A Test of the Martingale Hypothesis [1]

Joon Y. Park
Department of Economics
Rice University and Sungkyunkwan University

and

Yoon-Jae Whang
Department of Economics
Korea University

Abstract

This paper proposes a statistical test of the martingale hypothesis. It can be used to test whether a given time series is a martingale process against certain non-martingale alternatives. The class of alternative processes against which our test has power is very general and it encompasses many nonlinear non-martingale processes which may not be detected using traditional spectrum-based or variance-ratio tests. We look at the hypothesis of martingale, in contrast with other existing methods which test for the hypothesis of martingale difference. Two different types of test are considered: one is a generalized Kolmogorov-Smirnov test and the other is a Cramer-von Mises type test. For the processes that are first order Markovian in mean, in particular, our approach yields the test statistics that 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 stochastic process so that the critical values are easily tabulated. We prove consistency of our tests and further investigate their finite sample properties via 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, chaotic, and some of nonstationary processes.

This version: September 3, 2004

Key words and phrases: Martingale hypothesis, Kolmogorov-Smirnov test, Cramer-von Mises test, Non-linear time series.

---

[1] The authors thank for the editor and an anonymous referee for helpful comments. They are also grateful to Don Andrews, Herman Bierens, Yoosoon Chang, In Choi, Miguel Delgado, Steven Durlauf, Yongmiao Hong, 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. The first draft of this paper was written while the authors were visiting Cowles Foundation for Research in Economics, whose hospitality is gratefully acknowledged. Park thanks the Korea Research Foundation for research support, and Whang acknowldeges support provided by the Korea University Research Grant. Address correspondence to: Yoon-Jae Whang, Department of Economics, Korea University, Seoul 136-701, Korea. Phone: +82-2-3290-2224; Fax: +82-2-928-4948; E-mail: whang@korea.ac.kr

1. Introduction

In this paper, we introduce a statistical test of 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 (in the sense of least mean squared errors) 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 arguments 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 whether a given time series is a martingale process against certain non-martingale alternative processes. The class of alternative processes against which our tests have power is very general and it encompasses, for example, many interesting nonlinear non-martingale processes including exponential and threshold autoregressive processes, markov switching and chaotic processes (possibly with stochastic noise), and some of nonstationary processes, see Tong (1990) for more examples of nonlinear time series processes.

We consider two types of tests, which can be regarded as generalizations respectively of the Kolmogorov-Smirnov test and the Cramer-von Mises test of goodness of fit to the regression framework. Though they are expected to have discriminatory powers against a wide class of non-martingale processes, our tests are very simple to implement in practical applications. In particular, if used to test for the martingale hypothesis within the class of first order Markovian processes, the proposed tests become extremely simple to use: the test statistics are easy to compute and they 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 stochastic process. Since the distributions are free of any nuisance parameters, we provide a set of critical values which can be used readily in practical applications. For the test of the martingale hypothesis in a more general context, they are still free of any nuisance parameters if the test statistics are appropriately formulated, see Section 2 below for a discussion.

Our tests are closely related, among others, to the tests by Durlauf (1991), Hong (1999), Deo (2000), Dominguez and Lobato (2000), and Kuan and Lee (2003). Their tests are, however, not directly comparable to ours, since theirs are the tests of the *martingale difference hypothesis*. Durlauf (1991) 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. However, there are nonlinear non-Gaussian processes which are non-martingales with serially uncorrelated processes [see, e.g., Brockett, Hinich and Patterson (1988, p.658) for an example]. The Durlauf tests 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. On the other hand, Hong (1999) suggests a test for the martingale hypothesis based on the so-called generalized spectral derivative. His test does have powers against non-linear non-Gaussian non-martingales but it might be sensitive to choice of smoothing parameters in practice. Furthermore, the test requires the innovation sequence to be strictly stationary, whereas our tests allow some degree of heterogeneity of the innovations. See Section 2 for a more discussion on comparison between our tests and the existing tests of the martingale difference hypothesis.

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 a survey on the subject]. The unit root hypothesis, however, is obviously more general than our martingale hypothesis. Also, the alternatives considered by 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 which might be either stationary or nonstationary. Therefore, we believe that our tests would 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).[2] Some of 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 behavior of the nonlinearity tests for nonstationary processes has not yet been investigated. Our tests are also related to the model specification tests by Bierens(1990), Bierens and Ploberger (1997) and de Jong (1996), see the next section for more discussions.

The remainder of this paper is organized as follows. Section 2 introduces the null and alternative hypotheses and defines the test statistics. In Section 3, we derive the asymptotic null distributions of the test statistics and tabulate their critical values. Section 4 considers the consistency of our tests. In particular, we establish the consistency of our tests against general non-martingales that are asymptotically stationary. The test consistency against some nonstationary non-martingales is also discussed. Section 5 reports some simulation results, and 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 filtration to which (y_t) is adapted. The null hypothesis of interest is that (y_t) is a martingale process with respect to the filtration $(\mathcal{F}_t)$, i.e.,

$$H_0 : \mathbf{P}(\mathbf{E}(y_t|\mathcal{F}_{t-1}) = y_{t-1}) = 1 \tag{1}$$

for each $t \geq 1$, where $\mathbf{E}(\cdot|\mathcal{F}_{t-1})$ denotes as usual the conditional expectation given $\mathcal{F}_{t-1}$. The alternative hypothesis is the negation of (1). To test the hypothesis (1), it is of course essential to further specify the filtration $(\mathcal{F}_t)$. For many applications, the most relevant choice of $(\mathcal{F}_t)$ appears to be the natural filtration of (y_t), in which case $\mathcal{F}_t$ for each $t \geq 1$ is defined to be the σ-field generated by (y_s) for all $s \leq t$. Different specifications of $(\mathcal{F}_t)$, which

[2]Some of these tests are used to check the departures in each moment, while ours concentrate on testing for serial dependence in mean.

may in particular include other covariates, are also possible and can be more interesting choices.

In this paper, we mainly consider the simple case in which

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

for all $t \geq 1$, see below for a discussion on its multivariate generalization. Clearly, (2) holds if (y_t) is first order Markovian. We call (y_t) the *first order Markovian-in-mean* if it satisfies (2). Note that, as shown in, e.g., Billingsley (1995, Theorem 16.10 (iii), p.213),

$$\mathbf{E}(\triangle y_t|y_{t-1}) = 0 \text{ a.s. iff } \mathbf{E}\triangle y_t 1\{y_{t-1} \leq x\} = 0 \text{ for almost all } x \in \mathbf{R}, \tag{3}$$

where and elsewhere in the paper we denote by $\triangle$ the usual difference operator (i.e., $\triangle y_t = y_t - y_{t-1}$) and by $1\{\cdot\}$ the indicator function. On the other hand, (3) implies that, when $\mathbf{P}(\mathbf{E}(\triangle y_t|y_{t-1}) = 0) < 1$, i.e., when (1) is not true, we have $\mathbf{E}\triangle y_t 1\{y_{t-1} \leq x\} \neq 0$ for some $x \in \mathbf{R}$, see Section 4 below for more details. This motivates us to consider the following as the basis of our test statistics for the martingale hypothesis (1)[3]:

$$Q_n(x) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \triangle y_t 1\{y_{t-1} \leq x\}. \tag{4}$$

