

# Orthogonal samples for estimators in time series

Suhasini Subba Rao

Department of Statistics, Texas A&M University

College Station, TX, U.S.A.

`suhasini@stat.tamu.edu`

May 10, 2017

## Abstract

Inference for statistics of a stationary time series often involve nuisance parameters and sampling distributions that are difficult to estimate. In this paper, we propose the method of orthogonal samples, which can be used to address some of these issues. For a broad class of statistics, an orthogonal sample is constructed through a slight modification of the original statistic, such that it shares similar distributional properties as the centralised statistic of interest. We use the orthogonal sample to estimate nuisance parameters of weighted average periodogram estimators and  $L_2$ -type spectral statistics. Further, the orthogonal sample is utilized to estimate the finite sampling distribution of various test statistics under the null hypothesis. The proposed method is simple and computationally fast to implement. The viability of the method is illustrated with various simulations.

*Keywords* Nuisance parameters, orthogonal transformations, statistical tests, time series.

## 1 Introduction

In classical statistics, given the correct distribution it is often possible to define estimators and pivotal quantities which do not depend on any nuisance parameters, examples include the studentized t-statistic and log-likelihood ratio statistic. In time series analysis, due to dependence in the data and that the underlying distribution of the process is usually unknown, such statistics rarely arise. However, for inference it is necessary to estimate the variance of the statistic which can often involve a complicated function of higher order cumulants. For the past 30 years, the standard approach to the estimation of nuisance

parameters and finite sample distributions of statistics is to use the bootstrap. This is a simple method for mimicing the behaviour of the time series. There exists many methods for constructing the bootstrap. Classical examples include the block-type bootstrap (see Künsch [1989], Politis and Romano [1994], Romano and Thombs [1996], Politis et al. [1999], Lahiri [2003], Kirch and Politis [2011] and Kreiss and Lahiri [2012]), fixed-b bootstrap, which accounts for the influence of bandwidth in the block bootstrap (see Kiefer and Vogelsang [2005] and Shao and Politis [2013]), sieve bootstrap (see Kreiss [1992], Kreiss et al. [2010] and Jentsch and Politis [2013]) frequency domain bootstrap (Hurvich and Zeger [1987], Franke and Härdle [1992] and Dahlhaus and Janas [1996]) linear process bootstrap (McMurray and Politis [2010] and Jentsch and Politis [2015]) and the recently introduced moving average bootstrap (see Krampe et al. [2016]). An alternative, is the method of self-normalisation proposed in Lobato [2001] and Shao [2009], where the limiting distribution is non-standard but free of nuisance parameters.

The purpose of this paper is to propose an alternative approach to the estimation of nuisance parameters, which leads to pivotal statistics. We make no claims that the proposed method is better than any of the excellent methods mentioned above, but we believe it is based on some interesting properties which have not previously been explored. In many respects our approach is very classical. It is motivated by Fisher's definition of an ancillary variable from the 1930s and by the innovations in spectral analysis for time series developed during the 1950's and 1960's. An ancillary variable, is a statistic whose sampling distribution does not depend on the parameters of interest yet holds important information about the statistic of interest. For example, if  $\{X_i\}_{i=1}^n$  are iid random variables with mean  $\mu$  and variance  $\sigma^2$  and  $\bar{X}$  is the sample mean, then  $X_i - \bar{X}$  can be considered as an ancillary variable since its sampling distribution does not change with  $\mu$ , however, since  $\text{var}[X_i - \bar{X}] = (n-1)\sigma^2/n$  it does contain information about the variance  $\sigma^2$ . Thus the ancillary variables,  $\{X_i - \bar{X}\}_{i=1}^n$ , are used to estimate the variance of the sample mean. Ancillary variables rarely occur in time series analysis, however, our aim is to show that several estimators give rise to *asymptotic* ancillary variables, which can be used to estimate the variance of the estimator of interest and construct pivotal quantities. Since the asymptotic ancillary variables constructed in this paper are uncorrelated to each other in this paper we call them an *orthogonal sample*.

To illustrate the proposed method we consider a well known example in time series, where implicitly the notion of an orthogonal sample is used. Let  $\{X_t\}$  be a stationary, short memory, time series with mean  $\mu$ , autocovariance function  $\{c(j)\}$  and spectral density function  $f(\omega) = \frac{1}{2\pi} \sum_{r \in \mathbb{Z}} c(j) \exp(ij\omega)$  and  $i = \sqrt{-1}$ . We observe  $\{X_t\}_{t=1}^T$  and use the sample mean  $\bar{X}_T$  as the estimator of the mean. It is well know that the variance of the sample mean is asymptotically equal to the long run variance  $\text{var}[\sqrt{T}\bar{X}_T] \approx \sum_{j \in \mathbb{Z}} c(j)$ . We recall that

$\sum_{j \in \mathbb{Z}} c(j) = 2\pi f(0)$ , thus estimation of the long run variance is equivalent to estimating the spectral density function at frequency zero. A classical estimator of the spectral density is the local average of the periodogram (see Bartlett [1950] and Parzen [1957]). Applying this to long run variance estimation, this means using  $2\pi \hat{f}_M(0)$  as an estimator of  $2\pi f(0)$ , where

$$\hat{f}_M(0) = \frac{2\pi}{M} \sum_{k=1}^M |J_T(\omega_k)|^2 \text{ with } J_T(\omega_k) = \frac{1}{\sqrt{2\pi T}} \sum_{t=1}^T X_t \exp(it\omega_k) \text{ and } \omega_k = \frac{2\pi k}{T}.$$

We now take a step back and consider this estimator from a slightly different perspective, which fits with the notion of an ancillary variable, discussed above. We observe (i)  $\sqrt{2\pi} J_T(0) = \sqrt{T} \bar{X}$  (ii)  $E[J_T(\omega_k)] = 0$  if  $1 \leq k < T/2$ ,  $\text{cov}[J_T(\omega_{k_1}), J_T(\omega_{k_2})] = O(T^{-1})$  (if  $k_1 \neq k_2$ ) and (iii) if  $\omega_k$  is in a “neighbourhood” of zero  $\text{var}[J_T(\omega_k)] \approx f(0)$ . Thus  $M \ll T$   $\{J_T(\omega_k)\}_{k=1}^M$  can be considered as an orthogonal sample to the sample mean  $\sqrt{2\pi} J_T(0)$ ; it contains no mean information but shares the same (asymptotic) variance as the sample mean. Furthermore, if the random vector  $\{J_T(\omega_k)\}_{k=0}^M$  is asymptotically normal (cf. Brockwell and Davis [1991] in the case of linear stationary time series), then for fixed  $M$  we have the asymptotically pivotal statistic

$$\mathcal{T} = \frac{\sqrt{T}(\bar{X} - \mu)}{\sqrt{\frac{1}{M} \sum_{k=1}^M |J_T(\omega_k)|^2}} \xrightarrow{D} t_{2M}, \text{ as } T \rightarrow \infty \quad (1)$$

where  $t_{2M}$  denotes a t-distribution with  $2M$  degrees of freedom. This approximation has been extensively studied in Müller [2007], Sun [2011], Sun [2013], Sun [2014] and Müller and Watson [2016]. In particular, Sun [2014] derives an expression for the error between the finite sample distribution of  $\mathcal{T}$  in (1) and the  $t$ -distribution. Since many statistics, such as the sample covariance or parameter estimators, have a “mean-like” form the approximation in (1) can be applied to a wide class of statistics. Sun [2013] shows that the approximation in (1), where  $M$  is kept fixed but  $T \rightarrow \infty$ , is related to the fixed- $b$  approximations considered in Kiefer and Vogelsang [2005] and Sun et al. [2008]. In Kiefer and Vogelsang [2005] and Sun et al. [2008], the denominator in (1) is replaced with the “time domain” estimator of the long run variance based on weighting the sample covariances

$$\tilde{f}_b(0) = \sum_{s=-(T-1)}^{T-1} \lambda \left( \frac{s}{bT} \right) \hat{c}_T(s) \quad (2)$$

where  $\hat{c}_T(s) = T^{-1} \sum_{t=1}^{T-|s|} (X_t - \bar{X}_T)(X_{t+|s|} - \bar{X}_T)$  and  $\lambda$  is a lag window. Analogous to the result in (1), in the fixed- $b$  framework, where  $b$  is kept fixed and  $T \rightarrow \infty$ , Kiefer and Vogelsang [2005] show that the resulting t-statistic is an asymptotically pivotal, non-standard distribution that depends only on the lag window  $\lambda$ . Comparing  $\hat{f}_M$  with  $\tilde{f}_b$  we see that  $M$  and  $b^{-1}$  play much the same role. However, as pointed out by Sun [2013], a distinct advantage

of using  $\widehat{f}_M$  instead of  $\widetilde{f}_b$  is that the limiting pivotal distribution is a standard  $t$ -distribution, which makes its application far simpler than using a non-standard pivotal distribution.

The main focus in the literature reviewed above is on constructing  $t$ -statistics for “mean-like” estimators. By using a hybrid of fixed  $b$  and small  $b$  asymptotics (analogous to fixed  $M$  or letting  $M \rightarrow \infty$ ), the above authors show that the pivotal distribution approximations are more accurate at approximating the distribution of the  $t$ -statistic than the normal distribution. However, several useful statistics that are considered in time series, such as quadratic forms, do not have a “mean-like” representation. For quadratic forms, there does not exist a general method for estimating its variance. In this case, the result in (1) cannot be directly applied to such statistics.

Our objective in this paper is to consider statistics which commonly arise in time series, that may not have a “mean-like” representation. We obtain the orthogonal samples for such estimators, which we use to estimate the nuisance parameters associated to the statistic of interest. The critical insight in the construction of an orthogonal sample is that many statistics can be rewritten in terms of sums and products of Discrete Fourier transform. Using this formulation and well known properties of DFTs for stationary time series, the orthogonal samples associated with several statistics can easily be constructed. We use this result to quantify the uncertainty when constructing an estimator based on the orthogonal sample and use this to construct statistics which have a simple pivotal distribution when  $M$  is kept fixed and  $T \rightarrow \infty$  (so called fixed  $M$ -asymptotics).

Our main focus is the class of quadratic forms, which can be written as a weighted average periodogram

$$A_T(\phi) = \frac{1}{T} \sum_{k=1}^T \phi(\omega_k) |J_T(\omega_k)|^2. \quad (3)$$

This broad class of statistics was first introduced in Parzen [1957] and includes the sample autocovariance function, spectral density estimators and Whittle likelihood estimators. Furthermore, several well known  $L_2$ -function can be written as  $L_2$  functions of  $A_T(\phi)$ , such as the Ljung-Box test and  $L_2$ -statistics based on the spectral density function. In Section 2 we define the orthogonal sample associated with  $A_T(\phi)$  and show that asymptotically it has the same moments as a centralised version of  $A_T(\phi)$ . Using the orthogonal sample we obtain an estimator of the variance of  $A_T(\phi)$  and define an asymptotically pivotal statistic analogous to (1). Often a statistical test is formulated in terms of a  $L_2$ -statistic. But to apply the test, the mean and variance of the  $L_2$ -statistics under the null hypothesis are required. These are often very complicated and involve unknown nuisance parameters. As mentioned above, the orthogonal sample can be used to estimate the centralized moments of a statistic. In Section 2.4 we utilize this idea to estimate the mean and variance of  $L_2$ -spectral statistics and construct an asymptotically pivotal statistic.

In Section 3 we address the issue of testing. Specifically, since the orthogonal sample shares similar sampling properties with the centralised version of the statistic, we use it to estimate the finite sample distribution, and critical values, of the statistic under the null that the mean of the statistic is zero. In Section 4 we propose an average square criterion to select the number of terms in the orthogonal sample. Evaluation of the orthogonal sample requires only  $O(T \log T)$  computing operations, which makes the procedure extremely fast.

The purpose of this paper is to propose a methodology without rigorous proof. However, some proofs are given in the supplementary material.

