# ESTIMATING THE LONG-RUN IMPLICATIONS OF DYNAMIC ASSET PRICING MODELS

TIMOTHY M. CHRISTENSEN

*Department of Economics*
*Yale University*

Abstract. This paper introduces new econometric tools for studying the long-run implications of dynamic asset pricing models. The long-run implications of a model are jointly determined by the functional form of the SDF and the dynamic behavior of the variables in the model. The estimators introduced in this paper treat the dynamics as an unknown nuisance parameter. Nonparametric sieve estimators of the positive eigenfunction and eigenvalue used to decompose the stochastic discount factor (SDF) into its permanent and transitory components are proposed, as are estimators of the long-term yield and entropy of the permanent component of the SDF. The estimators are particularly simple to implement, and may be used to numerically compute the long-run implications of fully specified models for which analytical solutions are unavailable. Nonparametric identification conditions are presented. Consistency and convergence rates of the estimators are established. An approach for conducting asymptotic inference on the eigenvalue, long-term yield, and entropy of the permanent component of the SDF is provided. The semiparametric efficiency bounds for these parameters are derived and their estimators are shown to be efficient. The long-run implications of the consumption CAPM are investigated using these methods. The estimators, identification conditions, and large sample theory presented in this paper have broader application in economics including, for example, the nonparametric estimation of marginal utilities of consumption in representative agent models.

Keywords: Nonparametric estimation, Nonparametric identification, Sieve estimation, Stochastic discount factor, Permanent component, Long run, Entropy.

JEL codes: C13, C14, C58.

## JOB MARKET PAPER

*Date*: January 9, 2014.
*E-mail:* `timothy.christensen@yale.edu`
*URL:* `http://tmchristensen.wordpress.com`
I thank Xiaohong Chen and Peter Phillips who have been gracious with their advice, support and feedback. I would also like to thank Don Andrews, Caio Almeida, Tim Armstrong, Jiti Gao, Yuichi Kitamura, Enrique Sentana, and participants of the SKBI Summer Institute on Financial Economics (May 2012), the SoFiE Conference on Asset Pricing and Portfolio Allocation in the Long Run (December 2012), the North American and European Meetings of the Econometric Society (June and August 2013), and seminars at Monash University and Yale University. While writing this paper I was generously supported by a Carl Arvid Anderson Fellowship from the Cowles Foundation.

## 1. Introduction

Dynamic asset pricing models link the prices of future state-contingent payoffs with sources of risk, the payoff horizon, and the preferences of economic agents. The stochastic discount factor (SDF, or pricing kernel) within a dynamic asset pricing model assigns values to future state-contingent payoffs. Recent work in macroeconomics and asset pricing has shown how to extract information about the long-run pricing implications of a model by analyzing the permanent component of the SDF (see, for example, Alvarez and Jermann (2005); Hansen and Scheinkman (2009); Hansen (2012); Backus, Chernov, and Zin (2013)). As this work has highlighted, long-run implications provide a powerful and robust means with which to analyze dynamic asset pricing models. Different assumptions about the preferences of economic agents may, in some cases, result in different short-run implications but the same long-run implications. Consequently, the long-run implications of classes of asset pricing models may, in some cases, be inferred by studying just one model.

This paper introduces an econometric framework for extracting information about the permanent component of the SDF, and the pricing of long-horizon assets, from a dynamic asset pricing model. The permanent component of the SDF and the long-run implications of the model are jointly determined by both the functional form of the SDF and the short-run dynamics, or law of motion, of the variables in the model. The framework introduced in this paper treats the dynamics as an unknown nuisance parameter. Economic theory is often vague regarding the precise form that the dynamics should take. In practice, dynamics are usually specified parametrically in a way that makes analytical solution of the model feasible. Changing the dynamics can change the long-run implications of a model. One might, therefore, be concerned that the long-run implications of a model may be sensitive to the specification of the dynamics. Generalized method of moments (GMM) is a popular technique for estimating asset pricing models because it uses moment restrictions, typically based on an Euler equation or asset-pricing equation, which are derived from economic theory and places only weak assumptions on the dynamics of the data. The estimators proposed in this paper are based on the same Euler equation or asset-pricing equation one would use to estimate a model with GMM. Rather than placing parametric restrictions on the dynamics, the estimators proposed in this paper nonparametrically infer, from a time series of data, attributes of the dynamics from which the long-run implications of the model are obtained.

As shown by Hansen and Scheinkman (2009) and Hansen (2012), information about the permanent component of the SDF and the pricing of long-horizon assets can be extracted by studying a positive eigenfunction problem related to an appropriately chosen operator. Their analysis applies to economies in which there exists a Markov state process whose value at each point in time contains all the relevant information for valuation. The operator is determined jointly by the SDF and the dynamics of the state process. The positive eigenfunction characterizes the state dependence of the prices of long-horizon assets, and its eigenvalue is related to the yield on long-term zero-coupon bonds and the entropy of the permanent component of the SDF. The entropy of the permanent component of the SDF is a joint measure of the dispersion of the SDF and persistence of the SDF process. This metric can be used to place an upper bound on average excess returns on risky assets relative to long-term bonds (see Alvarez and Jermann (2005)). Whether or not the bound is satisfied by historical average returns on equity relative to long-term bonds provides a measure with which the predictions of the model may be evaluated.

The central focus of this paper is the nonparametric estimation of the positive eigenfunction and its eigenvalue, the long-term yield, and the entropy of the permanent component. The operator is unknown when the dynamics are unknown. Extracting the long-run implications of the model given a time series of data on the state process therefore requires estimating a positive eigenfunction of an unknown operator. A feasible nonparametric sieve estimator is proposed, inspired by earlier work of Chen, Hansen, and Scheinkman (2000). Sieve estimation methods are appealing in this context as they reduce an intractable infinite-dimensional eigenfunction problem to a simple matrix eigenvector problem. The matrix eigenvector problem is formed by instrumenting the Euler equation or asset-pricing equation in the model by a growing collection of basis functions. The estimators are particularly easy to implement: no simulation, optimization or numerical integration is required. By contrast, the use of kernel-based methods in this context would involve nonparametric estimation of a conditional density, numerical computation of an integral, and solution of an infinite-dimensional eigenfunction problem. The sieve estimators may also be used to numerically compute the long-run implications of fully specified asset pricing models for which analytical solutions are unavailable.

Large sample properties of the estimators are established. The eigenfunction estimators are consistent and converge at reasonable nonparametric rates under appropriate regularity conditions. The asymptotic distribution and semiparametric efficiency bounds of the eigenvalue, long-term yield and entropy of the permanent component are derived, and the estimators of these quantities are shown to be efficient. An approach to performing asymptotic inference is provided. The derivation of the large sample properties is nonstandard, as the eigenfunction and eigenvalue being estimated are defined implicitly by an unknown nonselfadjoint operator. Favorable small-sample performance of the estimators is illustrated in a Monte Carlo study. The large sample theory is first presented for the case in which the long-run implications of a given SDF are to be investigated. The large sample theory is then extended to the case in which the researcher first estimates a SDF, either parametrically or semi/nonparametrically, from a time series of data on returns and the state process, then estimates the long-run implications of the estimated SDF. Other extensions of the large sample theory are explored, including nonparametric sieve estimation of marginal utilities in representative agent models.

To simplify the econometric analysis, the scope of this paper is confined to discrete-time economies with finite-dimensional stationary state processes. Primitive nonparametric identification conditions for the positive eigenfunction in stationary discrete-time environments are provided. The conditions are formulated in terms of positivity and integrability conditions on the SDF and the stationary and transition densities of the state process. The existence and identification conditions complement those that Hansen and Scheinkman (2009) provide for general continuous-time environments. Existence of the positive eigenfunction is guaranteed under the identification conditions. A version of the long-run pricing result of Hansen and Scheinkman (2009) also obtains under the identification conditions.

The estimators are used to study the long-run implications of the consumption capital asset pricing model (CAPM). High levels of risk aversion are required to generate an estimated entropy of the permanent component of the SDF that is consistent with historical average returns on equities relative to long-term bonds. Moreover, when risk aversion is set high enough to rationalize this excess return the implied long-term yield is much larger than historical long-term yields. The long-run implications of the consumption CAPM are the same as a wider class of consumption-based representative agent models, including some habit formation models and limiting versions of recursive preference models. These empirical findings therefore have broader import beyond the standard consumption CAPM.

The estimators introduced in this paper cannot, in their present form, be used to study models with latent state variables. Latent variables are a useful modeling tool for incorporating features such as stochastic growth and stochastic volatility in an analytically tractable way. For example, the popular long-run risks model of Bansal and Yaron (2004) specifies that (log) consumption growth is the sum of a latent predictable component and a stochastic component, in which both the latent predictable component and stochastic volatility evolve as first-order Gaussian processes. The estimators introduced in this paper may be used to analyze models whose state processes exhibit time-variation in growth, conditional volatility, and other nonlinearities provided these features are state-dependent (instead of latent). Allowing for nonlinear, state-dependent dynamics goes some way to incorporate features that might otherwise be modeled by latent processes. So although the scope of the estimators is confined to models with observable variables, the restrictions this imposes on the dynamics the observable variables is less severe than it may first appear. Moreover, the version of the long-run pricing result presented in this paper applies equally to models with latent state variables, and the sieve approach may be used to numerically calculate the long-run implications of fully specified models with latent state variables.

The remainder of the paper is structured as follows. Section 2 discusses three other nonparametric eigenfunction problems in economics that the research developed in this paper could be used to analyze. Section 3 reviews the decomposition of the SDF into its permanent and transitory components using the positive eigenfunction and its eigenvalue, and introduces other quantities to be studied. Identification and a version of the long-term pricing result are presented in Section 4. The nonparametric sieve estimators are introduced in Section 5, and their large sample properties are derived. Section 6 discusses extension of the large sample theory to cover estimated SDFs, more general SDFs, and nonparametric sieve estimation of marginal utilities. Section 7 examines the performance of the estimators in a Monte Carlo exercise. Section 8 studies the consumption CAPM using the estimators introduced in this paper, and Section 9 concludes. An appendix contains supplementary results and all proofs.

## 2. Nonparametric eigenfunction problems in economics

The identification conditions, estimators, and large sample theory developed in this paper have broader application to nonparametric identification and estimation of economic models. Three other applications, namely nonparametric Euler equations, household consumption models, and transitory misspecification of asset pricing models, are now briefly outlined.

2.1. **Nonparametric Euler equations.** The Euler equation within consumption-based asset pricing models places restrictions on the comovement of asset returns and the marginal utility of consumption of economic agents. Such restrictions have been the basis for a vast literature on estimating consumption-based asset pricing models from time series of asset returns and consumption data. Recent work has shown that marginal utility of consumption may be represented as a positive eigenfunction of an appropriately chosen operator. This eigenfunction representation provides an alternative framework in which to study nonparametric identification and estimation of semi/nonparametric consumption-based asset pricing models.

Let $MU_t^h$ denote the marginal utility of consumption of agent $h$ at time $t$. Consider an economy in which the gross return on asset $i$ from time $t$ to $t+1$, denoted $R_{i,t+1}$, is determined by the Euler equation

$$MU_t^h = E[\beta MU_{t+1}^h R_{i,t+1} | \mathcal{I}_t^h] \tag{1}$$

where $\beta > 0$ is a time-preference parameter, and $\mathcal{I}_t^h$ is the information set of the agent at time $t$. Assume $MU_t^h$ is a function (known to the agent/s but unknown to the econometrician) of a vector of explanatory variables $X_t^h$

$$MU_t^h = MU(X_t^h)$$

and that the explanatory variables belong to the agent's information set, i.e. $\sigma(X_t^h) \subseteq \mathcal{I}_t^h$. By iterated expectations, the Euler equation (1) can be rewritten as

$$E[MU(X_{t+1}^h) R_{i,t+1} | X_t^h] = \beta^{-1} MU(X_t^h)\,. \tag{2}$$

Expression (2) defines $(MU, \beta)$ as the solution to nonparametric eigenfunction problem

$$\mathbb{T}_i MU = \beta^{-1} MU \tag{3}$$

where $\mathbb{T}_i f(X_t^h) = E[f(X_{t+1}^h) R_{i,t+1} | X_t^h]$. Marginal utility of consumption is typically assumed to be positive, in which case $MU$ is a positive eigenfunction of $\mathbb{T}_i$.

Linton, Lewbel, and Srisuma (2011) and Escanciano and Hoderlein (2012) use this positive eigenfunction representation of marginal utility to analyze identification in representative agent models.[1] Chen, Chernozhukov, Lee, and Newey (2013a) use a similar eigenfunction representation to provide nonparametric identification conditions in a representative agent model with external habit formation. The eigenfunction representation of marginal utility of consumption does not appear to have been used to study heterogeneous-agent models to date.

The sieve estimators introduced in this paper extend to the nonparametric estimation of the marginal utility function $MU$ and time-preference parameter $\beta$ of a representative agent, given a time series of data on $\{(X_t, R_{i,t+1})\}$ (see Section 6.3 for further details). This sieve-based approach is an alternative to the kernel-based procedure introduced in Linton, Lewbel, and Srisuma (2011).

[1]This approach also has some similarities with Ross (2013), who nonparametrically recovers the pricing kernel from panels of option prices by solving a positive eigenvector problem.

That the same pair $(MU, \beta)$ are the solution to (3) for each asset $i$ for which the Euler equation holds provides a source of over-identifying restrictions with which to test the model in both the representative- and heterogeneous-agent cases.

2.2. **Household consumption models.** Eigenfunction techniques may also be used to study semiparametric Euler equations. Consider a semiparametric variant of the preceding model, in which $MU_t^h$ is of the form

$$MU_t^h = [C_t^h]^{-\gamma} v(Z_t^h) \tag{4}$$

where $C_t^h$ is the consumption of household $h$ at time $t$, $Z_t^h$ is a vector of explanatory variables, and $v$ is a positive function. Attanasio and Weber (1993, 1995) use a model of this form to estimate preference parameters from household-level panel data. In their treatment, the function $v$ is used to correct for the effect that a household's demographic structure may have on the marginal utility of a given level of consumption expenditure. A common approach for estimating these models from household-level panel data is to (i) assume a parametric form for $v$, (ii) log linearize the Euler equation, and (iii) estimate the log-linearized model by a panel instrumental variables regression. Marginal utility of the form (4) could equally be a heterogeneous-agent variant of the semiparametric consumption CAPM with external habit formation studied by Chen and Ludvigson (2009) and Chen, Chernozhukov, Lee, and Newey (2013a). The function $v$ would represent an external habit formation component in this interpretation of (4). The identification conditions, estimators and large sample theory developed in this paper may be extended to provide an alternative means with which to study nonparametric identification and estimation of these models.

The function $v$ may be represented as the positive eigenfunction of an operator related to the Euler equation. Substituting $MU_t^h$ of the form (4) into the Euler equation (1) yields

$$[C_t^h]^{-\gamma} v(Z_t^h) = E[\beta [C_{t+1}^h]^{-\gamma} v(Z_{t+1}^h) R_{i,t+1} | \mathcal{I}_t^h] \,. \tag{5}$$

Let $G_{t+1}^h = C_{t+1}^h / C_t^h$ denote the growth in consumption of household $h$ from time $t$ to time $t+1$. When $\sigma((C_t^h, Z_t^h)) \subseteq \mathcal{I}_t^h$, the Euler equation (5) can be rewritten as

$$E[(G_{t+1}^h)^{-\gamma} R_{i,t+1} v(Z_{t+1}^h) | Z_t^h] = \beta^{-1} v(Z_t^h) \,.$$

Therefore, $(v, \beta^{-1})$ are the solution to the eigenfunction problem:

$$\mathbb{T}_{i,h} v = \beta^{-1} v \tag{6}$$

where $\mathbb{T}_{i,h} f(Z_t^h) = E[(G_{t+1}^h)^{-\gamma} R_{i,t+1} f(Z_{t+1}^h) | Z_t^h]$. The same $(v, \beta^{-1})$ must solve the eigenfunction relation (6) for each household $h$ and asset $i$, providing a source of overidentifying restrictions.

2.3. **Diagnosing transitory misspecifications in asset pricing models.** The recent literature on extracting the long-run implications of asset pricing models has highlighted the fact that classes of asset pricing model may yield the same long-run implications but different short-run implications (see, e.g., Bansal and Lehmann (1997); Hansen (2012); Hansen and Scheinkman (2013); Backus, Chernov, and Zin (2013)). This line of research may also be used to study transitory misspecifications of SDFs in asset pricing models.

Let the economy be characterized by discrete-time Markov state process $\{X_t\}$ and consider SDF misspecification of the form

$$m(X_t, X_{t+1}) = \alpha m_{mis}(X_t, X_{t+1}) \frac{h(X_{t+1})}{h(X_t)} \tag{7}$$

where $m$ is the true SDF and $m_{mis}$ is the misspecified SDF used by the econometrician. The constant $\alpha > 0$ in expression (7) plays the role of a discount rate distortion, and the function $h > 0$ captures transitory misspecification of the SDF. Hansen (2012) shows that both $m$ and $m_{mis}$ will have the same permanent component, but different transitory components. If $\alpha = 1$ then $m$ and $m_{mis}$ will imply the same long-run rate of return. If $h = 1$ then both $m$ and $m_{mis}$ will share the same positive eigenfunction.

Hansen and Scheinkman (2013) show that the true SDF $m$ may be recovered from the misspecified SDF $m_{mis}$ by solving a positive eigenfunction problem. Assume assets are priced using the true SDF $m$, i.e.

$$E\left[ m(X_t, X_{t+1}) R_{i,t+1} | \, X_t \right] = 1 \,. \tag{8}$$

Substitution of (7) into (8) yields

$$E\left[ m_{mis}(X_t, X_{t+1}) R_{i,t+1} h(X_{t+1}) | \, X_t \right] = \alpha^{-1} h(X_t) \,.$$

The transitory adjustment $h$ and multiplicative constant $\alpha$ are therefore the solution to the positive eigenfunction problem

$$\mathbb{T}_i h = \alpha^{-1} h \tag{9}$$

where $\mathbb{T}_i f(X_t) = E[m_{mis}(X_t, X_{t+1}) R_{i,t+1} f(X_{t+1}) | X_t]$. The techniques developed in this paper may be applied to study nonparametric identification and estimation of $(h, \alpha)$ from a time-series of data on $(X_t, R_{i,t+1})$, thereby providing a means with which to diagnose transitory misspecifications of SDFs. Over-identifying restrictions are again implicit since the same $(h, \alpha)$ must solve the eigenfunction relation (9) for each asset $i$ for which (8) holds.

## 3. Review of the long-run implications of dynamic asset pricing models

This section briefly reviews the positive eigenfunction problem and its relation to the long-run implications of asset pricing models, as exposited by Hansen and Scheinkman (2009) and Hansen (2012). The connection between these quantities and other metrics developed by Alvarez and Jermann (2005) and Backus, Chernov, and Zin (2013) is also discussed.

### 3.1. Model.

Consider a class of economy characterized by a discrete-time (first-order) Markov state process $\{X_t\}$ defined on a complete probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, \mathbb{P})$, where time is indexed by $t \in \mathbb{Z}$ and where $\mathcal{F}_t = \sigma(X_t, X_{t-1}, \ldots)$ denotes the completion of the $\sigma$-algebra generated by $\{X_t, X_{t-1}, \ldots\}$. Let $\{X_t\}$ have support $\mathcal{X} \subseteq \mathbb{R}^d$. Assume further that in this economy the date-$t$

price of a claim to the date-$(t+\tau)$ state-dependent payoff $Z_{t+\tau}$ is given by

$$E\left[\left(\prod_{s=t}^{t+\tau-1} m(X_s, X_{s+1})\right) Z_{t+\tau} \middle| X_t\right] \tag{10}$$

for some positive measurable function $m : \mathcal{X} \times \mathcal{X} \to \mathbb{R}_+$, for all $t \in \mathbb{Z}$ and $\tau \geq 1$. The function $m$ will be referred to generically as the SDF. The sequence $\{\ldots, m(X_{t-1}, X_t), m(X_t, X_{t+1}), \ldots\}$ forms a stochastic process called the SDF process, which is denoted $\{m(X_t, X_{t+1})\}$.

3.1.1. *Example: Consumption CAPM.* Consider the consumption CAPM with complete, frictionless markets and a representative agent who maximizes, subject to a budget constraint, expected utility given by

$$\sum_{s \geq 0} \beta^s E[u(C_{t+s})|\mathcal{I}_t]$$

where $\mathcal{I}_t$ is the information set at time $t$, $C_{t+s}$ is consumption of a representative good at date $t+s$, and $\beta$ is a time preference parameter. The SDF is given by

$$m(X_t, X_{t+1}) = \beta \frac{u'(C_{t+1})}{u'(C_t)} .$$

When $u(c) = (1-\gamma)^{-1}(c^{1-\gamma} - 1)$ the SDF takes the familiar form

$$m(X_t, X_{t+1}) = \beta G_{t+1}^{-\gamma}$$

where $G_{t+1} = C_{t+1}/C_t$ is aggregate consumption growth and $\gamma$ is the coefficient of relative risk aversion. Let $\{X_t\}$ be a strictly stationary and ergodic Markov state process and let $\mathcal{I}_t = \sigma(X_t)$. The consumption CAPM falls within the scope of the analysis of this paper when $G_t = g(X_{t-1}, X_t)$ for some known function $g : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$. For instance, one might take $X_t = (G_t, Y_t')'$.

3.1.2. *Example: External habit formation.* Following Abel (1990, 1999) and Galí (1994), consider an environment with complete, frictionless markets and a representative agent who maximizes, subject to a budget constraint, expected utility given by

$$\sum_{s \geq 0} \beta^s E[u(C_{t+s}, v_{t+s})|\mathcal{I}_t]$$

where $v_{t+s}$ is a benchmark level of consumption which the agent takes as exogenous. When

$$u(c_t, v_t) = \frac{(c_t/v_t)^{1-\gamma} - 1}{1-\gamma}, \quad v_t = C_t^{\gamma_0} C_{t-1}^{\gamma_1}$$

the SDF is of the form

$$m(X_t, X_{t+1}) = \beta G_{t+1}^{\lambda} G_t^{\alpha}$$

where $\lambda$ and $\alpha$ are functions of risk aversion $\gamma$ and the consumption externality parameters $\gamma_0$ and $\gamma_1$. This model falls within the scope of the analysis in this paper when $\{X_t\}$ is a strictly stationary and ergodic Markov state process, $\mathcal{I}_t = \sigma(X_t)$ and $G_t = g(X_t)$ for some known $g : \mathcal{X} \to \mathbb{R}$.

3.2. **The principal eigenpair.** As described in Hansen and Scheinkman (2009), the restriction of equation (10) to payoffs of the form $Z_{t+\tau} = \psi(X_{t+\tau})$ for suitable $\psi : \mathcal{X} \to \mathbb{R}$ defines a collection of linear operators $\{\mathbb{M}_\tau : \tau \geq 1\}$.[2] For each $\tau \geq 1$, the operator $\mathbb{M}_\tau$ is defined as

$$\mathbb{M}_\tau \psi(x) = E\left[ \left( \prod_{s=t}^{t+\tau-1} m(X_s, X_{s+1}) \right) \psi(X_{t+\tau}) \middle| X_t = x \right] . \tag{11}$$

Given a payoff function $\psi$, the operator $\mathbb{M}_\tau$ assigns a date-$t$ price to a claim to the date-$(t+\tau)$ state-dependent payoff $\psi(X_{t+\tau})$. For example, if $\iota(x) = 1$ for all $x \in \mathcal{X}$ then $\mathbb{M}_\tau \iota(X_t)$ is the date-$t$ price of a $\tau$-period zero-coupon bond.

As $\{X_t\}$ is a Markov process, the pricing operators factorize as $\mathbb{M}_\tau = \mathbb{M}_1^\tau$ for each $\tau \geq 1$. Let $\mathbb{M} := \mathbb{M}_1$ denote the 1-period pricing operator, i.e.

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

A function $\phi$ is an *eigenfunction* of the collection $\{\mathbb{M}_\tau : \tau \geq 1\}$ with *eigenvalue* $\rho$ if

$$\mathbb{M}_\tau \phi = \rho^\tau \phi \tag{12}$$

for all $\tau \geq 1$. If, in addition, $\phi$ is positive then $\phi$ is referred to as the *principal eigenfunction*, $\rho$ is the *principal eigenvalue*, and $(\rho, \phi)$ are the *principal eigenpair*.

Hansen and Scheinkman (2009) and Hansen (2012) show that principal eigenpairs may be used to decompose the SDF into its permanent and transitory components. That is,

$$m(X_t, X_{t+1}) = M_{t,t+1}^P M_{t,t+1}^T$$

where the permanent component of the SDF is

$$M_{t,t+1}^P = \rho^{-1} \frac{\phi(X_{t+1})}{\phi(X_t)} m(X_t, X_{t+1})$$

and the transitory component is

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

(cf. Equation (20) in Hansen (2012)).[3] The notion of permanent and transitory components employed here is different from that used in the study of nonstationary time series. For example, Beveridge and Nelson (1981) additively decompose a nonstationary time series into the sum of a random walk (martingale) permanent component and a stationary transitory component. In contrast, here the SDF process is multiplicatively decomposed into the product of $M_{t,t+1}^P$ and $M_{t,t+1}^T$ where the permanent component $M_{t,t+1}^P$ is a multiplicative martingale: $E[M_{t,t+1}^P|\mathcal{F}_t] = 1$ almost surely. By the definition of $M_{t,t+1}^P$, equation (11) may be rewritten as

$$\rho^{-\tau} \mathbb{M}_\tau \psi(x) = E\left[ \left( \prod_{s=t}^{t+\tau-1} M_{s,s+1}^P \right) \frac{\psi(X_{t+\tau})}{\phi(X_{t+\tau})} \middle| X_t = x \right] \phi(x)$$

[2]The set of "suitable" functions $\psi : \mathcal{X} \to \mathbb{R}$ will be defined subsequently.

[3]The definitions of the permanent and transitory components used here are the same as the definitions in Backus, Chernov, and Zin (2013), and correspond with what Alvarez and Jermann (2005) define as the growth in the permanent and transitory components of the pricing kernel process.

where $\rho^{-\tau}\mathbb{M}_\tau$ may be interpreted as an horizon-normalized price. Therefore, any differences between the horizon-normalized prices of claims to distinct future payoffs $\psi(X_{t+\tau})$ are due to differences between the covariation of the permanent component of the SDF and the scaled payoff $\psi(X_{t+\tau})/\phi(X_{t+\tau})$.

Under stochastic stability and integrability conditions, Hansen and Scheinkman (2009) and Hansen (2012) obtain a single-factor representation of the prices of long-horizon assets, namely

$$\lim_{\tau\to\infty}\rho^{-\tau}\mathbb{M}_\tau\psi(X_t) = \widetilde{E}[\psi(X)/\phi(X)]\phi(X_t) \tag{13}$$

where $\widetilde{E}[\cdot]$ denotes expectation under a "twisted" probability measure associated with the permanent component. Equation (13) shows that when $\tau$ is large, the yield implied by the date-$t$ price of a claim to $\psi(X_{t+\tau})$ is approximately $-\log\rho$, the *long-term yield*. Moreover, after discounting by $\rho$, state-dependence of the price is captured solely through $\phi(X_t)$. A restatement of (13) is provided in Theorem 4.2 below. This theorem shows how to calculate $\widetilde{E}[\cdot]$ in stationary discrete-time environments, and makes precise the sense in which the limit in (13) holds under the identification conditions presented in this paper.

3.3. **Entropies.** The *entropy of the permanent component of the SDF* is defined as

$$L(M^P_{t,t+1}) = \log E[M^P_{t,t+1}] - E[\log M^P_{t,t+1}]\,.$$

Backus, Chernov, and Zin (2013) refer to $L(M^P_{t,t+1})$ as the "long-horizon entropy". Alvarez and Jermann (2005) show that

$$L(M^P_{t,t+1}) \geq E[\log R_{t+1}] - E[\log R_{\infty,t+1}] \tag{14}$$

where $R_{t+1}$ is the gross return on a risky asset from time $t$ to $t+1$ and $R_{\infty,t+1}$ is the gross return on a risk-free bond with infinite maturity from time $t$ to $t+1$. For an asset pricing model to be consistent with observed returns on risky assets relative to long-term bonds, its permanent component must be large enough to satisfy the bound (14). In the stationary discrete-time environment considered in this paper, the entropy of the permanent component takes the convenient form

$$L(M^P_{t,t+1}) = \log\rho - E[\log m(X_t, X_{t+1})] \tag{15}$$

whenever $E[\log\phi(X_t)]$ and $E[\log m(X_t,X_{t+1})]$ are finite. Given $\rho$ and $m$, the premium on risky assets in excess of long-term bonds may be bounded by

$$\log\rho - E[\log m(X_t,X_{t+1})] \geq E[\log R_{t+1}] - E[\log R_{\infty,t+1}]$$

as a consequence of (14) and (15). By contrast, the entropy of the SDF is defined as

$$L(m(X_t,X_{t+1})) = \log E[m(X_t,X_{t+1})] - E[\log m(X_t,X_{t+1})] \tag{16}$$

and may be used to bound returns relative to short-term risk-free bonds:

$$L(m(X_t,X_{t+1})) \geq E[\log R_{t+1}] - E[\log R_{1,t+1}] \tag{17}$$

where $R_{1,t+1}$ is the gross return on a one-period risk-free bond from time $t$ to $t+1$ (Cochrane, 1992; Bansal and Lehmann, 1997; Backus, Chernov, and Martin, 2011).

The entropy of the SDF measures the "roughness" or "dispersion" of the SDF, whereas the entropy of the permanent component measures both the roughness of the SDF and the persistence of the SDF process $\{m(X_t, X_{t+1})\}$. This latter point is reflected by the bound (15), which shows that the entropy of the permanent component may be used to bound the return on risky assets relative to short-term bonds minus the term premium $E[\log R_{\infty,t+1}] - E[\log R_{1,t+1}]$. The return on risky assets relative to short-term bonds depends on the dispersion of the SDF (cf. expression (17)) whereas the term premium depends on the dynamics of the SDF process. If the SDF is i.i.d. (independent and identically distributed) each period, then the entropy of the SDF and the entropy of the permanent component of the SDF are equal and the term premium is zero.

3.4. **Robustness.** An attractive reason for focusing on the long-run implications of an asset pricing model is that different models can have the same long-run implications but different short-run implications. This property was first noted by Bansal and Lehmann (1997), and is explored further by Hansen (2012), Hansen and Scheinkman (2013) and Backus, Chernov, and Zin (2013). This robustness property makes the long-run implications of a model a powerful means with which to analyze dynamic asset pricing models.

Let $m$ and $m^*$ be two SDFs that differ by the ratio of two transitory terms, i.e.

$$m^*(X_t, X_{t+1}) = m(X_t, X_{t+1}) \frac{f(X_{t+1})}{f(X_t)}$$

for some positive function $f$. For instance, $m$ could be the SDF in the consumption CAPM and $f$ might be an external habit formation term or a term that represents a limiting version of recursive preferences (Hansen, 2012; Hansen and Scheinkman, 2013). Although the short-run implications of $m$ and $m^*$ may differ, the permanent components of $m$ and $m^*$ will be the same (and so the entropy of the permanent components of $m$ and $m^*$ will be the same), and $m$ and $m^*$ will imply the same long-term yield. This robustness property means that the long-run implications of classes of asset pricing models can be inferred from the analysis of one model.

## 4. Nonparametric identification

Nonparametric identification of the positive eigenfunction $\phi$ is a consequence of the law of motion of the state variables, the form of the SDF, and the space of functions to which the eigenfunction is assumed to belong. Hansen and Scheinkman (2009) study identification of the positive eigenfunction in continuous-time economies. They use Markov process theory to derive sufficient conditions for identification of the positive eigenfunction. This section presents nonparametric identification conditions for the positive eigenfunction in stationary discrete-time economies. The conditions are also sufficient for existence of the positive eigenfunction. A function-analytic approach is used to establish identification and existence: existence follows by application of the Perron-Frobenius theorem for positive integral operators, and identification is established by a version of the Kreĭn-Rutman theorem. A version of the long-term pricing result of Hansen and Scheinkman (2009) holds under the identification conditions.

Function-analytic methods have been used recently to study identification of positive eigenfunctions related to other operators in economics. Chen, Chernozhukov, Lee, and Newey (2013a) study nonparametric identification of a habit formation component in a semiparametric consumption CAPM using these methods. In ongoing work, Linton, Lewbel, and Srisuma (2011) and Escanciano and Hoderlein (2012) use related techniques to analyze nonparametric identification of marginal utilities of consumption in representative agent models. However, in each of these studies the model and operator analyzed is different from the operator studied here.

4.1. **Identification and existence.** The conditions presented below are sufficient for nonparametric identification and existence of the positive eigenfunction $\phi$. The conditions are stronger than required for identification, but are convenient for establishing both identification and the large sample properties of the estimators. Weaker nonparametric identification and existence conditions are presented in Appendix A. Alternative identification conditions for stationary discrete- and continuous-time environments are explored in Christensen (2013).

**Assumption 4.1.** *$\{X_t\}$ and $m$ satisfy the following conditions:*

*(i) $\{X_t\}$ is a strictly stationary and ergodic (first-order) Markov process with support $\mathcal{X} \subseteq \mathbb{R}^d$*
*(ii) the stationary distribution $Q$ of $\{X_t\}$ has density $q$ (wrt Lebesgue measure) s.t. $q(x) > 0$ almost everywhere*
*(iii) $(X_0, X_1)$ has joint density $f$ (wrt Lebesgue measure) s.t. $f(x_0, x_1) > 0$ almost everywhere and $f(x_0, x_1)/(q(x_0)q(x_1))$ is uniformly bounded away from infinity*
*(iv) $m : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ has $m(x_0, x_1) > 0$ almost everywhere and $E[m(X_0, X_1)^2] < \infty$ .*

Stationarity and ergodicity (Assumption 4.1(i)) is a stronger assumption than the irreducibility condition of Hansen and Scheinkman (2009). However, the requirement of stationarity is not necessarily restrictive. For example, consumption-based asset pricing models are typically written in terms of consumption growth to avoid potential nonstationarity in aggregate consumption (Hansen and Singleton, 1982; Gallant and Tauchen, 1989). Stationarity of the state process is also convenient for the derivation of the large sample properties of the estimators. Positivity of the joint density and boundedness of the ratio of joint to marginal densities (Assumptions 4.1(ii) and (iii)) is used both for identification and to develop the large sample theory. In particular, Assumption 4.1(iii) implies that $\{X_t\}$ is geometrically beta-mixing and geometrically rho-mixing. Boundedness of the ratio of the joint to marginal densities in Assumption 4.1(iii) is violated if $\{X_t\}$ is constructed by stacking a higher-order Markov process into a first-order process as the joint distribution of $(X_0, X_1)$ will be degenerate. Positivity of the SDF in Assumption 4.1(iv) is in line with the strict positivity of the SDF process assumed for identification in Hansen and Scheinkman (2009). Positivity of the SDF is satisfied in representative agent consumption-based asset pricing models for which

$$m(X_t, X_{t+1}) = \beta \frac{u'(C_{t+1})}{u'(C_t)}$$

provided the representative agent's marginal utility of consumption $u'(\cdot)$ is positive almost everywhere. Square integrability of the SDF is a standard assumption in asset pricing by no arbitrage (Hansen and Richard, 1987; Hansen and Renault, 2010).

Let $\mathscr{X}$ denote the Borel $\sigma$-algebra on $\mathcal{X}$ and let $L^2(Q) := L^2(\mathcal{X}, \mathscr{X}, Q)$ denote the space of all measurable functions $\psi : \mathcal{X} \to \mathbb{R}$ for which $\|\psi\| := E[\psi(X)^2]^{1/2} < \infty$. The inner product $\langle \psi_1, \psi_2 \rangle := E[\psi_1(X)\psi_2(X)]$ makes $L^2(Q)$ a Hilbert space. Under Assumption 4.1, the pricing operator $\mathbb{M} : L^2(Q) \to L^2(Q)$ may be rewritten as

$$\mathbb{M}\psi(x_0) = \int_{\mathcal{X}} m(x_0, x_1) \frac{f(x_0, x_1)}{q(x_0)q(x_1)} \psi(x_1) \, \mathrm{d}Q(x_1) \,. \tag{18}$$

Therefore $\mathbb{M}$ is an integral operator on $L^2(Q)$ of the form

$$\mathbb{M}\psi(x_0) = \int_{\mathcal{X}} \mathcal{K}(x_0, x_1)\psi(x_1) \, \mathrm{d}Q(x_1)$$

where the integral kernel $\mathcal{K} : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ is given by

$$\mathcal{K}(x_0, x_1) = m(x_0, x_1) \frac{f(x_0, x_1)}{q(x_0)q(x_1)} \,. \tag{19}$$

Assumption 4.1(ii)–(iv) implies that the kernel $\mathcal{K}$ is positive almost everywhere and

$$\int_{\mathcal{X}} \int_{\mathcal{X}} \mathcal{K}^2(x_0, x_1) \, \mathrm{d}Q(x_0) \, \mathrm{d}Q(x_1) < \infty \,.$$

Square-integrability of $\mathcal{K}$ implies $\mathbb{M}$ is Hilbert-Schmidt and therefore compact. The following identification and existence result is immediate by Theorems A.1 and A.2 in Appendix A.

