Computing the Distributions of Economic Models via Simulation

John Stachurski and Vance Martin[1]

We study a Monte Carlo algorithm for computing marginal and stationary densities of Markov models, establishing global asymptotic normality and $O_P(n^{-1/2})$ convergence. From these results we derive error bounds and a new nonparametric test for ergodic Markov processes. *Journal of Economic Literature* Classification Numbers: C15, C22, C63

Keywords: Markov processes, densities, simulation, ergodicity, numerical methods.

1. Introduction

When analyzing the dynamics of economic and econometric models, one often wishes to study the marginal and stationary distributions associated with the vector of state variables. For many models no closed form solution for these distributions exists, and numerical methods form the main bridge to quantitative applications. This paper studies one such method, proposed first by Glynn and Henderson (2001).

The problem can be introduced as follows. Let $\mathbb{X} \subset \mathbb{R}^k$, and let $p \colon \mathbb{X} \times \mathbb{X} \to \mathbb{R}$ be a *density kernel* on $\mathbb{X}$. That is, p is jointly measurable and

[1]This project has benefited from the comments of Peter Glynn, Shane Henderson, Sean Meyn, Kazuo Nishimura, Yoshihiko Nishiyama, Roberto Raimondo, co-editor Whitney Newey, an anonymous referee, and participants at the 12th International Conference on Computing in Economics and Finance, Cyprus 2006.

$p(x,y)\,\mathrm{d}y$ is a density on $\mathbb{X}$ for each $x \in \mathbb{X}$. Taking X_1 as given and recursively drawing

$$X_{t+1} \sim p(X_t, y)\,\mathrm{d}y \qquad (t \geq 1)$$

yields a discrete time Markov process $(X_t)_{t\geq 1}$ on $\mathbb{X}$.[2] It is well-known that for such a process, the (marginal) distribution of X_t can be represented by a density ψ_t on $\mathbb{X}$, and, moreover, the sequence $(\psi_t)_{t\geq 1}$ satisfies

$$\psi_{t+1}(y) = \int p(x,y)\psi_t(x)\,\mathrm{d}x \qquad (y \in \mathbb{X},\ t \geq 1) \tag{1}$$

Further, a density ψ_∞ on $\mathbb{X}$ is called *stationary* for the kernel p if

$$\psi_\infty(y) = \int p(x,y)\psi_\infty(x)\,\mathrm{d}x \qquad (y \in \mathbb{X}) \tag{2}$$

It is an equilibrium in the sense that if $X_1 \sim \psi_\infty$, then $X_t \sim \psi_\infty$ for all t, and in fact one can show that $(X_t)_{t\geq 1}$ is (strict sense) stationary.

In this paper we study how to compute numerical approximations to ψ_T (for some given $T \in \mathbb{N}$) and ψ_∞ when analytical expressions are unavailable. Previously a number of techniques have been suggested, including (i) discretization and (ii) simulation combined with histograms or nonparametric kernel density estimates. In what follows we analyze an alternative simulation-based technique which is both intuitively simple and computationally efficient.

To compute ψ_T, Glynn and Henderson (2001) propose the *marginal density look ahead estimator* (MDLAE) defined by

$$\psi_T^n(y) := \frac{1}{n}\sum_{i=1}^{n} p(X_{T-1}^i, y) \qquad (y \in \mathbb{X}) \tag{3}$$

where $(X_{T-1}^i)_{i=1}^n$ is n independent draws of the *lagged* state X_{T-1}. The intuition behind the estimator is straightforward: In view of (1) we have $\mathbb{E}\, p(X_{T-1}, y) = \psi_T(y)$. As $\psi_T^n(y)$ in (3) is by definition the sample mean of independent observations of $p(X_{T-1}, y)$, it follows that $\psi_T^n(y)$ is unbiased

[2]Given X_1 and p such a process $(X_t)_{t\geq 1}$ exists on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Conversely, given a model which defines the random process $(X_t)_{t\geq 1}$ directly, let $p(x,\, \mathrm{d}y)$ be the conditional distribution of X_{t+1} given $X_t = x$. We require that $p(x,\, \mathrm{d}y)$ can be represented by a density $p(x,y)\,\mathrm{d}y$ for all $x \in \mathbb{X}$.

and consistent for $\mathbb{E}\, p(X_{T-1}, y) = \psi_T(y)$. Moreover, when $\mathbb{E} p(X_{T-1}, y)^2$ is finite the Central Limit Theorem (CLT) implies that $\psi_T^n(y)$ is also $\sqrt{n}$-consistent for $\psi_T(y)$.[3]

