TESTING THE MARTINGALE HYPOTHESIS

By

Peter C.B. Phillips and Sainan Jin

September 2013

COWLES FOUNDATION DISCUSSION PAPER NO. 1912

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281

http://cowles.econ.yale.edu/

Testing the Martingale Hypothesis*

Peter C. B. Phillips
Yale University, University of Auckland,
University of Southampton & Singapore Management University

Sainan Jin
Singapore Management University

September 17, 2013

Abstract

We propose new tests of the martingale hypothesis based on generalized versions of the Kolmogorov-Smirnov and Cramér-von Mises tests. The tests are distribution free and allow for a weak drift in the null model. The methods do not require either smoothing parameters or bootstrap resampling for their implementation and so are well suited to practical work. The paper develops limit theory for the tests under the null and shows that the tests are consistent against a wide class of nonlinear, non-martingale processes. Simulations show that the tests have good finite sample properties in comparison with other tests particularly under conditional heteroskedasticity and mildly explosive alternatives. An empirical application to major exchange rate data finds strong evidence in favor of the martingale hypothesis, confirming much earlier research.

JEL Classification: C12

Keywords: Brownian functional, Martingale hypothesis, Kolmogorov-Smirnov test, Cramér-von Mises test, Explosive process, Exchange rates.

*The authors thank Joon Park and Yoon-Jae Whang for comments. Phillips thanks the NSF for support under Grant Nos.SES-0956687 and SES 12-58258. Jin acknowledges research support from the Sim Kee Boon Institute at Singapore Management University.

1 Introduction

Martingales underlie many important results in economics and finance. According to Hall (1978), for example, when individuals maximize expected utility the conditional expectation of their future marginal utility is under certain conditions a function of present consumption, and other past information is irrelevant, making the marginal utility of consumption a martingale. Similarly, the fundamental theorem of asset pricing shows that if the market is in equilibrium and there is no arbitrage opportunity, then properly normalized asset prices are martingales under some probability measure. Efficient markets are then defined when available information is "fully reflected" in market prices, leading to stochastic processes that are martingales (Fama,1970). Empirical demonstration that a stochastic process is a martingale is thus extremely useful as it justifies the use of models and assumptions that are fundamental in economic theory.

Given the current information set, the martingale hypothesis implies that the best predictor of future values of a time series, in the sense of least mean squared error, is simply the current value of the time series. So current values fully represent all the available information. Formally, for a given a time series X_t let $\mathcal{F}_t$ be the filtration to which X_t is adapted. The martingale hypothesis (MH) for X_t requires the conditional expectation with respect to the past information in $\mathcal{F}_{t-1}$ to satisfy

$$\mathbb{E}\left(X_t|\mathcal{F}_{t-1}\right) = X_{t-1} \tag{1}$$

almost surely (a.s.). Let $I_t = \{X_t, X_{t-1}, X_{t-2}, ...\}$. The natural choice for $\mathcal{F}_t$ is the σ-field generated by I_t and this may be extended by including other covariates of interest in the information set I_t. There have been many studies in the literature concerned with tests of the martingale hypothesis. Most of these concentrate on tests of the martingale difference hypothesis (MDH), viz

$$\mathbb{E}\left(\Delta X_t|\Omega_{t-1}\right) = \mu \tag{2}$$

for some unknown $\mu \in \mathbb{R}$ and where Δ is the difference operator, $\Delta X_t = X_t - X_{t-1}$ and $\Omega_t = \{\Delta X_t, \Delta X_{t-1}, \Delta X_{t-2}, ...\}$. The MDH is slightly modified in this formulation to allow for an unknown mean for ΔX_t and information set based on the differences. Typically the information set includes the infinite past history of the series and I_t

and Ω_t may be taken as equivalent in this case. If a finite number of lagged values is included in the conditioning set, some dependence structure in the process may be missed due to omitted lags. However, tests that are designed to cope with the infinite lag case may have very low power (e.g., de Jong (1996)) and may not be feasible in empirical applications.

Several procedures for MDH testing are currently popular. Since Lo and MacKinlay (1988) proposed a variance ratio (VR) test, this procedure has been widely used and undergone many improvements for testing market efficiency and return predictability – see Chow and Denning (1993), Choi (1999), Wright (2000), Chen and Deo (2006), and Kim (2006), among many others. An alternative test for return predictability is the Box-Pierce (BP) test proposed by Box and Pierce (1970) and Ljung and Box (1978) and later generalized by Lobato, Nankervis and Savin (2001, 2002) and Escanciano and Lobato (2009a). These two categories of tests are designed to test lack of serial correlation but not necessarily the MDH. The spectral shape tests proposed by Durlauf (1991) and Deo (2000) are powerful in testing for lack of correlations but may not be able to detect nonlinear non-martingales with zero correlations. Nankervis and Savin (2010) use another approach based on generalizing the Andrews-Ploberger tests and find these tests have good power compared to the generalized BP tests of Lobato Nankervis and Savin (2002) and the Deo (2000) tests. These tests are designed to test a linear dependence structure when the time series is uncorrelated but may be statistically dependent. In order to capture nonlinear dependence which has recently been shown to be evident in asset returns, some new MDH tests have been proposed – see Hong (1999), Domínguez and Lobato (2003, hereafter DL), Hong and Lee (2003, 2005), Kuan and Lee (2004), Escanciano and Velasco (2006), among others. Readers may refer to Escanciano and Lobato (2009b) for a comprehensive review.

All the above tests are martingale difference tests. Technically, it is often simple and convenient to deal with asset returns and test whether the asset returns follow a martingale difference sequence (MDS). Park and Whang (2005, hereafter PW) introduced some explicit statistical tests of the martingale hypothesis that are very different from the MDH tests. Drift is assumed to be zero and PW test for a pure martingale process. Simulations show that the tests are robust to conditional heteroskedasticity under the null and have power against some general alternatives including many interesting nonlinear non-martingale processes such as exponential and threshold autoregressive processes, markov switching and chaotic processes (possibly with stochastic noise) and some other nonstationary processes. However the PW tests appear to be inconsistent against explosive processes such as the simple AR(1) with explosive coefficient θ. In particular, the simulations in PW (Table 11) show that test power against a simple explosive alternative $H_1 : \theta > 1$ declines as $n \to \infty$ when $\theta = 1.05$ but increases when $\theta = 1.01$. One contribution of the present paper is to provide a limit theory that confirms these anomalous simulation findings, showing that the PW tests are inconsistent against explosive AR(1) alternatives. Also, some key results in PW need rigorous limit theory for their justification and new arguments to address the difficulties are provided here.

The present paper proposes some new martingale tests which can be regarded as generalizations of the Kolmogorov-Smirnov test and the Cramér-von Mises goodness of fit test. One sequence of tests proposed here (GKS_n and $GCVM_n$ defined in (11) below) modify the S_n and T_n tests in PW. The limiting forms of these tests are defined and new technical arguments are given in developing the weak convergence arguments to these limits. The other sequence of tests (GKS_n^* and $GCVM_n^*$ defined in (12)) explicitly take into account the possibility of drift in the null model, which may be relevant in some empirical applications. In particular, the model may involve a weak deterministic drift that captures mild departures from a martingale null. This type of weak drift, which can be modeled via an evaporating intercept of the form $\mu = \mu_0 n^{-\gamma}$, was studied in recent work by Phillips, Shi and Yu (2012; PSY) on real time bubble detection methods. Many financial and macroeconomic time series observed over short and medium terms display drift but the drift is often small, hard to detect and may not be the dominating component of the series, thereby justifying this type of formulation.

Martingale with a weak drift in the null satisfy

$$\mathbb{E}\left((X_t - \mu)|I_{t-1}\right) = X_{t-1}, \tag{3}$$

or, equivalently, the empirically appealing and convenient form

$$\mathbb{E}\left((\Delta X_t - \mu)|I_{t-1}\right) = 0, \tag{4}$$

with $\mu = \mu_0 n^{-\gamma}$. The magnitude of the drift depends on the sample size n and a localizing exponent parameter γ. Estimation of γ is discussed in PSY (2012). When γ is positive, the drift term is small relative to a linear trend. We develop asymptotic theory for tests of (4) over different ranges of γ. When $\gamma \in [0, 0.5)$ for which the drift dominates the stochastic trend, the test statistics are asymptotically distribution free. When $\gamma = 0.5$, where $n^{-1/2}X_t$ behaves asymptotically like a Brownian motion with drift, the limit theory is quite different from the previous case and bootstrap tests have to be used as the limit theory depends on nuisance parameters. Time series for which the drift dominates and $\gamma \in [0, 0.5]$ are not martingales and thus not of central interest to this paper. Instead, we focus on the case where $\gamma > 0.5$ and the drift is small relative to the martingale and stochastic trend. In this case the intercept does not affect the limit theory and test limit distributions are free of nuisance parameters. These limit distributions are easy to compute, do not require bootstrap procedures to obtain critical values, and the tests involve no bandwidth parameters. So they are well suited to practical work.

Our tests are consistent against a wide class of nonlinear, non-martingale processes including explosive AR(1) processes, exponential autoregressive processes, threshold autoregressive models, bilinear processes, and nonlinear moving average models. Simulations show that the GKS_n^* and $GCVM_n^*$ tests generally perform better than GKS and $GCVM_n$, while the GKS and $GCVM_n$ tests generally perform slightly better than the S_n and T_n tests introduced in PW. However, for some data generating processes the performance of the PW tests is particularly poor and the comparisons are more dramatic in those cases. A leading example is the case where the data are generated by explosive AR (1) processes. When the AR(1) coefficient is 1.05, the rejection probabilities of our tests and the PW tests are above 90% for various GARCH specifications when the sample size is small. But for large samples, the power of the PW T_n test declines to 50% when $n = 1,000$ whereas our tests have 100% power in that case. Another example is the near-unit root case where the performance of the PW tests is unsatisfactory especially when sample size is small. In particular, when the AR(1) coefficient is 0.95, the PW tests basically have no power when n is less than 500, and the rejection probabilities are 48.4% for S_n and 73.5% for T_n when $n = 1,000$; the GKS and $GCVM_n$ tests perform slightly better than the

PW tests, and the GKS_n^* and $GCVM_n^*$ tests have noticeably superior power. When $n = 250$, the rejection probabilities are around 30%, and they reach 100% for GKS_n^* and $GCVM_n^*$ when sample size rises to $1,000$.

Simulations show that our tests have good size control and are robust to GARCH and stochastic volatility structures in the errors when the drift is set to zero. When $\mu = \mu_0 n^{-\gamma}$, with $\gamma = 1$, the martingale component dominates the drift and test size is robust to thick tails. We also try to assess the sensitivity of our tests by setting $\gamma = 0.5$, where $n^{-1/2}X_t$ behaves asymptotically like a Brownian motion with drift. In this case, the tests GKS_n and $GCVM_n$ suffer large size distortions, while the tests GKS_n^* and $GCVM_n^*$ still work well with good size performance. This outcome is unsurprising since the GKS and $GCVM_n$ tests are based on the PW tests which are designed for null settings with $\mu = 0$, while the GKS_n^* and $GCVM_n^*$ tests are constructed to allow explicitly for drift in the data.

Our tests and the PW tests are closely related to the test proposed by DL. The former test the MH null (1), while the DL test is an MDH test and tests the null (2). As emphasized in PW, the former deal with levels and the latter relies only on first differences. Many popular models in economic and financial applications (including threshold autoregressive models, error correction models and various diffusion models) specify how the conditional mean changes as a function of lagged levels rather than lagged differences, thereby increasing the appeal of the martingale null (1). Tests of (1) lead to different asymptotics from those of tests of (2) mainly because the presence of lagged levels in the test statistics influences the limit theory. An advantage of these asymptotics for our tests is that they are distribution free and implementation does not require user-selected bandwidth parameters or bootstrapping even when a drift is present in the model. On the other hand, many MDS tests, including DL, require bootstrap resampling and/or smoothing parameter selection for their implementation. Direct analysis of differences between the limit theory of MH and MDH tests is not possible. But the finite sample performance and asymptotic characteristics of the new tests make them a useful addition to this literature.

We apply the tests to examine evidence for the martingale hypothesis in major exchange rate data, as studied recently in Escanciano and Lobato (2009b). The null martingale hypothesis is supported for all exchange rates at both daily and weekly frequencies with the exception of the (weekly) Japanese Yen, where there is a rejection at the 5% level with the $GCVM_n^*$ test – so the empirical results are inconclusive in that case. The MDS tests used in Escanciano and Lobato (2009b) find similar results with some minor differences. Their results indicate that exchange rate returns are martingale difference sequences with the exception of the daily Euro exchange rate return for which the MDS is rejected.

The rest of the paper is organized as follows. Section 2 introduces the model, formulates the hypothesis, and constructs the tests. Section 3 establishes limit theory under the null and Section 4 shows consistency. Simulations are reported in Section 5 and Section 6 provides an empirical application to major foreign exchange markets. Section 7 concludes. Proofs and additional technical results are given in the Appendix.

2 Hypotheses and Tests

The martingale null is formulated as

$$X_t = \mu + \theta X_{t-1} + u_t, \quad \text{with } \theta = 1, \tag{5}$$

so that $X_t = \mu t + \xi_t + X_0$ with $\xi_t = \Sigma_{s=1}^t u_s$ and initialization $X_0 = 0$ for convenience. Then under weak conditions on u_t

$$\mathbb{E}\left((\Delta X_t - \mu)|I_{t-1}\right) = 0. \tag{6}$$

The intercept is defined as $\mu = \mu_0 n^{-\gamma}$ so the deterministic drift in X_t is $\mu t = \mu_0 t/n^\gamma$, whose magnitude depends on the sample size and the localizing parameter γ. When $\gamma = 0$, the drift produces a linear trend μt component in X_t under the null. When γ is positive, the drift $\mu_0 t/n^\gamma$ is small relative to a linear trend as $n \to \infty$ but still dominates the stochastic trend component $\xi_t = \Sigma_{s=1}^t u_s$ in X_t when $\gamma \in (0, 0.5)$. When $\gamma = 0.5$, $n^{-1/2} X_t$ behaves asymptotically like a Brownian martingale with drift. When $\gamma > 0.5$, the drift is small relative to the stochastic trend and $n^{-1/2} X_t$ behaves like a Brownian martingale in the limit as $n \to \infty$ under very general conditions on u_t. This formulation suits many financial and macroeconomic time series for which a small (possibly negligible) drift may be present in the series but where the drift is not the dominant component and is majorized by the martingale component. Accordingly, hypothesis testing of the null (6) which allows for that possibility will often be empirically more appealing than a pure martingale null in which $\mu = 0$ is imposed.

