A bootstrapped spectral test for adequacy in weak ARMA models

BY KE ZHU
Institute of Applied Mathematics, Chinese Academy of Sciences, Haidian District, Zhongguancun, Beijing, China
kzhu@amss.ac.cn

AND WAI KEUNG LI
Department of Statistics and Actuarial Science, University of Hong Kong, Pokfulam Road, Hong Kong
hrntlwk@hku.hk

SUMMARY

This paper proposes a Cramér-von Mises (CM) test statistic to check the adequacy of weak ARMA models. Without posing a martingale difference assumption on the error terms, the asymptotic null distribution of the CM test is obtained. Moreover, this CM test is consistent, and has nontrivial power against the local alternative of order $n^{-1/2}$. Due to the unknown dependence of error terms and the estimation effects, a new block-wise random weighting method is constructed to bootstrap the critical values of the test statistic. The new method is easy to implement and its validity is justified. The theory is illustrated by a small simulation study and an application to S&P 500 stock index.

Some key words: Block-wise random weighting method; Diagnostic checking; Least squares estimation; Spectral test; Weak ARMA models; Wild bootstrap.

JEL classifications: C01; C12; C22.

1. INTRODUCTION

After the seminal work of Box and Pierce (1970) and Ljung and Box (1978), diagnostic checking has been an important step in the application of the following ARMA(p, q) model:

$$y_t = \sum_{i=1}^{p} \phi_i y_{t-i} + \sum_{i=1}^{q} \varphi_i \varepsilon_{t-i} + \varepsilon_t, \tag{1}$$

where ε_t are error terms with mean zero. As usual, we say that model (1) is weak when $\{\varepsilon_t\}$ is an uncorrelated sequence, and that model (1) is strong when $\{\varepsilon_t\}$ is an iid sequence; see, e.g., Francq and Zakoïan (1998). Up to now, the most famous diagnostic checking tools for model (1) are the portmanteau tests in Box and Pierce (1970) and Ljung and Box (1978). However, their asymptotic null distributions are only valid for strong ARMA models; see, e.g., Romano and Thombs (1996) and Francq, Roy, and Zakoïan (2005). Moreover, empirical studies in Franses and Van Dijk (1996) and Tsay (2005) demonstrated that many economic and financial series follow an ARMA model with uncorrelated errors (e.g., ARCH-type errors). In addition, Francq and Zakoïan (1998) and Francq, Roy, and Zakoïan (2005) indicated that many nonlinear models admit a weak ARMA representation. Thus, it is meaningful to consider diagnostic checking for weak ARMA models.

Based on either observable series (i.e., $p = q = 0$) or residual series, a huge literature so far has been focused on testing model adequacy in weak ARMA models. These existing tests are roughly categorized into two types: time domain correlation-based tests and frequency domain periodogram-based tests. The tests in the first category usually use the autocorrelations up to lag m (a user-chosen integer), so they are unable to detect serial correlations beyond lag m; see, e.g., Romano and Thombs (1996), Lobato (2001), and Horowitz, Lobato, Nankervis, and Savin (2006) for observable series, or Francq, Roy, and Zakoïan (2005) and Delgado and Velasco (2011) for residual series. To avoid selecting m, Escanciano and Lobato (2009) and Escanciano, Lobato, and Zhu (2013) derived a data-driven portmanteau test under the assumption that ε_t is a martingale difference sequence (MDS). However, it is unclear whether their tests are applicable if ε_t is not an MDS.

Since the correlation-based tests are inconsistent, the periodogram-based tests in the second category have drawn more attention in the literature; see, e.g., Durlauf (1991) and Deo (2000) for earlier works. Under the assumption that ε_t admits a linear process of iid innovations, the smoothing rescaled periodogram test in Paparoditis (2000, 2001) and generalised likelihood ratio test in Fan and Zhang (2004) can be used to check the whiteness of residual series; see, e.g., Fan and Jiang (2007) for more works on smoothed tests. However, the two aforementioned smoothed tests are not applicable when ε_t follows some often used non-linear models (e.g., ARCH-type models). Under the assumption that ε_t is an MDS, many spectral tests have been constructed by Delgado, Hidalgo, and Velasco (2005), Escanciano (2006, 2007), and Escanciano and Velasco (2006). Recently, Shao (2011a) proposed a spectral test for observable series without the MDS assumption on the error terms. A natural but important extension is to construct spectral tests for residual series when ε_t is not an MDS. Under the assumption that ε_t is GMC(8) (a condition weaker than MDS), Shao (2011b) proved the validation of the kernel-based spectral test in Hong (1996), where the definition of GMC is given in Remark 3 below, and the lag m as a bandwidth grows slowly with the sample size.

This paper proposes a Cramér-von Mises (CM) spectral test statistic to check the adequacy of weak ARMA models. Under certain conditions allowing for non-MDS error terms, the asymptotic null distribution of the CM test is obtained. Moreover, this CM test is consistent, and has nontrivial power against local alternatives of order $n^{-1/2}$. Due to the unknown dependence structure of error terms and the estimation effects, our null distribution is no longer asymptotically pivotal. This is also the main challenge for other spectral tests in weak ARMA models. To overcome it, a new block-wise random weighting (BRW) method is constructed to bootstrap critical values of the CM test. The new method is easy to implement and its validity is justified. The theory is illustrated by a small simulation study and an application to S&P 500 stock index.

This paper is organized as follows. Section 2 gives our test statistic and establishes its asymptotic theory. Section 3 proposes a BRW method and proves its validation. Simulation results are reported in Section 4. A real example is provided in Section 5. Concluding remarks are offered in Section 6. All of the proofs are given in the Appendix. Throughout the paper, A' is the transpose of matrix A, $|A| = (tr(A'A))^{1/2}$ is the Euclidean norm of a matrix A, $\|A\|_s = (E|A|^s)^{1/s}$ is the L^s-norm $(s \geq 1)$ of a random matrix, $o_p(1)(O_p(1))$ denotes a sequence of random numbers converging to zero (bounded) in probability, "$\rightarrow_d$" denotes convergence in distribution, and "$\rightarrow_p$" denotes convergence in probability.

2. Test statistic and asymptotic theory

Denote by $\gamma(j) = cov(\varepsilon_t, \varepsilon_{t+j})$. Let

$$f(\omega) = \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} \gamma(j) e^{-ij\omega} \quad \text{for } \omega \in [-\pi, \pi]$$

and $F(\lambda) = \int_0^\lambda f(\omega) d\omega$ for $\lambda \in [0, \pi]$ be the spectral density function and spectral distribution function of ε_t, respectively. Note that $F(\lambda) = \sum_{j=0}^{\infty} \gamma(j) \psi_j(\lambda)$, where

$$\psi_j(\lambda) = \begin{cases} \sin(j\lambda)/j\pi & \text{if } j \neq 0 \\ \lambda/2\pi & \text{if } j = 0 \end{cases}.$$

Then, following Shao (2011a), the sample spectral distribution function of ε_t is

$$F_n(\lambda) = \sum_{j=0}^{n-1} \widehat{\gamma}(j) \psi_j(\lambda),$$

where $\widehat{\gamma}(j) = n^{-1} \sum_{t=1+|j|}^{n} \varepsilon_t \varepsilon_{t-|j|}$ is the sample autocovariance function of ε_t at lag j. Since $F(\lambda) = \gamma(0)\psi_0(\lambda)$ under the null hypothesis

$$H_0 : y_t \text{ admits a weak ARMA}(p, q) \text{ model representation as in } (1),$$

the sample spectral distribution $F_n(\lambda)$ becomes $\widehat{\gamma}(0)\psi_0(\lambda)$ in this case. Thus, as in Shao (2011a), we consider the following Cramér von-Mises statistic

$$\text{CM}_n = \int_0^\pi S_n^2(\lambda) d\lambda$$

to detect H_0, where the process

$$S_n(\lambda) = \sqrt{n} \left\{ F_n(\lambda) - \widehat{\gamma}(0)\psi_0(\lambda) \right\} := \sum_{j=1}^{n-1} \sqrt{n} \widehat{\gamma}(j) \psi_j(\lambda)$$

measures the distance between $F_n(\lambda)$ and $\widehat{\gamma}(0)\psi_0(\lambda)$. However, the statistic CM_n is not feasible because ε_t is unobservable.

Next, let $\theta = (\phi_1, \cdots, \phi_p, \varphi_1, \cdots, \varphi_q)' \in \Theta$ be the unknown parameter of model (1). Then, given the observations $\{y_1, \cdots, y_n\}$, we can calculate the least squares estimator (LSE) θ_n defined by

$$\theta_n = \arg\min_{\Theta} \widetilde{L}_n(\theta) \quad \text{where } \widetilde{L}_n(\theta) = \frac{1}{n} \sum_{t=1}^{n} \widetilde{\varepsilon}_t^2(\theta) =: \frac{1}{n} \sum_{t=1}^{n} \widetilde{l}_t(\theta),$$

and $\widetilde{\varepsilon}_t(\theta)$ is calculated recursively by

$$\widetilde{\varepsilon}_t(\theta) = y_t - \sum_{i=1}^{p} \phi_i y_{t-i} - \sum_{i=1}^{q} \varphi_i \widetilde{\varepsilon}_{t-i}(\theta) \tag{2}$$

with $\widetilde{\varepsilon}_0(\theta) = \widetilde{\varepsilon}_{-1}(\theta) = \cdots = \widetilde{\varepsilon}_{-q+1}(\theta) = y_0 = y_{-1} = \cdots = y_{-p+1} = 0$. Now, by using the residual $\widetilde{\varepsilon}_t = \widetilde{\varepsilon}_t(\theta_n)$, we can propose a feasible Cramér von-Mises statistic as follows:

$$\widetilde{\text{CM}}_n = \int_0^\pi \widetilde{S}_n^2(\lambda) d\lambda,$$

where $\widetilde{S}_n(\lambda) = \sum_{j=1}^{n-1} \sqrt{n}\widetilde{\gamma}(j)\psi_j(\lambda)$ and $\widetilde{\gamma}(j) = n^{-1}\sum_{t=1+|j|}^{n} \widetilde{\varepsilon}_t\widetilde{\varepsilon}_{t-|j|}$.

In order to obtain the limiting distribution of $\widetilde{\text{CM}}_n$, we regard $\widetilde{S}_n(\lambda)$ as a random element in the Hilbert space $L_2[0, \pi]$ of all square integrable functions with the inner product

$$\langle f, g\rangle = \int_0^{\pi} f(\lambda)g^c(\lambda)d\lambda,$$

where $g^c(\lambda)$ denotes the complex conjugate of $g(\lambda)$. Here, $L_2[0, \pi]$ is endowed with the natural Borel σ-field induced by the norm $\|f\| = \langle f, f\rangle^{1/2}$; see Parthasarathy (1967). Since the "$\| \cdot \|$" functional is a continuous mapping from $L_2[0, \pi]$ to $\mathcal{R}$, the limiting distribution of $\widetilde{\text{CM}}_n$ follows directly from the weak convergence of $\widetilde{S}_n(\lambda)$ in $L_2[0, \pi]$; see, e.g., Politis and Romano (1994), Escanciano (2006), Shao (2011a), and many others.

Let $\varepsilon_t(\theta)$ be the parametric model (1), i.e., given initial values $\{y_0, y_{-1}, \cdots\}$ and observations $\{y_1, \cdots, y_n\}$, $\varepsilon_t(\theta)$ is iteratively constructed from

$$\varepsilon_t(\theta) = y_t - \sum_{i=1}^{p} \phi_i y_{t-i} - \sum_{i=1}^{q} \varphi_i \varepsilon_{t-i}(\theta).$$

To obtain the weak convergence of $\widetilde{S}_n(\lambda)$ in $L_2[0, \pi]$, we make the following three assumptions:

Assumption 1. (i) The parametric space $\Theta \subset \mathcal{R}^{p+q}$ is compact, and the true parameter θ_0 of model (1) belongs to the interior of Θ.

(ii) For each $\theta \in \Theta$, $\phi(z) := 1 - \sum_{i=1}^{p} \phi_i z^i \neq 0$ and $\varphi(z) := 1 + \sum_{i=1}^{q} \varphi_i z^i \neq 0$ when $|z| \leq 1$, and $\phi(z)$ and $\varphi(z)$ have no common root with $\phi_p \neq 0$ or $\varphi_q \neq 0$.

Assumption 2. $\{y_t\}$ is strictly stationary with $E|y_t|^{4+2\nu} < \infty$ and

$$(i)\ \sum_{k=0}^{\infty} \{\alpha_y(k)\}^{\nu/(2+\nu)} < \infty$$

for some $\nu > 0$, where $\{\alpha_y(k)\}$ is the sequence of strong mixing coefficients of $\{y_t\}$;

$$(ii)\quad \sum_{s_1,s_2,s_3=-\infty}^{\infty} |cum(y_0, y_{s_1}, y_{s_2}, y_{s_3})| < \infty.$$

Assumption 3. (i) There exists a unique interior point $\breve{\theta}_0 \in \Theta$ such that $\|\theta_n - \breve{\theta}_0\| = o_p(1)$.

(ii) The matrix $\Sigma = E\left[\partial^2 l_t(\breve{\theta}_0)/\partial\theta\partial\theta'\right]$ exists and is positive definite, where $l_t(\theta) = \varepsilon_t^2(\theta)$.

Assumption 1(i) is a basic set-up for model (1), and Assumption 1(ii) is the condition for the stationarity, invertibility and identifiability of model (1); see, e.g., Brockwell and Davis (1991) and Zhu and Ling (2012). Assumption 2(i) from Francq and Zakoïan (1998) is a technical condition for proving the asymptotic theory of θ_n. In addition, the mixing condition on y_t is valid for large classes of processes; see, e.g., Pham (1986) and Carrasco and Chen (2002). Assumption 2(ii) from Shao (2011a) is a cumulant summability condition, and a sufficient condition is given in Doukhan and León (1989), that is, there exists a $\nu_0 \in (0, 1]$ such that

$$\sum_{k=0}^{\infty} (k+1)^{s-2}\{\alpha_y(k)\}^{\nu_0/(s+\nu_0)} < \infty \text{ for } s = 1, \cdots, 4.$$

Assumption 3(i) from Escanciano (2006) guarantees the weak convergence of θ_n. Assumption 3(ii) ensures that the inverse of Σ exists. According to Theorem 1 in Francq and Zakoïan (1998), we know that $\breve{\theta}_0 = \theta_0$ under H_0. However, if H_0 fails, $\breve{\theta}_0$ and θ_0 may be different.

Let $\breve{\varepsilon}_t = \varepsilon_t(\breve{\theta}_0)$ and $e_{t,j} = \breve{\varepsilon}_t\breve{\varepsilon}_{t-j} + z_{tj}$, where

$$z_{tj} = -E\left[\frac{\partial(\breve{\varepsilon}_t\breve{\varepsilon}_{t-j})}{\partial\theta'}\right]\Sigma^{-1}\left[\frac{\partial l_t(\breve{\theta}_0)}{\partial\theta}\right]. \tag{3}$$

We are now ready to give our first main result:

THEOREM 1. *Assume that Assumptions* 1-3 *hold. Then, as* $n \to \infty$,

$$\widetilde{S}_n(\lambda) - E\{\breve{S}_n(\lambda)\} \Rightarrow S(\lambda),$$

where "⇒" stands for weak convergence in $L_2[0,\pi]$ *endowed with the norm metric,*

$$\breve{S}_n(\lambda) = \sum_{j=1}^{n-1}\sqrt{n}\breve{\gamma}(j)\psi_j(\lambda) \text{ with } \breve{\gamma}(j) = n^{-1}\sum_{t=1+|j|}^{n}\breve{\varepsilon}_t\breve{\varepsilon}_{t-|j|},$$

and $S(\lambda)$ *is a Gaussian process in* $C[0,\pi]$ *with mean zero and covariance function*

$$cov\{S(\lambda), S(\lambda')\} = \sum_{j=1}^{\infty}\sum_{k=1}^{\infty}\sum_{d=-\infty}^{\infty} cov(e_{t,j}, e_{t-d,k})\psi_j(\lambda)\psi_k(\lambda').$$

COROLLARY 1. *Assume that Assumptions* 1-3 *hold. Then, as* $n \to \infty$,

$$(i)\ \widetilde{CM}_n \to_d \int_0^{\pi} S^2(\lambda)d\lambda \ \ \text{under } H_0;$$

$$(ii)\ \frac{\widetilde{CM}_n}{n} \to_p \sum_{j=1}^{\infty}[E(\breve{\varepsilon}_t\breve{\varepsilon}_{t-j})]^2\int_0^{\pi}\psi_j^2(\lambda)d\lambda.$$

Remark 1. When $p = q = 0$, the Gaussian process $S(\lambda)$ is the same as the one in Theorem 2.1 of Shao (2011a). When some p or q is nonzero, the Gaussian process $S(\lambda)$ depends on z_{tj}, which is caused by the estimation effect. This phenomenon happens not only in our case but in most specification tests.

Remark 2. When ε_t follows a GARCH model, Ling (2007) showed that a finite fourth moment of y_t is necessary to prove the asymptotic normality of the LSE in ARMA-GARCH models. In view of this, our moment assumption on y_t is not restrictive.

Remark 3. In the proof of Theorem 1, we use a mixing condition of y_t to ensure the asymptotic normality theory; see Rosenblatt (1985). Recently, an alternative way for this is to use the physical dependence condition of y_t in Wu (2005) as done by Shao (2011a, b) and many others. In general, the physical dependence condition is implied by the geometric-moment contraction (GMC) condition defined as follows:

Definition [Wu (2005)]: Assume that $y_t = G(\cdots, \varepsilon_{t-1}, \varepsilon_t)$, where G is a measurable function. Let $\{\varepsilon'_k\}_{k\in\mathcal{Z}}$ be an iid copy of $\{\varepsilon_k\}_{k\in\mathcal{Z}}$, and $y'_t = G(\cdots, \varepsilon'_{-1}, \varepsilon'_0, \varepsilon_1, \cdots, \varepsilon_t)$ be a coupled version of y_t. Then, y_t is GMC(α) for $\alpha > 0$, if there exist $C > 0$ and $\rho = \rho(\alpha) \in (0, 1)$ such that $E\left(|y_t - y'_t|^{\alpha}\right) \le C\rho^t$ for $t \in \mathcal{Z}$.

The GMC condition indicates that the process $\{y_t\}$ forgets its past exponentially fast, and this can not be implied from the α-mixing condition. Shao and Wu (2007) and Shao (2011b) have verified the GMC condition for many nonlinear time series models such as GARCH models, all-pass ARMA models, bilinear models, to name a few. Particularly, if y_t satisfies the GMC(4) condition, Assumption 2(ii) holds according to Proposition 2 in Wu and Shao (2004). Needless to say, the concepts of α-mixing and GMC are two parallel tools to depict the dependence structure of y_t. In this paper, we mainly focus on the α-mixing condition, and our results could be obtained similarly in the GMC context.

