
Volume 162
Issue 2
June 2011
ISSN 0304-4076
ELSEVIER
JOURNAL OF
Econometrics
Editors:
TAKESHI AMEMIYA
A. RONALD GALLANT
JOHN F. GEWEKE
CHENG HSIAO
PETER ROBINSON
ARNOLD ZELLNER
Associate Editors:
YACINE AÏT-SAHALIA
BADI H. BALTAGI
RAVI BANSAL
MARCUS J. CHAMBERS
SONGNIAN CHEN
XIAOHONG CHEN
MIKHAIL CHERNOV
VICTOR CHERNOZHUKOV
MANFRED DEISTLER
MIGUEL A. DELGADO
YANQIN FAN
SYLVIA FRUHWIRTH-
-SCHNATTER
ERIC GHYSELS
JOHN C. HAM
JAVIER HIDALGO
HAN HONG
MICHAEL KEANE
YUICHI KITAMURA
GARY M. KOOP
NAOTO KUNITOMO
KAJAL LAHIRI
QI LI
TONG LI
ROSA L. MATZKIN
FRANCESCA MOLINARI
FRANZ C. PALM
DALE J. POIRIER
BENEDIKT PÖTSCHER
INGMAR PRUCHA
ERIC RENAULT
ROBIN SICKLES
FALLAW SOWELL
MARK STEEL
DAG BJARNE TJOESTHEIM
HERMAN VAN DIJK
QUANG H. VUONG
EDWARD VYTLACIL
TOM WANSBEEK
TAO ZHA
CONTENTS
O.E. Barndorff-Nielsen, P.R. Hansen, A. Lunde and N. Shephard, Multivariate realised kernels: Consistent positive semi-definite estimators of the covariation of equity prices with noise and non-synchronous trading 149
A. Lewbel, D. McFadden and O. Linton, Estimating features of a distribution from binomial data 170
Z. Song, A martingale approach for testing diffusion models based on infinitesimal operator 189
X. Shao, A bootstrap-assisted spectral test of white noise under unknown dependence 213
Z. Zhao, Nonparametric model validations for hidden Markov models with applications in financial econometrics 225
U. Hassler, Estimation of fractional integration under temporal aggregation 240
T. Oka and Z. Qu, Estimating structural changes in regression quantiles 248
Y. Fan, M. Gentry and T. Li, A new class of asymptotically efficient estimators for moment condition models 268
A. Holly, A. Monfort and M. Rockinger, Fourth order pseudo maximum likelihood methods 278
N. Sizova, Integrated variance forecasting: Model based vs. reduced form 294
CONTENTS CONTINUED ON BACK COVER

Journal of Econometrics 162 (2011) 213–224

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

A bootstrap-assisted spectral test of white noise under unknown dependence

Xiaofeng Shao *

Department of Statistics, University of Illinois at Urbana-Champaign, 725 South Wright St, Champaign, IL, 61820, United States

ARTICLE INFO

Article history:
Received 16 October 2008
Received in revised form
24 December 2010
Accepted 20 January 2011
Available online 27 January 2011

JEL classification:
C12
C20

Keywords:
Hypothesis testing
Spectral distribution function
Time series
White noise
Wild bootstrap

ABSTRACT

To test for the white noise null hypothesis, we study the Cramér–von Mises test statistic that is based on the sample spectral distribution function. Since the critical values of the test statistic are difficult to obtain, we propose a blockwise wild bootstrap procedure to approximate its asymptotic null distribution. Using a Hilbert space approach, we establish the weak convergence of the difference between the sample spectral distribution function and the true spectral distribution function, as well as the consistency of bootstrap approximation under mild assumptions. Finite sample results from a simulation study and an empirical data analysis are also reported.

1. Introduction

This article is concerned with testing if a sequence of observations is generated from a white noise process. Testing for white noise, or lack of serial correlation, is a classical problem in time series analysis and it is an integral part of the Box–Jenkins linear modeling framework. For example, if the series is white noise, then there is no need to fit an ARMA-type linear time series model to the data, which is mainly used to model nonzero auto-correlations. We consider a stationary real-valued time series $X_t, t \in \mathbb{Z}$. Denote by $\mathbb{E}(X_t) = \mu$ and $\gamma(k) = \mathrm{cov}(X_t, X_{t+k})$. Then the null and alternative hypothesis are

$$H_0 : \gamma(k) = 0, \quad k \in \mathbb{N} \quad \text{versus} \quad H_1 : \gamma(k) \neq 0$$

for some $k \in \mathbb{N}$.

Let

$$f(w) = (2\pi)^{-1} \sum_{k=-\infty}^{\infty} \gamma(k) \mathrm{e}^{-ikw}, \quad w \in [-\pi, \pi] \quad \text{and}$$

$$F(\lambda) = \int_0^{\lambda} f(w) \mathrm{d}w, \quad \lambda \in [0, \pi]$$

be the spectral density function and spectral distribution function respectively. Under the null hypothesis, $f(\lambda) = \gamma(0)/(2\pi)$ is a constant over $[-\pi, \pi]$. There is a huge literature on the white noise testing problem and the existing tests can be roughly categorized into two types: time domain correlation-based tests and frequency domain periodogram-based tests. One of the challenges for the existing methods is that the asymptotic null distributions of the test statistics have been obtained under the independent and identically distributed (iid) assumption, and they may be invalid when the series is uncorrelated but dependent. The distinction between iid and uncorrelated dependence is important in view of the popularity of conditional heteroscedastic models (e.g. GARCH models) as used in econometric and financial time series modeling. Also there are other nonlinear processes that are uncorrelated but not independent; see Remark 2.2 for more examples.

The discrepancy in asymptotic null distributions can be attributed to the unknown dependence of the series, which leads to severe size distortion if we use the critical values of the asymptotic null distribution obtained under iid assumptions. This phenomenon was first discovered for the popular Box and Pierce's (1970) (BP, hereafter) portmanteau test by Romano and Thombs (1996), who advocated the use of nonparametric bootstrap and subsampling methods to approximate the sampling distribution of the BP test statistic. Also see Lobato (2001), Lobato et al. (2002) and Horowitz et al. (2006) for more recent contributions along this line. The BP test statistic and its variants only test for zero correlations up to lag m, where m is a prespecified integer, so they have trivial power against correlations at lags beyond m. Recently, Escanciano and Lobato (2009) proposed a data-driven BP test for serial correlation. Their work is interesting as it avoids

* Tel.: +1 217 2447285; fax: +1 217 2447190.
E-mail address: xshao@uiuc.edu.

0304-4076/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.jeconom.2011.01.001

any user-chosen number (say, m), does not involve the use of bootstrap and is shown to have good size and power performance. However, the asymptotic null distribution of their test statistic was established under the martingale difference assumption and it is unclear whether the limiting theory for their test still holds under the weaker white noise hypothesis. For time series models that are white noise but are not martingale difference, see Remark 2.2.

The main focus of this paper is on the test statistic that is based on the sample spectral distribution function $F_n(\lambda) = \int_0^\lambda I_n(w)\mathrm{d}w$, where

$$I_n(w) = (2\pi n)^{-1} \sum_{t=1}^{n} |(X_t - \bar{X})e^{itw}|^2$$

is the periodogram and $\bar{X} = n^{-1}\sum_{t=1}^n X_t$ is the sample mean. The early work on spectra-based tests dates back to Bartlett (1955), who proposed famous U_p and T_p processes in the frequency domain. A rigorous asymptotic theory was obtained by Grenander and Rosenblatt (1957) for the iid processes. More recent developments can be found in Durlauf (1991), Anderson (1993) and Deo (2000), to name a few. On one hand, it has been well-established in the literature that the spectra-based test belongs to the test of omnibus type, and it has nontrivial power against local alternatives that are within $n^{-1/2}$-neighborhoods of the null hypothesis. On the other hand, as demonstrated in Deo (2000), the spectra-based test no longer has the usual asymptotic distribution under iid assumption when dependence (e.g. conditional heteroscedasticity) is present. To account for unknown conditional heteroscedasticity, the latter author proposed a modified test statistic and justified its use for a subclass of GARCH processes and stochastic volatility models. However, the validity of Deo's modification heavily relies on certain assumptions on the data generating process, which exclude some other commonly used GARCH models (e.g. EGARCH model) and uncorrelated nonlinear processes. Also those assumptions are difficult to check in practice.

In this paper, we propose a blockwise wild bootstrap approach to approximating the asymptotic null distribution of the spectra-based test statistic. The blockwise wild bootstrap is a simple variant of the well-known wild bootstrap (Wu, 1986) in that the auxiliary variables are generated in a blockwise fashion. We establish the consistency of the bootstrap approximation using the Hilbert space approach, which allows us to get the limiting distribution of the Cramér–von Mises statistic. Assisted with the blockwise wild bootstrap, our test is asymptotically valid for a very wide class of processes, including GARCH processes of various forms and other uncorrelated linear or nonlinear processes.

We now introduce some notation. For a column vector $x = (x_1, \ldots, x_q)' \in \mathbb{R}^q$, let $|x| = (\sum_{j=1}^q x_j^2)^{1/2}$. Let ξ be a random vector. Write $\xi \in \mathcal{L}^p$ $(p > 0)$ if $\|\xi\|_p := [\mathbb{E}(|\xi|^p)]^{1/p} < \infty$ and let $\|\cdot\| = \|\cdot\|_2$. The positive constant C is generic and it may vary from line to line. Denote by "$\rightarrow_D$" and "$\rightarrow_p$" convergence in distribution and in probability, respectively. The symbols $O_P(1)$ and $o_p(1)$ signify being bounded in probability and convergence to zero in probability, respectively. The paper is structured as follows. In Section 2, we present the spectra-based test and establish the weak convergence of the difference between the sample spectral distribution function and its population counterpart. Section 3 proposes a blockwise wild bootstrap approach and proves its consistency. We illustrate the finite sample performance of the test using a Monte Carlo experiment and real data sets in Section 4. Section 5 concludes. Proofs are gathered in the Appendix.

2. Test statistics and asymptotic theory

To illustrate the idea, we note the relationship $I_n(w) = (2\pi)^{-1}\sum_{k=1-n}^{n-1}\hat{\gamma}(k)e^{-ikw}$, where $\hat{\gamma}(j) = n^{-1}\sum_{t=1+|j|}^{n}(X_t-\bar{X})(X_{t-|j|}-\bar{X})$ is the sample auto-covariance function at lag $j = 0, \pm 1, \ldots, \pm(n-1)$. Write $F_n(\lambda) = \sum_{j=0}^{n-1}\hat{\gamma}(j)\Psi_j(\lambda)$ and $F(\lambda) = \sum_{j=0}^{\infty}\gamma(j)\Psi_j(\lambda)$, where $\Psi_j(\lambda) = \sin(j\lambda)/(j\pi)\mathbf{1}(j \neq 0) + \lambda/(2\pi)\mathbf{1}(j = 0)$. Under H_0, $F(\lambda) = \gamma(0)\Psi_0(\lambda)$, so we can construct a test based on the distance between $F_n(\lambda)$ and $\hat{\gamma}(0)\Psi_0(\lambda)$, the latter of which is a natural estimator of $F(\lambda)$ under the null hypothesis. This leads to the process

$$S_n(\lambda) = \sqrt{n}\{F_n(\lambda) - \hat{\gamma}(0)\Psi_0(\lambda)\} = \sum_{j=1}^{n-1}\sqrt{n}\hat{\gamma}(j)\Psi_j(\lambda).$$

Let $Y_t = X_t - \mu$. Under the null hypothesis, we expect that $S_n(\lambda)$ weakly converges to a Gaussian process $S(\lambda)$ in $C[0,\pi]$ (the space of continuous functions on $[0,\pi]$), where $S(\lambda)$ has mean zero and covariance function

$$\mathrm{cov}\{S(\lambda), S(\lambda')\} = \sum_{j=1}^{\infty}\sum_{k=1}^{\infty}\sum_{d=-\infty}^{\infty}\mathrm{cov}(Y_tY_{t-j}, Y_{t-d}Y_{t-d-k})\Psi_j(\lambda)\Psi_k(\lambda'). \tag{1}$$

The continuous mapping theorem yields the limiting distributions of the following two well-known statistics:

Kolmogorov–Smirnov (K–S) statistic

$$\sup_{\lambda\in[0,\pi]}|S_n(\lambda)| \rightarrow_D \sup_{\lambda\in[0,\pi]}|S(\lambda)|, \tag{2}$$

Cramér von-Mises (C–M) statistic

$$\mathrm{CM}_n = \int_0^\pi S_n^2(\lambda)\mathrm{d}\lambda \rightarrow_D \int_0^\pi S^2(\lambda)\mathrm{d}\lambda. \tag{3}$$

The covariance structure of $S(\lambda)$ can be simplified considerably if the series is iid Gaussian under H_0, and one may be able to exploit the Gaussian assumption to consistently approximate the limiting distributions of the K–S and C–M statistics. But for dependent white noise, such as a GARCH process, the critical values of $\sup_{\lambda\in[0,\pi]}|S(\lambda)|$ and $\int_0^\pi S^2(\lambda)\mathrm{d}\lambda$ are typically hard to obtain. To circumvent the difficulty, Deo (2000) proposed a modified version of K–S and C–M statistics for the standardized version of $S_n(\lambda)$, where $\hat{\gamma}(j)$ is replaced by its autocorrelation counterpart $\hat{\rho}(j) = \hat{\gamma}(j)/\hat{\gamma}(0)$. He established the validity of the modified test statistic under the assumptions that ensure asymptotic independence of $\hat{\rho}(j)$ and $\hat{\rho}(k)$, $j \neq k$. His method seems no longer (asymptotically) valid if the asymptotic independence does not hold and also his assumptions are hard to check in practice. Therefore, it is desirable to come up with a general solution that is capable of approximating the limiting null distributions of K–S and C–M statistics.

