中央研究院經濟所學術研討論文
IEAS Working Paper

A New Test of the martingale Difference Hypothesis

Chung-Ming Kuan and Wei-Ming Lee

IEAS Working Paper No. 03-A001

November, 2003

Institute of Economics
Academia Sinica
Taipei 115, TAIWAN
http://www.sinica.edu.tw/econ/



中央研究院　經濟研究所

INSTITUTE OF ECONOMICS, ACADEMIA SINICA

TAIWAN

copyright @ 2003 (Chung-Ming Kuan and Wei-Ming Lee)

A New Test of the Martingale Difference Hypothesis

Chung-Ming Kuan and **Wei-Ming Lee**

Institute of Economics
Academia Sinica

Abstract

In this paper we propose a new class of tests for the martingale difference hypothesis based on the moment conditions derived by Bierens (1982). In contrast with the existing consistent tests, the proposed test has a standard limiting distribution and is easy to implement. Comparing with the commonly used autocorrelation- and spectrum-based tests, it has power against a much larger class of alternatives that may be serially correlated or uncorrelated. Moreover, this test does not rely on the assumption of conditional homoskedasticity and requires a weaker moment condition. Our simulations confirm that the proposed test is powerful against various linear and nonlinear alternatives and is quite robust to the failure of higher-order moments. Our empirical study on exchange rate returns also shows that the conclusion resulted from the proposed test is different from that of the conventional tests.

JEL classification: C12, C22

Keywords: autocorrelation-based test, Bierens' equivalence result, martingale difference sequence, multivariate exponential distribution, spectrum-based test

† Author for correspondence: Chung-Ming Kuan, Institute of Economics, Academia Sinica, Taipei 115, Taiwan; ckuan@econ.sinica.edu.tw.

†† We would like to thank Yi-Ting Chen for many useful comments. C.-M. Kuan gratefully acknowledges the research support from the National Science Council of the Republic of China (NSC91-2415-H-001-016); W.-M. Lee's research is supported by a graduate fellowship under the same project.

1 Introduction

It is important to test if a time series is a martingale difference (MD) sequence in many economic and financial studies. For example, the martingale version of the market efficiency hypothesis requires the asset returns in an efficient market to follow an MD process, so that currently available information does not help improving the forecasts of future returns; see, e.g., Fama (1970, 1991) and LeRoy (1989). Hall (1978) also argued that changes in consumption between any two consecutive periods should be unpredictable. The concept of MD has also been used to define the correctness of econometric models. A time series regression model is said to be correctly specified (for the conditional mean) if the disturbances of the model follow an MD sequence. Therefore, a test of the MD hypothesis is useful in evaluating economic hypotheses as well as econometric models.

To test the MD hypothesis, it is well known that the tests based on the autocorrelation function and its spectral counterpart are not consistent against non-MD sequences that are serially uncorrelated. The autocorrelation-based Q test of Box and Pierce (1970) and Ljung and Box (1978) and the spectrum-based test of Durlauf (1991) are leading examples. Although these tests were subsequently refined to be robust to conditional heteroskedasticity (e.g., Lobato, Nankervis, and Savin, 2001; Deo, 2000), they still are unable to detect more general non-MD alternatives. There are several consistent tests of the MD hypothesis in the literature; see e.g., Bierens (1982, 1984), De Jong (1996), Bierens and Ploberger (1997), Dominguez and Lobato (2000), and Whang (2000, 2001). While consistency is an important property, these MD tests typically suffer from the drawback that their limiting distributions are data dependent. Implementing these tests are therefore practically cumbersome because their critical values can not be tabulated. An exception is the test proposed by Hong (1999); yet Hong's test is in effect a test of pairwise independence which is not necessary for the MD hypothesis.

In this paper, we propose a new class of MD tests based on a set of unconditional moment conditions that are equivalent to the MD hypothesis (Bierens, 1982). The proposed test has the following advantages, relative to existing tests. First, in contrast with the aforementioned consistent tests, it has a standard limiting distribution and is easy to implement. Second, although not consistent, it has power against a much larger class of alternatives than the commonly used autocorrelation- and spectrum-based tests. Third, the validity of this test does not rely on the assumption of conditional homoskedasticity and requires a weaker moment condition. This feature makes the proposed test a sensible tool for testing economic and financial time series. Our simulations confirm that the proposed test is indeed powerful against various linear and nonlinear alternatives and is quite robust to the failure of higher-order moment conditions. Our empirical study on

exchange rate returns also shows that the conclusion resulted from the proposed test is different from that of the conventional tests.

This paper proceeds as follows. In Section 2, we introduce the new class of MD tests and analyze its limiting behavior. We then discuss how to implement the proposed test in Section 3. Monte Carlo simulation results are reported in Section 4. An empirical study of exchange rate returns is presented in Section 5. Section 6 concludes this paper. All mathematical proofs are deferred to Appendix.

2 Testing the Martingale Difference Hypothesis

Let $\{Y_t\}$ be a sequence of integrable random variables defined on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and $\{\mathcal{F}_t\}$ be an increasing family of sub-σ-algebras of $\mathcal{F}$. $\{Y_t\}$ is said to be an MD sequence relative to $\{\mathcal{F}_t\}$ if Y_t is $\mathcal{F}_t$-measurable and $\mathbb{E}[Y_t \mid \mathcal{F}_{t-1}] = 0$. Whether a sequence is MD depends crucially on the associated filtration $\{\mathcal{F}_t\}$. In the context of time series analysis, we are primarily interested in the MD sequence $\{Y_t\}$ relative to $\{\sigma(Y_t, Y_{t-1}, \ldots, Y_1)\}$, the sequence of σ-algebras generated by $Y_t, Y_{t-1}, \ldots, Y_1$. In what follows, we will not explicitly mention this filtration for simplicity. Note also that the definition of the MD sequence restricts only its mean behavior but not other moments. Thus, a conditionally heteroskedastic sequence, such as a GARCH (generalized autoregressive conditionally heteroskedastic) process, may still be an MD sequence.

Let $\boldsymbol{\eta}_{t-1}$ denote a collection of the elements in $\{Y_{t-1}, \ldots, Y_1\}$. It is well known that $\{Y_t\}$ is an MD sequence if, and only if, $\mathbb{E}[Y_t h(\boldsymbol{\eta}_{t-1})] = 0$ for any measurable function h and any collection $\boldsymbol{\eta}_{t-1}$. In practice, it is typical to consider only the specific collection $\boldsymbol{\eta}_{t-1} = \boldsymbol{Y}_{t-1,k} := \{Y_{t-1}, \ldots, Y_{t-k}\}$ for a given k. While evaluating $\mathbb{E}[Y_t h(\boldsymbol{Y}_{t-1,k})] = 0$ for all measurable functions h is practically implausible, the well known autocorrelation-based tests focus on the linear function. Clearly, such tests ignore possible nonlinear relationships between Y_t and $\boldsymbol{Y}_{t-1,k}$ and hence are inconsistent against non-MD sequences with zero autocorrelations. Consider, for example, the following bilinear process:

$$Y_t = \varepsilon_t + a\,\varepsilon_{t-1} Y_{t-2},$$

and the nonlinear moving average process:

$$Y_t = \varepsilon_{t-1}\varepsilon_{t-2}(\varepsilon_{t-2} + \varepsilon_t + 1),$$

where ε_t are i.i.d. $\mathcal{N}(0,1)$ random variables. It is clear that $\operatorname{corr}(Y_t, Y_{t-j}) = 0$ for all j in both cases, yet these two processes are not MD sequences. The autocorrelation-based tests thus may mislead us to infer that this is an MD sequence. One may also

choose h as a nonlinear function and construct a nonlinearity test accordingly; see e.g., Keenan (1985), Tsay (1986), and Luukkonen et al. (1988). The resulting tests are still restrictive, as they may not be sensitive to other forms of nonlinearity.

Bierens (1982) shows that, for an integrable random variable Y and a random vector $\boldsymbol{X}$ in $\mathbb{R}^k$,

$$\mathbb{E}[Y \mid \sigma(\boldsymbol{X})] = 0 \text{ if and only if } \mathbb{E}[Y \exp(i\boldsymbol{\omega}'\boldsymbol{X})] = 0, \text{ for all } \boldsymbol{\omega} \in \mathbb{R}^k, \tag{1}$$

where $i = \sqrt{-1}$. In accordance with (1), the hypothesis $\mathbb{E}(Y_t \mid \boldsymbol{Y}_{t-1,k}) = 0$ can be expressed as

$$m_k(\boldsymbol{\omega}) := \mathbb{E}[Y_t \exp(i\boldsymbol{\omega}'\boldsymbol{Y}_{t-1,k})] = \mathbb{E}\{Y_t \exp[i(\omega_1 Y_{t-1} + \cdots + \omega_k Y_{t-k})]\} = 0, \tag{2}$$

for all $\boldsymbol{\omega} \in \mathbb{R}^k$. This equivalence result enables us to concentrate on the correlations between Y_t and the exponential functions of $\boldsymbol{Y}_{t-1,k}$. Other equivalence relationships between the conditional expectation and unconditional moment conditions can be found in, e.g., Bierens (1990), Bierens and Ploberger (1997), and Stinchcombe and White (1998).

As (2) is a collection of uncountably many moment conditions indexed by $\boldsymbol{\omega}$, we may "summarize" these conditions by introducing a weighting function g of $\boldsymbol{\omega}$ and integrate out $\boldsymbol{\omega}$. In what follows, let g be such that $\int g(\boldsymbol{\omega})\, \mathrm{d}\boldsymbol{\omega} < \infty$. For example, Bierens (1982, 1984) consider a condition equivalent to (2):

$$\int_{\mathbb{R}^k} m_k(\boldsymbol{\omega})^2 g(\boldsymbol{\omega})\, \mathrm{d}\boldsymbol{\omega} = 0. \tag{3}$$

The resulting test statistic is, however, difficult to implement and has a data-dependent limiting distribution; see also the discussion at the end of this section. Instead of (3), we follow Chen, Chou, and Kuan (2000) to consider an implication of (2):

$$\int_{\mathbb{R}^k} m_k(\boldsymbol{\omega}) g(\boldsymbol{\omega})\, \mathrm{d}\boldsymbol{\omega} = \int_{\mathbb{R}^k} [m_{c,k}(\boldsymbol{\omega}) + i\, m_{s,k}(\boldsymbol{\omega})] g(\boldsymbol{\omega})\, \mathrm{d}\boldsymbol{\omega} = 0,$$

where $m_{c,k}(\boldsymbol{\omega})$ and $m_{s,k}(\boldsymbol{\omega})$ are, respectively, the real and imaginary parts of $m_k(\boldsymbol{\omega})$. A problem with this condition is that it may be zero even when m_k is not identically zero; cf. (3). For example, when Y_t follows a stationary AR(1) process generated by Gaussian innovations, it can be shown that $m_{c,1}(\omega) = \mathbb{E}[Y_t \cos(\omega_1 Y_{t-1})] = 0$ and

$$m_{s,1}(\omega) = \mathbb{E}[Y_t \sin(\omega_1 Y_{t-1})] = \alpha\omega_1 \exp\Big(-\frac{1}{2}\omega_1^2\Big),$$

where α is the AR coefficient. Then, $m_{s,1}$ (and hence m_1) would be integrated to zero as long as g is a positive function symmetric about the vertical axis. This kind of problem may be alleviated to a large extent by choosing a proper weighting function.

