Testing for the Martingale Hypothesis [1]

Joon Y. Park
School of Economics
Seoul National University

and

Yoon-Jae Whang
Department of Economics
Ewha University

Abstract

This paper proposes general speci cation tests for the martingale hypothesis. They can be used to test the null hypothesis that a given time series is a martingale process, against the alternative hypothesis that it is a stationary ergodic non-martingale process. We consider tests of two different types: one is a generalized Kolmogorov-Smirnov test and the other is a Cramer-von Mises type test. The tests introduced in the paper are simple to compute, neither depend upon any smoothing parameter nor require any resampling procedure to simulate the null distributions. The null distributions of the test statistics are given by the functionals of a continuous martingale process, which are free of any nuisance parameter. We show that the tests are consistent against stationary ergodic non-martingale alternatives, and further investigate the nite sample properties of the test statistics through simulation. Our tests are found to be rather powerful in moderate size samples against a wide variety of non-martingales including exponential autoregressive, threshold autoregressive, markov switching, and chaotic processes.

First Draft: June 10, 1999
This version: December 1, 1999

Key words and phrases: Martingale hypothesis, Kolmogorov-Smirnov test, Cramer-von Mises test, speci cation test.

[1] The authors are grateful to Don Andrews, Herman Bierens, Yoosoon Chang, In Choi, Miguel Delgado, Steven Durlauf, Atsushi Inoue, Peter Phillips, Dongwan Shin and seminar participants at Cowles Foundation Econometrics Conference on New Developments in Time Series Econometrics, October 1999, and at 1999 Kamp Econometrics, Pusan, Korea for helpful discussions and comments. This paper was completed while the authors are visiting Cowles Foundation for Research in Economics, whose hospitality is gratefully acknowledged. Park thanks the Department of Economics at Rice University, where he is an Adjunct Professor, for its continuing hospitality and secretarial support. Address correspondence to: Yoon-Jae Whang, Cowles Foundation for Research in Economics, Yale University, New Haven CT 06520-8281, U.S.A. Phone: 203-432-3721; Fax: 203-432-6167; E-mail: yoon-jae.whang@yale.edu

1. Introduction

In this paper, we introduce the specification tests for the martingale hypothesis. The martingale hypothesis has been considered to be very important in economics and other related fields, since it implies that the best predictor of future values of a time series given the current information set is just the current value of the time series. See, e.g., Hall (1978) for some supportive arguements that consumption is a martingale. The reader is also referred to Durlauf (1991) for more discussions on the martingale hypothesis arising in other contexts of economic theory.

Our tests can be used to test the null hypothesis that a given time series is a martingale process, against the alternative hypothesis that it is a stationary non-martingale process. The alternative hypothesis is hence very general and it encompasses, for example, a broad class of nonlinear stationary processes including exponential and threshold autoregressive processes, markov switching processes, and some of the chaotic processes (possibly with stochastic noise). Tong (1990) provides an extensive set of examples for such nonlinear time series processes.

We consider two types of specification tests, which can be regarded as the generalizations respectively of the Kolmogorov-Smirnov test and the Cramer-von Mises test of goodness of fit to the regression framework. Our tests are simple to compute and neither depend upon any smoothing parameter nor require any resampling procedure to simulate the null distributions. Their null limiting distributions are nicely characterized as functionals of a continuous martingale process. In particular, they do not involve any nuisance parameter, and can readily be obtained.

Our tests are comparable, in particular, to the tests by Durlauf (1991). To test the martingale hypothesis for a given time series, he looks at the spectrum of the first differences and see whether it is constant. Naturally, his tests are designed to be powerful against all non-martingales generated by serially correlated innovations. For the Gaussian model, the absence of correlation in the first differences occurs when and only when the underlying process is a martingale. His test is thus consistent also against all Gaussian non-martingale processes. There are, however, nonlinear non-Gaussian processes which are non-martinagles with serially uncorrelated processes [see, e.g., Brockett, Hinich and Patterson (1988, p.658) for such an example]. The tests by Durlauf are not expected to have discriminatory powers against such non-martingale processes. Our tests do have powers against such nonlinear non-Gaussian non-martingales, and are more general than his in this respect.

We also note that there is a huge literature related to the testing problem considered here. One branch of the literature deals with testing for a unit root [see, e.g., Stock (1994) or Phillips (1997) for an excellent survey on the subject]. The unit root hypothesis, however, is obviously more general than our martingale hypothesis. Also, the alternatives considered for most of the existing unit root tests are much more restrictive than ours: Their alternatives are usually stationary *linear* autoregressive processes, whereas our alternatives allow general *nonlinear* processes. Therefore, we believe our tests might deliver further insight on the property of a given time series, especially when the underlying data generating mechanism is nonlinear. The other branch of the related literature consists of the nonlinearity tests for time series. Examples of such tests include, among others, An and Bing (1991), Brockett, Hinich, and Patterson (1988), Chan and Tong (1986), Hinich (1982), Hjellvik and Tjøstheim (1995) and Koul and Stute (1999), Luukkonen, Saikkonen, and Teräsvirta (1988). These tests are also consistent against general nonlinear alternatives, but they only look at *stationary* null and alternative hypotheses. Our tests consider *nonstationary* processes. To the best of our knowledge, the asymptotic behaviors of the nonlinearity tests for nonstationary processes have not yet been investigated. Our tests are also related to the model speci cation tests by Bierens (1990) and de Jong (1996).

The remainder of this paper is organized as follows. Section 2 introduces the null and alternative hypotheses and de nes the test statistics. Section 3 derives the asymptotic null distributions of the test statistics. Section 4 establishes the consistency of our tests. Section 5 reports the results from simulation experiments. Section 6 contains the proofs for the theorems in the main text.

2. The Hypotheses and Test Statistics

Let a time series (y_t) be given, and let $(\mathcal{F}_t)$ be a ltration to which (y_t) is adapted. The null hypothesis of interest is that (y_t) is a martingale with respect to the ltration $(\mathcal{F}_t)$, i.e.,

$$H_0 : \mathbf{E}(y_t|\mathcal{F}_{t-1}) = y_{t-1} \quad \text{a.s.} \tag{1}$$

for all $t \geq 1$, where $\mathbf{E}(\cdot|\mathcal{F}_{t-1})$ denotes as usual the conditional expectation given $\mathcal{F}_{t-1}$. The alternative hypothesis is that (y_t) is a stationary ergodic non-martingale process for which we have

$$H_1 : \mathbf{E}(y_t|\mathcal{F}_{t-1}) \neq y_{t-1} \tag{2}$$

with some positive probability.