To facilitate asymptotic analysis, we shall adopt the Hilbert space approach to establish the weak convergence of S_n. As demonstrated in Escanciano and Velasco (2006), the Hilbert space approach is superior to the usual "sup" norm approach in proving the weak convergence due to its lack of theoretical complication in general settings. In particular, the proof of the tightness using the Hilbert space approach is simpler than that for the "sup" norm approach. Here, we regard S_n as a random element in the Hilbert space $L_2[0,\pi]$ of all square integrable functions (with respect to the Lebesgue measure) with the inner product

$$\langle f, g\rangle = \int_{[0,\pi]} f(\lambda)g^c(\lambda)\mathrm{d}\lambda,$$

where $g^c(\lambda)$ denotes the complex conjugate of $g(\lambda)$. Note that $L_2[0,\pi]$ is endowed with the natural Borel σ-field induced by the norm $\|f\| = \langle f, f\rangle^{1/2}$; see Parthasarathy (1967). The Hilbert space approach has been utilized in time series analysis by Politis and Romano (1994), Chen and White (1996, 1998) and Escanciano and Velasco (2006), among others. From both methodological and technical perspectives, this paper is closely related to Escanciano and Velasco (2006), who proposed a generalized spectral distribution function type test for the null hypothesis of martingale difference and adopt the wild bootstrap to get a consistent approximation of the asymptotic null distribution. In our testing problem, a blockwise wild bootstrap is necessary to yield consistent approximation and our technical developments are considerably different from theirs since our white noise null hypothesis is weaker than their martingale difference hypothesis.

To obtain the weak convergence of S_n in $L_2[0,\pi]$, we need to impose some structural assumption on the process X_t. Throughout, we assume that

$$X_t = G(\ldots, \varepsilon_{t-1}, \varepsilon_t), \quad t \in \mathbb{Z}, \tag{4}$$

where G is a measurable function for which X_t is a well-defined random variable. The class of processes that (4) represents is huge; see Rosenblatt (1971), Kallianpur (1981), and Tong (1990, p. 204). Let $(\varepsilon'_k)_{k\in\mathbb{Z}}$ be an iid copy of $(\varepsilon_k)_{k\in\mathbb{Z}}$; define the physical dependence measure (Wu, 2005) $\delta_q(t) = \|X_t - X_t^*\|_q$, $q \ge 1$, where $X_t^* = G(\ldots, \varepsilon_{-1}, \varepsilon'_0, \varepsilon_1, \ldots, \varepsilon_t)$, $t \in \mathbb{N}$. For $\xi \in \mathcal{L}^1$ define projection operators $\mathcal{P}_k\xi = \mathbb{E}(\xi|\mathcal{F}_k) - \mathbb{E}(\xi|\mathcal{F}_{k-1})$, $k \in \mathbb{Z}$, where $\mathcal{F}_k = (\ldots, \varepsilon_{k-1}, \varepsilon_k)$.

Theorem 2.1. *Assume the process X_t admits the representation (4), $X_t \in \mathcal{L}^4$ and*

$$\sum_{k=0}^{\infty} \delta_4(k) < \infty. \tag{5}$$

Further we assume that

$$\sum_{k_1,k_2,k_3=-\infty}^{\infty} |\mathrm{cum}(X_0, X_{k_1}, X_{k_2}, X_{k_3})| < \infty. \tag{6}$$

Then $S_n(\lambda) - \mathbb{E}\{S_n(\lambda)\} \Rightarrow S(\lambda)$ in $L_2[0,\pi]$, where $S(\lambda)$ is a mean zero Gaussian process with the covariance specified in (1) and "$\Rightarrow$" stands for weak convergence in the Hilbert space $L_2[0,\pi]$ endowed with the norm metric.

Corollary 2.1. *Under the assumptions of Theorem 2.1, (3) holds under H_0. Under H_1, we have that*

$$\frac{\mathrm{CM}_n}{n} \to_p \sum_{j=1}^{\infty} \gamma^2(j) \int_0^{\pi} \Psi_j^2(\lambda) \mathrm{d}\lambda. \tag{7}$$

Thus the C–M test is consistent since at least one $\gamma(j) \neq 0$ for $j = 1, 2, \ldots$ under the alternative. The Hilbert space approach is very natural for studying the asymptotic properties of the statistic of Cramér–von Mises type. Unfortunately, it is not applicable to the statistic of Kolmogorov–Smirnov type, since the "sup" functional is no longer a continuous mapping from $L_2[0,\pi]$ to $\mathbb{R}$. This is a price we pay for the reduced technicality of the Hilbert space approach as compared to the "sup" norm approach.

Remark 2.1. Note that

$$\gamma(k) = \mathrm{cov}(X_0, X_k) = \mathbb{E}\left(\sum_{j\in\mathbb{Z}} \mathcal{P}_j X_0 \sum_{j'\in\mathbb{Z}} \mathcal{P}_{j'} X_k\right) = \sum_{j\in\mathbb{Z}} \mathbb{E}(\mathcal{P}_j X_0 \mathcal{P}_j X_k).$$

By the Cauchy–Schwarz inequality, we have

$$\sum_{k\in\mathbb{Z}} |\gamma(k)| \le \sum_{j,k\in\mathbb{Z}} |\mathbb{E}(\mathcal{P}_j X_0 \mathcal{P}_j X_k)| \le \sum_{j,k\in\mathbb{Z}} \|\mathcal{P}_j X_0\| \|\mathcal{P}_j X_k\| = \left(\sum_{j\in\mathbb{Z}} \|\mathcal{P}_0 X_j\|\right)^2.$$

By Theorem 1 of Wu (2005), $\|\mathcal{P}_0 X_j\| \le \delta_2(j) \le \delta_4(j)$ for $j \ge 0$. Therefore, the assumption (5) implies that $\sum_{k=-\infty}^{\infty} |\gamma(k)| < \infty$, i.e., X_t is a short memory process.

Remark 2.2. A natural question is how to verify the conditions (5) and (6) for time series models. It turns out that both (5) and (6) are implied by the GMC(4) condition; see Wu (2005) and Wu and Shao (2004). Let $(\varepsilon'_k)_{k\in\mathbb{Z}}$ be an iid copy of $(\varepsilon_k)_{k\in\mathbb{Z}}$; let $X'_n = G(\ldots, \varepsilon'_{-1}, \varepsilon'_0, \varepsilon_1, \ldots, \varepsilon_n)$ be a coupled version of X_n. We say that X_n is GMC(α), $\alpha > 0$, if there exist $C > 0$ and $\rho = \rho(\alpha) \in (0, 1)$ such that

$$\mathbb{E}(|X_n - X'_n|^{\alpha}) \le C\rho^n, \quad n \in \mathbb{N}. \tag{8}$$

The property (8) indicates that the process $\{X_n\}$ forgets its past exponentially fast, and it can be verified for many nonlinear time series models, such as threshold model (Tong, 1990), bilinear model (Subba Rao and Gabr, 1984), various forms of GARCH models (Bollerslev, 1986; Ding et al., 1993); see Wu and Min (2005) and Shao and Wu (2007) for more details.

Recently, there is a surge of interest in conditional heteroscedastic models (e.g. GARCH models) due to its empirical relevance in modeling financial time series. The GARCH process is a martingale difference with time varying conditional variance, so it is uncorrelated but dependent. In addition, there are a few commonly used models that are uncorrelated but are not martingale differences (MDS). Examples include all-pass ARMA models (Breidt et al., 2001), certain forms of bilinear models (Granger and Anderson, 1978) and nonlinear moving average models (Granger and Teräsvirta, 1993); see Lobato et al. (2002). These models also satisfy GMC conditions, as demonstrated in Shao (in press).

To study the local power, we define the spectral density under the local alternative as

$$H_{1n} : f_n(w) = (2\pi)^{-1}\gamma(0)(1 + g(w)/\sqrt{n}), \quad w \in [-\pi, \pi], \tag{9}$$

where g is a symmetric and 2π-periodic function that satisfies $\int_{-\pi}^{\pi} g(w)\mathrm{d}w = 0$, which ensures that f_n is a valid spectral density function for large n. In this case, we have $\gamma_n(j) = \int_{-\pi}^{\pi} f_n(w)e^{ijw}\mathrm{d}w = \gamma(0)\mathbf{1}(j = 0) + \mathbf{1}(j \neq 0)\gamma(0)/(2\pi\sqrt{n})\int_{-\pi}^{\pi} g(w)e^{ijw}\mathrm{d}w$. So under the local alternative,

$$S_n(\lambda) = \sum_{j=1}^{n-1} \sqrt{n}\{\hat{\gamma}(j) - \gamma_n(j)\}\Psi_j(\lambda) + \sum_{j=1}^{n-1} \sqrt{n}\gamma_n(j)\Psi_j(\lambda),$$

where the first term converges to $S(\lambda)$ in $L_2[0,\pi]$ and the second term approaches $\gamma(0)/(2\pi)\int_0^{\lambda} g(w)\mathrm{d}w$. This implies $CM_n \to_D \int_0^{\pi} \{S(\lambda) + \gamma(0)/(2\pi)\int_0^{\lambda} g(w)\mathrm{d}w\}^2 \mathrm{d}\lambda$. It is not hard to show that $\lim_{n\to\infty} P(\mathrm{CM}_n \text{ rejects } H_0|H_{1n})$, which is the limiting power of the C–M test, approaches 1 as $\|\int_0^{\lambda} g(w)\mathrm{d}w\|$ goes to ∞. This suggests that the test has nontrivial power against local alternative that is within $n^{-1/2}$-neighborhood of the null hypothesis.

3. Nonparametric blockwise wild bootstrap

In this section, we shall use a blockwise wild bootstrap method to approximate the limiting null distribution of CM_n. We describe the steps involved in forming bootstrapped statistic CM_n^* as follows:

1. Set a block size b_n, s.t. $1 \le b_n < n$. Denote the blocks by $B_s = \{(s-1)b_n + 1, \ldots, sb_n\}$, $s = 1, \ldots, L_n$, where the number of blocks $L_n = n/b_n$ is assumed to be an integer for the convenience of presentation.
2. Take iid random draws δ_s, $s = 1, 2, \ldots, L_n$, independent of the data, from a common distribution W, where $\mathbb{E}(W) = 0$, $\mathbb{E}(W^2) = 1$ and $\mathbb{E}(W^4) < \infty$. Define the auxiliary variables $w_t = \delta_s$, if $t \in B_s$, for $t = 1, \ldots, n$.
3. Let $\hat{\gamma}^*(j) = n^{-1}\sum_{t=j+1}^{n}\{(X_t - \bar{X})(X_{t-j} - \bar{X}) - \hat{\gamma}(j)\}w_t$ for $j = 1, \ldots, n-1$. Define the bootstrapped process
$$S_n^*(\lambda) = \sqrt{n}\sum_{j=1}^{n-1}\hat{\gamma}^*(j)\Psi_j(\lambda).$$
4. Compute the bootstrapped test statistic $\mathrm{CM}_n^* = \int_0^\pi \{S_n^*(\lambda)\}^2 \mathrm{d}\lambda$.
5. Repeat steps 2 and 3 M times and denote by $\mathrm{CM}_{n,\alpha}^*$ the empirical $100(1-\alpha)\%$ sample percentile of CM_n^* based on M bootstrapped values. Then we reject the null hypothesis at the significance level α if $\mathrm{CM}_n > \mathrm{CM}_{n,\alpha}^*$.