**Theorem 4.1.** *Under Assumption 4.1,*

*(i)* $\mathbb{M}$ *has a unique (to scale) eigenfunction* $\phi \in L^2(Q)$ *such that* $\phi > 0$ *(almost everywhere)*
*(ii)* $\rho$ *is positive, has multiplicity one, and is the largest element of the spectrum of* $\mathbb{M}$*.*

There are a number of important implications of Theorem 4.1 beyond identification. First, that $\rho$ has multiplicity one means that both $\rho$ and $\phi$ are continuous with respect to small perturbations of $\mathbb{M}$. This continuity property is exploited in the derivation of the large sample properties of the estimators. Moreover, the fact that $\rho$ is the largest eigenvalue of $\mathbb{M}$ is useful for estimation: if an estimator can be constructed that is close to $\mathbb{M}$ in an appropriate sense, then its maximum eigenvalue should be close to $\rho$.

4.2. **Time reversal.** Under Assumption 4.1 the time-reversed pricing operator $\mathbb{M}^* : L^2(Q) \to L^2(Q)$ is defined formally as the adjoint of $\mathbb{M}$ and is given by

$$\mathbb{M}^*\psi^*(x) = E[m(X_0, X_1)\psi^*(X_0)|X_1 = x] \,. \tag{20}$$

The reversed pricing operator might be interpreted as a pricing operator in the economy with time run backwards, but with the same SDF as if time were being run forwards.[4] Under Assumption 4.1, $\mathbb{M}^*$ has a unique (to scale) positive eigenfunction $\phi^* \in L^2(Q)$ such that

$$\mathbb{M}^*\phi^* = \rho\phi^* \tag{21}$$

[4]See Rosenblatt (1971) for a discussion of time reversal for Markov processes.

(see Theorem A.2 in the appendix). The adjoint eigenfunction $\phi^*$ is an important component of the asymptotic variance of the estimators, and also appears in the restatement of the long-term pricing result of Hansen and Scheinkman (2009) below.

**Remark 4.1.** *It is convenient to impose the normalizations* $E[\phi(X)^2] = 1$ *and* $E[\phi(X)\phi^*(X)] = 1$, *which define* $\phi$ *and* $\phi^*$ *uniquely. These normalizations will be maintained hereafter.*

4.3. **Asymptotic single-factor pricing.** Hansen and Scheinkman (2009) and Hansen (2012) show that the positive eigenfunction $\phi$ captures state-dependence of the prices of long-horizon assets via the asymptotic single-factor pricing formula

$$\lim_{\tau\to\infty} \rho^{-\tau}\mathbb{M}_\tau \psi(X_t) = \widetilde{E}[\psi(X)/\phi(X)]\phi(X_t)$$

(see Section 7 of Hansen and Scheinkman (2009) and Section 6 of Hansen (2012) for precise statements of this result). Although Hansen and Scheinkman (2009) and Hansen (2012) define $\widetilde{E}[\cdot]$ as an expectation under a "twisted" probability measure associated with the permanent component, they do not show how to calculate the "twisted" probability measure.

The following theorem shows that an asymptotic pricing result holds for stationary discrete-time environments under Assumption 4.1. This theorem also shows how to calculate the twisted expectation $\widetilde{E}[\cdot]$ in stationary discrete-time environments.

**Theorem 4.2.** *Under Assumption 4.1, there exists a* $c > 0$ *such that*

$$\sup_{\psi\in L^2(Q):E[\psi(X)^2]\leq 1} \int_{\mathcal{X}} \left(\rho^{-\tau}\mathbb{M}^\tau \psi(x) - E[\psi(X)\phi^*(X)]\psi(x)\right)^2 \,\mathrm{d}Q(x) = O(e^{-c\tau})$$

*as* $\tau \to \infty$.

Theorem 4.2 shows that the scaled price $\rho^{-\tau}\mathbb{M}^\tau\psi(x)$ converges in mean square, uniformly over all payoff functions with unit norm, to $E[\psi(X)\phi^*(X)]\phi(x)$. Moreover, the approximation error vanishes exponentially quickly in the horizon $\tau$. Let $\widetilde{Q}$ denote the twisted probability measure used to define the expectation $\widetilde{E}[\cdot]$. It follows by equating $E[\psi(X)\phi^*(X)]$ and $\widetilde{E}[\psi(X)/\phi(X)]$ that the Radon-Nikodym derivative of $\widetilde{Q}$ with respect to $Q$ is

$$\frac{\mathrm{d}\widetilde{Q}(x)}{\mathrm{d}Q(x)} = \phi(x)\phi^*(x)$$

under Assumption 4.1. Theorem 4.2 is generalized to other $L^p(Q)$ spaces in Appendix A.

## 5. Estimation

This section introduces estimators of the positive eigenfunction, its eigenvalue, the long-term yield and the entropy of the permanent component of the SDF and presents the large sample properties of the estimators. It is assumed in this section that the SDF $m$ is known and the researcher has available a time series $\{X_0, X_1, \ldots, X_n\}$ of data on the state process. Thus this section applies when the researcher is interested in investigating the long-run implications of a given SDF $m$. Extension to the case in which the SDF is first estimated from data is discussed in Section 6.1.

Sieve methods are used here to reduce the infinite-dimensional eigenfunction problem to a finite-dimensional matrix eigenvector problem. Implementation of the estimators is as simple as computing the eigenvectors and eigenvalues of two appropriately chosen matrices. The estimators introduced below may also be used to numerically compute the long-run implications of fully specified models for which analytical solutions are unavailable.

Chen, Hansen, and Scheinkman (2000) and Gobet, Hoffmann, and Reiß (2004) use sieve techniques to nonparametrically estimate an eigenfunction of the (selfadjoint) conditional expectation operator of a scalar diffusion process. In the present paper the operator $\mathbb{M}$ will typically be nonselfadjoint which introduces some additional technicalities.[5] For example, if $\mathbb{M}$ is selfadjoint then $\phi = \phi^*$. Moreover, if $\mathbb{M}$ is selfadjoint the pair $(\rho, \phi)$ are equivalently defined as the solution to an infinite-dimensional maximization problem. However, this equivalence does not hold in the nonselfadjoint case. The consistency and convergence rate calculations for the estimators of $\rho$ and $\phi$ follow by simple modification of the arguments in Gobet, Hoffmann, and Reiß (2004). Estimation of the adjoint eigenfunction $\phi^*$, derivation of the asymptotic distribution of the eigenvalue estimator and related estimators via a perturbation expansion, and the semiparametric efficiency bound calculations are all new.

5.1. **Operator approximation.** Let the sieve spaces $\{B_K : K \geq 1\} \subset L^2(Q)$ be a sequence of subspaces of $L^2(Q)$ of dimension $K$. For each $K$, let $b_{K1}, \ldots, b_{KK}$ denote the sieve basis functions that span $B_K$. Common examples of sieve basis functions include polynomial splines, wavelets, Fourier series and orthogonal polynomials (see Chen (2007) for an overview). Any function $\psi \in B_K$ may be written as

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

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

$$\mathbf{G}_K = E[b^K(X) b^K(X)'] \,.$$

The relation $\psi \mapsto c_K(\psi)$ makes the space $B_K$ isomorphic to $\mathbb{R}^K$ under the inner product induced by the Gram matrix because

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

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

The infinite-dimensional eigenfunction problem $\mathbb{M}\phi = \rho\phi$ in $L^2(Q)$ is approximated by a $K$-dimensional eigenfunction problem in $B_K$. Let $\Pi_K^b : L^2(Q) \to B_K$ denote the orthogonal projection onto $B_K$. Consider the eigenfunction problem

$$\Pi_K^b \mathbb{M} \phi_K = \rho_K \phi_K \tag{22}$$

where $\rho_K$ is the largest eigenvalue of $\Pi_K^b \mathbb{M}$. Under the regularity conditions stated below, for all $K$ sufficiently large the approximate eigenfunction $\phi_K$ will be unique (to scale) and $\rho_K$ will be real-valued and positive. The approximate eigenfunction $\phi_K$ must belong to the space $B_K$. Consequently,

[5]The operators in the other applications discussed in Section 2 will also typically be nonselfadjoint.

$\phi_K$ can be written as

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

where $c_K = c(\phi_K)$ to simplify notation. Whenever $\mathbf{G}_K$ is invertible (this is guaranteed under the regularity conditions below) the approximate eigenvalue problem (22) may be rewritten as

$$\mathbf{G}_K^{-1}\mathbf{M}_K c_K = \rho_K c_K \tag{23}$$

where

$$\mathbf{M}_K = E[b^K(X_0) m(X_0, X_1) b^K(X_1)'] \tag{24}$$

and $\rho_K$ is the largest eigenvalue of $\mathbf{G}_K^{-1}\mathbf{M}_K$.

Approximation of the adjoint positive eigenfunction $\phi^*$ is more subtle. When a solution to (22) exists with $\rho_K$ real-valued, the adjoint of $\Pi_K^b \mathbb{M}$ has an eigenfunction $\phi_K^*$ with eigenvalue $\rho_K$. That is, there exists a $\phi_K^*$ such that

$$E[\phi_K^*(X) \Pi_K^b \mathbb{M} \psi(X)] = \rho_K E[\phi_K^*(X) \psi(X)]$$

for all $\psi \in L^2(Q)$. Let $\Pi_K^b \mathbb{M}|_{B_K}$ denote the restriction of $\Pi_K^b \mathbb{M}$ to the sieve space $B_K$. This restriction defines a linear operator $\Pi_K^b \mathbb{M}|_{B_K} : B_K \to B_K$. When a solution to (22) exists with $\rho_K$ real-valued, the adjoint of $\Pi_K^b \mathbb{M}|_{B_K}$ has an eigenfunction $\phi_K^\star$ with eigenvalue $\rho_K$. That is,

$$E[\phi_K^\star(X) \Pi_K^b \mathbb{M} \psi_K(X)] = \rho_K E[\phi_K^\star(X) \psi_K(X)]$$

for all $\psi_K \in B_K$. The notation $\star$ in place of $*$ is used to denote that $\phi_K^\star$ is the eigenfunction of the adjoint of $\Pi_K^b \mathbb{M}|_{B_K}$ and that $\phi_K^*$ is the eigenfunction of the adjoint of $\Pi_K^b \mathbb{M}$. Although $\Pi_K^b \phi_K^* = \phi_K^\star$, it is not generally the case that $\phi_K^\star = \phi_K^*$. The approximate adjoint eigenfunction $\phi_K^\star$ belongs to the sieve space $B_K$. Therefore, $\phi_K^\star$ may be written as

$$\phi_K^\star(x) = b^K(x)' c_K^\star$$

where $c_K^\star = c_K(\phi_K^\star)$ to simplify notation. When $\rho_K$ is real-valued and $\mathbf{G}_K$ is invertible, the vector $c_K^\star$ solves

$$\mathbf{G}_K^{-1}\mathbf{M}_K' c_K^\star = \rho_K c_K^\star\,. \tag{25}$$

where $\rho_K$ is the largest eigenvalue of $\mathbf{G}_K^{-1}\mathbf{M}_K'$.

In summary, the infinite-dimensional eigenfunctions $\phi$ and $\phi^*$ are approximated by $\phi_K$ and $\phi_K^\star$, where

$$\begin{aligned} \phi_K(x) &= b^K(x)' c_K \\ \phi_K^\star(x) &= b^K(x)' c_K^\star \end{aligned}$$

and $c_K$ and $c_K^\star$ solve

$$\begin{aligned} \mathbf{G}_K^{-1}\mathbf{M}_K c_K &= \rho_K c_K \\ \mathbf{G}_K^{-1}\mathbf{M}_K' c_K^\star &= \rho_K c_K^\star \end{aligned}$$

where $\rho_K$ is the largest eigenvalue of both $\mathbf{G}_K^{-1}\mathbf{M}_K$ and $\mathbf{G}_K^{-1}\mathbf{M}_K'$. Under the regularity conditions below, unique solutions to these eigenvector problems exist for all $K$ sufficiently large. As, $\phi_K$, $\phi_K^*$ and $\phi_K^\star$ are only defined up to scale, it is convenient to impose the sign normalizations $E[\phi_K(X)\phi(X)] \geq 0$, $E[\phi_K^*(X)\phi(X)^*] \geq 0$ and $E[\phi_K^\star(X)\phi(X)^*] \geq 0$ and the scale normalizations $E[\phi_K(X)^2] = 1$, $E[\phi_K(X)\phi_K^*(X)] = 1$ and $E[\phi_K(X)\phi_K^\star(X)] = 1$. These sign- and scale normalizations will be maintained hereafter, and define $\phi_K$, $\phi_K^*$ and $\phi_K^\star$ uniquely.

5.2. **Estimators.** The matrices $\mathbf{G}_K$ and $\mathbf{M}_K$ can be estimated from data $\{X_0, X_1, \ldots, X_n\}$ by replacing the population expectations with their sample analogues, namely

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

and

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

The estimator $\widehat{\rho}$ of $\rho$ is the largest eigenvalue of $\widehat{\mathbf{G}}_K^{-1}\widehat{\mathbf{M}}_K$, i.e.

$$\widehat{\rho} = \lambda_{\max}(\widehat{\mathbf{G}}_K^{-1}\widehat{\mathbf{M}}_K).$$

When $\widehat{\rho}$ is real valued (which it is with probability approaching one under the regularity conditions below) let $\widehat{c}$ and $\widehat{c}^\star$ solve the matrix eigenvalue problems

$$\widehat{\mathbf{G}}_K^{-1}\widehat{\mathbf{M}}_K\widehat{c} = \widehat{\rho}\,\widehat{c} \tag{26}$$

$$\widehat{\mathbf{G}}_K^{-1}\widehat{\mathbf{M}}_K'\widehat{c}^\star = \widehat{\rho}\,\widehat{c}^\star. \tag{27}$$

The estimators of $\phi$ and $\phi^*$ are

$$\widehat{\phi}(x) = b^K(x)'\widehat{c} \tag{28}$$

$$\widehat{\phi}^\star(x) = b^K(x)'\widehat{c}^\star. \tag{29}$$

As $\widehat{\phi}$ and $\widehat{\phi}^\star$ are only defined up to sign and scale, impose the sign normalizations $E[\widehat{\phi}(X)\phi_K(X)] \geq 0$ and $E[\widehat{\phi}^\star(X)\phi_K^\star(X)] \geq 0$ and the scale normalizations $E[\widehat{\phi}(X)^2] = 1$ and $E[\widehat{\phi}(X)\widehat{\phi}^\star(X)] = 1$. The estimators $\widehat{\phi}$ and $\widehat{\phi}^\star$ are defined uniquely under these normalizations.

Recall that the long-term yield is $y = -\log\rho$ and the entropy of the permanent component of the SDF is $L = \log\rho - E[\log m(X_0, X_1)]$. In light of these definitions,

$$\widehat{y} = -\log\widehat{\rho} \tag{30}$$

and

$$\widehat{L} = \log\widehat{\rho} - \frac{1}{n}\sum_{t=0}^{n-1}\log m(X_t, X_{t+1}) \tag{31}$$

are natural estimators of $y$ and $L$.

5.3. **Regularity conditions and convergence rates.** The following regularity conditions, in conjunction with Assumption 4.1, are sufficient to establish consistency and convergence rates of the estimators. Estimation of the positive eigenfunctions of a collection of operators under higher-level conditions is discussed in Appendix C.

As in Newey (1997), define the sequence of constants $\zeta_0(K) = \|\sqrt{b^K(x)'b^K(x)}\|_\infty$. For example, $\zeta_0(K) = O(\sqrt{K})$ for polynomial spline, Fourier series and wavelet bases and $\zeta_0(K) = O(K)$ for polynomial bases on appropriate domains (see, e.g., Newey (1997)). Let $\widetilde{b}^K(x) = E[b^K(X)b^K(X)']^{-1}b^K(x)$ denote a vector of orthonormalized sieve basis functions. Define the orthonormalized estimators

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

and their orthonormalized population counterparts

$$\begin{aligned}\widetilde{\mathbf{G}}_K &= E[\widetilde{b}^K(X)\widetilde{b}^K(X)'] \\ \widetilde{\mathbf{M}}_K &= E[\widetilde{b}^K(X_0)m(X_0,X_1)\widetilde{b}^K(X_1)']\end{aligned}$$

where $\widetilde{\mathbf{G}}_K = I_K$ (the $K\times K$ identity matrix) by virtue of orthonormalization. The orthonormalized estimators are infeasible in practice because $Q$ is typically unknown; however it is convenient to define the regularity conditions in terms of these quantities.

Let $\|\cdot\|_2$ denote the matrix spectral norm when applied to matrices and the Euclidean norm when applied to vectors. That is, if $\mathbf{A}_K$ is a $K\times K$ matrix and $c = (c_1,\ldots,c_K)' \in \mathbb{R}_K$ then

$$\begin{aligned}\|\mathbf{A}_K\|_2 &= \sup\{\|\mathbf{A}_K c\|_2 : c\in\mathbb{R}^K, \|c\|_2 = 1\} \\ \|c\|_2 &= \left(\sum_{k=1}^K c_k^2\right)^{-1/2}.\end{aligned}$$

Recall that $\|\cdot\|$ denotes the $L^2(Q)$ norm when applied to functions in $L^2(Q)$. Let $\|\cdot\|$ also denote the operator norm when applied to linear operators on $L^2(Q)$. That is, if $A : L^2(Q)\to L^2(Q)$ is a linear operator then $\|A\| = \sup\{\|Af\| : f\in L^2(Q), \|f\|\le 1\}$.

Define the $K$-vectors $\widetilde{c}_K$ and $\widetilde{c}_K^\star$ such that $\phi_K(x) = \widetilde{b}^K(x)'\widetilde{c}_K$ and $\phi_K^\star(x) = \widetilde{b}^K(x)'\widetilde{c}_K^\star$. Let $\{\bar\eta_{n,K}, \eta_{n,K} : n, K\ge 1\}$ be sequences of positive real numbers such that

$$\max\left\{\|\widehat{\widetilde{\mathbf{G}}}_K - I_K\|_2, \|\widehat{\widetilde{\mathbf{M}}}_K - \widetilde{\mathbf{M}}_K\|_2\right\} = O_p(\bar\eta_{n,K})$$

and

$$\begin{aligned}\max\Big\{&\|(\widehat{\widetilde{\mathbf{G}}}_K - \widetilde{\mathbf{G}}_K)\widetilde{c}_K\|_2, \|(\widehat{\widetilde{\mathbf{G}}}_K - \widetilde{\mathbf{G}}_K)\widetilde{c}_K^\star/\|\widetilde{c}_K^\star\|_2\|_2, \\ &\|(\widehat{\widetilde{\mathbf{M}}}_K - \widetilde{\mathbf{M}}_K)\widetilde{c}_K\|_2, \|(\widehat{\widetilde{\mathbf{M}}}_K - \widetilde{\mathbf{M}}_K)\widetilde{c}_K^\star/\|\widetilde{c}_K^\star\|_2\|_2\Big\} = O_p(\eta_{n,K}).\end{aligned}$$

The inequality $\|\mathbf{A}_K c\|_2 \le \|\|\mathbf{A}_K\|_2\|c\|_2$ holds by definition of $\|\cdot\|_2$ and implies $\eta_{n,K} = O(\bar\eta_{n,K})$. Different values of $\bar\eta_{n,K}$ and $\eta_{n,K}$ will be obtained depending on the number of moments of $m(X_t,X_{t+1})$.

**Remark 5.1.** *Appendix C provides further details as to how to calculate $\bar{\eta}_{n,K}$ and $\eta_{n,K}$. If Assumption 4.1 holds, then $\bar{\eta}_{n,K} = \zeta_0(K)^2/\sqrt{n}$ and $\eta_{n,K} = \zeta_0(K)^2/\sqrt{n}$. If, in addition, $m$ is bounded and $\zeta_0(K)(\log n)/\sqrt{n} = o(1)$, then $\bar{\eta}_{n,K} = \zeta_0(K)(\log n)\sqrt{n}$ and $\eta_{n,K} = \zeta_0(K)/\sqrt{n}$.*

**Assumption 5.1.** *The following regularity conditions are satisfied:*

*(i)* $\|\Pi_K^b \mathbb{M} - \mathbb{M}\| = O(\delta_K)$ *where* $\delta_K = o(1)$ *as* $K \to \infty$
*(ii)* $\bar{\eta}_{n,K} = o(1)$ *as* $n, K \to \infty$
*(iii)* $\lambda_{\min}(\mathbf{G}_K) \geq \underline{\lambda} > 0$ *for each* $K \geq 1$
*(iv) there exists a sequence* $\{h_K^* : K \geq 1\}$ *with* $h_K^* \in B_K$ *such that* $\|\phi^* - h_K^*\| = O(\delta_K^*)$.

Assumption 5.1(i) requires that the range of $\mathbb{M}$ can be uniformly well approximated over the sieve space $B_K$, with the approximation error vanishing as the dimension of the sieve space increases. Assumption 5.1(i) also implies that $\|\Pi_K^b \phi - \phi\| = O(\delta_K)$. The weaker condition $\|\Pi_K^b \mathbb{M} - \mathbb{M}\| = o(1)$ and $\|\Pi_K^b \phi - \phi\| = O(\delta_K)$ suffices to calculate the following convergence rates for $\widehat{\rho}$ and $\widehat{\phi}$, however the stronger form presented in Assumption 5.1(i) is useful for derivation of the limit theory. Assumption 5.1(ii) is a condition on the maximum rate at which $K$ can increase with $n$ while maintaining consistency of the matrix estimators. Assumption 5.1(iii) is a standard condition for nonparametric estimation with a linear sieve space (see, e.g., Newey (1997); Chen and Pouzo (2012)). Assumption 5.1(iii) can be relaxed to allow $\lambda_{\min}(\mathbf{G}_K) \searrow 0$ as $K$ increases, but this may slow the convergence rates. Assumption 5.1(iv) requires that $\phi^*$ can be approximated by a sequence of elements of the sieve space, with the approximation error vanishing as $K$ increases. The condition on $\phi^*$ in Assumption 5.1(iv) can be dropped if $\mathbb{M}$ is selfadjoint (since $\phi = \phi^*$ in that case). When $\mathbb{M}$ is nonselfadjoint the separate treatment of $\phi$ and $\phi^*$ is required because Assumption 5.1(i) does not guarantee that $\mathbb{M}^*$, and therefore $\phi^*$, can be approximated well over $B_K$. Assumptions 5.1(i) and (iv) can be motivated by imposing smoothness conditions on the kernel $\mathcal{K}$ and choosing an appropriate sieve, as in the example below.

The following theorem establishes consistency of $\widehat{\rho}$, and mean square convergence rates of $\widehat{\phi}$ and $\widehat{\phi}^\star$ as $n \to \infty$.

**Theorem 5.1.** *Under Assumptions 4.1 and 5.1, there is a set whose probability approaches one on which $\widehat{\rho}$ is real and positive and has multiplicity one, and*

*(i)* $|\widehat{\rho} - \rho| = O_p(\delta_K + \eta_{n,K})$
*(ii)* $\|\widehat{\phi} - \phi\| = O_p(\delta_K + \eta_{n,K})$
*(iii)* $\|\widehat{\phi}^\star/\|\widehat{\phi}^*\| - \phi^*/\|\phi^*\|\| = O_p(\delta_K^* + \eta_{n,K})$.

The rates of convergence in Theorem 5.1 exhibit the standard bias-variance tradeoff in nonparametric estimation. The "bias term" is $O(\delta_K)$ (or $O(\delta_K^*)$ for $\phi^*$) which measures error in approximating $\phi$ and $\phi^*$ by their $K$-dimensional counterparts $\phi_K$ and $\phi_K^\star$. The "variance term" is $O_p(\eta_{n,K})$ which measures the difference between the estimators $\widehat{\phi}$ and $\widehat{\phi}^\star$ and their sample counterparts $\phi_K$ and $\phi_K^\star$.

Consistency and preliminary rates of convergence for $\widehat{y}$ and $\widehat{L}$ are established in the following Corollary. These estimators will be shown to be $\sqrt{n}$-consistent under stronger assumptions in Section 5.4.

**Corollary 5.1.** *Under the assumptions of Theorem 5.1,* $|\widehat{y} - y| = O_p(\delta_K + \eta_{n,K})$. *If, in addition,* $E[(\log m(X_0, X_1))^2] < \infty$, *then* $|\widehat{L} - L| = O_p(\delta_K + \eta_{n,K} + n^{-1/2})$.

Let $\|\cdot\|_\infty$ denote the sup norm. That is, if $f : \mathcal{X} \to \mathbb{R}$ then $\|f\|_\infty = \sup_x\{|f(x)| : x \in \mathcal{X}\}$. Sup-norm rates of convergence of $\widehat{\phi}$ and $\widehat{\phi}^\star$ follow from the $L^2(Q)$ rates by standard arguments for sieve estimation under a slight strengthening of Assumptions 5.1(i) and 5.1(iv). The sup-norm rates obtained in Corollary 5.2 below are useful for constructing estimators of the asymptotic variance of $\widehat{\rho}$, $\widehat{y}$, and $\widehat{L}$.

**Assumption 5.2.** *There exist sequences of functions* $\{g_K : K \geq 1\}$ *and* $\{g_K^* : K \geq 1\}$ *such that* $g_K \in B_K$ *and* $g_K^* \in B_K$ *for each* $K \geq 1$ *and:*

*(i)* $\|\phi - g_K\|_\infty = O(\delta_K)$
*(ii)* $\|\phi^* - g_K^*\|_\infty = O(\delta_K^*)$.

A sufficient condition for Assumption 5.1(i) and 5.2(i) is that $\mathbb{M}$ maps the $L^2(Q)$ unit ball to a subspace $\mathcal{S} \subset L^2(Q)$ over which the sieve has uniformly good approximation properties in sup-norm, i.e. $\{\mathbb{M}\psi : \|\psi\| \leq 1\} \subseteq \mathcal{S}$, and $\sup_{f \in \mathcal{S}} \inf_{b(f) \in B_K} \|f - b(f)\|_\infty = O(\delta_K)$. This condition can be motivated by imposing smoothness conditions on the integral kernel $\mathcal{K}$ and using an appropriate sieve, as in the example below. Assumption 5.2(ii) is a sufficient condition for Assumption 5.1(iv) by virtue of the relation $\|\phi^* - g_K^*\| \leq \|\phi^* - g_K^*\|_\infty$.

**Corollary 5.2.** *Under Assumptions 4.1, 5.1, and 5.2,*

*(i)* $\|\widehat{\phi} - \phi\|_\infty = O_p(\zeta_0(K)(\delta_K + \eta_{n,K}))$
*(ii)* $\|\widehat{\phi}^\star/\|\widehat{\phi}^\star\| - \phi^*/\|\phi^*\|\|_\infty = O_p(\zeta_0(K)(\delta_K^* + \eta_{n,K}))$.

5.3.1. *Example: Smooth kernel.* This example shows how to calculate $\delta_K$ and $\bar{\eta}_{n,K}$ under primitive smoothness conditions on the integral kernel $\mathcal{K}$ defined in expression (19). For any $p > 0$ let $[p]$ denote the maximum integer less than or equal to $p$. Let $C^{[p]}(\mathcal{X})$ denote the space of $[p]$-times continuously differentiable functions with support $\mathcal{X}$. Given a $d$-tuple $\alpha = (\alpha_1, \ldots, \alpha_d)$ of nonnegative integers, set $|\alpha| = \alpha_1 + \ldots + \alpha_d$ and let $D^\alpha$ denote the differential operator

$$D^\alpha = \frac{\partial^{|\alpha|}}{\partial x_1^{\alpha_1} \ldots \partial x_d^{\alpha_d}}\,.$$

Define the Hölder norm $\|\cdot\|_{\infty,p}$ on $C^{[p]}(\mathcal{X})$ by

$$\|f\|_{\infty,p} = \max_{|\alpha| \leq [p]} \sup_{x \in \mathcal{X}} |D^\alpha f(x)| + \max_{|\alpha| = [p]} \sup_{x, x' \in \mathcal{X}, x \neq x'} \frac{|D^\alpha f(x) - D^\alpha f(x')|}{\|x - x'\|^{p - [p]}}\,.$$

Let $\Lambda^p(\mathcal{X}) = \{f \in C^{[p]}(\mathcal{X})$ such that $\|f\|_{\infty,p} < \infty\}$ denote the Hölder space of $p$-smooth functions and let $\Lambda_c^p(\mathcal{X}) = \{f \in \Lambda^p(\mathcal{X})$ such that $\|f\|_{\infty,p} \leq c\}$ denote the Hölder ball of smoothness $p$ and radius $c$.

Let Assumptions 4.1 and 5.1(iii) hold and assume additionally that (i) $\mathcal{X} \subset \mathbb{R}^d$ is compact, rectangular and has nonempty interior, (ii) $q$ is continuous and uniformly bounded away from zero on $\mathcal{X}$, (iii) there is a $p > 0$ and finite constant $C$ such that

$$\int_{\mathcal{X}} (D^\alpha \mathcal{K}(x_0, x_1))^2 \,\mathrm{d}Q(x_1) \leq C^2$$

for each $|\alpha| \leq [p]$ and

$$\int_{\mathcal{X}} (D^\alpha \mathcal{K}(x, x_1) - D^\alpha \mathcal{K}(x', x_1))^2 \,\mathrm{d}Q(x_1) \leq C^2 \|x - x'\|^{2(p-[p])}$$

for each $|\alpha| = [p]$, and (iv) $B_K$ is a spanned by a (tensor product) of polynomial splines of degree $v > p$ with uniformly bounded mesh ratio (see Schumaker (2007)).

Conditions (ii) and (iii) imply that $\{\mathbb{M}f : \|f\| \leq 1\} \subset \Lambda^p_c(\mathcal{X})$ for some finite $c$ and, in particular, that $\phi \in \Lambda^p_c(\mathcal{X})$. Assumptions 5.1(i) and 5.2(i) are satisfied with $\delta_K = O(K^{-p/d})$ for a polynomial spline sieve under conditions (i) and (iv) (Schumaker, 2007, Chapter 12). Condition (iv) implies that $\zeta_0(K) = O(\sqrt{K})$ (see, for example, Newey (1997)), so $\eta_{n,K} = O(K/\sqrt{n})$ if $m$ is unbounded and $\eta_{n,K} = O(\sqrt{K}/\sqrt{n})$ if $m$ is bounded. The following mean-square and sup-norm convergence rates obtain:

(a) If $m$ is unbounded, choosing $K \asymp n^{d/(2p+2d)}$ yields $\|\widehat{\phi} - \phi\| = O_p(n^{-p/(2p+2d)})$. If $p > \frac{1}{2}d$ this choice of $K$ yields a sup-norm rate of convergence of $\|\widehat{\phi} - \phi\|_\infty = O_p(n^{(d/2-p)/(2p+2d)})$.
(b) If $m$ is bounded, choosing $K \asymp n^{d/(2p+d)}$ yields $\|\widehat{\phi} - \phi\| = O_p(n^{-p/(2p+d)})$. This is the same as the minimax optimal mean-square convergence rate for a nonparametric regression estimator of a $p$-smooth function of $d$ variables (Stone, 1982). If $p > \frac{1}{2}d$ this choice of $K$ yields a sup-norm rate of convergence of $\|\widehat{\phi} - \phi\|_\infty = O_p(n^{(d/2-p)/(2p+d)})$.

This example shows that reasonable mean-square convergence rates for $\widehat{\phi}$ can be obtained when $\mathcal{K}(x_0, x_1)$ is smooth in $x_0$. The kernel $\mathcal{K}$ does not necessarily need to be smooth in $x_1$ to attain these rates. For example, if $m(x_0, x_1) = m(x_1)$ then $\mathcal{K}$ may satisfy the above smoothness conditions provided $f(x_1|x_0)$ is sufficiently smooth in $x_0$ even if $m(x_1)$ is kinked or discontinuous in $x_1$. However, such kinks or discontinuities may affect how well $\phi^*$ can be approximated, because $\mathbb{M}^*$ is an integral operator with kernel $\mathcal{K}^*$ given by $\mathcal{K}^*(x_0, x_1) = \mathcal{K}(x_1, x_0)$.

5.4. **Asymptotic inference.** A feasible means of conducting asymptotic inference for the eigenvalue $\rho$, the long-term yield $y$, and the entropy of the permanent component of the SDF $L$ is now provided. The asymptotic distribution of the estimators is derived via a perturbations expansion. This approach is distinct from the usual Taylor-series arguments used in the derivation of the limit distribution of extremum estimators.

Under the regularity conditions below, the estimator $\widehat{\rho}$ is asymptotically linear and its influence function is formed from $m$, $\phi$, $\phi^*$, and $\rho$, i.e.

$$\sqrt{n}(\widehat{\rho} - \rho) = \frac{1}{\sqrt{n}} \sum_{t=0}^{n-1} \{\phi^*(X_t) m(X_t, X_{t+1}) \phi(X_{t+1}) - \rho \phi^*(X_t) \phi(X_t)\} + o_p(1)\,. \tag{32}$$

Conveniently, the summands in expression (32) form a martingale difference sequence with respect to $\mathcal{F}_t = \sigma(X_t, X_{t-1}, \ldots)$. The asymptotic distribution for $\widehat{\rho}$ then follows by applying a central limit theorem for martingales with stationary and ergodic differences. The asymptotic distributions of $\widehat{y}$ and $\widehat{L}$ also follow straightforwardly from the expansion (32). An additional assumption is needed to ensure the representation (32) is valid and that the asymptotic variances of the estimators are well defined.

**Assumption 5.3.** *The following moment and rate conditions are satisfied:*

*(i) Either (a) or (b) holds:*
  *(a) $m$ is bounded, $E[\phi(X)^4] < \infty$ and $E[\phi^*(X)^4] < \infty$*
  *(b) $E[m(X_0, X_1)^6] < \infty$, $E[\phi(X)^6] < \infty$ and $E[\phi^*(X)^6] < \infty$*

*(ii) $\bar{\eta}_{n,K} = o(n^{-1/4})$, $\delta_K = o(n^{-1/2})$, and $\zeta_0(K) \max\{\delta_K^*, \delta_K, \bar{\eta}_{n,K}\} = o(1)$*

*(iii) $E[|\log m(X_0, X_1)|^2] < \infty$.*

Assumption 5.3(i)(iii) guarantees that the asymptotic variance of the estimators are well defined. Assumption 5.3(ii) ensures the estimation and approximation errors vanish sufficiently quickly that the expansion (32) is valid. The condition $\bar{\eta}_{n,K} = o(n^{-1/4})$ is analogous to the requirement in semiparametric extremum estimation that the estimator of the nonparametric part converges at least as fast as $n^{-1/4}$ to obtain $\sqrt{n}$-consistency of the estimator of the parametric part. The condition $\delta_K = o(n^{-1/2})$ ensures the bias term $\rho_K - \rho$ vanishes sufficiently quickly that it does not affect the asymptotic distribution for $\widehat{\rho}$. The condition $\zeta_0(K) \max\{\delta_K^*, \delta_K, \bar{\eta}_{n,K}\} = o(1)$ is used, inter alia, to establish consistency of the asymptotic variance estimators introduced below.

If $\{Z_t\}$ is a real-valued stationary stochastic process, define the long-run variance of $\{Z_t\}$ as

$$\operatorname{lrvar}(Z_t) = \sum_{t=-\infty}^{\infty} E[Z_0 Z_t] \,.$$

Let

$$V_\rho = E\left[\{\phi^*(X_0) m(X_0, X_1) \phi(X_1) - \rho \phi^*(X_0) \phi(X_0)\}^2\right] \tag{33}$$

$$\begin{aligned} V_L = {} & \operatorname{lrvar}(\rho^{-1} \phi^*(X_t) m(X_t, X_{t+1}) \phi(X_{t+1}) - \phi^*(X_t) \phi(X_t) \\ & - \log m(X_t, X_{t+1}) + E[\log m(X_0, X_1)]) \,. \end{aligned} \tag{34}$$

Assumptions 4.1 and 5.1 provide that $V_\rho$ and $V_L$ are well defined.

**Theorem 5.2.** *Under Assumptions 4.1, 5.1, 5.2, and 5.3, if $V_\rho > 0$ and $V_L > 0$, then*

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

*(ii) $\sqrt{n}(\widehat{y} - y) \to_d N(0, \rho^{-2} V_\rho)$*

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

The conditions $V_\rho > 0$ and $V_L > 0$ exclude cases in which the limit distributions of the estimators are degenerate. For example, if $m(x_0, x_1) = c$ for some positive constant $c$ then $\phi(x) = 1$, $\phi^*(x) = 1$, and $\rho = c$ irrespective of the law of motion of the state variables, in which case $V_\rho = 0$ and $V_L = 0$.