## 2 Orthogonal samples and its applications

### 2.1 Notation and assumptions

A time series  $\{X_t\}$  is said to be  $p$ th-order stationary if all moments up to the  $p$ th moment are invariant to shift (for example, a strictly stationary time series with finite  $p$ -order moment satisfies such a condition). We denote the covariance and  $s$ -order cumulant as  $c(j) = \text{cov}(X_t, X_{t+j})$  and  $\kappa_s(j_1, \dots, j_{s-1}) = \text{cum}(X_t, X_{t+j_1}, \dots, X_{t+j_{s-1}})$ . Furthermore, we define the spectral density and  $s$ -order spectral density functions as

$$f(\omega) = \frac{1}{2\pi} \sum_{j \in \mathbb{Z}} c(j) e^{ij\omega} \text{ and } f_s(\omega_1, \dots, \omega_{s-1}) = \frac{1}{(2\pi)^s} \sum_{j_1, \dots, j_{s-1} \in \mathbb{Z}} \kappa_s(j_1, \dots, j_{s-1}) e^{i(j_1\omega_1 + \dots + j_{s-1}\omega_{s-1})}.$$

To simplify notation we will assume that  $\{X_t\}$  is a zero mean time series, noting that the same methodology also works when the mean of  $\{X_t\}$  is constant since the DFT of a constant mean is zero at most frequencies. We let  $\Re X$  and  $\Im X$  denote the real and imaginary parts of the variable  $X$ .

**Assumption 2.1** ( *$p$ th-order stationary and cumulant conditions*) *The time series  $\{X_t\}_t$  is  $p$ -order stationary, with  $E|X_t|^p < \infty$  and for all  $2 \leq s \leq p$  and  $1 \leq i \leq s$*

$$\sum_{j_1, \dots, j_{s-1} \in \mathbb{Z}} (1 + |j_i|) |\kappa_s(j_1, \dots, j_{s-1})| < \infty.$$

### 2.2 Construction of orthogonal samples and pivotal statistics

In this section the main focus will be on  $A_T(\phi)$  (defined in (3)). We start by reviewing some of the well known sampling properties of  $A_T(\phi)$ . If  $\{X_t\}$  is a fourth order stationary time series which satisfies Assumption 2.1 with  $p = 4$ , then it can be shown that  $A_T(\phi)$  is a mean squared consistent estimator of  $A(\phi)$ , where

$$A(\phi) = \frac{1}{2\pi} \int_0^{2\pi} \phi(\omega) f(\omega) d\omega.$$

Clearly, depending on the choice of  $\phi$ ,  $A_T(\phi)$  estimates several parameters of interest and we give some examples below.

**Example 2.1** (a) The sample autocovariance function at lag  $j$ , with  $\phi(\omega) = \exp(ij\omega)$ , corresponds to

$$\widehat{c}_T(j) = \frac{1}{T} \sum_{k=1}^T \exp(ij\omega_k) |J_T(\omega_k)|^2 = \widetilde{c}_T(j) + \widetilde{c}_T(T-j) = \widetilde{c}_T(j) + O_p\left(\frac{|j|}{T}\right), \quad (4)$$

where  $\widetilde{c}_T(j) = \frac{1}{T} \sum_{t=1}^{T-|j|} X_t X_{t+|j|}$ .

(b) The spectral density estimator with  $\phi(\omega) = b^{-1}W\left(\frac{\omega-\omega_k}{b}\right)$ .

(c) In order to test for goodness of fit of a model with spectral density function  $g(\omega; \theta)$ , Milhoj [1981] proposed estimating the  $j$ th autocovariance function of the residuals obtained by fitting the linear model corresponding to  $g(\omega; \theta)$  using

$$\widehat{\gamma}_T(j) = \frac{1}{T} \sum_{k=1}^T \frac{\exp(ij\omega_k)}{g(\omega_k; \theta)} |J_T(\omega_k)|^2.$$

In this case  $\widehat{\gamma}_T(j) = A_T(e^{ij\cdot} g(\cdot; \theta)^{-1})$  and  $\phi(\omega) = e^{ij\omega} g(\omega; \theta)^{-1}$ .

(d) The Whittle likelihood estimator (which is asymptotically equivalent to the quasi-Gaussian likelihood), where  $\widehat{\theta}_T = \arg \min_{\theta \in \Theta} \mathcal{L}_T(\theta)$  with

$$\mathcal{L}_T(\theta) = \frac{1}{T} \sum_{k=1}^T \left( \frac{|J_T(\omega_k)|^2}{f(\omega_k; \theta)} + \log f(\omega_k; \theta) \right)$$

and  $\Theta$  is a compact parameter space. For the purpose of estimation and testing usually the derivative of the likelihood is required, where

$$\nabla_{\theta} \mathcal{L}_T(\theta) = A_T(\phi) + \frac{1}{T} \sum_{k=1}^T \frac{1}{f(\omega_k; \theta)} \nabla_{\theta} f(\omega_k; \theta) \quad (5)$$

and  $\phi(\omega) = \nabla_{\theta} f(\omega_k; \theta)^{-1}$ .

(e) Consider the quadratic form  $A_T(\phi) = \frac{1}{T} \sum_{t,\tau=1}^T \Phi_{t-\tau} X_t X_{\tau}$ , if  $\Phi_s = \Phi_{-s}$  for all  $s \in \mathbb{Z}$ , then  $A_T(\phi)$  can be written as

$$A_T(\phi) = \frac{1}{T} \sum_{k=1}^T \phi_T(\omega_k) |J_T(\omega_k)|^2$$

where  $\phi_T(\omega) = \frac{1}{T} \sum_{s=0}^{T-1} \Phi_s e^{is\omega}$ .

Under stationarity and some additional mixing-type and regularity conditions it is easily shown that  $T\text{var}[A_T(\phi)] = V(0) + O(T^{-1})$ , where

$$\begin{aligned} V(0) &= \frac{1}{2\pi} \int_0^{2\pi} f(\omega)^2 \left( |\phi(\omega)|^2 + \phi(\omega)\overline{\phi(-\omega)} \right) d\omega + \\ &\quad \frac{1}{2\pi} \int_0^{2\pi} \int_0^{2\pi} \phi(\omega_1)\overline{\phi(\omega_2)} f_4(\omega_1, -\omega_1, -\omega_2) d\omega_1 d\omega_2. \end{aligned} \quad (6)$$

It is clear that the variance,  $V(0)$ , has a complicated structure and cannot be directly estimated. Instead, we obtain an orthogonal sample associated with  $A_T(\phi)$  to estimate  $V(0)$ . To do so, we recall some of the pertinent features of the orthogonal sample associated with the sample mean. To summarize,  $\{\sqrt{2}\Re J_T(\omega_k), \sqrt{2}\Im J_T(\omega_k); k = 1, \dots, M\}$  is a ‘near uncorrelated’ sequence which has similar distributional properties as a centralised version of  $\sqrt{T/2\pi}\bar{X}_T = J_T(0)$ . Returning to  $A_T(\phi)$  we observe that it is a weighted average of the periodogram  $|J_T(\omega_k)|^2$ . We now compare  $|J_T(\omega_k)|^2$  with  $J_T(\omega_k)\overline{J_T(\omega_{k+r})}$ . Using Theorem 4.3.2, Brillinger [1981], it is clear that  $|J_T(\omega_k)|^2$  and  $\{\sqrt{2}\Re J_T(\omega_k)\overline{J_T(\omega_{k+r})}, \sqrt{2}\Im J_T(\omega_k)\overline{J_T(\omega_{k+r})}\}$  are estimating very different quantities (the spectral density and zero respectively). However, they are almost uncorrelated and in the case that  $r$  is small and  $k > 0$  they have approximately the same variance. This suggests that in order to construct the orthogonal sample associated with  $A_T(\phi)$  we replace  $|J_T(\omega_k)|^2$  with  $J_T(\omega_k)\overline{J_T(\omega_{k+r})}$  and define

$$A_T(\phi; r) = \frac{1}{T} \sum_{k=1}^T \phi(\omega_k) J_T(\omega_k) \overline{J_T(\omega_{k+r})} \quad r \in \mathbb{Z}. \quad (7)$$

Note that  $A_T(\phi; 0) = A_T(\phi)$ . In the following lemmas we show that  $\{\sqrt{2}\Re A_T(\phi; r), \sqrt{2}\Im A_T(\phi; r)\}_{r=1}^M$  is an orthogonal sample to  $A_T(\phi)$ . We first show that in general  $A_T(\phi; 0)$  and  $A_T(\phi; r)$  ( $r > 0$ ) have differing means.

**Lemma 2.1** *Suppose that  $\{X_t\}$  satisfies Assumption 2.1 with  $p = 2$  and  $\phi(\cdot)$  is a Lipschitz continuous bounded function. Then we have*

$$E[A_T(\phi; r)] = \begin{cases} \frac{1}{2\pi} \int_0^{2\pi} \phi(\omega) f(\omega) d\omega + O(T^{-1}) & r = 0 \\ O(T^{-1}) & 0 < r < T/2 \end{cases}$$

**PROOF** In the Supplementary material. □

Despite  $A_T(\phi; r)$  having different expectations in the following lemma and corollary we show that they share similar second order properties.

**Theorem 2.1** *Suppose  $\{X_t\}$  satisfies Assumption 2.1 with  $p = 4$  and the function  $\phi : [0, 2\pi] \rightarrow \mathbb{R}$  is a Lipschitz continuous bounded function.*

(i) Then we have

$$T\text{var}[A_T(\phi)] = T\text{var}[A_T(\phi; 0)] = V(0) + O(T^{-1})$$

$$T\text{cov}[\Re A_T(\phi; r_1), \Re A_T(\phi; r_2)] = \begin{cases} \frac{1}{2}V(\omega_r) + O(T^{-1}) & 0 < r_1 = r_2 (= r) \\ O(T^{-1}) & 0 < r_1 \neq r_2 \leq T/2 \end{cases}$$

$$T\text{cov}[\Im A_T(\phi; r_1), \Im A_T(\phi; r_2)] = \begin{cases} \frac{1}{2}V(\omega_r) + O(T^{-1}) & 0 < r_1 = r_2 (= r) \\ O(T^{-1}) & 0 < r_1 \neq r_2 \leq T/2 \end{cases}$$

and

$$T\text{cov}[\Re A_T(\phi; r_1), \Im A_T(\phi; r_2)] = O(T^{-1}) \quad 0 < r_1, r_2 \leq T/2$$

(ii) Suppose, further that Assumption 2.1 holds with  $p = 8$ , then we have

$$\text{cov} \left\{ |\sqrt{T}A_T(\phi; r_1)|^2, |\sqrt{T}A_T(\phi; r_2)|^2 \right\} = \begin{cases} V(\omega_r)^2 + O(T^{-1}) & 0 < r_1 = r_2 (= r) \\ O(T^{-1}) & 0 \leq r_1 < r_2 < T/2 \end{cases} \quad (8)$$

where

$$\begin{aligned} V(\omega_r) &= \frac{1}{2\pi} \int_0^{2\pi} f(\omega) f(\omega + \omega_r) \left( |\phi(\omega)|^2 + \phi(\omega) \overline{\phi(-\omega - \omega_r)} \right) d\omega + \\ &\quad \frac{1}{2\pi} \int_0^{2\pi} \int_0^{2\pi} \phi(\omega_1) \overline{\phi(\omega_2)} f_4(\omega_1, -\omega_1 - \omega_r, -\omega_2) d\omega_1 d\omega_2. \end{aligned} \quad (9)$$

**PROOF** In the Supplementary material. □

We observe that Assumption 2.1 with  $p = 4$  implies that the spectral density function  $f(\cdot)$  and fourth order spectral density function  $f_4(\cdot)$  are Lipschitz continuous over each variable. These observations immediately lead to the following result.