The following example helps illustrate how ψ_T^n can be constructed in applications. Consider a model of the form

$$X_{t+1} = \mu(X_t) + \Sigma\, U_{t+1}, \quad (U_t)_{t \geq 1} \overset{\text{IID}}{\sim} N(0, \mathbb{I}_k) \tag{4}$$

where $\Gamma := \Sigma\Sigma^\top$ has positive determinant. The corresponding density kernel (i.e., conditional density of X_{t+1} given $X_t = x$) is

$$p(x, y) := \frac{1}{(2\pi)^{k/2} |\Gamma|^{1/2}} \exp\left\{ -\frac{1}{2} (y - \mu(x))^\top \Gamma^{-1} (y - \mu(x)) \right\} \tag{5}$$

An observation of $\psi_T^n(y)$ for this model can be generated using the algorithm below.

for *i in 1 to n* **do**
- draw X from the distribution of X_1 (which is given) ;
- **for** *t in 2 to* $T-1$ **do**
  - draw $U \sim N(0, \mathbb{I}_k)$;
  - set $X \leftarrow \mu(X) + \Sigma\, U$;
- **end**
- set $X_{T-1}^i \leftarrow X$;

end

return $\psi_T^n(y) := \frac{1}{n} \sum_{i=1}^n p(X_{T-1}^i, y)$, *where p is defined in* (5)

Next let us consider approximating the stationary density ψ_∞. Under the conditions on p in Section 3, a unique stationary density exists, and the associated Markov process $(X_t)_{t \geq 1}$ is ergodic in the sense that

$$\lim_{n \to \infty} \frac{1}{n} \sum_{t=1}^n h(X_t) = \int h(x) \psi_\infty(x)\, \mathrm{d}x \quad \text{with probability one} \tag{6}$$

for any initial X_1 and any ψ_∞-integrable function h.[4] Ergodicity implies

[3] In comparison, the nonparametric kernel density estimator generated from observations of X_T is biased and the error is $O_P((n\delta_n^k)^{-1/2})$, where $\delta_n \to 0$ is the bandwidth and k is the dimension of $\mathbb{X}$ (Yakowitz (1985)). The intuition behind the superior performance of the MDLAE is that the conditional density p in (3) subsumes the role of the kernel in the nonparametric estimator. While p always incorporates the dynamic structure contained in the original model, the nonparametric kernel and bandwidth do not.

[4] That is, any measurable $h \colon \mathbb{X} \to \mathbb{R}$ with $\int |h(x)| \psi_\infty(x)\, \mathrm{d}x < \infty$.

that sample moments contain information about ψ_∞. Based on this intuition, Glynn and Henderson (2001) propose approximating ψ_∞ via the *stationary density look ahead estimator* (SDLAE)

$$\psi_\infty^n(y) := \frac{1}{n}\sum_{t=1}^{n} p(X_t, y) \qquad (y \in \mathbb{X}) \tag{7}$$

where $(X_t)_{t=1}^n$ is a *time series* simulated from p and arbitrary X_1. Condition (6) now implies that with probability one,

$$\lim_{n\to\infty} \psi_\infty^n(y) = \lim_{n\to\infty} \frac{1}{n}\sum_{t=1}^{n} p(X_t, y) = \int p(x,y)\psi_\infty(x)\,\mathrm{d}x$$

In light of (2) this reads $\lim \psi_\infty^n(y) = \psi_\infty(y)$, and hence $\psi_\infty^n(y)$ is consistent for all $y \in \mathbb{X}$, independent of the initial condition X_1. Under some additional mixing conditions $\psi_\infty^n(y)$ is also $\sqrt{n}$-consistent for $\psi_\infty(y)$.

Returning to the model (4), with a growth restriction on μ (see below) the model is ergodic with unique stationary density ψ_∞. To approximate $\psi_\infty(y)$ using the SDLAE one can apply the following algorithm:

set $X_1 \leftarrow x$, where x is an arbitrary point in $\mathbb{X}$;
for *t in* $1, \ldots, n-1$ **do** `// generate` $X_{t+1} \sim p(X_t, y)\,\mathrm{d}y$
    draw $U \sim N(0, \mathbb{I}_k)$;
    set $X_{t+1} \leftarrow \mu(X_t) + \Sigma\, U$;
end
return $\psi_\infty^n(y) := \frac{1}{n}\sum_{t=1}^{n} p(X_t, y)$, *where p is defined in* (5)

We make the following contributions. Sections 2 and 3 extend Glynn and Henderson's analysis, emphasizing *global* convergence of ψ_T^n and ψ_∞^n to ψ_T and ψ_∞, respectively. Using a Hilbert space CLT, we prove that, when viewed as random functions, the deviations $\psi_T^n - \psi_T$ and $\psi_\infty^n - \psi_\infty$ are asymptotically normally distributed over a certain function space, and $\sqrt{n}$-consistent in the sense that the *norm* deviation is $O_P(n^{-1/2})$.

Section 4 discusses applications. While the look ahead estimators can be used in all fields where marginal and stationary densities of Markov models are calculated,[5] we focus on hypothesis testing, outlining how the

[5] As one example, Deaton and Laroque (1992) use discretization to compute the stationary distribution for prices associated with a competitive storage model. For the model they describe the SDLAE is applicable and trivial to implement. Further, with the theory developed below error bounds can be estimated.

asymptotic distribution of $\psi_\infty^n - \psi_\infty$ can be used to construct a nonparametric hypothesis test for ergodic Markov processes.