Remark 4. Let $r_0 = s_0 = 2 + 2\nu/(4+\nu)(\leq 4)$. Under Assumption 2(i), the Davydov's inequality in Davydov (1968) implies that

$$|cov(y_t, y_{t-k})| \leq O(1)\|y_t\|_{r_0}\|y_{t-k}\|_{s_0}\left[\alpha_y(k)\right]^{1-1/r_0-1/s_0}$$

for any $k \geq 0$. Thus, it follows that

$$\sum_{k=0}^{\infty}|cov(y_t, y_{t-k})|^2 \leq O(1)\sum_{k=0}^{\infty}\left[\alpha_y(k)\right]^{\nu/(2+\nu)} < \infty.$$

So, we know that $\sum_{k=-\infty}^{\infty}\left[\gamma(k)\right]^2 < \infty$, and hence $\sum_{k=-\infty}^{\infty}|\gamma(k)| < \infty$, i.e., y_t is a short memory process under Assumption 2(i).

In practice, since θ_0 is generally unknown, one may focus on the following alternative hypothesis H_1, where

$H_1 : y_t$ does not admit a weak ARMA(p, q) model representation as in (1) with parameter $\breve{\theta}_0$.

Since at least one $E(\breve{\varepsilon}_t\breve{\varepsilon}_{t-j}) \neq 0$ under H_1, the test statistic $\widetilde{\text{CM}}_n$ is consistent in detecting H_1 by Corollary 1(ii).

In the end, as in Shao (2011a), we consider a Pitman's local alternative as follows:

$$H_{1n} : f_n(\omega) = \frac{\gamma(0)}{2\pi}\left(1 + \frac{g(\omega)}{\sqrt{n}}\right),$$

where $\omega \in [-\pi, \pi]$, g is a symmetric and 2π-periodic function that satisfies $\int_{-\pi}^{\pi} g(\omega)d\omega = 0$. Clearly, f_n is a valid spectral density function, and under H_{1n},

$$\gamma_n(j) = \begin{cases} \frac{\gamma(0)}{2\pi\sqrt{n}}\int_{-\pi}^{\pi} g(\omega)e^{ij\omega}d\omega & \text{if } j \neq 0 \\ \gamma(0) & \text{if } j = 0 \end{cases}. \tag{4}$$

As in Escanciano (2006), we need one more assumption as follows:

Assumption 4. Under H_{1n}, $\|\theta_n - \theta_0\| = o_p(1)$ (i.e., $\theta_0 = \breve{\theta}_0$).

COROLLARY 2. *Assume that Assumptions* 1-4 *hold. Then, under H_{1n}, as $n \to \infty$,*

$$\widetilde{CM}_n \to_d \int_0^{\pi}\{S(\lambda) + \Pi(\lambda)\}^2 d\lambda,$$

where $\Pi(\lambda) = \frac{\gamma(0)}{2\pi}\int_0^{\lambda} g(\omega)d\omega$.

Corollary 2 shows that if the value of $\Pi(\lambda)$ deviates from zero, $\widetilde{\text{CM}}_n$ has nontrivial power against the local alternative of order $n^{-1/2}$. Note that the kernel-based spectral test T_n in

Hong (1996) and Shao (2011b) only has nontrivial power against the local alternative of order $(n/m_n^{1/2})^{-1/2}$, where

$$T_n = \sum_{j=1}^{n-1} K^2\left(\frac{j}{m_n}\right)\widetilde{\rho}^2(j), \tag{5}$$

with $\widetilde{\rho}(j) = \widetilde{\gamma}(j)/\widetilde{\gamma}(0)$ being the residual autocorrelation at lag j, $K(\cdot)$ being the kernel function satisfying Assumption 2.1 in Shao (2011b), and m_n being the bandwidth such that $\log n = o(m_n)$ and $m_n = o(n^{1/2})$. However, this does not guarantee that $\widetilde{\mathrm{CM}}_n$ is always more powerful than T_n under H_{1n}. To see the reason, on one hand, by (A20) in the Appendix, we have

$$\Pi(\lambda) \approx \sum_{j=1}^{n} \sqrt{n}\gamma_n(j)\psi_j(\lambda) = \sum_{j=1}^{n} \sqrt{n}\gamma_n(j)\frac{sin(j\pi)}{j\pi},$$

from which we know that the impact of $\gamma_n(j)$ to $\Pi(\lambda)$ is proportional to j^{-1}. So, it implies that if $\gamma_n(j)$ is only significantly different from zero at large lag j, the value of $\Pi(\lambda)$ may not deviate significantly from zero, and hence this will cause a low local power of $\widetilde{\mathrm{CM}}_n$. On the other hand, the local power of T_n tends to be proportional to $\|g/m_n^{1/4}\|$ under H_{1n} (see Theorem 4 in Hong (1996)). From (4), we know that the value of $\|g\|$ becomes large when the value of $\gamma_n(j)$ is significantly different from zero at any lag j. Thus, if the value of m_n is not big enough, T_n will not be deficient in local power even when $\gamma_n(j)$ is only significantly different from zero at large lag j. Generally speaking, under H_{1n}, $\widetilde{\mathrm{CM}}_n$ may be locally more (or less) powerful than T_n, when $\gamma_n(j)$ is significantly different from zero at small (or large) lag j; see, e.g., the simulation results for Examples 1 and 2 in Section 4 below. Similar phenomenon has been well documented by Eubank and LaRiccia (1992) and Paparoditis (2001). Moreover, if we consider Rosenblatt's (1975) sharp peak local alternative, the asymptotic theory of $\widetilde{\mathrm{CM}}_n$ and T_n is still unclear. The pioneering work in Ghosh and Huang (1991) and Paparoditis (2001) may be extended to both tests, and we leave it for future study.

3. BOOTSTRAPPED CRITICAL VALUES

Since the limiting distribution of $\widetilde{\mathrm{CM}}_n$ depends on the unknown model parameters and unknown structure of dependent innovation, we use a block-wise random weighting (BRW) method to bootstrap its critical values. The detailed steps are as follows:

1. Set a block size b_n, such that $1 \leq b_n < n$. Denote the blocks by $B_s = \{(s-1)b_n + 1, \cdots, sb_n\}$ for $s = 1, \cdots, L_n$, where $L_n = n/b_n$ is assumed to be an integer for the convenience of presentation.

2. Generate a sequence of positive i.i.d. random variables $\{\delta_1, \cdots, \delta_{L_n}\}$, independent of the data, from a common distribution W, where $E(W) = 1$ and $var(W) = 1$. Define the random weights $w_t^* = \delta_s$, if $t \in B_s$, for $t = 1, \cdots, n$. Calculate θ_n^* via

$$\theta_n^* = \arg\min_{\Theta} \widetilde{L}_n^*(\theta), \text{ where } \widetilde{L}_n^*(\theta) = \frac{1}{n}\sum_{t=1}^{n} w_t^*\widetilde{\varepsilon}_t^2(\theta) =: \frac{1}{n}\sum_{t=1}^{n}\tilde{l}_t^*(\theta).$$

3. Let $\widetilde{\varepsilon}_t^* = \widetilde{\varepsilon}_t(\theta_n^*)$ for $t = 1, \cdots, n$, with $\widetilde{\varepsilon}_t(\theta)$ being defined as in (2), and

$$\widetilde{S}_n^*(\lambda) = \sum_{j=1}^{n-1}\sqrt{n}\widetilde{\gamma}^*(j)\psi_j(\lambda) \text{ with } \widetilde{\gamma}^*(j) = \frac{1}{n}\sum_{t=1+j}^{n} w_t^*\widetilde{\varepsilon}_t^*\widetilde{\varepsilon}_{t-j}^*.$$

Define the bootstrapped process $\Delta_n(\lambda) = \widetilde{S}_n^*(\lambda) - \widetilde{S}_n(\lambda) - \widetilde{Z}_n(\lambda)$, where

$$\widetilde{Z}_n(\lambda) = \sum_{j=1}^{n-1} \left\{ \frac{1}{\sqrt{n}} \sum_{t=1+j}^{n} [(w_t^* - 1)\widetilde{\gamma}(j)] \right\} \psi_j(\lambda). \tag{6}$$

4. Compute the bootstrapped test statistic $\widetilde{\text{CM}}_n^* = \int_0^\pi \{\Delta_n(\lambda)\}^2 d\lambda$.

5. Repeat steps 2-4 J times and denote by $\widetilde{\text{CM}}_{n,\alpha}^*$ the empirical $100(1-\alpha)\%$ sample percentile of $\widetilde{\text{CM}}_n^*$ based on J bootstrapped values. Then we reject H_0 at the significance level α if $\widetilde{\text{CM}}_n > \widetilde{\text{CM}}_{n,\alpha}^*$.

We now offer some remarks on the BRW method. First, when $p = q = 0$, we set $\widetilde{\varepsilon}_t = \widetilde{\varepsilon}_t^* = y_t$ for all t in step 2, and our BRW method reduces to the wild bootstrap method in Shao (2011a). The novel feature of our BRW method is that it takes into account the estimation effect of the unknown model parameters in step 2. Second, the BRW method as a natural extension of the RW method in Jin, Ying, and Wei (2001) is related to the wild dependent bootstrap in Wu (1986) and Liu (1988), and the original RW method has been widely used for statistical inference in regression based on the least absolute deviation estimation; see, e.g., Chen, Ying, Zhang, and Zhao (2008) and Chen, Guo, Lin, and Ying (2010). Third, the BRW method has no need to generate the bootstrap pseudo series. Fourth, the terms $[\widetilde{S}_n^*(\lambda) - \widetilde{S}_n(\lambda)]$ and $\widetilde{Z}_n(\lambda)$ involved in $\Delta_n(\lambda)$ are used to mimic $\widetilde{S}_n(\lambda)$ and $E\{\breve{S}_n(\lambda)\}$ in Theorem 1, respectively; see Lemma A7 and (A40) in the Appendix. Thus, $\widetilde{Z}_n(\lambda)$ is a centering factor as in Shao (2011a). Without the use of $\widetilde{Z}_n(\lambda)$, our BRW bootstrap method is invalid under the alternative, because the term $Ee_{t,j}$ involved in the covariance function of $S(\lambda)$ can not be captured in this case.

Let d_ω be any metric that metricizes weak convergence in $L_2[0, \pi]$, and $\mathcal{L}(\xi_n|\chi_n)$ be the distribution of any random variable ξ_n given the sample $\chi_n := \{y_1, \cdots, y_n\}$; see Politis and Romano (1994). Denote by P^*, E^* and var^* the probability, expectation and variance conditional on χ_n; by $o_p^*(1)(O_p^*(1))$ a sequence of random variables converging to zero (bounded) in probability conditional on χ_n. We are now ready to present our second main result:

THEOREM 2. *Assume that (a) Assumptions* 1-3 *hold; (b)* $E|y_t|^{8+4\nu} < \infty$ *for some* $\nu > 0$ *and* $\lim_{k\to\infty} k^2[\alpha_y(k)]^{\nu/(2+\nu)} = 0$*; (c)* $b_n^{-1} = o(1)$ *and* $b_n = o(n^{1/3})$*; (d)* $E(w_t^*)^4 < \infty$*. Then, as* $n \to \infty$*,*

$$(i)\ d_\omega \left[\mathcal{L}\{\Delta_n(\lambda)|\chi_n\}, \mathcal{L}\{S(\lambda)\}\right] \to_p 0;$$

(ii) consequently,

$$\widetilde{CM}_n^* \to_d \int_0^\pi S^2(\lambda) d\lambda \ \textit{ in probability.}$$

Remark 5. An exponentially fast decaying $\alpha_y(k)$ is sufficient for the condition on $\alpha_y(k)$ in Theorem 2 to hold.

Compared to the conditions in Shao (2011a), our conditions in Theorem 2 are stronger. This is a price we pay for not assuming a stronger cumulant summability condition:

$$\sum_{s_1,\cdots,s_K=-\infty}^{\infty} |s_k||cum(y_0, y_{s_1} \cdots, y_{s_K})| < \infty, \ k = 1, \cdots, K, \tag{7}$$

for $K = 1, \cdots, 7$. Although (7) is implied by the GMC(8) condition of y_t according to Proposition 2 of Wu and Shao (2004), a sufficient condition for (7) in the context of α-mixing condition is still unknown. If (7) holds, following a similar proof in Shao (2011a, p.221-222), we can easily show that Theorem 2 holds under some weaker conditions. We summarize it in the following theorem:

THEOREM 3. *Assume that (a) Assumptions* 1-3 *and (7) hold; (b)* $Ey_t^8 < \infty$*; (c)* $b_n^{-1} = o(1)$ *and* $(\log n)b_n = o(n)$*; (d)* $E(w_t^*)^4 < \infty$*. Then, the conclusions in Theorem* 2 *hold.*

Remark 6. Theorems 2-3 guarantee that when J is large, our bootstrapped critical values from the BRW method are valid for $\widetilde{\mathrm{CM}}_n$ under the null or the alternative hypothesis. The reason why our bootstrap method works is probably because the conditional distribution of $\sqrt{n}(\theta_n^* - \theta_n)$ can always well mimic the distribution of $\sqrt{n}(\theta_n - \breve{\theta}_0)$, and this guarantees that we can handle the estimation effect successfully without generating any pseudo series. To see it clearly, from Lemma A3(ii) in the Appendix, we have

$$\sqrt{n}(\theta_n - \breve{\theta}_0) \to_d N(0, V) \text{ as } n \to \infty, \tag{8}$$

where $V = \Sigma^{-1}\Omega\Sigma^{-1}$ and

$$\Omega = \lim_{n\to\infty} var\left[\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{\partial l_t(\breve{\theta}_0)}{\partial\theta}\right] \tag{9}$$

is well defined by Lemma 3 in Francq and Zakoïan (1998). By (A39) in the Appendix, it is not hard to see that under the conditions of Theorem 2 or 3, we have

$$\mathcal{L}\left\{\sqrt{n}(\theta_n^* - \theta_n)|\chi_n\right\} \to_d N(0, V) \text{ in probability as } n \to \infty. \tag{10}$$

By (8) and (10), it follows that we can estimate V by $\widetilde{V}_n$ even when model (1) is misspecified (i.e., $\breve{\theta}_0 \neq \theta_0$) under the alternative hypothesis, where

$$\widetilde{V}_n := \text{sample variance-covariance matrix of } \{\sqrt{n}(\theta_{1n}^* - \theta_n), \cdots, \sqrt{n}(\theta_{Jn}^* - \theta_n)\}, \tag{11}$$

and θ_{in}^* for $i = 1, \cdots, J$ is calculated from step 2 in our BRW procedure. Particularly, by using $\widetilde{V}_n$, the classical Wald test is now applicable for model (1).

Remark 7. Since ε_t is unobservable with unknown dependent structure beyond MDS, another intuitive way to bootstrap the null distribution of $\widetilde{\mathrm{CM}}_n$ is using the residual-based block bootstrap in Paparoditis and Politis (2003) or the residual-based stationary bootstrap in Parker, Paparoditis, and Politis (2006). Both methods have been used for unit root testing. Their ideas are to generate pseudo series $\{y_t^{**}\}$ from model (1) with parameter θ_n, where the error sequence is obtained by sampling blocks of the residual series $\{\widetilde{\varepsilon}_t\}$ randomly with replacement. Then, we can calculate the bootstrapped samples $\{\widetilde{\mathrm{CM}}_{in}^{**}\}_{i=1}^{J}$ from the bootstrapped residuals $\{\widetilde{\varepsilon}_t^{**}(\theta_n^{**})\}$, where $\theta_n^{**} = \arg\min_{\Theta} n^{-1}\sum_{t=1}^{n}[\widetilde{\varepsilon}_t^{**}(\theta)]^2$, and

$$\widetilde{\varepsilon}_t^{**}(\theta) = y_t^{**} - \sum_{i=1}^{p}\phi_i y_{t-i}^{**} - \sum_{i=1}^{q}\psi_i\widetilde{\varepsilon}_t^{**}(\theta).$$

However, y_t^{**} may not well mimic y_t, because the residual series $\{\widetilde{\varepsilon}_t\}$ is correlated under the alternative hypothesis; see also Jentsch, Politis, and Paparoditis (2014) on a similar phenomenon for integrated processes. Based on this concern, the empirical distribution of $\{\widetilde{\mathrm{CM}}_{in}^{**}\}$ can not properly mimic the null distribution, and hence we do not follow the residual-based block methods in this paper.

Remark 8. In order to implement $\widetilde{\mathrm{CM}}_n$ in practice, we need to first select the orders p and q. When $\{\varepsilon_t\}$ is an iid sequence, this can be done by many well-known information criteria such as AIC, BIC, or EACF; see, e.g., Tsay and Tiao (1984). When $\{\varepsilon_t\}$ is an MDS, Chen, Min, and Chen (2013) proposed some order determination schemes based on the ACF and PACF of y_t. When $\{\varepsilon_t\}$ is an uncorrelated sequence, there is no valid method to select the orders so far, and we suggest using a Wald test Ξ_n to choose the orders by detecting the hypothesis $\Gamma\check{\theta}_0 = 0_{s\times 1}$, where Γ is a $s \times (p+q)$ constant matrix with rank s, and

$$\Xi_n = n(\Gamma\theta_n)'(\Gamma\widetilde{V}_n\Gamma')^{-1}(\Gamma\theta_n)$$

with $\widetilde{V}_n$ being defined as in (11). By using Ξ_n, it is now possible to reduce the orders from a vast fitted model, and this could enhance the power of our test.

Remark 9. By a repetitive but even simpler proof as in the Appendix, we can show that Theorems 2-3 hold if $b_n = 1$ when ε_t is an MDS. However, when ε_t is not an MDS, we need a block technique (i.e., $b_n \neq 1$) to capture the dependence of ε_t beyond MDS; see, e.g., Romano and Thombs (1996), Horowitz, Lobato, Nankervis, and Savin (2006), and Shao (2011a). Compared to the proof of Theorem 1 in the unconditional case, the proofs of Theorems 2-3 in the conditional case follow the same idea but with more nontrivial proofs caused by the block technique.

Finally, it is worth noting that Theorem 2 requires a stronger condition for b_n than Theorem 3. This demonstrates that if we allow for a more general structure of y_t, we may suffer from a smaller valid range of b_n. Hence, there is a tradeoff between the dependence structure of y_t and the theoretical valid range of b_n. Also, we should highlight that the often used block-wise wild bootstrap method applies the random weight w_t^* to the residual $\widetilde{\varepsilon}_t$ directly, but this method may fail in some cases as shown in Section 4.3 of Brüggemann, Jentsch, and Trenkler (2014). Our BRW method essentially applies the random weight w_t^* to the product of the residual $\widetilde{\varepsilon}_t\widetilde{\varepsilon}_{t+j}$ $(j \geq 1)$, and so it is different from the often used one. As one referee pointed out, this feature of our BRW method may lead to a robust size performance in terms of b_n, especially when the dependence structure of $\varepsilon_t\varepsilon_{t+j}(j \geq 1)$ is weak; see, e.g., Figure 1 in Section 4 below. Nevertheless, how to select the optimal b_n under certain "criterion" is unknown up to now. This is a familiar problem with all blocking methods. The heuristic work in Hall, Horowitz, and Jing (1995) and Politis, Romano, and Wolf (1999) may be extended in this case, and we leave it for future study.

