ECONOMETRICA

JOURNAL OF THE ECONOMETRIC SOCIETY

An International Society for the Advancement of Economic Theory in its Relation to Statistics and Mathematics

http://www.econometricsociety.org/

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

NONPARAMETRIC STOCHASTIC DISCOUNT FACTOR DECOMPOSITION

TIMOTHY M. CHRISTENSEN
Department of Economics, New York University

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

NONPARAMETRIC STOCHASTIC DISCOUNT FACTOR DECOMPOSITION

TIMOTHY M. CHRISTENSEN
Department of Economics, New York University

Stochastic discount factor (SDF) processes in dynamic economies admit a permanent-transitory decomposition in which the permanent component characterizes pricing over long investment horizons. This paper introduces an empirical framework to analyze the permanent-transitory decomposition of SDF processes. Specifically, we show how to estimate nonparametrically the solution to the Perron–Frobenius eigenfunction problem of Hansen and Scheinkman (2009). Our empirical framework allows researchers to (i) construct time series of the estimated permanent and transitory components and (ii) estimate the yield and the change of measure which characterize pricing over long investment horizons. We also introduce nonparametric estimators of the continuation value function in a class of models with recursive preferences by reinterpreting the value function recursion as a nonlinear Perron–Frobenius problem. We establish consistency and convergence rates of the eigenfunction estimators and asymptotic normality of the eigenvalue estimator and estimators of related functionals. As an application, we study an economy where the representative agent is endowed with recursive preferences, allowing for general (nonlinear) consumption and earnings growth dynamics.

KEYWORDS: Nonparametric estimation, sieve estimation, stochastic discount factor, permanent-transitory decomposition, nonparametric value function estimation.

1. INTRODUCTION

IN DYNAMIC ASSET PRICING MODELS, stochastic discount factors (SDFs) are stochastic processes that are used to infer equilibrium asset prices. Alvarez and Jermann (2005), Hansen and Scheinkman (2009), and Hansen (2012) introduced and studied a long-term factorization of SDF processes into permanent and transitory components. The *permanent component* is a martingale that induces an alternative probability measure which is used to characterize pricing over long investment horizons. The *transitory component* is related to the return on a discount bond of (asymptotically) long maturity. Alvarez and Jermann (2005) and Bakshi and Chabi-Yo (2012) have found that SDFs must have nontrivial permanent and transitory components in order to explain several salient features of historical returns data. Qin and Linetsky (2017) showed that the permanent-transitory decomposition obtains even in very general semimartingale environments, suggesting that the decomposition is a fundamental feature of arbitrage-free asset pricing models.

The pathbreaking work of Hansen and Scheinkman (2009) links SDF decomposition in Markovian environments to a Perron–Frobenius eigenfunction problem. Specifically, Hansen and Scheinkman (2009) showed that the permanent and transitory components are constructed from the SDF process, the Perron–Frobenius eigenfunction, and

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

This paper is based on my job market paper, which was titled "Estimating the Long-Run Implications of Dynamic Asset Pricing Models" and dated November 8, 2013. I am very grateful to my advisors Xiaohong Chen and Peter Phillips for their advice and support. I would like to thank the co-editor and three anonymous referees for insightful and helpful comments. I have benefited from feedback from Caio Almeida (discussant), Jarda Borovička, Tim Cogley, Jean-Pierre Florens (discussant), Nour Meddahi, Keith O'Hara, Eric Renault, Guillaume Roussellet, Tom Sargent, and participants of seminars at Chicago Booth, Columbia, Cornell, Duke, Indiana, Monash, Montréal, Northwestern, NYU, Penn, Princeton, Rutgers, Sydney, UCL, UNSW, Vanderbilt, Wisconsin, and several conferences. All errors are my own.

 DOI: 10.3982/ECTA11600

its eigenvalue. The eigenvalue determines the average yield on long-horizon payoffs and the eigenfunction characterizes dependence of the price of long-horizon payoffs on the Markov state. The probability measure that is relevant for pricing over long investment horizons may be expressed in terms of the eigenfunction and another eigenfunction that is obtained from a time-reversed Perron–Frobenius problem. See Hansen and Scheinkman (2012, 2017), Backus, Chernov, and Zin (2014), Borovička, Hansen, and Scheinkman (2016), and Qin and Linetsky (2016, 2017) for related theoretical developments.

This paper complements the existing theoretical literature by providing an empirical framework for extracting the permanent and transitory components of SDF processes. We show how to estimate the solution to the Perron–Frobenius eigenfunction problem of Hansen and Scheinkman (2009) from time-series data on state variables and a SDF process. By estimating directly the eigenvalue and eigenfunction, researchers can construct time series of the estimated permanent and transitory components and investigate their properties. The methodology also allows researchers to estimate the yield and the change of measure which characterize pricing over long investment horizons. This approach is fundamentally different from existing empirical methods for studying the permanent-transitory decomposition, which produce bounds on various moments of the permanent and transitory components as functions of asset returns (Alvarez and Jermann (2005), Bakshi and Chabi-Yo (2012), Bakshi, Chabi-Yo, and Gao Bakshi (2015a, 2015b)).[1] Although presented in the context of SDF decomposition, the methodology can be applied to more general processes such as the valuation and stochastic growth processes in Hansen, Heaton, and Li (2008), Hansen and Scheinkman (2009), and Hansen (2012).

The empirical framework is *nonparametric*, that is, it does not place any tight parametric restrictions on the law of motion of state variables or the joint distribution of the state variables and the SDF process. This approach is coherent with the existing literature in which bounds on moments of the permanent and transitory components are derived without placing any parametric restrictions on the joint distribution of the SDF, state variables, and asset returns. This approach is also in line with conventional moment-based estimation methods for asset pricing models, such as GMM (Hansen (1982)) and its various extensions.[2]

In structural macro-finance models, SDF processes (and their permanent and transitory components) are determined by both the preferences of economic agents and the dynamics of state variables. Several works have shown that standard preference and state-process specifications struggle to explain salient features of historical returns data. For instance, Backus, Chernov, and Zin (2014) found that certain specifications appear unable to generate a SDF whose permanent component is large enough to explain historical return premia without also generating unrealistically large spreads between long- and short-term yields. Bakshi and Chabi-Yo (2012) found that historical returns data support positive covariance between the permanent and transitory components, but that positive association cannot be replicated by workhorse models such as the long-run risks model of Bansal and Yaron (2004). Our nonparametric methodology may be used in conjunction with parametric methods to better understand the roles of dynamics and preferences in

[1] Recently, Qin, Linetsky, and Nie (2016) used a complementary parametric approach to recover the permanent component in a parametric term structure model.

[2] Examples include conditional moment-based estimation methodology of Ai and Chen (2003) which has been applied to estimate asset pricing models featuring habits (Chen and Ludvigson (2009)) and recursive preferences (Chen, Favilukis, and Ludvigson (2013)) and the extended method of moments methodology of Gagliardini, Gourieroux, and Renault (2011) which is particularly relevant for derivative pricing.

building models whose permanent and transitory components have empirically realistic properties.

Of course, if state dynamics are treated nonparametrically, then certain forward-looking components, such as the continuation value function under recursive preferences, are not available analytically. We therefore introduce nonparametric estimators of the continuation value function in models with Epstein and Zin (1989) recursive preferences with elasticity of intertemporal substitution (EIS) equal to unity. This class of preferences is used in prominent empirical work, such as Hansen, Heaton, and Li (2008), and may also be interpreted as risk-sensitive preferences as formulated by Hansen and Sargent (1995) (see Tallarini (2000)). We reinterpret the fixed-point problem solved by the value function as a *nonlinear* Perron–Frobenius problem. In so doing, we draw connections with the literature on nonlinear Perron–Frobenius theory following Solow and Samuelson (1953).

As an application, we study an environment similar to that studied by Hansen, Heaton, and Li (2008). We assume a representative agent with Epstein and Zin (1989) recursive preferences and unit EIS. However, instead of modeling consumption and earnings using a homoscedastic Gaussian VAR as in Hansen, Heaton, and Li (2008), we model consumption growth and earnings growth as a general (nonlinear) Markov process. We recover the time series of the SDF process and its permanent and transitory components without assuming any parametric law of motion for the state. The permanent component is large enough to explain historical returns on equities relative to long-term bonds, strongly countercyclical, and highly correlated with the SDF. We also show that the permanent component induces a probability measure that tilts the historical distribution of consumption and earnings growth towards regions of low earnings and consumption growth. To understand better the role of dynamics, we compare properties of the permanent and transitory components extracted nonparametrically with permanent and transitory components implied by two benchmark parametric models for state dynamics, namely, a Gaussian VAR and an AR process with stochastic volatility. We find that for certain values of preference parameters, the nonparametric permanent and transitory components can be positively correlated whereas the permanent and transitory components corresponding to the two parametric models are strongly negatively correlated. Overall, our findings suggest that nonlinear dynamics may have a useful role to play in explaining the long end of the term structure.

The sieve (or projection) estimators of the Perron–Frobenius eigenfunction and eigenvalue that we propose draw heavily on earlier work on nonparametric estimation of Markov diffusions by Chen, Hansen, and Scheinkman (2000) and Gobet, Hoffmann, and Reiß (2004) and are very simple to implement.[3] We also propose sieve estimators of the continuation value function in a class of models with recursive preferences. Implementing the sieve value function estimators requires solving a low-dimensional fixed-point problem for which we propose a computationally simple iterative scheme. Both estimation procedures sidestep nonparametric estimation of the state transition distribution.

The main theoretical contributions of the paper are as follows. First, we establish consistency and convergence rates of the Perron–Frobenius eigenfunction estimators and establish asymptotic normality of the eigenvalue estimator and estimators of related functionals. The large-sample properties are established in sufficient generality that they can

[3]See Darolles, Florens, and Renault (1998), Darolles, Florens, and Gourieroux (2004), and Carrasco, Florens, and Renault (2007) for kernel-based estimation of conditional expectation operators and Lewbel, Linton, and Srisuma (2011) and Escanciano, Hoderlein, Lewbel, Linton, and Srisuma (2015) for kernel-based estimation of marginal utility in nonparametric Euler equation models via eigenfunction methods.

accommodate SDF processes either of a known functional form or containing components that have been first estimated from data (such as preference parameters and continuation value functions). Second, semiparametric efficiency bounds for the eigenvalue and related functionals are derived for the case in which the SDF is of a known functional form and the estimators are shown to attain their bounds. Third, this paper is the first to establish consistency and convergence rates for sieve estimators of the continuation value function for a class of models with recursive preferences. Although the analysis is confined to models in which the state vector is observable, the main theoretical results on eigenfunction and continuation value function estimation apply equally to models in which components of the state are latent.

The remainder of the paper is as follows. Section 2 briefly reviews the theoretical framework in Hansen and Scheinkman (2009) and related literature and discusses the scope of the analysis and identification issues. Section 3 introduces the estimators of the Perron–Frobenius eigenvalue, eigenfunctions, and related functionals and establishes their large-sample properties. Nonparametric continuation value function estimation is studied in Section 4. Section 5 presents a simulation exercise, Section 6 presents the empirical application, and Section 7 concludes. Additional results on estimation and inference are deferred to the Appendix. The Supplemental Material (Christensen (2017)) contains proofs of all results in the main text and sufficient conditions for some assumptions appearing in the main text. An Online Appendix contains additional results on identification, further simulation results, and additional proofs.

2. SETUP

2.1. *Theoretical Framework*

This subsection summarizes the theoretical framework in Alvarez and Jermann (2005), Hansen and Scheinkman (2009) (HS hereafter), Hansen (2012), and Borovička, Hansen, and Scheinkman (2016) (BHS hereafter). We work in discrete time with T denoting the set of nonnegative integers.

In arbitrage-free environments, there is a positive *stochastic discount factor* process $M = \{M_t : t \in T\}$ that satisfies

$$\mathbb{E}\left[\frac{M_{t+\tau}}{M_t} R_{t,t+\tau} \middle| \mathcal{I}_t\right] = 1, \tag{1}$$

where $R_{t,t+\tau}$ is the (gross) return on a traded asset over the period from t to $t+\tau$, $\mathcal{I}_t$ denotes the information available to all investors at date t, and $\mathbb{E}[\cdot]$ denotes expectation with respect to investors' beliefs (see, e.g., Hansen and Renault (2010)). Throughout this paper, we impose rational expectations by assuming that investors' beliefs agree with the data-generating probability measure.

Alvarez and Jermann (2005) introduced the *permanent-transitory decomposition*:

$$\frac{M_{t+\tau}}{M_t} = \frac{M^P_{t+\tau}}{M^P_t} \frac{M^T_{t+\tau}}{M^T_t}. \tag{2}$$

The permanent component $M^P_{t+\tau}/M^P_t$ is a martingale: $\mathbb{E}[M^P_{t+\tau}/M^P_t | \mathcal{I}_t] = 1$ (almost surely). HS showed that the martingale induces an alternative probability measure which is used to characterize pricing over long investment horizons. The transitory component $M^T_{t+\tau}/M^T_t$ is the reciprocal of the return to holding a discount bond of (asymptotically) long maturity from date t to date $t+\tau$. Alvarez and Jermann (2005) provided conditions under which the permanent and transitory components exist. Qin and Linetsky (2017) showed that the decomposition obtains in very general semimartingale environments.

To formally introduce the framework in HS and BHS, consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ on which there is a time-homogeneous, strictly stationary, and ergodic Markov process $X = \{X_t : t \in T\}$ taking values in $\mathcal{X} \subseteq \mathbb{R}^d$. Let Q denote the stationary distribution of X. Let $\{\mathcal{F}_t : t \in T\} \subseteq \mathcal{F}$ be the filtration generated by the histories of X. It is assumed that X_t summarizes all information relevant for asset pricing at date t. When we consider payoffs depending only on future values of the state and allow trading at intermediate dates, we may assume the SDF process is a *positive multiplicative functional* of X. That is, M_t is adapted to $\mathcal{F}_t$, $M_t > 0$ for each $t \in T$ (almost surely) and

$$\frac{M_{t+\tau}}{M_t} = M_\tau(\theta_t)$$

with $\theta_t : \Omega \to \Omega$ the time-shift operator given by $X_\tau(\theta_t(\omega)) = X_{t+\tau}(\omega)$ for each $\tau, t \in T$ (see Section 2 of HS). Thus, M_τ is a function of $X_0, \dots, X_\tau$ and $M_\tau(\theta_t)$ is the same function applied to $X_t, \dots, X_{t+\tau}$. In particular, we have

$$\frac{M_{t+1}}{M_t} = m(X_t, X_{t+1}) \tag{3}$$

for some positive function m. For convenience, we occasionally refer to m as the SDF.

Given the Markovianity of X, we may define a one-period pricing operator $\mathbb{M}$ which assigns date-t prices to state-dependent payoffs at date $t+1$. That is, the date-t price of a claim to the payoff $\psi(X_{t+1})$ at time $t+1$ may be written as

$$\mathbb{M}\psi(x) = \mathbb{E}\big[m(X_t, X_{t+1})\psi(X_{t+1}) | X_t = x\big]. \tag{4}$$

Pricing operators may be defined analogously for payoff horizons $\tau \geq 1$. The operator $\mathbb{M}_\tau$ assigning date-t prices to date-$(t+\tau)$ payoffs $\psi(X_{t+\tau})$ is given by

$$\mathbb{M}_\tau\psi(x) = \mathbb{E}\left[\frac{M_{t+\tau}}{M_t}\psi(X_{t+\tau})\middle| X_t = x\right]. \tag{5}$$

It follows by Markovianity of the state and the multiplicative functional property of the SDF process that $\mathbb{M}_\tau = \mathbb{M}^\tau$ (i.e., $\mathbb{M}$ applied τ times) for each $\tau \geq 1$. Therefore, it suffices to study the one-period operator $\mathbb{M}$.

HS introduced and studied the Perron–Frobenius eigenfunction problem

$$\mathbb{M}\phi(x) = \rho\phi(x), \tag{6}$$

where the eigenvalue ρ is a positive scalar and the eigenfunction ϕ is positive.[4] Classical, finite-dimensional Perron–Frobenius theory says that a positive matrix has positive right and left eigenvectors corresponding to its spectral radius.[5] The Kreĭn and Rutman (1950) theorem and its well-known extensions generalize finite-dimensional Perron–Frobenius theory to infinite-dimensional Banach spaces. To introduce formally the left eigenfunction of $\mathbb{M}$ corresponding to ρ, we use a time-reversed version of the Perron–Frobenius

[4]We say a function is positive (nonnegative) if it is positive (nonnegative) Q-almost everywhere.
[5]See, for example, Theorem 8.2.2 in Horn and Johnson (2013).

problem (6). Recall that a first-order Markov process seen in reverse time is also a first-order Markov process (see Rosenblatt (1971, p. 4)). Define the time-reversed operator

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

In what follows, we will assume that $\mathbb{M}$ is a bounded linear operator on the Hilbert space $L^2 = \{\psi : \mathcal{X} \to \mathbb{R}$ such that $\int \psi^2 \,\mathrm{d}Q < \infty\}$ in which case $\mathbb{M}^*$ is defined formally as the adjoint of $\mathbb{M}$. The time-reversed Perron–Frobenius problem is

$$\mathbb{M}^*\phi^*(x) = \rho\phi^*(x), \tag{7}$$

where ρ is the eigenvalue from (6) and the eigenfunction ϕ^* is positive.