2. Global Convergence, Marginal Distribution

First let us consider global convergence of ψ_T^n to ψ_T. We use some facts concerning probability in Hilbert space. In what follows, let $\mathcal{H}$ be a separable Hilbert space with inner product $\langle g, h \rangle$ and norm $\|h\|_{\mathcal{H}} := \langle h, h\rangle^{1/2}$. If Y is a random variable taking values in $\mathcal{H}$ and $\mathbb{E}\|Y\|_{\mathcal{H}}$ is finite we can define $\mathcal{E}Y \in \mathcal{H}$ by the expression $\langle \mathcal{E}Y, h\rangle = \mathbb{E}\langle Y, h\rangle$, all $h \in \mathcal{H}$. This vector $\mathcal{E}Y$ is called the *expectation* of Y, and is necessarily unique.[6]

The CLT extends from $\mathbb{R}^k$ to general $\mathcal{H}$ almost unchanged: If $(Y_n)_{n\geq 1}$ is IID and $\mathbb{E}\|Y_1\|^2_{\mathcal{H}}$ is finite, then $\bar{Y}_n := n^{-1}\sum_{i=1}^n Y_i$ satisfies

$$\sqrt{n}(\bar{Y}_n - \mathcal{E}Y_1) \xrightarrow{\mathscr{D}} W \quad (n \to \infty) \tag{8}$$

where the random variable W is centered Gaussian on $\mathcal{H}$.[7] A corollary of this convergence in distribution is that $\|\bar{Y}_n - \mathcal{E}Y_1\|_{\mathcal{H}} = O_P(n^{-1/2})$.

The Hilbert space CLT can be used to study convergence of ψ_T^n to ψ_T. Let X_{T-1} be a random variable distributed according to ψ_{T-1}, and let $Y := p(X_{T-1}, \cdot\,)$ be the random function $y \mapsto p(X_{T-1}, y)$ from $\mathbb{X}$ to $\mathbb{R}$. An immediate consequence of this definition is that if $(X^i_{T-1})_{i=1}^n$ are IID copies of X_{T-1} then the sample mean

$$\bar{Y}_n := \frac{1}{n}\sum_{i=1}^n Y_i = \frac{1}{n}\sum_{i=1}^n p(X^i_{T-1}, \cdot\,) \tag{9}$$

is precisely ψ_T^n. Our asymptotic normality proof applies the CLT in (8) to $\bar{Y}_n = \psi_T^n$ in (9).

To employ the CLT in (8) three steps are necessary, the details of which are deferred to the appendix. The first step is to ensure that $Y = p(X_{T-1}, \cdot\,)$

[6] By the Cauchy-Schwartz inequality, $|\mathbb{E}\langle Y, h\rangle| \leq \mathbb{E}\|Y\|_{\mathcal{H}}\|h\|_{\mathcal{H}}$, and since $\mathbb{E}\|Y\|_{\mathcal{H}}$ is finite, $h \mapsto \mathbb{E}\langle Y, h\rangle$ is a bounded linear functional on $\mathcal{H}$. By the Riesz Representation Theorem, to such a functional there corresponds a vector $\mathcal{E}Y \in \mathcal{H}$ satisfying $\langle \mathcal{E}Y, h\rangle = \mathbb{E}\langle Y, h\rangle$, $h \in \mathcal{H}$. This $\mathcal{E}Y$ is defined to be the expectation of Y. In the present context all standard notions of vector-valued integration coincide (cf., e.g., Bosq (2000)).

[7] W is called centered Gaussian on $\mathcal{H}$ if, for every $h \in \mathcal{H}$, the real-valued random variable $\langle W, h\rangle$ has Gaussian distribution $N(0, \sigma_h^2)$ on $\mathbb{R}$ for some $\sigma_h^2 \geq 0$.

does in fact take values in a separable Hilbert space; in particular

$$\mathcal{H} = L_2(\mathbb{X}) := \left\{ \text{all measurable } h\colon \mathbb{X} \to \mathbb{R} \text{ s.t. } \int h(x)^2 \, \mathrm{d}x < \infty \right\}$$

with inner product $\langle g, h \rangle = \int gh$. This is done by placing a restriction on p in Theorem 1 below. The second step is to show that the moment condition $\mathbb{E}\|Y\|^2 < \infty$ is satisfied, where $\|\cdot\|$ is the norm on $L_2(\mathbb{X})$. The third step is to show that the expectation $\mathcal{E}Y$ of Y is ψ_T, in which case we have

$$\sqrt{n}(\bar{Y}_n - \mathcal{E}Y) = \sqrt{n}(\psi_T^n - \psi_T) \tag{10}$$

and the CLT in (8) can be applied:

THEOREM 1: *Let $(X_{T-1}^i)_{i=1}^n$ be* IID *copies of X_{T-1}, and let ψ_T^n be the MDLAE. If there exists a ψ_{T-1}-integrable function $V\colon \mathbb{X} \to \mathbb{R}$ such that*

$$\int p(x,y)^2 \, \mathrm{d}y \le V(x) \qquad (x \in \mathbb{X}) \tag{11}$$

then $\sqrt{n}(\psi_T^n - \psi_T)$ converges in distribution to a centered Gaussian random variable W taking values in $L_2(\mathbb{X})$.[8]

As a consequence we obtain the rate $\|\psi_T^n - \psi_T\| = O_P(n^{-1/2})$.

