1' - \alpha_2'A + [\alpha_1'VA' + \alpha_2'\Sigma](I - A')^{-1}V^{-1} \\
&= [\alpha_1'(I - A)^{-1} + \alpha_2'](I - A)V(I - A')^{-1}V^{-1}.
\end{aligned}$$

It therefore suffices to prove $(I - A)V(I - A')^{-1}V^{-1}(A'\Sigma^{-1}A + V^{-1})^{-1}V^{-1}(I - A)^{-1}V(I - A') = \Sigma$

or equivalently $V^{-1}(I-A)[V+VA'\Sigma^{-1}AV](I-A')V^{-1} = (I-A')\Sigma^{-1}(I-A')$ which holds because:

$$\begin{aligned} V^{-1}(I-A)[V+VA'\Sigma^{-1}AV](I-A')V^{-1} &= V^{-1}[V(I-A')\Sigma^{-1}(I-A)V]V^{-1} \\ &= (I-A')\Sigma^{-1}(I-A)\,. \end{aligned}$$

Moreover, the joint density of (X_t, X_{t+1}) and the unconditional density of X_t both exist and are strictly positive. The SDF is also strictly positive. Therefore Assumption 3.5 holds. To check boundedness of $\mathbb{M}$ on L^p we use the Schur test. Consider the test functions $\psi(x) = \exp(a'x)$ and $\zeta(x) = \exp(b'x)$. By similar calculations to the above, for $p \in (1,\infty)$ and $q = (1-p^{-1})^{-1}$ we have

$$\begin{aligned} \mathbb{M}\psi^q(x) &= c_0 \exp((\alpha_1' + qa'A)x) \\ \mathbb{M}^*\zeta^p(x) &= c_0^* \exp((\alpha_2' + [\alpha_1' - \alpha_2'A + pb'][V - VA'V^{-1}AV]A'\Sigma^{-1})x) \end{aligned}$$

where c_0, c_0^* are finite positive constants. For the Schur test to hold we need $\mathbb{M}\psi^q \leq C_0^*\zeta^q$ and $\mathbb{M}^*\zeta^p \leq C_0^*\psi^p$ for positive constants C_0, C_0^*. Equivalently, we require a solution to the system:

$$\begin{aligned} qb &= \alpha_1 + qA'a \\ pa &= \alpha_2 + \Sigma^{-1}A[V - VA'V^{-1}AV][\alpha_1 - A'\alpha_2 + pb]\,. \end{aligned}$$

Substituting qb into pa and letting δ denote the sum of all terms depending on α_1 and α_2 (and not on a), we have:

$$\begin{aligned} a &= \delta + \Sigma^{-1}A[V - VA'V^{-1}AV]A'a \\ &= \delta + \Sigma^{-1}(V-\Sigma)a - \Sigma^{-1}(V-\Sigma)V^{-1}(V-\Sigma)a \\ &= \delta + (\Sigma^{-1}V - I)a - (\Sigma^{-1}V - I)(I - V^{-1}\Sigma)a \quad = \quad \delta + (I - V^{-1}\Sigma)a \end{aligned}$$

therefore $a = \Sigma^{-1}V\delta$ and $b = q^{-1}\alpha_1 + A'\Sigma^{-1}V\delta$. ■

Proof of Lemma B.2. First consider ϕ. We substitute the ansatz $\phi(x) = \beta_0 \exp\{\beta_1'x\}$ into $\mathbb{M}\phi = \rho\phi$ and apply $\Gamma_1 + \frac{1}{2}\Gamma_2\Sigma\Gamma_2' = 0$ to obtain:

$$\begin{aligned} \rho\exp\{\beta_1'x_t\} &= \alpha_0 \exp\left\{(\alpha_1' + \beta_1'A)x_t + x_t'\Gamma_1 x_t + \frac{1}{2}(\alpha_2' + \beta_1' + x_t'\Gamma_2)\Sigma(\alpha_2 + \beta_1 + \Gamma_2'x_t)\right\} \\ &= \alpha_0 \exp\left\{(\alpha_1' + \beta_1'A)x_t + \frac{1}{2}(\alpha_2' + \beta_1')\Sigma(\alpha_2 + \beta_1) + (\alpha_2' + \beta_1')\Sigma\Gamma_2'x_t\right\}. \end{aligned}$$

The expressions for ϕ and ρ follow by equating coefficients of x_t on the left- and right-hand sides.

For ϕ^*, we substitute the ansatz $\phi^*(x) = \beta_0^* \exp\{\beta_1^{*\prime}x - x'B^*x\}$ into $\mathbb{M}^*\phi^* = \rho^*\phi^*$. Calculating $\mathbb{E}[\cdot|X_{t+1} = x_{t+1}]$ using the time-reversed conditional distribution in the proof of Lemma B.1 and applying the condition $\Gamma_1 + \frac{1}{2}\Gamma_2\Sigma\Gamma_2' = 0$, we obtain:

$$
\begin{aligned}
&\rho^* \exp\{\beta_1^{*\prime} x_{t+1} - x_{t+1}' B^* x_{t+1}\} \\
&\quad = \mathbb{E}[\alpha_0 \exp\{(\alpha_1' + \beta_1^{*\prime})X_t + \alpha_2'(x_{t+1} - AX_t) + X_t'(\Gamma_1 - B^*)X_t + X_t'\Gamma_2(x_{t+1} - AX_t)\}|X_{t+1} = x_{t+1}] \\
&\quad = \alpha_0 \exp\{\alpha_2' x_{t+1}\}\mathbb{E}[\exp\{(\varpi' + x_{t+1}'\Gamma_2')X_t + X_t'(-\tfrac{1}{2}(\Gamma_2\Sigma\Gamma_2' + \Gamma_2 A + A'\Gamma_2') - B^*)X_t\}|X_{t+1} = x_{t+1}] \\
&\quad = \frac{\alpha_0 \exp\{\alpha_2' x_{t+1} - \frac{1}{2}x_{t+1}'(\Sigma^{-1} - V^{-1})x_{t+1}\}}{\sqrt{|\Sigma||D|}} \exp\left\{\frac{1}{2}(x_{t+1}'S' + \varpi')D^{-1}(Sx_{t+1} + \varpi)\right\}
\end{aligned}
$$

where $\varpi = \alpha_1 - A'\alpha_2 + \beta_1^*$, $V = \sum_{j\geq 0} A^j \Sigma A'^j$, $S = A'\Sigma^{-1} + \Gamma_2$, $D = S\Sigma S' + V^{-1} + 2B^*$, $\beta_1^* = (I - S'D^{-1})^{-1}(\alpha_2 + SD^{-1}(\alpha_1 - A'\alpha_2))$, and B^* solves

$$0 = (2B^* + V^{-1}) - \Sigma^{-1} + S'(S\Sigma S' + (2B^* + V^{-1}))^{-1}S \tag{44}$$

which implies

$$D = S\Sigma S' + \Sigma^{-1} - S'D^{-1}S\,. \tag{45}$$

Therefore, $\varpi = (I - S'D^{-1})^{-1}(\alpha_1 + (I - A')\alpha_2)$. After some algebra, we obtain:

$$
\begin{aligned}
\rho &= \alpha_0 \exp\left\{\frac{1}{2}(\alpha_1' + \alpha_2'(I - A))(I - \Sigma S')^{-1}\Sigma(I - S\Sigma)^{-1}(\alpha_1 + (I - A')\alpha_2)\right\} \\
\rho^* &= \frac{\alpha_0}{\sqrt{|\Sigma||D|}} \exp\left\{\frac{1}{2}(\alpha_1' + \alpha_2'(I - A))(I - D^{-1}S)^{-1}D^{-1}(I - S'D^{-1})^{-1}(\alpha_1 + (I - A')\alpha_2)\right\}.
\end{aligned}
$$

We need $\rho = \rho^*$ to verify Assumption 3.3. Comparing the expressions for ρ and ρ^*, we see $\rho = \rho^*$ provided that both $(I - \Sigma S')^{-1}\Sigma(I - S\Sigma)^{-1} = (I - D^{-1}S)^{-1}D^{-1}(I - S'D^{-1})^{-1}$ and $|\Sigma||D| = 1$ hold. The first condition follows from taking the inverse of both its sides and using (45). Further, by (45) and the properties of determinants of block matrices we have

$$|\Sigma^{-1}|\begin{vmatrix} D & S \\ S' & \Sigma^{-1} \end{vmatrix} = |D - S\Sigma S'| = |\Sigma^{-1} - S'D^{-1}S| = |D|\begin{vmatrix} D & S \\ S' & \Sigma^{-1} \end{vmatrix}.$$

Therefore, $|\Sigma||D| = 1$ and so $\rho = \rho^*$.

Now, $\phi^* \in L^q$ provided $\lambda_{\max}(-qB^* - \frac{1}{2}V^{-1}) < 0$ or, equivalently, that $2B^* + V^{-1} > p^{-1}V^{-1}$ where $q^{-1} + p^{-1} = 1$ (the inequalities are in the sense of positive definite matrices). Rewrite (44) in terms of $\Omega := 2B^* + V^{-1}$ to give

$$\Omega = \Sigma^{-1} - S'(S\Sigma S' + \Omega)^{-1}S\,. \tag{46}$$

Consider the symmetric Stein equation

$$0 = (S\Sigma)'\Omega^{-1}(S\Sigma) - \Omega^{-1} + \Sigma\,. \tag{47}$$

There exists a unique positive definite solution for Ω^{-1} because all eigenvalues of $S\Sigma = A' + \Gamma_2\Sigma$ lie inside the unit circle and Σ is positive definite (Lancaster and Tismenetsky, 1985, Theorem 1,

p. 451).[25] We may rearrange (47) to obtain

$$\begin{aligned}\Omega &= (\Sigma + \Sigma S'\Omega^{-1}S\Sigma)^{-1} \\ &= \Sigma^{-1}(\Sigma^{-1} + S'\Omega^{-1}S)^{-1}\Sigma^{-1} \\ &= \Sigma^{-1}(\Sigma - \Sigma S'(\Omega + S\Sigma S')^{-1}S\Sigma)\Sigma^{-1}\end{aligned}$$

which implies that Ω is the unique positive definite solution of (46). Positive definiteness of Ω implies that $\Omega = 2B^* + V^{-1} > p^{-1}V^{-1}$ for all sufficiently large $p > 1$.

Assumption 3.5 holds by the same logic as the proof of Lemma B.1. To check boundedness of $\mathbb{M}$ on L^p we use the Schur test. Consider the test functions $\psi(x) = \exp(a'x)$ and $\zeta(x) = \exp(f'x - x'Fx)$. By similar calculations to the above, for $p \in (1,\infty)$ and $q = (1-p^{-1})^{-1}$ we have

$$\begin{aligned}\mathbb{M}\psi^q(x) &= c_0 \exp([\alpha_1' + qa'A + (\alpha_2' + qa')\Sigma\Gamma_2']x) \\ \mathbb{M}^*\zeta^p(x) &= c_0^* \exp\Big((\alpha_2' + (\alpha_1' - A'\alpha_2 + pf')[S\Sigma S' + V^{-1} + 2pF]^{-1}S)x \\ &\qquad + \frac{1}{2}x'[S'(S\Sigma S' + V^{-1} + 2pF)^{-1}S + V^{-1} - \Sigma^{-1}]x\Big)\end{aligned}$$

where c_0, c_0^* are finite positive constants. For the Schur test to hold we need $\mathbb{M}\psi^q \leq C_0^*\zeta^q$ and $\mathbb{M}^*\zeta^p \leq C_0^*\psi^p$ for positive constants C_0, C_0^*. These inequalities will hold if we can choose F such that F is negative definite and $S'(S\Sigma S' + V^{-1} + 2pF)^{-1}S + V^{-1} - \Sigma^{-1}$ is negative definite. Taking $F = -cI$ we see that $S'(S\Sigma S' + V^{-1} - 2pcI)^{-1}S + V^{-1} - \Sigma^{-1} < 0$ for sufficiently large c (because $V^{-1} - \Sigma^{-1} < 0$). Further, this choice of F is clearly negative definite. ■

Proof of Lemma B.3. Substituting the ansatz $\phi(g, v) = \beta_0 \exp\{\beta_1 g + \beta_2 v\}$ into $\mathbb{M}\phi = \rho\phi$:

$$\begin{aligned}&\rho \exp\{\beta_1 g_t + \beta_2 v_t\} \\ &\quad = \alpha_0 \exp\left\{(\alpha_1 + \beta_1)(1-\kappa)\bar{g} + (\alpha_1 + \beta_1)\kappa g_t + \left(\alpha_2 + \frac{1}{2}(\alpha_1 + \beta_1 + \alpha_4)^2\sigma^2\right)v_t\right\} \\ &\qquad \times \exp\left\{\varphi_v(\alpha_3 + \beta_2)(1 - (\alpha_3 + \beta_2)c_v)^{-1}v_t - \delta_v \log(1 - (\alpha_3 + \beta_2)c_v)\right\}.\end{aligned}$$

Collecting coefficients, we have:

$$\begin{aligned}\rho &= \alpha_0 \exp\{\alpha_1\bar{g} - \delta_v \log(1 - (\alpha_3 + \beta_2)c_v)\} \\ \beta_1 &= \frac{\alpha_1\kappa}{1-\kappa} \\ \beta_2 &= \alpha_2 + \frac{1}{2}(\alpha_1 + \beta_1 + \alpha_4)^2\sigma^2 + \frac{\varphi_v(\alpha_3 + \beta_2)}{1 - (\alpha_3 + \beta_2)c_v}.\end{aligned} \tag{48}$$

[25] In fact, the requirement that all eigenvalues of $A' + \Gamma_2\Sigma$ lie inside the unit circle is also necessary for existence of a positive definite solution for Ω^{-1}, and therefore for positive definiteness of $\Omega = 2B^* + V^{-1}$ (Lancaster and Tismenetsky, 1985, Theorem 1, p. 451).

