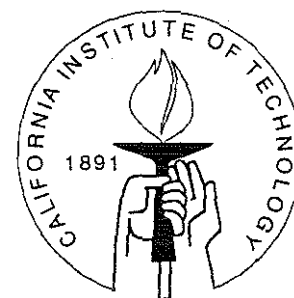DIVISION OF THE HUMANITIES AND SOCIAL SCIENCES

CALIFORNIA INSTITUTE OF TECHNOLOGY

PASADENA, CALIFORNIA 91125

A CONSISTENT TEST OF STATIONARY-ERGODICITY

Ian Domowitz
Northwestern University

Mahmoud A. El-Gamal
California Institute of Technology

SOCIAL SCIENCE WORKING PAPER 794

A Consistent Test of Stationary-Ergodicity

Ian Domowitz and Mahmoud A. El-Gamal

Abstract

A formal statistical test of stationary-ergodicity is developed for known Markovian processes on $\mathbb{R}^d$. This makes it applicable to testing models and algorithms, as well as estimated time series processes ignoring the estimation error. The analysis is conducted by examining the asymptotic properties of the Markov operator on density space generated by the transition in the state space. The test is developed under the null of stationary-ergodicity, and it is shown to be consistent against the alternative of non-stationary-ergodicity. The test can be easily performed using any of a number of standard statistical and mathematical computer packages.

A Consistent Test of Stationary-Ergodicity

Ian Domowitz and Mahmoud A. El-Gamal *

1   Introduction

Ergodicity conditions play an integral part in many estimation and modeling decisions. In this paper we assume that we have a known Markovian process on $\mathbb{R}^d$. Our null hypothesis is that the process is ergodic with a unique atomless stationary density having full support. Under that null hypothesis, we develop a consistent test. We know of no other similar results in the literature. If the hypothesis of interest is k-order Markovian, it is well known that it can be reduced to a k-dimensional first-order Markovian process, hence attention can be limited to the latter.

The idea of the test is very direct. Stationary ergodicity of the data generating process is equivalent to the convergence of Cesàro averages of the transition probabilities to a unique invariant measure. In the case of a discrete state space this simply follows trivially from the convergence of powers of the probability transition matrix. In more general state spaces in $\mathbb{R}^d$, one can define a Markov operator on the space of densities on $\mathbb{R}^d$ which corresponds to the transition kernel on the state space. Testing the convergence of the Cesàro averages of the transition kernel is then equivalent to testing the convergence of Cesàro averages of iterates of all initial densities – through the corresponding Markov operator – to the same unique stationary density. We construct algorithms to draw i.i.d. samples from the Cesàro averages of iterates of any two initial densities through our Markov operator. As the number of iterates we average goes to infinity, and then as our sample sizes go to infinity, applying a test of goodness of fit to the two samples yields a test with the appropriate size. We then construct an algorithm to randomly

choose our two initial densities, and show that the probability of drawing two densities whose Cesàro averaged iterates converge to the same limit under the alternative of non-stationary-ergodicity is zero, hence obtaining a consistent test.

In section 2, we explicitly define our null hypothesis, and show its equivalence to the Cesàro convergence of a Markov operator on a density space. In sections 3 and 4, we develop our consistent randomized test of that Cesàro convergence. In Section 5, we investigate the small sample behavior of the size and power of our test via Monte Carlo experiments. Section 6 concludes the paper with remarks regarding extensions of the test to the case of an estimated law of motion.

2 The Null Hypothesis of Stationary-Ergodicity

We assume to have a known stochastic process on $\mathbb{R}^d$ defined by the transition function $p(\xi, A)$ for $\xi \in X$, and $A \in \mathcal{B}(X)$, where $\mathcal{B}(X)$ is the Borel σ-algebra of subsets of X. We assume as in [10] that for a given ξ, $p(\xi, .)$ is a probability measure on $\mathcal{B}(X)$, and for a given $A \in \mathcal{B}(X)$, $p(., A)$ is a Borel measurable function. We shall refer to $p(., .)$ as the one-step transition probability. As usual, we define the s-step transition probability recursively by:

$$p^{(s)}(\xi, A) = \int_X p^{(s-1)}(\xi, d\eta) \ p(\eta, A)$$

We assume that the measure $p(\xi, .)$ is absolutely continuous.