Of course, our assumption (2) can be too restrictive for some applications. To deal with more general processes, we may wish to look at the case

$$\mathbf{E}(y_t|\mathcal{F}_{t-1}) = \mathbf{E}(y_t|y_{t-1}, \ldots, y_{t-\kappa}) \tag{5}$$

for all $t \geq 1$, with some $\kappa \geq 2$. Similarly as above, we may call (y_t) the *κ-th order Markovian-in-mean* if it satisfies (5). In this case, we may use the statistics based on

$$Q_n(x_1, \ldots, x_\kappa) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \triangle y_t 1\{y_{t-1} \leq x_1\} \cdots 1\{y_{t-\kappa} \leq x_\kappa\} \tag{6}$$

in place of $Q_n(x)$ introduced in (4) to more effectively discriminate our martingale null hypothesis against nonmartingale alternatives. Clearly, $Q_n(x)$ in (4) may be regarded as a special case $\kappa = 1$ of $Q_n(x_1, \ldots, x_\kappa)$ defined in (6). We may consider even more general cases where $(\mathcal{F}_t)$ includes the information from other covariates in a similar way. Moreover, it is also conceivable to increase κ in (5) and (6) as the sample size grows. All these generalizations and extensions, however, will not be pursued in this paper. They require some new development of the functional central limit theory, and will therefore be reported in our subsequent work.

---

[3]In the specification testing literature, Stinchcombe and White (1988, p.299) call the class of indicator functions $(1\{x \leq t\}, t \in \mathbf{R})$ as the *totally revealing set*. Examples of specification tests that are based on this class of functions include An and Bing (1991), Delgado (1993), Andrews (1997), Stute (1997) and Whang (2000). On the other hand, other choices of function classes are possible (e.g., the exponential functions used by Bierens (1990), but the indicator function has an advantage that it does not require an arbitrary choice of a nuisance parameter space.

We assume that the time series $(\triangle y_t)$ are martingale differences with non-vanishing variances, as will be more formally introduced in the next section. Roughly, this implies that (y_t) becomes a nonstationary integrated process. The formulation in (6) clearly distinguishes the tests of the *martingale hypothesis* from those of the *martingale difference hypothesis.* We have to deal with the levels for the former, while we may rely only on the first differences for the latter.[4] There is a large class of models used in economic and financial applications that specify the mean changes as functions of the lagged levels rather than the lagged differences, including, for instance, threshold autoregressive models, (both linear and nonlinear) error correction models and various diffusion models. As will be seen clearly in later sections, our tests for the martingale hypothesis yield asymptotics that are very different from those for the existing tests of the martingale difference hypothesis. This is mainly due to the presence of the lagged level in our test statistics.

We may construct two different types of statistics from $Q_n(x)$ defined in (4). A Kolmogorov-Smirnov type statistic is given by

$$S_n = \sup_{x \in \mathbf{R}} |Q_n(x)|. \tag{7}$$

Moreover, a Cramer-von Mises type statistic is defined as

$$T_n = \int Q_n^2(x)\, \mu_n(dx),$$

where μ_n denotes some measure. In this paper, we define μ_n to be the empirical distribution of (y_{t-1}), in which case T_n reduces to

$$T_n = \frac{1}{n} \sum_{t=1}^{n} Q_n^2(y_{t-1}). \tag{8}$$

See, e.g., Shorack and Wellner (1986) for other choices of μ_n.

The martingale hypothesis is intimately related to the unit root hypothesis, though strictly speaking none of them generally implies the other.[5] 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, which may be reformulated as

---

[4] After the first draft of our paper was written, Dominguez and Lobato (2000) proposed a test of the hypothesis $\mathbf{E}(\triangle y_t | \triangle y_{t-1}, \ldots, \triangle y_{t-p}) = 0$ a.s. against its negation. The test also uses the indicator function as the weight function, similary to ours. Their test can indeed be viewed as the test of the martingale difference hypothesis, corresponding to our tests of the martingale hypothesis.

[5] 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.

$\mathbf{E}(\triangle y_t|\mathcal{F}_{t-1}) \neq 0$. 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 look into many other nonlinear directions as well for the violation of the martingale hypothesis. Our generalization in (5), of course, can be similarly compared with the augmented Dickey-Fuller test.

If the martingale difference hypothesis, not the martingale hypothesis, is what we want to test, it can also be done using the approaches taken by Bierens (1990) and de Jong (1996) in a broader context of general model specifications. In particular, the test proposed by de Jong (1996) can be used to test the martingale difference hypothesis

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

with the natural filtration, if it is applied to the first differences $(\triangle y_t)$. However, de Jong (1996)'s test appears not to be very attractive in our context. His test is a bit too general, and hence has very low power in small samples as is indicated by the author. Furthermore, his test is computationally very demanding since it depends on high dimensional integration and cumbersome Monte Carlo simulations.

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 define $(\mathcal{F}_t)$ to be the filtration introduced earlier. Throughout this section, we suppose that (y_t) is first order Markovian-in-mean. The condition in (2) therefore holds. We assume

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

(a) $\displaystyle\frac{1}{n}\sum_{t=1}^{n} \mathbf{E}\left(u_t^2 \middle| \mathcal{F}_{t-1}\right) \to_p \sigma^2 > 0$, and

(b) $\displaystyle\sup_{t \geq 1} \mathbf{E}\left(u_t^4 \middle| \mathcal{F}_{t-1}\right) < K$ a.s. for some constant $K < \infty$.

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. It is satisfied for instance by the martingales driven by ARCH-type innovations. The part (b) of Assumption 3.1 requires that the fourth conditional moment is uniformly bounded.[6] The condition implies in particular the conditional version of Linder-

[6] Strictly speaking, Assumption 3.1(b) rules out the standard GARCH(1,1) process. However, this assumption is standard in the nonstationary time series literature (see, e.g. Stock (1994)) and we believe that this assumption is not entirely necessary for our asymptotic results to hold. Furthermore, our simulation experiments in Section 5 show that our tests have good size and power performance in the presence of GARCH errors.

berg condition, i.e.,

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

as $n\to\infty$ for any $\varepsilon>0$, which is routinely imposed to obtain the martingale limit theory.

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 $n\to\infty$.

In what follows, we assume that $\sigma^2=1$ and (u_t) is normalized so that $\sigma_n^2=1$. Of course, the normalization can be done by dividing (y_t) by σ_n.[7] This is to ease the exposition of our theory. Given the consistency of σ_n^2 in Lemma 3.2, the convention imposes no restriction on our subsequent theory.