With Theorem 5.2 in hand it remains to provide consistent variance estimators. The rates of convergence in Theorem 5.1 and 5.2 are for estimators whose scale has been normalized under the true distribution $Q$. Let $\widehat{\phi}^f$ and $\widehat{\phi}^{\star f}$ denote versions of $\widehat{\phi}$ and $\widehat{\phi}^\star$ normalized under the empirical measure, so that

$$\frac{1}{n}\sum_{t=0}^{n-1}\widehat{\phi}^f(X_t)^2 = 1, \quad \frac{1}{n}\sum_{t=0}^{n-1}\widehat{\phi}^f(X_t)\widehat{\phi}^{\star f}(X_t) = 1\,.$$

**Corollary 5.3.** *Under Assumptions 4.1 and 5.1,*

*(i)* $\|\widehat{\phi}^f - \phi\| = O_p(\delta_K + \bar{\eta}_{n,K})$
*(ii)* $\|\widehat{\phi}^{\star f} - \phi^*\| = O_p(\delta_K + \delta_K^* + \bar{\eta}_{n,K})\,.$

*If, in addition, Assumption 5.2 holds, then*

*(iii)* $\|\widehat{\phi}^f - \phi\|_\infty = O_p(\zeta_0(K)(\delta_K + \bar{\eta}_{n,K}))$
*(iv)* $\|\widehat{\phi}^{\star f} - \phi^*\|_\infty = O_p(\zeta_0(K)(\delta_K + \delta_K^* + \bar{\eta}_{n,K}))$.

The asymptotic variance estimators for $\widehat{\rho}$ and $\widehat{y}$ are constructed by replacing the population quantities in $V_\rho$ and $\rho^{-2}V_\rho$ by feasible sample analogues. To simplify notation, for any $f : \mathcal{X} \to \mathbb{R}$ let $f_t = f(X_t)$, and let $m_{t,t+1} = m(X_t, X_{t+1})$. The estimator of $V_\rho$ is

$$\widehat{V}_\rho = \frac{1}{n}\sum_{t=0}^{n-1}\left(\widehat{\phi}_t^{\star f} m_{t,t+1}\widehat{\phi}_{t+1}^f - \widehat{\rho}\widehat{\phi}_t^{\star f}\widehat{\phi}_t^f\right)^2\,. \tag{35}$$

No sample mean correction is required because $\sum_{t=0}^{n-1}(\widehat{\phi}_t^{\star f} m_{t,t+1}\widehat{\phi}_{t+1}^f - \widehat{\rho}\widehat{\phi}_t^{\star f}\widehat{\phi}_t^f) = 0$ by definition of $\widehat{\phi}$ and $\widehat{\phi}^\star$. The estimator of the asymptotic variance of $\widehat{y}$ is $\widehat{\rho}^{-2}\widehat{V}_\rho$.

Estimating $V_L$ requires estimating a long-run variance. The following approach for conducting inference on $L$ uses an orthogonal series long-run variance (OSLRV) estimator of Phillips (2005) in conjunction with fixed-bandwidth asymptotics as in Chen, Liao, and Sun (2012).[6] The estimator will be asymptotically $\chi^2$-distributed and therefore inconsistent. However, asymptotic inference for $L$ can still be performed using this OSLRV estimator and the asymptotic distribution for $\widehat{L}$ developed in Theorem 5.2: the only difference is that Gaussian critical values are replaced by $t$ critical values.

Let $\{h_j : j \geq 0\}$ be a continuously differentiable orthonormal basis for $L^2([0,1], \mathscr{B}([0,1]), Leb)$ (where $\mathscr{B}([0,1])$ denotes the Borel $\sigma$-algebra on $[0,1]$ and $Leb$ is Lebesgue measure), such as a cosine basis or a Legendre polynomial basis. Let $h_0 = 1$, whence $\int_0^1 h_j(u)\,\mathrm{d}u = 0$ for each $j \geq 1$ by

[6]There is a large literature on consistent long-run variance estimation using kernel-based truncated lag estimators following Parzen (1957) (standard econometric references include Newey and West (1987) and Andrews (1991)). To ensure consistency of these estimators, the truncation lag is required to increase at an appropriate rate with the sample size. Recent research has shown that, in some circumstances, asymptotic inference using consistent kernel-based truncated-lag estimators can suffer considerable size and power distortions in finite samples. To this end, a literature has developed that explores inference under alternative bandwidth asymptotics (see, e.g., Kiefer, Vogelsang, and Bunzel (2000); Jansson (2004); Müller (2007); Sun, Phillips, and Jin (2008)). Preliminary Monte Carlo simulations (not reported) revealed that the coverage probabilities of asymptotic confidence intervals for $L$ constructed using a consistent kernel-based truncated lag estimator were sensitive to both the choice of kernel and bandwidth.

orthogonality. For each $j = 1, \ldots, J$, define

$$\widehat{\Lambda}_j = \frac{1}{\sqrt{n}} \sum_{t=0}^{n-1} h_j \left( \frac{t+1}{n} \right) \widehat{\Delta}_{t,t+1} \,.$$

where

$$\begin{aligned} \widehat{\Delta}_{t,t+1} &= \widehat{\rho}^{-1} \left( \widehat{\phi}_t^{\star f} m_{t,t+1} \widehat{\phi}_{t+1}^{f} - \widehat{\rho} \widehat{\phi}_t^{\star f} \widehat{\phi}_t^{f} \right) - \left( \log m_{t,t+1} - \overline{\log m}_n \right) \\ \overline{\log m}_n &= n^{-1} \sum_{t=0}^{n-1} \log m_{t,t+1} \,. \end{aligned}$$

The OSLRV estimator $\widehat{V}_{L,J}^{os}$ for $V_L$ using $J$ basis functions is defined as

$$\widehat{V}_{L,J}^{os} = \frac{1}{J} \sum_{j=1}^{J} \widehat{\Lambda}_j^2 \,. \tag{36}$$

The estimator $\widehat{V}_{L,J}^{os}$ is, by definition, guaranteed to be non-negative.

An additional regularity condition is required for the derivation of the limit theory for the OSLRV estimator. To introduce this condition, define the shrinking neighborhood $N_K = \{(f, f^*) \in B_K \times B_K : \|f - \phi\| \leq (\delta_K + \bar{\eta}_{n,K}) \log(\log n) \text{ and } f^* \in B_K : \|f^* - \phi^*\| \leq (\delta_K + \delta_K^* + \bar{\eta}_{n,K}) \log(\log n)\}$.

**Assumption 5.4.** *The following equicontinuity conditions are satisfied:*

*(i)* $\sup_{(f,f^*) \in N_K} \sum_{t=0}^{n-1} h_j(\frac{t+1}{n}) \{\phi_t^* \phi_t - f_t^* f_t - E[\phi_t^* \phi_t - f_t^* f_t]\} = o_p(n^{1/2})$
*(ii)* $\sup_{(f,f^*) \in N_K} \sum_{t=0}^{n-1} h_j(\frac{t+1}{n}) \{m_{t,t+1}(\phi_t^* \phi_{t+1} - f_t^* f_{t+1}) - E[m_{t,t+1}(\phi_t^* \phi_{t+1} - f_t^* f_{t+1})]\} = o_p(n^{1/2})$.

Assumption 5.4 is essentially Assumption 5.2(i) of Chen, Liao, and Sun (2012) applied in this context. The definition of $N_K$ and the convergence rates of $\widehat{\phi}^f$ and $\widehat{\phi}^{f\star}$ established in Corollary 5.3 ensure that $(\widehat{\phi}^f, \widehat{\phi}^{f\star}) \in N_K$ with probability approaching one.

Consistency of the asymptotic variance estimators for $\widehat{\rho}$ and $\widehat{y}$ are now established, together with a means of performing asymptotic inference for $\widehat{L}$ based on the $t$ distribution using the OSLRV estimator $\widehat{V}_{L,J}^{os}$. Let $\chi_J^2$ and $t_J$ denote the $\chi^2$ and $t$ distributions with $J$ degrees of freedom.

**Theorem 5.3.** *Under Assumptions 4.1, 5.1 5.2, and 5.3*

*(i)* $\widehat{V}_\rho \to_p V_\rho$
*(ii)* $(\widehat{\rho}^{-2}) \widehat{V}_\rho \to_p \rho^{-2} V_\rho$.

*If, in addition, Assumption 5.4 holds,* $V_\rho > 0$ *and* $V_L > 0$*, then*

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

Theorems 5.2 and 5.3 together provide a means with which to perform feasible asymptotic inference on $\rho$, $y$, and $L$.

5.4.1. *Example: Smooth kernel (continued).* Here $\delta_K = O(K^{-p/d})$. Assume that $\phi^*$ belongs to a Hölder ball of smoothness $s$, so that $\delta^*_K = O(K^{-s/d})$.

(a) If $m$ is unbounded then $\bar{\eta}_{n,K} = O(K/\sqrt{n})$ and $K$ may be chosen so that the conditions $\bar{\eta}_{n,K} = o(n^{-1/4})$ and $\delta_K = o(n^{-1/2})$ are satisfied provided $p > 2d$.
(b) If $m$ is bounded then $\bar{\eta}_{n,K} = O(\zeta_0(K)(\log n)/n)$ and $K$ may be chosen so that the conditions $\bar{\eta}_{n,K} = o(n^{-1/4})$ and $\delta_K = o(n^{-1/2})$ are satisfied provided $p > d$.

In either case the remaining condition $\zeta_0(K)\max\{\delta^*_K, \delta_K\} = o(1)$ is satisfied if $s > \frac{1}{2}d$. If, for arguments sake, $\mathbb{M}$ is selfadjoint, then $\phi^* = \phi$ and the condition $\delta_K = o(n^{-1/2})$ can be relaxed to $\delta_K = o(n^{-1/4})$ (Gobet, Hoffmann, and Reiß, 2004, Remark 4.7), in which case it suffices that $p > d$ if $m$ is unbounded and $p > \frac{1}{2}d$ if $m$ is bounded.

5.5. **Semiparametric efficiency.** The semiparametric efficiency bounds for $\widehat{\rho}$, $\widehat{y}$ and $\widehat{L}$ are now derived, and it is shown that the estimators attain their efficiency bounds.[7] The efficiency bound derivations follow the arguments of Greenwood and Wefelmeyer (1995) and Wefelmeyer (1999) (see also Bickel and Kwon (2001)). A tangent space of admissible perturbations to the unknown transition distribution of the state process is first constructed. A nonparametric version of local asymptotic normality holds for the perturbed models. The parameters $\rho$, $y$ and $L$ are shown to be differentiable with respect to the perturbation of the transition density and their gradients are characterized. The efficient influence function of the estimators are determined by projecting their gradients onto the (closure of the) tangent space. The asymptotic variances of $\widehat{\rho}$, $\widehat{y}$ and $\widehat{L}$ are shown to coincide with the second moment of their efficient influence functions, whence efficiency obtains.

**Theorem 5.4.** *Under Assumptions 4.1, 5.1, 5.2 and 5.3, the semiparametric efficiency bounds for $\rho$, $y$ and $L$ are $V_\rho$, $\rho^{-2}V_\rho$ and $V_L$, and are achieved by $\widehat{\rho}$, $\widehat{y}$ and $\widehat{L}$.*

Now consider the somewhat artificial case in which the stationary distribution $Q$ is known but the dynamics of $\{X_t\}$ are still unknown. In this setting the Gram matrix $\mathbf{G}_K$ is known but $\mathbf{M}_K$ is unknown. An alternative estimator for $\rho$ is

$$\check{\rho} = \lambda_{\max}(\mathbf{G}_K^{-1}\widehat{\mathbf{M}}_K)\,.$$

One might expect the asymptotic variance of $\check{\rho}$ to be smaller than that of $\widehat{\rho}$ because $\check{\rho}$ appears to make use of the fact that $Q$ is known. The following theorem shows otherwise.

**Theorem 5.5.** *Under Assumptions 4.1, 5.1, 5.2 and 5.3(i)(ii), if $V_\rho > 0$ then*

$$\sqrt{n}(\check{\rho} - \rho) \to_d N(0, V_\rho + W_\rho)$$

*where* $W_\rho = 2\rho^2 E[(\phi^*(X_0)\phi(X_0) - 1)^2] + \rho^2 \mathrm{lrvar}((\phi^*(X_t)\phi(X_t) - 1))$.

[7] In practice the true SDF is unknown, so the term "limited information bound" may be more appropriate than "semiparametric efficiency bound".

Clearly $W_\rho \geq 0$, and the inequality is strict if $\phi(x)\phi^*(x)$ is non-constant on a set of positive probability. Therefore, the estimator $\widehat{\rho}$ is relatively more efficient than $\check{\rho}$, even though $\check{\rho}$ appears to incorporate the fact that the density is known.[8]

## 6. Extensions

The estimators and large-sample theory presented in Section 5 is now extended to study (i) the long-term implications of estimated SDFs, (ii) the long-term implications of SDFs with additional roughness, and (iii) nonparametric sieve estimation of the marginal utility of consumption of a representative agent.

6.1. **Plugging-in an estimated SDF.** Consider the two-stage problem of first estimating a SDF from data, then extracting its long-term implications. Let the data consist of a time series $\{(X_0, R_0), \ldots, (X_n, R_n)\}$ where $R_t = (R_{i,t}, \ldots, R_{d_R,t})'$ is a vector of returns on $d_R$ assets for each $t$. Assume that the researcher has estimated a SDF, say $\widehat{m}$, from the data. The SDF estimator $\widehat{m}$ could be parametric or semi/nonparametric. An example of a parametric SDF estimator $\widehat{m}$ is the consumption CAPM SDF $m(X_t, X_1; \beta, \gamma) = \beta G_{t+1}^{-\gamma}$ evaluated at $(\widehat{\beta}, \widehat{\gamma})$ where $(\widehat{\beta}, \widehat{\gamma})$ are estimated from $\{(X_0, R_0), \ldots, (X_n, R_n)\}$. Semi/nonparametric estimators include the semiparametric consumption CAPMs studied in Gallant and Tauchen (1989) and Fleissig, Gallant, and Seater (2000), nonparametric nonlinear factor models (Bansal and Viswanathan, 1993), models with nonparametric habit formation (Chen and Ludvigson, 2009), and models with recursive preferences and unknown dynamics (Chen, Favilukis, and Ludvigson, 2013b).

In this case the matrix $\mathbf{M}_K$ is estimated using

$$\widehat{\mathbf{M}}_K = \frac{1}{n}\sum_{t=0}^{n-1} b^K(X_t)\widehat{m}(X_t, X_{t+1})b^K(X_{t+1})' \,. \tag{37}$$

The eigenvalue $\rho$ and eigenfunctions $\phi$ and $\phi^*$ are estimated by solving the matrix eigenvalue problems (26) and (27) with $\widehat{\mathbf{M}}_K$ given by (37). The estimators of the long-term yield and entropy of the permanent component of the SDF are

$$\begin{aligned} \widehat{y} &= -\log \widehat{\rho} \\ \widehat{L} &= \log \widehat{\rho} - \frac{1}{n}\sum_{t=0}^{n-1} \log \widehat{m}(X_t, X_{t+1}) \end{aligned}$$

by analogy with (30) and (31). Consistency and convergence rates of the estimators follow under similar conditions to those described in Section 5.

**Theorem 6.1.** *Let Assumption 4.1 hold, and let Assumption 5.1 hold with $\widehat{\mathbf{M}}_K$ as in expression (37). Then there is a set whose probability approaches one on which $\widehat{\rho}$ is real and positive and has multiplicity one, and*

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

[8] If $Q$ is known the semiparametric efficiency bound for $\rho$ may be different from $V_\rho$. Consequently, $\widehat{\rho}$ may not be semiparametrically efficient when $Q$ is known.

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

The requirement that Assumption 5.1 hold with $\widehat{\mathbf{M}}_K$ as in expression (37) is an implicit condition on convergence of $\widehat{m}$ to $m$.

The asymptotic distribution for $\widehat{\rho}$, $\widehat{y}$ and $\widehat{L}$ will be distorted (relative to the case in which $m$ is known) by the error introduced by replacing $m$ with a first-stage estimator $\widehat{m}$. The form of the asymptotic distribution for $\widehat{\rho}$, $\widehat{y}$ and $\widehat{L}$ will therefore differ depending on the method used to construct $\widehat{m}$. The following high-level assumption is made to establish the asymptotic linear expansion for $\widehat{\rho}$ in this setting.

**Assumption 6.1.** $\sum_{t=0}^{n-1} |\widehat{m}(X_t, X_{t+1}) - m(X_t, X_{t+1})|(1 + \phi(X_{t+1}) + \phi^*(X_t)) = O_p(n^{1/2})$.

The following Theorem establishes the distortion to the limit distribution of $\widehat{\rho}$ that arises due to the first-stage estimator $\widehat{m}$. The limit distribution of $\widehat{\rho}$, $\widehat{y}$ and $\widehat{L}$ can then be derived from this expansion on a case-by-case basis.

**Theorem 6.2.** *Let Assumption 4.1, 5.1, 5.2, and 5.3 hold with $\widehat{\mathbf{M}}_K$ as in expression (37), and let Assumption 6.1 hold. Then*

$$\begin{aligned}\sqrt{n}(\widehat{\rho} - \rho) &= \frac{1}{\sqrt{n}} \sum_{t=0}^{n-1} \{\phi^*(X_t) m(X_t, X_{t+1}) \phi(X_{t+1}) - \rho \phi^*(X_t) \phi(X_t)\} \\ &\quad + \frac{1}{\sqrt{n}} \sum_{t=0}^{n-1} \phi^*(X_t)(\widehat{m}(X_t, X_{t+1}) - m(X_t, X_{t+1}))\phi(X_{t+1}) + o_p(1)\,.\end{aligned}$$

Comparing Theorem 6.2 with the expansion for $\widehat{\rho}$ when $m$ is known shows that the limit distribution of $\widehat{\rho}$ will be distorted (relative to the known SDF case) by an additional functional of $(\widehat{m} - m)$. The following remark deals with the case in which $\widehat{m}$ is estimated parametrically.

**Remark 6.1.** *Let $m$ be known up to a finite-dimensional parameter $\theta_0 \in \Theta \subseteq \mathbb{R}^{d_\theta}$ and let $m$ be estimated by plugging in a first-stage estimator $\widehat{\theta}$ of $\theta_0$, i.e.*

$$\begin{aligned} m(X_t, X_{t+1}) &= m(X_t, X_{t+1}; \theta_0) \\ \widehat{m}(X_t, X_{t+1}) &= m(X_t, X_{t+1}; \widehat{\theta})\,. \end{aligned}$$

*If (a) $\sqrt{n}(\widehat{\theta} - \theta_0) = O_p(1)$, (b) $\theta_0 \in \mathrm{int}(\Theta)$, (c) for all $(x_0, x_1) \in \mathcal{X}^2$, $m(x_0, x_1; \theta)$ is twice continuously differentiable in $\theta$ on a neighborhood $\Theta_0 \subset \mathrm{int}(\Theta)$ containing $\theta_0$, (d)*

$$E\left[\left\|\frac{\partial m(X_0, X_1; \theta_0)}{\partial \theta}\right\|_2^2\right] < \infty \quad \textit{and} \quad E\left[\left|\frac{\partial m(X_0, X_1; \theta_0)}{\partial \theta_i}\right| \phi^*(X_0)\phi(X_1)\right] < \infty$$

*for $i = 1, \ldots, d_\theta$, (e) there exists a $g : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ such that*

$$\sup_{\theta \in \Theta_0} \left\|\frac{\partial^2 m(x_0, x_1; \theta)}{\partial \theta \partial \theta'}\right\|_2 \le g(x_0, x_1)$$

*with $E[g(X_0,X_1)^2] < \infty$ and $E[g(X_0,X_1)\phi^*(X_0)\phi(X_1)] < \infty$. Then, Assumption 6.1 is satisfied and, under the remaining conditions of Theorem 6.2,*

$$\begin{aligned}\sqrt{n}(\widehat{\rho}-\rho) &= \frac{1}{\sqrt{n}}\sum_{t=0}^{n-1}\{\phi^*(X_t)m(X_t,X_{t+1})\phi(X_{t+1})-\rho\phi^*(X_t)\phi(X_t)\} \\ &\quad +E\left[\phi^*(X_0)\phi(X_1)\frac{\partial m(X_0,X_1;\theta_0)}{\partial\theta}\right]\sqrt{n}(\widehat{\theta}-\theta_0)+o_p(1)\,.\end{aligned}$$

The limit distribution for $\widehat{\rho}$ when $m$ is estimated semi/nonparametrically may be similarly derived using Theorem 6.2.

6.2. **Roughing-up the SDF.** Following Hansen and Scheinkman (2012, 2013), consider a class of economy characterized by a discrete-time (first-order) Markov state process $\{(X_t,Y_t)\}$ defined on a complete probability space $(\Omega,\mathcal{F},\{\mathcal{F}_t\},\mathbb{P})$ where time is again indexed by $t\in\mathbb{Z}$ and where $\mathcal{F}_t=\sigma(X_t,Y_t,X_{t-1},Y_{t-1},\ldots)$. Assume that the distribution of $(X_{t+1},Y_{t+1})$ conditioned on $(X_t,Y_t)$ is the same as the joint distribution of $(X_{t+1},Y_{t+1})$ conditioned on $X_t$. More compactly,

$$(X_{t+1},Y_{t+1})|(X_t,Y_t) =_d (X_{t+1},Y_{t+1})|X_t \tag{38}$$

for all $t$, where $=_d$ denotes equality in distribution. The "non-causality" condition (38) is convenient for dimension reduction: it allows for the SDF to be a function of $(X_t,X_{t+1},Y_{t+1})$ whilst restricting the class of eigenfunctions to be functions of $X$ only (not functions of $(X,Y)$).

Let $\{X_t\}$ have support $\mathcal{X}\subseteq\mathbb{R}^d$ and let $\{Y_t\}$ have support $\mathcal{Y}\subseteq\mathbb{R}^{d_y}$. Assume further that the date-$t$ 1-period SDF is now $m(X_t,X_{t+1},Y_{t+1})$ for some $m:\mathcal{X}\times\mathcal{X}\times\mathcal{Y}\to\mathbb{R}$. Define $\mathbb{M}:L^2(Q)\to L^2(Q)$ as the 1-period pricing operator given by

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

The adjoint operator $\mathbb{M}^*$ is defined as

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

The positive eigenfunction problems are again

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

with $\phi$ and $\phi^*$ positive (almost everywhere). The following regularity conditions are a straightforward extension of Assumption 4.1.

**Assumption 6.2.** *$\{(X_t,Y_t)\}$ and $m$ satisfy the following conditions:*

*(i) $\{(X_t,Y_t)\}$ is a strictly stationary and ergodic (first-order) Markov process which satisfies the non-causality condition (38), and which has support $\mathcal{X}\times\mathcal{Y}\subseteq\mathbb{R}^d\times\mathbb{R}^{d_y}$*

*(ii) the stationary distributions $Q$ of $\{X_t\}$ and $Q_y$ of $\{Y_t\}$ have densities $q$ and $q_y$ (wrt Lebesgue measure) s.t. $q(x)>0$ and $q_y(y)>0$ almost everywhere*

*(iii)* $(X_0, X_1, Y_1)$ *has joint density* $f$ *(wrt Lebesgue measure) s.t.* $f(x_0, x_1, y_1) > 0$ *almost everywhere and* $f(x_0, x_1, y_1)/(q(x_0)q(x_1)q_y(y_1))$ *is uniformly bounded away from infinity.*

*(iv)* $m : \mathcal{X} \times \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ *has* $m(x_0, x_1, y_1) > 0$ *almost everywhere and* $E[m(X_0, X_1, Y_1)^2] < \infty$.

Nonparametric identification of $\phi$ and $\phi^*$ in this environment follows similarly.

**Theorem 6.3.** *Under Assumption 6.2,*

*(i)* $\mathbb{M}$ *and* $\mathbb{M}^*$ *have unique (to scale) eigenfunctions* $\phi \in L^2(Q)$ *and* $\phi^* \in L^2(Q)$ *such that* $\phi > 0$ *and* $\phi^* > 0$ *(almost everywhere)*

*(ii)* $\rho$ *is positive, has multiplicity one, and is the largest element of the spectrum of* $\mathbb{M}$.

Given a candidate SDF $m$ and a time series of data $\{X_0, (X_1, Y_1), \ldots, (X_n, Y_n)\}$, the positive eigenfunctions $\phi$ and $\phi^*$ and the eigenvalue $\rho$ can be estimated by solving the matrix eigenvalue problems (26) and (27) as before, but with $\mathbf{M}_K$ and $\widehat{\mathbf{M}}_K$ given by

$$\mathbf{M}_K = E[b^K(X_0)m(X_0, X_1, Y_1)b^K(X_1)'] \tag{39}$$

$$\widehat{\mathbf{M}}_K = \frac{1}{n}\sum_{t=0}^{n-1} b^K(X_t)m(X_t, X_{t+1}, Y_{t+1})b^K(X_{t+1})' \,. \tag{40}$$

The long-run yield and entropy of the permanent component of the SDF are estimated with

$$\begin{aligned} \widehat{y} &= -\log \widehat{\rho} \\ \widehat{L} &= \log \widehat{\rho} - \frac{1}{n}\sum_{t=0}^{n-1} \log m(X_t, X_{t+1}, Y_{t+1}) \end{aligned}$$

by analogy with expressions (30) and (31).

Consistency, convergence rates, and the asymptotic distribution of the estimators follow by arguments identical to the case dealt with in Section 5. However, Assumption 6.2 does not characterize the joint weak-dependence properties of $\{(X_t, Y_t)\}$ so an extra assumption is required to establish the limit distribution of $\widehat{L}$. For the remainder of this subsection, let $V_\rho$ and $V_L$ be defined as in expressions (33) and (34), but with $m(X_t, X_{t+1}, Y_{t+1})$ in place of $m(X_t, X_{t+1})$. Let

$$\begin{aligned} \psi_L(X_t, X_{t+1}, Y_{t+1}) &= \rho^{-1}\phi^*(X_t)m(X_t, X_{t+1}Y_{t+1})\phi(X_{t+1}) - \phi^*(X_t)\phi(X_t) \\ &\quad - \log m(X_t, X_{t+1}, Y_{t+1}) + E[\log m(X_0, X_1, Y_1)] \\ V_L &= \operatorname{lrvar}(\psi_L(X_t, X_{t+1}, Y_{t+1})) \end{aligned}$$

The following high-level assumption is sufficient to establish the limit distribution of $\widehat{L}$.

**Assumption 6.3.** *The following regularity conditions hold:*

*(i)* $V_L < \infty$

*(ii)* $n^{-1/2}\sum_{t=0}^{n-1} \psi_L(X_t, X_{t+1}, Y_{t+1}) \to_d N(0, V_L)$.

Mean-square convergence rates of the eigenfunction estimators and $\sqrt{n}$-asymptotic normality of $\widehat{\rho}$, $\widehat{y}$ and $\widehat{L}$ are now established.

**Theorem 6.4.** *Let Assumption 6.2 hold, and let Assumption 5.1 hold with $\mathbf{M}_K$ and $\widehat{\mathbf{M}}_K$ as in expressions (39) and (40). Then there is a set whose probability approaches one on which $\widehat{\rho}$ is real and positive and has multiplicity one, and*

*(i)* $|\widehat{\rho} - \rho| = O_p(\delta_K + \eta_{n,K})$
*(ii)* $\|\widehat{\phi} - \phi\| = O_p(\delta_K + \eta_{n,K})$
*(iii)* $\|\widehat{\phi}^\star/\|\widehat{\phi}^*\| - \phi^*/\|\phi^*\|\| = O_p(\delta_K^* + \eta_{n,K})$.

*If, in addition, Assumptions 5.2 and 5.3 hold with $m(X_0, X_1, Y_1)$ in place of $m(X_0, X_1)$ and $V_\rho > 0$, then*

*(iv)* $\sqrt{n}(\widehat{\rho} - \rho) \to_d N(0, V_\rho)$
*(v)* $\sqrt{n}(\widehat{y} - y) \to_d N(0, \rho^{-2} V_\rho)$.

*If, in addition, Assumption 6.3 holds and $V_L > 0$, then*

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

The asymptotic variances $V_\rho$, $\rho^{-2}V_\rho$ and $V_L$ of $\widehat{\rho}$, $\widehat{y}$ and $\widehat{L}$ may be estimated analogously to the case dealt with in Section 5. That is, $\widehat{V}_\rho$ and $\widehat{V}_{L,J}^{os}$ are defined as in expression (35) and (36), respectively, but with $m(X_t, X_{t+1}, Y_{t+1})$ in place of $m(X_t, X_{t+1})$. The estimators $\widehat{V}_\rho$ and $\widehat{\rho}^{-2}\widehat{V}_\rho$ are consistent under the conditions of Theorem 6.4(iv)(v). Asymptotic inference based on $\widehat{V}_{L,J}^{os}$ follows under additional regularity.

6.3. **Application: nonparametric Euler equation estimation.** The results of Section 6.2 may be used to establish the large sample properties of nonparametric sieve estimators of the marginal utility of consumption of a representative agent, as outlined in Section 2.1. The sieve approach outlined below is an alternative to the kernel-based procedure analyzed by Linton, Lewbel, and Srisuma (2011). As in Linton, Lewbel, and Srisuma (2011), the process $\{(X_t, R_{i,t+1})\}$ is required to be stationary and ergodic. This requirement restricts the forms of utility compatible with this analysis. Consider, for example, $MU$ of the form

$$MU_t = MU(C_t, C_{t-1}, Z_t) \tag{41}$$

where $C_t$ is aggregate consumption at date $t$ and $X_t = (C_t, C_{t-1}, Z_t)$. A conventional assumption is that aggregate consumption $\{C_t\}$ is nonstationary but growth in aggregate consumption $\{C_t/C_{t-1}\}$ is stationary (see, e.g., Hansen and Singleton (1982); Gallant and Tauchen (1989)). Under this assumption, $MU$ of the form (41) is incompatible with the stationarity requirement. If $MU$ in expression (41) is homogeneous of degree zero in its first two arguments, $MU_t$ may be rewritten as

$$MU_t = MU(C_t/C_{t-1}, Z_t)\,. \tag{42}$$

Marginal utility of the form (42) may then be estimated as described below, provided the process $\{(C_t/C_{t-1}, Z_t, R_{i,t+1})\}$ is strictly stationary and ergodic.[9]

[9]Chen and Ludvigson (2009) use a similar homogeneity assumption to rewrite a semiparametric habit formation model in terms of consumption growth (rather than levels of consumption).

Assume $MU_t = MU(X_t)$ where $\{(X_t, R_{i,t})\}$ is strictly stationary and ergodic (here $\{(X_t, R_{i,t})\}$ does not need to be a Markov process). Given data $\{X_0, (X_1, R_{i,1}), \ldots, (X_n, R_{i,n})\}$, $\beta$ and $MU$ can be estimated by solving

$$\widehat{\mathbf{G}}_K^{-1}\widehat{\mathbf{T}}_{i,K}\widehat{c} = \widehat{\beta}^{-1}\widehat{c}$$

and setting $\widehat{MU}(x) = b^K(x)'\widehat{c}$, where $\widehat{\beta}^{-1}$ is the largest eigenvalue of $\widehat{\mathbf{G}}_K^{-1}\widehat{\mathbf{T}}_{i,K}$ and

$$\widehat{\mathbf{T}}_{i,K} = \frac{1}{n}\sum_{t=0}^{n-1} b^K(X_t)R_{i,t+1}b^K(X_{t+1})'\,.$$

Let $MU^*$ solve $E[R_{i,t+1}MU^*(X_t)|X_{t+1}] = \beta^{-1}MU^*(X_{t+1})$. Then $MU^*$ may be estimated by solving $\widehat{\mathbf{G}}_K^{-1}\widehat{\mathbf{T}}'_{i,K}\widehat{c}^\star = \widehat{\beta}^{-1}\widehat{c}^\star$ and setting $\widehat{MU}^\star(x) = b^K(x)'\widehat{c}^\star$.

The large sample properties of $\widehat{\beta}^{-1}$, $\widehat{MU}$ and $\widehat{MU}^\star$ follow from Theorem 6.4. Normalize $MU$ and $MU^*$ so that $E[MU(X)^2] = 1$, $E[MU(X)MU^*(X)] = 1$. Without confusion, let $Q$ denote the stationary distribution of $\{X_t\}$. Replace Assumption 6.2 with (a) $\mathbb{T}_i : L^2(Q) \to L^2(Q)$ is Hilbert-Schmidt, and (b) $\mathbb{T}_i MU = \beta^{-1}MU$ where $MU \in L^2(Q)$ and $\beta^{-1} > 0$ is the largest eigenvalue of $\mathbb{T}_i$ and has multiplicity one. Also let Assumptions 5.1, 5.2 and 5.3 hold with $\mathbb{T}_i$, $\widehat{\mathbf{T}}_{i,K}$, $\mathbf{T}_{i,K} = E[b^K(X_t)R_{i,t+1}b^K(X_t)']$, $MU$, and $MU^*$ in place of $\mathbb{M}$, $\widehat{\mathbf{M}}_K$, $\mathbf{M}_K$, $\phi$ and $\phi^*$. Theorem 6.4(i)–(iii) establishes consistency and convergence rates of $\widehat{\beta}^{-1}$, $\widehat{MU}$ and $\widehat{MU}^\star$. The limit distribution for $\widehat{\beta}$ is more subtle. Let $\mathcal{G}_t = \sigma(X_t, R_{i,t}, X_{t-1}, R_{i,t-1}, \ldots)$. If $(X_{t+1}, R_{i,t+1})|X_t =_d (X_{t+1}, R_{i,t+1})|\mathcal{G}_t$ for all $t$, then

$$\sqrt{n}(\widehat{\beta} - \beta) \to_d N(0, \beta^4 E[\{MU(X_{t+1})R_{i,t+1}MU^*(X_t) - \beta^{-1}MU(X_t)MU^*(X_t)\}^2])$$

by Theorem 6.4(iv) and the delta method. Otherwise, simple modification of the proof of Theorem 5.2 yields, under regularity,

$$\sqrt{n}(\widehat{\beta} - \beta) \to_d N(0, \beta^4 W_\beta)$$

where $W_\beta = \mathrm{lrvar}(\{MU(X_{t+1})R_{i,t+1}MU^*(X_t) - \beta^{-1}MU(X_t)MU^*(X_t)\})$.

## 7. Monte Carlo simulation

The following Monte Carlo (MC) exercise explores the performance of the estimators when applied to a stylized consumption CAPM. The SDF is

$$m(X_t, X_{t+1}) = \beta\exp(-\gamma g_{t+1})$$

where $\beta$ is the time preference parameter, $\gamma$ is the risk aversion parameter, and $g_{t+1}$ is log consumption growth from time $t$ to $t+1$. The state variable is simply $X_t = g_t$. The data are constructed to be somewhat representative of U.S. real monthly aggregate consumption growth. Log consumption growth evolves as the Gaussian AR(1)

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

where the $e_t$ are i.i.d. $N(0,1)$ random variables. Gaussianity of the disturbances ensures that the process $\{g_t\}_{t=-\infty}^{\infty}$ is time reversible (see, e.g., Weiss (1975)), which is used to obtain a closed-form

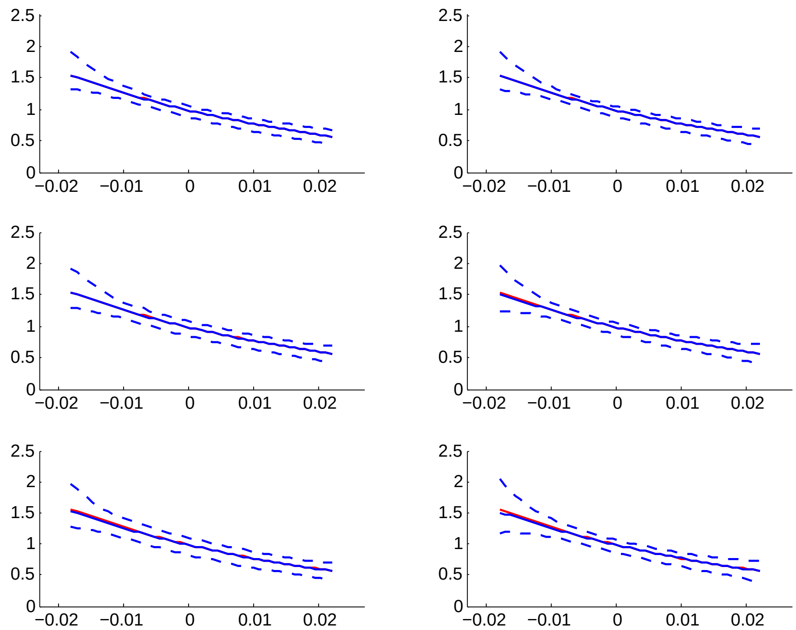


FIGURE 1. MC plots for $\widehat{\phi}^f$ with $\gamma = 25$. Each panel shows the true $\phi$ (solid red line), pointwise MC median (solid blue line), and pointwise MC 90% confidence bands (dashed lines). Results are presented for Hermite polynomial (left) and B-spline (right) sieves of dimension 6 (top), 10 (middle) and 14 (bottom).

solution for $\phi^*$. The positive eigenfunction and adjoint positive eigenfunction are

$$\begin{aligned}\phi(g) &= \exp\left(-\frac{\gamma\kappa}{1-\kappa}g + \frac{\mu\gamma\kappa}{1-\kappa} - \frac{\gamma^2\kappa^2\sigma^2}{(1-\kappa^2)(1-\kappa)^2}\right)\\ \phi^*(g) &= \exp\left(-\frac{\gamma}{1-\kappa}g + \frac{\mu\gamma}{1-\kappa} + \frac{\gamma^2\sigma^2}{(1-\kappa^2)(1-\kappa)^2}\left\{\kappa^2 - \frac{1}{2}(1+\kappa)^2\right\}\right)\end{aligned}$$

where both $\phi$ and $\phi^*$ have been normalized so that $E[\phi^2(g)] = 1$ and $E[\phi(g)\phi^*(g)] = 1$. Their eigenvalue $\rho$ is

$$\rho = \beta \exp\left(-\gamma\mu + \frac{1}{2}\frac{\gamma^2}{(1-\kappa)^2}\sigma^2\right).$$

The parameters for the simulation are $\mu = 0.002$, $\kappa = 0.3$, and $\sigma = 0.01/\sqrt{1-\kappa^2}$, which are similar in magnitude to the parameters of the U.S. real per capita consumption growth series investigated in the next section. The sample length is set to 500, and 10000 simulations are performed. The time preference parameter $\beta$ is set to 0.998, and $\gamma$ is varied from 0 to 30. Two choices of sieve are used, namely Hermite polynomials and cubic B-splines, with dimension $K = 6$, 10 and 14. For each simulation, the Hermite polynomial sieve was 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. Cosine bases of dimension $J = 10$ and 15 were used to compute the OSLRV estimator.