Starting from an initial density $f_0(x)$ on the state space $\mathbb{R}^d$, the probability of the process falling in any Borel set A at period s can easily be defined by:

$$Pr_{f_0}\{x_s \in A\} = \int_A f_0(\xi).p^{(s)}(\xi, d\eta) \equiv \int_A f_s(\eta) \ d\eta$$

This implicitly defines the Markov operator $P: D(\mathbb{R}^d) \to D(\mathbb{R}^d)$ (via $f_s(.) = P^s f_0(.)$), where $D(.)$ is the space of densities. We assume that there exists at least one stationary density f^* such that $Pf^* = f^*$.

Under that assumption of the existence of a stationary density for P, following [15, p. 89], we say the stochastic process defined by $p : \mathbb{R}^d \times \mathcal{B}(\mathbb{R}^d) \to [0, 1]$, or alternatively by $P : D(\mathbb{R}^d) \to D(\mathbb{R}^d)$, is *stationary-ergodic* if there exists a unique measure π with a corresponding density f^* such that

$$\lim_{s \uparrow \infty} \frac{1}{s} \sum_{i=0}^{s-1} p^{(i)}(\xi, A) = \pi(A)$$

for all sets $A \in \mathcal{B}(X)$, or alternatively

$$\lim_{s \uparrow \infty} \frac{1}{s} \sum_{i=0}^{s-1} P^i f(x) = f^*(x)$$

2

for all $x \in X$, for all $f \in D(\mathbb{R}^d)$.

Remarks:

- It is known that this definition of stationary-ergodicity is equivalent to the more standard definition of uniqueness of the invariant measure (e.g. [6, Theorem 2, p.39]).

- The restriction to the stationary-ergodic case may be a limitation. However, since ergodicity is mainly used to ensure that time series sample moments converge to the moments under the unique stationary measure, this limitation does not seem very severe. Our test will be shown to have asymptotic power 1 against the non-stationary-ergodic alternative.

- Typical sufficient conditions in the literature such as Doob's (D) and (Σ) conditions [10, pp. 192, 195] for obtaining the above convergence to hold are typically too strong. The (D), or Doeblin, condition guarantees the stronger uniform convergence of the Cesàro averages from all initial ξ. As seen from [10, example 4, p. 218], our weaker criterion of (not necessarily uniform) convergence can be satisfied in cases where condition (D) is violated.

3   An Operational Test of Stationary-Ergodicity

We wish to test (H_0:P is stationary-ergodic) versus (H_A: P is not stationary-ergodic). The null will fail to hold if and only if one of the following two conditions occurs:

1. P is non-stationary. Technically this means that there does not exist an $f^\star$ such that $Pf^\star = f^\star$, or alternatively that $p(.,.)$ has no invariant measure. Notice that non-stationarity in this sense is different from the common usage of the term in time-series contexts. In particular, any given *sequence* of draws need not be stationary since the initial draw may not come from the stationary measure. The important thing about existence of a stationary measure is the existence of Cesàro limits of iterates of any initial density.

2. P is non-ergodic. This means that there is more than one stationary density, or more than one invariant measure. This will be the case, for example, for processes with multiple ergodic subclasses. The mass in each of the ergodic subclasses will converge to the unique stationary measure for the process restricted to that subclass. The Cesàro limits starting from different initial measures will then be different convex combinations of the ergodic measures on each of the ergodic subclasses, depending on the initial mass in each of them.

The idea of our test is quite simple. Under the null, the average of $P^i f$ should converge for all initial $f \in D(\mathbb{R}^d)$ to some unknown $f^\star$. We can then look at different initial

densities and see if the averages of P^i iterates on those densities converge to the same limit. In fact, we will start with two initial densities f and g, and compare $\frac{1}{s}\sum_{i=0}^{s-1} P^i f(x)$ with $\frac{1}{s}\sum_{i=0}^{s-1} P^i g(x)$ as s gets large.

We require an i.i.d. sample (of size n) from $\frac{1}{s}\sum_{i=0}^{s-1} P^i f(x)$ and another from $\frac{1}{s}\sum_{i=0}^{s-1} P^i g(x)$. We generate those samples as follows.[1] For each $j = 1, ..., n$, we randomly draw an x from f, and an $i \in \{0, 1, ..., s-1\}$ taking on each of the values with probability $1/s$. We then draw $x_j^f \sim p^{(i)}(x, .)$. The resulting collection of n draws are a random sample from $\frac{1}{s}\sum_{i=0}^{s-1} P^i f(x)$. The same procedure can now be followed where the x_j^g's are drawn from g. Under the null hypothesis, as $s \uparrow \infty$, the two distributions from which our two samples are drawn should become arbitrarily similar. Hence, we can apply any of a number of non-parametric procedures to test the null hypothesis as the sample size $n \uparrow \infty$.