Let

$$W_n(r)=\frac{1}{\sqrt{n}}\sum_{t=1}^{[nr]}u_t \tag{9}$$

for $0\le r\le 1$, where $[z]$ is the largest integer which does not exceed z. Under Assumption 3.1, invariance principle holds, as shown by, e.g., Hall and Heyde (1980, Theorem 4.1, p. 99). More precisely, we have the weak convergence

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

in $D[0,1]$, the space of cadlag functions on $[0,1]$, endowed with the Skorohod topology, where W is the standard Brownian motion.

Now we define

$$M_n(x)=Q_n(x\sqrt{n}). \tag{11}$$

We may write S_n as

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

Moreover, we have

$$T_n=\int_0^1 M_n^2(W_n(r))\,dr, \tag{13}$$

where W_n is the process introduced in (9).

From now on, we regard M_n defined 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

[7] The normalized sequences should be more precisely denoted by (y_{nt}) and (u_{nt}), since they depend upon n. We will, however, continue to use (y_t) and (u_t) to simplify the notation

endow $D(\mathbf{R})$ also with the Skorohod topology. We now write the process M_n introduced in (11) as

$$\begin{aligned} M_n(x) &= \frac{1}{\sqrt{n}} \sum_{t=1}^{n} u_t \, 1\left\{ \frac{y_{t-1}}{\sqrt{n}} \le x \right\} \\ &= \int_0^1 1\{W_n(r) \le x\} \, dW_n(r) \end{aligned}$$

for $x \in \mathbf{R}$. We may extend the definition of M_n to $\pm\infty$ by putting

$$M_n(-\infty) = 0 \quad \text{and} \quad M_n(\infty) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} u_t = W_n(1).$$

Then, M_n becomes a process taking values in $D[-\infty, \infty]$ which, up to a strictly increasing continuous transformation, is the same as $D[0,1]$.

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 as $n \to \infty$ in $D[-\infty, \infty]$ to M defined by

$$M(x) = \int_0^1 1\{W(r) \le x\} \, dW(r) \tag{14}$$

for $x \in \mathbf{R}$ with $M(-\infty) = 0$ and $M(\infty) = W(1)$. The weak convergence is presented in the following lemma.

3.3 Lemma Under Assumption 3.1, we have $M_n \to_d M$ in $D[-\infty, \infty]$ 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

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

as $n \to \infty$.

The proof of the above theorems and some of our subsequent results involves 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 defined as

$$L(t,s) = \lim_{\varepsilon \to 0} \frac{1}{2\varepsilon} \int_0^t 1\{|W(r) - s| \le \varepsilon\} \, dr \tag{15}$$

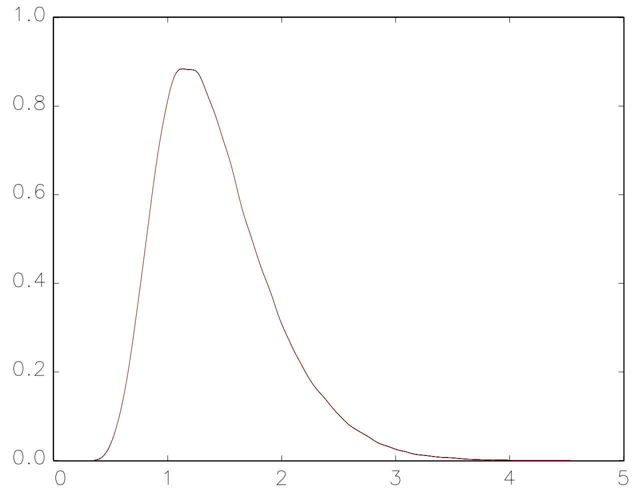

Figure 1: Probability Density of S

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

$$\int_0^t F(W(r))\,dr = \int_{-\infty}^{\infty} F(s)L(t,s)\,ds$$

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 We have for any $p \geq 2$ and $x, y \in \mathbf{R}$,

$$\mathbf{E}|M(x) - M(y)|^p \leq c_p|x - y|^{p/2},$$

where c_p is a constant depending only upon p.

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

We may therefore assume that M is a continuous stochastic process.

Since the process M is continuous and effectively stopped at

$$s_{\min} = \inf_{r\in[0,1]} W(r) \quad \text{and} \quad s_{\max} = \sup_{r\in[0,1]} W(r),$$

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.

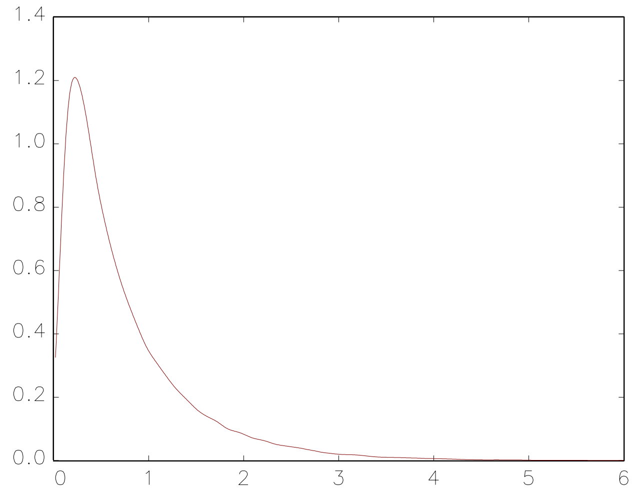

Figure 2: Probability Density of T

Moreover, the process M is a.s. of locally integrable sample path, and therefore we have

$$\int_0^1 M^2(W(r))\,dr = \int_{-\infty}^{\infty} M^2(s)L(1,s)\,ds$$

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

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 defined in (7) and (8) respectively, are free of any nuisance parameters. They can readily be obtained through simulations and their probability densities are sketched 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.

As we have noted above, our tests have asymptotic null distributions that are distribution-free and do not require any resampling procedure to simulate the critical values. Therefore, they are extremely simple to use in practical applications. This is in sharp contrast with other existing tests for the martingale difference hypothesis, whose critical values are heavily dependent upon the underlying distribution and have to be estimated by bootstrap or by any other methods that may substitute bootstrap. The distributional results for our tests are, of course, not directly comparable to those for the existing martingale difference tests. The former include the lagged level that is nonstationary, while the latter only consider the lagged differences that are assumed to be stationary. The distribution-free nature of our tests is not particularly due to the first order Markovian-in-mean structure of the model that we consider in the paper. They continue to be independent of the underlying distribution if we consider the statistic (6) for the general κ-th order Markovian-in-mean model (5). The details will be reported in our subsequent work.

4. Consistency of the Tests

In this section, we establish consistency of our tests based on the statistics S_n and T_n against certain non-martingale alternatives.

Suppose, for now, that (y_t) is strictly stationary. By definition, (y_t) is in the alternative hypothesis if it satisfies

$$\mathbf{P}(\mathbf{E}(\triangle y_t | y_{t-1}) \neq 0) > 0. \tag{16}$$

Note that (16) is equivalent to (17):

$$\begin{aligned} \mathbf{E}\triangle y_t 1\{y_{t-1} \leq x\} &= \int \mathbf{E}(\triangle y_t | y_{t-1} = z) 1\{z \leq x\} \, d\mathrm{P}(z) \\ &\neq 0 \text{ for some } x \in \mathbf{R}, \end{aligned} \tag{17}$$

where P denotes the time invariant stationary distribution of (y_t). Therefore, we can see that the tests based on the sample analogue of (17) might be consistent againt general alternatives satisfying (16). This is shown in Theorem 4.4 below.

