KIER DISCUSSION PAPER SERIES

KYOTO INSTITUTE
OF
ECONOMIC RESEARCH

Discussion Paper No.787

"A GOODNESS OF FIT TEST
FOR ERGODIC MARKOV PROCESSES"

VANCE MARTIN, YOSHIHIKO NISHIYAMA
AND JOHN STACHURSKI

October 2011

KYOTO UNIVERSITY

KYOTO, JAPAN

A GOODNESS OF FIT TEST FOR ERGODIC MARKOV PROCESSES

VANCE MARTIN[A], YOSHIHIKO NISHIYAMA[B]
AND JOHN STACHURSKI[C]

We introduce a goodness of fit test for ergodic Markov processes. Our test compares the data against the set of stationary densities implied by the class of models specified in the null hypothesis, and rejects if no model in the class yields a stationary density that matches with the data. No alternative needs to be specified in order to implement the test. Although our test compares densities it involves no smoothing parameters, and is powerful against $1/\sqrt{n}$ local alternatives.

KEYWORDS: Specification test, goodness of fit, Markov processes.

1. INTRODUCTION

For a dynamic stochastic model used in some particular application, an overriding concern is whether or not the dynamics of the model are consistent with the time series being modeled. To give one of many possible examples, most valuations of interest rate derivative securities depend on the underlying model used to represent the interest rate. If the model fit is poor, in the sense that probabilities implied by the model are inconsistent with actual interest rate dynamics, then the resulting valuation will be unreliable.

In testing model specification for random dynamic models, one potential problem is that, in general, stochastic processes are relatively complex objects, described by high-dimensional joint distributions. As a result, the power of any specification test risks being dispersed over a large space of possible alternatives. To accommodate this, some tests require a tight specification of the alternative hypothesis. An obvious problem here is that, in many settings, theory says little about the set of possible alternatives.

To test without imposing strict assumptions on the alternative and yet still retain power in appropriate directions, a natural way to proceed is to compare the data with partic-

[1]The authors are grateful to Peter Glynn and seminar participants at the 2009 NCER conference at Princeton and the 2011 Australasian Meetings of the Econometric society for many helpful comments. Our research was supported in part by Australian Research Council Grant DP0987589, and by Japan Society for the Promotion of Science Grants-in-Aid 22330067.

[a]Department of Economics, The University of Melbourne
[b]Institute of Economic Research, Kyoto University
[c]Research School of Economics, Australian National University

ular features of the theoretical data generating process (DGP). A classic example of this approach is the Hansen–Sargan J-test (Hansen, 1982), which compares the data against theoretical moment restrictions on the DGP. Another is the nonparametric test proposed by Aït-Sahalia (1996), where a nonparametric kernel density estimate of the stationary distribution is compared to the stationary distribution of the model.

In this paper, we consider a specification test for stationary dynamic models with the Markov property. In the manner of Aït-Sahalia (1996), our test compares the data against the set of stationary densities corresponding to the set of models contained in the null hypothesis, and rejects if no model in this class yields a stationary density that matches with the data. No alternative needs to be specified in order to implement the test. Unlike Aït-Sahalia's test, our test involves no smoothing parameters, and is powerful against $1/\sqrt{n}$ local alternatives.[1]

While our test is in many senses a very natural test for stationary Markov processes, the attractive features of the test mentioned above do not constitute a free lunch. The test makes explicit use of the correlation structure in the null hypothesis, and this involves two costs. The first cost is that there is no direct analogue of our test in the case of IID observations. (Although our test is well-defined for an IID null hypothesis, it has zero power in all directions.) The second is that the conditional transition density associated with the null hypothesis forms part of the test statistic, and as such this density must be relatively tractable.

1.1. *Goodness of Fit*

To describe our test in an abstract setting, we begin with a goodness of fit test for a single model p, where p is a Markov transition density kernel. Heuristically, $p(x, y)dy$ is the probability of transitioning from state $x \in \mathbb{X}$ to state $y \in \mathbb{X}$ over one unit of time. Suppose that p is ergodic with a unique stationary ψ. By definition, ψ satisfies

$$(1) \qquad \int p(x,y)\psi(x)dx = \psi(y) \qquad (y \in \mathbb{X})$$

Our interest is in testing whether some given $\mathbb{X}$-valued time series $\{X_t\}_{t=1}^n$ is generated by p. To test the validity of this null hypothesis, consider the deviation

$$(2) \qquad \left| \frac{1}{n} \sum_{t=1}^{n} p(X_t, y) - \psi(y) \right|$$

[1]In addition, our test is formulated for Markov processes of arbitrary dimension, whereas Aït-Sahalia's test is formulated for univariate data. (In fact, for our test the state space is an arbitrary measure space, so in theory the dimension may be infinite.) Aït-Sahalia's test could no doubt be extended to the multivariate case, but the performance will be compromised because nonparametric kernel density estimators are known to degrade rapidly as the dimension of the state space increases.

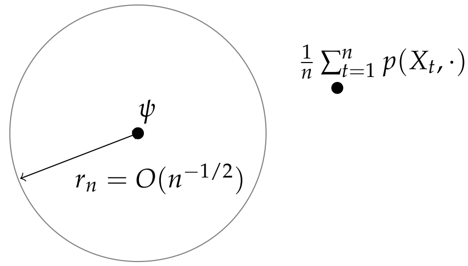

FIGURE 1.— Reject if $\frac{1}{n}\sum_{t=1}^n p(X_t, \cdot) \notin B(r_n, \psi)$

When the null holds, the sequence $\{X_t\}_{t=1}^n$ is stationary and ergodic with common density ψ, and hence, for large n,

$$\frac{1}{n}\sum_{t=1}^n p(X_t, y) - \psi(y) \approx \mathbf{E} p(X_t, y) - \psi(y) = \int p(x, y)\psi(x)dx - \psi(y) = 0$$

In other words, the deviation in (2) should be small for large n. Moreover, since this argument is valid for any given y, we can adopt a functional approach, regarding $\frac{1}{n}\sum_{t=1}^n p(X_t, \cdot) - \psi(\cdot)$ as a random element taking values in the function space L_2, and rejecting the null when its norm is large—that is, when its realization lies outside a sphere $B(r_n, 0)$ centered on the origin of L_2. The radius r_n of the sphere is computed from an L_2 central limit theorem to produce a test of given size.

Another way to phrase the test is that we reject the null if $\frac{1}{n}\sum_{t=1}^n p(X_t, \cdot)$ lies outside a sphere of the same radius r_n centered on ψ. This perspective is illustrated in figure 1. The test compares the theoretical stationary distribution ψ against an estimate $\frac{1}{n}\sum_{t=1}^n p(X_t, \cdot)$ that is $\sqrt{n}$-consistent for ψ under the null. The radius r_n is of the form $c/\sqrt{n}$, where c depends on the size of the test. The fact that the radius is $O(1/\sqrt{n})$ suggests that the test will have nontrivial power against $1/\sqrt{n}$ local alternatives. This intuition is confirmed in section 3.2.[2]

1.2. Estimated Parameters

The goodness of fit test described above is mainly of theoretical interest. In practical situations our models usually contain unknown parameters, and we wish to test whether or not our parametric class of models can represent the data. We extend to this setting

[2]The consistency of $\frac{1}{n}\sum_{t=1}^n p(X_t, \cdot)$ for ψ has also been studied in a computational (rather than statistical) setting. The idea is that if ψ is intractable but p is known and $\{X_t\}$ can be simulated from p, then $\frac{1}{n}\sum_{t=1}^n p(X_t, \cdot)$ can be used as an approximation of ψ. In this setting, $\frac{1}{n}\sum_{t=1}^n p(X_t, \cdot)$ is called the look-ahead estimator of ψ, and was introduced by Henderson and Glynn (2001). Functional $\sqrt{n}$-consistency of the look-ahead estimator was proved by Stachurski and Martin (2008).

by taking $\mathcal{P}$ to be a parametric family of Markov models, indexed by a vector $\theta \in \Theta$. In particular, for each θ, we take $p(\theta, x, y) \in \mathcal{P}$ to be a density kernel, and let $\psi(\theta, y)$ be the corresponding stationary density. Taking $\{\hat{\theta}_n\}$ to be a $\sqrt{n}$-consistent estimator of the true parameter under the null, we propose a test statistic based on the L_2 norm of

$$\frac{1}{n}\sum_{t=1}^{n} p(\hat{\theta}_n, X_t, \cdot) - \psi(\hat{\theta}_n, \cdot)$$

Conveniently, the asymptotic distribution of the test statistic turns out to be independent of the asymptotic distribution of the estimator $\hat{\theta}_n$. This is in contrast to the Kolmogorov and Cramér-von Mises statistics with estimated parameters. For those tests, the asymptotic distribution of the test statistic depends in a nontrivial way on the asymptotic distribution of $\sqrt{n}(\hat{\theta}_n - \theta_0)$ except in certain special cases.

While our test with estimated parameters is similar in spirit to the test proposed by Aït-Sahalia (1996), from a theoretical perspective our test is better understood as an infinite-dimensional Hansen–Sargan J-test. To see this, recall that the Hansen–Sargan test begins with a moment restriction of the form $\mathbf{E} g(X_t, \theta) = 0$ for some function g. The null hypothesis of the test is

$$H_0\colon\ \exists\, \theta \in \Theta \ \text{ such that } \ \mathbf{E} g(X_t, \theta) = 0 \tag{3}$$

The null hypothesis is rejected if

$$n \left\| \frac{1}{n}\sum_{t=1}^{n} g(X_t, \hat{\theta}_n) \right\|_W^2 \tag{4}$$

is large relative to a particular χ^2 distribution, where $\| \cdot \|_W$ is a weighted euclidean norm.

To formulate our test in a parallel manner, let $\bar{p}(\theta, x, y) := p(\theta, x, y) - \psi(\theta, y)$. In the parametric case we consider here, the null hypothesis that the data is generated by $p(\theta_0, x, y)$ for some $\theta_0 \in \Theta$. If $\{X_t\}$ has this property and is stationary, then $X_t \sim \psi(\theta_0, \cdot)$, and hence (1) implies that

$$\mathbf{E} p(\theta_0, X_t, y) - \psi(\theta_0, y) = \mathbf{E} \bar{p}(\theta_0, X_t, y) = 0$$

Treating all y simultaneously, we can write this restriction as

$$\exists\, \theta \in \Theta \ \text{ such that } \ \mathcal{E} \bar{p}(\theta, X_t, \cdot) = 0 \tag{5}$$

where $\mathcal{E}$ is a functional expectation for random elements of L_2, and the zero on the right-hand side is the origin of L_2. This is an infinite-dimensional version of (3), and our test statistic is analogous to (4) when the norm in (4) is replaced with the L_2 norm.[3]

[3]Infinite-dimensional Hansen-Sargan J-tests were considered previously in the excellent paper of Car-

1.3. Other Literature

Specification and goodness of fit tests dated back to the χ^2 test of Pearson. Pearson-type χ^2 tests discard information through discretization. This limitation motivated the development of the Kolmogorov-Smirnov and the Cramér-von Mises tests, which match entire distributions. These and other related tests—such as the Anderson-Darling test and Kuiper's test—are distinguished by the metric that they use to assess deviation between the distribution implied by the model and the distribution implied by the data. They have been extended to the case where the hypothesized distribution contains unknown parameters by many authors, including Darling (1955), Durbin (1973), Pollard (1984) and (for a semiparametric conditional Kolmogorov test) Andrews (1997). A recent survey can be found in del Barrio *et al.* (2007).

Our interest is in time series models, where observations are dependent. For extensions of Kolmogorov-Smirnov and Craémer-von Mises type tests to the dependent setting, see, for example, Weiss (1978) and Chicheportiche and Bouchaud (2011).[4] These tests can be used to measure the fit of the data to the stationary distribution of a hypothesized DGP. The test proposed in Aït-Sahalia (1996) performs a similar comparison. The main difference is that in Aït-Sahalia (1996), the comparison is in terms of L_p deviation between densities, with nonparametric kernel density estimation used to form the empirical stationary density.

1.4. Comments on the test statistic

As in Aït-Sahalia (1996), our test compares steady state implications of the model and the data based on L_p deviation between densities. Thus, our test concentrates power against alternatives with stationary distributions that differ from the stationary distribution of the null.[5] Use of L_p deviation between stationary densities is rather natural in a decision-theoretic setting. For example, consider a setting where an agent chooses an optimal action a^* by minimizing an expected loss function of the form

$$\ell(a, \psi) = \int L(a, y)\psi(y)dy$$

---

rasco and Florens (2000). Their focus is mainly on estimation and IID observations. (Ours is on testing a rather specific class of null hypotheses with dependent observations.) Unfortunately, their theoretical results on the Hansen-Sargan J-test cannot be used here, since we permit the state space for the Markov process to be multidimensional, we permit the parameters to be estimated by any $\sqrt{n}$-consistent technique and our data is explicitly Markovian under the null.

[4]Other approaches to goodness of fit tests for dependent data can be found in Bai (2003), Chen *et al.* (2008) and Neumann and Paparoditis (2008).

[5]The Hansen–Sargan J-test can also be regarded as a test of steady state implications in the time-series setting. The potential benefits of matching steady state implications are discussed in Aït-Sahalia, Hansen and Scheinkman (2010).

where ψ is the density of a vector of relevant state variables, and $L(a, y)$ is subjective loss from choosing action a when the realized state is y (see, e.g., Diebold *et al.*, 1998). Suppose now that we have a given model m which implies density ψ_m for the state. The true and unknown density for the state we denote by ψ_0. In this setting, we wish to know whether the optimal action $\operatorname{argmin}_a \ell(a, \psi_m)$ implied by the model is close to the optimal action $\operatorname{argmin}_a \ell(a, \psi_0)$ under the true distribution ψ_0. Bounding the deviation between these minimizers requires a uniform bound on the deviation between $\ell(\cdot, \psi_m)$ and $\ell(\cdot, \psi_0)$. Such a bound can be obtained via Hölder's inequality, which yields

$$\begin{aligned}\sup_a |\ell(a, \psi_m) - \ell(a, \psi_0)| &= \sup_a \left| \int L(a, y)(\psi_m(y) - \psi_0(y)) dy \right| \\ &\leq \sup_a \left[\int L(a, y)^q dy \right]^{1/q} \left[\int (\psi_m(y) - \psi_0(y))^p dy \right]^{1/p}\end{aligned}$$

where q is a constant satsifying $1/q + 1/p = 1$. The term on the far right is the L_p deviation between ψ_m and ψ_0.[6]

2. SET UP

We consider stochastic processes taking values in an arbitrary state space $\mathbb{X}$, with countably generated σ-algebra $\mathcal{X}$ and σ-finite measure $\mu \colon \mathcal{X} \to \mathbb{R}_+$. To simplify notation, we use symbols such as dx and dy to indicate integration with respect to μ, rather than $\mu(dx)$ and $\mu(dy)$. Two common settings are where

1. $\mathbb{X}$ is a Borel subset of $\mathbb{R}^k$ and μ is Lebesgue measure.
2. $\mathbb{X}$ is discrete and μ is the counting measure.