To conduct the test for the transition $p(\xi, .)$ on $\mathbb{R}^d$, we can follow one of two procedures:

1. We can use multidimensional versions of the Cramèr-von Mises or Kolmogorov-Smirnov goodness of fit tests à la Khmaladze [14] using critical values from Piterbarg and Fatalov [16].

2. Alternatively, and this is the approach we follow, we can find a map $\psi\colon \mathbb{R}^d \to [0, 1]$ through which the transition $p(\xi, .)$ on $\mathbb{R}^d$ induces a transition $p_\psi(x, .)$ on $[0, 1]$, and then show that studying the latter transition is sufficient for our purposes. The function $\psi(.)$ is defined as follows:

   (a) Let the components of the d-dimensional vector x be $(x_1, ..., x_d)$. Define the mapping $\tau\colon \mathbb{R}^d \to [0, 1]^d$ by

   $$\tau((x_1, ..., x_d)) = (\Phi(x_1), ..., \Phi(x_d)),$$

   where $\Phi(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-y^2/2}\,dy$ is the standard normal cumulative distribution function. The evaluation of $\Phi(.)$ and $\Phi^{-1}(.)$ is standardized in most statistical and numerical packages.

   (b) For $(\xi_1, ..., \xi_d) \in [0, 1]^d$, let the decimal expansion of ξ_i be $\xi_i = 0.a_1^i a_2^i...$, for $i = 1, ..., d$.

   (c) Define $\rho\colon [0, 1]^d \to [0, 1]$ by

   $$\rho(\xi_1, ..., \xi_d) = 0.a_1^1 a_1^2 ... a_1^d a_2^1 ... a_2^d ...$$

   Note that ρ is one-to-one and onto and preserves Lebesgue-measure. ρ can be easily implemented as follows:

---

[1]This version of our algorithm (where we randomly draw the i's) was suggested to us by Donald W. K. Andrews.

i. For $i = 1, ..., d$ and $j = 1, 2, ...$, extract the integers $a_j^i \in \{0, 1, ..., 9\}$ as follows:
$$a_j^i = floor(10 \times remainder(10^{j-1} \times \xi_i)),$$
where $floor(x)$ is the integer part of x, and $remainder(x) = x - floor(x)$.

ii. Construct $\rho(\xi_1, ..., \xi_n) = \sum_{j=1}^{\infty} \sum_{i=1}^{d} a_j^i / (10^{(j-1)d+i})$.

Obviously, the implementation of $\rho^{-1}(.)$ is equally easy. For $t = 0.t_1 t_2 ... \in [0, 1]$, we extract $a_j^i = floor(10 \times remainder(10^{(j-1)d+i-1} \times t))$, and then construct $\xi_i = \sum_{j=1}^{\infty} a_j^i / (10^j)$, for $i = 1, ..., d$.

(d) Define the mapping $\psi: \mathbb{R}^d \dashrightarrow [0, 1]$ by $\psi(x) = \rho \circ \tau(x)$. The inverse of this mapping $\psi^{-1}(t) = \tau^{-1} \circ \rho^{-1}(t)$ is a well defined function since ρ, and Φ are both one-to-one and onto transformations.

In the following lemma, we prove that the transition $p(\xi, .)$ on $\mathbb{R}^d$ is stationary-ergodic if and only if the induced transition $p_\psi(x, .)$ on $[0, 1]$ is stationary-ergodic. Notice that the mapping ψ is itself one-to-one and onto, and is quite easy to implement in practice, and hence, we can limit attention to transitions on $[0, 1]$.

Lemma 1 *Given $\psi: \mathbb{R}^d \dashrightarrow [0, 1]$, defined above, the transition $p(\xi, A)$ on $\mathbb{R}^d$ is stationary-ergodic if and only if the induced transition $p_\psi(x, B) = p(\psi^{-1}(x), \psi^{-1}(B))$ on $[0, 1]$ is stationary-ergodic.*

Proof: The proof follows immediately from the one-to-one and onto properties of ψ. Let $p(., .)$ be stationary-ergodic. Then, by definition, there exists a unique measure $\pi(.)$ such that
$$\lim_{s \uparrow \infty} \frac{1}{s} \sum_{i=0}^{s-1} p^{(i)}(\xi, A) = \pi(A),$$
for all $\xi \in \mathbb{R}^d$, and $A \in \mathcal{B}(\mathbb{R}^d)$. Define the measure $\pi_\psi(.)$ by $\pi_\psi(B) = \pi(\psi^{-1}(B))$, for all $B \in \mathcal{B}([0, 1])$. Since $\psi(.)$ is one-to-one and onto, this uniquely defines the measure $\pi_\psi(.)$.