We now relax the assumption of strict stationarity. To allow for some degree of heterogeneity of alternative processes, we write explicitly the random variables (y_t) to be triangular arrays, i.e., (y_{nt}) for $n \geq 1$ and $1 \leq t \leq n$. By definition, (y_{nt}) is in the alternative hypothesis if it satisfies:

4.1 Assumption Assume that we have, for all $z \in \mathbf{R}$, $(1/n)\sum_{t=1}^n \mathbf{E}(\triangle y_{nt} | y_{n,t-1} = z) \to H(z)$ as $n \to \infty$, where H is a measurable function on $\mathbf{R}$, and that we have, for any Borel set $A \subset \mathbf{R}$, $(1/n)\sum_{t=1}^n \mathrm{P}_{nt}(A) \to \mathrm{P}(A)$ as $n \to \infty$, where P is a probability measure on $\mathbf{R}$ and P_{nt} are the distributions of (y_{nt}) for $1 \leq t \leq n$, $n \geq 1$. Furthermore, we let $\int 1\{H(z) \neq 0\} \, d\mathrm{P}(z) > 0$.

Clearly, Assumption 4.1 includes (16) as a special case and is easy to check in practice. For example, suppose (y_t) is a stationary AR(1) process, i.e., $y_t = \alpha y_{t-1} + \varepsilon_t$, where $|\alpha| < 1$ and (ε_t) are i.i.d. $(0, \sigma^2)$. Then we have $H(z) = \mathbf{E}(\triangle y_t | y_{t-1} = z) = (\alpha - 1)z$ and P becomes the time invariant stationary distribution of (y_t). In this case, Assumption 4.1 holds unless P is degenerate and puts mass 1 at the origin (in which case we have $y_t = 0$ a.s. for all t). Similarly, suppose (y_t) is a deterministically trending process, i.e., $y_t = \alpha_0 + \alpha_1(t/n) + \varepsilon_t$, where (ε_t) are i.i.d. $U[-1, 1]$. Then, $H(z) = \lim_{n\to\infty}(1/n)\sum_{t=1}^n \mathbf{E}(\triangle y_t | y_{t-1} = z) = -z + \alpha_0 + \alpha_1/2$ and P is given by the uniform distribution $U[\alpha_0 - (1-\alpha_1)/2,\ 1 + \alpha_0 - (1-\alpha_1)/2]$. In this case also, Assumption 4.1 holds with $\int 1\{H(z) \neq 0\} \, d\mathrm{P}(z) = 1$

We further assume that the triangular array of random variables (y_{nt}) is weakly dependent and satisfies a moment condition.

4.2 Assumption Assume that (y_{nt}) is a strong mixing triangular array that satisfies $\sup_{n\geq 1, 1\leq t\leq n} \mathbf{E}|\triangle y_{nt}|^p < \infty$ for some $p \in [1,\infty]$.

This assumption might be relaxed along the lines discussed below, if needed.

For our consistency result, we need the following uniform Weak Law of Large Numbers (WLLN):

4.3 Lemma Under Assumption 4.2, we have

$$\sup_{x\in\mathbf{R}} \left| \frac{1}{n} \sum_{t=1}^{n} [\triangle y_{nt} 1\{y_{n,t-1} \leq x\} - \mathbf{E}\triangle y_{nt} 1\{y_{n,t-1} \leq x\}] \right| \to_p 0 \tag{18}$$

as $n \to \infty$.

Consistency of our tests is established in the following theorem:

4.4 Theorem Suppose that Assumptions 4.1 and 4.2 hold with $p \geq 2$. Then, we have

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

as $n \to \infty$.

Theorem 4.4 shows that the tests S_n and T_n are consistent if we reject the null hypothesis when they take large values.

4.5 Remarks (a) The strong mixing assumption and L^p-boundedness condition in Assumption 4.2 were assumed to use the pointwise WLLN result of Andrews (1988, example 4, p.462), see proof of Theorem 4.4 and Lemma 4.3 below. They can be relaxed if needed. For example, to allow for trending random variables, one can use the result of de Jong (1995, Theorems 1 or 3) to verify the pointwise WLLN which requires the triangular array of random variables $(\triangle y_{nt} 1\{y_{n,t-1} \leq x\})$ and $(\triangle y_{nt}^2)$ minus their respective means are L_q-mixingale or L_q-near epoch dependent on some strong mixing sequence with $1 \leq q \leq 2$ and satisfy other additional moment conditions in the Theorems. In this case, the bracketing condition (33) in the proof of Lemma 4.3 can be verified under the assumption $\limsup_{n\to\infty} (1/n) \sum_{t=1}^{n} \mathbf{E}|\triangle y_{nt}|^p < \infty$ for some $p > 1$.

(b) Lemma 4.3 gives a uniform WLLN for unbounded and non-differentiable functions of weakly dependent and non-identically distributed random variables. To the best of our knowledge, such result is not yet available in the literature and hence would be of separate interest. This lemma also differs from the uniform WLLN of Koul and Stute (1999, equation (4.1)) who assume stationarity of the random variables whereas we allow for heterogeneous random variables.

The alternatives we consider in Assumptions 4.1 and 4.2 are processes that are essentially stationary. Though we allow for quite flexible forms of nonstationarity there, it is required that the nonstationarity be vanished asymptotically. Unfortunately, it does not seem possible to obtain any general theoretical results for the powers of our tests against the non-martingale processes with non-vanishing nonstationarity. Our tests may or may not have powers against such non-martingales that are intrinsically nonstationary. Among the processes we consider in our simulations reported in the next section, S_n appears to have desirable power against the explosive process that is intrinsically nonstationary. On the other hand, both S_n and T_n fail to have effective powers against many non-martingale unit root processes. For the intrinsically nonstationary models, (16) does not warrant the consistency of our tests, even if it holds for all $t \geq 1$. If (y_t) is nonstationary even asymptotically, our tests may become inconsistent against the non-martingale alternatives. We may indeed show that the basis of our tests Q_n, introduced in (4), does not diverge under many unit root non-martingale alternatives.[8]

To see this, we first consider the simple random walk (y_t) given by $\triangle y_t = u_t$, where (u_t) is an i.i.d. innovation sequence with mean zero and unit variance. As shown in Chang and Park (2004), we have for this process

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} u_t 1\{y_t \leq 0\} \to_d M(0) + KL(1,0), \tag{19}$$

where M is the process defined in (14), $K > 0$ is some constant and L is the local time given in (15). We may compare the result in (19) with Lemma 3.3 to understand the effects of the presence of dependency in the innovation u_t and the argument y_t in the indicator function. The dependency does not change the rate of convergence. However, it alters the limit distribution, and in particular, it shifts the limit distribution to the right by $KL(1,0)$. Note that $L(1,0) > 0$ a.s.