The tests we construct are based on the following equivalence (see e.g. Billingsley (1995, page 213, Theorem 16.10 (iii))

$$\mathbb{E}\left((\Delta X_t - \mu)|I_{t-1}\right) = 0 \text{ a.s. iff } \mathbb{E}(\Delta X_t - \mu)W(I_{t-1}) = 0 \tag{7}$$

where $W(\cdot)$ represents any $\mathcal{F}_{t-1}$ measurable weighting function. A convenient choice of weight function W is the indicator function $\mathbf{1}(\cdot)$, as is common in work on econometric specification, such as Andrews (1997), Stute (1997), Koul and Stute (1999), and Whang (2000). Other classes of functions, such as complex exponential functions considered in Bierens (1984, 1990) and Bierens and Ploberger (1997), might be used instead. None of the weighting function classes dominate, but the indicator function has the advantage that it is particularly convenient for use with integrated time series (as shown in Park and Phillips, 2000, 2001) and does not require selection of an arbitrary nuisance parameter space.

As in PW we concentrate on the simple case where

$$\mathbb{E}((X_t - \mu)|\mathcal{F}_{t-1}) = \mathbb{E}((X_t - \mu)|X_{t-1}),$$

and thus

$$\mathbb{E}\left((\Delta X_t - \mu)|X_{t-1}\right) = 0 \text{ a.s. iff } \mathbb{E}(\Delta X_t - \mu)\mathbf{1}(X_{t-1} \leq x) = 0, \tag{8}$$

for almost all $x \in \mathbb{R}$. The formulation (8) may be restrictive in some applications and it may be desirable to deal with more general processes in which

$$\mathbb{E}((\Delta X_t - \mu)|\mathcal{F}_{t-1}) = \mathbb{E}((\Delta X_t - \mu)|X_{t-1}, X_{t-2}, ...X_{t-p}, Z_{t-1}, Z_{t-2}, ..., Z_{t-k}) \tag{9}$$

for all $t \geq 1$, with some $p \geq 2$, $k \geq 1$. The DL test for the MDH works from a form different from (9) in which

$$\mathbb{E}((\Delta X_t - \mu)|\mathcal{F}_{t-1}) = \mathbb{E}((\Delta X_t - \mu)|\Delta X_{t-1}, \Delta X_{t-2}, ...\Delta X_{t-p}, W_{t-1}, W_{t-2}, ..., W_{t-k}) \tag{10}$$

The conditioning information set includes lagged differences instead of lagged levels, and conditioning variables are there assumed to be strictly stationary and ergodic. Extension of our framework to the general case (9) is technically challenging and is not pursued in the present paper.

Under the null, $(u_t, \mathcal{F}_t)$ is a martingale difference sequence assumed to satisfy the following conditions, as in PW.

Assumption 1 *(a)* $\frac{1}{n}\sum_{t=1}^{n}\mathbb{E}(u_t^2|\mathcal{F}_{t-1}) \to_p \sigma^2 > 0,$ *and*

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

Part (a) allows the innovation sequence to be conditionally heteroskedastic with variation that averages out in the limit. Part (b) requires uniformly bounded fourth conditional moments. This assumption might be relaxed at the cost of greater complications. Simulations show that the tests considered here have good size and power performance in the presence of conditional heteroskedasticity even when (b) may not apply (e.g. GARCH errors) as in PW.

Define the least squares (OLS) residual $\widehat{u}_t = X_t - \widehat{\mu} - \widehat{\theta}X_{t-1}$, where $(\widehat{\mu}, \widehat{\theta})$ is known to be consistent for (μ, θ) under quite general conditions (Phillips, 1987), and the following holds by straightforward arguments.

Lemma 1 *Let Assumption 1 hold. Under the null, we have* $\sigma_n^2 = \frac{1}{n}\sum_{t=1}^{n}u_t^2 \to_p \sigma^2$, $\widehat{\sigma}_n^2 = \frac{1}{n}\sum_{t=1}^{n}\widehat{u}_t^2 \to_p \sigma^2$ *as* $n \to \infty$.

The following self normalized quantities form the basis of the test statistics that we consider here

$$\Gamma_n(x) = \frac{\sum_{t=1}^{n}\Delta X_t \mathbf{1}(X_{t-1} \leq x)}{\left(\sum_{t=1}^{n}\widehat{u}_t^2\right)^{1/2}} = \frac{\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\Delta X_t \mathbf{1}(X_{t-1} \leq x)}{\left(\frac{1}{n}\sum_{t=1}^{n}\widehat{u}_t^2\right)^{1/2}},$$

and

$$\Gamma_n^*(x) = \frac{\sum_{t=1}^{n}\left(\Delta X_t - \overline{\Delta X}\right)\mathbf{1}(X_{t-1} \leq x)}{\left(\sum_{t=1}^{n}\widehat{u}_t^2\right)^{1/2}} = \frac{\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left(\Delta X_t - \overline{\Delta X}\right)\mathbf{1}(X_{t-1} \leq x)}{\left(\frac{1}{n}\sum_{t=1}^{n}\widehat{u}_t^2\right)^{1/2}},$$

where $\overline{\Delta X} = \frac{1}{n}\sum_{t=1}^{n}\Delta X_t$. Define

$$J_n(a) = \Gamma_n(a\sqrt{n}) = \Gamma_n(x), \quad \text{and } J_n^*(a) = \Gamma_n^*(a\sqrt{n}) = \Gamma_n^*(x),$$

for $x = a\sqrt{n}$. The quantities $J_n(a)$ and $J_n^*(a)$ are stochastic processes with parameter $a \in \mathbb{R}$ taking values in the space of RCLL functions. We consider two specific types of tests, which extend the Kolmogorov-Smirnov test and the Cramér-von Mises test of goodness of fit to this regression framework:

$$GKS_n = \sup_{a\in\mathbb{R}} |J_n(a)| = \sup_{x\in\mathbb{R}} |\Gamma_n(x)|, \quad \text{and } GCVM_n = \frac{1}{n}\sum_{t=1}^{n} (\Gamma_n(X_{t-1}))^2, \tag{11}$$

and

$$GKS_n^* = \sup_{a\in\mathbb{R}} |J_n^*(a)| = \sup_{x\in\mathbb{R}} |\Gamma_n^*(x)|, \quad \text{and } GCVM_n^* = \frac{1}{n}\sum_{t=1}^{n} (\Gamma_n^*(X_{t-1}))^2. \tag{12}$$

Remark 1. Under the null, $\Delta X_t = u_t + \mu$ and $\Delta X_t - \overline{\Delta X} = u_t - \overline{u}$, where $\overline{u} = \frac{1}{n}\sum_{t=1}^{n} u_t$. We normalize $\frac{1}{\sqrt{n}}\sum_{t=1}^{n} \Delta X_t \mathbf{1}(X_{t-1} \leq x) = \frac{1}{\sqrt{n}}\sum_{t=1}^{n}(u_t+\mu)\mathbf{1}(X_{t-1} \leq x)$ and $\frac{1}{\sqrt{n}}\sum_{t=1}^{n}(\Delta X_t - \overline{\Delta X})\mathbf{1}(X_{t-1} \leq x) = \frac{1}{\sqrt{n}}\sum_{t=1}^{n}(u_t - \overline{u})\mathbf{1}(X_{t-1} \leq x)$ by $\left(\frac{1}{n}\sum_{t=1}^{n}\widehat{u}_t^2\right)^{1/2}$, a consistent estimator of σ. A natural alternative normalization of the numerator is a consistent standard error estimator such as $\left(\frac{1}{n}\sum_{t=1}^{n}\widehat{u}_t^2\mathbf{1}(X_{t-1} \leq x)\right)^{1/2}$. But simulations show that test statistics based on this normalization, viz.,

$$\frac{\sum_{t=1}^{n}(\Delta X_t - \overline{\Delta X})\mathbf{1}(X_{t-1} \leq x)}{\left(\sum_{t=1}^{n}\widehat{u}_t^2\mathbf{1}(X_{t-1} \leq x)\right)^{1/2}},$$

tend to have size distortions when the errors exhibit strong conditional heteroskedasticity. For this reason we normalize by $\left(\frac{1}{n}\sum_{t=1}^{n}\widehat{u}_t^2\right)^{1/2}$.

Remark 2. PW set $\mu = 0$, so $\Delta X_t = u_t$, and assume $\sigma^2 = 1$ so that u_t is self normalized with $\sigma_n^2 = 1$. That normalization can be achieved in practice by dividing X_t with $\sigma_n = \left(\frac{1}{n}\sum_{t=1}^{n}u_t^2\right)^{1/2}$. The test statistics used in PW are defined as

$$S_n = \sup_{a\in\mathbb{R}} |M_n(a)| = \sup_{x\in\mathbb{R}} |Q_n(x)|, \text{ and } T_n = \frac{1}{n}\sum_{t=1}^{n} Q_n^2(X_{t-1}),$$

where $M_n(a) = Q_n(a\sqrt{n}) = Q_n(x)$ and

$$Q_n(x) = \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{\Delta X_t}{\sigma_n}\mathbf{1}\left(\frac{X_{t-1}}{\sigma_n} \leq x\right).$$

As shown in Section 4 below, under an explosive AR(1) alternative, the PW test statistics normalized by σ_n are inconsistent, which explains some of the anomalous power findings reported in PW for the explosive case. Our statistics are normalized by $\left(\frac{1}{n}\sum_{t=1}^{n}\widehat{u}_t^2\right)^{1/2}$ and are shown to be divergent when $n \to \infty$ giving consistent tests under explosive alternatives.

3 Asymptotic Distribution under H_0

3.1 Asymptotic behavior of GKS_n and GKS_n^*

Under the null, $X_t = \mu t + \xi_t$,where $\xi_t = \sum_{s=1}^{t} u_s$ is first order Markovian. The process $W_n(r) = \frac{1}{\sqrt{n}}\sum_{t=1}^{[nr]}\frac{u_t}{\sigma_n}$ satisfies the functional law

$$W_n(r) \Rightarrow W(r), \quad r \in [0,1] \tag{13}$$

where $W(\cdot)$ is the standard Brownian motion. The weak convergence of $J_n(a)$ and $J_n^*(a)$ is presented in the following lemma.

Lemma 2 *Let Assumption 1 hold. Under the null, when $\mu = \mu_0 n^{-\gamma}$ with $\gamma \geq 0.5$, we have*

$$J_n(a) \Rightarrow J(a), \text{ and } J_n^*(a) \Rightarrow J^*(a),$$

where

$$J(a) = \int_0^1 1\{W(s) \leq a\}\, dW(s), \quad J^*(a) = \int_0^1 1\{W(s) \leq a\}\, dB(s), \tag{14}$$

when $\gamma > 0.5$; and

$$J(a) = \int_0^1 1\{W(s) + \mu_0 s \leq a\}\, d\left(W(s) + \mu_0 s\right), \quad J^*(a) = \int_0^1 1\{W(s) + \mu_0 s \leq a\}\, dB(s), \tag{15}$$

when $\gamma = 0.5$. If $\mu = \mu_0 n^{-\gamma}$ and $\gamma < 0.5$

$$\frac{1}{n^{0.5-\gamma}}\Gamma_n(bn^{1-\gamma}) = \frac{1}{n^{0.5-\gamma}}\Gamma_n(x) \to_p \mu_0 \int_0^1 1\{s \leq b\}\, ds,$$
$$\Gamma_n^*(bn^{1-\gamma}) = \Gamma_n^*(x) = \int_0^1 1\{s \leq b\}\, dB(s),$$

for $x = bn^{1-\gamma}$. $B(s) = W(s) - sW(1)$ is a standard Brownian bridge on the unit interval.

Remark 3. As discussed previously, martingale tests are of little interest when the drift term dominates or has the same magnitude as the martingale component because neither the time series nor the limit process is a martingale in these cases. Hence, in the following we focus on the case $\gamma > 0.5$ where the drift is small relative to the stochastic trend so the limit process is a martingale.

Theorem 3 *Let Assumption 1 hold. Under the null with $\mu = \mu_0 n^{-\gamma}$ and $\gamma > 0.5$, as $n \to \infty$*

$$GKS_n \Rightarrow \sup_{a \in \mathbb{R}} J(a), \quad GKS_n^* \Rightarrow \sup_{a \in \mathbb{R}} J^*(a)$$

where $J(a)$ and $J^*(a)$ are given in (14).

Remark 4. When $\mu = \mu_0 n^{-\gamma}$ with $\gamma > 0.5$, the intercept does not affect the asymptotics and the limit distributions are free of nuisance parameters. These features facilitate computation and no bootstrap resampling or smoothing parameter selection is needed for implementation. Asymptotic critical values of the test statistics when $\mu = \mu_0 n^{-\gamma}$ with $\gamma > 0.5$ are displayed in Table 1 in Section 5. The simulation results in Section 5 show that these tests have good size performance and are robust to GARCH or stochastic volatility structures of the errors when the drift $\mu = 0$ or has the form $\mu_0 n^{-\gamma}$ with $\gamma > 0.5$. When $\gamma = 0.5$, GKS_n^* and $GCVM_n^*$ using the critical values from Table 1 still work very well and these tests have good size performance.