4. Simulation studies

In this section, we examine the finite-sample performance of $\widetilde{\mathrm{CM}}_n$ for several weak ARMA models. As a comparison, we also consider the kernel-based test T_n in (5). Under H_0 and certain conditions, Shao (2011b) showed that

$$\frac{nT_n - m_n C(K)}{\sqrt{2m_n D(K)}} \to_d N(0,1) \quad \text{as } n \to \infty,$$

where $C(K) = \int_0^\infty K^2(x)dx$ and $D(K) = \int_0^\infty K^4(x)dx$ with the kernel function $K(\cdot)$ and the bandwidth m_n being chosen as in (5). So, we reject H_0 at the significance level α, if $nT_n > \sqrt{2m_n D(K)}c_\alpha + m_n C(K)$, where c_α is the $(1-\alpha)$-th percentile of $N(0,1)$. Under H_1 and certain conditions, Shao (2011b) proved that $P(nT_n > \sqrt{2m_n D(K)}c_\alpha + m_n C(K)) = 1$ as $n \to \infty$, and hence this spectral test is consistent (see also Hong (1996)).

Next, we introduce our basic set-up. In all calculations, we generate 1000 replications of sample size $n = 400$ and 1000 from each specified model in Examples 1-4 below, and choose the significance level $\alpha = 1\%, 5\%$ or 10%. For $\widetilde{\text{CM}}_n$, we use 500 bootstrap samples in each replication with block size $b_n = n^{1/5}, 2n^{1/5}, \sqrt{n}/2, \sqrt{n}$ or $2\sqrt{n}$ to obtain its corresponding critical value for every aforementioned significance level α. These choices of set-up deliver $b_n = 3, 6, 10, 20, 40$ for $n = 400$ and $3, 7, 15, 31, 63$ for $n = 1000$. Here, δ_t is employed from the following Bernoulli distribution:

$$P\left(\delta_t = \frac{3-\sqrt{5}}{2}\right) = \frac{1+\sqrt{5}}{2\sqrt{5}} \text{ and } P\left(\delta_t = \frac{3+\sqrt{5}}{2}\right) = 1 - \frac{1+\sqrt{5}}{2\sqrt{5}},$$

although other choices like the standard exponential distribution are also suitable for δ_t. For T_n, we use the Parzen kernel $K(x)$ defined as

$$K(x) = \begin{cases} 1 - 6x^2 + 6|x|^3 & \text{for } 0 \le |x| \le 1/2, \\ 2(1-|x|)^3 & \text{for } 1/2 \le |x| \le 1, \\ 0 & \text{otherwise.} \end{cases}$$

In general, since there is no clear objective procedure for optimally choosing the bandwidth m_n, we carry out the calculation for $m_n = 1, \cdots, 20$ when $n = 400$ and $1, \cdots, 32$ when $n = 1000$. In most cases of m_n, we find that the sizes of T_n are distorted (see Figure 1 below). Hence, only the results in which the sizes are close to their nominal ones are reported, although the corresponding choices of m_n may not be optimal in some sense.

Example 1. Consider the following weak ARMA$(1, 1)$ model:

$$y_t = \kappa y_{t-1} + 0.8\varepsilon_{t-1} + \varepsilon_t \text{ and } \varepsilon_t = \eta_t^2 \eta_{t-1}, \tag{12}$$

where η_t is a sequence of iid N$(0, 1)$ random variables, and $\kappa \in \{0.0, 0.1, 0.2, 0.3, 0.4\}$. Clearly, ε_t in (12) are uncorrelated but non-MDS. Next, we use $\widetilde{\text{CM}}_n$ and T_n to detect whether a weak MA(1) model is adequate to fit the data sample generated from model (12). The empirical power and sizes of both tests are reported in Table 1, and the sizes correspond to the cases where $\kappa = 0.0$.

Example 2. Consider the following weak AR(2) model:

$$y_t = 0.5y_{t-1} + \kappa y_{t-2} + \varepsilon_t \text{ and } \varepsilon_t = \eta_t^2 \eta_{t-1}, \tag{13}$$

where η_t is a sequence of iid N$(0, 1)$ random variables, and $\kappa \in \{0.0, 0.1, 0.2, 0.3, 0.4\}$. We use $\widetilde{\text{CM}}_n$ and T_n to detect whether a weak AR(1) model is adequate to fit the data sample generated from model (13). The empirical power and sizes of both tests are reported in Table 2, and the sizes correspond to the cases where $\kappa = 0.0$.

Example 3. Consider the following switching-regime Markov model (see, e.g., Hamilton (1994)):

$$y_t = \kappa y_{t-1} + \eta_t + (0.2 + 0.3\Delta_t)\eta_{t-1}, \tag{14}$$

where Δ_t is a sequence of Bernoulli random variables with $P(\Delta_t = 0) = 1/3$ and $P(\Delta_t = 1) = 2/3$, η_t is a sequence of iid N$(0, 1)$ random variables, and $\kappa \in \{0.0, 0.05, 0.1, 0.15, 0.2\}$. Here, we assume that Δ_t and η_t are independent. When $\kappa = 0.0$, Francq and Zakoïan (1998) showed that model (14) admits a weak MA(1) representation: $y_t = \varepsilon_t + \varphi\varepsilon_{t-1}$, where ε_t are uncorrelated but non-MDS. Thus, we can use $\widetilde{\text{CM}}_n$ and T_n to detect whether a weak MA(1) model is adequate to fit the data sample generated from model (14). The empirical power and sizes of both tests are reported in Table 3, and the sizes correspond to the cases where $\kappa = 0.0$.

Table 1. *Empirical sizes and power ($\times 100$) for $\widetilde{CM}_n$ and T_n in model (12).*

			$\kappa = 0.0$			$\kappa = 0.1$			$\kappa = 0.2$			$\kappa = 0.3$			$\kappa = 0.4$		
Tests	n	$b_n(m_n)$	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%
$\widetilde{\text{CM}}_n$	400	3	1.3	6.8	12.5	3.9	14.1	26.0	22.0	49.0	64.4	54.9	80.2	89.1	80.1	93.7	96.8
		6	1.1	5.5	11.5	3.3	14.0	26.5	19.9	44.1	59.7	50.2	77.8	87.3	73.2	91.2	95.5
		10	1.6	5.5	10.9	4.2	15.3	27.1	22.0	47.3	60.7	49.6	75.6	87.1	68.6	88.0	95.6
		20	1.3	6.6	13.3	5.4	17.1	26.2	21.8	46.8	59.7	47.9	72.4	82.7	64.9	85.7	93.7
		40	3.2	7.8	13.3	8.4	16.8	25.0	25.1	44.3	56.4	48.5	68.4	80.1	63.8	80.5	89.9
T_n		3	1.4	2.0	3.8	8.9	12.9	16.6	37.4	46.5	52.1	80.2	86.0	89.3	97.1	98.3	98.6
		4	3.1	6.6	8.2	15.5	20.7	24.6	53.8	61.4	65.8	88.4	91.2	92.9	98.1	99.0	99.5
$\widetilde{\text{CM}}_n$	1000	3	1.2	5.1	11.6	13.2	35.6	48.1	63.8	82.7	88.8	94.4	98.4	99.2	99.1	99.8	99.9
		7	1.0	4.3	9.3	13.9	31.9	46.0	60.1	82.1	89.6	93.5	97.8	99.2	98.9	99.8	99.9
		15	1.2	5.3	11.8	13.8	33.4	44.8	62.6	82.7	90.5	91.5	97.8	99.0	97.9	99.7	99.8
		31	0.9	6.2	12.5	13.2	34.3	47.9	62.9	83.9	91.1	90.2	98.7	99.7	94.6	99.2	99.8
		63	2.1	6.3	11.7	17.1	31.6	46.2	65.7	82.3	88.4	86.5	95.8	97.9	88.5	96.6	99.0
T_n		3	2.9	4.9	6.2	21.5	30.2	35.5	79.3	84.1	86.7	98.9	99.5	99.7	100	100	100
		4	5.4	8.2	11.1	33.0	41.2	46.2	87.3	91.2	92.6	99.9	100	100	100	100	100

Table 2. *Empirical sizes and power ($\times 100$) for $\widetilde{CM}_n$ and T_n in model (13).*

			$\kappa = 0.0$			$\kappa = 0.1$			$\kappa = 0.2$			$\kappa = 0.3$			$\kappa = 0.4$		
Tests	n	$b_n(m_n)$	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%
$\widetilde{\text{CM}}_n$	400	3	0.5	4.6	9.9	19.5	35.9	46.8	63.7	80.9	86.2	89.6	95.2	96.9	97.0	98.7	99.2
		6	2.0	4.4	10.1	18.0	35.2	46.0	61.4	79.0	86.0	86.2	93.3	95.9	94.4	98.5	99.3
		10	1.1	4.8	10.4	20.0	39.8	51.5	66.3	83.4	88.6	86.7	95.1	97.1	95.6	98.5	99.0
		20	1.6	6.1	12.7	22.6	41.4	52.1	65.6	81.5	87.8	87.4	96.1	97.5	94.1	98.5	99.1
		40	2.6	6.4	13.1	25.4	37.7	47.3	64.2	78.7	84.9	85.4	94.8	97.4	92.4	98.3	98.8
T_n		10	1.9	3.4	4.9	15.1	23.6	28.5	72.2	78.2	82.5	97.0	97.7	98.0	99.9	99.9	99.9
		15	2.4	4.3	6.8	16.9	24.6	30.4	72.5	79.5	82.7	96.8	98.4	98.7	99.9	99.9	100
$\widetilde{\text{CM}}_n$	1000	3	1.3	5.6	11.1	45.2	65.7	75.3	93.8	98.2	98.6	99.6	99.9	100	100	100	100
		7	1.1	6.3	10.8	48.7	67.9	74.7	93.2	97.1	98.4	99.7	99.9	99.9	99.9	100	100
		15	1.2	6.4	12.2	48.3	67.1	75.9	92.4	97.6	98.7	99.7	99.9	100	99.7	100	100
		31	1.2	5.2	11.7	48.1	66.6	74.9	92.6	96.6	98.0	99.3	99.9	99.9	99.8	100	100
		63	1.6	6.4	11.5	49.9	66.7	75.5	94.6	97.8	98.8	99.4	99.8	100	99.8	100	100
T_n		9	2.0	3.7	5.6	44.7	56.0	62.5	97.8	98.9	99.4	100	100	100	100	100	100
		13	2.9	5.4	7.5	51.0	60.3	64.9	98.1	99.0	99.3	100	100	100	100	100	100

Example 4. Consider the following bilinear model (see, e.g., Granger and Andersen (1978) and Pham (1986)):

$$y_t = \kappa\eta_{t-1} + \eta_t + 0.2 y_{t-1}\eta_{t-2}, \tag{15}$$

where η_t is a sequence of iid N$(0,1)$ random variables, and $\kappa \in \{0.0, 0.05, 0.1, 0.15, 0.2\}$. When $\kappa = 0.0$, Francq and Zakoïan (1998) showed that model (15) admits a weak MA(3) representation: $y_t = \varepsilon_t + \varphi\varepsilon_{t-3}$, where ε_t are uncorrelated but non-MDS. Thus, we can use $\widetilde{\text{CM}}_n$ and T_n to detect whether a weak MA(3) model is adequate to fit the data sample generated from model (15). The empirical power and sizes of both tests are reported in Table 4, and the sizes correspond to the cases where $\kappa = 0.0$.

Table 3. *Empirical sizes and power* ($\times 100$) *for* $\widetilde{CM}_n$ *and* T_n *in model* (14).

			$\kappa=0.0$			$\kappa=0.05$			$\kappa=0.1$			$\kappa=0.15$			$\kappa=0.2$		
Tests	n	$b_n(m_n)$	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%
$\widetilde{\text{CM}}_n$	400	3	1.1	5.4	10.4	1.9	8.1	13.9	4.2	14.2	22.4	12.7	32.7	44.1	29.5	53.8	65.6
		6	1.7	5.7	12.4	2.0	7.3	14.4	3.7	13.5	22.2	14.8	32.5	45.5	31.6	55.2	67.9
		10	1.7	6.9	11.8	2.0	7.6	13.6	4.8	13.7	21.5	15.0	32.0	43.4	31.8	55.4	66.8
		20	2.4	7.1	12.1	3.1	9.0	15.2	6.7	14.8	23.9	16.9	32.3	43.4	33.9	53.3	65.3
		40	3.6	7.8	13.0	4.6	10.6	18.6	9.8	19.1	28.9	21.9	36.9	47.7	40.0	57.6	69.5
T_n		19	0.7	1.9	3.3	0.4	2.4	3.7	1.4	3.6	6.1	6.3	11.3	16.3	19.8	28.7	35.5
		20	0.9	2.1	3.4	0.8	2.3	4.4	2.2	4.8	8.3	7.2	13.7	17.6	16.7	28.0	34.7
$\widetilde{\text{CM}}_n$	1000	3	0.9	5.8	10.8	2.7	9.5	17.3	15.2	33.4	44.9	39.6	63.1	75.2	79.7	91.6	94.9
		7	1.6	5.1	10.5	4.6	10.9	17.5	14.5	29.8	42.1	40.9	63.6	75.1	79.2	91.3	95.7
		15	1.3	4.7	10.1	3.9	11.2	18.4	14.7	32.5	44.3	43.8	65.7	74.8	79.2	90.8	95.1
		31	1.7	6.1	10.6	4.2	11.4	17.3	16.5	33.9	45.1	47.4	69.4	79.5	79.1	90.5	94.7
		63	3.7	8.9	13.6	4.0	11.5	18.6	20.3	36.1	46.7	48.5	67.1	75.4	81.4	91.9	95.5
T_n		21	0.9	2.4	4.0	1.9	4.0	6.5	7.7	12.7	17.2	24.4	37.0	44.5	61.7	74.8	79.6
		22	1.1	2.5	4.9	1.6	3.9	5.7	6.0	11.3	15.4	24.2	35.9	44.7	60.6	73.8	80.6

Table 4. *Empirical sizes and power* ($\times 100$) *for* $\widetilde{CM}_n$ *and* T_n *in model* (15).

			$\kappa=0.0$			$\kappa=0.05$			$\kappa=0.1$			$\kappa=0.15$			$\kappa=0.2$		
Tests	n	$b_n(m_n)$	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%
$\widetilde{\text{CM}}_n$	400	3	1.0	4.4	9.1	5.7	17.2	25.1	20.6	43.4	53.9	51.9	77.5	85.0	83.9	94.4	97.1
		6	2.4	7.9	12.7	4.9	15.6	24.0	21.8	43.3	55.3	53.8	76.3	83.5	82.5	95.2	97.9
		10	1.4	5.8	10.6	5.6	16.3	25.8	21.5	43.6	55.2	52.1	76.2	84.3	82.9	94.3	96.9
		20	2.9	8.6	15.9	5.2	14.0	22.6	26.4	46.6	57.5	58.7	78.9	86.7	82.2	93.7	97.1
		40	3.6	10.4	16.7	9.4	18.3	25.9	26.9	44.9	57.7	61.0	76.4	86.2	85.8	95.1	97.9
T_n		16	1.1	3.2	5.9	4.9	7.7	10.7	19.6	30.0	35.8	48.2	61.9	68.0	76.2	85.5	89.2
		17	1.1	3.5	5.2	3.0	7.9	10.7	19.1	28.5	33.3	46.2	58.7	65.1	75.8	84.6	88.7
$\widetilde{\text{CM}}_n$	1000	3	1.0	5.0	8.9	12.8	30.1	41.4	60.9	81.3	88.1	94.6	99.4	99.7	100	100	100
		7	0.8	5.5	10.9	13.2	31.7	44.2	58.5	80.6	88.0	94.7	98.5	99.3	100	100	100
		15	1.2	6.7	12.0	14.3	29.4	39.2	61.5	81.5	88.7	95.2	98.9	99.5	99.8	100	100
		31	2.3	7.3	11.8	15.1	30.5	42.6	62.2	81.7	89.2	94.8	98.6	99.6	99.7	99.9	99.9
		63	3.3	8.2	13.3	20.1	34.9	45.1	63.7	81.9	89.6	94.7	98.1	99.3	99.7	100	100
T_n		29	1.4	4.5	6.2	7.7	14.2	19.3	42.1	54.5	63.5	88.9	93.1	95.2	99.2	99.6	99.7
		30	1.5	4.2	6.9	8.4	15.1	19.8	43.7	57.1	64.9	87.6	93.0	95.3	99.2	99.7	99.8

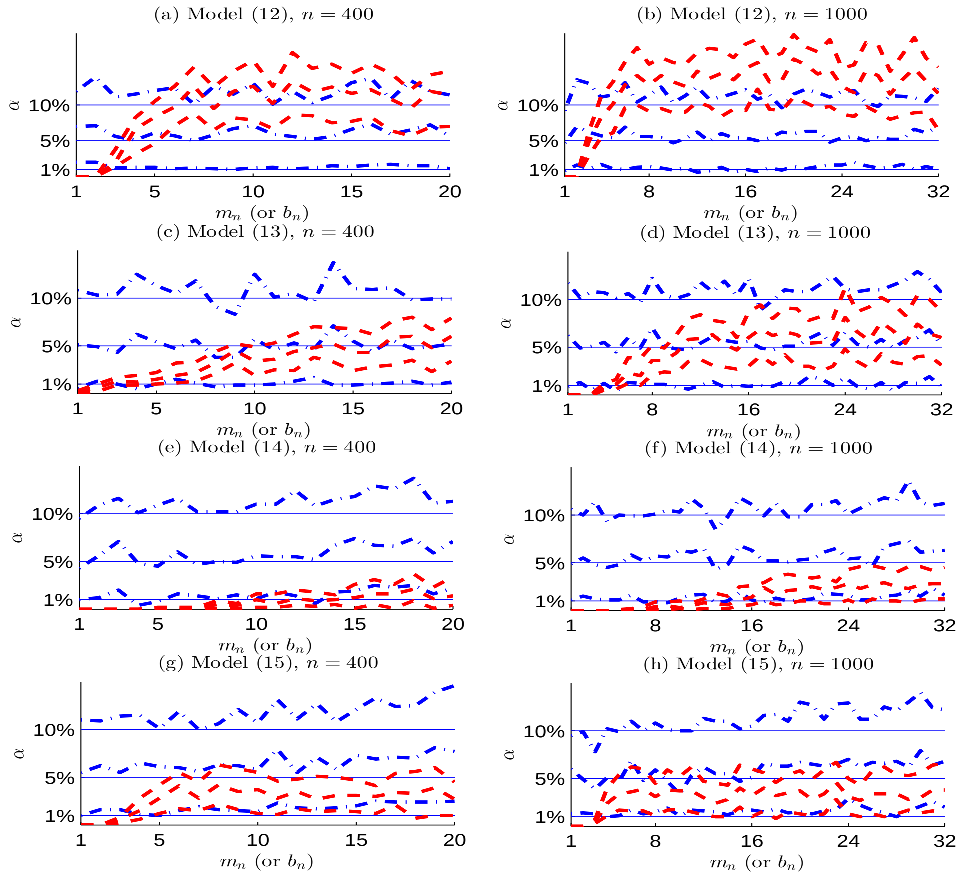

Fig. 1. In each panel, the dash (or dash-dot) lines from top to bottom are the sizes of T_n (or $\widetilde{\text{CM}}_n$) at the significance level $\alpha = 10\%, 5\%$ and 1% for each model with $\kappa = 0.0$, based on different values of m_n (or b_n), where the solid lines from top to bottom are the nominal significance level $\alpha = 10\%, 5\%$ and 1%.