We now look at the non-martingale unit root process (y_t) generated by $\triangle y_t = u_t$ with (u_t) that is serially correlated. The result in (19) for the simple random walk gives us an obvious clue on how our tests would behave for this class of nonmartingales. Note that u_t is correlated with y_{t-1} when (u_t) are serially correated, so in this case we may expect that our tests have the asymptotics similar to (19). Therefore, it is clear that our tests are generally inconsistent for the unit root nonmartingales driven by serially correlated innovations. Yet, we may predict that our tests would have some nontrivial powers against such non-martingales, since the presence of serial correlation in (u_t) would shift the limit distributions of our tests. The appearance of the additional term involving $L(1,0)$ in (19) is due to the nonzero correlation of u_t and y_t, and we may see that a similar term will appear in our case here. In fact, this is exactly what we observe in our simulation study.

[8]For the unit root non-martingales, the conditional expectation $\mathbf{E}(\triangle y_t|\mathcal{F}_{t-1})$ is generally given as a function of the lagged differences $\triangle y_{t-1}, \triangle y_{t-2}, \ldots$, and in particular, our maintained assumption (2) does not hold. Therefore, strictly speaking, they are not allowed in our framework. To test the martingale hypothesis against such alternatives, it seems preferred to use any of the existing martingale difference tests.

5. Simulation Results

In this section, we examine the finite sample performance of our tests in a small scale simulation experiment. We choose ten different models as 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	$y_t = y_{t-1} + u_t;\ u_t = \sigma_t\varepsilon_t,\ \sigma_t^2 = 1 + \theta_1 u_{t-1}^2 + \theta_2\sigma_{t-1}^2$
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	$y_t = \theta_1 y_{t-1}1\{\lvert y_{t-1}\rvert < \theta_2\} + 0.9y_{t-1}1\{\lvert y_{t-1}\rvert \geq \theta_2\} + \varepsilon_t$
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 $1,\ \mu_0 = 0,\ \mu_1 = 1.$
	$\theta_2 = P(s_t = 0\lvert s_{t-1} = 0) = P(s_t = 1\lvert s_{t-1} = 1)$
FM	$y_t = m_t + u_t;\ m_t = \theta_1 y_{t-1}(1 - y_{t-1}),\ u_t = \theta_2 v_t\eta_t,$
	$v_t = \min\{m_t,\ 1 - m_t\},\ \eta_t \sim i.i.d.$ Uniform$(0,1)$
EXP	$y_t = \theta_1 y_{t-1} + u_t;\ \lvert\theta_1\rvert > 1,\ u_t = \sigma_t\varepsilon_t,\ \sigma_t^2 = 1 + \theta_2 u_{t-1}^2 + \theta_3\sigma_{t-1}^2$
UNIT	$y_t = y_{t-1} + u_t;\ u_t = \theta_1 u_{t-1} + \varepsilon_t$
TREND	$y_t = \theta_1 + \theta_2(t/n) + y_{t-1} + \varepsilon_t$

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).[9] 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 "fixed 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. 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. Model EXP is an explosive AR(1) model and Model UNIT is a unit root process with an AR(1) innovation sequence. Finally, Model TREND is a random walk model with a deterministic trend.

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

[9] 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.

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 3-13 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.[10]

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.[11]

Tables 4-13 report the finite 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 especially when the sample size is small. When the autoregressive coefficient 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 comparison, 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 size increases. When there is a moving average component, i.e., $\theta_2 \neq 0$, the performances of all three tests become slightly worse but, nevertheless, 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 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

[10] 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.

[11] This result is not surprising because it is now well known that CVM_n is not robust to volatility clustering, see Deo (2000) for this point.

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 switching 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 finite 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.

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 11 presents the power performance against an explosive AR(1) process. Although the latter process violates our Assumption 4.2, both S_n and T_n are consistent against the midly explosive alternatives (i.e., $\theta_2 = 1.01$) and more powerful than CVM_n. However, when the process becomes more explosive (i.e., $\theta_2 = 1.05$), T_n does not appear to be consistent.

Table 12 shows that our tests are not consistent against a unit root process with AR(1) disturbances, except the case when the AR coefficient $\theta_1 = 1$. This is expected because, even if this process satisfies $\mathbf{E}(\triangle y_t | y_{t-1}) \neq 0$ with positive probability, the correlation of $u_t = \triangle y_t$ and y_{t-1} (which is nonstationary) merely shifts the limiting distributions of our test statistics, see Section 4 for details. However, our tests do have some nontrivial powers against this alternative process and their powers tend to increase as we have more persistency in the innovation sequence, i.e. as θ_1 gets larger.

Finally, Table 13 shows that both S_n and T_n have satisfactory power performance against a nonstationary process with a deterministic trend. As expected, CVM_n does not have any distinguishing power against such alternative.

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

Table 11. Rejection Probabilities (DGP: EXP)

$(\theta_1,\theta_2,\theta_3)$	n	S_n	T_n	CVM_n	$(\theta_1,\theta_2,\theta_3)$	n	S_n	T_n	CVM_n
	100	.237	.216	.035		100	.955	.928	.931
$(1.01,0,0)$	250	.661	.614	.141	$(1.05,0,0)$	250	1.00	501	1.00
	500	.962	.941	.910		500	1.00	.492	1.00
	1000	1.00	1.00	1.00		1000	1.00	.523	1.00
	100	.222	.196	.311		100	.948	.913	.928
$(1.01,.9,0)$	250	.626	.562	.483	$(1.05,.9,0)$	250	1.00	502	1.00
	500	.948	.933	.946		500	1.00	.487	1.00
	1000	1.00	1.00	.999		1000	1.00	.505	1.00
	100	.232	.207	.119		100	.937	.914	.921
$(1.01,.3,.4)$	250	.648	.597	.221	$(1.05,.3,.4)$	250	1.00	498	1.00
	500	.957	.938	.908		500	1.00	.480	1.00
	1000	1.00	.999	1.00		1000	1.00	.515	1.00

Table 12. Rejection Probabilities (DGP: UNIT)

θ_1	n	S_n	T_n	CVM_n	θ_1	n	S_n	T_n	CVM_n
	100	.062	.064	.112		100	.367	.343	1.00
0.1	250	.059	.060	.303	0.7	250	.347	.302	1.00
	500	.056	.060	.560		500	.360	.312	1.00
	1000	.061	.059	.874		1000	.338	.302	1.00
	100	.119	.113	.774		100	.647	.613	1.00
0.3	250	.113	.102	.998	0.9	250	.612	.572	1.00
	500	.110	.099	1.00		500	.601	.568	1.00
	1000	.109	.099	1.00		1000	.596	.560	1.00
	100	.219	.206	.996		100	.900	.870	1.00
0.5	250	.209	.185	1.00	1.0	250	.941	.934	1.00
	500	.207	.185	1.00		500	.958	.943	1.00
	1000	.194	.177	1.00		1000	.972	.960	1.00