Now, consider for any $y \in [0, 1]$ and $B \in \mathcal{B}([0, 1])$ the Cesàro limit
$$\lim_{s \uparrow \infty} \frac{1}{s} \sum_{i=0}^{s-1} p_\psi^{(i)}(y, B) = \lim_{s \uparrow \infty} \frac{1}{s} \sum_{i=0}^{s-1} p^{(i)}(\psi^{-1}(y), \psi^{-1}(B)) = \pi(\psi^{-1}(B)) = \pi_\psi(B)$$
where the first equality holds by the definition of $p_\psi(., .)$, the second holds by the postulated stationary-ergodicity of $p(., .)$, and the third holds by the definition of $\pi_\psi(.)$. This establishes the existence of a unique probability measure on $([0, 1], \mathcal{B}([0, 1]))$ (namely $\pi_\psi(.)$) to which the Cesàro averages of transitions from any initial condition converge, which is our definition of the stationary-ergodicity of $p_\psi(., .)$.

By the one-to-one and onto properties of $\psi(.)$, the other direction follows symmetrically by replacing $\psi(.)$ by $\psi^{-1}(.)$, and $\psi^{-1}(.)$ by $(\psi^{-1})^{-1}(.) = \psi(.)$ in the above argument. This concludes the proof of the lemma. ∎

4   A consistent testing procedure

A test of the type described above can be performed for any pair of initial densities to obtain the required size. Power considerations, however, must take the statement $\forall f \in D$ seriously. A natural way to perform such a test would be to parametrize a particular class of densities D by a set of parameters $\theta \in \Theta$ where Θ is usually taken to be a compact space. One then tests if the maximum over all $\theta \in \Theta$ of the distances between the sample and the warranted distribution is zero. In such circumstances (e.g. [7]) even though the distribution of the *max* cannot be computed, we may be able to find an upper bound on the probability of the *max* being larger than some number. Unfortunately, such techniques will usually require the numerical computation of a multiple integral depending on the dimensionality of Θ. Numerical computation of integrals over a space of high dimensionality is highly costly and unreliable, and will generically only provide upper bounds for the desired probabilities. Our testing procedure bears some similarity to that parametrization approach, but instead of actually trying to maximize over all values of the density parameters, it will be a randomized test in the same spirit of [2],[3],[4], and [5].

Given two original densities f and g, we draw samples $y_1^f, ..., y_n^f$ and $y_1^g, ..., y_n^g$ as described in the previous section, where y_i^f is a random draw from $P^j f$ with probability $1/s$, $j = 0, 1, ..., s - 1$. We then transform the y_i^f and y_i^g data via the transformation ψ described above. Denote the resulting samples $x_1^f, ..., x_n^f$ and $x_1^g, ..., x_n^g$ where $x_i^f = \psi(y_i^f)$ and $x_i^g = \psi(y_i^g)$. The resulting x's are i.i.d. draws from F_ψ and G_ψ, respectively. We can now compute the uniform process

$$U_n^{f,g}(t) = \frac{1}{\sqrt{n}} \left[\#\{x_i^f | x_i^f < t; 1 \le i \le n\} - \#\{x_i^g | x_i^g < t; 1 \le i \le n\} \right]$$

Many tests of equality of f and g can now be used by testing $U_n^{f,g}(.) \equiv 0$. The Kolmogorov-Smirnov test derives the distribution of the sup-norm, the Cramèr-von Mises and other tests work with the L^2 norm, and test if that norm is small. See [18] for a number of those tests based on the empirical distribution function.

The idea of our test is to randomly select a pair of densities $(\tilde{f}, \tilde{g}) \in D^2(\mathbb{R}^d)$. We then show that under the null of stationary-ergodicity,

$$\|U_n^{\tilde{f},\tilde{g}}\| \xrightarrow{d} \text{Null\_Dist} \tag{4.1}$$

and under the alternative of non-stationary-ergodicity

$$\|U_n^{\tilde{f},\tilde{g}}\| \xrightarrow{p} \infty \tag{4.2}$$

where Null_Dist is a random variable with the distribution of our test statistic of goodness of fit under the null, $\|.\|$ is the norm used by that test, and the limits are taken as $s \uparrow \infty$ and then as $n \uparrow \infty$.