To simplify the choice of the weighting function, we shall concentrate on the following implication of (2):

$$\int_{\mathbb{R}^{k+}} m_k(\boldsymbol{\omega}) g(\boldsymbol{\omega})\, \mathrm{d}\boldsymbol{\omega} = 0, \tag{4}$$

where $\mathbb{R}^{k+} = (\mathbb{R}^+)^k$. Although (4) may still be zero when m_k is not identically zero on $\mathbb{R}^{k+}$, choosing a proper g function in this case is relatively easy; we will elaborate more on the choice of g in Section 3. By changing the orders of integration, (4) is equivalent to

$$\mathbb{E}[Y_t\, \varphi_g(\boldsymbol{Y}_{t-1,k})] = 0, \tag{5}$$

where $\varphi_g(\boldsymbol{Y}_{t-1,k})$ is an indicator of the weighted exponential functions of $\boldsymbol{Y}_{t-1,k}$:

$$\varphi_g(\boldsymbol{Y}_{t-1,k}) = \int_{\mathbb{R}^{k+}} \exp(i\boldsymbol{\omega}'\boldsymbol{Y}_{t-1,k})\, g(\boldsymbol{\omega})\, \mathrm{d}\boldsymbol{\omega}. \tag{6}$$

In the light of Bierens' equivalence result (1), the correlation between Y_t and $\varphi_g(\boldsymbol{Y}_{t-1,k})$ may be interpreted as a summary measure of the correlations between Y_t and all possible functions of $\boldsymbol{Y}_{t-1,k}$. As compared with the autocorrelation-based tests and the spectrum-based tests that check only autocorrelations among Y_t, a proper test of (5) ought to have power against a much larger class of alternatives.

Let $\varphi_{j,g}(\boldsymbol{Y}_{t-1,k})$, $j = c, s$, denote the real and imaginary parts of $\varphi_g(\boldsymbol{Y}_{t-1,k})$:

$$\varphi_{c,g}(\boldsymbol{Y}_{t-1,k}) = \int_{\mathbb{R}^{k+}} \cos(\boldsymbol{\omega}'\boldsymbol{Y}_{t-1,k})\, g(\boldsymbol{\omega})\, \mathrm{d}\boldsymbol{\omega},$$

$$\varphi_{s,g}(\boldsymbol{Y}_{t-1,k}) = \int_{\mathbb{R}^{k+}} \sin(\boldsymbol{\omega}'\boldsymbol{Y}_{t-1,k})\, g(\boldsymbol{\omega})\, \mathrm{d}\boldsymbol{\omega}.$$

The hypothesis (5) is equivalent to the joint hypothesis:

$$H_0 \colon \mathbb{E}[\psi_{j,g}(\boldsymbol{Y}_{t-1,k})] = 0, \qquad j = c, s, \tag{7}$$

where $\psi_{j,g}(Y_t, \boldsymbol{Y}_{t-1,k}) = Y_t \varphi_{j,g}(\boldsymbol{Y}_{t-1,k})$, $j = c, s$, are the real and imaginary parts of $\psi_g(Y_t, \boldsymbol{Y}_{t-1,k})$. It is now natural to base a test of (7) on its sample counterpart:

$$\bar{\psi}_{j,g} = \frac{1}{T-k} \sum_{t=k+1}^{T} \psi_{j,g}(Y_t, \boldsymbol{Y}_{t-1,k}). \qquad j = c, s.$$

This approach is similar in spirit to that of Chen et al. (2000) and Chen (2002) for constructing characteristic-function-based tests of time reversibility and goodness of fit.

For asymptotic analysis, we consider one of the following two conditions on Y_t.

[S1] Y_t are stationary and ergodic with finite variance σ_Y^2.

[M1] Y_t are either uniform mixing with the mixing coefficient of size $r/(2r-1)$ for some $r \geq 1$ or strong mixing with the mixing coefficient of size $r/(r-1)$ for some $r > 1$, and Y_t are L_{2r+a}-bounded for some $a > 0$.

These conditions regulate the behavior of Y_t in different ways. While the condition [S1] imposes stationarity and ergodicity, the mixing condition [M1] permits Y_t to be weakly dependent and heterogeneously distributed. Note that Y_t may still be conditionally heteroskedastic under [S1]. Note also that [S1] requires only finite variance, yet the moment condition in [M1] must be strengthened when the dependence among Y_t becomes stronger. These moment conditions are, nonetheless, almost "minimal" to ensure a central limit theorem in the current context; see e.g., White (1984). Moreover, $\psi_{j,g}(Y_t, \boldsymbol{Y}_{t-1,k})$ are also stationary and ergodic under [S1] or mixing with the same mixing coefficient under [M1].

As the sine and cosine functions are bounded between ± 1, $\varphi_{j,g}(\boldsymbol{Y}_{t-1,k})$ are also bounded. It follows that $\psi_{j,g}(Y_t, \boldsymbol{Y}_{t-1,k})$ have the same moment property as Y_t and hence have bounded variance under [S1] and [M1]. We therefore define $\sigma_{j,g}^2 = \mathrm{var}\big(\sqrt{T-k}\,\bar{\psi}_{j,g}\big)$ and $\sigma_{cs,g} = \mathrm{cov}\big(\sqrt{T-k}\,\bar{\psi}_{c,g}, \sqrt{T-k}\,\bar{\psi}_{s,g}\big)$, which are also bounded. Moreover, when $\{Y_t\}$ is an MD sequence, $\{\psi_{c,g}(Y_t, \boldsymbol{Y}_{t-1,k})\}$ and $\{\psi_{s,g}(Y_t, \boldsymbol{Y}_{t-1,k})\}$ are also MD sequences with respect to the same filtration. Thus, $\psi_{c,g}(Y_t, \boldsymbol{Y}_{t-1,k})$ and $\psi_{s,g}(Y_\tau, \boldsymbol{Y}_{\tau-1,k})$ must be uncorrelated for all $t \neq \tau$. It follows that

$$\sigma_{j,g}^2 = \frac{1}{T-k} \sum_{t=k+1}^{T} \mathbb{E}[\psi_{j,g}(Y_t, \boldsymbol{Y}_{t-1,k})^2], \qquad j = c, s,$$

$$\sigma_{cs,g} = \frac{1}{T-k} \sum_{t=k+1}^{T} \mathbb{E}[\psi_{c,g}(Y_t, \boldsymbol{Y}_{t-1,k})\psi_{s,g}(Y_t, \boldsymbol{Y}_{t-1,k})],$$

when $\{Y_t\}$ is an sequence.

We are now ready to state the main distribution result of normalized $\bar{\psi}_{j,g}$.

Theorem 2.1 *Given* [S1] *or* [M1], *suppose that* $\mathbb{E}(Y_t|\mathcal{F}_{t-1}) = 0$. *If* $\sigma_{c,g}^2\sigma_{s,g}^2 - \sigma_{cs,g}^2 > \epsilon > 0$ *uniformly in* T, *then*

$$\begin{bmatrix} \sigma_{c,g}^2 & \sigma_{cs,g} \\ \sigma_{cs,g} & \sigma_{s,g}^2 \end{bmatrix}^{-1/2} \begin{bmatrix} \sqrt{T-k}\,\bar{\psi}_{c,g} \\ \sqrt{T-k}\,\bar{\psi}_{s,g} \end{bmatrix} \xrightarrow{D} \mathcal{N}(0,\, \boldsymbol{I}_2),$$

where $\xrightarrow{D}$ *stands for convergence in distribution and* $\boldsymbol{I}_2$ *is the* 2×2 *identity matrix.*

Remark: It is possible to allow k to grow with T at a slower rate so that we are able to test asymptotically the MD property with respect to all past information. Proving such a result is, however, technically much more involved. For example, $\psi_{j,g}(Y_t, \boldsymbol{Y}_{t-1,k})$ need not be mixing when k grows with T; more technical conditions are then needed to ensure that $\psi_{j,g}(Y_t, \boldsymbol{Y}_{t-1,k})$ are well behaved. We will not pursue this possibility here.

The sample counterparts of $\sigma^2_{j,g}$ and $\sigma_{cs,g}$ are:

$$\hat{\sigma}^2_{j,g} = \frac{1}{T-k}\sum_{t=k+1}^{T} \psi_{j,g}(Y_t, \boldsymbol{Y}_{t-1,k})^2, \qquad j = c, s,$$

$$\hat{\sigma}_{cs,g} = \frac{1}{T-k}\sum_{t=k+1}^{T} \psi_{c,g}(Y_t, \boldsymbol{Y}_{t-1,k})\psi_{s,g}(Y_t, \boldsymbol{Y}_{t-1,k}),$$

Under [S1], $\hat{\sigma}^2_{j,g}$ and $\hat{\sigma}_{cs,g}$ are strongly consistent for $\sigma^2_{j,g}$ and $\sigma_{cs,g}$ by the ergodic theorem. The consistency also holds under [M1] by Corollary 3.48 of White (1984). The proposed test statistic for (7) is then:

$$\begin{aligned} \mathcal{J}_g &= \begin{bmatrix} \sqrt{T-k}\,\bar{\psi}_{c,g} \\ \sqrt{T-k}\,\bar{\psi}_{s,g} \end{bmatrix}' \begin{bmatrix} \hat{\sigma}^2_{c,g} & \hat{\sigma}_{cs,g} \\ \hat{\sigma}_{cs,g} & \hat{\sigma}^2_{s,g} \end{bmatrix}^{-1} \begin{bmatrix} \sqrt{T-k}\,\bar{\psi}_{c,g} \\ \sqrt{T-k}\,\bar{\psi}_{s,g} \end{bmatrix}, \\ &= \frac{T-k}{\hat{\sigma}^2_{c,g}\hat{\sigma}^2_{s,g} - \hat{\sigma}^2_{cs,g}} \left[\hat{\sigma}^2_{s,g}\bar{\psi}^2_{c,g} + \hat{\sigma}^2_{c,g}\bar{\psi}^2_{s,g} - 2\hat{\sigma}_{cs,g}\bar{\psi}_{c,g}\bar{\psi}_{s,g}\right]. \end{aligned} \tag{8}$$

The result below follows immediately from Theorem 2.1.