We now de ne our test statistics. Let

$$Q_n(y) = \frac{1}{n}\sum_{t=1}^{n} q_t(y) = \frac{1}{n}\sum_{t=1}^{n} \triangle y_t 1\{y_{t-1} \leq y\}, \tag{3}$$

where and elsewhere in the paper we denote by $\triangle$ the usual difference operator and by $1\{\cdot\}$ the indicator function. Note that, under the null hypothesis (1), we have $\mathbf{E}Q_n(y) = 0$ for all $y \in \mathbf{R}$. In contrast, we expect that there exists a point $y \in \mathbf{R}$ for which $\mathbf{E}Q_n(y) \neq 0$ under the alternative hypothesis (2), see the discussion below and Section 4. Therefore, we consider $Q_n(y)$ as the basis of our test statistics for the martingale hypothesis (1). We let

$$P_n(y) = (\sqrt{n}/\sigma_n)Q_n(y) \tag{4}$$

where σ_n^2 is the usual variance estimate for $(\triangle y_t)$, i.e., $\sigma_n^2 = \sum_{t=1}^{n}(\triangle y_t)^2/n$, and de ne a Kolmogorov-Smirnov type statistic

$$S_n = \sup_{y\in\mathbf{R}} |P_n(y)| \tag{5}$$

We may also look at a Cramer-von Mises type statistic. In particular, we consider the statistic given by

$$T_n = \int P_n^2(y) d\mu_n(y) = \frac{1}{n}\sum_{t=1}^{n} P_n^2(y_{t-1}) \tag{6}$$

where P_n is as defined in (4), and μ_n denotes the empirical distribution of (y_{t-1}).

If applied to the first-order Markovian processes, the tests based on the statistics S_n and T_n are expected to be powerful against all stationary and ergodic alternatives, i.e., (2). This is because for all such processes $\mathbf{E}Q_n(y) \neq 0$ at some $y \in \mathbf{R}$. It will indeed be shown later that our tests are consistent against such alternatives. Our tests, of course, are consistent against non-Markovian alternatives as long as $\mathbf{E}Q_n(y) \neq 0$ for some $y \in \mathbf{R}$, see Section 4 below. However, it is expected that they have most effective discriminatory powers against first-order Markovian non-martingales.[2] It might be also possible to consider the statistics based on

$$Q_n(x_1, \ldots, x_\kappa) = \frac{1}{n}\sum_{t=1}^{n} \triangle y_t 1\{\triangle y_{t-1} \leq x_1\} \cdots 1\{\triangle y_{t-\kappa+1} \leq x_{\kappa-1}\} 1\{y_{t-\kappa} \leq x_\kappa\} \tag{7}$$

in place of $Q_n(y)$ introduced in (3) to more effectively discriminate our martingale null hypothesis against nonmartingale alternatives that are not first-Markovian. The extension along this line, however, would not be pursued in this paper both for clarity of the exposition and for some technical difficulties.[3]

The martingale hypothesis is intimately related to the unit root hypothesis, though strictly speaking none of them generally implies the other.[4] It therefore seems interesting to compare our tests with the unit root test by Dickey and Fuller (1979). Their test is most commonly used to test for the unit root. The test relies on the t-statistic on the coefficient β in the regression

$$\triangle y_t = \beta y_{t-1} + \varepsilon_t$$

where (ε_t) is assumed to be martingale differences. We may thus expect that the test has some discriminatory powers against our alternatives (2), which may be reformulated as $\mathbf{E}(\triangle y_t | \mathcal{F}_{t-1}) \neq 0$. The test, however, concentrates on one possible violation of the martingale hypothesis, i.e., the one into the direction spanned linearly by y_{t-1}, as is the case for the stationary first order autoregression. In contrast, our tests

[2] The stationarity and ergodicity assumption of (y_t) under the alternative hypothesis is also not crucial. It is made just to simplify the proofs for test consistency. We may well expect that our tests are also powerful against some of nonstationary and nonergodic non-martingales.

[3] The relevant asymptotic theory becomes much more involved and the approach used in the paper is not directly extendable to deal with the tests based on the multi-parameter stochastic process $Q_n(x_1, \ldots, x_\kappa)$ defined in (7). See also Koul and Stute(1999, Remark 2.4) and Khmaladze(1988) for discussions on related issues.

[4] Here we use the term unit root as defined in Stock (1994) or Phillips (1997). The unit root process with correlated innovations is in general not a martingale. Conversely, the martingale whose differences vanishing asymptotically is not a unit root process. We, however, are mostly concerned with the unit root martingales in the paper.

look into many other nonlinear directions as well for the violation of the martingale hypothesis.

The martingale hypothesis considered here can also be tested using the approaches taken by Bierens (1990) and de Jong (1996) in a broader context of general model speci cations. In particular, the test proposed by de Jong (1996) can be used to test our hypothesis if it is applied to the rst differeces ($\triangle y_t$) of the given time series (y_t), and if it is made conditional on y_0. Note that the martingale hypothesis (1) with the natural ltration ($\mathcal{F}_t$) for (y_t) can be written as

$$\mathbf{E}(\triangle y_t | \triangle y_{t-1}, \ldots, \triangle y_1) = 0 \quad \text{a.s.}$$

if we assume that y_0 is xed. Our tests, however, look more directly into the martingale hypothesis, and are more closely related in their aim to the tests developed by Durlauf (1991).

3. The Null Distributions

In this section, we derive the null distributions of the test statistics S_n and T_n introduced in the previous section. We let

$$u_t = \triangle y_t$$

and de ne ($\mathcal{F}_t$) to be the ltration introduced earlier. We assume

3.1 Assumption $(u_t, \mathcal{F}_t)$ is a martingale difference sequence such that

(a) $\displaystyle \frac{1}{n}\sum_{t=1}^{n} \mathbf{E}(u_t^2|\mathcal{F}_{t-1}) \to_p \sigma^2$, and

(b) $\displaystyle \frac{1}{n}\sum_{t=1}^{n} \mathbf{E}\left(u_t^2 1\{|u_t| > \varepsilon\sqrt{n}\}\big|\,\mathcal{F}_{t-1}\right) \to_p 0$ for any $\varepsilon > 0$

as $n \to \infty$.

Note that the condition in the part (a) of Assumption 3.1 allows the innovation sequence (u_t) to be heteroskedastic, conditionally and/or unconditionally, as long as it is averaged out in the limit. Our subsequent theory is therefore applicable in particular for the martingales driven by ARCH-type innovations. The part (b) of Assumption 3.1 is the conditional version of Linderberg condition, which is routinely imposed to obtain the martingale limit theory. It is satis ed if, for instance, $\sup_{t\geq 1} \mathbf{E}(|u_t|^p|\mathcal{F}_{t-1}) < \infty$ a.s. for some $p > 2$.

Under Assumption 3.1, the usual variance estimator $\sigma_n^2 = (1/n)\sum_{t=1}^n u_t^2$ of (u_t) is consistent for its asymptotic variance σ^2, i.e., $\sigma_n^2 \to_p \sigma^2$, which we state formally as a lemma.

3.2 Lemma Let Assumption 3.1 hold. Then we have $\sigma_n^2 \to_p \sigma^2$ as $n \to \infty$.

Moreover, an invariance principle applies to the particial sum of (u_t). Under Assumption 3.1, the appropriately normalized partial sum process indeed converges in distribution to Brownian motion. This is well known [see Theorem 4.1 and the subsequent discussions in Hall and Heyde (1980)]. For the development of our theory, however, it is more convenient to use a direct embedding of the standardized partial sum process into a Brownian motion in an extended probability space. The embedding will be introduced below.