Given ρ and ϕ which solve the Perron–Frobenius problem (6), HS defined

$$\frac{M^P_{t+\tau}}{M^P_t} = \rho^{-\tau}\frac{M_{t+\tau}}{M_t}\frac{\phi(X_{t+\tau})}{\phi(X_t)}, \qquad \frac{M^T_{t+\tau}}{M^T_t} = \rho^{\tau}\frac{\phi(X_t)}{\phi(X_{t+\tau})}. \tag{8}$$

It follows from (6) that $\mathbb{E}[M^P_{t+\tau}/M^P_t|\mathcal{F}_t] = 1$ (almost surely) for each $\tau, t \in T$. HS showed that although there may exist multiple solutions to the Perron–Frobenius problem, only one solution leads to processes M^P and M^T that may be interpreted correctly as permanent and transitory components. Such a solution has a martingale term that induces a change of measure under which X is *stochastically stable*; see Condition 3 in BHS for sufficient conditions. Loosely speaking, stochastic stability requires that conditional expectations under the distorted probability measure converge (as the horizon increases) to an unconditional expectation $\widetilde{\mathbb{E}}[\cdot]$. The expectation $\widetilde{\mathbb{E}}[\cdot]$ will typically be different from the expectation $\mathbb{E}[\cdot]$ associated with the stationary distribution of X. Under stochastic stability, the one-factor representation

$$\lim_{\tau\to\infty} \rho^{-\tau}\mathbb{M}_\tau\psi(x) = \widetilde{\mathbb{E}}\left[\frac{\psi(X_t)}{\phi(X_t)}\right]\phi(x) \tag{9}$$

holds for each ψ for which $\widetilde{\mathbb{E}}[\psi(X_t)/\phi(X_t)]$ is finite (see, e.g., Proposition 7.1 in HS). When a long-run approximation like (9) holds, we may interpret $M^P_{t+\tau}/M^P_t$ and $M^T_{t+\tau}/M^T_t$ from (8) as the permanent and transitory components of the SDF process. Moreover, the scalar $-\log\rho$ may be interpreted as the *long-run yield*. The long-run approximation (9) also shows that ϕ captures state dependence of long-horizon asset prices. The function ϕ^* is itself of interest as it will play a role in characterizing the expectation $\widetilde{\mathbb{E}}[\cdot]$ and will also appear in the asymptotic variance of estimators of ρ.

The theoretical framework of HS may be used to characterize properties of the permanent and transitory components analytically by solving the Perron–Frobenius eigenfunction problem. Below, we describe an empirical framework to estimate the eigenvalue ρ and eigenfunctions ϕ and ϕ^* from time-series data on X and a SDF process.

2.2. *Scope of the Analysis*

The Markov state vector X_t is assumed throughout to be observable to the econometrician. However, we do not constrain the transition law of X to be of any parametric form. This approach is similar to that taken by Gagliardini, Gourieroux, and Renault (2011), who also presumed the existence of a stationary, time-homogeneous Markov state process that is observable to the econometrician but did not constrain its transition law to be of any parametric form.

We assume the SDF function m is either observable or known up to some parameter which is first estimated from data on X (and possibly asset returns).

Case 1: SDF is observable. Here the functional form of m is known ex ante. For example, consider a representative agent model with power utility in which the time preference parameter β and risk aversion parameter γ are both pre-specified by the researcher. Here we would simply take $m(X_t, X_{t+1}) = \beta G_{t+1}^{-\gamma}$ provided consumption growth G_{t+1} is of the form $G_{t+1} = G(X_t, X_{t+1})$ for some known function G.

Case 2: SDF is estimated. In this case, we assume that $m(X_t, X_{t+1}; \alpha_0)$ where the functional form of m is known up to a parameter α_0, which could be of several forms:

- A finite-dimensional vector of preference parameters in structural models (e.g., Hansen and Singleton (1982) and Hansen, Heaton, and Yaron (1996)) or risk-premium parameters in reduced-form models (e.g., Gagliardini, Gourieroux, and Renault (2011)).
- A vector of parameters θ_0 together with a function h_0, so $\alpha_0 = (\theta_0, h_0)$. One example is models with recursive preferences, as the continuation value function is not known when state dynamics are modeled nonparametrically (see Chen, Favilukis, and Ludvigson (2013) and the application in Section 6). For such models, θ_0 would consist of discount, risk aversion, and intertemporal substitution parameters and h_0 would be the continuation value function. Another example is semiparametric habit formation models, where θ_0 would consist of time discount and homogeneity parameters and h_0 would be a nonparametric internal or external habit formation component (see Chen and Ludvigson (2009)).
- We could also take α_0 to be m itself, in which case $\hat{\alpha}$ would be a nonparametric estimate of the SDF. Prominent examples include Bansal and Viswanathan (1993), Aït-Sahalia and Lo (1998), and Rosenberg and Engle (2002).

In Case 2, we consider a two-step approach to SDF decomposition. In the first step, α_0 is estimated from time-series data on the state and possibly also asset returns. In the second step, we plug the first-stage estimator $\hat{\alpha}$ into the nonparametric procedure to estimate ρ, ϕ, ϕ^*, and related quantities.

2.3. *Identification*

In this section, we present some sufficient conditions that ensure there is a unique solution to the Perron–Frobenius problems (6) and (7). The conditions also ensure that a long-run approximation of the form (9) holds. Therefore, the resulting M^P and M^T constructed from ρ and ϕ as in (8) may be interpreted correctly as the permanent and transitory components. For estimation, all that we require is for the conclusions of Proposition 2.1 below to hold. Therefore, the following conditions could be replaced by other sets of sufficient conditions.

HS and BHS established very general identification, existence, and long-run approximation results using Markov process theory. The operator-theoretic conditions that we use are more restrictive than the conditions in HS and BHS, but they are convenient for deriving the large-sample theory that follows. In particular, the conditions ensure certain continuity properties of ρ, ϕ, and ϕ^* with respect to perturbations of the operator $\mathbb{M}$. Our results are derived for the specific parameter (function) space that is relevant for estimation, whereas the results in HS and BHS apply to a larger class of functions. Connections between our conditions and the conditions in HS and BHS are discussed in detail in the Online Appendix, which also treats separately the issues of identification, existence, and long-run approximation.

We take the cone of all positive functions in L^2 as the parameter space for ϕ. Let $\|\cdot\|$ and $\langle\cdot,\cdot\rangle$ denote the L^2 norm and inner product. We say that $\mathbb{M}$ is *bounded* if $\|\mathbb{M}\| :=$

$\sup\{\|\mathbb{M}\psi\| : \psi \in L^2, \|\psi\| = 1\} < \infty$ and *compact* if $\mathbb{M}$ maps bounded sets into relatively compact sets. Finally, let $Q \otimes Q$ denote the product measure on $\mathcal{X}^2$.

ASSUMPTION 2.1: *Let the operators* $\mathbb{M}$ *in display* (4) *and* $\mathbb{M}_\tau$ *in display* (5) *satisfy the following*:

(a) $\mathbb{M}$ *is a bounded linear operator of the form*

$$\mathbb{M}\psi(x_t) = \int \mathcal{K}_m(x_t, x_{t+1})\psi(x_{t+1})\,\mathrm{d}Q(x_{t+1})$$

for some $\mathcal{K}_m : \mathcal{X}^2 \to \mathbb{R}$ *that is positive* ($Q \otimes Q$*-almost everywhere*),

(b) $\mathbb{M}_\tau$ *is compact for some* $\tau \geq 1$.

Discussion of Assumptions: Part (a) are mild boundedness and positivity conditions. If the unconditional density $f(x_t)$ and the transition density $f(x_{t+1}|x_t)$ of X exist, then $\mathcal{K}_m$ is of the form

$$\mathcal{K}_m(x_t, x_{t+1}) = m(x_t, x_{t+1})\frac{f(x_{t+1}|x_t)}{f(x_{t+1})}.$$

Part (b) is weaker than requiring $\mathbb{M}$ to be compact. A sufficient condition for compactness of $\mathbb{M}$ is the Hilbert–Schmidt condition $\int\int \mathcal{K}_m(x_t, x_{t+1})^2\,\mathrm{d}Q(x_t)\,\mathrm{d}Q(x_{t+1}) < \infty$.

To introduce the identification result, let $\sigma(\mathbb{M}) \subset \mathbb{C}$ denote the spectrum of $\mathbb{M}$.[6] We say that ρ is *simple* if it has a unique eigenfunction (up to scale) and *isolated* if there exists a neighborhood N of ρ such that $\sigma(\mathbb{M}) \cap N = \{\rho\}$. As ϕ and ϕ^* are defined up to scale, we say that ϕ and ϕ^* are unique if they are unique up to scale. Normalizing ϕ and ϕ^* so that $\mathbb{E}[\phi(X_t)\phi^*(X_t)] = 1$, we may define a probability measure $\widetilde{Q}$ that is absolutely continuous with respect to Q by the change of measure:

$$\frac{\mathrm{d}\widetilde{Q}}{\mathrm{d}Q} = \phi\phi^*. \tag{10}$$

Finally, let $\widetilde{\mathbb{E}}[\cdot]$ denote expectation under $\widetilde{Q}$, that is, for any indicator function χ, we have

$$\widetilde{\mathbb{E}}\big[\chi(X_t)\big] = \mathbb{E}\big[\chi(X_t)\phi(X_t)\phi^*(X_t)\big], \tag{11}$$

where the expectation on the right-hand side is taken under the stationary distribution Q.

PROPOSITION 2.1: *Let Assumption* 2.1 *hold*. *Then*:

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

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

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

(d) *The representation* (9) *holds for all* $\psi \in L^2$ *with* $\widetilde{\mathbb{E}}[\cdot]$ *defined in* (11).

Parts (a) and (b) are existence and identification results, respectively. This is a well-known extension of the classical Krein–Rutman theorem Schaefer (1974, Theorems V.5.2

[6]See, for example, Dunford and Schwartz (1958, Chapter VII.3) for definitions.

and V.6.6). Recently, similar operator-theoretic results have been applied to study identification in semi/nonparametric Euler equation models (see Escanciano and Hoderlein (2012), Lewbel, Linton, and Srisuma (2011), Chen, Chernozhukov, Lee, and Newey (2014), and Escanciano et al. (2015)). Identification under slightly weaker but related conditions was studied in Christensen (2015).

Part (c) guarantees that ρ is isolated and simple, which is used extensively in the derivation of the large-sample theory. Part (d) says that ρ and ϕ are the relevant eigenvalue-eigenfunction pair for constructing the permanent and transitory components and links the expectation $\widetilde{\mathbb{E}}$ to ϕ^*. Note, in particular, that $\widetilde{\mathbb{E}}[\psi(X_t)/\phi(X_t)] = \mathbb{E}[\psi(X_t)\phi^*(X_t)]$. Estimating ϕ and ϕ^* therefore allows one to estimate the Radon–Nikodym derivative of $\widetilde{Q}$ with respect to Q.

3. ESTIMATION

This section introduces the estimators of the Perron–Frobenius eigenvalue ρ and eigenfunctions ϕ and ϕ^* and presents the large-sample properties of the estimators.

3.1. *Sieve Estimation*

We follow Chen, Hansen, and Scheinkman (2000) and Gobet, Hoffmann, and Reiß (2004) in using a sieve approach in which the infinite-dimensional eigenfunction problem is approximated by a low-dimensional matrix eigenvector problem. Let $b_{k1}, \dots, b_{kk} \in L^2$ be a dictionary of linearly independent basis functions (e.g., polynomials, splines, wavelets, or a Fourier basis) and let $B_k \subset L^2$ denote the linear subspace spanned by $b_{k1}, \dots, b_{kk}$. The sieve dimension $k < \infty$ is a smoothing parameter chosen by the econometrician and should increase with the sample size.

To describe the approximation, let $\Pi_k : L^2 \to B_k$ denote the orthogonal projection onto B_k. Consider the projected eigenfunction problem

$$(\Pi_k \mathbb{M})\phi_k = \rho_k \phi_k, \tag{12}$$

where ρ_k is the largest real eigenvalue of $\Pi_k\mathbb{M}$ and $\phi_k : \mathcal{X} \to \mathbb{R}$ is its eigenfunction. Under regularity conditions, the problem (12) has a unique solution for all k large enough (see Lemma A.1). As the function ϕ_k belongs to the space B_k, we have that $\phi_k(x) = b^k(x)'c_k$ for a vector $c_k \in \mathbb{R}^k$, where $b^k(x) = (b_{k1}(x), \dots, b_{kk}(x))'$. The eigenfunction problem (12) may be written in matrix form as

$$\mathbf{G}_k^{-1}\mathbf{M}_k c_k = \rho_k c_k,$$

where the $k \times k$ matrices $\mathbf{G}_k$ and $\mathbf{M}_k$ are given by

$$\mathbf{G}_k = \mathbb{E}\big[b^k(X_t)b^k(X_t)'\big], \tag{13}$$

$$\mathbf{M}_k = \mathbb{E}\big[b^k(X_t)m(X_t, X_{t+1})b^k(X_{t+1})'\big], \tag{14}$$

and where ρ_k is the largest real eigenvalue of $\mathbf{G}_k^{-1}\mathbf{M}_k$ and c_k is its eigenvector (we assume throughout that $\mathbf{G}_k$ is nonsingular). We refer to $\phi_k(x) = b^k(x)'c_k$ as the *approximate solution* for ϕ. The approximate solution for ϕ^* is $\phi_k^*(x) = b^k(x)'c_k^*$, where c_k^* is the eigenvector of $\mathbf{G}_k^{-1}\mathbf{M}_k'$ corresponding to ρ_k. Together, (ρ_k, c_k, c_k^*) solve the generalized eigenvector problem

$$\mathbf{M}_k c_k = \rho_k \mathbf{G}_k c_k, \qquad c_k^{*\prime}\mathbf{M}_k = \rho_k c_k^{*\prime}\mathbf{G}_k, \tag{15}$$

where ρ_k is the largest real generalized eigenvalue of the pair $(\mathbf{M}_k, \mathbf{G}_k)$. We suppress dependence of $\mathbf{M}_k$ and $\mathbf{G}_k$ on k hereafter to simplify notation.

To estimate ρ, ϕ, and ϕ^*, we solve the sample analogue of (15), namely,

$$\widehat{\mathbf{M}}\hat{c} = \hat{\rho}\widehat{\mathbf{G}}\hat{c}, \qquad \hat{c}^{*\prime}\widehat{\mathbf{M}} = \hat{\rho}\hat{c}^{*\prime}\widehat{\mathbf{G}}, \tag{16}$$

where $\widehat{\mathbf{M}}$ and $\widehat{\mathbf{G}}$ are defined below and where $\hat{\rho}$ is the maximum real generalized eigenvalue of the matrix pair $(\widehat{\mathbf{M}}, \widehat{\mathbf{G}})$. The estimators $\hat{\rho}$, $\hat{c}$, and $\hat{c}^*$ may be computed simultaneously using, for example, the `eig` function in Matlab. The estimators of ϕ and ϕ^* are

$$\hat{\phi}(x) = b^k(x)'\hat{c}, \qquad \hat{\phi}^*(x) = b^k(x)'\hat{c}^*.$$

Under the regularity conditions below, the eigenvalue $\hat{\rho}$ and its right- and left-eigenvectors $\hat{c}$ and $\hat{c}^*$ will be unique with probability approaching 1 (see Lemma A.3).

Given a time series of data $\{X_0, X_1, \dots, X_n\}$, a natural estimator of $\mathbf{G}$ is

$$\widehat{\mathbf{G}} = \frac{1}{n}\sum_{t=0}^{n-1} b^k(X_t)b^k(X_t)'. \tag{17}$$

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

Case 1: SDF is observable. First, consider the case in which the function $m(X_t, X_{t+1})$ is specified by the researcher. In this case,

$$\widehat{\mathbf{M}} = \frac{1}{n}\sum_{t=0}^{n-1} b^k(X_t)m(X_t, X_{t+1})b^k(X_{t+1})'. \tag{18}$$

Case 2: SDF is estimated. Now suppose that the SDF is of the form $m(X_t, X_{t+1}; \alpha_0)$, where the functional form of m is known up to the parameter α_0 which is to be estimated first from the data on X and possibly also asset returns. Let $\hat{\alpha}$ denote this first-stage estimator. In this case,

$$\widehat{\mathbf{M}} = \frac{1}{n}\sum_{t=0}^{n-1} b^k(X_t)m(X_t, X_{t+1}; \hat{\alpha})b^k(X_{t+1})'. \tag{19}$$

3.1.1. *Other Functionals*

Recall that the *long-run yield* is $y \equiv -\log\rho$. We may estimate y using

$$\hat{y} = -\log\hat{\rho}. \tag{20}$$

Another functional of interest is the *entropy* of the permanent component, namely, $L \equiv \log\mathbb{E}[M^P_{t+1}/M^P_t] - \mathbb{E}[\log(M^P_{t+1}/M^P_t)]$, which is bounded from below by the expected excess return of any traded asset relative to a discount bond of (asymptotically) long maturity (see Alvarez and Jermann (2005, Proposition 2)). Previous empirical work has estimated this bound from data on equity returns and proxies for holding period returns on long-maturity discount bonds (see, e.g., Alvarez and Jermann (2005) and Bakshi and Chabi-Yo (2012)). Here we take a complementary approach by assuming the SDF process is identifiable and estimate the entropy of its permanent component directly.

In Markovian environments, the entropy has the simple form $L = \log\rho - \mathbb{E}[\log m(X_t, X_{t+1})]$ (see Hansen (2012) and Backus, Chernov, and Zin (2014)). Given $\hat{\rho}$, a natural estimator of L is

$$\hat{L} = \log\hat{\rho} - \frac{1}{n}\sum_{t=0}^{n-1}\log m(X_t, X_{t+1}) \tag{21}$$

in Case 1; in Case 2, we replace $m(X_t, X_{t+1})$ by $m(X_t, X_{t+1}; \hat{\alpha})$ in (21). The size of the permanent component may also be measured by other types of statistical discrepancies besides entropy (e.g., Cressie–Read divergences) which may be computed from the time series of the permanent component recovered empirically using $\hat{\rho}$ and $\hat{\phi}$. We confine our attention to entropy because the theoretical literature has typically used entropy to measure the size of SDFs and their permanent components over different horizons (see, e.g., Hansen (2012) and Backus, Chernov, and Zin (2014)) and for sake of comparison with the empirical literature on bounds.