Remark 5. The asymptotic distributions of GKS_n and GKS_n^* are easily obtained when $\mu = \mu_0 n^{-\gamma}$ with $\gamma \leq 0.5$. but as discussed in Remark 3, these results are not of direct interest in the present paper. When $\mu = \mu_0 n^{-\gamma}$ with $\gamma = 0.5$, as $n \to \infty$ we have

$$GKS_n \Rightarrow \sup_{a \in \mathbb{R}} J(a), \quad GKS_n^* \Rightarrow \sup_{a \in \mathbb{R}} J^*(a)$$

where $J(a)$ and $J^*(a)$ are given in (15). When $\mu = \mu_0 n^{-\gamma}$ with $\gamma < 0.5$, as $n \to \infty$ we have

$$\frac{1}{n^{0.5-\gamma}} \sup_{b \in \mathbb{R}} \left| \Gamma_n(bn^{1-\gamma}) \right| = \sup_{x \in \mathbb{R}} |\Gamma_n(x)| \to_p \sup_{b \in \mathbb{R}} \mu_0 \int_0^1 1\{s \leq b\}\, ds = \max[\mu_0, 0],$$
$$\sup_{b \in \mathbb{R}} \left| \Gamma_n^*(bn^{1-\gamma}) \right| = \sup_{x \in \mathbb{R}} |\Gamma_n^*(x)| \to_p \sup_{b \in \mathbb{R}} \int_0^1 1\{s \leq b\}\, dB(s).$$

Remark 6. PW set $\mu = 0$ and thus $X_t = \xi_t$. As discussed in Remark 2, the PW tests are defined as

$$S_n = \sup_{a \in \mathbb{R}} |M_n(a)| = \sup_{x \in \mathbb{R}} |Q_n(x)|, \quad \text{and } T_n = \frac{1}{n}\sum_{t=1}^{n} Q_n^2(X_{t-1}) = \int_0^1 M_n^2(W_n(r))dr,$$

where $M_n(a) = Q_n(a\sqrt{n}) = Q_n(x)$ and

$$Q_n(x) = \frac{1}{\sqrt{n}}\sum_{t=1}^{n} \frac{\Delta X_t}{\sigma_n}\mathbf{1}\left(\frac{X_{t-1}}{\sigma_n} \leq x\right) = \frac{1}{\sqrt{n}}\sum_{t=1}^{n} \frac{\Delta \xi_t}{\sigma_n}\mathbf{1}\left(\frac{\xi_{t-1}}{\sigma_n} \leq x\right).$$

Theorem 3.4 in PW shows that

$$S_n \Rightarrow S = \sup_{a \in \mathbb{R}} |M(a)| \tag{16}$$

where

$$M(a) = \int_0^1 1\{W(s) \leq a\}\, dW(s) \tag{17}$$

The result in (16) follows readily by continuous mapping because $M_n(a) \Rightarrow M(a)$ under H_0. PW also indicate that

$$T_n \Rightarrow T = \int_0^1 M^2(W(r))dr. \tag{18}$$

Proving (18) rigorously causes some difficulty. First $M(a)$ is defined in (17) as a stochastic integral, namely as an L_2 limit involving Riemann sums of the form

$$M_n(a) = \frac{1}{\sqrt{n}}\sum_{t=1}^{n} u_t \mathbf{1}\left(\frac{X_{t-1}}{\sigma_n\sqrt{n}} \leq a\right) = \frac{1}{\sqrt{n}}\sum_{t=1}^{n} u_t \mathbf{1}\left(\frac{\xi_{t-1}}{\sigma_n\sqrt{n}} \leq a\right),$$

an argument that requires the 'integrand' to be $\mathcal{F}_{t-1}-$measurable. For a given fixed a the quantity $\mathbf{1}\left(\frac{\xi_{t-1}}{\sigma_n\sqrt{n}} \leq a\right)$ is $\mathcal{F}_{t-1}-$measurable. However, the limit process given in (18) involves (by virtue of plugging in the stochastic process argument $a = W(r)$)

$$M(W(r)) = \int_0^1 1\{W(s) \leq W(r)\}\, dW(s). \tag{19}$$

Here, the integrand $1\{W(s) \leq W(r)\}$ is $\mathcal{F}_s-$measurable only when $0 \leq r \leq s$. Hence, $M(W(r))$ cannot be defined directly as a conventional stochastic integral. The definition of $M(W(r))$ is not discussed in PW. Performing a "plug in" with $M(a = W(r))$ may be interpreted as a functional composition $M(W(r)) := (M \circ W)(r)$ in which the process $M(a)$ for $a \in \mathbb{R}$ is composed with the stochastic process $W(r)$ which takes values in $\mathbb{R}$ for any given r. An alternate definition is given in the section below. A separate argument that takes account of the composite nature of $M(W(r))$ is needed to show the weak convergence of

$$T_n = \frac{1}{n}\sum_{t=1}^{n} Q_n^2(X_{t-1}) = \frac{1}{n}\sum_{t=1}^{n} Q_n^2(\xi_{t-1}) \Rightarrow \int_0^1 M^2(W(r))dr,$$

given in (18). A second difficulty in the PW argument is that the occupation time formula is used to derive the expression

$$\int_0^1 M^2(W(r))dr = \int_{-\infty}^{\infty} M^2(s)L(1,s)ds. \tag{20}$$

However, as is apparent from the definition (17)

$$M(a) = M_W(a) = \int_0^1 \mathbf{1}\{W(s) \leq a\}\, dW(s), \tag{21}$$

the functional M itself also depends on the same stochastic process $\{W(s)\}_0^1$, as is emphasized in the alternate notation $M_W(a)$ given in (21). The simple occupation time formula (20) is not justified here because of this functional dependence in the argument $M(a) = M_W(a)$. These two technical difficulties will be resolved in the following section.

3.2 Asymptotic behavior of $GCVM_n$ and $GCVM_n^*$

In order to justify the technical arguments leading to (19) and (20), it is helpful to show that $M_n(a) \Rightarrow M(a)$ uniformly over $a \in \mathbb{R}$. To achieve this result and study the asymptotic behavior of T_n, $GCVM_n$ and $GCVM_n^*$, we make the following assumptions.

Assumption 1b *(i)* $E(u_t^2|F_{t-1}) = \sigma^2$ *a.s. for all* $t = 1, ..., n$, *and*

(ii) $\sup_{t\geq 1} E(u_t^4|F_{t-1}) < K$ *a.s. for some constant* $K < \infty$.

Part (i) introduces a more restrictive condition on the innovations than Assumption 1(i) in order to make use of results on uniform convergence to stochastic integrals as discussed below. The condition might be relaxed to allow for conditional heteroskedasticity in the errors with more complicated arguments here and in the uniform convergence results we utilize but we do not undertake those extensions in the present paper. As demonstrated in the simulations reported below, our tests are found to have good finite sample properties that are robust to GARCH, EGARCH, and stochastic volatility formulations.

Assumption 2 *(a) $g(x, a)$ is H-regular as defined in Park and Phillips (1999), with asymptotic order $\kappa(\lambda, a)$, limit homogeneous function $h(x, a)$, and residual $R(x, \lambda, a)$, where $\lambda \in \mathbb{R}^+$. Then $g(\lambda x, a) = \kappa(\lambda, a)h(x, a) + R(x, \lambda, a)$, with $\kappa^{-1}(\lambda, a)R(x, \lambda, a) = o(1)$ for all a in a compact set A as $\lambda \to \infty$.*

(b) There exists a function v: $\mathbb{R} \to \mathbb{R}^+$ such that for all $x \in \mathbb{R}$ and a, $a' \in A$,

$$\sup_{\lambda \geq 1} \left| \kappa^{-1}(\lambda, a)g(\lambda x, a) - \kappa^{-1}(\lambda, a')g(\lambda x, a') \right| \leq v(x) \left| a - a' \right|$$

where function $v(x)$ is symmetric and bounded, $v(|x|)$ is increasing in $|x|$, with $\mathbb{E}v^2 (|W(1)| + c) < \infty$ for some $c > 0$, and there exists an $a \in A$, such that

$$\lim \sup_{n \to \infty} \frac{1}{n} \sum_{t=1}^{n} \mathbb{E}h^2 \left(\frac{\xi_{t-1}}{\sqrt{n}}, a \right) < \infty$$

A useful result concerning uniform weak convergence to stochastic integrals involving nonlinear homogeneous functions $g(\xi_{t-1}, a)$ that includes functions like $1\left(\frac{\xi_{t-1}}{\sqrt{n}} \leq a\right)$ of integrated processes is stated in Lemma 4 below. The result is based on Lemma 5.2 of Shi and Phillips (2012) and holds under weak conditions that apply here.

Lemma 4 *Let Assumptions 1b and 2 hold. Then, uniformly in $a \in A$, and in a suitably expanded probability space*

$$n^{-1/2}\kappa^{-1}(n^{1/2}, a) \sum_{t=1}^{n} g\left(\xi_{t-1}, a\right) u_t \to_p \sigma \int h(W, a)dW$$

under the null hypothesis.

Remark 7. From Lemma 4, it is straightforward to show that

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} u_t \mathbf{1} \left(\frac{\xi_{t-1}}{\sigma_n \sqrt{n}} \leq a \right) \to_p \int 1(W(s) \leq a)dW,$$

uniformly in $a \in A$ under the null hypothesis. Thus for all compact sets A we have

$$M_n(a) \to_p M(a) \tag{22}$$

uniformly in $a \in A$, and correspondingly on a suitable probability space we have

$$M_n(a) \to_{a.s.} M(a) \tag{23}$$

uniformly in $a \in A$. We would now like to show that uniformly for $a \in \mathbb{R}$, $M_n(a) \Rightarrow M(a)$.

Recall that

$$W_n(r) = \frac{1}{\sqrt{n}} \sum_{t=1}^{[nr]} \frac{u_t}{\sigma_n} \Rightarrow W(r), \tag{24}$$

and

$$M_n(a) \Rightarrow M(a), \tag{25}$$

for any $a \in \mathbb{R}$. Let

$$S_{nt} = \sup_{r \leq t} W_n(r), \quad \text{and } S_t = \sup_{r \leq t} W(r).$$

Note from (24) that

$$S_{nt} \Rightarrow S_t. \tag{26}$$

Take a probability space where (24), (25), and (26) apply almost surely and then on the expanded probability space

$$W_n(r) \rightarrow_{a.s.} W(r), \text{ and } S_{nt} \rightarrow_{a.s.} S_t, \tag{27}$$

from which it follows that

$$P(S_{nt} \geq b) \rightarrow P(S_t \geq b)$$

for some large $b > 0$. Note that (e.g., Proposition 3.7 in Revuz and Yor (1999))

$$P(S_t \geq b) = 2P(W_t \geq b) = P(|W_t|) \geq b),$$

where $W_t = BM(1)$. Using the boundary crossing probability

$$P(W(t) \geq b \text{ for some } t \in [0,1]) = O\left(e^{-2b^2}\right),$$

as $b \rightarrow \infty$ (see e.g., Siegmund (1986), Wang and Potzelberger(1997)), we therefore have

$$P(S_{nt} \geq b) \rightarrow P(S_t \geq b) = O\left(e^{-\alpha b^2}\right) \tag{28}$$

for some $\alpha > 0$ and $b \rightarrow \infty$.

The boundary crossing probability in (28) faciliates the development of the following uniform results.

Theorem 5 *Let Assumptions 1b and 2 hold. Let $\mu = 0$. Then, uniformly in $a \in \mathbb{R}$,*

$$M_n(a) \Rightarrow M(a), \tag{29}$$

where

$$M_n(a) = Q_n(a\sqrt{n}) = Q_n(x)$$

and $Q_n(x) = \frac{1}{\sqrt{n}}\sum_{t=1}^{n} \frac{\Delta X_t}{\sigma_n}\mathbf{1}\left(\frac{X_{t-1}}{\sigma_n} \leq x\right) = \frac{1}{\sqrt{n}}\sum_{t=1}^{n} \frac{\Delta \xi_t}{\sigma_n}\mathbf{1}\left(\frac{\xi_{t-1}}{\sigma_n} \leq x\right)$ *as defined in PW.*

It follows that when $\mu = \mu_0 n^{-\gamma}$ with $\gamma \geq 0.5$

$$J_n(a) \Rightarrow J(a), \quad \text{and } J_n^*(a) \Rightarrow J^*(a),$$

uniformly in $a \in \mathbb{R}$. Thus

$$J_n(X_n(r)) \Rightarrow J(W(r)), \text{ and } J_n^*(X_n(r)) \Rightarrow J^*(W(r)) \tag{30}$$

as required.

3.2.1 Definition of $M(W(r))$

As discussed in Remark 6, definition of $\int_0^1 M^2(W(r))dr$ requires definition of the stochastic integral

$$M(W(r)) = \int_0^1 1\{W(s) \leq W(r)\}\, dW(s) \tag{31}$$

that appears in the integrand. For this purpose it is convenient to use Tanaka's formula for local time (e.g. Revuz and Yor, 1999) that for all $a \in \mathbb{R}$

$$\begin{aligned}\int_0^t 1\{W(s) \leq a\}\, dW(s) &= \frac{1}{2}L_W(t,a) - \left\{(W(t)-a)^- - (W(0)-a)^-\right\} \\ &= \frac{1}{2}L_W(t,a) - \left\{(W(t)-a)^- - (-a)^-\right\}.\end{aligned}$$

It follows that we can write

$$M(a) = \int_0^1 1\{W(s) \leq a\}\, dW(s) = \frac{1}{2}L_W(1,a) - \left\{(W(1)-a)^- - (-a)^-\right\}.$$

This formulation enables us to define (31) directly as follows

$$M(W(r)) := \frac{1}{2}L_W(1,W(r)) - \left\{(W(1)-W(r))^- - (-W(r))^-\right\}. \tag{32}$$

In this expression, $L_W(1, W(r))$ is the local time that the process $\{W(s) : s \in [0,1]\}$ spends at $W(r)$, i.e. the local time that W over $[0,1]$ has spent at the current position $W(r)$. This concept appears in the probability literature in Aldous (1986) and is used in Phillips (2009). It is also related to the concept of self intersection local time used in Wang and Phillips (2012). With this approach all the quantities $\int_0^1 M^2(W(r))dr$, $\int_0^1 J^2(W(r))dr$, and $\int_0^1 J^{*2}(W(r))dr$ are well defined.