Due to the Skorohod representation theorem [see, e.g., Hall and Heyde (1980, Theorem A.1, p269)], there exists a probability space supporting a standard Brownian motion and a time change $(\tau_i)_{i\geq 0}$ such that

$$W_{\tau_{k-1}/n} =_d \frac{1}{\sigma\sqrt{n}} \sum_{i=1}^{k} u_i \tag{8}$$

for $1 \leq k \leq n$ and $n \geq 1$. In particular, under Assumption 3.1, we may choose the time change $(\tau_i)_{i\geq 0}$ so that

$$\max_{1\leq i\leq n} \left| \frac{\tau_i - i}{n} \right| \to_p 0$$

as shown in Hall and Heyde (1980, Lemma 4.4, p106). Therefore, if we de ne a process W_n by

$$W_{nt} = \sum_{i=1}^{n} W_{\tau_{i-1}/n} 1\left\{ \frac{\tau_{i-1}}{n} \leq t < \frac{\tau_i}{n} \right\} \tag{9}$$

then it follows from the continuity of the Brownian motion sample path [see, e.g., Hida (1980, Theorem 2.6)] that

$$\sup_{t\in[0,1]} |W_{nt} - W_t| \to_p 0$$

as $n \to \infty$. Of course, the weak convergence

$$W_n \to_d W \tag{10}$$

holds as a consequence in $D[0,1]$, the space of cadlag functions on $[0,1]$, endowed with the supremum norm. For the rest of the paper, we will assume that the distributional equality in (8) is indeed the equality. This causes no loss in generality, since our theory only involves the distributional results for test statistics. Yet, the convention greatly simplies and clari es our subsequent theoretical developments.

We now rewrite $Q_n(y)$ as

$$Q_n(y) = \frac{1}{n} \sum_{t=1}^{n} u_t 1\{y_{t-1} \leq y\}$$

Moreover, for the expositional brevity, we let σ^2 be known and P_n be de ned with σ^2 in place of σ_n^2, i.e., $P_n(y) = (\sqrt{n}/\sigma)Q_n(y)$. It is quite clear that the replacement of σ_n^2 by σ^2 would not change any of our subsequent results as long as $\sigma_n^2 \to_p \sigma^2$. De ne

$$M_n(x) = P_n(\sigma x\sqrt{n}) \tag{11}$$

Then we may write S_n as

$$S_n = \sup_{x\in\mathbf{R}} |M_n(x)| \tag{12}$$

Moreover, we have

$$T_n = \int_0^1 M_n^2(W_{nt})\, dt \tag{13}$$

where W_n is the process introduced in (9).

From now on, we regard M_n de ned in (11) as a stochastic process with parameter $x \in \mathbf{R}$. It takes values in $D(\mathbf{R})$, i.e., the space of cadlag functions on $\mathbf{R}$. As before, we endow $D(\mathbf{R})$ also with the uniform topology. We now write the process M_n introduced in (11) as

$$\begin{aligned} M_n(x) &= \frac{1}{\sigma\sqrt{n}}\sum_{t=1}^n u_t\, 1\left\{\frac{y_{t-1}}{\sigma\sqrt{n}} \leq x\right\} \\ &= \int_0^1 1\{W_{nt} \leq x\}\, dW_t + o_p(1) \end{aligned}$$

which holds uniformly in $x \in \mathbf{R}$. Given the weak convergence (10) of W_n to W in $D[0,1]$, it is well expected that the stochastic process M_n weakly converges in $D(\mathbf{R})$ to M de ned by

$$M(x) = \int_0^1 1\{W_t \leq x\}\, dW_t \tag{14}$$

as $n \to \infty$. The weak convergence is presented in the following lemma.

3.3 Lemma Under Assumption 3.1, we have $M_n \to_d M$ as $n \to \infty$.

The asymptotic distributions of the statistics S_n and T_n can now be readily derived from the result in Lemma 3.3 and the continuous mapping theorem, since they are continuous functionals of M_n.

3.4 Theorem Suppose that Assumption 3.1 holds. Then we have

$$\begin{aligned} S_n &\to_d S = \sup_{x\in\mathbf{R}} |M(x)| \\ T_n &\to_d T = \int_0^1 M^2(W_t)\, dt \end{aligned}$$

as $n \to \infty$.

The proofs of the above theorems and some of our subsequent results rely on the local time of the limit Brownian motion W, which we denote by $L(t, s)$ with t and s

signifying respectively the time and space parameters. It may be de ned as

$$L(t,s) = \lim_{\varepsilon \to 0} \frac{1}{2\varepsilon} \int_0^t 1\{|W_r - s| \le \varepsilon\}\, dr$$

and can be interpreted as the time spent by W, up to time t, in the immediate vicinity of the level s. The local time L yields the equality

$$\int_0^t F(W_t)\, dt = \int_{-\infty}^{\infty} F(s) L(t,s)\, ds$$

for any locally integrable function $F : \mathbf{R} \to \mathbf{R}$, which is known as the occupation times formula. The reader is referred to Chung and Williams (1990) for an introduction to the Brownian local time and occupation times formula.

We also need to further investigate the properties of the limit process M to fully understand the asymptotic properties of the test statistics S_n and T_n.

3.5 Lemma The limit process M is a martingale (with respect to its natural l-tration). Moreover, for any $r \ge 2$ and $x, y \in \mathbf{R}$,

$$\mathbf{E}|M(x) - M(y)|^r \le c_r |x - y|^{r/2}$$

where c_r is a constant depending only upon r.

3.6 Proposition There is a modi cation of M, whose paths are Hölder continuous of order $r \in [0, 1/2)$.

The results in Lemma 3.5 and Proposition 3.6 imply in particular that the limit process M is a continuous martingale.

It is interesting to compare our results with the one established for the stationary ergodic case by, e.g., Koul and Stute (1999). If we consider a strictly stationary and ergodic sequence (y_t) generated as

$$y_t = m(y_{t-1}) + u_t$$

with some function m, then we have in $D(\mathbf{R})$

$$P_n(\sigma y) \to_d B(\tau^2(y)) \tag{15}$$

with $\tau^2(y) = \mathbf{P}\{(y_t/\sigma) \le y\}$, where B is a standard Brownian motion. This is shown in Koul and Stute (1999). In contrast, here we consider a nonstationary and nonergodic process (y_t), and establish the weak convergence in $D(\mathbf{R})$

$$P_n(\sigma x \sqrt{n}) \to_d B([M](x)) \tag{16}$$

where M, $[M](x) = \int_0^1 1\{W_t \le x\} dt$, is the quadratic variation of M de ned in (14). Note that the continuous martingale $M(\cdot)$ can be written as a time-changed Brownian

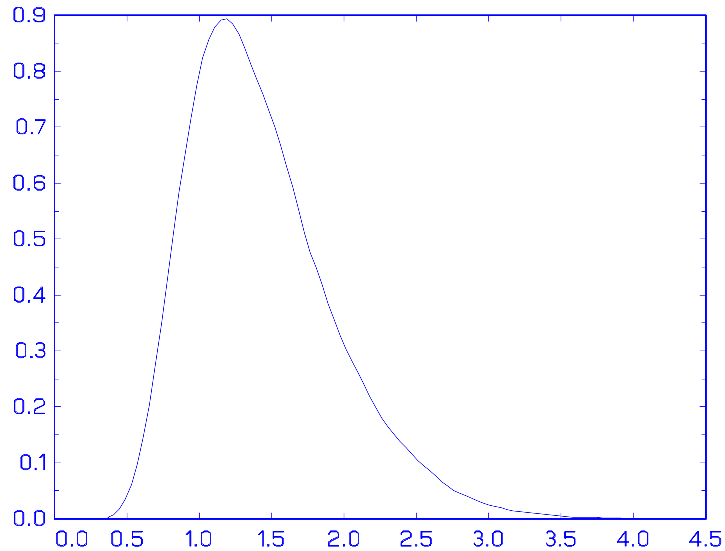

Figure 1: Probability Density of S

motion $B([M](\cdot))$, where B is the DDS (Dambis-Dubins-Schwarz) Brownian motion [see, e.g., Revuz and Yor (1994)], as we presented above in (16).