3.2. *Consistency and Convergence Rates*

Here we establish consistency of the estimators and derive the convergence rates of the eigenfunction estimators under mild regularity conditions.

ASSUMPTION 3.1: $\mathbb{M}$ *is bounded and conclusions* (a)–(d) *of Proposition* 2.1 *hold*.

ASSUMPTION 3.2: $\|\Pi_k\mathbb{M} - \mathbb{M}\| = o(1)$.

Let $\mathbf{G}^{-1/2}$ denote the inverse of the positive definite square root of $\mathbf{G}$ and let $\mathbf{I}$ denote the $k \times k$ identity matrix. Define the "orthogonalized" matrices $\mathbf{M}^o = \mathbf{G}^{-1/2}\mathbf{M}\mathbf{G}^{-1/2}$, $\widehat{\mathbf{G}}^o = \mathbf{G}^{-1/2}\widehat{\mathbf{G}}\mathbf{G}^{-1/2}$, and $\widehat{\mathbf{M}}^o = \mathbf{G}^{-1/2}\widehat{\mathbf{M}}\mathbf{G}^{-1/2}$. Let $\|\cdot\|$ also denote the Euclidean norm when applied to vectors and the operator norm (largest singular value) when applied to matrices. Note that $\widehat{\mathbf{G}}^o$ and $\widehat{\mathbf{M}}^o$ are a proof device and do not need to be calculated in practice.

ASSUMPTION 3.3: $\|\widehat{\mathbf{G}}^o - \mathbf{I}\| = o_p(1)$ *and* $\|\widehat{\mathbf{M}}^o - \mathbf{M}^o\| = o_p(1)$.

Discussion of Assumptions: Assumption 3.2 requires that the space B_k be chosen such that it approximates well the range of $\mathbb{M}$ (as $k \to \infty$). This assumption also implicitly requires that $\mathbb{M}$ is compact, as has been assumed previously in the literature on sieve estimation of eigenfunctions (see, e.g., Gobet, Hoffmann, and Reiß (2004)).[7] Assumption 3.3 ensures that the sampling error in estimating $\mathbf{G}^{-1}\mathbf{M}$ vanishes asymptotically. This condition implicitly restricts the maximum rate at which k can grow with n. Appendix C.1 in the Supplemental Material (Christensen (2017)) presents some sufficient conditions for Assumption 3.3.

Before presenting the main result on convergence rates, we first introduce sequences of constants that bound the approximation bias and sampling error. As eigenfunctions are only normalized up to scale, impose the normalizations $\|\phi\| = 1$ and $\|\phi^*\| = 1$. Define

$$\delta_k = \|\Pi_k\phi - \phi\| \quad \text{and} \quad \delta_k^* = \|\Pi_k\phi^* - \phi^*\|. \tag{22}$$

[7] If $\mathbb{M}$ is not compact but $\mathbb{M}_\tau$ is compact for some $\tau \geq 2$, then one can apply the estimators to $\mathbb{M}_\tau$ in place of $\mathbb{M}$ and estimate the solution (ρ^τ, ϕ) to $\mathbb{M}_\tau\phi = \rho^\tau\phi$ and similarly for ϕ^*. Large-sample properties of the estimators of ρ^τ, ϕ, and ϕ^* would then follow directly from Theorems 3.1–3.5.

Here δ_k and δ_k^* measure the bias incurred by approximating ϕ and ϕ^* by elements of B_k. Bounds for δ_k and δ_k^* are available for commonly used bases when ϕ and ϕ^* belong a Hölder, Sobolev, or Besov class (see, e.g., Chen (2007)). Let $\tilde{c}_k = \mathbf{G}^{1/2} c_k$ and $\tilde{c}_k^* = \mathbf{G}^{1/2} c_k^*$ and normalize c_k and c_k^* so that $\|\tilde{c}_k\| = \|\tilde{c}_k^*\| = 1$ (this is equivalent to $\|\phi_k\| = \|\phi_k^*\| = 1$). Under Assumption 3.3, we may choose positive sequences $\eta_{n,k}$ and $\eta_{n,k}^*$ which are both $o(1)$, so that

$$\big\|\big((\widehat{\mathbf{G}}^o)^{-1}\widehat{\mathbf{M}}^o - \mathbf{M}^o\big)\tilde{c}_k\big\| = O_p(\eta_{n,k}) \quad \text{and} \quad \big\|\big((\widehat{\mathbf{G}}^o)^{-1}\widehat{\mathbf{M}}^{o\prime} - \mathbf{M}^{o\prime}\big)\tilde{c}_k^*\big\| = O_p(\eta_{n,k}^*). \tag{23}$$

Appendix C.1 presents bounds on $\eta_{n,k}$ and $\eta_{n,k}^*$.

THEOREM 3.1: *Let Assumptions 3.1–3.3 hold. Then*:
(a) $|\hat{\rho} - \rho| = O_p(\delta_k + \eta_{n,k})$,
(b) $\|\hat{\phi} - \phi\| = O_p(\delta_k + \eta_{n,k})$,
(c) $\|\hat{\phi}^* - \phi^*\| = O_p(\delta_k^* + \eta_{n,k}^*)$,
where δ_k and δ_k^* are defined in (22) and $\eta_{n,k}$ and $\eta_{n,k}^*$ are defined in (23). The convergence rates for $\hat{\phi}$ and $\hat{\phi}^*$ should be understood to hold under the scale normalizations $\|\phi\| = 1$, $\|\hat{\phi}\| = 1$, $\|\phi^*\| = 1$, and $\|\hat{\phi}^*\| = 1$ and sign normalizations $\langle \phi, \hat{\phi}\rangle \geq 0$ and $\langle \phi^*, \hat{\phi}^*\rangle \geq 0$.

It is worth noting that Theorem 3.1 holds for $\hat{\rho}$, $\hat{\phi}$, and $\hat{\phi}^*$ calculated from any estimators $\widehat{\mathbf{G}}$ and $\widehat{\mathbf{M}}$ that satisfy Assumption 3.3. Indeed, Theorem 3.1 is sufficiently general that it applies to models with latent state vectors without modification: all that is required is that one can construct estimators of $\mathbf{G}$ and $\mathbf{M}$ that satisfy Assumption 3.3.

Theorem 3.1 displays the usual bias-variance tradeoff encountered in nonparametric estimation. The bias terms δ_k and δ_k^* will be decreasing in k (since ϕ and ϕ^* are approximated over increasingly rich subspaces as k increases). On the other hand, the variance terms $\eta_{n,k}$ and $\eta_{n,k}^*$ will typically be increasing in k (larger matrices) and decreasing in n (more data). Choosing k to balance the bias and variance terms will yield the best convergence rate. As an illustration, we now establish the convergence rates of $\hat{\phi}$ and $\hat{\phi}^*$ in Case 1, where $\widehat{\mathbf{G}}$ and $\widehat{\mathbf{M}}$ are as in (17) and (18), under standard conditions from the statistics literature on optimal convergence rates. Although the following conditions are not necessarily appropriate in an asset pricing context, the result is informative about the convergence properties of $\hat{\phi}$ and $\hat{\phi}^*$. Let $W^p = \{f \in L^2 : \sum_{|a|\leq p} \|D^a f\| < \infty\}$ with $D^a f = \frac{\partial^{a_1+\cdots+a_d}}{\partial^{a_1} x_1 \cdots \partial^{a_d} x_d} f$ and $|a| = a_1 + \cdots + a_d$ denote a Sobolev space of smoothness $p \in \mathbb{N}$ equipped with the norm $\|f\|_{W^p} = \sum_{|a|\leq p} \|D^a f\|$.

COROLLARY 3.1: *Let Assumption 3.1 and the following conditions hold*: (i) $\mathcal{X} \subset \mathbb{R}^d$ *is compact and rectangular*; (ii) Q *has a continuous density bounded away from zero*; (iii) $\phi, \phi^* \in W^p$ *and* $\mathbb{M}$ *is a bounded linear operator from* L^2 *into* $W^{\bar{p}}$ *for some* $p \geq \bar{p} > 0$; (iv) B_k *is spanned by tensor-product B-splines of order* $\nu > p$ *with equally spaced interior knots*; (v) $\mathbb{E}[m(X_0, X_1)^r] < \infty$ *for some* $r > 2$; (vi) $k^{2+2/r}/n = o(1)$; (vii) X *is exponentially rho-mixing. Then*: *Assumptions 3.2 and 3.3 hold and we may take* $\delta_k, \delta_k^* = O(k^{-p/d})$, *and* $\eta_{n,k}, \eta_{n,k}^* = O(k^{(r+2)/2r}/\sqrt{n})$. *Choosing* $k \asymp n^{\frac{rd}{2rp+(2+r)d}}$ *yields*

$$\|\hat{\phi} - \phi\| = O_p\big(n^{-\frac{rp}{2rp+(2+r)d}}\big), \qquad \big\|\hat{\phi}^* - \phi^*\big\| = O_p\big(n^{-\frac{rp}{2rp+(2+r)d}}\big).$$

If m is bounded, the rates become $n^{-p/(2p+d)}$ which is the optimal L^2-norm rate for nonparametric regression estimators when the regression function belongs to W^p.

Sieve methods may also be used to numerically compute ρ, ϕ, and ϕ^* in models for which analytical solutions are unavailable. For such models, the matrices $\mathbf{M}$ and $\mathbf{G}$ may be computed directly (e.g., via simulation or numerical integration) and ρ_k, ϕ_k, and ϕ_k^* can be obtained by solving (15). Lemma A.2 gives the rates $|\rho_k - \rho| = O(\delta_k)$, $\|\phi_k - \phi\| = O(\delta_k)$, and $\|\phi_k^* - \phi^*\| = O(\delta_k^*)$.

We close this subsection with a remark relating δ_k and δ_k^* under an additional condition on the sieve basis B_k. Assumption 3.2 implies that $\mathbb{M}$ is compact. Therefore, $\mathbb{M}$ has a singular value decomposition $\{(\mu_n, \varphi_n, g_n) : n \in \mathbb{N}\}$ where $\{\mu_n : n \in \mathbb{N}\}$ are the singular values of $\mathbb{M}$ arranged in non-increasing order (i.e., $\mu_n \geq \mu_{n+1} \searrow 0$) and $\{\varphi_n : n \in \mathbb{N}\}$ and $\{g_n : n \in \mathbb{N}\}$ are orthonormal bases for L^2 with $\mathbb{M}\varphi_n = \mu_n g_n$ and $\mathbb{M}^* g_n = \mu_n \varphi_n$ for each $n \in \mathbb{N}$ (see, e.g., Chapter 15.4 in Kress (1989)).

REMARK 3.1: Let Assumption 3.2 hold and let B_k span the linear subspaces generated by $\{\varphi_n : 1 \leq n \leq k\}$ and $\{g_n : 1 \leq n \leq k\}$. Then δ_k and δ_k^* are both $O(\mu_{k+1})$.

For example, if X is a scalar Gaussian AR(1), $m(X_t, X_{t+1})$ is exponentially affine in (X_t, X_{t+1}), and the basis functions are Hermite polynomials, then δ_k and δ_k^* are $O(e^{-ck})$ for some $c > 0$. Similar spanning assumptions are often made in the literature on sieve estimation of nonparametric instrumental variables models (see, e.g., Blundell, Chen, and Kristensen (2007)).

3.3. *Asymptotic Normality*

In this section, we establish the asymptotic normality of $\hat{\rho}$. The semiparametric efficiency bound in Case 1 is also derived and $\hat{\rho}$ is shown to be efficient in this case. Related results on asymptotic normality and semiparametric efficiency of the estimator of the entropy of the permanent component are presented in Appendix B.

3.3.1. *Asymptotic Normality in Case 1*

To establish asymptotic normality of $\hat{\rho}$, we derive the representation

$$\sqrt{n}(\hat{\rho} - \rho) = \frac{1}{\sqrt{n}} \sum_{t=0}^{n-1} \psi_\rho(X_t, X_{t+1}) + o_p(1), \tag{24}$$

where the influence function ψ_ρ is given by

$$\psi_\rho(x_0, x_1) = \phi^*(x_0) m(x_0, x_1) \phi(x_1) - \rho \phi^*(x_0) \phi(x_0), \tag{25}$$

with ϕ and ϕ^* normalized so that $\|\phi\| = 1$ and $\langle \phi, \phi^* \rangle = 1$. The process $\{\psi_\rho(X_t, X_{t+1}) : t \in T\}$ is a martingale difference sequence (relative to the filtration $\{\mathcal{F}_t : t \in T\}$). Therefore, the asymptotic distribution of $\hat{\rho}$ follows from (24) by a central limit theorem for martingale differences. To formalize this argument, we make the following assumption.

ASSUMPTION 3.4: *Let the following hold*:
(a) $\delta_k = o(n^{-1/2})$ *and* $\delta_k^* = o(n^{-1/2})$,
(b) $\|\widehat{\mathbf{G}}^o - \mathbf{I}\| = o_p(n^{-1/4})$ *and* $\|\widehat{\mathbf{M}}^o - \mathbf{M}^o\| = o_p(n^{-1/4})$,
(c) $\mathbb{E}[(\phi^*(X_t) m(X_t, X_{t+1}) \phi(X_{t+1}))^2] < \infty$.

Discussion of Assumptions: Assumption 3.4(a) is an under-smoothing condition which ensures that the leading bias term $\sqrt{n}(\rho_k - \rho)$ and higher-order bias terms involving ϕ_k, ϕ_k^*, and ρ_k are asymptotically negligible. Assumption 3.4(b) ensures that $\widehat{\mathbf{G}}$ and $\widehat{\mathbf{M}}$ converge fast enough that $\sqrt{n}(\hat{\rho} - \rho_k)$ may be written in an asymptotically linear form similar to (24)–(25) but with ϕ_k, ϕ_k^*, and ρ_k in place of ϕ, ϕ^*, and ρ. This result, in view of the asymptotic negligibility of the leading and higher-order bias terms under Assumption 3.4(a), leads to the representation (24). Sufficient conditions for Assumption 3.4(b) are presented in Appendix C.1 in the Supplemental Material. Assumption 3.4(c) allows a CLT for square-integrable martingale differences to be applied to the martingale difference sequence $\{\psi_\rho(X_t, X_{t+1}) : t \in T\}$. Let $V_\rho = \mathbb{E}[\psi_\rho(X_0, X_1)^2]$.

THEOREM 3.2: *Let Assumptions 3.1–3.4 hold. Then the asymptotic linear expansion (24) holds and* $\sqrt{n}(\hat{\rho} - \rho) \to_d N(0, V_\rho)$.

It follows directly from Theorem 3.2 that $\sqrt{n}(\hat{y} - y) \to_d N(0, \rho^{-2} V_\rho)$.

We conclude by deriving the semiparametric efficiency bounds for Case 1. We require a further technical condition to characterize the tangent space (see Appendix B).

THEOREM 3.3: *Let Assumptions 3.1–3.4 and B.1 hold. Then the semiparametric efficiency bound for* ρ *is* V_ρ *and* $\hat{\rho}$ *is semiparametrically efficient.*

3.3.2. *Asymptotic Normality in Case 2*

For Case 2, we obtain the following expansion (under regularity conditions):

$$\sqrt{n}(\hat{\rho} - \rho) = \frac{1}{\sqrt{n}} \sum_{t=0}^{n-1} \big(\psi_\rho(X_t, X_{t+1}) + \psi_{\alpha,k}(X_t, X_{t+1})\big) + o_p(1), \tag{26}$$

where ψ_ρ is from display (25) with $m(x_0, x_1) = m(x_0, x_1; \alpha_0)$ and where

$$\psi_{\alpha,k}(x_0, x_1) = \phi_k^*(x_0)\big(m(x_0, x_1; \hat{\alpha}) - m(x_0, x_1; \alpha_0)\big)\phi_k(x_1). \tag{27}$$

The expansion (26) shows that the asymptotic distribution of $\hat{\rho}$ and related functionals will depend on the properties of the first-stage estimator $\hat{\alpha}$. The following regularity conditions are deliberately general so as to accommodate a wide class of estimators.

We first suppose that α_0 is a finite-dimensional parameter and the plug-in estimator $\hat{\alpha}$ is root-n consistent and asymptotically normal. Let $\psi_{\rho,t} = \psi_\rho(X_t, X_{t+1})$.

ASSUMPTION 3.5: *Let the following hold*:

(a) $\sqrt{n}(\hat{\alpha} - \alpha_0) = \frac{1}{\sqrt{n}} \sum_{t=0}^{n-1} \psi_{\alpha,t} + o_p(1)$ *for some* $\mathbb{R}^{d_\alpha}$*-valued random process* $\{\psi_{\alpha,t} : t \in T\}$,

(b) $\frac{1}{\sqrt{n}} \sum_{t=0}^{n-1} (\psi_{\rho,t}, \psi_{\alpha,t}')' \to_d N(0, V_{[2a]})$ *for some finite matrix* $V_{[2a]}$,

(c) $m(x_0, x_1; \alpha)$ *is continuously differentiable in* α *on a neighborhood* N *of* α_0 *for all* $(x_0, x_1) \in \mathcal{X}^2$ *and there exists some function* $\bar{m} : \mathcal{X}^2 \to \mathbb{R}$ *with* $\mathbb{E}[\bar{m}(X_t, X_{t+1})^s] < \infty$ *for some* $s \geq 2$ *such that*

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

(d) $\mathbb{E}[(\phi(X_t)\phi^*(X_t))^{s/(s-1)}] < \infty$.