A density on $\mathbb{X}$ is any $\mathcal{X}$-measurable $f \colon \mathbb{X} \to \mathbb{R}_+$ with $\int f(x)\, dx = 1$. A *density kernel* on $\mathbb{X}$ is an $\mathcal{X} \otimes \mathcal{X}$-measurable function $p \colon \mathbb{X} \times \mathbb{X} \to \mathbb{R}_+$ such that $p(x, \cdot)$ is a density on $\mathbb{X}$ for all $x \in \mathbb{X}$. An $\mathbb{X}$-valued stochastic process $\{X_t\}$ will be called *p-Markov* if it is a stationary Markov process with transition density p, in the sense that $p(X_t, \cdot)$ is the conditional density of X_{t+1} given X_t for all t.[7]

EXAMPLE 2.1 Let $\mathbb{X} = \mathbb{R}^k$, let $\mathcal{X}$ be the Borel sets, and let μ be Lebesgue measure. Consider a stationary nonlinear AR(1) process

$$X_{t+1} = g(X_t) + W_{t+1} \qquad (W_t)_{t \geq 1} \overset{\text{IID}}{\sim} \phi \tag{6}$$

[6] In this paper, we focus on the case $q = p = 2$, placing our analysis in a Hilbert space setting where a general asymptotic theory can be constructed.

[7] Here and below, all random variables are defined on a common probability space $(\Omega, \mathscr{F}, \mathbf{P})$.

where ϕ is a density on $\mathbb{R}^k$ and g is a measurable function from $\mathbb{R}^k$ to itself. The sequence $\{X_t\}$ in (6) is p-Markov for

$$p(x,y) := \phi(y - g(x)) \qquad ((x,y) \in \mathbb{R}^k \times \mathbb{R}^k) \tag{7}$$

EXAMPLE 2.2 Let $\mathbb{X} = \{1, \dots, N\}$, let p be a stochastic $N \times N$ matrix,[8] and let $\{X_t\}$ be a stationary Markov chain on $\mathbb{X}$ satisfying

$$\mathbf{P}\{X_{t+1} = y \mid X_t = x\} = p(x,y) \qquad ((x,y) \in \mathbb{X} \times \mathbb{X})$$

If $\mathcal{X} := \{B : B \subset \mathbb{X}\}$ and μ is the counting measure, then p is a density kernel on $\mathbb{X}$, and $\{X_t\}$ is p-Markov.

EXAMPLE 2.3 Let $\mathbb{X} = \mathbb{R}$, let $\mathcal{X}$ be the Borel sets and let μ be Lebesgue measure. Under the Vasicek model, the rate of interest X_t follows

$$dX_t = \kappa(b - X_t)dt + \sigma dW_t \tag{8}$$

where κ, b and σ are parameters, and W_t is Brownian motion. The transition probability function associated with this process is

$$q(t,x,y) := \{2\pi v(t)\}^{-1/2} \exp\left\{\frac{-(y - m(t,x))^2}{2v(t)}\right\} \tag{9}$$

where $v(t) := \sigma^2(1 - e^{-2\kappa t})/(2\kappa)$ and $m(t,x) := b + (x - b)e^{-\kappa t}$. One unit of time corresponds to one year. If $\{X_t\}_{t=1}^n$ is a sequence of monthly observations from the process (8), then $\{X_t\}_{t=1}^n$ is p-Markov for $p(x,y) := q(1/12, x, y)$.

Returning to the general case, let kernel p be given, and consider a p-Markov process $\{X_t\}$ on $\mathbb{X}$. The conditional distribution of X_t given $X_0 = x$ is represented by the t-th order density $p^t(x, \cdot)$, where $p^1 := p$ and

$$p^t(x,y) := \int p(x,z)p^{t-1}(z,y)dz \qquad ((x,y) \in \mathbb{X} \times \mathbb{X})$$

A density ψ on $\mathbb{X}$ is called *stationary* with respect to p if $\psi(y) = \int p(x,y)\psi(x)dx$ for all $y \in \mathbb{X}$. In all cases we consider, p will have a unique stationary density ψ. In this setting, we define

$$\bar{p}^t(x,y) := p^t(x,y) - \psi(y) \qquad (t \in \mathbb{N}, \ \ (x,y) \in \mathbb{X} \times \mathbb{X})$$

If $\{X_t\}$ is p-Markov, then $X_t \sim \psi$ for all $t \geq 0$.

[8] I.e., $p(x,y) \geq 0$ for each $(x,y) \in \mathbb{X} \times \mathbb{X}$, and $\sum_{y \in \mathbb{X}} p(x,y) = 1$ for each $x \in \mathbb{X}$.

2.1. *Ergodicity Assumptions*

Our test relies on a central limit theorem, which in turn depends on the properties of the underlying Markov process. We will assume that that process is V-uniformly ergodic. V-uniformly ergodicity is satisfied in a wide variety of applications.[9] It also implies relatively general laws of large numbers and central limit theorems. To define V-uniform ergodicity, let V be an $\mathcal{X}$-measurable map from $\mathbb{X}$ to $[1, \infty)$. For measurable $f : \mathbb{X} \to \mathbb{R}$, let

$$\|f\|_V := \sup_{|h| \leq V} \left| \int f(x) h(x) dx \right|$$

Given V, a kernel p with stationary density ψ is called V-*unformly ergodic* if it has a unique stationary density ψ and

$$\sup_{x \in \mathbb{X}} \frac{\|p^t(x, \cdot) - \psi\|_V}{V(x)} \to 0 \qquad (t \to \infty) \tag{10}$$

(See, e.g., Meyn and Tweedie, 2009, p. 392). If $V \equiv 1$, then the term to the right of the supremum reduces to the L_1 deviation between the time t density $p^t(x, \cdot)$ and the stationary density ψ. This is the uniformly ergodic case. If p is V-uniformly ergodic, then ψ satisfies

$$\int V(x) \psi(x) dx < \infty \tag{11}$$

We focus on density kernels satisfying the following assumption:

ASSUMPTION 2.1 There exists a measurable function $V : \mathbb{X} \to [1, \infty)$ such that p is V-uniformly ergodic and

$$\int p(x, y)^2 dy \leq V(x) \qquad \forall\, x \in \mathbb{X} \tag{12}$$

EXAMPLE 2.4 Consider p in example 2.1. Let $\| \cdot \|$ be any norm on $\mathbb{R}^k$. If g and ϕ are both continuous, ϕ is strictly positive on $\mathbb{R}^k$, and there exist constants $\alpha \in [0, 1)$ and $\beta \in \mathbb{R}_+$ such that $\|g(x)\| \leq \alpha \|x\| + \beta$ for all $x \in \mathbb{R}^k$, then, given any constant $c \geq 1$, the kernel p is V-uniformly ergodic for $V(x) := \|x\| + c$.[10] If $s := \int \phi(y)^2 dy < \infty$, then assumption 2.1 is satisfied when $c \geq \max\{s, 1\}$.

[9] Kristensen (2007) gives conditions for a number of popular time-series models, including (nonlinear) ARMA, bilinear, GARCH and random coefficient models. Nishimura and Stachurski (2005) demonstrate V-uniform ergodicity of the one-sector stochastic optimal growth model under the classical assumptions. Meyn and Tweedie (2009, chapter 16) provide a general treatment.

[10] For details, see Meyn and Tweedie, 2009, prop 6.1.5, thm. 6.2.9, and thm. 16.1.2. The continuity and positivity assumptions can be weaked significantly.

EXAMPLE 2.5 If the discrete Markov chain in example 2.2 is irreducible and aperiodic, then assumption 2.1 is satisfied.

EXAMPLE 2.6 The Vasicek density kernel p satisfies assumption 2.1 whenever $\kappa > 0$. The unique stationary density ψ is $N(b, \sigma^2/(2\kappa))$.

2.2. *The Test Variance-Covariance Function*

Let L_2 be the set of all $\mathcal{X}$-measurable functions h mapping $\mathbb{X}$ to $\mathbb{R}$ with $\int h(x)^2 dx < \infty$. As usual, elements of L_2 equal μ-almost everywhere are identified. The inner product and norm are defined by

$$\langle g, h \rangle = \int g(x)h(x)dx \quad \text{and} \quad \|h\| := \langle h, h \rangle^{1/2}$$

respectively. Since we have assumed that $\mathcal{X}$ is countably generated, the space $(L_2, \|\cdot\|)$ is separable.

LEMMA 2.1 *Let p be a density kernel, and let ψ be its stationary density. If p satisfies assumption 2.1, then $\psi \in L_2$, $p(x, \cdot) \in L_2$ and $\bar{p}(x, \cdot) \in L_2$ for all $x \in \mathbb{X}$. Moreover, if X is any $\mathbb{X}$-valued random variable, then $y \mapsto p(X, y)$ is an L_2-valued random variable.*[11]

Each p satisfying assumption 2.1 defines a real-valued function γ on $\mathbb{X} \times \mathbb{X}$ by

$$\begin{aligned}(13) \qquad \gamma(y, y') := &\int \bar{p}(x, y)\bar{p}(x, y')\psi(x)dx \\ &+ \sum_{t=2}^{\infty} \left\{ \int \bar{p}(x, y)\bar{p}^t(x, y')\psi(x)dx + \int \bar{p}(x, y')\bar{p}^t(x, y)\psi(x)dx \right\}\end{aligned}$$

As shown below, γ is the covariance function of the mapping $y \mapsto n^{-1/2} \sum_t^n \bar{p}(X_t, y)$ under the null hypothesis of our test. This covariance function plays a key role in what follows. An example is given in figure 2, corresponding to the Vasicek density kernel

$$(14) \qquad p(x, y) := q(1/12, x, y), \quad \kappa = 0.85837, \ b = 0.089102, \ \sigma^2 = 0.0021854$$

where q is defined in (9). These parameter values are estimated from US short rate data in Aït Sahalia (1996).

Let $C \colon L_2 \to L_2$ be the integral operator corresponding to γ. That is,

$$(15) \qquad Ch(y') := \int \gamma(y, y')h(y)dy \qquad (h \in L_2)$$

[11]The statement that $p(X, \cdot)$ is an L_2-valued random variable includes the claim that $\Omega \ni \omega \mapsto p(X(\omega), \cdot) \in L_2$ is also measurable. See the proof for details.

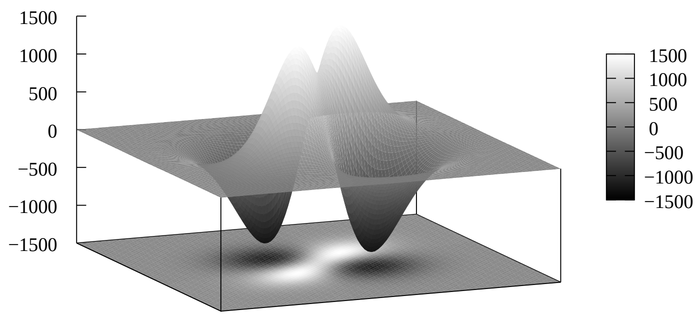

FIGURE 2.— The function $\gamma(y, y')$ for the Vasicek model

LEMMA 2.2 *The operator C is positive, symmetric and Hilbert-Schmidt.*

As a consequence of lemma 2.2, we can apply the spectral theorem in L_2 to obtain the decomposition

$$Ch = \sum_{\ell=1}^{\infty} \lambda_\ell \langle h, v_\ell \rangle v_\ell \qquad (h \in L_2) \tag{16}$$

Here $(v_\ell)_{\ell \geq 1}$ is an orthonormal basis of L_2 consisting of eigenfunctions of C, and $(\lambda_\ell)_{\ell \geq 1}$ is the corresponding eigenvalues (i.e., $Cv_\ell = \lambda_\ell v_\ell$ for all ℓ). The eigenvalues are real, nonnegative, and satisfy $\sum_{\ell \geq 1} \lambda_\ell < \infty$. The function γ, the operator C, the eigenfunctions $(v_\ell)_{\ell \geq 1}$ and the eigenvalues $(\lambda_\ell)_{\ell \geq 1}$ are all determined by the density kernel p.

3. GOODNESS OF FIT FOR MARKOV PROCESSES

We begin discussion of the test in this section by looking at a simple null hypothesis, corresponding to the statement that the data is generated by a particular density kernel p. (The case of simple null is mainly of theoretical interest. The case of composite null is treated in from section 4 on.) Taking p to be a fixed density kernel satisfying assumption 2.1, suppose that we have n observations of an $\mathbb{X}$-valued stochastic process $\{X_t\}$. Our null hypothesis is:

$$H_0 : \{X_t\}_{t=1}^{n} \text{ is } p\text{-Markov} \tag{17}$$

THEOREM 3.1 *Let $\{Z_\ell\}_{\ell \geq 1}$ be an IID sequence of standard normal random variables, and let $\{\lambda_\ell\}_{\ell \geq 1}$ be the eigenvalues of the operator C defined in* (15). *Under H_0 we have*

$$\frac{1}{n} \int \left\{ \sum_{t=1}^{n} \bar{p}(X_t, y) \right\}^2 dy \xrightarrow{d} \sum_{\ell=1}^{\infty} \lambda_\ell Z_\ell^2 \qquad \text{as } n \to \infty \tag{18}$$

It follows from (18) that if $\alpha \in (0,1)$ and c_α is the $1-\alpha$ quantile of $\sum_\ell \lambda_\ell Z_\ell^2$, then the test

$$\text{(19)} \qquad \text{Reject } H_0 \text{ if } \quad \frac{1}{n}\int \left\{ \sum_{t=1}^{n} \bar{p}(X_t, y) \right\}^2 dy > c_\alpha$$

is asymptotically of size α. The integral on the left-hand size of (19) can be computed by numerical integration. Computation of c_α is discussed in section 3.1.

REMARK 3.1 The asymptotic distribution of our test statistic is an infinite weighted sum of independent $\chi^2(1)$ random variables, where the weights correspond to the spectrum of a certain covariance operator. The Cramér-von Mises test statistic also has this property, both in the independent (Smirnov, 1936) and dependent (Chicheportiche and Bouchaud, 2011) observation cases. The difference for all of these tests is in the particular covariance operator—in our case, it is the operator C defined in (15).

REMARK 3.2 It is interesting to note that, unlike the classical goodness of fit tests such as the Pearson, Kolmogorov and Cramér-von Mises tests, our test has no obvious equivalent in the IID case. In particular, while our test is formally well-defined when the null hypothesis states that the data is IID, an IID null corresponds to the case $p(x,\cdot) = \psi(\cdot)$ for all x, or $\bar{p} = 0$. When $\bar{p} = 0$ the test statistic in (19) is identically equal to zero, and the test has zero power against all alternatives.