Substituting $\phi^*(g,v) = \beta_0 \exp\{\beta_1^* g + \beta_2^* v\}$ into $\mathbb{M}^*\phi^* = \rho^*\phi^*$:

$$
\begin{aligned}
&\rho^* \exp\{\beta_1^* g_{t+1} + \beta_2^* v_{t+1}\} \\
&\quad = \quad \alpha_0 \exp\{[-\alpha_4(1-\kappa) + (\beta_1^* - \alpha_4\kappa)(1-\kappa)]\bar{g} + [\alpha_1 + \alpha_4 + \beta_1^* - \alpha_4\kappa] g_{t+1} + \alpha_3 v_{t+1}\} \\
&\qquad \times \exp\left\{\varphi_v \left(\alpha_2 + \beta_2^* + \frac{1}{2}(\beta_1^* - \alpha_4\kappa)^2\sigma^2\right)\left(1 - \left(\alpha_2 + \beta_2^* + \frac{1}{2}(\beta_1^* - \alpha_4\kappa)^2\sigma^2\right) c_v\right)^{-1} v_{t+1}\right\} \\
&\qquad \times \exp\left\{\delta_v \log\left(1 - \left(\alpha_2 + \beta_2^* + \frac{1}{2}(\beta_1^* - \alpha_4\kappa)^2\sigma^2\right) c_v\right)\right\} .
\end{aligned}
$$

Collecting coefficients, we have:

$$
\begin{aligned}
\rho^* &= \alpha_0 \exp\left\{\alpha_1 \bar{g} - \delta_v \log\left(1 - \left(\alpha_2 + \beta_2^* + \frac{1}{2}(\beta_1^* - \alpha_4\kappa)^2\sigma^2\right) c_v\right)\right\} \\
\beta_1^* &= \frac{\alpha_1}{1-\kappa} + \alpha_4(1+\kappa) \\
\beta_2^* &= \alpha_3 + \varphi_v \left(\alpha_2 + \beta_2^* + \frac{1}{2}(\beta_1^* - \alpha_4\kappa)^2\sigma^2\right)\left(1 - \left(\alpha_2 + \beta_2^* + \frac{1}{2}(\beta_1^* - \alpha_4\kappa)^2\sigma^2\right) c_v\right)^{-1} . \qquad (49)
\end{aligned}
$$

Comparing ρ in the forward calculation and ρ^* in the and reverse calculation, it is clear that $\rho = \rho^*$ provided

$$
\alpha_3 + \beta_2 = \alpha_2 + \beta_2^* + \frac{1}{2}(\beta_1^* - \alpha_4\kappa)^2\sigma^2 \qquad (50)
$$

holds. Substituting (50) into (49) and rearranging yields precisely (48). Therefore, equality (50) holds and so $\rho = \rho^*$.

It remains to solve the quadratic equations (48) and (49) for β_2 and β_2^*, respectively. Consider (48) first. Assuming the discriminant is positive there are two real solutions for β_2. When $\sigma = 0$ we require that $\beta_2 = 0$, so we take the negative root. Similar logic applies for β_2^*. ■

D.3 Proof of results in Appendix C

Proof of Lemma C.1. First note that ρ is a simple isolated eigenvalue of $\mathbb{M}$ under Assumption 4.1 (by Theorem 3.2(c)). Therefore, there exists an $\epsilon > 0$ such that $|\lambda - \rho| > 2\epsilon$ for all $\lambda \in \sigma(\mathbb{M})$. In what follows, 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 $C_\mathcal{R} := \sup_{z\in\Gamma} \|\mathcal{R}(\mathbb{M}, z)\| < \infty$ because $\mathcal{R}(\mathbb{M}, z)$ is a holomorphic function on Γ and Γ is compact.

By Assumption 4.2(a) there exists $\bar{K} \in \mathbb{N}$ such that $\|\Pi_K \mathbb{M} - \mathbb{M}\| < C_\mathcal{R}^{-1}$ for all $K \geq \bar{K}$. Therefore, for all $K \geq \bar{K}$ the inequality

$$
\|\Pi_K \mathbb{M} - \mathbb{M}\| \sup_{z\in\Gamma} \|\mathcal{R}(\mathbb{M}, z)\| \leq C_\mathcal{R} \|\Pi_K \mathbb{M} - \mathbb{M}\| < 1
$$

holds. It follows by Theorem IV.3.18 on p. 214 of Kato (1980) that whenever $K \geq \bar{K}$: (i) the operator $\Pi_K \mathbb{M}$ has precisely one simple eigenvalue ρ_K inside Γ; (ii) $\Gamma \subset (\mathbb{C} \setminus \sigma(\Pi_K \mathbb{M}))$; and (iii) $\sigma(\Pi_K \mathbb{M}) \setminus \{\rho\}$ lies on the exterior of Γ. Note that ρ_K must be real whenever $K \geq \bar{K}$ because complex eigenvalues come in conjugate pairs. Thus, if ρ_K were complex-valued then its conjugate would also be in Γ, which would contradict that ρ_K is the unique eigenvalue of $\Pi_K \mathbb{M}$ on the interior of Γ. This proves (a), (b) and (c).

The proof of (e) follows some arguments from the proof of Proposition 4.2 of Gobet et al. (2004). Take $K \geq \bar{K}$ and let $P_K = (\phi_K \otimes \phi_K^*)$ denote the spectral projection of $\Pi_K \mathbb{M}$ corresponding to ρ_K. By Lemma 6.4 on p. 279 of Chatelin (1983), we have that

$$\phi - P_K \phi \leq \left(\frac{-1}{2\pi \mathrm{i}} \int_\Gamma \frac{\mathcal{R}(\Pi_K \mathbb{M}, z)}{\rho - z} \, \mathrm{d}z \right) (\Pi_K \mathbb{M} - \mathbb{M}) \phi$$

and so

$$\begin{aligned} \|\phi - P_K \phi\| &\leq \frac{1}{2\pi} \left\| \left(\int_\Gamma \frac{\mathcal{R}(\Pi_K \mathbb{M}, z)}{\rho - z} \, \mathrm{d}z \right) (\Pi_K \mathbb{M} - \mathbb{M}) \phi \right\| \\ &\leq \frac{1}{2\pi} (2\pi\epsilon) \frac{\sup_{z \in \Gamma} \|\mathcal{R}(\Pi_K \mathbb{M}, z)\|}{\epsilon} \|(\Pi_K \mathbb{M} - \mathbb{M}) \phi\| \\ &\leq \left(\sup_{z \in \Gamma} \|\mathcal{R}(\Pi_K \mathbb{M}, z)\|\right) \|(\Pi_K \mathbb{M} - \mathbb{M}) \phi\| . \end{aligned} \tag{51}$$

Moreover, for each $z \in \Gamma$ we have

$$\|\mathcal{R}(\Pi_K \mathbb{M}, z)\| \leq \frac{\|\mathcal{R}(\mathbb{M}, z)\|}{1 - \|\Pi_K \mathbb{M} - \mathbb{M}\| C_{\mathcal{R}}} \leq \frac{C_{\mathcal{R}}}{1 - \|\Pi_K \mathbb{M} - \mathbb{M}\| C_{\mathcal{R}}} = O(1) \tag{52}$$

where the first inequality is by Theorem IV.3.17 on p. 214 of Kato (1980) and the second is by definition of $C_{\mathcal{R}}$. This inequality holds uniformly for $z \in \Gamma$. Substituting (52) into (51) yields

$$\begin{aligned} \|\phi - P_K \phi\| &= O(\|(\Pi_K \mathbb{M} - \mathbb{M}) \phi\|) \\ &= \rho \times O(\|\Pi_K \phi - \phi\|) \\ &= O(\delta_K) \end{aligned}$$

where the final line is by Assumption 4.2(b). Note that $P_K = \phi_K \otimes \phi_K^*$. Again by the proof of Proposition 4.2 of Gobet et al. (2004),

$$\|\phi - \phi_K\|^2 \leq 2\|\phi - (\phi_K \otimes \phi_K)\phi\|^2 \leq 2\|\phi - P_K \phi\|^2 \tag{53}$$

and so $\|\phi - \phi_K\| = O(\delta_K)$, proving (d).

The proof of part (d) is similar to the proof of Corollary 4.3 of Gobet et al. (2004). By the triangle inequalities, we have

$$
\begin{aligned}
|\rho_K - \rho| &= |\|\Pi_K \mathbb{M} \phi_K\| - \|\mathbb{M}\phi\|| \\
&\leq \|\Pi_K \mathbb{M} \phi_K - \mathbb{M}\phi\| \\
&\leq \|\Pi_K \mathbb{M} \phi_K - \Pi_K \mathbb{M}\phi\| + \|\Pi_K \mathbb{M}\phi - \mathbb{M}\phi\| \\
&\leq \|\Pi_K \mathbb{M}\| \|\phi_K - \phi\| + \rho \|\Pi_K \phi - \phi\|
\end{aligned}
$$

which is $O(\delta_K)$ because $\|\Pi_K \mathbb{M}\|\|\phi_K - \phi\| = O(\delta_K)$ by part (e) ($\|\Pi_K \mathbb{M}\| = O(1)$ because $\mathbb{M}$ is bounded and Π_K is a (weak) contraction) and $\|\Pi_K \phi - \phi\| = O(\delta_K)$ by Assumption 4.2(b).

For part (f), note that the spectral projection of $(\Pi_K \mathbb{M})^* = \mathbb{M}^* \Pi_K$ corresponding to the eigenvalue ρ_K is given by $P_K^* = (\phi_K^* \otimes \phi_K)$. Also note that $\|\mathcal{R}(\mathbb{M}^*, z)\| = \|\mathcal{R}(\mathbb{M}, \bar{z})\|$ holds for all $z \in \Gamma$ (where $\bar{z}$ denotes the conjugate of z) because $\mathcal{R}(\mathbb{M}^*, z) = \mathcal{R}(\mathbb{M}, \bar{z})^*$ (Kato, 1980, Theorem 6.22, p. 184) and an operator and its adjoint have the same norm. Similarly, $\|\mathcal{R}((\Pi_K \mathbb{M})^*, z)\| = \|\mathcal{R}(\Pi_K \mathbb{M}, \bar{z})\|$ holds for all $z \in \Gamma$ whenever $K \geq \bar{K}$. Thus, by identical arguments to the proof of part (e), we have

$$
\begin{aligned}
\|\phi^* - P_K^* \phi^*\| &\leq O(1) \times \|((\Pi_K \mathbb{M})^* - \mathbb{M}^*)\phi^*\| \\
&\leq \|\mathbb{M}^*\| \|\Pi_K \phi^* - \phi^*\| \\
&= O(\delta_K^*)
\end{aligned}
\tag{54}
$$

where the final line is by Assumption 4.2(c) and boundedness of $\mathbb{M}$. Now observe that

$$
\begin{aligned}
\left\| \frac{\phi^*}{\|\phi^*\|} - \frac{\phi_K^*}{\|\phi_K^*\|} \right\|^2 &\leq 2 \left\| \frac{\phi^*}{\|\phi^*\|} - \left\langle \frac{\phi_K^*}{\|\phi_K^*\|}, \frac{\phi^*}{\|\phi^*\|} \right\rangle \frac{\phi_K^*}{\|\phi_K^*\|} \right\|^2 \\
&\leq 2 \left\| \frac{\phi^*}{\|\phi^*\|} - \left\langle \phi_K \|\phi_K^*\|, \frac{\phi^*}{\|\phi^*\|} \right\rangle \frac{\phi_K^*}{\|\phi_K^*\|} \right\|^2 \\
&= \frac{2}{\|\phi^*\|^2} \times \|\phi^* - P_K^* \phi^*\|^2
\end{aligned}
$$

which is $O((\delta_K^*)^2)$ by (54).