MC results $\widehat{\phi}^f$ and $\widehat{\phi}^{\star f}$ for $\gamma = 25$ are presented in Figures 1 and 2, respectively. Each panel shows the true $\phi$ (or $\phi^*$) for $g \in [\mu - 2\sigma, \mu + 2\sigma]$ (solid red lines) together with the pointwise MC median $\widehat{\phi}$ (or $\widehat{\phi}^\star$) (solid blue lines) and pointwise MC 90% confidence bands (the pointwise

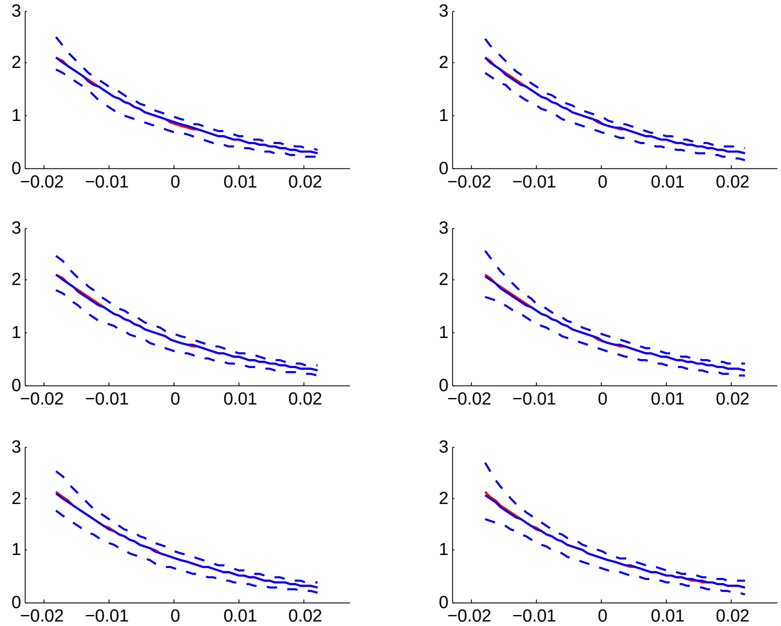


FIGURE 2. MC plots for $\widehat{\phi}^{\star f}$ with $\gamma = 25$. Each panel shows the true $\phi$ (solid red line), pointwise MC median (solid blue line), and pointwise MC 90% confidence bands (dashed lines). Results are presented for Hermite polynomial (left) and B-spline (right) sieves of dimension 6 (top), 10 (middle) and 14 (bottom).

.05 and .95 quantiles of the estimator approximated by simulation; dashed lines). Both $\widehat{\phi}$ and $\widehat{\phi}^{\star}$ are normalized feasibly as in Corollary 5.3. The estimators have negligible bias. The width of the confidence bands increases with the sieve dimension $K$, which illustrates that the "variance term" $\bar{\eta}_{n,K}$ is increasing in $K$. Other simulations (not reported) show that increasing/decreasing $\gamma$ also increases/decreases the width of the MC confidence bands.

Table 1 shows the MC coverage probabilities for 90% and 95% confidence intervals (CIs) for $\rho$ and $y$, and Table 2 shows the MC coverage probabilities for 90% and 95% CIs for $L$. To construct the MC coverage probabilities, for each simulation $\rho$, $y$, and $L$ were estimated and their 90% and 95% confidence intervals estimated using the variance estimators $\widehat{V}_\rho$, $\widehat{\rho}^{-2}\widehat{V}_\rho$, and $\widehat{V}^{os}_{L,J}$. Gaussian critical values were used for the CIs for $\rho$ and $y$, and $t_J$ critical values were used for the CIs for $L$. The MC coverage probabilities are the proportion of simulations for which the estimated CIs contained the true parameter values. Table 1 shows that the 90% and 95% CIs for $\rho$ and $y$ have MC coverage probabilities that are very close to their nominal coverage probabilities, for all sieve choices and all levels of $\gamma$. The MC coverage probabilities for $L$ presented in Table 2 show that the CIs corresponding to a B-spline sieve are too narrow, especially at high values of $\gamma$. The coverage probabilities for the CIs corresponding to a Hermite polynomial sieve are close to their nominal values with both $J = 10$ and $J = 15$. The MC coverage probabilities for $\rho$, $y$ and $L$ appear generally robust to the dimension $K$ of the sieve space, especially when a Hermite polynomial sieve is used.

| | 90% CI for $\rho$ | | | 95% CI for $\rho$ | | |
|---|---|---|---|---|---|---|
| | $K=6$ | $K=10$ | $K=14$ | $K=6$ | $K=10$ | $K=14$ |
| $\gamma=5$ | 89.60 | 89.64 | 89.75 | 94.82 | 94.78 | 94.79 |
| H-Pol $\gamma=15$ | 89.86 | 89.91 | 90.04 | 94.52 | 94.54 | 94.60 |
| $\gamma=25$ | 89.43 | 89.53 | 89.39 | 93.63 | 93.59 | 93.61 |
| $\gamma=5$ | 89.63 | 89.60 | 89.54 | 94.82 | 94.75 | 94.74 |
| B-Spl $\gamma=15$ | 89.80 | 89.63 | 89.57 | 94.50 | 94.51 | 94.39 |
| $\gamma=25$ | 89.28 | 88.64 | 88.41 | 93.51 | 93.27 | 92.95 |
| | 90% CI for $y$ | | | 95% CI for $y$ | | |
| | $K=6$ | $K=10$ | $K=14$ | $K=6$ | $K=10$ | $K=14$ |
| $\gamma=5$ | 89.70 | 89.66 | 89.74 | 94.81 | 94.80 | 94.77 |
| H-Pol $\gamma=15$ | 89.91 | 89.90 | 90.02 | 94.62 | 94.60 | 94.61 |
| $\gamma=25$ | 89.44 | 89.58 | 89.50 | 93.78 | 93.76 | 93.79 |
| $\gamma=5$ | 89.68 | 89.58 | 89.39 | 94.80 | 94.68 | 94.53 |
| B-Spl $\gamma=15$ | 89.78 | 89.53 | 89.42 | 94.53 | 94.50 | 94.18 |
| $\gamma=25$ | 89.33 | 88.68 | 88.24 | 93.67 | 93.35 | 92.91 |

TABLE 1. Monte Carlo coverage probabilities for 90% and 95% asymptotic confidence intervals for $\rho$ and $y$ based on the asymptotic distribution in Theorem 5.2 and the consistent variance estimators $\widehat{V}_\rho$ and $\widehat{\rho}^{-2}\widehat{V}_\rho$. Results are presented for Hermite polynomial (H-Pol) and B-spline (B-Spl) sieves of varying dimension $K$.

| | 90% CI for $L$ | | | 95% CI for $L$ | | |
|---|---|---|---|---|---|---|
| | $K=6$ | $K=10$ | $K=14$ | $K=6$ | $K=10$ | $K=14$ |
| $\gamma=5$ | 91.66 | 91.70 | 91.80 | 95.92 | 95.96 | 95.90 |
| H-Pol, $J=10$ $\gamma=15$ | 90.35 | 90.34 | 90.30 | 94.91 | 94.96 | 94.98 |
| $\gamma=25$ | 88.69 | 88.47 | 88.28 | 93.67 | 93.53 | 93.18 |
| $\gamma=5$ | 92.95 | 93.06 | 93.03 | 96.74 | 96.80 | 96.84 |
| H-Pol, $J=15$ $\gamma=15$ | 91.11 | 91.20 | 91.08 | 95.47 | 95.52 | 95.53 |
| $\gamma=25$ | 89.41 | 89.11 | 88.94 | 94.27 | 93.98 | 93.86 |
| $\gamma=5$ | 88.26 | 88.20 | 87.84 | 92.98 | 92.71 | 92.62 |
| B-Spl, $J=10$ $\gamma=15$ | 86.69 | 86.12 | 85.87 | 91.55 | 91.16 | 90.80 |
| $\gamma=25$ | 84.47 | 83.63 | 83.09 | 89.36 | 88.96 | 88.51 |
| $\gamma=5$ | 89.72 | 89.67 | 89.51 | 93.99 | 93.83 | 93.70 |
| B-Spl, $J=15$ $\gamma=15$ | 87.49 | 86.98 | 86.75 | 92.29 | 91.83 | 91.66 |
| $\gamma=25$ | 85.29 | 84.58 | 84.10 | 89.93 | 89.48 | 88.97 |

TABLE 2. Monte Carlo coverage probabilities for 90% and 95% asymptotic confidence intervals for $L$ based on the asymptotic distribution in Theorem 5.3 and the OSLRV estimator $\widehat{V}^{os}_{L,J}$, which was computed with a cosine basis of dimension $J = 10$ and $J = 15$. Results are presented for Hermite polynomial (H-pol) and B-spline (B-Spl) sieves for of varying dimension $K$.

## 8. EMPIRICAL ILLUSTRATION

The long-run implications of the consumption CAPM are now investigated using the tools introduced in this paper. The consumption CAPM has been the basis for a vast amount of research, from the seminal works of Hansen and Singleton (1982) and Mehra and Prescott (1985) though to recent rare disasters-based investigations of Backus, Chernov, and Martin (2011) and Julliard and Ghosh (2012). The SDF to be investigated is

$$m(X_t, X_{t+1}; \beta, \gamma) = \beta \exp(-\gamma g_{t+1}) \tag{43}$$

where $\beta$ is the time preference parameter, $\gamma$ is the risk aversion parameter, and $g_{t+1}$ is log consumption growth from time $t$ to $t+1$. As shown in Bansal and Lehmann (1997), Hansen (2012) and Backus, Chernov, and Zin (2013), the SDF (43) has the same permanent component (and therefore entropy of the permanent component) and implies the same long-term yield as SDFs of the form

$$m(X_t, X_{t+1}; \beta, \gamma) = \beta \exp(-\gamma g_{t+1}) \frac{h(X_{t+1})}{h(X_t)} \tag{44}$$

where $h$ is a positive function. For example, $h(X_t)$ may capture a limiting version of recursive preferences as in Hansen (2012). Alternatively, $h(X_t)$ may be an external habit formation component as in Chen and Ludvigson (2009). The following analysis therefore applies to a wider class of consumption-based asset pricing models than simply the consumption CAPM.

Three specifications of the state process are investigated, namely $X_t = g_t$, $X_t = (g_t, g_{e,t})'$ where $g_{e,t}$ denotes the growth in corporate earnings from time $t-1$ to time $t$, and $X_t = (g_t, r_{f,t})'$ where $r_{f,t} = \log R_{1,t+1}$ denotes the short-term risk-free rate at date $t$. Corporate earnings growth is included as a state variable in line with Hansen, Heaton, and Li (2008) who, in a different but related application, model log consumption and log corporate earnings jointly using a Gaussian vector autoregression.[10] The risk-free rate is included in the state process for comparison with Case I of Bansal and Yaron (2004), in which log consumption growth is modeled as

$$\begin{aligned} g_{t+1} &= \mu + x_t + \sigma_g e_{t+1} \\ x_{t+1} &= \rho x_t + \sigma_x \eta_{t+1} \end{aligned} \tag{45}$$

where $x_t$ is a latent predictable component of consumption growth and $e_{t+1}$ and $\eta_{t+1}$ are mutually independent and i.i.d. $N(0,1)$. When $\rho \in (-1,1)$ the state vector $X_t = (g_t, x_t)'$ is a strictly stationary and ergodic first-order Markov process. The risk-free rate in Case I of Bansal and Yaron (2004) is an affine function of $x_t$. Therefore, in Case I of Bansal and Yaron (2004) the observable vector $(g_t, r_{f,t})'$ and the partially latent vector $(g_t, x_t)'$ contain the same information: one can simply rewrite $x_t$ as an affine function of $r_{f,t}$. Both Hansen, Heaton, and Li (2008) and Bansal and Yaron (2004) assume a representative agent with Epstein-Zin-Weil recursive preferences. The SDF (43) is a restricted parameterization of the recursive preferences SDF used in these models (obtained by setting the elasticity of intertemporal substitution equal to $\gamma^{-1}$). As is common practice, it is assumed that the household decision interval coincides with the sampling interval.

8.1. **Data.** The data span 1947:Q2 to 2012:Q4 (263 observations). Data on aggregate consumption, corporate earnings, and population size were sourced from the National Income and Product Accounts (NIPA) tables. The consumption growth series is formed by taking seasonally adjusted consumption of nondurables and services data (NIPA Table 2.3.5), deflating by the implicit price

[10] Hansen, Heaton, and Li (2008) model log consumption and log corporate earnings as a cointegrated, fifth-order Gaussian vector autoregression. By contrast, here log consumption growth and log earnings growth are modeled jointly as a first-order (possibly nonlinear) Markov process.

| | $g$ | $r_f$ | $g_e$ |
|---|---|---|---|
| Mean | 0.0053 | 0.0030 | 0.0115 |
| Std Dev | 0.0055 | 0.0069 | 0.0733 |
| Skewness | -0.4631 | -0.4747 | -0.1722 |
| Kurtosis | 4.2770 | 5.3039 | 8.2600 |
| AR(1) | 0.2859 | 0.7446 | 0.0016 |

TABLE 3. Summary statistics for quarterly U.S. per capita (log) growth in consumption of nondurables and services $g$, risk-free rate $r_f$, and quarterly growth in corporate earnings $g_e$. AR(1) denotes the first order autocorrelation coefficient. The data span 1947:Q2–2012:Q4.

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). After-tax corporate earnings data (NIPA Table 1.12) were used because dividends are paid out to stockholders on an after-tax basis (Longstaff and Piazzesi, 2004). The risk-free rate was taken as the three-month Treasury bill rate (from CRSP). The earnings growth and risk-free series were both converted to real rates using the PCE deflator data.

Summary statistics of the consumption growth, risk-free rate and corporate earnings growth series are presented in Table 3. All series exhibit negative skewness and excess kurtosis. The consumption growth and risk-free rate series are positively autocorrelated. Earnings growth exhibits little persistence, but is much more volatile than consumption growth.

8.2. **Implementation.** The time preference parameter was set at $\beta = 0.998^3$. The risk aversion parameter was varied between $\gamma = 0$ (risk neutrality) and $\gamma = 30$. Hermite polynomial bases were formed for each series (centering and scaling by the sample mean and standard deviation of each series). A sieve spaces of dimension $K = 8$ is used for $X_t = g_t$, and sieve spaces of dimension $K = 16$ are used for $X_t = (g_t, r_{f,t})'$ and $X_t = (g_t, g_{e,t})'$. The sieve spaces for bivariate state vectors are formed by taking the tensor product of two univariate bases of dimension four. The OSLRV estimator $\widehat{V}^{os}_{L,J}$ was implemented with a cosine basis of dimension $J = 10$. The estimates were insensitive to both the dimension of the sieve space and the dimension of the basis used to compute the OSLRV estimators.

8.3. **Results.** Figure 3 displays the estimates $\widehat{\phi}$ and $\widehat{\phi}^\star$ with $\gamma = 5$, $\gamma = 15$, and $\gamma = 25$ for the case $X_t = g_t$. The estimates $\widehat{\phi}$ and $\widehat{\phi}^\star$ are more acutely sloped for higher levels of $\gamma$. The estimated positive eigenfunctions are decreasing in $g$, which implies that the price of long-horizon assets is a decreasing function of aggregate consumption growth.

The estimated long-run yield $\widehat{y}$ and entropy of the permanent component $\widehat{L}$ are plotted in Figure 4 for $X_t = g_t$ and in Figure 5 for $X_t = (g_t, g_{e,t})'$. The entropy of the permanent component of the SDF is independent of $\beta$. The long-run yield depends on both $\beta$ and $\gamma$. The solid blue lines present the pointwise estimates, and the dashed blue lines are 90% pointwise confidence bands. Comparison of Figures 4 and 5 show that very similar estimated long-run yields and permanent

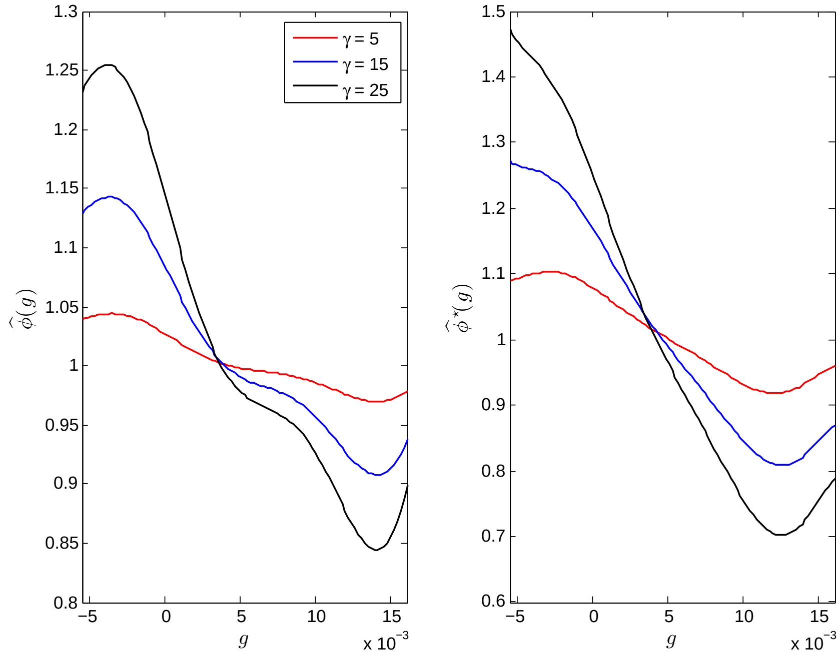


FIGURE 3. Estimated $\phi$ and $\phi^*$ for the consumption CAPM at different levels of risk aversion $\gamma$. The state variable is $X_t = g_t$, where $g_t$ is quarterly real U.S. per capita (log) growth in consumption of nondurables and services.

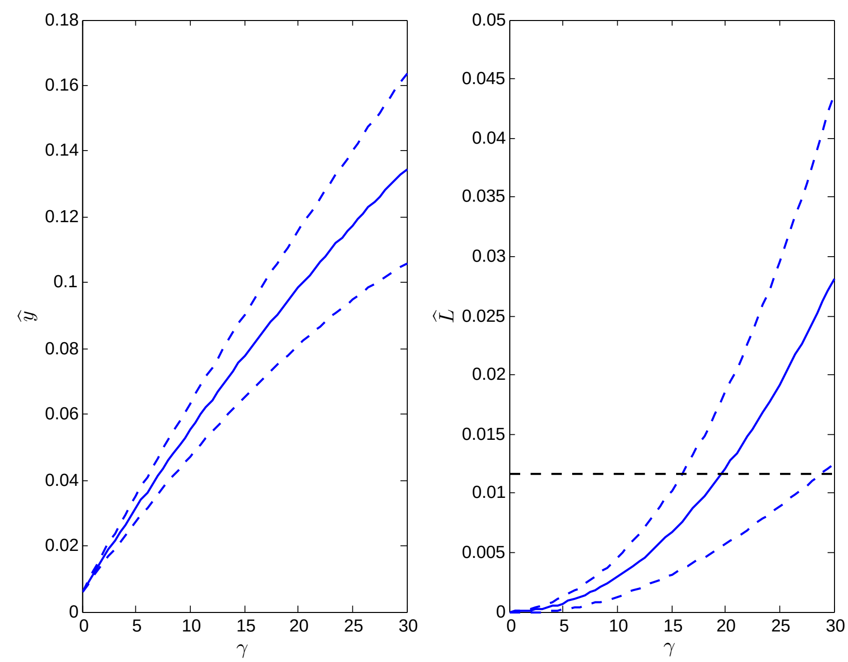


FIGURE 4. Estimated long-run yield $\widehat{y}$ and entropy of the permanent component of the SDF $\widehat{L}$ for the consumption CAPM at different levels of risk aversion $\gamma$, for $\beta = 0.998^3$ (solid blue lines). The state variable is $X_t = g_t$, where $g_t$ is quarterly real U.S. per capita (log) consumption growth. Dashed blue lines are pointwise asymptotic 90% confidence bands. The dashed black line represents the estimated average quarterly excess return on equities relative to long-term bonds.

component entropies are obtained for $X_t = g_t$ or $X_t = (g_t, g_{e,t})'$. Similar results are also obtained for the specification $X_t = (g_t, r_{f,t})$ (not presented).

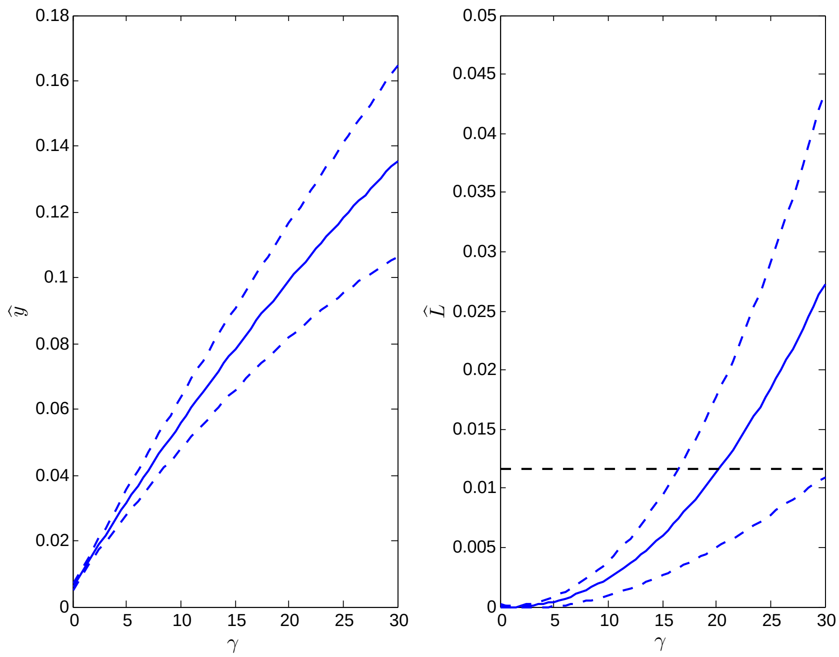


FIGURE 5. Estimated long-run yield $\widehat{y}$ and entropy of the permanent component of the SDF $\widehat{L}$ for the consumption CAPM at different levels of risk aversion $\gamma$, for $\beta = 0.998^3$ (solid blue lines). The state variable is $X_t = (g_t, g_{e,t})'$, where $g_t$ is quarterly real U.S. per capita (log) consumption growth and $g_{e,t}$ is quarterly real U.S. corporate earnings growth. Dashed blue lines are pointwise asymptotic 90% confidence bands. The dashed black line represents the estimated average quarterly excess return on equities relative to long-term bonds.

The entropy of the permanent component of the SDF is an upper bound for the average return on risky assets relative to long-term bonds (see equation (14)). An average excess return of 1.17% per quarter was estimated from the quarterly return on the NYSE/AMEX/NASDAQ combined index, including dividends, relative to the quarterly return on 30-year bonds over the sample period (both return series were sourced from CRSP).[11] The estimates presented in Figures 4 and 5 show that one requires $\gamma \geq 15$ for the estimated entropy of the permanent component of the consumption CAPM SDF to exceed 1.17%. As the bound (14) applies to all risky assets (not just the aggregate market), the lower bound for the entropy of the permanent component of the SDF would be at least as large as 1.17% if information on other assets was taken into account. A larger $\gamma$ would then be required to generate and estimated entropy that was compatible with a higher bound. This analysis suggests that the level of risk aversion required for the entropy of the permanent component of the consumption CAPM SDF to be compatible with historical average returns on equity to relative to long-term bonds is substantially higher than the threshold of 10 imposed by Mehra and Prescott (1985).

[11]The effect of coupon payments is ignored. Ignoring coupon payments is unlikely to have any substantial effect on the qualitative implications of these findings, however. The estimated quarterly premium for the combined index in excess of the three-month T-bill rate was 1.46% over the sample period. Historical data show that the term premium is small. For example, Backus, Chernov, and Zin (2013) suggest that the absolute value of the average yield spread is unlikely to exceed 0.1% monthly (see also Alvarez and Jermann (2005)).

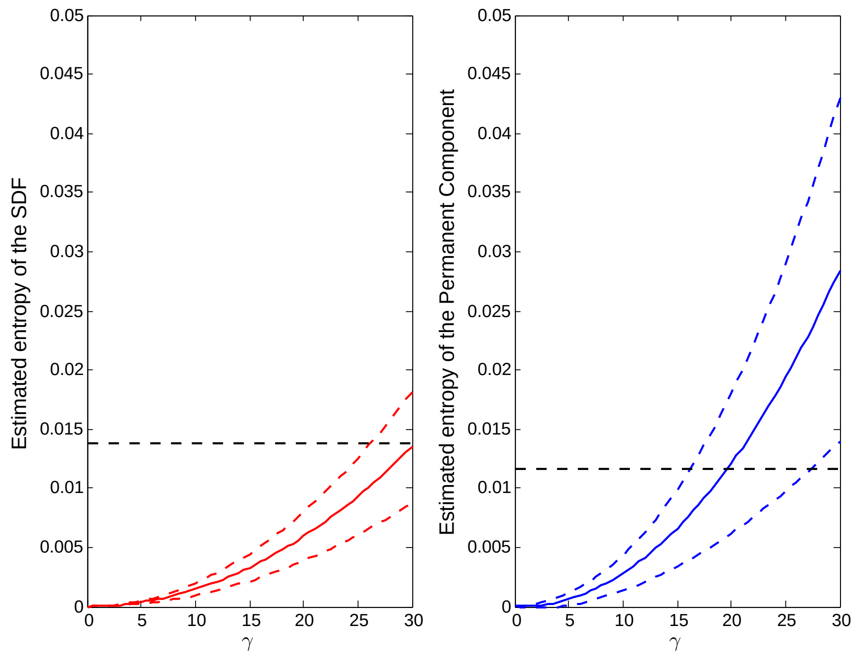


FIGURE 6. Estimated entropy of the SDF and entropy of the permanent component of the SDF for the consumption CAPM at different levels of risk aversion $\gamma$, for $\beta = 0.998^3$ (solid red and blue lines). The state variable is $X_t = g_t$, where $g_t$ is quarterly real U.S. per capita (log) consumption growth. Dashed lines are pointwise asymptotic 90% confidence bands. The dashed black lines represent historical average quarterly excess returns on equities relative to short-term bonds (left panel) and relative to long-term bonds (right panel).

As Figures 4 and 5 show, the estimated long-term yield is much larger than historical average long-term yields when $\gamma$ is set sufficiently high to rationalize the average return on equities relative to long-term bonds. For $\gamma \geq 15$ the estimated long-term *quarterly* yield is at least 6% per quarter. By contrast, the average real yield on the longest maturity Treasury bond over the period February 1959 to December 2012 was 0.76% per quarter.[12] Decreasing the time preference parameter $\beta$ further increases the estimated long-term yields. Under the restriction $\gamma \geq 15$, estimates of the long-term yield in line with historical average yields on long-term bonds were only obtainable with $\beta > 1$. Again, very similar results are obtained with $X_t = (g_t, r_{f,t})'$ (not presented)

These findings provide evidence of a long-term version of the equity premium and risk-free rate puzzles under the restrictions $\beta \in (0, 1)$ and $\gamma \in [0, 10]$ imposed by Mehra and Prescott (1985), at least to the extent that U.S. aggregate consumption growth can be described as a stationary Markov process with low-dimensional state vector. Similar qualitative results are obtained with monthly data.[13]

[12]This yield is estimated by taking the maximum treasury constant maturity yield available each month (either 20 or 30 years) from the Federal Reserve H-15 release, adjusting to a quarterly yield, and deflating using the implicit price deflator for personal consumption expenditures in the NIPA tables.

[13]Bakshi and Chabi-Yo (2012) estimate a monthly return premium in excess of long-term bonds of 0.41% per month, from U.S. market data spanning January 1932 to December 2010. With monthly data (spanning February 1959–December 2012), the level of $\gamma$ required to generate an estimated entropy compatible with this bound was in excess of 20 for $X_t = g_t$ and $X_t = (g_t, r_{f,t})'$ (corporate earnings are not available at the monthly frequency). Moreover, the

How does the entropy of the SDF in the consumption CAPM compare with the entropy of its permanent component? Figure 6 presents estimates of the entropy of the consumption CAPM SDF (left panel) and the entropy of its permanent component (right panel), together with their 90% pointwise confidence bands.[14] The dashed horizontal lines are the estimated average returns on equities relative to short-term bonds (left panel) and relative to long-term bonds (right panel) over the sample period.[15] The entropy of the SDF is an upper bound for the average return on risky assets relative to short-term bonds (see expression (17)). Figure 6 shows that the level of risk aversion required to generate an entropy of the SDF that rationalizes the historical average return on equities relative to short-term bonds is considerably larger than the level required to generate an entropy of the permanent component that rationalizes the historical average return on equities relative to long-term bonds. It may be possible to augment the consumption CAPM SDF by a term of the form $h(X_{t+1})/h(X_t)$ as in expression (44) so as to rationalize the premium relative to short term bonds at lower levels of risk aversion. However, such transitory distortions will not alter the permanent component of the SDF.

## 9. Conclusion

The long-run implications of a dynamic asset pricing model are jointly determined by both the functional form of the SDF and the short-run dynamics, or law of motion, of the variables in the model. The econometric framework introduced in this paper treats the dynamics as an unknown nuisance parameter. This paper introduces nonparametric sieve estimators of the positive eigenfunction and its eigenvalue (which are used to decompose the SDF into its permanent and transitory components), the long-term yield, and the entropy of the permanent component of the SDF. The sieve estimators are particularly easy to implement, and may also be used to numerically compute the long-run implications of fully specified models for which analytical solutions are unavailable. Consistency and convergence rates of the estimators are established, together with a means of performing asymptotic inference on the eigenvalue, long-run yield and entropy of the permanent component of the SDF. The estimators of the eigenvalue, long-run yield and entropy of the permanent component are shown to be semiparametrically efficient. Nonparametric identification conditions are presented for the positive eigenfunction in stationary discrete-time environments, and a version of the long-run pricing result of Hansen and Scheinkman (2009) is shown to obtain under the identification conditions.

There are several ways in which the research reported in this paper may be extended. One such extension is to study nonparametric identification and estimation in environments in which the state variable contains latent components or is measured with error. Data-driven methods for choosing the sieve space dimension would provide a more objective means for choosing the sieve

---

estimated long-run yields were at least 3% per month for values of $\gamma$ larger than 20. By contrast, the average real yield on the longest maturity Treasury bond over the sample period is around 0.25% per month.

[14]The entropy of the SDF is estimated by replacing the expected values in expression (16) by their sample averages. Confidence bands for the estimated entropy of the SDF are computed using the OSLRV estimator with a cosine basis of dimension $J = 10$.

[15]The return relative to short-term bonds is estimated from the quarterly return on the NYSE/AMEX/NASDAQ combined index, including dividends, relative to the three-month T-bill rate.

space dimension than the ad hoc approach used in this paper. Confidence bands for the estimated eigenfunction and the asymptotic distribution of functionals of the estimated eigenfunction would be useful for performing inference on the estimated eigenfunction. These extensions are currently being investigated by the author.

## Appendix

This appendix contains supplementary results and proofs. Appendix A presents high-level sufficient conditions for nonparametric identification of the positive eigenfunction. Versions of the long-term pricing result of Hansen and Scheinkman (2009) are also shown to obtain under the identification conditions. Appendix B briefly reviews some relevant concepts from spectral theory. Appendix C establishes consistency and convergence rates for nonparametric sieve estimators of the positive eigenfunctions of a collection of operators and their adjoints. Useful results on the convergence of random matrices are also presented in Appendix C. All proofs are in Appendix D.

Notation: Let $(\mathcal{A}, \mathscr{A}, \mu)$ be a measure space and let $p \in [1, \infty]$. The space $L^p(\mathcal{A}, \mathscr{A}, \mu)$ is abbreviated as $L^p(\mu)$. Let $\| \cdot \|_{L^p(\mu)}$ denote the $L^p(\mu)$ norm when applied to functions and the operator norm when applied to linear operators on the space $L^p(\mu)$. Let a.e.-$[\mu]$ denote almost everywhere with respect to the measure $\mu$ and a.e.-$[\mu \otimes \mu]$ denote almost everywhere with respect to the product measure $\mu \otimes \mu$. Let $\Gamma(\delta, \lambda)$ denote the positively oriented circle (in the complex plane) centered at $\lambda$ with radius $\delta$. Let $B(\delta, \lambda)$ denote the open ball (in the complex plane) centered at $\lambda$ with radius $\delta$. Finally, let $d(z, A) = \inf_{\zeta \in A} |z - \zeta|$ for $z \in \mathbb{C}$ and $A \subset \mathbb{C}$.

## Appendix A. Identification and long-term pricing

This appendix provides primitive sufficient conditions for the identification of the positive eigenfunction and adjoint eigenfunction in stationary discrete-time environments for which $\mathbb{M}$ may be represented as an integral operator with positive kernel. Existence is achieved by application of the Perron-Frobenius theorem for positive integral operators. Identification is then established by a type of Kreĭn-Rutman theorem. A restatement of the long-term pricing result of Hansen and Scheinkman (2009) is shown to obtain under these conditions.

Basic regularity conditions are first introduced. Let $\mathscr{X}$ be the Borel $\sigma$-algebra on $\mathcal{X}$. The following conditions are sufficient for existence and nonparametric identification of $\phi$.

**Assumption A.1.** $\{X_t\}$ *is a strictly stationary and ergodic (first-order) Markov process supported on a Borel set* $\mathcal{X} \subseteq \mathbb{R}^d$*, and its stationary distribution* $Q$ *has density* $q$ *with respect to Lebesgue measure, with* $q(x) > 0$ *for almost every* $x \in \mathcal{X}$*.*

**Assumption A.2.** $\mathbb{M} : L^p(Q) \to L^p(Q)$ *is bounded and* $\mathbb{M}_\tau$ *is compact for some* $\tau \geq 1$*.*

**Assumption A.3.** $\mathbb{M}$ *may be written as in (18) with integral kernel* $\mathcal{K}$ *as in (19) such that:*

*(i)* $\mathcal{K}(x, y) \geq 0$ *a.e.-*$[Q \otimes Q]$
*(ii)* $\int_{A^c} \int_A m(x, y) f(x, y) \, \mathrm{d}x \, \mathrm{d}y > 0$ *for every* $A \in \mathscr{X}$ *with* $0 < Q(A) < 1$*.*

Assumption A.1 is the same as Assumption 4.1(i). Assumptions A.2 and A.3 are higher-level conditions in place of Assumptions 4.1(ii)–(iv). Assumption A.2 only requires some power of $\mathbb{M}$ to be compact and is weaker than requiring $\mathbb{M}$ itself to be compact.[16] Assumption A.3 is trivially satisfied if $\mathcal{K}(x,y) > 0$ a.e.-$[Q \otimes Q]$. Assumption 4.1 is sufficient for Assumptions A.1, A.2, and A.3 for $p = 2$.

