TESTING THE MARTINGALE HYPOTHESIS

BY

Peter C. B. Phillips and Sainan Jin

COWLES FOUNDATION PAPER NO. 1443

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

2014

http://cowles.econ.yale.edu/
Testing the Martingale Hypothesis

Peter C. B. Phillips
Yale University, New Haven, CT 06520, and University of Auckland, Auckland 1010, New Zealand
(peter.phillips@yale.edu)

Sainan Jin
School of Economics, Singapore Management University, Singapore 178903 (snjin@smu.edu.sg)

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 article develops limit theory for the tests under the null and shows that the tests are consistent against a wide class of nonlinear, nonmartingale processes. Simulations show that the tests have good finite sample properties in comparison with other tests particularly under conditional heteroscedasticity 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.

KEY WORDS: Brownian functional; Cramér-von Mises test; Exchange rates; Explosive process; Kolmogorov–Smirnov test.

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 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

\[
E(X_t|\mathcal{F}_{t-1}) = X_{t-1}
\]

almost surely (a.s.). Let \( I_t = \{X_1, X_{t-1}, X_{t-2}, \ldots\} \). The natural choice for \( \mathcal{F}_t \) is the \( \sigma \)-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 focus on tests of the martingale difference hypothesis (MDH), vis-à-vis

\[
E(\Delta X_t|\Omega_{t-1}) = \mu
\]

for some unknown \( \mu \in \mathbb{R} \) and where \( \Delta \) is the difference operator, \( \Delta X_t = X_t - X_{t-1} \) and \( \Omega_t = \{\Delta X_t, \Delta X_{t-1}, \Delta X_{t-2}, \ldots\} \). The MDH is slightly modified in this formulation to allow for an unknown mean for \( \Delta X_t \) and information set based on the differences. Typically, the information set includes the infinite past history of the series and \( I_t \) and \( \Omega_t \) 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 has 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 nonmartingales with zero correlations. Nankervis and Savin (2010) used another approach based on generalizing the Andrews–Ploberger tests and found 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. 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 Lo- bato (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 previous tests are martingale difference tests. Techni- cally, 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 al- ternatives including many interesting nonlinear nonmartingale processes such as exponential and threshold autoregressive processes, Markov switching and chaotic processes (possibly with stochastic noise), and some other nonstationary processes. How- ever, the PW tests appear to be inconsistent against explosive processes such as the simple AR(1) with explosive coefficient \( \theta \). 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 article 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 article 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) later) 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 a recent work by Phillips, Shi, and Yu (2014; 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 domi- nating component of the series, thereby justifying this type of formulation.

Martingales with a weak drift in the null satisfy

\[
\mathbb{E}((X_t - \mu)|I_{t-1}) = X_{t-1},
\]

(3)

or, equivalently, the empirically appealing and convenient form

\[
\mathbb{E}((\Delta X_t - \mu)|I_{t-1}) = 0,
\]

(4)

with \( \mu = \mu_0 n^{-\gamma} \). The magnitude of the drift depends on the sample size \( n \) and a localizing exponent parameter \( \gamma \). Estimation of \( \gamma \) is discussed in PSY (2014). When \( \gamma \) is positive, the drift term is small relative to a linear trend. We develop asymptotic theory for tests of (4) over different ranges of \( \gamma \). 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 article. 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, nonmartingale 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\(_n^*\) and GCVM\(_n^*\), while the GKS\(_n^*\) 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 are around 30%, and they reach 100% when sample size is small. But for large samples, the power of the PW \( T_n \) test declines to 50% when \( n = 1000 \) 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 \( n = 1000 \); the GKS\(_n^*\) 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 1000.

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\(_n\) 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 pro- posed by DL. The former test the MH null (1), while the DL
HYPOTHESES AND TESTS

The martingale null is formulated as

\[ X_t = \mu + \theta X_{t-1} + u_t, \quad \text{with } \theta = 1, \]  

(5)

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

\[ \mathbb{E}(\Delta X_t | I_{t-1}) = 0. \]  

(6)

The intercept is defined as \( \mu = \mu_0 e^{-\gamma} \) so the deterministic drift in \( X_t \) is \( \mu t \) with drift \( \mu_0 \) of the same magnitude, which depends on the specific model and the localizing parameter \( \gamma \). When \( \gamma = 0 \), the drift produces a linear trend \( \mu t \) component in \( X_t \) under the null. When \( \gamma > 0 \) the drift is small relative to the stochastic trend \( n^{-1/2} X_t \) 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, p. 213, Theorem 16.10 (iii))

\[ \mathbb{E}(\Delta X_t - \mu | I_{t-1}) = 0 \quad \text{a.s. iff } \mathbb{E}(\Delta X_t - \mu) W(I_{t-1}) = 0, \]  

(7)

where \( W(\cdot) \) represents any \( F_{t-1} \) measurable weighting function. A convenient choice of weight function is the indicator function \( 1(\cdot) \), 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

\[ \mathbb{E}(X_t | F_{t-1}) = \mathbb{E}(X_t | F_{t-1}), \]  

and thus

\[ \mathbb{E}(\Delta X_t | F_{t-1}) = 0 \quad \text{a.s. iff } \mathbb{E}(\Delta X_t | F_{t-1}) I(X_{t-1} \leq x) = 0, \]  

(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

\[ \mathbb{E}(\Delta X_t | F_{t-1}) = \mathbb{E}(\Delta X_t | X_{t-1}, X_{t-2}, \ldots X_{t-p}, 
Z_{t-1}, Z_{t-2}, \ldots, Z_{t-k}) \]  

(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

\[ \mathbb{E}(\Delta X_t | F_{t-1}) = \mathbb{E}(\Delta X_t | X_{t-1}, \Delta X_{t-2}, \ldots, \Delta X_{t-p}, W_{t-1}, W_{t-2}, \ldots, W_{t-k}) \]  

(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 article.

Under the null, \((u_t, F_t)\) is a martingale difference sequence assumed to satisfy the following conditions, as in PW.

\textbf{Assumption 1.}

(a) \[ \frac{1}{n} \sum_{t=1}^{n} \mathbb{E}(u_t^2 | F_{t-1}) \to_p \sigma^2 > 0, \]  

and

(b) \[ \sup_{i \geq 1} \mathbb{E}(u_t^2 | F_{t-1}) < K \quad \text{a.s. for some constant } K < \infty. \]

Part (a) allows the innovation sequence to be conditionally heteroscedastic 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 heteroscedasticity even when (b) may not apply (e.g., GARCH errors) as in PW.

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

**Lemma 1.** Let Assumption 1 hold. Under the null, we have

\[
\sigma_n^2 = \frac{1}{n} \sum_{t=1}^{n} u_t^2 \rightarrow p \sigma^2, \quad \hat{\sigma}_n^2 = \frac{1}{n} \sum_{t=1}^{n} \hat{\epsilon}_t^2 \rightarrow p \sigma^2 \quad \text{as} \quad n \rightarrow \infty.
\]

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

\[
\Gamma_n(x) = \frac{\sum_{t=1}^{n} \Delta X_t 1(X_{t-1} \leq x)}{\left( \frac{1}{n} \sum_{t=1}^{n} \hat{\epsilon}_t^2 \right)^{1/2}},
\]

and

\[
\Gamma_n^*(x) = \frac{\sum_{t=1}^{n} \Delta X_t - \Delta \bar{X} 1(X_{t-1} \leq x)}{\left( \frac{1}{n} \sum_{t=1}^{n} \hat{\epsilon}_t^2 \right)^{1/2}},
\]

where \( \Delta \bar{X} = \frac{1}{n} \sum_{t=1}^{n} \Delta X_t \).