We consider the class of densities $f \in D[0, 1]$ that can be written in the form

$$f(x) = c_0 + c_1 x + ... + c_k x^k \tag{4.3}$$

which we consider to be sufficiently general for our analysis. For large k, this class closely approximates most densities.[2] To constrain the f to the space of densities $D[0,1]$, we require that

1. $c_0 \geq - \sum_{i:c_i<0} c_i.$ (non-negativity)

2. $\int_X f(x) = 1.$

To set up the algorithm, first notice that if we set

$$p_0 = c_0 + \sum_{i:c_i<0} c_i \tag{4.4}$$

and, for $i = 1, ..., k,$

$$p_i = \frac{c_i}{(i+1)} \qquad c_i \geq 0 \tag{4.5}$$

$$p_i = \frac{-ic_i}{(i+1)} \qquad c_i < 0 \tag{4.6}$$

then the vector $p_0, p_1, ..., p_k$ is a probability vector. This has been shown by Ahrens and Dieter [1], and used also in [8] to generate numbers from a particular $\tilde{f}(x) = \sum_{i=0}^{k} c_i x^i$. Their algorithm generates a discrete random variable Z from the multinomial distribution with the probability vector $p_0, ..., p_k$, and then generates x as

1. $x = U^{\frac{1}{Z+1}}$ if $c_Z \geq 0.$

2. $x = U_1^{\frac{1}{Z+1}} U_2$ otherwise.

then x is distributed according to $\tilde{f}$ [8, pp. 71-73]. This sets up our algorithm for selecting random coefficients $c_i, i = 1, ..., k.$

Algorithm A

1. Choose k large.

2. Generate $U_1, ..., U_k$ i.i.d. U[0,1].

3. Generate the order statistics $U_{(0)}, ..., U_{(k+1)}$ by sorting $U_1, ..., U_k$, and setting $U_{(0)} = 0$, and $U_{(k+1)} = 1.$

4. Generate $p_i = U_{(i+1)} - U_{(i)}, i = 0, ..., k.$ This is a random vector on the kth dimensional unit simplex. We now use those to find the coefficients of the polynomial.

---

[2]See for example [12], in which such representations are studied asymptotically but implemented for very small k.

5. Compute $c_0, ..., c_k$

   (a) for $i = 1, ..., k$, generate V_i (i.i.d. U[0,1])

   (b) $c_i = (i + 1)p_i$ if $V_i > 0.5$.

   (c) $c_i = \frac{-(i+1)}{i}p_i$ otherwise.

   (d) $c_0 = p_0 - \sum_{i:c_i < 0} c_i$.

6. Generate a discrete variate Z from the multinomial distribution with probability vector $p_0, ..., p_k$.

7. Now generate x from $\tilde{f}(x) = \sum_{i=0}^{k} c_i x^i$, generate U, U_1, U_2, i.i.d. U[0,1],

   (a) $x = U^{\frac{1}{Z+1}}$ if $c_Z \geq 0$.

   (b) $x = U_1^{\frac{1}{Z+1}} U_2$ otherwise.

8. Repeat steps 6, and 7 to generate as many data points as necessary from $\tilde{f}(x) = \sum_{i=0}^{k} c_i x^i$.

Now the procedure we follow to test the stationary-ergodicity of the known transition kernel $p(\xi, A)$ can be summarized by the following algorithm:

Algorithm B

1. Draw 2 initial densities f, g using algorithm A's steps 1-5.

2. Using steps 6-8 of algorithm A, draw 2 random samples of size n each from f and g. Label the two samples $\xi_1^f, ..., \xi_n^f$ and $\xi_1^g, ..., \xi_n^g$.

3. Transform each of the ξ's using the transformation ψ^{-1} of Lemma 1. Label the transformed data $y_i^f = \psi^{-1}(\xi_i^f)$ and $y_i^g = \psi^{-1}(\xi_i^g)$, for $i = 1, ..., n$.

4. For $i = 1, ..., n$, randomly select $j \in \{0, ..., s - 1\}$ taking on each of its potential values with probability $1/s$, then let $x_i^f \sim p^{(j)}(y_i^f, .)$. Now $(x_1^f, ..., x_n^f)$ is an i.i.d. sample from $\frac{1}{s}\sum_{i=0}^{s-1} P^i f(x)$. Similarly, generate $(x_1^g, ..., x_n^g)$ as an i.i.d. sample from $\frac{1}{s}\sum_{i=0}^{s-1} P^i g(x)$.

5. Transform all the x^f's and the x^g's via the mapping $x \mapsto \psi(x)$. This results in two samples of size n on $[0, 1]$. Apply your favorite 2-sample test of goodness of fit to those two samples.