Let $\mathrm{spr}(\mathbb{M})$ denote the spectral radius of $\mathbb{M}$ (see Appendix B). Existence of $\phi$ follows by Theorem V.6.6 of Schaefer (1974).

**Theorem A.1** (Schaefer (1974))**.** *If Assumptions A.1, A.2 and A.3 hold for some $p \in [1,\infty]$, then there exists a $\phi \in L^p(Q)$ with $\phi(x) > 0$ a.e.-$[Q]$ such that $\mathbb{M}\phi = \rho\phi$ with $\rho = \mathrm{spr}(\mathbb{M}) > 0$, and $\phi$ is the unique (to scale) eigenfunction of $\mathbb{M}$ corresponding to the eigenvalue $\rho$.*

*If, in addition, $\mathcal{K} > 0$ a.e.-$[Q \otimes Q]$ then any other eigenvalue $\lambda$ of $\mathbb{M}$ has modulus $|\lambda| < \rho$.*

If $1 \leq p < \infty$ let the dual index $p'$ for $L^p(Q)$ be defined as $p^{-1} + p'^{-1} = 1$ with $p' = \infty$ if $p = 1$. The dual space of $L^p(Q)$ can be identified with the space $L^{p'}(Q)$ under the evaluation $E[\psi(X)\psi^*(X)]$ for $\psi \in L^p(Q)$, $\psi^* \in L^{p'}(Q)$. The adjoint operator $\mathbb{M}^* : L^{p'}(Q) \to L^{p'}(Q)$ is defined such that

$$E[\psi(X)\mathbb{M}^*\psi^*(X)] = E[\psi^*(X)\mathbb{M}\psi(X)]$$

for $\psi \in L^p(Q)$ and $\psi^* \in L^{p'}(Q)$. Let $(\mathcal{X}, \mathscr{X}_1, Q_1)$ denote the completion of $(\mathcal{X}, \mathscr{X}, Q)$ as described on p. 296 of Dunford and Schwartz (1958). The dual space of $L^\infty(Q)$ can be identified with the space $\mathrm{ba}(\mathcal{X}, \mathscr{X}_1, Q_1)$ of signed measures on $(\mathcal{X}, \mathscr{X}_1)$ which are absolutely continuous with respect to $Q_1$, under the evaluation $\psi^*(\psi) = \int_{\mathcal{X}} \psi \, \mathrm{d}\nu_{\psi^*}$ for $\psi \in L^\infty(Q)$ and $\nu_{\psi^*} \in \mathrm{ba}(\mathcal{X}, \mathscr{X}_1, Q_1)$ (Dunford and Schwartz, 1958, p. 296).

It follows from Theorem A.1 by a version of the Kreĭn-Rutman theorem due to Schaefer (1960) that $\phi$ is the unique non-negative eigenfunction of $\mathbb{M}$.

**Theorem A.2.** *If Assumptions A.1, A.2 and A.3 hold for some $p \in [1,\infty]$, then $\rho$ is an eigenvalue of $\mathbb{M}$ of multiplicity one, $\phi$ is the unique (to scale) eigenfunction of $\mathbb{M}$ with $\phi(x) \geq 0$ a.e.-$[Q]$ and:*

- *(i) If $p \in [1,\infty)$ there exists a $\phi^* \in L^{p'}(Q)$ with $\phi^*(x) > 0$ a.e.-$[Q]$ such that $\mathbb{M}^*\phi^* = \rho\phi^*$, and $\phi^*$ is the unique (to scale) eigenfunction of $\mathbb{M}^*$ with $\phi^*(x) \geq 0$ a.e.-$[Q]$*
- *(ii) If $p = \infty$ there exists a unique (to scale) nonzero $\Phi_1^* \in \mathrm{ba}(\mathcal{X}, \mathscr{X}_1, Q_1)$ such that $\Phi_1^*(A) \geq 0$ for all $A \in \mathscr{X}$ with $Q(A) > 0$ and*

$$\int_{\mathcal{X}} \mathbb{M}\psi(x) \, \mathrm{d}\Phi_1^*(x) = \rho \int \psi(x) \, \mathrm{d}\Phi_1^*(x)$$

*for all $\psi \in L^\infty(Q)$.*

Versions of the long-run pricing result

$$\lim_{\tau \to \infty} \rho^{-\tau}\mathbb{M}_\tau \psi(X_t) = \widetilde{E}[\psi(X)/\phi(X)]\phi(X_t)$$

of Hansen and Scheinkman (2009) hold under the identification conditions just presented. First consider the case $1 \leq p < \infty$. The positive eigenfunction and adjoint positive eigenfunction exist under

[16]There are several sufficient conditions for this compactness condition. For $1 < p < \infty$ this is satisfied if there is a $\tau \geq 1$ such that $\mathbb{M}_\tau$ maps $L^p(Q)$ into $L^\infty(Q)$, for $p = 1$ if there is a $\tau \geq 1$ such that $\mathbb{M}_\tau$ maps $L^1(Q)$ into $L^r(Q)$ for some $r > 1$, and for $p = \infty$ if there is a $\tau \geq 1$ such that $\mathbb{M}_\tau$ has a continuous extension that maps $L^r(Q)$ into $L^\infty(Q)$ for some $r < \infty$ (Schaefer, 1974, p. 337).

Assumptions A.1, A.2 and A.3. Impose the normalizations $E[\phi(X)^p] = 1$ and $E[\phi(X)\phi^*(X)] = 1$, and let $P : L^p(Q) \to L^p(Q)$ be defined as

$$P\psi(x) = E[\psi(X)\phi^*(X)]\phi(x) .$$

**Theorem A.3.** *If Assumptions A.1, A.2 and A.3 hold for some $p \in [1, \infty)$ with $\mathcal{K}(x, y) > 0$ a.e.-$[Q \otimes Q]$, then there exists $c > 0$ such that $\|\rho^{-\tau}\mathbb{M}^\tau - P\|_{L^p(Q)} = O(e^{-c\tau})$ as $\tau \to \infty$.*

Now consider the space $L^\infty(Q)$. Under Assumptions A.1, A.2 and A.3 the positive eigenfunction exists, together with a nonzero measure $\Phi_1^* \in \mathrm{ba}(\mathcal{X}, \mathscr{X}_1, Q_1)$ such that $\Phi_1^*(A) \geq 0$ for all $A \in \mathscr{X}$ with $Q(A) > 0$. Normalize $\phi$ and $\Phi_1^*$ so that $\Phi_1^*(\mathcal{X}) = 1$ (making $\Phi_1^*$ a probability measure) and $\int_{\mathcal{X}} \phi \, \mathrm{d}\Phi_1^* = 1$. Let $P : L^\infty(Q) \to L^\infty(Q)$ be defined as

$$P\psi(x) = \left( \int_{\mathcal{X}} \psi \, \mathrm{d}\Phi_1^* \right) \phi(x) .$$

**Theorem A.4.** *If Assumptions A.1, A.2 and A.3 hold for $p = \infty$ with $\mathcal{K}(x, y) > 0$ a.e.-$[Q \otimes Q]$, then there exists $c > 0$ such that $\|\rho^{-\tau}\mathbb{M}^\tau - P\|_{L^\infty(Q)} = O(e^{-c\tau})$.*

Both Theorem A.3 and Theorem A.4 show that versions of the long-run pricing result hold in operator norm with the approximation error vanishing exponentially with $\tau$. The theorems also show how to calculate the twisted probability measure $\widetilde{Q}$ used to calculate the unconditional expectation $\widetilde{E}$ when $1 \leq p < \infty$. Specifically, the Radon-Nikodym derivative of the twisted measure $\widetilde{Q}$ with respect to $Q$ is

$$\frac{\mathrm{d}\widetilde{Q}(x)}{\mathrm{d}Q(x)} = \phi(x)\phi^*(x) . \tag{46}$$

Therefore, the twisted expectation $\widetilde{E}$ may be recovered by solving the eigenfunction problems $\mathbb{M}\phi = \rho\phi$ and $\mathbb{M}^*\phi^* = \rho\phi^*$.

## Appendix B. Brief review of spectral theory

Relevant concepts from spectral theory are briefly reviewed. Definitions are as in Kato (1980), unless stated otherwise. Let $E$ be a Banach space and $T : E \to E$ be a bounded linear operator. The definitions are presented in the case that $E$ is a Banach space over $\mathbb{C}$. When $E$ is a Banach space over $\mathbb{R}$, the definitions are applied to the complex extension of $T$, given by $T(x+iy) = T(x)+iT(y)$ for $x, y \in E$, on the complexification of $E$, namely $E + iE$ of $E$.

The resolvent set $\mathrm{res}(T) \subseteq \mathbb{C}$ of $T$ is the set of all $z \in \mathbb{C}$ for which the resolvent operator $R(T, z) := (T - zI)^{-1}$ is a bounded linear operator on $E$ (where $I : E \to E$ is the identity). The spectrum $\sigma(T)$ is defined as the complement in $\mathbb{C}$ of $\mathrm{res}(T)$, i.e. $\sigma(T) := (\mathbb{C} \setminus \mathrm{res}(T))$. If $S : E \to E$ is another bounded linear operator and $z \in \mathrm{res}(T) \cup \mathrm{res}(S)$ then the so-called second resolvent equation obtains:

$$R(S, z) - R(T, z) = R(S, z)(T - S)R(T, z) . \tag{47}$$

The spectral radius $\mathrm{spr}(T)$ of $T$ is $\mathrm{spr}(T) = \sup\{|\lambda| : \lambda \in \sigma(T)\}$. The Gelfand formula shows that $\mathrm{spr}(T) = \lim_{n\to\infty} \|T^n\|^{1/n}$. The point spectrum $\pi(T) \subseteq \sigma(T)$ of $T$ is the set of all $z \in \mathbb{C}$ for which the null space of $(T - zI)$ is not 0. When $\pi(T)$ is nonempty, each $\lambda \in \pi(T)$ is an eigenvalue of $T$ and any nonzero $\psi$ in the null space of $(T - \lambda I)$ is an eigenvector of $T$ corresponding to $\lambda$. The dimension of the null space of $(T - \lambda I)$ is the geometric multiplicity of the eigenvalue $\lambda$.

An eigenvalue $\lambda \in \pi(T)$ is said to be isolated if $\inf_{z\in\sigma(T):z\neq\lambda} |z - \lambda| \geq 2\epsilon$ for some $\epsilon > 0$, in which case the spectral projection of $T$ corresponding to $\lambda$ can be written as

$$P = \frac{-1}{2\pi i} \int_{\Gamma(\epsilon,\lambda)} R(T, z)\, \mathrm{d}z\,. \tag{48}$$

The dimension of $PE$ is the algebraic multiplicity of $\lambda$. The algebraic multiplicity is the order of the pole of $R(T, z)$ at $z = \lambda$ and is at least as large as the geometric multiplicity of $\lambda$ (Chatelin, 1983). The term multiplicity is used for eigenvalues whose algebraic and geometric multiplicity are equal. Suppose that $\lambda$ is an isolated real eigenvalue of $T$. Then $\lambda$ is an isolated real eigenvalue of the adjoint $T^*$ of $T$, and the algebraic and geometric multiplicities of $\lambda$ are the same for $T$ and $T^*$. If, in addition, $\lambda$ has multiplicity one, then $P = x \otimes x^*$ with where $(x \otimes x^*)\psi = x^*(\psi)x$ and where $x$ and $x^*$ are eigenvectors of $T$ and $T^*$ corresponding to $\lambda$, and $x^*(x) = 1$ (Chatelin, 1983, p. 113). If, in addition, $E$ is a Hilbert space then $P = (x \otimes x^*)$ is given by $(x \otimes x^*)\psi = \langle \psi, x^*\rangle x$ where $x^*$ is an eigenvector of $T^*$ corresponding to $\lambda$, $\|x\| = 1$ and $\langle x, x^*\rangle = 1$. In this case $\|P\| = \|x^*\| \geq 1$, with $x^* = x$ and $\|P\| = 1$ if $P$ is an orthogonal projection (which is the case when $T$ is selfadjoint).

## Appendix C. Additional results for estimation

C.1. **Estimation of eigenvalues and eigenfunctions.** Let $\{X_t\}$ be a strictly stationary (not necessarily Markov) process with stationary distribution $Q$ and whose support is a Borel set $\mathcal{X} \subseteq \mathbb{R}^d$. Consider a set of operators $\{\mathbb{M}_\alpha : \alpha \in \mathcal{A}\}$ indexed by an arbitrary parameter $\alpha \in \mathcal{A}$, where each $\mathbb{M}_\alpha : L^2(Q) \to L^2(Q)$ is given by

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

for some $m(\cdot, \cdot; \alpha) : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$. This setup trivially nests the case dealt with in the body of the paper in which $\mathcal{A}$ is a singleton. Suppose each $\mathbb{M}_\alpha$ has a isolated eigenvalue $\rho_\alpha = \mathrm{spr}(\mathbb{M}_\alpha)$ and unique positive eigenfunction $\phi_\alpha$ corresponding to $\rho_\alpha$ (so each $\mathbb{M}_\alpha^*$ has a unique adjoint eigenfunction $\phi_\alpha^*$). Uniform (in $\alpha$) convergence rates of nonparametric sieve estimators of $(\rho_\alpha, \phi_\alpha, \phi_\alpha^*)$ are now established.

The following analysis is conducted in $L^2(Q)$ as in the body of the paper. Let $B_K$ be the sieve space spanned by the basis functions $\{b_{K1}, \ldots, b_{KK}\}$ and let $\Pi_K^b$ be the orthogonal projection onto $B_K$. Under regularity conditions, for $K$ sufficiently large the largest eigenvalue $\rho_{\alpha,K}$ of $\Pi_K^b \mathbb{M}_\alpha$ will be real and positive and have multiplicity one for all $\alpha \in \mathcal{A}$. Let $\phi_{\alpha,K} \in B_K$ be an eigenfunction of $\Pi_K^b \mathbb{M}_\alpha$ corresponding to $\rho_{\alpha,K}$. Similarly, the adjoint in $L^2(Q)$ of $\Pi_K^b \mathbb{M}_\alpha$ will have a unique eigenfunction $\phi_{\alpha,K}^*$ corresponding to $\rho_{\alpha,K}$, and the adjoint in $B_K$ of $\Pi_K^b \mathbb{M}_\alpha|_{B_K} : B_K \to B_K$ will have an eigenfunction $\phi_{\alpha,K}^\star \in B_K$ corresponding to $\rho_{\alpha,K}$. As all quantities are defined up to sign and scale, impose the sign normalizations $E[\phi_{\alpha,K}(X)\phi_\alpha(X)] \geq 0$, $E[\phi^*_{\alpha,K}(X)\phi^*_\alpha(X)] \geq 0$ and $E[\phi^\star_{\alpha,K}(X)\phi^*_\alpha(X)] \geq 0$ and the scale normalizations $E[\phi_{\alpha,K}(X)^2] = 1$, $E[\phi_{\alpha,K}(X)\phi^*_{\alpha,K}(X)] = 1$ and $E[\phi_{\alpha,K}(X)\phi^\star_{\alpha,K}(X)] = 1$.

Let the Gram matrix $\mathbf{G}_K$ and its estimator $\widehat{\mathbf{G}}_K$ be as in the body of the paper. For each $\alpha \in \mathcal{A}$ let $\mathbf{M}_{\alpha,K}$ be as defined as in (24) with $m(\cdot,\cdot;\alpha)$ in place of $m(\cdot,\cdot)$ and let $\widehat{\mathbf{M}}_{\alpha,K}$ be a $K \times K$ matrix estimator of $\mathbf{M}_{\alpha,K}$ (i.e. a measurable function of the sample data). Under regularity conditions, with probability approaching one $\widehat{\mathbf{G}}_K$ is invertible and for each $\alpha \in \mathcal{A}$ the eigenvector problems

$$\begin{aligned} \widehat{\mathbf{G}}_K^{-1}\widehat{\mathbf{M}}_{\alpha,K}\widehat{c}_\alpha &= \widehat{\rho}_\alpha\widehat{c}_\alpha \\ \widehat{\mathbf{G}}_K^{-1}\widehat{\mathbf{M}}'_{\alpha,K}\widehat{c}^\star_\alpha &= \widehat{\rho}_\alpha\widehat{c}^\star_\alpha \end{aligned}$$

are solvable, where $\widehat{\rho}_\alpha = \lambda_{\max}(\widehat{\mathbf{G}}_K^{-1}\widehat{\mathbf{M}}_{\alpha,K})$ is real and positive. Then for each $\alpha \in \mathcal{A}$, $\widehat{\rho}_\alpha$ is the estimator of $\rho_\alpha$, $\widehat{\phi}_\alpha = b^K(x)'\widehat{c}_\alpha$ is the estimator of $\phi_\alpha$, and $\widehat{\phi}^\star_\alpha = b^K(x)'\widehat{c}^\star_\alpha$ is the estimator of $\phi^*_\alpha$. As these eigenfunction estimators are only defined up to scale, impose the sign normalizations $E[\widehat{\phi}_\alpha(X)\phi_{\alpha,K}(X)] \geq 0$ and $E[\widehat{\phi}^\star_\alpha(X)\phi^\star_{\alpha,K}(X)] \geq 0$ and the scale normalizations $\|\widehat{\phi}_\alpha\| = 1$ and $E[\widehat{\phi}_\alpha(X)\widehat{\phi}^\star_\alpha(X)] = 1$.

Some definitions are required before introducing the regularity conditions. As in Section 5.2, let $\widetilde{b}^K$ denote the vector of orthonormalized basis functions and let

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

For each $\alpha \in \mathcal{A}$ define

$$\widehat{\widetilde{\mathbf{M}}}_{\alpha,K} = \mathbf{G}_K^{-1/2}\widehat{\mathbf{M}}_{\alpha,K}\mathbf{G}_K^{-1/2}\,.$$

The orthonormalized estimators $\widehat{\widetilde{\mathbf{G}}}_K$ and $\widehat{\widetilde{\mathbf{M}}}_{\alpha,K}$ are infeasible and do not actually need to be constructed, but it makes the asymptotic arguments easier to work with them in place of $\widehat{\mathbf{G}}_K$ and $\widehat{\mathbf{M}}_{\alpha,K}$. Note that any $\psi \in B_K$ can be written as $\psi_K(x) = \widetilde{c}_K(\psi)'\widetilde{b}^K(x)$ for some $\widetilde{c}_K(\psi) \in \mathbb{R}^K$. The space $B_K$ is therefore isomorphic to $\mathbb{R}^K$ endowed with the Euclidean inner (dot) product, since

$$E[\psi_1(X)\psi_2(X)] = \widetilde{c}_K(\psi_1)'E[\widetilde{b}^K(x)\widetilde{b}^K(x)']\widetilde{c}_K(\psi_2) = \widetilde{c}_K(\psi_1)'\widetilde{c}_K(\psi_2)\,.$$

Therefore the matrix spectral norm $\|\cdot\|_2$ when applied to the orthonormalized matrices in $\mathbb{R}^{K\times K}$ is isomorphic to the operator norm for linear operators on $B_K$.

The spectrum $\sigma(\cdot)$ and resolvent $R(\cdot,z)$ are defined in Appendix B. Let $\widetilde{c}_{\alpha,K}, \widetilde{c}^\star_{\alpha,K} \in \mathbb{R}^K$ be such that $\widetilde{b}^K(x)'\widetilde{c}_{\alpha,K} = \phi_{\alpha,K}$ and $\widetilde{b}^K(x)'\widetilde{c}^\star_{\alpha,K} = \phi^\star_{\alpha,K}$. Let $\{\delta_K, \delta^*_K, \bar{\eta}_{1,n,K}, \bar{\eta}_{2,n,K}, \eta_{1,n,K}, \eta_{2,n,K} : n, K \geq 1\}$ be sequences of positive real values to be defined in the following assumptions. In the event of measurability issues, outer probabilities are used below implicitly in place of probabilities.

**Assumption C.1.** *The set of operators* $\{\mathbb{M}_\alpha : \alpha \in \mathcal{A}\}$ *satisfies:*

*(i) for each* $\alpha \in \mathcal{A}$, $\mathbb{M}_\alpha : L^2(Q) \to L^2(Q)$ *is a bounded linear operator and* $\rho_\alpha = \mathrm{spr}(\mathbb{M}_\alpha)$ *is an isolated eigenvalue of* $\mathbb{M}_\alpha$ *of multiplicity one*

*(ii)* $\sup_{\alpha\in\mathcal{A}}\|\mathbb{M}_\alpha\| < \infty$ *and* $\bar{\epsilon} := \inf_{\alpha\in\mathcal{A}} d(\rho_\alpha, \sigma(\mathbb{M}_\alpha) \setminus \{\rho_\alpha\}) > 0$.

**Assumption C.2.** *The sieve approximation error satisfies:*

*(i)* $\sup_{\alpha\in\mathcal{A}} \|\Pi_K^b \mathbb{M}_\alpha - \mathbb{M}_\alpha\| = O(\bar\delta_K)$ *where* $\bar\delta_K = o(1)$ *as* $K \to \infty$

*(ii)* $\sup_{\alpha\in\mathcal{A}} \|(\Pi_K^b \mathbb{M}_\alpha - \mathbb{M}_\alpha)\phi_\alpha\| = O(\delta_K)$, $\sup_{\alpha\in\mathcal{A}} \|(\mathbb{M}_\alpha^*\Pi_K^b - \mathbb{M}_\alpha^*)\phi_\alpha^*/\|\phi_\alpha^*\|\| = O(\delta_K^*)$.

**Assumption C.3.** *There exists a continuous decreasing function* $r : (0,\infty) \to (0,\infty)$ *such that for each* $\alpha \in \mathcal{A}$*:*

*(i)* $\|R(\mathbb{M}_\alpha, z)\| \le r(d(z, \sigma(\mathbb{M}_\alpha)))$ *for all* $z \in (B(\bar\epsilon, \rho_\alpha) \setminus \sigma(\mathbb{M}_\alpha))$

*(ii)* $\|R(\Pi_K^b \mathbb{M}_\alpha|_{B_K}, z)\| \le r(d(z, \sigma(\Pi_K^b \mathbb{M}_\alpha|_{B_K})))$ *for all* $z \in (B(\bar\epsilon, \rho_\alpha) \setminus \sigma(\Pi_K^b \mathbb{M}_\alpha|_{B_K}))$.

**Assumption C.4.** *The matrix estimators and their population counterparts are such that:*

*(i)* $\lambda_{\min}(\mathbf{G}_K) > 0$ *for every* $K \ge 1$

*(ii)*

$$\begin{aligned} \|\widehat{\widetilde{\mathbf{G}}}_K - I_K\|_2 &= O_p(\bar\eta_{1,n,K}) \\ \sup_{\alpha\in\mathcal{A}} \|\widehat{\widetilde{\mathbf{M}}}_{\alpha,K} - \widetilde{\mathbf{M}}_{\alpha,K}\|_2 &= O_p(\bar\eta_{2,n,K}) \end{aligned}$$

*where* $\bar\eta_{n,K} = \max\{\bar\eta_{1,n,K}, \bar\eta_{2,n,K}\} = o(1)$ *as* $n, K \to \infty$.

*(iii)*

$$\begin{aligned} \sup_{\alpha\in\mathcal{A}} \|(\widehat{\widetilde{\mathbf{G}}}_K - \widetilde{\mathbf{G}}_K)\widetilde{c}_{\alpha,K}\|_2 &= O_p(\eta_{1,n,K}) \\ \sup_{\alpha\in\mathcal{A}} \|(\widehat{\widetilde{\mathbf{G}}}_K - \widetilde{\mathbf{G}}_K)\widetilde{c}^\star_{\alpha,K}/\|\widetilde{c}^\star_{\alpha,K}\|_2\|_2 &= O_p(\eta_{1,n,K}) \\ \sup_{\alpha\in\mathcal{A}} \|(\widehat{\widetilde{\mathbf{M}}}_{\alpha,K} - \widetilde{\mathbf{M}}_{\alpha,K})\widetilde{c}_{\alpha,K}\|_2 &= O_p(\eta_{2,n,K}) \\ \sup_{\alpha\in\mathcal{A}} \|(\widehat{\widetilde{\mathbf{M}}}'_{\alpha,K} - \widetilde{\mathbf{M}}'_{\alpha,K})\widetilde{c}^\star_{\alpha,K}/\|\widetilde{c}^\star_{\alpha,K}\|_2\|_2 &= O_p(\eta_{2,n,K})\,. \end{aligned}$$

Assumption C.1(i) ensures the positive eigenfunction of $\mathbb{M}_\alpha$ exists and is identified for each $\alpha \in \mathcal{A}$. Part (ii) of Assumption C.1 ensures the operators are uniformly bounded and the eigenvalues $\{\rho_\alpha : \alpha \in \mathcal{A}\}$ are uniformly well separated from the rest of the spectrum of $\{\mathbb{M}_\alpha : \alpha \in \mathcal{A}\}$. Assumption C.1(ii) is implicitly satisfied by Assumption C.1(i) if $\mathcal{A}$ has finite cardinality. Assumption C.2 ensures the ranges of the operators $\mathbb{M}_\alpha$ are uniformly well approximated over the sieve space as $K$ increases. Assumption C.3 is required to ensure the spectrum of each $\Pi_K^b \mathbb{M}_\alpha$ remains sufficiently continuous as the dimension of the sieve space increases, and is trivially satisfied with $r(x) = x^{-1}$ if $\mathbb{M}_\alpha$ and $\Pi_K^b \mathbb{M}_\alpha|_{B_K}$ are normal or selfadjoint operators. Bounds are also available for common classes of compact operators, such as Hilbert-Schmidt and other Schatten-class operators (see Bandtlow (2004)). If $T$ is a linear operator on a Hilbert space the lower bound $\|R(T,z)\| \ge d(z, \sigma(T))^{-1}$ for all $z \in \mathrm{res}(T)$ obtains generically. Assumption C.4(i) is a standard condition for nonparametric estimation with a linear sieve space and is made to ensure that $\mathbf{G}_K$ is invertible uniformly in $K$. Assumption C.4(ii) defines the rate of convergence of the matrix estimators. Assumption C.4(ii) is sufficient for Assumption C.4(iii) with $\eta_{1,n,K} = \bar\eta_{1,n,K}$ and $\eta_{2,n,K} = \bar\eta_{2,n,K}$ (by the relation between the spectral and Euclidean norms) but may lead to improved rates of convergence for $\widehat\phi_\alpha$ and $\widehat\phi_\alpha^*$ in certain circumstances.

The following two Theorems calculate the "bias" and "variance" components of the rates of convergence separately. These are proved by extending arguments in Gobet, Hoffmann, and Reiß (2004) to estimate the eigenfunction and adjoint eigenfunctions of nonselfadjoint operators.

**Theorem C.1.** *Under Assumptions C.1, C.2, and C.3(i), there exists $\bar{K}$ sufficiently large such that for each $K \geq \bar{K}$, $\rho_{\alpha,K}$ is real and positive and has multiplicity one and $\phi_{\alpha,K}$ and $\phi^*_{\alpha,K}$ are unique for each $\alpha \in \mathcal{A}$, and*

*(i)* $\sup_{\alpha\in\mathcal{A}} |\rho_\alpha - \rho_{\alpha,K}| = O(\delta_K)$
*(ii)* $\sup_{\alpha\in\mathcal{A}} \|\phi_\alpha - \phi_{\alpha,K}\| = O(\delta_K)$
*(iii)* $\sup_{\alpha\in\mathcal{A}} \|\phi^*_\alpha/\|\phi^*_\alpha\| - \phi^*_{\alpha,K}/\|\phi^*_{\alpha,K}\|\| = O(\delta^*_K)$
*(iv)* $\sup_{\alpha\in\mathcal{A}} \|\phi^*_\alpha - \phi^*_{\alpha,K}\| = O(\bar{\delta}_K)$.

**Theorem C.2.** *Under Assumptions C.1, C.2, C.3 and C.4, there is a set whose probability approaches one on which $\widehat{\rho}_\alpha$ is real and positive and $\widehat{\phi}_\alpha$ and $\widehat{\phi}^*_\alpha$ are unique for each $\alpha \in \mathcal{A}$, and*

*(i)* $\sup_{\alpha\in\mathcal{A}} |\widehat{\rho}_\alpha - \rho_{\alpha,K}| = O_p(\eta_{n,K})$
*(ii)* $\sup_{\alpha\in\mathcal{A}} \|\widehat{\phi}_\alpha - \phi_{\alpha,K}\| = O_p(\eta_{n,K})$
*(iii)* $\sup_{\alpha\in\mathcal{A}} \|\widehat{\phi}^\star_\alpha/\|\widehat{\phi}^\star_\alpha\| - \phi^\star_{\alpha,K}/\|\phi^\star_{\alpha,K}\|\| = O_p(\eta_{n,K})$
*(iv)* $\sup_{\alpha\in\mathcal{A}} \|\widehat{\phi}^\star_\alpha - \phi^\star_{\alpha,K}\| = O_p(\bar{\eta}_{n,K})$.

The assumptions of Theorem C.2 are sufficient to establish a uniform asymptotic expansion of the eigenvalue estimators $\widehat{\rho}_\alpha$.

**Theorem C.3.** *Under Assumptions C.1, C.2, C.3 and C.4,*

$$\sup_{\alpha\in\mathcal{A}} \left| \widehat{\rho}_\alpha - \rho_{\alpha,K} - \widetilde{c}^{\star\prime}_{\alpha,K}(\widehat{\widetilde{\mathbf{G}}}_K^{-1}\widehat{\widetilde{\mathbf{M}}}_{\alpha,K} - \widetilde{\mathbf{M}}_{\alpha,K})\widetilde{c}_{\alpha,K} \right| = O_p(\bar{\eta}^2_{n,K})\,.$$

C.2. **Additional results on convergence of the matrix estimators.** The following Lemmas are useful to verify Assumption C.4. Let

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

The results are presented under different weak-dependence conditions and different assumptions on the number of moments of $m(X_0, X_1)$.

**Assumption C.5.** $\lambda_{\min}(\mathbf{G}_K) \geq \underline{\lambda} > 0$ *for all $K \geq 1$.*

**Assumption C.6.** $\{X_t\}_{t\in\mathbb{Z}}$ *is strictly stationary and geometrically beta-mixing.*

**Assumption C.7.** $\{X_t\}_{t\in\mathbb{Z}}$ *is strictly stationary and geometrically rho-mixing.*

Assumption C.5 is a standard assumption in nonparametric estimation with a linear sieve. Assumptions C.6 and C.7 are standard weak dependence conditions. Lemma D.1 provides primitive sufficient conditions under which both of these assumptions are satisfied.

As in the body of the text, let $\zeta_0(K) = \sqrt{\|b^K(x)'b^K(x)\|_\infty}$ denote a measure of roughness of the sieve basis functions. The results for beta-mixing data use an exponential inequality for sums of weakly-dependent random matrix random matrices developed in Chen and Christensen (2013). The results for rho-mixing data follow arguments similar to Gobet, Hoffmann, and Reiß (2004) with the necessary modifications.

**Lemma C.1.** *Under Assumptions C.5 and C.6, if $m$ is bounded, then*

$$\|\widehat{\widetilde{\mathbf{M}}}_K - \widetilde{\mathbf{M}}_K\|_2 = O_p\left(\frac{\zeta_0(K)\log n}{\sqrt{n}}\right)$$

*provided $\zeta_0(K)\log n/\sqrt{n} = o(1)$.*

**Lemma C.2.** *Under Assumptions C.5 and C.7, if $E[m(X_0,X_1)^2] < \infty$, then*

$$\|\widehat{\widetilde{\mathbf{M}}}_K - \widetilde{\mathbf{M}}_K\|_2 = O_p\left(\frac{\zeta_0(K)^2}{\sqrt{n}}\right).$$

*If, in addition, $m$ is bounded, then*

$$\|\widehat{\widetilde{\mathbf{M}}}_K - \widetilde{\mathbf{M}}_K\|_2 = O_p\left(\frac{\zeta_0(K)\sqrt{K}}{\sqrt{n}}\right).$$

**Lemma C.3.** *Under Assumptions C.5 and C.7, if $E[m(X_0,X_1)^2] < \infty$ and $\{v_K : K \geq 1\}$ is a sequence of deterministic constants with $v_K \in \mathbb{R}^K$ and $\sup_K \|v_K\|_2 < \infty$, then*

$$\|(\widehat{\widetilde{\mathbf{M}}}_K - \widetilde{\mathbf{M}}_K)v_K\|_2 = O_p\left(\frac{\zeta_0(K)^2}{\sqrt{n}}\right).$$

*If, in addition, $m$ is bounded, then*

$$\|(\widehat{\widetilde{\mathbf{M}}}_K - \widetilde{\mathbf{M}}_K)v_K\|_2 = O_p\left(\frac{\zeta_0(K)}{\sqrt{n}}\right).$$

*Moreover, the same rates obtain for $\|(\widehat{\widetilde{\mathbf{M}}}'_K - \widetilde{\mathbf{M}}'_K)v_K\|_2$.*

Rates for the estimator of the Gram matrix are also available. These are proved using arguments similar to Lemmas C.1 and C.2 so their proofs are omitted.

**Lemma C.4.** *Under Assumptions C.5 and C.6,*

$$\|\widehat{\widetilde{\mathbf{G}}}_K - I_K\|_2 = O_p\left(\frac{\zeta_0(K)\log n}{\sqrt{n}}\right)$$

*provided $\zeta_0(K)\log n/\sqrt{n} = o(1)$.*

**Lemma C.5.** *Under Assumptions C.5 and C.7, if $\{v_K : K \geq 1\}$ is a sequence of deterministic constants with $v_K \in \mathbb{R}^K$ and $\sup_K \|v_K\|_2 < \infty$, then*

$$\|(\widehat{\widetilde{\mathbf{G}}}_K - I_K)v_K\|_2 = O_p\left(\frac{\zeta_0(K)}{\sqrt{n}}\right).$$

## Appendix D. Proofs

D.1. **Proofs for Section 5.3.** Weak-dependence properties of $\{X_t\}$ are first established under Assumption 4.1.

**Lemma D.1.** *Under Assumption 4.1(i)–(iii), $\{X_t\}$ is geometrically phi-mixing.*

*Proof of Lemma D.1.* Let $\mathbb{E}$ denote the conditional expectation operator associated with $\{X_t\}$, i.e. $\mathbb{E}\psi(x) = E[\psi(X_1)|X_0 = x]$. A sufficient condition for $\{X_t\}$ to be geometrically phi-mixing is

$$\sup_{\psi \in L^\infty(Q):\psi \neq 0, E[\psi(X)]=0} \frac{\|\mathbb{E}^\tau \psi\|_{L^\infty(Q)}}{\|\psi\|_{L^\infty(Q)}} \to 0$$

as $\tau \to \infty$ (Doukhan, 1994, pp. 88–89). The inequality

$$\sup_{\psi \in L^\infty(Q):\psi \neq 0, E[\psi(X)]=0} \frac{\|\mathbb{E}^\tau \psi\|_{L^\infty(Q)}}{\|\psi\|_{L^\infty(Q)}} \leq \sup_{\psi \in L^\infty(Q):\psi \neq 0} \frac{\|\mathbb{E}^\tau \psi - E[\psi(X)]\|_{L^\infty(Q)}}{\|\psi\|_{L^\infty(Q)}} \tag{49}$$

is immediate. It is therefore sufficient to establish that the right-hand side of (49) is $O(e^{-c\tau})$ for some $c > 0$.

Theorem A.4 will be applied to $\mathbb{E}$ (by setting $m(x_0, x_1) \equiv 1$). It is enough to show that $\mathbb{E}$ is compact. First observe that $\mathbb{E} : L^\infty(Q) \to L^\infty(Q)$ may be continuously extended to have domain $L^1(Q)$. For $\psi \in L^1(Q)$,

$$\begin{aligned}
\|\mathbb{E}\psi\|_{L^\infty(Q)} &\leq \sup_{x_0,x_1} \left| \frac{f(x_0,x_1)}{q(x_0)q(x_1)} \right| \int_{\mathcal{X}} |\psi(x_1)| \, \mathrm{d}Q(x_1) \\
&\leq C\|\psi\|_{L^1(Q)}
\end{aligned}$$

for some finite positive constant $C$, under Assumption 4.1(i)–(iii). Therefore $\mathbb{E} : L^1(Q) \to L^\infty(Q)$ is continuous, and so $\mathbb{E} : L^\infty(Q) \to L^\infty(Q)$ is compact (Schaefer, 1974, p. 337).