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

\[
\frac{\sum_{t=1}^{n} (\Delta X_t - \Delta \bar{X}) 1(X_{t-1} \leq x)}{\left( \frac{1}{n} \sum_{t=1}^{n} \hat{\epsilon}_t^2 1(X_{t-1} \leq x) \right)^{1/2}},
\]
tend to have size distortions when the errors exhibit strong conditional heteroscedasticity. For this reason, we normalize by \( \left( \frac{1}{n} \sum_{t=1}^{n} \hat{\epsilon}_t^2 \right)^{1/2} \).

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

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

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

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

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

# 3. ASYMPTOTIC DISTRIBUTION UNDER H₀

## 3.1 Asymptotic Behavior of GKS₁ₙ and GKSⁿₙ

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

\[
W_n(r) \rightharpoonup W(r), \quad r \in [0,1],
\]

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

\[
J_n(a) \rightharpoonup J(a), \quad \text{and} \quad J_n^*(a) \rightharpoonup J^*(a),
\]

where

\[
J(a) = \int_{0}^{1} \left\{ W(s) \leq a \right\} dW(s),
\]

\[
J^*(a) = \int_{0}^{1} \left\{ W(s) \leq a \right\} dB(s),
\]

when \( \gamma > 0.5 \); and

\[
J(a) = \int_{0}^{1} \left\{ W(s) + \frac{\mu_0}{\sigma} s \leq a \right\} d \left( W(s) + \frac{\mu_0}{\sigma} s \right),
\]

\[
J^*(a) = \int_{0}^{1} \left\{ W(s) + \frac{\mu_0}{\sigma} s \leq a \right\} dB(s),
\]

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

\[
\frac{1}{n^{0.5-\gamma}} \Gamma_n(bn^{1-\gamma}) = \frac{1}{n^{0.5-\gamma}} \Gamma_n(x) \rightharpoonup \int_{0}^{1} 1 \{ s \leq b_1 \} ds,
\]

\[
\Gamma_n^*(bn^{1-\gamma}) = \Gamma_n^*(x) = \int_{0}^{1} 1 \{ s \leq b_1 \} dB(s),
\]
for \( x = bn^{1−γ} \), where \( b_1 = \frac{b}{\sqrt{n}} \), and \( B(s) = W(s) − sW(1) \) 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 \( γ > 0.5 \) 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 \( μ = μ_0n^{−γ} \) and \( γ > 0.5 \), as \( n \to \infty \)

\[
\text{GKS}_n \to \sup_{a \in \mathbb{R}} J(a), \quad \text{GKS}_n^* \to \sup_{a \in \mathbb{R}} J^*(a)
\]

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

Remark 4. When \( μ = μ_0n^{−γ} \) with \( γ > 0.5 \), 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 \( μ = μ_0n^{−γ} \) with \( γ > 0.5 \) 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 \( μ = 0 \) or has the form \( μ_0n^{−γ} \) with \( γ > 0.5 \). When \( γ = 0.5 \), \( \text{GKS}_n^* \) and \( \text{GCVM}_n^* \) using the critical values from Table 1 still work very well and these tests have good size performance.

Remark 5. The asymptotic distributions of \( \text{GKS}_n \) and \( \text{GKS}_n^* \) are easily obtained when \( μ = μ_0n^{−γ} \) with \( γ ≤ 0.5 \), but as discussed in Remark 3, these results are not of direct interest in the present article. When \( μ = μ_0n^{−γ} \) with \( γ = 0.5 \), as \( n \to \infty \) we have

\[
\text{GKS}_n \to \sup_{a \in \mathbb{R}} J(a), \quad \text{GKS}_n^* \to \sup_{a \in \mathbb{R}} J^*(a),
\]

where \( J(a) \) and \( J^*(a) \) are given in (15). When \( μ = μ_0n^{−γ} \) with \( γ < 0.5 \), as \( n \to \infty \) we have

\[
\frac{1}{n^{0.5−γ}} \sup_{b \in \mathbb{R}} |\Gamma_n(bn^{1−γ})| = \sup_{x \in \mathbb{R}} |\Gamma_n(x)| \to_r \sup_{b_1 \in \mathbb{R}} \frac{μ_0}{σ} \int_0^1 1 \{ s ≤ b_1 \} ds = \max \left[ \frac{μ_0}{σ}, 0 \right], \quad \text{sup}_{b \in \mathbb{R}} |\Gamma_n^*(bn^{1−γ})| = \sup_{x \in \mathbb{R}} |\Gamma_n^*(x)| \to_r \sup_{b_1 \in \mathbb{R}} \int_0^1 1 \{ s ≤ b_1 \} dB(s).
\]

Remark 6. PW set \( μ = 0 \) 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

\[
Q_n(x) = \frac{1}{\sqrt{n}} \sum_{t=1}^n \frac{ΔX_t}{σ_n} I\left( \frac{X_{t−1}}{σ_n} ≤ x \right) = \frac{1}{\sqrt{n}} \sum_{t=1}^n \frac{Δ\xi_t}{σ_n} I\left( \frac{ξ_{t−1}}{σ_n} ≤ x \right).
\]

Theorem 3.4 in PW shows that

\[
S_n \Rightarrow S = \sup_{a \in \mathbb{R}} |M(a)|,
\]

where

\[
M(a) = \int_0^1 1 \{ W(s) ≤ a \} dW(s).
\]

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

\[
T_n \Rightarrow T = \int_0^1 M^2(W(r))dr.
\]

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

\[
M_n(a) = \frac{1}{\sqrt{n}} \sum_{t=1}^n u_t I\left( \frac{X_{t−1}}{σ_n\sqrt{n}} ≤ a \right) = \frac{1}{\sqrt{n}} \sum_{t=1}^n u_t I\left( \frac{ξ_{t−1}}{σ_n\sqrt{n}} ≤ a \right),
\]

an argument that requires the ‘integrand’ to be \( F_{r−1} \)-measurable. For a given fixed \( a \), the quantity \( I\left( \frac{ξ_{t−1}}{σ_n\sqrt{n}} ≤ a \right) \) is \( F_{r−1} \)-measurable. However, the limit process given in (18) involves (by virtue of plugging in the stochastic process argument \( a = W(r) \))

\[
M(W(r)) = \int_0^1 1 \{ W(s) ≤ W(r) \} dW(s).
\]

Here, the integrand \( 1 \{ W(s) ≤ W(r) \} \) is \( F_r \)-measurable only when \( 0 ≤ r ≤ s \). 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 next section. A separate argument that takes account of the composite nature of \( M(W(r)) \) is needed to show the weak convergence of

\[
T_n = \frac{1}{n} \sum_{t=1}^n Q_n^2(X_{t−1}) = \frac{1}{n} \sum_{t=1}^n Q_n^2(ξ_{t−1}) \Rightarrow \int_0^1 M^2(W(r))dr,
\]

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

\[
\int_0^1 M^2(W(r))dr = \int_{−∞}^∞ M^2(s)L(1,s)ds.
\]

(20)
However, as is apparent from the definition (17)
\[ M(a) = M_{W}(a) = \int_{0}^{1} 1 \{ W(s) \leq a \} dW(s), \] (21)
the functional \( M \) itself also depends on the same stochastic process \( \{ W(s) \}^{\infty}_{t=0} \) 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\(_{n}^{*}\)

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 \( t = 1, \ldots, n \), and

(ii) \( \sup_{t \geq 1} E(u_{t}^{2}|F_{t-1}) = K \) a.s. for some constant \( K < \infty \).