REMARK 3.3 Our test is not distribution free: the asymptotic distribution of the test statistic depends on the null hypothesis. For other distributional goodness of fit tests such as the Kolmogorov and Cramér-von Mises tests, the tests are either distribution free or modifications have been proposed that generate this property. However, this is for the IID case. When dependence is present, the distribution free property is more problematic (see, e.g., Chicheportiche and Bouchaud, 2011). Moreover, practical problems usually involve estimated parameters, and when estimated parameters are present all of these tests lack the distribution free property, except in very special cases.

3.1. Computing Critical Values

As shown in theorem 3.1, the asymptotic distribution of the test statistic depends on $\{\lambda_\ell\}_{\ell \geq 1}$, the eigenvalues of the operator C corresponding to the function γ defined in (13). In principle, these eigenvalues can be calculated using standard numerical techniques for solving linear operator equations, such as Galerkin projection. However, the simplest technique for computing the critical value c_α in (19) is to simulate the test statistic under the null, as in algorithm 1.

Algorithm 1: Approximates c_α corresponding to given α and kernel p

Fix M, T to be large integers ;
for $m \in \{1, \ldots, M\}$ **do**
  simulate a p-Markov time series $X_1^s, \ldots, X_T^s$;
  set $J_m \leftarrow T^{-1} \int \{\sum_{t=1}^T \bar{p}(X_t^s, y)\}^2 dy$;
end
return the $1 - \alpha$ quantile of $J_1, \ldots, J_M$;

A size-adjusted test is produced by setting $T = n$. In this case the observations $\{J_m\}$ produced by the algorithm are IID draws from the distribution of the test statistic under the null hypothesis, and hence the $1 - \alpha$ empirical quantile of these observations converges in probability to the $1 - \alpha$ quantile of its distribution as $M \to \infty$.

3.2. *Local Alternatives*

In this section we investigate the power of the test against $1/\sqrt{n}$ local alternatives. In particular, the test we consider is

$$H_0 : \{X_t\}_{t=1}^n \text{ is } p\text{-Markov} \qquad \text{vs} \qquad H_L : \{X_t\}_{t=1}^n \text{ is } p_n\text{-Markov}$$

where p is a fixed kernel satisfying assumption 2.1, and $\{p_n\}$ is the sequence of kernels $p_n(x, y) := p(x, y) + k(x, y)/\sqrt{n}$ for some fixed $k \colon \mathbb{X} \times \mathbb{X} \to \mathbb{R}$. To ensure that p_n is a density kernel, we require $\int k(x, y) dy = 0$ for all x. We set

$$Y_n(y) := \frac{1}{\sqrt{n}} \sum_{t=1}^n \bar{p}(X_t, y) \qquad (y \in \mathbb{X})$$

Y_n is a random element of L_2, and the squared norm of Y_n is the test statistic in (19).

ASSUMPTION 3.1 The functions k and p jointly satisfy the third moment condition

$$\mathbf{E} \sup_{\delta \in [0,1]} \frac{|k(X_1, X_2)|^3}{|p(X_1, X_2) + \delta\, k(X_1, X_2)|^3} < \infty$$

THEOREM 3.2 *Let τ be the element of L_2 defined by*

$$\tau(y) := \sum_{t=1}^{\infty} \mathbf{E} \left\{ \bar{p}(X_{t+1}, y) \frac{k(X_1, X_2)}{p(X_1, X_2)} \right\}$$

and let C be the operator in (15) *corresponding to p. If H_L and assumption 3.1 both hold, then $\{Y_n\}$ converges in distribution to $N(\tau, C)$.*

REMARK 3.4 Theorem 3.2 implies non-trivial power for the test (19) whenever $\tau \neq 0$. The reason is that the test statistic is the squared norm of Y_n. In the proof of theorem 3.1 it is shown that under H_0 the sequence $\{Y_n\}$ converges in distribution to $N(0, C)$. Theorem 3.2 tells us that under H_L it converges instead to $N(\tau, C)$.

In assumption 3.1 and in the definition of τ in theorem 3.2, the expectation is taken under H_0. The exact meaning of the claim in theorem 3.2 can be clarified as follows: Let $(\Omega_n, \mathscr{F}_n)$ be the product space $\mathbb{X}^n := \times_{t=1}^n \mathbb{X}$ with its product σ-algebra, let $X_t \colon \Omega_n \to \mathbb{X}$ be the projection $X_t(x_1, \ldots, x_n) = x_t$, let $\mathbf{P}_n$ be the distribution of $(X_1, \ldots, X_n)$ over $\mathbb{X}^n$ constructed from p in H_0, and let $\mathbf{Q}_n$ be the distribution on $\mathbb{X}^n$ constructed from the local alternative p_n. (Construction of $\mathbf{P}_n$ and $\mathbf{Q}_n$ from their respective kernels is via the standard definition—see, e.g., Meyn and Tweedie, ch. 3, 2009.) The claim in theorem 3.2 is that, for all continuous bounded $g \colon L_2 \to \mathbb{R}$, we have $\int g(Y_n) d\mathbf{Q}_n \to \int g d\nu$ as $n \to \infty$, where ν is the L_2 Gaussian $N(\tau, C)$.

The proof of theorem 3.2 uses a contiguity argument, based on an Hilbert space extension of Le Cam's third lemma. All additional details are given in section 7.

3.3. *Simulation of ψ*

In applications, the stationary density ψ that forms part of the test statistic (18) may be intractable. In this case, one possibility is to approximate ψ via simulation. To implement this idea, consider again the setting of theorem 3.1. Fix $k \in \mathbb{N}$, and let $\{X_t'\}_{t=1}^{kn}$ be a simulated p-Markov sequence that is independent of the data $\{X_t\}_{t=1}^n$. For each $k \in \mathbb{N}$ we have the following result:

THEOREM 3.3 *Let $\{Z_\ell\}_{\ell \geq 1}$ be an* IID *sequence of standard normal random variables. If the conditions of theorem 3.1 hold, then, as $n \to \infty$,*

$$\frac{1}{n} \int \left\{ \sum_{t=1}^{n} p(X_t, y) - \frac{1}{k} \sum_{t=1}^{nk} p(X_t', y) \right\}^2 dy \xrightarrow{d} (1 + 1/k) \sum_{\ell=1}^{\infty} \lambda_\ell Z_\ell^2 \tag{20}$$

In particular, the limit $(1 + 1/k) \sum_{\ell=1}^{\infty} \lambda_\ell Z_\ell^2$ of the simulation-based test statistic converges almost surely to that of the original test statistic (29) as the length of the simulation run converges to infinity.

4. A SPECIFICATION TEST FOR PARAMETRIC CLASSES

The test (19) corresponds to the simple null H_0 in (17), and represents a goodness of fit test for individual models. This test is mainly of theoretical interest. A more practical setting is where we have a parametric class of models, and we wish to test the hypothesis that the data is generated by some model in this class. In this case we need to augment our asymptotic theory to accommodate estimates of parameters.

4.1. *The Test*

Let Θ be a compact convex subset of $\mathbb{R}^M$, and let $\{p_\theta\}_{\theta \in \Theta}$ be a parametric family of density kernels, all satisfying assumption 2.1. Let ψ_θ be the unique stationary density corresponding to p_θ. When convenient, we write $p(\theta, x, y)$ instead of $p_\theta(x, y)$, and $\psi(\theta, y)$ in place of $\psi_\theta(y)$. In addition, let

$$\bar{p}(\theta, x, y) := p(\theta, x, y) - \psi(\theta, y) \qquad (\theta \in \Theta, \;\; (x, y) \in \mathbb{X} \times \mathbb{X})$$

Finally, we use the following notation: For each $\theta \in \Theta$,

- $C(\theta)$ is the operator (15) corresponding to p_θ, and
- $(\lambda_\ell(\theta))_{\ell \geq 1}$ is the sequence of eigenvalues for $C(\theta)$, as defined by (16).

Consider the null hypothesis

$$H_0 \colon \text{ the data } \{X_t\}_{t=1}^n \text{ is } p_\theta\text{-Markov for some } \theta \in \Theta \tag{21}$$

When the null is assumed to hold, we let $\theta_0 \in \Theta$ denote the true value of θ. In the assumptions that follow, $\| \cdot \|_E$ denotes the Euclidean norm in $\mathbb{R}^M$, as opposed to $\| \cdot \|$, the norm in L_2, and V is the function corresponding to $p(\theta_0, \cdot, \cdot)$ in assumption 2.1.

ASSUMPTION 4.1 There exists a open neighborhood U of θ_0 such that the vector of partial derivatives

$$D\bar{p}(\theta, x, y) := [D_m \bar{p}(\theta, x, y)]_{m=1}^M := \left[\frac{\partial}{\partial \theta_m} \bar{p}(\theta, x, y) \right]_{m=1}^M$$

exists for $x, y \in \mathbb{X} \times \mathbb{X}$ and all $\theta \in U$.

ASSUMPTION 4.2 There exists a function $K_1 \colon \mathbb{X} \times \mathbb{X} \to \mathbb{R}$ with $\int K_1(x, y)^2 dy \leq V(x)$ and $\|D\bar{p}(\theta_0, x, y)\|_E \leq K_1(x, y)$ for all $(x, y) \in \mathbb{X} \times \mathbb{X}$.

ASSUMPTION 4.3 There exists an $\alpha > 0$ and $K_2 \colon \mathbb{X} \times \mathbb{X} \to \mathbb{R}$ with $\int K_2(x, y)^2 dy \leq V(x)$ and $\|D\bar{p}(\theta, x, y) - D\bar{p}(\theta', x, y)\|_E \leq K_2(x, y) \|\theta - \theta'\|_E^\alpha$ for all $\theta, \theta' \in U$ and all $(x, y) \in \mathbb{X} \times \mathbb{X}$.

We can now state our main result concerning asymptotic distributions:

THEOREM 4.1 *Let $\{\hat{\theta}_n\}$ be a $\sqrt{n}$-consistent sequence of estimators for θ_0, in the sense that $(\hat{\theta}_n - \theta_0) = O_P(n^{-1/2})$ whenever H_0 is true. Let $\{Z_\ell\}_{\ell \geq 1}$ be an IID sequence of standard normal random variables, and let assumptions 4.1–4.3 all hold. Under the null hypothesis H_0 we have*

$$n \int \left\{ \frac{1}{n} \sum_{t=1}^{n} \bar{p}(\hat{\theta}_n, X_t, y) \right\}^2 dy \xrightarrow{d} \sum_{\ell=1}^{\infty} \lambda_\ell(\theta_0) Z_\ell^2 \qquad \text{as } n \to \infty \tag{22}$$

For each $\theta \in \Theta$, let $c_\alpha(\theta)$ be the $1-\alpha$ quantile of the random variable $\sum_{\ell=1}^{\infty} \lambda_\ell(\theta) Z_\ell^2$, where $(Z_\ell)_{\ell \geq 1}$ is IID and standard normal. In view of theorem 4.1, a test rejecting H_0 when the left-hand side of (22) exceeds $c_\alpha(\theta_0)$ is asymptotically of size α. However, θ_0 is not observable, and hence $c_\alpha(\theta_0)$ cannot be evaluated. Instead, we approximate it with $c_\alpha(\hat{\theta}_n)$. This gives the test

$$\text{Reject } H_0 \text{ if } \quad \frac{1}{n} \int \left\{ \sum_{t=1}^{n} \bar{p}(\hat{\theta}_n, X_t, y) \right\}^2 dy > c_\alpha(\hat{\theta}_n) \tag{23}$$

THEOREM 4.2 *If the conditions of theorem 4.1 hold and c_α is continuous at θ_0, then the test* (23) *is asymptotically of size α.*

REMARK 4.1 The critical value $c_\alpha(\hat{\theta}_n)$ in (23) can be computed by the numerical methods discussed in section (3.1), replacing p with $p_{\hat{\theta}_n}$.

REMARK 4.2 As mentioned in the introduction and verified in theorem 4.1, a convenient feature of this test is that the asymptotic distribution of the test statistic does not depend on the asymptotic distribution of $\sqrt{n}(\hat{\theta}_n - \theta_0)$.

4.2. Consistency of the Test

The test (23) is not consistent against all alternatives in the negation of the null hypothesis specified in (21). In essence, the test compares Markov models by their stationary distribution, and models with identical stationary distributions cannot be distinguished. However, if we consider our test as an infinite dimension Hansen–Sargan test, with null hypothesis given in 5, and the alternative by

$$H_1\colon \inf_{\theta \in \Theta} \| \mathcal{E} \bar{p}(\theta, X_t, \cdot) \| > 0$$

then the test becomes consistent whenever the following assumptions hold:

ASSUMPTION 4.4 Under H_1, the sequence $\{X_t\}$ is stationary and ergodic. In particular, the sample mean $\frac{1}{n}\sum_{t=1}^{n} h(X_t)$ converges in probability to the expectation $\mathcal{E}h(X_t)$ for all measurable $h\colon \mathbb{X} \to L_2$ such that $\mathcal{E}h(X_t)$ exists.

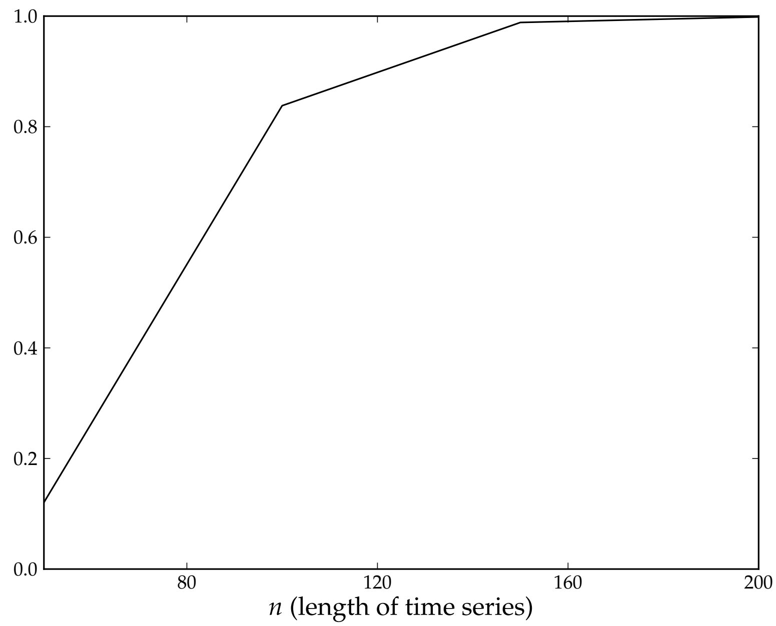

FIGURE 3.— Rejection frequency, nonstationary alternative

ASSUMPTION 4.5 The sequence $\hat{\theta}_n$ converges in probability under H_1.