Let $f(x) = 1$ for all $x \in \mathcal{X}$. The function $f \in L^\infty(Q)$ and is an eigenfunction of $\mathbb{E}$ with eigenvalue 1 because $\mathbb{E}f = f$. The functional $e^* : L^\infty(Q) \to \mathbb{R}$ defined by $e^*(\psi) = E[\psi(X)] = \int_{\mathcal{X}} \psi(x) \, \mathrm{d}Q(x)$ is clearly bounded and linear. Let $x^* : L^\infty(Q) \to \mathbb{R}$ be a bounded linear functional. The adjoint $\mathbb{E}^*$ of $\mathbb{E}$ is defined by

$$(\mathbb{E}^* x^*)(\psi) = x^*(\mathbb{E}\psi)$$

for all $\psi \in L^\infty(Q)$ and all $x^*$ in the dual space of $L^\infty(Q)$. By iterated expectations

$$(\mathbb{E}^* e^*)(\psi) = e^*(\mathbb{E}\psi) = \int_{\mathcal{X}} \mathbb{E}\psi(x) \, \mathrm{d}Q(x) = E[\psi(X)] = e^*(\psi)$$

for all $\psi \in L^\infty(Q)$. Therefore $e^*$ is an eigenfunction of $\mathbb{E}^*$ with eigenvalue 1. Define $P : L^\infty(Q) \to L^\infty(Q)$ by $P\psi(x) = E[\psi(X)]$ for all $x \in \mathcal{X}$. Theorem A.4 applied to $\mathbb{E} : L^\infty(Q) \to L^\infty(Q)$ yields the desired result. $\square$

**Remark D.1.** *Phi-mixing implies other notions of weak dependence. Let $\alpha_\tau$, $\varrho_\tau$, $\beta_\tau$ and $\varphi_\tau$ denote the alpha-, rho-, beta-, and phi-mixing coefficients of $\{X_t\}$. Lemma D.1 and the relations*

$$\begin{aligned}
2\alpha_\tau &\leq \beta_\tau \leq \varphi_\tau \\
4\alpha_\tau &\leq \rho_\tau \leq 2\sqrt{\varphi_\tau}
\end{aligned}$$

*imply that $\{X_t\}$ is geometrically alpha-, rho- and beta-mixing under Assumption 4.1(i)–(iii).*

**Lemma D.2.** *Under Assumption 4.1, there exists a $\delta > 0$ such that for any $f : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ and $g : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ with $E[f(X_0, X_1)] = E[g(X_0, X_1)] = 0$, $E[f(X_0, X_1)^2] < \infty$ and $E[g(X_0, X_1)^2] < \infty$, and any $t \geq 1$,*

$$|E[f(X_0, X_1)g(X_t, X_{t+1})]| \leq e^{-\delta(t-1)} E[f(X_0, X_1)^2]^{1/2} E[g(X_0, X_1)^2]^{1/2} .$$

*Proof of Lemma D.2.* $\{X_t\}$ is geometrically rho-mixing under Assumption 4.1 (see Remark D.1), so there exists a $\delta > 0$ such that $|\mathrm{Cov}(f_1(X_t), g_1(X_{t+\tau}))| \leq \exp^{-\delta\tau} E[f_1(X_0)^2]^{1/2} E[g_1(X_0)^2]^{1/2}$ for each $\tau \geq 1$ and each $f_1 : \mathcal{X} \to \mathbb{R}$ and $g_1 : \mathcal{X} \to \mathbb{R}$ with $E[f_1(X_0)^2] < \infty$ and $E[g_1(X_0)^2] < \infty$ by the covariance inequality for rho-mixing random variables (Doukhan, 1994, p. 9)

By the Markov property,

$$\begin{aligned} E[f(X_0, X_1)g(X_t, X_{t+1})] &= E[f(X_0, X_1)E[g(X_t, X_{t+1})|\mathcal{F}_t]] \\ &= E[g(X_0, X_1)E[g(X_t, X_{t+1})|X_t]] \\ &= E[E[f(X_0, X_1)|\mathcal{G}_1]E[g(X_t, X_{t+1})|X_t]] \\ &= E[E[f(X_0, X_1)|X_1]E[g(X_t, X_{t+1})|X_t]] \end{aligned}$$

Therefore, by the covariance inequality,

$$|E[f(X_0, X_1)g(X_t, X_{t+1})]| = e^{-\delta(t-1)} E[E[f(X_0, X_1)|X_1]^2]^{1/2} E[E[g(X_0, X_1)|X_0]^2]^{1/2}$$

and the result follows by Jensen's inequality. □

*Proof of Theorem 5.1.* First verify the conditions of Theorems C.1 and C.2 for $\mathbb{M}$.

Assumption C.1 is satisfied for $\mathbb{M}$ under Assumption 4.1 by Theorem 4.1.

Assumption C.2(i) and the part of Assumption C.2(ii) pertaining to $\phi$ is trivially satisfied by Assumptions 5.1(i). The remaining condition in Assumption C.2(ii) is satisfied by Assumption 5.1(iv) because

$$\|(\mathbb{M}^* \Pi_K^b - \mathbb{M}^*)\phi^*\| \leq \|\mathbb{M}^*\| \|\Pi_K^b \phi^* - \phi^*\| = \|\mathbb{M}\| \|\Pi_K^b \phi^* - \Pi_K^b h_K^* + h_K^* - \phi^*\| \leq 2\|\mathbb{M}\| \|\phi^* - h_K^*\|$$

which is $O(\delta_K^*)$.

Let $\| \cdot \|_{HS}$ denote the Hilbert-Schmidt norm and recall $\|\mathbb{M}\|_{HS} < \infty$ under Assumption 4.1. The bound

$$\|R(z, \mathbb{M})\| \leq \frac{1}{d(z, \sigma(\mathbb{M}))} \exp\left(\frac{1}{2} + 2\frac{\|\mathbb{M}\|_{HS}^2}{d(z, \sigma(\mathbb{M}))^2}\right)$$

obtains for any $z \in \mathbb{C} \setminus \sigma(\mathbb{M})$ (see, e.g. Bandtlow (2004)). Let $\{e_k : k \geq 1\}$ be an orthonormal basis for $L^2(Q)$ such that $\{e_k : 1 \leq k \leq K\}$ are an orthonormal basis for $B_K$. As Hilbert-Schmidt norms are invariant to the choice of basis,

$$\|\Pi_K^b \mathbb{M}|_{B_K}\|_{HS}^2 = \sum_{k=1}^{K} \|\Pi_K^b \mathbb{M} e_k\|^2 \leq \sum_{k=1}^{\infty} \|\Pi_K^b \mathbb{M} e_k\|^2 \leq \sum_{k=1}^{\infty} \|\mathbb{M} e_k\|^2 = \|\mathbb{M}\|_{HS}^2 .$$

Therefore, $\Pi_K^b \mathbb{M}|_{B_K}$ is Hilbert-Schmidt and the bound

$$\|R(z, \Pi_K^b \mathbb{M}|_{B_K})\| \leq \frac{1}{d(z, \sigma(\Pi_K^b \mathbb{M}|_{B_K}))} \exp\left(\frac{1}{2} + 2\frac{\|\mathbb{M}\|_{HS}^2}{d(z, \sigma(\Pi_K^b \mathbb{M}|_{B_K}))^2}\right)$$

obtains for any $z \in \mathbb{C} \setminus \sigma(\Pi_K^b \mathbb{M}|_{B_K})$. The function $r : (0, \infty) \to (0, \infty)$ given by

$$r(x) = \frac{1}{x} \exp\left(\frac{1}{2} + 2\frac{\|\mathbb{M}\|_{HS}}{x^2}\right)$$

is continuous and strictly decreasing, verifying Assumption C.3.

Assumption C.4(i) is trivially satisfied by Assumption 5.1(iii). Assumption C.4(ii) and (iii) are satisfied by Assumption 5.1(ii) and definition of $\bar{\eta}_{n,K}$ and $\eta_{n,K}$.

Parts (i) and (ii) are straightforward applications of Theorems C.1 and C.2. For part (iii) it is enough to show that $\|\phi_K^*/\|\phi_K^*\| - \phi_K^\star/\|\phi_K^\star\|\| = O(\delta_K^*)$, which follows from Assumptions 5.1(iv). $\square$

*Proof of Corollary 5.1.* The rate of convergence of $\widehat{y}$ follows immediately from Theorem 5.1 by continuity of log on a neighborhood of $\rho$. For $\widehat{L}$, first write

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

It is enough to show that the second term on the right-hand side is $O_p(n^{-1/2})$. This follows by Chebychev's inequality, using the condition $E[\log m(X_0, X_1)^2] < \infty$ and Lemma D.2. $\square$

*Proof of Corollary 5.2.* For any $f_K \in B_K$, the sup and $L^2(Q)$ norms are related by

$$\|f_K\|_\infty^2 \leq \underline{\lambda}^{-1} \zeta_0(K)^2 \|f_K\|^2 .$$

By Assumption 5.2(i) and the triangle inequality

$$\begin{aligned}
\|\phi - \widehat{\phi}\|_\infty &\leq \|\phi - g_K\|_\infty + \|g_K - \widehat{\phi}\|_\infty \\
&\leq O(\delta_K) + \zeta_0(K)\underline{\lambda}^{-1/2}\|g_K - \widehat{\phi}\| \\
&\leq O(\delta_K) + \zeta_0(K)\underline{\lambda}^{-1/2}(\|g_K - \phi\| + \|\phi - \widehat{\phi}\|) \\
&= O_p(\zeta_0(K)(\delta_K + \eta_{n,K}))
\end{aligned}$$

where the final line is by Theorem 5.1 and the fact that the sup norm dominates the $L^2(Q)$ norm. This proves part (i); the proof of part (ii) is similar. $\square$

D.2. **Proofs for Section 5.4.** Several lemmas are first required before proving Theorem 5.2. Define the remainder term

$$\begin{aligned}
\tau_{n,K} &= \frac{1}{n}\sum_{t=0}^{n-1} \xi_{K,t} - \xi_t \\
\xi_{K,t} &= \phi_K^\star(X_t)\phi_K(X_{t+1})m(X_t, X_{t+1}) - \rho_K \phi_K^\star(X_t)\phi_K(X_t) \\
\xi_t &= \phi^*(X_t)\phi(X_{t+1})m(X_t, X_{t+1}) - \rho\phi^*(X_t)\phi(X_t) .
\end{aligned} \tag{50}$$

**Lemma D.3.** *Under Assumptions 4.1, 5.1, 5.2, and 5.3(i)(ii),* $\tau_{n,K} = O_p(\zeta_0(K)(\delta_K^* + \delta_K)/\sqrt{n})$.

*Proof of Lemma D.3.* First write

$$\tau_{n,K} = \frac{1}{\sqrt{n}} S_{n,K}$$

where $S_{n,K} = \sqrt{n}\tau_{n,K}$. Note that the summands in $S_{n,K}$ have mean zero and finite second moment. By Lemma D.2 and the inequality $(a+b)^2 \leq 2a^2 + 2b^2$, there exists a finite positive constant $C$ such that

$$\begin{aligned}
E[S_{n,K}^2] &\leq CE[(\xi_{K,0} - \xi_0)^2] \\
&\leq 2CE\left[(\phi_K^\star(X_0)\phi_K(X_1) - \phi^*(X_0)\phi(X_1))^2 m(X_0,X_1)^2\right] \\
&\quad +2CE[(\rho_K\phi_K^\star(X_0)\phi_K(X_0) - \rho\phi^*(X_0)\phi(X_0))^2] \\
&\leq 4C\big(E[\{(\phi_K^\star(X_0) - \phi^*(X_0))^2\phi_K(X_1)^2 + (\phi_K(X_1) - \phi(X_1))^2\phi^*(X_0)^2\}m(X_0,X_1)^2] \\
&\quad +2\rho_K^2 E[(\phi_K^\star(X_0) - \phi^*(X_0))^2\phi_K(X_1)^2] + 2\rho_K^2 E[\phi^*(X_0)^2(\phi_K(X_1) - \phi(X_1))^2] \\
&\quad +2(\rho_K - \rho)^2 E[\phi^*(X_0)^2\phi(X_1)^2]\big)\,.
\end{aligned} \tag{51}$$

Assumptions 4.1 and 5.1 are sufficient to apply Theorem C.1 to $\mathbb{M}$ (see the proof of Theorem 5.1). This yields $\rho_K - \rho = O(\delta_K)$, $\|\phi_K - \phi\| = O(\delta_K)$ and $\|\phi_K^* - \phi^*\| = O(\delta_K)$. Under Assumption 5.2(ii),

$$\|\phi_K^\star - \phi_K^*\| = \|\Pi_K^b(\phi_K^* - g_K^*) + g_K^* - \phi_K^*\| \leq 2\|\phi_K^* - g_K^*\| \leq 2\|\phi_K^* - \phi^*\| + 2\|\phi^* - g_K^*\|$$

which is $O(\delta_K + \delta_K^*)$ by Assumption 5.2(ii) and Theorem C.1. It follows by Assumption 5.2 using similar arguments to the proof of Corollary 5.2 that $\|\phi_K^\star - \phi^*\|_\infty = O(\zeta_0(K)(\delta_K^* + \delta_K))$ and $\|\phi_K - \phi\|_\infty = O(\zeta_0(K)\delta_K)$. Plugging these rates into (51) and using the Hölder inequality yields

$$E[S_{n,K}^2] = O(\zeta_0(K)^2(\delta_K^* + \delta_K)^2)\,.$$

The result follows by Chebychev's inequality. □

**Lemma D.4.** *Under Assumptions 4.1, 5.1, 5.2, and 5.3(i)(ii),*

$$\sqrt{n}(\widehat{\rho} - \rho) = \frac{1}{\sqrt{n}} \sum_{t=0}^{n-1} \{\phi^*(X_t)m(X_t,X_{t+1})\phi(X_{t+1}) - \rho\phi^*(X_t)\phi(X_t)\} + o_p(1)\,.$$

*Proof of Lemma D.4.* Assumptions 4.1 and 5.1 are sufficient for the assumptions of Theorem C.3. Application of Theorem C.3 yields

$$\widehat{\rho} - \rho_K = (\mathbf{G}_K^{1/2}c_K^\star)'(\widehat{\widetilde{\mathbf{G}}}_K^{-1}\widehat{\widetilde{\mathbf{M}}}_K - \widetilde{\mathbf{M}}_K)(\mathbf{G}_K^{1/2}c_K) + O_p(\bar{\eta}_{n,K}^2)\,. \tag{52}$$

First, using the fact that $\widehat{\widetilde{\mathbf{G}}}_K^{-1} = I_K - \widehat{\widetilde{\mathbf{G}}}_K^{-1}(\widehat{\widetilde{\mathbf{G}}}_K - I_K)$ gives

$$\begin{aligned}
&\widehat{\widetilde{\mathbf{G}}}_K^{-1}\widehat{\widetilde{\mathbf{M}}}_K - \widetilde{\mathbf{M}}_K \\
={}& \widehat{\widetilde{\mathbf{M}}}_K - \widetilde{\mathbf{M}}_K - \widehat{\widetilde{\mathbf{G}}}_K^{-1}(\widehat{\widetilde{\mathbf{G}}}_K - I_K)\widetilde{\mathbf{M}}_K - \widehat{\widetilde{\mathbf{G}}}_K^{-1}(\widehat{\widetilde{\mathbf{G}}}_K - I_K)(\widehat{\widetilde{\mathbf{M}}}_K - \widetilde{\mathbf{M}}_K) \\
={}& \widehat{\widetilde{\mathbf{M}}}_K - \widetilde{\mathbf{M}}_K - (I_K - \widehat{\widetilde{\mathbf{G}}}_K^{-1}(\widehat{\widetilde{\mathbf{G}}}_K - I_K))(\widehat{\widetilde{\mathbf{G}}}_K - I_K)\widetilde{\mathbf{M}}_K - \widehat{\widetilde{\mathbf{G}}}_K^{-1}(\widehat{\widetilde{\mathbf{G}}}_K - I_K)(\widehat{\widetilde{\mathbf{M}}}_K - \widetilde{\mathbf{M}}_K) \\
={}& \widehat{\widetilde{\mathbf{M}}}_K - \widehat{\widetilde{\mathbf{G}}}_K\widetilde{\mathbf{M}}_K + \widehat{\widetilde{\mathbf{G}}}_K^{-1}(\widehat{\widetilde{\mathbf{G}}}_K - I_K)^2\widetilde{\mathbf{M}}_K - \widehat{\widetilde{\mathbf{G}}}_K^{-1}(\widehat{\widetilde{\mathbf{G}}}_K - I_K)(\widehat{\widetilde{\mathbf{M}}}_K - \widetilde{\mathbf{M}}_K)\,.
\end{aligned}$$

The leading term in (52) is then

$$\begin{aligned}(\mathbf{G}_K^{1/2}c_K^\star)'(\widehat{\widetilde{\mathbf{M}}}_K - \widehat{\widetilde{\mathbf{G}}}_K\widetilde{\mathbf{M}}_K)(\mathbf{G}_K^{1/2}c_K) &= c_K^{\star\prime}\widehat{\mathbf{M}}_K c_K - c_K^{\star\prime}\widehat{\mathbf{G}}_K\mathbf{G}_K^{-1}\mathbf{M}_K c_K \\ &= c_K^{\star\prime}\widehat{\mathbf{M}}_K c_K - \rho_K c_K^{\star\prime}\widehat{\mathbf{G}}_K c_K\end{aligned}$$

where the second line is by equation (23). It remains to show that the remaining part of expression (52) is $O_p(\bar{\eta}_{n,K}^2)$. By the Cauchy-Schwarz inequality,

$$\begin{aligned}&|(\mathbf{G}_K^{1/2}c_K^\star)'(\widehat{\widetilde{\mathbf{G}}}_K^{-1}(\widehat{\widetilde{\mathbf{G}}}_K - I_K)^2\widetilde{\mathbf{M}}_K - \widehat{\widetilde{\mathbf{G}}}_K^{-1}(\widehat{\widetilde{\mathbf{G}}}_K - I_K)(\widehat{\widetilde{\mathbf{M}}}_K - \widetilde{\mathbf{M}}_K))(\mathbf{G}_K^{1/2}c_K)| \\ \leq\ & \|\phi_K^\star\|\|\phi_K\|\left\|\widehat{\widetilde{\mathbf{G}}}_K^{-1}(\widehat{\widetilde{\mathbf{G}}}_K - I_K)^2\widetilde{\mathbf{M}}_K - \widehat{\widetilde{\mathbf{G}}}_K^{-1}(\widehat{\widetilde{\mathbf{G}}}_K - I_K)(\widehat{\widetilde{\mathbf{M}}}_K - \widetilde{\mathbf{M}}_K)\right\|_2 \\ =\ & (\|\phi^*\| + o(1))(\|\phi\| + o(1)) \times O_p(\bar{\eta}_{n,K}^2)\end{aligned}$$

where the final line is because $\|\phi_K - \phi\| = o(1)$ and $\|\phi_\star^K - \phi_K^*\| = o(1)$ (see the proof of Lemma D.3), and the $O_p(\bar{\eta}_{n,K}^2)$ term follows by the same arguments as the proof of Lemma D.8. The expansion (52) may therefore be reexpressed as

$$\widehat{\rho} - \rho_K = \frac{1}{n}\sum_{t=0}^{n-1}\{\phi_K^\star(X_t)m(X_t,X_{t+1})\phi_K(X_{t+1}) - \rho_K\phi_K^\star(X_t)\phi_K(X_t)\} + O_p(\bar{\eta}_{n,K}^2)\,.$$

Rearranging yields

$$\widehat{\rho} - \rho = \rho_K - \rho + \frac{1}{n}\sum_{t=0}^{n-1}\{\phi^*(X_t)m(X_t,X_{t+1})\phi(X_{t+1}) - \rho\phi^*(X_t)\phi(X_t)\} + \tau_{n,K} + O_p(\bar{\eta}_{n,K}^2)\,.$$

where $\tau_{n,K}$ is defined in expression (50). Lemma D.3 and Assumption 5.3(ii) together imply that $\sqrt{n}(\tau_{n,K} + O_p(\bar{\eta}_{n,K}^2)) = o_p(1)$. Finally, $|\rho_K - \rho| = O(\delta_K) = o(n^{-1/2})$ by the proof of Lemma D.3 and the condition $\delta_K = o(n^{-1/2})$. □

*Proof of Theorem 5.2.* Part (i): Lemma D.4 shows that the representation (32) is valid. Assumption 5.3(i) and the Hölder inequality imply $V_\rho$ is finite. The central limit theorem for martingales with stationary and ergodic differences (Billingsley, 1961) then yields

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

whenever $V_\rho > 0$.

Part(ii): The asymptotic distribution for $\widehat{y}$ then follows by the delta method.

Part (iii): Let $\bar{\ell}_n = n^{-1}\sum_{t=1}^n \ell(X_t, X_{t+1})$ where

$$\ell(X_t,X_{t+1}) = \log m(X_t,X_{t+1}) - E[\log m(X_t,X_{t+1})]\,.$$

By definition of $\widehat{L}$ and the asymptotic linear expansion for $\widehat{\rho}$,

$$\begin{aligned}\sqrt{n}(\widehat{L} - L) &= \sqrt{n}(\log\widehat{\rho} - \log\rho - \bar{\ell}_n) \\ &= \frac{1}{\sqrt{n}}\sum_{t=0}^{n-1}\left\{\rho^{-1}\phi^*(X_t)m(X_t,X_{t+1})\phi(X_{t+1}) - \phi^*(X_t)\phi(X_t) - \ell(X_t,X_{t+1})\right\} + o_p(1)\,.\end{aligned}$$

The summands are strictly stationary geometrically phi-mixing random variables by Assumption 4.1 and Lemma D.1, have mean zero, and have and finite second moment by Assumptions 5.3. Application of Lemma D.2 provides that

$$V_L = \lim_{n\to\infty} \frac{1}{n} E\left[\left(\sum_{t=0}^{n-1} \left\{\rho^{-1}\phi^*(X_t)m(X_t,X_{t+1})\phi(X_{t+1}) - \phi^*(X_t)\phi(X_t) - \ell(X_t,X_{t+1})\right\}\right)^2\right] < \infty\,.$$

The result follows by a CLT for strictly stationary phi-mixing sequences (Peligrad, 1985, Corollary 2.2). $\square$

*Proof of Corollary 5.3.* Part (i): Write $\|\widehat{\phi}^f - \phi\| \leq \|\widehat{\phi}^f - \widehat{\phi}\| + \|\widehat{\phi} - \phi\|$ where $\|\widehat{\phi}\| = 1$. Theorem 5.1 gives $\|\widehat{\phi} - \phi\| = O_p(\delta_K + \eta_{n,K})$ so it remains to control $\|\widehat{\phi}^f - \widehat{\phi}\|$. Note that $\widehat{\phi} = \widehat{c}'b^K$ where $\widehat{c}'\mathbf{G}_K\widehat{c} = 1$ and $\widehat{\phi}^f = \widehat{c}^{f\prime}b^K$ where $\widehat{c}^f = \widehat{c}/(\widehat{c}'\widehat{\mathbf{G}}_K\widehat{c})^{1/2}$. Therefore,

$$\begin{aligned}\|\widehat{\phi}^f - \widehat{\phi}\| &= \left|\frac{1}{(\widehat{c}'\widehat{\mathbf{G}}_K\widehat{c})^{1/2}} - 1\right| \\ &= \left|\frac{1}{(\widehat{c}'\mathbf{G}_K^{1/2}\widehat{\widehat{\mathbf{G}}}_K\mathbf{G}_K^{1/2}\widehat{c})^{1/2}} - 1\right|\,.\end{aligned}$$

The minimax characterization of eigenvalues of symmetric matrices (Kato, 1980, Section I.6.10) implies that

$$\lambda_{\min}(\widehat{\widehat{\mathbf{G}}}_K) \leq (\widehat{c}'\mathbf{G}_K^{1/2}\widehat{\widehat{\mathbf{G}}}_K\mathbf{G}_K^{1/2}\widehat{c})^{1/2} \leq \lambda_{\max}(\widehat{\widehat{\mathbf{G}}}_K)\,.$$

Moreover,

$$\begin{aligned}\max\{|\lambda_{\max}(\widehat{\widehat{\mathbf{G}}}_K) - 1|, |\lambda_{\min}(\widehat{\widehat{\mathbf{G}}}_K) - 1|\} &= \max\{|\lambda_{\max}(\widehat{\widehat{\mathbf{G}}}_K - I_K)|, |\lambda_{\min}(\widehat{\widehat{\mathbf{G}}}_K - I_K)|\} \\ &= \|\widehat{\widehat{\mathbf{G}}}_K - I_K\|_2 \\ &= O_p(\bar{\eta}_{n,K})\end{aligned}$$

by definition of $\bar{\eta}_{n,K}$. This proves $\|\widehat{\phi}^f - \widehat{\phi}\| = O_p(\bar{\eta}_{n,K})$.

Part (iii): By the relation between the $L^2(Q)$ norm and sup norm on $B_K$ and Assumption 5.2(i),

$$\begin{aligned}\|\widehat{\phi}^f - \phi\|_\infty &\leq \underline{\lambda}^{-1/2}\zeta_0(K)\|\widehat{\phi}^f - \widehat{\phi}\| + \|\widehat{\phi} - \phi\|_\infty \\ &\leq \underline{\lambda}^{-1/2}\zeta_0(K)\|\widehat{\phi}^f - \widehat{\phi}\| + \|\widehat{\phi} - g_K\|_\infty + \|g_K - \phi\|_\infty \\ &\leq \underline{\lambda}^{-1/2}\zeta_0(K)\|\widehat{\phi}^f - \widehat{\phi}\| + \underline{\lambda}^{-1/2}\zeta_0(K)\|\widehat{\phi} - g_K\| + \|g_K - \phi\|_\infty \\ &\leq \underline{\lambda}^{-1/2}\zeta_0(K)\|\widehat{\phi}^f - \widehat{\phi}\| + \underline{\lambda}^{-1/2}\zeta_0(K)\|\widehat{\phi} - \phi\| + \underline{\lambda}^{-1/2}\zeta_0(K)\|\phi - g_K\| + \|g_K - \phi\|_\infty\,.\end{aligned}$$

The result follows by Part (i), Assumption 5.2(i), and Theorem 5.1.

Part (ii): Write $\|\widehat{\phi}^{\star f} - \phi^*\| \leq \|\widehat{\phi}^{\star f} - \widehat{\phi}^\star\| + \|\widehat{\phi}^\star - \phi^*\|$ where $E[\widehat{\phi}(X)\widehat{\phi}^\star(X)] = 1$. Theorems C.1 and C.2 show that $\|\widehat{\phi}^\star - \phi^*\| = O_p(\delta_K + \bar{\eta}_{n,K})$. It remains to control $\|\widehat{\phi}^{\star f} - \widehat{\phi}^\star\| = O_p(\bar{\eta}_{n,K})$. Note that $\widehat{\phi}^\star = \widehat{c}^{\star\prime}b^K$ where $\widehat{c}^{\star\prime}\mathbf{G}_K\widehat{c} = 1$ and $\widehat{\phi}^f = \widehat{c}^{\star f\prime}b^K$ where $\widehat{c}^{\star f} = \widehat{c}^\star(\widehat{c}'\widehat{\mathbf{G}}_K\widehat{c})^{1/2}/(\widehat{c}^{\star\prime}\widehat{\mathbf{G}}_K\widehat{c})$. Therefore,

$$\|\widehat{\phi}^{\star f} - \widehat{\phi}^\star\| = \left|\frac{(\widehat{c}'\widehat{\mathbf{G}}_K\widehat{c})^{1/2}}{\widehat{c}^{\star\prime}\widehat{\mathbf{G}}_K\widehat{c}} - 1\right| \|\widehat{\phi}^\star\|$$

where $\|\widehat{\phi}^\star\| = O_p(1)$ by Theorem 5.1(ii), and the proof of Part (i) shows $(\widehat{c}'\widehat{\mathbf{G}}_K\widehat{c})^{1/2} = 1 + O_p(\bar{\eta}_{n,K})$. Moreover,

$$\begin{aligned} \widehat{c}^{\star\prime}\widehat{\mathbf{G}}_K\widehat{c} &= \widehat{c}^{\star\prime}\mathbf{G}_K\widehat{c} + \widehat{c}^{\star\prime}(\widehat{\mathbf{G}}_K - \mathbf{G}_K)\widehat{c} \\ &= 1 + \widehat{c}^{\star\prime}\mathbf{G}_K^{1/2}(\widehat{\widetilde{\mathbf{G}}}_K - I_K)\mathbf{G}_K^{1/2}\widehat{c} \\ &\le 1 + \|\widehat{\phi}^\star\|\|\widehat{\widetilde{\mathbf{G}}}_K - I_K\|_2 \\ &= 1 + O_p(1) \times O_p(\bar{\eta}_{n,K}) \end{aligned}$$

by Theorem 5.1(iii), definition of $\bar{\eta}_{n,K}$, and the normalization $\|\widehat{\phi}\| = 1$. Thus $\|\widehat{\phi}^{\star f} - \widehat{\phi}^\star\| = O_p(\bar{\eta}_{n,K})$.

Part (iv): Arguing as in the proof of Part (iii) yields

$$\|\widehat{\phi}^{\star f} - \phi^*\|_\infty \le \underline{\lambda}^{-1/2}\zeta_0(K)\|\widehat{\phi}^{f\star} - \widehat{\phi}^\star\| + \underline{\lambda}^{-1/2}\zeta_0(K)\|\widehat{\phi}^\star - \phi^*\| + \underline{\lambda}^{-1/2}\zeta_0(K)\|\phi^* - g_K^*\| + \|g_K^* - \phi^*\|_\infty .$$

The result follows by Part (ii), Assumption 5.2(ii), and $\|\widehat{\phi}^\star - \phi^*\| = O_p(\delta_K + \bar{\eta}_{n,K})$. □

*Proof of Theorem 5.3.* Part (i): By addition and subtraction of terms,

$$\widehat{V}_\rho - V_\rho = \frac{1}{n}\sum_{t=0}^{n-1}\left(\widehat{\phi}_t^{\star f2} m_{t,t+1}^2 \widehat{\phi}_{t+1}^{f2} - \phi_t^{*2} m_{t,t+1}^2 \phi_{t+1}^2\right)$$

(53) $$+\frac{1}{n}\sum_{t=0}^{n-1}\phi_t^{*2} m_{t,t+1}^2 \phi_{t+1}^2 - E[\phi^*(X_0)^2 m(X_0,X_1)^2 \phi(X_1)^2]$$

(54) $$+\frac{1}{n}\sum_{t=0}^{n-1}\left(\widehat{\rho}^2\widehat{\phi}_t^{\star f2}\widehat{\phi}_t^{f2} - \rho^2\phi_t^{*2}\phi_t^2\right)$$

(55) $$+\frac{1}{n}\sum_{t=0}^{n-1}\rho^2\phi_t^{*2}\phi_t^2 - \rho^2 E[\phi^*(X_0)\phi(X_0)]$$

(56) $$-\frac{2}{n}\sum_{t=0}^{n-1}\left(\widehat{\rho}\widehat{\phi}_t^{\star f2} m_{t,t+1}\widehat{\phi}_t^f\widehat{\phi}_{t+1}^f - \rho\phi_t^{*2} m_{t,t+1}\phi_t\phi_{t+1}\right)$$

(57) $$-\frac{2}{n}\sum_{t=0}^{n-1}\rho\phi_t^{*2} m_{t,t+1}\phi_t\phi_{t+1} + 2\rho E[\phi^*(X_0)^2 m(X_0,X_1)\phi(X_0)\phi(X_1)]$$

Terms (53), (55) and (57) are all $o_{a.s}(1)$ by the ergodic theorem (the expectations exist by Assumption 5.3(i)). For term (53),

$$\begin{aligned} (53) &= \frac{1}{n}\sum_{t=0}^{n-1}\phi_t^{*2} m_{t,t+1}^2[\widehat{\phi}_{t+1}^{f2} - \phi_{t+1}^2] + \frac{1}{n}\sum_{t=0}^{n-1}[\widehat{\phi}_t^{\star f2} - \phi_t^{*2}] m_{t,t+1}^2[\widehat{\phi}_{t+1}^{f2} - \phi_{t+1}^2] \\ &\quad + \frac{1}{n}\sum_{t=0}^{n-1}[\widehat{\phi}_t^{\star f2} - \phi_t^{*2}] m_{t,t+1}^2 \phi_{t+1}^2 . \end{aligned}$$

Using the relation $(a^2-b^2)=(a+b)(a-b)$, the triangle inequality, and the sup-norm convergence rates established in Corollary 5.3,

$$\begin{aligned}|(53)| &\leq O_p(\zeta_0(K)(\delta^*_K+\delta_K+\bar\eta_{n,K}))\Big(\frac{1}{n}\sum_{t=0}^{n-1}\phi_t^{*2}m_{t,t+1}^2|\widehat{\phi}^f_{t+1}+\phi_{t+1}| \\ &\quad +\frac{1}{n}\sum_{t=0}^{n-1}|\widehat{\phi}^{\star f}_t+\phi^*_t|m_{t,t+1}^2|\widehat{\phi}^f_{t+1}+\phi_{t+1}|+\frac{1}{n}\sum_{t=0}^{n-1}|\widehat{\phi}^{\star f}_t+\phi^*_t|m_{t,t+1}^2\phi_{t+1}^2\Big).\end{aligned}$$

Writing

$$|\widehat{\phi}^f_{t+1}+\phi_{t+1}|\leq 2\phi_{t+1}+\|\widehat{\phi}^f-\phi\|_\infty$$

and similarly for $\widehat{\phi}^{\star f}$, the condition $\zeta_0(K)(\delta^*_K+\delta_K+\bar\eta_{n,K})=o(1)$ (by Assumption 5.3(ii)) and sup-norm convergence rates in Corollary 5.2 yield

$$\begin{aligned}|(53)| &\leq o_p(1)\Big(2\frac{1}{n}\sum_{t=0}^{n-1}\phi_t^{*2}m_{t,t+1}^2\phi_{t+1}+o_p(1)\frac{1}{n}\sum_{t=0}^{n-1}\phi_t^{*2}m_{t,t+1}^2+4\frac{1}{n}\sum_{t=0}^{n-1}\phi_t^{*}m_{t,t+1}^2\phi_{t+1} \\ &\quad +o_p(1)\frac{1}{n}\sum_{t=0}^{n-1}m_{t,t+1}^2+o_p(1)\frac{1}{n}\sum_{t=0}^{n-1}\phi^*_t m_{t,t+1}^2+o_p(1)\frac{1}{n}\sum_{t=0}^{n-1}m_{t,t+1}^2\phi_{t+1} \\ &\quad +2\frac{1}{n}\sum_{t=0}^{n-1}\phi^*_t m_{t,t+1}^2\phi_{t+1}^2+o_p(1)\frac{1}{n}\sum_{t=0}^{n-1}m_{t,t+1}^2\phi_{t+1}^2\Big).\end{aligned}$$

All sample averages in this display are of the form

$$\frac{1}{n}\sum_{t=0}^{n-1}\phi^*(X_t)^k m(X_t,X_{t+1})^2\phi(X_{t+1})^l$$

with $0\leq k,l\leq 2$, and are therefore all $O_{a.s.}(1)$ by the ergodic theorem (all moments exist by Assumption 5.3(i)). Therefore term (53) is $o_p(1)$. Similar arguments show that terms (54) and (56) are both $o_p(1)$.

Part (ii): Immediate from Part (i) and consistency of $\widehat{\rho}$.

Parts (iii) and (iv): For each $j=1,\ldots,J$, write

$$\begin{aligned}\widehat{\Lambda}_j &= \frac{1}{\sqrt{n}}\sum_{t=0}^{n-1}h_j\left(\frac{t+1}{n}\right)\Delta_{t,t+1}+\frac{1}{\sqrt{n}}\sum_{t=0}^{n-1}h_j\left(\frac{t+1}{n}\right)(\widehat{\Delta}_{t,t+1}-\Delta_{t,t+1}) \\ &=: \Lambda_j+\frac{1}{\sqrt{n}}\sum_{t=0}^{n-1}h_j\left(\frac{t+1}{n}\right)(\widehat{\Delta}_{t,t+1}-\Delta_{t,t+1})\end{aligned}$$

where

$$\Delta_{t,t+1}=\rho^{-1}\left(\phi^{\star f}_t m_{t,t+1}\phi^f_{t+1}-\rho\phi^{\star f}_t\phi^f_t\right)-(\log m_{t,t+1}-E[\log m(X_0,X_1)])$$

Writing out term-by-term gives

$$\widehat{\Lambda}_j - \Lambda_j = \left(\frac{1}{n}\sum_{t=0}^{n-1} h_j\left(\frac{t+1}{n}\right)\right)\sqrt{n}(\overline{\log m}_n - E[\log m(X_0,X_1)]) \tag{58}$$

$$+\frac{1}{\sqrt{n}}\sum_{t=0}^{n-1} h_j\left(\frac{t+1}{n}\right)(\phi_t^*\phi_t - \widehat{\phi}_t^\star\widehat{\phi}_t) \tag{59}$$

$$+\widehat{\rho}^{-1}\frac{1}{\sqrt{n}}\sum_{t=0}^{n-1} h_j\left(\frac{t+1}{n}\right)(\widehat{\phi}_t^\star\widehat{\phi}_{t+1} - \phi_t^*\phi_{t+1})m_{t,t+1} \tag{60}$$

$$+\sqrt{n}(\widehat{\rho}^{-1} - \rho^{-1}) \times \frac{1}{n}\sum_{t=0}^{n-1} h_j\left(\frac{t+1}{n}\right)\phi_t^*\phi_{t+1}m_{t,t+1}\,. \tag{61}$$

Term (58) is $o_p(1)$ because $\frac{1}{n}\sum_{t=0}^{n-1} h_j\left(\frac{t+1}{n}\right) = O(n^{-1/2})$ (by numerical integration, using $\int_0^1 h(u)\,\mathrm{d}u = 0$) and $\sqrt{n}(\overline{\log m}_n - E[\log m(X_0,X_1)]) = O_p(1)$ (by Markov's inequality, using the fact that $\{X_t\}$ is geometrically rho-mixing under Assumption 4.1, and that enough moments exist by Assumption 5.3(iii)).