Corollary 2.2 *Given* [S1] *or* [M1]*, suppose that* $\mathbb{E}(Y_t|\mathcal{F}_{t-1}) = 0$*. When* $\sigma^2_{j,g}$ *are bounded away from zero,*

$$\mathcal{J}_g \xrightarrow{D} \chi^2(2).$$

When $\mathbb{E}[\psi_{j,g}(Y_t, \boldsymbol{Y}_{t-1,k})] = \delta_j \neq 0$*,* $\mathcal{J}_g$ *diverges in probability.*

The proposed test (8) differs from the autocorrelation-based tests and the spectrum-based tests in the following respects. First, it has power against non-MD sequences that are serially uncorrelated. Second, it requires a weaker moment condition under [S1]. Recall that the autocorrelation-based tests and the spectrum-based tests require the data to possess finite fourth or even eighth moment; the consistent tests of Hong (1999) and Dominguez and Lobato (2000) also impose similar moment conditions. Third, there is no restriction on the second moment of the series so that the proposed test is robust to conditional heteroskedasticity. The last two features are particularly important for testing financial time series because such series are, in general, conditionally heterosekdastic and may even lack high-order moments (Loretan and Phillips, 1994; de Lima, 1997).

It is also interesting to compare our test with the integrated conditional moment test of Bierens (1982, 1984) based on (3). In contrast with (4), (3) does not permit changing the orders of integration and hence yields a complex test statistic. Worse still is that the limiting distribution of Bierens' test depends on the data generating process so that the critical values can not be tabulated. Other consistent tests of the MD hypothesis (e.g., Dominguez and Lobato, 2000; Whang, 2000, 2001) and the characteristic-function-based tests of goodness of fit (e.g., Feuerverger and Mureika, 1977) suffer from the same drawback. By contrast, the proposed test is easy to implement and asymptotically pivotal with the $\chi^2(2)$ distribution in the limit. Note also that the degrees of freedom of this distribution does not depend on the number of lags k. Clearly, these attractive properties are obtained at the expense of test consistency and render the proposed test a convenient diagnostic tool.

3 The Choice of Weighting Functions

The precise expression of the proposed $\mathcal{J}_g$ test depends on the weighting function g. While a weighting function has no effect under the null hypothesis, it does affect the test power. Determining a proper weighting function is practically difficult because m_k varies with data and is unknown to researchers. In this section, we discuss how the weighting function may be chosen.

From (4) we can see that a weighting function g should be chosen such that m_k will not be integrated to zero unless it is identically zero on $\mathbb{R}^{k+}$. To better understand the behavior of m_k, we consider four non-MD sequences and plot $m_{c,1}$ and $m_{s,1}$, the real and imaginary parts of m_1, in Figures 1. The sequences being analyzed are: an AR(1) process, a SETAR (self-exciting threshold AR) process, a bilinear process and an all-pass process; see Table 2 in Section 4 for their data generating processes. As discussed in the preceding section, $m_{c,1}$ of the AR(1) process is zero and

$$m_{s,1} = \alpha\omega_1 \exp\Big(-\frac{1}{2}\omega_1^2\Big),$$

which is positive for all $\omega > 0$, as shown in Figure 1 For the other non-MD sequences, $m_{c,1}$ and $m_{s,1}$ are obtained via simulations. In particular, $m_{c,1}$ and $m_{s,1}$ of the SETAR and bilinear processes are damped (cosine and sine) waves around zero, whereas $m_{c,1}$ and $m_{s,1}$ of the all-pass process are all non-negative. A common feature of these plots is that $m_{c,1}$ and $m_{s,1}$, if non-zero, take large values for small ω and eventually approach zero as ω increases. Thus, to avoid $m_{c,1}$ and $m_{s,1}$ being integrated to zero, we may choose g as a positive, monotonically decreasing function on $\mathbb{R}^+$ such that it takes large values for ω close to the origin and approaches zero when ω tends to infinity. Although characterizing

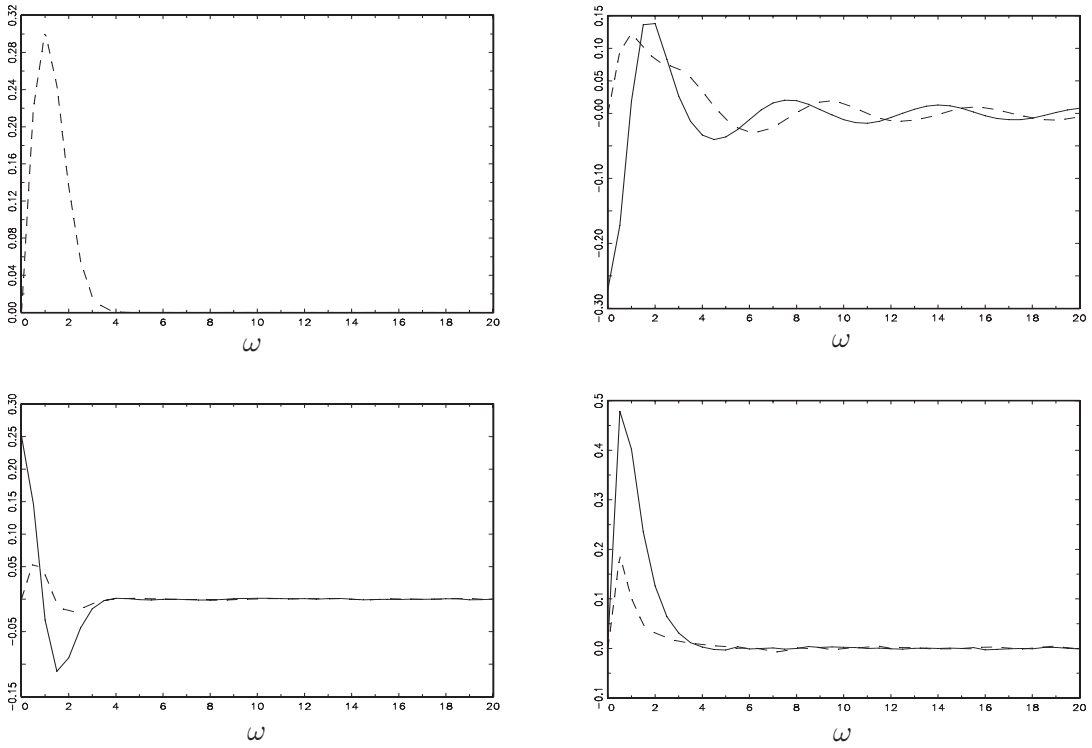


Figure 1: Simulated $m_{c,1}(\omega)$ and $m_{s,1}(\omega)$ (dashed line) for AR(1) (upper left), SETAR (upper right), bilinear (lower left), and all-pass (lower right) processes.

$m_{c,k}$ and $m_{s,k}$ for $k > 1$ is not easy, it is reasonable to require g to have these properties in the one-dimensional case ($k = 1$).

As far as application is concerned, we would like to choose g such that $\mathcal{J}_g$ can be easily computed. Observe that the indicator $\varphi_g(\boldsymbol{Y}_{t-1,k})$ involves k-dimensional integration. Thus, it would be very convenient if φ_g has an analytic form. When g is a density function on $\mathbb{R}^{k+}$, $\varphi_g(\boldsymbol{Y}_{t-1,k})$ in (6) may be viewed as the characteristic function of g with the parameter $\boldsymbol{Y}_{t-1,k}$. This suggests to choose g as the density function on $\mathbb{R}^{k+}$ that has an analytic characteristic function. With such a g function, φ_g has a closed form, and so do $\varphi_{c,g}$ and $\varphi_{s,g}$. Consequently, $\psi_{c,g}(Y_t, \boldsymbol{Y}_{t-1,k})$ and $\psi_{s,g}(Y_t, \boldsymbol{Y}_{t-1,k})$ can be easily computed. A potential candidate of g is the density of the multivariate exponential distribution which has the characteristic function:

$$\phi(\boldsymbol{\nu}) := \mathbb{E}[\exp(i\boldsymbol{\nu}'\boldsymbol{\omega})] = \{\det[\boldsymbol{I}_k - \beta\, i\boldsymbol{R}\boldsymbol{D}_{\boldsymbol{\nu}}]\}^{-1},$$

where $\boldsymbol{D}_{\boldsymbol{\nu}} = \mathrm{diag}[\nu_1, \ldots, \nu_k]$, $\boldsymbol{R}$ is a correlation matrix, and β is the parameter. The precise expressions of $\phi(\boldsymbol{\nu})$ with $k = 1, 2, 3, 4$ are given in Appendix; more detailed discussion of multivariate exponential distributions can be found in Johnson and Kotz (1972,

Chapter 41). Note that the one-dimensional, exponential density function satisfies the requirements discussed in the preceding paragraph.