ASSUMPTION 4.6 The vector of partial derivatives $Dp(\theta, x, y)$ exists for all x, y in $\mathbb{X}$ and all $\theta \in \Theta$. Moreover, there exists a function $\Lambda \colon \mathbb{X} \times \mathbb{X} \to \mathbb{R}$ such that $\mathbf{E} \int \Lambda(X_t, y)^2 dy$ is finite under H_1 and $\|Dp(\theta, x, y)\|_E \leq \Lambda(x, y)$ for all $(x, y) \in \mathbb{X} \times \mathbb{X}$ and $\theta \in \Theta$.

ASSUMPTION 4.7 The expectation $\mathbf{E} \int p(\theta, X_t, y)^2 dy$ is finite under H_1 for all $\theta \in \Theta$.

THEOREM 4.3 *If H_1 is valid and assumptions 4.4–4.7 hold, then the probability that the test* (23) *rejects H_0 converges to one as $n \to \infty$.*

The proof of theorem 4.3 can be found in section 7. Note that the conditions of the theorem are sufficient but by no means necessary for consistency. While assumption 4.4 requires a stationary and ergodic alternative, intuition suggests that most nonstationary alternatives are likely to be rejected with probability one when the sample size is large. For example, let p in H_0 be the fixed kernel given by (14), and for the alternative take the same model but with $\kappa = 0$. The rejection probabilities for data sizes $n = 50, 100, 150, 200$ are shown in figure 3. By $n = 200$ the rejection probability is one.[12]

[12]Rejection probabilities were calculated by averaging over 2,000 observations.

5. DISCUSSION

In this section we present simulations that illustrate several features of the test.

5.1. *Properties of the Test under H_0*

In the introduction we indicated the similarities between our test and the test proposed by Aït-Sahalia (1996), both of which evaluate ergodic Markov models based on L_2 comparison of densities. Aït-Sahalia's test is a seminal contribution to the literature and his results have initiated an important line of research. However, it was argued by Pritsker (1998) that Aït-Sahalia's test statistic requires very large data sizes to attain its asymptotic distribution, causing excessively high rejection rates both under the null and under the alternative when the asymptotic critical value is adopted.

Our test provides a new perspective on this problem. On one hand, our test implements essentially the same idea as Aït-Sahalia's test (i.e, comparison of stationary densities using L_2 norm). On the other hand, intuition suggests that it will have lower size distortion in finite samples, since our test statistic contains information about the autocorrelation structure of the null via the density kernel p, and the test has nontrivial power against $1/\sqrt{n}$ local alternatives (theorem 3.2).

To investigate this idea and compare size distortions in our test and Aït-Sahalia's test, we conducted an experiment to re-examine the critique of Pritsker (1998). The experiment investigated rejection rates under a true null when the sample size is relatively small and the asymptotic critical value is used. Following Pritsker, the underlying model in the experiment was the Vasicek model of interest rates. For the DGP that generates the data $\{X_t\}$ we chose the particular Vasicek model given in (14), while for H_0 we hypothesed (correctly) that the data was generated by some Vasicek model.

Beginning with Aït-Sahalia's test, we computed the asymptotic critical value of the test (see Aït-Sahalia, 1996, p. 393) when $\alpha = 0.05$, set $n = 264$ (corresponding to 22 years of monthly observations), generated 2,000 time series of length n from the DGP, and evaluated the test on each time series. Consistent with Pritsker (1998), we found that Aït-Sahalia's test rejected the true null in over 50% of our samples.[13] On the other hand, when we repeated the experiment with our test in place of Aït-Sahalia's test, our test rejected the true null in 4.7% of our samples. Thus, at least for this particular problem, the size distortion was essentially resolved by our test.[14]

[13]The bandwidth used was the optimal bandwidth for estimating the stationary density of the Vasicek model with the true parameters. We experimented with other bandwidths but all choices gave a rejection rate in excess of 50%.

[14]As with Aït-Sahalia's test, we took $\alpha = 0.05$ and $n = 264$. The DGP was the particular Vasicek

5.2. Properties of the Test under H_1

Next, we ran an experiment to study the power of our test. As above, we took H_0 as the hypothesis that the data was generated by some model in the Vasicek class. We then generated data using the level effects interest rate model

$$dX_t = \kappa(b - X_t)dt + \sigma X_t^{\delta} dB_t \qquad (0 \leq \delta \leq 0.5) \tag{24}$$

for different values of δ, and other parameters held fixed at the values given in (14). When $\delta = 0$ the Vasicek null hypothesis is true. For all other values of δ the null hypothesis is false. If $\delta = 1/2$, then (24) corresponds to the CIR model of Cox, Ingersoll and Ross (1985). Fixing $n = 264$ and $\alpha = 0.05$, we computed the power function with for values of δ ranging over the interval $[0, 0.5]$. The results are shown as the unbroken line in Figure 4. As $\delta \to 0.5$, the power converges to one.[15]

For comparison, we also ran a conditional moment test of the same null hypothesis, with the same sample size and the data generated by the same set of alternatives. After estimating the parameters of the Vasicek model using maximum likelihood and then running the regression

$$\frac{\partial \mathcal{L}(X_{t+1}, X_t)}{\partial \sigma} = \beta_0 + \beta_1 X_t + u_{t+1}$$

where $\mathcal{L}(X_{t+1}, X_t)$ is the log likelihood of (X_{t+1}, X_t), we conducted a two-sided test of $\beta_1 = 0$, an equality that holds whenever the Vasicek null hypothesis is true.[16] This test was repeated 2,000 times at each δ, and the resulting rejection rates are shown as the dashed line in figure 4. For this particular experiment, the power of the conditional moment test is much lower than that of the test proposed in this paper.[17]

model given in (14), while H_0 was that the data was generated by *some* Vasicek model. In running the experiment, we first computed the critical value c_α on the right-hand side of (23) for $\alpha = 0.05$, using algorithm 1 applied to the baseline Vasicek density kernel with parameters given in (14). (Since we wanted to compute the asymptotic critical value, we used the exact parameter values rather than estimates. When applying algorithm 1, we set $M = 2500$ and $T = 10^5$.) Next we simulated 2,000 times series $\{X\}_{t=1}^n$ from the DGP, where $n = 264$. For each of these time series, we used OLS to obtain an estimate $\hat{\theta}_n$ for the vector of parameters of the Vasicek model. With the resulting density kernel $p_{\hat{\theta}_n}$, we evaluated the test statistic on the left-hand side of (23) and compared it to the critical value. Of the 2,000 times series we generated, 4.7% of the test outcomes were rejections.

[15] Figure 4 gives the rejection frequency over 2,000 simulated time series . At each iteration, parameters of the Vasicek model were estimated from the time series using OLS, and the test given in (23) was evaluated. Since OLS estimates are consistent for the parameters of the Vasicek model under the null, theorem 4.1 applies.

[16] Further discussion of the test is given in Pritsker (1998, p. 462).

[17] In the conditional moment test we used a standard t-test with OLS standard errors. Following Pritsker, we also repeated our experiment with GMM standard errors but obtained almost identical results. (The power of the test was slighty lower.)

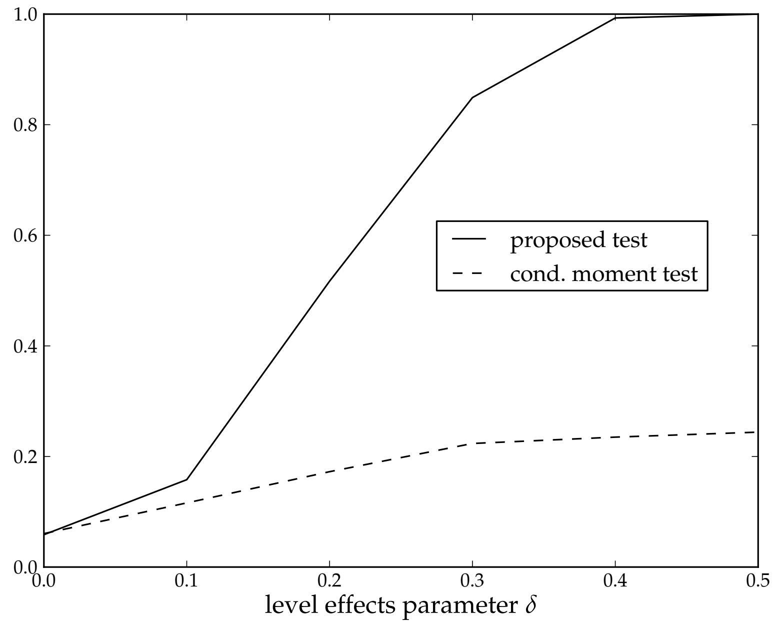

FIGURE 4.— Power against level effects model when $\alpha = 0.05$ and $n = 264$

These results are interesting for two reasons. First, Pritsker's results led him to conclude that the conditional moment test was "far more powerful [than Aït-Sahalia's test] for distinguishing between the Vasicek null hypothesis and the CIR alternative." He went on to say that "The modest power of Aït-Sahalia's test suggests that the marginal [stationary] density may not be estimated precisely enough for a test based on the marginal density alone to be able to distinguish among various short rate models" (Pritsker, 1989, p. 462). While our results in no way contradict Pritsker's findings, they do appear to validate Aït-Sahalia's intuition that tests based on the stationary density can be useful to distinguish between short rate models.

A second point of interest is that the conditional moment test was chosen by Pritsker because the structure of the test suggests that it should be powerful when testing a Vasicek null against a CIR alternative (Pritsker, 1989, p. 462). On the other hand, our test is a general test, which is not weighted towards any particular alternative. Nevertheless, at least for these parameter values, the power of our test is much higher.

In fact, we can further increase the power of our test against particular alternatives by using a weighting function. Recall that, although we used the symbols dx, dy to indicate integration, these symbols were introduced as short hand for the more general expressions $\mu(dx)$, $\mu(dy)$, where μ is an arbitrary σ-finite measure. (See the discussion at the start of section 2.) If we take $\mu(dy) = w(y)\lambda(dy)$ where λ is Lebesgue measure

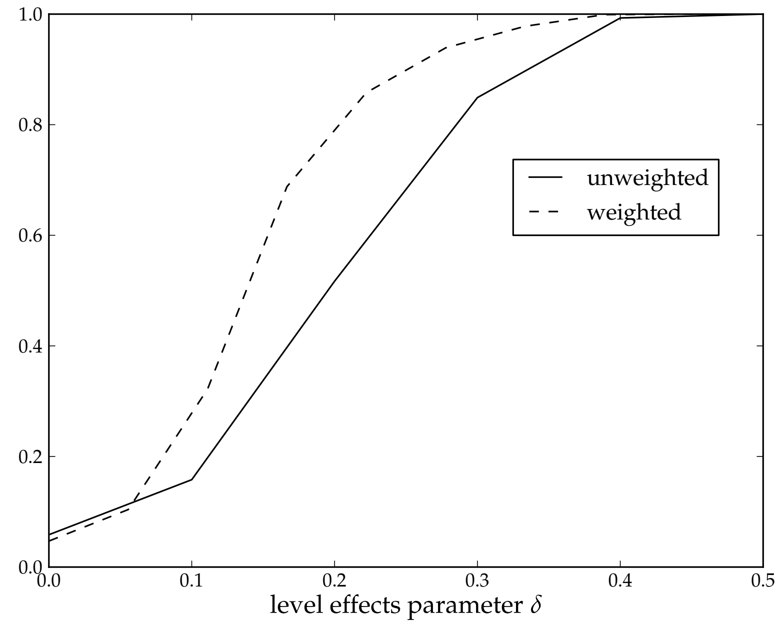

FIGURE 5.— Comparison of weighted and unweighted tests

and w is a nonnegative Lebesgue-integrable function, then all of our theory remains valid. Implementation requires only that dy is replaced by $w(y)dy$ in (23) and algorithm 1. The function w can be used as a weight function to direct power towards certain alternatives. For example, in the dashed line in figure 5, we show the effect of using the stationary density of the CIR model as the weight function w. (All other aspects of the experiment were identical to the unweighted case that was used to produce in the unbroken line shown in figure 4. The same line is replicated in figure 5.) As can be seen from figure 5, use of the CIR stationary density as a weight function leads to a considerable increase in power against the level effect alternatives.[18]

6. CONCLUSION

In this paper we proposed a natural goodness of fit test for ergodic Markov processes. The test can be used to test the hypothesis that a given time series is generated by a parametric class of ergodic Markov models. No alternative needs to be specified in order to implement the test. Although the test is based on comparison of densities, the test statistic contains no smoothing parameters, and the test has nontrivial power

[18]The stationary distribution of the CIR model is gamma with shape parameter $2\kappa b/\sigma^2$ and scale parameter $\sigma^2/(2\kappa)$. We used the values in (14) for the weight function.

against $1/\sqrt{n}$ local alternatives. While the test is not distribution free, the critical value can be computed consistently by simulation, and the asymptotic distribution of the test does not depend on the asymptotic distribution of the parameter estimators. Simulations in section 5 showed favorable finite sample properties.

In section 5 we briefly discussed the possibility of using weighting functions to obtain additional power against certain alternatives. The ability to apply different weighting functions should add to the usefulness of the test. Further investigation of this topic is left to future research.

7. PROOFS

We begin with a brief discussion of random variables in a separable Hilbert space $\mathscr{H}$. An $\mathscr{H}$-valued random variable F on $(\Omega, \mathscr{F}, \mathbf{P})$ is a measurable map from $(\Omega, \mathscr{F})$ into $\mathscr{H}$ paired with its Borel sets. If $\mathbf{E}\|F\| < \infty$, where $\mathbf{E}$ is the ordinary scalar expectation, then the (Pettis vector) expectation $\mathcal{E}F$ of F is the unique element of $\mathscr{H}$ satisfying $\langle \mathcal{E}F, h\rangle = \mathbf{E}\langle F, h\rangle$ for all $h \in \mathscr{H}$.[19] If $\mathbf{E}\|F\|^2 < \infty$ and $\mathcal{E}F = 0$, then its covariance operator $C \colon \mathscr{H} \to \mathscr{H}$ is defined by $Ch = \mathcal{E}\langle F, h\rangle F$ for $h \in \mathscr{H}$. Equivalently,

$$\langle g, Ch\rangle = \mathbf{E}\langle g, F\rangle\langle h, F\rangle \qquad (g, h \in \mathscr{H}) \tag{25}$$

Any covariance operator is linear, positive, symmetric and Hilbert-Schmidt.[20]

An $\mathscr{H}$-valued random variable G is called Gaussian if $\langle h, G\rangle$ is normally distributed on $\mathbb{R}$ for every $h \in \mathscr{H}$. We say that $G \sim N(m, C)$ if G is Gaussian on $\mathscr{H}$ with mean $m \in \mathscr{H}$ and covariance operator C. This is known to be equivalent to the statement $\mathbf{E}\exp(i\langle h, G\rangle) = \exp\{i\langle h, m\rangle - \langle h, Ch\rangle/2\}$ for all $h \in \mathscr{H}$, from which it is simple to obtain the characterization

$$G \sim N(m, C) \text{ on } L_2 \iff \langle G, h\rangle \sim N(\langle h, m\rangle, \langle h, Ch\rangle) \text{ for all } h \in \mathscr{H} \tag{26}$$

PROOF OF LEMMA 2.1: Evidently (12) implies that $p(x, \cdot) \in L_2$ for each $x \in \mathbb{X}$. Regarding the claim that $\psi \in L_2$, the definition of stationarity and Jensen's inequality give

$$\int \psi(y)^2 dy = \int \left[\int p(x,y)\psi(x)dx\right]^2 dy \leq \int \int p(x,y)^2\psi(x)dx\, dy$$

From assumption 2.1 and (11), we then have

$$\int \psi(y)^2 dy \leq \int \int p(x,y)^2 dy \psi(x)dx \leq \int V(x)\psi(x)dx < \infty$$

[19] Here $\mathbf{E}$ and $\mathcal{E}$ denote scalar and vector-valued expectation respectively. Existence of $\mathcal{E}F$ follows directly from the Riesz representation theorem.