Theorem 1 *Let $(\tilde{f}, \tilde{g}) \in D^2$ be generated by algorithm A, and apply a test of goodness of fit to the generated samples using algorithm B. Then (4.1) and (4.2) hold under the null of stationary-ergodicity and the alternative of non-stationary-ergodicity, respectively.*

8

To prove this theorem, we will first need the following result. If P is not stationary-ergodic, we define the set β_P of bad draws for our randomized test as the set:

$$\beta_P = \{(\tilde{f}, \tilde{g}) \in D^2 : \frac{1}{s} \sum_{i=0}^{s-1} [P^i \tilde{f}(x) - P^i \tilde{g}(x)] \to 0, \text{ for almost all } x \in \mathbb{R}^d\} \qquad (4.7)$$

Lemma 2 *Under the conditions of Theorem 1, if P is not stationary-ergodic*

$$Pr_A\{\beta_P\} = 0$$

where Pr_A is the probability under algorithm A.

Proof: Let $\tilde{f}$ and $\tilde{g}$ be drawn using algorithm A, then, for c_i's and d_i's that satisfy restrictions (4.4)-(4.6), $\tilde{f}(x) = c_0 + c_1 x + ... + c_k x^k$, and $\tilde{g}(x) = d_0 + d_1 x + ... + d_k x^k$. Now, we know that under the alternative of non-stationary-ergodicity, Harris recurrence [13, p.115] must be violated. Harris recurrence states that $\exists$ a measure m, such that all sets with positive m-measure will be visited infinitely often with probability one starting from any initial condition. Theorem 1 of Harris [13, p. 116], states that Harris recurrence is sufficient for there being a unique stationary measure which is absolutely continuous with respect to m. Since we restrict the unique stationary measure to be absolutely continuous with respect to Lebesgue measure, the failure of Harris recurrence must mean that the Markov chain is decomposable, i.e., there exists a set B with $0 < \mu(B) < 1$ and another set A with $0 < \mu(A) < 1$ such that $p^{(n)}(x, B) = 0$ for all $x \in A$, and all $n \geq 1$, where $\mu(.)$ is Lebesgue measure. In other words, there must be a set of positive measure A such that the iterates from some initial condition x_0 can fall in A with positive probability and the iterates from all points in A will never reach B. Let C be the union of all the Borel sets A whose iterates will never reach B. Then $E = [0,1] \setminus C$ is the set of points that will reach B with positive probability. No mass can escape from C to E by the definition of C. Hence, all the mass in E at all time periods must have started there.

We have now shown that for some set E with $1 > \mu(E) > 0$, for all i, $\int_E P^i \tilde{f}(x).dx = \int_E \tilde{f}(x).dx$. Similarly, for all i, $\int_E P^i \tilde{g}(x).dx = \int_E \tilde{g}(x).dx$. It follows that for all s:

$$\sum_{i=0}^{s} \int_E P^i \tilde{f}(x).dx = \sum_{i=0}^{s} \int_E P^i \tilde{g}(x).dx$$

if and only if $\int_E \tilde{f}(x).dx = \int_E \tilde{g}(x).dx$. Now, E is a Borel set and can be written as the (finite or countably infinite) union of intervals $E = \cup_{j=1}^{\infty}(a_j, b_j)$, and the coefficients of the polynomials $\tilde{f}$ and $\tilde{g}$ are randomly drawn using algorithm A,

$$Pr_A\{\beta_P\} \leq Pr_A\{\sum_{j=1}^{\infty} \int_{a_j}^{b_j} c_0 + c_1 x + ... + c_k x^k .dx = \sum_{j=1}^{\infty} \int_{a_j}^{b_j} d_0 + d_1 x + ... + d_k x^k .dx\}$$

The event in the R.H.S. probability requires a linear restriction on the values the c_i and d_i coefficients to hold. From step 5 of algorithm A, the probability of drawing coefficients that satisfy that linear restriction is clearly zero, and the proof is complete. ∎

9

Proof of Theorem 1: Under the null, we consider the characteristic function of $\|U_n^{\tilde{f}, \tilde{g}}\|$, and we can easily see that by bounded convergence,

$$\lim_{n \uparrow \infty} \lim_{s \uparrow \infty} E e^{it\|U_n^{\tilde{f}, \tilde{g}}\|} = \lim_{n \uparrow \infty} \lim_{s \uparrow \infty} E[E[e^{it\|U_n^{\tilde{f}, \tilde{g}}\|} | \tilde{f}, \tilde{g}]] = E e^{it \cdot \text{Null\_Dist}} \tag{4.8}$$