3. GLOBAL CONVERGENCE, STATIONARY DISTRIBUTION

Next we consider convergence of the SDLAE ψ_∞^n in (7) to ψ_∞. As for the case of local convergence (i.e., $\psi_\infty^n(y) \to \psi_\infty(y)$ for fixed y), global convergence of ψ_∞^n to ψ_∞ requires a form of ergodicity. We suppose that p is V-uniformly ergodic (V-UE); viz., there exists a measurable function $V\colon \mathbb{X} \to [1, \infty)$ and positive constants $\alpha < 1$ and $R < \infty$ with

$$\sup_{|h| \le V} \left| \int h(y) p^t(x,y) \, \mathrm{d}y - \int h(y) \psi_\infty(y) \, \mathrm{d}y \right| \le \alpha^t R V(x)$$

for all $x \in \mathbb{X}$ and all $t \ge 1$. Here p^t refers to the t-th order kernel: $p^t(x, \cdot)$ is the density of X_{k+t} when $X_k = x$.[9] Thus, $\int h(y) p^t(x,y) \, \mathrm{d}y$ is the expectation of $h(X_{t+1})$ conditional on $X_1 = x$.

[8] For example, if $x \mapsto \int p(x,y)^2 \, \mathrm{d}y$ is bounded on $\mathbb{X}$ then the conditions of the theorem are always satisfied.

[9] The kernels are defined by $p^1 = p$ and $p^{t+1}(x,y) = \int p(x,z) p^t(z,y) \, \mathrm{d}z$.

V-UE implies that $\int h(y) p^t(x, y)\, \mathrm{d}y$ converges geometrically to the expectation of h with respect to the stationary distribution. It also implies total variation (and hence L_1) convergence of $p^t(x, \cdot)$ to ψ_∞, as well as uniqueness of ψ_∞ and ergodicity as in (6).[10]

The V-UE property is closely related to geometric ergodicity, and sufficient conditions are well understood. For example, the model given by (4) and (5) is V-UE whenever μ satisfies

$$\exists\, a \in [0, 1) \text{ and } b \in \mathbb{R}_+ \text{ s.t. } \|\mu(x)\| \leq a\|x\| + b \quad (x \in \mathbb{X}) \tag{12}$$

for some norm $\|\cdot\|$ on $\mathbb{X}$. Kristensen (2006, Theorem 2) gives a useful set of sufficient conditions for geometric ergodicity, which he applies to linear and nonlinear ARMA, random coefficient and GARCH models. These conditions are in fact sufficient for the V-UE property.

With some modifications, the Hilbert space CLT in (8) can be used to prove asymptotic normality of the SDLAE. Let

$$L_2(\mathbb{X}, \psi_\infty) := \left\{ \text{all measurable } h \colon \mathbb{X} \to \mathbb{R} \text{ s.t. } \int h(x)^2 \psi_\infty(x)\, \mathrm{d}x < \infty \right\}$$

let $\langle g, h \rangle_{\psi_\infty} = \int g(x) h(x) \psi_\infty(x)\, \mathrm{d}x$ be the inner product on $L_2(\mathbb{X}, \psi_\infty)$, and let $\|\cdot\|_{\psi_\infty}$ denote the norm. Adding mild restrictions to p (see below), the densities $p(x, \cdot)$, ψ_∞^n and ψ_∞ all take values in $L_2(\mathbb{X}, \psi_\infty)$.

Now let $(X_t)_{t \geq 1}$ be a time series generated by p, and let Y_t be the $L_2(\mathbb{X}, \psi_\infty)$ valued random variable $p(X_t, \cdot)$. It follows that the sample mean $\bar{Y}_n$ is precisely ψ_∞^n. As discussed in the appendix, if $(X_t)_{t \geq 1}$ is stationary then the expectation $\mathcal{E} Y_1 = \mathcal{E} p(X_1, \cdot)$ is equal to ψ_∞, which yields

$$\sqrt{n}(\bar{Y}_n - \mathcal{E} Y_1) = \sqrt{n}(\psi_\infty^n - \psi_\infty) \tag{13}$$

The Hilbert space CLT in (8) does not immediately apply, as $(Y_t)_{t \geq 1}$ is now a correlated process. However, it is known that for Hilbert space valued functions of V-UE processes the CLT continues to hold (Stachurski (2006)). This gives the foundations of the following result:

THEOREM 2: *Let $(X_t)_{t \geq 1}$ be a Markov process on $\mathbb{X}$ with V-UE density kernel p. If*

$$\int p(x, y)^2 \psi_\infty(y)\, \mathrm{d}y \leq V(x) \qquad (x \in \mathbb{X}) \tag{14}$$

[10] In addition, V-UE implies aperiodicity, irreducibility and geometric mixing. Interested readers should consult Meyn and Tweedie (1993, Chapter 16).

then $\sqrt{n}(\psi_\infty^n - \psi_\infty)$ converges in distribution to a centered Gaussian random variable W on $L_2(\mathbb{X}, \psi_\infty)$ with covariance function

$$\begin{aligned}\Gamma(y,y') = \int p(x,y)p(x,y')\psi_\infty(x)\,\mathrm{d}x - \psi_\infty(y)\psi_\infty(y') \\ + \sum_{t\geq 1}^{\infty}\left[\int p(x,y)p^{t+1}(x,y')\psi_\infty(x)\,\mathrm{d}x - \psi_\infty(y)\psi_\infty(y')\right] \\ + \sum_{t\geq 1}^{\infty}\left[\int p(x,y')p^{t+1}(x,y)\psi_\infty(x)\,\mathrm{d}x - \psi_\infty(y)\psi_\infty(y')\right]\end{aligned}$$

The covariance function $\Gamma(y,y')$ can be viewed as the infinite dimensional analogue of a variance-covariance matrix.[11]

From Theorem 2 we obtain the asymptotic distribution of the error, measured in terms of the norm distance between ψ_∞^n and ψ_∞.

COROLLARY 1: *Under the hypotheses of Theorem 2 we have*

$$n\|\psi_\infty^n - \psi_\infty\|_{\psi_\infty}^2 \xrightarrow{\mathscr{D}} \sum_{\ell\geq 1}^{\infty}\lambda_\ell Z_\ell^2 \qquad (n\to\infty)$$

where $(\lambda_\ell)_{\ell\geq 1}$ are the eigenvalues of the covariance function Γ in Theorem 2, and $(Z_\ell)_{\ell\geq 1}$ are independent standard normal.[12]

Here $n\|\psi_\infty^n - \psi_\infty\|_{\psi_\infty}^2$ is the square of $\|\sqrt{n}(\psi_\infty^n - \psi_\infty)\|_{\psi_\infty}$, and Corollary 1 is an infinite dimensional version of the well-known fact that if $Y \sim N(0,C)$ in $\mathbb{R}^k$, then $\|Y\|^2$ has the same distribution as $\sum_{\ell=1}^k \lambda_\ell Z_\ell^2$, where $\|\cdot\|$ is the norm on $\mathbb{R}^k$, λ_ℓ is the ℓ-th eigenvalue of C and $(Z_\ell)_{\ell=1}^k$ are IID and $N(0,1)$. An immediate consequence of Corollary 1 is global $\sqrt{n}$-consistency. In particular, $\|\psi_\infty^n - \psi_\infty\|_{\psi_\infty} = O_P(n^{-1/2})$.

A final remark on Theorem 2 is that if p is V-UE and bounded then the conclusion of the theorem holds without (14). For example, p in (5) satisfies all the conditions of the theorem when (12) holds.

[11]Note that in fact we do not need $X_1 \sim \psi_\infty$. The result holds for $X_1 = x \in \mathbb{X}$, where x is arbitrary. This is important for implementation. It means that when simulating $(X_t)_{t\geq 1}$ to construct ψ_∞^n one can start at any $x \in \mathbb{X}$.

[12]More correctly, $(\lambda_\ell)_{\ell\geq 1}$ are the eigenvalues of the covariance operator C defined by the function Γ. For $h \in L_2(\mathbb{X}, \psi_\infty)$, Ch is given by $Ch(y') := \int \Gamma(y,y')h(y)\psi_\infty(y)\,\mathrm{d}y$.

4. APPLICATION

In this section we outline how the SDLAE can be used to construct a nonparametric hypothesis test for Markov processes. In brief, if $(X_t)_{t=1}^n$ is data, assumed to be generated from a V-UE kernel p, and if that same data is used to build the SDLAE ψ_∞^n in (7), then Theorem 2 provides the asymptotic distribution of $\psi_\infty^n - \psi_\infty$. This distribution can be used to test the hypothesis that the data is in fact generated by p.

Since we intend only to illustrate potential applications of the look ahead technique, we focus on one particular case, namely a nonparametric test for continuous time interest rate models. A related test was studied by Aït-Sahalia (1996), to which the reader is referred for further background.