Finally, for part (h) we have

$$
\begin{aligned}
\left\| \frac{\phi_K^+}{\|\phi_K^+\|} - \frac{\phi^*}{\|\phi^*\|} \right\|^2 &\leq 2 \left\| \frac{\phi^*}{\|\phi^*\|} - \left\langle \frac{\phi_K^+}{\|\phi_K^+\|}, \frac{\phi^*}{\|\phi^*\|} \right\rangle \frac{\phi_K^+}{\|\phi_K^+\|} \right\|^2 \\
&\leq 2 \left\| \frac{\phi^*}{\|\phi^*\|} - \left\langle f, \frac{\phi^*}{\|\phi^*\|} \right\rangle \frac{\phi_K^+}{\|\phi_K^+\|} \right\|^2
\end{aligned}
$$

for any $f \in L^2$. Substituting $f = \phi^* \frac{\|\phi_K^+\|}{\|\phi^*\|\|\phi_K^*\|}$ and taking the square root of both sides,

$$\begin{aligned}
\left\| \frac{\phi_K^+}{\|\phi_K^+\|} - \frac{\phi^*}{\|\phi^*\|} \right\| &\leq \sqrt{2} \left\| \frac{\phi^*}{\|\phi^*\|} - \frac{\phi_K^+}{\|\phi_K^*\|} \right\| \\
&\leq \sqrt{2} \left(\left\| \frac{\phi^*}{\|\phi^*\|} - \frac{\Pi_K \phi^*}{\|\phi^*\|} \right\| + \left\| \frac{\Pi_K \phi^*}{\|\phi^*\|} - \frac{\phi_K^+}{\|\phi_K^*\|} \right\| \right) \\
&\leq \sqrt{2} \left(\frac{\|\phi^* - \Pi_K \phi^*\|}{\|\phi^*\|} + \left\| \frac{\phi^*}{\|\phi^*\|} - \frac{\phi_K^*}{\|\phi_K^*\|} \right\| \right)
\end{aligned}$$

where the final line uses the fact that $\phi_K^+ = \Pi_K \phi_K^*$ and $\|\Pi_K\| = 1$. By Assumption 4.2(c) and part(f), we obtain

$$\left\| \frac{\phi_K^+}{\|\phi_K^+\|} - \frac{\phi^*}{\|\phi^*\|} \right\| \leq \sqrt{2}(O(\delta_K^*) + O(\delta_K^*))$$

as required. ■

Proof of Lemma C.2. As in the proof of Lemma C.1, let Γ denote a positively oriented circle in $\mathbb{C}$ centered at ρ with radius ϵ which separates ρ from $\sigma(\mathbb{M}) \setminus \{\rho\}$. Since the nonzero eigenvalues of $\Pi_K \mathbb{M}$ and $\Pi_K \mathbb{M}|_{B_K}$ are the same, it follows from the proof of Lemma C.1 that Γ separates ρ_K from $\sigma(\Pi_K \mathbb{M}|_{B_K}) \setminus \{\rho_K\}$ for all $K \geq \bar{K}$.

Claim 1: $\|\mathcal{R}(\Pi_K \mathbb{M}|_{B_K}, z)\| \leq \|\mathcal{R}(\Pi_K \mathbb{M}, z)\|$ for all z for which $z \in \mathbb{C} \setminus (\sigma(\Pi_K \mathbb{M}) \cup \sigma(\Pi_K \mathbb{M}|_{B_K}))$.

Proof of Claim 1: Fix such a z. Then 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 such that $\psi_K = (\Pi_K \mathbb{M} - zI)\zeta_K(\psi_K)$. Similarly, 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 such that $\psi = (\Pi_K \mathbb{M} - zI)\zeta(\psi)$. Therefore, for any $\psi_K \in B_K$ we must have $\zeta_K(\psi_K) = \zeta(\psi_K)$, i.e., $\mathcal{R}(\Pi_K \mathbb{M}|_{B_K}, z)\psi_K = \mathcal{R}(\Pi_K \mathbb{M}, z)\psi_K$ for all $\psi_K \in B_K$. Therefore,

$$\begin{aligned}
\|\mathcal{R}(\Pi_K \mathbb{M}|_{B_K}, z)\| &:= \sup_{\psi_K \in B_K : \|\psi_K\| = 1} \|\mathcal{R}(\Pi_K \mathbb{M}|_{B_K}, z)\psi_K\| \\
&= \sup_{\psi_K \in B_K : \|\psi_K\| = 1} \|\mathcal{R}(\Pi_K \mathbb{M}, z)\psi_K\| \\
&\leq \sup_{\psi \in L^2 : \|\psi\| = 1} \|\mathcal{R}(\Pi_K \mathbb{M}, z)\psi\| \quad =: \quad \|\mathcal{R}(\Pi_K \mathbb{M}, z)\|
\end{aligned}$$

which proves the claim.

Note $\mathbf{G}^{-1}\mathbf{M}$ is isomorphic to the restriction of $\Pi_K \mathbb{M}$ to B_K, denoted $\Pi_K \mathbb{M}|_{B_K}$, under the inner product $(u, v) \mapsto u'\mathbf{G}v$ on $\mathbb{R}^K$. Taking $K \geq \bar{K}$, it follows from Claim 1 and (52) in the proof of Lemma C.1, that

$$\sup_{z \in \Gamma} \|\mathcal{R}(\Pi_K \mathbb{M}, z)\| = O(1)\,.$$

Therefore, the inequality

$$\|\widehat{\mathbf{G}}^{-1}\widehat{\mathbf{M}} - \mathbf{G}^{-1}\mathbf{M}\|_{\mathbf{G}} \sup_{z\in\Gamma} \|\mathcal{R}(\Pi_K \mathbb{M}, z)\| < 1 \tag{55}$$

holds wpa1 by Assumption 4.3(a). By Claim 1, the inequality

$$\|\widehat{\mathbf{G}}^{-1}\widehat{\mathbf{M}} - \mathbf{G}^{-1}\mathbf{M}\|_{\mathbf{G}} \sup_{z\in\Gamma} \|\mathcal{R}(\Pi_K \mathbb{M}|_{B_K}, z)\| < 1$$

also hold whenever (55) holds. It follows by Theorem IV.3.18 on p. 214 of Kato (1980) that whenever (55) holds: (i) $\widehat{\mathbf{G}}^{-1}\widehat{\mathbf{M}}$ has precisely one simple eigenvalue $\widehat{\rho}$ inside Γ; (ii) $\Gamma \subset (\mathbb{C} \setminus \sigma(\widehat{\mathbf{G}}^{-1}\widehat{\mathbf{M}}))$; and (iii) $\sigma(\widehat{\mathbf{G}}^{-1}\widehat{\mathbf{M}}) \setminus \{\widehat{\rho}\}$ lies on the exterior of Γ. Again, $\widehat{\rho}$ must be real whenever (55) holds because complex eigenvalues come in conjugate pairs. This proves (a), (b) and (c).

For the remainder of the proof we work on the set on which (55) holds. Let $\widehat{P}_K$ denote the spectral projection of $\widehat{\mathbf{G}}^{-1}\widehat{\mathbf{M}}$ corresponding to the eigenvalue $\widehat{\rho}$. Note that $\widehat{P}_K$ is given by $\widehat{P}_K u = \widehat{c}(u'\mathbf{G}\widehat{c}^*)$ for $u \in \mathbb{R}^K$. Because $\mathbb{R}^K$ endowed with $(u,v) \mapsto u'\mathbf{G}v$ is isomorphic to B_K under the L^2 inner product, we have, by similar arguments to the proof of Lemma C.1(d),

$$c_K - \widehat{P}_K c_K \leq \left(\frac{-1}{2\pi \mathrm{i}} \int_\Gamma \frac{\mathcal{R}(\widehat{\mathbf{G}}^{-1}\widehat{\mathbf{M}}, z)}{\rho_K - z} \,\mathrm{d}z \right) (\widehat{\mathbf{G}}^{-1}\widehat{\mathbf{M}} - \mathbf{G}^{-1}\mathbf{M}) c_K$$

and so

$$\|c_K - \widehat{P}_K c_K\|_{\mathbf{G}} \leq \frac{\epsilon}{\inf_{z\in\Gamma} |z - \rho_K|} (\sup\nolimits_{z\in\Gamma} \|\mathcal{R}(\widehat{\mathbf{G}}^{-1}\widehat{\mathbf{M}}, z)\|_{\mathbf{G}}) \|(\widehat{\mathbf{G}}^{-1}\widehat{\mathbf{M}} - \mathbf{G}^{-1}\mathbf{M}) c_K\|_{\mathbf{G}} \,. \tag{56}$$

Note that $\inf_{z\in\Gamma} |z - \rho_K| \to \epsilon$ because Γ is centered at ρ and $|\rho - \rho_K| = o(1)$ by Lemma C.1. Further, when (55) holds, for each $z \in \Gamma$ we have

$$\begin{aligned} \|\mathcal{R}(\widehat{\mathbf{G}}^{-1}\widehat{\mathbf{M}}, z)\|_{\mathbf{G}} &\leq \frac{\|\mathcal{R}(\mathbf{G}^{-1}\mathbf{M}, z)\|_{\mathbf{G}}}{1 - \|\widehat{\mathbf{G}}^{-1}\widehat{\mathbf{M}} - \mathbf{G}^{-1}\mathbf{M}\|_{\mathbf{G}} \sup_{z\in\Gamma} \|\mathcal{R}(\mathbf{G}^{-1}\mathbf{M}, z)\|_{\mathbf{G}}} \\ &= \frac{\|\mathcal{R}(\Pi_K \mathbb{M}|_{B_K}, z)\|}{1 - \|\widehat{\mathbf{G}}^{-1}\widehat{\mathbf{M}} - \mathbf{G}^{-1}\mathbf{M}\|_{\mathbf{G}} \sup_{z\in\Gamma} \|\mathcal{R}(\Pi_K \mathbb{M}|_{B_K}, z)\|} \\ &\leq \frac{\|\mathcal{R}(\Pi_K \mathbb{M}, z)\|}{1 - \|\widehat{\mathbf{G}}^{-1}\widehat{\mathbf{M}} - \mathbf{G}^{-1}\mathbf{M}\|_{\mathbf{G}} \sup_{z\in\Gamma} \|\mathcal{R}(\Pi_K \mathbb{M}, z)\|} \quad = \quad O_p(1) \end{aligned} \tag{57}$$

by Assumption 4.3(a) and (52). Therefore, $\|c_K - \widehat{P}_K c_K\|_{\mathbf{G}} = O_p(\eta_{n,K})$ by (56), (57), and Assumption 4.3(b). Similar to (53) we have

$$\|\widehat{\phi} - \phi_K\|^2 = \|\widehat{c} - c_K\|_{\mathbf{G}}^2 \leq \|c_K - \widehat{c}(\widehat{c}, c_K)\|_{\mathbf{G}}^2 \leq \|c_K - \widehat{P}_K c_K\|_{\mathbf{G}}^2 = O_p(\eta_{n,K}^2) \tag{58}$$

which proves (e).

For part (d), we use the a similar to the proof of Lemma C.1(d) to obtain

$$\begin{aligned}
|\widehat{\rho} - \rho_K| &= \Big| \|\widehat{\mathbf{G}}^{-1}\widehat{\mathbf{M}}\widehat{c}\|_{\mathbf{G}} - \|\mathbf{G}^{-1}\mathbf{M}c_K\|_{\mathbf{G}} \Big| \\
&\leq \|\widehat{\mathbf{G}}^{-1}\widehat{\mathbf{M}}\widehat{c} - \mathbf{G}^{-1}\mathbf{M}c_K\|_{\mathbf{G}} \\
&\leq \|\widehat{\mathbf{G}}^{-1}\widehat{\mathbf{M}}(\widehat{c} - c_K)\|_{\mathbf{G}} + \|(\widehat{\mathbf{G}}^{-1}\widehat{\mathbf{M}} - \mathbf{G}^{-1}\mathbf{M})c_K\|_{\mathbf{G}} \\
&\leq \|\widehat{\mathbf{G}}^{-1}\widehat{\mathbf{M}}\|_{\mathbf{G}}\|\widehat{c} - c_K\|_{\mathbf{G}} + O_p(\eta_{n,K}) \qquad (59)
\end{aligned}$$

where the final line is by Assumption 4.3(b). Moreover, Assumption 4.3(a) implies that $\|\widehat{\mathbf{G}}^{-1}\widehat{\mathbf{M}}\|_{\mathbf{G}} = O_p(1)$, and so we obtain $|\widehat{\rho} - \rho_K| = O_p(\eta_{n,K})$ by substituting (58) into (59).