Let $h_{[2a]} = (1, \mathbb{E}[\phi^*(X_t)\phi(X_{t+1})\frac{\partial m(X_t,X_{t+1};\alpha_0)}{\partial\alpha'}])'$ and define $V_\rho^{[2a]} = h'_{[2a]}V_{[2a]}h_{[2a]}$.

THEOREM 3.4: *Let Assumptions 3.1–3.5 hold. Then* $\sqrt{n}(\hat{\rho} - \rho) \to_d N(0, V_\rho^{[2a]})$.

We now suppose that α_0 is an infinite-dimensional parameter. The parameter space is $\mathcal{A} \subseteq \mathbb{A}$ (a Banach space) equipped with some norm $\|\cdot\|_{\mathcal{A}}$. This includes the case in which (1) α is a function, that is, $\alpha = h$ with $\mathbb{A} = \mathbb{H}$ a function space, and (2) α consists of both finite-dimensional and function parts, that is, $\alpha = (\theta, h)$ with $\mathbb{A} = \Theta \times \mathbb{H}$ with $\Theta \subseteq \mathbb{R}^{\dim(\theta)}$. For example, under recursive preferences the vector θ could consist of discount, risk aversion, and EIS parameters and h could be the continuation value function.

Inference in this case involves the (typically nonlinear) functional $\ell : \mathcal{A} \to \mathbb{R}$, given by

$$\ell(\alpha) = \mathbb{E}\big[\phi^*(X_t)\phi(X_{t+1})m(X_t, X_{t+1}; \alpha)\big].$$

We focus on the case in which $\ell(\alpha_0)$ is root-n estimable. We say the functional $\ell : \mathcal{A} \to \mathbb{R}$ is *pathwise differentiable* at α_0 if $\lim_{\tau\to 0^+}(\ell(\alpha_0 + \tau[\alpha - \alpha_0]) - \ell(\alpha_0))/\tau$ exists for every fixed $\alpha \in \mathcal{A}$. If so, we denote the derivative by $\dot{\ell}_{\alpha_0}[\alpha - \alpha_0]$. Define $\mathcal{G} = \{g_\alpha : \alpha \in \mathcal{A}\}$ where $g_\alpha(x_t, x_{t+1}) = \phi^*(x_t)\phi(x_{t+1})(m(x_t, x_{t+1}; \alpha) - m(x_t, x_{t+1}; \alpha_0))$. Let $\mathcal{Z}_n$ denote the centered empirical process on $\mathcal{G}$. We say $\mathcal{G}$ is *Donsker* if $\sum_{t\in\mathbb{Z}} \mathrm{Cov}(g(X_0, X_1), g(X_t, X_{t+1}))$ is absolutely convergent over $\mathcal{G}$ to a nonnegative quadratic form $\mathbb{K}(g, g)$ and there exists a sequence of Gaussian processes $\mathcal{Z}^{(n)}$ indexed by $\mathcal{G}$ with covariance function $\mathbb{K}$ and a.s. uniformly continuous sample paths such that $\sup_{g\in G} |\mathcal{Z}_n(g) - \mathcal{Z}^{(n)}(g)| \to_p 0$ as $n \to \infty$ (see Doukhan, Massart, and Rio (1995)). Finally, let $\|\cdot\|_p$ denote the L^p norm $\|\psi\|_p = (\int |\psi|^p \,\mathrm{d}Q)^{1/p}$ for any $1 \le p < \infty$ (note that $\|\cdot\|_2 = \|\cdot\|$ in our earlier notation).

ASSUMPTION 3.6: *Let the following hold*:

(a) *$\mathcal{G}$ is Donsker,*

(b) *ℓ is pathwise differentiable at α_0 and* $|\ell(\alpha) - \ell(\alpha_0) - \dot{\ell}_{\alpha_0}[\alpha - \alpha_0]| = O(\|\alpha - \alpha_0\|_{\mathcal{A}}^2)$,

(c) $\sqrt{n}\dot{\ell}_{\alpha_0}[\hat{\alpha} - \alpha_0] = \frac{1}{\sqrt{n}}\sum_{t=0}^{n-1}\psi_{\ell,t} + o_p(1)$ *for some $\mathbb{R}$-valued random process* $\{\psi_{\ell,t} : t \in T\}$, $\|\hat{\alpha} - \alpha_0\|_{\mathcal{A}} = o_p(n^{-1/4})$, *and* $\mathbb{K}(g_{\hat{\alpha}}, g_{\hat{\alpha}}) = o_p(1)$,

(d) $\frac{1}{\sqrt{n}}\sum_{t=0}^{n-1}(\psi_{\rho,t}, \psi_{\ell,t})' \to_d N(0, V_{[2b]})$ *for some finite matrix* $V_{[2b]}$,

(e) $\mathbb{E}[\sup_{\alpha\in\mathcal{A}} m(X_t, X_{t+1}; \alpha)^s] < \infty$ *and either* $\|\phi_k\|_{2s/(s-2)} = O(1)$ *and* $\|\phi^*\|_{2s/(s-2)} < \infty$ *or* $\|\phi_k^*\|_{2s/(s-2)} = O(1)$ *and* $\|\phi\|_{2s/(s-2)} < \infty$ *holds for some* $s > 2$.

Discussion of Assumptions: Sufficient conditions for the class $\mathcal{G}$ to be Donsker are well known (see, e.g., Doukhan, Massart, and Rio (1995)). Parts (b) and (c) are standard conditions for inference in nonlinear semiparametric models (see, e.g., Theorem 4.3 in Chen (2007)). Part (d) is a mild CLT condition and part (e) is a mild higher-than-second-moments condition.

For the following theorem, let $h_{[2b]} = (1, 1)'$ and define $V_\rho^{[2b]} = h'_{[2b]}V_{[2b]}h_{[2b]}$.

THEOREM 3.5: *Let Assumptions 3.1–3.4 and 3.6 hold. Then* $\sqrt{n}(\hat{\rho} - \rho) \to_d N(0, V_\rho^{[2b]})$.

4. VALUE FUNCTION RECURSION AS A NONLINEAR PERRON–FROBENIUS PROBLEM

This section describes how to estimate nonparametrically the continuation value function and SDF in a class of models with recursive preferences by solving a *nonlinear* Perron–Frobenius eigenfunction problem. We focus on models in which a representative agent has Epstein and Zin (1989) recursive preferences with unit elasticity of intertemporal substitution (EIS). This class of preferences may also be interpreted as risk-sensitive preferences as formulated by Hansen and Sargent (1995) (see Tallarini (2000)). After describing the setup, we present some regularity conditions for local identification. We then introduce the estimators and derive their large-sample properties.

4.1. *Setup*

Under Epstein–Zin preferences, the date-t continuation value of the representative agent is defined via the recursion

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

where C_t is date-t consumption, $1/\theta$ is the EIS, $\beta \in (0,1)$ is the time discount parameter, and $\gamma > 1$ is the relative risk aversion parameter. We maintain the assumption of a Markov state process X. Let consumption growth, namely, $G_{t+1} = C_{t+1}/C_t$, be a measurable function of (X_t, X_{t+1}). Hansen, Heaton, and Li (2008) showed that the scaled continuation value V_t/C_t may be written as $V(X_t)$, where

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

With unit EIS (i.e., $\theta = 1$), the fixed-point equation (28) reduces to

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

with $v(x) = \log V(x)$. Analytical solutions for v are typically only available when the conditional moment-generating function of the Markov state is exponentially affine and $\log G_{t+1}$ is affine in (X_t, X_{t+1}). Assuming frictionless markets, the SDF is

$$\frac{M_{t+1}}{M_t} = \beta G_{t+1}^{-1}\frac{(V_{t+1})^{1-\gamma}}{\mathbb{E}\left[(V_{t+1})^{1-\gamma}|X_t\right]}. \tag{30}$$

The dynamics of X determine both the value function and the conditional expectation in the denominator of the SDF. The value function and conditional expectation are therefore unknown when the dynamics of X are treated nonparametrically.

Consider the following reformulation of the fixed-point problem in display (29) as a nonlinear Perron–Frobenius problem. Setting $h(x) = \exp(\frac{1-\gamma}{\beta}v(x))$ and rearranging, we obtain the fixed-point equation $\mathbb{T}h = h$, where

$$\mathbb{T}\psi(x) = \mathbb{E}\left[G_{t+1}^{1-\gamma}\left|\psi(X_{t+1})\right|^{\beta}|X_t = x\right].$$

As we seek a positive solution, taking an absolute value inside the conditional expectation in the preceding display does not change the fixed point. Dividing $\mathbb{T}h = h$ by $\|h\|$ and using the fact that $\mathbb{T}$ is positive homogeneous of degree β, we obtain the nonlinear Perron–Frobenius problem

$$\mathbb{T}\chi(x) = \lambda\chi(x), \tag{31}$$

where $\chi(x) = h(x)/\|h\|$ is a positive eigenfunction of $\mathbb{T}$ and $\lambda = \|h\|^{1-\beta}$ is its eigenvalue. Throughout this section, we normalize the eigenfunction χ to have unit norm. Unlike with linear operators, here changing the scaling of h changes the corresponding eigenvalue: $c\chi$ is a positive eigenfunction of $\mathbb{T}$ with eigenvalue $c^{\beta-1}\lambda$ for any $c > 0$.

Reformulation of the recursion as a nonlinear Perron–Frobenius problem also leads to a convenient representation of the SDF. Rewriting the SDF from display (30) in terms of h, we obtain

$$\frac{M_{t+1}}{M_t} = \beta G_{t+1}^{-\gamma} \frac{\big(h(X_{t+1})\big)^{\beta}}{\mathbb{T}h(X_t)}.$$

Rescaling by $\|h\|$ and using (31) yields

$$\frac{M_{t+1}}{M_t} = \frac{\beta}{\lambda} G_{t+1}^{-\gamma} \frac{\big(\chi(X_{t+1})\big)^{\beta}}{\chi(X_t)}. \tag{32}$$

In what follows, we show how to estimate χ and λ from time-series data on X. The estimates $\hat{\chi}$ and $\hat{\lambda}$ can be plugged into (32) to obtain nonparametric estimates of the SDF process (i.e., without assuming a parametric law of motion for X).

4.2. *Local Identification*

In this section, we provide sufficient conditions for local identification of the fixed point h and its corresponding eigenfunction χ. We establish the results for the parameter (function) space L^2 because it is convenient for sieve estimation. One cannot establish (global) identification using contraction mapping arguments because $\mathbb{T}$ is not a contraction on L^2.[8] Some of the regularity conditions we require for estimation are sufficient for $\mathbb{T}$ to satisfy a local ergodicity property which, in turn, is sufficient for local identification.

To describe the local ergodicity property, first choose some (nonzero) function $\psi \in L^2$ and set $\chi_1(\psi) = \psi$. Then consider the sequence defined iteratively by

$$\chi_{n+1}(\psi) = \frac{\mathbb{T}\chi_n(\psi)}{\|\mathbb{T}\chi_n(\psi)\|}$$

for $n \geq 1$. Proposition 4.1 below shows that the sequence $\chi_n(\psi)$ converges to χ for any starting value ψ in a suitably defined region. This is similar to various "stability" results in the literature on balanced growth following Solow and Samuelson (1953).[9] There, $\mathbb{T} : \mathbb{R}^K \to \mathbb{R}^K$ is a homogeneous input-output system, $\chi_n \in \mathbb{R}^K$ lists the proportions of commodities in the economy in period n, and $\mathbb{T}\chi_n$ is normalized by its ℓ^1 norm so that $\chi_{n+1} := \mathbb{T}\chi_n / \|\mathbb{T}\chi_n\|_{\ell^1}$ lists the proportions in period $n + 1$. "Stability" concerns convergence of the sequence χ_n to a positive eigenvector χ of $\mathbb{T}$ (representing balanced growth proportions).

[8]Suppose that $\mathbb{T}$ has a positive fixed point $h \in L^2$. The function $\bar{h} \equiv 0$ is also a fixed point. Therefore, $\mathbb{T}$ is not a contraction on L^2 (else the Banach contraction mapping theorem would yield a unique fixed point).

[9]The literature on infinite-dimensional Perron–Frobenius theory has typically dealt with function spaces for which the cone of nonnegative functions has nonempty interior (see Krause (2015) for a recent overview). The nonnegative cone in L^2 has empty interior. If $\mathcal{X}$ is bounded, then these previous results may be used to derive (global) identification conditions in the space $C(\mathcal{X})$. However, bounded support seems inappropriate for common choices of state variable, such as consumption growth and dividend growth.

Write $\mathbb{T} = \mathbb{G}\mathbb{F}$ where $\mathbb{F}$ is the nonlinear operator $\mathbb{F}\psi(x) = |\psi(x)|^\beta$ and $\mathbb{G}$ is the linear operator

$$\mathbb{G}\psi(x) = \mathbb{E}\big[G_{t+1}^{1-\gamma}\psi(X_{t+1})|X_t = x\big].$$

The operator $\mathbb{T}$ is bounded (respectively, compact) on L^2 whenever $\mathbb{G}$ is bounded (compact) on L^2 (see Chapter 5 of Krasnosel'skii, Zabreiko, Pustylnik, and Sbolevskii (1976)). We say that $\mathbb{G}$ is *positive* if $\mathbb{G}\psi$ is positive for any nonnegative $\psi \in L^2$ that is not identically zero. Positivity of $\mathbb{G}$ ensures that the sequence $\chi_n(\psi)$ is well defined and that any nonzero fixed point of $\mathbb{T}$ is positive. We say that $\mathbb{T}$ is *Fréchet differentiable* at h if there exists a bounded linear operator $\mathbb{D}_h : L^2 \to L^2$ such that

$$\big\|\mathbb{T}(h+\psi) - \mathbb{T}h - \mathbb{D}_h\psi\big\| = o\big(\|\psi\|\big) \quad \text{as } \|\psi\| \to 0.$$

If it exists, the Fréchet derivative $\mathbb{D}_h$ of $\mathbb{T}$ is given by

$$\mathbb{D}_h\psi(x) = \mathbb{E}\big[\beta G_{t+1}^{1-\gamma} h(X_t)^{\beta-1}\psi(X_{t+1})|X_t = x\big].$$

Let $r(\mathbb{D}_h)$ denote the spectral radius of $\mathbb{D}_h$.

PROPOSITION 4.1: *Let $\mathbb{G}$ be positive and bounded and let $\mathbb{T}$ be Fréchet differentiable at h with $r(\mathbb{D}_h) < 1$. Then there exist finite positive constants C, c and a neighborhood N of χ such that*

$$\big\|\chi_{n+1}(\psi) - \chi\big\| \le Ce^{-cn}$$

for any initial point ψ in the cone $\{aN : a \in \mathbb{R}, a \neq 0\}$.

We say that χ is *locally identified* if there exists a neighborhood N of χ such that χ is the unique eigenfunction of $\mathbb{T}$ belonging to $N \cap S_1$ where S_1 denotes the unit sphere in L^2 (recall we normalize eigenfunctions of $\mathbb{T}$ to have unit norm). Similarly, we say that h is locally identified if h is the unique fixed point of $\mathbb{T}$ belonging to some neighborhood N' of h. To see why local identification follows from Proposition 4.1, suppose $\bar{\chi}$ is a positive eigenfunction of $\mathbb{T}$ belonging to $N \cap S_1$. Proposition 4.1 implies that $\|\chi_{n+1}(\bar{\chi}) - \chi\| = \|\bar{\chi} - \chi\| \le Ce^{-cn}$ for each n, hence $\bar{\chi} = \chi$. Local identification of h follows similarly.

COROLLARY 4.1: *h and χ are locally identified under the conditions of Proposition 4.1.*

In fact, local identification of χ and positive homogeneity of $\mathbb{T}$ imply that h is the unique fixed point of $\mathbb{T}$ in the cone $\{a(N \cap S_1) : a \in \mathbb{R}, a \neq 0\}$.

Existence and (global) identification of value functions in models with recursive preferences have been studied previously (see Marinacci and Montrucchio (2010), Hansen and Scheinkman (2012), and references therein). The most closely related work to ours is Hansen and Scheinkman (2012), who studied existence and uniqueness of value functions for Markovian environments in L^1 spaces (whose cones of nonnegative functions also have empty interior). Hansen and Scheinkman (2012) provided conditions under which a fixed point may exist when the EIS is equal to unity, but did not establish its uniqueness. Their existence conditions are based, in part, on existence of a positive eigenfunction of the operator $\mathbb{G}$.

There is also a connection between Corollary 4.1 and the literature on local identification of nonlinear, nonparametric econometric models. We can write $\mathbb{T}h = h$ as the conditional moment restriction:

$$\mathbb{E}\left[G_{t+1}^{1-\gamma}\left|h(X_{t+1})\right|^{\beta} - h(X_t)|X_t\right] = 0$$

(almost surely). The conditions of Proposition 4.1 ensure that the above moment restriction is Fréchet differentiable at h with derivative $\mathbb{D}_h - I$. The condition $r(\mathbb{D}_h) < 1$ implies that $\mathbb{D}_h - I$ is invertible on L^2. The conditions in Proposition 4.1 are therefore similar to the differentiability and rank conditions that Chen et al. (2014) used to study local identification in nonlinear conditional moment restriction models.