Part (i) introduces a more restrictive condition on the innovations than Assumption 1(i) to make use of results on uniform convergence to stochastic integrals as discussed next. 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 use but we do not undertake those extensions in the present article. As demonstrated in the simulations reported next, 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, \( g(x, a) = \kappa(\lambda, a)h(x, a) + R(x, \lambda, a) \), 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 \( a, a' \in A \),
\[
\sup_{\lambda \geq 1} \kappa^{-1}(\lambda, a)g(\lambda x, a) - \kappa^{-1}(\lambda, a')g(\lambda x, a') \\
\leq v(x)|a - a'|,
\]
where function \( v(x) \) is symmetric and bounded, \( v(|x|) \) is increasing in \( |x| \), with \( E v^{2}(|W(1)| + c) < \infty \) for some \( c > 0 \), and there exists an \( a \in A \), such that
\[
\limsup_{n \to \infty} \frac{1}{n} \sum_{t=1}^{n} E h^{2} \left( \frac{\hat{\xi}_{t-1}}{\sqrt{n}}, a \right) < \infty.
\]

A useful result concerning uniform weak convergence to stochastic integrals involving nonlinear homogeneous functions \( g(\xi_{t-1}, a) \) that includes functions like \( 1(\frac{\xi_{t-1}}{\sqrt{n}} \leq a) \) of integrated processes is stated in Lemma 3 next. The result is based on Lemma 5.2 of Shi and Phillips (2012) and holds under weak conditions that apply here.

**Lemma 3.** Let Assumptions 1b and 2 hold. Then, uniformly in \( a \in A \), and in a suitably expanded probability space
\[
n^{-1/2} \kappa^{-1}(n^{1/2}, a) \sum_{t=1}^{n} g(\xi_{t-1}, a) u_{t} \to_{p} \sigma \int h(W, a) dW
\]
under the null hypothesis.

**Remark 7.** From Lemma 3, it is straightforward to show that
\[
\frac{1}{\sqrt{n}} \sum_{t=1}^{n} u_{t} \mathbb{1} \left( \frac{\hat{\xi}_{t-1}}{\sigma_{n} \sqrt{n}} \leq a \right) \to_{p} \int 1(W(s) \leq a) dW(s),
\]
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
\[
W_{n}(r) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} u_{t} \mathbb{1} \Rightarrow W(r), \tag{24}
\]
and
\[
M_{n}(a) \Rightarrow M(a), \tag{25}
\]
for any \( a \in \mathbb{R} \). Let
\[
S_{nt} = \sup_{r \leq t} W_{n}(r), \quad \text{and} \quad S_{t} = \sup_{r \leq t} W(r).
\]
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
\[
W_{n}(r) \to_{a.s.} W(r), \quad \text{and} \quad S_{nt} \to_{a.s.} S_{t}, \tag{27}
\]
from which it follows that
\[
P(S_{nt} \geq b) \to P(S_{t} \geq b)
\]
for some large \( b > 0 \). Note that (e.g., Proposition 3.7 in Revuz and Yor 1999)
\[
P \left( S_{t} \geq b \right) = 2P \left( W_{t} \geq b \right) = P(\left| W_{t} \right| \geq b),
\]
where \( W_{t} = BM(1) \). Using the boundary crossing probability
\[
P \left( W_{t} \geq b \right. \text{ for some } t \in [0,1]) = O(e^{-2b^{2}}),
\]
as \( b \to \infty \) (see, e.g., Siegmund 1986; Wang and Potzelberger 1997), we therefore have
\[
P(S_{nt} \geq b) \to P \left( S_{t} \geq b \right) = O(e^{-a b^{2}}) \tag{28}
\]
for some \( a > 0 \) and \( b \to \infty \).
The boundary crossing probability in (28) facilitates the development of the following uniform results.

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

\[
M_n(a) \Rightarrow M(a),
\]

where

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

and

\[
Q_n(x) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\Delta x_i}{\sigma_i} 1(x_{i+1} \leq x) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\Delta x_i}{\sigma_i} 1(\frac{x_i}{\sigma_i} \leq x) \text{ as defined in } PW.
\]

It follows that when \( \mu = \mu \omega^{n^{-\gamma}} \) with \( \gamma \geq 0.5 \)

\[
J_n(a) \Rightarrow J(a), \text{ and } J^*_n(a) \Rightarrow J^*(a),
\]

uniformly in \( a \in \mathbb{R} \). Thus,

\[
J_n(X_n(r)) \Rightarrow J(W(r)), \text{ and } J^*_n(X_n(r)) \Rightarrow J^*(W(r))
\]

as required.

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

\[
M(W(r)) = \int_0^1 1 \{ W(s) \leq W(r) \} dW(s)
\]

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 \( a \in \mathbb{R} \)

\[
\int_0^1 1 \{ W(s) \leq a \} dW(s)
= \frac{1}{2} L_W(t, a) - [(W(t) - a)^- - (W(0) - a)^-]
= \frac{1}{2} L_W(t, a) - [(W(t) - a)^- - (-a)^-].
\]

It follows that we can write

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

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

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

In this expression, \( L_W(1, W(r)) \) is the local time that the process \( \{ W(s) : s \in [0, 1] \} \) spends at \( W(r) \), that is, 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 \( \int_0^1 J^*(W(r))dr \) are well defined.

Using Theorem 5, we can establish the limit theory for GCVM and GCVM*.

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

\[
\text{GCVM}_n \Rightarrow \int_0^1 J^2(W(r))dr, \text{ GCVM}^*_n \Rightarrow \int_0^1 J^{*2}(W(r))dr,
\]

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 \( A \subset R \), \( \frac{1}{n} \sum_{t=1}^{n} P_n(A) \to P(A) \) as \( n \to \infty \) where \( P \) is a probability measure on \( R \) and \( P_n \) is the distribution of \( X_{nt} \) for \( 1 \leq t \leq n, n \geq 1 \). Also, \( \frac{1}{n} \sum_{t=1}^{n} E|\Delta X_{nt}|^q \to H(x) \) 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

\[
\sup_{x \in \mathbb{R}} \left| \frac{1}{n} \sum_{t=1}^{n} I(\Delta X_{nt} \leq x) \right| \to 0.
\]

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

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

\[
\text{GKS}_n, \text{GCVM}_n, \text{GKS}^*_n, \text{GCVM}^*_n \to \infty
\]

as \( n \to \infty \).

The proof of Lemma 4 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, and the divergence rates of the test statistics are as follows:

\[
\text{GKS}_n = O_p(n^{1/2}), \text{ GKS}^*_n = O_p(n^{1/2}),
\]

\[
\text{GCVM}_n = O_p(n), \text{ GCVM}^*_n = O_p(n)
\]

(33)
following the proof of Theorem 4.4 in Park and Whang (2005). 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 the test power of \( T_n \) against a simple explosive alternative (with \( AR(1) \) processes).