**Corollary 2.1** *Suppose Assumption 2.1 with  $p = 4$  holds and  $\phi$  is Lipschitz continuous. Let  $V(\cdot)$  be defined as in (9). Then we have*

$$|V(\omega_r) - V(0)| \leq K|r|T^{-1},$$

where  $K$  is a finite constant that does not depend on  $r$  or  $T$ .

Theorem 2.1(i) and Lemma 2.1 together imply for  $M \ll T$ , that the sequence  $\{\sqrt{2}\Re A_T(\phi; r), \sqrt{2}\Im A_T(\phi; r); r = 1, \dots, M\}$  are ‘near uncorrelated’ random variables with



approximately the same variance,  $V(0)$ . Based on these observations we propose the following estimator of  $V(0)$

$$\widehat{V}_M(0) = \frac{T}{2M} \sum_{r=1}^M (2|\Re A_T(\phi; r)|^2 + 2|\Im A_T(\phi; r)|^2) = \frac{T}{M} \sum_{r=1}^M |A_T(\phi; r)|^2. \quad (10)$$

Below we obtain the orthogonal sample associated with each of the estimators described in Example 2.1.

**Example 2.2** (a) *We recall that the sample covariance is*

$$A_T(e^{ij\cdot}) = \frac{1}{T} \sum_{k=1}^T |J_T(\omega_k)|^2 e^{ij\omega_k} \approx \frac{1}{T} \sum_{t=1}^{T-j} X_t X_{t+j}$$

and the orthogonal sample is approximately

$$A_T(e^{ij\cdot}; r) = \frac{1}{T} \sum_{k=1}^T J_T(\omega_k) \overline{J_T(\omega_{k+r})} e^{ij\omega_k} \approx \frac{1}{T} \sum_{t=1}^{T-j} X_t X_{t+j} e^{-it\omega_r}.$$

Thus the sample covariance is a sample mean, and, as expected, the orthogonal sample is analogous to the DFT of  $\{X_t\}_{t=1}^T$ , but with  $\{X_t X_{t+j}\}$  replacing  $\{X_t\}$ .

(b) *The orthogonal sample for the spectral density estimator is*

$$\widehat{f}(\omega; r) = \frac{1}{bT} \sum_{k=1}^T W\left(\frac{\omega - \omega_k}{b}\right) J_T(\omega_k) \overline{J_T(\omega_{k+r})}.$$

Note that since  $\phi = \frac{1}{b} W\left(\frac{\omega - \omega_k}{b}\right)$  is not a bounded function (over  $b$ ) the rates in Lemma 2.1 and Theorem 2.1 do not hold, and some adjustment of the rate is necessary.

(c) *The orthogonal sample for  $\widehat{\gamma}_T(j)$  is  $A_T(e^{ij\cdot} g(\cdot; \theta)^{-1}; r) = T^{-1} \sum_{k=1}^T e^{ij\omega_k} g(\omega_k; \theta)^{-1} J_T(\omega_k) \overline{J_T(\omega_{k+r})}$ .*

(d) *If the objective is to test  $H_0 : \theta = \theta_0$  versus  $H_A : \theta \neq \theta_0$  using the score test based on the Whittle likelihood,  $\sqrt{T} \nabla_{\theta_0} \mathcal{L}_T(\theta)$ , then we require an estimator of the variance*

$$V = \lim_{T \rightarrow \infty} \text{var}[\sqrt{T} \nabla_{\theta_0} \mathcal{L}_T(\theta)] = \lim_{T \rightarrow \infty} \text{var}[\sqrt{T} A_T(\nabla_{\theta_0} f(\omega_k; \theta)^{-1})].$$

In this case the orthogonal sample is

$$A_T(\nabla_{\theta_0} f(\omega_k; \theta)^{-1}; r) = T^{-1} \sum_{k=1}^T \nabla_{\theta_0} f(\omega_k; \theta)^{-1} J_T(\omega_k) \overline{J_T(\omega_{k+r})}.$$

(e) *If  $\sum_{s=0}^{\infty} |\Phi_s| < \infty$  and  $A_T(\phi) = \frac{1}{T} \sum_{t,\tau=1}^T \Phi_{t-\tau} X_t X_\tau = \frac{1}{T} \sum_{k=1}^T \phi_T(\omega_k) |J_T(\omega_k)|^2$ , then the corresponding orthogonal sample can be written as  $A_T(\phi; r) = \frac{1}{T} \sum_{k=1}^T \phi_T(\omega_k) J_T(\omega_k) \overline{J_T(\omega_{k+r})} = \frac{1}{T} \sum_{t,\tau=1}^T \Phi_{t-\tau} X_t X_\tau e^{-it\omega_r}$ .*

We return to the Whittle likelihood considered in Example 2.1(d) and 2.2(d). In Example 2.2(d), we use the Whittle likelihood for hypothesis testing. However, the Whittle likelihood is also used in estimation, where  $\hat{\theta}_T = \arg \min \mathcal{L}_T(\theta)$  is an estimator of the true parameter,  $\theta_0$ . By using (5) and the Taylor series expansion it is well known that

$$\sqrt{T} \left( \hat{\theta}_T - \theta_0 \right) \xrightarrow{\mathcal{D}} \mathcal{N} \left( 0, W_{\theta_0}^{-1} V_{\theta_0} W_{\theta_0}^{-1} \right),$$

where  $W_{\theta_0} = \frac{1}{2\pi} \int_0^{2\pi} f(\omega)^{-2} \nabla_{\theta_0} f_{\theta}(\omega; \theta) \nabla_{\theta_0} f_{\theta}(\omega; \theta)' d\omega$  and  $V_{\theta_0} = \lim_{T \rightarrow \infty} T \text{var}[A_T(\nabla_{\theta_0} f(\omega_k; \theta)^{-1})]$ . However,  $\theta_0$  is unknown, and we only have an estimator  $\hat{\theta}_T$ . Instead, we replace  $\theta_0$  with  $\hat{\theta}_T$  and use

$$A_T(\nabla_{\hat{\theta}_T} f(\cdot; \hat{\theta}_T)^{-1}; r) = \frac{1}{T} \sum_{k=1}^T \nabla_{\hat{\theta}_T} f(\omega_k; \hat{\theta}_T)^{-1} J_T(\omega_k) \overline{J_T(\omega_{k+r})}$$

as the orthogonal sample and

$$\hat{V}_{\hat{\theta}_T, M}(0) = \frac{T}{M} \sum_{r=1}^M \left| A_T(\nabla_{\hat{\theta}_T} f(\cdot; \hat{\theta}_T)^{-1}; r) \right|^2 \quad (11)$$

as an estimator of  $V_{\theta_0}(0) = \lim_{T \rightarrow \infty} T \text{var}[A_T(\nabla_{\theta_0} f(\cdot; \theta_0)^{-1})]$ .

It is straightforward to generalize this idea to estimate  $V_{\theta_0}(0) = \text{var} \left[ \sqrt{T} A_T(\phi_{\theta_0}) \right]$ , where  $\phi_{\theta}$  is a known function but  $\theta_0$  is unknown. Given an estimator  $\hat{\theta}_T$  of  $\theta_0$ , an obvious estimator of  $V_{\theta_0} = \lim_{T \rightarrow \infty} T \text{var}[A_T(\phi_{\theta_0})]$  is

$$\hat{V}_{\hat{\theta}_T, M}(0) = \frac{T}{M} \sum_{r=1}^M |A_T(\phi_{\hat{\theta}_T}; r)|^2, \quad \text{where} \quad A_T(\phi_{\hat{\theta}_T}; r) = \frac{1}{T} \sum_{k=1}^T \phi(\omega_k; \hat{\theta}_T) J_T(\omega_k) \overline{J_T(\omega_{k+r})}. \quad (12)$$

In Lemmas 2.2 and 2.3, below, we show that  $\hat{V}_M(0)$  and  $\hat{V}_{\hat{\theta}_T, M}(0)$  are consistent estimators of the variance. However, in most applications variance estimation is mainly required in the construction of confidence intervals or to test a hypothesis. In which case the main object of interest is the studentized statistic

$$T_M = \frac{\sqrt{T}[A_T(\phi) - A(\phi)]}{\sqrt{\hat{V}_M(0)}}. \quad (13)$$

Since  $\hat{V}_M(0)$ , is a consistent estimator of  $V(0)$  as  $M/T \rightarrow 0$  with  $M, T \rightarrow \infty$ , under suitable conditions we would expect  $T_M \xrightarrow{\mathcal{D}} N(0, 1)$ . However, such an approximation does not take into account that  $\hat{V}_M(0)$  is *only* an estimator of the variance. We want a better finite sample approximation that takes into account that  $M$  is fixed. This requires the following theorem.

**Theorem 2.2** *Let us suppose that  $\{X_t\}$  is a stationary  $\alpha$ -mixing time series, where the  $\alpha$ -mixing coefficient  $\alpha(t)$  is such that  $\alpha(t) \leq K|t|^{-s}$  (for  $|t| \neq 0$ ), where  $s > 6$  and  $K < \infty$  and for some  $r > 4s/(s - 6)$  we have  $E|X_t|^r < \infty$ .*

Let  $A_T(\phi)$  and  $A_T(\phi; r)$  be defined as in (3) and (7) respectively. We assume that  $\phi : [0, 2\pi] \rightarrow \mathbb{R}$  has a bounded second derivative and  $A_T(\phi)$  is a real-valued random variable. Let  $A_{M,T} = (A_T(\phi; 1), \dots, A_T(\phi; M))'$ . Then for  $M$  fixed we have

$$\sqrt{\frac{T}{V(0)}} \begin{pmatrix} A_T(\phi) - A(\phi) \\ \sqrt{2}\Re A_{M,T} \\ \sqrt{2}\Im A_{M,T} \end{pmatrix} \xrightarrow{\mathcal{D}} \mathcal{N}(0, I_{2M+1}), \quad (14)$$

where  $I_{2M+1}$  denotes the identity matrix of dimension  $2M + 1$ .

**PROOF** The proof immediately follows from Theorem 2.1, Lee and Subba Rao [2016].

Note that the stated conditions imply that Assumption 2.1 holds with  $p = 4$ , see Statulevicius and Jakimavicius [1988], Theorem 3, part (2) and Remark 3.1, Neumann [1996].  $\square$

Using the asymptotic independence of  $\Re A_{M,T}$  and  $\Im A_{M,T}$  (proved in the above theorem) we observe that for fixed  $M$

$$M \frac{\widehat{V}_M(0)}{V(0)} \xrightarrow{\mathcal{D}} \chi_{2M}^2$$

as  $T \rightarrow \infty$ , where  $\chi_{2M}^2$  denotes a chi-square distribution with  $2M$  degrees of freedom. Furthermore, since  $A_T(\phi)$  is asymptotically independent of  $\Re A_{M,T}$  and  $\Im A_{M,T}$  for fixed  $M$  we have

$$T_M \xrightarrow{\mathcal{D}} t_{2M} \quad (15)$$

as  $T \rightarrow \infty$ , where  $T_M$  is defined in (13) and  $t_{2M}$  denotes the  $t$ -distribution with  $2M$  degrees of freedom. Observe that this result is analogous to the fixed  $M$ -asymptotic result in (1) and Sun [2013].