Table 13. Rejection Probabilities (DGP: TREND)

(θ_1, θ_2)	n	S_n	T_n	CVM_n	(θ_1, θ_2)	n	S_n	T_n	CVM_n
	100	.083	.101	.037		100	.158	.164	.037
(.01, .1)	250	.150	.131	.036	(.05, .1)	250	.325	.287	036
	500	.249	.204	.031		500	.583	.498	.031
	1000	.447	.320	.036		1000	.858	.751	.036
	100	.283	.229	.040		100	.443	.339	.040
(.01, .3)	250	.635	.444	.039	(.05, .3)	250	.833	.668	039
	500	.914	.727	.039		500	.987	.916	.039
	1000	.998	.957	.049		1000	1.00	.997	.049
	100	.638	.419	.041		100	.754	.577	.041
(.01, .5)	250	.961	.812	.048	(.05, .5)	250	.989	.922	048
	500	1.00	.985	.066		500	1.00	.997	.066
	1000	1.00	1.00	.099		1000	1.00	1.00	.099

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 finite dimensional distribution of M_n to that of M, and tightness of (M_n). To prove the first part, we let (c_i) and (x_i) be finite sets of numbers that are given arbitrarily, and consider the transformation Π defined by

$$\Pi(f)(r) = \sum_i c_i 1\{f(r) \leq x_i\}$$

on $D[0,1]$. It is straightforward to see that the transformation Π is continuous on $C[0,1] \subset D[0,1]$ a.s. Note that the Skorohod metric coincides with the uniform norm if restricted to the set of continuous functions $C[0,1]$ defined on $[0,1]$. It now follows from the continuous mapping theorem that

$$\begin{aligned} \Pi(W_n) &= \sum_i c_i 1\{W_n(\cdot) \leq x_i\} \\ &\to_d \Pi(W) = \sum_i c_i 1\{W(\cdot) \leq x_i\} \end{aligned} \tag{20}$$

in $D[0,1]$, and therefore, we have

$$\begin{aligned} \sum_i c_i M_n(x_i) &= \int_0^1 \sum_i c_i 1\{W_n(r) \leq x_i\}\, dW_n(r) \\ &\to_d \int_0^1 \sum_i c_i 1\{W(r) \leq x_i\}\, dW(r) \\ &= \sum_i c_i M(x_i) \end{aligned} \tag{21}$$

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

To establish the tightness, we show that Chentsov criterion [see, e.g., Billingsley (1968, Theorem 15.6)] holds. Fix $-\infty \leq x < y \leq \infty$ and let w be an arbitrary number between x and y. We consider

$$\begin{aligned} &\Big(M_n(x) - M_n(w)\Big)^2 \Big(M_n(w) - M_n(y)\Big)^2 \\ &= \frac{1}{n^2} \left(\sum_{t=1}^n u_t 1\left\{x < \frac{y_{t-1}}{\sqrt{n}} \leq w\right\} \right)^2 \left(\sum_{t=1}^n u_t 1\left\{w < \frac{y_{t-1}}{\sqrt{n}} \leq y\right\} \right)^2 \\ &= \frac{1}{n^2} \sum_{i,j,k,\ell} u_i u_j u_k u_\ell \, 1\left\{x < \frac{y_{i-1}}{\sqrt{n}}, \frac{y_{j-1}}{\sqrt{n}} < w\right\} 1\left\{w < \frac{y_{k-1}}{\sqrt{n}}, \frac{y_{\ell-1}}{\sqrt{n}} < y\right\} \end{aligned}$$

It can be easily deduced that

$$\begin{aligned} &\mathbf{E}\Big(M_n(x) - M_n(w)\Big)^2 \Big(M_n(w) - M_n(y)\Big)^2 \\ &= \frac{1}{n^2} \sum_{i,j<k} \mathbf{E}\, u_i u_j u_k^2 \, 1\left\{x < \frac{y_{i-1}}{\sqrt{n}}, \frac{y_{j-1}}{\sqrt{n}} < w\right\} 1\left\{w < \frac{y_{k-1}}{\sqrt{n}} < y\right\} \\ &\quad + \frac{1}{n^2} \sum_{i,j<k} \mathbf{E}\, u_i u_j u_k^2 \, 1\left\{x < \frac{y_{k-1}}{\sqrt{n}} < w\right\} 1\left\{w < \frac{y_{i-1}}{\sqrt{n}}, \frac{y_{j-1}}{\sqrt{n}} < y\right\} \end{aligned} \tag{22}$$

using the fact that $(u_t, \mathcal{F}_t)$ is a martingale difference sequence and (y_t) is adapted to $(\mathcal{F}_t)$. We will only consider the first term in (22). The treatment of the second term is entirely analogous. For the first term in (22), we have

$$\begin{aligned}
&\frac{1}{n^2}\sum_{i,j<k}\mathbf{E}\,u_iu_ju_k^2\,1\left\{x<\frac{y_{i-1}}{\sqrt{n}},\frac{y_{j-1}}{\sqrt{n}}<w\right\}1\left\{w<\frac{y_{k-1}}{\sqrt{n}}<y\right\}\\
&=\frac{1}{n^2}\sum_{j=1}^{n}\mathbf{E}\left[\left(\sum_{i=1}^{j-1}u_i1\left\{x<\frac{y_{i-1}}{\sqrt{n}}\le w\right\}\right)^2u_j^21\left\{w<\frac{y_{j-1}}{\sqrt{n}}<y\right\}\right]\\
&=\frac{1}{n^2}\sum_{j=1}^{n}\mathbf{E}\left[\left(\sum_{i=1}^{j-1}u_i1\left\{x<\frac{y_{i-1}}{\sqrt{n}}\le w\right\}\right)^2\mathbf{E}(u_j^2|\mathcal{F}_{j-1})1\left\{w<\frac{y_{j-1}}{\sqrt{n}}<y\right\}\right]\\
&\le\frac{1}{n^2}\left[\mathbf{E}\sum_{j=1}^{n}\left(\sum_{i=1}^{j-1}u_i1\left\{x<\frac{y_{i-1}}{\sqrt{n}}\le w\right\}\right)^4\right]^{1/2}\left[\mathbf{E}\sum_{t=1}^{n}\left(\mathbf{E}(u_t^2|\mathcal{F}_{t-1})\right)^21\left\{w<\frac{y_{t-1}}{\sqrt{n}}\le y\right\}\right]^{1/2}.
\end{aligned}\tag{23}$$

The inequality in the last line, in particular, follows from Cauchy-Schwarz inequality.