From Tables 1-4, we find that the sizes of $\widetilde{\text{CM}}_n$ are close to their nominal ones when b_n is small (e.g., $b_n = n^{1/5}$ or $2n^{1/5}$). When b_n gets large, $\widetilde{\text{CM}}_n$ tends to be oversized in general, but the size distortion becomes weaker as n increases. This finding is consistent with the one in Shao (2011a). For T_n, we find that its size performance is very sensitive to the choice of m_n in models (12) and (13). A visual understanding of this phenomenon can be obtained in Figure 1(a)-(d), where we plot all the empirical sizes of T_n for different choices of m_n. As a comparison, the empirical sizes of $\widetilde{\text{CM}}_n$ for different choices of b_n are also plotted in Figure 1(a)-(d). It is clear that when m_n is larger, the sizes of T_n are seriously distorted at each significance level α, and when m_n is small, T_n tends to be seriously undersized at significance levels $\alpha = 5\%$ and 10%. This drawback of T_n is unchanged even when n becomes larger. By using other kernels (e.g., the Bartlett kernel and the quadratic spectral kernel), similar result holds for T_n, and hence they are not reported. Compared to T_n, the sizes of $\widetilde{\text{CM}}_n$ are much more robust at each significance level especially when b_n is small.

Furthermore, it is worth noting that unlike models (12) and (13), T_n is always undersized for different choices of m_n in models (14)-(15). This problem becomes extremely serious when m_n

is small. However, like models (12) and (13), the size performance of $\widetilde{\mathrm{CM}}_n$ is much more robust in those cases; see also Figure 1(e)-(h) for more visual evidence. From Figure 1, we find that the size performance of $\widetilde{\mathrm{CM}}_n$ is more satisfactory when $b_n = 1$ or $b_n \approx 1$. This is probably because our $\widetilde{\mathrm{CM}}_n$ test applies the random weight w_t^* to $\widetilde{\varepsilon}_t\widetilde{\varepsilon}_{t+h}$ not to $\widetilde{\varepsilon}_t$, and the dependence structure of $\widetilde{\varepsilon}_t\widetilde{\varepsilon}_{t+h}$ for $h \geq 1$ in models (12)-(15) is very weak (e.g., the autocovariance of $\varepsilon_t\varepsilon_{t+h}$ at each lag $j(> 0)$ is zero in models (12)-(13)). This feature of our $\widetilde{\mathrm{CM}}_n$ test may help us to explain its robust size performance, and a rigorous justification on this conjecture is an interesting topic for future study. Overall, we know that the sizes of $\widetilde{\mathrm{CM}}_n$ are precise especially when b_n is small, while the sizes of T_n could be seriously undersized or oversized in most cases of m_n. It means that the performance of T_n is heavily relied on whether we can obtain an optimal m_n, but this is not the case for $\widetilde{\mathrm{CM}}_n$. Considering the difficulty of selecting the optimal bandwidth in most of the nonparametric methods for practitioners, $\widetilde{\mathrm{CM}}_n$ has a size advantage over T_n in this direction.

Next, we consider the power performances for $\widetilde{\mathrm{CM}}_n$ and T_n, and the conclusion is generally as expected. First, all the powers become large as n increases. Second, $\widetilde{\mathrm{CM}}_n$ is generally more powerful than T_n for all examined alternatives in models (14)-(15), while T_n has a power advantage over $\widetilde{\mathrm{CM}}_n$ in model (12) and model (13) with a large κ. The power advantage of T_n over $\widetilde{\mathrm{CM}}_n$ in model (12) is probably because the residuals of a fitted MA(1) model exhibit certain autocorrelations at some large lags. The power advantage of T_n over $\widetilde{\mathrm{CM}}_n$ in model (13) is quite reasonable, since the sample autocovariance of the residuals from the fitted AR(1) model becomes more significantly different from zero when κ is larger. Overall, although $\widetilde{\mathrm{CM}}_n$ does not have a consistent power advantage over T_n, it is reasonable to recommend $\widetilde{\mathrm{CM}}_n$ in practice since it has a very robust size performance especially when the block size is small.

5. Application to S&P 500 stock index

In this section, we revisit the real example on S&P 500 stock index in Escanciano and Velasco (2006). We consider two sample periods for the S&P 500 stock index. The first period is from 3 January 1994 until 31 December 1997 with a total of 1011 observations. The second period is from 2 January 1998 until 28 August 2002 with a total of 1170 observations. Denote the log-return of both series (after mean-adjusted) by y_{1t} and y_{2t}, respectively. The generalized spectral tests in Escanciano and Velasco (2006, p.172) indicate that y_{1t} is non-MDS at the significance level $\alpha = 5\%$, while y_{2t} is non-MDS at the significance level $\alpha = 10\%$. Thus, we are of interest to test whether y_{1t} or y_{2t} is a weak white noise (i.e., an uncorrelated sequence) by using $\widetilde{\mathrm{CM}}_n$. As in Section 4, we choose $b_n = n^{1/5}, 2n^{1/5}, \sqrt{n}/2, \sqrt{n}$ or $2\sqrt{n}$, and it delivers $b_n = 3, 7, 15, 31$ for y_{1t} and $4, 8, 16, 32$ for y_{2t}. The corresponding results for $\widetilde{\mathrm{CM}}_n$ are listed in Table 5, from which we can not reject the hypothesis that y_{1t} or y_{2t} is a weak white noise at the 5% significance level, and this conclusion is unchanged for all choices of b_n. Thus, a weak but non-MDS process should be suitable to fit y_{1t} or y_{2t}.

Next, we use $\widetilde{\mathrm{CM}}_n$ to check whether a weak MA(3) model defined as $y_t = \varepsilon_t + \varphi\varepsilon_{t-3}$ for $|\varphi| < 1$, is adequate to fit y_{1t} or y_{2t}. Based on LS estimation, the fitted weak MA(3) models for y_{1t} and y_{2t} are as follows:

$$y_{1t} = \varepsilon_{1t} - 0.0482\varepsilon_{1t-3}, \tag{16}$$

$$y_{2t} = \varepsilon_{2t} - 0.0423\varepsilon_{2t-3}, \tag{17}$$

Table 5. *p-values of $\widetilde{\mathrm{CM}}_n$ for testing the adequacy of a weak white noise on two S&P 500 stock indexes*

		b_n				
Series		$n^{1/5}$	$2n^{1/5}$	$\sqrt{n}/2$	$\sqrt{n}$	$2\sqrt{n}$
y_{1t}	p-value†	0.6900	0.6537	0.5050	0.6257	0.5637
y_{2t}	p-value	0.5110	0.5180	0.4017	0.4157	0.2783

† p-values bootstrapped by the BRW method with $J = 3000$.

where the estimated values of $\sigma^2_{\varepsilon_1} = 6.2 \times 10^{-5}$ and $\sigma^2_{\varepsilon_2} = 1.8 \times 10^{-4}$. The p-values of $\widetilde{\mathrm{CM}}_n$ in Table 6 indicate that models (16)-(17) are adequate at the 5% significance level, while the p-values of the Ljung-Box test statistics $Q(M)$ and Li-Mak test statistics $Q^2(M)$ in Table 7 imply that models (16)-(17) are not strong at the same significance level. Note that a Bilinear model like (15) with $\kappa = 0$ has a weak MA(3) representation. Thus, it motivates us to fit y_{1t} or y_{2t} by the following Bilinear-GARCH model:

$$\begin{cases} y_t = \eta_t + u y_{t-1}\eta_{t-2}, \\ \eta_t = \sqrt{h_t}\nu_t \quad \text{and} \quad h_t = \omega + \alpha\eta^2_{t-1} + \beta h_{t-1}, \end{cases} \tag{18}$$

where $|u| < 1$, $\omega > 0$, $\alpha, \beta \geq 0$ and ν_t is an iid re-scaled error sequence. For each series, model (18) is estimated by using the QMLE method (see, e.g., Ling (2007) and Francq and Zakoïan (2010)). The related results are summarized in Table 8, from which we know that model (18) is adequate to fit y_{2t}, while a marginal autocorrelation up to lag 6 is detected in the fitted conditional mean model for y_{1t}. Based on this, we re-fit y_{1t} by another Bilinear-GARCH model:

$$\begin{cases} y_t = v\eta_{t-1} + \eta_t + u y_{t-1}\eta_{t-2}, \\ \eta_t = \sqrt{h_t}\nu_t \quad \text{and} \quad h_t = \omega + \alpha\eta^2_{t-1} + \beta h_{t-1}, \end{cases} \tag{19}$$

where $|v| < 1, |u| < 1$, $\omega > 0$, $\alpha, \beta \geq 0$ and ν_t is an iid re-scaled error sequence. The related results for the fitted model (19) are given in Table 8, from which we know that model (19) is adequate in fitting y_{1t}.

Table 6. *p-values of $\widetilde{\mathrm{CM}}_n$ for testing the adequacy of a weak MA(3) model on two S&P 500 stock indexes*

		b_n				
Series		$n^{1/5}$	$2n^{1/5}$	$\sqrt{n}/2$	$\sqrt{n}$	$2\sqrt{n}$
y_{1t}	p-value†	0.9087	0.8923	0.8637	0.9707	0.9627
y_{2t}	p-value	0.8420	0.8630	0.6720	0.5560	0.4940

† p-values bootstrapped by the BRW method with $J = 3000$.

Table 7. *p-values of $Q(M)$ and $Q^2(M)$ for testing the adequacy of a strong MA(3) model on two S&P 500 stock indexes*

Series		$Q(6)$	$Q(12)$	$Q(24)$	$Q^2(6)$	$Q^2(12)$	$Q^2(24)$
y_{1t}	p-value	0.3453	0.0106	0.0588	0.0000	0.0000	0.0000
y_{2t}	p-value	0.2756	0.1774	0.2689	0.0000	0.0000	0.0000

Table 8. *QMLE-fitted model and its corresponding portmanteau tests on two S&P 500 stock indexes*

		$QMLE$									
	Series	v_n	u_n	ω_n	α_n	β_n	σ_ν^2	$Q(6)$	$Q(24)$	$Q^2(6)$	$Q^2(24)^\dagger$
Model (18)	y_{1t}	$---$	0.9961	0.0000	0.1045	0.8686	0.9984	0.0461	0.2591	0.9517	0.9945
	y_{2t}	$---$	0.8004	0.0000	0.1129	0.8213	0.9984	0.4106	0.3525	0.2549	0.6193
Model (19)	y_{1t}	0.0703	0.8001	0.0000	0.1083	0.8650	0.9971	0.4310	0.6353	0.9614	0.9951

† p-values for the Ljung-Box test statistics $Q(6)$ and $Q(24)$, and the Li-Mak test statistics $Q^2(6)$ and $Q^2(24)$.

6. CONCLUDING REMARKS

In this paper, we study the asymptotic property of a CM-type spectral test statistic $\widetilde{\text{CM}}_n$ for checking the adequacy of an ARMA model with uncorrelated errors. By releasing the martingale difference assumption on the error terms, $\widetilde{\text{CM}}_n$ is applicable to a large class of uncorrelated nonlinear processes. Since we do not specify the form of error terms, the limiting distribution of $\widetilde{\text{CM}}_n$ is not pivotal, and so a novel BRW method is necessary to bootstrap the critical values of $\widetilde{\text{CM}}_n$. Simulation studies show that the size and power performances of $\widetilde{\text{CM}}_n$ are robust to the selection of block size b_n in BRW method especially when the sample size is large, while the size of kernel-based test T_n in Shao (2011b) is always sensitive to the choice of the bandwidth m_n. In addition, $\widetilde{\text{CM}}_n$ has a power advantage over T_n under most of the examined alternatives. By revisiting two S&P 500 stock index series in Escanciano and Velasco (2006), $\widetilde{\text{CM}}_n$ suggests that the Bilinear-GARCH models are adequate to fit both series. This empirical example illustrates that although some economic or financial series is not a martingale difference sequence, it is still very likely to be an uncorrelated sequence. Our test statistic $\widetilde{\text{CM}}_n$ now gives us a way to check for the adequacy of ARMA models driven by an uncorrelated error sequence. Moreover, once a weak ARMA model is found to be adequate in fitting the given series, some non-linear processes with a weak ARMA representation may also be considered to fit this series adequately. This point of view should be important for practitioners.

ACKNOWLEDGEMENT

The authors greatly appreciate the very constructive suggestions and comments of P. Doukhan, three anonymous referees, the Associate Editor, and the Editor J. Fan that greatly improve this article. This work is supported by Research Grants Council of the Hong Kong SAR Government, GRF grant HKU703711P, National Natural Science Foundation of China (No.11201459), President Fund of Academy of Mathematics and Systems Science, CAS (No.2014-cjrwlzx-zk), and Key Laboratory of RCSDS, CAS.

APPENDIX: PROOFS

Denote by $W_h(j) = \int_0^\pi h(\lambda)\psi_j(\lambda)d\lambda$ for any $h \in L_2[0,\pi]$; by $P_j = \int_0^\pi \psi_j^2(\lambda)d\lambda$ for $j \in \mathcal{N}$; by C a positive generic constant which may vary from place to place. Note that $P_j \le Cj^{-2}$ uniformly in $j \in \mathcal{N}$, and $\int_0^\pi \psi_j(\lambda)\psi_k(\lambda)d\lambda = 0$ when $j \neq k$ and $j, k \in \mathcal{N}$. In order to prove Theorem 1, we rewrite

$$\begin{aligned}\widetilde{S}_n(\lambda) &= \left[\widetilde{S}_n(\lambda) - \breve{S}_n(\lambda)\right] + \breve{S}_n(\lambda)\\ &= \left[\widetilde{S}_n(\lambda) - \breve{S}_n(\lambda)\right] + \left[\breve{S}_n(\lambda) - \check{S}_n(\lambda)\right] + \check{S}_n(\lambda)\\ &= I_{1n}(\lambda) + I_{2n}(\lambda) + \check{S}_n(\lambda) \text{ say.} \qquad \text{(A1)}\end{aligned}$$

where $\breve{S}_n(\lambda) = \sum_{j=1}^{n-1}\sqrt{n}\breve{\gamma}(j)\psi_j(\lambda)$ with $\breve{\gamma}(j) = n^{-1}\sum_{t=1+|j|}^n \breve{\varepsilon}_t\breve{\varepsilon}_{t-|j|}$ and $\breve{\varepsilon}_t = \varepsilon_t(\theta_n)$. Then, we need the following five lemmas:

LEMMA A1. *Suppose that Assumption* 1 *holds. Then, there exist* $C > 0$ *and* $\rho \in (0,1)$ *such that*

$$(i)\ \sup_\Theta |\varepsilon_t(\theta)| < C\xi_{\rho t},\ \sup_\Theta \left\|\frac{\partial \varepsilon_t(\theta)}{\partial\theta}\right\| < C\xi_{\rho t-1},\ \textit{and} \sup_\Theta \left\|\frac{\partial^2 \varepsilon_t(\theta)}{\partial\theta\partial\theta'}\right\| < C\xi_{\rho t-1};$$

$$(ii)\ \sup_\Theta |\varepsilon_t(\theta) - \widetilde{\varepsilon}_t(\theta)| \le O(\rho^t)\xi_{\rho 0},\ \sup_\Theta \left\|\frac{\partial \varepsilon_t(\theta)}{\partial\theta} - \frac{\partial \widetilde{\varepsilon}_t(\theta)}{\partial\theta}\right\| \le O(\rho^t)\xi_{\rho 0},$$

$$\textit{and} \sup_\Theta \left\|\frac{\partial^2 \varepsilon_t(\theta)}{\partial\theta\partial\theta'} - \frac{\partial^2 \widetilde{\varepsilon}_t(\theta)}{\partial\theta\partial\theta'}\right\| \le O(\rho^t)\xi_{\rho 0},$$

where $\xi_{\rho t} = 1 + \sum_{i=0}^\infty \rho^i |y_{t-i}|$.

Proof. The proof follows directly from Lemmas A.1 and A.4 in Ling (2007). □

LEMMA A2. *Suppose that Assumptions* 1-2 *hold. Then,* $\|I_{1n}(\lambda)\|^2 = o_p(1)$.