Basing on the expressions of $\phi(\boldsymbol{\nu})$ in Appendix and replacing $\boldsymbol{\nu}$ with $\boldsymbol{Y}_{t-1,k}$, we obtain the real and imaginary parts of $\psi_g(Y_t, \boldsymbol{Y}_{t-1,k})$. For the special case that $\boldsymbol{R}$ is an identity matrix, we have for $k = 1$,

$$\psi_{c,\exp(\beta)}(Y_t, Y_{t-1,1}) = \frac{Y_t}{1 + \beta^2 Y_{t-1}^2},$$
$$\psi_{s,\exp(\beta)}(Y_t, Y_{t-1,1}) = \frac{\beta Y_t Y_{t-1}}{1 + \beta^2 Y_{t-1}^2},$$

where $Y_{t-1,1}$ is just Y_{t-1}. Similarly, for $k = 2$ we have

$$\psi_{c,\exp(\beta)}(Y_t, \boldsymbol{Y}_{t-1,2}) = \frac{Y_t \left(1 - \beta^2 Y_{t-1} Y_{t-2}\right)}{\left(1 + \beta^2 Y_{t-1}^2\right)\left(1 + \beta^2 Y_{t-2}^2\right)},$$
$$\psi_{s,\exp(\beta)}(Y_t, \boldsymbol{Y}_{t-1,2}) = \frac{\beta Y_t \left(Y_{t-1} + Y_{t-2}\right)}{\left(1 + \beta^2 Y_{t-1}^2\right)\left(1 + \beta^2 Y_{t-2}^2\right)},$$

and for $k = 3$,

$$\psi_{c,\exp(\beta)}(Y_t, \boldsymbol{Y}_{t-1,3}) = \frac{Y_t \left(1 - \beta^2 (Y_{t-1} Y_{t-2} + Y_{t-2} Y_{t-3} + Y_{t-1} Y_{t-3})\right)}{\left(1 + \beta^2 Y_{t-1}^2\right)\left(1 + \beta^2 Y_{t-2}^2\right)\left(1 + \beta^2 Y_{t-3}^2\right)},$$
$$\psi_{s,\exp(\beta)}(Y_t, \boldsymbol{Y}_{t-1,3}) = \frac{Y_t \left[\beta(Y_{t-1} + Y_{t-2} + Y_{t-3}) - \beta^3 Y_{t-1} Y_{t-2} Y_{t-3}\right]}{\left(1 + \beta^2 Y_{t-1}^2\right)\left(1 + \beta^2 Y_{t-2}^2\right)\left(1 + \beta^2 Y_{t-3}^2\right)}.$$

Observe from these expressions that the numerators involve the cross-product terms of $Y_t, \ldots, Y_{t-k}$. When k increases, the derivation of $\psi_{j,\exp(\beta)}$ becomes laborious. Using an identity of trigonometry it is easy to obtain:

$$\varphi_{c,g}(\boldsymbol{Y}_{t-1,k}) = \varphi_{c,g}(\boldsymbol{Y}_{t-1,k-1})\varphi_{c,g}(Y_{t-k}) - \varphi_{s,g}(\boldsymbol{Y}_{t-1,k-1})\varphi_{s,g}(Y_{t-k}),$$
$$\varphi_{s,g}(\boldsymbol{Y}_{t-1,k}) = \varphi_{c,g}(\boldsymbol{Y}_{t-1,k-1})\varphi_{s,g}(Y_{t-k}) + \varphi_{s,g}(\boldsymbol{Y}_{t-1,k-1})\varphi_{c,g}(Y_{t-k});$$

see Appendix for detail. Thus, $\varphi_{j,g}(\boldsymbol{Y}_{t-1,k})$ and hence $\psi_{j,g}(Y_t, \boldsymbol{Y}_{t-1,k})$ can be calculated recursively.

For $\psi_{c,\exp(\beta)}$ and $\psi_{s,\exp(\beta)}$ given above, we must also determine the parameter β. As will be seen in next section, the choice of β affects test power and may have to vary with data. There is, however, no simple rule to select an appropriate β. A rule of thumb is to set β as the reciprocal of the sample standard deviation of Y_t. This choice amounts to "standardizing" Y_t.

4 Monte Carlo Simulations

In this section, the finite sample performance of the proposed $\mathcal{J}_{\exp(\beta)}$ test is evaluated using Monte Carlo simulations. For each simulation experiment, the number of replications is 1000, and the sample sizes T are 100 and 500. To examine the effect of β on the performance of the proposed test, we consider four different values of β: $0.1, 1.0, 2.0$ and $\tilde{\beta} = (T^{-1}\sum_{t=1}^{T} Y_t^2)^{-1/2}$. For comparison, we simulate the modified, spectrum-based test of Deo (2000):

$$\mathrm{CVM}_m = \int_0^1 |U_w(\lambda)|^2 \, \mathrm{d}\lambda,$$

where

$$U_w(\lambda) = \frac{\sqrt{2}}{\pi} \sum_{j=1}^{T-1} w_T(j)\sqrt{T}\hat{a}_j \, \frac{\sin(j\pi\lambda)}{j},$$

with w_T a weighting function,

$$\hat{a}_j = \hat{\sigma}^2 \, \hat{\rho}_j \left[\frac{1}{T-j} \sum_{t=j+1}^{T} (Y_t - \bar{Y})^2 (Y_{t-j} - \bar{Y})^2 \right]^{-1/2},$$

and $\hat{\rho}_j$ are sample autocorrelation coefficients. As shown by Deo (2000), the limiting distribution of CVM_m is a functional of the standard Brownian bridge; the asymptotic critical values at the significance levels 5% and 1% are 0.46 and 0.74, respectively. To allow for conditional heteroskedasticity, Deo (2000) also proposed a modified version of the Box and Pierce (1970) test which is closely related to the Q^* test of Lobato, Nankervis, and Savin (2001). The modified Box-Pierce test statistic is

$$\mathrm{BP}_{m,k} = T \sum_{j=1}^{k} \hat{a}_j^2.$$

Under some regularity conditions, $\mathrm{BP}_{m,k}$ follows a chi-squared distribution with k degrees of freedom asymptotically.

For size simulations, we generate data from i.i.d. $\mathcal{N}(0,1)$, student's $t(3)$, and the symmetric stable distribution with the characteristic exponent being 1.5, denoted by $S(1.5)$. As $t(3)$ does not have finite third moment and $S(1.5)$ does not have finite variance, we may assess the effect of moment failure on the test performance. We also generate data from an ARCH(1) process and a GARCH(1,1) process: $Y_t = \sigma_t \varepsilon_t$, where $\varepsilon_t \sim$ i.i.d. $\mathcal{N}(0,1)$, and σ_t^2 has an ARCH(1) structure:

$$\sigma_t^2 = 1.0 + 0.5\, Y_{t-1}^2,$$

or a GARCH(1,1) structure:

$$\sigma_t^2 = 0.001 + 0.09\, Y_{t-1}^2 + 0.9\sigma_{t-1}^2.$$

Such data allow us to evaluate how conditional heteroskedasticity affects test performance. The GARCH(1,1) process given above was also considered by Dominguez and Lobato (2000); a feature of this process is that it does not have finite sixth moment. The empirical sizes of various tests are reported in Table 1. It can be seen that all tests perform quite well under the null hypothesis. The empirical sizes are close to the nominal size 5% for most cases, yet they are somewhat under-sized when the data are distributed as $S(1.5)$. Note also that conditional heteroskedasticity has little effect on the size performance of these tests.

To evaluate the power performance, we consider seven non-MD alternatives, including an AR(1) process, an MA(1) process, two self-exciting threshold autoregressive processes (SETAR-1, SETAR-2), two bilinear processes (BL-1, BL-2), and an all-pass process. The data generating processes of these alternatives are summarized in Table 2. To shed light on the effect of higher-order moment failure, we again generate the innovations of the first six sequences from $\mathcal{N}(0,1)$, $t(3)$, and $S(1.5)$ distributions; for the all-pass process, innovations are generated from the centered standard log-normal distribution. Clearly, the AR(1) and MA(1) processes have only linear dependence, whereas the SETAR and BL processes exhibit both linear and nonlinear dependence. It is of particular interest to note that the BL-2 process and the all-pass process are serially uncorrelated and, therefore, cannot be detected by the tests on serial correlations. An all-pass process is in fact the autoregressive-moving average (ARMA) process whose AR roots are reciprocals of the MA roots. Since this process is serially uncorrelated, it is also serially independent when the innovations are Gaussian. On the other hand, an all-pass process with non-Gaussian innovations is a linear process with a nonlinear pattern (i.e., volatility clustering). For more discussion on the all-pass process, we refer to Breidt et al. (2000).

The empirical powers are summarized in Tables 3–5. We first examine the power performance in Table 3 in which the innovations are $\mathcal{N}(0,1)$. For the AR(1) and MA(1) processes, the powers of the proposed $\mathcal{J}_{\exp(\beta)}$ test are comparable with those of the $\mathrm{BP}_{m,k}$ and CVM_m tests. We also observe that, as k increases, the power of $\mathcal{J}_{\exp(\beta)}$ decreases rapidly for the MA(1) process. This is due to the fact that Y_t and Y_{t-k} are independent for all $k > 1$ when Y_t are generated by an MA(1) model with i.i.d. innovations. For the SETAR-1, SETAR-2, and BL-1 processes which exhibit nonlinear dependence, it is expected that the proposed test performs better than the $\mathrm{BP}_{m,k}$ and CVM_m tests which are designed to capture only linear correlations. Table 3 shows that our tests indeed dominate the $\mathrm{BP}_{m,k}$ tests for both SETAR-1 and SETAR-2 and outperform the CVM_m

Table 1: Empirical sizes of the martingale difference tests.