[20] For definitions and a proof see, for example, Bosq (2000, theorem 1.7).

We can now see that $\bar{p}(x, \cdot) \in L_2$ for any $x \in \mathbb{X}$, because

$$\|\bar{p}(x, \cdot)\| = \|p(x, \cdot) - \psi(\cdot)\| \leq \|p(x, \cdot)\| + \|\psi\|$$

To show that $p(X, \cdot)$ is an L_2-valued random variable, we need to prove that $\Omega \ni \omega \mapsto p(X(\omega), \cdot) \in L_2$ is also measurable, in the sense that preimages of Borel subsets of L_2 are measurable in Ω. Since L_2 is separable, it follows from the Pettis measurability theorem that any mapping $\Omega \ni \omega \mapsto g(\omega) \in L_2$ is measurable whenever $\Omega \ni \omega \mapsto \langle g(\omega), h\rangle \in \mathbb{R}$ is measurable for each $h \in L_2$. Using this fact, the measurability of $\omega \mapsto p(X(\omega), \cdot)$ is easily verified. This concludes the proof of lemma 2.1. *Q.E.D.*

For the proof of theorem 3.1, we need the preliminary result that if $X \sim \psi$, then $\mathcal{E}\bar{p}(X, \cdot) = 0$. Taking $X \sim \psi$, this amounts to the claim that, for any $h \in L_2$ we have

$$\mathbf{E} \int \bar{p}(X, y) h(y) dy = 0$$

Note that for each $y \in \mathbb{X}$ we have

$$\mathbf{E}\bar{p}(X, y) = \int p(x, y)\psi(x) dx - \psi(y) = \psi(y) - \psi(y) = 0 \tag{27}$$

As a consequence, if Fubini's theorem is valid, then

$$\mathbf{E} \int \bar{p}(X, y) h(y) dy = \int \mathbf{E}\bar{p}(X, y) h(y) dy = 0 \tag{28}$$

as claimed. To check the validity of Fubini's theorem, observe that

$$\int |\bar{p}(x, y) h(y)| dy \leq \|\bar{p}(x, \cdot)\| \|h\| \leq (\|p(x, \cdot)\| + \|\psi\|) \|h\|$$

Applying assumption 2.1 and (11), we obtain

$$\mathbf{E}\|p(X, \cdot)\| \leq \left[\mathbf{E} \int p(X, y)^2 dy\right]^{1/2} \leq \left[\int V(x)\psi(x) dx\right]^{1/2} < \infty$$

Hence $\mathbf{E} \int |\bar{p}(X, y) h(y)| dy < \infty$, and Fubini's theorem is valid.

LEMMA 7.1 *If $\{X_t\}$ is a V-uniformly ergodic Markov process on $\mathbb{X}$, then $M_t = (X_t, X_{t+1})$ is a V-uniformly ergodic Markov process on $\mathbb{X} \times \mathbb{X}$.*

PROOF: To see that $\{M_t\}$ is Markov, pick any bounded measurable $h \colon \mathbb{X} \times \mathbb{X} \to \mathbb{R}$. We have

$$\mathbf{E}[h(M_t) \mid M_{t-1}, \dots, M_1] = \mathbf{E}[h(X_t, X_{t+1}) \mid X_t, X_{t-1}, \dots, X_1]$$

Applying the Markov property of $\{X_t\}$, the right-hand side of this equation is equal to $\mathbf{E}[h(X_t, X_{t+1}) \mid X_t, X_{t-1}]$. Hence $\{M_t\}$ is Markov as claimed, since we have shown that

$$\mathbf{E}[h(M_t) \mid M_{t-1}, \dots, M_1] = \mathbf{E}[h(M_t) \mid M_{t-1}]$$

Next consider ergodicity. By assumption, $\{X_t\}$ is V-uniformly ergodic, and hence there exists a function $V \colon \mathbb{X} \to [1, \infty)$ such that (10) holds.[21] Given this function V, define $\hat{V}$ on $\mathbb{X}^2$ by $\hat{V}(x, y) = V(y)$. Pick any $h \colon \mathbb{X}^2 \to \mathbb{R}$ such that $|h| \leq \hat{V}$. Fix $(x_1, x_2) \in \mathbb{X}$. Let ψ_t be the marginal density of X_t when $X_2 = x_2$. Observing that $\hat{\psi}(x, y) = \psi(x)p(x, y)$ is the stationary density of $\{M_t\}$ and $\hat{\psi}_t(x, y) = \psi_t(x)p(x, y)$ is the marginal density of M_t given $M_1 = (x_1, x_2)$, we then have

$$\begin{aligned}\left| \int h\hat{\psi}_t - \int h\hat{\psi} \right| &= \left| \int \int h(x, y)\psi_t(x)p(x, y)dxdy - \int \int h(x, y)\psi(x)p(x, y)dxdy \right| \\ &= \left| \int g\psi_t - \int g\psi \right|\end{aligned}$$

for $g(x) := \int h(x, y)p(x, y)dy$. Since

$$|g(x)| \leq \int |h(x, y)|p(x, y)dy \leq \int V(y)p(x, y)dy \leq cV(x)$$

for some constant c,[22] we can apply the definition of V-uniform ergodicity of $\{X_t\}$ to obtain

$$\left| \int h\hat{\psi}_t - \int h\hat{\psi} \right| = \left| \int g\psi_t - \int g\psi \right| = c \left| \int (g/c)\psi_t - \int (g/c)\psi \right| \leq c\|\psi_t - \psi\|_V$$

Since h is an arbitrary function satisfying $|h| \leq \hat{V}$, this implies that

$$\|\hat{\psi}_t - \hat{\psi}\|_{\hat{V}} \leq c\|\psi_t - \psi\|_V$$

Dividing through by $\hat{V}(x_1, x_2)$, we obtain

$$\frac{\|\hat{\psi}_t - \hat{\psi}\|_{\hat{V}}}{\hat{V}(x_1, x_2)} \leq c\frac{\|\psi_t - \psi\|_V}{\hat{V}(x_1, x_2)} = c\frac{\|\psi_t - \psi\|_V}{V(x_2)} \leq c \sup_{x_2 \in \mathbb{X}} \frac{\|\psi_t - \psi\|_V}{V(x_2)}$$

Taking the sup of the left hand side over all (x_1, x_2) in $\mathbb{X}^2$, and observing that the right-hand side converges to zero in t by V-uniform ergodicity of $\{X_t\}$, we conclude that $\{M_t\}$ is $\hat{V}$-uniformly ergodic. *Q.E.D.*

[21]See Meyn and Tweedie (2009, chapter 16, p. 392) for the definition of V-uniform ergodicity.

[22]Since $\{X_t\}$ is V-uniformly ergodic, existence of this constant c may be obtained by appealing to Meyn and Tweedie (2009, theorem 16.0.1 part (iv) and the bound (V4) on p. 376).

7.1. Theorems 3.1 and 3.3

PROOF OF THEOREM 3.1: We begin the proof by showing that

$$n^{-1/2} \sum_{t=1}^{n} \bar{p}(X_t, \cdot) \xrightarrow{d} \sum_{\ell \geq 1} \lambda_\ell^{1/2} Z_\ell v_\ell \tag{29}$$

in L_2.[23] To simplify notation, let $F(X_t) := \bar{p}(X_t, \cdot)$. We have established that $F(X_t)$ is an L_2-valued random variable satisfying $\mathcal{E}F(X_t) = 0$ and $\|F(x)\|^2 \leq V(x)$ for all $x \in \mathbb{X}$. By Stachurski (2010, theorem 3.1), we have

$$n^{-1/2} \sum_{t=1}^{n} F(X_t) \xrightarrow{d} N(0, C_0) \qquad (n \to \infty) \tag{30}$$

where the covariance operator C_0 is defined by

$$\begin{aligned} \langle g, C_0 h \rangle = \mathbf{E}\langle g, F(X_1^*)\rangle\langle h, F(X_1^*)\rangle \\ + \sum_{t \geq 2} \{\mathbf{E}\langle g, F(X_1^*)\rangle\langle h, F(X_t^*)\rangle + \mathbf{E}\langle h, F(X_1^*)\rangle\langle g, F(X_t^*)\rangle\} \end{aligned} \tag{31}$$

Here g and h are arbitrary elements of L_2, and $\{X_t^*\}$ is any stationary p-Markov process. We claim that $C_0 = C$, the operator defined in (15). This amounts to the claim that $\langle g, C_0 h\rangle = \langle g, Ch\rangle$ for any such g and h, or

$$\langle g, C_0 h\rangle = \int \int \gamma(y, y') g(y) h(y') dy dy'$$

where γ is the function in (13). To verify this, pick any $g, h \in L_2$, and define

$$\kappa_{st}(y, y') := \int \bar{p}^s(x, y) \bar{p}^t(x, y') \psi(x) dx$$

Letting $\langle g, \kappa_{st} h\rangle := \int \int g(y) \kappa_{st}(y, y') h(y') dy dy'$ and using the definition of γ in (13), we can now write

$$\int \int \gamma(y, y') g(y) h(y') dy dy' = \langle g, \kappa_{11} h\rangle + \sum_{t \geq 2} \{\langle g, \kappa_{1t} h\rangle + \langle g, \kappa_{t1} h\rangle\}$$

We need to show that the individual terms in this expression agree with the corresponding terms on the right-hand side of (31). We will only check that

$$\langle g, \kappa_{1t} h\rangle = \mathbf{E}\langle g, F(X_1^*)\rangle\langle h, F(X_t^*)\rangle \tag{32}$$

[23] As usual, if E is a metric space, then a sequence of E-valued random variables (Y_n) converges in distribution to an E-valued random variable Y if $\mathbf{E}g(Y_n) \to g(Y)$ for every continuous bounded $g \colon E \to \mathbb{R}$. The limit on the right-hand side of (29) is an L_2 limit, and the following proof shows that this limit exists almost surely.

since other terms are similar. By the law of iterated expectations, we have

$$\mathbf{E}\langle g, F(X_1^*)\rangle\langle h, F(X_t^*)\rangle = \mathbf{E}[\langle g, F(X_1^*)\rangle \mathbf{E}[\langle h, F(X_t^*)\rangle \mid X_1^*]] \tag{33}$$

Using the Markov property, we see that $\mathbf{E}[\langle h, F(X_t^*)\rangle \mid X_1^*]$ is equal to

$$\begin{aligned}\int \langle h, F(z)\rangle p^{t-1}(X_1^*, z)dz &= \int\int h(y)\bar{p}(z,y)dyp^{t-1}(X_1^*,z)dz \\ &= \int h(y)\left[\int \bar{p}(z,y)p^{t-1}(X_1^*,z)dz\right]dy\end{aligned}$$

Applying the definition of the k-th order density, we obtain

$$\begin{aligned}\int \bar{p}(z,y)p^{t-1}(X_1^*,z)dz &= \int\{p(z,y)-\psi(y)\}p^{t-1}(X_1^*,z)dz \\ &= \int p(z,y)p^{t-1}(X_1^*,z)dz - \psi(y) = \bar{p}^t(X_1^*,y)\end{aligned}$$

Combining these equalities, we have now shown that

$$\mathbf{E}[\langle h, F(X_t^*)\rangle \mid X_1^*] = \int h(y)\bar{p}^t(X_1^*,y)dy$$

Inserting into (33) and using the definition of κ_{1t}, we now have

$$\begin{aligned}\mathbf{E}\langle g, F(X_1^*)\rangle\langle h, F(X_t^*)\rangle &= \mathbf{E}\left[\int g(y)\bar{p}(X_1^*,y)dy\int h(y')\bar{p}^t(X_1^*,y')dy'\right] \\ &= \int\left[\int\int \bar{p}(x,y)\bar{p}^t(x,y')g(y)h(y')dydy'\right]\psi(x)dx \\ &= \int\int\left[\int \bar{p}(x,y)\bar{p}^t(x,y')\psi(x)dx\right]g(y)h(y')dydy' \\ &= \int\int \kappa_{1t}(y,y')g(y)h(y')dydy' = \langle g, \kappa_{1t}h\rangle\end{aligned}$$

The proof of (32) is now complete.

The preceding argument has shown that $C_0 = C$. Combining this result with (30), we have shown that

$$n^{-1/2}\sum_{t=1}^n F(X_t) = n^{-1/2}\sum_{t=1}^n \bar{p}(X_t,\cdot) \xrightarrow{d} N(0,C)$$

To finish the proof of (29), it remains to show that

$$G := \sum_{\ell\geq 1}\lambda_\ell^{1/2}Z_\ell v_\ell \sim N(0,C)$$

To show that G is a zero-mean Gaussian element of L_2, we must show that $\langle G, h \rangle$ is zero-mean Gaussian in $\mathbb{R}$ for every $h \in L_2$. It suffices to show that this property holds on an orthonormal subset of L_2.[24] Choosing (v_ℓ) as our orthonormal subset, we have

$$\langle G, v_\ell \rangle = \sum_k \lambda_k^{1/2} Z_k \langle v_k, v_\ell \rangle = \lambda_\ell^{1/2} Z_\ell \sim N(0, \lambda_\ell)$$

To prove that $G \sim N(0, C)$, we also need to prove that C is the covariance operator of G, or $\mathbf{E}\langle G, g \rangle \langle G, h \rangle = \langle g, Ch \rangle$ for any $g, h \in L_2$. It suffices to show the same for $h, g \in \{v_\ell\}_{\ell \in \mathbb{N}}$. Fixing $j, k \in \mathbb{N}$, we have

$$\mathbf{E}\langle G, v_j \rangle \langle G, v_k \rangle = \mathbf{E} \lambda_j^{1/2} Z_j \lambda_k^{1/2} Z_k = \lambda_k \mathbb{1}\{j = k\}$$

On the other hand, since v_k is an eigenfunction of C with eigenvalue λ_k, we have $\langle v_j, C v_k \rangle = \langle v_j, \lambda_k v_k \rangle = \lambda_k \mathbb{1}\{j = k\}$. Hence C is the covariance operator of G, and $G \sim N(0, C)$ as claimed.