Proof. By direct calculation, we have

$$E\|I_{1n}(\lambda)\|^2 = \frac{1}{n}\sum_{j=1}^{n-1} E\left(\sum_{t=1+j}^n b_{tj}(\theta_n)\right)^2 P_j,$$

where $b_{tj}(\theta) = \varepsilon_t(\theta)\varepsilon_{t-j}(\theta) - \widetilde{\varepsilon}_t(\theta)\widetilde{\varepsilon}_{t-j}(\theta)$. By Minkowski inequality, it follows that

$$\begin{aligned}E\|I_{1n}(\lambda)\|^2 &\le \frac{1}{n}\sum_{j=1}^{n-1}\left(\sum_{t=1+j}^n \left\{E\left[b_{tj}(\theta_n)\right]^2\right\}^{1/2}\right)^2 P_j\\ &\le \frac{1}{n}\sum_{j=1}^{n-1}\left(\sum_{t=1+j}^n \left\{E\left[\sup_\Theta \|b_{tj}(\theta)\|\right]^2\right\}^{1/2}\right)^2 P_j. \qquad \text{(A2)}\end{aligned}$$

By Lemma A1, we know that there exists a constant $\rho \in (0,1)$ such that

$$\begin{aligned}\sup_\Theta \|b_{tj}(\theta)\| &\le \sup_\Theta \|\left[\varepsilon_t(\theta) - \widetilde{\varepsilon}_t(\theta)\right]\varepsilon_{t-j}(\theta)\| + \sup_\Theta \|\widetilde{\varepsilon}_t(\theta)\left[\varepsilon_{t-j}(\theta) - \widetilde{\varepsilon}_{t-j}(\theta)\right]\|\\ &\le O(\rho^t)\xi_{\rho 0}\xi_{\rho t-j} + O(\rho^{t-j})\xi_{\rho 0}\xi_{\rho t}.\end{aligned}$$

Note that $E|\xi_{\rho t}|^4 < \infty$ by Assumption 2. Thus, from (A2), by Hölder inequality, we can show that

$$
\begin{aligned}
E\|I_{1n}(\lambda)\|^2 &\leq \frac{1}{n}\sum_{j=1}^{n-1}\left(\sum_{t=1+j}^{n}\left\{O(\rho^{2t})E\left[\xi_{\rho 0}\xi_{\rho t-j}\right]^2 + O(\rho^{2(t-j)})E\left[\xi_{\rho 0}\xi_{\rho t}\right]^2\right\}^{1/2}\right)^2 P_j \\
&\leq \frac{1}{n}\sum_{j=1}^{n-1}\left(\sum_{t=1+j}^{n}\left\{O(\rho^{2t})\left(E\left[\xi_{\rho 0}\right]^4 E\left[\xi_{\rho t-j}\right]^4\right)^{1/2}\right.\right. \\
&\qquad \left.\left. +O(\rho^{2(t-j)})\left(E\left[\xi_{\rho 0}\right]^4 E\left[\xi_{\rho t}\right]^4\right)^{1/2}\right\}^{1/2}\right)^2 P_j \\
&\leq \frac{1}{n}\sum_{j=1}^{n-1}\left(\sum_{t=1+j}^{n}\left\{O(\rho^{t}) + O(\rho^{t-j})\right\}\right)^2 P_j = O\left(\frac{1}{n}\right),
\end{aligned}
$$

which implies that $\|I_{1n}(\lambda)\|^2 = o_p(1)$. □

LEMMA A3. *Suppose that Assumptions* 1-3 *hold. Then,*

(i) $E\left[\frac{\partial l_t(\breve{\theta}_0)}{\partial \theta}\right] = 0;$

(ii) $\sqrt{n}(\theta_n - \breve{\theta}_0) = O_p(1)$ *with* $\sqrt{n}(\theta_n - \breve{\theta}_0) = -\Sigma^{-1}\left[\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{\partial l_t(\breve{\theta}_0)}{\partial\theta}\right] + o_p(1),$

where $l_t(\theta)$ *is defined as in Assumption* 3*(ii).*

Proof. (i) By Lemma A1, it is not hard to show that

$$
\sup_{\Theta}\left\|\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left[\frac{\partial l_t(\theta)}{\partial\theta} - \frac{\partial \widetilde{l}_t(\theta)}{\partial\theta}\right]\right\| = o_p(1), \tag{A3}
$$

$$
\sup_{\Theta}\left\|\frac{1}{n}\sum_{t=1}^{n}\left[\frac{\partial^2 l_t(\theta)}{\partial\theta\partial\theta'} - \frac{\partial^2 \widetilde{l}_t(\theta)}{\partial\theta\partial\theta'}\right]\right\| = o_p(1). \tag{A4}
$$

Then, since $\partial\widetilde{l}_t(\theta_n)/\partial\theta = 0$, by Taylor's expansion and (A3)-(A4), we have

$$
\begin{aligned}
\theta_n - \breve{\theta}_0 &= -\left[\frac{1}{n}\sum_{t=1}^{n}\frac{\partial^2\widetilde{l}_t(\zeta_n)}{\partial\theta\partial\theta'}\right]^{-1}\left[\frac{1}{n}\sum_{t=1}^{n}\frac{\partial\widetilde{l}_t(\breve{\theta}_0)}{\partial\theta}\right] \\
&= -\left[\frac{1}{n}\sum_{t=1}^{n}\frac{\partial^2 l_t(\zeta_n)}{\partial\theta\partial\theta'}\right]^{-1}\left[\frac{1}{n}\sum_{t=1}^{n}\frac{\partial l_t(\breve{\theta}_0)}{\partial\theta}\right] + o_p(1),
\end{aligned} \tag{A5}
$$

where ζ_n lies between θ_n and $\breve{\theta}_0$. By Lemma A1(i) and Assumption 2, we know that

$$
E\sup_{\Theta}\left\|\frac{\partial^2 l_t(\theta)}{\partial\theta\partial\theta'}\right\| \leq CE\left(\xi_{\rho t-1}^2 + \xi_{\rho t}\xi_{\rho t-1}\right) < \infty
$$

for some $\rho \in (0, 1)$. Thus, by Theorem 3.1 in Ling and McAleer (2003), we have

$$
\frac{1}{n}\sum_{t=1}^{n}\frac{\partial^2 l_t(\zeta_n)}{\partial\theta\partial\theta'} = E\left[\frac{\partial^2 l_t(\zeta_n)}{\partial\theta\partial\theta'}\right] + o_p(1) = \Sigma + o_p(1), \tag{A6}
$$

where the last equality holds by the dominated convergence theorem and the fact that $\xi_n \to_p \breve{\theta}_0$ as $n \to \infty$ by Assumption 3. By (A5)-(A6) and the ergodic theorem, it follows that

$$\theta_n - \breve{\theta}_0 = -\Sigma^{-1}\left[\frac{1}{n}\sum_{t=1}^{n}\frac{\partial l_t(\breve{\theta}_0)}{\partial \theta}\right] + o_p(1) = -\Sigma^{-1}E\left[\frac{\partial l_t(\breve{\theta}_0)}{\partial \theta}\right] + o_p(1).$$

Since $\theta_n - \breve{\theta}_0 = o_p(1)$ by Assumption 3, it implies that (i) holds.

(ii) By (A3)-(A5), it is not hard to see that

$$\sqrt{n}(\theta_n - \breve{\theta}_0) = -\left[\frac{1}{n}\sum_{t=1}^{n}\frac{\partial^2 l_t(\zeta_n)}{\partial \theta \partial \theta'}\right]^{-1}\left[\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{\partial l_t(\breve{\theta}_0)}{\partial \theta}\right] + o_p(1).$$

Note that $\partial l_t(\breve{\theta}_0)/\partial\theta = 2\breve{\varepsilon}_t(\partial\breve{\varepsilon}_t/\partial\theta)$. Thus, by Assumptions 1 and 2(i), Lemmas 3-4 in Francq and Zakoïan (1998) implies that $n^{-1/2}\sum_{t=1}^{n}\partial l_t(\breve{\theta}_0)/\partial\theta = O_p(1)$. By (A6), it follows that (ii) holds. □

LEMMA A4. *Suppose that Assumptions* 1-3 *hold. Then,*

$$\|I_{2n}(\lambda) - A_n'(\lambda)[\sqrt{n}(\theta_n - \breve{\theta}_0)]\|^2 = o_p(1),$$

where

$$A_n(\lambda) = \sum_{j=1}^{n-1} E\left[\frac{\partial(\breve{\varepsilon}_t\breve{\varepsilon}_{t-j})}{\partial\theta}\right]\psi_j(\lambda).$$

Proof. By Taylor's expansion, we have $\widehat{\varepsilon}_t - \breve{\varepsilon}_t = (\partial\varepsilon_t(\zeta_n)/\partial\theta')(\theta_n - \breve{\theta}_0)$, where ζ_n lies between θ_n and $\breve{\theta}_0$. Then, it follows that

$$I_{2n}(\lambda) = \sum_{j=1}^{n-1}\left\{\frac{1}{n}\sum_{t=1+j}^{n}\left[\frac{\partial\varepsilon_t(\zeta_n)}{\partial\theta'}\widehat{\varepsilon}_{t-j} + \breve{\varepsilon}_t\frac{\partial\varepsilon_{t-j}(\zeta_n)}{\partial\theta'}\right]\psi_j(\lambda)\right\}[\sqrt{n}(\theta_n - \breve{\theta}_0)],$$

which entails

$$I_{2n}(\lambda) = \left\{I_{2n}^{(1)}(\lambda,\zeta_n,\theta_n) + I_{2n}^{(2)}(\lambda,\zeta_n) + I_{2n}^{(3)}(\lambda)\right\}[\sqrt{n}(\theta_n - \breve{\theta}_0)], \tag{A7}$$

where

$$I_{2n}^{(1)}(\lambda,\theta_1,\theta_2) = \sum_{j=1}^{n-1}\left\{\frac{1}{n}\sum_{t=1+j}^{n}\left[\frac{\partial\varepsilon_t(\theta_1)}{\partial\theta'}\varepsilon_{t-j}(\theta_2) - E\left(\frac{\partial\breve{\varepsilon}_t}{\partial\theta'}\breve{\varepsilon}_{t-j}\right)\right]\psi_j(\lambda)\right\},$$

$$I_{2n}^{(2)}(\lambda,\theta_1) = \sum_{j=1}^{n-1}\left\{\frac{1}{n}\sum_{t=1+j}^{n}\left[\breve{\varepsilon}_t\frac{\partial\varepsilon_{t-j}(\theta_1)}{\partial\theta'} - E\left(\breve{\varepsilon}_t\frac{\partial\breve{\varepsilon}_{t-j}}{\partial\theta'}\right)\right]\psi_j(\lambda)\right\},$$

$$I_{2n}^{(3)}(\lambda) = \sum_{j=1}^{n-1}\left\{\frac{n-j}{n}\left[E\left(\frac{\partial\breve{\varepsilon}_t}{\partial\theta'}\breve{\varepsilon}_{t-j}\right) + E\left(\breve{\varepsilon}_t\frac{\partial\breve{\varepsilon}_{t-j}}{\partial\theta'}\right)\right]\psi_j(\lambda)\right\}.$$

We first consider $I_{2n}^{(1)}(\lambda,\zeta_n,\theta_n)$. By direct calculation, we have

$$E\|I_{2n}^{(1)}(\lambda,\zeta_n,\theta_n)\|^2 = \sum_{j=1}^{n-1}(Ec_{nj}^2)P_j, \tag{A8}$$

where

$$c_{nj} = \frac{1}{n}\sum_{t=1+j}^{n}\left[\frac{\partial\varepsilon_t(\zeta_n)}{\partial\theta'}\varepsilon_{t-j}(\theta_n) - E\left(\frac{\partial\breve{\varepsilon}_t}{\partial\theta'}\breve{\varepsilon}_{t-j}\right)\right].$$

As for (A6), by Assumptions 2 and 3(i) and Lemma A1(i), we can show that uniformly in $j \in \{1, \cdots, n-1\}$, $Ec_{nj}^2 = o(1)$. Thus, since $\sum_{j=1}^{\infty} P_j < \infty$, by (A8), it is straightforward to see that

$$E\|I_{2n}^{(1)}(\lambda, \zeta_n, \theta_n)\|^2 = \sum_{j=1}^{n-1} o(P_j) = o(1),$$

which implies that $\|I_{2n}^{(1)}(\lambda, \zeta_n, \theta_n)\|^2 = o_p(1)$. Similarly, $\|I_{2n}^{(2)}(\lambda, \zeta_n)\|^2 = o_p(1)$.

Next, we consider $I_{2n}^{(3)}(\lambda)$. By direct calculation and the fact $P_j = O(j^{-2})$, we have

$$E\|I_{2n}^{(3)}(\lambda) - A_n(\lambda)\|^2 = \sum_{j=1}^{n-1} \frac{j^2}{n^2} \left[E\left(\frac{\partial \breve{\varepsilon}_t}{\partial \theta'} \breve{\varepsilon}_{t-j} \right) + E\left(\breve{\varepsilon}_t \frac{\partial \breve{\varepsilon}_{t-j}}{\partial \theta'} \right) \right]^2 P_j = O(n^{-1}).$$

Now, the conclusion follows from (A7) and Lemma A3(ii). □

LEMMA A5. *Suppose that Assumptions* 1-3 *hold. Then,*

$$\left\| \sum_{j=1}^{n-1} \left(\frac{1}{\sqrt{n}} \sum_{t=1}^{j} z_{tj} \right) \psi_j(\lambda) \right\|^2 = o_p(1),$$

where z_{tj} *is defined as in (3).*

Proof. First, by Lemma A3(i), we have $Ez_{tj} = 0$. Then, as for (A6), by Assumptions 1-2, it is not hard to show that

$$E\left[\frac{1}{j} \sum_{t=1}^{j} z_{tj} \right]^2 \to 0 \quad \text{as} \quad j \to \infty.$$

Thus, $\forall \varepsilon > 0$, there exists a $n_0(\varepsilon)$ such that when $j \geq n_0$,

$$E\left[\frac{1}{j} \sum_{t=1}^{j} z_{tj} \right]^2 < \varepsilon.$$

Next, by direct calculation, for $n \geq \max(n_0 + 1, \lfloor \varepsilon^{-1} \rfloor)$, we have

$$\begin{aligned}
& E\left\| \sum_{j=1}^{n-1} \left(\frac{1}{\sqrt{n}} \sum_{t=1}^{j} z_{tj} \right) \psi_j(\lambda) \right\|^2 \\
&= \frac{1}{n} \sum_{j=1}^{n-1} j^2 E\left[\frac{1}{j} \sum_{t=1}^{j} z_{tj} \right]^2 P_j \\
&= \frac{1}{n} \sum_{j=1}^{n_0-1} j^2 E\left[\frac{1}{j} \sum_{t=1}^{j} z_{tj} \right]^2 P_j + \frac{1}{n} \sum_{j=n_0}^{n-1} j^2 E\left[\frac{1}{j} \sum_{t=1}^{j} z_{tj} \right]^2 P_j \\
&\leq O\left(\frac{1}{n} \right) + \frac{\varepsilon}{n} \sum_{j=n_0}^{n-1} j^2 P_j \\
&= O\left(\frac{1}{n} \right) + O(\varepsilon) = O(\varepsilon).
\end{aligned}$$

Thus, it follows that the conclusion holds. □

PROOF OF THEOREM 1. By (A1) and Lemmas A2, A4 and A5, it suffices to show that $\overline{S}_n(\lambda) - E\{\overline{S}_n(\lambda)\} \Rightarrow S(\lambda)$ as $n \to \infty$, where $\overline{S}_n(\lambda) = \sum_{j=1}^{n-1} \sqrt{n}\overline{\lambda}(j)\psi_j(\lambda)$ with $\overline{\lambda}_n(\lambda) = n^{-1} \sum_{t=1+|j|}^{n} e_{t,j}$.

Here, we have used the fact that $E\{\breve{S}_n(\lambda)\} = E\{\overline{S}_n(\lambda)\}$ by Lemma A3(i). For each fixed integer $K \in \{1, \cdots, n-1\}$, we rewrite

$$\overline{S}_n(\lambda) = \sum_{j=1}^{K} \sqrt{n}\overline{\lambda}(j)\psi_j(\lambda) + \sum_{j=K+1}^{n-1} \sqrt{n}\overline{\lambda}(j)\psi_j(\lambda) =: \overline{S}_n^K(\lambda) + R_n^K(\lambda).$$

Then, as in Shao (2011a), the conclusion holds from the following three claims:

(a). For any $h \in L_2[0, \pi]$, the finite dimensional distributions of $\langle \overline{S}_n^K - E(\overline{S}_n^K), h\rangle$ converge to those of $\langle S^K(\lambda), h\rangle$, where $S^K(\lambda)$ is a Gaussian process with zero mean and asymptotic projected variances

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

(b). The sequence $\{\overline{S}_n^K(\lambda)\}$ is tight.
(c). For $\forall\varepsilon > 0$, $\lim_{K\to\infty}\lim_{n\to\infty} P\left(\|R_n^K(\lambda) - E\{R_n^K(\lambda)\}\| > \varepsilon\right) = 0.$ Q.E.D.

PROOF OF CLAIM (a). By a direct calculation, we can show that

$$\begin{aligned}
\langle \overline{S}_n^K - E(\overline{S}_n^K), h\rangle &= \frac{1}{\sqrt{n}}\sum_{j=1}^{K}\sum_{t=j+1}^{n}\{e_{t,j} - E(e_{t,j})\}W_h(j) \\
&= \frac{1}{\sqrt{n}}\sum_{t=2}^{K+1}\sum_{j=1}^{t-1}\{e_{t,j} - E(e_{t,j})\}W_h(j) \\
&\quad + \frac{1}{\sqrt{n}}\sum_{t=K+2}^{n}\sum_{j=1}^{K}\{e_{t,j} - E(e_{t,j})\}W_h(j), \qquad \text{(A9)}
\end{aligned}$$

where the first summand above is $o_p(1)$ since K is finite. Rewrite

$$\begin{aligned}
Y_t := \sum_{j=1}^{K} e_{t,j}W_h(j) &= 1'_{K+1} \times \left(\breve{\varepsilon}_t\breve{\varepsilon}_{t-1}W_h(1), \cdots, \breve{\varepsilon}_t\breve{\varepsilon}_{t-K}W_h(K), \kappa\breve{\varepsilon}_t\frac{\partial\breve{\varepsilon}_t}{\partial\theta'}\right)' \\
&=: 1'_{K+1} \times v_t, \qquad \text{(A10)}
\end{aligned}$$

where $1_{K+1} = (1, \cdots, 1)' \in \mathcal{R}^{(K+1)\times 1}$ and $\kappa = -2\sum_{j=1}^{K} E\left[\partial(\breve{\varepsilon}_t\breve{\varepsilon}_{t-j})/\partial\theta'\right]W_h(j)$. By the finiteness of $W_h(j)$ and κ and the same argument as in Francq, Roy, and Zakoïan (2005, page 543), we have

$$\frac{1}{\sqrt{n}}\sum_{t=K+2}^{n}(v_t - Ev_t) \to_d N(0, \Omega^*) \text{ as } n \to \infty, \text{ where } \Omega^* = \lim_{n\to\infty} var\left[\frac{1}{\sqrt{n}}\sum_{t=K+2}^{n} v_t\right] < \infty.$$

Hence, it follows that for the second summand, $n^{-1/2}\sum_{t=K+2}^{n}(Y_t - EY_t) \to_d N(0, \check{I})$ as $n \to \infty$, where

$$\begin{aligned}
\check{I} &= \lim_{n\to\infty} var\left(\frac{1}{\sqrt{n}}\sum_{t=K+2}^{n} Y_t\right) \\
&= \lim_{n\to\infty}\frac{1}{n}\sum_{j=1}^{K}\sum_{k=1}^{K}\left(\sum_{t=K+2}^{n}\sum_{t'=K+2}^{n} cov(e_{t,j}, e_{t',k})\right) W_h(j)W_h(k) \\
&= \lim_{n\to\infty}\frac{1}{n}\sum_{j=1}^{K}\sum_{k=1}^{K}\left(\sum_{d=K+2-n}^{n-K-2}\sum_{t=K+2+\max(0,d)}^{n+\min(0,d)} cov(e_{t,j}, e_{t-d,k})\right) W_h(j)W_h(k) \\
&= \lim_{n\to\infty}\sum_{j=1}^{K}\sum_{k=1}^{K}\left(\sum_{d=K+2-n}^{n-K-2}\frac{n-K-2-|d|}{n} cov(e_{t,j}, e_{t-d,k})\right) W_h(j)W_h(k) \\
&= \sigma^2_{h,K}. \qquad \text{(A11)}
\end{aligned}$$

Thus, it follows that claim (a) holds. Q.E.D.