**Remark 2.1** *The above method can be generalized to estimate the covariance between several estimators which take the form (3). Let  $\mathbf{A} = (A(\phi_1), \dots, A(\phi_p))$  denote a  $p$ -dimensional parameter and  $\mathbf{A}_T = (A_T(\phi_1), \dots, A_T(\phi_p))$  their corresponding estimators. Further, let  $\mathbf{A}_T(r) = (A_T(\phi_1; r), \dots, A_T(\phi_p; r))$  denote the orthogonal sample vector associated with  $\mathbf{A}_T$ . It can be shown that  $\text{var}[\sqrt{T}\mathbf{A}_T] = \Sigma + O(T^{-1})$  where*

$$\begin{aligned} \Sigma_{j_1, j_2} &= \frac{1}{2\pi} \int_0^{2\pi} f(\omega)^2 \left[ \phi_{j_1}(\omega) \overline{\phi_{j_2}(\omega)} + \phi_{j_1}(\omega) \overline{\phi_{j_2}(-\omega)} \right] d\omega + \\ &\quad \frac{1}{2\pi} \int_0^{2\pi} \int_0^{2\pi} \phi_{j_1}(\omega_1) \overline{\phi_{j_2}(\omega_2)} f_4(\omega_1, -\omega_1, -\omega_2) d\omega_1 d\omega_2. \end{aligned} \quad (16)$$

Based on similar ideas to those presented above we estimate the variance  $\Sigma$  with

$$\widehat{\Sigma}_M = \frac{T}{M} \sum_{r=1}^M \{\Re \mathbf{A}_T(r) \mathbf{A}_T(r)^* + \Im \mathbf{A}_T(r) \mathbf{A}_T(r)^*\}$$

where  $\mathbf{A}_T(r)^*$  denote the complex conjugate and transpose of  $\mathbf{A}_T(r)$ . Furthermore, if  $M$  is kept fixed, the statistic

$$T (\mathbf{A}_T - \mathbf{A})' \widehat{\Sigma}_M^{-1} (\mathbf{A}_T - \mathbf{A}) \xrightarrow{\mathcal{D}} T_{p,2M}^2, \quad T \rightarrow \infty$$

where  $T_{p,2M}^2$  denotes Hotelling's  $T$ -squared distribution with  $2M$ -degrees of freedom.

Finally, we show that  $\widehat{V}_M(0)$  is a mean square consistent estimator of  $V$ , if  $M \rightarrow \infty$  as  $T \rightarrow \infty$ .

**Lemma 2.2** *Suppose Assumption 2.1 with  $p = 8$  is satisfied and  $\phi$  is Lipschitz continuous. Let  $\widehat{V}_M(0)$  be defined as in (10). Then we have  $|\mathbb{E}[\widehat{V}_M(0)] - V(0)| = O(M/T)$  and*

$$\mathbb{E} \left( \widehat{V}_M(0) - V(0) \right)^2 = O \left( \frac{M^2}{T^2} + \frac{1}{M} \right), \quad (17)$$

where  $M \rightarrow \infty$  as  $T \rightarrow \infty$ .

**PROOF** In the Supplementary material. □

It is interesting to note that the estimator of  $\widehat{V}_M(0)$  is analogous to kernel estimators in nonparametric regression, where  $M$  plays the role of window width. From Lemma 2.2 we observe if  $M$  is large, then  $\widehat{V}_M(0)$  can have a large bias. On the other hand, if  $M$  is small the bias is small but the variance is large. However, by using a small  $M$ , we can correct for the large variance of  $\widehat{V}_M(0)$  by using the  $t$ -distribution approximation given in (15). The cost of using small  $M$  are slightly larger critical values.

Using the above we show that  $\widehat{V}_{\widehat{\theta},M}(0)$  consistently estimates  $V_\theta$ .

**Lemma 2.3** *Suppose Assumption 2.1 with  $p = 8$  is satisfied. Let  $\widehat{\theta}_T$  be a consistent estimator of  $\theta$  such that  $|\widehat{\theta}_T - \theta| = O_p(T^{-1/2})$ ,  $\sup_{\theta,\omega} \left| \frac{\partial \phi(\omega;\theta)}{\partial \omega} \right| < \infty$  and  $\sup_{\theta,\omega} \left| \frac{\partial^2 \phi(\omega;\theta)}{\partial \theta^2} \right| < \infty$ . Let  $\widehat{V}_{\widehat{\theta}_T,M}(0)$  be defined as in (12) and  $V_\theta = \lim_{T \rightarrow \infty} T \text{var}[A_T(\phi_\theta)]$ . Then we have*

$$\left| \widehat{V}_{\widehat{\theta}_T,M}(0) - V_\theta \right| = O_p \left( \frac{M}{T} + \frac{1}{\sqrt{M}} \right),$$

where  $M \rightarrow \infty$  as  $T \rightarrow \infty$ .

**PROOF** In the Supplementary material. □

**Example 2.3** *The above result can be used to show consistency of the variance estimator of the Whittle likelihood estimator defined in (11). We observe that if  $f(\omega; \theta)$  is uniformly bounded away from zero and uniformly bounded from above for all  $\theta \in \Theta$  and  $\omega \in [0, 2\pi]$ , and its first and second derivatives with respect to  $\theta$  and  $\omega$  are uniformly bounded, then (11) is a consistent estimator of  $V_\theta$  if  $M/T \rightarrow 0$  as  $M \rightarrow \infty$  and  $T \rightarrow \infty$ .*

### 2.3 Example 1

We illustrate the result in (15) with some simulations. Let  $\hat{c}_T(1) = A_T(e^{i\cdot})$  denote the estimator of the covariance at lag one (defined in (4)) and  $\{A_T(e^{i\cdot}; r)\}$  the corresponding orthogonal sample. We use  $M = 5$  and define the studentized statistic

$$T_{10} = \frac{A_T(e^{i\cdot}) - c(1)}{\sqrt{\frac{1}{5} \sum_{r=1}^5 |A_T(e^{i\cdot}; r)|^2}}. \quad (18)$$

We focus on models where there is no correlation, thus  $c(j) = 0$  for  $j \neq 0$ , but possible higher order dependence. For models (ii) and (iii) we consider below, we define the uncorrelated time series

$$X_t = \sum_{j=0}^{\infty} 0.6^j \varepsilon_{t-j} - \frac{0.6}{1 - 0.6^2} \varepsilon_{t+1} \quad (19)$$

where  $\{\varepsilon_t\}$  are uncorrelated random variables. The models we consider are

- (i)  $\{X_t\}$  are independent, identically distributed (iid) normal random variables
- (ii)  $\{X_t\}$  satisfies (19) where the innovations  $\varepsilon_t$  are iid t-distributed random variables with 5df (thus  $\{X_t\}$  is an uncorrelated, non-causal linear time series with a finite fourth moment)
- (iii)  $\{X_t\}$  satisfies (19) where the innovations  $\varepsilon_t$  satisfy the ARCH(1) representation  $\varepsilon_t = \sigma_t Z_t$  with  $\sigma_t^2 = 1 + 0.7\varepsilon_{t-1}^2$  and  $\{Z_t\}$  are Gaussian random variables.  $\{X_t\}$  is a nonlinear, uncorrelated time series whose fourth moment is not finite, thus  $A_T(e^{i\cdot})$  will not have a finite variance.

For each model a time series of size  $T = 100$  and  $200$  is generated and  $T_{10}$  evaluated (see equation (18)). This is done over 1000 replications. The QQplot of  $T_{10}$  against the quantiles of a t-distribution with 10 df are given in Figures 1 (for model (i)) 2 (for model (ii)) and 3 (for model (iii)). It is reassuring to see that even when the sample size is relatively small ( $T = 100$ ), for model (i) and (ii), the finite sample distribution of  $T_{10}$  is close to  $t_{10}$ . Furthermore, the small deviation seen in the tails when  $T = 100$  is reduced when the sample size is increased to  $T = 200$ . These simulations suggest the assertions in (15) (and

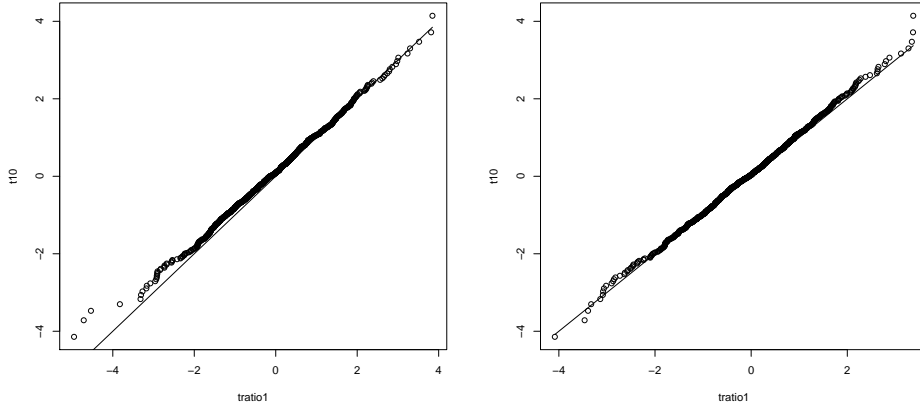


Figure 1:  $X_t$  are iid standard normal random variables. Left:  $T = 100$ . Right  $T = 200$ .

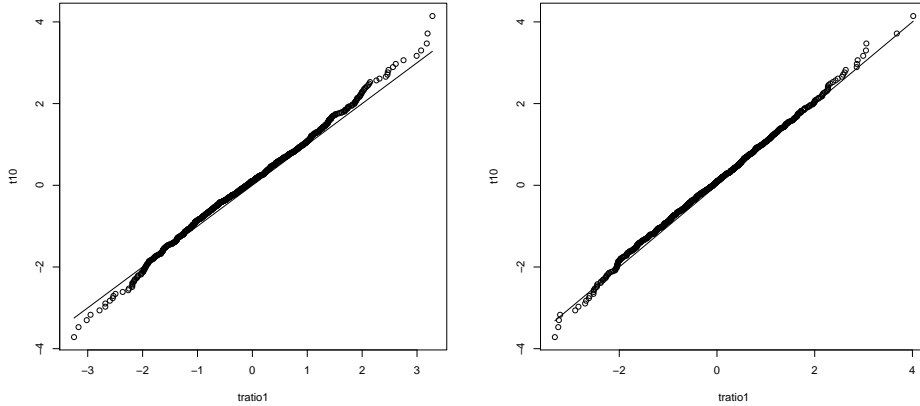


Figure 2:  $X_t$  satisfies (19) where the innovations are from  $t$ -distribution with 5df. Left:  $T = 100$ . Right  $T = 200$ .

Sun [2013], since the sample covariance is a “mean-type” estimator) are reasonable. On the other hand, for model (iii)  $E[X_t^4]$  is not finite, thus  $\text{var}[A_T(e^i; r)]$  is not finite and the assumptions which underpin the result in (15) do not hold. This is apparent in Figure 3, where the  $t$ -distribution seems inappropriate. It is interesting to investigate the limiting distribution of  $T_{10}$ , in particular whether it is pivotal, when the variance of the numerator is not finite. We leave this for future research.

## 2.4 Example 2: Testing for equality of spectral densities

Suppose that  $\{X_t\}$  and  $\{Y_t\}$  are two time series which are jointly stationary and with univariate spectral densities  $f_X$  and  $f_Y$  respectively. We now apply the above methodology to testing for equality of spectral densities i.e.  $H_0 : f_X(\omega) = f_Y(\omega)$  for all  $\omega \in [0, 2\pi]$  against

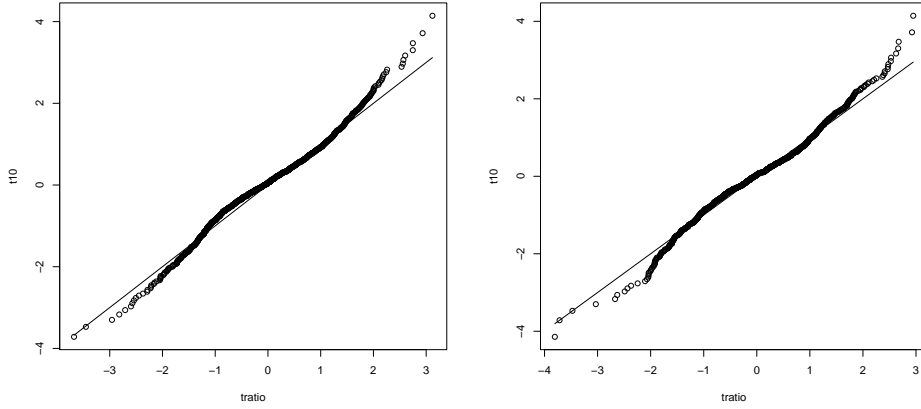


Figure 3:  $X_t$  satisfies (19) where the innovations are an ARCH process (the fourth moment does not exist). Left:  $T = 100$ . Right:  $T = 200$ .