4.3. *Estimation*

We again use a sieve approach to reduce the infinite-dimensional problem to a low-dimensional (nonlinear) eigenvector problem. Consider the projected fixed-point problem

$$(\Pi_k \mathbb{T})h_k = h_k, \tag{33}$$

where $\Pi_k : L^2 \to B_k$ is the orthogonal projection onto the sieve space defined in Section 3. Lemma A.5 in the Appendix guarantees existence of a solution h_k to (33) on a neighborhood of h for all k sufficiently large. As $h_k \in B_k$, we have $h_k = b^k(x)'v_k$ for some vector $v_k \in \mathbb{R}^k$ which solves

$$\mathbf{G}_k^{-1}\mathbf{T}_k v_k = v_k, \tag{34}$$

where $\mathbf{T}_k v = \mathbb{E}[b^k(X_t)G_{t+1}^{1-\gamma}|b^k(X_{t+1})'v|^{\beta}]$. To simplify notation, we drop dependence of $\mathbf{G}_k$ and $\mathbf{T}_k$ on k hereafter. For estimation, we solve a sample analogue of (34), namely,

$$\widehat{\mathbf{G}}^{-1}\widehat{\mathbf{T}}\hat{v} = \hat{v}, \tag{35}$$

where $\widehat{\mathbf{G}}$ is defined in display (17) and $\widehat{\mathbf{T}} : \mathbb{R}^k \to \mathbb{R}^k$ is given by

$$\widehat{\mathbf{T}}v = \frac{1}{n}\sum_{t=0}^{n-1} b^k(X_t)G_{t+1}^{1-\gamma}\left|b^k(X_{t+1})'v\right|^{\beta}.$$

Under the regularity conditions below, a solution $\hat{v}$ on a neighborhood of v_k necessarily exists wpa1 (see Lemma A.7 in the Appendix). The estimators of h, χ, and λ are

$$\hat{h}(x) = b^k(x)'\hat{v}, \qquad \hat{\chi}(x) = \frac{b^k(x)'\hat{v}}{\left(\hat{v}'\widehat{\mathbf{G}}\hat{v}\right)^{1/2}}, \qquad \hat{\lambda} = \left(\hat{v}'\widehat{\mathbf{G}}\hat{v}\right)^{\frac{1-\beta}{2}}. \tag{36}$$

The estimators $\hat{\chi}$ and $\hat{\lambda}$ can then be plugged into display (32) to obtain an estimate of the SDF consistent with preference parameters (β, γ) and the observed law of motion of the state.

ASSUMPTION 4.1: *Let the following hold*:
(a) $\mathbb{T}$ *has a unique positive fixed point* $h \in L^2$,
(b) $\mathbb{G}$ *is positive and compact*,
(c) $\mathbb{T}$ *is Fréchet differentiable at* h *with* $r(\mathbb{D}_h) < 1$.

ASSUMPTION 4.2: *Let the following hold*:
(a) $\|\Pi_k \mathbb{D}_h - \mathbb{D}_h\| = o(1)$,
(b) $\sup_{\psi \in L^2 : \|\psi\| \le c} \|\Pi_k \mathbb{T}\psi - \mathbb{T}\psi\| = o(1)$ *for each* $c > 0$.

Let $\widehat{\mathbf{G}}^o$ be as in Assumption 3.3. Let $\mathbf{T}^o v = \mathbf{G}^{-1/2}\mathbf{T}(\mathbf{G}^{-1/2}v)$ and $\widehat{\mathbf{T}}^o v = \mathbf{G}^{-1/2}\widehat{\mathbf{T}}(\mathbf{G}^{-1/2}v)$. Note that $\widehat{\mathbf{G}}^o$ and $\widehat{\mathbf{T}}^o$ are a proof device and do not need to be calculated in practice.

ASSUMPTION 4.3: $\|\widehat{\mathbf{G}}^o - \mathbf{I}\| = o_p(1)$ *and* $\sup_{v \in \mathbb{R}^k : \|v\| \le c} \|\widehat{\mathbf{T}}^o v - \mathbf{T}^o v\| = o_p(1)$ *for each* $c > 0$.

Discussion of Assumptions: Assumption 4.1(a) is a global identification assumption; parts (b) and (c) impose some mild structure on $\mathbb{T}$ which ensures that fixed points of $\mathbb{T}$ are continuous under perturbations. Assumptions 4.2(a) and (b) are analogous to Assumption 3.2. Assumption 4.3 is similar to Assumption 3.3 and restricts the rate at which the sieve dimension k can grow with n; sufficient conditions are presented in Appendix C.2.

Let $\tau_k = \|\Pi_k h - h\|$ denote the bias in approximating h by an element of the sieve space. Assumption 4.2(b) implies that $\tau_k = o(1)$. To control the sampling error, fix any small $\varepsilon > 0$. By Assumption 4.3, we may choose a sequence of positive constants $\nu_{n,k}$ with $\nu_{n,k} = o(1)$ such that

$$\sup_{v \in \mathbb{R}^k : \|v' b^k - h\| \le \varepsilon} \left\| \left(\widehat{\mathbf{G}}^o\right)^{-1} \widehat{\mathbf{T}}^o v - \mathbf{T}^o v \right\| = O_p(\nu_{n,k}). \tag{37}$$

Appendix C.2 presents bounds on $\nu_{n,k}$.

THEOREM 4.1: *Let Assumptions* 4.1–4.3 *hold*. *Then*:
(a) $|\hat{\lambda} - \lambda| = O_p(\tau_k + \nu_{n,k})$,
(b) $\|\hat{\chi} - \chi\| = O_p(\tau_k + \nu_{n,k})$,
(c) $\|\hat{h} - h\| = O_p(\tau_k + \nu_{n,k})$.

The convergence rates obtained in Theorem 4.1 again exhibit a bias-variance tradeoff. The bias term τ_k is decreasing in k, whereas the variance term $\nu_{n,k}$ is typically increasing in k but decreasing in n. Choosing k to balance the terms will lead to the best convergence rate.

For implementation, we propose the following iterative scheme based on Proposition 4.1. Set $z_1 = \widehat{\mathbf{G}}^{-1}(\frac{1}{n}\sum_{t=0}^{n-1} b^k(X_t))$, then calculate

$$a_k = \frac{z_k}{\left(z_k' \widehat{\mathbf{G}} z_k\right)^{1/2}}, \qquad z_{k+1} = \widehat{\mathbf{G}}^{-1}\widehat{\mathbf{T}} a_k$$

for $k \ge 1$. If the sequence $\{(a_k, z_k) : k \ge 1\}$ converges to $(\hat{a}, \hat{z})$ (say), we then set

$$\hat{h}(x) = \hat{\lambda}^{\frac{1}{1-\beta}} b^k(x)'\hat{a}, \qquad \hat{\chi}(x) = b^k(x)'\hat{a}, \qquad \hat{\lambda} = \left(\hat{z}'\widehat{\mathbf{G}}\hat{z}\right)^{1/2}.$$

This iterative scheme proved to be a computationally efficient procedure for solving the sample fixed-point problem (35) in the simulations and empirical application.

5. SIMULATION EVIDENCE

The following Monte Carlo experiment illustrates the performance of the estimators in consumption-based models with power utility and recursive preferences. The state variable is log consumption growth, that is, $X_t = g_t$, which evolves as a Gaussian AR(1) process:

$$g_{t+1} - \mu = \kappa(g_t - \mu) + \sigma e_{t+1}, \quad e_t \sim \text{i.i.d. N}(0, 1).$$

The parameters for the simulation are $\mu = 0.005$, $\kappa = 0.6$, and $\sigma = 0.01$. The data are constructed to be somewhat representative of quarterly growth in U.S. real per capita consumption of nondurables and services (for which $\kappa \approx 0.3$ and $\sigma \approx 0.005$). However, we make the consumption growth process twice as persistent to produce more nonlinear eigenfunctions and twice as volatile to produce a more challenging estimation problem.

We consider a power utility design in which $m(X_t, X_{t+1}) = \beta G_{t+1}^{-\gamma}$ and a design with recursive preferences with unit EIS, whose SDF is presented in display (32). For both designs, we set $\beta = 0.994$ and $\gamma = 15$. The parameterization $\beta = 0.994$ and $\gamma = 10$ is typically used in calibrations of long-run risk models; here we take $\gamma = 15$ so that the eigenfunctions and continuation value function are more nonlinear. For each design, we generate 50,000 samples of length 400, 800, 1600, and 3200. Results reported in this section use a Hermite polynomial basis of dimension $k = 8$. Further experimentation with other sieve dimensions showed that the results were reasonably insensitive to the dimension of the sieve space. Similar results were obtained using B-splines (see the Online Appendix).

We estimate ϕ, ϕ^*, ρ, y, and L for both designs and χ and λ for the recursive preference design. We use the estimator $\widehat{\mathbf{G}}$ in (17) for both preference specifications. For power utility, we use the estimator $\widehat{\mathbf{M}}$ in (18). For recursive preferences, we first estimate (λ, χ) using the method described in the previous section, then construct the estimator $\widehat{\mathbf{M}}$ as in display (19), using

$$m(X_t, X_{t+1}; \hat{\lambda}, \hat{\chi}) = \frac{\beta}{\hat{\lambda}} G_{t+1}^{-\gamma} \frac{\left(\hat{\chi}(X_{t+1})\right)^{\beta}}{\hat{\chi}(X_t)}$$

based on the first-stage estimators $(\hat{\lambda}, \hat{\chi})$ of (λ, χ). We impose the scale normalizations $\frac{1}{n}\sum_{t=0}^{n-1} \hat{\phi}(X_t)^2 = 1$, $\frac{1}{n}\sum_{t=0}^{n-1} \hat{\phi}(X_t)\hat{\phi}^*(X_t) = 1$, and $\frac{1}{n}\sum_{t=0}^{n-1} \hat{\chi}(X_t)^2 = 1$.

The bias and RMSE of the estimators are presented in Tables I and II.[10] Table I shows that ϕ, ϕ^*, and χ may be estimated with small bias and RMSE using a reasonably low-dimensional sieve. Table II presents similar results for $\hat{\rho}$, $\hat{y}$, $\hat{L}$, and $\hat{\lambda}$. The RMSEs for $\hat{\phi}$ and $\hat{\rho}$ under recursive preferences are typically smaller than the RMSEs for $\hat{\phi}$ and $\hat{\rho}$ under power utility, even though with recursive preferences the continuation value must be first estimated nonparametrically. In contrast, the RMSE for $\hat{\phi}^*$ is larger under recursive preferences, which is likely due to the fact that ϕ^* is much more curved for that design (as evident from comparing the vertical scales Figures 1(b) and 1(d)). The results in Table I

[10]To calculate the RMSE of $\hat{\phi}$, $\hat{\phi}^*$, and $\hat{\chi}$, for each replication we calculate the L^2 distance between the estimators and their population counterparts, then take the average over the MC replications. To calculate the bias, we take the average of the estimators across the simulations to produce $\bar{\phi}(x)$, $\bar{\phi}^*(x)$, and $\bar{\chi}(x)$ (say), then compute the L^2 distance between $\bar{\phi}$, $\bar{\phi}^*$, and $\bar{\chi}$ and the true ϕ, ϕ^*, and χ. The use of the "bias" here is not to be confused with the bias term in the convergence rate calculations: here, "bias" of an estimator refers to the distance between the parameter and the average of its estimates across the simulations. Bias for $\hat{\rho}$, $\hat{y}$, $\hat{L}$, and $\hat{\lambda}$ is the average of the estimates across simulations minus the true parameter values.

TABLE I

SIMULATION RESULTS FOR $\hat{\phi}$, $\hat{\phi}^*$, AND $\hat{\chi}$ WITH A HERMITE POLYNOMIAL SIEVE OF DIMENSION $k = 8$

		Power Utility		Recursive Preferences		
	n	$\hat{\phi}$	$\hat{\phi}^*$	$\hat{\phi}$	$\hat{\phi}^*$	$\hat{\chi}$
Bias	400	0.0144	0.0129	0.0027	0.0247	0.0119
	800	0.0115	0.0129	0.0020	0.0187	0.0090
	1600	0.0084	0.0104	0.0016	0.0128	0.0062
	3200	0.0058	0.0077	0.0014	0.0095	0.0040
RMSE	400	0.1136	0.1683	0.0458	0.4068	0.1034
	800	0.0872	0.1060	0.0413	0.3513	0.0760
	1600	0.0681	0.0837	0.0361	0.1763	0.0577
	3200	0.0552	0.0677	0.0317	0.1591	0.0455

also show that χ may be estimated with a reasonably small degree of bias and RMSE in moderate samples.

Figures 1(a)–1(e) also present (pointwise) confidence intervals for ϕ, ϕ^*, and χ computed across simulations of different sample sizes. For each figure, the true function lies approximately in the center of the pointwise confidence intervals, and the widths of the intervals shrink noticeably as the sample size n increases. Corresponding plots using a B-spline basis are presented in the Online Appendix and are very similar to Figures 1(a)–1(e).

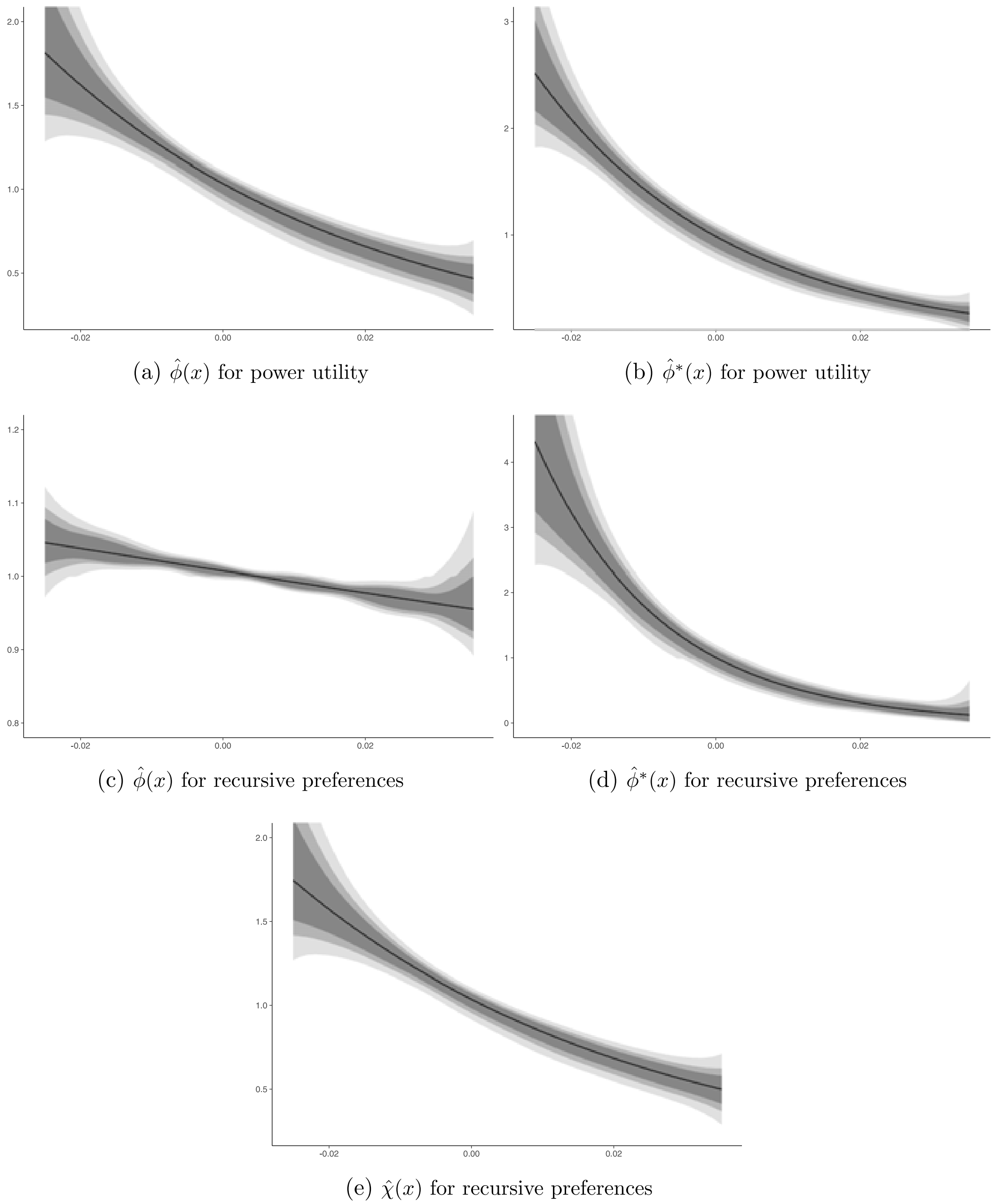
6. EMPIRICAL APPLICATION

In this section, we study an economy similar to that in Hansen, Heaton, and Li (2008). We assume a representative agent with Epstein and Zin (1989) recursive preferences with unit EIS and specify a two-dimensional state process in consumption and earnings growth. Our analysis may be summarized as follows. First, with discount and risk aversion parameters estimated from asset returns data ($\hat{\beta} \approx 0.985$ and $\hat{\gamma} \approx 24.5$), we show that this bivariate specification is able to generate a permanent component which implies a long-run equity premium (i.e., return on assets relative to long-term discount bonds) of approximately 2% per quarter. Second, we document the business cycle properties of the

TABLE II

SIMULATION RESULTS FOR $\hat{\rho}$, $\hat{y}$, $\hat{L}$, AND $\hat{\lambda}$ WITH A HERMITE POLYNOMIAL SIEVE OF DIMENSION $k = 8$

		Power Utility			Recursive Preferences			
	n	$\hat{\rho}$	$\hat{y}$	$\hat{L}$	$\hat{\rho}$	$\hat{y}$	$\hat{L}$	$\hat{\lambda}$
Bias	400	0.0035	−0.0029	0.0029	0.0010	−0.0008	0.0034	0.0040
	800	0.0027	−0.0024	0.0024	0.0011	−0.0010	0.0027	0.0022
	1600	0.0020	−0.0018	0.0018	0.0010	−0.0008	0.0020	0.0014
	3200	0.0014	−0.0013	0.0012	0.0010	−0.0009	0.0016	0.0009
RMSE	400	0.0358	0.0338	0.0282	0.0216	0.0179	0.0420	0.1005
	800	0.0264	0.0251	0.0214	0.0217	0.0172	0.0299	0.0318
	1600	0.0204	0.0192	0.0168	0.0190	0.0151	0.0227	0.0179
	3200	0.0159	0.0149	0.0133	0.0192	0.0155	0.0204	0.0123



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

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

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

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

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

FIGURE 1.—Simulation results for a Hermite polynomial basis with $k = 8$. Panels (a)–(d) display pointwise 90% confidence intervals for ϕ and ϕ^* across simulations (light, medium, and dark correspond to $n = 400$, 800, and 1600, respectively; the true ϕ and ϕ^* plotted as solid lines). Panel (e) displays results for the positive eigenfunction χ of the continuation value operator.

permanent and transitory components. Third, we describe the wedge required to tilt the distribution of the state to that which is relevant for long-run pricing. Finally, we show that, unlike the linear-Gaussian case, allowing for flexible treatment of the state process can lead to different behavior of long-run yields and different signs of correlation between the permanent and transitory components for different preference parameters. This suggests that nonlinearities in dynamics can be important in explaining the long end of the yield curve.