For term (59), write

$$(59) = \frac{1}{\sqrt{n}}\sum_{t=0}^{n-1} h_j\left(\frac{t+1}{n}\right)\{\phi_t^*\phi_t - \widehat{\phi}_t^{\star f}\widehat{\phi}_t^f - E[\phi_t^*\phi_t - \widehat{\phi}_t^{\star f}\widehat{\phi}_t^f]\} + E[\phi_t^*\phi_t - \widehat{\phi}_t^{\star f}\widehat{\phi}_t^f]\frac{1}{\sqrt{n}}\sum_{t=0}^{n-1} h_j\left(\frac{t+1}{n}\right)$$

The first term in this display is $o_p(1)$ by Assumption 5.4(i). The second term in this display is $o_p(1)$ because $\frac{1}{n}\sum_{t=0}^{n-1} h_j\left(\frac{t+1}{n}\right) = O(n^{-1/2})$ and

$$\begin{aligned}|E[\phi_t^*\phi_t - \widehat{\phi}_t^{\star f}\widehat{\phi}_t^f]| &\leq \|\phi^*\|\|\phi - \widehat{\phi}^f\| + \|\widehat{\phi}^f\|\|\phi^* - \widehat{\phi}^{\star f}\| \\ &= O_p(\delta_K + \delta_K^* + \bar{\eta}_{n,K})\end{aligned}$$

with the first line by the triangle and Cauchy-Schwarz inequalities, and the second line by Corollary 5.3(i)(ii).

A similar argument shows term (60) is $o_p(1)$, using Assumption 5.4(ii) and Corollary 5.3(i)(ii).

For term (61), $\sqrt{n}(\widehat{\rho}^{-1} - \rho^{-1}) = O_p(1)$ by Theorem 5.2 and the delta method. For the remaining term, write

$$\frac{1}{n}\sum_{t=0}^{n-1} h_j\left(\frac{t+1}{n}\right)\phi_t^*\phi_{t+1}m_{t,t+1} = \frac{1}{n}\sum_{t=0}^{n-1} h_j\left(\frac{t+1}{n}\right)\{\phi_t^*\phi_{t+1}m_{t,t+1} - \rho E[\phi^*(X_0)\phi(X_0)]\} \tag{62}$$

$$+\rho E[\phi^*(X_0)\phi(X_0)] \times \frac{1}{n}\sum_{t=0}^{n-1} h_j\left(\frac{t+1}{n}\right)\,. \tag{63}$$

Term (63) is $O(n^{-1/2})$. Consider the process $\{\phi_t^*\phi_{t+1}m_{t,t+1} - \rho E[\phi^*(X_0)\phi(X_0)]\}$ and let $V_\phi$ denote its long-run variance. $V_\phi$ is finite by geometric rho-mixing of $\{X_t\}$ and the moment assumptions in Assumption 5.3(i). Moreover, straightforward calculation shows

$$V_\phi = V_\rho + 2\rho^2 E[(\phi^*(X_0)\phi(X_0) - 1)^2] + \rho^2 \textstyle\sum_{t=-\infty}^{\infty} E[(\phi^*(X_0)\phi(X_0) - 1)(\phi^*(X_t)\phi(X_t) - 1)]$$

whence $V_\phi \geq V_\rho > 0$. Therefore, the process $\{\phi_t^* \phi_{t+1} m_{t,t+1} - \rho E[\phi^*(X_0)\phi(X_0)]\}$ satisfies an invariance principle under Assumption 4.1 (which implies the process is phi-mixing); see Corollary 2.2 of Peligrad (1985). Functional limit and Wiener integration arguments yield

$$\frac{1}{\sqrt{n}} \sum_{t=0}^{n-1} h_j \left( \frac{t+1}{n} \right) \{\phi_t^* \phi_{t+1} m_{t,t+1} - \rho E[\phi^*(X_0)\phi(X_0)]\} \to_d N(0, V_\phi)$$

since $\int_0^1 h(u)^2 \, \mathrm{d}u = 1$. Therefore, term (62) is $O_p(n^{-1/2})$, and so term (61) is $o_p(1)$.

Finally, the process $\{\Delta_{t,t+1}\}$ satisfies an invariance principle under Assumption 4.1 (which implies the process is geometrically phi-mixing) and Assumption 5.3 (which guarantees enough moments); see Corollary 2.2 of Peligrad (1985). Therefore, by the functional limit and Wiener integration arguments in Phillips (2005),

$$\begin{aligned} \begin{pmatrix} \sqrt{n}(\widehat{L} - L) & \widehat{\Lambda}_1 & \cdots & \widehat{\Lambda}_J \end{pmatrix}' &= \begin{pmatrix} \frac{1}{\sqrt{n}} \sum_{t=0}^{n-1} \Delta_{t,t+1} & \Lambda_1 & \cdots & \Lambda_J \end{pmatrix}' + o_p(1) \\ &\to_d \quad N(0, V_L \times I_{J+1}) \end{aligned}$$

and the result follows by definition of the $\chi_J^2$ and $t_J$ distributions. □

## D.3. **Proofs for Section 5.5.**

*Proof of Theorem 5.4.* Efficiency bound for $\rho$: The tangent space is first characterized as in Greenwood and Wefelmeyer (1995) (see also Wefelmeyer (1999); Greenwood, Schick, and Wefelmeyer (2001)). Let $\mathcal{BM}(\mathcal{X} \times \mathcal{X})$ denote the space of all real-valued bounded measurable functions on $\mathcal{X} \times \mathcal{X}$,[17] and define

$$\mathcal{T} = \{h \in \mathcal{BM}(\mathcal{X} \times \mathcal{X}) : \, E[h(X_0, X_1)|X_0 = x] = 0 \text{ for all } x \in \mathcal{X}\} \, .$$

Let $f(x_1|x_0)$ denote the true transition density of $X_1 = x_1$ given $X_0 = x_0$ (this exists by Assumption 4.1). For any $h \in \mathcal{T}$ there is $N_h \in \mathbb{N}$ such that for all $n \geq N_h$ the function

$$f_{n,h}(x_1|x_0) = f(x_1|x_0)\{1 + n^{-1/2} h(x_0, x_1)\}$$

is non-negative and integrates to 1 for every $x_0$, and is therefore a legitimate transition density.

Let $P_{n,h}$ denote the distribution of the sample $\{X_0, X_1, \ldots, X_n\}$ under the perturbed transition density $f_{n,h}$ and $P_{n,0}$ denote the distribution of the sample $\{X_0, X_1, \ldots, X_n\}$ under the true transition density $f$. A version of local asymptotic normality is known to obtain for this set of perturbed transition densities, i.e.

$$\log \frac{\mathrm{d}P_{n,h}}{\mathrm{d}P_{n,0}} = \frac{1}{\sqrt{n}} \sum_{i=0}^{n-1} h(X_i, X_{i+1}) - \frac{1}{2} E[h(X_0, X_1)^2] + o_{P_{n,0}}(1)$$

[17] It suffices to consider bounded measurable functions as the are dense in the space $\{f : \mathcal{X} \times \mathcal{X} \to \mathbb{R} \text{ s.t. } f \text{ is measurable and } E[f(X_0, X_1)^2] < \infty\}$ (Wefelmeyer, 1999).

(see Greenwood and Wefelmeyer (1995); Wefelmeyer (1999); Greenwood, Schick, and Wefelmeyer (2001)) where

$$\frac{1}{\sqrt{n}}\sum_{i=0}^{n-1} h(X_i, X_{i+1}) \to_d N(0, E[h(X_0, X_1)^2])$$

by a central limit theorem for martingales with stationary and ergodic differences (Billingsley, 1961).

The gradient of $\rho$ is now characterized in terms of the transition density. For any $h \in \mathcal{T}$ define the perturbed pricing operator $\mathbb{M}_{n,h} : L^2(Q) \to L^2(Q)$ by

$$\mathbb{M}_{n,h}\psi(x) = \int_{\mathcal{X}} m(x,y)\frac{f_{n,h}(y|x)}{q(y)}\psi(y)\,\mathrm{d}Q(y)$$

and let its kernel be defined as

$$\mathcal{K}_{n,h}(y,x) = m(x,y)\frac{f_{n,h}(y|x)}{q(y)}\,.$$

Whenever $h \in \mathcal{T}$

$$\int_{\mathcal{X}}\int_{\mathcal{X}} (\mathcal{K}(y,x) - K_{n,h}(y,x))^2\,\mathrm{d}Q(x)\mathrm{d}Q(y) \leq Cn^{-1}E[m(X_0,X_1)^2 h(X_0,X_1)^2]$$

for some finite positive constant $C$ under Assumption 4.1. This implies that $\|\mathbb{M} - \mathbb{M}_{n,h}\| = O(n^{-1/2})$ since the Hilbert-Schmidt norm dominates the operator norm. Application of Lemma D.6 shows that for $n$ sufficiently large, the maximum eigenvalue $\rho_{n,h}$ of $\mathbb{M}_{n,h}$ is real and positive and

$$\rho_{n,h} = \rho + E[\phi^*(X_0)(\mathbb{M}_{n,h} - \mathbb{M})\phi(X_0)] + o(n^{-1/2})\,.$$

By this and the law of iterated expectations,

$$\sqrt{n}(\rho_{n,h} - \rho) = E[\phi^*(X_0)m(X_0,X_1)\phi(X_1)h(X_0,X_1)] + o(1)$$

where the expectation is finite for all $h \in \mathcal{T}$ under Assumption 5.3(i). The gradient of $\rho$ is $\phi^*(x_0)m(x_0,x_1)\phi(x_1)$ and its projection onto (the closure of) $\mathcal{T}$ is

$$\begin{aligned}
\tilde{\psi}_\rho(x_0,x_1) &= \phi^*(x_0)m(x_0,x_1)\phi(x_1) - E[\phi^*(X_0)m(X_0,X_1)\phi(X_1)|X_0 = x_0] \\
&= \phi^*(x_0)m(x_0,x_1)\phi(x_1) - \phi^*(x_0)\mathbb{M}\phi(x_0) \\
&= \phi^*(x_0)m(x_0,x_1)\phi(x_1) - \rho\phi^*(x_0)\phi(x_0)\,.
\end{aligned}$$

Therefore $\tilde{\psi}_\rho(x_0,x_1)$ is the efficient influence function and $E[\tilde{\psi}_\rho(X_0,X_1)^2] = V_\rho$ is the efficiency bound for $\rho$. Theorem 5.2 shows $\widehat{\rho}$ attains this bound. Efficiency bound for $y$: follows (by continuity) from the efficiency bound for $\rho$.

Efficiency bound for $L$: As shown in Greenwood and Wefelmeyer (1995) and Wefelmeyer (1999),[18] the efficient influence function for estimating $E[\log m(X_0,X_1)]$ is

$$\begin{aligned}\tilde{\psi}_m(x_0,x_1) &= \log m(x_0,x_1) - E[\log m(X_0,X_1)|X_0=x_0] \\ &\quad + \sum_{t=1}^{\infty} \left( E[\log m(X_t,X_{t+1})|X_1=x_1] - E[\log m(X_t,X_{t+1})|X_0=x_0] \right) .\end{aligned}$$

The efficient influence function for $L$ is therefore, by linearity and continuity of log,

$$\tilde{\psi}_L(x_0,x_1) = \rho^{-1}\tilde{\psi}_\rho(x_0,x_1) - \tilde{\psi}_m(x_0,x_1)\,.$$

Note that

$$V_L = \rho^{-2}V_\rho - 2\rho^{-1}C_{\rho m} + V_m'$$

where

$$\begin{aligned} V_m &= \textstyle\sum_{t=-\infty}^{\infty} E[(\log m(X_0,X_1) - E[\log m(X_0,X_1)])(\log m(X_t,X_{t+1}) - E[\log m(X_0,X_1)])] \\ C_{\rho m} &= E[(\phi^*(X_0)m(X_0,X_1)\phi(X_1) - \rho\phi^*(X_0)\phi(X_0))\log m(X_0,X_1)]\,. \end{aligned}$$

The efficiency bound for $L$ is then

$$\begin{aligned} E[\tilde{\psi}_L(X_0,X_1)^2] &= \rho^{-2}V_\rho + E[\tilde{\psi}_m(X_0,X_1)^2] - 2\rho^{-1}E[\tilde{\psi}_\rho(X_0,X_1)\tilde{\psi}_m(X_0,X_1)] \\ &= \rho^{-2}V_\rho + V_m - 2\rho^{-1}E[\tilde{\psi}_\rho(X_0,X_1)\tilde{\psi}_m(X_0,X_1)] \end{aligned}$$

since $E[\tilde{\psi}_m(X_0,X_1)^2] = V_m$ (Wefelmeyer, 1999). Using the fact that $E[\tilde{\psi}_\rho(X_0,X_1)|X_0]=0$,

$$\begin{aligned} & E[\tilde{\psi}_\rho(X_0,X_1)\tilde{\psi}_m(X_0,X_1)] \\ = \;& E[\log m(X_0,X_1)\tilde{\psi}_\rho(X_0,X_1)] \\ & + \sum_{t=1}^{\infty} E[E[\log m(X_t,X_{t+1})|X_1]E[\tilde{\psi}_\rho(X_0,X_1)|X_1]] \\ = \;& E[\log m(X_0,X_1)\{\phi^*(X_0)m(X_0,X_1)\phi(X_1) - \rho\phi^*(X_0)\phi(X_0)\}] \\ & + \sum_{t=1}^{\infty} \rho E[E[\log m(X_t,X_{t+1})|X_1](\phi^*(X_1)\phi(X_1) - E[\phi^*(X_0)\phi(X_0)|X_1])] \\ = \;& E[\log m(X_0,X_1)\phi^*(X_0)m(X_0,X_1)\phi(X_1)] - \rho \lim_{t\to\infty} E[\log m(X_t,X_{t+1})\phi^*(X_0)\phi(X_0)] \\ = \;& C_{\rho m} \end{aligned}$$

where the second equality is by definition of $\phi^*$, the fourth is by telescoping series, and the fifth is by Lemma D.2. Therefore, $E[\tilde{\psi}_m(X_0,X_1)^2] = V_L$. Theorem 5.2 shows $\widehat{L}$ attains this bound. □

*Proof of Theorem 5.5.* Application of Theorem C.3 with $\widetilde{\mathbf{G}}_K$ in place of $\widehat{\widehat{\mathbf{G}}}_K$ yields

$$\check{\rho} - \rho_K = (\mathbf{G}_K^{1/2}c_K^\star)'(\widetilde{\mathbf{G}}_K^{-1}\widehat{\widetilde{\mathbf{M}}}_K - \widetilde{\mathbf{M}}_K)(\mathbf{G}_K^{1/2}c_K) + O_p(\bar{\eta}_{n,K}^2)\,.$$

[18]Greenwood and Wefelmeyer (1995) prove efficiency of sample averages for estimating the expectation of bounded measurable functions of $(X_0,X_1)$. Wefelmeyer (1999) extends this to the class of functions of $(X_0,X_1)$ with finite second moment.

Using the fact that $\widetilde{\mathbf{G}}_K^{-1} = I_K$ and $\mathbf{G}_K^{-1}\mathbf{M}_K c_K = \rho_K c_K$ yields

$$\check{\rho} - \rho_K = \frac{1}{n}\sum_{t=0}^{n-1} \phi_K^\star(X_t) m(X_t, X_{t+1})\phi_K(X_{t+1}) - \rho_K + O_p(\bar{\eta}_{n,K}^2)\,.$$

Applying the same arguments as in the proof of Theorem 5.2 yields

$$\sqrt{n}(\check{\rho} - \rho) = \frac{1}{\sqrt{n}}\sum_{t=0}^{n-1}\{\phi_t^* m_{t,t+1}\phi_{t+1} - \rho\} + o_p(1)\,.$$

The summands are strictly stationary phi-mixing random variables by Assumption 4.1 and Lemma D.1, have mean zero, and have and finite second moment by Assumption 5.3. It follows by Corollary 2.2 of Peligrad (1985) that

$$\sqrt{n}(\check{\rho} - \rho) \to_d N(0, \mathrm{lrvar}(\phi_t^* m_{t,t+1}\phi_{t+1} - \rho))\,.$$

By definition,

$$\begin{aligned} \mathrm{lrvar}(\phi_t^* m_{t,t+1}\phi_{t+1} - \rho) &= E[(\phi_0^* m_{0,1}\phi_1 - \rho)^2] \\ &\quad + 2\textstyle\sum_{t=1}^\infty E[(\phi_0^* m_{0,1}\phi_1 - \rho)(\phi_t^* m_{t,t+1}\phi_{t+1} - \rho)] \end{aligned} \tag{64}$$

where

$$E[(\phi_0^* m_{0,1}\phi_1 - \rho)^2] = V_\rho + \rho^2 E[(\phi^*(X_0)\phi(X_0) - 1)^2]$$

and, for each $t \geq 1$,

$$\begin{aligned} E[(\phi_0^* m_{0,1}\phi_1 - \rho)(\phi_t^* m_{t,t+1}\phi_{t+1} - \rho)] &= \rho^2 E[(\phi_1^*\phi_1 - 1)(\phi_t^*\phi_t - 1)] \\ &= \rho^2 E[(\phi_0^*\phi_0 - 1)(\phi_{t-1}^*\phi_{t-1} - 1)]\,. \end{aligned}$$

Substituting into (64) yields

$$\mathrm{lrvar}(\phi_t^* m_{t,t+1}\phi_{t+1} - \rho) = V_\rho + 2\rho^2 E[(\phi_0^*\phi_0 - 1)^2] + \rho^2 \mathrm{lrvar}((\phi_t^*\phi_t - 1))$$

as required. □

### D.4. **Proofs for Section 6.1.**

*Proof of Theorem 6.1.* Follows identical arguments to the proofs of Theorem 5.1. □

*Proof of Theorem 6.2.* Repeating the arguments in Lemma D.4 shows

$$\widehat{\rho} - \rho_K = \frac{1}{n}\sum_{t=0}^{n-1}\{\phi_K^\star(X_t)\widehat{m}(X_t, X_{t+1})\phi_K(X_{t+1}) - \rho_K \phi_K^\star(X_t)\phi_K(X_t)\} + O_p(\bar{\eta}_{n,K}^2)\,.$$

Therefore,

$$\widehat{\rho} - \rho = \rho_K - \rho + \frac{1}{n}\sum_{t=0}^{n-1}\{\phi_K^\star(X_t) m(X_t, X_{t+1})\phi_K(X_{t+1}) - \rho_K\phi_K^\star(X_t)\phi_K(X_t)\} + O_p(\bar{\eta}_{n,K}^2) \tag{65}$$

$$+\frac{1}{n}\sum_{t=0}^{n-1}\phi_K^\star(X_t)(\widehat{m}(X_t, X_{t+1}) - m(X_t, X_{t+1}))\phi_K(X_{t+1})\,. \tag{66}$$

Expression (65) is controlled as in the proof of Lemma D.4. It follows from the uniform convergence rates established in Corollary 5.2 and Assumption 6.1 that

$$(66) = \frac{1}{n}\sum_{t=0}^{n-1} \phi^*(X_t)(\widehat{m}(X_t, X_{t+1}) - m(X_t, X_{t+1}))\phi(X_{t+1}) + o_p(1)\,.$$

The result follows. □

D.5. **Proofs for Section 6.2.**

*Proof of Theorem 6.3.* The proof is analogous to the proof of Theorem 4.1. Note that $\mathbb{M} : L^2(Q) \to L^2(Q)$ can be represented as an integral operator with integral kernel $\mathcal{K}(x_0, x_1)$ given by

$$\begin{aligned}
\mathcal{K}(x_0, x_1) &= \left\{ \int_{\mathcal{Y}} m(x_0, x_1, y_1) f(x_1, y_1|x_0)\,\mathrm{d}y_1 \right\} \frac{1}{q(x_1)} \\
&= \left\{ \int_{\mathcal{Y}} m(x_0, x_1, y_1) \frac{f(x_0, x_1, y_1)}{q(x_0)q(x_1)q_y(y_1)}\,\mathrm{d}Q_y(y_1) \right\}.
\end{aligned}$$

The positivity conditions in Assumption 6.2 imply that $\mathcal{K}(x_0, x_1) > 0$ a.e.-$[Q \otimes Q]$.

To check square-integrability of $\mathcal{K}$, observe that

$$\begin{aligned}
& \int_{\mathcal{X}}\int_{\mathcal{X}} \mathcal{K}^2(x_0, x_1)\,\mathrm{d}Q(x_0)\,\mathrm{d}Q(x_1) \\
=\;& \int_{\mathcal{X}}\int_{\mathcal{X}} \left\{ \int_{\mathcal{Y}} m(x_0, x_1, y_1) \frac{f(x_0, x_1, y_1)}{q(x_0)q(x_1)q_y(y_1)}\,\mathrm{d}Q_y(y_1) \right\}^2 \mathrm{d}Q(x_0)\,\mathrm{d}Q(x_1) \\
\leq\;& \int_{\mathcal{X}}\int_{\mathcal{X}}\int_{\mathcal{Y}} m(x_0, x_1, y_1)^2 \frac{f(x_0, x_1, y_1)^2}{q(x_0)q(x_1)q_y(y_1)}\,\mathrm{d}y_1\,\mathrm{d}x_0\,\mathrm{d}x_1 \\
\leq\;& C\int_{\mathcal{X}}\int_{\mathcal{X}}\int_{\mathcal{Y}} m(x_0, x_1, y_1)^2 f(x_0, x_1, y_1)\,\mathrm{d}y_1\,\mathrm{d}x_0\,\mathrm{d}x_1 \\
=\;& CE[m(X_0, X_1, Y_1)^2] \quad < \quad \infty
\end{aligned}$$

for some finite positive $C$, by virtue of boundedness of $f(x_0, x_1, y_1)/(q(x_0)q(x_1)q_y(y_1))$, and Assumption 6.2(iv). The result follows by Theorem A.1. □

*Proof of Theorem 6.4.* Follows identical arguments to the proofs of Theorems 5.1 and 5.2, noting that in this case $\mathbb{M}$ may be rewritten as

$$\mathbb{M}\psi(x) = E[E[m(X_0, X_1, Y_1)|X_0, X_1]\psi(X_1)|X_0 = x]$$

where clearly $E[m(X_0, X_1, Y_1)|X_0, X_1]$ is a function of $(X_0, X_1)$. □

D.6. **Proofs for Appendix A.** All of the following definitions are as in Schaefer (1999). Let $E$ denote the Banach lattice $L^p(Q)$ for $1 \leq p \leq \infty$, and let $E^*$ denote its dual space. For $f \in E$, $f^* \in E^*$, define the evaluation $\langle f, f^* \rangle := f^*(f)$. Let $E_+$ denote the positive cone of $E$, i.e. $E_+ = \{f \in E : f \geq 0 \text{ a.e.-}[Q]\}$. An element $f \in E_+$ belongs to the quasi-interior $E_{++}$ of $E_+$ if $\{g \in E : 0 \leq g \leq f\}$ is a total subset of $E$. If $E = L^p(Q)$ with $1 \leq p < \infty$ and $f \in E$ is such that $f > 0$ a.e.-$[Q]$ then $f \in E_{++}$. If $E = L^\infty(Q)$ and $f \in E$ is such that $\operatorname{ess\,inf} f > 0$ then $f \in E_{++}$. The dual cone $E_+^* := \{f^* \in E^* : \langle f, f^* \rangle \geq 0$ whenever $f \in E_+\}$ is the set of positive linear functionals on $E$. An element $f^* \in E_+^*$ is strictly positive if $f \in E_+$ and $f \neq 0$ implies $\langle f, f^* \rangle > 0$. The set of all strictly positive elements of $E_+^*$ is denoted $E_{++}^*$. Recall that the adjoint $\mathbb{M}^* : E^* \to E^*$ of $\mathbb{M}$ is defined as $\mathbb{M}^*(f^*) = f^* \circ \mathbb{M}$, i.e. $f^* \circ \mathbb{M} : E \to \mathbb{R}$ is a bounded linear functional for each $f^* \in E^*$. The operator $\mathbb{M}$ is irreducible if $\mathbb{M}R(\mathbb{M}, z) : (E_+ \setminus \{0\}) \to E_{++}$ for each $z \in (\mathrm{spr}(\mathbb{M}), \infty)$ where $R(\mathbb{M}, z)$ is the resolvent of $T$. In what follows, unique means unique up to scale.

*Proof of Theorem A.2.* Theorem A.1 shows that $\mathbb{M}$ has an eigenfunction $\phi \in E$ corresponding to the eigenvalue $\rho = \mathrm{spr}(\mathbb{M})$. Moreover $\mathrm{spr}(\mathbb{M})$ an isolated eigenvalue because $\mathbb{M}^\tau$ is compact for some $\tau \geq 1$ under Assumption A.2 (Dunford and Schwartz, 1958, Theorem 6, p. 579).

Part (i), $1 \leq p < \infty$: $\mathbb{M}$ is irreducible (by the proof of Theorem V.6.6 in Schaefer (1974)) and so $\phi \in E_{++}$ and $\phi^* \in E_{++}^*$ and $\mathrm{spr}(\mathbb{M})$ is a pole of $R(\mathbb{M}, z)$ of order one, so $\mathrm{spr}(\mathbb{M})$ has algebraic and geometric multiplicity one (Schaefer, 1999, p. 318). Therefore, $\mathrm{spr}(\mathbb{M})$ is an eigenvalue of multiplicity one of $\mathbb{M}$ and $\mathbb{M}^*$ (algebraic and geometric multiplicities are preserved by taking adjoints: see Kato (1980), Remark III.6.23).

Suppose $\psi \in (E_+ \setminus \{0\})$ is a nonnegative eigenfunction of $\mathbb{M}$ with eigenvalue $\lambda$ such that $\psi$ and $\phi$ are linearly independent. Note that $\lambda \neq \mathrm{spr}(\mathbb{M})$ because $\mathrm{spr}(\mathbb{M})$ is an eigenvalue of $\mathbb{M}$ of multiplicity one. Also note that $\langle \psi, \phi^* \rangle > 0$ because $\phi^* \in E_{++}^*$ and $\psi \in (E_+ \setminus \{0\})$. Then,

$$\lambda \langle \psi, \phi^* \rangle = \langle \mathbb{M}\psi, \phi^* \rangle = (\phi^* \circ \mathbb{M})(\psi) = \mathbb{M}^*(\phi^*)(\psi) = \mathrm{spr}(\mathbb{M})\phi^*(\psi) = \mathrm{spr}(\mathbb{M})\langle \psi, \phi^* \rangle$$

which contradicts $\lambda \neq \mathrm{spr}(\mathbb{M})$. A similar argument shows that $\phi^*$ is the unique eigenfunction of $\mathbb{M}^*$ belonging to $E_+^*$.

Part (ii) $p = \infty$: the preadjoint of $\mathbb{M}$ on $L^1(Q)$ is irreducible (by the proof of Theorem V.6.6 in Schaefer (1974)). The proof for $L^1(Q)$ applied to the preadjoint of $\mathbb{M}$ provides that $\phi \in E_{++}$, $\phi$ is the unique eigenfunction of $\mathbb{M}$ belonging to $E_+$, and that $\rho$ is an eigenvalue of $\mathbb{M}$ of multiplicity one. $\square$

*Proof of Theorem A.3.* The positive eigenfunction and adjoint eigenfunction are unique (by Theorems A.1 and A.2), and $\rho$ is an eigenvalue of $\mathbb{M}$ of multiplicity one.

Let $\overline{\mathbb{M}} = \rho^{-1}\mathbb{M}$. The condition $\mathcal{K}(x, y) > 0$ a.e.-$[Q \otimes Q]$ implies that any eigenvalue $\lambda$ of $\overline{\mathbb{M}}$ with $\lambda \neq 1$ has $|\lambda| < 1$ (by Theorem A.1). By construction, $P$ is the spectral projection of $\overline{\mathbb{M}}$ corresponding to the eigenvalue 1.

Consider the bounded linear operator $\overline{\mathbb{M}} - P$. Let $\epsilon = \inf_{z \in \sigma(\overline{\mathbb{M}}) : z \neq 1} |z - 1|$ and note that $\epsilon > 0$. Define $g : \mathbb{C} \to \mathbb{R}$ such that $g(z) = 1$ for all $z \in \mathbb{C}$ with $|z - 1| \leq \epsilon/2$ and $g(z) = 0$ for all $z \in \mathbb{C}$ with $|z - 1| > \epsilon/2$. Let $f : \mathbb{C} \to \mathbb{R}$ be given by $f(z) = z - g(z)$. Then

$$\begin{aligned} \overline{\mathbb{M}} - P &= \frac{-1}{2\pi i} \int_{\Gamma(1+\epsilon,0)} z R(\overline{\mathbb{M}}, z) \, \mathrm{d}z - \frac{-1}{2\pi i} \int_{\Gamma(\frac{1}{2}\epsilon,1)} R(\overline{\mathbb{M}}, z) \, \mathrm{d}z \\ &= \frac{-1}{2\pi i} \int_{\Gamma(1+\epsilon,0)} f(z) R(\overline{\mathbb{M}}, z) \, \mathrm{d}z \end{aligned}$$

(Dunford and Schwartz, 1958, Theorem 10, p. 560) since the only singularity of $R(\mathbb{M}, z)$ within $\Gamma(\frac{1}{2}\epsilon, 1)$ is at $z = 1$. By the spectral mapping theorem (Dunford and Schwartz, 1958, Theorem 11, p. 569) $\sigma(\overline{\mathbb{M}} - P) = f(\sigma(\overline{\mathbb{M}}))$. Therefore $\mathrm{spr}(\overline{\mathbb{M}} - P) < 1$ because $f(1) = 0$ and $f(\lambda) = \lambda$ for any $\lambda \in \sigma(\overline{\mathbb{M}})$ with $\lambda \neq 1$.

By the Gelfand formula (Dunford and Schwartz, 1958, p. 567),

$$\mathrm{spr}(\overline{\mathbb{M}} - P) = \lim_{\tau \to \infty} \|(\overline{\mathbb{M}} - P)^{\tau}\|_{L^p(Q)}^{1/\tau} < 1 \,. \tag{67}$$

Let $\{\tau_k : k \geq 1\} \subseteq \mathbb{N}$ be the maximal subset of $\mathbb{N}$ for which $\|(\overline{\mathbb{M}} - P)^{\tau_k}\|_{L^p(Q)} > 0$ for each $\tau_k$. If this subsequence is finite then the proof is complete for suitable choice of $c$. If this subsequence is infinite, then by expression (67),

$$\limsup_{\tau_k \to \infty} \frac{\log \|(\overline{\mathbb{M}} - P)^{\tau_k}\|_{L^p(Q)}}{\tau_k} < 0 \,.$$

Therefore, there exists a finite positive constant $C$ such that for all $\tau_k$ large enough,

$$\log \|(\overline{\mathbb{M}} - P)^{\tau_k}\|_{L^p(Q)} \leq -C\tau_k \,.$$

Finally observe that $(\overline{\mathbb{M}} - P)^{\tau} = \overline{\mathbb{M}}^{\tau} - P = \overline{\mathbb{M}}_{\tau} - P$ since $\overline{\mathbb{M}}$ and $P$ commute (Kato, 1980, pp. 178–179) and $\overline{\mathbb{M}}P\psi = E[\psi(X)\phi^*(X)]\overline{\mathbb{M}}\phi = P\psi$ for all $\psi \in L^p(Q)$. □

*Proof of Theorem A.4.* Follows the same arguments as the proof of Theorem A.3. □

D.7. **Proofs for Appendix C.** Several lemmas are needed first before Theorems C.1, C.2 and C.3 are proved. Parts (i) and (ii) of the following Lemma are a straightforward modification of two results in Gobet, Hoffmann, and Reiß (2004); parts (iii) and (iv) deal with estimation of the adjoint eigenfunction and are new. Lemma D.6 is the key lemma from which the asymptotic linear expansion is derived.

**Lemma D.5.** *Let $\{T_\alpha, T_{\alpha,\epsilon} : \alpha \in \mathcal{A}\}$ be a collection of linear operators on a real Hilbert space such that $T_\alpha$ has an isolated real eigenvalue $\lambda_\alpha$ of multiplicity one for each $\alpha \in \mathcal{A}$. Let $f_\alpha$ denote the eigenfunction corresponding to $\lambda_\alpha$ normalized so that $\|f_\alpha\| = 1$. Suppose there exists a $\bar{T} < \infty$ such that $\sup_{\alpha \in \mathcal{A}} \|T_\alpha\| \leq \bar{T}$ and there exists a $\delta > 0$ such that $\inf_{z \in \sigma(T_\alpha): z \neq \lambda_\alpha} |z - \lambda_\alpha| > \delta$ for all $\alpha \in \mathcal{A}$. Let $\bar{r} = (\sup_{\alpha \in \mathcal{A}} \sup_{z \in \Gamma(\delta, \lambda_\alpha)} \|R(T_\alpha, z)\|)^{-1}$. If $\bar{r} < \infty$ and $\sup_{\alpha \in \mathcal{A}} \|T_\alpha - T_{\alpha,\epsilon}\| < \frac{1}{2}\bar{r}$, then*

- *(i) The only element of $\sigma(T_{\alpha,\epsilon})$ within $\Gamma(\delta, \lambda_\alpha)$ is an real eigenvalue $\lambda_{\alpha,\epsilon}$ of multiplicity one, and $\sup_{\alpha \in \mathcal{A}} |\lambda_\alpha - \lambda_{\alpha,\epsilon}| \leq ((\bar{T} + \frac{1}{2}\bar{r})\sqrt{8}\bar{r}^{-1} + 1) \sup_{\alpha \in \mathcal{A}} \|(T_\alpha - T_{\alpha,\epsilon})f_\alpha\|$*
- *(ii) Each $T_{\alpha,\epsilon}$ has an eigenfunction $f_{\alpha,\epsilon}$ corresponding to $\lambda_{\alpha,\epsilon}$ normalized so that $\|f_{\alpha,\epsilon}\| = 1$, and $\sup_{\alpha \in \mathcal{A}} \|f_\alpha - f_{\alpha,\epsilon}\| \leq \sqrt{8}\bar{r}^{-1} \sup_{\alpha \in \mathcal{A}} \|(T_\alpha - T_{\alpha,\epsilon})f_\alpha\|$*
- *(iii) Each $T^*_{\alpha,\epsilon}$ has an eigenfunction $f^*_{\alpha,\epsilon}$ corresponding to $\lambda_{\alpha,\epsilon}$ normalized so that $\langle f^*_{\alpha,\epsilon}, f_{\alpha,\epsilon}\rangle = 1$, and $\sup_{\alpha \in \mathcal{A}} \|f^*_\alpha/\|f^*_\alpha\| - f^*_{\alpha,\epsilon}/\|f^*_{\alpha,\epsilon}\|\| \leq \sqrt{8}\bar{r}^{-1} \sup_{\alpha \in \mathcal{A}} \|(T^*_\alpha - T^*_{\alpha,\epsilon})f^*_\alpha/\|f^*_\alpha\|\|$*
- *(iv) Moreover, $\sup_{\alpha \in \mathcal{A}} \|f^*_\alpha - f^*_{\alpha,\epsilon}\| \leq 2\delta\bar{r}^{-1}(\sqrt{2}\bar{r}^{-1} + 1) \sup_{\alpha \in \mathcal{A}} \|T_\alpha - T_{\alpha,\epsilon}\|$.*

*Proof of Lemma D.5.* Parts (i) and (ii) are a straightforward modification of Proposition 4.2 and Corollary 4.3 of Gobet, Hoffmann, and Reiß (2004). Note that

$$\sup_{\alpha \in \mathcal{A}} \sup_{z \in \Gamma(\delta, \lambda_\alpha)} \|R(T_\alpha, z)\| \|T_\alpha - T_{\alpha,\epsilon}\| \leq \frac{1}{2}$$

holds. This implies that $\Gamma(\delta, \lambda_\alpha)$ contains precisely one eigenvalue of $T_{\alpha,\epsilon}$ by Theorem IV.3.18 of Kato (1980) and the discussion in Section IV.3.5 of Kato (1980). Therefore each $\lambda_{\alpha,\epsilon}$ must be real-valued (if it were complex-valued its conjugate would also be in $\Gamma(\delta, \lambda_\alpha)$, which would contradict there being only one eigenvalue of $T_{\alpha,\epsilon}$ in $\Gamma(\delta, \lambda_\alpha)$).