		$\mathcal{N}(0,1)$		$t(3)$		S(1.5)		ARCH(1)		GARCH(1,1)	
	k	$T=100$	500	100	500	100	500	100	500	100	500
$\mathcal{J}_{\exp(0.1)}$	1	6.5	4.6	4.8	4.1	2.8	3.4	4.0	5.8	5.2	4.9
	2	5.3	4.7	4.8	4.5	3.5	3.4	4.7	4.8	5.2	4.8
	3	5.4	4.5	3.9	5.7	3.8	3.9	4.5	4.2	4.9	4.9
	4	5.5	4.6	5.2	4.7	3.5	3.6	3.5	4.4	5.2	5.4
$\mathcal{J}_{\exp(1.0)}$	1	5.6	4.5	4.6	3.9	2.9	3.0	5.3	5.2	7.1	4.4
	2	4.0	3.6	4.5	4.2	3.4	2.8	4.5	5.3	5.2	5.2
	3	4.5	4.9	3.8	3.8	2.8	3.6	4.1	4.4	4.3	5.0
	4	3.6	4.4	3.8	4.5	2.2	3.2	5.7	5.1	5.1	5.4
$\mathcal{J}_{\exp(2.0)}$	1	5.8	6.5	6.3	3.6	2.9	3.5	5.2	5.1	6.3	3.8
	2	3.9	3.3	3.3	4.0	2.8	3.3	3.9	3.6	4.6	5.0
	3	3.9	4.2	2.9	4.6	2.6	3.2	5.1	5.9	4.7	5.5
	4	3.7	5.2	3.1	4.4	2.2	4.2	4.8	5.1	4.6	3.1
$\mathcal{J}_{\exp(\tilde{\beta})}$	1	5.5	5.1	4.6	3.9	3.2	4.6	3.9	4.4	4.0	4.0
	2	4.2	5.5	4.1	5.6	3.8	3.0	4.7	4.9	4.7	4.1
	3	4.0	4.8	4.9	6.0	3.3	3.2	5.4	5.9	6.6	4.1
	4	3.0	4.4	4.0	4.5	3.7	2.8	4.6	4.7	4.8	5.0
$\mathrm{BP}_{m,5}$		5.5	4.7	4.3	3.9	3.2	3.2	2.9	3.3	5.5	5.5
$\mathrm{BP}_{m,10}$		4.0	5.1	3.4	2.6	4.4	2.5	4.1	4.6	3.5	4.4
CVM_m		4.3	4.6	4.5	3.4	3.0	3.1	3.9	4.7	4.1	4.5

Note: The entries are rejection frequencies in percentage when the nominal size is 5%; $\tilde{\beta}$ is the reciprocal of the sample standard deviation of the process.

test for SETAR-1. For the BL-1 alternative, it can also be seen that both the $\mathrm{BP}_{m,k}$ and CVM_m tests perform poorly even when $T = 500$, yet the proposed test with $\beta = 0.1$ has much better power performance. For the alternatives of the BL-2 and all-pass processes which are serially uncorrelated, the $\mathrm{BP}_{m,k}$ and CVM_m tests virtually have no power, but the proposed test with a proper β is much more powerful. For example, the power of $\mathcal{J}_{\exp(1.0)}$ against the all-pass process can be as high as 100.0% when the sample size is 500. For the BL-2 alternative, it is interesting to see that $\mathcal{J}_{\exp(\beta)}$ does not have power for $k = 1$. This may be due to the fact that Y_t depends on Y_{t-2} but not on Y_{t-1}. We also observe that $\mathcal{J}_{\exp(\tilde{\beta})}$ performs quite well in most cases, though it need not be the best among all $\mathcal{J}_{\exp(\beta)}$ tests. Finally, note that our test may be superior to that of Dominguez and Lobato (2000). In their simulations against the BL-1 process, the test powers (with

Table 2: The data generating processes of non-martingale difference alternatives.

AR(1) process:	$Y_t = 0.5\, Y_{t-1} + \varepsilon_t$
MA(1) process:	$Y_t = \varepsilon_t + 0.5\, \varepsilon_{t-1}$
SETAR-1 process:	$Y_t = \left(-0.5 I_{\{Y_{t-1} \geq 1\}} + 0.4 I_{\{Y_{t-1} < 1\}}\right) Y_{t-1} + \varepsilon_t$
SETAR-2 process:	$Y_t = \left(-0.5 I_{\{Y_{t-2} \geq 1\}} + 0.4 I_{\{Y_{t-2} < 1\}}\right) Y_{t-1} + \varepsilon_t$
BL-1 process:	$Y_t = \varepsilon_t + 0.25\, \varepsilon_{t-1} Y_{t-1} + 0.15\, \varepsilon_{t-1} Y_{t-2}$
BL-2 process:	$Y_t = \varepsilon_t + 0.7\, \varepsilon_{t-1} Y_{t-2}$
All-Pass process:	$Y_t = 0.8\, Y_{t-1} + u_t - 1.25 u_{t-1}$

Note: $I_{\{\cdot\}}$ is the indicator function; ε_t are i.i.d. $\mathcal{N}(0,1)$, $t(3)$, or $S(1.5)$; u_t are i.i.d. centered standard log-normal.

lag = 1) are, respectively, 23.8% and 98.1% for $T = 100$ and 500, whereas our test with $\beta = 0.1$ has power 48.7% and 99.9% for these two samples.

In Tables 4 and 5, the innovations are generated from $t(3)$ and $S(1.5)$, respectively. When innovations are generated from $t(3)$, $\mathcal{J}_{\exp(\beta)}$ is valid, but the $\mathrm{BP}_{m,k}$ and CVM_m tests are not. Table 4 shows, however, that $\mathrm{BP}_{m,k}$ and CVM_m still have power against AR(1) and MA(1) alternatives even when $T = 100$, and these two tests have power against SETAR-2 when $T = 500$. In other cases, $\mathrm{BP}_{m,k}$ and CVM_m have little power. The proposed test with a suitable β value performs much better than the $\mathrm{BP}_{m,k}$ and CVM_m tests in all cases, especially when the alternatives are SETAR-1, BL-1, and BL-2. The same conclusion also holds for Table 5. Even though the proposed test is not valid when the data are generated from $S(1.5)$, Table 5 suggests that it is very robust to the failure of higher-order moment condition. The robustness property of this test is particularly apparent when β is small (e.g., $\beta = 0.1$).

From the simulation results we also find that for smaller β, the powers of our test are quite stable across different k. Moreover, for smaller β, say $\beta = 0.1$, our test has much better power against the first five processes; larger β values may be needed to capture the dependence of BL-2 and the all-pass processes, however. To illustrate the effect of β, we plot the powers of $\mathcal{J}_\beta$ against β in Figure 2. It is clear from this figure that the power performance of the proposed test may vary dramatically with β. Except for the all-pass process, these powers decrease with β, suggesting that a smaller β value should be preferred. As there may not exist a general criterion to determine a proper β, we may have to compute the proposed test with different β values in practice.

Table 3: Empirical powers of the martingale difference tests ($\varepsilon_t \sim \mathcal{N}(0,1)$).

		AR(1)		MA(1)		SETAR-1		SETAR-2		BL-1		BL-2		All-Pass	
	k	$T=100$	500	100	500	100	500	100	500	100	500	100	500	100	500
$\mathcal{J}_{\exp(0.1)}$	1	99.4	100.0	96.9	100.0	61.0	100.0	41.2	96.6	48.7	99.9	6.4	8.7	9.2	31.7
	2	94.6	100.0	53.0	99.9	57.6	99.9	29.1	82.8	43.9	99.9	7.5	8.8	11.0	44.5
	3	85.8	100.0	34.4	94.0	55.2	99.7	27.0	86.4	44.3	99.7	8.9	8.9	11.2	51.1
	4	75.0	100.0	26.4	88.4	50.1	99.8	19.3	79.5	43.0	99.6	9.6	7.1	12.1	55.6
$\mathcal{J}_{\exp(1.0)}$	1	98.9	100.0	91.6	100.0	42.2	97.7	37.6	98.0	17.5	57.4	5.8	5.9	55.0	100.0
	2	81.5	100.0	32.5	95.2	17.1	65.7	38.0	97.8	7.6	27.3	22.1	82.0	54.3	100.0
	3	44.4	99.0	17.0	64.9	10.3	40.4	23.1	83.3	7.5	23.2	15.9	65.7	36.1	97.8
	4	23.5	82.0	11.4	34.4	8.4	22.2	13.9	50.9	7.0	17.6	11.1	45.2	18.9	78.0
$\mathcal{J}_{\exp(2.0)}$	1	93.9	100.0	78.1	100.0	34.2	96.6	33.7	94.0	11.2	22.0	5.0	6.1	52.5	99.8
	2	48.4	100.0	18.2	72.4	11.5	39.6	26.8	89.4	4.5	5.2	22.9	83.5	29.5	98.3
	3	14.0	60.7	7.4	25.9	6.4	13.4	11.4	49.8	3.4	4.5	17.4	63.4	10.1	54.5
	4	4.6	17.5	5.5	9.6	5.5	7.0	4.9	19.8	5.0	6.9	7.7	22.5	5.1	14.4
$\mathcal{J}_{\exp(\tilde{\beta})}$	1	97.2	100.0	91.7	100.0	42.4	98.0	41.5	98.6	17.8	64.7	7.1	6.2	36.3	98.5
	2	85.2	100.0	36.6	96.4	18.4	69.0	40.5	98.1	9.8	30.7	14.0	68.1	49.8	99.9
	3	50.4	99.8	20.6	74.3	11.9	43.8	24.1	83.3	9.0	26.8	14.7	58.7	51.7	99.8
	4	25.3	91.4	11.0	44.1	7.7	26.0	15.5	59.5	5.9	20.6	11.2	44.1	47.7	100.0
$\mathrm{BP}_{m,5}$		96.3	100.0	88.3	100.0	12.5	48.7	25.3	93.9	7.3	19.8	6.6	6.6	0.5	0.5
$\mathrm{BP}_{m,10}$		91.0	100.0	70.3	100.0	9.8	40.1	19.3	86.2	6.4	12.6	6.5	6.3	1.3	0.3
CVM_m		99.8	100.0	98.7	100.0	14.5	46.8	40.9	97.3	8.9	31.4	7.6	9.1	0.9	0.3

Note: The entries are rejection frequencies in percentage when the nominal size is 5%; $\tilde{\beta}$ is the reciprocal of the sample standard deviation of the process.

5 Application to Exchange Rate Returns

In this section, we apply the proposed test to re-examine the MD hypothesis for five weekly exchange rate returns studied by Liu and He (1991) and Fong and Ouliaris (1995). These data consist of 760 weekly returns on the Canadian dollar (Can), the German deutschmark (Dm), the French franc (Fr), the sterling pound (£), and the Japanese yen (¥), from August 14, 1974 to March 29, 1989. The summary statistics of these data are given in Table 6. The last column of Table 6 shows that none of these means are significantly different from zero. Moreover, the sample kurtosis coefficients reveal that the distributions of these return data may have fatter tails than the normal distribution.

Fong and Ouliaris (1995) examined these data using a family of the spectrum-based

Table 4: Empirical powers of the martingale difference tests ($\varepsilon_t \sim t(3)$).