All data are quarterly and span the period 1947:Q1 to 2016:Q1 (277 observations). Data on consumption, dividends, inflation, and population are sourced from the National Income and Product Accounts (NIPA) tables. Real per capita consumption and dividend growth series are formed by taking seasonally adjusted consumption of nondurables and services (Table 2.3.5, lines 8 plus 13) and dividends (Table 1.12, line 16), deflating by the personal consumption implicit price deflator (Table 2.3.4, line 1), then converting to per capita growth rates using population data (Table 2.1, line 40). The resulting state variable is $X_t = (g_t, d_t)$, where g_t and d_t are real per capita consumption growth and dividend growth in quarter t, respectively. Similar results are obtained replacing d_t with real per capita growth in corporate earnings (using after-tax profits from line 15 of Table 1.12) and with real per capita growth in a four-quarter geometric moving average of dividends, as in Hansen, Heaton, and Li (2008).

We also use data on seven asset returns, namely, the returns on the six value-weighted portfolios sorted on size and book-to-market values (sourced from Kenneth French's website) and the 90-day Treasury bill rate. All asset returns series are converted to real returns using the implicit price deflator for personal consumption expenditures.

We estimate the preference parameters (β, γ) and the pair (λ, χ) using the data on X_t and the time series of seven asset returns. This falls into the setup of "Case 2" with $\alpha = (\beta, \gamma, \lambda, \chi)$. We estimate the parameters (β, γ) using a series conditional moment estimation procedure (Ai and Chen (2003)). This methodology was used recently in a similar context by Chen, Favilukis, and Ludvigson (2013).[11] For each (β, γ), we estimate the solution to the nonlinear eigenfunction problem, namely, $(\hat{\lambda}_{(\beta,\gamma)}, \hat{\chi}_{(\beta,\gamma)})$, using the procedure introduced in Section 4. Here we make explicit the dependence of (λ, χ) on β and γ, since different preference parameters will correspond to different continuation value functions. We then form

$$m\big(X_t, X_{t+1}; (\beta, \gamma, \hat{\lambda}_{(\beta,\gamma)}, \hat{\chi}_{(\beta,\gamma)})\big) = \frac{\beta}{\hat{\lambda}_{(\beta,\gamma)}} G_{t+1}^{-\gamma} \frac{\big(\hat{\chi}_{(\beta,\gamma)}(X_{t+1})\big)^{\beta}}{\hat{\chi}_{(\beta,\gamma)}(X_t)}.$$

Let $\mathbf{R}_{t+1}$ denote a vector of (gross) asset returns from time t to $t+1$ and $\mathbf{1}$ and $\mathbf{0}$ denote conformable vectors of ones and zeros. As the Euler equation $\mathbb{E}[m(X_t, X_{t+1})\mathbf{R}_{t+1} - \mathbf{1} | X_t] = \mathbf{0}$ holds conditionally, we instrument the generalized residuals, namely,

$$m\big(X_t, X_{t+1}; (\beta, \gamma, \hat{\lambda}_{(\beta,\gamma)}, \hat{\chi}_{(\beta,\gamma)})\big)\mathbf{R}_{t+1} - \mathbf{1},$$

[11]The differences between our estimator and that of Chen, Favilukis, and Ludvigson (2013) are as follows. First, we focus on the EIS = 1 case, whereas Chen, Favilukis, and Ludvigson (2013) treat the EIS as a free parameter. Second, we exploit the eigenfunction representation of the continuation value recursion. This allows us to "profile out" continuation value function estimation by solving for (λ, χ) by computationally simple iterative methods. Third, the continuation value is a function of the Markov state in our analysis, whereas the continuation value function in Chen, Favilukis, and Ludvigson (2013) depends on contemporaneous consumption growth and the lagged continuation value.

by basis functions of X_t to form a criterion function which exploits the conditional nature of the Euler equation. This leads to the criterion function

$$\Delta_n(\beta, \gamma) = \frac{1}{n} \sum_{t=0}^{n-1} \left\| e_n(X_t, \beta, \gamma) \right\|^2,$$

where

$$e_n(x, \beta, \gamma) = \left(\frac{1}{n} \sum_{t=0}^{n-1} \big(m\big(X_t, X_{t+1}; (\beta, \gamma, \hat{\lambda}_{(\beta,\gamma)}, \hat{\chi}_{(\beta,\gamma)})\big) \mathbf{R}_{t+1} - \mathbf{1} \big) b^k(X_t)' \right) \widehat{\mathbf{G}}^{-} b^k(x).$$

We minimize $\Delta_n(\beta, \gamma)$ to obtain $(\hat{\beta}, \hat{\gamma})$ and we set $\hat{\alpha} = (\hat{\beta}, \hat{\gamma}, \hat{\lambda}_{(\hat{\beta},\hat{\gamma})}, \hat{\chi}_{(\hat{\beta},\hat{\gamma})})$. We then estimate ρ, ϕ, ϕ^* and related quantities using the estimator $\widehat{\mathbf{M}}$ in display (19) for this choice of $\hat{\alpha}$.

To implement the procedure, we use the same basis functions for estimation of (ρ, ϕ, ϕ^*) and (λ, χ). We form fifth-order univariate Hermite polynomial bases for the g_t and d_t series. We then construct a tensor product basis from the univariate bases, discarding any tensor product polynomials whose total degree is order six or higher. The resulting sparse basis has dimension 15, whereas a tensor product basis would have dimension 25. We sometimes compare with the univariate state process $X_t = g_t$ for which we use an eighth-order Hermite polynomial basis. We instrument the generalized residuals with a lower-dimensional vector of basis functions to form the criterion function (dimension 6 in the univariate case and 10 in the bivariate case). Similar results are obtained with different dimensional bases.

Table III presents the estimates and bootstrapped 90% confidence intervals.[12] There are several notable aspects. First, both state specifications generate a permanent component whose entropy is consistent with a return premium of around 2% per quarter

TABLE III

PARAMETER ESTIMATES

	$X_t = (g_t, d_t)$	$X_t = g_t$	$X_t = (g_t, d_t)$		
$\hat{\rho}$	0.9812 (0.9733,0.9902)	0.9817 (0.9726,0.9893)	0.9859 (0.9851,0.9872)	0.9861 (0.9850,0.9881)	0.9860 (0.9842,0.9913)
$\hat{y}$	0.0190 (0.0098,0.0270)	0.0184 (0.0107,0.0277)	0.0142 (0.0129,0.0150)	0.0140 (0.0120,0.0151)	0.0141 (0.0087,0.0159)
$\hat{L}$	0.0193 (0.0000,0.0381)	0.0215 (0.0000,0.0426)	0.0128 (0.0090,0.0185)	0.0203 (0.0146,0.0295)	0.0297 (0.0198,0.0435)
$\hat{\beta}$	0.9851 (0.9784,0.9926)	0.9853 (0.9771,0.9921)	0.99	0.99	0.99
$\hat{\gamma}$	24.4712 (0.6850,44.7570)	27.4838 (0.0000,50.4619)	20	25	30
$\hat{\lambda}$	0.8999 (0.8146,0.9922)	0.8872 (0.7927,0.9888)	0.9154 (0.9008,0.9324)	0.8983 (0.8789,0.9205)	0.8834 (0.8579,0.9111)

[a]Left panel: Estimates of ρ, y, and L corresponding to $(\hat{\beta}, \hat{\gamma}, \hat{\lambda}, \hat{\chi})$. Right panel: estimates of ρ, y, and L corresponding to prespecified (β, γ) and estimated $(\hat{\lambda}, \hat{\chi})$. 90% bootstrap confidence intervals are in parentheses.

[12]We resample the data 1000 times using the stationary bootstrap with an expected block length of six quarters. In the left panel, we re-estimate β, γ, λ, χ, ρ, y, and L for each bootstrap replication. We discard the tiny fraction of replications in which the estimator of (β, γ) failed to converge. In the right panel, we fix β and γ and re-estimate λ, χ, ρ, y, and L for each bootstrap replication.

relative to the long bond, which is in the ballpark of empirically reasonable estimates. Second, the estimated long-run yield of around 1.9% per quarter is too large, which is explained by the low value of $\hat{\beta}$. Third, the estimated entropy of the SDF, namely, $\log(\frac{1}{n}\sum_{t=0}^{n-1} m(X_t, X_{t+1}; \hat{\alpha})) - \frac{1}{n}\sum_{t=0}^{n-1}\log(m(X_t, X_{t+1}; \hat{\alpha}))$, is 0.0191 for the bivariate specification and 0.0208 for the univariate specification. Therefore, the estimated horizon dependence (the difference between the entropy of the permanent component and the entropy of the SDF) is within the bound of $\pm 0.1\%$ per month that Backus, Chernov, and Zin (2014) argued is required to match the spread in average yields between short- and long-term bonds. Finally, the estimates of γ are quite imprecise, in agreement with previous studies (e.g., Chen, Favilukis, and Ludvigson (2013)). The confidence intervals for ρ, y, and L in the left panel of Table III are rather wide, which reflects, in large part, the uncertainty in estimating β and γ. Experimentation with different sieve dimensions resulted in estimates of γ between 20 and 30.[13] The right panel of Table III presents estimates of ρ, y, and L fixing $\beta = 0.99$ and $\gamma = 20$, 25, and 30 (χ and λ are still estimated nonparametrically). It is clear that the resulting confidence intervals are much narrower once the uncertainty from estimating β and γ is shut down.

We now turn to analyzing the time-series properties of the permanent and transitory components. The upper two panels of Figure 2 display time-series plots for the bivariate state specification of the SDF $m(X_t, X_{t+1}; \hat{\alpha})$ and its permanent component, constructed as

$$\frac{\hat{M}^P_{t+1}}{\hat{M}^P_t} = \hat{\rho}^{-1} m(X_t, X_{t+1}; \hat{\alpha}) \frac{\hat{\phi}(X_{t+1})}{\hat{\phi}(X_t)}.$$

As can be seen, the SDF and its permanent component evolve closely over time and exhibit strong countercyclicality. The transitory component (not plotted) is small, consistent with the literature on bounds which finds that the transitory component should be substantially smaller than the permanent component. The correlation of the permanent component series $\hat{M}^P_{t+1}/\hat{M}^P_t$ and GDP growth is approximately -0.39, whereas the correlation of the transitory component series $\hat{M}^T_{t+1}/\hat{M}^T_t$ and GDP growth is approximately 0.05. The lower panels of Figure 2 display time-series plots of the SDF and permanent component obtained under power utility using the same $(\hat{\beta}, \hat{\gamma})$ as in the recursive preference specification. These panels show that the permanent component, which is similar to that obtained under recursive preferences, is much more volatile than the SDF series. The large difference between the SDF and permanent component series under power utility is due to a very volatile transitory component, which implies a counterfactually large spread in average yields between short- and long-term bonds (Backus, Chernov, and Zin (2014)).

To understand further the long-run pricing implications of the estimates of ρ, ϕ, and ϕ^*, Figures 3(a)–3(d) plot the estimated ϕ and ϕ^* under recursive preferences for both the bivariate and univariate state specifications. It is evident from the vertical scales in Figures 3(a) and 3(b) that both estimates of ϕ are relatively flat, which explains the small transitory component obtained with recursive preferences. However, the estimated ϕ^* has a pronounced downward slope in g. The estimated ϕ^* in the bivariate specification is also downward-sloping in d at low levels of consumption growth. Recall that Proposition 2.1

[13]Chen, Favilukis, and Ludvigson (2013) obtained $\hat{\gamma} \approx 60$ using aggregate consumption data and $\hat{\gamma} \approx 20$ using stockholder consumption data. Further, with stockholder data, they find that the EIS is not significantly different from unity.

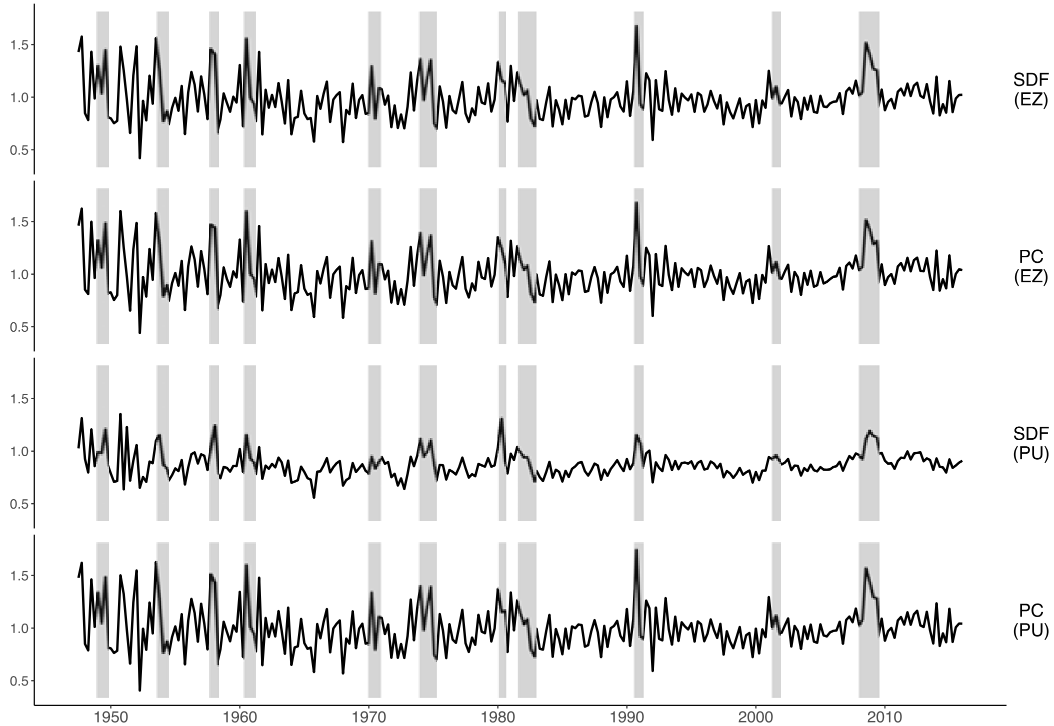
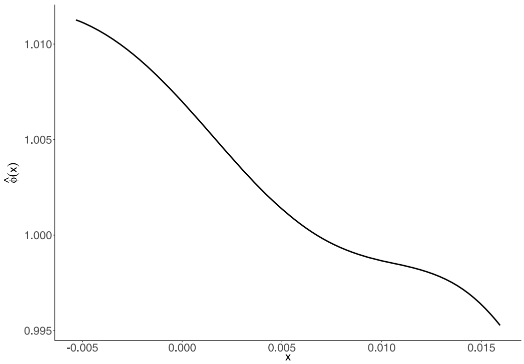


FIGURE 2.—Recovered time series of the SDF and and its permanent component (PC) when $X_t = (g_t, d_t)$. Upper panels are with Epstein and Zin (1989) recursive preferences with unit EIS (EZ); lower panels are with power utility (PU). Both use the estimated preference parameters $(\hat{\beta}, \hat{\gamma}) = (0.985, 24.471)$. Shaded areas denote NBER recession periods.

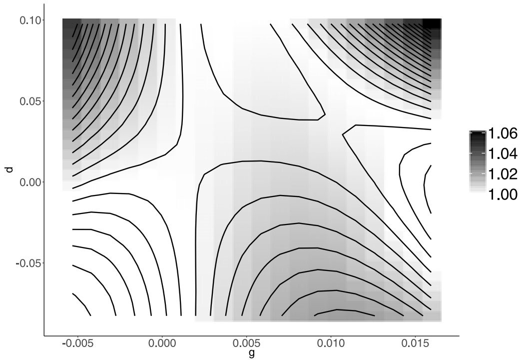
shows that $\phi\phi^*$ is the Radon–Nikodym derivative of $\widetilde{Q}$ with respect to Q. Figures 3(e)–3(f) plot the estimated change of measure for the two specifications of the state process. As the estimate of ϕ is relatively flat, the estimated change of measure is characterized largely by $\hat{\phi}^*$. In the bivariate case, the distribution $\widetilde{Q}$ assigns relatively more mass to regions of the state space in which there is low dividend and consumption growth than the stationary distribution Q, and relatively less mass to regions with high consumption growth.