We now consider two terms appearing in (23). To analyze the first term, we may apply a maximal inequality for martingale [see, e.g., Revuz and Yor (1994, Corollary 1.6, pp 50-51)] to get

$$\mathbf{E}\left[\max_{1\le j\le n}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{j}u_i1\left\{x<\frac{y_{i-1}}{\sqrt{n}}\le w\right\}\right)^4\right]\le(4/3)^4\mathbf{E}\left(\frac{1}{\sqrt{n}}\sum_{t=1}^{n}u_t1\left\{x<\frac{y_{t-1}}{\sqrt{n}}\le w\right\}\right)^4.\tag{24}$$

Moreover, it can be deduced from Rosenthal's inequality [see, e.g., Hall and Heyde (1980, Theorem 2.12, pp 23-24)],

$$\begin{aligned}
&\mathbf{E}\left(\frac{1}{\sqrt{n}}\sum_{t=1}^{n}u_t1\left\{x<\frac{y_{t-1}}{\sqrt{n}}\le w\right\}\right)^4\\
&\quad\le\;K\left[\mathbf{E}\left(\frac{1}{n}\sum_{t=1}^{n}\mathbf{E}(u_t^2|\mathcal{F}_{t-1})1\left\{x<\frac{y_{t-1}}{\sqrt{n}}\le w\right\}\right)^2+\frac{1}{n^2}\sum_{t=1}^{n}\mathbf{E}u_t^4\right]
\end{aligned}\tag{25}$$

for some absolute constant K. Under the condition given in Assumption 3.1(b), the second term in (25) is of order $O_p(n^{-1})$ uniformly in w. Therefore, it will be ignored in our subsequent derivation. We also have under Assumption 3.1(b)

$$\sup_{t\ge1}\left(\mathbf{E}\left(u_t^2\middle|\mathcal{F}_{t-1}\right)\right)^2\le\sup_{t\ge1}\mathbf{E}\left(u_t^4\middle|\mathcal{F}_{t-1}\right)<K\quad\text{a.s.}\tag{26}$$

due to the conditional Jensen's inequality. Consequently,

$$\sup_{t\ge1}\mathbf{E}\left(u_t^2\middle|\mathcal{F}_{t-1}\right)<K^{1/2}\quad\text{a.s.}$$

Therefore, it follows from (24) and (25) that

$$\mathbf{E}\left[\frac{1}{n}\sum_{j=1}^{n}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{j-1}u_i 1\left\{x<\frac{y_{i-1}}{\sqrt{n}}\leq w\right\}\right)^4\right]\leq K\left[\mathbf{E}\left(\frac{1}{n}\sum_{t=1}^{n}1\left\{x<\frac{y_{t-1}}{\sqrt{n}}\leq w\right\}\right)^2\right] \quad (27)$$

for some constant K. To deal with the second term in (23), we use (26) to deduce that

$$\mathbf{E}\left[\frac{1}{n}\sum_{t=1}^{n}(\mathbf{E}(u_t^2|\mathcal{F}_{t-1}))^2 1\left\{w<\frac{y_{t-1}}{\sqrt{n}}\leq y\right\}\right]\leq K\left[\mathbf{E}\left(\frac{1}{n}\sum_{t=1}^{n}1\left\{w<\frac{y_{t-1}}{\sqrt{n}}\leq y\right\}\right)\right] \quad (28)$$

for some constant K.

Let $k = 1, 2$. For any fixed x and y, we have

$$\begin{aligned}\left(\frac{1}{n}\sum_{t=1}^{n}1\left\{x<\frac{y_{t-1}}{\sqrt{n}}\leq y\right\}\right)^k &= \left(\int_0^1 1\{x<W_n(r)\leq y\}\,dr\right)^k \\ &\to_d \left(\int_0^1 1\{x<W(r)\leq y\}\,dr\right)^k\end{aligned}$$

which holds as a special case of (20). However, since

$$\left(\frac{1}{n}\sum_{t=1}^{n}1\left\{x<\frac{y_{t-1}}{\sqrt{n}}\leq y\right\}\right)^k\leq 1$$

and bounded, we have

$$\mathbf{E}\left(\frac{1}{n}\sum_{t=1}^{n}1\left\{x<\frac{y_{t-1}}{\sqrt{n}}\leq y\right\}\right)^k\to\mathbf{E}\left(\int_0^1 1\{x<W(r)\leq y\}\,dr\right)^k \quad (29)$$

as $n \to \infty$.

By the occupation times formula, we have

$$\int_0^1 1\{x<W(r)\leq y\}\,dr=\int_{-\infty}^{\infty}1\{x<s\leq y\}L(1,s)\,ds$$

where L is the local time of the standard Brownian motion. Therefore, it follows that

$$\begin{aligned}\mathbf{E}\left(\int_0^1 1\{x<W(r)\leq y\}\,dr\right)^k &= \mathbf{E}\left(\int_x^y L(1,s)\,ds\right)^k \\ &= (y-x)^k\mathbf{E}\left(\frac{1}{y-x}\int_x^y L(1,s)\,ds\right)^k \\ &\leq (y-x)^k\mathbf{E}\left(\frac{1}{y-x}\int_x^y L(1,s)^k ds\right) \\ &\leq (y-x)^k\sup_{s\in\mathbf{R}}\mathbf{E}L(1,s)^k, \end{aligned} \quad (30)$$

where the last inequality is due to Fubini's theorem.

Let a and b be constants such that

$$\mathbf{P}\left\{a \leq \min_{0\leq r\leq 1} W(r), \max_{0\leq r\leq 1} W(r) \leq b\right\} > 1-\epsilon$$

for $\epsilon > 0$ arbitrarily small. For $x, y \in [a, b]$, we now have from (22), (23), (27), (28), (29) and (30) that

$$\begin{aligned}\mathbf{E}\Big(M_n(x) - M_n(w)\Big)^2\Big(M_n(w) - M_n(y)\Big)^2 &\leq K(w-x)(y-w)^{1/2} \\ &\leq K(y-x)^{3/2} \qquad (31)\end{aligned}$$

for some constant K. This establishes Chenstov condition for tightness. The tightness result in (31), together with the weak convergence of the finite dimensional distributions shown in (21), proves the stated result. ■

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 Let $x < y$, and note that

$$\begin{aligned}|M(x) - M(y)|^p &= \left|\int_0^1 1\{W(r) \leq x\}\, dW(r) - \int_0^1 1\{W(r) \leq y\}\, dW(r)\right|^p \\ &= \left|\int_0^1 1\{x < W(r) \leq y\}\, dW(r)\right|^p.\end{aligned}$$

We have

$$\mathbf{E}\left|\int_0^1 1\{x < W(r) \leq y\}\, dW(r)\right|^p \leq c\,\mathbf{E}\left|\int_0^1 1\{x < W(r) \leq y\}\, dr\right|^{p/2}$$

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

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

Consequently, it follows that

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

and we may simply let

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

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, p. 25). ■

6.6 Proof of Lemma 4.3 Let $F_{nt}(\cdot)$ denote the distribution function of y_{nt}. For an integer $K > 1$, let

$$\xi_{ntm} = \inf\left\{x \in \mathbf{R} : F_{nt}(x) \geq \frac{m}{K}\right\} \text{ for } m = 1, ..., K-1,$$

and also let $\xi_{nt0} = -\infty$ and $\xi_{ntK} = +\infty$. Define $\mathcal{F} = \{\triangle y_{nt} 1\{y_{n,t-1} \leq x\} : x \in \mathbf{R}\}$ to be a class of functions and we denote a uniform analogue of the L^1- norm by $\rho(f) = \sup_{n,t} \mathbf{E}|f(x_{nt})|$ for $f \in \mathcal{F}$, where $x_{nt} = (y_{nt}, y_{n,t-1})'$.