Using Theorem 5, we can establish the limit theory for $GCVM_n$ and $GCVM_n^*$.

Theorem 6 *Let Assumptions 1b and 2 hold. Under the null, when* $\mu = \mu_0 n^{-\gamma}$ *with* $\gamma > 0.5$, *we have*

$$GCVM_n \Rightarrow \int_0^1 J^2(W(r))dr, \quad GCVM_n^* \Rightarrow \int_0^1 J^{*2}(W(r))dr,$$

as $n \to \infty$, *where the quantities*

$$J(W(r)) = \int_0^1 1\{W(s) \leq W(r)\} dW(s),$$

and

$$J^*(W(r)) = \int_0^1 1\{W(s) \leq W(r)\} dB(s),$$

are defined as in (32).

4 Power Asymptotics

This section shows consistency of the new tests against non-martingale alternatives. The approach here follows PW. We first consider stationary-side alternatives to the null and replace the time series X_t by triangular arrays X_{nt} for $1 \leq t \leq n$, $n \geq 1$, making the following two assumptions.

Assumption 3 *The array* X_{nt} *is strong mixing satisfying* $\sup_{1 \leq t \leq n,\ n \geq 1} E|\Delta X_{nt}|^q < \infty$ *for some* $q \geq 2$.

Assumption 4 *For any Borel set* $A \subset R$, $\frac{1}{n}\sum_{t=1}^n P_{nt}(A) \to P(A)$ *as* $n \to \infty$ *where* P *is a probability measure on* R *and* P_{nt} *is the distribution of* X_{nt} *for* $1 \leq t \leq n$, $n \geq 1$. *Also,* $\frac{1}{n}\sum_{t=1}^n E(\Delta X_{nt}|X_{n,t-1} = x) \to H(x)$ *for all* $x \in R$ *as* $n \to \infty$, *where* H *is a measurable function on* R. $\int I(H(x) \neq 0)\, dP(x) > 0$.

Assumptions 3 and 4 are similar to Assumptions 4.2 and 4.1 in PW where their relevance and applicability are discussed. PW show that the following uniform weak law of large numbers holds under Assumption 3

$$\sup_{x\in\mathbb{R}}\left|\frac{1}{n}\sum_{t=1}^{n}\left[\Delta X_{nt}\mathbf{1}\left(X_{n,t-1}\leq x\right)-\mathbb{E}\Delta X_{nt}\mathbf{1}\left(X_{n,t-1}\leq x\right)\right]\right|\rightarrow_p 0.$$

The following lemma establishes the consistency of the tests under these conditions.

Lemma 7 *Suppose Assumptions 3 and 4 hold. We have*

$$GKS_n, GCVM_n, GKS_n^*, GCVM_n^* \rightarrow \infty$$

as $n\rightarrow\infty$.

The proof of Lemma 7 follows the proof of Theorem 4.4 in PW and is therefore omitted. Both our tests and the PW tests are consistent for the alternatives we consider here in Assumptions 3 and 4. As shown in the simulations reported in PW, the tests allow for quite flexible forms of nonstationarity. These simulations (Table 11 in PW) show that test power of T_n against a simple explosive alternative (with AR coefficient θ and $H_1:\theta>1$) declines as $n\rightarrow\infty$ when $\theta=1.05$ but increases when $\theta=1.01$. By contrast, tests based on GKS, $GCVM_n$, GKS_n^*, and $GCVM_n^*$ are consistent against an explosive AR (1) model with $\theta>1$ as shown in the following result, which remains true under more general weakly dependent errors u_t as can be shown using the results in Phillips and Magdalinos (2008, 2009). To simplify the exposition here, we maintain Assumption 1.

Theorem 8 *Under* $H_1:\theta>1$*, we have as* $n\rightarrow\infty$

$$GKS_n, GCVM_n, GKS_n^*, GCVM_n^* \rightarrow \infty.$$

Remark 9. The proof is given in the Appendix. Under explosive alternatives we find that $\sum_{t=1}^{n}\Delta X_t\mathbf{1}(X_{t-1}\leq x)=O_p\left(\theta^n\right)$, $\frac{1}{n}\sum_{t=1}^{n}\widehat{u}_t^2=O_p\left(1\right)$, and thus

$$\Gamma_n(x)=\frac{\sum_{t=1}^{n}\Delta X_t\mathbf{1}(X_{t-1}\leq x)}{\left(\sum_{t=1}^{n}\widehat{u}_t^2\right)^{1/2}}=O\left(\frac{\theta^n}{\sqrt{n}}\right),$$

$$\Gamma_n^*(x)=\frac{\sum_{t=1}^{n}(\Delta X_t-\overline{\Delta X})\mathbf{1}(X_{t-1}\leq x)}{\left(\sum_{t=1}^{n}\widehat{u}_t^2\right)^{1/2}}=O\left(\frac{\theta^n}{\sqrt{n}}\right),$$

so that tests based on $\Gamma_n(x)$ and $\Gamma_n^*(x)$ are consistent, explaining the results in the theorem. However, under explosive alternatives with $\theta > 1$, $\Delta X_t \neq u_t$, so $\sigma_n^2 = \frac{1}{n}\sum_{t=1}^n (\Delta X_t)^2$ as defined in PW does not equal $\frac{1}{n}\sum_{t=1}^n u_t^2$. Following similar arguments to those in the proof, it is easy to show that $\sigma_n^2 = \frac{1}{n}\sum_{t=1}^n (\Delta X_t)^2 = O_p\left(\frac{\theta^{2n}}{n}\right)$, and thus $Q_n(x) = \frac{1}{\sqrt{n}}\sum_{t=1}^n \frac{\Delta X_t}{\sigma_n}\mathbf{1}\left(\frac{X_{t-1}}{\sigma_n} \leq x\right) = O_p(1)$. Thus, the PW tests based on $Q_n(x)$ are not consistent against explosive AR (1) processes.

PW also look at the non-martingale unit root process generated by $\Delta X_t = u_t$ where u_t is serially correlated. Chang and Park (2011) show that

$$\frac{1}{\sqrt{n}}\sum_{t=1}^n u_t\mathbf{1}(X_t \leq 0) \to_d \int_0^1 1\left(W(r) \leq 0\right) dW(r) + L(1,0) \qquad (33)$$

where u_t is *iid* with zero mean and unit variance and $\Delta X_t = u_t$. When u_t is correlated with X_{t-1}, u_t is serially correlated. In that event, our tests and the PW tests have asymptotics related to (33). As pointed out in PW, the presence of serial correlation in u_t will therefore tend to shift the limit distributions of the tests by an additional term involving $L(1,0)$ as it appears in (33), but the tests are generally not consistent in this case. Simulation results not reported here show that, like the PW tests, our tests do have nontrivial power against such non-martingales when there is some dependence in the innovation sequence.

5 Simulation Evidence

This section reports results of simulations conducted to evaluate the finite sample performance of the tests given here. The limit distributions of the test statistics GKS, $GCVM_n$, GKS_n^*, and $GCVM_n^*$ are free of nuisance parameters when $\mu = \mu_0 n^{-\gamma}$ with $\gamma > 0.5$ and these distributions are readily obtained by simulation. Table 1 gives the asymptotic critical values of the test statistics GKS, $GCVM_n$, GKS_n^*, and $GCVM_n^*$ when $\mu = \mu_0 n^{-\gamma}$ with $\gamma > 0.5$. These critical values were generated for $n = 1,000$ observations using 50,000 replications and a Gaussian *iid* $N(0,1)$ null.

As noted in Section 3 when $\gamma < 0.5$, the limit distributions of the test statistics are also free of nuisance parameters, but the drift dominates the martingale process in this case and the case is not of direct interest. When $\gamma = 0.5$, the limit distributions depend on nuisance parameters and bootstrap versions of tests are needed. But the

Table 1: Asymptotic Critical Values of Test Statistics

sig. level	0.99	0.95	0.90	0.10	0.05	0.01
GKS_n	0.5930	0.7465	0.8462	2.0877	2.3519	2.8885
$GCVM_n$	0.0551	0.0999	0.1422	1.6118	2.1250	3.3667
GKS_n^*	0.5164	0.6481	0.7341	1.5326	1.6716	1.9300
$GCVM_n^*$	0.0426	0.0752	0.1066	0.7619	0.9214	1.2871

Note: asymptotic critical values of the test statistics are computed from simulations with 50,000 replications, iid $N(0,1)$ errors and $n = 1,000$

tests are not martingale tests in that case. The simulation experiments described below consider cases where $\mu = 0$ and $\mu \neq 0$ under the null. For $\mu \neq 0$,we set $\gamma = 1$. We also used $\gamma = 0.5$ to assess the sensitivity of the tests in that case.

5.1 Experimental design

We use the following data generating processes (DGPs) under the martingale null.

1. Random walk process (NULL1): $X_t = \mu + X_{t-1} + u_t$, $\mu = 0$, with
   (a) independent and identically distributed N(0,1) errors (IID): $u_t \sim iidN(0,1)$.
   (b) GARCH errors as used in PW (GARCH): $u_t = \sigma_t \varepsilon_t$, $\sigma_t^2 = 1 + \theta_1 u_{t-1}^2 + \theta_2 \sigma_{t-1}^2$, $\varepsilon_t \sim iidN(0,1)$, and $(\theta_1, \theta_2) = (0.3, 0)$, $(0.9, 0)$, $(0.2, 0.3)$, $(0.3, 0.4)$, and $(0.7, 0.2)$.
   (c) stochastic volatility model (SV1) considered in Escanciano and Velasco (2006): $u_t = \exp(\sigma_t)\varepsilon_t$, $\sigma_t = 0.936\sigma_{t-1} + 0.32v_t$, $\varepsilon_t \sim iidN(0,1)$, $v_t \sim iidN(0,1)$, and ε_t are independent of v_t.
   (d) stochastic volatility model (SV2) considered in Charles, Darne, and Kim (2011): $u_t = \exp(0.5\sigma_t)\varepsilon_t$, $\sigma_t = 0.95\sigma_{t-1} + v_t$, $\varepsilon_t \sim iidN(0,1)$, $v_t \sim iidN(0,1)$, ε_t independent of v_t.
2. Random walk process (NULL2): $X_t = \mu + X_{t-1} + u_t$, $\mu = \mu_0 n^{-\gamma}$, $\mu_0 = 1$, $\gamma = 1$. The errors u_t follow (a) IID, (b) GARCH, (c) SV1, and (d) SV2 (all as above).
3. Random walk process (NULL3): $X_t = \mu + X_{t-1} + u_t$, $\mu = \mu_0 n^{-\gamma}$, $\mu_0 = 1$, $\gamma = 0.5$. The errors u_t follow (a) IID, (b) GARCH, (c) SV1, and (d) SV2 (all as above).

Seven different models taken from PW are chosen to generate simulated data under the alternative.

4. Explosive AR(1) model (EXP1): $X_t = \theta X_{t-1} + u_t$, $\theta = 1.01$. The errors u_t follow (a) IID, (b) GARCH, (c) SV1, and (d) SV2 (all as above).

5. Explosive AR(1) model (EXP2): $X_t = \theta X_{t-1} + u_t$, $\theta = 1.05$. The errors u_t follow (a) IID, (b) GARCH, (c) SV1, and (d) SV2 (all as above).

6. Autoregressive moving average model of order (1,1) (ARMA): $X_t = \theta_1 X_{t-1} + \theta_2 \varepsilon_{t-1} + \varepsilon_t$, with parameter values $(\theta_1, \theta_2) = (0.3, 0)$, $(0.5, 0)$, $(0.95, 0)$, $(0.3, 0.2)$, $(0.5, 0.2)$, and $(0.7, 0.2)$.

7. Exponential autoregressive model (EXAR): $X_t = \theta_1 X_{t-1} + \theta_2 X_{t-1} \exp(-0.1 |X_{t-1}|) + \varepsilon_t$, with parameter values $(\theta_1, \theta_2) = (0.6, 0.2)$, $(0.6, 0.3)$, $(0.6, 0.4)$, $(0.9, 0.2)$, $(0.9, 0.3)$, and $(0.9, 0.4)$.

8. Threshold autoregressive model of order 1 (TAR): $X_t = \theta_1 X_{t-1} \mathbf{1}(|X_{t-1}| < \theta_2) + 0.9 X_{t-1} \mathbf{1}(|X_{t-1}| \geq \theta_2) + \varepsilon_t$, with parameter values $(\theta_1, \theta_2) = (0.3, 1.0)$, $(0.5, 1.0)$, $(0.7, 1.0)$, $(0.3, 2.0)$, $(0.5, 2.0)$, and $(0.5, 2.0)$.

9. Bilinear processes: $X_t = \theta_1 X_{t-1} + \theta_2 X_{t-1} \varepsilon_{t-1} + \varepsilon_t$, with parameter values $(\theta_1, \theta_2) = (0.4, 0.1)$, $(0.4, 0.2)$, $(0.4, 0.3)$, $(0.8, 0.1)$, $(0.8, 0.2)$, and $(0.8, 0.3)$.