Let \( H_1 : \theta > 1 \) declines as \( n \to \infty \) when \( \theta = 1.05 \) but increases when \( \theta = 1.01 \). By contrast, tests based on \( GKS_n \), \( GCVM_n \), \( GKS_n^* \), 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 \( \epsilon_t \), 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 \( n \to \infty \)

\[
GKS_n, \ GCVM_n, \ GKS_n^*, \ GCVM_n^* \to \infty.
\]

**Remark 8.** The proof is given in the Appendix. Under explosive alternatives, we find that \( \sum_{i=1}^{n} \Delta X_t I(X_{t-1} \leq x) = O_p(\theta^n), \sum_{i=1}^{n} \Delta X_t I(X_{t-1} \leq x) = O_p(1), \) and thus

\[
\Gamma_n(x) = \frac{\sum_{i=1}^{n} \Delta X_t I(X_{t-1} \leq x)}{(\sum_{i=1}^{n} \Delta X_t)^{1/2}} = O \left( \frac{\theta^n}{n} \right),
\]

\[
\Gamma_n^*(x) = \frac{\sum_{i=1}^{n} \Delta X_t - \Delta \bar{X}_t I(X_{t-1} \leq x)}{(\sum_{i=1}^{n} \Delta X_t)^{1/2}} = O \left( \frac{\theta^n}{n} \right),
\]

\[
GKS_n = O_p \left( \frac{\theta^n}{n} \right), \ GKS_n^* = O_p \left( \frac{\theta^n}{n} \right),
\]

\[
GCVM_n = O_p \left( \frac{\theta^{2n}}{n} \right), \ GCVM_n^* = O_p \left( \frac{\theta^{2n}}{n} \right),
\]

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 \( \sigma_n^2 = \frac{1}{n} \sum_{i=1}^{n} (\Delta X_t)^2 \) as defined in PW does not equal \( \frac{1}{n} \sum_{i=1}^{n} u_t^2 \). Following similar arguments to those in the proof, it is easy to show that \( \sigma_n^2 = \frac{1}{n} \sum_{i=1}^{n} (\Delta X_t)^2 - O_p \left( \frac{\theta^n}{n} \right) \), and thus \( Q_n(x) = \frac{1}{n} \sum_{i=1}^{n} \frac{\Delta X_t I(X_{t-1} \leq x)}{\sigma_n^2} = O_p(1) \). Thus, the test powers based on \( Q_n(x) \) are not consistent against explosive \( AR(1) \) processes.

PW also look at the nonmartingale unit root process generated by \( \Delta X_t = u_t \) where \( u_t \) is serially correlated. Chang and Park (2011) showed that

\[
\frac{1}{\sqrt{n}} \sum_{i=1}^{n} u_t I(X_t \leq x) \to d \int_0^1 \left[ W(r) \leq 0 \right] dW(r) + L(1, 0),
\]

(34)

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 (34). 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 (34), 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 nonmartingales when there is some dependence in the innovation sequence.

### Table 1. Asymptotic critical values of test statistics

<table>
<thead>
<tr>
<th>Sig. level</th>
<th>0.99</th>
<th>0.95</th>
<th>0.90</th>
<th>0.10</th>
<th>0.05</th>
<th>0.01</th>
</tr>
</thead>
<tbody>
<tr>
<td>GKS_n</td>
<td>0.5930</td>
<td>0.7465</td>
<td>0.8462</td>
<td>2.0877</td>
<td>2.3519</td>
<td>2.8885</td>
</tr>
<tr>
<td>GCVM_n</td>
<td>0.0551</td>
<td>0.0999</td>
<td>0.1422</td>
<td>1.6118</td>
<td>2.1250</td>
<td>3.3667</td>
</tr>
<tr>
<td>GKS_n^*</td>
<td>0.5164</td>
<td>0.6481</td>
<td>0.7341</td>
<td>1.5326</td>
<td>1.6716</td>
<td>1.9300</td>
</tr>
<tr>
<td>GCVM_n^*</td>
<td>0.0426</td>
<td>0.0752</td>
<td>0.1066</td>
<td>0.7619</td>
<td>0.9214</td>
<td>1.2871</td>
</tr>
</tbody>
</table>

**NOTE:** Asymptotic critical values of the test statistics are computed from simulations with 50,000 replications, iid \( N(0, 1) \) errors and \( n = 1000 \).

### 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 \( GKS_n \), \( GCVM_n \), \( GKS_n^* \), and \( GCVM_n^* \) are free of nuisance parameters when \( \mu = \mu_0(n - 0^+) \) with \( \gamma > 0 \) and these distributions are readily obtained by simulation. Table 1 gives the asymptotic critical values of the test statistics \( GKS_n \), \( GCVM_n \), \( GKS_n^* \), and \( GCVM_n^* \) when \( \mu = \mu_0(n - 0^+) \) with \( \gamma > 0 \). These critical values were generated for \( n = 1000 \) 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 tests are not martingale tests in that case. The simulation experiments described next consider cases where \( \mu = 0 \) and \( \mu 
eq 0 \) under the null. For \( \mu 
eq 0 \), 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); \( u_t \sim iid N(0, 1) \).
   (b) GARCH errors as used in PW (GARCH); \( u_t = \sigma_t \varepsilon_t, \sigma_t^2 = 1 + \theta_1 u_{t-1}^2 + \theta_2 \sigma_{t-1}^2, \varepsilon_t \sim iid N(0, 1), \) 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): \( u_t = \exp(\sigma_t) \varepsilon_t, \sigma_t = 0.936 \sigma_{t-1} + 0.32 \varepsilon_t, \varepsilon_t \sim iid N(0, 1), \) \( v_t \sim iid N(0, 1), \) \( \theta_1, \theta_2 \) independent of \( u_t \).
   (d) Stochastic volatility model (SV2) considered in Charles, Darne, and Kim (2011): \( u_t = \exp(0.5 \sigma_t) \varepsilon_t, \sigma_t = 0.95 \sigma_{t-1} + v_t, \varepsilon_t \sim iid N(0, 1), \) \( v_t \sim iid N(0, 1), \) \( \theta_1 \) independent of \( v_t \).

2. Random walk process (NULL2): \( X_t = \mu + X_{t-1} + u_t, \mu = \mu_0(n - 0^+) \), \( \gamma = 1 \). The errors \( u_t \) follow (a) IID, (b) GARCH, (c) SV1, and (d) SV2 (all as earlier).
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 earlier).

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 earlier).

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 earlier).

6. Autoregressive moving average model of order (1,1) (ARMA): $X_t = \theta_1 X_{t-1} + \theta_2 \epsilon_{t-1} + \epsilon_t$, 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): $X_t = \theta_1 X_{t-1} + \theta_2 \epsilon_{t-1} \exp(-0.1|X_{t-1}|) + \epsilon_t$, 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): $X_t = \begin{cases} \theta_1 X_{t-1} | X_{t-1} | < \theta_2 & + \frac{0.9 X_{t-1} I (|X_{t-1}| \geq \theta_2) + \epsilon_t,} {\theta_2} \\ \text{with parameter values } (\theta_1, \theta_2) = (0.3, 1.0), (0.5, 1.0), (0.7, 1.0), & (0.3, 2.0), (0.5, 2.0), \text{ and } (0.5, 2.0). \end{cases}$

9. Bilinear processes: $X_t = \theta_1 X_{t-1} + \theta_2 \epsilon_{t-1} \epsilon_{t-2} + \epsilon_t$, 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): $X_t = \theta_1 X_{t-1} + \theta_2 \epsilon_{t-1} \epsilon_{t-2} + \epsilon_t$, 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 $(0.8, 0.6)$.

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 nominal size 0.05. The results corresponding to different nominal sizes are qualitatively similar and are not reported.

Table 2 reports the empirical size of the test statistics when $\mu$ 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. 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$\gamma$ and GCVM$\gamma$ have large size distortions in most cases, confirming asymptotic theory, whereas GKS$\gamma^*$ and GCVM$\gamma^*$ still work well and the tests have good size performance in most cases, again corroborating the asymptotics. The findings for GKS$\gamma$ and GCVM$\gamma$ are unsurprising because these tests are based on the PW tests which are designed for the case where $\mu = 0$, whereas the GKS$\gamma^*$ and GCVM$\gamma^*$ 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 nonmartingale alternatives at the 5% nominal level. The tests are consistent in all of the cases we consider here and GKS$\gamma^*$ and GCVM$\gamma^*$ generally perform much better than GKS$\gamma$ and GCVM$\gamma$ tests except for one case (the mildly explosive AR(1) process with $\theta = 1.01$), and GKS$\gamma$ and GCVM$\gamma$ generally perform slightly better or similar to the PW tests.