$H_A : f_X(\omega) \neq f_Y(\omega)$  for some  $\omega$  (with non-zero measure). Eichler [2008] and Dette and Paparoditis [2009] propose testing for equality of the spectral densities using an  $L_2$ -distance, this requires estimators for  $f_X$  and  $f_Y$ . Define

$$\begin{aligned}\widehat{f}_X(\omega_l; r) &= \frac{1}{bT} \sum_{k=1}^T W\left(\frac{\omega_l - \omega_k}{b}\right) J_{X,T}(\omega_k) \overline{J_{X,T}(\omega_{k+r})} \\ \widehat{f}_Y(\omega_l; r) &= \frac{1}{bT} \sum_{k=1}^T W\left(\frac{\omega_l - \omega_k}{b}\right) J_{Y,T}(\omega_k) \overline{J_{Y,T}(\omega_{k+r})},\end{aligned}$$

where  $J_{X,T}(\omega_k)$  and  $J_{Y,T}(\omega_k)$  denote the DFT of  $\{X_t\}_{t=1}^T$  and  $\{Y_t\}_{t=1}^T$  respectively and  $W(\cdot)$  is a spectral window. It is clear that

$$\widehat{f}_X(\omega_l) = \widehat{f}_X(\omega_l; 0) \text{ and } \widehat{f}_Y(\omega_l) = \widehat{f}_Y(\omega_l; 0) \quad (20)$$

are estimators of the spectral density and, from Example 2.2(b),  $\widehat{f}_X(\omega_l; r)$  and  $\widehat{f}_Y(\omega_l; r)$  ( $r \neq 0$ ) are the corresponding orthogonal sample.

An obvious method for testing equality of the spectral densities is to use the  $L_2$ -statistic

$$S_T = \frac{2}{T} \sum_{j=1}^{T/2} \left| \widehat{f}_X(\omega_j) - \widehat{f}_Y(\omega_j) \right|^2, \quad (21)$$

where  $\widehat{f}_X(\omega)$  and  $\widehat{f}_Y(\omega)$  are the spectral density estimators defined in (20). Let  $\mu_T$  and  $\sigma_T^2$  denote the mean and variance of  $S_T$  under the null hypothesis. Under the null hypothesis and suitable mixing conditions it can be shown that

$$\frac{S_T - \mu_T}{\sigma_T} \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1) \text{ as } T \rightarrow \infty, \quad (22)$$

where expressions for the mean and variance,  $\mu_T$  and  $\sigma_T^2$ , can be deduced from Eichler [2008], Theorem 3.11. Under the alternative hypothesis, the mean of  $S_T$  is different and it is this dichotomy that gives the test power. However, from a practical perspective the expressions for  $\mu_T$  and  $\sigma_T$  are rather complicated functions of the spectrum and cross-spectrum of  $\{X_t, Y_t\}$ , which makes estimating these expressions difficult. Alternatively, a rather painless method is to use the orthogonal sample to estimate the mean and variance. The critical insight, is that under the null hypothesis of equal spectral densities we have

$$\mathbb{E}[\widehat{f}_X(\omega_k) - \widehat{f}_Y(\omega_k)] = 0 \text{ for all } \omega \in [0, 2\pi].$$

On the other hand, regardless of whether the null holds or not,  $\mathbb{E}[\widehat{f}_X(\omega_k; r)] = O(T^{-1})$  and  $\mathbb{E}[\widehat{f}_Y(\omega_k; r)] = O(T^{-1})$ , thus

$$\mathbb{E}[\widehat{f}_X(\omega_k; r) - \widehat{f}_Y(\omega_k; r)] \approx 0.$$

Therefore under the null hypothesis both  $\widehat{f}_X(\omega_k) - \widehat{f}_Y(\omega_k)$  and  $\widehat{f}_X(\omega_k; r) - \widehat{f}_Y(\omega_k; r)$  share (approximately) the same mean. Furthermore, regardless of whether the null or alternative hypothesis is true,  $\widehat{f}_X(\omega_k) - \widehat{f}_Y(\omega_k)$ ,  $\sqrt{2}\Re[\widehat{f}_X(\omega_k; r) - \widehat{f}_Y(\omega_k; r)]$  and  $\sqrt{2}\Im[\widehat{f}_X(\omega_k; r) - \widehat{f}_Y(\omega_k; r)]$  asymptotically have the same variance. These observations allow us to build the orthogonal sample associated with  $S_T$ . Let

$$\begin{aligned} S_{R,T}(r) &= \frac{4}{T} \sum_{j=1}^{T/2} |\Re \widehat{f}_X(\omega_j; r) - \Re \widehat{f}_Y(\omega_j; r)|^2 \\ \text{and } S_{I,T}(r) &= \frac{4}{T} \sum_{j=1}^{T/2} |\Im \widehat{f}_X(\omega_j; r) - \Im \widehat{f}_Y(\omega_j; r)|^2. \end{aligned}$$

Tedious calculations show that if  $M \ll T$ , then  $\{S_{R,T}(r), S_{I,T}(r)\}_{r=1}^M$  have asymptotically the same mean and variance. Furthermore, if the null is true then the mean and variance of  $\{S_{R,T}(r), S_{I,T}(r)\}_{r=1}^M$  and  $S_T$  are asymptotically the same. We define the  $(2M + 1)$ -dimensional vector

$$\mathbf{S}_{T,M} = (S_T, S_{R,T}(1), S_{I,T}(1), \dots, S_{R,T}(M), S_{I,T}(M)).$$

Under sufficient mixing conditions and the null hypothesis, it can be shown that

$$\sigma_T^{-1} (\mathbf{S}_{T,M} - \mu_T \mathbf{1}) \xrightarrow{\mathcal{D}} \mathcal{N}(0, I_{2M+1}), \quad (23)$$

as  $b \rightarrow 0$ ,  $bT \rightarrow \infty$  and  $T \rightarrow \infty$ , where  $\mathbf{1}$  is a  $(2M + 1)$ -dimensional vector of ones. We estimate the mean and variance  $\mu_T$  and  $\sigma_T^2$  using the sample mean and variance of the



orthogonal sample  $\{S_{R,T}(r), S_{I,T}(r)\}_{r=1}^M$ . In particular,

$$\begin{aligned}\widehat{\mu}_T &= \frac{1}{2M} \sum_{r=1}^M [S_{R,T}(r) + S_{I,T}(r)] \\ \text{and } \widehat{\sigma}_T^2 &= \frac{1}{2M} \sum_{r=1}^M [(S_{R,T}(r) - \widehat{\mu}_T)^2 + (S_{I,T}(r) - \widehat{\mu}_T)^2].\end{aligned}\quad (24)$$

Under the null it can be shown that  $\widehat{\mu}_T$  and  $\widehat{\sigma}_T^2$  consistently estimate the mean and variance of  $S_T$ ,  $\mu_T$  and  $\sigma_T^2$  if  $M/T \rightarrow 0$  as  $M \rightarrow \infty$  and  $T \rightarrow \infty$ . This implies that under the null

$$\frac{S_T - \widehat{\mu}_T}{\widehat{\sigma}_T} \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1) \quad \text{as } M, T \rightarrow \infty.$$

Of course, it is more realistic to assume  $M$  is fixed and let  $T \rightarrow \infty$ . To derive the distribution for fixed  $M$  we note that asymptotically  $S_T$  and  $\widehat{\mu}_T$  are independent, this together with (23) gives

$$\frac{\sqrt{T}}{\sigma_T} \begin{pmatrix} S_T - \mu_T \\ \widehat{\mu}_T - \mu_T \end{pmatrix} \xrightarrow{\mathcal{D}} \mathcal{N} \left( 0, \begin{pmatrix} 1 & 0 \\ 0 & (2M)^{-1} \end{pmatrix} \right) \quad \text{as } T \rightarrow \infty.$$

Further more by using (23) we have (i)  $(2M - 1)\widehat{\sigma}_T^2 \xrightarrow{\mathcal{D}} \chi_{2M-1}^2$  and (ii)  $(2M - 1)\widehat{\sigma}_T^2$  is asymptotically independent of  $(S_T - \mu_T)$ . This gives for fixed  $M$

$$\frac{S_T - \widehat{\mu}_T}{\widehat{\sigma}_T} \xrightarrow{\mathcal{D}} \left(1 + \frac{1}{2M}\right)^{1/2} t_{2M-1}, \quad (25)$$

as  $T \rightarrow \infty$ .

**Remark 2.2** For finite  $T$ ,  $S_T$  is positive and the finite sample distribution of  $S_T$  tends to be right skewed. To make  $S_T$  more normal Chen and Deo [2004] (see also Terrell [2003]), propose the power transform,  $S_T^\beta$ , where  $\beta \in \mathbb{R}$  (typically  $\beta < 1$ ), which makes the distribution of  $S_T^\beta$  more normal than  $S_T$ . Since  $S_T$  is asymptotically normal (see (22)) by using Chen and Deo [2004] we have

$$\frac{S_T^\beta - \mu_{(\beta)}(\mu_T, \sigma_T)}{\sigma_{(\beta)}(\sigma_T)} \xrightarrow{\mathcal{D}} N(0, 1)$$

where

$$\mu_{(\beta)}(\mu_T, \sigma_T) = \mu_T^\beta + \frac{1}{2}\beta(\beta - 1)\mu_T^{\beta-2}\sigma_T^2 \quad \text{and} \quad \sigma_{(\beta)}(\sigma_T) = \beta\mu_T^{\beta-1}\sigma_T.$$

Since  $\mu_T$  and  $\sigma_T$  are unknown we replace the above by the estimators defined in (24). Using the same arguments as those used in the derivation of (25) we have

$$\frac{S_T^\beta - \mu_{(\beta)}(\widehat{\mu}_T, \widehat{\sigma}_T)}{\sigma_{(\beta)}(\widehat{\sigma}_T)} \xrightarrow{\mathcal{D}} \left(1 + \frac{1}{2M}\right)^{1/2} t_{2M-1}. \quad (26)$$

Chen and Deo [2004] propose selecting  $\beta$  to minimize the skewness of the statistic, so that the centralized third-order moment is zero. Following the same calculations as those used in Chen and Deo [2004] gives

$$\beta = 1 - \frac{\mu_T \mathbb{E}[S_T - \mu_T]^3}{3\sigma_T^4}. \quad (27)$$

As  $\mu_T$ ,  $\sigma_T$  and  $\mathbb{E}[S_T - \mu_T]^3$  are unknown we estimate them using the orthogonal sample  $\{S_{R,T}(r), S_{I,T}(r)\}_{r=1}^M$ .

We illustrate the above procedure with some simulations. Following Dette and Paparoditis [2009], we use the linear bivariate time series

$$X_t = 0.8X_{t-1} + \varepsilon_t \quad \text{and} \quad Y_t = 0.8Y_{t-1} + \delta Y_{t-2} + \eta_t, \quad (28)$$

where  $\{(\varepsilon_t, \eta_t)\}_t$  are iid bivariate Gaussian random variables with  $\text{var}[\varepsilon_t] = 1$ ,  $\text{var}[\eta_t] = 1$  and  $\text{cov}[\varepsilon_t, \eta_t] = \rho$ . If  $\delta = 0$  then the spectral densities of  $\{X_t\}$  and  $\{Y_t\}$  are the same and the null hypothesis is true. For the alternative hypothesis we use  $\delta = 0.1$  and  $-0.1$ . We use the test statistic  $S_T$  defined in (21) and (26), where the spectral density is estimated using the Daniell kernel. We use (i)  $\beta = 0.25$  and (ii) an estimate of (27) (which we denote as  $\hat{\beta}$ ). In the simulations we use  $T = 128, 512$  and  $1024$  over 500 replications.