10. Nonlinear moving average model (NLMA): $X_t = \theta_1 X_{t-1} + \theta_2 \varepsilon_{t-1} \varepsilon_{t-2} + \varepsilon_t$, with parameter values $(\theta_1, \theta_2) = (0.4, 0.2)$, $(0.4, 0.4)$, $(0.4, 0.6)$, $(0.8, 0.2)$, $(0.8, 0.4)$, and $(0.8, 0.6)$.[1]

5.2 Results

For each experiment we set initial values to be zero and use 50,000 replications. We take $n = 100$, 250, 500, 1000 and report for each n the rejection probabilities of the tests with norminal size 0.05. The results corresponding to different nominal sizes are qualitatively similar and are not reported.

[1] PW also consider a Markov switching model and Feigenbaum maps with system noise. We found that the results are similar for these models and so they are not reported.

Table 2 reports the empirical size of the test statistics when μ is set to be zero. We find that the new tests have reasonably good size performance and are robust to both GARCH and stochastic volatility structures in the errors[2]. Table 3 reports the empirical size of the tests when $\mu = \mu_0 n^{-\gamma}$, with $\mu_0 = 1$, $\gamma = 1$. When $\mu \neq 0$ but the martingale process dominates the drift term, the empirical size properties of the tests are appropriate and seem robust to thick tails. When $\gamma = 0.5$, where $n^{-1/2}X_t$ behaves asymptotically like a Brownian motion with drift, the limit theory depends on nuisance parameters. We see in Table 4 that, using the asymptotic critical values given in Table 1, GKS_n and $GCVM_n$ have large size distortions in most cases, confirming asymptotic theory, whereas GKS_n^* and $GCVM_n^*$ still work well and the tests have good size performance in most cases, again corroborating the asymptotics. The findings for GKS and $GCVM_n$ are unsurprising because these tests are based on the PW tests which are designed for the case where $\mu = 0$, whereas the GKS_n^* and $GCVM_n^*$ tests are constructed under the explicit assumption that there may be a mild drift in the data.

Tables 5-11 report finite sample powers of the tests against various non martingale alternatives at the 5% nominal level. The tests are consistent in all of the cases we consider here and GKS_n^* and $GCVM_n^*$ generally perform much better than GKS and $GCVM_n$ tests except for one case (the mildly explosive AR(1) process with $\theta = 1.01$), and GKS and $GCVM_n$ generally perform slightly better or similar to the PW tests. We draw special attention to three aspects of Tables 5-11. First, as shown in Table 6 here, when the data are generated from an explosive AR(1) process with $\theta = 1.05$, our tests have superior power to T_n (see also Table 11 in PW for T_n). The rejection probabilities are above 90% with different GARCH specifications for all the tests when the sample size is small ($n = 100$). Our test power quickly jumps to 100% as the sample size rises whereas the test power of T_n declines as $n \to \infty$. When $n = 1,000$, for example, the rejection probabilities of T_n drop to around 50% in all cases.

Second, for the ARMA case, Table 4 in PW shows that the performance of the PW tests against near-unit root processes is not satisfactory especially when sample size is small. For example, when the AR(1) coefficient is 0.95, the PW tests basically have

[2] We also tried EGARCH models as in Fong and Ouliaris (1995) and the results are again similar.

no power when n is less than 500, while the rejection probabilities of the PW tests are 48.4% for S_n and 73.5% for T_n when $n = 1,000$. Table 7 here shows that GKS and $GCVM_n$ perform slightly better than the PW tests, but GKS_n^* and $GCVM_n^*$ both have substantially higher power in this case. When $n = 250$, the rejection probabilities are around 30%, and they reach 100% for GKS_n^* and $GCVM_n^*$ when sample size increases to $1,000$. When there is a moving average component, our tests continue to outpreform the PW tests: for example, when $(\theta_1, \theta_2) = (0.7, 0.2)$, the PW tests basically have zero power when $n = 100$, the rejection probabilities reach 53% for S_n and 85.5% for T_n when $n = 250$, and the power increases to 100% when $n = 500$; on the other hand, the rejection probabilities are 65.74% for GKS_n^* and 84.98% for $GCVM_n^*$ when $n = 100$ and these powers quickly jump to 100% as the sample size rises to 250.

Third, for other data generating processes including exponential autoregressive processes, threshold autoregressive models, bilinear processes, and nonlinear moving average models, our tests continue to perform well and outperform the PW tests. In particular, there are many cases where the performance of the PW tests is disappointing, and in these cases the comparison is more dramatic especially when the sample size is small. For example, Table 6 in PW shows that the rejection probabilities of the PW tests are around zero when $(\theta_1, \theta_2) = (0.3, 1.0)$, $(0.5, 1.0)$, $(0.7, 1.0)$ and $(0.7, 2.0)$ for TAR when sample size is $n = 100$ and the power improves only slowly as the sample size increases to 250 (less than 1% in the worst scenario when $(\theta_1, \theta_2) = (0.7, 1.0)$ and less than 50% in the best scenario when $(\theta_1, \theta_2) = (0.7, 2.0)$); by contrast, the GKS_n^* and $GCVM_n^*$ tests have effective discriminatory power in all these cases. Table 9 shows that when $n = 100$, rejection probabilities range from around 30% when $(\theta_1, \theta_2) = (0.7, 1.0)$ to 60% when $(\theta_1, \theta_2) = (0.7, 0.2)$. When the sample size increases to 250, the rejection probabilities quickly rise to 90% for $(\theta_1, \theta_2) = (0.7, 1.0)$ and 99% for $(\theta_1, \theta_2) = (0.7, 2.0)$.

6 Empirical Applications

If foreign exchange markets are efficient, nominal exchange rates are expected to follow a martingale. Numerous studies tested the martingale hypothesis in major

Table 2: Empirical Size (DGP: NULL1)

	n	IID	$GARCH$ (θ_1, θ_2)					SV	
			$(0.3, 0)$	$(0.9, 0)$	$(0.2, 0.3)$	$(0.3, .0.4)$	$(0.7, 0.2)$	$SV1$	$SV2$
GKS_n	100	0.0438	0.0436	0.0376	0.0443	0.0432	0.0388	0.0351	0.0269
	250	0.0460	0.0459	0.0403	0.0456	0.0457	0.0421	0.0400	0.0304
	500	0.0488	0.0476	0.0424	0.0484	0.0479	0.0448	0.0420	0.0321
	1000	0.0501	0.0506	0.0457	0.0500	0.0499	0.0462	0.0468	0.0353
$GCVM_n$	100	0.0460	0.0464	0.0443	0.0456	0.0442	0.0436	0.0461	0.0467
	250	0.0471	0.0479	0.0459	0.0473	0.0473	0.0469	0.0482	0.0488
	500	0.0480	0.0478	0.0473	0.0473	0.0475	0.0483	0.0480	0.0476
	1000	0.0502	0.0507	0.0505	0.0501	0.0508	0.0513	0.0493	0.0487
GKS_n^*	100	0.0428	0.0423	0.0343	0.0414	0.0398	0.0358	0.0351	0.0308
	250	0.0476	0.0448	0.0356	0.0458	0.0447	0.0375	0.0380	0.0314
	500	0.0479	0.0460	0.0374	0.0484	0.0452	0.0428	0.0410	0.0328
	1000	0.0509	0.0495	0.0426	0.0498	0.0491	0.0452	0.0428	0.0353
$GCVM_n^*$	100	0.0494	0.0509	0.0512	0.0484	0.0474	0.0490	0.0561	0.0617
	250	0.0494	0.0487	0.0514	0.0479	0.0489	0.0510	0.0547	0.0561
	500	0.0494	0.0498	0.0500	0.0504	0.0496	0.0511	0.0540	0.0555
	1000	0.0512	0.0514	0.0527	0.0522	0.0519	0.0542	0.0533	0.0548

Note: Each row gives the empirical size of the test statistics for a fixed sample size n and nominal test size is 5%. The results are based on simulations with 50,000 replications.

Table 3: Empirical Size (DGP: NULL2)

	n	IID	$GARCH$ (θ_1, θ_2)					SV	
			$(0.3, 0)$	$(0.9, 0)$	$(0.2, 0.3)$	$(0.3, 0.4)$	$(0.7, 0.2)$	$SV1$	$SV2$
GKS_n	100	0.0473	0.0461	0.0377	0.0458	0.0443	0.0381	0.0378	0.0277
	250	0.0506	0.0506	0.0419	0.0505	0.0494	0.0428	0.0411	0.0280
	500	0.0515	0.0490	0.0434	0.0505	0.0500	0.0449	0.0416	0.0328
	1000	0.0494	0.0478	0.0375	0.0478	0.0469	0.0397	0.0456	0.0352
$GCVM_n$	100	0.0534	0.0518	0.0471	0.0502	0.0485	0.0468	0.0500	0.0495
	250	0.0521	0.0521	0.0482	0.0506	0.0498	0.0485	0.0509	0.0499
	500	0.0510	0.0508	0.0487	0.0497	0.0498	0.0498	0.0496	0.0485
	1000	0.0525	0.0523	0.0514	0.0520	0.0519	0.0520	0.0504	0.0489
GKS_n^*	100	0.0420	0.0421	0.0319	0.0421	0.0412	0.0340	0.0333	0.0307
	250	0.0438	0.0444	0.0334	0.0421	0.0437	0.0386	0.0340	0.0326
	500	0.0461	0.0457	0.0401	0.0458	0.0459	0.0416	0.0405	0.0317
	1000	0.0490	0.0517	0.0398	0.0515	0.0479	0.0425	0.0441	0.0315
$GCVM_n^*$	100	0.0490	0.0503	0.0517	0.0486	0.0478	0.0497	0.0558	0.0608
	250	0.0491	0.0489	0.0514	0.0483	0.0482	0.0518	0.0549	0.0560
	500	0.0491	0.0502	0.0506	0.0502	0.0494	0.0521	0.0535	0.0556
	1000	0.0517	0.0504	0.0537	0.0521	0.0525	0.0552	0.0534	0.0551

Note: Each row gives the empirical size of the test statistics for a fixed sample size n and nominal test size is 5%. The results are based on simulations with 50,000 replications.

Table 4: Empirical Size (DGP: NULL3)

	n	IID	$GARCH$ (θ_1, θ_2)					SV	
			$(0.3, 0)$	$(0.9, 0)$	$(0.2, 0.3)$	$(0.3, 0.4)$	$(0.7, 0.2)$	$SV1$	$SV2$
GKS_n	100	0.1754	0.1385	0.0784	0.1148	0.0896	0.0733	0.1124	0.0831
	250	0.1945	0.1519	0.0809	0.1282	0.0993	0.0751	0.1040	0.0518
	500	0.2036	0.1621	0.0832	0.1356	0.1086	0.0774	0.0915	0.0448
	1000	0.2024	0.1630	0.0785	0.1369	0.1081	0.0783	0.0901	0.0432
$GCVM_n$	100	0.1930	0.1571	0.0924	0.1315	0.1043	0.0855	0.1226	0.0985
	250	0.2022	0.1645	0.0927	0.1367	0.1091	0.0864	0.1095	0.0699
	500	0.2051	0.1672	0.0924	0.1398	0.1118	0.0865	0.1017	0.0616
	1000	0.2117	0.1719	0.0941	0.1447	0.1145	0.0891	0.0994	0.0589
GKS_n^*	100	0.0361	0.0374	0.0310	0.0377	0.0375	0.0333	0.0265	0.0237
	250	0.0403	0.0400	0.0334	0.0409	0.0420	0.0380	0.0341	0.0297
	500	0.0427	0.0423	0.0367	0.0439	0.0440	0.0415	0.0364	0.0322
	1000	0.0465	0.0463	0.0409	0.0470	0.0479	0.0436	0.0414	0.0337
$GCVM_n^*$	100	0.0418	0.0435	0.0464	0.0439	0.0441	0.0463	0.0400	0.0416
	250	0.0436	0.0453	0.0482	0.0459	0.0464	0.0496	0.0456	0.0487
	500	0.0447	0.0450	0.0498	0.0464	0.0462	0.0515	0.0469	0.0520
	1000	0.0478	0.0484	0.0530	0.0491	0.0499	0.0538	0.0501	0.0542

Note: Each row gives the empirical size of the test statistics for a fixed sample size n and nominal test size is 5%. The results are based on simulations with 50,000 replications.

Table 5: Power (DGP: EXP1)

	n	IID	$GARCH$ (θ_1, θ_2)					SV	
			$(0.3, 0)$	$(0.9, 0)$	$(0.2, 0.3)$	$(0.3, 0.4)$	$(0.7, 0.2)$	$SV1$	$SV2$
GKS_n	100	0.2271	0.2238	0.2075	0.2222	0.2178	0.2057	0.2073	0.1947
	250	0.6691	0.6689	0.6472	0.6669	0.6633	0.6458	0.6319	0.5840
	500	0.9597	0.9591	0.9534	0.9589	0.9577	0.9520	0.9491	0.9242
	1000	0.9995	0.9997	0.9996	0.9997	0.9995	0.9995	0.9994	0.9990
$GCVM_n$	100	0.1964	0.1920	0.1808	0.1907	0.1854	0.1769	0.1797	0.1714
	250	0.6233	0.6215	0.5964	0.6192	0.6146	0.5932	0.5773	0.5259
	500	0.9449	0.9448	0.9365	0.9436	0.9418	0.9352	0.9311	0.8969
	1000	0.9993	0.9994	0.9992	0.9993	0.9991	0.9993	0.9991	0.9982
GKS_n^*	100	0.0300	0.0276	0.0208	0.0279	0.0258	0.0213	0.0197	0.0174
	250	0.2829	0.2809	0.2664	0.2795	0.2750	0.2636	0.2608	0.2514
	500	0.9412	0.9400	0.9314	0.9387	0.9379	0.9310	0.9241	0.8884
	1000	0.9995	0.9996	0.9995	0.9996	0.9995	0.9994	0.9994	0.9988
$GCVM_n^*$	100	0.0333	0.0330	0.0291	0.0321	0.0298	0.0281	0.0276	0.0271
	250	0.2744	0.2748	0.2687	0.2735	0.2699	0.2634	0.2666	0.2628
	500	0.9368	0.9363	0.9289	0.9347	0.9340	0.9281	0.9227	0.8898
	1000	0.9994	0.9996	0.9994	0.9995	0.9995	0.9994	0.9993	0.9987

Note: Each row gives the empirical power of the test statistics for a fixed sample size n and nominal test size is 5%. The results are based on simulations with 50,000 replications.