The weak limits of P_n for the stationary ergodic case and for the nonstationary nonergodic case given respectively in (15) and (16) are thus quite similar. Both limit processes are represented by time-changed Brownian motions. For the ergodic time series (y_t), the required time change is given by a deterministic function τ^2. We must, however, use a stochastic time change to represent the limit process if the time series (y_t) is nonergodic. In this sense, our results here can be regarded as an extension of Koul and Stute (1999) to nonstationary nonergodic autoregressive models. To define P_n for the nonstationary nonergodic (y_t), we only need the normalization of the ordinate $y = x\sqrt{n}$, so that it grows at the same rate as the underlying process.

Since the process M is continuous and effectively stopped at

$$s_{\min} = \inf_{t\in[0,1]} W_t \quad \text{and} \quad s_{\max} = \sup_{t\in[0,1]} W_t$$

i.e., $M(x) = M(s_{\min}) = 0$ for all $x \leq s_{\min}$ and $M(x) = M(s_{\max}) = W_1$ for all $x \geq s_{\max}$, it is obvious that the limit random variable S introduced in Theorem 3.4 is a.s. well defined. Moreover, the process M is a.s. of locally integrable sample path, and therefore we have

$$\int_0^1 M^2(W_t)\,dt = \int_{-\infty}^{\infty} M^2(x)L(1,x)\,dx$$

due to the occupation times formula. This shows that the limit random variable T in Theorem 3.4 is also well defined a.s.

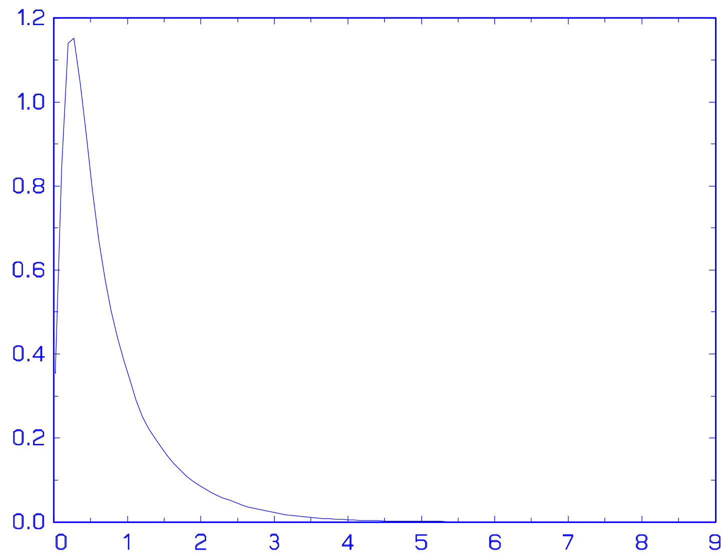

Figure 2: Probability Density of T

Table 1: Asymptotic Critical Values of S_n and T_n

sig. level (α)	0.99	0.95	0.90	0.10	0.05	0.01
S_n	0.612	0.765	0.865	2.119	2.388	2.911
T_n	0.055	0.101	0.145	1.650	2.165	3.328

The distributions of S and T, i.e., the limit distributions of the test statistics S_n and T_n de ned in (5) and (6) respectively, can readily be obtained through simulations and their probability densities are skeched in Figures 1 and 2. Approximately, the distribution of S (T) has mean 1.433 (0.746), median 1.350 (0.520), standard deviation 0.502 (0.704) and excess kurtosis 1.044 (7.274) and is skewed to the right with skewness 0.911 (2.198). The asymptotic critical values of the tests S_n and T_n are given in Table 1.

4. Consistency of the Tests

In this section, we establish the consistency of our tests based on the statistics S_n and T_n. Under the alternative hypothesis, we suppose the following assumption holds.

4.1 Assumption (y_t) is a strictly stationary and ergodic rst-order Markov process that is square integrable.

4.2 Theorem Suppose that Assumption 4.1 holds. Then, under the alternative hypothesis H_1,

$$S_n,\ T_n \to_p \infty$$

as $n \to \infty$.

Theorem 4.2 shows that the tests S_n and T_n are consistent if we reject the null hypothesis when they take large values.

The strict stationarity in Assumption 4.1 is sufficient, but not necessary. It is assumed here mainly to ease the proofs and expositions. We may certainly allow for more general weakly dependent and mildly heterogeneous processes under appropriate moment and mixing conditions.

The rst-order Markovian assumption can also be relaxed. In fact, it is easy to see from the proof of Theorem 4.2 that our tests are consistent against all (possibly non-Markovian) alternatives that are strictly stationary, ergodic, and square intergrable, as long as they satisfy the condition $\mathbf{E}\triangle y_t 1\{y_{t-1} \leq y\} \neq 0$ for some $y \in \mathbf{R}$. Note that the required condition holds if and only if $\mu(\mathbf{E}(\triangle y_t | y_{t-1}) \neq 0) > 0$, where we denote by μ the distribution of (y_t). If (y_t) is rst-order Markovian, the latter condition is equivalent to the alternative hypothesis (2). If (y_t) is not rst-order Markovian, however, the condition is stronger than (2) in general.

5. Simulation Results

In this section, we examine the nite sample performance of our tests in a small scale simulation experiment. We choose the eight different models described in Table 2 to generate simulated data. Model NULL generates random walk processes possibly with GARCH errors and is considered to evaluate the size performance of our tests. The other models are considered to see the power performance of our tests.

Table 2. Data Generating Processes

Model	DGP ($\varepsilon_t \sim i.i.d.\ N(0,1)$)
NULL	$y_t = y_{t-1} + u_t$, where $u_t = \sigma_t \varepsilon_t,\ \sigma_t^2 = 1 + \theta_1 u_{t-1}^2 + \theta_2 \sigma_{t-1}^2$
ARMA	$y_t = \theta_1 y_{t-1} + \theta_2 \varepsilon_{t-1} + \varepsilon_t$
EXAR	$y_t = \theta_1 y_{t-1} + \theta_2 y_{t-1} \exp(-0.1\,\lvert y_{t-1}\rvert) + \varepsilon_t$
TAR	$y_t = \theta_1 y_{t-1} 1\{\lvert y_{t-1}\rvert < \theta_2\} + 0.9 y_{t-1} 1\{\lvert y_{t-1}\rvert \geq \theta_2\} + \varepsilon_t$
BL	$y_t = \theta_1 y_{t-1} + \theta_2 y_{t-1} \varepsilon_{t-1} + \varepsilon_t$
NLMA	$y_t = \theta_1 y_{t-1} + \theta_2 \varepsilon_{t-1} \varepsilon_{t-2} + \varepsilon_t$
MARKOV	$y_t - \mu_{s_t} = \theta_1 (y_{t-1} - \mu_{s_{t-1}}) + \varepsilon_t,\ s_t = 0$ or $1,\ \mu_0 = 0,\ \mu_1 = 1,$ $\theta_2 = P(s_t = 0 \mid s_{t-1} = 0) = P(s_t = 1 \mid s_{t-1} = 1)$
FM	$y_t = m_t + u_t$, where $m_t = \theta_1 y_{t-1}(1 - y_{t-1}),\ u_t = \theta_2 v_t \eta_t,$ $v_t = \min\{m_t,\ 1 - m_t\}$, and $\eta_t \sim i.i.d.$ Uniform$(0,1)$