		AR(1)		MA(1)		SETAR-1		SETAR-2		BL-1		BL-2	
	k	$T=100$	500	100	500	100	500	100	500	100	500	100	500
$\mathcal{J}_{\exp(0.1)}$	1	98.4	100.0	87.5	100.0	58.6	99.8	22.0	66.1	55.2	99.7	4.4	5.7
	2	95.1	100.0	46.8	98.8	54.8	99.7	19.7	52.9	53.1	99.6	6.3	9.2
	3	86.7	99.9	32.6	94.3	53.7	99.5	22.8	71.8	50.5	99.7	7.0	9.3
	4	74.9	99.9	27.4	81.8	54.4	99.3	17.8	61.1	55.5	99.3	5.4	8.6
$\mathcal{J}_{\exp(1.0)}$	1	89.2	100.0	69.4	99.8	28.6	79.0	21.9	66.3	12.5	34.0	5.2	3.9
	2	57.3	99.1	21.6	78.6	10.3	32.3	22.7	73.9	5.7	11.8	23.5	77.2
	3	20.9	79.3	8.6	35.0	8.2	17.0	13.9	49.2	5.7	12.3	16.3	61.2
	4	10.9	34.4	5.7	19.0	5.5	10.8	9.1	26.2	5.7	10.4	7.7	26.8
$\mathcal{J}_{\exp(2.0)}$	1	73.3	99.9	48.4	97.4	20.4	69.1	17.8	50.6	8.6	13.8	4.8	3.9
	2	28.0	83.5	11.0	41.0	6.6	15.5	14.4	54.2	2.4	4.5	19.3	64.3
	3	6.8	26.6	5.9	13.4	3.7	5.7	7.0	23.2	4.1	5.2	11.4	34.2
	4	4.8	8.4	4.1	5.6	4.7	4.2	4.5	9.6	3.6	3.4	4.6	11.5
$\mathcal{J}_{\exp(\tilde{\beta})}$	1	95.7	100.0	81.1	100.0	31.5	92.2	23.5	73.0	26.9	93.3	4.4	5.2
	2	77.4	100.0	33.7	92.6	17.2	68.4	24.8	77.9	19.0	87.6	11.8	29.4
	3	49.1	99.4	18.3	67.5	14.4	58.2	18.4	64.3	21.8	83.4	10.5	26.0
	4	26.0	88.6	12.4	44.0	11.1	43.0	10.9	44.4	16.2	82.1	7.7	18.4
$\mathrm{BP}_{m,5}$		87.1	99.2	60.0	95.8	5.1	11.2	15.9	61.7	4.0	4.1	4.1	4.8
$\mathrm{BP}_{m,10}$		81.0	99.1	50.2	92.7	5.2	10.6	11.9	50.6	3.6	4.5	4.2	4.0
CVM_m		94.6	99.4	85.3	97.4	9.8	21.4	23.8	62.2	4.7	4.8	5.2	5.0

Note: The entries are frequencies of rejection in percentage and the nominal size is 5%.

tests proposed by Durlauf (1991). Deo (2000) demonstrated that Durlauf's test is not robust to conditional heteroskedasticity and may over-reject the MD hypothesis. In our study, we consider the $\mathcal{J}_{\exp(\beta)}$ tests with β the reciprocal of the sample standard deviation of the data, the test of Durlauf (1991) (denoted as CVM), and the modified test of Deo (2000) (denoted as CVM_m). The empirical results are reported in Table 7. While CVM accepts only £ as a serially uncorrelated sequence, CVM_m fails to reject the null hypothesis for another two return series: Can and Fr. By contrast, $\mathcal{J}_{\exp(\tilde{\beta})}$ are significant at 5% level for all series. If the results of CVM_m are more reliable, we may infer that the return series of Can, Fr, and £ are serially uncorrelated, yet they are not MD sequences. Thus, the returns of Can, Fr, and £ may not be linearly predictable but may be predicted using some nonlinear functions of the past returns. This conclusion is different from that of Fong and Ouliaris (1995).

Table 5: Empirical powers of the martingale difference tests ($\varepsilon_t \sim S(1.5)$).

		AR(1)		MA(1)		SETAR-1		SETAR-2		BL-1		BL-2	
	k	$T = 100$	500	100	500	100	500	100	500	100	500	100	500
$\mathcal{J}_{\exp(0.1)}$	1	97.1	99.5	81.1	98.4	46.7	87.5	14.4	33.2	50.7	87.8	2.8	4.3
	2	88.5	99.1	41.7	90.4	43.7	83.8	14.5	31.4	47.9	89.2	7.3	23.5
	3	77.5	97.0	29.7	78.1	44.5	83.8	14.5	49.0	47.3	87.4	5.1	23.7
	4	63.6	94.0	23.9	66.5	40.9	81.7	12.1	37.2	42.2	85.2	4.1	13.2
$\mathcal{J}_{\exp(1.0)}$	1	71.3	92.9	50.5	81.3	20.3	40.8	10.9	22.3	10.8	16.0	3.6	4.0
	2	35.2	74.9	13.2	38.2	6.6	13.5	14.0	36.4	4.4	6.1	27.1	69.5
	3	12.0	36.8	7.0	15.2	5.3	7.8	8.4	23.0	4.9	5.2	18.8	49.8
	4	5.0	14.1	3.6	8.2	4.0	4.0	6.3	11.4	4.5	5.9	6.0	14.8
$\mathcal{J}_{\exp(2.0)}$	1	53.5	86.8	31.7	71.4	14.0	33.1	9.3	16.5	7.2	8.0	3.7	2.1
	2	15.6	42.3	6.4	19.3	4.1	6.7	10.0	25.9	4.1	4.1	22.9	50.3
	3	5.2	9.7	4.6	6.1	2.8	3.6	4.9	12.2	3.2	4.2	11.8	28.6
	4	2.2	5.9	2.2	4.7	3.3	3.2	3.1	5.7	3.4	3.4	2.8	7.9
$\mathcal{J}_{\exp(\tilde{\beta})}$	1	92.9	99.4	72.2	97.5	29.5	81.1	13.7	31.4	34.8	86.2	3.5	2.7
	2	73.3	98.7	29.9	80.6	20.8	65.3	15.0	38.0	34.0	87.7	8.1	22.5
	3	48.3	93.5	19.6	63.7	16.4	60.8	12.8	38.0	32.4	88.2	8.4	25.3
	4	31.9	83.4	12.7	41.4	15.1	54.8	8.6	25.0	31.2	84.5	5.4	10.4
$\mathrm{BP}_{m,5}$		70.2	85.9	34.3	56.7	4.6	5.1	10.5	35.9	2.5	1.5	2.7	3.1
$\mathrm{BP}_{m,10}$		65.9	84.5	30.4	50.4	3.0	5.4	10.2	29.2	3.2	2.5	3.6	3.3
CVM_m		77.3	83.7	56.9	66.6	5.5	6.4	10.6	17.5	1.9	1.5	4.1	2.4

Note: The entries are frequencies of rejection in percentage and the nominal size is 5%.

It is also interesting to see from Table 7 that for the return series of £ and ¥, $\mathcal{J}_{\exp(\tilde{\beta})}$ increases with k. For Dm and Fr, $\mathcal{J}_{\exp(\tilde{\beta})}$ first increases with k and then decreases when $k \geq 4$; for Can, $\mathcal{J}_{\exp(\tilde{\beta})}$ starts increasing from $k = 2$. These are different from the simulation results reported in the last section, where the test statistics decrease with k in general. Thus, these exchange rate returns may depend more on the information in the distant past in a complex way. Fong and Ouliaris (1995) also found that the rejection of the null hypothesis for Dm and ¥ may be due to long memory in data. Granger and Teräsvirta (1999), however, pointed out that a simple nonlinear process with short memory may behave like a long-memory process. As our test is capable of detecting any neglected nonlinearity, the increasing test statistics may be an indication that more lagged returns are needed to capture the dynamics of these return series.

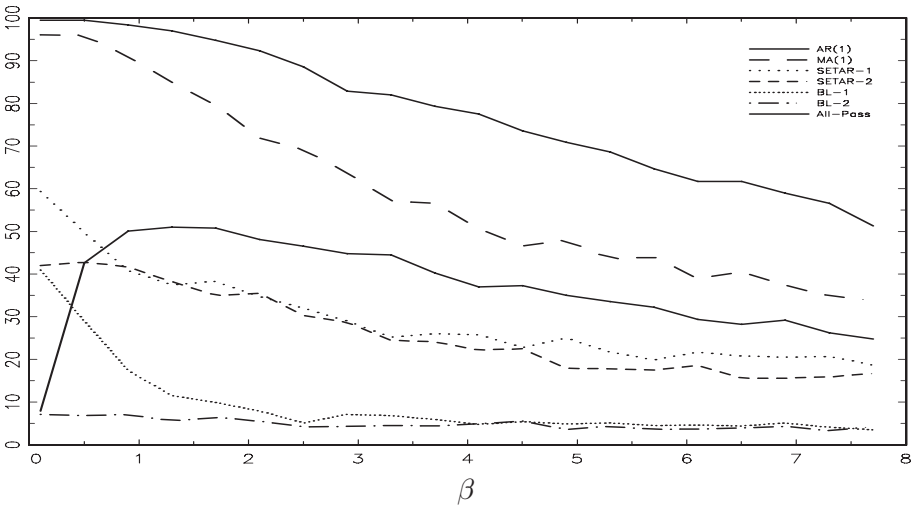


Figure 2: Powers of $\mathcal{J}_{\exp(\beta)}$ with $T = 100$.

Table 6: Summary statistics of exchange rate returns.

	T	Mean	Variance	LRV	Skewness	Kurtosis	t-ratio
Can	760	0.01150	0.06255	0.05303	0.41126	7.67075	1.37725
Dm	760	-0.01785	0.37928	0.55157	-0.20676	4.44518	-0.65266
Fr	760	0.01644	0.36886	0.49808	0.03850	5.20738	0.64228
£	760	0.01974	0.39198	0.57844	-0.22776	5.87089	0.71542
¥	760	-0.04664	0.37003	0.55228	-0.58848	5.49540	-1.73012

Note: The returns are measured in percentages. LRV denotes the long-run variance.

6 Conclusions

In this paper, we introduce a new class of MD tests based on an equivalence relationship between conditional expectation and unconditional moment conditions due to Bierens (1982). The proposed test is easy to compute and is asymptotically pivotal with a standard limiting distribution, whereas the existing consistent tests are not. Moreover, it has power against a much broader class of non-MD alternatives than do the autocorrelation- and spectrum-based tests. Our simulations confirm that this test is capable of detecting a wide variety of non-MD sequences when the conventional tests, such as $\mathrm{BP}_{m,k}$ and CVM_m, can not. Thus, the proposed test is a useful complement to existing tests and allows us to identify nonlinear predictability.