1PW 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.

2We also tried EGARCH models as in Fong and Ouliaris (1995) and the results are again similar.
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 $n = 1000$, 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 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 $n = 1000$. Table 7 here shows that GKS and GCVM perform slightly better than the PW tests, but GKS* and GCVM* both have substantially higher power in this case. When $n = 250$, the rejection probabilities are around 30%, and they reach 100% for GKS* and GCVM* when sample size increases to 1000. When there is a moving average component, our tests continue to outperform 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

<table>
<thead>
<tr>
<th>$n$</th>
<th>IID</th>
<th>GARCH ($\theta_1, \theta_2$)</th>
<th>SV</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>(0.3, 0)</td>
<td>(0.9, 0)</td>
</tr>
<tr>
<td>GKS</td>
<td>100</td>
<td>0.0473</td>
<td>0.0461</td>
</tr>
<tr>
<td></td>
<td>250</td>
<td>0.0506</td>
<td>0.0506</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>0.0515</td>
<td>0.0490</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>0.0494</td>
<td>0.0478</td>
</tr>
<tr>
<td>GCVM</td>
<td>100</td>
<td>0.0534</td>
<td>0.0518</td>
</tr>
<tr>
<td></td>
<td>250</td>
<td>0.0521</td>
<td>0.0521</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>0.0510</td>
<td>0.0508</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>0.0525</td>
<td>0.0523</td>
</tr>
<tr>
<td>GKS*</td>
<td>100</td>
<td>0.0420</td>
<td>0.0421</td>
</tr>
<tr>
<td></td>
<td>250</td>
<td>0.0438</td>
<td>0.0444</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>0.0461</td>
<td>0.0457</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>0.0490</td>
<td>0.0517</td>
</tr>
<tr>
<td>GCVM*</td>
<td>100</td>
<td>0.0490</td>
<td>0.0503</td>
</tr>
<tr>
<td></td>
<td>250</td>
<td>0.0491</td>
<td>0.0489</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>0.0491</td>
<td>0.0502</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>0.0517</td>
<td>0.0504</td>
</tr>
</tbody>
</table>

Table 3. Empirical size (DGP: NULL2)

Table 4. Empirical size (DGP: NULL3)

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.
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.

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

### Table 5. Power (DGP: EXP1)

<table>
<thead>
<tr>
<th>( n )</th>
<th>IID</th>
<th>((0.3, 0))</th>
<th>((0.9, 0))</th>
<th>((0.2, 0.3))</th>
<th>((0.3, 0.4))</th>
<th>((0.7, 0.2))</th>
<th>( SV )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( GKS_n )</td>
<td>100</td>
<td>0.2271</td>
<td>0.2238</td>
<td>0.2075</td>
<td>0.2222</td>
<td>0.2178</td>
<td>0.2057</td>
</tr>
<tr>
<td></td>
<td>250</td>
<td>0.6691</td>
<td>0.6689</td>
<td>0.6472</td>
<td>0.6669</td>
<td>0.6633</td>
<td>0.6458</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>0.9597</td>
<td>0.9591</td>
<td>0.9534</td>
<td>0.9589</td>
<td>0.9577</td>
<td>0.9520</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>0.9995</td>
<td>0.9997</td>
<td>0.9996</td>
<td>0.9997</td>
<td>0.9995</td>
<td>0.9995</td>
</tr>
<tr>
<td>( GCVM_n )</td>
<td>100</td>
<td>0.1964</td>
<td>0.1920</td>
<td>0.1808</td>
<td>0.1907</td>
<td>0.1854</td>
<td>0.1769</td>
</tr>
<tr>
<td></td>
<td>250</td>
<td>0.6233</td>
<td>0.6215</td>
<td>0.5964</td>
<td>0.6192</td>
<td>0.6146</td>
<td>0.5932</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>0.9449</td>
<td>0.9448</td>
<td>0.9365</td>
<td>0.9436</td>
<td>0.9418</td>
<td>0.9352</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>0.9993</td>
<td>0.9994</td>
<td>0.9992</td>
<td>0.9993</td>
<td>0.9991</td>
<td>0.9993</td>
</tr>
<tr>
<td>( GKS_n^* )</td>
<td>100</td>
<td>0.0300</td>
<td>0.0276</td>
<td>0.0208</td>
<td>0.0279</td>
<td>0.0258</td>
<td>0.0213</td>
</tr>
<tr>
<td></td>
<td>250</td>
<td>0.2829</td>
<td>0.2809</td>
<td>0.2664</td>
<td>0.2795</td>
<td>0.2750</td>
<td>0.2636</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>0.9412</td>
<td>0.9400</td>
<td>0.9314</td>
<td>0.9387</td>
<td>0.9379</td>
<td>0.9310</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>0.9995</td>
<td>0.9996</td>
<td>0.9995</td>
<td>0.9996</td>
<td>0.9995</td>
<td>0.9994</td>
</tr>
<tr>
<td>( GCVM_n^* )</td>
<td>100</td>
<td>0.0333</td>
<td>0.0330</td>
<td>0.0291</td>
<td>0.0321</td>
<td>0.0298</td>
<td>0.0281</td>
</tr>
<tr>
<td></td>
<td>250</td>
<td>0.2744</td>
<td>0.2748</td>
<td>0.2687</td>
<td>0.2735</td>
<td>0.2699</td>
<td>0.2634</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>0.9368</td>
<td>0.9363</td>
<td>0.9289</td>
<td>0.9347</td>
<td>0.9340</td>
<td>0.9281</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>0.9994</td>
<td>0.9996</td>
<td>0.9994</td>
<td>0.9995</td>
<td>0.9995</td>
<td>0.9994</td>
</tr>
</tbody>
</table>

### Table 6. Power (DGP: EXP2)

<table>
<thead>
<tr>
<th>( n )</th>
<th>IID</th>
<th>((0.3, 0))</th>
<th>((0.9, 0))</th>
<th>((0.2, 0.3))</th>
<th>((0.3, 0.4))</th>
<th>((0.7, 0.2))</th>
<th>( SV )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( GKS_n )</td>
<td>100</td>
<td>0.9556</td>
<td>0.9537</td>
<td>0.9438</td>
<td>0.9526</td>
<td>0.9501</td>
<td>0.9419</td>
</tr>
<tr>
<td></td>
<td>250</td>
<td>1.0000</td>
<td>1.0000</td>
<td>0.9999</td>
<td>1.0000</td>
<td>1.0000</td>
<td>0.9999</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td>( GCVM_n )</td>
<td>100</td>
<td>0.9411</td>
<td>0.9377</td>
<td>0.9250</td>
<td>0.9370</td>
<td>0.9334</td>
<td>0.9230</td>
</tr>
<tr>
<td></td>
<td>250</td>
<td>0.9999</td>
<td>1.0000</td>
<td>0.9998</td>
<td>1.0000</td>
<td>1.0000</td>
<td>0.9998</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td>( GKS_n^* )</td>
<td>100</td>
<td>0.9350</td>
<td>0.9319</td>
<td>0.9178</td>
<td>0.9305</td>
<td>0.9259</td>
<td>0.9190</td>
</tr>
<tr>
<td></td>
<td>250</td>
<td>1.0000</td>
<td>1.0000</td>
<td>0.9999</td>
<td>1.0000</td>
<td>1.0000</td>
<td>0.9999</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td>( GCVM_n^* )</td>
<td>100</td>
<td>0.9330</td>
<td>0.9293</td>
<td>0.9188</td>
<td>0.9290</td>
<td>0.9256</td>
<td>0.9173</td>
</tr>
<tr>
<td></td>
<td>250</td>
<td>0.9999</td>
<td>0.9999</td>
<td>0.9999</td>
<td>0.9999</td>
<td>1.0000</td>
<td>0.9999</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
</tbody>
</table>