Throughout the paper, we assume that $b_n^{-1} + b_n/n \to 0$ as $n \to \infty$. As seen from the above procedure, the blockwise wild bootstrap is a simple variant of the traditional wild bootstrap (Wu, 1986; Liu, 1988; Mammen, 1993), which was originally proposed to deal with independent and heteroscedastic errors in the regression problems. The main difference from the traditional wild bootstrap is that the auxiliary variables $\{w_t\}_{t=1}^n$ are the same within each block of size b_n and iid across different blocks. This helps to capture the dependence between $\{(X_t - \bar{X})(X_{t-j} - \bar{X})\}$ and $\{(X_{t'} - \bar{X})(X_{t'-k} - \bar{X})\}$ for $j, k = 1, \ldots, n-1$ when t, t' belong to the same block. Here we mention that the wild bootstrap has been widely used in the context of specification testing to achieve a consistent approximation of the limiting null distribution of the test statistics that depends on the data generating mechanism; see Stute et al. (1998), Li et al. (2003) and Escanciano and Velasco (2006), among others. No centering such as the one appeared in $\hat{\gamma}^*(j)$ has been used in the above-mentioned papers.

Let d_w be any metric that metricizes weak convergence in $L_2[0, \pi]$; see Politis and Romano (1994). To describe the convergence of CM_n^*, conditional on the sample, we adopt the notion "in distribution in probability" introduced by Li et al. (2003). Given the observations $\{X_t\}_{t=1}^n$ and a statistic ξ_n, $\mathcal{L}(\xi_n|X_1, X_2, \ldots)$ (i.e., the distribution of ξ_n given $X_1, X_2, \ldots$) is said to converge to $\mathcal{L}(\xi)$ in distribution in probability if for any subsequence $\xi_{n'}$, there exists a further subsequence $\xi_{n''}$ such that $\mathcal{L}(\xi_{n''}|X_1, X_2, \ldots)$ converges to $\mathcal{L}(\xi)$ for almost every sequence $(X_1, X_2, \ldots)$. Denote by P^*, $\mathbb{E}^*$, var^* and cov^* the probability, expectation, variance and covariance conditional on the data.

Theorem 3.1. *Assume (5) and*
$$\sum_{k_1,\ldots,k_j=-\infty}^{\infty} |k_j||\mathrm{cum}(X_0, X_{k_1}, \ldots, X_{k_j})| < \infty, \quad j = 1, \ldots, J, \qquad (10)$$
for $J = 1, \ldots, 7$. Under the null hypothesis, under any fixed alternative hypothesis, or under the local alternatives (9), we have that
$$d_w[\mathcal{L}\{S_n^*(\lambda)|\mathcal{X}_n\}, \mathcal{L}\{S(\lambda)\}] \to 0 \qquad (11)$$
in probability as $n \to \infty$, where $\mathcal{L}(S_n^*(\lambda)|\mathcal{X}_n)$ stands for the distribution of $S_n^*(\lambda)$ given the sample $\mathcal{X}_n$. Consequently, the bootstrapped test statistic CM_n^* converges to $\int_0^\pi S^2(\lambda)\mathrm{d}\lambda$ in distribution in probability.

Summability conditions on joint cumulants (e.g. (10)) are common in spectral analysis; see Brillinger (1975). For a linear process $X_t = \sum_{j\in\mathbb{Z}} a_j\varepsilon_{t-j}$ with ε_j being iid, (10) holds if $\sum_{j\in\mathbb{Z}} |ja_j| < \infty$ and $\varepsilon_1 \in \mathcal{L}^8$. If X_t admits the form (4), then (10) is true provided that X_t satisfies GMC(8) (see Wu and Shao, 2004, Proposition 2). The assumption of eighth moment condition is a bit strong, but seems necessary in our technical argument.

Remark 3.1. If the process X_t is a sequence of martingale differences satisfying the conditions of Theorem 3.1, then
$$\mathrm{cov}\{S(\lambda), S(\lambda')\} = \sum_{j,k=1}^{\infty} \mathbb{E}(Y_t^2 Y_{t-j}Y_{t-k})\Psi_j(\lambda)\Psi_k(\lambda')$$
(compare (1)) and taking $b_n = 1$ (i.e., using traditional wild bootstrap) still leads to a consistent bootstrap approximation to the limiting null distribution of CM_n. Since the proof is basically a repetition of the argument in the proof of Theorem 3.1, we omit the details.

Therefore, in view of Theorem 2.1, Corollary 2.1 and the discussions at the end of Section 2, Theorem 3.1 suggests that the bootstrap-assisted test (i.e. use bootstrapped critical values) has a correct asymptotic level, is consistent, and has nontrivial power to detect the alternatives tending to the null hypothesis at $n^{-1/2}$ rate.

4. Finite sample performance

In the section, we shall investigate the finite sample performance of our bootstrap-assisted test in comparison with three other methods: (i) Deo's corrected C–M test; (ii) The $\widetilde{Q}_K$ test as proposed in Lobato et al. (2002) to account for dependence; (iii) Subsampling-based test (Politis et al., 1999). Deo (2000) reported through simulation studies that the uncorrected standardized C–M test statistic has severe size distortion when conditional heteroscedasticity is present, the corrected standardized C–M test statistic retains the nominal size even when the correction is unnecessary, and using the corrected statistic does not result in a loss of power compared to its uncorrected counterpart. So we do not include the uncorrected standardized C–M test statistic in the comparison. For the $\widetilde{Q}_K$ test, a bandwidth parameter is involved in the consistent estimation of asymptotic covariance matrix of the first K sample correlations. Here we adopt an automatic procedure as used in Lobato et al. (2002), i.e., we employ the AR(1) prewhitening on each series and select the bandwidth using formula (2.2) of Newey and West (1994) with weights equal to one and lag truncation equal to $2(n/100)^{2/9}$. We consider $K = 1, 5, 10$. For the subsampling method, the key idea is to approximate the sampling distribution of CM_n with the subsampling analogue $\{CM_l(t)\}_{t=1}^{n-l+1}$, where l is the subsample window width and $CM_l(t)$ is the subsampled counterpart of CM_n and is based on the t-th subsample $(X_t, \ldots, X_{t+l-1})$ for $t = 1, \ldots, n-l+1$. We let $l = \lfloor\sqrt{n}/2\rfloor$, $\lfloor\sqrt{n}\rfloor$, $\lfloor 2\sqrt{n}\rfloor$, $\lfloor 4\sqrt{n}\rfloor$, where $\lfloor a\rfloor$ stands for the integer part of a. Among these three methods, the theoretical validity under general dependence assumption (including non-martingale difference sequence) has been proved only for $\widetilde{Q}_K$ test. Presumably one can show that the subsampling method also works under certain mixing assumptions in view of its wide applicability. Three series lengths ($n = 100$, 400 and 1000) are investigated. For our bootstrap-assisted test, we let $b_n = 1$, $\lfloor\sqrt{n}/2\rfloor$, $\lfloor\sqrt{n}\rfloor$ and $\lfloor 2\sqrt{n}\rfloor$. These choices deliver $b_n = 1, 5, 10, 20$ for $n = 100$, $b_n = 1, 10, 20, 40$ for $n = 400$ and $b_n = 1, 15, 31, 63$ for

$n = 1000$. The following Bernoulli distribution is employed to generate δ_t:

$$P(\delta_t = 0.5(1-\sqrt{5})) = \frac{1+\sqrt{5}}{2\sqrt{5}} \quad \text{and}$$

$$P(\delta_t = 0.5(1+\sqrt{5})) = 1 - \frac{1+\sqrt{5}}{2\sqrt{5}}.$$

500 bootstrap samples were taken in the calculation of empirical rejection percentage.

4.1. Size

Let $\varepsilon_t \sim$ iid $N(0, 1)$ unless specified and B denotes the backward shift operator. To examine the empirical size, we consider the following models:

M_1: IID normal, $X_t = \varepsilon_t$.
M_2: GARCH(1, 1), $X_t = \varepsilon_t\sigma_t$, where $\sigma_t^2 = 0.001 + 0.09\sigma_{t-1}^2 + 0.89X_{t-1}^2$.
M_3: IGARCH(1, 1), $X_t = \varepsilon_t\sigma_t$, where $\sigma_t^2 = 0.001 + 0.09\sigma_{t-1}^2 + 0.91X_{t-1}^2$.
M_4: EGARCH, $X_t = \varepsilon_t\sigma_t$, where $\log(\sigma_t^2) = 1 + (1+0.8B)[0.5\varepsilon_t + 2\{|\varepsilon_t| - \mathbb{E}|\varepsilon_t|\}]$.
M_5: Non-MDS 1 (non-martingale difference): $X_t = \varepsilon_t^2\varepsilon_{t-1}$.
M_6: Non-MDS 2: $X_t = \varepsilon_t^2\varepsilon_{t-1}^2\varepsilon_{t-2}^2\varepsilon_{t-3}$.
M_7: All-pass ARMA(1, 1) model: $X_t = 0.8X_{t-1} + \varepsilon_t - (1/0.8)\varepsilon_{t-1}$, where $\varepsilon_t \sim t(10)$.
M_8: Bilinear model: $X_t = \varepsilon_t + 0.5\varepsilon_{t-1}X_{t-2}$.

The models M_1–M_4 are martingale differences, whereas the models M_5–M_8 are uncorrelated but are not martingale differences. The models M_2 and M_3 have been used in Lobato et al. (2001) in the examination of the finite sample performance of a modified Box–Pierce test. Among the non-MDS examples, M_6 exhibits stronger dependence than M_5. The models M_7 and M_8 have been investigated in Lobato et al. (2002) for $\widetilde{Q}_K$ test. Note that the IGARCH model M_3 does not admit a finite second moment and the GARCH model M_2 admits a finite second moment but the fourth moment is infinite; see Davidson (2004) and Ling and McAleer (2002) for sufficient and necessary conditions on the existence of higher order moments for GARCH models. According to examples 2.1 and 2.2 in Shao (in press), the models M_7 and M_8 both satisfy GMC(8), so all the moment and weak dependence assumptions in Theorems 2.1 and 3.1 hold. The GMC(8) property also holds for the models M_1, M_5 and M_6, as can be easily seen. For the EGARCH model M_4, Min (2004) showed that it is GMC(q) for any $q \in \mathbb{N}$ if $\varepsilon_t \sim$ iid $N(0, 1)$.

Table 1(a)–(c) shows the proportion (in percentage) of 5000 replications in which the null hypothesis was rejected at 1%, 5% and 10% nominal significance levels for $n = 100$, $n = 400$ and $n = 1000$ respectively. For all sample sizes, Deo's test delivers accurate size for the models M_1, M_2, M_3 and M_7 but the size distortion is noticeable for other models. Note that the asymptotic validity of Deo's test statistic requires the asymptotic independence of $\hat{\rho}(j)$ and $\hat{\rho}(k)$, $j \neq k$ and that the distorted size for the model M_4 might be due to the fact that this assumption does not hold for M_4 although it is a MDS. For the bootstrap-assisted test, the size is accurate for the models M_1–M_3 at $b_n = 1$ since as commented in Remark 3.1, taking $b_n = 1$ leads to consistent bootstrap approximation when the series is a MDS. When b_n gets large, the size is more distorted at all levels, but the size distortion decreases as n increases. For the model M_4, taking $b_n = \sqrt{n}$ ($2\sqrt{n}$) corresponds to least size distortion when $n = 100$ ($n = 1000$). For non-MDS models, the optimal b_n in terms of size depends on the model and sample size; compare the models M_5 and M_6. The subsampling based test seems quite sensitive to the choice of the subsampling width, with $\sqrt{n}/2$ a suboptimal choice for almost all models at all sample sizes and $2\sqrt{n}$ corresponding to accurate size for most models. The $\widetilde{Q}_K$ test performs reasonably well when $K = 1$ for the models M_1, M_2, M_3, M_7 and M_8, but less satisfactory for other models. A plausible explanation is that the automatic bandwidth selection procedure may not perform well in all the situations. The $\widetilde{Q}_K$ test performs very poorly for $K = 5, 10$ when $n = 100$, which indicates that the asymptotic approximation is less accurate when K is relatively large. The size distortion of $\widetilde{Q}_K$ at $K = 5, 10$ improves for some models as sample size increases, but not uniformly. It is a little surprising to see that for some models (e.g. M_6) the size gets more distorted as sample size gets larger. This might be related to the finite sample bias in the data-driven bandwidth selection algorithm. In particular, the use of the truncation lag $2(n/100)^{2/9}$ is somewhat arbitrary, and it may not work well for all the models under examination. Hence, for some models, the bias in the estimation of optimal bandwidth could get large as sample size increases, thus affects the size negatively. It is also worth noting that our bootstrap-based test performs reasonably well for the model M_3, especially at $n = 1000$, although the second moment is infinite.