Model ARMA generates an autoregressive moving average process of order (1,1). Model EXAR is an exponential autoregressive model. Model TAR is a threshold autoregressive model of order 1. This model can capture the possibility of asymmetric movements in a time series, see Tong (1990, Section 3.3).[5] Model BL is a bilinear model. This model introduces coefficients that are linear function of the error term and is considered to lie somewhere between the xed coefficient autoregressive models and the random coefficient autoregressive models, see also Tong (1990, p.114). Model NLMA is a nonlinear moving average model. Model MARKOV is a markov switching model, see Hamilton (1989) for motivation. Finally, Model FM is a Feigenbaum map with system noise. When $\theta_1 = 4$, this map generates a chaotic process which is a globally bounded but locally explosive stationary process, see for example Whang and Linton (1999) and the references therein for discussions about chaotic processes.

In each of the model, we generate (ε_t) independently from the standard normal distribution and set the initial values, e.g., y_0, $\varepsilon_0, \varepsilon_{-1}$ to zero. A total of 1,000 replications are used for each experiment. We take $n = 100, 250, 500, 1000$ and report for each n the rejection probabilities of the test with nominal size $\alpha = 0.05$. The results corresponding to different nominal sizes were similar and hence are not reported.

Tables 3-10 present the rejection probabilities of our tests based on the statistics S_n and T_n. We compare the performance of our tests with the Cramer-von Mises type test of the martingale hypothesis proposed by Durlauf (1991), denoted as CVM_n[6]. Table 3 shows that our tests, designated as S_n and T_n, have reasonably good size performance and the size performance is little affected by the GARCH structure of the errors. On the other hand, the test CVM_n tends to over-reject when the errors follow GARCH processes.

Table 4-10 report the nite sample performances of our tests against a wide variety of alternative non-martingale processes. The performances of our tests are reasonably good in general, but they are somewhat critically dependent upon the underlying data generating processes.

Table 4 considers the case of the ARMA(1,1) process. The overall performance of our tests against the stationary ARMA processes appears to be reasonably good. However, the performances of our tests against the near-unit root process are somewhat unsatisfactory when the sample size is small. When the autoregressive coef- cient is close to unity, i.e., $\theta_1 = .95$, our tests indeed do not seem to have any discriminatory power in samples of size less than $n = 250$. Though it is also far from being satisfactory, the Durlauf test has better powers than our tests in small samples. The comparsion, however, is reversed drastically as the sample size increases. For the samples as large as $n = 1,000$, our tests S_n and T_n, especially the one based on T_n, have effective discriminating powers against the near-unit root alternative. The power of the Durlauf CVM_n test, however, improves only very slowly as the sample

[5] We have also considered momentum threshold autoregressive models (or MTAR models), which are introduced by Enders and Granger (1998), but the simulation results were similar to those of TAR and hence are not reported here.

[6] In our simulation experiment, we also considered the Kolomogorov-Smirnov type test KS_n of Durlauf (1991). But the test was unambiguously dominated by CVM_n in both size and power performance in almost all the cases we considered and hence the results for KS_n are not reported here.

size increases. When there is a moving average component, i.e., $\theta_2 \neq 0$, the performances of all three tests become slightly worse. Neverthless, the comparison between our tests S_n and T_n with the Durlauf CVM_n remains to be largely the same.

Table 5 gives the rejection probabilities when the data are generated from exponential autoregressive processes. It shows that both S_n and T_n perform well for samples of moderately large size. In particular, their performances are substantially better than that of CVM_n in large samples. For samples of small size, however, CVM_n performs better than S_n and T_n in several cases. As for the case of the stationary ARMA alternatives, performances of our tests S_n and T_n improve rapidly as the sample size increases. This is not so for the Durlauf CVM_n test. The power of CVM_n increases only very slowly.

Table 6 shows that our tests are consistent against the threshold autoregressive models. The rejection probabilities increase as θ_1 decreases (i.e., more asymmetry exists) or as θ_2 increases (i.e., the regime with high frequency movements occurs more often). The results also show that our tests have superior power to CVM_n especially when n is large. Table 7 reports the power performance of the tests against bilinear models. Our tests are consistent in all of the cases we considered and have generally better performance than CVM_n except for a few cases with small sample sizes.

Table 8 presents the results for nonlinear moving average models. Our tests exhibit substantially better power performance than CVM_n in relatively large samples, as the coefficient for the linear autoregressive part θ_1 gets close to unity. The results for the markov swithcing models are reported in Table 9. All three tests appear to have satisfactory discriminatory powers against the nonmartingale markov switching models unless they have the autoregressive coefficient θ_1 close to unity. The nite sample powers of our tests S_n and T_n against the nonmartingale markov switching models with the near-unity autoregressive coefficient can be quite low, when the sample size is small. However, they increase rapidly as the sample size increases. For the CVM_n test, the rate of increase in powers with respect to sample size is much slower, as is for many other cases considered here.

Finally, Table 10 shows that our tests are consistent against the Feigenbaum map with noise. It shows that the powers increase as the process becomes chaotic (i.e., $\theta_1 = 4$) and as the process has more system noise (i.e., as θ_2 increases). One can see that our tests perform better than CVM_n when $\theta_1 = 2.5$, while all the tests have complete distinguishing power against the case $\theta_2 = 4$.

Table 3. Rejection Probabilities (DGP: NULL)

(θ_1,θ_2)	n	S_n	T_n	CVM_n	(θ_1,θ_2)	n	S_n	T_n	CVM_n
	100	.043	.044	.033		100	.042	.047	.076
(.0, .0)	250	.048	.047	.030	(.2, .3)	250	.049	.042	.090
	500	.028	.041	.031		500	.039	.045	.082
	1000	.042	.042	.035		1000	.040	.044	.095
	100	.043	.050	.101		100	.039	.050	.116
(.3, .0)	250	.049	.043	.114	(.3, .4)	250	.050	.041	.145
	500	.040	.047	.118		500	.040	.044	.164
	1000	.039	.049	.114		1000	.039	.038	.171
	100	.032	.054	.310		100	.035	.051	.271
(.9, .0)	250	.035	.049	.453	(.7, .2)	250	.043	.045	.403
	500	.038	.047	.535		500	.040	.051	.483
	1000	.043	.051	.656		1000	.043	.051	.578

Table 4. Rejection Probabilities (DGP: ARMA)