The results are reported in Table 1, where all the tests are done at the 5% level. From Table 1 we see that under the null the test statistic keeps the 5% level relatively well when  $T = 512$  and  $T = 1024$ . We note there is some over rejection when  $T = 128$ . In the case  $\rho = 0.1$  and  $-0.1$  (thus the alternative is true), the test has power which grows as  $T$  grows. We note that the estimated power transform  $\hat{\beta}$  tends, on average, to be larger than 0.25. Thus the distribution of  $S_T^{\hat{\beta}}$  tends to be more right skewed than  $S_T^{0.25}$ . This may explain why the proportion of rejection levels under the null using  $S_T^{\hat{\beta}}$  are a little larger than those with  $S_T^{0.25}$ . Comparing our results to those reported in Dette and Paparoditis [2009], Table 1, we see their frequency bootstrap procedure has better power. There are two possible explanations for this (i) they use a different test statistic based on ratios (ii) the power transform may make the test a little conservative in rejecting the null. However, it is interesting to note that a procedure with very little computational expensive performs well even against highly sophisticated bootstrap procedures.

In the following section, we look again at the issue of testing. We have seen that by construction the orthogonal sample has approximately the same variance as the statistic of interest. Furthermore, in many testing procedures, under the null, both the test statistic and the orthogonal sample have the same distribution. In this section, this property was exploited to estimate the mean and variance of the test statistic. This in conjunction with the t-distribution (with the appropriate degrees of freedom) was used to obtain the p-value

of the test statistic. An alternative approach is to use the orthogonal sample to estimate the sampling distribution of the test statistic under the null hypothesis. We investigate this idea in the following section.

$\rho$	$\delta$	$T = 128$ and $b = 0.15$		$T = 512$ and $b = 0.1$		$T = 1024$ and $b = 0.1$	
		$\hat{\beta}$	$\beta = 0.25$	$\hat{\beta}$	$\beta = 0.25$	$\hat{\beta}$	$\beta = 0.25$
0.9	0.0	14.4	9.6	6.2	5.4	5.6	4.8
	0.1	54.4	39.6	93.2	91.2	100	99.6
	-0.1	32.8	24.2	74.8	73.8	95.6	97.8
0.5	0.0	13.6	8.4	5.0	4.0	3.2	3.0
	0.1	32.8	27.2	57.4	51.6	81.8	81.2
	-0.1	16.6	12.4	30.2	27.6	48	47
0.0	0.0	12.8	8.6	4.4	3.8	4.8	4.4
	0.1	26.8	20.4	47.2	43	72.4	70
	-0.1	13.2	9.2	19.4	17.7	33	31.6
-0.5	0.0	12.2	8.0	7.4	5.2	3.6	2.8
	0.1	31.6	24.8	54	49.2	80	80.4
	-0.1	16.8	13.0	23.8	20.8	43.8	43.6
-0.9	0.0	10.8	8.0	7.6	5.8	4.6	4.2
	0.1	56.8	45.6	90	87.4	98.4	98.4
	-0.1	32.6	24.8	74.2	69.6	93.4	94

Table 1:  $\delta$  and  $\rho$  are defined in (28). When  $\delta = 0$   $H_0$  is true. When  $T = 128$  we use  $M = 6$ , when  $T = 512$  we use  $M = 12$  and when  $T = 1024$  we use  $M = 18$ . Note that the choice of  $M$  was arbitrary and may not be optimal.

### 3 Testing in Time Series

Many test statistics in time series can be formulated in terms of the parameters  $\{A(\phi_j)\}_j$  for some particular set of functions  $\{\phi_j\}$ , where under the null hypothesis  $A(\phi_j) = 0$  for  $j = 1, \dots, L$  and under the alternative  $A(\phi_j) \neq 0$ . Thus under the null, by using Lemma 2.1 we have  $E[A_T(\phi_j)] = O(T^{-1})$ . This motivates the popular  $\ell_2$  test statistic

$$S_T = T \sum_{j=1}^L |A_T(\phi_j)|^2. \quad (29)$$

In this section we use orthogonal samples to estimate the distribution of  $S_T$  under the null hypothesis.

By using the results in Section 2 and Remark 2.1, equation (16), we observe that  $\{A_T(\phi_j)\}_{j=1}^L$  and  $\{\sqrt{2}\Re A_T(\phi_j; r)\}_{j=1}^L$  and  $\{\sqrt{2}\Im A_T(\phi_j; r)\}_{j=1}^L$ , for  $r \ll T$ , asymptotically have the same variance matrix. In addition, under the null hypothesis that  $A(\phi_j) = 0$  for  $j = 1, \dots, L$ , asymptotically  $\{\sqrt{2}\Re A_T(\phi_j; r), \sqrt{2}\Im A_T(\phi_j; r)\}$  and  $A_T(\phi_j)$  have the same mean and limiting Gaussian distribution. This suggests that the distribution of  $S_T$  under

the null can be approximated by the empirical distribution of the corresponding orthogonal sample. Based on the above observations we define the orthogonal sample associated with  $S_T$  as

$$S_{R,T}(r) = 2T \sum_{j=1}^L |\Re A_T(\phi_j; r)|^2 \text{ and } S_{I,T}(r) = 2T \sum_{j=1}^L |\Im A_T(\phi_j; r)|^2 \text{ for } 1 \leq r \leq M. \quad (30)$$

In the theorem below we show that under the null hypothesis  $H_0 : A(\phi_j) = 0$  for  $1 \leq j \leq L$ , the asymptotic sampling properties of  $S_T$ ,  $S_{R,T}(r)$  and  $S_{I,T}(r)$  are equivalent.

**Theorem 3.1** *Suppose Assumption 2.1 holds with  $p = 16$ . Furthermore, we assume  $\{\phi_j\}$  are Lipschitz continuous functions and  $\Re A_T(\phi_j) = A_T(\phi_j)$ . Let  $S_T$ ,  $S_{R,T}(r)$  and  $S_{I,T}(r)$  be defined as in (29) and (30) respectively and define  $V_{j_1, j_2}$  as*

$$\begin{aligned} V_{j_1, j_2} &= \frac{1}{2\pi} \int_0^{2\pi} f(\omega)^2 \left[ \phi_{j_1}(\omega) \overline{\phi_{j_2}(\omega)} + \phi_{j_1}(\omega) \overline{\phi_{j_2}(-\omega)} \right] d\omega + \\ &\quad \frac{1}{2\pi} \int_0^{2\pi} \int_0^{2\pi} \phi_{j_1}(\omega_1) \overline{\phi_{j_2}(\omega_2)} f_4(\omega_1, -\omega_1, -\omega_2) d\omega_1 d\omega_2 + O(T^{-1}). \end{aligned}$$

Then we have

(i) The mean

(a) Under the null hypothesis that  $A(\phi_j) = 0$  for  $1 \leq j \leq L$  we have

$$\mathbb{E}[S_T] = \sum_{j=1}^L V_{j,j} + O(T^{-1}).$$

However, if for at least one  $1 \leq j \leq L$   $A(\phi_j) \neq 0$ , then  $\mathbb{E}[S_T] = O(T)$ .

(b) Under both the null and alternative and for  $0 < r < T/2$  we have

$$\mathbb{E}[S_{R,T}(r)] = \sum_{j=1}^L V_{j,j} + O(|r|T^{-1}) \text{ and } \mathbb{E}[S_{I,T}(r)] = \sum_{j=1}^L V_{j,j} + O(|r|T^{-1}).$$

(ii) The covariance

(a) Under the null hypothesis,  $\text{var}[S_T] = 2 \sum_{j_1, j_2=1}^L V_{j_1, j_2}^2 + O(T^{-1})$

(b) Under both the null and alternative hypothesis where  $1 \leq r_1, r_2 < T/2$  we have

$$\begin{aligned} \text{cov}[S_{R,T}(r_1), S_{R,T}(r_2)] &= \begin{cases} 2 \sum_{j_1, j_2=1}^L V_{j_1, j_2}^2 + O(|r|T^{-1}) & r_1 = r_2 (= r) \neq 0 \\ O(T^{-1}) & \end{cases} \\ \text{cov}[S_{I,T}(r_1), S_{I,T}(r_2)] &= \begin{cases} 2 \sum_{j_1, j_2=1}^L V_{j_1, j_2}^2 + O(|r|T^{-1}) & r_1 = r_2 (= r) \neq 0 \\ O(T^{-1}) & \end{cases} \\ \text{cov}[S_{R,T}(r_1), S_{I,T}(r_2)] &= O(T^{-1}) \quad \text{for all } r_1, r_2 \in \mathbb{Z}. \end{aligned}$$

(iii) Higher order cumulants Suppose Assumption 2.1 holds with the order  $2p$ . Let  $\text{cum}_p$  denote the  $p$ th order cumulant of a random variable. Then under the null hypothesis

$$|\text{cum}_p(S_T) - \text{cum}_p(S_{R,T}(r))| = O(|r|T^{-1}), \quad |\text{cum}_p(S_T) - \text{cum}_p(S_{I,T}(r))| = O(|r|T^{-1}).$$

We observe that the above theorem implies under the null hypothesis,  $S_T$ ,  $S_{R,T}(r)$  and  $S_{R,I}(r)$  asymptotically have equivalent mean, variance and higher order cumulants. Furthermore, under the alternative hypothesis the asymptotic mean and variance of  $S_{R,T}$  and  $S_{I,T}$  are finite and bounded as  $M \rightarrow \infty$  and  $T \rightarrow \infty$ . Therefore motivated by these results we define the empirical distribution

$$\widehat{F}_{M,T}(x) = \frac{1}{2M} \left( \sum_{r=1}^M [I(S_{R,T}(r) \leq x) + I(S_{I,T}(r) \leq x)] \right). \quad (31)$$

To do the test we use  $\widehat{F}_{M,T}(x)$  as an approximation of the distribution of  $S_T$  under the null hypothesis. We reject the null at the  $\alpha\%$ -level if  $1 - \widehat{F}_{M,T}(S_T) < \alpha\%$ . We note that under the alternative hypothesis that at least one  $j = 1, \dots, L$   $A(\phi_j) \neq 0$ , then  $S_T = O_p(T)$ . By Theorem 3.1(ii) the variance of  $S_{R,T}(r)$  and  $S_{I,T}(r)$  is finite and uniformly bounded for all  $r$  and  $T$ . This implies that  $1 - \widehat{F}_{M,T}(S_T) \xrightarrow{\mathcal{P}} 0$  as  $M$  and  $T \rightarrow \infty$ , thus giving the procedure power. To show a Glivenko-Cantelli type result of the form  $\sup_{x \in \mathbb{R}} |\widehat{F}_{M,T}(x) - F(x)| \xrightarrow{\text{a.s.}} 0$  as  $M \rightarrow \infty$  and  $T \rightarrow \infty$ , where  $F$  denotes the limiting distribution of  $S_T$  under the null hypothesis is beyond the scope of the current paper. However, based on the simulations in Section 3.3, we conjecture that this result is true.

We illustrate this procedure to test for uncorrelatedness and goodness of fit.

### 3.1 A Portmanteau test for uncorrelatedness

Let us suppose we observe the stationary time series  $\{X_t\}$ . The classical test for serial correlation assumes that under the null hypothesis the observations are iid random variables. In this case the classical Box-Pierce statistic is defined as

$$\tilde{Q}_T = \frac{T}{\tilde{c}_T(0)^2} \sum_{j=1}^L |\tilde{c}_T(j)|^2, \quad (32)$$

where  $\tilde{c}_T(j)$  is defined in Example 2.1(a). If the null holds, then  $\tilde{Q}_T$  is asymptotically a chi-square distribution with  $L$  degrees of freedom. However, if the intention is to test for uncorrelatedness without the additional constraint of independence, then it can be shown that

$$T \text{cov}[\tilde{c}_T(j_1), \tilde{c}_T(j_2)] = c(0)^2 \delta_{j_1, j_2} + c(0)^2 \delta_{j_1, j_2} \delta_{j_1, 0} + \sum_{k=-\infty}^{\infty} \kappa_4(j_1, k, k + j_2), \quad (33)$$

where  $\delta_{j_1, j_2}$  is the dirac-delta function. Consequently, under the null of uncorrelatedness, the distribution of  $\tilde{Q}_T$  is not a standard chi-square.