PROOF OF CLAIM (b). First, as for (A9), we have

$$\begin{aligned}
\overline{S}_n^K - E(\overline{S}_n^K) &= \frac{1}{\sqrt{n}}\sum_{j=1}^{K}\sum_{t=j+1}^{n}\{e_{t,j} - E(e_{t,j})\}\psi_j(\lambda) \\
&= \frac{1}{\sqrt{n}}\sum_{t=2}^{K+1}\sum_{j=1}^{t-1}\{e_{t,j} - E(e_{t,j})\}\psi_j(\lambda) + \frac{1}{\sqrt{n}}\sum_{t=K+2}^{n} G_t^K, \qquad \text{(A12)}
\end{aligned}$$

where the first term in (A12) is tight since each summand is tight, and

$$G_t^K = \sum_{j=1}^{K}\{e_{t,j} - E(e_{t,j})\}\psi_j(\lambda).$$

Next, we use Theorem 2.1 in Politis and Romano (1994) to prove the tightness of the second term in (A12). Note that G_t^K is independent of n. We only need to verify that

(i) $E\|G_t^K\|^2 < \infty$;

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

(iii) $\lim_{n\to\infty} var\left[\langle \bar{S}_n^K - E(\bar{S}_n^K), h\rangle\right] = \sigma^2_{h,K}$.

The proof of (i) is trivial, and the proof of (iii) is directly from the one as for (A11). We now consider the proof of (ii). Note that

$$\sum_{t=K+2}^{\infty}\left|E\left[\langle G_{K+2}^K, G_t^K\rangle\right]\right| = \sum_{t=K+2}^{\infty}\left|\sum_{j=1}^{K} cov(e_{t,j}, e_{K+2,j})P_j\right|. \qquad \text{(A13)}$$

Using the same argument as for Lemma 3 in Francq and Zakoïan (1998), it is not hard to show that for each $j \in \{1, \cdots, K\}$, there exists a $\rho \in (0,1)$ such that

$$|cov(e_{t,j}, e_{K+2,j})| \le C\left\{\rho^{|t-K-2|/2} + \left[\alpha_y\left(\left\lfloor\frac{|t-K-2|}{2}\right\rfloor\right)\right]^{\nu/(2+\nu)}\right\}. \qquad \text{(A14)}$$

By (A13)-(A14), it follows that

$$\sum_{t=K+2}^{\infty} \left| E\left[\langle G_{K+2}^K, G_t^K \rangle\right] \right| \le C \left(\sum_{j=1}^{K} P_j\right) \sum_{s=0}^{\infty} \left\{ \rho^{|s|/2} + \left[\alpha_y\left(\left\lfloor \frac{|s|}{2} \right\rfloor\right)\right]^{\nu/(2+\nu)} \right\} < \infty,$$

which implies that (ii) holds. This completes the proof of claim (b). Q.E.D.

PROOF OF CLAIM (c). First, by direct calculation, we have

$$E\|R_n^K(\lambda) - E\{R_n^K(\lambda)\}\|^2 = \frac{1}{n} \sum_{j=K+1}^{n-1} \sum_{t,t'=j+1}^{n} cov(e_{t,j}, e_{t',j}) P_j. \tag{A15}$$

Since $e_{t,j} = \breve{\varepsilon}_t \breve{\varepsilon}_{t-j} + z_{tj}$, there are four terms in $cov(e_{t,j}, e_{t',j})$. For simplicity, we only prove the conclusion for the term $cov(z_{tj}, z_{t'j})$, since the proofs for other terms are similar. Note that for any $m \in \{1, \cdots, p+q\}$, the m-th entry of z_{tj} satisfies that

$$z_{tj,m} = O(1)\breve{\varepsilon}_t \frac{\partial \varepsilon_t(\breve{\theta}_0)}{\partial \theta_m} = O(1)\left[\sum_{i=0}^{\infty} c_i y_{t-i}\right]\left[\sum_{k=0}^{\infty} c_{k,m} y_{t-k}\right], \tag{A16}$$

where $c_i = O(\rho^i)$ and $c_{i,m} = O(\rho^i)$ for some $\rho \in (0,1)$. Then, for any $(m, m') \in \{1, \cdots, p+q\}^2$, we have

$$\begin{aligned}
&\frac{1}{n} \sum_{j=K+1}^{n-1} \sum_{t,t'=j+1}^{n} cov(z_{tj,m}, z_{t'j,m'}) \\
&\le O\left(\frac{1}{n}\right) \sum_{j=K+1}^{n-1} \sum_{t,t'=j+1}^{n} \sum_{i,k,i',k' \ge 0} |c_i c_{k,m'} c_{i'} c_{k',m'}|\, |cov(y_{t-i} y_{t-k}, y_{t'-i'} y_{t'-k'})| P_j \\
&\le O(1) \sum_{i,k,i',k' \ge 0} |c_i c_{k,m'} c_{i'} c_{k',m'}| \sum_{j=K+1}^{n-1} \left\{ \frac{1}{n} \sum_{t,t'=j+1}^{n} |cov(y_0 y_{i-k}, y_{t'-t+i-i'} y_{t'-t+i-k'})| \right\} P_j.
\end{aligned}$$

Furthermore, by Assumption 2, we can show that for any i, k, i', k', j,

$$\begin{aligned}
&\frac{1}{n} \sum_{t,t'=j+1}^{n} |cov(y_0 y_{i-k}, y_{t'-t+i-i'} y_{t'-t+i-k'})| \\
&\le \frac{1}{n} \sum_{t,t'=j+1}^{n} \{|cum(y_0, y_{i-k}, y_{t'-t+i-i'}, y_{t'-t+i-k'})| \\
&\quad + |\gamma(t'-t+i-i')\gamma(t'-t+k-k')| + |\gamma(t'-t+i-k')\gamma(t'-t+k-i')|\} \\
&\le \sum_{d=-(n-1-j)}^{n-1-j} \frac{n-1-j-|d|}{n} \{|cum(y_0, y_{i-k}, y_{d+i-i'}, y_{d+i-k'})| \\
&\quad + |\gamma(d+i-i')\gamma(d+k-k')| + |\gamma(d+i-k')\gamma(d+k-i')|\} \\
&\le \sum_{s_1,s_2,s_3=-\infty}^{\infty} |cum(y_0, y_{s_1}, y_{s_2}, y_{s_3})| + 2\sum_{s=-\infty}^{\infty} [\gamma(s)]^2 < \infty.
\end{aligned}$$

Thus, it follows that

$$\frac{1}{n} \sum_{j=K+1}^{n-1} \sum_{t,t'=j+1}^{n} cov(z_{tj,m}, z_{t'j,m'}) \le O(1) \sum_{j=K+1}^{\infty} P_j \to 0 \text{ as } K \to \infty. \tag{A17}$$

By (A15) and (A17), we know that $\lim_{K\to\infty}\lim_{n\to\infty} E\|R_n^K(\lambda) - E\{R_n^K(\lambda)\}\|^2 = 0$. Now, claim (c) follows directly from Chebyshev's inequality. Q.E.D.

PROOF OF COROLLARY 1. Under H_0, we have $\theta_0 = \breve{\theta}_0$ and $\gamma(j) = 0$ for $j \geq 1$. Then, it is straightforward to see that $\|E\{\breve{S}_n(\lambda)\}\| = o(1)$. Thus, (i) follows directly from continuous mapping theorem. For (ii), since $n^{-1/2}\widetilde{S}_n(\lambda) - E\left\{n^{-1/2}\breve{S}_n(\lambda)\right\} \Rightarrow 0$ in $L_2[0,\pi]$ by Theorem 1, it follows that

$$
\begin{aligned}
\frac{\widetilde{\mathrm{CM}}_n}{n} &= \left\| E\left[\frac{\breve{S}_n(\lambda)}{\sqrt{n}}\right]\right\| + \left\|\frac{\widetilde{S}_n(\lambda)}{\sqrt{n}}\right\| - \left\| E\left[\frac{\breve{S}_n(\lambda)}{\sqrt{n}}\right]\right\| \\
&= \sum_{j=1}^{\infty} \left[E(\breve{\varepsilon}_t\breve{\varepsilon}_{t-j})\right]^2 P_j + o(1),
\end{aligned}
$$

which entails that (ii) holds. Q.E.D.

PROOF OF COROLLARY 2. Rewrite

$$
\begin{aligned}
\widetilde{S}_n(\lambda) &= \overline{S}_n(\lambda) + \left[\widetilde{S}_n(\lambda) - \overline{S}_n(\lambda)\right] \\
&= \left[\overline{S}_n(\lambda) - E\left\{\overline{S}_n(\lambda)\right\}\right] + E\left\{\overline{S}_n(\lambda)\right\} + \left[\widetilde{S}_n(\lambda) - \overline{S}_n(\lambda)\right]. \qquad \text{(A18)}
\end{aligned}
$$

On one hand, by Assumptions 1-3, from the proof of Theorem 1, we have

$$
\overline{S}_n(\lambda) - E\left\{\overline{S}_n(\lambda)\right\} \Rightarrow S(\lambda) \ \text{ and } \ E\|\widetilde{S}_n(\lambda) - \overline{S}_n(\lambda)\|^2 \to 0 \quad \text{as } n\to\infty. \qquad \text{(A19)}
$$

On the other hand, since $\breve{\theta}_0 = \theta_0$ by Assumption 4, we can show that under H_{1n},

$$
\begin{aligned}
E\left\{\overline{S}_n(\lambda)\right\} &= E\left\{\breve{S}_n(\lambda)\right\} \\
&= E\left[\sum_{j=1}^{n-1}\sqrt{n}\widehat{\gamma}(j)\psi_j(\lambda)\right] \\
&= \sum_{j=1}^{n-1}\sqrt{n}\gamma_n(j)\psi_j(\lambda) \\
&= \sum_{j=1}^{n-1}\frac{\gamma(0)}{2\pi}\int_{-\pi}^{\pi}\left[g(\omega)e^{ij\omega}d\omega\right]\psi_j(\lambda) \to \frac{\gamma(0)}{2\pi}\int_0^{\lambda} g(\omega)d\omega \qquad \text{(A20)}
\end{aligned}
$$

as $n\to\infty$, where $\gamma_n(j)$ is defined as in (5). Now, the conclusion holds from (A18)-(A20) and continuous mapping theorem. Q.E.D.

Next, in order to prove Theorem 2, we need three more lemmas:

LEMMA A6. *Assume that Assumptions* 1-3 *hold and* $b_n^{-1} = o(1)$. *Then,*

(*i*) $\|\theta_n^* - \breve{\theta}_0\| = o_p^*(1)$;

(*ii*) $\sqrt{n}(\theta_n^* - \breve{\theta}_0) = O_p^*(1)$ *with* $\sqrt{n}(\theta_n^* - \breve{\theta}_0) = -\Sigma^{-1}\left[n^{-1/2}\sum_{t=1}^{n} w_t^*\frac{\partial l_t(\breve{\theta}_0)}{\partial\theta}\right] + o_p^*(1)$.

Proof. By the definition of θ_n^*, we have

$$\begin{aligned}\theta_n^* - \breve{\theta}_0 &= -\left[\frac{1}{n}\sum_{t=1}^{n}\frac{\partial^2 \tilde{l}_t^*(\zeta_n)}{\partial\theta\partial\theta'}\right]^{-1}\left[\frac{1}{n}\sum_{t=1}^{n}\frac{\partial \tilde{l}_t^*(\breve{\theta}_0)}{\partial\theta}\right] \\ &= -\left[\frac{1}{n}\sum_{t=1}^{n}w_t^*\frac{\partial^2 \tilde{l}_t(\zeta_n)}{\partial\theta\partial\theta'}\right]^{-1}\left[\frac{1}{n}\sum_{t=1}^{n}(w_t^*-1)\frac{\partial \tilde{l}_t(\breve{\theta}_0)}{\partial\theta} + \frac{1}{n}\sum_{t=1}^{n}\frac{\partial \tilde{l}_t(\breve{\theta}_0)}{\partial\theta}\right] \\ &=: -\left[s_{1n}\right]^{-1}\left[s_{2n}+s_{3n}\right],\end{aligned}$$

where ζ_n lies between θ_n^* and $\breve{\theta}_0$. First, by Lemma A1, it is straightforward to see that

$$E^*\|s_{1n}\| \le \frac{1}{n}\sum_{t=1}^{n}E^*(w_t^*)\sup_{\Theta}\left\|\frac{\partial^2\tilde{l}_t(\theta)}{\partial\theta\partial\theta'}\right\| = \frac{1}{n}\sum_{t=1}^{n}\sup_{\Theta}\left\|\frac{\partial^2 l_t(\theta)}{\partial\theta\partial\theta'}\right\| + o_p(1) = O_p(1),$$

which entails $s_{1n} = O_p^*(1)$. Next, by direct calculation and Lemma A1, we have

$$\begin{aligned}E^*\left[s_{2n}s_{2n}'\right] &= \frac{1}{n^2}\sum_{s=1}^{L_n}\left[\sum_{t,t'\in B_s}\frac{\partial\tilde{l}_t(\breve{\theta}_0)}{\partial\theta}\frac{\partial\tilde{l}_{t'}(\breve{\theta}_0)}{\partial\theta'}\right] \\ &= \frac{1}{n^2}\sum_{s=1}^{L_n}\left[\sum_{t,t'\in B_s}\frac{\partial l_t(\breve{\theta}_0)}{\partial\theta}\frac{\partial l_{t'}(\breve{\theta}_0)}{\partial\theta'}\right] + o_p(1) =: s_{4n} + o_p(1), \qquad \text{(A21)}\end{aligned}$$

where B_s is defined in step 1 of the BRW procedure in Section 3. Moreover, since $l_t(\theta)$ is stationary and $E\partial l_t(\breve{\theta}_0)/\partial\theta = 0$ by Lemma A3(i), it is straightforward to see that

$$E\left(s_{4n}\right) = \frac{b_n}{n^2}\sum_{s=1}^{L_n} var\left[\frac{1}{\sqrt{b_n}}\sum_{t\in B_s}\frac{\partial l_t(\breve{\theta}_0)}{\partial\theta}\right] = \frac{b_n L_n}{n^2}var\left[\frac{1}{\sqrt{b_n}}\sum_{t=1}^{b_n}\frac{\partial l_t(\breve{\theta}_0)}{\partial\theta}\right] = O\left(\frac{1}{n}\right), \qquad \text{(A22)}$$

where the last equality holds due to the fact that $b_n^{-1} = o(1)$ and (9). Then, by (A21)-(A22), we have $s_{2n} = O_p^*(n^{-1/2})$. Note that $s_{3n} = o_p(1)$ by the ergodic theorem and Lemmas A1 and A3(i). Thus, it follows that (i) holds. Consequently, by Lemma A1, it is not hard to show that

$$\frac{1}{n}\sum_{t=1}^{n}(w_t^*-1)\frac{\partial^2\tilde{l}_t(\zeta_n)}{\partial\theta\partial\theta'} = o_p^*(1) \quad \text{and} \quad \frac{1}{n}\sum_{t=1}^{n}\frac{\partial^2\tilde{l}_t(\zeta_n)}{\partial\theta\partial\theta'} = \Sigma + o_p^*(1),$$

and hence $s_{1n} = \Sigma + o_p^*(1)$. Note that $\sqrt{n}s_{2n} = O_p^*(1)$ and $\sqrt{n}s_{3n} = O_p(1)$ by Lemma A3(ii). Thus, it follows that (ii) holds. □

LEMMA A7. *Assume that Assumptions* 1-3 *hold,* $b_n^{-1} = o(1)$, *and* $b_n n^{-1} = o(1)$. *Then,* $E^*\|\widetilde{Z}_n(\gamma) - \overline{Z}_n(\gamma)\|^2 = o_p(1)$, *where* $\widetilde{Z}_n(\gamma)$ *is defined in (6), and*

$$\overline{Z}_n(\gamma) = \sum_{j=1}^{n-1}\left[\frac{1}{\sqrt{n}}\sum_{t=1+j}^{n}(w_t^*-1)E(e_{t,j})\right]\psi_j(\lambda).$$

Proof. Note that

$$E^*\|\widetilde{Z}_n(\gamma) - \overline{Z}_n(\gamma)\|^2 \le 2E^*\|\widetilde{Z}_n(\gamma) - \breve{Z}_n(\gamma)\|^2 + 2E^*\|\breve{Z}_n(\gamma) - \overline{Z}_n(\gamma)\|^2, \qquad \text{(A23)}$$

where

$$\breve{Z}_n(\gamma) = \sum_{j=1}^{n-1}\left[\frac{1}{\sqrt{n}}\sum_{t=1+j}^{n}\frac{(w_t^*-1)(n-j)}{n}E\left(\breve{\varepsilon}_t\breve{\varepsilon}_{t-j}\right)\right]\psi_j(\lambda).$$

By direct calculation, we have

$$\begin{aligned} E^*\|\widetilde{Z}_n(\gamma) - \breve{Z}_n(\gamma)\|^2 &= \sum_{j=1}^{n-1} \left\{ \frac{1}{n} E^* \left[\sum_{t=1+j}^{n} (w_t^* - 1) d_{nj} \right]^2 \right\} P_j \\ &= \sum_{j=1}^{n-1} \left\{ \frac{1}{n} \sum_{s=1}^{L_n} \left[\sum_{t \in B_s \cap [1+j, n]} d_{nj} \right]^2 \right\} P_j \\ &\le \sum_{j=1}^{n-1} \left\{ \frac{L_n b_n^2}{n} d_{nj}^2 \right\} P_j \\ &= \frac{b_n}{n} \sum_{j=1}^{n-1} \left(\sqrt{n} d_{nj} \right)^2 P_j, \end{aligned} \tag{A24}$$

where $d_{nj} = n^{-1} \sum_{t'=1+j}^{n} [\widetilde{\varepsilon}_{t'} \widetilde{\varepsilon}_{t'-j} - E(\breve{\varepsilon}_{t'} \breve{\varepsilon}_{t'-j})]$. By Lemma A1, it is straightforward to see that

$$\sqrt{n} d_{nj} = \frac{1}{\sqrt{n}} \sum_{t=1+j}^{n} [\check{\varepsilon}_t \check{\varepsilon}_{t-j} - E(\breve{\varepsilon}_t \breve{\varepsilon}_{t-j})] + o_p(1). \tag{A25}$$