Table 6: Power (DGP: EXP2)

	n	IID	$GARCH$ (θ_1, θ_2)					SV	
			$(0.3, 0)$	$(0.9, 0)$	$(0.2, 0.3)$	$(0.3, 0.4)$	$(0.7, 0.2)$	$SV1$	$SV2$
GKS_n	100	0.9556	0.9537	0.9438	0.9526	0.9501	0.9419	0.9386	0.9081
	250	1.0000	1.0000	0.9999	1.0000	1.0000	0.9999	0.9999	0.9998
	500	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	1000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
$GCVM_n$	100	0.9411	0.9377	0.9250	0.9370	0.9334	0.9230	0.9190	0.8764
	250	0.9999	1.0000	0.9998	1.0000	1.0000	0.9998	0.9999	0.9996
	500	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	1000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
GKS_n^*	100	0.9350	0.9319	0.9178	0.9305	0.9259	0.9190	0.9096	0.8641
	250	1.0000	1.0000	0.9999	1.0000	1.0000	0.9999	0.9999	0.9997
	500	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	1000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
$GCVM_n^*$	100	0.9330	0.9293	0.9188	0.9290	0.9256	0.9173	0.9156	0.8820
	250	0.9999	0.9999	0.9999	0.9999	1.0000	0.9999	0.9999	0.9997
	500	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	1000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Note: Each row gives the empirical power of the test statistics for a fixed sample size n and nominal test size is 5%. The results are based on simulations with 50,000 replications.

Table 7: Power (DGP: ARMA)

	n	(θ_1, θ_2)					
		$(0.3, 0)$	$(0.5, 0)$	$(0.95, 0)$	$(03, 0.2)$	$(05.0.2)$	$(0.7, 0.2)$
GKS_n	100	0.9918	0.7878	0.0011	0.8829	0.3895	0.0348
	250	1.0000	1.0000	0.0040	1.0000	1.0000	0.7773
	500	1.0000	1.0000	0.0572	1.0000	1.0000	1.0000
	1000	1.0000	1.0000	0.5725	1.0000	1.0000	1.0000
$GCVM_n$	100	0.9996	0.9357	0.0008	0.9787	0.5974	0.0455
	250	1.0000	1.0000	0.0050	1.0000	1.0000	0.9500
	500	1.0000	1.0000	0.0729	1.0000	1.0000	1.0000
	1000	1.0000	1.0000	0.7854	1.0000	1.0000	1.0000
GKS_n^*	100	1.0000	0.9990	0.0820	0.9998	0.9854	0.6574
	250	1.0000	1.0000	0.2672	1.0000	1.0000	1.0000
	500	1.0000	1.0000	0.7443	1.0000	1.0000	1.0000
	1000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
$GCVM_n^*$	100	1.0000	1.0000	0.0965	1.0000	0.9996	0.8498
	250	1.0000	1.0000	0.3495	1.0000	1.0000	1.0000
	500	1.0000	1.0000	0.9044	1.0000	1.0000	1.0000
	1000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Note: Each row gives the empirical power of the test statistics for a fixed sample size n and nominal test size is 5%. The results are based on 50,000 replications.

Table 8: Power (DGP: EXAR)

	n	(θ_1, θ_2)					
		$(0.6, 0.2)$	$(0.6, 0.3)$	$(0.6, 0.4)$	$(0.9, 0.2)$	$(0.9.0.3)$	$(0.9, 0.4)$
GKS_n	100	0.0589	0.0095	0.0019	0.0072	0.0947	0.3059
	250	0.8262	0.3220	0.0341	0.0202	0.2212	0.4494
	500	0.9999	0.9863	0.4562	0.0834	0.6451	0.8660
	1000	1.0000	1.0000	0.9995	0.2143	0.9740	1.0000
$GCVM_n$	100	0.0950	0.0135	0.0020	0.0199	0.2247	0.4605
	250	0.9707	0.5482	0.0540	0.0384	0.3726	0.5092
	500	1.0000	0.9999	0.8278	0.1052	0.7406	0.9498
	1000	1.0000	1.0000	1.0000	0.1995	0.9734	0.9999
GKS_n^*	100	0.7269	0.4012	0.1724	0.1215	0.3374	0.6541
	250	1.0000	0.9852	0.7187	0.3061	0.8700	0.9937
	500	1.0000	1.0000	0.9995	0.4649	0.9880	1.0000
	1000	1.0000	1.0000	1.0000	0.6785	1.0000	1.0000
$GCVM_n^*$	100	0.8947	0.5749	0.2290	0.1525	0.4441	0.7984
	250	1.0000	0.9999	0.9418	0.3683	0.9527	0.9997
	500	1.0000	1.0000	1.0000	0.4552	0.9871	1.0000
	1000	1.0000	1.0000	1.0000	0.6617	0.9981	1.0000

Note: Each row gives the empirical power of the test statistics for a fixed sample size n and nominal test size is 5%. The results are based on 50,000 replications.

Table 9: Power (DGP: TAR)

	n	(θ_1, θ_2)					
		$(0.3, 1.0)$	$(0.5, 1.0)$	$(0.7, 1.0)$	$(0.3, 2.0)$	$(0.5, 2.0)$	$(0.7, 2.0)$
GKS_n	100	0.0506	0.0203	0.0065	0.7304	0.3545	0.0589
	250	0.5391	0.3184	0.1475	0.9999	0.9808	0.6231
	500	0.9895	0.9395	0.7967	1.0000	1.0000	0.9971
	1000	1.0000	1.0000	0.9999	1.0000	1.0000	1.0000
$GCVM_n$	100	0.0331	0.0155	0.0066	0.7571	0.3990	0.0709
	250	0.3866	0.2549	0.1517	0.9992	0.9764	0.6547
	500	0.9759	0.9433	0.8837	1.0000	1.0000	0.9982
	1000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
GKS_n^*	100	0.4765	0.3482	0.2541	0.9830	0.8940	0.5649
	250	0.9663	0.9170	0.8418	1.0000	1.0000	0.9903
	500	1.0000	0.9998	0.9992	1.0000	1.0000	1.0000
	1000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
$GCVM_n^*$	100	0.4646	0.3816	0.3066	0.9830	0.9061	0.6266
	250	0.9775	0.9609	0.9356	1.0000	0.9999	0.9953
	500	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	1000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Note: Each row gives the empirical power of the test statistics for a fixed sample size n and nominal test size is 5%. The results are based on 50,000 replications.

Table 10: Power (DGP: BL)

	n	(θ_1, θ_2)					
		$(0.4, 0.1)$	$(0.4, 0.2)$	$(0.4, 0.3)$	$(0.8, 0.1)$	$(0.8, 0.2)$	$(0.8, 0.3)$
GKS_n	100	0.9388	0.8999	0.8073	0.0317	0.0239	0.0123
	250	1.0000	1.0000	1.0000	0.5894	0.3740	0.1509
	500	1.0000	1.0000	1.0000	0.9992	0.9681	0.6402
	1000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9887
$GCVM_n$	100	0.9909	0.9784	0.9349	0.0474	0.0373	0.0241
	250	1.0000	1.0000	1.0000	0.8009	0.5872	0.3116
	500	1.0000	1.0000	1.0000	1.0000	0.9983	0.8971
	1000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9985
GKS_n^*	100	1.0000	1.0000	0.9984	0.5400	0.4082	0.2494
	250	1.0000	1.0000	1.0000	0.9960	0.9540	0.7254
	500	1.0000	1.0000	1.0000	1.0000	1.0000	0.9800
	1000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9994
$GCVM_n^*$	100	1.0000	1.0000	0.9999	0.7127	0.5739	0.4153
	250	1.0000	1.0000	1.0000	1.0000	0.9955	0.9181
	500	1.0000	1.0000	1.0000	1.0000	1.0000	0.9973
	1000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999

Note: Each row gives the empirical power of the test statistics for a fixed sample size n and nominal test size is 5%. The results are based on 50,000 replications.

Table 11: Power (DGP: NLMA)

	n	(θ_1, θ_2)					
		$(0.4, 0.2)$	$(0.4, 0.4)$	$(0.4, 0.6)$	$(0.8, 0.2)$	$(0.8, 0.4)$	$(0.8, 0.6)$
GKS_n	100	0.9486	0.9515	0.9447	0.0296	0.0311	0.0315
	250	1.0000	1.0000	1.0000	0.6461	0.6693	0.6942
	500	1.0000	1.0000	1.0000	0.9997	0.9997	0.9999
	1000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
$GCVM_n$	100	0.9932	0.9928	0.9909	0.0461	0.0495	0.0582
	250	1.0000	1.0000	1.0000	0.8560	0.8704	0.8753
	500	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	1000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
GKS_n^*	100	1.0000	0.9999	0.9998	0.5899	0.5991	0.6001
	250	1.0000	1.0000	1.0000	0.9986	0.9985	0.9987
	500	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	1000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
$GCVM_n^*$	100	1.0000	1.0000	1.0000	0.7633	0.7727	0.7757
	250	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	500	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	1000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Note: Each row gives the empirical power of the test statistics for a fixed sample size n and nominal test size is 5%. The results are based on 50,000 replications.

foreign exchanges rates since Meese and Rogoff (1983) showed that structural and other time series models of exchange rates generally perform poorly in terms of out-of-sample forecasting accuracy compared to a random walk model. Among others, Liu and He (1991), Fong, Koh, and Ouliaris (1997), Wright (2000), Yilmaz (2003), and Belaire-Franch and Opong (2005) use various variance ratio tests proposed originally by Lo and MacKinlay (1988) to examine the MDH in major exchange rates. Similarly, Hsieh (1988), Lobato, Nankervis, and Savin (2001), Horowitz, Lobato, Nankervis, and Savin (2006), Escanciano and Lobato (2009a, 2009b), and Charles and Kim (2011) study foreign exchange rates applying Box-Pierce type autocorrelation tests. In other work, Fong and Ouliaris (1995), Hong and Lee (2003), Kuan and Lee (2004), and Escanciano and Velasco (2006) analyze foreign exchange rates using spectral shape tests. All of the above are MDS tests and examine whether exchange rate returns are predictable based on past return information. The findings from these studies are partly mixed and sometimes inconclusive.

To complement this work using the tests developed here we examine the martingale properties of major exchanges rates that have been studied in recent work by Escanciano and Lobato (2009b). The data consist of four daily and weekly exchange rates on the Euro (EUR), Canadian dollar (CAD), British pound (GBP), and the Japanese yen (JPY) relative to the US dollar. The daily data cover the period from January 2, 2004 to August 17, 2007, and comprise a total of 908 observations. The weekly data have a total of 382 observations observed on Wednesday or on the next trading day if the Wednesday observations are missing. The nominal exchange rate data are obtained from http://www.federalreserve.gov/Releases/h10/hist.

The empirical findings are given in Table 12. The results support the martingale null hypothesis for all exchange rates at both frequencies, daily and weekly, with the exception of the weekly Japanese yen, which is rejected at the 5% level by the $GCVM_n^*$ test – so the outcome is inconclusive in this case. The MDS tests used in Escanciano and Lobato (2009b) find similar results with only a slight difference. They find that the exchange rate returns are martingale differences with the exception of the daily Euro exchange rate return, for which their test rejects the null.

Table 12: Testing the Martingale of Exchange Rates

	Daily				Weekly			
P-values	EUR	GBP	CAD	JPY	EUR	GBP	CAD	JPY
GKS_n	0.8371	0.7989	0.5667	0.4901	0.9476	0.9472	0.3636	0.1945
$GCVM_n$	0.8527	0.9257	0.7170	0.4078	0.9719	0.9676	0.3329	0.1366
GKS_n^*	0.3378	0.2465	0.9494	0.5715	0.5548	0.6188	0.9730	0.1202
$GCVM_n^*$	0.4211	0.4580	0.9264	0.4195	0.7456	0.8283	0.9852	0.0307

7 Conclusion

New martingale hypothesis tests are developed based on versions of the Kolmogorov-Smirnov and Cramér-von Mises tests extended to the regression framework. The tests are distribution free even when a drift is present in the model so there is no need to choose bandwidth parameters or obtain bootstrap versions of the tests in implementation. We develop limit theory under the null and show that test consistency against a wide class of nonlinear non-martingale processes. Simulation performance is encouraging and shows that the new tests have good finite sample properties in terms of size and power. An empirical application confirms that major exchange rates are best modeled as martingale processes, confirming much earlier research.

The present work overcomes some of the limitations of the PW tests, particularly against explosive alternatives, but also shares some of their shortcomings. In particular, the new tests focus on whether a univariate first-order Markovian process follows a martingale. To deal with more general cases, multivariate processes might be considered where martingale hypothesis tests become non pivotal and some resampling procedure is necessary, as discussed in Escanciano (2007). We may also want to mount tests to assess whether a κ-th order Markovian process follows a martingale, i.e.,

$$\mathbb{E}((X_t - \mu)|\mathcal{F}_{t-1}) = \mathbb{E}((X_t - \mu)|X_{t-1}, X_{t-2}, ..., X_{t-\kappa}),$$

for all $t \geq 1$ with some $\kappa > 1$, and other covariates might be included in the information set. The distribution-free nature of the tests continues to hold for the κ-th order Markovian process. The extension, as pointed out in PW, requires some new limit theory and is left for future work.