(θ_1,θ_2)	n	S_n	T_n	CVM_n	(θ_1,θ_2)	n	S_n	T_n	CVM_n
	100	.822	.989	.990		100	.457	.818	.657
(.3, .0)	250	1.00	1.00.	1.00	(.3, .2)	250	1.00	1.00	1.00
	500	1.00	1.00	1.00		500	1.00	1.00	1.00
	1000	1.00	1.00	1.00		1000	1.00	1.00	1.00
	100	.337	.688	.755		100	.094	.229	.160
(.5, .0)	250	1.00	1.00	.998	(.5, .2)	250	.993	1.00	.700
	500	1.00	1.00	1.00		500	1.00	1.00	.995
	1000	1.00	1.00	1.00		1000	1.00	1.00	1.00
	100	.000	.000	.038		100	.001	.008	.045
(.95, .0)	250	.003	.001	.045	(.7, .2)	250	.530	.855	.182
	500	.040	.040	.067		500	1.00	1.00	.588
	1000	.484	.735	.103		1000	1.00	1.00	.986

Table 5. Rejection Probabilities (DGP: EXAR)

(θ_1,θ_2)	n	S_n	T_n	CVM_n	(θ_1,θ_2)	n	S_n	T_n	CVM_n
	100	.010	.022	.188		100	.005	.021	.030
(.6, .2)	250	.598	.905	.515	(.9, .2)	250	.020	.028	.044
	500	1.00	1.00	.905		500	.071	.118	.067
	1000	1.00	1.00	1.00		1000	.176	.177	.096
	100	.001	.001	.111		100	.086	.224	.031
(.6, .3)	250	.148	.308	.257	(.9, .3)	250	.185	.342	.061
	500	.947	1.00	.582		500	.609	.703	.145
	1000	1.00	1.00	.931		1000	.976	.976	.319
	100	.000	.000	.070		100	.299	.474	.021
(.6, .4)	250	.012	.024	.114	(.9, .4)	250	.427	.505	.057
	500	.307	.692	.278		500	.837	.937	.198
	1000	1.00	1.00	.556		1000	1.00	1.00	.536

Table 6. Rejection Probabilities (DGP: TAR)

(θ_1, θ_2)	n	S_n	T_n	CVM_n	(θ_1, θ_2)	n	S_n	T_n	CVM_n
	100	.016	.011	.090		100	.500	.616	.655
$(.3, 1.0)$	250	.379	.270	.166	$(.3, 2.0)$	250	.997	.997	.948
	500	.972	.950	.349		500	1.00	1.00	1.00
	1000	1.00	1.00	.664		1000	1.00	1.00	1.00
	100	.005	.004	.071		100	.151	.211	.328
$(.5, 1.0)$	250	.207	.167	.128	$(.5, 2.0)$	250	.956	.957	.688
	500	.880	.894	.276		500	1.00	1.00	.952
	1000	1.00	1.00	.547		1000	1.00	1.00	1.00
	100	.001	.000	.060		100	.017	.029	.118
$(.7, 1.0)$	250	.073	.085	.102	$(.7, 2.0)$	250	.456	.499	.249
	500	.690	.810	.221		500	.990	.994	.525
	1000	1.00	1.00	.441		1000	1.00	1.00	.869

Table 7. Rejection Probabilities (DGP:BL)

(θ_1, θ_2)	n	S_n	T_n	CVM_n	(θ_1, θ_2)	n	S_n	T_n	CVM_n
	100	.606	.908	.892		100	.001	.013	.117
$(.4, .1)$	250	1.00	1.00	1.00	$(.8, .1)$	250	.361	.630	.185
	500	1.00	1.00	1.00		500	.997	1.00	.392
	1000	1.00	1.00	1.00		1000	1.00	1.00	.762
	100	.563	.865	.793		100	.001	.008	.178
$(.4, .2)$	250	1.00	1.00	.997	$(.8, .2)$	250	.220	.438	.296
	500	1.00	1.00	1.00		500	.938	.996	.509
	1000	1.00	1.00	1.00		1000	1.00	1.00	.818
	100	.460	.758	.621		100	.001	.004	.263
$(.4, .3)$	250	1.00	1.00	.981	$(.8, .3)$	250	.070	.217	.566
	500	1.00	1.00	1.00		500	.521	.852	.860
	1000	1.00	1.00	1.00		1000	.985	.999	.979

Table 8. Rejection Probabilities (DGP: NLMA)

(θ_1, θ_2)	n	S_n	T_n	CVM_n	(θ_1, θ_2)	n	S_n	T_n	CVM_n
	100	.598	.914	.922		100	.005	.009	.135
$(.4, .2)$	250	1.00	1.00	1.00	$(.8, .2)$	250	.410	.671	.352
	500	1.00	1.00	1.00		500	.999	1.00	.710
	1000	1.00	1.00	1.00		1000	1.00	1.00	.986
	100	.596	.926	.912		100	.006	.010	.146
$(.4, .4)$	250	1.00	1.00	1.00	$(.8, .4)$	250	.453	.709	.349
	500	1.00	1.00	1.00		500	.995	1.00	.710
	1000	1.00	1.00	1.00		1000	1.00	1.00	.968
	100	.577	.918	.902		100	.007	.011	.168
$(.4, .6)$	250	1.00	1.00	1.00	$(.8, .6)$	250	.463	.742	.368
	500	1.00	1.00	1.00		500	.998	1.00	.697
	1000	1.00	1.00	1.00		1000	1.00	1.00	.957

Table 9. Rejection Probabilities (DGP: MARKOV)

(θ_1, θ_2)	n	S_n	T_n	CVM_n	(θ_1, θ_2)	n	S_n	T_n	CVM_n
	100	.906	.995	.991		100	.796	.984	.961
(.3, .3)	250	1.00	1.00	1.00	(.3, .7)	250	1.00	1.00	1.00
	500	1.00	1.00	1.00		500	1.00	1.00	1.00
	1000	1.00	1.00	1.00		1000	1.00	1.00	1.00
	100	.432	.790	.813		100	.386	.708	.665
(.5, .3)	250	1.00	1.00	.999	(.5, .7)	250	1.00	1.00	996
	500	1.00	1.00	1.00		500	1.00	1.00	1.00
	1000	1.00	1.00	1.00		1000	1.00	1.00	1.00
	100	.001	.001	.060		100	.001	.001	.057
(.9, .3)	250	.033	.051	.091	(.9, .7)	250	.028	.053	.089
	500	.472	.730	.199		500	.482	.732	.187
	1000	1.00	1.00	.410		1000	.999	1.00	.368

Table 10. Rejection Probabilities (DGP: FM)

(θ_1, θ_2)	n	S_n	T_n	CVM_n	(θ_1, θ_2)	n	S_n	T_n	CVM_n
	100	.151	1.00	.000		100	1.00	1.00	1.00
(2.5, .04)	250	1.00	1.00	.006	(4.0, .04)	250	1.00	1.00	1.00
	500	1.00	1.00	.998		500	1.00	1.00	1.00
	1000	1.00	1.00	1.00		1000	1.00	1.00	1.00
	100	.680	1.00	.000		100	1.00	1.00	1.00
(2.5, .05)	250	1.00	1.00	.506	(4.0, .05)	250	1.00	1.00	1.00
	500	1.00	1.00	1.00		500	1.00	1.00	1.00
	1000	1.00	1.00	1.00		1000	1.00	1.00	1.00
	100	.947	1.00	.000		100	1.00	1.00	1.00
(2.5, .06)	250	1.00	1.00	.969	(4.0, .06)	250	1.00	1.00	1.00
	500	1.00	1.00	1.00		500	1.00	1.00	1.00
	1000	1.00	1.00	1.00		1000	1.00	1.00	1.00