Under the alternative, we know that the two Cesàro averages $\frac{1}{s} \sum_{i=0}^{s-1} P^i \tilde{f}(x)$ and $\frac{1}{s} \sum_{i=0}^{s-1} P^i \tilde{g}(x)$ do not converge to the same density, hence, for some $\epsilon > 0$, the two empirical distributions resulting from our two size n samples must be more than ϵ apart, i.e.

$$\lim_{n \uparrow \infty} \lim_{s \uparrow \infty} Pr_A \{ \frac{1}{\sqrt{n}} \|U_n^{\tilde{f}, \tilde{g}}\| > \epsilon \} = \lim_{n \uparrow \infty} \lim_{s \uparrow \infty} Pr_A \{ \|\tilde{F}_n - \tilde{G}_n\| > \epsilon \} = Pr_A \{ D \setminus \beta_P \} = 1$$

Where $\tilde{F}_n$ and $\tilde{G}_n$ are the two empirical distribution functions from our two samples of size n. It follows that $Pr\{\|U_n^{\tilde{f}, \tilde{g}}\| > \epsilon \times \sqrt{n}\}$ converges to one, and hence, $\|U_n^{\tilde{f}, \tilde{g}}\| \xrightarrow{p} \infty$ and the desired result is proven. ■

5 Monte Carlo Investigation of small sample behavior

In this section, we report on Monte Carlo results investigating the small sample properties of our suggested testing procedure. For our investigation of the size of our test, we use Model 1:

$$x_{t+1} = 2x_t + \epsilon_t \ (mod\ 1)$$

where ϵ_t are distributed i.i.d. $U[0, 0.01]$. This law of motion is clearly ergodic (as a matter of fact, even with $\epsilon_t \equiv 0$, this would be an ergodic law of motion with the unique stationary density being $U[0, 1]$).

For our investigations of the power of the test, we use Model 2:

$$x_{t+1} = 2x_t + \epsilon_t \ (mod\ 0.5) \quad \text{if } x_t \leq 0.5$$

$$x_{t+1} = 0.5 + [2x_t + \epsilon_t \ (mod\ 0.5)] \quad \text{if } x_t > 0.5$$

where, again, ϵ_t are i.i.d. $U[0, 0.01]$. This law of motion has two ergodic subclasses $[0, 0.5]$, and $(0.5, 1.0)$.

For each of the models, we conduct the test 1000 times. Each test consists of randomly drawing two densities $\tilde{f}$ and $\tilde{g}$ (each of which is a kth order polynomial), and drawing n sample points from $(1/s) \sum_{i=0}^{s-1} P^i \tilde{f}$ and $(1/s) \sum_{i=0}^{s-1} P^i \tilde{g}$ following algorithms A and B. We then apply a two-sample Kolmogorov-Smirnov test to the so generated samples. For random number generation, we used Press et al's [17], subroutine *ran1()*, and we initialized it with the clock time each time we ran a new Monte Carlo at different values

of k, n, and s. For the Kolmogorov-Smirnov test, we used Press et al's [17] subroutine *kstwo()*, with its accompanying subroutines *probks()* and *sort()*. The rest of the code was written in C, and compiled, vectorized, and run on a Cray YMP2E/116. The following table shows the proportion of time we rejected the null hypothesis of stationary-ergodicity at the 0.01, 0.05, and 0.10 levels for various values of k, n and s. The size of the test seems quite reasonable, and its power performance (if not ideal) is still quite encouraging at such small values of n and s, and is significantly better for larger values of n and s. One should be careful since the theory proves the consistency of the test (and its having appropriate size) only as $s \uparrow \infty$ and then as $n \uparrow \infty$. Choosing s too small relative to n (e.g. $s = 10$ and $n = 100$, and $s = 40$ and $n = 500$ in Table I) results in an upward size distortion. Choosing s too large relative to n (e.g. $s = 100$, $n = 100$) results in a downward size distortion. A ratio of $n/s = 5$ (e.g. $s = 100$, $n = 500$ and to a lesser extent $s = 20$, $n = 100$) seems to yield excellent size behavior. The power of the test generally behaves in the right direction, increasing as n and s grow at appropriate rates.