APPENDIX

A Proof of Lemma 2

Following a similar argument to Lemma 3.3 of PW, and using $\widehat{\sigma}_n^2 = \frac{1}{n}\sum_{t=1}^n \widehat{u}_t^2 \to_p \sigma^2$ as shown in Lemma 1, we obtain

$$J_n(a) \Rightarrow J(a), \quad J_n^*(a) \Rightarrow J^*(a)$$

as $n \to \infty$ when $\mu = \mu_0 n^{-\gamma}$ with $\gamma \geq 0.5$. When $\mu = \mu_0 n^{-\gamma}$ with $\gamma < 0.5$, the proof is straightforward and omitted.

B Proof of Theorem 3

The results follow directly from the continuous mapping theorem and the weak convergence of $J_n(a)$ to $J(a)$ and $J_n^*(a)$ to $J^*(a)$ established in Lemma 2.

C Proof of Theorem 5

We prove that when $\mu = 0$, $M_n(a) = \frac{1}{\sqrt{n}}\sum_{t=1}^n \frac{u_t}{\sigma_n} 1\left(\frac{X_{t-1}}{\sigma_n\sqrt{n}} \leq a\right) \Rightarrow M(a)$, uniformly for any $a \in \mathbb{R}$. Let $A = [-b, b]$ for some large $b > 0$. We consider the two cases $a \geq b$ and $a \leq -b$ separately.

For $a \geq b$ we have

$$\begin{aligned} M_n(a) &= \frac{1}{\sqrt{n}}\sum_{t=1}^n \frac{u_t}{\sigma_n} 1\left(\frac{X_{t-1}}{\sigma_n\sqrt{n}} \leq a\right) \\ &= \frac{1}{\sqrt{n}}\sum_{t=1}^n \frac{u_t}{\sigma_n} 1\left(\frac{X_{t-1}}{\sigma_n\sqrt{n}} \leq b\right) + \frac{1}{\sqrt{n}}\sum_{t=1}^n \frac{u_t}{\sigma_n} 1\left(b < \frac{X_{t-1}}{\sigma_n\sqrt{n}} \leq a\right). \end{aligned} \tag{34}$$

For the second term of (34), by virtue of (28) we have $1\left(\sup_t \frac{X_{t-1}}{\sigma_n\sqrt{n}} > b\right) = O_p\left(e^{-\alpha b^2}\right)$

as $b \to \infty$, so that

$$\begin{aligned}
&\mathrm{Var}\left\{ \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \frac{u_t}{\sigma_n} 1\left(b < \frac{X_{t-1}}{\sigma_n \sqrt{n}} \leq a \right) \right\} \\
&= \frac{1}{n} \sum_{t=1}^{n} P\left(b < \frac{X_{t-1}}{\sqrt{n}} \leq a \right) \\
&= O\left(P\left(\sup_{t \leq n} \frac{X_{t-1}}{\sigma_n \sqrt{n}} > b \right) \right) = O\left(e^{-\alpha b^2} \right),
\end{aligned}$$

uniformly in $a \geq b$. Hence

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \frac{u_t}{\sigma_n} 1\left(b < \frac{X_{t-1}}{\sigma_n \sqrt{n}} \leq a \right) = O_p\left(e^{-\alpha b^2} \right), \tag{35}$$

uniformly in $a \geq b$ and so is exponentially small for large b. Thus, (34) becomes

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \frac{u_t}{\sigma_n} 1\left(\frac{X_{t-1}}{\sigma_n \sqrt{n}} \leq b \right) + O_p\left(e^{-\alpha b^2} \right) \tag{36}$$

We then have

$$M_n(a) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \frac{u_t}{\sigma_n} 1\left(\frac{X_{t-1}}{\sigma_n \sqrt{n}} \leq b \right) + O_p\left(e^{-\alpha b^2} \right) = M_n(b) + O_p\left(e^{-\alpha b^2} \right), \tag{37}$$

uniformly in $a \geq b$, so that

$$|M_n(a) - M(a)| \leq |M_n(b) - M(b)| + |M(b) - M(a)| + O_p\left(e^{-\alpha b^2} \right).$$

Note that

$$\begin{aligned}
|M(a) - M(b)| &= \left| \int_0^1 1\{W(s) \leq a\}\, dW(s) - \int_0^1 1\{W(s) \leq b\}\, dW(s) \right| \\
&= O_p\left(\left| \int_0^1 1\{b < W(s) \leq a\}\, dW(s) \right| \right) \\
&= O_p\left(P\left(\sup_{t \leq 1} W(t) > b \right) \right) = O\left(e^{-\alpha b^2} \right),
\end{aligned} \tag{38}$$

uniformly in $a \geq b$. We also have, from (23), $M_n(b) \to_{a.s.} M(b)$. Thus,

$$\sup_{a \geq b} |M_n(a) - M(a)| \leq |M_n(b) - M(b)| + |M(b) - M(a)| + O_p\left(e^{-\alpha b^2} \right) = o_{a.s}(1) + O_p\left(e^{-\alpha b^2} \right),$$

which is negligibly different from zero for large enough b as $n \to \infty$. Hence

$$M_n(a) \to_p M(a)$$

uniformly in $a \geq b$ as $n \to \infty$ and $b \to \infty$.

For $a \leq -b$, we have

$$M_n(a) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \frac{u_t}{\sigma_n} 1\left(\frac{X_{t-1}}{\sigma_n \sqrt{n}} \leq a\right)$$
$$= \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \frac{u_t}{\sigma_n} 1\left(\frac{X_{t-1}}{\sigma_n \sqrt{n}} \leq -b\right) - \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \frac{u_t}{\sigma_n} 1\left(a \leq \frac{X_{t-1}}{\sigma_n \sqrt{n}} < -b\right), \quad (39)$$

and again by virtue of (28)

$$\mathrm{Var}\left\{\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \frac{u_t}{\sigma_n} 1\left(a \leq \frac{X_{t-1}}{\sigma_n \sqrt{n}} < -b\right)\right\}$$
$$= \frac{1}{n} \sum_{t=1}^{n} P\left(a \leq \frac{X_{t-1}}{\sigma_n \sqrt{n}} < -b\right)$$
$$= O\left(P\left(\inf_{t \leq n} \frac{X_{t-1}}{\sigma_n \sqrt{n}} \leq -b\right)\right) = O\left(e^{-\alpha b^2}\right),$$

uniformly in $a \leq -b$ for some $\alpha > 0$ and $b \to \infty$. Then

$$M_n(a) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \frac{u_t}{\sigma_n} 1\left(\frac{X_{t-1}}{\sigma_n \sqrt{n}} \leq -b\right) + O_p\left(e^{-\alpha b^2}\right)$$
$$= \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \frac{u_t}{\sigma_n} 1\left(\frac{X_{t-1}}{\sigma_n \sqrt{n}} \leq -b\right) + O_p\left(e^{-\alpha b^2}\right)$$
$$= M_n(-b) + O_p\left(e^{-\alpha b^2}\right) \quad (40)$$

uniformly in $a \leq -b$, so that

$$|M_n(a) - M(a)| \leq |M_n(-b) - M(-b)| + |M(-b) - M(a)| + O_p\left(e^{-\alpha b^2}\right).$$

As in (38)

$$|M(-b) - M(a)| = O_p\left(P\left(\inf_{t \leq 1} W(t) < -b\right)\right) = O\left(e^{-\alpha b^2}\right),$$

uniformly in $a \leq -b$. Using (23)

$$M_n(-b) \to_{a.s.} M(-b),$$

and it follows that

$$\sup_{a \leq -b} |M_n(a) - M(a)| \leq o_{a.s}(1) + O_p\left(e^{-\alpha b^2}\right),$$

which is negligibly different from zero for large enough b as $n \to \infty$. Hence

$$M_n(a) \to_p M(a),$$

uniformly for $a \leq -b$ as $n \to \infty$ and $b \to \infty$. It follows that $M_n(a) \to_p M(a)$ uniformly in both $a \in A$ and $a \in A^c$, i.e., on the expanded probability space

$$M_n(a) \to_p M(a) \quad \text{uniformly for any } a \in \mathbb{R}. \tag{41}$$

Hence, on the original space we have

$$M_n(a) \Rightarrow M(a) \quad \text{uniformly for any } a \in \mathbb{R}.$$

D Proof of Theorem 8

When the model is an explosive AR(1) process with $\theta > 1$, we have

$$X_t(\theta) = \frac{X_t}{\theta^t} = \sum_{j=1}^{t} \frac{u_j}{\theta^j} \to_{a.s.} X(\theta) = \sum_{j=1}^{\infty} \frac{u_j}{\theta^j}.$$

By the MGCT and under Gaussianity,

$$X(\theta) \equiv N\left(0, \frac{\sigma^2}{\theta^2 - 1}\right).$$

Under H_1, we have $X_t = \theta^t X(\theta)(1 + o_{a.s.}(1))$ so that

$$\begin{aligned} \triangle X_t &= \theta^{t-1} X(\theta)(\theta - 1)(1 + o_{a.s.}(1)), \\ \overline{\Delta X} &= \frac{1}{n}(X_n - X_0) = \frac{\theta^n}{n} X(\theta)(1 + o_{a.s.}(1)). \end{aligned}$$

Hence

$$\begin{aligned} \frac{1}{\theta^n} \sum_{t=1}^{n} \triangle X_t 1(X_{t-1} \leq x) &= \frac{1}{\theta^n} \sum_{t=1}^{n} \frac{\Delta X_t}{\theta^{t-1}} \theta^{t-1} 1\left(\frac{X_{t-1}}{\theta^{t-1}} \leq \frac{x}{\theta^{t-1}}\right) \\ &= \frac{1}{\theta^n} \sum_{t=L}^{n} \theta^{t-1} X(\theta)(\theta - 1) 1\left(X(\theta) \leq \frac{x}{\theta^{t-1}}\right)(1 + o_p(1)) + \frac{1}{\theta^n} \sum_{t=1}^{L-1} \triangle X_t 1(X_{t-1} \leq x) \\ &= \frac{1}{\theta^n} \sum_{t=L}^{n} \theta^{t-1} X(\theta)(\theta - 1) 1\left(X(\theta) \leq \frac{x}{\theta^{t-1}}\right)(1 + o_p(1)) + O_p\left(\frac{\theta^L}{\theta^n} L\right) \\ &= \frac{1}{\theta^n} \sum_{t=L}^{n} \theta^{t-1} 1\left(X(\theta) \leq \frac{x}{\theta^{t-1}}\right) X(\theta)(\theta - 1)(1 + o_p(1)) + o_p(1), \end{aligned}$$

for some L satisfying $\frac{1}{L}+\frac{L}{n}\to 0$. For all fixed x we have $\frac{x}{\theta^{t-1}}=o(1)$ as $t\geq L\to\infty$ and so we can add the frontal sum in $\frac{1}{\theta^n}\sum_{t=1}^{L-1}\theta^{t-1}1\left(X\left(\theta\right)\leq 0\right)X\left(\theta\right)\left(\theta-1\right)=o_p(1)$ without affecting the limit. Thus

$$\begin{aligned}\frac{1}{\theta^n}\sum_{t=1}^{n}\triangle X_t1(X_{t-1}\leq x)&=\frac{1}{\theta^n}\sum_{t=1}^{n}\theta^{t-1}1\left(X\left(\theta\right)\leq 0\right)X\left(\theta\right)\left(\theta-1\right)(1+o_p(1))\\&=\frac{1}{\theta^n}\frac{\theta^n-1}{\theta-1}1\left(X\left(\theta\right)\leq 0\right)X\left(\theta\right)\left(\theta-1\right)(1+o_p(1))\\&=X\left(\theta\right)1\left(X\left(\theta\right)\leq 0\right)\left(1+o_p(1)\right).\end{aligned}$$

The same argument holds for $x=a\sqrt{n}$ because $\frac{a\sqrt{n}}{\theta^L}=o(1)$ and therefore

$$\frac{1}{\theta^n}\sum_{t=1}^{n}\triangle X_t1(X_{t-1}\leq x)=X\left(\theta\right)1\left(X\left(\theta\right)\leq 0\right)\left(1+o_p(1)\right)$$

for any $x\in(-A_n,B_n)$ with $A_n,B_n=o\left(\theta^L\right)$ for some $\frac{1}{L}+\frac{L}{n}\to 0$, and

$$\begin{aligned}\sum_{t=1}^{n}\Delta X_t1(X_{t-1}\leq x)&=\theta^n\frac{1}{\theta^n}\sum_{t=1}^{n}\frac{\Delta X_t}{\theta^{t-1}}\theta^{t-1}1\left(\frac{X_{t-1}}{\theta^{t-1}}\leq\frac{x}{\theta^{t-1}}\right)\\&=\theta^n\left[\frac{1}{\theta^n}\sum_{t=1}^{n}\theta^{t-1}1\left(X\left(\theta\right)\leq\frac{x}{\theta^{t-1}}\right)X\left(\theta\right)\left(\theta-1\right)(1+o_p(1))\right]\\&=\theta^nX\left(\theta\right)\left(\theta-1\right)\left[\frac{1}{\theta^n}\sum_{t=1}^{n}\theta^{t-1}1\left(X\left(\theta\right)\leq\frac{x}{\theta^{t-1}}\right)\left(1+o_p(1)\right)\right].\end{aligned}\tag{42}$$

For the denominator, $\widehat{u}_t$ is a consistent estimator for u_t under the explosive alternative with $\widehat{u}_t=u_t+O_p\left(\frac{1}{\theta^n}\right)$ – see Phillips and Magdalinos (2008) for details. We therefore have

$$\frac{1}{n}\sum_{t=1}^{n}\widehat{u}_t^2=O_p(1),\qquad\frac{1}{n}\sum_{t=1}^{n}\widehat{u}_t^21(X_{t-1}\leq x)=O_p(1).\tag{43}$$

Hence,

$$\begin{aligned}GKS_n=\sup_{x\in\mathbb{R}}\left|\Gamma_n(x)\right|&=\sup_{x\in\mathbb{R}}\frac{\sum_{t=1}^{n}\Delta X_t1(X_{t-1}\leq x)}{\left(\sum_{t=1}^{n}\widehat{u}_t^2\right)^{1/2}}\\&=\frac{\frac{\theta^n}{\sqrt{n}}\left|X\left(\theta\right)\right|\left(\theta-1\right)\left[\frac{1}{\theta^n}\sum_{t=1}^{n}\theta^{t-1}1\left(X\left(\theta\right)\leq\infty\right)\left(1+o_p(1)\right)\right]}{\left(\frac{1}{n}\sum_{t=1}^{n}\widehat{u}_t^2\right)^{1/2}}\\&=O_p\left(\frac{\theta^n}{\sqrt{n}}\right),\end{aligned}$$

Similarly

$$
\begin{aligned}
GKS_n^* = \sup_{x\in\mathbb{R}} |\Gamma_n^*(x)| &= \sup_{x\in\mathbb{R}} \frac{\sum_{t=1}^n \left(\Delta X_t - \overline{\Delta X}\right) 1(X_{t-1} \le x)}{\left(\sum_{t=1}^n \widehat{u}_t^2\right)^{1/2}} \\
&= O_p\left(\frac{\theta^n}{\sqrt{n}}\right),
\end{aligned}
$$

as $n \to \infty$. Hence GKS_n and GKS_n^* are consistent against H_1.