4.2. Power

We report the empirical powers of the above-mentioned tests in a small Monte Carlo study. The following alternatives are considered:

M_9: MA(1), $X_t = u_t + \rho u_{t-1}$, where u_t follows the GARCH model M_2 and $\rho = 0.05, 0.1, \ldots, 0.3$.
M_{10}: MA(1), $X_t = u_t + \rho u_{t-1}$, where u_t follows the bilinear model M_8 and $\rho = 0.05, 0.1, \ldots, 0.3$.
M_{11}: ARFIMA(0, d, 0), $(1-B)^d X_t = \varepsilon_t$, where $d = -0.2, -0.15, \ldots, 0.2$.

We present the size-adjusted power in Table 2(a), (b) and (c) for the models M_9, M_{10} and M_{11} at sample size $n = 400$. For bootstrap and subsampling based tests, the calculation of size-adjusted power does not appear to as obvious as that for an asymptotic test. Here we use the technique described in Dominguez and Lobato (2001), who proposed an alternative way of calculating size-corrected power for bootstrap tests. The extension to subsampling based test is straightforward, so we omit the details. As shown in Table 2, the power for the bootstrap-assisted test seems not very sensitive to the choice of b_n. It is comparable to the power for Deo's test and $\widetilde{Q}_1$ test, and is superior to that for $\widetilde{Q}_5$ and $\widetilde{Q}_{10}$ for all the examined alternatives. In a few cases, the bootstrap-based test outperforms all the other alternative methods in power. In particular, our bootstrap test is more powerful than the $\tilde{Q}_K$ for $K = 1, 5, 10$ for the FARIMA(0, d, 0) alternative when $d = 0.05, 0.1, 0.15$ and 0.2 (i.e., in the long memory region). For the subsampling method, larger subsampling width corresponds to less power. The power for $l = \sqrt{n}/2$ is comparable to those delivered by Deo's corrected C–M test and our bootstrap test, and the power corresponding to the other three widths are substantially lower. However, the size distortion reaches the largest when $l = \sqrt{n}/2$, so the small width is also not preferred. In view of the overall size and power performance, it is fair to conclude that the subsampling method does not perform as well as the bootstrap-based test, which has very good size and power properties provided that the block size is suitably chosen.

Table 1
Rejection rates in percentage under the null hypothesis for sample sizes (a) $n = 100$, (b) $n = 400$ and (c) $n = 1000$. The symbols BOOT(l) and SUB(l) stand for the bootstrap-assisted test and the subsampling based test with the bandwidth parameter equal to l. The largest standard error is 0.7%.

(a) $n = 100$	M_1			M_2			M_3			M_4		
	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%
Deo	0.6	5	9.6	0.8	4.8	10	0.8	4.8	10.1	0.2	2.1	6.9
BOOT(1)	1.1	5.8	10.8	1.3	5.4	10.9	1.4	5.6	10.9	0.4	3.4	11.0
BOOT($\sqrt{n}/2$)	2.1	7.3	12.7	1.8	7.1	12.7	1.7	6.8	12.5	0.4	3.3	9.4
BOOT($\sqrt{n}$)	4.2	9.3	15.0	3.5	8.5	14.6	3.2	8.6	14.6	1.0	4.5	10.4
BOOT($2\sqrt{n}$)	11.8	15.2	21.1	10.0	13.7	19.1	10.2	13.6	19.0	5.2	9.7	13.2
SUB($\sqrt{n}/2$)	3.7	12.6	26.2	3.6	11.6	27.5	3.7	11.9	28.0	2.9	26.5	60.8
SUB($\sqrt{n}$)	2.4	7.1	15.4	2.1	6.3	14.6	2.0	6.1	14.9	0.9	5.8	30.3
SUB($2\sqrt{n}$)	2.8	6.6	11.7	1.9	5.3	11.3	2.0	5.5	11.2	1.2	6.1	12.0
SUB($4\sqrt{n}$)	3.9	8.1	12.5	3.2	7.1	12.2	2.9	6.9	12.4	2.5	8.1	13.3
$\tilde{Q}_1$	0.3	4.0	9.3	0.4	3.8	8.9	0.4	3.6	9.2	0.6	1.6	3.8
$\tilde{Q}_5$	0.1	1.8	5.3	0.3	1.7	5.2	0.2	2.0	5.4	3.0	3.9	4.6
$\tilde{Q}_{10}$	0.1	0.6	1.5	0.5	1.4	2.6	0.8	2.0	3.3	6.4	7.6	8.7
$n = 100$	M_5			M_6			M_7			M_8		
	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%
Deo	0.2	2.4	7.6	0.1	1.1	4.8	0.6	3.5	8.3	1.8	7.9	13.9
BOOT(1)	0.3	4.1	10.6	0.2	2.5	11.0	0.9	4.5	9.7	2.6	9.0	15.2
BOOT($\sqrt{n}/2$)	0.6	4.0	11.0	0.1	1.3	7.2	1.7	6.6	12.1	3.3	9.7	16.6
BOOT($\sqrt{n}$)	1.5	5.9	12.6	0.2	1.9	6.8	3.8	9.1	14.6	4.6	10.8	17.7
BOOT($2\sqrt{n}$)	7.0	11.3	16.1	2.3	6.5	8.7	11.1	15.2	20.6	12.9	16.8	21.8
SUB($\sqrt{n}/2$)	1.8	15.5	42.0	1.8	31.9	73.8	3.0	11.1	26.3	5.4	15.4	35.0
SUB($\sqrt{n}$)	1.1	4.8	22.8	0.8	6.4	40.3	2.3	6.4	14.6	2.7	8.3	18.8
SUB($2\sqrt{n}$)	1.6	5.8	11.4	1.1	6.9	14.0	2.9	6.1	11.5	2.7	7.1	12.1
SUB($4\sqrt{n}$)	3.5	8.0	13.5	2.6	8.3	13.7	4.2	8.2	12.7	4.2	9.0	14.2
$\tilde{Q}_1$	0.3	1.6	5.5	0.9	1.1	2.7	0.3	3.6	8.4	1.1	6.5	13.3
$\tilde{Q}_5$	1.8	2.7	3.4	5.9	6.7	7.1	0.1	1.7	4.9	0.3	2.2	5.9
$\tilde{Q}_{10}$	5.5	7.1	8.2	14.6	16.0	17.1	0.1	0.6	1.8	0.6	1.6	2.7
(b) $n = 400$	M_1			M_2			M_3			M_4		
	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%
Deo	0.8	4.9	10.4	0.8	5	10.3	0.8	5	10.3	0.2	2.7	7.7
BOOT(1)	1.1	5.2	10.7	1.2	5.2	10.6	1.1	5.3	10.6	0.4	3.6	9.9
BOOT($\sqrt{n}/2$)	1.6	6.2	12.2	1.3	5.6	11.2	1.4	5.6	11.4	0.5	3.6	10.1
BOOT($\sqrt{n}$)	2.4	7.4	13.2	1.7	6.2	11.8	1.6	5.7	11.1	0.6	3.8	10.5
BOOT($2\sqrt{n}$)	4.4	9.8	15.2	3.3	7.9	13.3	2.6	7.0	12.3	1.7	5.9	11.5
SUB($\sqrt{n}/2$)	1.0	6.6	16.9	0.5	5.1	17.6	0.5	4.8	18.0	0.5	32.2	61.8
SUB($\sqrt{n}$)	1.2	5.1	12.7	0.3	2.7	11.3	0.4	2.1	11.0	0.6	12.5	40.2
SUB($2\sqrt{n}$)	1.8	4.9	10.4	0.5	2.2	6.8	0.4	2.1	6.9	0.8	3.9	8.5
SUB($4\sqrt{n}$)	2.2	5.5	9.5	1.0	3.2	7.7	0.7	3.1	7.5	1.6	5.4	9.9
$\tilde{Q}_1$	0.7	4.5	10.4	0.5	4.4	10.0	0.5	4.1	9.6	0.2	1.9	6.3
$\tilde{Q}_5$	0.5	3.5	8.2	0.3	2.3	6.0	0.2	2.2	5.6	0.5	0.7	1.1
$\tilde{Q}_{10}$	0.3	1.9	4.9	0.2	1.0	2.8	0.2	1.0	2.9	1.0	1.3	1.7
$n = 400$	M_5			M_6			M_7			M_8		
	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%
Deo	0.4	4.1	8.8	0.1	1.6	5.7	0.7	4.1	9.1	2	7.4	13.9
BOOT(1)	0.6	4.9	10.4	0.1	2.7	11.3	0.9	4.5	9.2	2.5	7.7	14.3
BOOT($\sqrt{n}/2$)	1.0	4.6	10.6	0.1	1.8	8.8	1.5	6.2	11.6	2.1	6.9	12.5
BOOT($\sqrt{n}$)	1.2	5.9	11.5	0.1	1.9	8.4	2.2	7.0	12.3	2.6	7.3	13.2
BOOT($2\sqrt{n}$)	2.6	7.7	13.8	0.4	3.4	8.5	4.0	8.9	14.3	4.2	9.1	14.6
SUB($\sqrt{n}/2$)	0.2	11.6	33.9	0.5	47.1	81.6	0.8	6.0	16.3	0.6	7.8	21.9
SUB($\sqrt{n}$)	0.2	4.9	21.6	0.6	15.7	53.3	0.6	4.4	11.9	0.6	4.1	14.2
SUB($2\sqrt{n}$)	0.6	2.4	6.7	0.8	4.0	9.5	1.3	4.2	9.1	0.9	3.4	8.4
SUB($4\sqrt{n}$)	1.3	3.9	8.2	1.3	5.2	11.3	1.8	4.5	9.1	1.7	4.5	8.4
$\tilde{Q}_1$	0.2	3.0	8.0	0.2	0.9	3.6	0.6	4.4	10.0	1.2	6.4	12.9
$\tilde{Q}_5$	0.1	0.4	1.2	1.0	1.2	1.4	0.6	3.3	7.2	0.5	3.2	7.2
$\tilde{Q}_{10}$	0.2	0.6	1.0	2.8	3.3	3.4	0.2	1.6	4.6	0.2	1.9	4.2

(*continued on next page*)

4.3. Empirical illustration

In this section, we apply all the aforementioned tests to the 780 monthly returns on the CRSP-NYSE equal-weighted and value-weighted portfolios from January 1926 to December 1990. This data set has been analyzed by Deo (2000) and Durlauf (1991) using the corrected and uncorrected standardized C–M test statistic respectively. Table 3 reports the p-values for a range of block sizes

Table 1 (*continued*)