Suppose we wish to test the null hypothesis that the short rate of interest $(X_t)_{t \geq 1}$ follows the Vasicek model

$$\mathrm{d}X_t = \kappa(\theta - X_t)\,\mathrm{d}t + \sigma\,\mathrm{d}B_t \tag{15}$$

where B_t is a standard Brownian motion and the parameters are set to $\kappa = 0.85837$, $\theta = 0.089102$, and $\sigma^2 = 0.0021854$.[13] Using the well-known formula for the conditional transition densities associated with this process, the density kernel under the null hypothesis is

$$p_0(x,y) := \frac{1}{\sqrt{2\pi}\sigma_v} \exp\left\{ \frac{-(y - m(x))^2}{2\sigma_v^2} \right\} \tag{16}$$

where $\sigma_v^2 := v_E(1 - e^{-2\kappa\Delta})$, $v_E := \sigma^2/(2\kappa)$, $m(x) := \theta + (x - \theta)e^{-\kappa\Delta}$ and $\Delta := 1/12$ is the time interval for the transition density, which is set to one month. The unique stationary density ψ_0 of p_0 is $N(\theta, v_E)$. The kernel p_0 is V-UE for $V(x) = |x| + 1$.

Let $(X_t)_{t=1}^n$ be n monthly observations of the short rate of interest, generated by *unknown* density kernel p, and let ψ_∞ be the stationary density of p. Define φ_n to be the function $n^{-1}\sum_{t=1}^n p_0(X_t, \cdot)$, where $(X_t)_{t=1}^n$ is the data. If the null hypothesis is true then $p_0 = p$, $\psi_0 = \psi_\infty$, φ_n is an observation of the SDLAE ψ_∞^n, and, by Corollary 1,

$$n\|\varphi_n - \psi_0\|_{\psi_0}^2 = n\|\psi_\infty^n - \psi_\infty\|_{\psi_\infty}^2 := n \int (\psi_\infty^n(y) - \psi_\infty(y))^2 \psi_\infty(y)\,\mathrm{d}y \tag{17}$$

[13] As estimated from US short rate data using GMM by Aït Sahalia (1996).

is asymptotically distributed as $\sum_{\ell \geq 1} \lambda_\ell Z_\ell^2$, where $(\lambda_\ell)_{\ell \geq 1}$ are the eigenvalues of the covariance function Γ in Theorem 2, and the $(Z_\ell)_{\ell \geq 1}$ are IID and standard normal. A test of size α rejects the null hypothesis if the left hand side of (17) exceeds the $(1-\alpha) \times 100\%$-quantile of the distribution of $\sum_{\ell \geq 1} \lambda_\ell Z_\ell^2$.

Our test is somewhat analogous to the one proposed by Aït-Sahalia (1996), the test statistic for which is similar to (17) but uses a standard nonparametric kernel density estimator (NPKDE) in place of ψ_∞^n. However, note that for the NPKDE none of the results developed in this paper apply, and Aït-Sahalia constructs his asymptotic theory using very different methods.

It has been argued (cf., e.g., Pritsker (1998)) that as the asymptotic distribution of Aït-Sahalia's test statistic depends only on the stationary density (see Aït-Sahalia (1996, p. 422)), it is insensitive to the amount of correlation in the underlying process, and as a result the test needs a relatively large amount of data to attain its asymptotic distribution. In contrast, the asymptotic distribution of (17) depends on the eigenvalues $(\lambda_\ell)_{\ell \geq 1}$ of the covariance function Γ, which, in turn, is defined by the whole sequence of higher order kernels $(p^t)_{t \geq 1}$, thereby capturing the full correlation structure of the process. Combined with the global $\sqrt{n}$-consistency of ψ_∞^n, this suggests that the SDLAE based statistic may attain its asymptotic distribution with less data.

While thorough investigation of these issues requires its own paper and is left for future research, we briefly describe two Monte Carlo experiments which illustrate the test. In the first experiment we study the size of the test under a true null hypothesis. In the second experiment we study the power of the test under false null hypotheses.[14]

To study the test under a true null hypothesis we generated 5,000 time series from the Vasicek model (16), where each time series is 22 years, giving $22 \times 12 = 264$ observations. For each series, we computed the test statistic (17) based on the true parameters. From these 5,000 samples of the test statistic we obtained the empirical rejection frequency for the true null when the size is $\alpha = 0.05$.

In doing so it was necessary to calculate the 95% quantile of the random

[14]In what follows, when evaluating the test statistic on the left hand side of (17) and computing its asymptotic distribution under the null hypothesis, we use the exact parameters given after (15). Hence the results of our Monte Carlo experiments do not account for estimated parameters in the density kernel p_0 and stationary density ψ_0.