Finally, for part (f), by identical arguments to the proof of (d) we have

$$\|c_K^* - \widehat{P}_K^* c_K\|_{\mathbf{G}} \leq O_p(1) \times \|(\widehat{\mathbf{G}}^{-1}\widehat{\mathbf{M}}' - \mathbf{G}^{-1}\mathbf{M}')c_K^*\|_{\mathbf{G}} \qquad (60)$$

from which it follows that

$$\begin{aligned}
\left\| \frac{\widehat{\phi}^*}{\|\widehat{\phi}^*\|} - \frac{\phi_K^+}{\|\phi_K^+\|} \right\|^2 &\leq 2 \left\| \frac{\phi_K^+}{\|\phi_K^+\|} - \left\langle \frac{\widehat{\phi}^*}{\|\widehat{\phi}^*\|}, \frac{\phi_K^+}{\|\phi_K^+\|} \right\rangle \frac{\widehat{\phi}^*}{\|\widehat{\phi}^*\|} \right\|^2 \\
&\leq 2 \left\| \frac{\phi_K^+}{\|\phi_K^+\|} - \left\langle \widehat{\phi}\|\widehat{\phi}^*\|, \frac{\phi_K^+}{\|\phi_K^+\|} \right\rangle \frac{\widehat{\phi}^*}{\|\widehat{\phi}^*\|} \right\|^2 \\
&= 2 \left\| \frac{c_K^*}{\|c_K^*\|_{\mathbf{G}}} - \widehat{P}_K^* \left(\frac{c_K^*}{\|c_K^*\|_{\mathbf{G}}} \right) \right\|_{\mathbf{G}}^2 \\
&= O_p(1) \times \|(\widehat{\mathbf{G}}^{-1}\widehat{\mathbf{M}}' - \mathbf{G}^{-1}\mathbf{M}')c_K^*/\|c_K^*\|_{\mathbf{G}}\|_{\mathbf{G}}^2
\end{aligned}$$

by (60). The result follows by Assumption 4.3(c). ■

We first present a general result for verifying Assumption 4.3. The estimators $\widehat{\rho}$, $\widehat{\phi}$ and $\widehat{\phi}^*$ are invariant under an invertible linear transformation of the basis functions $b_{K1}, \ldots, b_{KK}$. Let $\mathbf{M}^o$, $\widehat{\mathbf{G}}^o$, and $\widehat{\mathbf{M}}^o$ by obtained by pre- and post-multiplying $\mathbf{M}$, $\widehat{\mathbf{G}}$ and $\widehat{\mathbf{M}}$ by $\mathbf{G}^{-1/2}$ (where $\mathbf{G}^{-1/2}$ denotes the inverse of the positive definite square root of $\mathbf{G}$). Under this orthogonalization:

$$\begin{aligned}
\|\widehat{\mathbf{G}}^{-1}\widehat{\mathbf{M}} - \mathbf{G}^{-1}\mathbf{M}\|_{\mathbf{G}} &= \|(\widehat{\mathbf{G}}^o)^{-1}\widehat{\mathbf{M}}^o - \mathbf{M}^o\| \\
\|(\widehat{\mathbf{G}}^{-1}\widehat{\mathbf{M}} - \mathbf{G}^{-1}\mathbf{M})c_K\|_{\mathbf{G}} &= \|((\widehat{\mathbf{G}}^o)^{-1}\widehat{\mathbf{M}}^o - \mathbf{M}^o)v_K\| \qquad (61) \\
\|(\widehat{\mathbf{G}}^{-1}\widehat{\mathbf{M}}' - \mathbf{G}^{-1}\mathbf{M}')c_K^*/\|c_K^*\|_{\mathbf{G}}\|_{\mathbf{G}} &= \|((\widehat{\mathbf{G}}^o)^{-1}\widehat{\mathbf{M}}^{o\prime} - \mathbf{M}^{o\prime})v_K^*/\|v_K^*\|\|
\end{aligned}$$

where $v_K = \mathbf{G}^{1/2}c_K$, $v_K^* = \mathbf{G}^{1/2}c_K^*$, and $\|\cdot\|$ denotes the Euclidean norm on $\mathbb{R}^K$ when applied to vectors and the spectral norm (largest singular value) when applied to matrices. Note that $\|v_K\| = 1$ under the normalization $\|\phi_K\| = 1$.

The following general result shows that the convergence rates of the terms in Assumption 4.3 may be bounded by the individual convergence rates of $\widehat{\mathbf{G}}^o$ and $\widehat{\mathbf{M}}^o$. In what follows, let $I = I_K$ denote the $K \times K$ identity matrix and let $a \vee b = \max\{a, b\}$.

Lemma D.2 *Let $\|\widehat{\mathbf{G}}^o - I\| = o_p(1)$, let $\|\widehat{\mathbf{M}}^o - \mathbf{M}^o\| = o_p(1)$ and $\|\mathbf{M}^o\| = O(1)$. Then:*

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

(b) If, in addition, $\|(\widehat{\mathbf{G}}^o - I)v_K\| = O_p(\eta_{n,K,1})$ and $\|(\widehat{\mathbf{M}}^o - \mathbf{M}^o)v_K\| = O_p(\eta_{n,K,2})$ then:

$$\|(\widehat{\mathbf{G}}^{-1}\widehat{\mathbf{M}} - \mathbf{G}^{-1}\mathbf{M})c_K\|_{\mathbf{G}} = O_p(\eta_{n,K,1} \vee \eta_{n,K,2})$$

and if $\|(\widehat{\mathbf{G}}^o - I)v_K^*/\|v_K^*\|\| = O_p(\eta_{n,K,1}^*)$ and $\|(\widehat{\mathbf{M}}^{o\prime} - \mathbf{M}^{o\prime})v_K^*/\|v_K^*\|\| = O_p(\eta_{n,K,2}^*)$ then:

$$\|(\widehat{\mathbf{G}}^{-1}\widehat{\mathbf{M}}' - \mathbf{G}^{-1}\mathbf{M}')c_K^*/\|c_K^*\|_{\mathbf{G}}\|_{\mathbf{G}} = O_p(\eta_{n,K,1}^* \vee \eta_{n,K,2}^*)\,.$$

The condition $\|\mathbf{M}^o\| = O(1)$ is trivially satisfied whenever $\mathbb{M}$ is a bounded operator on L^2.

Proof of Lemma D.2. Follows directly from (61) and Lemma D.3. ■

Lemma D.3 *Let $\|\widehat{\mathbf{G}}^o - I\| = o_p(1)$. Then the following inequalities hold with probability approaching one:*

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

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

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

Proof of Lemma D.3. The condition $\|\widehat{\mathbf{G}}^o - I\| = o_p(1)$ implies that the smallest and largest eigenvalues of $\widehat{\mathbf{G}}^o$ are bounded between $\frac{1}{2}$ and 2 wpa1. Whenever $\frac{1}{2} \leq \lambda_{\min}(\widehat{\mathbf{G}}^o) \leq \lambda_{\max}(\widehat{\mathbf{G}}^o) \leq 2$, we have

$$\begin{aligned}
(\widehat{\mathbf{G}}^o)^{-1}\widehat{\mathbf{M}}^o - \mathbf{M}^o &= (I - (\widehat{\mathbf{G}}^o)^{-1}(\widehat{\mathbf{G}}^o - I))\widehat{\mathbf{M}}^o - \mathbf{M}^o \\
&= \widehat{\mathbf{M}}^o - \mathbf{M}^o - (\widehat{\mathbf{G}}^o)^{-1}(\widehat{\mathbf{G}}^o - I)\mathbf{M}^o - (\widehat{\mathbf{G}}^o)^{-1}(\widehat{\mathbf{G}}^o - I)(\widehat{\mathbf{M}}^o - \mathbf{M}^o)\,. \quad (62)
\end{aligned}$$

Part (a) follows by the triangle inequality, noting that $\|(\widehat{\mathbf{G}}^o)^{-1}\| \leq 2$ whenever $\lambda_{\min}(\widehat{\mathbf{G}}^o) \geq \frac{1}{2}$. Post-multiplying (62) by v_K and using the identity $\mathbf{M}^o v_K = \rho_K v_K$ yields:

$$((\widehat{\mathbf{G}}^o)^{-1}\widehat{\mathbf{M}}^o - \mathbf{M}^o)v_K = (\widehat{\mathbf{M}}^o - \mathbf{M}^o)v_K - \rho_K(\widehat{\mathbf{G}}^o)^{-1}(\widehat{\mathbf{G}}^o - I)v_K - (\widehat{\mathbf{G}}^o)^{-1}(\widehat{\mathbf{G}}^o - I)(\widehat{\mathbf{M}}^o - \mathbf{M}^o)v_K$$

from which (b) follows by the triangle inequality. Part (c) follows similarly by replacing $\widehat{\mathbf{M}}^o$ and $\mathbf{M}^o$ in (62) by their transposes and using the identity $\mathbf{M}^{o\prime}v_K^* = \rho_K v_K^*$. ■

Lemma D.4 *Let $\{X_t\}$ be strictly stationary and exponentially beta-mixing and let $\xi_K \lambda_K (\log n)/\sqrt{n} = O(1)$. Then:*

(a) $\|\widehat{\mathbf{G}}^o - I\| = O_p(\xi_K \lambda_K (\log n)/\sqrt{n})$

(b) $\|(\widehat{\mathbf{G}}^o - I)v_K\| = O_p(\xi_K \lambda_K (\log n)/\sqrt{n})$ *for each deterministic sequence of vectors $v_K \in \mathbb{R}^K$ with $\|v_K\| = 1$.*

Proof of Lemma D.4. Part (a) is just Lemma 2.2 of Chen and Christensen (2014); part (b) follows directly by definition of the spectral norm. ■

Lemma D.5 *Let $\{X_t\}$ be strictly stationary and exponentially beta-mixing, let $\mathbb{E}[|m(X_0, X_1)|^r] < \infty$ for some $2 \leq r \leq \infty$, and let $\xi_K \lambda_K (\log n)/\sqrt{n} = O(1)$. Then:*

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

(b) $\|(\widehat{\mathbf{M}}^o - \mathbf{M}^o)v_K\| = O_p((\xi_K \lambda_K)^{1+2/r} (\log n)/\sqrt{n})$ *and* $\|(\widehat{\mathbf{M}}^{o\prime} - \mathbf{M}^{o\prime})v_K\| = O_p((\xi_K \lambda_K)^{1+2/r} (\log n)/\sqrt{n})$ *for each deterministic sequence of vectors $v_K \in \mathbb{R}^K$ with $\|v_K\| = 1$.*

Proof of Lemma D.5. For part (a) we use a truncation argument in conjunction with an exponential inequality for weakly dependent random matrices due to Chen and Christensen (2014). Let $\{T_n : n \geq 1\}$ be a sequence of positive constants to be defined subsequently, and write

$$\widehat{\mathbf{M}}^o - \mathbf{M}^o = \sum_{t=0}^{n-1} \Xi_{1,t,n} + \sum_{t=0}^{n-1} \Xi_{2,t,n}$$

where

$$\begin{aligned}
\Xi_{1,t,n} &= n^{-1} \widetilde{b}^K(X_t) m(X_t, X_{t+1}) \widetilde{b}^K(X_{t+1})' \chi_{\{\|\widetilde{b}^K(X_t) m(X_t, X_{t+1}) \widetilde{b}^K(X_{t+1})'\| \leq T_n\}} \\
&\quad - n^{-1} \mathbb{E}[\widetilde{b}^K(X_t) m(X_t, X_{t+1}) \widetilde{b}^K(X_{t+1})' \chi_{\{\|\widetilde{b}^K(X_t) m(X_t, X_{t+1}) \widetilde{b}^K(X_{t+1})'\| \leq T_n\}}] \\
\Xi_{2,t,n} &= n^{-1} \widetilde{b}^K(X_t) m(X_t, X_{t+1}) \widetilde{b}^K(X_{t+1})' \chi_{\{\|\widetilde{b}^K(X_t) m(X_t, X_{t+1}) \widetilde{b}^K(X_{t+1})'\| > T_n\}} \\
&\quad - n^{-1} \mathbb{E}[\widetilde{b}^K(X_t) m(X_t, X_{t+1}) \widetilde{b}^K(X_{t+1})' \chi_{\{\|\widetilde{b}^K(X_t) m(X_t, X_{t+1}) \widetilde{b}^K(X_{t+1})'\| > T_n\}}]
\end{aligned}$$

and χ_Λ denotes the indicator function of the event Λ. Note that $\mathbb{E}[\Xi_{1,t,n}] = 0$ and $\|\Xi_{1,t,n}\| \leq 2n^{-1} T_n$ by construction. Further, for any $0 \leq t, s \leq n-1$ we have:

$$\begin{aligned}
\mathbb{E}[\Xi_{1,t,n} \Xi_{1,s,n}'] &\leq n^{-2} \xi_K^2 \lambda_K^2 \mathbb{E}[\widetilde{b}^K(X_t) m(X_t, X_{t+1}) m(X_s, X_{s+1}) \widetilde{b}^K(X_s)'] \\
\mathbb{E}[\Xi_{1,t,n}' \Xi_{1,s,n}] &\leq n^{-2} \xi_K^2 \lambda_K^2 \mathbb{E}[\widetilde{b}^K(X_{t+1}) m(X_t, X_{t+1}) m(X_s, X_{s+1}) \widetilde{b}^K(X_{s+1})']
\end{aligned}$$

and so, by the variational characterization of the spectral norm and the generalized Hölder inequality, we obtain:

$$\begin{aligned}
\|\mathbb{E}[\Xi_{1,t,n}\Xi_{1,s,n}']\| &\leq \sup_{u,v\in\mathbb{R}^K:\|u\|,\|v\|=1} u'\mathbb{E}[\Xi_{1,t,n}\Xi_{1,s,n}']v \\
&\leq n^{-2}\xi_K^2\lambda_K^2\mathbb{E}[|m(X_t,X_{t+1})|^r]^{1/r}E[|m(X_s,X_{s+1})|^r]^{1/r} \\
&\quad \times \sup_{u,v\in\mathbb{R}^K:\|u\|,\|v\|=1} \mathbb{E}[|(u'\widetilde{b}^K(X_t))|^q]^{1/q}\mathbb{E}[|(v'\widetilde{b}^K(X_s))|^q]^{1/q} \\
&\leq n^{-2}\xi_K^2\lambda_K^2\mathbb{E}[|m(X_0,X_1)|^r]^{2/r} \sup_{u\in\mathbb{R}^K:\|u\|=1} \mathbb{E}[|(u'\widetilde{b}^K(X_0))|^q]^{2/q} \\
&\leq n^{-2}\xi_K^2\lambda_K^2\mathbb{E}[|m(X_0,X_1)|^r]^{2/r} \sup_{u\in\mathbb{R}^K:\|u\|=1} (\|u\|\times\xi_K\lambda_K)^{2(q-2)/q}\mathbb{E}[(u'\widetilde{b}^K(X_0))^2]^{2/q} \\
&= O(n^{-2}(\xi_K\lambda_K)^{(2r+4)/r})
\end{aligned}$$

because $1=\frac{2}{r}+\frac{2}{q}$, with the usual modification if $r=2$ or $r=\infty$. This bound holds uniformly for $0\leq t,s\leq n-1$, and also holds for $\|\mathbb{E}[\Xi_{1,t,n}'\Xi_{1,s,n}]\|$. It follows by Corollary 4.2 of Chen and Christensen (2014) that

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

provided $n^{-1}T_n\log n = O((\xi_K\lambda_K)^{1+2/r}/\sqrt{n})$.