By construction, for each $x \in R$, there exists $m \in \{1, ..., K-1\}$ such that

$$|F_{nt}(x) - F_{nt}(\xi_{ntm})| \leq \frac{1}{K},$$

so that

$$\begin{aligned}
& |\triangle y_{nt} [1\{y_{n,t-1} \leq x\} - 1\{y_{n,t-1} \leq \xi_{ntm}\}]| \\
\leq \ & |\triangle y_{nt}| \sup_{y : |F_{nt}(y) - F_{nt}(\xi_{ntm})| \leq \frac{1}{K}} |1\{y_{n,t-1} \leq y\} - 1\{y_{n,t-1} \leq \xi_{ntm}\}| \\
\equiv \ & b_{ntm}, \text{ say.} \qquad (32)
\end{aligned}$$

This result implies that for any function $\triangle y_{nt} 1\{y_{n,t-1} \leq x\}$ in $\mathcal{F}$, there exists $m \in \{1, ..., K-1\}$ such that

$$l_m \leq \triangle y_{nt} 1\{y_{n,t-1} \leq x\} \leq u_m \ ,$$

where

$$\begin{aligned}
l_m &= 1\{y_{n,t-1} \leq \xi_{ntm}\} - b_{ntm}, \\
u_m &= 1\{y_{n,t-1} \leq \xi_{ntm}\} + b_{ntm}.
\end{aligned}$$

Note that we have

$$\begin{aligned}
\rho(b_{ntm}) &= \sup_{n,t} \mathbf{E} |\triangle y_{nt}| \sup_{y : |F_{nt}(y) - F_{nt}(\xi_{ntm})| \leq \frac{1}{K}} |1\{y_{n,t-1} \leq y\} - 1\{y_{n,t-1} \leq \xi_{ntm}\}| \\
&\leq \sup_{n,t} \mathbf{E} |\triangle y_{nt}| \, 1\{\xi_{nt,m-1} \leq y_{n,t-1} \leq \xi_{nt,m+1}\} \\
&\leq \sup_{n,t} (\mathbf{E}|\triangle y_{nt}|^p)^{1/p} \sup_{n,t} |F_{nt}(\xi_{nt,m+1}) - F_{nt}(\xi_{nt,m-1})|^{1-1/p} \\
&= C_1 \left(\frac{1}{K}\right)^{1-1/p}, \qquad (33)
\end{aligned}$$

where $C_1 = 2^{1-1/p} \sup_{n,t} \mathbf{E}|\triangle y_{nt}|^p < \infty$ by Assumption 4.2 and the second inequality holds by Hölder's inequality. Therefore,

$$\{[l_m, \ u_m] : m = 1, ..., K-1\}\}$$

forms an $\varepsilon = 2C_1K^{1/p-1}$ bracket for $(\mathcal{F}, \rho)$. Hence, the bracketing number (see, e.g., van der Vaart and Wellner (1996, p.83) for the definition) satisfies

$$N(\varepsilon, \mathcal{F}, \rho) \leq \left(\frac{2C_1}{\varepsilon}\right)^{p/(p-1)} < \infty. \tag{34}$$

This result and pointwise WLLN of Andrews (1988, example 4, p.462) give the desired result using an argument similar to Lemma 2.4.1 of van der Vaart and Wellner (1996, p.123). ■

6.7 Proof of Theorem 4.4 Under Assumption 4.2, we have

$$\sigma_n \to_p \left(\lim_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} \mathbf{E} \left(\triangle y_{nt}\right)^2\right)^{1/2} \equiv \sigma^* < \infty \tag{35}$$

by WLLN of Andrews (1988, example 4, p.462). Define

$$Q(y) = \int H(z)1(z \leq y)\, d\mathrm{P}(z). \tag{36}$$

Then, by Lemma 4.3, (35) and rearranging terms, we have

$$n^{-1/2}S_n \to_p (1/\sigma^*) \sup_{y\in\mathbf{R}} |Q(y)| \tag{37}$$

and

$$n^{-1}T_n \to_p (1/\sigma^*) \int Q^2(y)\, d\mathrm{P}(y). \tag{38}$$

The stated result now follows since the right hand sides of (37) and (38) are positive under Assumption 4.1. ■

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. (1988). ”Laws of large numbers for dependent non-identically distributed random variables,” *Econometric Theory*, 4, 458-467.

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.

Andrews, D. W. K. (1997). ”A conditional Kolmogorov test,” *Econometrica*, 65, 1097-1128.

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

Billingsley, P. (1995). *Probability and Measure*, 3rd ed. John Wiley: New York.

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

Bierens, H.J. and W. Ploberger (1997). "Asymptotic theory of integrated conditional moment tests," *Econometrica*, 65, 1129-1152.

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.

Chang, Y. and J.Y. Park (2004). "Endogeneity in nonlinear regressions with integrated time series," unpublished manuscript, Department of Economics, Rice University.

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

de Jong, R. M. (1995). "Law of large numbers for dependent heterogeneous processes," *Econometric Theory,* 11, 347-358.

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

Delgado, M. (1993). "Testing the equality of nonparametric regression curves," *Statistics and Probability Letters*, 17, 199-204.

Deo, R. S. (2000). "Spectral tests of the martingale hypothesis under conditional heteroscedasticity," *Journal of Econometrics*, 99, 291-315.

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.

Dominguez, M. A. and I. N. Lobato (2000). "A consistent test for the martingale difference hypothesis," unpublished manuscript, Instituto Tecnológico Autónomo de México.

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.

Hong, Y. (1999). "Hypothesis testing in time series via the empirical characteristic function: A generalized spectral density approach," *Journal of the American Statistical Association*, 84, 1201-1220.

Khmaladze, E. V. (1988), "An innovation approach to goodness-of-fit 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.

Kuan, C. -M. and W. -M. Lee (2003), "A new test of the martingale difference hypothesis," unpublished manuscript, Institute of Economics, Academia Sinica.

Kurtz, T.G. and P. 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.

Shorack, G. R. and J. A. Wellner (1986). *Empirical Processes with Applications to Statistics*, Wiley: New York.

Stinchcombe, M. B. and H. White (1998). "Consistent specification testing with nuisance parameters present only under the alternative," *Econometric Theory* 14, 295-325.

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.

Stute, W. (1997). "Nonparametric model checks for regression," *Annals of Statistics*, 25, 613-641.

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

van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes.* Springer-Verlag: New York.

Whang, Y. -J. (2000). "Consistent bootstrap tests of parametric regression functions," *Journal of Econometrics*, 98, 27-46.

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.