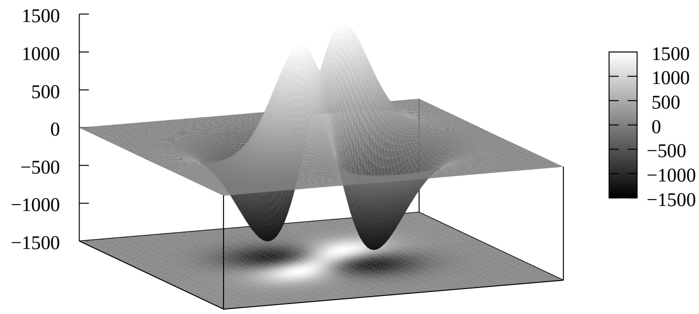

Figure 1: Covariance function $\Gamma(y, y')$, Vasicek model

variable $\sum_{\ell \geq 1} \lambda_\ell Z_\ell^2$ to which the statistic (17) converges. This required computing the covariance function Γ using the expression given in Theorem 2.[15] The function is shown in Figure 1. As expected, it is symmetric and nonnegative along the main diagonal. From Γ we computed a finite subset of the eigenvalues $(\lambda_\ell)_{\ell \geq 1}$. By simulating independent standard normals, we calculated the 95% quantile of $\sum_{\ell \geq 1} \lambda_\ell Z_\ell^2$ to be $\simeq$ 3,397.[16]

Of the 5,000 samples, some 4.257% exceeded the critical value, which is relatively close to the true test size of 5%. To give some context we repeated the same experiment but using the NPKDE-based test of Aït-Sahalia (1996, p. 393) instead of our test. The Aït-Sahalia test rejected the true null in over 50% of the samples with the same amount of data.[17]

Finally, we also considered the distribution of the statistic (17) under a false null by generating data using the level effects interest rate model

$$\mathrm{d}X_t = \kappa(\theta - X_t)\,\mathrm{d}t + \sigma X_t^\gamma\,\mathrm{d}B_t \tag{18}$$

where $0 \leq \gamma \leq 0.5$ and the other parameters are as before. Setting $\gamma = 0$ recovers the Vasicek null in (15).

[15] The infinite sums were truncated at 100 for numerical calculation.

[16] To compute eigenvalues a Galerkin projection technique was employed, projecting $L_2(\mathbb{X}, \psi_\infty)$ into a finite dimensional space spanned by 30 Hermite polynomials. Details of the method and computer code are available from the authors.

[17] The bandwidth used was the optimal bandwidth for estimating ψ_0, the stationary density of the Vasicek model. We experimented with other bandwidths but all choices gave a rejection rate in excess of 50%. Our results are consistent with those of Pritsker (1998), who found that Aït-Sahalia's test rejected the true null in over 50% of samples using the same Vasicek model and 22 years of *daily* data.

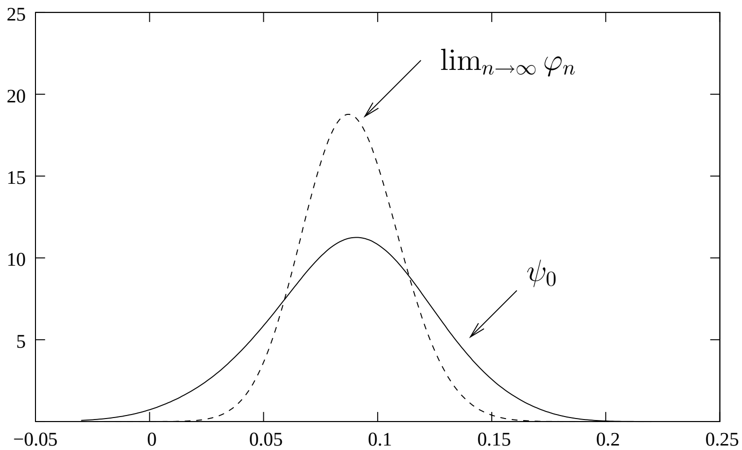

Figure 2: Limiting densities, false null

Suppose that $\gamma \neq 0$, so the Vasicek null hypothesis is false. Consider the test statistic $n\|\varphi_n - \psi_0\|^2_{\psi_0}$, where ψ_0 is the stationary distribution of the Vasicek null and $\varphi_n(y) = n^{-1}\sum_{t=1}^n p_0(X_t, y)$ is the SDLAE formed from the density kernel p_0 of the Vasicek null hypothesis and data $(X_t)_{t=1}^n$ generated from the level effects model (18). As the data generating process is not p_0 the limit of φ_n is (in general) distinct from ψ_0, and the test is rejected for sufficiently large n. Figure 2 shows ψ_0 and $\lim_{n\to\infty}\varphi_n$ when $\gamma = 0.3$.

To illustrate the ability of the test to reject the Vasicek null hypothesis when data is generated from the level effects model (18), we computed the power function (i.e., empirical rejection frequency) with 22 years of simulated data ($n = 264$) as γ ranges over the interval $[0, 0.5]$. Figure 3 shows the rejection frequency over 5,000 simulated time series with respect to the critical value 3,397. As $\gamma \to 0.5$ the rejection frequency converges to one.