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)

<table>
<thead>
<tr>
<th>$n$</th>
<th>$(0.3, 0)$</th>
<th>$(0.5, 0)$</th>
<th>$(0.95, 0)$</th>
<th>$(0.3, 0.2)$</th>
<th>$(0.5, 0.2)$</th>
<th>$(0.7, 0.2)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>GKS</td>
<td>0.0589</td>
<td>0.0095</td>
<td>0.0019</td>
<td>0.0072</td>
<td>0.0947</td>
<td>0.3059</td>
</tr>
<tr>
<td>100</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>250</td>
<td>0.8262</td>
<td>0.3220</td>
<td>0.0341</td>
<td>0.0202</td>
<td>0.2212</td>
<td>0.4494</td>
</tr>
<tr>
<td>500</td>
<td>0.9999</td>
<td>0.9863</td>
<td>0.4562</td>
<td>0.0834</td>
<td>0.6451</td>
<td>0.8660</td>
</tr>
<tr>
<td>1000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>0.9995</td>
<td>0.2143</td>
<td>0.9740</td>
<td>1.0000</td>
</tr>
<tr>
<td>GCVM</td>
<td>0.0950</td>
<td>0.0135</td>
<td>0.0020</td>
<td>0.0199</td>
<td>0.2247</td>
<td>0.4605</td>
</tr>
<tr>
<td>100</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>250</td>
<td>0.9707</td>
<td>0.5482</td>
<td>0.0540</td>
<td>0.0384</td>
<td>0.3726</td>
<td>0.5092</td>
</tr>
<tr>
<td>500</td>
<td>1.0000</td>
<td>0.9999</td>
<td>0.8278</td>
<td>0.1052</td>
<td>0.7406</td>
<td>0.9498</td>
</tr>
<tr>
<td>1000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>0.9995</td>
<td>0.9734</td>
<td>0.9999</td>
</tr>
<tr>
<td>$GKS^{*}$</td>
<td>0.7269</td>
<td>0.4012</td>
<td>0.1724</td>
<td>0.1215</td>
<td>0.3374</td>
<td>0.6541</td>
</tr>
<tr>
<td>100</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>250</td>
<td>1.0000</td>
<td>0.9852</td>
<td>0.7187</td>
<td>0.3061</td>
<td>0.8700</td>
<td>0.9937</td>
</tr>
<tr>
<td>500</td>
<td>1.0000</td>
<td>1.0000</td>
<td>0.9995</td>
<td>0.4649</td>
<td>0.9880</td>
<td>1.0000</td>
</tr>
<tr>
<td>1000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>0.6785</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td>$GCVM^{*}$</td>
<td>0.8947</td>
<td>0.5749</td>
<td>0.2290</td>
<td>0.1525</td>
<td>0.4441</td>
<td>0.7984</td>
</tr>
<tr>
<td>100</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>250</td>
<td>1.0000</td>
<td>0.9999</td>
<td>0.9418</td>
<td>0.3683</td>
<td>0.9527</td>
<td>0.9997</td>
</tr>
<tr>
<td>500</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>0.4552</td>
<td>0.9871</td>
<td>1.0000</td>
</tr>
<tr>
<td>1000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>0.6617</td>
<td>0.9981</td>
<td>1.0000</td>
</tr>
</tbody>
</table>

Table 8. Power (DGP: EXAR)

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

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 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 foreign exchange 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) used 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 et al. (2006), Escanciano and Lobato (2009a, 2009b), and Charles,
Darne, and Kim (2011) studied 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) analyzed foreign exchange rates using spectral shape tests. All of these 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 exchange 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.

### Table 9. Power (DGP: TAR)

<table>
<thead>
<tr>
<th>n</th>
<th>(0.3, 1.0)</th>
<th>(0.5, 1.0)</th>
<th>(0.7, 1.0)</th>
<th>(0.3, 2.0)</th>
<th>(0.5, 2.0)</th>
<th>(0.7, 2.0)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GKS</td>
<td>100</td>
<td>0.0506</td>
<td>0.0203</td>
<td>0.0065</td>
<td>0.7304</td>
<td>0.3545</td>
</tr>
<tr>
<td></td>
<td>250</td>
<td>0.5391</td>
<td>0.3184</td>
<td>0.1475</td>
<td>0.9999</td>
<td>0.9808</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>0.9895</td>
<td>0.9395</td>
<td>0.7967</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>0.9999</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td>GCMV</td>
<td>100</td>
<td>0.0331</td>
<td>0.0155</td>
<td>0.0066</td>
<td>0.7571</td>
<td>0.3992</td>
</tr>
<tr>
<td></td>
<td>250</td>
<td>0.3866</td>
<td>0.2549</td>
<td>0.1517</td>
<td>0.9992</td>
<td>0.9764</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>0.9759</td>
<td>0.9433</td>
<td>0.8837</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
</tbody>
</table>

**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)

<table>
<thead>
<tr>
<th>n</th>
<th>(0.4, 0.1)</th>
<th>(0.4, 0.2)</th>
<th>(0.4, 0.3)</th>
<th>(0.8, 0.1)</th>
<th>(0.8, 0.2)</th>
<th>(0.8, 0.3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GKS</td>
<td>100</td>
<td>0.9388</td>
<td>0.8999</td>
<td>0.8073</td>
<td>0.0317</td>
<td>0.0239</td>
</tr>
<tr>
<td></td>
<td>250</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>0.5894</td>
<td>0.3740</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>0.9992</td>
<td>0.9681</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td>GCMV</td>
<td>100</td>
<td>0.9909</td>
<td>0.9784</td>
<td>0.9349</td>
<td>0.0474</td>
<td>0.0373</td>
</tr>
<tr>
<td></td>
<td>250</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>0.8009</td>
<td>0.5872</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>0.9983</td>
<td>0.9681</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td>GKS*</td>
<td>100</td>
<td>1.0000</td>
<td>1.0000</td>
<td>0.9984</td>
<td>0.5400</td>
<td>0.4082</td>
</tr>
<tr>
<td></td>
<td>250</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>0.9986</td>
<td>0.9540</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td>GCMV*</td>
<td>100</td>
<td>1.0000</td>
<td>1.0000</td>
<td>0.9999</td>
<td>0.7127</td>
<td>0.5739</td>
</tr>
<tr>
<td></td>
<td>250</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>0.9955</td>
<td>0.9181</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
</tbody>
</table>

**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)

<table>
<thead>
<tr>
<th>n</th>
<th>(0.4, 0.2)</th>
<th>(0.4, 0.4)</th>
<th>(0.4, 0.6)</th>
<th>(0.8, 0.2)</th>
<th>(0.8, 0.4)</th>
<th>(0.8, 0.6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GKS_n</td>
<td>100</td>
<td>0.9486</td>
<td>0.9515</td>
<td>0.9447</td>
<td>0.0296</td>
<td>0.0311</td>
</tr>
<tr>
<td></td>
<td>250</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>0.6461</td>
<td>0.6693</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>0.9997</td>
<td>0.9997</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td>GCVM_n</td>
<td>100</td>
<td>0.9932</td>
<td>0.9928</td>
<td>0.9909</td>
<td>0.0461</td>
<td>0.0495</td>
</tr>
<tr>
<td></td>
<td>250</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>0.8560</td>
<td>0.8704</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td>GKS*_n</td>
<td>100</td>
<td>1.0000</td>
<td>0.9999</td>
<td>0.9998</td>
<td>0.5899</td>
<td>0.5991</td>
</tr>
<tr>
<td></td>
<td>250</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>0.9986</td>
<td>0.9985</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td>GCVM*_n</td>
<td>100</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>0.7633</td>
<td>0.7727</td>
</tr>
<tr>
<td></td>
<td>250</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
</tbody>
</table>