Next, by Taylor's expansion, we have

$$\check{\varepsilon}_t \check{\varepsilon}_{t-j} = \breve{\varepsilon}_t \breve{\varepsilon}_{t-j} + \frac{\partial(\breve{\varepsilon}_t \breve{\varepsilon}_{t-j})}{\partial \theta'} (\theta_n - \breve{\theta}_0) + (\theta_n - \breve{\theta}_0)' \left[\frac{1}{2} \frac{\partial^2 (\varepsilon_t(\theta) \varepsilon_{t-j}(\theta))}{\partial \theta \partial \theta'} |_{\theta = \zeta_n} \right] (\theta_n - \breve{\theta}_0),$$

where ζ_n lies between θ_n and $\breve{\theta}_0$. Note that $\sqrt{n}(\theta_n - \breve{\theta}_0) = O_p(1)$ by Lemma A3(ii). Thus, by (A25) it follows that for all $j \in \{1, \cdots, n-1\}$,

$$\begin{aligned} \sqrt{n} d_{nj} &= \frac{1}{\sqrt{n}} \sum_{t=1+j}^{n} [\breve{\varepsilon}_t \breve{\varepsilon}_{t-j} - E(\breve{\varepsilon}_t \breve{\varepsilon}_{t-j})] \\ &\quad + \frac{1}{n} \sum_{t=1+j}^{n} \frac{\partial(\breve{\varepsilon}_t \breve{\varepsilon}_{t-j})}{\partial \theta'} [\sqrt{n}(\theta_n - \breve{\theta}_0)] + o_p(1) \\ &= \frac{1}{\sqrt{n}} \sum_{t=1+j}^{n} [\breve{\varepsilon}_t \breve{\varepsilon}_{t-j} - E(\breve{\varepsilon}_t \breve{\varepsilon}_{t-j})] + O_p(1). \end{aligned} \tag{A26}$$

As for (A17), we can show that for all $j \in \{1, \cdots, n-1\}$,

$$E \left\{ \frac{1}{\sqrt{n}} \sum_{t=1+j}^{n} [\breve{\varepsilon}_t \breve{\varepsilon}_{t-j} - E(\breve{\varepsilon}_t \breve{\varepsilon}_{t-j})] \right\}^2 = \frac{1}{n} \sum_{t,t'=1+j}^{n} cov(\breve{\varepsilon}_t \breve{\varepsilon}_{t-j}, \breve{\varepsilon}_{t'} \breve{\varepsilon}_{t'-j}) = O(1).$$

Thus, by (A26), we know that $\sqrt{n} d_{nj} = O_p(1)$ for all j. Since $b_n n^{-1} = o(1)$ and $\sum_{j=1}^{\infty} P_j < \infty$, by (A24), it entails that

$$E^*\|\widetilde{Z}_n(\gamma) - \breve{Z}_n(\gamma)\|^2 = \frac{b_n}{n} \sum_{j=1}^{n-1} O_p(P_j) = o_p(1). \tag{A27}$$

Next, since $E(e_{t,j}) = E(\breve{\varepsilon}_t\breve{\varepsilon}_{t-j})$ and $b_n n^{-1} = o(1)$, it is straightforward to see that

$$\begin{aligned} E^*\|\breve{Z}_n(\gamma) - \overline{Z}_n(\gamma)\|^2 &= E^*\left\|\sum_{j=1}^{n-1}\left[\frac{j}{n^{3/2}}\sum_{t=1+j}^{n}(w_t^*-1)E\left(\breve{\varepsilon}_t\breve{\varepsilon}_{t-j}\right)\right]\psi_j(\lambda)\right\|^2 \\ &= \sum_{j=1}^{n-1}\frac{j^2}{n^3}E^*\left[\sum_{t=1+j}^{n}(w_t^*-1)E\left(\breve{\varepsilon}_t\breve{\varepsilon}_{t-j}\right)\right]^2 P_j \\ &= \sum_{j=1}^{n-1}\frac{j^2}{n^3}\sum_{s=1}^{L_n}\left[\sum_{t\in B_s\cap[1+j,n]}E\left(\breve{\varepsilon}_t\breve{\varepsilon}_{t-j}\right)\right]^2 P_j \\ &\le \sum_{j=1}^{n-1}\frac{j^2}{n^3}L_n b_n^2 P_j \\ &= O\left(b_n n^{-1}\right) = o(1). \end{aligned} \tag{A28}$$

Now, the conclusion follows directly from (A23) and (A27)-(A28). □

LEMMA A8. *Suppose that Assumptions* 1-3 *hold,* $b_n^{-1} = o(1)$*, and* $(\log n)b_n n^{-1} = o(1)$*. Then,*

$$E^*\left\|\sum_{j=1}^{n-1}\left[\frac{1}{\sqrt{n}}\sum_{t=1}^{j}(w_t^*-1)\widetilde{z}_{tj}\right]\psi_j(\lambda)\right\|^2 = o_p(1),$$

where $\widetilde{z}_{tj}$ *is defined in the same way as* z_{tj} *in (3) with* $\widetilde{l}_t(\breve{\theta}_0)$ *replacing* $l_t(\breve{\theta}_0)$*.*

Proof. By direct calculation, we have

$$\begin{aligned} E^*\left\|\sum_{j=1}^{n-1}\left[\frac{1}{\sqrt{n}}\sum_{t=1}^{j}(w_t^*-1)\widetilde{z}_{tj}\right]\psi_j(\lambda)\right\|^2 &= \sum_{j=1}^{n-1}\frac{1}{n}E^*\left(\sum_{t=1}^{j}(w_t^*-1)\widetilde{z}_{tj}\right)^2 P_j \\ &= \sum_{j=1}^{n-1}\frac{1}{n}\sum_{s=1}^{L_n}\left(\sum_{t\in B_s\cap[1,j]}\widetilde{z}_{tj}\right)^2 P_j. \end{aligned}$$

By Lemma A1, it is straightforward to see that

$$\sum_{j=1}^{n-1}\frac{1}{n}\sum_{s=1}^{L_n}\left(\sum_{t\in B_s\cap[1,j]}\widetilde{z}_{tj}\right)^2 P_j = \sum_{j=1}^{n-1}\frac{1}{n}\sum_{s=1}^{L_n}\left(\sum_{t\in B_s\cap[1,j]}z_{tj}\right)^2 P_j + o_p(1) =: H_n + o_p(1).$$

Note that $\sum_{j=1}^{\infty} P_j < \infty$. For $\forall\varepsilon > 0$, there exists a $j_0(\varepsilon) > 0$ such that

$$\sum_{j=j_0+1}^{\infty} P_j < \varepsilon.$$

Since $b_n \to \infty$ as $n \to \infty$, we rewrite

$$
\begin{aligned}
H_n &= \sum_{j=1}^{j_0} \frac{1}{n} \sum_{s=1}^{L_n} \left(\sum_{t \in B_s \cap [1,j]} z_{tj} \right)^2 P_j + \sum_{j=j_0+1}^{b_n} \frac{1}{n} \sum_{s=1}^{L_n} \left(\sum_{t \in B_s \cap [1,j]} z_{tj} \right)^2 P_j \\
&\quad + \sum_{j=b_n+1}^{n-1} \frac{1}{n} \sum_{s=1}^{L_n} \left(\sum_{t \in B_s \cap [1,j]} z_{tj} \right)^2 P_j \\
&=: H_{1n} + H_{2n} + H_{3n}. \qquad \text{(A29)}
\end{aligned}
$$

First, for H_{1n}, we know that as n is large enough,

$$
EH_{1n} \le \sum_{j=1}^{j_0} \frac{1}{n} \sum_{s=1}^{L_n} O(j_0^2) P_j = O\left(\frac{L_n}{n} \right) < \varepsilon. \qquad \text{(A30)}
$$

Next, for H_{2n}, direct calculation gives us that

$$
H_{2n} = \sum_{j=j_0+1}^{b_n} \frac{1}{n} \sum_{s=1}^{1} \left(\sum_{t \in B_s \cap [1,j]} z_{tj} \right)^2 P_j = \sum_{j=j_0+1}^{b_n} \frac{1}{n} \left(\sum_{t \in B_1} z_{tj} \right)^2 P_j.
$$

By Lemma 3 in Francq and Zakoïan (1998), it follows that as n is large enough,

$$
EH_{2n} = \sum_{j=j_0+1}^{b_n} \frac{b_n}{n} E\left(\frac{1}{\sqrt{b_n}} \sum_{t=1}^{b_n} z_{tj} \right)^2 P_j = \sum_{j=j_0+1}^{b_n} \frac{b_n}{n} O(P_j) \le O\left(\frac{b_n}{n} \varepsilon \right) < \varepsilon. \qquad \text{(A31)}
$$

Third, for H_{3n}, we truncate it as

$$
\begin{aligned}
H_{3n} &= \frac{1}{n} \sum_{s'=2}^{L_n} \sum_{j \in B_{s'}} \sum_{s=1}^{L_n} \left(\sum_{t \in B_s \cap [1,j]} z_{tj} \right)^2 P_j \\
&= \frac{1}{n} \sum_{s'=2}^{L_n} \sum_{j \in B_{s'}} \left(\sum_{s<s'} + \sum_{s=s'} \right) \left(\sum_{t \in B_s \cap [1,j]} z_{tj} \right)^2 P_j, \qquad \text{(A32)}
\end{aligned}
$$

where the summand in the case of $s > s'$ is zero since $B_s \cap [1,j] = \emptyset$. As for (A31), by the stationarity of z_{tj}, we can show that

$$
\begin{aligned}
E\left[\frac{1}{n} \sum_{s'=2}^{L_n} \sum_{j \in B_{s'}} \sum_{s<s'} \left(\sum_{t \in B_s \cap [1,j]} z_{tj} \right)^2 P_j \right] &= \frac{1}{n} \sum_{s'=2}^{L_n} \sum_{j \in B_{s'}} \sum_{s<s'} E\left(\sum_{t \in B_s} z_{tj} \right)^2 P_j \\
&= \frac{b_n}{n} \sum_{s'=2}^{L_n} \sum_{j \in B_{s'}} \sum_{s<s'} O(P_j) \\
&\le \frac{b_n L_n}{n} \sum_{s'=2}^{L_n} \sum_{j \in B_{s'}} O(P_j) \\
&\le \sum_{j=j_0+1}^{\infty} O(P_j) < \varepsilon. \qquad \text{(A33)}
\end{aligned}
$$

Furthermore, since $(\log n) b_n n^{-1} = o(1)$, it is not hard to see that

$$\begin{aligned}
E\left[\frac{1}{n}\sum_{s'=2}^{L_n}\sum_{j\in B_{s'}}\sum_{s=s'}\left(\sum_{t\in B_s\cap[1,j]} z_{tj}\right)^2 P_j\right] &= \frac{1}{n}\sum_{s'=2}^{L_n}\sum_{j\in B_{s'}} O(b_n^2)P_j \\
&= \frac{1}{n}\sum_{j=b_n+1}^{n-1} O(b_n^2)\frac{1}{j^2} \\
&\le \frac{b_n}{n}\sum_{j=b_n+1}^{n-1} O(1)\frac{1}{j} \\
&= O\left(\frac{b_n \log n}{n}\right) < \varepsilon. \qquad \text{(A34)}
\end{aligned}$$

Now, the conclusion follows from (A29)-(A34). □

PROOF OF THEOREM 2. By Taylor's expansion we have

$$\widetilde{\varepsilon}_t^*\widetilde{\varepsilon}_{t-j}^* = \widetilde{\varepsilon}_t\widetilde{\varepsilon}_{t-j} + \frac{\partial(\widetilde{\varepsilon}_t\widetilde{\varepsilon}_{t-j})}{\partial\theta'}(\theta_n^* - \theta_n) + (\theta_n^* - \theta_n)'\left[\frac{1}{2}\frac{\partial^2(\widetilde{\varepsilon}_t(\theta)\widetilde{\varepsilon}_{t-j}(\theta))}{\partial\theta\partial\theta'}|_{\theta=\zeta_n}\right](\theta_n^* - \theta_n),$$

where ζ_n lies between θ_n^* and θ_n. Then, it follows that

$$\begin{aligned}
\widetilde{S}_n^*(\lambda) - \widetilde{S}_n(\lambda) = &\sum_{j=1}^{n-1}\frac{1}{\sqrt{n}}\left[\sum_{t=1+j}^{n}(w_t^* - 1)\widetilde{\varepsilon}_t\widetilde{\varepsilon}_{t-j}\right]\psi_j(\lambda) + I_{1n}^*(\lambda)[\sqrt{n}(\theta_n^* - \theta_n)] \\
&+ [\sqrt{n}(\theta_n^* - \theta_n)']I_{2n}^*(\lambda)[\sqrt{n}(\theta_n^* - \theta_n)], \qquad \text{(A35)}
\end{aligned}$$

where

$$\begin{aligned}
I_{1n}^*(\lambda) &= \sum_{j=1}^{n-1}\frac{1}{n}\sum_{t=1+j}^{n} w_t^*\frac{\partial(\widetilde{\varepsilon}_t\widetilde{\varepsilon}_{t-j})}{\partial\theta'}\psi_j(\lambda), \\
I_{2n}^*(\lambda) &= \sum_{j=1}^{n-1}\frac{1}{n^{3/2}}\sum_{t=1+j}^{n} w_t^*\left[\frac{1}{2}\frac{\partial^2(\widetilde{\varepsilon}_t(\theta)\widetilde{\varepsilon}_{t-j}(\theta))}{\partial\theta\partial\theta'}|_{\theta=\zeta_n}\right]\psi_j(\lambda).
\end{aligned}$$

By Lemma A4, we can easily show that

$$E^*\left\|I_{1n}^*(\lambda) - \sum_{j=1}^{n-1}E\left[\frac{\partial(\breve{\varepsilon}_t\breve{\varepsilon}_{t-j})}{\partial\theta'}\right]\psi_j(\lambda)\right\|^2 = O_p\left(b_n n^{-1}\right). \qquad \text{(A36)}$$

On the other hand, it is straightforward to see that

$$E^*\left\|I_{2n}^*(\lambda)\right\|^2 = O_p(n^{-1}). \qquad \text{(A37)}$$

Since $\sqrt{n}(\theta_n^* - \theta_n) = O_p^*(1)$ by Lemma A3(ii) and Lemma A6(ii), under (A35)-(A37), we have

$$\begin{aligned}
\widetilde{S}_n^*(\lambda) - \widetilde{S}_n(\lambda) = &\sum_{j=1}^{n-1}\frac{1}{\sqrt{n}}\left[\sum_{t=1+j}^{n}(w_t^* - 1)\widetilde{\varepsilon}_t\widetilde{\varepsilon}_{t-j}\right]\psi_j(\lambda) \\
&+ \left\{\sum_{j=1}^{n-1}E\left[\frac{\partial(\breve{\varepsilon}_t\breve{\varepsilon}_{t-j})}{\partial\theta'}\right]\psi_j(\lambda)\right\}[\sqrt{n}(\theta_n^* - \theta_n)] + \text{negligible terms}. \qquad \text{(A38)}
\end{aligned}$$

Moreover, by Lemma A3(ii), Lemma A6(ii) and (A3), we have

$$\begin{aligned}\sqrt{n}(\theta_n^* - \theta_n) &= -\Sigma^{-1}\left[\frac{1}{\sqrt{n}}\sum_{t=1}^{n}(w_t^* - 1)\frac{\partial l_t(\breve{\theta}_0)}{\partial\theta}\right] + o_p^*(1)\\ &= -\Sigma^{-1}\left[\frac{1}{\sqrt{n}}\sum_{t=1}^{n}(w_t^* - 1)\frac{\partial \widetilde{l}_t(\breve{\theta}_0)}{\partial\theta}\right] + o_p^*(1). \qquad \text{(A39)}\end{aligned}$$

By (A38)-(A39) and Lemma A8, it follows that

$$\widetilde{S}_n^*(\lambda) - \widetilde{S}_n(\lambda) = \sum_{j=1}^{n-1}\frac{1}{\sqrt{n}}\left[\sum_{t=1+j}^{n}(w_t^* - 1)\widetilde{e}_{t,j}\right]\psi_j(\lambda) + \text{negligible terms}, \qquad \text{(A40)}$$

where $\widetilde{e}_{t,j} = \widetilde{\varepsilon}_t\widetilde{\varepsilon}_{t-j} + \widetilde{z}_{tj}$ and $\widetilde{z}_{tj}$ is defined as in Lemma A8.

Let $\breve{\gamma}^*(j) = n^{-1}\{\sum_{t=1+j}^{n}(w_t^* - 1)\left[\widetilde{e}_{t,j} - E(e_{t,j})\right]\}$. By (A40) and Lemma A7, it follows that

$$\Delta_n(\lambda) = \sqrt{n}\sum_{j=1}^{n-1}\breve{\gamma}^*(j)\psi_j(\lambda) + \text{negligible terms} =: \breve{S}_n^*(\lambda) + \text{negligible terms}.$$

Finally, for each fixed integer $K \in \{1, \cdots, n-1\}$, we rewrite

$$\breve{S}_n^*(\lambda) = \sqrt{n}\sum_{j=1}^{K}\breve{\gamma}^*(j)\psi_j(\lambda) + \sqrt{n}\sum_{j=K+1}^{n-1}\breve{\gamma}^*(j)\psi_j(\lambda) =: \breve{S}_n^{K*}(\lambda) + \breve{R}_n^{K*}(\lambda).$$

Then, as in Shao (2011a), the conclusion holds from the following three claims:

(d). For any $h \in L_2[0,\pi]$, the finite dimensional distributions of $\langle \breve{S}_n^{K*}, h\rangle$ converge to those of $\langle S^K(\lambda), h\rangle$ in probability conditional on χ_n.

(e). For $\forall\varepsilon > 0$, $\lim_{K\to\infty}\lim_{n\to\infty} P^*\left(\|\breve{R}_n^{K*}(\lambda)\| > \varepsilon\right) = 0$ in probability conditional on χ_n.

(f). The sequence $\{\breve{S}_n^*(\lambda)\}$ is tight in probability conditional on χ_n.

The proofs of claims (e) and (f) are similar to these of part (a,ii) and part (b) in Shao (2011a, p.222). Thus, we only need to prove claim (d). Q.E.D.