Now consider the remaining term. When $r=\infty$ we can set $\Xi_{2,t,n}=0$ for all $0\leq t\leq n-1$ and all n by taking $T_n=C(\xi_K\lambda_K)^2$ for sufficiently large C. Now consider the case $2<r<\infty$. By the triangle and Jensen inequalities,

$$\begin{aligned}
\mathbb{E}\left[\left\|\sum_{t=0}^{n-1}\Xi_{2,t,n}\right\|\right] &\leq 2n^{-1}\sum_{t=0}^{n-1}\mathbb{E}[\|\widetilde{b}^K(X_t)m(X_t,X_{t+1})\widetilde{b}^K(X_{t+1})'\|\chi_{\{\|\widetilde{b}^K(X_t)m(X_t,X_{t+1})\widetilde{b}^K(X_{t+1})'\|>T_n\}}] \\
&\leq \frac{2}{nT_n^{r-1}}\sum_{t=0}^{n-1}\mathbb{E}[\|\widetilde{b}^K(X_t)m(X_t,X_{t+1})\widetilde{b}^K(X_{t+1})'\|^r\chi_{\{\|\widetilde{b}^K(X_t)m(X_t,X_{t+1})\widetilde{b}^K(X_{t+1})'\|>T_n\}}] \\
&\leq \frac{2(\xi_K\lambda_K)^{2r}}{T_n^{r-1}}\mathbb{E}[|m(X_0,X_1)|^r]\,.
\end{aligned}$$

Markov's inequality then yields

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

We choose T_n so that

$$\frac{(\xi_K\lambda_K)^{2r}}{T_n^{r-1}} = \frac{(\xi_K\lambda_K)^{1+2/r}(\log n)}{\sqrt{n}}$$

so that $\|\sum_{t=0}^{n-1}\Xi_{2,t,n}\| = O_p((\xi_K\lambda_K)^{1+2/r}(\log n)/\sqrt{n})$. The condition $n^{-1}T_n\log n = O((\xi_K\lambda_K)^{1+2/r}/\sqrt{n})$

holds for this choice of T_n provided $\xi_K \lambda_K (\log n)/\sqrt{n} = o(1)$.

Part (b) follows from part (a) because $\|\mathbf{A}v\| \leq \|\mathbf{A}\|\|v\|$. ■

Proof of Lemma C.3. Follows from Lemmas D.3, D.4, and D.5. ■

Lemma D.6 *Let $\{X_t\}$ be strictly stationary and exponentially rho-mixing. Then:*

(a) $\|\widehat{\mathbf{G}}^o - I\| = O_p(\xi_K \lambda_K \sqrt{K}/\sqrt{n})$

(b) $\|(\widehat{\mathbf{G}}^o - I)v_K\| = O_p(\xi_K \lambda_K/\sqrt{n})$ *for each deterministic sequence of vectors* $v_K \in \mathbb{R}^K$ *with* $\|v_K\| = 1$.

Proof of Lemma D.6. Parts (a) and (b) may be proved by a slight generalization of the proof of Lemmas 4.8 and 4.12 of Gobet et al. (2004), using $\sup_x \|\widetilde{b}^K(x)\| \leq \lambda_K \sup_x \|b^K(x)\| = \xi_K \lambda_K$. ■

Lemma D.7 *Let $\{X_t\}$ be strictly stationary and exponentially rho-mixing and let $\mathbb{E}[|m(X_0, X_1)|^r] < \infty$ for some $2 \leq r \leq \infty$. Then:*

(a) $\|\widehat{\mathbf{M}}^o - \mathbf{M}^o\| = O_p((\xi_K \lambda_K)^{1+2/r} \sqrt{K}/\sqrt{n})$

(b) $\|(\widehat{\mathbf{M}}^o - \mathbf{M}^o)v_K\| = O_p((\xi_K \lambda_K)^{1+2/r}/\sqrt{n})$ *and* $\|(\widehat{\mathbf{M}}^{o\prime} - \mathbf{M}^{o\prime})v_K\| = O_p((\xi_K \lambda_K)^{1+2/r}/\sqrt{n})$ *for each deterministic sequence of vectors* $v_K \in \mathbb{R}^K$ *with* $\|v_K\| = 1$.

Proof of Lemma D.7. We first prove part (b) using similar arguments to Lemmas 4.8 and 4.9 of Gobet et al. (2004). By the covariance inequality for exponentially rho-mixing processes, we obtain:

$$\begin{aligned}
&\mathbb{E}[\|(\widehat{\mathbf{M}}^o - \mathbf{M}^o)v_K\|^2] \\
&= \frac{1}{n^2}\sum_{l=1}^{K} \mathbb{E}\left[\left(\sum_{t=0}^{n-1} \widetilde{b}_{Kl}(X_t) m(X_t, X_{t+1})(\widetilde{b}^K(X_{t+1})'v_K) - \mathbb{E}[\widetilde{b}_{Kl}(X_t) m(X_t, X_{t+1})(\widetilde{b}^K(X_{t+1})'v_K)]\right)^2\right] \\
&\leq \frac{C}{n}\sum_{l=1}^{K} \mathbb{E}\left[\widetilde{b}_{Kl}(X_t)^2 m(X_t, X_{t+1})^2 (\widetilde{b}^K(X_{t+1})'v_K)^2\right] \\
&\leq \frac{C(\xi_K\lambda_K)^2}{n} \mathbb{E}[|m(X_0, X_1)|^r]^{2/r} \mathbb{E}[(\widetilde{b}^K(X_0)'v_K)^{2r/(r-2)}]^{(r-2)/r} \\
&\leq \frac{C(\xi_K\lambda_K)^{2+4/r}}{n} \mathbb{E}[|m(X_0, X_1)|^r]^{2/r} \qquad (63)
\end{aligned}$$

where the constant C depends only upon the rho-mixing coefficients, $\widetilde{b}_{K1}(x), \ldots, \widetilde{b}_{KK}(x)$ denote the elements of $\widetilde{b}^K(x)$, and the final line is because $\|v_K\| = 1$ and $E[(\widetilde{b}^K(X_0)'v_K)^2] = \|v_K\|^2 = 1$. Chebyshev's inequality and (63) imply $\|(\widehat{\mathbf{M}}^o - \mathbf{M}^o)v_K\| = O_p((\xi_K \lambda_K)^{1+2/r}/\sqrt{n})$. An identical argument proves the result for $\|(\widehat{\mathbf{M}}^{o\prime} - \mathbf{M}^{o\prime})v_K\|$. This completes the proof of (b). For part (a), let

$u_1,\dots,u_K$ be an orthonormal basis for $\mathbb{R}^K$ (with respect to the Euclidean inner product). Using the fact that the Frobenius norm $\|\cdot\|_F$ dominates the L^2 norm, we have

$$\begin{aligned} \mathbb{E}[\|\widehat{\mathbf{M}}^o - \mathbf{M}^o\|^2] &\leq \mathbb{E}[\|\widehat{\mathbf{M}}^o - \mathbf{M}^o\|_F^2] \\ &= \sum_{k=1}^K \mathbb{E}[\|(\widehat{\mathbf{M}}^o - \mathbf{M}^o)u_k\|^2] \\ &\leq \frac{CK(\xi_K\lambda_K)^{2+4/r}}{n}\mathbb{E}[|m(X_0,X_1)|^r]^{2/r} \end{aligned}$$

where the final line is by (63). Part (a) follows by Chebyshev's inequality. ■

Proof of Lemma C.4. Follows from Lemmas D.3, D.6, and D.7. ■

Lemma D.8 *Let $\{X_t\}$ be strictly stationary and exponentially beta-mixing, let $\mathbb{E}[|m(X_0,X_1)|^r] < \infty$ for some $2 \leq r \leq \infty$, let $(\frac{1}{n}\sum_{t=0}^{n-1}(\widehat{m}(X_t,X_{t+1}) - m(X_t,X_{t+1}))^2)^{1/2} = O_p(\nu_n)$ where $\nu_n = o(1)$, and let $\xi_K\lambda_K(\log n)/\sqrt{n} = o(1)$. Then:*

(a) $\|\widehat{\mathbf{M}}^o - \mathbf{M}^o\| = O_p((\xi_K\lambda_K)^{1+2/r}\sqrt{K}/\sqrt{n} + \xi_K\lambda_K\nu_n)$

(b) $\|(\widehat{\mathbf{M}}^o - \mathbf{M}^o)v_K\| = O_p((\xi_K\lambda_K)^{1+2/r}/\sqrt{n} + \xi_K\lambda_K\nu_n)$ *and* $\|(\widehat{\mathbf{M}}^{o\prime} - \mathbf{M}^{o\prime})v_K\| = O_p((\xi_K\lambda_K)^{1+2/r}/\sqrt{n} + \xi_K\lambda_K\nu_n)$ *for each deterministic sequence of vectors $v_K \in \mathbb{R}^K$ with $\|v_K\| = 1$.*

Proof of Lemma D.8. For part (a), by Lemma D.5 and the triangle inequality we have

$$\|\widehat{\mathbf{M}}^o - \mathbf{M}^o\| \leq \left\|\frac{1}{n}\sum_{t=0}^{n-1}\widetilde{b}^K(X_t)[\widehat{m}(X_t,X_{t+1}) - m(X_t,X_{t+1})]\widetilde{b}^K(X_{t+1})'\right\| + O_p\left(\frac{(\xi_K\lambda_K)^{1+2/r}\log n}{\sqrt{n}}\right).$$

To control the leading term,

$$\begin{aligned} &\left\|\frac{1}{n}\sum_{t=0}^{n-1}\widetilde{b}^K(X_t)[\widehat{m}(X_t,X_{t+1}) - m(X_t,X_{t+1})]\widetilde{b}^K(X_{t+1})'\right\| \\ &= \sup_{v_1,v_2\in\mathbb{R}^K:\|v_1\|,\|v_2\|=1}\left|\frac{1}{n}\sum_{t=0}^{n-1}(v_1'\widetilde{b}^K(X_t))[\widehat{m}(X_t,X_{t+1}) - m(X_t,X_{t+1})](v_2'\widetilde{b}^K(X_{t+1}))\right| \\ &\leq \xi_K\lambda_K \sup_{v_1\in\mathbb{R}^K:\|v_1\|=1}\frac{1}{n}\sum_{t=0}^{n-1}|(v_1'\widetilde{b}^K(X_t))(\widehat{m}(X_t,X_{t+1}) - m(X_t,X_{t+1}))| \\ &\leq \xi_K\lambda_K \times O_p(\nu_n) \times \sup_{v_1\in\mathbb{R}^K:\|v_1\|=1}\left(\frac{1}{n}\sum_{t=0}^{n-1}(v_1'\widetilde{b}^K(X_t))^2\right)^{1/2} \\ &= \xi_K\lambda_K \times O_p(\nu_n) \times \sup_{v_1\in\mathbb{R}^K:\|v_1\|=1}\left(v_1'\widehat{\mathbf{G}}^o v_1\right)^{1/2} = \xi_K\lambda_K \times O_p(\nu_n) \times O_p(1) \end{aligned}$$

where the first inequality is by Hölder's inequality, the second is by the Cauchy-Schwarz inequality, and the final line is because $\|\widehat{\mathbf{G}}^o - I\| = o_p(1)$ under the condition $\xi_K \lambda_K (\log n)/\sqrt{n} = o(1)$ (by Lemma D.4). Part (b) follows directly. ■

Proof of Lemma C.5. Follows from Lemmas D.3, D.4, and D.8. ■

Lemma D.9 *Let $\{X_t\}$ be strictly stationary and exponentially rho-mixing, let $\mathbb{E}[|m(X_0, X_1)|^r] < \infty$ for some $2 \leq r \leq \infty$, let $(\frac{1}{n}\sum_{t=0}^{n-1}(\widehat{m}(X_t, X_{t+1}) - m(X_t, X_{t+1}))^2)^{1/2} = O_p(\nu_n)$ where $\nu_n = o(1)$, and let $\xi_K \lambda_K \sqrt{K}/\sqrt{n} = o(1)$. Then:*

(a) $\|\widehat{\mathbf{M}}^o - \mathbf{M}^o\| = O_p((\xi_K \lambda_K)^{1+2/r}\sqrt{K}/\sqrt{n} + \xi_K \lambda_K \nu_n)$

(b) $\|(\widehat{\mathbf{M}}^o - \mathbf{M}^o)v_K\| = O_p((\xi_K \lambda_K)^{1+2/r}/\sqrt{n} + \xi_K \lambda_K \nu_n)$ *and* $\|(\widehat{\mathbf{M}}^{o\prime} - \mathbf{M}^{o\prime})v_K\| = O_p((\xi_K \lambda_K)^{1+2/r}/\sqrt{n} + \xi_K \lambda_K \nu_n)$ *for each deterministic sequence of vectors $v_K \in \mathbb{R}^K$ with $\|v_K\| = 1$.*

Proof of Lemma D.9. Follows by similar arguments to the proof of Lemma D.8. ■

Proof of Lemma C.6. Follows from Lemmas D.3, D.6, and D.9. ■

We present a simple lemma which is used in the proof of Lemma C.7.