Next consider $GCVM_n$ and $GCVM_n^*$

$$
\begin{aligned}
GCVM_n &= \frac{1}{n}\sum_{t=1}^n \Gamma_n^2(X_{t-1}) \\
&= \frac{\frac{1}{n}\sum_{t=1}^n \left\{\frac{1}{\sqrt{n}}\sum_{s=1}^n \Delta X_s 1\left(X_{s-1} \le X_{t-1}\right)\right\}^2}{\frac{1}{n}\sum_{t=1}^n \widehat{u}_t^2}.
\end{aligned}
$$

Note that $X(\theta)$ may be positive or negative and for $s, t > L$, $1\left(X_{s-1} \le X_{t-1}\right)$ iff $1\left(\theta^{s-1}X(\theta) \le \theta^{t-1}X(\theta)\right)$ iff $1\left(s < t \text{ and } X(\theta) > 0\right)$ or $1\left(s > t \text{ and } X(\theta) < 0\right)$. As in (42), we find that for $t \ge L$

$$
\begin{aligned}
A_n &\equiv \frac{1}{\sqrt{n}}\sum_{s=1}^n \Delta X_s I(X_{s-1} \le X_{t-1}) \\
&= \frac{\theta^n}{\sqrt{n}}\frac{1}{\theta^n}\sum_{s=1}^n \frac{\Delta X_s}{\theta^{s-1}}\theta^{s-1} 1\left(\frac{X_{s-1}}{\theta^{s-1}} \le \frac{X_{t-1}}{\theta^{s-1}}\right) \\
&= \frac{\theta^n}{\sqrt{n}} X(\theta)(\theta-1)\left[\frac{1}{\theta^n}\sum_{s=L}^n \theta^{s-1} 1\left(X(\theta) \le \frac{X_{t-1}}{\theta^{s-1}}\right)(1+o_p(1)) + o_p(1)\right] \\
&= \frac{\theta^n}{\sqrt{n}} X(\theta)(\theta-1)\left[\frac{1}{\theta^n}\sum_{s=L}^n \theta^{s-1} 1\left(X(\theta) \le \frac{X_{t-1}}{\theta^{t-1}}\frac{\theta^{t-1}}{\theta^{s-1}}\right)(1+o_p(1)) + o_p(1)\right] \\
&= \frac{\theta^n}{\sqrt{n}} X(\theta)(\theta-1)\left[\frac{1}{\theta^n}\sum_{s=L}^n \theta^{s-1}\left(1\left\{s < t \text{ and } X(\theta) > 0\right\}\right)(1+o_p(1)) + o_p(1)\right] \\
&+ \frac{\theta^n}{\sqrt{n}} X(\theta)(\theta-1)\left[\frac{1}{\theta^n}\sum_{s=L}^n \theta^{s-1}\left(1\left\{s > t \text{ and } X(\theta) < 0\right\}\right)(1+o_p(1)) + o_p(1)\right].
\end{aligned}
$$

Now we evaluate $GCVM_n$ for $s > t$ and $X(\theta) < 0$

$$\begin{aligned} GCVM_n &= \frac{\frac{1}{n}\sum_{t=1}^{n}\left\{\frac{1}{\sqrt{n}}\sum_{s=1}^{n}\Delta X_s 1\left(X_{s-1} \leq X_{t-1}\right)\right\}^2}{\frac{1}{n}\sum_{t=1}^{n}\widehat{u}_t^2} \\ &= O_p\left(\frac{1}{n^2}\sum_{t=1}^{n}\left\{\sum_{s=t+1}^{n}\theta^{s-1}\left|X(\theta)\right|(\theta-1)1\left(X(\theta)<0\right)\left(1+o_p(1)\right)\right\}^2\right) \\ &= O_p\left(\frac{\theta^{2n}}{n}\right), \end{aligned}$$

so that $GCVM_n$ is divergent for $1\left(X(\theta) < 0\right)$.

Evaluating $GCVM_n$ for $s \leq t$ and $X(\theta) > 0$ we have

$$\begin{aligned} GCVM_n &= \frac{\frac{1}{n}\sum_{t=1}^{n}\left\{\frac{1}{\sqrt{n}}\sum_{s=1}^{n}\Delta X_s 1\left(X_{s-1} \leq X_{t-1}\right)\right\}^2}{\frac{1}{n}\sum_{t=1}^{n}\widehat{u}_t^2} \\ &= O_p\left(\frac{1}{n^2}\sum_{t=1}^{n}\left[\left\{\sum_{s=1}^{t}\theta^{s-1}\left|X(\theta)\right|(\theta-1)1\left(X(\theta)>0\right)\left(1+o_p(1)\right)\right\}^2\right]\right) \\ &= O_p\left(\frac{\theta^{2n}}{n}\right) \end{aligned}$$

Thus, $GCVM_n$ is divergent for $1\left(X(\theta) > 0\right)$. It follows that the test $GCVM_n$ is consistent against explosive AR(1) alternatives. In a similar way we have $GCVM_n^* = O_p\left(\frac{\theta^{2n}}{n}\right)$ and the test $GCVM_n^*$ is consistent against explosive AR(1) alternatives.

References

Aldous, D.J., 1986. Self-intersections of 1-dimensional random walks. Probability Theory and Related Fields, 72, 559–587.

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

Belaire-Franch, J., Opong, K.K., 2005. Some evidence of random walk behaviour of euro exchange rates using ranks and signs. Journal of Banking and Finance 29, 1631–1643.

Billingsley, P., 1995. Probability and measure, 3rd ed. John Wiley: New York.

Bierens, H.J., 1984. Model specification testing of time series regressions. Journal of Econometrics 26, 323-353.

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

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

Box, G., Pierce, D., 1970. Distribution of residual autocorrelations in autoregressive integrated moving average time series models. Journal of American Statistical Association 65, 1509–1527.

Charles, A., Darne O., Kim, J.H., 2011. Small sample properties of alternative tests for martingale difference hypothesis. Economics Letters 110, 151–154.

Chen, W.W., Deo, R.S., 2006. The variance ratio statistic at large horizons. Econometric Theory 22, 206–234.

Choi, I., 1999. Testing the random walk hypothesis for real exchange rates. Journal of Applied Econometrics 14, 293–308.

Chow, K.V., Denning, K.C., 1993. A simple multiple variance ratio test. Journal of Econometrics 58, 385–401.

Chang, Y., Park, J.Y., 2010. Endogeneity in nonlinear regressions with integrated time series. Econometric Reviews 30, 51-87.

de Jong, R.M., 1996. The Bierens' tests under data dependence. Journal of Econometrics 72, 1–32.

Deo, R.S., 2000. Spectral tests of the martingale hypothesis under conditional heteroskedasticity. Journal of Econometrics 99, 291–315.

Domínguez, M.A., Lobato, I.N., 2003. A consistent test for the martingale difference hypothesis. Econometric Reviews 22, 351–377.

Durlauf, S., 1991. Spectral-based test for the martingale hypothesis. Journal of Econometrics 50, 1–19.

Escanciano, J.C., 2007. Weak convergence of non-stationary multivariate marked processes with applications to martingale testing. Journal of Multivariate Analysis 98, 1321-1336.

Escanciano, J.C., Lobato, I.N., 2009a. An automatic portmanteau test for serial correlation. Journal of Econometrics 151, 140–149.

Escanciano, J.C., Lobato, I.N., 2009b. Testing the martingale hypothesis. In: Patterson, K., Mills, T.C. (Eds.), Palgrave Hand-book of Econometrics. Palgrave MacMillan, New York, pp. 972–1003.

Escanciano, J.C., Velasco, C., 2006. Generalized spectral tests for the martingale difference hypothesis. Journal of Econometrics 134, 151–185.

Fama, E., 1970. Efficient capital markets: a review of theory and empirical work. Journal of Finance 25, 383–417.

Fong,W.M., Ouliaris, S., 1995. Spectral tests of the martingale hypothesis for exchange rates. Journal of Applied Econometrics 10, 255–271.

Fong, W.M., Koh, S.K., Ouliaris, S., 1997. Joint variance-ratio tests of the martingale hypothesis for exchange rates. Journal of Business and Economic Statistics 15, 51–59.

Hall, R.E., 1978. Stochastic implications of the life cycle-permanent income hypothesis: theory and evidence. Journal of Political Economy 86, 971–987.

Hong, Y., 1999. Hypothesis testing in time series via the empirical characteristic function: a generalized spectral density approach. Journal of the American Statistical Association 94, 1201–1220.

Hong, Y., Lee, T.H., 2003. Inference on predictability of foreign exchange rate changes via generalized spectrum and nonlinear time series models. Review of Economics and Statistics 85, 1048–1062.

Hong, Y., Lee, Y.J., 2005. Generalized spectral tests for conditional mean models in time series with conditional heteroskedasticity of unknown form. Review of Economic Studies 72, 499–541.

Horowitz, J.L., Lobato, I.N., Nankervis, J.C., Savin, N.E., 2006. Bootstrapping the Box–Pierce Q test: a robust test of uncorrelatedness. Journal of Econometrics 133, 841–862.

Hsieh, D.,1988. The statistical property of daily foreign exchange rates: 1974–1983. Journal of International Economics 24,129–145.

Kim, J.H., 2006. Wild bootstrapping variance ratio tests. Economics Letters 92,38–43.

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

Kuan, C.-M., Lee, W., 2004. A new test of the martingale difference hypothesis. Studies in Nonlinear Dynamics and Econometrics 8, 1–24.

Liu, C.Y., He, J., 1991. A variance ratio test of random walks in foreign exchange rate. Journal of Finance 46, 773–785.

Ljung, G.M., Box, G.E.P., 1978. On a measure of lack of fit in time series models. Biometrika 65, 297–303.

Lo, A.W., MacKinlay, A.C., 1988. Stock market prices do not follow random walk: evidence from a simple specification test. The Review of Financial Studies 1, 41–66.

Lobato, I.N., Nankervis, J.C., Savin, N.E., 2001. Testing for autocorrelation using a modified Box–Pierce Q test. International Economic Review 42, 187–205.

Lobato, I.N., Nankervis, J.C., Savin, N.E., 2002. Testing for zero autocorrelation in the presence of statistical dependence. Econometric Theory 18, 730–743.

Meese, R.A., Rogoff, K., 1983. Empirical exchange rate models of the seventies: do they fit out of sample? Journal of International Economics 14, 3–24.

Nankervis, J.C., Savin, N.E., 2010. Testing for serial correlation: generalized Andrews–Ploberger tests. Journal of Economic and Business Statistics 28, 246–255.

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

Park, J.Y., Phillips, P.C.B., 2000. Nonstationary binary choice. Econometrica, 68, 1249-1280.

Park, J.Y., Phillips, P.C.B., 2001. Nonlinear regressions with integrated time series. Econometrica 69, 117–161.

Park, J.Y., Whang, Y.J., 2005. Testing for the martingale hypothesis. Studies in Nonlinear Dynamics and Econometrics 9, 1-30.

Phillips, P.C.B., 1987. Time series regression with a unit root. Econometrica 55, 277–301.

Phillips, P.C.B., 2009. Local limit theory and spurious nonparametric regression. Econometric Theory 25, 1466-1467.

Phillips, P.C.B., Magdalinos, T., 2008. Limit theory for explosively cointegrated systems. Econometric Theory 24, 865-887.

Phillips, P.C.B., Magdalinos, T., 2009. Limit theory for cointegrated systems with moderately integrated and moderately explosive regressors. Econometric Theory 25, 482-526.

Phillips, P.C.B., Shi, S., Yu, J., 2013. Specification sensitivities in right-tailed unit root testing for explosive behavior, Oxford Bulletin of Economics and Statistics, forthcoming.

Revuz, D., Yor, M., 1999. Continuous Martingale and Brownian Motion, 3rd ed. Springer–Verlag: New York.

Shi, X., Phillips, P.C.B., 2012. Nonlinear cointegrating regression under weak identification. Econometric Theory 28, 509–547.

Siegmund, D., 1986. Boundary crossing probabilities and statistical applications. Annals of Statistics 14, 361-404.

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

Wang, L., Potzelberger, K., 1997. Boundary crossing probability for Brownian motion and general boundaries. Journal of Applied Probability, 34, 54-65.

Wang, Q., Phillips, P.C.B., 2012. A specification test for nonlinear nonstationary models. Annals of Statistics 40, 727–758.

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

Wright, J.H., 2000. Alternative variance-ratio tests using ranks and signs. Journal of Business and Economic Statistics 18, 1–9.

Yilmaz, K., 2003. Martingale property of exchange rates and central bank intervention. Journal of Business and Economic Statistics 21, 383–395.