(c) $n = 1000$	M_1			M_2			M_3			M_4		
	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%
Deo	1.1	5.1	10	0.8	4.6	10.4	1.1	5.3	10.6	0.2	2.8	7
BOOT(1)	1.2	5.3	10.4	1.1	4.8	10.4	1.3	5.7	10.7	0.4	3.4	9.4
BOOT($\sqrt{n}/2$)	1.7	5.7	10.9	1.3	5.5	10.7	1.3	5.5	10.5	0.5	3.4	9.5
BOOT($\sqrt{n}$)	1.9	6.8	12.0	1.4	5.8	10.8	1.4	5.5	10.4	0.6	3.4	9.7
BOOT($2\sqrt{n}$)	3.0	7.7	13.0	1.9	6.5	12.1	1.5	5.3	11.1	1.0	4.5	10.1
SUB($\sqrt{n}/2$)	0.9	5.6	13.3	0.1	3.8	16.3	0.0	3.6	17.4	0.1	34.4	62.2
SUB($\sqrt{n}$)	0.9	4.9	11.3	0.0	2.3	11.7	0.1	1.7	11.8	0.2	17.3	45.7
SUB($2\sqrt{n}$)	1.2	4.8	10.2	0.2	1.6	8.2	0.1	1.2	7.7	0.4	2.6	28.7
SUB($4\sqrt{n}$)	1.4	4.7	9.1	0.4	2.0	5.7	0.2	1.7	5.6	0.9	3.9	8.1
$\widetilde{Q}_1$	1.0	5.1	10.3	0.7	4.7	9.8	0.8	4.9	9.9	0.2	1.8	6.0
$\widetilde{Q}_5$	0.7	4.4	9.4	0.3	2.4	6.2	0.3	2.0	4.7	0.1	0.2	0.5
$\widetilde{Q}_{10}$	0.5	3.3	7.3	0.1	1.4	3.6	0.1	1.2	3.1	0.3	0.4	0.6
$n = 1000$	M_5			M_6			M_7			M_8		
	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%
Deo	0.7	4	9	0.1	2.1	6.8	0.8	4.6	9.4	2	8.4	14.6
BOOT(1)	0.9	4.6	9.7	0.2	3.4	10.8	1.1	4.8	9.4	2.4	8.4	14.6
BOOT($\sqrt{n}/2$)	1.2	4.7	9.8	0.1	2.1	9.2	1.5	6.4	11.6	1.7	7.1	12.9
BOOT($\sqrt{n}$)	1.3	5.2	10.6	0.1	2.3	8.6	2.0	7.1	12.2	2.0	7.6	12.9
BOOT($2\sqrt{n}$)	2.1	6.6	12.0	0.4	2.6	8.6	2.9	8.7	14.0	3.0	8.7	13.8
SUB($\sqrt{n}/2$)	0.0	8.4	29.3	0.3	51.5	82.7	0.6	5.2	14.1	0.3	6.4	18.6
SUB($\sqrt{n}$)	0.1	4.8	18.8	0.4	25.2	63.5	0.8	5.0	11.7	0.5	4.6	13.9
SUB($2\sqrt{n}$)	0.3	2.0	12.5	0.6	3.3	38.1	1.2	4.7	10.5	0.7	3.6	11.5
SUB($4\sqrt{n}$)	0.8	3.0	7.1	0.8	3.9	9.1	1.4	4.2	9.3	1.2	4.0	8.5
$\widetilde{Q}_1$	0.4	3.4	8.5	0.1	1.0	4.4	0.9	5.2	10.7	0.9	5.9	11.7
$\widetilde{Q}_5$	0.0	0.6	2.1	0.3	0.4	0.4	0.6	4.1	8.4	0.9	4.2	8.8
$\widetilde{Q}_{10}$	0.0	0.3	1.0	0.7	0.8	0.9	0.6	2.7	6.2	0.3	3.1	6.7

Table 2
Rejection rates in percentage under the alternative hypothesis for (a) M_9, (b) M_{10} and (c) M_{11}. The symbols BOOT(l) and SUB(l) stand for the bootstrap-assisted test and the subsampling based test with the bandwidth parameter equal to l. The significance level is 5%.

$n = 400$	(a)						(b)					
ρ	0.05	0.1	0.15	0.2	0.25	0.3	0.05	0.1	0.15	0.2	0.25	0.3
Deo	10.4	28.5	55.4	80	94	98.6	12.9	37.3	69.6	90.4	98.4	99.8
Boot(1)	11.4	30.7	57.6	81.9	94.6	98.9	12.7	36.1	69.1	90.0	98.3	99.7
Boot($\sqrt{n}/2$)	10.1	29.5	58.3	82.4	95.3	99.0	13.0	37.1	69.2	90.5	98.2	99.6
Boot($\sqrt{n}$)	10.9	30.6	59.1	82.4	95.1	98.9	13.9	37.8	68.8	89.6	97.8	99.5
Boot($2\sqrt{n}$)	13.7	33.0	60.0	82.9	94.9	98.8	15.4	39.3	69.4	89.1	97.1	98.9
SUB($\sqrt{n}/2$)	12.5	32.8	58.8	82.4	94.5	98.8	10.6	31.5	62.1	84.7	95.5	99.0
SUB($\sqrt{n}$)	8.4	24.3	48.7	70.9	86.4	93.2	7.7	24.6	50.9	74.3	88.1	93.6
SUB($2\sqrt{n}$)	7.4	20.6	42.7	63.6	79.2	87.4	6.6	20.4	41.4	61.5	74.7	82.0
SUB($4\sqrt{n}$)	7.5	20.0	40.1	60.7	75.9	84.8	7.9	21.4	39.3	56.7	67.4	74.3
$\widetilde{Q}_1$	8.1	24.9	53.0	78.6	93.3	98.0	13.1	37.4	69.7	90.7	98.2	99.7
$\widetilde{Q}_5$	6.3	13.9	32.6	56.9	78.0	90.4	7.1	19.0	40.4	66.4	84.9	93.2
$\widetilde{Q}_{10}$	6.3	11.8	24.7	42.5	62.5	77.7	6.7	13.3	28.0	48.0	65.3	77.9

$n = 400$	(c)							
d	−0.2	−0.15	−0.1	−0.05	0.05	0.1	0.15	0.2
Deo	95.2	78.5	46.9	16.2	17.6	56.2	90.4	99.4
Boot(1)	95.5	78.7	47.6	16.9	17.6	56.4	90.4	99.4
Boot($\sqrt{n}/2$)	96.1	81.1	50.9	18.3	17.2	52.5	87.7	98.6
Boot($\sqrt{n}$)	96.0	81.8	51.4	19.5	18.9	52.8	87.3	98.6
Boot($2\sqrt{n}$)	95.3	82.2	55.1	22.4	21.0	54.6	87.0	97.9
SUB($\sqrt{n}/2$)	91.2	69.0	37.2	11.2	22.6	62.8	92.6	99.4
SUB($\sqrt{n}$)	88.7	66.8	35.6	10.9	20.7	59.4	90.8	98.9
SUB($2\sqrt{n}$)	83.9	62.6	33.4	11.2	18.2	52.3	84.9	97.0
SUB($4\sqrt{n}$)	78.6	58.3	32.4	11.2	15.8	44.6	75.0	91.1
$\widetilde{Q}_1$	94.1	78.4	48.0	17.4	15.2	49.0	84.4	98.0
$\widetilde{Q}_5$	96.3	78.1	42.5	13.4	9.1	26.8	58.5	84.3
$\widetilde{Q}_{10}$	92.4	67.8	33.4	11.4	8.3	19.6	42.2	64.7

$b_n = 1, 13, 26, 52, 104$, with the latter four values approximately equal to $\sqrt{n}/2$, $\sqrt{n}$, $2\sqrt{n}$ and $4\sqrt{n}$ respectively. We use 3000 Bootstrap replications. For the equal-weighted return, the p-values of our bootstrap-assisted test are quite sensitive to the choice of block size, with smaller p-values at $b_n = 1$. A possible explanation for this sensitivity is that the series may not be stationary, which

Table 3
The p-values of four types of tests (Deo's corrected C–M test, our bootstrap-assisted test, the subsampling-based test and $\widetilde{Q}_K$ test) applied to the equal-weighted and value-weighted returns for different block sizes. The symbols BOOT(l) and SUB(l) stand for the bootstrap-assisted test and the subsampling based test with the bandwidth parameter equal to l.

	Equal-weighted	Value-weighted
Deo	(2.5%, 5%)	(5%, 10%)
BOOT(1)	2.23%	9.5%
BOOT(13)	19.2%	26.2%
BOOT(26)	14.7%	15.8%
BOOT(52)	16.7%	20.8%
SUB(13)	3.8%	7.3%
SUB(26)	5.6%	6.9%
SUB(52)	8.9%	10.8%
SUB(104)	11.7%	11.5%
$\widetilde{Q}_1$	5.9%	15.0%
$\widetilde{Q}_5$	12.2%	23.1%
$\widetilde{Q}_{10}$	86.1%	73.5%

is required in our theory. The sensitivity of p-values with respect to block size can also be seen for the subsampling-based test. The p-values of $\widetilde{Q}_K$ test suggest that the lag-1 autocorrelation is marginally significant, and the autocorrelation at other lags (up to lag 10) are all zero. For the value-weighted returns, we cannot reject the null hypothesis at 5% level for all the tests at all the block sizes, which is consistent with the result obtained by Deo (2000).

5. Conclusion

In summary, we propose a blockwise wild bootstrap procedure to approximate the limiting null distribution of the C–M test statistic. Since in practice we do not know the data generating process of the series at hand, it seems preferable to use the bootstrap-based test proposed here, which has been justified to be widely applicable to a large class of uncorrelated nonlinear processes. The comparison of our test with three other methods in terms of the finite sample performance shows that our test has overall good size and power with superiority ability to deal with unknown dependence. In addition, it is encouraging to see that the size and power of our bootstrap-based test is not very sensitive to the choice of block size for large sample size. Through an empirical illustration, we show that the evidence against the white noise null hypothesis for the two series, which were tested by Durlauf (1991) and Deo (2000), is considerably weakened after accounting for unknown dependence. The problem of selecting optimal block size remains open, and more work is needed in that respect. It would be interesting to investigate if the block size selection methodology proposed in the block bootstrap and subsampling literature (cf. Hall et al., 1995; Bühlmann and Künsch, 1999; Politis et al., 1999) can be extended to our setting. We leave it for future research.

Acknowledgements

I would like to thank Rohit Deo for providing me with the data sets used in the paper. I am also grateful to an associate editor and two referees for constructive comments that led to substantial improvement of the paper. The research is supported in part by NSF grant DMS-0804937.

Appendix

Let $\tilde{\gamma}(j) = n^{-1}\sum_{t=1+|j|}^{n} Y_tY_{t-|j|}$ for $j = 0, \pm 1, \ldots, \pm(n-1)$. Define $\tilde{S}_n(\lambda) = \sum_{j=1}^{n-1}\sqrt{n}\tilde{\gamma}(j)\Psi_j(\lambda)$. Denote by $a \wedge b = \min(a, b)$ and $a \vee b = \max(a, b)$; by $W_h(j) = \int_\Pi h(\lambda)\Psi_j(\lambda)\mathrm{d}\lambda$ for any $h \in L_2(\Pi)$, where $\Pi = [0, \pi]$; by $P_j = \int_\Pi \Psi_j^2(\lambda)\mathrm{d}\lambda$, $j \in \mathbb{N}$. Note that $P_j \le Cj^{-2}$ uniformly in $j \in \mathbb{N}$. Further we note that $\int_\Pi \Psi_j(\lambda)\Psi_k(\lambda)\mathrm{d}\lambda = 0$ when $j \neq k$, $j, k \in \mathbb{N}$.