Table I Monte Carlo Investigations of Size and Power Null of Stationary-Ergodicity 1000 simulations								
k	s	n	proportion rejected under Model 1 at level			proportion rejected under Model 2 at level		
			0.01	0.05	0.10	0.01	0.05	0.10
5	10	100	0.080	0.204	0.291	0.544	0.600	0.627
5	20	100	0.014	0.069	0.129	0.523	0.566	0.605
10	40	100	0.009	0.059	0.118	0.504	0.535	0.573
10	100	100	0.004	0.023	0.063	0.458	0.482	0.508
10	40	500	0.031	0.125	0.188	0.589	0.659	0.711
10	100	500	0.011	0.047	0.096	0.626	0.680	0.727

6 Concluding Remarks

We have constructed a test of stationary-ergodicity of a known Markovian transition kernel on $\mathbb{R}^d$. By known, all we mean is that one can generate random draws from some stochastic transition $x_{t+1} \sim p(x_t, .)$. This accommodates among other things simulations from models where closed form solutions cannot be explicitly written. There is no reason in principle why one cannot use this test on estimated laws of motion $\hat{p}_T(\xi, .)$ which are believed to be consistent estimators of some true $p(\xi, .)$ under the maintained hypothesis of stationary-ergodicity. Clearly, as $T \uparrow \infty$, the stationary ergodic or otherwise behavior of the transition $\hat{p}_T(\xi, .)$ will mimic that of the original $p(\xi, .)$. More work needs to be done, however, to generalize our test to that case, taking into consideration the estimation error.

Our Monte Carlo studies suggest that the choice of s and n for conducting the test must be made judiciously. The theory proves the consistency of the test as $s \uparrow \infty$ first, and then as $n \uparrow \infty$. If in practice n is too large or too small compared to s (and that may depend on the actual law of motion being tested), the size of the test will be distorted. As with all tests whose optimality properties are showed asymptotically, one should proceed cautiously when applying the test for finite values of the parameters. On the other hand, our test parameters k, n, and s are within our control, bounded only by computational limitations.

References

1. Ahrens, J. and Dieter, U. Computer Methods for Sampling From Gamma, Beta, Poisson and Binomial Distributions. *Computing* 12 (1974): 223-246.

2. Bierens, H. ARMAX Model Specification Testing, with an Application to Unemployment in the Netherlands. *Journal of Econometrics* 35 (1987): 161-190.

3. Bierens, H. ARMA Memory Index Modeling of Economic Time Series (with discussion). *Econometric Theory* 4 (1988): 35–59.

4. Bierens, H. A Consistent Conditional Moment Test of Functional Form. *Econometrica* 58, (1990): 1443-1458.

5. Bierens, H. and J. Hartog Non-Linear Regression with Discrete Explanatory Variables, with an Application to the Earnings Function. *Journal of Econometrics* 38 (1988): 269–299.

6. Cornfeld, I., S. Fomin and Ya. Sinai *Ergodic Theory.* Berlin: Springer-Verlag, 1982.

7. Davies, R. Hypothesis Testing When a Nuisance Parameter Is Present Only Under the Alternative. *Biometrica* 64 (1977): 247-54.

8. Devroye, L. *Non-Uniform Random Variate Generation.* New York: Springer-Verlag, 1986.

9. Domowitz, I. and M. El-Gamal A Test of the Harris Ergodicity of Stationary Dynamical Economic Models. RCER working paper # 157, University of Rochester, 1988.

10. Doob, J. *Stochastic Processes.* New York: John Wiley and Sons Inc., 1953.

11. El-Gamal, M. *Estimation in Economic Systems Characterized by Deterministic Chaos*, Ph.D Dissertation, Northwestern University, 1988.

12. Gallant, A. R. and G. Tauchen Seminonparametric Estimation of Conditionally Constrained Heterogeneous Processes. *Econometrica* 57 (1989).

13. Harris, T. E. The Existence of Stationary Measures for Certain Markov Processes. In J. Neyman (ed.), *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability.* Berkeley: Univ. of California Press, 1956.

14. Khmaladze, E. An Innovation Approach to Goodness-of-Fit Tests in $\mathbb{R}^m$. *The Annals of Statistics* 16 (1988): 1503-1516.

15. Loève, M. *Probability Theory II, 4th Edition.* New York: Springer-Verlag, 1978.

16. Piterbarg, V. and Fatalov, V. Exact Asymptotics for Probabilities of Large Deviations for Certain Gaussian Fields of Use in Statistics. In (A. N. Kolmogorov, ed.) *Probabilistic Methods of Research.* Moscow: Moscow University Press, 1983.

17. Press, W.H., B. P. Flannery, S. A. Teukolsky, W. T. Vetterling *Numerical Recipes in C: The Art of Scientific Computing.* Cambridge: Cambridge Univ. Press, 1988.

18. Shorack, G. and J. Wellner *Empirical Processes with Applications to Statistics.* New York: John Wiley and Sons, 1986.