In this section we utilize orthogonal samples to estimate the distribution of the Portman-teau statistic under the null hypothesis of no correlation. We recall from Example 2.1(a) that  $A_T(e^{ij\cdot})$  is an estimator of the autocovariance  $\hat{c}_T(j)$ . Therefore, to test for uncorrelatedness at lag  $j = 1, \dots, L$  we define the test statistic

$$Q_T = T \sum_{j=1}^L |A_T(e^{ij\cdot})|^2. \quad (34)$$

Using  $\{\sqrt{2}\Re A_T(e^{ij\cdot}; r), \sqrt{2}\Im A_T(e^{ij\cdot}; r); r = 1, \dots, M\}$  we define the orthogonal sample associated with  $Q_T$  as

$$Q_{R,T}(r) = 2T \sum_{j=1}^L |\Re A_T(e^{ij\cdot}; r)|^2 \text{ and } Q_{I,T}(r) = 2T \sum_{j=1}^L |\Im A_T(e^{ij\cdot}; r)|^2 \text{ for } 1 \leq r \leq M.$$

The above orthogonal sample is used to define the estimated empirical distribution, similar to that defined in (31). We denote the estimated empirical distribution as  $\hat{F}_{Q,M,T}(x)$ . We reject the null at the  $\alpha\%$ -level if  $1 - \hat{F}_{Q,M,T}(Q_T) < \alpha\%$ . Results of the corresponding simulation study is given in Section 3.3, where we apply the proposed methodology to a wide class of uncorrelated processes.

### 3.2 Testing for goodness of fit

In this section we describe how orthogonal samples can be applied to testing for goodness of fit. Given that  $f$  is the spectral density of the observed time series, our objective is to test  $H_0 : f(\omega) = g(\omega; \theta)$  for all  $\omega \in [0, 2\pi]$  against  $H_A : f(\omega) \neq g(\omega; \theta)$  for some  $\omega \in [0, 2\pi]$ . Typically this is done by fitting the model to the data and applying the Portmanteau test to the residuals – in either the time or frequency domain (cf. Milhoj [1981], Hong [1996]). In Example 2.1(c) it was observed that  $A_T(e^{ij\cdot} g(\cdot; \theta)^{-1})$  is an estimator of the covariance of the residuals at lag  $j$ . Thus, under the null hypothesis that  $g$  is the true spectral density, the residual covariance  $A_T(e^{ij\cdot} g(\cdot; \theta)^{-1})$  for  $j \neq 0$  is estimating zero. Using this observation Milhoj [1981] defines the statistic

$$G_T = T \sum_{j=1}^L |A_T(e^{ij\cdot} g(\cdot; \theta)^{-1})|^2 \quad (35)$$

to test for goodness of fit.  $G_T$  gives rise to the orthogonal sample

$$G_{R,T}(r) = 2T \sum_{j=1}^L |\Re A_T(e^{ij\cdot} g(\cdot; \theta)^{-1}; r)|^2 \text{ and } G_{I,T}(r) = 2T \sum_{j=1}^L |\Im A_T(e^{ij\cdot} g(\cdot; \theta)^{-1}; r)|^2.$$

Using Theorem 3.1, under the null hypothesis,  $G_T$ ,  $G_{R,T}(r)$  and  $G_{I,T}(r)$  asymptotically share the same sampling properties when  $r$  is small. We use (31) to define the corresponding empirical distribution, which we denote as  $\widehat{F}_{G,M,T}(x)$ . We reject the null, that the spectral density is  $g(\cdot)$  at the  $\alpha\%$ -level if  $1 - \widehat{F}_{G,M,T}(G_T) < \alpha\%$ .

### 3.3 Simulations

In the following section we assess the tests described above through some simulations. All tests are done at the  $\alpha = 5\%$  and  $\alpha = 10\%$  nominal levels. We compare the orthogonal samples method to block bootstrap method. More precisely, the critical values are obtained using the method described in Romano and Thombs [1996], where a centralized version of the bootstrap statistic is obtained (using 1000 bootstrap replications) and the corresponding empirical distribution and critical value evaluated. We use the block lengths  $B = 5, 10$  and  $20$ .

Throughout this section we let  $\{Z_t\}$  and  $\{\varepsilon_t\}$  denote iid standard normal and chi-square (with one degree of freedom) random variables.

### 3.4 Example 3: Testing for uncorrelatedness

We illustrate the test for uncorrelatedness using the orthogonal sample method described in Section 3.1. We use the test statistic  $Q_T$  (defined in (34)), using  $L = 5$ , and obtain the critical values using the empirical distribution function,  $\widehat{F}_{Q,M,T}$  defined in (31). To select  $M$  we use the average squared criterion described in Section 4. More precisely, we focus on the sample covariance at lag one and choose  $M = \arg \min_{M \in \mathcal{S}} C_\phi(M)$ , where

$$C_\phi(M) = \frac{4}{T} \sum_{r=1}^{T/4} \left( \frac{|\sqrt{T}A_T(e^i; r)|^2}{\widehat{V}_M(\omega_r)} - 1 \right)^2 \quad \text{with} \quad \widehat{V}_M(\omega_r) = \frac{T}{M} \sum_{s=1+r}^{M+r} |A_T(e^i; s)|^2. \quad (36)$$

and  $\mathcal{S} = \{10, \dots, 30\}$ .

We compare our method with the regular Box-Pierce statistic (defined in (32)) and the robust Portmanteau test statistic defined in Diebold [1986], Weiss [1986], Robinson [1991], Bera and Higgins [1993] and Escanciano and Lobato [2009]. The robust Portmanteau test is checking for correlation in the time series, but the test statistics is constructed under the assumption that the time series are martingale differences and  $E[X_0^2 X_{-j_1}, X_{-j_2}] = 0$  when  $j_1 \neq j_2 \geq 1$ . These conditions induce asymptotic uncorrelatedness between the sample covariances. Based on this observation they propose the robust Portmanteau test

$$Q_T^* = T \sum_{j=1}^L \frac{|\tilde{c}_T(j)|^2}{\widehat{\tau}_j}, \quad (37)$$

where  $\widehat{\tau}_j = \frac{1}{T-j} \sum_{t=j+1}^T (X_t - \bar{X})^2 (X_{t-j} - \bar{X})^2$ . Under the null of martingale differences  $Q_T^*$

is asymptotically  $\chi^2$ -distributed with  $L$ -degrees of freedom. In the simulations we use  $L = 5$ .

Lobato et al. [2002] propose a Box-Pierce test for no correlation that does not make the martingale assumption. In this case, there will be correlations between the sample correlations, which Lobato et al. [2002] estimate using kernel-type estimator. They use this to construct a test statistic, which asymptotically under the null hypothesis has a chi-square distribution. We mention that the correlations between the sample covariances can be estimated using the orthogonal sample and a T-statistic similar to that defined in Remark 2.1 can be used as the test statistic, which, for fixed  $M$  but  $T \rightarrow \infty$ , has a Hotelling T-squared distribution (under the null hypothesis).

#### *Models under the null of no correlation*

The first two models we consider are iid random variables which follow a standard normal distribution and a t-distribution with five degrees of freedom. The third model is the uncorrelated two-dependent model  $X_{3,t} = Z_t Z_{t-1}$ . The fourth model we consider is the non-linear, non-martingale, uncorrelated model, defined in Lobato [2001], where  $X_{4,t} = Z_{t-1} Z_{t-2} (Z_{t-1} + Z_t + 1)$ . The fifth model we consider is the ARCH(1) process  $X_{5,t}$ , where

$$X_{5,t} = \sigma_{5,t} Z_t \quad \sigma_{5,t}^2 = 1 + 0.8 X_{5,t-1}^2.$$

The sixth model is  $X_{6,t} = |X_{5,t}| V_t$  where  $\{X_{5,t}\}$  and  $\{V_t\}$  are independent of each other,  $X_{5,t}$  is the ARCH process described above and  $V_t$  is an uncorrelated non-causal time series defined by

$$V_t = \sum_{j=0}^{\infty} a^j \varepsilon_{t-j} - \frac{a}{1-a^2} \varepsilon_{t+1},$$

where  $a = 0.8$ . The seventh model we consider is a ‘pseudo-linear’ non-causal, uncorrelated time series with ARCH innovations defined by

$$X_{7,t} = \sum_{j=0}^{\infty} b_1^j U_{1,t-j} - \frac{b_1}{1-b_1^2} U_{1,t+1}, \quad U_{1,t} = \sum_{j=0}^{\infty} b_2^j U_{2,t-j} - \frac{b_2}{1-b_2^2} U_{2,t+1}$$

where  $U_{2,t} = \sigma_{2,t} Z_t$  with  $\sigma_t = 1 + 0.5 U_{2,t-1}^2$ ,  $b_1 = -0.8$  and  $b_2 = -0.6$ . Finally, the eighth model we consider is the periodically stationary model defined in Politis et al. [1997],  $X_{7,t} = s_t X_{3,t}$  and  $s_t$  is a deterministic sequence of period 12, where the elements are  $\{1, 1, 1, 2, 3, 1, 1, 1, 1, 2, 4, 6\}$  (this time series does not satisfy our stationary assumptions). We used the sample sizes  $T = 100$  and  $T = 500$ .

The results are given in Tables 2 and 3. We observe that overall the orthogonal sampling method keeps to the nominal level. However, there is a mild inflation of the type I error



for independent data (normal and t-distribution), which is probably because the 5th and 10th quantile is estimated using a maximum of 60 points, since the order selection set is  $\mathcal{S} = \{10, \dots, 30\}$  (often it is a lot less than 60). As expected, the regular Box-Pierce statistic keeps the nominal level well for the iid data, but cannot control the type I error when the data is uncorrelated but not iid. The robust Portmanteau test is able to keep the type I error in most cases, the exception being the pseudo-linear model,  $X_{7,t}$ , where there is a mild inflation of the type I error. In the case of the Block Bootstrap for  $T = 100$  the performance depends on the size of the block. For  $B = 5$  and  $B = 10$  the type I error is below the nominal level, whereas for  $B = 20$  the type I error tends to be around and above the nominal level. However, when  $T = 500$  the block bootstrap is consistently below the nominal level for  $B = 5, 10$  and  $20$ . This suggests a larger block length should be used. It is quite possible that more accurate critical values, which are less sensitive to block length, can be obtained using the fixed- $b$  bootstrap. However, the fixed- $b$  bootstrap was not included in the study as the aim in this section is to compare different procedures which are simple and fast to implement using routines that already exist in R.

Model	Orthogonal $Q_T$		Regular $\tilde{Q}_T$		Robust $Q_T^*$		Block Bootstrap					
	5%	10%	5%	10%	5%	10%	B=5		B=10		B=20	
	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
Normal	6.52	11.1	4.1	8.34	5.42	10.18	0.0	0.1	1.14	4.36	5.14	11.94
$t_5$	6.34	11.42	4.08	8.46	4.92	10.5	0.0	0.08	0.96	4.12	4.74	10.70
$X_{3,t} = Z_t Z_{t-1}$	5.02	9.44	10.66	16.64	4.38	9.52	0.14	0.66	1.2	4.68	4.14	11.14
$X_{4,t}$	0.86	1.82	3.5	4.86	1.1	2.32	0.06	0.3	0.22	1.0	0.64	2.26
$X_{5,t}$	4.26	8.14	23.56	31.6	5.4	9.98	0.14	1.06	1.22	4.96	3.54	11.00
$X_{6,t}$	3.16	6.42	17.64	24.22	4.20	8.38	0.08	0.4	0.6	2.92	2.08	7.78
$X_{7,t}$	5.1	10.46	13.22	20.46	6.88	12.8	0.16	0.86	1.18	5	4.56	11.88
$X_{8,t}$	4.46	8.36	8.2	13.18	3.5	8.38	0.04	0.58	0.74	4	3.1	10.48

Table 2: Test for uncorrelatedness, under the null hypothesis,  $T = 100$  over 5000 replications.