6. Proofs

6.1 Proof of Lemma 3.2 The stated result follows directly from Theorem 2.23 of Hall and Heyde (1980), which shows that

$$\left| \frac{1}{n} \sum_{t=1}^{n} u_t^2 - \frac{1}{n} \sum_{t=1}^{n} \mathbf{E}(u_t^2 | \mathcal{F}_{t-1}) \right| \to_p 0$$

as $n \to \infty$. ■

6.2 Proof of Lemma 3.3 The proof for the weak convergence of M_n to M consists of two parts: weak convergence of nite dimensional distribution of M_n to that of M, and stochastic equicontinuity of (M_n). To prove the rst part, we let (c_i) and (x_i)

be nite sets of numbers that are given arbitrarily, and consider the transformation

$$\sum_i c_i 1\{\cdot \le x_i\} \tag{17}$$

in $D[0,1]$. Here we endow $D[0,1]$ with the Skorohod topology [see, e.g., Billingsley (1968, pp 111-114)] instead of the uniform topology. This change of topology is necessary to effectively deal with the transformation in (17). It is indeed easy to see that the transformation is continuous with respect to the Skorohod topology, but not with respect to the uniform topology. However, the Skorohod topology is coarser than the uniform topology in $D[0,1]$, and therefore, the weak convergence (10) continues to hold under the Skorohod topology. It now follows from the continuous mapping theorem that

$$\sum_i c_i 1\{W_{nt} \le x_i\} \to_d \sum_i c_i 1\{W_t \le x_i\}$$

and therefore we have

$$\begin{aligned} \sum_i c_i M_n(x_i) &= \int_0^1 \sum_i c_i 1\{W_{nt} \le x_i\}\, dW_t \\ &\to_d \int_0^1 \sum_i c_i 1\{W_t \le x_i\}\, dW_t \\ &= \sum_i c_i M(x_i) \end{aligned}$$

due to the result in Kurtz and Protter (1991).

To establish the stochastic equicontinuity of (M_n), it is required to show that for any $\varepsilon > 0$ there exists $\delta > 0$ satisfying

$$\limsup_{n\to\infty} \mathbf{P}\left\{ \sup_{|x-y|<\delta} |M_n(x) - M_n(y)| > \varepsilon \right\} < \varepsilon \tag{18}$$

[see, e.g., Andrews (1994)]. To prove (18), we rst note that

$$M_n(x) - M_n(y) = \int_0^1 1\{x < W_{nt} \le y\}\, dW_t \tag{19}$$

For the subsequent proof, we let $\varepsilon > 0$ be given arbitrarily and set $\delta > 0$ so that

$$0 < \delta < \frac{\varepsilon^3}{6\mathbf{E}\left(\sup_{s\in\mathbf{R}} L(1,s)\right)}$$

Since $W_n \to_p W$ uniformly,

$$\mathbf{P}\left\{ \sup_{t\in[0,1]} |W_{nt} - W_t| > \delta \right\} \le \frac{\varepsilon^3}{2}$$

for all large n.

Since M_n is a rightcontinuous martingale for each $n \geq 1$, we may apply Doob s L^p-inequality [see, e.g., Revuz and Yor (1994, Theorem 1.7, p52)] to the right hand side of (19) and get

$$\begin{aligned}
&\varepsilon^2 \mathbf{P}\left\{\sup_{|x-y|<\delta}\left|\int_0^1 1\{x < W_{nt} \leq y\}\, dW_t\right| > \varepsilon\right\} \\
&\leq \sup_{|x-y|<\delta} \mathbf{E}\left(\int_0^1 1\{x < W_{nt} \leq y\}\, dW_t\right)^2 \\
&\leq \sup_{|x-y|<\delta} \mathbf{E}\left(\int_0^1 1\{x < W_{nt} \leq y\}\, dt\right) \qquad (20)
\end{aligned}$$

Moreover, since

$$\int_0^1 1\{x < W_{nt} \leq y\}\, dt \leq 1$$

we have

$$\begin{aligned}
&\mathbf{E}\left(\int_0^1 1\{x < W_{nt} \leq y\}\, dt\right) \\
&\leq \mathbf{P}\left\{\sup_{t\in[0,1]} |W_{nt} - W_t| > \delta\right\} + \mathbf{E}\left(\int_0^1 1\{x < W_{nt} \leq y\}\, dt\right) 1\left\{\sup_{t\in[0,1]} |W_{nt} - W_t| \leq \delta\right\} \\
&\leq \frac{\varepsilon^3}{2} + \mathbf{E}\left(\int_0^1 1\{x < W_{nt} \leq y\}\, dt\right) 1\left\{\sup_{t\in[0,1]} |W_{nt} - W_t| \leq \delta\right\} \qquad (21)
\end{aligned}$$

for all large n.

However, if we let

$$\sup_{t\in[0,1]} |W_{nt} - W_t| \leq \delta$$

then it follows that

$$\begin{aligned}
\sup_{|x-y|<\delta} \int_0^1 1\{x < W_{nt} \leq y\}\, dt &\leq \sup_{|x-y|<3\delta} \int_0^1 1\{x < W_t \leq y\}\, dt \\
&= \sup_{|x-y|<3\delta} |x-y| \left(\sup_{s\in\mathbf{R}} L(1,s)\right) \\
&= 3\delta \left(\sup_{s\in\mathbf{R}} L(1,s)\right)
\end{aligned}$$

Therefore, we have

$$\begin{aligned}
&\sup_{|x-y|<\delta} \mathbf{E}\left(\int_0^1 1\{x < W_{nt} \leq y\}\, dt\right) 1\left\{\sup_{t\in[0,1]} |W_{nt} - W_t| \leq \delta\right\} \\
&\leq 3\delta\, \mathbf{E}\left(\sup_{s\in\mathbf{R}} L(1,s)\right) \leq \frac{\varepsilon^3}{2} \qquad (22)
\end{aligned}$$

for all sufficiently large n. The inequality in (18) now follows immediately from (19) (22). ■

6.3 Proof of Theorem 3.4 The stated results follow directly from the continuous mapping theorem, given the weak convergence of M_n to M that is established in Lemma 3.3. ■

6.4 Proof of Lemma 3.5 To see that M is a martingale, we rst note that

$$M(y) - M(x) = \int_0^1 1\{x \le W_t \le y\}\, dW_t$$

for any $x < y$ is uncorrelated with

$$M(w) = \int_0^1 1\{W_t \le w\}\, dW_t$$

for all values of $w \le x$. However, due to the strong Markov property of Brownian motion, they are Gaussian and therefore independent. To see this, let $x > 0$ be xed and observe that we may represent $M(x)$ as

$$M(x) = W_{\tau_1} + (W_{\tau_2} - W_{\sigma_2}) + \cdots + (W_{\tau_m} - W_{\sigma_m})$$

where $(\tau_i)_{i=1}^m$ and $(\sigma_i)_{i=1}^m$ are sequences of stopping times de ned recursively, up to their existence, as

$$\begin{aligned} \tau_i &= \inf\{\sigma_i < t \le 1 | W_t > x\} \\ \sigma_{i+1} &= \inf\{\tau_i < t \le 1 | W_t \le x\} \end{aligned}$$

starting from $\sigma_1 = 0$. It is obvious that a similar representation of $M(x)$ is possible for any $x \le 0$. Consequently, if we let $(\mathcal{F}_x)$, $\mathcal{F}_x = \sigma(M(w), w \le x)$, be the natural ltration of M, then it follows that

$$\mathbf{E}(M(y)|\mathcal{F}_x) = M(x)$$

as was to be shown.