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 12. Testing the martingale of exchange rates

<table>
<thead>
<tr>
<th></th>
<th>EUR</th>
<th>GBP</th>
<th>CAD</th>
<th>JPY</th>
<th>EUR</th>
<th>GBP</th>
<th>CAD</th>
<th>JPY</th>
</tr>
</thead>
<tbody>
<tr>
<td>GKS_n</td>
<td>0.8371</td>
<td>0.7989</td>
<td>0.5667</td>
<td>0.4901</td>
<td>0.9476</td>
<td>0.9472</td>
<td>0.3636</td>
<td>0.1945</td>
</tr>
<tr>
<td>GCVM_n</td>
<td>0.8527</td>
<td>0.9257</td>
<td>0.7170</td>
<td>0.4078</td>
<td>0.9719</td>
<td>0.9676</td>
<td>0.3329</td>
<td>0.1366</td>
</tr>
<tr>
<td>GKS*_n</td>
<td>0.3378</td>
<td>0.2465</td>
<td>0.9494</td>
<td>0.5715</td>
<td>0.5548</td>
<td>0.6188</td>
<td>0.9730</td>
<td>0.1202</td>
</tr>
<tr>
<td>GCVM*_n</td>
<td>0.4211</td>
<td>0.4580</td>
<td>0.9264</td>
<td>0.4195</td>
<td>0.7456</td>
<td>0.8283</td>
<td>0.9852</td>
<td>0.0307</td>
</tr>
</tbody>
</table>

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* test—so the outcome is inconclusive in this case. The MDS tests used by Escanciano and Lobato (2009b) found similar results with only a slight difference. They found 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.

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 nonmartingale 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 nonpivotal and some resampling procedure is necessary, as discussed in Escanciano (2007). We may also want to mount tests to assess whether a κth-order Markovian process follows a martingale, that is,

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

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 $\kappa$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 $\hat{\sigma}_n^2 = \frac{1}{2} \sum_{i=1}^{\sigma} \tilde{a}_i^2 \overset{p}{\rightarrow} \sigma^2$ 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 \( \mu = \mu_0n^{-\gamma} \) with \( \gamma \geq 0.5 \). When \( \mu = \mu_0n^{-\gamma} \) 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 \( \mu = 0 \), \( M_n(a) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \frac{u_t}{\sigma_n} (X_{t-1} \leq a) \) for \( a \in \mathbb{R} \). 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

\[
M_n(a) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \frac{u_t}{\sigma_n} (X_{t-1} \leq a) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \frac{u_t}{\sigma_n} (X_{t-1} \leq b) + \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \frac{u_t}{\sigma_n} (b < X_{t-1} \leq a).
\]

(C.1)

For the second term of (C.1), by virtue of (28) we have \( \sup \frac{X_{t-1}}{\sigma_n} > b = O_p(e^{-ab^2}) \) as \( b \to \infty \), so that

\[
\text{var} \left\{ \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \frac{u_t}{\sigma_n} \left( b < \frac{X_{t-1}}{\sigma_n} \leq a \right) \right\} = \frac{1}{n} \sum_{t=1}^{n} P \left( b < \frac{X_{t-1}}{\sigma_n} \leq a \right) = O \left( P \left( \sup \frac{X_{t-1}}{\sigma_n} > b \right) \right) = O(e^{-ab^2}),
\]

uniformly in \( a \geq b \). Hence,

\[
\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \frac{u_t}{\sigma_n} \left( b < \frac{X_{t-1}}{\sigma_n} \leq a \right) = O_p(e^{-ab^2}),
\]

(C.2)

uniformly in \( a \geq b \) and so is exponentially small for large \( b \). Thus, (C.1) becomes

\[
\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \frac{u_t}{\sigma_n} \left( \frac{X_{t-1}}{\sigma_n} \leq b \right) + O_p(e^{-ab^2}).
\]

(C.3)

We then have

\[
M_n(a) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \frac{u_t}{\sigma_n} \left( \frac{X_{t-1}}{\sigma_n} \leq b \right) + O_p(e^{-ab^2}) = M_n(b) + O_p(e^{-ab^2}).
\]

(C.4)

uniformly in \( a \geq b \), so that

\[
|M_n(a) - M(a)| \leq |M_n(b) - M(b)| + |M(b) - M(a)| + O_p(e^{-ab^2}).
\]

Note that

\[
|M(a) - M(b)|
\]

\[
= \left| \int_{0}^{1} \left\{ \int_{0}^{1} 1 \{ W(s) \leq a \} dW(s) - \int_{0}^{1} 1 \{ W(s) \leq b \} dW(s) \right\} dt \right|
\]

\[
= O_p \left( \left| \int_{0}^{1} \left\{ \int_{0}^{1} 1 \{ b < W(s) \leq a \} dW(s) \right\} dt \right| \right)
\]

\[
= O_p \left( P \left( \sup_{t \leq 1} W(t) > b \right) \right) = O(e^{-ab^2}),
\]

(C.5)

uniformly in \( a \geq b \). We also have, from (23), \( M_n(b) \to_p M(b) \). Thus,

\[
\sup_{a \geq b} |M_n(a) - M(a)| \leq |M_n(b) - M(b)| + |M(b) - M(a)| + O_p(e^{-ab^2})
\]

\[
= o_{a \to \infty} (1) + O_p(e^{-ab^2}),
\]

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

\[
M_n(a) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \frac{u_t}{\sigma_n} \left( \frac{X_{t-1}}{\sigma_n} \leq a \right)
\]

\[
= \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \frac{u_t}{\sigma_n} \left( \frac{X_{t-1}}{\sigma_n} \leq -b \right)
\]

\[
= - \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \frac{u_t}{\sigma_n} \left( a \leq \frac{X_{t-1}}{\sigma_n} \leq -b \right),
\]

(C.6)

and again by virtue of (28)

\[
\text{var} \left\{ \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \frac{u_t}{\sigma_n} \left( a \leq \frac{X_{t-1}}{\sigma_n} \leq -b \right) \right\} = \frac{1}{n} \sum_{t=1}^{n} P \left( a \leq \frac{X_{t-1}}{\sigma_n} \leq -b \right) = O \left( P \left( \inf_{t \leq \sigma_n} \frac{X_{t-1}}{\sigma_n} \leq -b \right) \right) = O(e^{-ab^2}),
\]

uniformly in \( a \leq -b \) for some \( a > 0 \) and \( b \to \infty \). Then,

\[
M_n(a) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \frac{u_t}{\sigma_n} \left( \frac{X_{t-1}}{\sigma_n} \leq -b \right) + O_p(e^{-ab^2})
\]

\[
= \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \frac{u_t}{\sigma_n} \left( \frac{X_{t-1}}{\sigma_n} \leq -b \right) + O_p(e^{-ab^2})
\]

\[
= M_n(-b) + O_p(e^{-ab^2})
\]

(C.7)

uniformly in \( a \leq -b \), so that

\[
|M_n(a) - M(a)| \leq |M_n(-b) - M(-b)| + |M(-b) - M(a)| + O_p(e^{-ab^2})
\]

As in (C.5)

\[
|M(-b) - M(a)| = O_p \left( P \left( \inf_{t \leq \sigma_n} W(t) < -b \right) \right) = O(e^{-ab^2}),
\]