Table 7: Empirical results of various martingale difference tests.

	k	Can	Dm	Fr	£	¥
$\mathcal{J}_{\exp(\tilde{\beta})}$	1	7.50*	11.32*	8.14*	6.90*	13.51*
	2	6.65*	22.03*	20.01*	11.97*	15.11*
	3	12.87*	32.31*	26.12*	18.54*	22.11*
	4	13.68*	26.55*	19.68*	24.80*	23.02*
CVM		0.75*	0.93*	0.51*	0.17	1.23*
CVM_m		0.42	0.73*	0.36	0.11	0.80*

Note: * denotes significance at 5% level. The estimated $\tilde{\beta}$ for Can, Dm, Fr, £, and ¥ are, respectively, 3.99, 1.62, 1.64, 1.60, and 1.64.

The proposed $\mathcal{J}_{\exp(\beta)}$ test may be further improved. We have seen that the power of this test heavily depends on the parameter β for the weighting function and that different data series may require different choices of β. It is not clear at this time how to find an "optimal" β for each series. More generally, there may be other weighting functions that could generate even more powerful tests. The weighting function considered in this paper is chosen mainly for convenience. It would be desirable if there is a better criterion to determine the weighting function. Note that existing tests that require a weighting function, such as those of Chen, Chou, and Kuan (2000) and Bierens (1982), also suffer from the same difficulty. These topics are quite challenging and currently being investigaed.

Appendix

Proof of Theorem 2.1:

We know that $\{\psi_{j,g}(Y_t, \boldsymbol{Y}_{t-1,k})\}$ are two MD sequences whenever $\{Y_t\}$ is. We also know that $\psi_{j,g}(Y_t, \boldsymbol{Y}_{t-1,k})$ are stationary and ergodic under [S1] and that $\{\psi_{j,g}(Y_t, \boldsymbol{Y}_{t-1,k})\}$ are mixing sequences with the same mixing coefficients as Y_t under [M1]. Making use of the Cramér-Wold device we consider

$$\begin{aligned} Z_{t,k} &= \begin{bmatrix} \lambda_1 \\ \lambda_2 \end{bmatrix}' \begin{bmatrix} \sigma_{c,g}^2 & \sigma_{cs,g} \\ \sigma_{cs,g} & \sigma_{s,g}^2 \end{bmatrix}^{-1/2} \begin{bmatrix} \psi_{c,g}(Y_t, \boldsymbol{Y}_{t-1,k}) \\ \psi_{s,g}(Y_t, \boldsymbol{Y}_{t-1,k}) \end{bmatrix} \\ &=: \theta_c \psi_{c,g}(Y_t, \boldsymbol{Y}_{t-1,k}) + \theta_s \psi_{s,g}(Y_t, \boldsymbol{Y}_{t-1,k}), \end{aligned}$$

where λ_1 and λ_2 are such that $\lambda_1^2 + \lambda_2^2 = 1$. Given [S1] and the MD property,

$$\frac{1}{T-k} \operatorname{var}\left(\sum_{t=k+1}^{T} Z_{t,k} \right) = \operatorname{var}(Z_{t,k}) = 1,$$

so that $(T-k)^{-1/2} \sum_{t=k+1}^{T} Z_{t,k} \xrightarrow{D} \mathcal{N}(0,1)$ by Theorem 5.15 of White (1984). Given [M1] and the MD property,

$$\frac{1}{T-k} \operatorname{var}\left(\sum_{t=k+1}^{T} Z_{t,k} \right) = \frac{1}{T-k} \sum_{t=k+1}^{T} \mathbb{E}(Z_{t,k}^2) = 1,$$

so that $(T-k)^{-1/2} \sum_{t=k+1}^{T} Z_{t,k} \xrightarrow{D} \mathcal{N}(0,1)$ by Corollary 5.25 of White (1984). As the results above hold for arbitrary λ_1 and λ_2 such that $\lambda_1^2 + \lambda_2^2 = 1$, the asserted asymptotic normality follows immediately. □

Proof of Corollary 2.2:

Given the strong consistency of $\hat{\sigma}_{j,g}^2$ and $\hat{\sigma}_{cs,g}$ under [S1] or [M1], the limiting $\chi^2(2)$ distribution of $\mathcal{J}_g$ follows immediately from Theorem 2.1 and the continuous mapping theorem. When $\mathbb{E}[\psi_{j,g}(Y_t, \boldsymbol{Y}_{t-1,k})] = \delta_j \neq 0$,

$$\begin{bmatrix} \hat{\sigma}_{c,g}^2 & \hat{\sigma}_{cs,g} \\ \hat{\sigma}_{cs,g} & \hat{\sigma}_{s,g}^2 \end{bmatrix} \xrightarrow{\mathbb{P}} \begin{bmatrix} \sigma_{c,g}^2 & \sigma_{cs,g} \\ \sigma_{cs,g} & \sigma_{s,g}^2 \end{bmatrix} + \begin{bmatrix} \delta_c^2 & \delta_c \delta_s \\ \delta_c \delta_s & \delta_s^2 \end{bmatrix} =: \boldsymbol{V} + \boldsymbol{M},$$

where $\xrightarrow{\mathbb{P}}$ stands for convergence in probability. It is then easy to verify that

$$\begin{bmatrix} \hat{\sigma}_{c,g}^2 & \hat{\sigma}_{cs,g} \\ \hat{\sigma}_{cs,g} & \hat{\sigma}_{s,g}^2 \end{bmatrix}^{-1/2} \begin{bmatrix} \bar{\psi}_{c,g}(Y_t, \boldsymbol{Y}_{t-1,k}) \\ \bar{\psi}_{s,g}(Y_t, \boldsymbol{Y}_{t-1,k}) \end{bmatrix} \xrightarrow{\mathbb{P}} (\boldsymbol{V} + \boldsymbol{M})^{-1/2} \boldsymbol{d},$$

where $\boldsymbol{d}' = (\delta_c, \delta_s)$. Consequently,

$$\frac{1}{T-k}\mathcal{J}_g \xrightarrow{\mathbb{P}} \boldsymbol{d}'(\boldsymbol{V}+\boldsymbol{M})^{-1}\boldsymbol{d},$$

proving the asserted divergence under the alternative. □

Multivariate Exponential Distribution

We now derive the exact expressions of

$$\phi(\boldsymbol{\nu}) := \mathbb{E}\left[\exp\left(i\sum_{j=1}^{k}\nu_j\omega_j\right)\right] = \{\det[\boldsymbol{I}_k - \beta\, i\boldsymbol{R}\boldsymbol{D}_{\boldsymbol{\nu}}]\}^{-1},$$

for $k = 1, 2, 3, 4$.

A.1 $k = 1$

In this case, $\boldsymbol{I}_k = \boldsymbol{R} = 1$ and $\boldsymbol{D}_{\omega} = \omega_1$. The characteristic function is

$$\phi(\nu_1) = \mathbb{E}\left[\exp\left(i\nu_1\omega_1\right)\right] = \frac{1}{1-\beta i\nu_1} = \frac{1}{1+\beta^2\nu_1^2} + i\,\frac{\beta\nu_1}{1+\beta^2\nu_1^2}.$$

Thus, ω_1 is simply an exponential random variable with mean being β.

A.2 $k = 2$

Consider mpw the case that ω_1 and ω_2 have a bivariate exponential distribution. Let

$$\boldsymbol{R} = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix},$$

where ρ is the correlation between ω_1 and ω_2. Then,

$$\boldsymbol{I}_2 - \beta\, i\boldsymbol{R}\boldsymbol{D}_{\boldsymbol{\nu}} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \beta\, i\begin{bmatrix} \nu_1 & \rho\nu_2 \\ \rho\nu_1 & \nu_1 \end{bmatrix} = \begin{bmatrix} 1-\beta\, i\nu_1 & -\beta\, i\nu_2 \\ -\beta\, i\rho\nu_1 & 1-\beta\, i\rho\nu_1 \end{bmatrix},$$

with the determinant $\left[1-\beta^2\left(1-\rho^2\right)\nu_1\nu_2\right] - \beta\, i(\nu_1+\nu_2)$. Hence,

$$\begin{aligned}\phi(\boldsymbol{\nu}) &= \left\{\left[1-\beta^2(1-\rho^2)\nu_1\nu_2\right] - \beta\, i(\nu_1+\nu_2)\right\}^{-1} \\ &= \frac{1-\beta^2(1-\rho^2)\nu_1\nu_2}{\lambda} + i\frac{\beta(\nu_1+\nu_2)}{\lambda},\end{aligned}$$

where

$$\lambda = 1 + 2\beta^2\rho^2\nu_1\nu_2 + \beta^4\left(1-\rho^2\right)^2\nu_1^2\nu_2^2 + \beta^2(\nu_1^2+\nu_2^2).$$

Note that when $\rho = 0$, $\lambda = (1+\beta^2\nu_1^2)(1+\beta^2\nu_2^2)$ and ω_1 and ω_2 are independent because $\mathbb{E}[\exp(i\nu_1\omega_1 + i\nu_2\omega_2)] = \mathbb{E}[\exp(i\nu_1\omega_1)]\,\mathbb{E}[\exp(i\nu_2\omega_2)]$ in this case.