Let $x < y$, and note that

$$\begin{aligned} |M(x) - M(y)|^r &= \left| \int_0^1 1\{W_t \le x\}\, dW_t - \int_0^1 1\{W_t \le y\}\, dW_t \right|^r \\ &= \left| \int_0^1 1\{x < W_t \le y\}\, dW_t \right|^r \end{aligned}$$

We have

$$\mathbf{E} \left| \int_0^1 1\{x < W_t \le y\}\, dW_t \right|^r \le c\, \mathbf{E} \left| \int_0^1 1\{x < W_t \le y\}\, dt \right|^{r/2}$$

for some constant c, as shown in, e.g., Revuz and Yor (1994, Proposition 4.3, p154), and

$$\begin{aligned} \int_0^1 1\{x < W_t \le y\}\, dt &= \int_{-\infty}^{\infty} 1\{x < s \le y\} L(1,s)\, ds \\ &= |x - y| \sup_{s \in \mathbf{R}} L(1,s) \end{aligned}$$

Consequently, it follows that

$$\mathbf{E}|M(x)-M(y)|^r \le c\,|x-y|^{r/2}\mathbf{E}\left(\sup_{s\in\mathbf{R}} L(1,s)\right)^{r/2}$$

and we may simply let

$$c_r = c\,\mathbf{E}\left(\sup_{s\in\mathbf{R}} L(1,s)\right)^{r/2}$$

to get the stated result. ■

6.5 Proof of Proposition 3.6 The result follows from Lemma 3.5. See, for instance, Revuz and Yor (1994, Theorem 2.1, p25). ■

6.6 Proof of Theorem 4.2 Let $Q(y) = \mathbf{E}q_t(y)$, where $q_t(\cdot)$ is as de ned in (3). By the uniform convergence result of Koul and Stute (1999, equation (4.1), p.224), we have

$$\sup_{y\in\mathbf{R}} |Q_n(y)-Q(y)| \to 0 \ \text{ a.s} \tag{23}$$

since Assumption 4.1 implies that $(\triangle y_t)$ is a strictly stationary square integrable ergodic sequence. Also, by the ergodic theorem and Assumption 4.1, we have

$$\sigma_n \to_p \sigma. \tag{24}$$

Combining (23) and (24) gives

$$n^{-1/2}S_n \to_p (1/\sigma)\sup_{y\in\mathbf{R}} |Q(y)|$$

and

$$n^{-1}T_n \to_p (1/\sigma)\int Q^2(y)d\mu(y)$$

where $\mu(\cdot)$ denotes the distribution of y_t. The stated result now follows, since under Assumption 4.1 and the alternative hypothesis (2), there exists some $y \in \mathbf{R}$ for which

$$\begin{aligned} Q(y) &= \mathbf{E}(\triangle y_t 1\{y_{t-1}\le y\}) \\ &= \int_{-\infty}^{y} \mathbf{E}(\triangle y_t | y_{t-1}=x)d\mu(x) \ne 0 \end{aligned}$$

where $\mu(\cdot)$ denotes the destribution of (y_t). ■

References

An, H.-Z. and C. Bing (1991). A Kolmogorov-Smirnov type statistic with application to test for nonlinearity in time series, *International Statistical Review*, 59, 287-307.

Andrews, D. W. K. (1994). Empirical process methods in econometrics, In R.F. Engle and D. McFadden, eds., *Handbook of Econometrics*, Vol.4, Elsevier: Amsterdam.

Billingsley, P. (1968). *Convergence of Probability Measures.* John Wiley: New York.

Bierens, H.J. (1990). A consistent conditional moment test of functional form, *Econometrica*, 58, 1443-1458.

Brockett, P.L., M. Hinich, and D. Patterson (1988). Bispectral based tests for the detection of Gaussianity and linearity in time series, *Journal of the American Statistical Association*, 83, 657-664.

Chan, W.S. and H. Tong (1986). On tests for non-linearity in time series analysis, *Journal of Forecasting*, 5, 217-228.

Chung, K.L. and R.T. Williams (1990). *Introduction to Stochastic Integration.* Birkhäuser: Boston.

de Jong, R.M. (1996). The Bierens test under data dependence, *Journal of Econometrics*, 72, 1-32.

Dickey, D. A. and W. A. Fuller (1979). Distribution of estimators for autoregressive time series with a unit root, *Journal of the American Statistical Association*, 74, 427-431.

Durlauf, S.N. (1991). Spectral based testing of the martingale hypothesis, *Journal of Econometrics*, 50, 355-376.

Enders, W. and C. W. J. Granger (1998). Unit root tests and asymmetric adjustment with an example using the term structure of interest rates, *Journal of Business and Economic Statistics*, 16, 304-311.

Hall, R. E. (1978). Stochastic implications of the life cycle-permanent income hypothesis: Theory and evidence, *Journal of Polical Economy*, 86, 971-987.

Hall, P. and C.C. Heyde (1980). *Martingale Limit Theory and Its Application.* Academic Press: New York.

Hida, T. (1980). *Brownian Motion.* Springer-Verlag: New York.

Hinich, M. (1982). Testing for Gaussianity and linearity of a stationary time series, *Journal of Time Series Analysis*, 3, 169-176.

Hjellvik, V. and D. Tjøstheim (1995). Nonparametric tests of linearity for time series, *Biometrika*, 82, 351-368.

Khmaladze, E. V. (1988), An innovation approach to goodness-of- t tests in $\mathbf{R}^m$," *Annals of Statistics*, 16, 1503-1516.

Koul, H. and W. Stute (1999), Nonparametric model checks for time series, *Annals of Statistics*, 27, 204-236.

Kurtz and Protter (1991). Weak limit theorems for stochastic integrals and stochastic differential equations, *Annals of Probability*, 19, 1035-1070.

Luukkonen, R., P. Saikkonen, and T. Teräsvirta (1988). Testing linearity against smooth transition autoregression, *Biometrika*, 75, 491-499.

Phillips, P. C. B. (1997). Unit root tests. In *Encyclopedia of Statistical Sciences* Vol. 12, Wiley: New York.

Park, J.Y. and P.C.B. Phillips (1999). Asymptotics for nonlinear transformations of integrated time series, *Econometric Theory*, 15, 269-298.

Revuz, D. and M. Yor (1994). *Continuous Martingale and Brownian Motion*, 2nd ed. Springer Verlag: New York.

Stock, J.H. (1994). Unit roots and structural breaks, In R.F. Engle and D. McFadden, eds., *Handbook of Econometrics*, Vol. 4, 2739-2841, Elsevier: Amsterdam.

Tong, H. (1990). *Nonlinear Time Series: A Dynamical Systems Approach*. Clarendon Press: Oxford.

Whang, Y. -J. and O. Linton (1999). The asymptotic distribution of nonparametric estimates of the Lyapunov exponent for stochastic time series, *Journal of Econometrics*, 91, 1-42.