Finally, consider the claim (18). We have just shown that $n^{-1/2} \sum_{t=1}^n \bar{p}(X_t, \cdot) \xrightarrow{d} G$. The continuous mapping theorem and Parseval's identity now give

$$\frac{1}{n} \int \left\{ \sum_{t=1}^n \bar{p}(X_t, y) \right\}^2 dy = \left\| n^{-1/2} \sum_{t=1}^n \bar{p}(X_t, \cdot) \right\|^2 \xrightarrow{d} \|G\|^2 = \sum_{\ell=1}^\infty \langle G, v_\ell \rangle^2 = \sum_{\ell=1}^\infty \lambda_\ell Z_\ell^2$$

The proof of theorem 3.1 is now complete. *Q.E.D.*

PROOF OF THEOREM 3.3: Let p and $\{X_t\}_{t=1}^n$ be as in (17). Define

$$\psi_n(y) := \frac{1}{n} \sum_{t=1}^n p(X_t, y) \quad \text{and} \quad \psi'_{kn}(y) := \frac{1}{n} \sum_{t=1}^{kn} p(X'_t, y) \tag{34}$$

Let $(U_\ell)_{\ell \geq 1}$ and $(U'_\ell)_{\ell \geq 1}$ be mutually independent IID sequences of standard normal random variables. Fix $k \in \mathbb{N}$, and consider the decomposition

$$n^{1/2}(\psi_n - \psi'_{kn}) = n^{1/2}(\psi_n - \psi) - k^{-1/2}(kn)^{1/2}(\psi'_{kn} - \psi)$$

Note that $n^{1/2}(\psi_n - \psi)$ and $(kn)^{1/2}(\psi'_{kn} - \psi)$ are independent random functions in L_2. By theorem 3.1, we have

$$n^{1/2}(\psi_n - \psi) \xrightarrow{d} \sum_\ell \lambda_\ell^{1/2} U_\ell v_\ell \quad \text{and} \quad (kn)^{1/2}(\psi'_{kn} - \psi) \xrightarrow{d} \sum_\ell \lambda_\ell^{1/2} U'_\ell v_\ell$$

[24] Let (e_ℓ) be any orthonormal subset of L_2, and suppose that $\langle G, e_\ell \rangle$ is zero-mean Gaussian in $\mathbb{R}$ for each $\ell \in \mathbb{N}$. Pick any $h \in L_2$. Then $\langle G, h \rangle = \sum_\ell \langle G, e_\ell \rangle \langle h, e_\ell \rangle$. The right-hand side is the almost sure limit of zero mean Gaussians, and hence is itself zero-mean Gaussian.

By independence and continuity of addition and scalar multiplication in L_2, we then have

$$n^{1/2}(\psi_n - \psi'_{nk}) \xrightarrow{d} \sum_\ell \lambda_\ell^{1/2} U_\ell - k^{-1/2} \sum_\ell \lambda_\ell^{1/2} U'_\ell v_\ell = \sum_\ell \lambda_\ell^{1/2}(U_\ell - k^{-1/2} U'_\ell) v_\ell$$

Applying the continuous mapping theorem and the Pythagorean law, we obtain

$$n\|\psi_n - \psi'_{nk}\|^2 \xrightarrow{d} \|\sum_\ell \lambda_\ell^{1/2}(U_\ell - k^{-1/2}U'_\ell)v_\ell\|^2 = \sum_\ell \lambda_\ell (U_\ell - k^{-1/2}U'_\ell)^2$$

The left-hand side of this equation is equal to the left-hand side of (20). Moreover, if Z_ℓ is standard normal, then $(1+1/k)Z_\ell^2$ and $(U_\ell - k^{-1/2}U'_\ell)^2$ have the same law. This completes the proof of (20). *Q.E.D.*

7.2. Local Alternatives: Theorem 3.2

Let $\mathscr{H}$ be defined as $L_2 \times \mathbb{R}$, with inner product

$$\langle g, h\rangle = \langle g_1, h_1\rangle + g_2 h_2 \qquad (g = (g_1, g_2) \text{ and } h = (h_1, h_2))$$

(Here $\langle g, h\rangle$ is the inner product in $\mathscr{H}$ and $\langle g_1, h_1\rangle$ is the inner product in L_2. The notation does not distinguish between them, but the meaning will be clear from context.) With the norm $\|h\| = \sqrt{\langle h, h\rangle}$, the space $\mathscr{H}$ is a Hilbert space, and the norm topology of $\mathscr{H}$ corresponds to the product topology of $L_2 \times \mathbb{R}$.

The next result is an extension of the Cramér-Wold theorem to $\mathscr{H}$:

LEMMA 7.2 *Let $U_n := (Y_n, \ell_n)$ be a random sequence in $\mathscr{H}$, where Y_n is a random element of L_2 and ℓ_n is a random variable for all n. Let U be a Gaussian random element of $\mathscr{H}$ with distribution $N(m, S)$. If $\{Y_n\}$ is tight in L_2 and*

$$\langle U_n, h\rangle \xrightarrow{d} N(\langle h, m\rangle, \langle h, Sh\rangle) \text{ in } \mathbb{R} \text{ for all } h \in \mathscr{H}$$

then $U_n \xrightarrow{d} U$ in $\mathscr{H}$.

PROOF: Suppose for the moment that $\{U_n\}$ is tight in $\mathscr{H}$. In this case, to show that U_n converges in distribution to U in $\mathscr{H}$, we need only show that $\langle U_n, h\rangle$ converges in distribution to $\langle U, h\rangle$ in $\mathbb{R}$ for all $h \in \mathscr{H}$ (Bosq, 2000, theorem 2.3). This is immediate from the hypotheses of the lemma and the characterization (26), which tells us that $\langle U, h\rangle$ has distribution $N(\langle h, m\rangle, \langle h, Sh\rangle)$.

It remains to show that $\{U_n\}$ is tight in $\mathscr{H}$. To see that this is so, note that, as required in the lemma, $\langle U_n, h\rangle$ converges in distribution for all $h \in \mathscr{H}$. Choosing $h = (0, 1)$, we see that ℓ_n converges in distribution, and is therefore tight.

Now fix $\epsilon > 0$. Since $\{Y_n\}$ and $\{\ell_n\}$ are both tight, we can find compact sets $K_a \subset L_2$ and $K_b \subset \mathbb{R}$ with $\mathbf{P}\{Y_n \notin K_a\} < \epsilon/2$ and $\mathbf{P}\{\ell_n \notin K_b\} < \epsilon/2$ for all n. The set $K_a \times K_b$ is compact in the product topology on $\mathscr{H}$, and we have

$$\mathbf{P}\{U_n \notin K_a \times K_b\} \leq \mathbf{P}\{Y_n \notin K_a\} \cup \{\ell_n \notin K_b\} \leq \mathbf{P}\{Y_n \notin K_a\} + \mathbf{P}\{\ell_n \notin K_b\} < \epsilon$$

We conclude that $\{U_n\}$ is tight in $\mathscr{H}$, completing the proof of lemma 7.2. *Q.E.D.*

LEMMA 7.3 *Let Y_n, $\mathbf{Q}_n$ and $\mathbf{P}_n$ be as defined in section 3.2, let*

$$r(x,y) := \frac{k(x,y)}{p(x,y)}, \qquad \sigma^2 := \mathbf{E} r(X_t, X_{t+1})^2 = \int r(X_t, X_{t+1})^2 d\mathbf{P}_n$$

and let $\ell_n \colon \mathbb{X}^n \to \mathbb{R}$ be the log likelihood ratio

$$\ell_n = \log \frac{d\mathbf{Q}_n}{d\mathbf{P}_n} = \log \left\{ \frac{\prod_{t=2}^n p_n(X_{t-1}, X_t)}{\prod_{t=2}^n p(X_{t-1}, X_t)} \right\}$$

If $U_n := (Y_n, \ell_n)$ and $U = (Y, \ell)$ has distribution $N(m, S)$ for $m = (0, -\sigma^2/2)$ and S satisfying

$$\langle h, Sh \rangle = \langle h_1, Ch_1 \rangle + 2\langle \tau, h_1 \rangle h_2 + h_2^2 \sigma^2 \qquad (h = (h_1, h_2) \in \mathscr{H})$$

then $U_n \xrightarrow{d} U$ under $\mathbf{P}_n$.

PROOF: In what follows, all probabilities and expectations are evaluated under $\mathbf{P}_n$. We begin by obtaining a more convenient expression for the likelihood ratio ℓ_n. Writing p_t for $p(X_{t-1}, X_t)$ and k_t for $k(X_{t-1}, X_t)$, we have

$$\ell_n = \sum_{t=2}^n \{\log(p_t + k_t/\sqrt{n}) - \log(p_t)\}$$

Expanding the log function around p_t yields

$$\ell_n = \frac{1}{\sqrt{n}} \sum_{t=2}^n \frac{k_t}{p_t} - \frac{1}{2n} \sum_{t=2}^n \frac{k_t^2}{p_t^2} + \frac{1}{3n^{3/2}} \sum_{t=2}^n \frac{k_t^3}{[p_t + \lambda n^{-1/2} k_t]^3}$$

For some $\lambda \in [0,1]$. Since (X_{t-1}, X_t) is itself ergodic (see lemma 7.1) and

$$\frac{k_t^3}{[p_t + \lambda n^{-1/2} k_t]^3} \leq \sup_{\delta \in [0,1]} \frac{|k(X_{t-1}, X_t)|^3}{|p(X_{t-1}, X_t) + \delta\, k(X_{t-1}, X_t)|^3}$$

it follows from assumption 3.1 that $\frac{1}{n} \sum_{t=2}^n k_t^3 / [p_t + \lambda n^{-1/2} k_t]^3 = O_P(1)$, and hence

$$\ell_n = \frac{1}{\sqrt{n}} \sum_{t=2}^n r(X_{t-1}, X_t) - \frac{1}{2n} \sum_{t=2}^n r(X_{t-1}, X_t)^2 + o_P(1) \tag{35}$$

Now we return to the proof of lemma 7.3. Taking into account lemma 7.2 and the fact that $\{Y_n\}$ is tight in L_2—as implied by the convergence in (29)—it suffices to show that

$$\langle U_n, h\rangle \xrightarrow{d} N(-h_2\sigma^2/2, \langle h_1, Ch_1\rangle + 2\langle \tau, h_1\rangle h_2 + h_2^2\sigma^2) \tag{36}$$

for arbitrary $h \in \mathscr{H}$. Fixing such an $h = (h_1, h_2)$, the definition of U_n and our expression for ℓ_n in (35) gives

$$\langle U_n, h\rangle = \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\langle h_1, \bar{p}(X_t, \cdot)\rangle + h_2\frac{1}{\sqrt{n}}\sum_{t=2}^{n} r(X_{t-1}, X_t) - h_2\frac{1}{2n}\sum_{t=2}^{n} r(X_{t-1}, X_t)^2 + o_P(1)$$

Since (X_{t-1}, X_t) is ergodic (lemma 7.1) we have

$$\frac{h_2}{2}\frac{1}{n}\sum_{t=2}^{n} r(X_{t-1}, X_t)^2 \to h_2\frac{\sigma^2}{2} \quad \text{in probability}$$

As a result of this convergence and Slutsky's theorem, the result (36) will be confirmed if we show that

$$\frac{1}{\sqrt{n}}\sum_{t=2}^{n} q(X_{t-1}, X_t) \xrightarrow{d} N(0, \langle h_1, Ch_1\rangle + 2\langle \tau, h_1\rangle h_2 + h_2^2\sigma^2) \tag{37}$$

for $q(X_{t-1}, X_t) := \langle h_1, \bar{p}(X_t, \cdot)\rangle + h_2 r(X_{t-1}, X_t)$. To see that this is indeed the case, observe first that $\mathbf{E}\langle h_1, \bar{p}(X_t, \cdot)\rangle = 0$ under H_0, as shown in (28), and also that

$$\mathbf{E}r(X_{t-1}, X_t) = \int\int \frac{k(x,y)}{p(x,y)}\psi(x)p(x,y)dxdy = \int\left[\int k(x,y)dy\right]\psi(x)dx = 0$$

It follows that $\mathbf{E}q(X_{t-1}, X_t) = 0$, and, as a result of V-uniform ergodicity of (X_{t-1}, X_t) (lemma 7.1) and the CLT for V-uniformly ergodic Markov processes (Meyn and Tweedie, 2009, theorem 17.0.1), we have

$$\frac{1}{\sqrt{n}}\sum_{t=2}^{n} q(X_{t-1}, X_t) \xrightarrow{d} N(0, v), \quad v := \mathbf{E}q(X_1, X_2)^2 + 2\sum_{t=2}^{\infty}\mathbf{E}q(X_1, X_2)q(X_t, X_{t+1})$$

It remains only to show that $v = \langle h_1, Ch_1\rangle + 2\langle \tau, h_1\rangle h_2 + h_2^2\sigma^2$, which is the right-hand side of the variance in (37). Prior to proving this, we observe that all of the following statements are valid, and will be used without comment below:

- $\mathbf{E}r(X_1, X_2) = 0$ and $\mathbf{E}[r(X_t, X_{t+1}) \mid X_1] = 0$ for all $t \geq 1$.
- $\mathbf{E}[r(X_1, X_2)r(X_t, X_{t+1})] = 0$ for all $t \geq 2$.
- $\mathbf{E}[\langle h_1, \bar{p}(X_1, \cdot)\rangle r(X_t, X_{t+1}) = 0$ for all $t \geq 1$.