Institute of Economic Research, Kyoto Univeristy, Yoshida honmachi, Sakyo-ku, Kyoto 606–8501, Japan; john@keir.kyoto-u.ac.jp

and

Department of Economics, University of Melbourne VIC 3010, Australia; vance@unimelb.edu.au

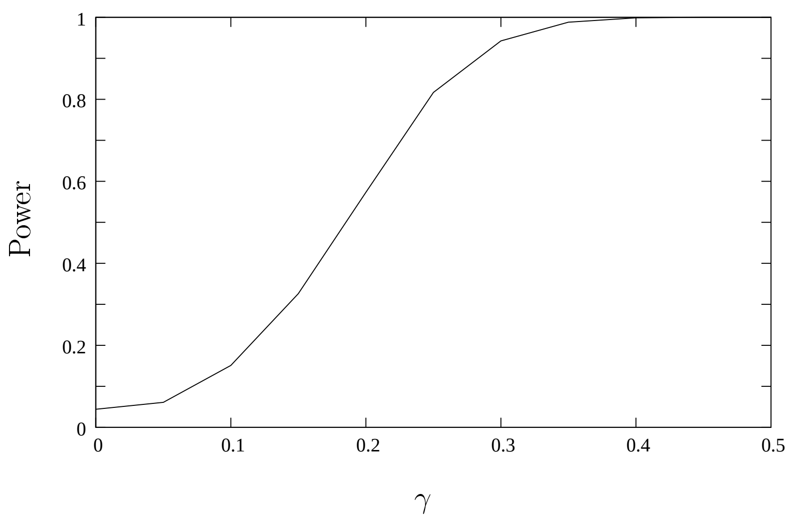

Figure 3: Rejection frequency, false null

Appendix

Regarding Theorem 1, in order to employ the CLT in (8), we must establish (i) that $Y = p(X_{T-1}, \cdot\,)$ takes values $L_2(\mathbb{X})$, (ii) that $\mathbb{E}\|Y\|^2 < \infty$, and (iii) that $\mathcal{E}Y = \psi_T$. In fact (i) is immediate from (11), as is (ii) because

$$\|Y\|^2 = \int p(X_{T-1}, y)^2 \, \mathrm{d}y \leq V(X_{T-1})$$

and $\mathbb{E}V(X_{T-1})$ is finite by assumption. To prove (iii) we must show that $\langle \psi_T, h \rangle = \mathbb{E}\langle Y, h \rangle$ for any $h \in L_2(\mathbb{X})$. Since $\psi_T(y) = \mathbb{E}p(X_{T-1}, y)$, for such an h we have

$$\langle \psi_T, h \rangle := \int \psi_T(y) h(y) \, \mathrm{d}y = \int \mathbb{E}p(X_{T-1}, y) h(y) \, \mathrm{d}y$$

On the other hand, an application of Fubini's theorem gives

$$\mathbb{E}\langle Y, h \rangle = \mathbb{E} \int p(X_{T-1}, y) h(y) \, \mathrm{d}y = \int \mathbb{E}p(X_{T-1}, y) h(y) \, \mathrm{d}y$$

Hence $\langle \psi_T, h \rangle = \mathbb{E}\langle Y, h \rangle$ for all $h \in L_2(\mathbb{X})$, and $\mathcal{E}Y = \psi_T$ as claimed.

Regarding Theorem 2, the fact that $\mathcal{E}Y_1 = \mathcal{E}p(X_1, \cdot) = \psi_\infty$ when $(X_t)_{t \geq 1}$ is stationary (and hence $X_1 \sim \psi_\infty$) can be proved in an almost identical manner to the proof of (iii) above. The sufficiency of (14) and the expression for Γ follow directly from Stachurski (2006, Theorem 3.1).

References

[1] Aït-Sahalia, Y. (1996): "Testing Continuous-Time Models of the Spot Interest Rate," *Review of Financial Studies*, 9 (2), 385–426.

[2] Bosq, D. (2000): *Linear Processes in Function Space*, Springer-Verlag.

[3] Deaton, A. and G. Laroque (1992): "On the Behavior of Commodity Prices," *Review of Economic Studies*, 59 (1), 1–23.

[4] Glynn, P. W. and S. G. Henderson (2001): "Computing Densities for Markov Chains via Simulation," *Mathematics of Operations Research*, 26, 375–400.

[5] Kristensen, D. (2006): "Geometric Ergodicity of a Class of Markov Chains with Applications to Time Series Models," mimeo, University of Wisconsin.

[6] Meyn, S. P. and Tweedie, R. L. (1993): *Markov Chains and Stochastic Stability*, Springer-Verlag: London.

[7] Pritsker, M. (1998): "Nonparametric Density Estimation and Tests of Continuous Time Interest Rate Models," *Review of Financial Studies*, 11 (3), 449–487.

[8] Stachurski, J. (2006): "A Hilbert Space Central Limit Theorem for Geometrically Ergodic Markov Chains," mimeo, Kyoto University.

[9] Yakowitz, S. J. (1985): "Nonparametric Density Estimation, Prediction and Regression for Markov Sequences," *Journal of the American Statistical Association*, Vol. 80, No. 389, 215–221.