Proof of Theorem 2.1. By Lemma 5.1, it suffices to show that $\tilde{S}_n(\lambda) - \mathbb{E}\{\tilde{S}_n(\lambda)\} \Rightarrow S(\lambda)$. For each fixed K, $1 \le K \le n-1$,

$$\tilde{S}_n(\lambda) = \sum_{j=1}^{K}\sqrt{n}\tilde{\gamma}(j)\Psi_j(\lambda) + \sum_{j=K+1}^{n-1}\sqrt{n}\tilde{\gamma}(j)\Psi_j(\lambda) = \tilde{S}_n^K(\lambda) + R_n^K(\lambda).$$

Then we want to show

(a). For an arbitrary but fixed integer K the finite dimensional distributions of $\tilde{S}_n^K(\lambda) - \mathbb{E}\{\tilde{S}_n^K(\lambda)\}$, $\langle \tilde{S}_n^K - \mathbb{E}\{\tilde{S}_n^K\}, h\rangle$, converge to those of $S^K(\lambda)$, $\langle S^K(\lambda), h\rangle$, for any $h \in L_2(\Pi)$, where $S^K(\lambda)$ is a Gaussian process with zero mean and asymptotic projected variances

$$\sigma_{h,K}^2 = \mathrm{var}[\langle S^K, h\rangle] = \sum_{j,k=1}^{K}\sum_{d=-\infty}^{\infty}\mathrm{cov}(Y_tY_{t-j}, Y_{t-d}Y_{t-d-k})W_h(j)W_h(k).$$

Note that

$$\langle \tilde{S}_n^K - \mathbb{E}\{\tilde{S}_n^K\}, h\rangle = n^{-1/2}\sum_{j=1}^{K}\sum_{t=j+1}^{n}\{Y_tY_{t-j} - \gamma(j)\}W_h(j) = n^{-1/2}\sum_{t=2}^{K+1}\sum_{j=1}^{t-1}\{Y_tY_{t-j} - \gamma(j)\}W_h(j) + n^{-1/2}\sum_{t=K+2}^{n}\sum_{j=1}^{K}\{Y_tY_{t-j} - \gamma(j)\}W_h(j),$$

where the first summand above is $o_p(1)$ since K is finite. For the second summand, since $\sum_{k=0}^{\infty}\delta_4(k) < \infty$ and

$$\|\mathcal{P}_0(Y_tY_{t-j})\| \le \|Y_tY_{t-j} - Y_t'Y_{t-j}'\| \le C\{\delta_4(t) + \delta_4(t-j)\},$$

we have that $\sum_{k=0}^{\infty}\|\mathcal{P}_0(Y_kY_{k-j})\| < \infty$ for any $j = 1, \ldots, K$. Consequently, the distributional convergence follows from Theorem 1 in Hannan (1973) or Lemma 1 in Wu and Min (2005).

(b). For an arbitrary but fixed integer K the sequence $\{\tilde{S}_n^K(\lambda)\}$ is tight.

$$\tilde{S}_n^K(\lambda) - \mathbb{E}\{\tilde{S}_n^K(\lambda)\} = n^{-1/2}\sum_{j=1}^{K}\sum_{t=j+1}^{n}\{Y_tY_{t-j} - \gamma(j)\}\Psi_j(\lambda) = n^{-1/2}\sum_{t=2}^{n}\sum_{j=1}^{(t-1)\wedge K}\{Y_tY_{t-j} - \gamma(j)\}\Psi_j(\lambda) = n^{-1/2}\sum_{t=2}^{K+1}G_{n,t}^K + n^{-1/2}\sum_{t=K+2}^{n}G_{n,t}^K,$$

where $G_{n,t}^K$ is implicitly defined. The tightness of $n^{-1/2}\sum_{t=2}^{K+1}G_{n,t}^K$ follows from the fact that it is a finite sum and each summand is tight. For the second term $n^{-1/2}\sum_{t=K+2}^{n}G_{n,t}^K$, we note that it is a sum of stationary, mean zero random elements such that

$\mathbb{E}\|\sum_{j=1}^{K}\{Y_tY_{t-j}-\gamma(j)\}\Psi_j(\lambda)\|^2<\infty$. Thus it suffices to verify the assumptions in Politis and Romano (1994) listed below:

(i) $\mathbb{E}\|G_{n,t}^K\|^2<\infty$.

(ii) For each integer $m\ge 2$, the random elements $(G_{n,K+2}^K, G_{n,K+3}^K,\ldots,G_{n,K+m}^K)$ converge in distribution to $(G_{K+2}^K, G_{K+3}^K,\ldots,G_{K+m}^K)$.

(iii) For each integer $m\ge 2$, $\mathbb{E}[\langle G_{n,K+2}^K, G_{n,K+m}^K\rangle]\to\mathbb{E}[\langle G_{K+2}^K, G_{K+m}^K\rangle]$ as $n\to\infty$.

(iv) $\lim_{n\to\infty}\sum_{t=K+2}^{n}\mathbb{E}[\langle G_{n,K+2}^K, G_{n,t}^K\rangle]=\sum_{t=K+2}^{\infty}\mathbb{E}[\langle G_{K+2}^K, G_t^K\rangle]<\infty$, and the last series converges absolutely.

(v) $\mathrm{var}[\langle\tilde S_n^K-\mathbb{E}(\tilde S_n^K),h\rangle]\to\sigma_{h,K}^2$.

Let $G_t^K=G_{n,t}^K=\sum_{j=1}^{K}\{Y_tY_{t-j}-\gamma(j)\}\Psi_j(\lambda)$. Then (i), (ii) and (iii) are trivially satisfied. For (iv), we have

$$\begin{aligned}\sum_{t=K+2}^{\infty}&|\mathbb{E}[\langle G_{K+2}^K,G_t^K\rangle]|\\&=\sum_{t=K+2}^{\infty}\left|\sum_{j=1}^{K}\mathrm{cov}(Y_tY_{t-j},Y_{K+2}Y_{K+2-j})P_j\right|\\&\le\sum_{t=K+2}^{\infty}\sum_{j=1}^{K}P_j\{|\mathrm{cum}(X_t,X_{K+2},X_{t-j},X_{K+2-j})|\\&\quad+\gamma^2(t-K-2)+|\gamma(t-K-2+j)\gamma(t-K-2-j)|\}\\&<\infty\end{aligned}$$

in view of the assumptions (5), (6) and Remark 2.1.

Further we note that

$$\begin{aligned}\mathrm{var}[\langle\tilde S_n^K-\mathbb{E}(\tilde S_n^K),h\rangle]&=n^{-1}\mathrm{var}\left(\sum_{j=1}^{K}\sum_{t=j+1}^{n}Y_tY_{t-j}W_h(j)\right)\\&=n^{-1}\sum_{j,k=1}^{K}\sum_{t=j+1}^{n}\sum_{t'=k+1}^{n}\mathrm{cov}(Y_tY_{t-j},Y_{t'}Y_{t'-k})W_h(j)W_h(k)\\&=n^{-1}\sum_{j,k=1}^{K}\sum_{d=(j-n)\wedge(k-n)+1}^{(n-j)\vee(n-k)-1}\mathrm{cov}(Y_tY_{t-j},Y_{t-d}Y_{t-d-k})W_h(j)W_h(k)\\&\quad\{(n-j)\vee(n-k)-|d|\}\to\sigma_{h,K}^2.\end{aligned}$$

Thus (v) is shown and this completes the proof of part (b).

(c). The process $R_n^K(\lambda)$ satisfies that for all $\epsilon>0$,

$$\lim_{K\to\infty}\lim_{n\to\infty}P(\|R_n^K(\lambda)-\mathbb{E}\{R_n^K(\lambda)\}\|>\epsilon)=0.$$

Under the assumptions (5) and (6), we have

$$\begin{aligned}\mathbb{E}\|R_n^K(\lambda)-\mathbb{E}\{R_n^K(\lambda)\}\|^2&=n\sum_{j=K+1}^{n-1}\mathrm{var}\{\tilde\gamma(j)\}P_j\\&\le n^{-1}\sum_{j=K+1}^{n-1}\sum_{t,t'=j+1}^{n}P_j\{|\mathrm{cum}(X_0,X_{-j},X_{t'-t},X_{t'-t-j})|\\&\quad+|\gamma^2(t'-t)|+|\gamma(t'-t+j)\gamma(t'-t-j)|\}\\&\le C\sum_{j=K+1}^{\infty}j^{-2}\to 0\quad\text{as }K\to\infty.\quad\square\end{aligned}$$

Proof of Corollary 2.1. Under the assumptions of Theorem 2.1, it follows from Lemma 5.1 and a straightforward calculation that $\|\mathbb{E}\{S_n(\lambda)\}-\sqrt{n}\sum_{j=1}^{\infty}\gamma(j)\Psi_j(\lambda)\|=o(1)$. Therefore, $S_n(\lambda)\Rightarrow S(\lambda)$ under H_0 and (3) follows from the continuous mapping theorem. Similarly, the assertion (7) also follows. □

Define $\tilde S_n^*(\lambda)=\sqrt{n}\sum_{j=1}^{n-1}\tilde\gamma^*(j)\Psi_j(\lambda)$, where $\tilde\gamma^*(j)=n^{-1}\sum_{t=j+1}^{n}\{Y_tY_{t-j}-\gamma(j)\}w_t$.

Proof of Theorem 3.1. By Lemma 5.2, it suffices to show that

(a) The process $\tilde S_n^*(\lambda)$ has the same asymptotic finite projections as the process $S_n(\lambda)$. To this end, we decompose $\tilde S_n^*(\lambda)$ as $\tilde S_n^*(\lambda)=\tilde S_n^{K*}(\lambda)+R_n^{K*}(\lambda)$, where $\tilde S_n^{K*}(\lambda)=\sqrt{n}\sum_{j=1}^{K}\tilde\gamma^*(j)\Psi_j(\lambda)$. We shall show that

(i). The finite dimensional distributions of $\tilde S_n^{K*}(\lambda)$, $\langle\tilde S_n^{K*},h\rangle$, converge to those of $S^K(\lambda)$, $\langle S^K(\lambda),h\rangle$ for any $h\in L_2(\Pi)$, in probability conditional on the original sample. Write $\tilde S_n^{K*}(\lambda)=n^{-1/2}\sum_{t=2}^{(K+1)}G_{n,t}^{K*}+n^{-1/2}\sum_{t=K+2}^{n}G_{n,t}^{K*}$, where $G_{n,t}^{K*}=\sum_{j=1}^{(t-1)\wedge K}[Y_tY_{t-j}-\gamma(j)]w_t\Psi_j(\lambda)$. Note that $\mathbb{E}[\mathbb{E}^*\{n^{-1/2}\sum_{t=2}^{(K+1)}G_{n,t}^{K*}\}^2]=o(1)$ since it only involves finite number of terms. It suffices to show the asymptotic normality of $\langle n^{-1/2}\sum_{t=K+2}^{n}G_{n,t}^{K*},h\rangle$. Write

$$\begin{aligned}J_n^{K*}&:=n^{-1/2}\sum_{t=K+2}^{n}\langle G_{n,t}^{K*},h\rangle\\&=n^{-1/2}\sum_{t=K+2}^{n}\sum_{j=1}^{K}\{Y_tY_{t-j}-\gamma(j)\}w_tW_h(j)\\&=\sum_{s=1}^{L_n}n^{-1/2}\delta_s\sum_{t\in B_s\cap[K+2,n]}\sum_{j=1}^{K}\{Y_tY_{t-j}-\gamma(j)\}W_h(j)\\&=\sum_{s=1}^{L_n}H_{sn}^*,\end{aligned}$$

where H_{sn}^* is implicitly defined. Then

$$\mathrm{var}^*(J_n^{K*})=n^{-1}\sum_{s=1}^{L_n}\left[\sum_{t\in B_s\cap[K+2,n]}\sum_{j=1}^{K}\{Y_tY_{t-j}-\gamma(j)\}W_h(j)\right]^2.$$

In the sequel, we shall show that $\mathrm{var}^*(J_n^{K*})\to_p\sigma_{h,K}^2$. To this end, we calculate $\mathbb{E}\{\mathrm{var}^*(J_n^{K*})\}$ and $\mathrm{var}\{\mathrm{var}^*(J_n^{K*})\}$ as follows. First, we note that

$$\begin{aligned}\mathbb{E}&\{\mathrm{var}^*(J_n^{K*})\}\\&=n^{-1}\sum_{s=1}^{L_n}\sum_{t,t'\in B_s\cap[K+2,n]}\sum_{j,j'=1}^{K}\mathrm{cov}(Y_tY_{t-j},Y_{t'}Y_{t'-j'})W_h(j')W_h(j)\\&\to\sigma_{h,K}^2.\end{aligned}$$