(The first of these statements has already been established above, and the proofs of the rest are similar.) Turning now to the evaluation of v, note that

$$\begin{aligned}\mathbf{E}q(X_1,X_2)^2 &= \mathbf{E}\{\langle h_1,\bar{p}(X_1,\cdot)\rangle^2 + 2\langle h_1,\bar{p}(X_1,\cdot)\rangle h_2 r(X_1,X_2) + h_2^2 r(X_1,X_2)^2\} \\ &= \mathbf{E}\langle h_1,\bar{p}(X_1,\cdot)\rangle^2 + h_2^2\sigma^2\end{aligned}$$

while, for any given $t \geq 2$,

$$\mathbf{E}q(X_1,X_2)q(X_t,X_{t+1}) = \mathbf{E}\langle h_1,\bar{p}(X_1,\cdot)\rangle\langle h_1,\bar{p}(X_t,\cdot)\rangle + \mathbf{E}\langle h_1,\bar{p}(X_t,\cdot)h_2 r(X_1,X_2)$$

As a result, we have

$$\begin{aligned}v = \mathbf{E}\langle h_1,\bar{p}(X_1,\cdot)\rangle^2 + 2\sum_{t=2}^{\infty}\mathbf{E}\langle h_1,\bar{p}(X_1,\cdot)\rangle\langle h_1,\bar{p}(X_t,\cdot)\rangle \\ + 2\sum_{t=2}^{\infty}\mathbf{E}\langle h_1,\bar{p}(X_t,\cdot)h_2 r(X_1,X_2) + h_2^2\sigma^2\end{aligned}$$

Applying (31) (and recalling that C_0 and C were shown to be equal below that equation) we obtain

$$\mathbf{E}\langle h_1,\bar{p}(X_1,\cdot)\rangle^2 + 2\sum_{t=2}^{\infty}\mathbf{E}\langle h_1,\bar{p}(X_1,\cdot)\rangle\langle h_1,\bar{p}(X_t,\cdot)\rangle = \langle h_1, Ch_1\rangle$$

Finally, using the definition of τ, we obtain $v = \langle h_1, Ch_1\rangle + 2\langle \tau, h_1\rangle h_2 + h_2^2\sigma^2$. This verifies (37), and completes the proof of lemma 7.3. *Q.E.D.*

LEMMA 7.4 *For ℓ_n, ℓ defined in lemma 7.3, we have $\ell_n \xrightarrow{d} \ell$ and $\mathbf{E}\exp(\ell) = 1$ under $\mathbf{P}_n$.*

PROOF: We saw in lemma 7.3 that, under $\mathbf{P}_n$, we have $\langle h, U_n\rangle \xrightarrow{d} \langle h, U\rangle$ for all $h \in \mathscr{H}$, where $U \sim N(m,S)$ for m and S defined in lemma 7.3. Specializing to $h = (0,1)$ obtains the first claim in lemma 7.4. Regarding the second claim in lemma 7.4, for this same h we have $\ell = \langle h, U\rangle = N(\langle m,h\rangle, \langle h, Sh\rangle)$, and given the definitions of m and S in lemma 7.3,

$$N(\langle m,h\rangle, \langle h,Sh\rangle) = N(-\sigma^2/2, \sigma^2)$$

$$\therefore \quad \mathbf{E}\exp(\ell) = \exp\left(-\frac{\sigma^2}{2} + \frac{\sigma^2}{2}\right) = 1$$

when expectation is taken under $\mathbf{P}_n$. This completes the proof. *Q.E.D.*

We are now ready to complete the proof of theorem 3.2.

PROOF OF THEOREM 3.2: We saw in lemmas 7.3 and 7.4 that if $\ell_n = d\mathbf{Q}_n / d\mathbf{P}_n$ is the log likelihood ratio, then under $\mathbf{P}_n$ we have

$$\begin{pmatrix} Y_n \\ \ell_n \end{pmatrix} \xrightarrow{d} \begin{pmatrix} Y \\ \ell \end{pmatrix} \sim N(m, S)$$

and, moreover, $\mathbf{E} \exp(\ell) = 1$. Applying the abstract version of Le Cam's third lemma presented in van der Vaart and Wellner (1996, theorem 3.10.7), we then have

$$Y_n \xrightarrow{d} \pi \text{ under } \mathbf{Q}_n$$

when π is the probability measure on L_2 defined by

$$\pi(f) = \mathbf{E} \exp(\ell) f(Y) \quad \text{for all bounded measurable } f \colon \mathbb{X} \to \mathbb{R}$$

To complete the proof of theorem 3.2, we need only show that

$$\pi := N(\tau, C) \tag{38}$$

To see that this equality holds, let V be a random element on L_2 with $V \sim \pi$. In view of (26), to verify (38) it sufficies to show that, for arbitrary fixed $a \in L_2$, we have

$$\langle a, V \rangle \sim N(\langle a, \tau \rangle, \langle a, Ca \rangle) \tag{39}$$

To establish (39), observe that, from the definition of π, the moment generating function of $\langle a, V \rangle$ is

$$M(t) := \mathbf{E} \exp(t \langle a, V \rangle) = \mathbf{E} \exp(\ell) \exp(t \langle a, Y \rangle) = \mathbf{E} \exp(t \langle a, Y \rangle + \ell)$$

If $h \in \mathscr{H}$ is defined as $h = (ta, 1)$, then $\langle h, U \rangle$ is precisely $t\langle a, Y \rangle + \ell$. Since $\langle h, U \rangle$ is always Gaussian, we know that $t\langle a, Y \rangle + \ell$ is Gaussian. Its expectation and variance are given by

$$\mathbf{E}(t\langle a, Y \rangle + \ell) = \mathbf{E}\langle (ta, 1), U \rangle = \langle (ta, 1), m \rangle = \langle (ta, 1), (0, -\sigma^2/2) \rangle = -\sigma^2/2$$

and

$$\mathbb{V}\langle (ta, 1), U \rangle = \langle (ta, 1), S(ta, 1) \rangle = t^2 \langle a, Ca \rangle + 2t \langle \tau, a \rangle + \sigma^2$$

where the final expression follows from the definition of S given in the statement of lemma 7.3. To finish the proof, we observe that, since $t\langle a, Y \rangle + \ell$ is Gaussian with mean and variance as derived above, we must have

$$\mathbf{E} \exp(t\langle a, Y \rangle + \ell) = \exp \left\{ -\frac{\sigma^2}{2} + t \langle \tau, a \rangle + t^2 \frac{\langle a, Ca \rangle}{2} + \frac{\sigma^2}{2} \right\}$$

Cancelling the two instances of $\sigma^2/2$, we find that the moment generating function of $\langle a, V \rangle$ is

$$M(t) = \exp\left\{ t\langle \tau, a\rangle + t^2 \frac{\langle a, Ca\rangle}{2} \right\} \qquad (t \in \mathbb{R})$$

This is precisely the moment generating function for the $N(\langle a, \tau\rangle, \langle a, Ca\rangle)$ distribution, and hence we have established (39). This completes the proof of theorem 3.2. *Q.E.D.*

7.3. Theorems 4.1 and 4.2

LEMMA 7.5 *Let the conditions of theorem 4.1 hold, and let X be a random variable on $\mathbb{X}$ having distribution $\psi(\theta_0, \cdot)$. For all $m \in \{1, \dots, M\}$, we have*

$$\mathcal{E} D_m \bar{p}(\theta_0, X, \cdot) := \mathcal{E} \frac{\partial}{\partial \theta^m} \bar{p}(\theta_0, X, \cdot) = 0 \tag{40}$$

PROOF: To establish (40), we must first show that the functional expectation $\mathcal{E} D_m \bar{p}(\theta_0, X_t, \cdot)$ is well defined. A sufficient condition is that

$$\mathbf{E}\|D_m\bar{p}(\theta_0, X, \cdot)\|^2 = \int \left[\int D_m \bar{p}(\theta_0, x, y)^2 dy\right] \psi(\theta_0, x) dx < \infty \tag{41}$$

From assumption 4.2 we have $|D_m\bar{p}(\theta_0, x, y)| \leq K_1(x, y)$, and hence

$$\int D_m \bar{p}(\theta_0, x, y)^2 dy \leq \int K_1(x, y)^2 dy \leq V(x) \tag{42}$$

From (11) we see that $\int V(x)\psi(\theta_0, x)dx$ is finite, which in turn gives the restriction in (41).

The second step is to show that $\mathcal{E} D_m \bar{p}(\theta_0, X_t, \cdot) = 0$. From the definition of the functional expectation, we need to show that

$$\mathbf{E} \int D_m \bar{p}(\theta_0, X_t, y) h(y) dy = 0 \tag{43}$$

for every $h \in L_2$. To see this, pick any $h \in L_2$. In view of assumption 4.2, Fubini's theorem applies, and

$$\mathbf{E} \int D_m \bar{p}(\theta_0, X_t, y) h(y) dy = \int \mathbf{E}\, D_m \bar{p}(\theta_0, X_t, y) h(y) dy$$

Using assumption 4.2 again, we can interchange expectation and differentiation to obtain

$$\mathbf{E}\, D_m \bar{p}(\theta_0, X_t, y) = D_m \mathbf{E}\, \bar{p}(\theta_0, X_t, y) = 0$$

where the last equality is due to H_0 and (27). The validity of (43) is now established. *Q.E.D.*

PROOF OF THEOREM 4.1: Assume the conditions of theorem 4.1. We begin by showing that

$$n^{-1/2} \sum_{t=1}^{n} \bar{p}(\hat{\theta}_n, X_t, \cdot) \xrightarrow{d} \sum_{\ell \geq 1} \lambda_\ell(\theta_0)^{1/2} Z_\ell v_\ell(\theta_0) \tag{44}$$

in L_2. To do so, let

$$Y_n(y) := n^{-1/2} \sum_{t=1}^{n} \bar{p}(\theta_0, X_t, y) \quad \text{and} \quad \hat{Y}_n(y) := n^{-1/2} \sum_{t=1}^{n} \bar{p}(\hat{\theta}_n, X_t, y)$$

Since H_0 is assumed true, theorem 3.1 gives $Y_n \xrightarrow{d} \sum_{\ell \geq 1} \lambda_\ell^{1/2}(\theta_0) Z_\ell v_\ell(\theta_0)$. The convergence in (44) amounts to the claim that $\hat{Y}_n$ converges in distribution to the same limit. This will hold whenever

$$\|Y_n - \hat{Y}_n\| \xrightarrow{p} 0 \qquad (n \to \infty) \tag{45}$$

(cf., e.g., Dudley, 2002, lemma 11.9.4). In order to establish (45), we write

$$\bar{p}(\hat{\theta}_n, x, y) - \bar{p}(\theta_0, x, y) = D\bar{p}(\theta_0, x, y)^\top (\hat{\theta}_n - \theta_0) + R(\hat{\theta}_n, x, y) \tag{46}$$

where R is the remainder term and $\top$ indicates inner product in $\mathbb{R}^M$. We then have

$$\begin{aligned}
\hat{Y}_n(y) - Y_n(y) &= n^{-1/2} \sum_{t=1}^{n} [\bar{p}(\hat{\theta}_n, X_t, y) - \bar{p}(\theta_0, X_t, y)] \\
&= n^{-1/2} \sum_{t=1}^{n} \left[D\bar{p}(\theta_0, X_t, y)^\top (\hat{\theta}_n - \theta_0) + R(\hat{\theta}_n, X_t, y) \right]
\end{aligned}$$

Using the triangle inequality for the L_2 norm, we obtain

$$\|\hat{Y}_n - Y_n\| \leq \left\| n^{-1/2} \sum_{t=1}^{n} D\bar{p}(\theta_0, X_t, \cdot)^\top (\hat{\theta}_n - \theta_0) \right\| + \left\| n^{-1/2} \sum_{t=1}^{n} R(\hat{\theta}_n, X_t, \cdot) \right\| \tag{47}$$

Consider the first term on the right-hand side of (47). Using the triangle inequality again, we see that this term is bounded above by

$$\sum_{m=1}^{M} |\hat{\theta}_n^m - \theta_0^m| \left\| n^{-1/2} \sum_{t=1}^{n} D_m \bar{p}(\theta_0, X_t, \cdot) \right\| = \sum_{m=1}^{M} |\hat{\theta}_n^m - \theta_0^m| n^{1/2} \left\| \frac{1}{n} \sum_{t=1}^{n} D_m \bar{p}(\theta_0, X_t, \cdot) \right\|$$

Fix $m \in \{1, \ldots, M\}$. From lemma 7.5, (42) and the CLT of Stachurski (2010), the sequence of random elements $n^{1/2} \frac{1}{n} \sum_{t=1}^{n} D_m \bar{p}(\theta_0, X_t, \cdot)$ converges in distribution to a centered Gaussian in L_2. Applying the continuous mapping theorem, the norm of this random function also converges in distribution, and hence

$$n^{1/2}\left\|\frac{1}{n}\sum_{t=1}^{n} D_m\bar{p}(\theta_0, X_t, \cdot)\right\| = O_P(1)$$

Since $|\hat{\theta}_n^m - \theta_0^m| = o_P(1)$ by assumption, we then have

$$|\hat{\theta}_n^m - \theta_0^m| n^{1/2}\left\|\frac{1}{n}\sum_{t=1}^{n} D_m\bar{p}(\theta_0, X_t, \cdot)\right\| = o_P(1)O_P(1) = o_P(1)$$

for each $m \in \{1, \ldots, M\}$.

$$\therefore \quad \sum_{m=1}^{M} n^{1/2}|\hat{\theta}_n^m - \theta_0^m|\left\|\frac{1}{n}\sum_{t=1}^{n} D_m\bar{p}(\theta_0, X_t, \cdot)\right\| = o_P(1)$$

We have now shown that the first term on the right-hand side of (47) is $o_P(1)$. It remains to show that the second term is also $o_P(1)$. In other words, we aim to show that

$$\left\|n^{-1/2}\sum_{t=1}^{n} R(\hat{\theta}_n, X_t, \cdot)\right\| = o_p(1) \tag{48}$$

Using the mean value theorem, we can write

$$R(\hat{\theta}_n, X_t, y) = \{D\bar{p}(\tilde{\theta}, X_t, y) - D\bar{p}(\theta_0, X_t, y)\}^\top(\hat{\theta}_n - \theta_0)$$

where $\tilde{\theta}$ lies on the line segment between θ_0 and $\hat{\theta}_n$. It follows that

$$\left|n^{-1/2}\sum_{t=1}^{n} R(\hat{\theta}_n, X_t, y)\right| = \left[\frac{1}{n}\sum_{t=1}^{n}\{D\bar{p}(\tilde{\theta}, X_t, y) - D\bar{p}(\theta_0, X_t, y)\}\right]^\top n^{1/2}(\hat{\theta}_n - \theta_0)$$

Applying the Cauchy-Schwartz inequality in $\mathbb{R}^M$, we obtain

$$\left|n^{-1/2}\sum_{t=1}^{n} R(\hat{\theta}_n, X_t, y)\right| \leq F_n(y)\, n^{1/2}\|\hat{\theta}_n - \theta_0\|_E \tag{49}$$

where $\|\cdot\|_E$ is the norm in $\mathbb{R}^M$, and

$$F_n(y) := \left\|\frac{1}{n}\sum_{t=1}^{n}\{D\bar{p}(\tilde{\theta}, X_t, y) - D\bar{p}(\theta_0, X_t, y)\}\right\|_E$$

From (49) we obtain the L_2 norm inequality

$$\left\| n^{-1/2} \sum_{t=1}^{n} R(\hat{\theta}_n, X_t, \cdot) \right\| \leq \|F_n\| \cdot O_P(1)$$