PROOF OF CLAIM (d). Let $G_t^{K*} = \sum_{j=1}^{K}(w_t^* - 1)\left[\widetilde{e}_{t,j} - E(e_{t,j})\right]\psi_j(\lambda)$. As for (A9), it suffices to show the asymptotic normality of J_n^{K*}, where

$$\begin{aligned}J_n^{K*} = \sum_{t=K+2}^{n}\frac{1}{\sqrt{n}}\langle G_t^{K*}, h\rangle &= \sum_{t=K+2}^{n}\frac{1}{\sqrt{n}}\sum_{j=1}^{K}(w_t^* - 1)\left[\widetilde{e}_{t,j} - E(e_{t,j})\right]W_h(j)\\ &= \sum_{s=1}^{L_n}\frac{\delta_s - 1}{\sqrt{n}}\sum_{t\in B_s\cap[K+2,n]}\sum_{j=1}^{K}\left[\widetilde{e}_{t,j} - E(e_{t,j})\right]W_h(j)\\ &=: \sum_{s=1}^{L_n} H_{sn}^*.\end{aligned}$$

Note that conditional on χ_n, $\{H_{sn}^*\}$ is a sequence of independent random variables. Thus, we only need to verify that

$$(i)\ \lim_{n\to\infty} var^*\left(J_n^{K*}\right) \to_p \sigma_{h,K}^2;$$

$$(ii)\ \lim_{n\to\infty}\sum_{s=1}^{L_n} E^*\left\{|H_{sn}^*|^2 I(|H_{sn}^*| > \varepsilon)\right\} \to_p 0.$$

Without loss of generality, we assume that $K+2 \leq b_n$. For (i), by Lemma A1, Taylor's expansion, and Lemma A3(ii), it is not hard to show that

$$
\begin{aligned}
var^*\left(J_n^{K*}\right) &= \frac{1}{n}\sum_{s=1}^{L_n}\left\{\sum_{t\in B_s\cap[K+2,n]}\sum_{j=1}^{K}\left[\widetilde{e}_{t,j}-E(e_{t,j})\right]W_h(j)\right\}^2 \\
&= \frac{1}{L_n}\sum_{s=2}^{L_n}\left\{\frac{1}{\sqrt{b_n}}\sum_{t\in B_s}\sum_{j=1}^{K}\left[\breve{e}_{t,j}-E(e_{t,j})\right]W_h(j)\right\}^2 + o_p(1) \\
&= \frac{1}{L_n}\sum_{s=2}^{L_n}\left\{\frac{1}{\sqrt{b_n}}\sum_{t\in B_s}\sum_{j=1}^{K}\left[e_{t,j}-E(e_{t,j})\right]W_h(j)\right\}^2 + O_p\left(\frac{b_n}{n}\right) + o_p(1) \\
&=: Z_n + o_p(1),
\end{aligned}
$$

where $\breve{e}_{t,j} = \breve{\varepsilon}_t\breve{\varepsilon}_{t-j} + z_{tj}$. As for (A11), we have $EZ_n \to \sigma_{h,k}^2$ as $n \to \infty$. Thus, we only need to prove that $var(Z_n) \to 0$ as $n \to \infty$. By direct calculation, we have

$$
\begin{aligned}
var(Z_n) &= \frac{1}{n^2}\sum_{s,s'=1}^{L_n}\sum_{t_1,t_2\in B_s}\sum_{t_1',t_2'\in B_{s'}}\sum_{j_1,j_2=1}^{K}\sum_{j_1',j_2'=1}^{K} C(t_1,t_2,t_1',t_2',j_1,j_2,j_1',j_2') \\
&\quad \times W_h(j_1)W_h(j_2)W_h(j_1')W_h(j_2') \\
&=: \frac{1}{n^2}\sum_{s,s'=1}^{L_n} z(s,s'),
\end{aligned}
$$

where $\gamma_e(j) = E(e_{t,j})$ and $C(t_1,t_2,t_1',t_2',j_1,j_2,j_1',j_2')$ equals to

$$
cov\left\{\left[\left(e_{t_1,j_1}-\gamma_e(j_1)\right)\left(e_{t_2,j_2}-\gamma_e(j_2)\right)\right],\left[\left(e_{t_1',j_1'}-\gamma_e(j_1')\right)\left(e_{t_2',j_2'}-\gamma_e(j_2')\right)\right]\right\}.
$$

Rewrite

$$
var(Z_n) = \frac{1}{n^2}\sum_{s=1}^{L_n}\sum_{|s'-s|\leq 1} z(s,s') + \frac{1}{n^2}\sum_{s=1}^{L_n}\sum_{|s'-s|>1} z(s,s'). \tag{A41}
$$

Fort the first summand in (A41), since $b_n = o(n^{1/3})$, it is straightforward to see that

$$
\frac{1}{n^2}\sum_{s=1}^{L_n}\sum_{|s'-s|\leq 1} z(s,s') = O\left(\frac{L_n b_n^4}{n^2}\right) = O\left(\frac{b_n^3}{n}\right) = o(1). \tag{A42}
$$

Next, for the second summand in (A41), $C(t_1,t_2,t_1',t_2',j_1,j_2,j_1',j_2')$ can be divided into 16 terms, since $e_{t,j} = \breve{\varepsilon}_t\breve{\varepsilon}_{t-j} + z_{tj}$. We only consider the proof for the term $cov\left(z_{t_1j_1}z_{t_2j_2}, z_{t_1'j_1'}z_{t_2'j_2'}\right)$, because the proofs for other terms are similar. In view of (A16), for any $(m_1, m_2, m_1', m_2') \in \{1, p+q\}^4$, we have

$$
\begin{aligned}
&\left|cov\left[z_{t_1 j_1, m_1} z_{t_2 j_2, m_2}, z_{t_1' j_1', m_1'} z_{t_2' j_2', m_2'}\right]\right| \\
&= \left| \sum_{i_1, k_1, i_2, k_2, i_1', k_1', i_2', k_2'} c_{i_1} c_{k_1, m_1} c_{i_2} c_{k_2, m_2} c_{i_1'} c_{k_1', m_1'} c_{i_2'} c_{k_2', m_2'} M(i_1, k_1, i_2, k_2, i_1', k_1', i_2', k_2') \right| \\
&\leq \Bigg[\sum_{i_1 > b_n/4} + \sum_{k_1 > b_n/4} + \sum_{i_2 > b_n/4} + \sum_{k_2 > b_n/4} + \sum_{i_1' > b_n/4} + \sum_{k_1' > b_n/4} + \sum_{i_2' > b_n/4} + \sum_{k_2' > b_n/4} \\
&\quad \sum_{i_1, k_1, i_2, k_2, i_1', k_1', i_2', k_2' \leq b_n/4} \Bigg] \left| c_{i_1} c_{k_1, m_1} c_{i_2} c_{k_2, m_2} c_{i_1'} c_{k_1', m_1'} c_{i_2'} c_{k_2', m_2'} M(i_1, k_1, i_2, k_2, i_1', k_1', i_2', k_2') \right| \\
&=: \sum_{i=1}^{9} g_i,
\end{aligned}
$$

where $M(i_1, k_1, i_2, k_2, i_1', k_1', i_2', k_2') = cov\left(y_{t_1 - i_1} y_{t_1 - k_1} y_{t_2 - i_2} y_{t_2 - k_2}, y_{t_1' - i_1'} y_{t_1' - k_1'} y_{t_2' - i_2'} y_{t_2' - k_2'}\right)$. By Cauchy-Schwarz inequality, we can show that

$$
\begin{aligned}
|M(i_1, k_1, i_2, k_2, i_1', k_1', i_2', k_2')| &\leq \sqrt{E\left(y_{t_1 - i_1} y_{t_1 - k_1} y_{t_2 - i_2} y_{t_2 - k_2}\right)^2 E\left(y_{t_1' - i_1'} y_{t_1' - k_1'} y_{t_2' - i_2'} y_{t_2' - k_2'}\right)^2} \\
&\leq E y_t^8 < \infty.
\end{aligned}
$$

Since $c_i = O(\rho^i)$ and $c_{i,m} = O(\rho^i)$ for some $\rho \in (0, 1)$, it is straightforward to see that

$$g_i \leq C \rho^{b_n/4}, \quad \text{for } 1 \leq i \leq 8.$$

Furthermore, the Davydov's inequality in Davydov (1968) implies that

$$
\begin{aligned}
g_9 &\leq C \sum_{i_1, k_1, i_2, k_2, i_1', k_1', i_2', k_2' \leq b_n/4} \| y_{t_1 - i_1} y_{t_1 - k_1} y_{t_2 - i_2} y_{t_2 - k_2} \|_{2+\nu} \| y_{t_1' - i_1'} y_{t_1' - k_1'} y_{t_2' - i_2'} y_{t_2' - k_2'} \|_{2+\nu} \\
&\quad \times \left[\alpha_y\left(\left\lfloor \frac{b_n}{2} \right\rfloor\right)\right]^{\nu/(2+\nu)} \left| c_{i_1} c_{k_1, m_1} c_{i_2} c_{k_2, m_2} c_{i_1'} c_{k_1', m_1'} c_{i_2'} c_{k_2', m_2'} \right| \\
&\leq C \left(E y_t^{8+4\nu}\right) \left[\alpha_y\left(\left\lfloor \frac{b_n}{2} \right\rfloor\right)\right]^{\nu/(2+\nu)} \\
&\quad \times \sum_{i_1, k_1, i_2, k_2, i_1', k_1', i_2', k_2' \leq b_n/4} \left| c_{i_1} c_{k_1, m_1} c_{i_2} c_{k_2, m_2} c_{i_1'} c_{k_1', m_1'} c_{i_2'} c_{k_2', m_2'} \right| \\
&\leq C \left[\alpha_y\left(\left\lfloor \frac{b_n}{2} \right\rfloor\right)\right]^{\nu/(2+\nu)}.
\end{aligned}
$$

Therefore, since $\lim_{k \to \infty} k^2 [\alpha_y(k)]^{\nu/(2+\nu)} = 0$, it follows that

$$\frac{1}{n^2} \sum_{s=1}^{L_n} \sum_{|s'-s|>1} z(s, s') \leq O\left(\frac{L_n^2 b_n^4}{n^2}\right) \left[\rho^{b_n/4} + \left[\alpha_y\left(\left\lfloor \frac{b_n}{2} \right\rfloor\right)\right]^{\nu/(2+\nu)}\right] = o(1). \quad \text{(A43)}$$

By (A41)-(A43), we know that (i) holds.

For (ii), since $E(w_t^*)^4 < \infty$, by Holder's inequality and the fact that $b_n = o(n^{1/3})$, we have

$$\begin{aligned}\sum_{s=1}^{L_n} E\left\{E^*\left[|H_{sn}^*|^2 I(|H_{sn}^*| > \varepsilon)\right]\right\} &\le C\sum_{s=1}^{L_n} E\left(E^*|H_{sn}^*|^4\right)\\ &= O\left(\frac{1}{n^2}\right)\sum_{s=1}^{L_n} E\left\{\sum_{t\in B_s}\sum_{j=1}^{K}[e_{t,j} - E(e_{t,j})]\right\}^4\\ &= O\left(\frac{L_n b_n^4}{n^2}\right) = o(1),\end{aligned}$$

i.e., (ii) holds. This completes the proof of claim (d). Q.E.D.

REFERENCES

[1] BOX, G.E.P. & PIERCE, D.A. (1970). Distribution of the residual autocorrelations in autoregressive integrated moving average time series models. *Journal of the American Statistical Association* **65**, 1509–1526.

[2] BROCKWELL, P.J. & DAVIS, R.A. (1991). Time Series: Theory and Methods. (2nd ed.). Springer-Verlag, New York.

[3] BRÜGGEMANN, R., JENTSCH, C. & TRENKLER, C. (2014). Inference in VARs with conditional heteroskedasticity of unknown form. Working paper. University of Mannheim.

[4] CARRASCO, M. & CHEN, X. (2002). Mixing and moment properties of various GARCH and stochastic volatility models. *Econometric Theory* **18**, 17–39.

[5] CHEN, K., YING, Z., ZHANG, H. & ZHAO, L. (2008). Analysis of least absolute deviation. *Biometrika* **95**, 107–122.

[6] CHEN, K., GUO, S., LIN, Y. & YING, Z. (2010). Least absolute relative error estimation. *Journal of the American Statistical Association* **105**, 1104–1112.

[7] CHEN, S., MIN, W. & CHEN, R. (2013). Model identification for time series with dependent innovations. *Statistica Sinica* **23**, 873–899.

[8] DAVYDOV, Y.A. (1968). Convergence of distributions generated by stationary stochastic processes. *Theory of Probability & Its Applications* **13**, 691–696.

[9] DELGADO, M.A., HIDALGO, J. & VELASCO, C. (2005). Distribution free goodness-of-fit tests for linear processes. *Annals of Statistics* **33**, 2568–2609.

[10] DELGADO, M.A. & VELASCO, C. (2011). An asymptotically pivotal transform of the residuals sample autocorrelations with application to model checking. *Journal of the American Statistical Association* **106**, 946–958.

[11] DEO, R.S. (2000). Spectral tests of the martingale hypothesis under conditional heteroskedasticity. *Journal of Econometrics* **99**, 291–315.

[12] DOUKHAN, P. & LEÓN, J.R. (1989). Cumulants for stationary mixing random equences and applications to empirical spectral density. *Probability and Mathematical Statistics* **10**, 11–26.

[13] DURLAUF, S.N. (1991). Spectral-based testing of the martingale hypothesis. *Journal of Econometrics* **50**, 355–376.

[14] ESCANCIANO, J.C. (2006). Goodness-of-fit tests for linear and non-linear time series models. *Journal of the American Statistical Association* **101**, 531–541.

[15] ESCANCIANO, J.C. (2007). Model checks using residual marked emprirical processes. *Statistica Sinica* **17**, 115–138.

[16] ESCANCIANO, J.C. and LOBATO, I.N. (2009). An automatic Portmanteau test for serial correlation. *Journal of Econometrics* **151**, 140–149.

[17] ESCANCIANO, J.C., LOBATO, I.N. & ZHU, L. (2013). Automatic specification testing for vector autoregressions and multivariate nonlinear time series models. *Journal of Business & Economic Statistics* **31**, 426–437.

[18] ESCANCIANO, J.C. & VELASCO, C. (2006). Generalized spectral tests for the martingale difference hypothesis. *Journal of Econometrics* **134**, 151–185

[19] EUBANK, R.L. & LARICCIA, V.N. (1992). Asymptotic comparison of Cramer-von Mises and nonparametric function estimation techniques for testing goodness-of-fit. *Annals of Statistics* **20**, 2071–2086.

[20] FAN, J. & JIANG, J. (2007). Nonpatamteric inference with generalized likelihood ratio tests (with discussion). *Test* **16**, 409–478.

[21] FAN, J. & ZHANG, W. (2004). Generalised likelihood ratio tests for spectral density. *Biometrika* **91**, 195–209.

[22] FRANCQ, C., ROY, R. & ZAKOÏAN, J.M. (2005). Diagnostic checking in ARMA models with uncorrelated errors. *Journal of the American Statistical Association* **100**, 532–544.

[23] Francq, C. & Zakoïan, J.M. (1998). Estimating linear representations of nonlinear processes. *Journal of Statistical Planning and Inference* **68**, 145–165.
[24] Francq, C. & Zakoïan, J.M. (2010). *GARCH Models: Structure, Statistical Inference and Financial Applications*. Wiley, Chichester, UK.
[25] Franses, P.H. & Van Dijk, R. (1996). Forecasting stock market volatility using (non-linear) Garch models. *Journal of Forecasting* **15**, 229–235.
[26] Ghosh, B.K. & Huang, W.-M. (1991). The power and optimal kernel of the Bickel-Rosenblatt test for goodness of fit. *Annals of Statistics* **19**, 999–1009.
[27] Granger, C.W.J. & Anderson, A.P. (1978). *An introduction to Bilinear Time Series Models*. Vandenhoek and Ruprecht, Gottinger.
[28] Hall, P., Horowitz, J.L. & Jing, B.-Y. (1995). On blocking rules for the bootstrap with dependent data. *Biometrica* **82**, 561–574.
[29] Hamilton, J.D. (1994). *Time Series Analysis*. Princeton University Press.
[30] Hong, Y. (1996). Consistent testing for serial correlation of unknown form. *Econometrica* **64**, 837–864.
[31] 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.
[32] Jentsch, C., Politis, D.N. & Paparoditis, E. (2014). Block bootstrap theory for multivariate integrated and cointegrated processes. Forthcoming in *Journal of Time Series Analysis*.
[33] Jin, Z., Ying, Z. & Wei, L.J. (2001). A simple resampling method by perturbing the minimand. *Biometrika* **88**, 381–390.
[34] Ling, S. (2007). Self-weighted and local quasi-maximum likelihood estimators for ARMA-GARCH/IGARCH models. *Journal of Econometrics* **140**, 849–873.
[35] Ling, S. & McAleer, M. (2003). Asymptotic theory for a new vector ARMA-GARCH model. *Econometric Theory* **19**, 280–310.
[36] Liu, R.Y. (1988). Bootstrap procedure under some non-i.i.d. models. *Annals of Statistics* **16**, 1696–1708.
[37] Ljung, G.M. & Box, G.E.P. (1978). On a measure of lack of fit in time series models. *Biometrika* **65**, 297–303.
[38] Lobato, I.N. (2001). Testing that a dependent process is uncorrelated. *Journal of the American Statistical Association* **96**, 1066–1076.
[39] Paparoditis, E. (2000). Spectral density based goodness-of-fit tests in time series models. *Scandinavian Journal of Statistics* **27**, 143–176.
[40] Paparoditis, E. (2001). On some power properties of goodness of fit tests in time series, in Probability and Statistical Models with Applications, Ed. by Charalambides/Koutras/Balakrishnan, Chapman and Hall, p.333–348.
[41] Paparoditis, E. & Politis, D.N. (2003). Residual-based block bootstrap for unit root testing. *Econometrica* **71**, 813–855.
[42] Parker, C., Paparoditis, E. & Politis, D.N. (2006). Unit root testing via the stationary bootstrap. *Journal of Econometrics* **133**, 601–638.
[43] Parthasarathy, K.R. (1967). *Probability measures on metric spaces*. Academic Press. New York.
[44] Pham, D.T. (1986). The mixing property of bilinear and generalized random coefficient autoregressive models. *Stochastic Processes and Their Applications* **23**, 291–300.
[45] Politis, D.N. & Romano, J. (1994). Limit theorems for weakly dependent Hilbert space valued random variables with application to the stationary boostrap. *Statistica Sinica* **4**, 461–476.
[46] Romano, J.L. & Thombs, L.A. (1996). Inference for autocorrelations under weak assumptions. *Journal of the American Statistical Association* **91**, 590–600.
[47] Rosenblatt, M. (1975). A quadratic measure of deviation of two dimensional density estimates and a test of independence. *Annals of Statistics* **3**, 1–14.
[48] Rosenblatt, M. (1985). Stationary Sequences and Random Fields. Birkhäuser, Boston.
[49] Politis, D.N., Romano, J. & Wolf, M. (1999). *Subsampling*. Springer-Verla, New York.
[50] Shao, X. (2011a). A bootstrap-assisted spectral test of white noise under unknown dependence. *Journal of Econometrics* **162**, 213–224.
[51] Shao, X. (2011b). Testing for white noise under unknown dependence and its applications to diagnostic checking for time series models. *Econometric Theory* **27**, 312–343.
[52] Shao, X. & Wu, W.B. (2007). Asymptotic spectral theory for nonlinear time series. *Annals of Statistics* **35**, 1773–1801.
[53] Tsay, R.S. (2005). *Analysis of financial time series (2nd ed.)*. New York: John Wiley & Sons, Incorporated.
[54] Tsay, R.S. & Tiao, G.C. (1984). Consistent estimates of autoregressive parameters and extended sample autocorrelation function for stationary and nonstationary ARMA models. *Journal of the American Statistical Association* **79**, 84–96.
[55] Wu, C.F.J. (1986). Jacknife, bootstrap and other resampling methods in regression analysis (with discussion). *Annals of Statistics* **14**, 1261–1350.

[56] WU, W.B. (2005). Nonlinear system theory: another look at dependence. *Proceedings of National Academy of Science, USA* **102**, 14150–14154.

[57] WU, W.B. & SHAO, X. (2004). Limit theorems for iterated random functions. *Journal of Applied Probability* **41**, 425–436.

[58] ZHU, K. & LING, S. (2012). The global weighted LAD estimators for finite/infinite variance ARMA(p, q) models. *Econometric Theory* **28**, 1065–1086.