For part (iii), existence of the $f^*_{\alpha,\epsilon}$ follows from the fact that each $\lambda_{\alpha,\epsilon}$ is an eigenvalue of multiplicity one. Applying part (ii) with $T^*_\alpha$ in place of $T_\alpha$ (using the fact that an operator and its adjoint have the same norm, and that $R(T^*_\alpha, z) = R(T_\alpha, \bar{z})$) yields part (iii).

For part (iv), let $P_\alpha$ denote the spectral projection of $T_\alpha$ corresponding to $\lambda_\alpha$ and let $P_{\alpha,\epsilon}$ be the projection of $T_{\alpha,\epsilon}$ corresponding to $\lambda_{\alpha,\epsilon}$. Note that both of these may be expressed as a contour integral over $\Gamma(\delta, \lambda_\alpha)$ (cf. equation (48)). Therefore,

$$\begin{aligned}
\sup_{\alpha \in \mathcal{A}} \|P_\alpha - P_{\alpha,\epsilon}\| &= \sup_{\alpha \in \mathcal{A}} \left\| \frac{-1}{2\pi i} \int_{\Gamma(\delta,\lambda_\alpha)} R(T_\alpha, z) - R(T_{\alpha,\epsilon}, z)\, \mathrm{d}z \right\| \\
&\leq \frac{1}{2\pi} 2\pi\delta \sup_{\alpha \in \mathcal{A}} \sup_{z \in \Gamma(\delta,\lambda_\alpha)} \|R(T_\alpha, z) - R(T_{\alpha,\epsilon}, z)\| \\
&\leq \delta \sup_{\alpha \in \mathcal{A}} \sup_{z \in \Gamma(\delta,\lambda_\alpha)} \|R(T_\alpha, z)\| \|T_\alpha - T_{\alpha,\epsilon}\| \|R(T_{\alpha,\epsilon}, z)\|
\end{aligned}$$

where the second inequality is by expression (47). The condition $\sup_{\alpha \in \mathcal{A}} \|T_\alpha - T_{\alpha,\epsilon}\| < \frac{1}{2}\bar{r}$, together with expression (47), also implies that

$$\|R(T_{\alpha,\epsilon}, z)\| \leq 2\|R(T_\alpha, z)\|$$

and so

$$\sup_{\alpha \in \mathcal{A}} \|P_\alpha - P_{\alpha,\epsilon}\| \leq 2\delta\bar{r}^{-1} \sup_{\alpha \in \mathcal{A}} \|T_\alpha - T_{\alpha,\epsilon}\|. \tag{68}$$

Note that $P_\alpha = f_\alpha \otimes f^*_\alpha$ and $P_{\alpha,\epsilon} = f_{\alpha,\epsilon} \otimes f^*_{\alpha,\epsilon}$ with $\|P_\alpha\| = \|f^*_\alpha\|$ and $\|P_{\alpha,\epsilon}\| = \|f^*_{\alpha,\epsilon}\|$. It follows by the forward and reverse triangle inequalities and (68) and part (iii) that

$$\begin{aligned}
\|f^*_\alpha - f^*_{\alpha,\epsilon}\| &\leq \|f^*_\alpha\| \left\| \frac{f^*_\alpha}{\|f^*_\alpha\|} - \frac{f^*_{\alpha,\epsilon}}{\|f^*_{\alpha,\epsilon}\|} \right\| + |\|f^*_\alpha\| - \|f^*_{\alpha,\epsilon}\|| \\
&\leq \left( \sup_{\alpha \in \mathcal{A}} \|f^*_\alpha\| \right) \sqrt{8}\bar{r}^{-1} \sup_{\alpha \in \mathcal{A}} \|(T^*_\alpha - T^*_{\alpha,\epsilon}) f^*_\alpha / \|f^*_\alpha\|\| + 2\delta\bar{r}^{-1} \sup_{\alpha \in \mathcal{A}} \|T_\alpha - T_{\alpha,\epsilon}\| \\
&\leq 2\delta\bar{r}^{-1}(\sqrt{2}\bar{r}^{-1} + 1) \sup_{\alpha \in \mathcal{A}} \|T_\alpha - T_{\alpha,\epsilon}\|
\end{aligned}$$

where the final line uses the fact that $\|f^*_\alpha\| = \|P_\alpha\| \leq \delta\bar{r}^{-1}$ and the definition of the operator norm. □

**Lemma D.6.** *Let $\{T_\alpha, T_{\alpha,\epsilon} : \alpha \in \mathcal{A}\}$ be a collection of bounded linear operators on a real Hilbert space such that $T_\alpha$ has an isolated real eigenvalue $\lambda_\alpha$ of multiplicity one for each $\alpha \in \mathcal{A}$. Let $f_\alpha$ and*

*$f_\alpha^*$ denote the eigenfunctions of $T_\alpha$ and $T_\alpha^*$ corresponding to $\lambda_\alpha$ normalized so that $\|f_\alpha\| = 1$ and $\langle f_\alpha, f_\alpha^*\rangle = 1$. Suppose there exists a $\delta > 0$ such that $\inf_{z\in\sigma(T_\alpha):z\neq\lambda_\alpha} |z - \lambda_\alpha| > \delta$ for each $\alpha \in \mathcal{A}$. Let $\bar{r} = \inf_{\alpha\in\mathcal{A}} \inf_{z\in\Gamma(\delta,\lambda_\alpha)}(\|R(T_\alpha, z)\|\|T_\alpha - T_{\alpha,\epsilon}\|)^{-1}$. If $\bar{r} > 1$ then the only element of $\sigma(T_{\alpha,\epsilon})$ within $\Gamma(\delta, \lambda_\alpha)$ is a real eigenvalue $\lambda_{\alpha,\epsilon}$ of multiplicity one, and*

$$\sup_{\alpha\in\mathcal{A}} |\lambda_{\alpha,\epsilon} - \lambda_\alpha - \langle f_\alpha^*, (T_{\alpha,\epsilon} - T_\alpha)f_\alpha\rangle| \leq \frac{\delta}{\bar{r}(\bar{r}-1)}\,.$$

*Proof of Lemma D.6.* For each $x \in \mathbb{R}$ and $\alpha \in \mathcal{A}$ define $T_{\alpha,\epsilon}(x) = T_\alpha + x(T_{\alpha,\epsilon} - T_\alpha)$. By the discussion on p. 379 of Kato (1980), the unique element of $\sigma(T_{\alpha,\epsilon}(x))$ within $\Gamma(\delta, \lambda_\alpha)$ is an eigenvalue of multiplicity one, say $\lambda_{\alpha,\epsilon}(x)$, for each $|x| < \bar{r}$, for each $\alpha \in \mathcal{A}$. Let $P_\alpha$ denote the spectral projection of $T_\alpha$ corresponding to $\lambda_\alpha$. By the error estimates in Section II.3.1 of Kato (1980),

$$|\lambda_{\alpha,\epsilon}(x) - \lambda_\alpha - x\mathrm{tr}\{(T_{\alpha,\epsilon} - T_\alpha)P_\alpha\}| \leq \frac{|x|^2 \sup_{z\in\Gamma(\delta,\lambda_\alpha)} |z - \lambda_\alpha|}{\bar{r}(\bar{r} - |x|)} = \frac{|x|^2\delta}{\bar{r}(\bar{r} - |x|)}$$

for each $|x| < \bar{r}$, for each $\alpha \in \mathcal{A}$ (each $T_\alpha$ and $T_{\alpha,\epsilon}$ are bounded so the results from finite-dimensional perturbation theory can be applied, see Section VII.3.2 of Kato (1980)). The result follows by setting $\delta = 1$ and using the relation $\mathrm{tr}\{(T_{\alpha,\epsilon} - T_\alpha)P_\alpha\} = \{(T_{\alpha,\epsilon} - T_\alpha)(f_\alpha \otimes f_\alpha^*)\} = \langle f_\alpha^*, (T_{\alpha,\epsilon} - T_\alpha)f_\alpha\rangle$. □

**Lemma D.7.** *Under Assumptions C.1, C.2, and C.3(i), there exists $\bar{K}$ sufficiently large such that for each $K \geq \bar{K}$, $\inf_{\alpha\in\mathcal{A}} \inf_{z\in\sigma(\Pi_K^b\mathbb{M}_\alpha):z\neq\rho_{\alpha,K}} |z - \rho_{\alpha,K}| \geq \frac{1}{2}\bar{\epsilon}$.*

*Proof of Lemma D.7.* By Theorem C.1 there is $K_0$ sufficiently large that $\sup_{\alpha\in\mathcal{A}} |\rho_\alpha - \rho_{\alpha,K}| \leq \frac{1}{4}\bar{\epsilon}$ for all $K \geq K_0$. By Theorem IV.3.18 of Kato (1980) and the discussion in Section IV.3.5 of Kato (1980), each $\Gamma(\frac{3}{4}, \rho_\alpha)$ encloses precisely one eigenvalue of $\Pi_K^b\mathbb{M}_\alpha$ provided

$$\sup_{\alpha\in\mathcal{A}} \sup_{z\in\Gamma(\frac{3}{4}\bar{\epsilon},\rho_\alpha)} \|R(\mathbb{M}_\alpha, z)\|\|\Pi_K^b\mathbb{M}_\alpha - \mathbb{M}_\alpha\| < 1\,.$$

By Assumptions C.1(ii), C.3(i) and C.2

$$\sup_{\alpha\in\mathcal{A}} \sup_{z\in\Gamma(\frac{3}{4}\bar{\epsilon},\rho_\alpha)} \|R(\mathbb{M}_\alpha, z)\|\|\Pi_K^b\mathbb{M}_\alpha - \mathbb{M}_\alpha\| \leq r(\tfrac{1}{4}\bar{\epsilon}) \times o(1) = o(1)$$

and so $\sup_{\alpha\in\mathcal{A}} \sup_{z\in\Gamma(\frac{3}{4}\bar{\epsilon},\rho_\alpha)} \|R(\mathbb{M}_\alpha, z)\|\|\Pi_K^b\mathbb{M}_\alpha - \mathbb{M}_\alpha\| < 1$ for all $K \geq K_1$ for some $K_1$. The result follows by setting $\bar{K} = \max\{K_0, K_1\}$. □

**Lemma D.8.** *Under Assumptions C.1(ii) and C.4(ii),*

$$\begin{aligned}
\sup_{\alpha\in\mathcal{A}} \left\|\hat{\widetilde{\mathbf{G}}}_K^{-1}\widehat{\widetilde{\mathbf{M}}}_{\alpha,K} - \widetilde{\mathbf{M}}_{\alpha,K}\right\|_2 &= O_p(\bar{\eta}_{n,K}) \\
\sup_{\alpha\in\mathcal{A}} \left\|\hat{\widetilde{\mathbf{G}}}_K^{-1}\widehat{\widetilde{\mathbf{M}}}'_{\alpha,K} - \widetilde{\mathbf{M}}'_{\alpha,K}\right\|_2 &= O_p(\bar{\eta}_{n,K})\,.
\end{aligned}$$

*Proof of Lemma D.8.* First note that since $\widetilde{\mathbf{M}}_{\alpha,K}$ is isomorphic to $\Pi_K^b\mathbb{M}_\alpha|_{B_K}$,

$$\|\widetilde{\mathbf{M}}_{\alpha,K}\|_2 = \|\Pi_K^b\mathbb{M}_\alpha|_{B_K}\| \leq \|\Pi_K^b\mathbb{M}_\alpha\| \leq \|\mathbb{M}_\alpha\|\,.$$

Therefore, $\sup_{\alpha\in\mathcal{A}} \|\widetilde{\mathbf{M}}_{\alpha,K}\|_2$ is bounded uniformly in $K$ by Assumption C.1(ii).

The condition $\|\widehat{\widehat{\mathbf{G}}}_K - I_K\|_2 = o_p(1)$ implies the eigenvalues of $\widehat{\widehat{\mathbf{G}}}_K$ are bounded between $\frac{1}{2}$ and 2 on a set whose probability is approaching one. Working on this set,

$$\begin{aligned}\widehat{\widehat{\mathbf{G}}}_K^{-1}\widehat{\widetilde{\mathbf{M}}}_{\alpha,K} - \widetilde{\mathbf{M}}_{\alpha,K} &= \left(I_K - \widehat{\widehat{\mathbf{G}}}_K^{-1}(\widehat{\widehat{\mathbf{G}}}_K - I_K)\right)\widehat{\widetilde{\mathbf{M}}}_{\alpha,K} - \widetilde{\mathbf{M}}_{\alpha,K} \\ &= \widehat{\widetilde{\mathbf{M}}}_{\alpha,K} - \widetilde{\mathbf{M}}_{\alpha,K} - \widehat{\widehat{\mathbf{G}}}_K^{-1}(\widehat{\widehat{\mathbf{G}}}_K - I_K)\widetilde{\mathbf{M}}_{\alpha,K} - \widehat{\widehat{\mathbf{G}}}_K^{-1}(\widehat{\widehat{\mathbf{G}}}_K - I_K)(\widehat{\widetilde{\mathbf{M}}}_{\alpha,K} - \widetilde{\mathbf{M}}_{\alpha,K})\end{aligned}$$

for each $\alpha \in \mathcal{A}$. The result follows by the triangle inequality and Assumption C.4(ii), noting that $\|\widehat{\widehat{\mathbf{G}}}_K^{-1}\|_2 \leq 2$ whenever the eigenvalues of $\widehat{\widehat{\mathbf{G}}}_K$ are bounded between $\frac{1}{2}$ and 2.

The proof for $\widehat{\widehat{\mathbf{G}}}_K^{-1}\widehat{\widetilde{\mathbf{M}}}'_{\alpha,K}$ follows similar arguments, using the fact that an operator and its adjoint have the same (operator) norm. □

**Lemma D.9.** *Under Assumptions C.1(ii) and C.4(iii), if $\bar{\eta}_{n,K} = o(1)$ then*

$$\begin{aligned}\sup_{\alpha\in\mathcal{A}}\left\|(\widehat{\widehat{\mathbf{G}}}_K^{-1}\widehat{\widetilde{\mathbf{M}}}_{\alpha,K} - \widetilde{\mathbf{M}}_{\alpha,K})\tilde{c}_{\alpha,K}\right\|_2 &= O_p(\eta_{n,K}) \\ \sup_{\alpha\in\mathcal{A}}\left\|(\widehat{\widehat{\mathbf{G}}}_K^{-1}\widehat{\widetilde{\mathbf{M}}}'_{\alpha,K} - \widetilde{\mathbf{M}}'_{\alpha,K})\tilde{c}^\star_{\alpha,K}/\|\tilde{c}^\star_{\alpha,K}\|_2\right\|_2 &= O_p(\eta_{n,K})\,.\end{aligned}$$

*Proof of Lemma D.9.* The same arguments as the proof of Lemma D.8 give

$$\sup_{\alpha\in\mathcal{A}}\left\|(\widehat{\widehat{\mathbf{G}}}_K^{-1}\widehat{\widetilde{\mathbf{M}}}_{\alpha,K} - \widetilde{\mathbf{M}}_{\alpha,K})\tilde{c}_{\alpha,K}\right\|_2 = O_p(\eta_{2,n,K}) + O_p(\eta_{1,n,K}) + O_p(\bar{\eta}_{1,n,K}) \times O_p(\eta_{2,n,K})\,.$$

The result follows by definition of $\eta_{n,K}$ and $\bar{\eta}_{n,K}$ and the condition $\bar{\eta}_{n,K} = o(1)$. The proof with $\widehat{\widetilde{\mathbf{M}}}'_{\alpha,K}$ is the same. □

*Proof of Theorem C.1.* Apply of Lemma D.5 with $\mathbb{M}_\alpha = T_\alpha$, $\Pi_K^b\mathbb{M}_\alpha = T_{\alpha,\epsilon}$ on the Hilbert space $L^2(Q)$. Set $\Gamma(\delta,\lambda_\alpha) = \Gamma(\frac{1}{2}\bar{\epsilon},\rho_\alpha)$. The resolvent bound in Assumption C.3(i) shows that for each $\alpha \in \mathcal{A}$ and $z \in \Gamma(\frac{1}{2}\bar{\epsilon},\rho_\alpha)$

$$2\bar{\epsilon}^{-1} = \frac{1}{d(z,\rho_\alpha)} \leq \|R(\mathbb{M}_\alpha, z)\| \leq r(d(z,\rho_\alpha)) \leq r(\tfrac{1}{2}\bar{\epsilon})$$

which implies

$$0 < r(\tfrac{1}{2}\bar{\epsilon})^{-1} \leq \bar{r} \leq \tfrac{1}{2}\bar{\epsilon} < \infty\,.$$

Assumption C.2(i) implies that $\sup_{\alpha\in\mathcal{A}}\|\Pi_K^b\mathbb{M} - \mathbb{M}_\alpha\| = o(1)$ thus $\sup_{\alpha\in\mathcal{A}}\|\Pi_K^b\mathbb{M} - \mathbb{M}_\alpha\| \leq \frac{1}{2}\bar{r}$ holds for all $K$ sufficiently large. □

*Proof of Theorem C.2.* Apply Lemma D.5 with $\widetilde{\mathbf{M}}_{\alpha,K} = T_\alpha$, $\widehat{\widehat{\mathbf{G}}}_K^{-1}\widehat{\widetilde{\mathbf{M}}}_{\alpha,K} = T_{\alpha,\epsilon}$ on the Hilbert space $\mathbb{R}^K$ (with the Euclidian inner (dot) product). Set $\Gamma(\delta,\lambda_\alpha) = \Gamma(\frac{1}{4}\bar{\epsilon},\rho_{\alpha,K})$. By Lemma D.7 there is a $\bar{K}$ sufficiently large that $\inf_{\alpha\in\mathcal{A}}\inf_{z\in\sigma(\Pi_K^b\mathbb{M}_\alpha):z\neq\rho_{\alpha,K}}|z-\rho_{\alpha,K}| \geq \frac{1}{2}\bar{\epsilon}$ for all $K \geq \bar{K}$. Take $K \geq \bar{K}$. The fact that $\widetilde{\mathbf{M}}_{\alpha,K}$ and $\Pi_K^b\mathbb{M}_\alpha|_{B_K}$ are isomorphic and the resolvent bound in Assumption C.3(ii) shows that for each $\alpha \in \mathcal{A}$ and $z \in \Gamma(\frac{1}{4}\bar{\epsilon},\rho_{\alpha,K})$

$$4\bar{\epsilon}^{-1} = \frac{1}{d(z,\rho_{\alpha,K})} \leq \|R(\Pi_K^b\mathbb{M}_\alpha|_{B_K}, z)\| \leq r(d(z,\rho_\alpha)) \leq r(\tfrac{1}{4}\bar{\epsilon})$$

which implies

$$0 < r(\tfrac{1}{4}\bar{\epsilon})^{-1} \leq \bar{r} \leq \tfrac{1}{4}\bar{\epsilon} < \infty\,.$$

Lemma D.8 provides that $\sup_{\alpha\in\mathcal{A}} \|\widehat{\widetilde{\mathbf{G}}}_K^{-1}\widehat{\widetilde{\mathbf{M}}}_{\alpha,K} - \widetilde{\mathbf{M}}_{\alpha,K}\|_2 = O_p(\bar{\eta}_{n,K})$. This and the condition $\bar{\eta}_{n,K} = o(1)$ implies the condition

$$\sup_{\alpha\in\mathcal{A}} \left\| \widehat{\widetilde{\mathbf{G}}}_K^{-1}\widehat{\widetilde{\mathbf{M}}}_{\alpha,K} - \widetilde{\mathbf{M}}_{\alpha,K} \right\|_2 \leq \frac{1}{2}\bar{r}$$

holds on a set whose probability is approaching one. Application of Lemma D.5 on this set proves parts (i) and (ii), with $\eta_{n,K}$ given by Lemma D.9.

Part (iii) follows by repeating this argument with $\widehat{\widetilde{\mathbf{M}}}'_{\alpha,K}$and $\widetilde{\mathbf{M}}'_{\alpha,K}$ in place of $\widehat{\widetilde{\mathbf{M}}}_{\alpha,K}$ and $\widetilde{\mathbf{M}}_{\alpha,K}$.

For part (iv), write

$$\begin{aligned}
\sup_{\alpha\in\mathcal{A}} \|\widehat{\phi}^\star_\alpha - \phi^\star_{\alpha,K}\| &\leq \sup_{\alpha\in\mathcal{A}} \|\phi^\star_{\alpha,K}\| \left\| \frac{\widehat{\phi}^\star_\alpha}{\|\widehat{\phi}^\star_\alpha\|} - \frac{\phi^\star_{\alpha,K}}{\|\phi^\star_{\alpha,K}\|} \right\| + \sup_{\alpha\in\mathcal{A}} |\|\widehat{\phi}^\star_{\alpha,K}\| - \|\phi^\star_{\alpha,K}\|| \\
&\leq \sup_{\alpha\in\mathcal{A}} \|\phi^\star_{\alpha,K}\| \times O_p(\eta_{n,K}) + \sup_{\alpha\in\mathcal{A}} |\|\widehat{\phi}^\star_{\alpha,K}\| - \|\phi^\star_{\alpha,K}\||
\end{aligned}$$

where the second line is by part (iii). Theorem C.1 shows that $\sup_{\alpha\in\mathcal{A}} \|\phi^\star_{\alpha,K}\| \leq \sup_{\alpha\in\mathcal{A}} \|\phi^*_\alpha\| + o(1)$ and it is easy to see that $\sup_{\alpha\in\mathcal{A}} \|\phi^*_\alpha\| < \infty$. The remaining term $\sup_{\alpha\in\mathcal{A}} |\|\widehat{\phi}^\star_{\alpha,K}\| - \|\phi^\star_{\alpha,K}\||$ can be shown to be $O_p(\bar{\eta}_{n,K})$ using a similar argument to the proof of part (iv) of Lemma D.5. □

*Proof of Theorem C.3.* First apply Lemma D.5 with $T_\alpha = \widetilde{\mathbf{M}}_{\alpha,K}$ and $T_{\alpha,\epsilon} = \widehat{\widetilde{\mathbf{G}}}_K^{-1}\widehat{\widetilde{\mathbf{M}}}_{\alpha,K}$. By Lemma D.7 there is a $\bar{K}$ sufficiently large that $\inf_{\alpha\in\mathcal{A}} \inf_{z\in\sigma(\Pi_K^b\mathbb{M}_\alpha):z\neq\rho_{\alpha,K}} |z - \rho_{\alpha,K}| \geq \frac{1}{2}\bar{\epsilon}$ for all $K \geq \bar{K}$. Take $K \geq \bar{K}$. Then $\|R(\widetilde{\mathbf{M}}_{\alpha,K}, z)\| \leq r(\frac{1}{4}\bar{\epsilon})$ for all $z \in \Gamma(\frac{1}{4}\bar{\epsilon}, \rho_{\alpha,K})$ by Assumption C.3(ii). The condition

$$r(\tfrac{1}{4}\bar{\epsilon}) > \|\widehat{\widetilde{\mathbf{G}}}_K^{-1}\widehat{\widetilde{\mathbf{M}}}_{\alpha,K} - \widetilde{\mathbf{M}}_{\alpha,K}\|_2$$

holds on a set whose probability is approaching one, since $\|\widehat{\widetilde{\mathbf{G}}}_K^{-1}\widehat{\widetilde{\mathbf{M}}}_{\alpha,K} - \widetilde{\mathbf{M}}_{\alpha,K}\|_2 = o_p(1)$ by Lemma D.8 and the condition $\bar{\eta}_{n,K} = o(1)$. Therefore, the condition

$$r_{n,K} := \inf_{\alpha\in\mathcal{A}} \inf_{z\in\Gamma(\frac{1}{4}\bar{\epsilon},\rho_{\alpha,K})} (\|R(\widetilde{\mathbf{M}}_{\alpha,K}, z)\|_2 \|\widehat{\widetilde{\mathbf{G}}}_K^{-1}\widehat{\widetilde{\mathbf{M}}}_{\alpha,K} - \widetilde{\mathbf{M}}_{\alpha,K}\|_2)^{-1} > 1$$

holds on a set whose probability is approaching one, on which Lemma D.5 provides that

$$\sup_{\alpha\in\mathcal{A}} \left| \widehat{\rho}_{\alpha,K} - \rho_\alpha - \widetilde{c}^{\star\prime}_{\alpha,K}(\widehat{\widetilde{\mathbf{G}}}_K^{-1}\widehat{\widetilde{\mathbf{M}}}_{\alpha,K} - \widetilde{\mathbf{M}}_{\alpha,K})\widetilde{c}_{\alpha,K} \right| \leq \frac{\bar{\epsilon}}{4r_{n,K}(r_{n,K}-1)}$$

uniformly for $\alpha \in \mathcal{A}$. The result follows by noticing that

$$\frac{1}{r_{n,K}(r_{n,K}-1)} = O_p(\bar{\eta}^2_{n,K})$$

by definition of $r_{n,K}$ and Lemma D.8. □

D.7.1. *Proofs of additional results on convergence of the matrix estimators.*

*Proof of Lemma C.1.* Let $\bar{M}$ be a finite positive constant such that $m(x_0,x_1) \le \bar{M}$. Let $m_{t,t+1} = m(X_t, X_{t+1})$. Consider the $K \times K$ random matrix

$$\Xi_{t,n} = n^{-1}\left(\widetilde{b}^K(X_t) m_{t,t+1} \widetilde{b}^K(X_{t+1}) - E[\widetilde{b}^K(X_0) m_{0,1} \widetilde{b}^K(X_1)]\right)$$

where clearly $E[\Xi_{t,n}] = 0$. Assumption C.5 and definition of $\zeta_0(K)$ imply that

$$\|\Xi_{t,n}\|_2 \le \frac{2\zeta_0(K)^2 \bar{M}}{\underline{\lambda} n} \,. \tag{69}$$

By the triangle and Cauchy-Schwarz inequalities, for any $u, v \in \mathbb{R}^K$ with $u'u = 1$ and $v'v = 1$,

$$\begin{aligned}
n^2 E[u'\Xi_{t,n}\Xi_{s,n}' v] &\le |E[u'\widetilde{b}^K(X_t) m_t \widetilde{b}^K(X_{t+1})' \widetilde{b}^K(X_{s+1}) m_s \widetilde{b}^K(X_s)' v]| \\
&\quad + |E[u'\widetilde{b}^K(X_0) m_0 \widetilde{b}^K(X_1)'] E[\widetilde{b}^K(X_1) m_0 \widetilde{b}^K(X_0)' v]| \\
&\le \bar{M}^2 E[|u'\widetilde{b}^K(X_t)||\widetilde{b}^K(X_{t+1})'\widetilde{b}^K(X_{s+1})||\widetilde{b}^K(X_s)'v|] \\
&\quad + \bar{M}^2 \underline{\lambda}^{-1} \zeta_0(K)^2 E[(u'\widetilde{b}^K(X_0))^2]^{1/2} E[(\widetilde{b}^K(X_0)'v)^2]^{1/2} \\
&\le 2\underline{\lambda}^{-1} \bar{M}^2 \zeta_0(K)^2 E[(u'\widetilde{b}^K(X_0))^2]^{1/2} E[(\widetilde{b}^K(X_0)'v)^2]^{1/2} \\
&\le 2\underline{\lambda}^{-1} \bar{M}^2 \zeta_0(K)^2
\end{aligned}$$

where the final line is because $E[(u'\widetilde{b}^K(X_0))^2] = u'E[\widetilde{b}_K(X_0)\widetilde{b}_K(X_0)']u = u'u = 1$ for any $u \in \mathbb{R}^K$ with $u'u = 1$. Since $\|A\|_2 = \sup_{u,v \in \mathbb{R}^K : u'u=1, v'v=1} u'Av$ for any $K \times K$ matrix $A$,

$$\left\| E\left[\Xi_{t,n}\Xi_{s,n}'\right]\right\|_2 \le \frac{2\bar{M}^2 \zeta_0(K)^2}{\underline{\lambda} n^2}$$

and similarly for $\|E[\Xi_{t,n}'\Xi_{s,n}]\|_2$.

The result follows by Corollary 5.2 of Chen and Christensen (2013). □

*Proof of Lemma C.2.* By geometric rho-mixing there exists a finite positive $C$ such that

$$\operatorname{Var}\left[\sum_{t=0}^{n-1} b(X_t, X_{t+1})\right] \le CnE[b(X_0,X_1)^2]$$

uniformly for all measurable $b : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ such that $E[b(X_0,X_1)^2] < \infty$ (see Lemma D.2). By the relation between the spectral and Frobenius norms,

$$\begin{aligned}
E[\|\widehat{\widetilde{\mathbf{M}}}_K - \widetilde{\mathbf{M}}_K\|_2^2] &\le \frac{1}{n^2}\sum_{k=1}^K \sum_{l=1}^K \operatorname{Var}\left[\sum_{t=0}^{n-1} \widetilde{b}_k^K(X_t)\widetilde{b}_l^K(X_{t+1}) m(X_t, X_{t+1})\right] \\
&\le \frac{C}{n}\sum_{k=1}^K \sum_{l=1}^K E\left[(\widetilde{b}_k^K(X_0)\widetilde{b}_l^K(X_1) m(X_0,X_1))^2\right] \\
&\le \frac{C\zeta_0(K)^2}{\underline{\lambda} n} \sum_{k=1}^K E\left[(\widetilde{b}_k^K(X_0) m(X_0,X_1))^2\right]
\end{aligned}$$

where

$$\sum_{k=1}^K E\left[(\widetilde{b}_k^K(X_0) m(X_0,X_1))^2\right] \le \begin{cases} \underline{\lambda}^{-1}\zeta_0(K)^2 E[m(X_0,X_1)^2] & \\ K\sup_{x_0,x_1}|m(x_0,x_1)|^2 & \text{if } m \text{ is bounded.}\end{cases}$$

The result follows by Markov's inequality. $\square$

*Proof of Lemma C.3.* By the arguments in the Proof of Lemma C.2

$$\begin{aligned}
E[\|(\widehat{\widetilde{\mathbf{M}}}_K - \widetilde{\mathbf{M}}_K)v_K\|_2^2] &\leq \frac{1}{n^2}\sum_{k=1}^{K} \mathrm{Var}\left[\sum_{t=0}^{n-1} \tilde{b}_k^K(X_t)(\tilde{b}^K(X_{t+1})'v_K)m(X_t,X_{t+1})\right] \\
&\leq \frac{C}{n}\sum_{k=1}^{K} E\left[(\tilde{b}_k^K(X_0)(\tilde{b}^K(X_1)'v_K)m(X_0,X_1))^2\right] \\
&\leq \frac{C\zeta_0(K)^2}{\underline{\lambda} n} E\left[((\tilde{b}^K(X_1)'v_K)^2 m(X_0,X_1)^2\right]
\end{aligned}$$

for some finite positive constant $C$, where

$$E\left[((\tilde{b}^K(X_1)'v_K)^2 m(X_0,X_1)^2\right] \leq \begin{cases} \underline{\lambda}^{-1}\zeta_0(K)^2\|v_K\|_2^2 E[m(X_0,X_1)^2] & \\ \|v_K\|_2^2 \sup_{x_0,x_1}|m(x_0,x_1)|^2 & \text{if } m \text{ is bounded.} \end{cases}$$

The result follows by Markov's inequality. The proof with $\widehat{\widetilde{\mathbf{M}}}_K'$ is identical. $\square$

## References

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

——— (1999): "Risk Premia and Term Premia in General Equilibrium," *Journal of Monetary Economics*, 43, 3–33.

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

Andrews, D. W. K. (1991): "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica*, 59, 817–858.

Attanasio, O. P. and G. Weber (1993): "Consumption Growth, the Interest Rate and Aggregation," *Review of Economic Studies*, 60, 631–649.

——— (1995): "Is Consumption Growth Consistent with Intertemporal Optimization? Evidence from the Consumer Expenditure Survey," *Journal of Political Economy*, 103, 1121–1157.

Backus, D., M. Chernov, and I. Martin (2011): "Disasters Implied by Equity Index Options," *Journal of Finance*, 66, 1969–2012.

Backus, D., M. Chernov, and S. Zin (2013): "Sources of Entropy in Representative Agent Models," *Journal of Finance*, forthcoming.

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.

Bandtlow, O. F. (2004): "Estimates for Norms of Resolvents and an Application to the Perturbation of Spectra," *Mathematische Nachrichten*, 267, 3–11.

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

Bansal, R. and S. Viswanathan (1993): "No Arbitrage and Arbitrage Pricing: A New Approach," *Journal of Finance*, 48, 1231–1262.

Bansal, R. and A. Yaron (2004): "Risks for the Long Run: A Potential Resolution of Asset Pricing Puzzles," *Journal of Finance*, 59, 1481–1509.

Beveridge, S. and C. R. Nelson (1981): "A New Approach to Decomposition of Economic Time Series into Permanent and Transitory Components with Particular Attention to Measurement of the Business Cycle," *Journal of Monetary Economics*, 7, 151–174.

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., V. Chernozhukov, S. Lee, and W. K. Newey (2013a): "Local Identification of Nonparametric and Semiparametric Models," *Econometrica*, forthcoming.

Chen, X. and T. M. Christensen (2013): "Optimal Uniform Convergence Rates for Sieve Nonparametric Instrumental Variables Regression," Cowles Foundation Discussion Paper d1923.

Chen, X., J. Favilukis, and S. C. Ludvigson (2013b): "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, Univeristy of Chicago.

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. 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. (2013): "Nonparametric Identification of Positive Eigenfunctions," Working paper, Yale University.

Cochrane, J. H. (1992): "Explaining the Variance of Price-Dividend Ratios," *Review of Financial Studies*, 5, 243–280.

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

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

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

Fleissig, A. R., A. R. Gallant, and J. J. Seater (2000): "Separability, Aggregation and Euler Equation Estimation," *Macroeconomic Dynamics*, 4, 547–572.

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

Gallant, A. R. and G. Tauchen (1989): "Seminonparametric Estimation of Conditionally Constrained Heterogeneous Processes: Asset Pricing Applications," *Econometrica*, 57, 1091–1120.

Gobet, E., M. Hoffmann, and M. Reiss (2004): "Nonparametric Estimation of Scalar Diffusions Based on Low Frequency Data," *Annals of Statistics*, 32, 2223–2253.

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

Greenwood, P. E. and W. Wefelmeyer (1995): "Efficiency of Empirical Estimators for Markov Chains," *Annals of Statistics*, 32, 132–143.

Hansen, L. P. (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 E. Renault (2010): *Encyclopedia of Quantitative Finance*, John Wiley & Sons, Ltd, chap. Pricing Kernels.

Hansen, L. P. and S. F. Richard (1987): "The Role of Conditioning Information in Deducing Testable Restrictions Implied by Dynamic Asset Pricing Models," *Econometrica*, 55, 587–613.

Hansen, L. P. and J. A. Scheinkman (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.

Hansen, L. P. and K. J. Singleton (1982): "Generalized Instrumental Variables Estimation of Nonlinear Rational Expectations Models," *Econometrica*, 50, 1269–1286.

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

Julliard, C. and A. Ghosh (2012): "Can Rare Events Explain the Equity Premium Puzzle?" *Review of Financial Studies*, 25, 3037–3076.

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.

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

Longstaff, F. A. and M. Piazzesi (2004): "Corporate Earnings and the Equity Premium," *Journal of Financial Economics*, 74, 401–421.

Mehra, R. and E. C. Prescott (1985): "The Equity Premium: A Puzzle," *Journal of Monetary Economics*, 15, 145–161.

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.

Newey, W. K. and K. D. West (1987): "A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55, 703–708.

Parzen, E. (1957): "On Consistent Estimates of the Spectrum of a Stationary Time Series," *Annals of Mathematical Statistics*, 28, 329–348.

Peligrad, M. (1985): "An Invariance Principle for Phi-Mixing Sequences," *Annals of Probability*, 13, 1304–1313.

Phillips, P. C. B. (2005): "HAC Estimation by Automated Regression," *Econometric Theory*, 21, 116–142.

Rosenblatt, M. (1971): *Markov Processes. Structure and Asymptotic Behavior*, Springer-Verlag, Berlin.

Ross, S. A. (2013): "The Recovery Theorem," *Journal of Finance*, forthcoming.

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.

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

Sun, Y., P. C. B. Phillips, and S. Jin (2008): "Optimal Bandwidth Selection in Heteroskedasticity-Autocorrelation Robust Testing," *Econometrica*, 76, 175–194.

Wefelmeyer, W. (1999): "Efficient Estimation in Markov Chain Models: An Introduction," in *Asymptotics, Nonparametrics and Time Series*, ed. by S. Ghosh, Dekker, New York, no. 158 in Statistics: Textbooks and Monographs.

Weiss, G. (1975): "Time-Reversibility of Linear Stochastic Processes," *Journal of Applied Probability*, 12, 831–836.