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

\[
M_n(-b) \to_p M(-b),
\]
and it follows that

$$\sup_{a \leq -b} |M_n(a) - M(a)| \leq o_{a,a} (1) + O_p(e^{-n\phi})$$

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 $a \in A'$, that is, on the expanded probability space

$$M_n(a) \to_p M(a) \quad \text{uniformly for any } a \in \mathbb{R}. \quad (C.8)$$

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

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

By the martingale convergence theorem and under Gaussianity,

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

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

$$\Delta X_i = \theta^{i-1} X(\theta)(\theta - 1)(1 + o_{a.s.}(1)),$$

$$\overline{\Delta X} = \frac{1}{n} \sum_{i=1}^{n} \Delta X_i = \frac{\theta^n}{n} X(\theta)(1 + o_{a.s.}(1)).$$

Hence,

$$\frac{1}{\theta^n} \sum_{i=1}^{n} \Delta X_i 1(X_{i-1} \leq x) = \frac{1}{\theta^n} \sum_{i=1}^{n} \theta^{i-1} \Delta X_i 1\left(\frac{X_{i-1}}{\theta^{i-1}} \leq \frac{x}{\theta^{i-1}}\right)$$

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

$$\frac{1}{\theta^n} \sum_{i=1}^{n} \Delta X_i 1(X_{i-1} \leq x) = \frac{1}{\theta^n} \sum_{i=1}^{n} \theta^{i-1} X(\theta)(\theta - 1)(1 + o_{p}(1))$$

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

$$\sum_{i=1}^{n} \Delta X_i 1(X_{i-1} \leq x) = X(\theta) 1(X(\theta) \leq 0)(1 + o_{p}(1)).$$

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

$$\frac{1}{\theta^n} \sum_{i=1}^{n} \Delta X_i 1(X_{i-1} \leq x) = X(\theta)(1 + o_{p}(1))$$

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

$$\sum_{i=1}^{n} \Delta X_i 1(X_{i-1} \leq x) = X(\theta)(\theta - 1)(1 + o_{p}(1))$$

so that

$$\sum_{i=1}^{n} \Delta X_i 1\left(\frac{X_{i-1}}{\theta^{i-1}} \leq \frac{x}{\theta^{i-1}}\right)$$

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

$$\sum_{i=1}^{n} \Delta X_i 1\left(\frac{X_{i-1}}{\theta^{i-1}} \leq \frac{x}{\theta^{i-1}}\right) = X(\theta)(\theta - 1)(1 + o_{p}(1)),$$

for all fixed $x$ we have $\frac{\theta^n}{\theta^{i-1}} = o(1)$ as $i \geq L \to \infty$ and so we can add the frontal sum in $\frac{1}{\theta^n} \sum_{i=1}^{L-1} \theta^{i-1} X(\theta)(\theta - 1) = o(1)$ without affecting the limit. Thus,
Magdalinos (2008) for details. We, therefore, have

\[
\frac{1}{n} \sum_{t=1}^{n} \tilde{a}_{t} = O_p(1), \quad \frac{1}{n} \sum_{t=1}^{n} \tilde{a}_{t}^{2} I(X_{t-1} \leq x) = O_p(1).
\]

Hence,

\[
GKS_{n} = \sup_{x \in \mathbb{R}} |\Gamma_{1}(x)| = \sup_{x \in \mathbb{R}} \frac{\sum_{t=1}^{n} \Delta X_{t} I(X_{t-1} \leq x)}{\left(\sum_{t=1}^{n} \tilde{a}_{t}^{2}\right)^{1/2}} = O_p \left(\frac{\hat{\theta}_{n}}{\sqrt{n}}\right),
\]

as \( n \to \infty \). Hence, \( GKS_{n} \) and \( GKS_{n}^{*} \) are consistent against \( H_{1} \).

Next, consider \( GCVM_{n} \) and \( GCVM_{n}^{*} \)

\[
GCVM_{n} = \frac{1}{n} \sum_{t=1}^{n} \Gamma_{1}^{2}(X_{t-1}) = \frac{1}{n} \sum_{t=1}^{n} \left\{ \frac{\sum_{t=1}^{n} \Delta X_{t} I(X_{t-1} \leq X_{t-1})}{\left(\sum_{t=1}^{n} \tilde{a}_{t}^{2}\right)^{1/2}} \right\}^{2},
\]

Note that \( X(\theta) \) may be positive or negative and for \( s,t > L, I(X_{s-1} \leq X_{t-1}) \) iff \( 1(\hat{\theta}^{s-1} X(\theta) \leq \hat{\theta}^{s-1} X(\theta)) \) iff \( 1(\hat{\theta}^{s-1} X(\theta) < 0) \) or \( 1(s > t \text{ and } X(\theta) > 0) \) or \( 1(t < s \text{ and } X(\theta) \leq 0) \). As shown in (D.1), we find that for \( t \geq L \)

\[
\begin{align*}
A_{n} &= \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \Delta X_{t} I(X_{t-1} \leq X_{t-1}) \\
&= \frac{\hat{\theta}_{n}}{\sqrt{n}} \frac{1}{\hat{\theta}_{n}} \sum_{t=1}^{n} \Delta X_{t} \hat{\theta}_{n}^{s-1} I(X_{t-1} \leq X_{t-1}) \\
&= \frac{\hat{\theta}_{n}}{\sqrt{n}} X(\theta)(\theta - 1) \left[ \frac{1}{\hat{\theta}_{n}} \sum_{t=1}^{n} \hat{\theta}_{n}^{s-1} I(X(\theta) \leq X_{t-1} \hat{\theta}_{n}^{s-1}) \right] \times (1 + o_{p}(1)) + o_{p}(1) \\
&= \frac{\hat{\theta}_{n}}{\sqrt{n}} X(\theta)(\theta - 1) \left[ \frac{1}{\hat{\theta}_{n}} \sum_{t=1}^{n} \hat{\theta}_{n}^{s-1} I(X(\theta) \leq X_{t-1} \hat{\theta}_{n}^{s-1}) \right] \times (1 + o_{p}(1)) + o_{p}(1).
\end{align*}
\]

Now we evaluate \( GCVM_{n} \) for \( s > t \) and \( X(\theta) < 0 \)

\[
GCVM_{n} = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left\{ \frac{\sum_{t=1}^{n} \Delta X_{t} I(X_{t-1} \leq X_{t-1})}{\left(\sum_{t=1}^{n} \tilde{a}_{t}^{2}\right)^{1/2}} \right\}^{2} = O_{p} \left(\frac{\hat{\theta}_{n}}{\sqrt{n}}\right),
\]

so that \( GCVM_{n} \) is divergent for \( 1(X(\theta) < 0) \).

Evaluating \( GCVM_{n} \) for \( s \leq t \) and \( X(\theta) > 0 \), we have

\[
GCVM_{n} = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left\{ \frac{\sum_{t=1}^{n} \Delta X_{t} I(X_{t-1} \leq X_{t-1})}{\left(\sum_{t=1}^{n} \tilde{a}_{t}^{2}\right)^{1/2}} \right\}^{2} = O_{p} \left(\frac{\hat{\theta}_{n}}{\sqrt{n}}\right).
\]

Thus, \( GCVM_{n} \) is divergent for \( 1(X(\theta) > 0) \). It follows that the test \( GCVM_{n} \) is consistent against explosive AR(1) alternatives. In a similar way, we have \( GCVM_{n}^{*} = O_{p}(\frac{\hat{\theta}_{n}}{\sqrt{n}}) \) and the test \( GCVM_{n}^{*} \) is consistent against explosive AR(1) alternatives.

ACKNOWLEDGMENTS

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