Finally, we investigate the role of nonlinearities and non-Gaussianity in explaining certain features of the long end of the term structure. Figure 4 presents nonparametric estimates of (a) the (quarterly) long-run yield and (b) the correlation between the logarithm of the permanent and transitory components, namely, $\hat{m}^P_{t+1} = \log(\hat{M}^P_{t+1}/\hat{M}^P_t)$ and $\hat{m}^T_{t+1} = \log(\hat{M}^T_{t+1}/\hat{M}^T_t)$, recovered from the data on $X_t = (g_t, d_t)$ with $\beta = 0.994$ and γ increased from $\gamma = 1$ to $\gamma = 35$. The nonparametric estimates are presented alongside estimates for two parametric specifications of the state process. The first assumes $X_t = (g_t, d_t)$ is a Gaussian VAR(1), that is, $X_t - \mu = A(X_t - \mu) + e_{t+1}$ where the e_{t+1} are i.i.d. $N(0, \Sigma)$. The second is a Gaussian AR(1) for log consumption growth with stochastic volatility

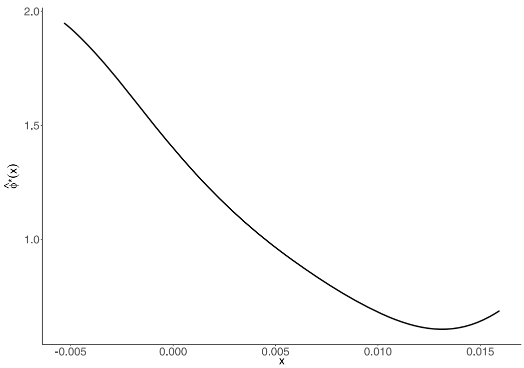
$$g_{t+1} - \mu = \kappa(g_t - \mu) + \sqrt{v_t}e_{t+1}, \quad e_{t+1} \sim \text{ i.i.d. N}(0, 1),$$



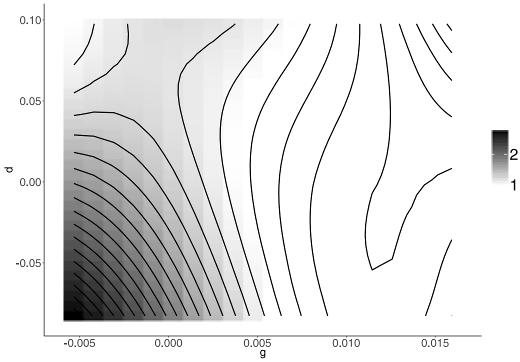
(a) Plot of $\hat{\phi}(x)$ for $X_t = g_t$



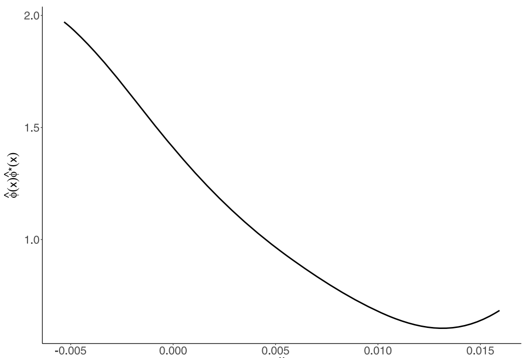
(b) Contour plot of $\hat{\phi}(x)$ for $X_t = (g_t, d_t)$



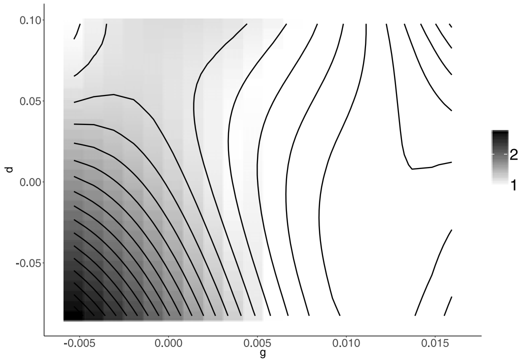
(c) Plot of $\hat{\phi}^*(x)$ for $X_t = g_t$



(d) Contour plot of $\hat{\phi}^*(x)$ for $X_t = (g_t, d_t)$



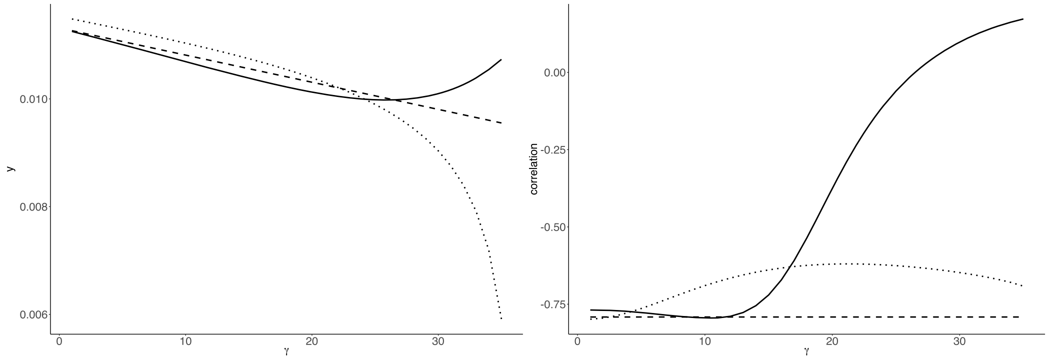
(e) Plot of $\hat{\phi}(x)\hat{\phi}^*(x)$ for $X_t = g_t$



(f) Contour plot of $\hat{\phi}(x)\hat{\phi}^*(x)$ for $X_t = (g_t, d_t)$

FIGURE 3.—Plots of $\hat{\phi}$ (upper panels), $\hat{\phi}^*$ (middle panels), and the estimated change of measure $\hat{\phi}(x)\hat{\phi}^*(x)$ between the stationary distribution Q and the distribution $\widetilde{Q}$ corresponding to $\widetilde{\mathbb{E}}$ (lower panels) under recursive preferences using the estimated preference parameters in the left panel of Table III.

where $\{v_t\}$ is a first-order autoregressive gamma process (a discrete-time version of the Feller square-root process; see Gourieroux and Jasiak (2006)) so the state vector is $X_t = (g_t, v_t)$. We refer to the second specification as SV-AR(1). The long-run yield and the correlation between $m^P_{t+1} = \log(M^P_{t+1}/M^P_t)$ and $m^T_{t+1} = \log(M^T_{t+1}/M^T_t)$ were obtained



(a) Long-run yield (b) Correlation between $\hat{m}^P_{t+1}$ and $\hat{m}^T_{t+1}$

FIGURE 4.—Solid lines: nonparametric estimates of the quarterly long-run yield and correlation between $\hat{m}^P_{t+1}$ and $\hat{m}^T_{t+1}$ under recursive preferences with $\beta = 0.994$ for different γ with $X_t = (g_t, d_t)$. Also displayed: parametric estimates obtained from fitting a Gaussian VAR(1) to $X_t = (g_t, d_t)$ (dashed lines) and fitting a SV-AR(1) to g_t (dotted lines).

analytically as functions of β, γ, and the estimates of the VAR(1) and SV-AR(1) parameters.[14]

Figure 4(a) shows that the nonparametric estimates of the long-run yield are nonmonotonic, whereas the parametric estimates are monotonically decreasing. This nonmonotonicity is not apparent in the nonparametric estimates using $X_t = g_t$. It is also clear that the nonparametric estimates of the long-run yield are much larger for high γ than the SV-AR(1) model.

Figure 4(b) displays the sample correlation of the nonparametric estimates $\hat{m}^P_{t+1}$ and $\hat{m}^T_{t+1}$ of the log permanent and transitory component series for different values of γ. This is compared with the correlation of the log permanent and transitory components m^P_{t+1} and m^T_{t+1} for the two parametric state-process specifications. The estimated correlation of the nonparametric estimates is negative for low to moderate values of γ, but becomes positive for larger values of γ. Similar results are obtained using lower- and higher-dimensional bases. In contrast, the correlations for the parametric state process specifications are around the same level as the nonparametric estimates for low values of γ, but remain negative for larger values of γ. A recent literature has emphasized the role of positive dependence between the permanent and transitory components in explaining excess returns of long-term bonds (Bakshi and Chabi-Yo (2012), Bakshi, Chabi-Yo, and Gao Bakshi (2015a, 2015b)). Positive dependence also features in models in which the term structure of risk prices is downward sloping (see the example presented in Section 7.2 in Borovička and Hansen (2016)). However, positive dependence is known to be difficult to generate via conventional preference specifications in workhorse models with exponentially-affine dynamics. Although the correlation is estimated imprecisely for large values of γ, this finding at least suggests that nonlinearities in state dynamics may have a role to play in explaining salient features of the long end of the yield curve.

[14]The VAR(1) parameters are estimated by OLS. The SV-AR(1) parameters are estimated via indirect inference using an AR(1) with GARCH(1, 1) errors as an auxiliary model. Derivations and further details on estimation are in the Supplemental Material.

7. CONCLUSION

This paper introduces an empirical framework to analyze the permanent and transitory components of SDF processes in the long-run factorization of Alvarez and Jermann (2005), Hansen and Scheinkman (2009), and Hansen (2012). We show how to estimate nonparametrically the solution to the Perron–Frobenius eigenfunction problem of Hansen and Scheinkman (2009) from time-series data on state variables and a SDF process. Our empirical framework allows researchers to (i) recover the time series of the estimated permanent and transitory components and investigate their properties and (ii) estimate the yield and the change of measure which characterize pricing over long investment horizons. This represents a useful contribution relative to existing empirical works which have established bounds on various moments of the permanent and transitory components as functions of asset returns, but have not extracted the components directly from data. The methodology is nonparametric in that it does not impose tight parametric restrictions on the dynamics of the state variables or the joint distribution of the state variables and the SDF process.

The main theoretical contributions of the paper are as follows. First, we establish consistency and convergence rates of the Perron–Frobenius eigenfunction estimators. Second, we establish asymptotic normality and some efficiency properties of the eigenvalue estimator and estimators of related functionals. Third, we introduce nonparametric estimators of the continuation value function in a class of models with recursive preferences by reinterpreting the value function recursion as a nonlinear Perron–Frobenius problem and we establish consistency and convergence rates of the value function estimators.

The econometric methodology may be extended and applied in several different ways. First, the methodology can be applied to study more general multiplicative functional processes such as the valuation and stochastic growth processes in Hansen, Heaton, and Li (2008), Hansen and Scheinkman (2009), and Hansen (2012). Second, the methodology can be applied to models with latent state variables. The main consistency and convergence rate results (Theorems 3.1 and 4.1) are sufficiently general that they apply equally to such cases. Finally, our analysis was conducted within the context of structural models in which the SDF process was linked tightly to preferences. A further extension would be to apply the methodology to SDF processes which are extracted flexibly from panels of asset returns data.

APPENDIX A: ADDITIONAL RESULTS ON ESTIMATION

A.1. *Bias and Variance Calculations for Theorem 3.1*

The results in this subsection draw heavily on arguments from Gobet, Hoffmann, and Reiß (2004). The first result shows that the approximate solutions ρ_k, ϕ_k, and ϕ_k^* from the eigenvector problem (15) are well defined and unique (i.e., up to sign and scale normalization) for all k sufficiently large.

LEMMA A.1: *Let Assumptions 3.1 and 3.2 hold. Then there exists $K \in \mathbb{N}$ such that for all $k \geq K$, the maximum eigenvalue ρ_k of the eigenvector problem (15) is real and simple, and hence $(\mathbf{M}, \mathbf{G})$ has unique right- and left-eigenvectors c_k and c_k^* corresponding to ρ_k.*

LEMMA A.2: *Let Assumptions 3.1 and 3.2 hold. Then:*
(a) $|\rho_k - \rho| = O(\delta_k)$,
(b) $\|\phi_k - \phi\| = O(\delta_k)$,

(c) $\|\phi_k^* - \phi^*\| = O(\delta_k^*)$,
where δ_k *and* δ_k^* *are defined in display* (22). *The rates should be understood to hold under the scale normalizations* $\|\phi\| = 1$, $\|\phi_k\| = 1$, $\|\phi^*\| = 1$, *and* $\|\phi_k^*\| = 1$ *and the sign normalizations* $\langle \phi_k, \phi \rangle \geq 0$ *and* $\langle \phi_k^*, \phi^* \rangle \geq 0$.

The following result shows that the solutions $\hat{\rho}$, $\hat{c}$, and $\hat{c}^*$ to the sample eigenvector problem (16) are well defined and unique with probability approaching 1 (wpa1).

LEMMA A.3: *Let Assumptions* 3.1–3.3 *hold*. *Then*, *wpa*1, *the maximum eigenvalue* $\hat{\rho}$ *of the generalized eigenvector problem* (16) *is real and simple*, *and hence* $(\widehat{\mathbf{M}}, \widehat{\mathbf{G}})$ *has unique right- and left-eigenvectors* $\hat{c}$ *and* $\hat{c}^*$ *corresponding to* $\hat{\rho}$.

LEMMA A.4: *Let Assumptions* 3.1–3.3 *hold*. *Then*:
(a) $|\hat{\rho} - \rho_k| = O_p(\eta_{n,k})$,
(b) $\|\hat{\phi} - \phi_k\| = O_p(\eta_{n,k})$,
(c) $\|\hat{\phi}^* - \phi_k^*\| = O_p(\eta_{n,k}^*)$,
where $\eta_{n,k}$ *and* $\eta_{n,k}^*$ *are defined in display* (23). *The rates should be understood to hold under the scale normalizations* $\|\hat{\phi}\| = 1$, $\|\phi_k\| = 1$, $\|\hat{\phi}^*\| = 1$, *and* $\|\phi_k^*\| = 1$ *and the sign normalizations* $\langle \hat{\phi}, \phi_k \rangle \geq 0$ *and* $\langle \hat{\phi}^*, \phi_k^* \rangle \geq 0$.

A.2. *Bias and Variance Calculations for Theorem 4.1*

The following two lemmas apply known results from the literature on the solution of nonlinear equations by projection methods (see, e.g., Chapter 19 of Krasnosel'skii, Vainikko, Zabreiko, Rutitskii, and Stetsenko (1972)). The first result shows that h_k is well defined for all k sufficiently large.

LEMMA A.5: *Let Assumptions* 4.1 *and* 4.2(b) *hold*. *Then there exist* $\varepsilon > 0$ *and* $K \in \mathbb{N}$ *such that for all* $k \geq K$, *the projected fixed-point problem* (33) *has at least one solution* h_k *in the ball* $N_k = \{\psi \in B_k : \|\psi - h\| < \varepsilon\}$.

REMARK A.1: Although the ball N_k may contain multiple solutions h_k of the projected fixed-point problem (33), under the conditions of Lemma A.5 we have that $\sup_{h_k \in H_k} \|h_k - h\| = o(1)$ where H_k denotes the set of all solutions to (33) in N_k.

REMARK A.2: If Assumption 4.1(c) is strengthened to require that $\mathbb{T}$ is continuously Fréchet differentiable at h with $r(\mathbb{D}_h) < 1$, then there exist $K \in \mathbb{N}$ and $\varepsilon > 0$ such that for all $k \geq K$, the projected fixed-point problem (33) has a unique solution h_k in the ball N_k.

In view of Remark A.1, in what follows we let h_k be any one of the solutions to (33) in N_k (or the unique solution under the additional assumption of continuous Fréchet differentiability of $\mathbb{T}$ at h). Let $\chi_k = h_k / \|h_k\|$ and $\lambda_k = \|\Pi_k \mathbb{T} \chi_k\|$.

LEMMA A.6: *Let Assumptions* 4.1 *and* 4.2 *hold*. *Then*:
(a) $|\lambda_k - \lambda| = O(\tau_k)$,
(b) $\|\chi_k - \chi\| = O(\tau_k)$,
(c) $\|h_k - h\| = O(\tau_k)$.

We now show that, wpa1, the sample fixed-point problem has a solution $\hat{v}$ for which $\hat{h}(x) = \hat{v}' b^k(x)$ belongs to N_k. We then derive convergence rates of the estimators formed using $\hat{v}$ (see display (36)). The following two results are new.

LEMMA A.7: *Let Assumptions 4.1–4.3 hold. Then, wpa1, there exists a fixed point $\hat{v}$ of $\widehat{\mathbf{G}}^{-1}\widehat{\mathbf{T}}$ such that the function $\hat{h}(x) = b^k(x)'\hat{v}$ belongs to N_k. Moreover, $\|\hat{h} - h\| = o_p(1)$.*

REMARK A.3: Although there may exist multiple fixed points $\hat{v}$ of $\widehat{\mathbf{G}}^{-1}\widehat{\mathbf{T}}$ for which $\hat{h}(x) = \hat{v}' b^k(x)$ belongs to N_k, under the conditions of Lemma A.7 we have that $\sup_{\hat{h}_k \in \hat{H}_k} \|\hat{h} - h\| = o_p(1)$ where $\hat{H}_k$ denotes the set of all such $b^k(x)'\hat{v}$ belonging to N_k.

In view of Remark A.3, the following lemma applies to estimators $\hat{\lambda}$, $\hat{\chi}$, and $\hat{h}$ in (36) formed from any fixed point $\hat{v}$ of $\widehat{\mathbf{G}}^{-1}\widehat{\mathbf{T}}$ for which $b^k(x)'\hat{v} \in N_k$.

LEMMA A.8: *Let Assumptions 4.1–4.3 hold. Then:*
(a) $|\hat{\lambda} - \lambda_k| = O_p(\nu_{n,k}) + o_p(\tau_k)$,
(b) $\|\hat{\chi} - \chi_k\| = O_p(\nu_{n,k}) + o_p(\tau_k)$,
(c) $\|\hat{h} - h_k\| = O_p(\nu_{n,k}) + o_p(\tau_k)$.

APPENDIX B: ADDITIONAL RESULTS ON INFERENCE

B.1. *Asymptotic Normality of Long-Run Entropy Estimators*

Here we consider the asymptotic distribution of the estimator $\hat{L}$ of the entropy of the permanent component of the SDF. In Case 1, the estimator of the long-run entropy is

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

Recall that $\psi_{\rho,t} = \psi_\rho(X_t, X_{t+1})$ where the influence function ψ_ρ is defined in (25). Define

$$\psi_{lm}(x_t, x_{t+1}) = \log m(x_t, x_{t+1}) - \mathbb{E}\big[\log m(X_t, X_{t+1})\big];$$

set $\psi_{lm,t} = \psi_{lm}(X_t, X_{t+1})$. Let $\hbar = (\rho^{-1}, -1)'$.