Lemma D.10 *Let Assumptions 4.1, 4.2 and 4.3 hold and let $\|\mathbf{G}^{-1}\mathbf{M} - \widehat{\mathbf{G}}^{-1}\widehat{\mathbf{M}}\|_{\mathbf{G}} = O_p(\bar{\eta}_{n,K})$. Then:*

$$\widehat{\rho} - \rho_K = c_K^{*\prime}\mathbf{G}(\widehat{\mathbf{G}}^{-1}\widehat{\mathbf{M}} - \mathbf{G}^{-1}\mathbf{M})c_K + O_p(\eta_{n,K}\bar{\eta}_{n,K})\,.$$

Proof of Lemma D.10. By the proof of Lemma C.2, we know that $\widehat{\mathbf{G}}^{-1}\widehat{\mathbf{M}}$ has a unique simple eigenvalue $\widehat{\rho}$ in the interval $[\rho - \epsilon, \rho + \epsilon]$ whenever (55) holds, which it does wpa1 under Assumption 4.3(a). For the remainder of the proof we will work on the set on which (55) holds.

Recall that the spectral projection of $\widehat{\mathbf{G}}^{-1}\widehat{\mathbf{M}}$ associated with the eigenvalue $\widehat{\rho}$ is given by $\widehat{P}_K$ $\widehat{P}_K u = \widehat{c}(u'\mathbf{G}\widehat{c}^*)$ for $u \in \mathbb{R}^K$. Similarly, the spectral projection of $\mathbf{G}^{-1}\mathbf{M}$ associated with the eigenvalue ρ_K is given by $P_K^+ u = c_K(u'\mathbf{G}c_K^*)$ for $u \in \mathbb{R}^K$. Also recall that c_K, c_K^*, $\widehat{c}$ and $\widehat{c}^*$ are normalized such that $c_K^{*\prime}\mathbf{G}c_K = 1$ and $\widehat{c}^{*\prime}\mathbf{G}\widehat{c} = 1$.

Working on the set on which (55) holds, we have

$$\begin{aligned}
\widehat{\rho} - \rho_K &= \mathrm{Tr}(\widehat{P}_K \widehat{\mathbf{G}}^{-1}\widehat{\mathbf{M}} - P_K^+ \mathbf{G}^{-1}\mathbf{M}) \\
&= \mathrm{Tr}(P_K^+(\widehat{\mathbf{G}}^{-1}\widehat{\mathbf{M}} - \mathbf{G}^{-1}\mathbf{M}) + (\widehat{P}_K - P_K^+)\widehat{\mathbf{G}}^{-1}\widehat{\mathbf{M}}) \\
&= c_K^{*\prime}\mathbf{G}(\widehat{\mathbf{G}}^{-1}\widehat{\mathbf{M}} - \mathbf{G}^{-1}\mathbf{M})c_K + \mathrm{Tr}((\widehat{P}_K - P_K^+)\widehat{\mathbf{G}}^{-1}\widehat{\mathbf{M}})
\end{aligned}$$

by linearity of trace. It remains to show that $\mathrm{Tr}((\widehat{P}_K - P_K^+)\widehat{\mathbf{G}}^{-1}\widehat{\mathbf{M}}) = O_p(\eta_{n,K}\bar{\eta}_{n,K})$. Using the inner product $(u,v) = u'\mathbf{G}v$,

$$
\begin{aligned}
&\mathrm{Tr}((\widehat{P}_K - P_K^+)\widehat{\mathbf{G}}^{-1}\widehat{\mathbf{M}}) \\
&\quad = \quad \widehat{\rho} - c_K^* \mathbf{G}\widehat{\mathbf{G}}^{-1}\widehat{\mathbf{M}}\widehat{P}_K c_K + c_K^{*\prime}\mathbf{G}\widehat{\mathbf{G}}^{-1}\widehat{\mathbf{M}}(\widehat{P}_K c_K - c_K) \\
&\quad = \quad \widehat{\rho}(c_K^*, c_K - \widehat{P}_K c_K) + c_K^{*\prime}\mathbf{G}\widehat{\mathbf{G}}^{-1}\widehat{\mathbf{M}}(\widehat{P}_K c_K - c_K) \\
&\quad = \quad (\widehat{\rho} - \rho_K)(c_K^*, c_K - \widehat{P}_K c_K) + \rho_K c_K^{*\prime}\mathbf{G}(c_K - \widehat{P}_K c_K) + c_K^{*\prime}\mathbf{G}\widehat{\mathbf{G}}^{-1}\widehat{\mathbf{M}}(\widehat{P}_K c_K - c_K) \\
&\quad = \quad (\widehat{\rho} - \rho_K)(c_K^*, c_K - \widehat{P}_K c_K) + c_K^{*\prime}\mathbf{G}\mathbf{G}^{-1}\mathbf{M}(c_K - \widehat{P}_K c_K) + c_K^{*\prime}\mathbf{G}\widehat{\mathbf{G}}^{-1}\widehat{\mathbf{M}}(\widehat{P}_K c_K - c_K) \\
&\quad = \quad (\widehat{\rho} - \rho_K)(c_K^*, c_K - \widehat{P}_K c_K) + (c_K^*, (\mathbf{G}^{-1}\mathbf{M} - \widehat{\mathbf{G}}^{-1}\widehat{\mathbf{M}})(c_K - \widehat{P}_K c_K))\,.
\end{aligned}
$$

By the Cauchy-Schwarz inequality $(u,v) \leq \|u\|_{\mathbf{G}}\|v\|_{\mathbf{G}}$, we have

$$
|\mathrm{Tr}((\widehat{P}_K - P_K^+)\widehat{\mathbf{G}}^{-1}\widehat{\mathbf{M}})| \leq \|c_K^*\|_{\mathbf{G}}\|c_K - \widehat{P}_K c_K\|_{\mathbf{G}}\left(|\widehat{\rho} - \rho_K| + \|\mathbf{G}^{-1}\mathbf{M} - \widehat{\mathbf{G}}^{-1}\widehat{\mathbf{M}}\|_{\mathbf{G}}\right)\,.
$$

Finally, $|\widehat{\rho} - \rho_K| = O_p(\eta_{n,K})$ by Lemma C.2(d), $\|c_K - \widehat{P}_K c_K\|_{\mathbf{G}} = O_p(\eta_{n,K})$ by the proof of Lemma C.2(e), $\|\mathbf{G}^{-1}\mathbf{M} - \widehat{\mathbf{G}}^{-1}\widehat{\mathbf{M}}\|_{\mathbf{G}} = O_p(\bar{\eta}_{n,K})$ by assumption. Finally, let Γ be as in the proof of Lemmas C.1 and C.2. Using the integral representation for P_K^+ (Kato, 1980, expression (6.19) on p. 178), we have

$$
\|c_K^*\|_{\mathbf{G}} = \|P_K^+\| = \left\|\frac{1}{2\pi\mathrm{i}}\int_\Gamma \mathcal{R}(\Pi_K\mathbb{M}|_{B_K}, z)\,\mathrm{d}z\right\| \leq \epsilon \times \sup_{z\in\Gamma}\|R(\Pi_K\mathbb{M}|_{B_K}, z)\|
$$

which is $O(1)$ by the proof of Lemma C.2. ■

Proof of Lemma C.7. Expression (61) and Lemma D.3 together imply that $\|\widehat{\mathbf{G}}^{-1}\widehat{\mathbf{M}} - \mathbf{G}^{-1}\mathbf{M}\|_{\mathbf{G}} = O_p(\bar{\eta}_{n,K,1} \vee \bar{\eta}_{n,K,2})$. Lemma D.10 then provides that

$$
\widehat{\rho} - \rho_K = c_K^{*\prime}\mathbf{G}(\widehat{\mathbf{G}}^{-1}\widehat{\mathbf{M}} - \mathbf{G}^{-1}\mathbf{M})c_K + O_p(\eta_{n,K} \times (\bar{\eta}_{n,K,1} \vee \bar{\eta}_{n,K,2}))\,. \tag{64}
$$

By rotational invariance, we have

$$
c_K^{*\prime}\mathbf{G}(\widehat{\mathbf{G}}^{-1}\widehat{\mathbf{M}} - \mathbf{G}^{-1}\mathbf{M})c_K = c_K^{*\prime}\mathbf{G}^{1/2}(\widehat{\mathbf{G}}^{o-1}\widehat{\mathbf{M}}^o - \mathbf{M}^o)\mathbf{G}^{1/2}c_K\,. \tag{65}
$$

As in the proof of Lemma D.3, the condition $\|\widehat{\mathbf{G}}^o - I\| = o_p(1)$ implies that, wpa1, the minimum and maximum eigenvalues of $\widehat{\mathbf{G}}^o$ are between $\frac{1}{2}$ and 2. We work on this set for the remainder of the proof. Repeated substitution of $\widehat{\mathbf{G}}^{o-1} = I - \widehat{\mathbf{G}}^{o-1}(\widehat{\mathbf{G}}^o - I)$ yields

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

where

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

Substituting into (65) yields

$$\begin{aligned} c_K^{*\prime}\mathbf{G}(\widehat{\mathbf{G}}^{-1}\widehat{\mathbf{M}} - \mathbf{G}^{-1}\mathbf{M})c_K &= c_K^{*\prime}(\widehat{\mathbf{M}} - \widehat{\mathbf{G}}\mathbf{G}^{-1}\mathbf{M})c_K + O_p(\bar{\eta}_{n,K,1} \times (\bar{\eta}_{n,K,1} \vee \bar{\eta}_{n,K,2})) \\ &= c_K^{*\prime}(\widehat{\mathbf{M}} - \rho_K\widehat{\mathbf{G}})c_K + O_p(\bar{\eta}_{n,K,1} \times (\bar{\eta}_{n,K,1} \vee \bar{\eta}_{n,K,2})) \end{aligned}$$

and the result follows by substituting into (64). ■

Proof of Theorem C.1. We first prove part (a). By addition and subtraction of terms,

$$\begin{aligned} \widehat{V}_\rho - V_\rho &= \frac{1}{n}\sum_{t=0}^{n-1}\left(\widehat{\phi}_t^{*f2}m_{t,t+1}^2\widehat{\phi}_{t+1}^{f2} - \phi_t^{*2}m_{t,t+1}^2\phi_{t+1}^2\right) && (66) \\ &\quad + \frac{1}{n}\sum_{t=0}^{n-1}\phi_t^{*2}m_{t,t+1}^2\phi_{t+1}^2 - \mathbb{E}[\phi^*(X_0)^2 m(X_0,X_1)^2\phi(X_1)^2] && (67) \\ &\quad + \frac{1}{n}\sum_{t=0}^{n-1}\left(\widehat{\rho}^2\widehat{\phi}_t^{*f2}\widehat{\phi}_t^{f2} - \rho^2\phi_t^{*2}\phi_t^2\right) && (68) \\ &\quad + \frac{1}{n}\sum_{t=0}^{n-1}\rho^2\phi_t^{*2}\phi_t^2 - \rho^2\mathbb{E}[\phi^*(X_0)^2\phi(X_0)^2] && (69) \\ &\quad - \frac{2}{n}\sum_{t=0}^{n-1}\left(\widehat{\rho}\widehat{\phi}_t^{*f2}m_{t,t+1}\widehat{\phi}_t^f\widehat{\phi}_{t+1}^f - \rho\phi_t^{*2}m_{t,t+1}\phi_t\phi_{t+1}\right) && (70) \\ &\quad - \frac{2}{n}\sum_{t=0}^{n-1}\rho\phi_t^{*2}m_{t,t+1}\phi_t\phi_{t+1} + 2\rho\mathbb{E}[\phi^*(X_0)^2 m(X_0,X_1)\phi(X_0)\phi(X_1)] && (71) \end{aligned}$$