Second,

$$\begin{aligned}\mathrm{var}&\{\mathrm{var}^*(J_n^{K*})\}=n^{-2}\\&\times\sum_{s,s'=1}^{L_n}\sum_{t_1,t_2\in B_s\cap[K+2,n]}\sum_{t_1',t_2'\in B_{s'}\cap[K+2,n]}\sum_{j_1,j_2=1}^{K}\\&\times\sum_{j_1',j_2'=1}^{K}C(t_1,t_2,t_1',t_2',j_1,j_2,j_1',j_2')\\&\times W_h(j_1)W_h(j_2)W_h(j_1')W_h(j_2'),\end{aligned}\tag{12}$$

where $C(t_1,t_2,t_1',t_2',j_1,j_2,j_1',j_2')$ equals to

$$\begin{aligned}&\mathrm{cov}[\{Y_{t_1}Y_{t_1-j_1}-\gamma(j_1)\}\{Y_{t_2}Y_{t_2-j_2}-\gamma(j_2)\},\\&\quad\{Y_{t_1'}Y_{t_1'-j_1'}-\gamma(j_1')\}\{Y_{t_2'}Y_{t_2'-j_2'}-\gamma(j_2')\}]\\&=\mathrm{cum}(Y_{t_1}Y_{t_1-j_1},Y_{t_2}Y_{t_2-j_2},Y_{t_1'}Y_{t_1'-j_1'},Y_{t_2'}Y_{t_2'-j_2'})\\&\quad+\mathrm{cov}(Y_{t_1}Y_{t_1-j_1},Y_{t_1'}Y_{t_1'-j_1'})\\&\quad\times\mathrm{cov}(Y_{t_2}Y_{t_2-j_2},Y_{t_2'}Y_{t_2'-j_2'})+\mathrm{cov}(Y_{t_1}Y_{t_1-j_1},Y_{t_2'}Y_{t_2'-j_2'})\\&\quad\times\mathrm{cov}(Y_{t_2}Y_{t_2-j_2},Y_{t_1'}Y_{t_1'-j_1'})\\&=T_1+T_2+T_3.\end{aligned}$$

By Theorem II.2 in Rosenblatt (1985), we have

$$T_1 = \sum_{v} \text{cum}(X_{i_j}, i_j \in v_1) \cdots \text{cum}(X_{i_j}, i_j \in v_p),$$

where the summation is over all indecomposable partitions $v = v_1 \cup \cdots \cup v_p$ of the two-way table

$$\begin{matrix} t_1 & t_1 - j_1 \\ t_2 & t_2 - j_2 \\ t_1' & t_1' - j_1' \\ t_2' & t_2' - j_2'. \end{matrix}$$

For example, one such term in T_1 is $\gamma(t_2 - j_2 - t_1)\gamma(t_1' - j_1' - t_2)\gamma(t_2' - t_1' - j_2')\gamma(t_2' - t_1 + j_1)$. Under (10), it is easily seen that its contribution to the sum (12) is of order $o(1)$. The contributions of other terms in T_1 are also $o(1)$ due to (10). For T_2, we can write

$$\begin{aligned} T_2 = {} & \{\text{cum}(X_0, X_{-j_1}, X_{t_1'-t_1}, X_{t_1'-t_1-j_1'}) \\ & + \gamma(t_1' - t_1)\gamma(t_1' - t_1 - j_1' + j_1) \\ & + \gamma(t_1' - j_1' - t_1)\gamma(t_1' - t_1 + j_1)\} \\ & \times \{\text{cum}(X_0, X_{-j_2}, X_{t_2'-t_2}, X_{t_2'-t_2-j_2'}) \\ & + \gamma(t_2' - t_2)\gamma(t_2' - t_2 - j_2' + j_2) \\ & + \gamma(t_2' - j_2' - t_2)\gamma(t_2' - t_2 + j_2)\}. \end{aligned}$$

When $s = s'$ in the summation of (12), the contribution of T_2 to $\text{var}\{\text{var}^*(J_n^{K*})\}$ is $O(n^{-2}L_n b_n^2) = o(1)$. In the case of $s \neq s'$, then (10) implies that the magnitude of all such terms is bounded by $Cn^{-2}L_n^2 = o(1)$. The same argument applies to T_3. Hence, $\text{var}\{\text{var}^*(J_n^{K*})\} = o(1)$.

Further, we note that $H_{sn}^*, s = 1, \ldots, L_n$ are independent conditional on the data. We verify Lindeberg–Feller's condition as follows:

$$\begin{aligned} & \sum_{s=1}^{L_n} \mathbb{E}[\mathbb{E}^*\{|H_{sn}^*|^2 \mathbf{1}(|H_{sn}^*| > \epsilon)\}] \le C \sum_{s=1}^{L_n} \mathbb{E}\{\mathbb{E}^*|H_{sn}^*|^4\} \\ & \le Cn^{-2} \sum_{s=1}^{L_n} \mathbb{E}\left[\sum_{t\in B_s\cap[K+2,n]} \sum_{j=1}^{K} \{Y_t Y_{t-j} - \gamma(j)\} W_h(j)\right]^4 \\ & \le Cn^{-2} \sum_{s=1}^{L_n} \sum_{t_1,t_2,t_3,t_4\in B_s\cap[K+2,n]} \sum_{j_1,j_2,j_3,j_4=1}^{K} |W_h(j_1)W_h(j_2) \\ & \quad \times W_h(j_3)W_h(j_4)||\mathbb{E}[\{Y_{t_1}Y_{t_1-j_1} - \gamma(j_1)\}\{Y_{t_2}Y_{t_2-j_2} - \gamma(j_2)\} \\ & \quad \times \{Y_{t_3}Y_{t_3-j_3} - \gamma(j_3)\}\{Y_{t_4}Y_{t_4-j_4} - \gamma(j_4)\}]|. \end{aligned}$$

Following a similar argument as used in proving $\text{var}\{\text{var}^*(J_n^{K*})\} = o(1)$, we can show that the above expression is of order $O(n^{-2}L_n b_n^2) = o(1)$. We omit the details.

(ii). For any $\epsilon > 0$, $\lim_{K\to\infty} \lim_{n\to\infty} P^*(\|R_n^{K*}(\lambda)\| \ge \epsilon) = 0$ in probability conditional on the sample. Write

$$\begin{aligned} \mathbb{E}^*\|R_n^{K*}(\lambda)\|^2 &= n \sum_{j=K+1}^{n-1} \mathbb{E}^*\{\tilde{\gamma}^*(j)^2\} P_j \\ &= n^{-1} \sum_{j=K+1}^{n-1} \sum_{t,t'=j+1}^{n} \{Y_t Y_{t-j} - \gamma(j)\}\{Y_{t'}Y_{t'-j} - \gamma(j)\} \\ &\quad \times \mathbb{E}^*(w_t w_{t'}) P_j \\ &= n^{-1} \sum_{j=K+1}^{n-1} \sum_{s=1}^{L_n} \left[\sum_{t\in B_s\cap[j+1,n]} \{Y_t Y_{t-j} - \gamma(j)\}\right]^2 P_j. \end{aligned}$$

So as $n \to \infty$,

$$\begin{aligned} & \mathbb{E}(\mathbb{E}^*\|R_n^{K*}(\lambda)\|^2) \\ & = n^{-1} \sum_{j=K+1}^{n-1} \sum_{s=1}^{L_n} \mathbb{E}\left(\sum_{t\in B_s\cap[j+1,n]} \{Y_t Y_{t-j} - \gamma(j)\}\right)^2 P_j \\ & = n^{-1} \sum_{j=K+1}^{n-1} \sum_{s=1}^{L_n} \sum_{t,t'\in B_s\cap[j+1,n]} \text{cov}(Y_t Y_{t-j}, Y_{t'}Y_{t'-j}) P_j \\ & \to \sum_{j=K+1}^{\infty} P_j \sigma_Y^2(j), \end{aligned}$$

where $\sigma_Y^2(j) = 2\pi f_{Y_t Y_{t-j}}(0)$. Note that

$$\begin{aligned} |\sigma_Y^2(j)| &\le \sum_{k=-\infty}^{\infty} |\text{cov}(Y_t Y_{t-j}, Y_{t+k}Y_{t+k-j})| \\ &\le \sum_{k=-\infty}^{\infty} \{|\text{cum}(Y_t, Y_{t-j}, Y_{t+k}, Y_{t+k-j})| \\ &\quad + |\gamma^2(k) + \gamma(k+j)\gamma(k-j)|\} < \infty \end{aligned}$$

uniformly in j. Then we get $\lim_{K\to\infty} \lim_{n\to\infty} \mathbb{E}(\mathbb{E}^*\|R_n^{K*}(\lambda)\|^2) = 0$.

(b) The process $\tilde{S}_n^*(\lambda)$ is tight in probability conditional on the sample. Write

$$\begin{aligned} \tilde{S}_n^*(\lambda) &= n^{-1/2} \sum_{t=2}^{n} \sum_{j=1}^{t-1} \{Y_t Y_{t-j} - \gamma(j)\} w_t \Psi_j(\lambda) \\ &= L_n^{-1/2} \sum_{s=1}^{L_n} b_n^{-1/2} \delta_s \sum_{t\in B_s\cap[2,n]} \sum_{j=1}^{t-1} \{Y_t Y_{t-j} - \gamma(j)\} \Psi_j(\lambda) \\ &= L_n^{-1/2} \sum_{s=1}^{L_n} M_{sn}^*(\lambda), \end{aligned}$$

where $M_{sn}^*(\lambda)$ is implicitly defined. Note that $M_{sn}^*(\lambda)$ and $M_{s'n}^*(\lambda)$, $s \neq s'$ are independent given the sample. In view of the argument in the proof of Theorem 4 of Escanciano and Velasco (2006), it suffices to verify that $\mathbb{E}^*\|M_{sn}^*(\lambda)\|^2 < \infty$ almost surely; also see Example 1.8.5 of Van der Vaart and Wellner (1996). To this end, write

$$\mathbb{E}\{\mathbb{E}^*\|M_{sn}^*(\lambda)\|^2\} = b_n^{-1} \sum_{t,t'\in B_s} \sum_{j=1}^{t\wedge t'-1} \text{cov}(Y_t Y_{t-j}, Y_{t'}Y_{t'-j}) P_j.$$

Under (5) and (6), the above expression is bounded by

$$\begin{aligned} & b_n^{-1} \sum_{t,t'\in B_s} \sum_{j=1}^{t\wedge t'-1} \{|\text{cum}(X_t, X_{t-j}, X_{t'}, X_{t'-j})| + |\gamma^2(t'-t)| \\ & + |\gamma(t'-t-j)\gamma(t'-t+j)|\} P_j < \infty, \end{aligned}$$

which implies that $\mathbb{E}^*\|M_{sn}^*(\lambda)\|^2$ is finite almost surely. The conclusion is established. □

The lemma below shows that the difference between $S_n(\lambda) - \mathbb{E}\{S_n(\lambda)\}$ and $\tilde{S}_n(\lambda) - \mathbb{E}\{\tilde{S}_n(\lambda)\}$ is negligible.

Lemma 5.1. *Under the assumptions* (5) *and* (6), *we have*

$$\|S_n(\lambda) - \mathbb{E}\{S_n(\lambda)\} - [\tilde{S}_n(\lambda) - \mathbb{E}\{\tilde{S}_n(\lambda)\}]\|^2 \to_p 0.$$

Proof of Lemma 5.1. Let $\bar{Y} = \bar{X} - \mu$. Since for $j = 0, 1, \ldots, n-1$,

$$\hat{\gamma}(j) = \tilde{\gamma}(j) - \bar{Y} n^{-1} \sum_{t=j+1}^{n} (Y_t + Y_{t-j}) + \frac{n-j}{n} \bar{Y}^2,$$

we have $S_n(\lambda) = \tilde{S}_n(\lambda) + M_n(\lambda)$, where

$$M_n(\lambda) = -n^{-1/2}\bar{Y}\sum_{j=1}^{n-1}\sum_{t=j+1}^{n}(Y_t + Y_{t-j})\Psi_j(\lambda) + n^{-1/2}\sum_{j=1}^{n-1}(n-j)\bar{Y}^2\Psi_j(\lambda).$$

Therefore,

$$\|M_n(\lambda)\|^2 \le \frac{3\bar{Y}^2}{n}\sum_{j=1}^{n-1}\left(\sum_{t=j+1}^{n}Y_t\right)^2 P_j + \frac{3\bar{Y}^2}{n}\sum_{j=1}^{n-1}\left(\sum_{t=j+1}^{n}Y_{t-j}\right)^2 P_j + \frac{3\bar{Y}^4}{n}\sum_{j=1}^{n-1}(n-j)^2P_j = I_1 + I_2 + I_3.$$

Under (5) and (6), it is easy to show $\mathbb{E}(\bar{Y}^4) < Cn^{-2}$, which implies that $\mathbb{E}(I_3) \le Cn^{-1}\mathbb{E}(\bar{Y}^4)\sum_{j=1}^{n-1}(n-j)^2P_j \le C/n$. By the Cauchy–Schwarz inequality, we have that

$$\mathbb{E}(I_1) \le Cn^{-1}\mathbb{E}^{1/2}(\bar{Y}^4)\mathbb{E}^{1/2}\left\{\sum_{j=1}^{n-1}\left(\sum_{t=j+1}^{n}Y_t\right)^2P_j\right\}^2 \le Cn^{-2}\left(\sum_{j,k=1}^{n-1}P_jP_k\sum_{t_1,t_2=j+1}^{n}\sum_{t_3,t_4=k+1}^{n}\mathbb{E}\{Y_{t_1}Y_{t_2}Y_{t_3}Y_{t_4}\}\right)^{1/2}.$$

Expressing $\mathbb{E}\{Y_{t_1}Y_{t_2}Y_{t_3}Y_{t_4}\} = \gamma(t_1-t_2)\gamma(t_3-t_4)+\gamma(t_1-t_3)\gamma(t_2-t_4)+\gamma(t_1-t_4)\gamma(t_2-t_3)+\mathrm{cum}(Y_{t_1},Y_{t_2},Y_{t_3},Y_{t_4})$, we can derive that $\mathbb{E}(I_1) \le C/n$ under (5) and (6). By a similar argument, $\mathbb{E}(I_2) \le C/n$. Subsequently, $\mathbb{E}\|M_n(\lambda)\|^2 = o(1)$. By the Cauchy–Schwarz inequality, we have $\|\mathbb{E}\{M_n(\lambda)\}\|^2 = o(1)$. The conclusion follows. □

The following lemma shows that the difference between $S_n^*(\lambda)$ and $\tilde{S}_n^*(\lambda)$ is negligible.