A.3 $k = 3$

Let ρ_j, $j = 1, 2$, denote the correlations between ω_1 and ω_{1+j} and

$$\boldsymbol{R} = \begin{bmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_1 \\ \rho_2 & \rho_1 & 1 \end{bmatrix},$$

It can be verified that

$$\begin{aligned} \boldsymbol{I}_3 - \beta\, i\boldsymbol{R}\boldsymbol{D}_{\boldsymbol{\nu}} &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} - \beta\, i \begin{bmatrix} \nu_1 & \rho_1\nu_2 & \rho_2\nu_3 \\ \rho_1\nu_1 & \nu_2 & \rho_1\nu_3 \\ \rho_2\nu_1 & \rho_1\nu_2 & \nu_3 \end{bmatrix} \\ &= \begin{bmatrix} 1-\beta\, i\nu_1 & -\beta\, i\rho_1\nu_2 & -\beta\, i\rho_2\nu_3 \\ -\beta\, i\rho_1\nu_1 & 1-\beta\, i\nu_2 & -\beta\, i\rho_1\nu_3 \\ -\beta\, i\rho_2\nu_1 & -\beta\, i\rho_1\nu_2 & 1-\beta\, i\nu_3 \end{bmatrix}. \end{aligned}$$

The determinant of $\boldsymbol{I}_3 - \beta\, i\boldsymbol{R}\boldsymbol{D}_{\boldsymbol{\nu}}$ then is $\gamma_c + i\gamma_s$, where

$$\begin{aligned} \gamma_c &= 1 - \beta^2\big[(1-\rho_1^2)(\nu_1\nu_2 + \nu_2\nu_3) + (1-\rho_2^2)\nu_1\nu_3\big], \\ \gamma_s &= \beta^3\nu_1\nu_2\nu_3\big[1 - 2\rho_1^2(1-\rho_2) - \rho_2^2\big] - \beta(\nu_1+\nu_2+\nu_3). \end{aligned}$$

Therefore, the resulting characteristic function can be expressed as

$$\phi(\boldsymbol{\nu}) = \frac{\gamma_c}{\gamma_c^2+\gamma_s^2} - i\frac{\gamma_s}{\gamma_c^2+\gamma_s^2}.$$

Note that if $\rho_1 = \rho$ and $\rho_2 = \rho^2$, then γ_c and γ_s become

$$\begin{aligned} \gamma_c &= 1 - \beta^2\left(1-\rho^2\right)\left[\nu_1\nu_2 + \nu_2\nu_3 + \left(1+\rho^2\right)\nu_1\nu_3\right], \\ \gamma_s &= \beta^3\nu_1\nu_2\nu_3\left[1 - \rho^2\left(2-\rho^2\right)\right] - \beta(\nu_1+\nu_2+\nu_3). \end{aligned}$$

A.4 $k = 4$

Let $\rho_j, j = 1, 2, 3$, denote the correlations between ω_1 and ω_{1+j}. Then the real and imaginary parts of $\det[\boldsymbol{I}_4 - \beta\, i\boldsymbol{R}\boldsymbol{D}_{\boldsymbol{\nu}}]$ are

$$\begin{aligned} \gamma_c &= 1 - \beta^2\big[(\nu_1+\nu_3)(\nu_2+\nu_4) + (1-\rho_2^2)(\nu_1\nu_3+\nu_2\nu_4)\big] \\ &\quad + \beta^4\big(1 + 2\rho_1^3\rho_3 + \rho_2^4 - \rho_1^2\rho_3^2 - \rho_1^4 - 2\rho_2^2\big)\nu_1\nu_2\nu_3\nu_4, \\ \gamma_s &= \beta^3(1-\rho_2^2)\big[(\nu_1+\nu_2)\nu_3\nu_4 + (\nu_3+\nu_4)\nu_1\nu_2\big] - \beta(\nu_1+\nu_2+\nu_3+\nu_4), \end{aligned}$$

and

$$\phi(\boldsymbol{\nu}) = \frac{\gamma_c}{\gamma_c^2+\gamma_s^2} - i\frac{\gamma_s}{\gamma_c^2+\gamma_s^2}.$$

An Algorithm for Calculating $\psi_{c,g}$ and $\psi_{s,g}$

A major difficulty in calculating $\psi_{j,g}(Y_t, \boldsymbol{Y}_{t-1,k}) = Y_t \varphi_{c,g}(\boldsymbol{Y}_{t-1,k})$ is that the derivation of the precise form of $\varphi_{j,g}$ is quite involved when k is large. For simplicity, we consider the weighting function: $g(\boldsymbol{\omega}) = \prod_{i=1}^{k} g_i(\omega_i)$, so that

$$\varphi_{c,g}(\boldsymbol{Y}_{t-1,k}) = \int_{\mathbb{R}^{k+}} \cos(\boldsymbol{\omega}'\boldsymbol{Y}_{t-1,k}) \prod_{i=1}^{k} g_i(\omega_i)\, \mathrm{d}\boldsymbol{\omega},$$

$$\varphi_{s,g}(\boldsymbol{Y}_{t-1,k}) = \int_{\mathbb{R}^{k+}} \sin(\boldsymbol{\omega}'\boldsymbol{Y}_{t-1,k}) \prod_{i=1}^{k} g_i(\omega_i)\, \mathrm{d}\boldsymbol{\omega}.$$

Using an identity of the sine and cosine functions, we have

$$\begin{aligned}
\cos\left(\sum_{j=1}^{k} \omega_j Y_{t-j}\right) &= \cos(\omega_k Y_{t-k}) \cos\left(\omega_1 Y_{t-1} + \cdots + \omega_{k-1} Y_{t-k+1}\right) \\
&\quad - \sin(\omega_k Y_{t-k}) \sin\left(\omega_1 Y_{t-1} + \cdots + \omega_{k-1} Y_{t-k+1}\right), \\
\sin\left(\sum_{j=1}^{k} \omega_j Y_{t-j}\right) &= \sin(\omega_k Y_{t-k}) \cos\left(\omega_1 Y_{t-1} + \cdots + \omega_{k-1} Y_{t-k+1}\right) \\
&\quad + \cos(\omega_k Y_{t-k}) \sin\left(\omega_1 Y_{t-1} + \cdots + \omega_{k-1} Y_{t-k+1}\right).
\end{aligned}$$

Accordingly, it can be easily shown that

$$\begin{aligned}
\varphi_{c,g}(\boldsymbol{Y}_{t-1,k}) =& \varphi_{c,g}(\boldsymbol{Y}_{t-1,k-1}) \int_{\mathbb{R}^{+}} \cos(\omega_k Y_{t-k}) g_k(\omega_k)\, \mathrm{d}\omega_k \\
&- \varphi_{s,g}(\boldsymbol{Y}_{t-1,k-1}) \int_{\mathbb{R}^{+}} \sin(\omega_k Y_{t-k}) g_k(\omega_k)\, \mathrm{d}\omega_k \\
=& \varphi_{c,g}(\boldsymbol{Y}_{t-1,k-1}) \varphi_{c,g_k}(Y_{t-k}) - \varphi_{s,g}(\boldsymbol{Y}_{t-1,k-1}) \varphi_{s,g_k}(Y_{t-k}), \\
\varphi_{s,g}(Y_{t-1,k}) =& \varphi_{c,g}(\boldsymbol{Y}_{t-1,k-1}) \int_{\mathbb{R}^{+}} \sin(\omega_k Y_{t-k}) g_k(\omega_k)\, \mathrm{d}\omega_k \\
&+ \varphi_{s,g}(\boldsymbol{Y}_{t-1,k-1}) \int_{\mathbb{R}^{+}} \cos(\omega_k Y_{t-k}) g_k(\omega_k)\, \mathrm{d}\omega_k \\
=& \varphi_{c,g}(\boldsymbol{Y}_{t-1,k-1}) \varphi_{s,g_k}(Y_{t-k}) + \varphi_{s,g}(\boldsymbol{Y}_{t-1,k-1}) \varphi_{c,g_k}(Y_{t-k}).
\end{aligned}$$

Thus, $\varphi_{c,g}(\boldsymbol{Y}_{t-1,k})$ and $\varphi_{s,g}(\boldsymbol{Y}_{t-1,k})$ can be obtained recursively even without explicit formulas. The algorithm in the text is a special case with g_i being the same function.

Bierens, H. J. (1982). Consistent model specification tests, *Journal of Econometrics*, **20**, 105–134.

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. and W. Ploberger (1997). Asymptotic theory of integrated conditional moment tests, *Econometrica*, **65**, 1129–1151.

Box, G. E. P. and D. A. Pierce (1970). Distribution of residual autocorrelations in autoregressive-integrated moving average time series models, *Journal of the American Statistical Association*, **65**, 1509–1526.

Breidt, F. J., R. A. Davis, and A. Trindade (2000). Least absolute deviation estimation for all-pass time series models, Working Paper, Iowa State University.

Chen, Y.-T. (2002). A new class of characteristic-function-based distribution tests and its application to GARCH model, Working paper, ISSP, Academia Sinica.

Chen, Y.-T., R. Y. Chou, and C.-M. Kuan (2000). Testing time reversibility without moment restrictions, *Journal of Econometrics*, **95**, 199–218.

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

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

de Lima, P. J. F. (1997). On the robustness of nonlinearity tests to moment condition failure, *Journal of Econometrics*, **76**, 251–280.

Dominguez, M. A. and I. N. Lobato (2000). A consistent test for the martingale difference hypothesis, Working Paper, Instituto Tecnologico Autonomo de Mexico.

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

Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work, *Journal of Finance*, **25**, 383–417.

Fama, E. F. (1991), Efficient capital markets: II, *Journal of Finance*, **46**, 1575–1618.

Feuerverger, A. and R. A. Mureika (1977). The empirical characteristic function and its

applications, *The Annals of Statistics*, **5**, 88–97.

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

Granger, C. W. J. and T. Teräsvirta (1999). A simple nonlinear time series model with misleading linear properties, *Economics Letters*, **62**, 161–165.

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.

Johnson, N. L. and S. Kotz (1972). *Distributions in Statistics: Continuous Multivariate Distributions*, New York: John Wiley.

Keenan, D. M. (1985). A Tukey nonadditivity-type test for time series nonlinearity, *Biometrika*, **72**, 39–44.

LeRoy, S. F. (1989). Efficient capital markets and martingales, *Journal of Economic Literature*, **27**, 1583–1621.

Liu, C. Y. and J. He (1991). A variance-ratio test of random walks in foreign exchange rates, *Journal of Finance*, **36**, 773–785.

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

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

Loretan, M. and P. C. B. Phillips (1994). Testing the covariance stationarity of heavy-tailed time series, *Journal of Empirical Finance*, **1**, 211–248.

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

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

Tsay, R. S. (1986). Nonlinearity tests for time series, *Biometrika*, **73**, 461–466.

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

Whang, Y.-J. (2001). Consistent specification testing for conditional moment restrictions, *Economics Letters*, **71**, 299–306.

White, H. (1984). Asymptotic Theory for Econometricians, Orlando: Academic Press.

Number	Author(s)	Title	Date
03-A001	Chung-Ming Kuan Wei-Ming Lee	A New Test of the martingale Difference Hypothesis	11/03