Terms (67), (69) and (71) are all $o_{a.s.}(1)$ by the ergodic theorem (the expectations exist by Assumption C.1(i)). Consider term (66), expanded as

$$\frac{1}{n}\sum_{t=0}^{n-1}\phi_t^{*2}m_{t,t+1}^2[\widehat{\phi}_{t+1}^{f2} - \phi_{t+1}^2] + \frac{1}{n}\sum_{t=0}^{n-1}[\widehat{\phi}_t^{*f2} - \phi_t^{*2}]m_{t,t+1}^2[\widehat{\phi}_{t+1}^{f2} - \phi_{t+1}^2] + \frac{1}{n}\sum_{t=0}^{n-1}[\widehat{\phi}_t^{*f2} - \phi_t^{*2}]m_{t,t+1}^2\phi_{t+1}^2 .$$

Let $\|\cdot\|_\infty$ denote the sup norm, and observe that $\|f\|_\infty \leq \xi_K\lambda_K\|f\|$ uniformly for all $f \in B_K$. By the relation $(a^2 - b^2) = (a+b)(a-b)$, we have

$$\begin{aligned} |(66)| &\leq \xi_K\lambda_K(\|\widehat{\phi}^f - \phi\| \vee \|\widehat{\phi}^{*f} - \phi^*\|)\Big\{\frac{1}{n}\sum_{t=0}^{n-1}\phi_t^{*2}m_{t,t+1}^2|\widehat{\phi}_{t+1}^f + \phi_{t+1}| \\ &\quad + \frac{1}{n}\sum_{t=0}^{n-1}|\widehat{\phi}_t^{*f} + \phi_t^*|m_{t,t+1}^2|\widehat{\phi}_{t+1}^f + \phi_{t+1}| + \frac{1}{n}\sum_{t=0}^{n-1}|\widehat{\phi}_t^{*f} + \phi_t^*|m_{t,t+1}^2\phi_{t+1}^2\Big\} . \end{aligned}$$

Also note that $|\widehat{\phi}_t^f + \phi_t| \leq 2\phi_t + \|\widehat{\phi}^f - \phi\|_\infty \leq 2\phi_t + \xi_K \lambda_K \|\widehat{\phi}^f - \phi\|$; a similar bound applies to $\widehat{\phi}^{*f}$. It follows by substituting into the above display and applying Assumption C.1(a)(b) that the term in braces is $O_p(1)$, and so:

$$|(66)| \quad \leq \quad \xi_K \lambda_K (\|\widehat{\phi}^f - \phi\| \vee \|\widehat{\phi}^{*f} - \phi^*\|) \times O_p(1) = o_p(1)$$

by Assumption C.1(b). Similar arguments may be applied to show that terms (68) and (70) are both $o_p(1)$. Therefore, $\widehat{V}_\rho \to_p V_\rho$ and so $\sqrt{n}\widehat{V}_\rho^{-1/2}(\widehat{\rho} - \rho) \to_d N(0,1)$ by Theorem 4.2(a) and the continuous mapping theorem. Part (b) now follows immediately from part (a).

For part (c), we first write

$$\widehat{\Lambda}_j = \Lambda_j + \frac{1}{\sqrt{n}} \sum_{t=0}^{n-1} h_j \left(\frac{t+1}{n} \right) (\widehat{\psi}_L(X_t, X_{t+1}) - \psi_L(X_t, X_{t+1}))$$

where $\Lambda_j = \frac{1}{\sqrt{n}} \sum_{t=0}^{n-1} h_j(\frac{t+1}{n}) \psi_L(X_t, X_{t+1})$. Let $h_{j,t} = h_j\left(\frac{t+1}{n}\right)$ to simplify notation. Writing out term-by-term, for each $j = 1, \ldots, J$ we have:

$$\begin{aligned} \widehat{\Lambda}_j - \Lambda_j \quad = \quad & \left(\frac{1}{\sqrt{n}} \sum_{t=0}^{n-1} h_{j,t} \right) (\overline{lm}_n - \mathbb{E}[\log m(X_0, X_1)]) & (72) \\ & + \frac{1}{\sqrt{n}} \sum_{t=0}^{n-1} h_{j,t} (\phi_t^* \phi_t - \widehat{\phi}_t^{*f} \widehat{\phi}_t^f) & (73) \\ & + \widehat{\rho}^{-1} \frac{1}{\sqrt{n}} \sum_{t=0}^{n-1} h_{j,t} (\widehat{\phi}_t^{*f} \widehat{\phi}_{t+1}^f - \phi_t^* \phi_{t+1}) m_{t,t+1} & (74) \\ & + \sqrt{n}(\widehat{\rho}^{-1} - \rho^{-1}) \times \frac{1}{n} \sum_{t=0}^{n-1} h_{j,t} \phi_t^* m_{t,t+1} \phi_{t+1} \,. & (75) \end{aligned}$$

Note that $n^{-1} \sum_{t=0}^{n-1} h_{j,t} = O(n^{-1})$ (because $\int_0^1 h_j(u)\,\mathrm{d}u = 0$ and h_j is continuously differentiable). Term (72) is $o_p(1)$ because and $\overline{lm}_n - \mathbb{E}[\log m(X_0, X_1)] = o_{a.s.}(1)$ by the ergodic theorem and because $n^{-1/2} \sum_{t=0}^{n-1} h_{j,t} = O(n^{-1/2})$. Term (73) may be rewritten as

$$\frac{1}{\sqrt{n}} \sum_{t=0}^{n-1} h_{j,t} \{ \phi_t^* \phi_t - \widehat{\phi}_t^{*f} \widehat{\phi}_t^f - \mathbb{E}[\phi_t^* \phi_t - \widehat{\phi}_t^{*f} \widehat{\phi}_t^f] \} + \mathbb{E}[\phi_t^* \phi_t - \widehat{\phi}_t^{*f} \widehat{\phi}_t^f] \frac{1}{\sqrt{n}} \sum_{t=0}^{n-1} h_{j,t}$$

where the first term is $o_p(1)$ by Assumption C.1(c) and the second term is $o_p(1)$ since $\frac{1}{\sqrt{n}} \sum_{t=0}^{n-1} h_{j,t} = O(n^{-1/2})$ and $|\mathbb{E}[\phi_t^* \phi_t - \widehat{\phi}_t^{*f} \widehat{\phi}_t^f]| \leq \|\phi^*\| \|\phi - \widehat{\phi}^f\| + \|\widehat{\phi}^f\| \|\phi^* - \widehat{\phi}^{*f}\| = o_p(1)$ by Assumption C.1(b). An identical argument shows (74) is $o_p(1)$. For term (75), $\sqrt{n}(\widehat{\rho}^{-1} - \rho^{-1}) = O_p(1)$ by Theorem 4.2 and the delta method. It may be deduced from Assumption C.1(d) that $\frac{1}{n} \sum_{t=0}^{n-1} h_{j,t} \phi_t^* m_{t,t+1} \phi_{t+1} = o_p(1)$, which implies that (75) is $o_p(1)$.

Thus, by the proof of Theorem 4.2(c) and Assumption C.1(e), we have

$$
\begin{aligned}
(\sqrt{n}(\widehat{L}-L), \widehat{\Lambda}_1, \ldots, \widehat{\Lambda}_J)' &= (n^{-1/2} \textstyle\sum_{t=0}^{n-1} \psi_L(X_t, X_{t+1}), \Lambda_1, \ldots, \Lambda_J)' + o_p(1) \\
&\to_d N(0, V_L \times I_{J+1})
\end{aligned}
$$

and the result follows by definition of the χ^2_J and t_J distributions. ■

E Additional Monte Carlo results

Here we present additional MC results for the design in the body of the text. Tables 4 and 5 present the results for cubic B-spline sieves for b^K, p^{K_1} and ψ^{K_2}. The MC mean and MC confidence bands together with the true functions for ϕ and ϕ^* under both preference specifications are plotted in Figures 5 (Hermite polynomials) and 6 (B-splines; note that the vertical scale is different for each subplot). Figures 5 and 6 show that ϕ and ϕ^* may be estimated to a high degree of accuracy in small samples under both preference specifications. Comparing the results in Tables 4 and 5 with those in the main text, we see that the overall behavior of the estimates with Hermite polynomials and B-splines is very similar.

Tables 6 and 7 present further MC results for the same design but with $\kappa = 0.30$ and $\sigma = 0.005$, which are roughly the parameters obtained by fitting a Gaussian AR(1) to quarterly real per capita consumption growth. Comparing the results in Tables 6 against the results presented in Tables 1 and 4, we see that the bias and RMSE for $\widehat{\phi}$, $\widehat{\phi}^*$ and $\widehat{v}$ are roughly one third of that obtained under the parameterization $\kappa = 0.60$ and $\sigma = 0.01$. Table 7 shows that the bias for $\widehat{\rho}$, $\widehat{y}$ and $\widehat{L}$ order or smaller 10^{-4}. Surprisingly, the RMSE of $\widehat{\rho}$, $\widehat{y}$ and $\widehat{L}$ with $\kappa = 0.30$ and $\sigma = 0.005$ obtained using a Hermite polynomial sieve when $n = 400$ are larger than under the more volatile and persistent specification. Table 7 also shows that the Bias and RMSE of $\widehat{\rho}$, $\widehat{y}$ and $\widehat{L}$ for B-splines with $n = 400$ and for both bases with $n > 400$ are of a smaller order of magnitude than the Bias and RMSE with $\kappa = 0.30$ and $\sigma = 0.005$ presented in Tables 2 and 5.

n	CRRA $\widehat{\phi}$	CRRA $\widehat{\phi}^*$	Epstein-Zin $\widehat{\phi}$	Epstein-Zin $\widehat{\phi}^*$	Epstein-Zin $\widehat{v}$
	Bias				
400	0.0382	0.0486	0.0361	0.0869	0.1145
800	0.0182	0.0240	0.0179	0.0399	0.0465
1600	0.0051	0.0067	0.0067	0.0128	0.0265
3200	0.0010	0.0013	0.0025	0.0049	0.0070
	RMSE				
400	0.0884	0.1081	0.0653	0.1730	0.2853
800	0.0538	0.0664	0.0358	0.1177	0.1909
1600	0.0325	0.0395	0.0165	0.0747	0.1364
3200	0.0183	0.0219	0.0069	0.0444	0.0930

Table 4: Bias and RMSE of $\widehat{\phi}$ and $\widehat{\phi}^*$ under both preference specifications, as well as bias and RMSE of $\widehat{v}$ under Epstein-Zin preferences. Results are obtained from 1000 replications of the MC design with $\kappa = 0.60$ and $\sigma = 0.01$ for different sample sizes using cubic B-spline bases for b^K, p^{K_1} and ψ^{K_2} with $K = 8$, $K_1 = 6$ and $K_2 = 12$.

n	CRRA $\widehat{\rho}$	CRRA $\widehat{y}$	CRRA $\widehat{L}$	Epstein-Zin $\widehat{\rho}$	Epstein-Zin $\widehat{y}$	Epstein-Zin $\widehat{L}$
	Bias					
400	0.0004	-0.0002	0.0006	0.0030	-0.0030	0.0005
800	0.0006	-0.0005	0.0003	0.0017	-0.0017	0.0005
1600	0.0003	-0.0003	0.0004	0.0010	-0.0010	0.0005
3200	0.0003	-0.0003	0.0001	0.0004	-0.0004	0.0001
	RMSE					
400	0.0180	0.0177	0.0119	0.0086	0.0082	0.0099
800	0.0102	0.0104	0.0058	0.0046	0.0045	0.0065
1600	0.0081	0.0082	0.0052	0.0041	0.0039	0.0054
3200	0.0050	0.0051	0.0025	0.0009	0.0010	0.0025

Table 5: Bias and RMSE of $\widehat{\rho}$, $\widehat{y}$ and $\widehat{L}$ under both preference specifications. Results are obtained from 1000 replications of the MC design with $\kappa = 0.60$, $\sigma = 0.01$ for different sample sizes using cubic B-spline bases for b^K, p^{K_1} and ψ^{K_2} with $K = 8$, $K_1 = 6$ and $K_2 = 12$.