Lemma 5.2. *Under the assumptions of* (5) *and* (6), $\mathbb{E}^*\|S_n^*(\lambda) - \tilde{S}_n^*(\lambda)\|^2 = o_p(1)$.

Proof of Lemma 5.2. According to the definitions of $\hat{\gamma}^*(j)$ and $\tilde{\gamma}^*(j)$, we can write $S_n^*(\lambda) - \tilde{S}_n^*(\lambda) = M_{1n}^*(\lambda) + M_{2n}^*(\lambda)$, where

$$M_{1n}^*(\lambda) = -n^{-1/2}\sum_{j=1}^{n-1}\{\hat{\gamma}(j)-\gamma(j)\}\Psi_j(\lambda)\sum_{t=j+1}^{n}w_t,$$

$$\begin{aligned}M_{2n}^*(\lambda) &= n^{-1/2}\sum_{j=1}^{n-1}\Psi_j(\lambda)\sum_{t=j+1}^{n}\{(X_t-\bar{X})\times(X_{t-j}-\bar{X}) - Y_tY_{t-j}\}w_t\\ &= -n^{-1/2}\bar{Y}\sum_{j=1}^{n-1}\sum_{t=j+1}^{n}Y_tw_t\Psi_j(\lambda) - n^{-1/2}\bar{Y}\sum_{j=1}^{n-1}\sum_{t=j+1}^{n}Y_{t-j}w_t\Psi_j(\lambda) + \bar{Y}^2n^{-1/2}\sum_{j=1}^{n-1}\sum_{t=j+1}^{n}w_t\Psi_j(\lambda)\\ &= \bar{Y}I_1^*(\lambda) + \bar{Y}I_2^*(\lambda) + \bar{Y}^2I_3^*(\lambda).\end{aligned}$$

For $M_{1n}^*(\lambda)$, we have

$$\mathbb{E}^*\|M_{1n}^*(\lambda)\|^2 = n^{-1}\sum_{j=1}^{n-1}\{\hat{\gamma}(j)-\gamma(j)\}^2\mathbb{E}^*\left(\sum_{t=j+1}^{n}w_t\right)^2P_j.$$

Following the same argument as presented in the proof of Lemma 5.1, we can show

$$\mathbb{E}\{\mathbb{E}^*\|M_{1n}^*(\lambda)\|^2\} \le Cn^{-1}\sum_{j=1}^{n-1}P_j\mathbb{E}\{\hat{\gamma}(j)-\gamma(j)\}^2(n-j)b_n \le Cb_n/n = o(1).$$

Hence, $\mathbb{E}^*\|M_{1n}^*(\lambda)\|^2 = o_p(1)$. Regarding $M_{2n}^*(\lambda)$, we have

$$\mathbb{E}\{\mathbb{E}^*\|I_1^*(\lambda)\|^2\} = n^{-1}\sum_{j=1}^{n-1}\mathbb{E}\left\{\mathbb{E}^*\left(\sum_{t=j+1}^{n}Y_tw_t\right)^2P_j\right\} = n^{-1}\sum_{j=1}^{n-1}\sum_{s=1}^{L_n}\mathbb{E}\left(\sum_{t\in B_s\cap[j+1,n]}Y_t\right)^2P_j = O(1),$$

which implies that $\mathbb{E}^*\|I_1^*(\lambda)\|^2 = O_p(1)$. By a similar argument, $\mathbb{E}^*\|I_2^*(\lambda)\|^2 = O_p(1)$. Concerning $I_3^*(\lambda)$, we have that

$$\mathbb{E}^*\|I_3^*(\lambda)\|^2 = n^{-1}\sum_{j=1}^{n-1}P_j\mathbb{E}^*\left(\sum_{t=j+1}^{n}w_t\right)^2 \le Cb_n.$$

By Corollary 2(ii) (ii) of Wu (2007), $\sum_{t=1}^{n}Y_t = o(n^{1/2}\log n)$ almost surely under the assumption (5). So $\mathbb{E}^*\|M_{2n}^*(\lambda)\|^2 = o_p(1)$ and then $\mathbb{E}^*\|S_{1n}^*(\lambda) - \tilde{S}_n^*(\lambda)\|^2 = o_p(1)$. This completes the proof. □

References

Anderson, T.W., 1993. Goodness of fit tests for spectral distributions. Annals of Statistics 21, 830–847.

Bartlett, M.S., 1955. An Introduction to Stochastic Processes. Cambridge University Press, Cambridge.

Bollerslev, T., 1986. Generalized autoregressive conditional heteroscedasticity. Journal of Econometrics 31, 307–327.

Box, G., Pierce, D., 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., Davis, R.A., Trindade, A.A., 2001. Least absolute deviation estimation for all-pass time series models. Annals of Statistics 29, 919–946.

Brillinger, D.R., 1975. Time Series: Data Analysis and Theory. Holden-Day, San-Francisco.

Bühlmann, P., Künsch, H.R., 1999. Block length selection in the bootstrap for time series. Computational Statistics and Data Analysis 31, 295–310.

Chen, X., White, H., 1996. Laws of large numbers for Hilbert space-valued mixingales with applications. Econometric Theory 12, 284–304.

Chen, X., White, H., 1998. Central limit and functional central limit theorems for Hilbert-valued dependent heterogeneous arrays with applications. Econometric Theory 14, 260–284.

Davidson, J., 2004. Moment and memory properties of linear conditional heteroscedasticity models and a new model. Journal of Business & Economic Statistics 22, 1629.

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

Ding, Z., Granger, C., Engle, R., 1993. A long memory property of stock market returns and a new model. Journal of Empirical Finance 1, 83–106.

Dominguez, M.A., Lobato, I.N., 2001. Size corrected power for bootstrap tests. Working Paper. Centro de Investigacion Economica, CIE, Instituto Tecnologico Autonomo de Mexico, ITAM. Downloaded from http://ftp.itam.mx/pub/academico/inves/lobato/01-02.pdf.

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

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

Escanciano, J., Lobato, I.N., 2009. An automatic data-driven portmanteau test for testing for serial correlation. Journal of Econometrics 151, 140–149.

Granger, C.W.J., Anderson, A.P., 1978. An Introduction to Bilinear Time Series Models. Vandenhoek and Ruprecht, Gottinger.

Granger, C.W.J., Teräsvirta, T., 1993. Modelling Nonlinear Economic Relationships. Oxford University Press, New York.

Grenander, U., Rosenblatt, M., 1957. Statistical Analysis of Stationary Time Series. Wiley, New York.
Hall, P., Horowitz, J.L., Jing, B.-Y., 1995. On blocking rules for the bootstrap with dependent data. Biometrika 82, 561–574.
Hannan, E.J., 1973. Central limit theorems for time series regression. Zeitschrift fur Wahrscheinlichkeitstheorie und Verwandte Gebiete 26, 157–170.
Horowitz, J.L., Lobato, I.N., Nankervis, J.C., Savin, N.E., 2006. Bootstrapping the Box–Pierce Q test: a robust test of uncorrelatedness. Journal of Econometrics 133, 841–862.
Kallianpur, G., 1981. Some ramifications of Wiener's ideas on nonlinear prediction. In: Wiener, N., Masani, P. (Eds.), Norbert Wiener, Collected Works with Commentaries. MIT Press, Cambridge, MA, pp. 402–424.
Li, Q., Hsiao, C., Zinn, J., 2003. Consistent specification tests for semiparametric/nonparametric models based on series estimation methods. Journal of Econometrics 112, 295–325.
Ling, S., McAleer, M., 2002. Necessary and sufficient moment conditions for the GARCH(r, s) and asymmetric power GARCH(r, s) models. Econometric Theory 18, 722–729.
Liu, R.Y., 1988. Bootstrap procedures under some non-i.i.d. models. Annals of Statistics 16, 1696–1708.
Lobato, I.N., 2001. Testing that a dependent process is uncorrelated. Journal of the American Statistical Association 96, 1066–1076.
Lobato, I.N., Nankervis, J.C., Savin, N.E., 2001. Testing for autocorrelation using a modified Box–Pierce Q test. International Economic Review 42, 187–205.
Lobato, I.N., Nankervis, J.C., Savin, N.E., 2002. Testing for zero autocorrelation in the presence of statistical dependence. Econometric Theory 18, 730–743.
Mammen, E., 1993. Bootstrap and wild bootstrap for high dimensional linear models. Annals of Statistics 21, 255–285.
Min, W., 2004. Inference on time series driven by dependent innovations. Ph.D. Dissertation, University of Chicago.
Newey, W.K., West, K.D., 1994. Automatic lag selection in covariance matrix estimation. Review of Economic Studies 61, 631–653.
Parthasarathy, K.R., 1967. Probability Measures on Metric Spaces. Academic Press, New York.
Politis, D.N., Romano, J., 1994. Limit theorems for weakly dependent Hilbert space valued random variables with application to the stationary bootstrap. Statistica Sinica 4, 461–476.
Politis, D.N., Romano, J.P., Wolf, M., 1999. Subsampling. Springer-Verlag, New York.
Romano, J.L., Thombs, L.A., 1996. Inference for autocorrelations under weak assumptions. Journal of the American Statistical Association 91, 590–600.
Rosenblatt, M., 1971. Markov Processes, Structure and Asymptotic Behavior. Springer, New York.
Rosenblatt, M., 1985. Stationary Sequences and Random Fields. Birkhäuser, Boston.
Shao, X., 2010. Testing for white noise under unknown dependence and its applications to diagnostic checking for time series models. Econometric Theory (in press).
Shao, X., Wu, W.B., 2007. Asymptotic spectral theory for nonlinear time series. Annals of Statistics 4, 1773–1801.
Stute, W., Gonzalez-Manteiga, W., Presedo-Quindimil, M., 1998. Bootstrap approximations in model checks for regression. Journal of the American Statistical Association 93, 141–149.
Subba Rao, T., Gabr, M.M., 1984. An Introduction to Bispectral Analysis and Bilinear Time Series Models. In: Lecture Notes in Statistics, vol. 24. Springer-Verlag, New York.
Tong, H., 1990. Non-linear Time Series: A Dynamical System Approach. Oxford University Press.
Van der Vaart, A.W., Wellner, J.A., 1996. Weak Convergence and Empirical Processes. Springer, New York.
Wu, C.F.J., 1986. Jackknife, bootstrap and other resampling methods in regression analysis (with discussion). Annals of Statistics 14, 1261–1350.
Wu, W.B., 2005. Nonlinear system theory: another look at dependence. Proceedings of National Academy of Science, USA 102, 14150–14154.
Wu, W.B., 2007. Strong invariance principles for dependent random variables. Annals of Probability 35, 2294–2320.
Wu, W.B., Min, M., 2005. On linear processes with dependent innovations. Stochastic Processes and Their Applications 115, 939–958.
Wu, W.B., Shao, X., 2004. Limit theorems for iterated random functions. Journal of Applied Probability 41, 425–436.