PROPOSITION B.1: *Let the assumptions of Theorem 3.2 hold and $\frac{1}{\sqrt{n}} \sum_{t=0}^{n-1} (\psi_{\rho,t}, \psi_{lm,t})' \to_d N(0, W)$ for some finite matrix W. Then*

$$\sqrt{n}(\hat{L} - L) \to_d N(0, V_L),$$

where $V_L = \hbar' W \hbar$.

In the preceding proposition, V_L will be the long-run variance:

$$V_L = \sum_{t \in \mathbb{Z}} \mathrm{Cov}\big(\psi_L(X_0, X_1), \psi_L(X_t, X_{t+1})\big),$$

where $\psi_L(X_t, X_{t+1}) = \rho^{-1}\psi_\rho(X_t, X_{t+1}) - \psi_{lm}(X_t, X_{t+1})$. Theorem B.1 below shows that V_L is the semiparametric efficiency bound for L.

In Case 2, the estimator of the long-run entropy is

$$\hat{L} = \log \hat{\rho} - \frac{1}{n}\sum_{t=0}^{n-1} \log m(X_t, X_{t+1}, \hat{\alpha}).$$

As with asymptotic normality of $\hat{\rho}$, the asymptotic distribution of $\hat{L}$ will depend on the manner in which $\hat{\alpha}$ was estimated. For brevity, we just consider the parametric case studied in Theorem 3.4. Let ψ_{lm} and $\psi_{lm,t}$ be as previously defined with $m(x_t, x_{t+1}) = m(x_t, x_{t+1}, \alpha_0)$. Recall $\psi_{\alpha,t}$ from Assumption 3.5 and define

$$\hbar_{[2a]} = \left(\rho^{-1}, \mathbb{E}\left[\left(\frac{\phi^*(X_t)\phi(X_{t+1})}{\rho} - \frac{1}{m(X_t, X_{t+1}, \alpha)}\right)\frac{\partial m(X_t, X_{t+1}, \alpha)}{\partial \alpha'}\right], -1\right)'.$$

PROPOSITION B.2: *Let the assumptions of Theorem 3.4 hold. Also let* (a) *there exist a neighborhood N_1 of α_0 upon which the function $\log m(x_0, x_1, \alpha)$ is continuously differentiable in α for all $(x_0, x_1) \in \mathcal{X}^2$ with*

$$\mathbb{E}\left[\sup_{\alpha \in N_1}\left\| \frac{1}{m(x_0, x_1, \alpha)} \frac{\partial m(x_0, x_1, \alpha)}{\partial \alpha}\right\|\right] < \infty,$$

and (b) $\frac{1}{\sqrt{n}}\sum_{t=0}^{n-1}(\psi_{\rho,t}, \psi'_{\alpha,t}, \psi_{lm,t})' \to_d N(0, W_{[2a]})$ *for some finite matrix $W_{[2a]}$. Then*

$$\sqrt{n}(\hat{L} - L) \to_d N\big(0, V_L^{[2a]}\big),$$

where $V_L^{[2a]} = \hbar'_{[2a]} W_{[2a]} \hbar_{[2a]}$.

B.2. *Semiparametric Efficiency Bounds in Case 1*

Let $P_n(x, A) = \Pr(X_{t+n} \in A | X_t = x)$ denote the n-step transition probability of X for any Borel set A. We say that $\{X_t\}_{t\in\mathbb{Z}}$ is *uniformly ergodic* if

$$\lim_{n\to\infty} \sup_{x\in\mathcal{X}} \big\| P_n(x, \cdot) - Q \big\|_{\mathrm{TV}} = 0,$$

where $\|\cdot\|_{\mathrm{TV}}$ denotes total variation norm and Q denotes the stationary distribution of X.

ASSUMPTION B.1: *$\{X_t\}_{t\in\mathbb{Z}}$ is uniformly ergodic.*

Sufficient conditions for Assumption B.1, such as Doeblin's condition, are well known. Assumption B.1 also implies that $\{X_t\}_{t\in\mathbb{Z}}$ is exponentially phi-mixing (see, e.g., Ibragimov and Linnik (1971, pp. 367–368)), and therefore exponentially beta- and rho-mixing.

THEOREM B.1:
(1) *Let Assumptions 3.1, 3.4*(c), *and B.1 hold and let $h : \mathbb{R} \to \mathbb{R}$ be continuously differentiable at ρ with $h'(\rho) \neq 0$. Then the efficiency bound for $h(\rho)$ is $h'(\rho)^2 V_\rho$.*
(2) *If, in addition, $\mathbb{E}[(\log m(X_t, X_{t+1}))^2] < \infty$, then the efficiency bound for L is V_L.*

B.3. *Sieve Perturbation Expansion*

The following result shows that $\hat{\rho} - \rho_k$ behaves as a linear functional of $\widehat{\mathbf{M}} - \rho_k \widehat{\mathbf{G}}$ and is used to derive the asymptotic distribution of $\hat{\rho}$ in Theorem 3.2. It follows from Assumption 3.3 that we can choose sequences of positive constants $\eta_{n,k,1}$ and $\eta_{n,k,2}$ such that

$$\left\|\widehat{\mathbf{G}}^o - \mathbf{I}\right\| = O_p(\eta_{n,k,1}) \quad \text{and} \quad \left\|\widehat{\mathbf{M}}^o - \mathbf{M}^o\right\| = O_p(\eta_{n,k,2})$$

with $\eta_{n,k,1} = o(1)$ and $\eta_{n,k,2} = o(1)$ as $n, k \to \infty$. Let c_k and c_k^* be normalized so that $\|\mathbf{G}^{1/2} c_k\| = 1$ and $c_k^{*\prime} \mathbf{G} c_k = 1$ (equivalent to $\|\phi_k\| = 1$ and $\langle \phi_k^*, \phi_k \rangle = 1$).

LEMMA B.1: *Let Assumptions 3.1–3.3 hold. Then*

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

In particular, if $\|\widehat{\mathbf{G}}^o - \mathbf{I}\| = o_p(n^{-1/4})$ *and* $\|\widehat{\mathbf{M}}^o - \mathbf{M}^o\| = o_p(n^{-1/4})$, *then*

$$\sqrt{n}(\hat{\rho} - \rho_k) = \sqrt{n} c_k^{*\prime}(\widehat{\mathbf{M}} - \rho_k \widehat{\mathbf{G}}) c_k + o_p(1).$$

REFERENCES

AI, C., AND X. CHEN (2003): "Efficient Estimation of Models With Conditional Moment Restrictions Containing Unknown Functions," *Econometrica*, 71 (6), 1795–1843. [1502,1524]

AÏT-SAHALIA, Y., AND A. W. LO (1998): "Nonparametric Estimation of State-Price Densities Implicit in Financial Asset Prices," *Journal of Finance*, 53 (2), 499–547. [1507]

ALVAREZ, F., AND U. J. JERMANN (2005): "Using Asset Prices to Measure the Persistence of the Marginal Utility of Wealth," *Econometrica*, 73 (6), 1977–2016. [1501,1502,1504,1505,1510,1530]

BACKUS, D., M. CHERNOV, AND S. ZIN (2014): "Sources of Entropy in Representative Agent Models," *Journal of Finance*, 69 (1), 51–99. [1502,1511,1526]

BAKSHI, G., AND F. CHABI-YO (2012): "Variance Bounds on the Permanent and Transitory Components of Stochastic Discount Factors," *Journal of Financial Economics*, 105 (1), 191–208. [1501,1502,1510,1529]

BAKSHI, G., F. CHABI-YO, AND X. GAO BAKSHI (2015a): "An Inquiry Into the Nature and Sources of Variation in the Expected Excess Return of a Long-Term Bond," SSRN Working Paper 2600097. [1502,1529]

——— (2015b): "A Recovery That We Can Trust? Deducing and Testing the Restrictions of the Recovery Theorem," Technical Report, University of Maryland. [1502,1529]

BANSAL, R., AND S. VISWANATHAN (1993): "No Arbitrage and Arbitrage Pricing: A New Approach," *Journal of Finance*, 48 (4), 1231–1262. [1507]

BANSAL, R., AND A. YARON (2004): "Risks for the Long Run: A Potential Resolution of Asset Pricing Puzzles," *Journal of Finance*, 59 (4), 1481–1509. [1502]

BLUNDELL, R., X. CHEN, AND D. KRISTENSEN (2007): "Semi-Nonparametric IV Estimation of Shape-Invariant Engel Curves," *Econometrica*, 75 (6), 1613–1669. [1513]

BOROVIČKA, J., AND L. P. HANSEN (2016): "Term Structure of Uncertainty in the Macroeconomy," in *Handbook of Macroeconomics*, Vol. 2. [1529]

BOROVIČKA, J., L. P. HANSEN, AND J. A. SCHEINKMAN (2016): "Misspecified Recovery," *Journal of Finance*, 71 (6), 2493–2544. [1502,1504]

CARRASCO, M., J.-P. FLORENS, AND E. RENAULT (2007): "Linear Inverse Problems in Structural Econometrics Estimation Based on Spectral Decomposition and Regularization," in *Handbook of Econometrics*, Vol. 6, ed. by J. J. Heckman and E. E. Leamer. Elsevier, 5633–5751. Part B, Chapter 77. [1503]

CHEN, X. (2007): "Large Sample Sieve Estimation of Semi-Nonparametric Models," in *Handbook of Econometrics*, Vol. 6, ed. by J. J. Heckman and E. E. Leamer. Elsevier, 5549–5632. Part B, Chapter 76. [1512,1515]

CHEN, X., AND S. C. LUDVIGSON (2009): "Land of Addicts? An Empirical Investigation of Habit-Based Asset Pricing Models," *Journal of Applied Econometrics*, 24 (7), 1057–1093. [1502,1507]

CHEN, X., V. CHERNOZHUKOV, S. LEE, AND W. K. NEWEY (2014): "Local Identification of Nonparametric and Semiparametric Models," *Econometrica*, 82 (2), 785–809. [1509,1519]

CHEN, X., J. FAVILUKIS, AND S. C. LUDVIGSON (2013): "An Estimation of Economic Models With Recursive Preferences," *Quantitative Economics*, 4 (1), 39–83. [1502,1507,1524,1526]

CHEN, X., L. P. HANSEN, AND J. A. SCHEINKMAN (2000): "Shape-Preserving Estimation of Diffusions," Working Paper, University of Chicago. [1503,1509]

CHRISTENSEN, T. M. (2015): "Nonparametric Identification of Positive Eigenfunctions," *Econometric Theory*, 31 (6), 1310–1330. [1509]

——— (2017): "Supplement to 'Nonparametric Stochastic Discount Factor Decomposition'," *Econometrica Supplementary Material*, 85, http://dx.doi.org/10.3982/ECTA11600. [1504,1511]

DAROLLES, S., J.-P. FLORENS, AND C. GOURIEROUX (2004): "Kernel-Based Nonlinear Canonical Analysis and Time Reversibility," *Journal of Econometrics*, 119 (2), 323–353. [1503]

DAROLLES, S., J.-P. FLORENS, AND E. RENAULT (1998): "Nonlinear Principal Components and Inference on a Conditional Expectation Operator With Applications to Markov Processes," Working Paper, Paris–Berlin Confernece, Garchy. [1503]

DOUKHAN, P., P. MASSART, AND E. RIO (1995): "Invariance Principles for Absolutely Regular Empirical Processes," *Annales de l'Institut Henri Poincaré (B) Probabilités et Statistiques*, 31, 393–427. [1515]

DUNFORD, N., AND J. T. SCHWARTZ (1958): *Linear Operators, Part I: General Theory*. New York: Interscience Publishers. [1508]

EPSTEIN, L. G., AND S. E. ZIN (1989): "Substitution, Risk Aversion, and the Temporal Behavior of Consumption and Asset Returns: A Theoretical Framework," *Econometrica*, 57 (4), 937–969. [1503,1516,1522,1527]

ESCANCIANO, J. C., AND S. HODERLEIN (2012): "Nonparametric Identification of Euler Equations," Working Paper, Indiana University. [1509]

ESCANCIANO, J. C., S. HODERLEIN, A. LEWBEL, O. B. LINTON, AND S. SRISUMA (2015): "Nonparametric Euler Equation Identification and Estimation," Cemmap Working Paper cwp61/15. [1503,1509]

GAGLIARDINI, P., C. GOURIEROUX, AND E. RENAULT (2011): "Efficient Derivative Pricing by the Extended Method of Moments," *Econometrica*, 79 (4), 1181–1232. [1502,1506,1507]

GOBET, E., M. HOFFMANN, AND M. REISS (2004): "Nonparametric Estimation of Scalar Diffusions Based on Low Frequency Data," *The Annals of Statistics*, 32, 2223–2253. [1503,1509,1511,1530]

GOURIEROUX, C., AND J. JASIAK (2006): "Autoregressive Gamma Processes," *Journal of Forecasting*, 25 (2), 129–152. [1528]

HANSEN, L. P. (1982): "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, 50 (4), 1029–1054. [1502]

——— (2012): "Dynamic Valuation Decomposition Within Stochastic Economies," *Econometrica*, 80 (3), 911–967. [1501,1502,1504,1511,1530]

HANSEN, L. P., AND E. RENAULT (2010): *Encyclopedia of Quantitative Finance, Chapter Pricing Kernels*. John Wiley & Sons, Ltd. [1504]

HANSEN, L. P., AND T. J. SARGENT (1995): "Discounted Linear Exponential Quadratic Gaussian Control," *IEEE Transactions on Automatic Control*, 40 (5), 968–971. [1503,1516]

HANSEN, L. P., AND J. A. SCHEINKMAN (2009): "Long-Term Risk: An Operator Approach," *Econometrica*, 77 (1), 177–234. [1501,1502,1504,1530]

——— (2012): "Recursive Utility in a Markov Environment With Stochastic Growth," *Proceedings of the National Academy of Sciences*, 109, 11967–11972. [1502,1518]

——— (2017): "Stochastic Compounding and Uncertain Valuation," in *After the Flood*, ed. by E. Glaeser, T. Santos, and G. Weyl. University of Chicago Press. [1502]

HANSEN, L. P., AND K. J. SINGLETON (1982): "Generalized Instrumental Variables Estimation of Nonlinear Rational Expectations Models," *Econometrica*, 50 (5), 1269–1286. [1507]

HANSEN, L. P., J. HEATON, AND A. YARON (1996): "Finite-Sample Properties of Some Alternative GMM Estimators," *Journal of Business & Economic Statistics*, 14 (3), 262–280. [1507]

HANSEN, L. P., J. C. HEATON, AND N. LI (2008): "Consumption Strikes Back? Measuring Long-Run Risk," *Journal of Political Economy*, 116, 260–302. [1502,1503,1516,1522,1524,1530]

HORN, R. A., AND C. R. JOHNSON (2013): *Matrix Analysis* (Second Ed.). Cambridge University Press. [1505]

IBRAGIMOV, I. A., AND Y. V. LINNIK (1971): *Independent and Stationary Sequences of Random Variables*. Wolters-Noordhoff. [1533]

KRASNOSEL'SKII, M. A., G. M. VAINIKKO, P. P. ZABREIKO, Y. B. RUTITSKII, AND V. Y. STETSENKO (1972): *Approximate Solution of Operator Equations*. Groningen: Wolters-Noordhoff. [1531]

KRASNOSEL'SKII, M. A., P. P. ZABREIKO, E. I. PUSTYLNIK, AND P. E. SBOLEVSKII (1976): *Integral Operators in Spaces of Summable Functions*. Noordhoff. [1518]

KRAUSE, U. (2015): *Positive Dynamical Systems in Discrete Time*. de Gruyter. [1517]

KREĬN, M. G., AND M. A. RUTMAN (1950): *Linear Operators Leaving Invariant a Cone in a Banach Space*. New York: American Mathematical Society. [1505]

KRESS, R. (1989): *Linear Integral Equations*. Springer-Verlag. [1513]

LEWBEL, A., O. B. LINTON, AND S. SRISUMA (2011): "Nonparametric Euler Equation Identification and Estimation," Working Paper, Boston College and London School of Economics. [1503,1509]

MARINACCI, M., AND L. MONTRUCCHIO (2010): "Unique Solutions for Stochastic Recursive Utilities," *Journal of Economic Theory*, 145 (5), 1776–1804. [1518]

QIN, L., AND V. LINETSKY (2016): "Positive Eigenfunctions of Markovian Pricing Operators: Hansen–Scheinkman Factorization, Ross Recovery, and Long-Term Pricing," *Operations Research*, 64 (1), 99–117. [1502]

——— (2017): "Long Term Risk: A Martingale Approach," *Econometrica*, 85 (1), 299–312. [1501,1502,1505]

QIN, L., V. LINETSKY, AND Y. NIE (2016): "Long Forward Probabilities, Recovery and the Term Structure of Bond Risk Premiums," SSRN Working Paper 2721366. [1502]

ROSENBERG, J. V., AND R. F. ENGLE (2002): "Empirical Pricing Kernels," *Journal of Financial Economics*, 64 (3), 341–372. [1507]

ROSENBLATT, M. (1971): *Markov Processes. Structure and Asymptotic Behavior*. Berlin: Springer-Verlag. [1506]

SCHAEFER, H. H. (1974): *Banach Lattices and Positive Operators*. Berlin: Springer-Verlag. [1508]

SOLOW, R. M., AND P. A. SAMUELSON (1953): "Balanced Growth Under Constant Returns to Scale," *Econometrica*, 21 (3), 412–424. [1503,1517]

TALLARINI, T. D. (2000): "Risk-Sensitive Real Business Cycles," *Journal of Monetary Economics*, 45 (3), 507–532. [1503,1516]

Co-editor Lars Peter Hansen handled this manuscript.

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