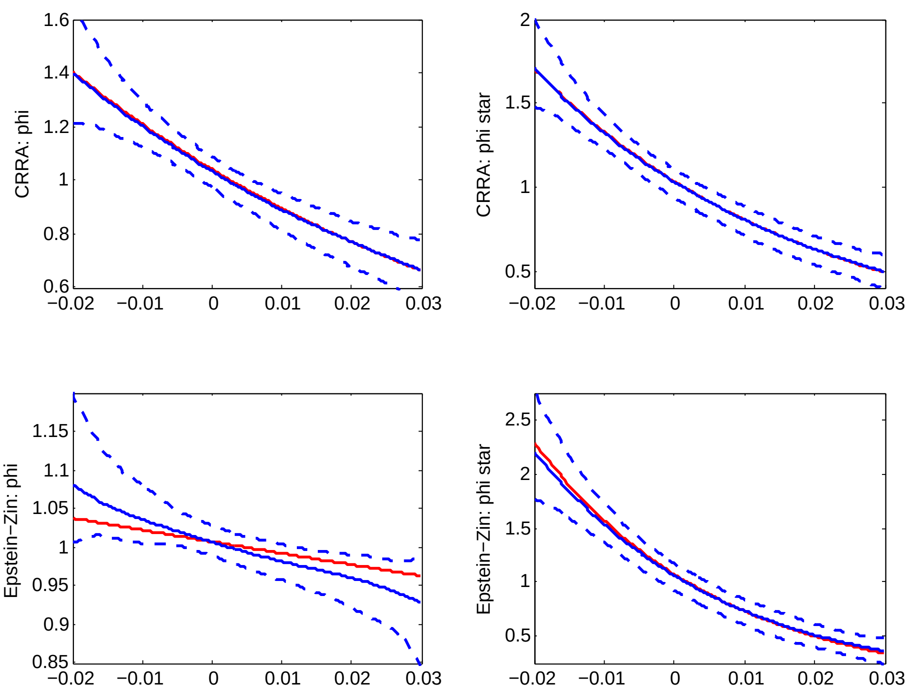

Figure 5: MC results for $\widehat{\phi}$ and $\widehat{\phi}^*$ under both preference specifications for the sample size $n = 400$. Dashed lines are pointwise 95% MC confidence intervals, solid red line is the true ϕ or ϕ^*, and solid blue line is the pointwise mean across MC replications. Results are obtained from 1000 replications of the MC design with $\kappa = 0.60$ and $\sigma = 0.01$ using Hermite polynomial bases for b^K, p^{K_1} and ψ^{K_2} with $K = 8$, $K_1 = 6$ and $K_2 = 12$. Note the vertical scales are different for each subplot.

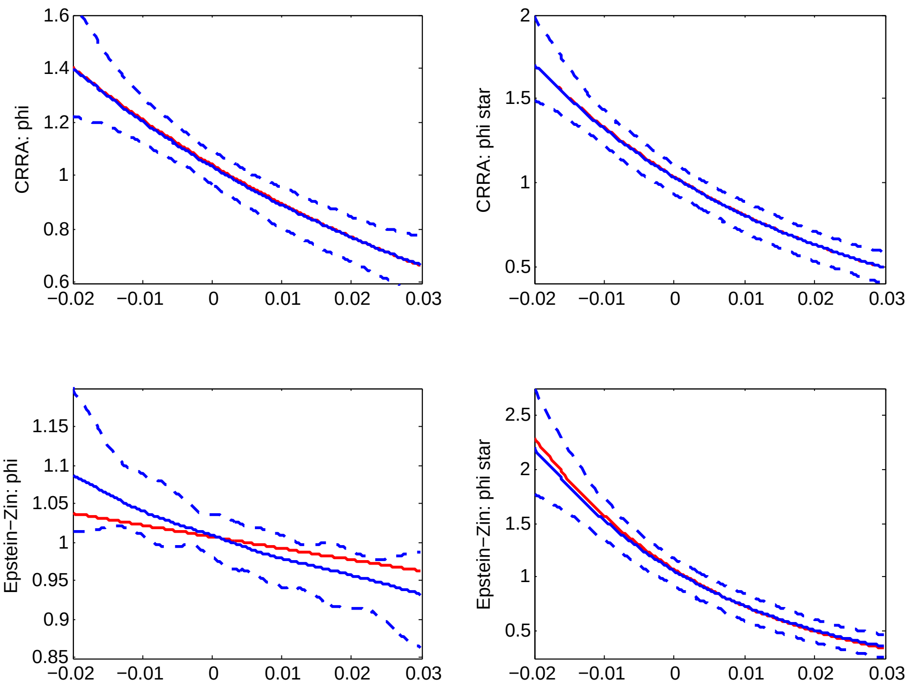

Figure 6: MC plots for $\widehat{\phi}$ and $\widehat{\phi}^*$ under both preference specifications for the sample size $n = 400$. Dashed lines are 95% MC confidence bands, solid red line is the true ϕ or ϕ^*, and solid blue line is the pointwise mean across MC replications. Results are obtained from 1000 replications of the MC design with $\kappa = 0.60$ and $\sigma = 0.01$ using cubic B-spline bases for b^K, p^{K_1} and ψ^{K_2} with $K = 8$, $K_1 = 6$ and $K_2 = 12$. Note the vertical scales are different for each subplot.

Sieve	n	CRRA $\widehat{\phi}$	CRRA $\widehat{\phi}^*$	Epstein-Zin $\widehat{\phi}$	Epstein-Zin $\widehat{\phi}^*$	Epstein-Zin $\widehat{v}$
		Bias				
HPol	400	0.0029	0.0139	0.0040	0.0139	0.0106
HPol	800	0.0002	0.0006	0.0009	0.0012	0.0013
HPol	1600	0.0001	0.0002	0.0005	0.0005	0.0027
HPol	3200	0.0001	0.0001	0.0002	0.0003	0.0017
Bspl	400	0.0286	0.0279	0.0286	0.0286	0.0360
Bspl	800	0.0126	0.0131	0.0126	0.0135	0.0151
Bspl	1600	0.0033	0.0034	0.0034	0.0034	0.0047
Bspl	3200	0.0004	0.0004	0.0004	0.0005	0.0018
		RMSE				
HPol	400	0.0153	0.0312	0.0112	0.0351	0.1046
HPol	800	0.0069	0.0096	0.0037	0.0126	0.0748
HPol	1600	0.0046	0.0056	0.0022	0.0080	0.0519
HPol	3200	0.0031	0.0037	0.0014	0.0053	0.0371
Bspl	400	0.0484	0.0533	0.0467	0.0562	0.1268
Bspl	800	0.0237	0.0255	0.0220	0.0277	0.0842
Bspl	1600	0.0094	0.0103	0.0074	0.0123	0.0545
Bspl	3200	0.0037	0.0044	0.0021	0.0059	0.0373

Table 6: Bias and RMSE of $\widehat{\phi}$ and $\widehat{\phi}^*$ under both preference specifications, as well as bias and RMSE of $\widehat{v}$ under Epstein-Zin preferences. Results are obtained from 1000 replications of the MC design with $\kappa = 0.30$ and $\sigma = 0.005$ for different sample sizes and sieves for b^K, p^{K_1} and ψ^{K_2} with $K = 8$, $K_1 = 6$ and $K_2 = 12$.

Sieve	n	CRRA $\widehat{\rho}$	CRRA $\widehat{y}$	CRRA $\widehat{L}$	Epstein-Zin $\widehat{\rho}$	Epstein-Zin $\widehat{y}$	Epstein-Zin $\widehat{L}$
		Bias					
HPol	400	0.0013	-0.0008	0.0009	0.0014	-0.0009	0.0008
HPol	800	0.0000	-0.0000	0.0000	0.0000	-0.0000	0.0000
HPol	1600	-0.0000	0.0000	0.0000	0.0000	-0.0000	0.0000
HPol	3200	0.0001	-0.0001	-0.0000	0.0000	-0.0000	-0.0000
Bspl	400	0.0000	0.0000	0.0001	0.0001	-0.0001	0.0001
Bspl	800	0.0000	-0.0000	-0.0000	0.0000	-0.0000	0.0000
Bspl	1600	-0.0000	0.0000	0.0000	0.0000	-0.0000	0.0000
Bspl	3200	0.0001	-0.0001	-0.0000	0.0000	-0.0000	-0.0000
		RMSE					
HPol	400	0.0463	0.0297	0.0296	0.0419	0.0269	0.0269
HPol	800	0.0024	0.0025	0.0003	0.0003	0.0003	0.0003
HPol	1600	0.0017	0.0018	0.0002	0.0002	0.0002	0.0002
HPol	3200	0.0012	0.0013	0.0001	0.0001	0.0001	0.0001
Bspl	400	0.0050	0.0051	0.0038	0.0033	0.0032	0.0032
Bspl	800	0.0024	0.0025	0.0003	0.0003	0.0003	0.0003
Bspl	1600	0.0017	0.0018	0.0002	0.0002	0.0002	0.0002
Bspl	3200	0.0012	0.0013	0.0001	0.0001	0.0001	0.0001

Table 7: Bias and RMSE of $\widehat{\rho}$, $\widehat{y}$ and $\widehat{L}$ under both preference specifications. Results are obtained from 1000 replications of the MC design with $\kappa = 0.30$, $\sigma = 0.005$ for different sample sizes and sieves for b^K, p^{K_1} and ψ^{K_2} with $K = 8$, $K_1 = 6$ and $K_2 = 12$.

References

Backus, D., M. Chernov, and S. Zin (2014): "Sources of Entropy in Representative Agent Models," *Journal of Finance*, 69, 51–99.

Beare, B. K. (2010): "Copulas and Temporal Dependence," *Econometrica*, 78, 395–410.

Bickel, P. J. and J. Kwon (2001): "Inference for Semiparametric Models: Some Questions and an Answer (with Discussion)," *Statistica Sinica*, 11, 863–960.

Billingsley, P. (1961): "The Lindeberg-Lévy Theorem for Martingales," *Proceedings of the American Mathematical Society*, 12, 788–792.

Chatelin, F. (1983): *Spectral Approximation of Linear Operators*, Academic Press, New York.

Chen, X. (2007): "Large Sample Sieve Estimation of Semi-Nonparametric Models," in *Handbook of Econometrics*, ed. by J. J. Heckman and E. E. Leamer, Elsevier, vol. 6, Part B, chap. 76, 5549–5632.

Chen, X. and T. M. Christensen (2014): "Optimal Uniform Convergence Rates and Asymptotic Normality for Series Estimators Under Weak Dependence and Weak Conditions," *Journal of Econometrics*, forthcoming.

Chen, X., L. P. Hansen, and M. Carrasco (2010): "Nonlinearity and Temporal Dependence," *Journal of Econometrics*, 155, 155–169.

Chen, X., Z. Liao, and Y. Sun (2012): "Sieve Inference on Possibly Misspecified Semi-Nonparametric Time Series Models," Cowles Foundation Discussion Paper d1849.

Chen, X., W. B. Wu, and Y. Yi (2009): "Efficient estimation of copula-based semiparametric Markov models," *The Annals of Statistics*, 37, 4214–4253.

Doukhan, P. (1994): *Mixing: Properties and Examples*, Springer, New York.

Duffee, G. R. (2002): "Term Premia and Interest Rate Forecasts in Affine Models," *The Journal of Finance*, 57, 405–443.

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.

Gourieroux, C. and J. Jasiak (2006): "Autoregressive Gamma Processes," *Journal of Forecasting*, 25, 129–152.

Greenwood, P. E., A. Schick, and W. Wefelmeyer (2001): "Comment [on Bickel and Kwon, 2001]," *Statistica Sinica*, 11, 892–906.

Hansen, L. P., J. C. Heaton, and N. Li (2008): "Consumption Strikes Back? Measuring Long-Run Risk," *Journal of Political Economy*, 116, 260–302.

Hansen, L. P. and J. A. Scheinkman (2009): "Long-Term Risk: An Operator Approach," *Econometrica*, 77, 177–234.

Jansson, M. (2004): "The Error in Rejection Probability of Simple Autocorrelation Robust Tests," *Econometrica*, 72, 937–946.

Kato, T. (1980): *Perturbation Theory for Linear Operators*, Springer-Verlag, Berlin.

Kiefer, N. M., T. J. Vogelsang, and H. Bunzel (2000): "Simple Robust Testing of Regression Hypotheses," *Econometrica*, 68, 695–714.

Lancaster, P. and M. Tismenetsky (1985): *The Theory of Matrices: With Applications*, Academic Press.

Lettau, M. and J. A. Wachter (2007): "Why Is Long-Horizon Equity Less Risky? A Duration-Based Explanation of the Value Premium," *The Journal of Finance*, 62, 55–92.

——— (2011): "The Term Structures of Equity and Interest Rates," *Journal of Financial Economics*, 101, 90–113.

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

Müller, U. K. (2007): "A Theory of Robust Long-Run Variance Estimation," *Journal of Econometrics*, 141, 1331–1352.

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

Phillips, P. C. B. (2005): "HAC Estimation by Automated Regression," *Econometric Theory*, 21, 116–142.

Schaefer, H. H. (1960): "Some Spectral Properties of Positive Linear Operators," *Pacific Journal of Mathematics*, 10, 1009–1019.

——— (1974): *Banach Lattices and Positive Operators*, Springer-Verlag, Berlin.

——— (1999): *Topological Vector Spaces*, Springer-Verlag, New York.

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

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