Hence, to establish (48), it suffices to prove that $\|F_n\| = o_P(1)$. By the definition of F_n, we have

$$\|F_n\| \leq \frac{1}{n} \sum_{t=1}^{n} \left[\int \|D\bar{p}(\tilde{\theta}, X_t, y) - D\bar{p}(\theta_0, X_t, y)\|_E^2 \, dy \right]^{1/2}$$

Let U be an open ball centered on θ_0 defined in assumption 4.3. Note from the definition of $\tilde{\theta}$ that $\tilde{\theta} \in U$ whenever $\hat{\theta}_n \in U$. As a result, on the set $\{\hat{\theta}_n \in U\}$ we have

$$\begin{aligned} \|F_n\| &\leq \frac{1}{n} \sum_{t=1}^{n} \left[\int K_2(X_t, y)^2 \, dy \, \|\tilde{\theta} - \theta_0\|_E^{2\alpha} \right]^{1/2} \\ &\leq \|\tilde{\theta} - \theta_0\|_E^{\alpha} \frac{1}{n} \sum_{t=1}^{n} V(X_t)^{1/2} \leq \|\hat{\theta}_n - \theta_0\|_E^{\alpha} \frac{1}{n} \sum_{t=1}^{n} V(X_t)^{1/2} \end{aligned}$$

Fixing $\delta > 0$, we have

$$\begin{aligned} \mathbf{P}\{\|F_n\| > \delta\} &\leq \mathbf{P}\{\hat{\theta}_n \notin U\} + \mathbf{P}\{\|F_n\| > \delta \text{ and } \hat{\theta}_n \in U\} \\ &\leq o(1) + \mathbf{P}\left\{ \|\hat{\theta}_n - \theta_0\|_E^{\alpha} \frac{1}{n} \sum_{t=1}^{n} V(X_t)^{1/2} > \delta \right\} \end{aligned}$$

By assumption, $\|\hat{\theta}_n - \theta_0\|_E^{\alpha} = o_P(1)$. Moreover, $\int V(x)^{1/2} \psi(\theta_0, x) dx < \infty$ by Jensen's inequality and (11), so by the scalar law of large numbers for V-uniformly ergodic Markov processes (Meyn and Tweedie, theorem 17.1.7), we have $n^{-1} \sum_{t=1}^{n} V(X_t)^{1/2} = O_P(1)$.

$$\therefore \quad \|\hat{\theta}_n - \theta_0\|_E^{\alpha} \frac{1}{n} \sum_{t=1}^{n} V(X_t)^{1/2} = o_P(1)$$

$$\therefore \quad \mathbf{P}\{\|F_n\| > \delta\} = o(1) + o(1) = o(1)$$

We conclude that $\|F_n\| = o_P(1)$, and hence (48) is valid.

Looking back, we have shown that (45) and hence (44) is true. To complete the proof of theorem 4.1, we need to prove (22). Rewriting (44), we know that

$$\hat{Y}_n := n^{-1/2} \sum_{t=1}^{n} \bar{p}(\hat{\theta}_n, X_t, \cdot) \xrightarrow{d} \sum_{\ell=1}^{\infty} \lambda_\ell(\theta_0)^{1/2} Z_\ell v_\ell(\theta_0) =: G \tag{50}$$

in L_2. The continuous mapping theorem and Parseval's identity now give

$$\frac{1}{n}\int\left\{\sum_{t=1}^{n}\bar{p}(\hat{\theta}_n X_t, y)\right\}^2 dy = \|\hat{Y}_n\|^2 \xrightarrow{d} \|G\|^2 = \sum_{\ell=1}^{\infty}\langle G, v_\ell(\theta_0)\rangle^2$$

Using the definition of G in (50), we have $\sum_{\ell=1}^{\infty}\langle G, v_\ell(\theta_0)\rangle^2 = \sum_{\ell=1}^{\infty}\lambda_\ell(\theta_0)Z_\ell^2$. The proof of theorem 4.1 is now complete. *Q.E.D.*

PROOF OF THEOREM 4.2: Let T_n be the test statistic on the left-hand side of (23). The claim in the theorem amounts to

$$\lim_{n\to\infty}\mathbf{P}\{T_n \le c_\alpha(\hat{\theta}_n)\} = 1-\alpha \tag{51}$$

We have $T_n \xrightarrow{d} \sum_\ell \lambda_\ell(\theta_0)Z_\ell^2$ and $c_\alpha(\hat{\theta}_n) \xrightarrow{p} c_\alpha(\theta_0)$, where the second result is due to continuity of c_α at θ_0. Slutsky's theorem yields $T_n - c_\alpha(\hat{\theta}_n) + c_\alpha(\theta_0) \xrightarrow{d} \sum_\ell \lambda_\ell(\theta_0)Z_\ell^2$. As a result,

$$\begin{aligned}\lim_{n\to\infty}\mathbf{P}\{T_n \le c_\alpha(\hat{\theta}_n)\} &= \lim_{n\to\infty}\mathbf{P}\{T_n - c_\alpha(\hat{\theta}_n) + c_\alpha(\theta_0) \le c_\alpha(\theta_0)\}\\ &= \mathbf{P}\left\{\sum_\ell \lambda_\ell(\theta_0)Z_\ell^2 \le c_\alpha(\theta_0)\right\} = 1-\alpha\end{aligned}$$

Here the last equality is valid by the definition of $c_\alpha(\theta_0)$. *Q.E.D.*

7.4. Consistency: Theorem 4.3

PROOF OF THEOREM 4.3: Recalling that the test statistic is $n\|n^{-1}\sum_{t=1}^{n}\bar{p}(\hat{\theta}_n, X_t, \cdot)\|^2$ and observing that

$$\epsilon \le \|\mathcal{E}\bar{p}(\hat{\theta}_n, X_t, \cdot)\| \le \left\|\frac{1}{n}\sum_{t=1}^{n}\bar{p}(\hat{\theta}_n, X_t, \cdot) - \mathcal{E}\bar{p}(\hat{\theta}_n, X_t, \cdot)\right\| + \left\|\frac{1}{n}\sum_{t=1}^{n}\bar{p}(\hat{\theta}_n, X_t, \cdot)\right\|$$

where $\epsilon > 0$ is the value of the infimum in the definition of H_1, we see that the claim in the theorem will be valid whenever

$$\left\|\frac{1}{n}\sum_{t=1}^{n}\bar{p}(\hat{\theta}_n, X_t, \cdot) - \mathcal{E}\bar{p}(\hat{\theta}_n, X_t, \cdot)\right\| = \left\|\frac{1}{n}\sum_{t=1}^{n}p(\hat{\theta}_n, X_t, \cdot) - \mathcal{E}p(\hat{\theta}_n, X_t, \cdot)\right\| \tag{52}$$

converges in probability to zero as $n \to \infty$. The term in (52) is bounded above by $(I) + (II) + (III)$ where

$$(I) := \left\|\frac{1}{n}\sum_{t=1}^{n}\{p(\hat{\theta}_n, X_t, \cdot) - p(\theta_1, X_t, \cdot)\}\right\|, \quad (II) := \left\|\frac{1}{n}\sum_{t=1}^{n}p(\theta_1, X_t, \cdot) - \mathcal{E}p(\theta_1, X_t, \cdot)\right\|$$

and $(III) := \left\|\mathcal{E}p(\theta_1, X_t, \cdot) - \mathcal{E}p(\hat{\theta}_n, X_t, \cdot)\right\|$. Here θ_1 is the in probability limit of $\hat{\theta}_n$. We claim that all of these terms converge to zero. To begin, consider first the term (I). Evidently

$$(I) \leq \frac{1}{n}\sum_{t=1}^{n} \left\|\{p(\hat{\theta}_n, X_t, \cdot) - p(\theta_1, X_t, \cdot)\}\right\|$$

By the mean value theorem in $\mathbb{R}^M$ we have

$$|p(\hat{\theta}_n, X_t, \cdot) - p(\theta_1, X_t, \cdot)| \leq \|Dp(\bar{\theta}, X_t, \cdot)\|_E \cdot \|\hat{\theta}_n - \theta_1\|_E \tag{53}$$

where $\bar{\theta}$ lies on the line segment between θ_1 and $\hat{\theta}_n$. Taking the the L_2 norm and then the average over t, we obtain

$$(I) \leq \|\hat{\theta}_n - \theta_1\|_E \, \frac{1}{n}\sum_{t=1}^{n}\left[\int \|Dp(\bar{\theta}, X_t, y)\|_E^2 \, dy\right]^{1/2}$$

Applying assumption 4.6 and Jensen's inequality, we obtain

$$\mathbf{E}\left[\int \|Dp(\bar{\theta}, X_t, y)\|_E^2 \, dy\right]^{1/2} \leq \left[\mathbf{E}\int \Lambda(X_t, y)^2 \, dy\right]^{1/2} < \infty$$

Since this expectation is finite and $\{X_t\}$ is assumed to be stationary and ergodic, the ergodic law of large numbers implies that

$$\frac{1}{n}\sum_{t=1}^{n}\left[\int \|Dp(\bar{\theta}, X_t, y)\|_E^2 \, dy\right]^{1/2} \xrightarrow{p} \mathbf{E}\left[\int \|Dp(\bar{\theta}, X_t, y)\|_E^2 \, dy\right]^{1/2}$$

Hence we have $(I) \leq o_P(1)O_P(1) = o_p(1)$ as claimed.

Turning to the term (II), the claim that this is $o_P(1)$ follows directly from assumption 4.4, provided that the expectation $\mathcal{E}p(\theta_1, X_t, \cdot)$ exists. This L_2 expectation exists whenever the scalar expectation of the norm of $p(\theta_1, X_t, \cdot)$ is finite. Finiteness of this scalar expectation is a direct consequence of assumption 4.7.

Regarding (III), another application of (53) gives

$$(III) \leq \mathbf{E}\,\|p(\theta_1, X_t, \cdot) - p(\hat{\theta}_n, X_t, \cdot)\| \leq \mathbf{E}\,\|\hat{\theta}_n - \theta_1\|_E \, \left[\int \|Dp(\bar{\theta}, X_t, y)\|_E^2 \, dy\right]^{1/2}$$

Using assumption 4.6 and the Cauchy-Schwartz inequality, we obtain

$$(III) \leq \mathbf{E}\,\|\hat{\theta}_n - \theta_1\|_E \, \left[\int \Lambda(X_t, y)^2 \, dy\right]^{1/2} \leq \left[\mathbf{E}\,\|\hat{\theta}_n - \theta_1\|_E^2 \; \mathbf{E}\int \Lambda(X_t, y)^2 \, dy\right]^{1/2}$$

The term $\mathbf{E}\int \Lambda(X_t, y)^2 \, dy$ is finite by assumption 4.6. Moreover $\|\hat{\theta}_n - \theta_1\|_E = o_P(1)$ implies $\|\hat{\theta}_n - \theta_1\|_E^2 = o_P(1)$, and the latter is uniformly bounded as a result of the compactness of Θ. Hence $\mathbf{E}\,\|\hat{\theta}_n - \theta_1\|_E^2$ converges to zero.

We conclude that $(I) + (II) + (III) = o_P(1) + o_P(1) + o(1) = o_P(1)$. Hence the term in (52) is $o_P(1)$ and the proof is done. *Q.E.D.*

REFERENCES

[1] Andrews, D.K. (1997): "A Conditional Kolmogorov Test," *Econometrica*, 65 (5), 1097–1128.

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

[3] Aït-Sahalia, Y., L.P. Hansen and J.A. Scheinkman (2010): "Operator Methods for Continuous-Time Markov Processes," in *Handbook of Financial Econometrics*, (Y. Aït-Sahalia and L.P. Hansen, eds.), North Holland.

[4] Bai, J. (2003): "Testing Parametric Conditional Distributions of Dynamic Models," *The Review of Economics and Statistics*, 85 (3), 531–549.

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

[6] Chen, S. X., J. Gao and C. Y. Tang (2008): "A Test for Model Specification of Diffusion Processes," *Annals of Statistics* 36 (1), 167–198.

[7] Chicheportiche, R. and J-P. Bouchaud (2011): "Goodness of Fit Tests with Dependent Observations," mimeo, Ecole Centrale Paris.

[8] Cox, J.C., J.E. Ingersoll and S.A. Ross (1985). "A Theory of the Term Structure of Interest Rates," *Econometrica*, 53: 385–407.

[9] Darling, A. (1955): "The Cramer-Smirnov Test in the Parametric Case," *The Annals of Mathematical Statistics*, 26 (1), 1–20.

[10] Diebold, F. X., T. A. Gunther and A. S. Tay (1998): "Evaluating Density Forecasts with Applications to Financial Risk Management," *International Economic Review*, 39 (4) pg. 863–883.

[11] Dudley, R. M. (2002): *Real Analysis and Probability*, Cambridge Studies in Advanced Mathematics No. 74, Cambridge University Press.

[12] Durbin, J. (1973): "Weak Convergence of the Sample Distribution Function when Parameters are Estimated," *The Annals of Statistics*, 1 (2), 279–290.

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

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

[15] Kristensen, D. (2007): "Geometric Ergodicity of a Class of Markov Chains with Applications to Time Series Models," mimeo, University of Wisconsin.

[16] Meyn, S. P. and Tweedie, R. L. (2009): *Markov Chains and Stochastic Stability*, Springer-Verlag: London.

[17] Neumann, M. H. and E. Paparoditis (2008): "Goodness-of-fit tests for Markovian time series models: Central limit theory and bootstrap approximations," *Bernoulli,* 14 (1), 14–46.

[18] Nishimura, K. and J. Stachurski (2005): "Stability of Stochastic Optimal Growth Models: A New Approach," *Journal of Economic Theory,* 122 (1), 100–118.

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

[20] Pollard, D. (1984): *Convergence of Stochastic Processes,* Springer-Verlag, New York.

[21] Smirnov, N. V. (1936): "Sur la Distribution de W^2," *Comp. Rend. Acad. Sci.,* 202, 449–452.

[22] Stachurski, J. and V. L. Martin (2008): "Computing the Distributions of Economic Models via Simulation," *Econometrica,* 76 (2), 443–450.

[23] Stachurski, J. (2010): "A Hilbert Space Central Limit Theorem for Geometrically Ergodic Markov Chains," mimeo, Australian National University.

[24] van der Vaart, A. W. and J. A. Wellner (1996): *Weak convergence and empirical processes,* Springer.

[25] Vasicek, O. (1977): "An Equilibrium Characterisation of the Term Structure," *Journal of Financial Economics,* 5 (2), 177–188.

[26] Weiss, M. S. (1978): "Modification of the Kolmogorov-Smirnov Test for use with Correlated Data," *Journal of the American Statistical Association*, 73 (364), 872–875.