### *Models under the alternative of correlation*

To assess the empirical power of the test we consider three different models. The first model is the Gaussian autoregressive process  $Y_{1,t}$ , where  $Y_{1,t} = -0.2Y_{1,t-1} + Z_t$ . The second model is  $Y_{2,t} = Y_{1,t}|U_{2,t}|$ , where  $\{Y_{1,t}\}$  and  $\{U_{2,t}\}$  are independent of each other,  $\{Y_{1,t}\}$  is defined above and  $\{U_{2,t}\}$  is the ARCH model defined in the previous section. Finally, the third model is  $Y_{3,t} = U_{3,t}|U_{2,t}|$ , where  $\{U_{2,t}\}$  and  $\{U_{3,t}\}$  are independent of each other,  $\{U_{2,t}\}$  is the ARCH model defined in the previous section and  $\{U_{3,t}\}$  is the Gaussian autoregressive process  $U_{3,t} = 0.5U_{3,t-1} + Z_t$ . We used the sample sizes  $T = 100$ ,  $T = 200$  and  $T = 500$ .

The result are given in Tables 4, 5 and 6. The power for most of the methods are relatively

Model	Orthogonal $Q_T$		Regular $\tilde{Q}_T$		Robust $Q_T^*$		Block Bootstrap					
	5%	10%	5%	10%	5%	10%	B=5		B=10		B=20	
	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
Normal	5.9	11.1	4.56	9.44	4.74	9.86	0.08	0.44	1.48	4.22	3.7	8.78
$t_5$	6.1	10.82	4.8	9.58	4.92	9.98	0.04	0.34	1.24	4.18	3.3	8.66
$X_{3,t} = Z_t Z_{t-1}$	5.00	9.82	15.26	22.26	4.9	9.16	0.7	2.52	2.5	6.56	3.6	9.16
$X_{4,t}$	1.0	1.86	30.52	38.56	5.6	11.16	1.56	4.48	2.72	7.94	3.1	10.02
$X_{5,t}$	3.76	7.06	49.86	58.76	4.82	9.34	1.06	4.10	2.08	6.9	2.52	7.96
$X_{6,t}$	2.88	6.22	42.48	50.46	3.64	7.24	0.64	2.4	1.24	4.68	1.5	6.62
$X_{7,t}$	4.48	8.88	20.38	28.32	6.02	11.46	0.78	2.58	1.52	5.04	2.78	7.5
$X_{8,t}$	5.28	9.46	15.08	20.46	4.12	8.54	1.28	3.82	3.04	7.36	4.34	9.92

Table 3: Test for uncorrelatedness, under the null hypothesis,  $T = 500$  over 5000 replications.

close. Though it is not surprising that the regular Box-Pierce statistic has the largest power, since it also has the largest inflated type I errors. Overall, in terms of power, the orthogonal sampling test and the robust Portmanteau test tend to have more power than the Block Bootstrap test, especially when the sample size is small.

Model	Orthogonal $Q_T$		Regular $\tilde{Q}_T$		Robust $Q_T^*$		Block Bootstrap					
	5%	10%	5%	10%	5%	10%	B=5		B=10		B=20	
	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
$Y_{1,t}$	27.06	38.88	29.52	40.84	28.60	40.50	2.62	9.4	14.12	29.34	25.74	43.22
$Y_{2,t}$	12.68	20.58	21.78	30.6	11.08	18.68	0.5	3.3	3.94	13.84	9.96	24.02
$Y_{3,t}$	55.7	68.6	71.94	79.58	61.8	72.32	18.36	41.46	41.62	65.14	52.86	74.30

Table 4: Test for uncorrelatedness, under the alternative hypothesis,  $T = 100$  over 5000 replications.

## 4 Selection of $M$

All the methods described above rely on a predetermined choice of  $M$ , which can influence the analysis. From Lemma 2.2, we observe if  $M$  is too large, then the variance of the orthogonal sample will not be the same which induces a bias. Therefore a “safe” option is to choose small  $M$  (which avoids the issues of bias). The down side of this approach is if  $M$  is small then the corresponding critical values will be large to account for the increased variance in the estimation of the nuisance parameters. As pointed out by a referee, Sun et al. [2008] and Sun [2014] propose a novel solution to the selection of tuning parameters (such as  $M$  or  $b$  defined in (2)) for “mean-type” estimators in the context of testing. Sun et al. [2008], Sun [2013] and Sun [2014] obtain an expression for the error in approximating the finite sample distribution of the test statistic with the corresponding pivotal distribution

Model	Orthogonal $Q_T$		Regular $\tilde{Q}_T$		Robust $Q_T^*$		Block Bootstrap					
	5%	10%	5%	10%	5%	10%	B=5		B=10		B=20	
	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
$Y_{1,t}$	54.70	67.64	58	69.06	55.78	67.18	18.58	36.16	41.10	59.84	48.64	66.54
$Y_{2,t}$	21.98	32.4	37.44	47.18	18.04	27.68	4.14	13.32	11.34	27.04	16.30	33.74
$Y_{t,3}$	87.04	92.64	95.72	97.28	85.84	91.22	74.5	87.96	84.42	93.74	84.54	94.58

Table 5: Test for uncorrelatedness under the alternative hypothesis,  $T = 200$  over 5000 replications.

Model	Orthogonal $Q_T$		Regular $\tilde{Q}_T$		Robust $Q_T^*$		Block Bootstrap					
	5%	10%	5%	10%	5%	10%	B=5		B=10		B=20	
	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
$Y_{1,2}$	94.86	97.44	95.94	97.84	95.24	97.60	82.30	91.74	92.96	96.48	93.76	97.28
$Y_{2,t}$	49.50	60.58	69.60	77.24	37	49.90	29.92	45.30	41.48	58.28	42.2	60.98
$Y_{3,t}$	98.86	99.36	99.94	99.98	98.5	99.16	99.2	99.68	98.98	99.82	97.94	99.82

Table 6: Test for uncorrelatedness under the alternative hypothesis,  $T = 500$  over 5000 replications.

(using an Edgeworth expansion). They use these results to approximate the type I error and the type II error for a certain local alternative. Based on these approximations they propose selecting the tuning parameter which minimises a convex sum of the type I and type II error. It is plausible that such an analysis can also be applied to the estimators defined (3) and (21). However, as the results in this paper are aimed at both constructing confidence intervals and testing, in this section we focus on selecting  $M$  that balances the bias and variance of the variance estimator.

We propose estimating the mean squared error for estimators of  $V(\omega_r)$  in the proximity of frequency zero. Our proposed method is based on the results derived in Section 2. Using Theorem 2.1 we note that  $\{\sqrt{T}A_T(\phi; r); 1 \leq r < T/2\}$  is an almost uncorrelated, zero mean sequence with variance  $\text{var}[\sqrt{T}A_T(\phi; r)] = V(\omega_r) + O(T^{-1})$  (where  $V(\omega_r)$  is defined in (9)). These observations together with (8) imply that

$$\left( \frac{|\sqrt{T}A_T(\phi; r)|^2}{V(\omega_r)} - 1 \right) \quad 1 \leq r < T/2, \quad (38)$$

is an almost uncorrelated sequence with mean zero and variance one. We use this sequence as the building blocks of the criterion. In order to select  $M$  we extend the estimator defined in (10) to all frequencies  $V(\omega_r)$ . More precisely, we use  $\hat{V}_M(\omega_r)$  as an estimator of  $V(\omega_r)$  where

$$\hat{V}_M(\omega_r) = \frac{T}{M} \sum_{s=1+r}^{M+r} |A_T(\phi; s)|^2,$$

noting that for  $r = 0$  we have the estimator defined in (10). Furthermore, by construction  $\widehat{V}_M(\omega_r)$  is asymptotically uncorrelated to  $|\sqrt{T}A_T(\phi; r)|^2$  (see Theorem 2.1). The suggested scheme is based on choosing the  $M$  which minimises the mean squared error of  $\{\widehat{V}_M(\omega_r); r = 1, \dots, T/p\}$ . This is analogous to bandwidth selection in nonparametric regression. Based on the above, we define the average squared error criterion

$$\mathcal{C}_\phi(M) = \frac{p}{T} \sum_{r=1}^{T/p} \left( \frac{|\sqrt{T}A_T(\phi; r)|^2}{\widehat{V}_M(\omega_r)} - 1 \right)^2.$$

To select  $M$  we use  $\widehat{M} = \arg \min_{M \in \mathcal{S}} \mathcal{C}_\phi(M)$ , where  $\mathcal{S}$  is the set over which we do the selection.

To understand what  $\mathcal{C}_\phi(M)$  is estimating and why it estimates the mean squared error, we make a Taylor expansion of  $\widehat{V}_M(\omega_r)^{-1}$  about  $\mathbb{E}[\widehat{V}_M(\omega_r)]^{-1}$  to give

$$\begin{aligned} & \left( \frac{|\sqrt{T}A_T(\phi; r)|^2}{\widehat{V}_M(\omega_r)} - 1 \right) \\ \approx & \frac{|\sqrt{T}A_T(\phi; r)|^2 - \mathbb{E}[|\sqrt{T}A_T(\phi; r)|^2]}{\mathbb{E}[\widehat{V}_M(\omega_r)]} - \frac{\mathbb{E}[|\sqrt{T}A_T(\phi; r)|^2]}{\mathbb{E}[\widehat{V}_M(\omega_r)]^2} \left( \widehat{V}_M(\omega_r) - \mathbb{E}[\widehat{V}_M(\omega_r)] \right) \\ & + \left( \frac{\mathbb{E}[|\sqrt{T}A_T(\phi; r)|^2]}{\mathbb{E}[\widehat{V}_M(\omega_r)]} - 1 \right). \end{aligned}$$

Taking the expectation squared of the above and using Lemma 2.1 and that  $|\sqrt{T}A_T(\phi; r)|^2$  and  $\widehat{V}_M(\omega_r)$  are asymptotically uncorrelated we have

$$\mathbb{E} \left( \frac{|\sqrt{T}A_T(\phi; r)|^2}{\widehat{V}_M(\omega_r)} - 1 \right)^2 \approx \frac{V(\omega_r)^2}{B_{r,M}^2} \underbrace{\text{var} \left( \widehat{V}_M(\omega_r) \right)}_{\text{variance}} + \underbrace{\left( \frac{V(\omega_r)}{\mathbb{E}[\widehat{V}_M(\omega_r)]} - 1 \right)^2}_{\text{bias}} + \frac{V(\omega_r)^2}{B_{r,M}}$$

where  $B_{r,M} = \mathbb{E} \left[ \widehat{V}_M(\omega_r) \right]^2$ . Substituting the above into  $\mathcal{C}_\phi(M)$  we observe that it is estimating

$$\sum_{r=1}^{T/p} \left\{ \frac{V(\omega_r)^2}{B_{r,M}^2} \text{var} \left[ \widehat{V}_M(\omega_r) \right] + \left( \frac{V(\omega_r)}{\mathbb{E}[\widehat{V}_M(\omega_r)]} - 1 \right)^2 + \frac{V(\omega_r)^2}{B_{r,M}} \right\}.$$

Thus  $\mathcal{C}_\phi(M)$  can be viewed as a linear combination of the bias and variance. If  $M$  is small, the bias is small but the variance is large, conversely if  $M$  large the bias is large but the variance is small. It seems reasonable to choose the  $M$  which balances these two terms. To illustrate how the criterion behaves, in Figure 4 we plot  $\mathcal{C}_\phi(M)$  over  $M$ , for the function  $\phi(\omega) = e^{i\omega}$  (which corresponds to the statistic which estimates the autocovariance at lag one). As expected  $\mathcal{C}_\phi(M)$  is large when  $M$  is both small and large. In this example, selecting  $M$  anywhere between 7 and 13 seems to be reasonable.

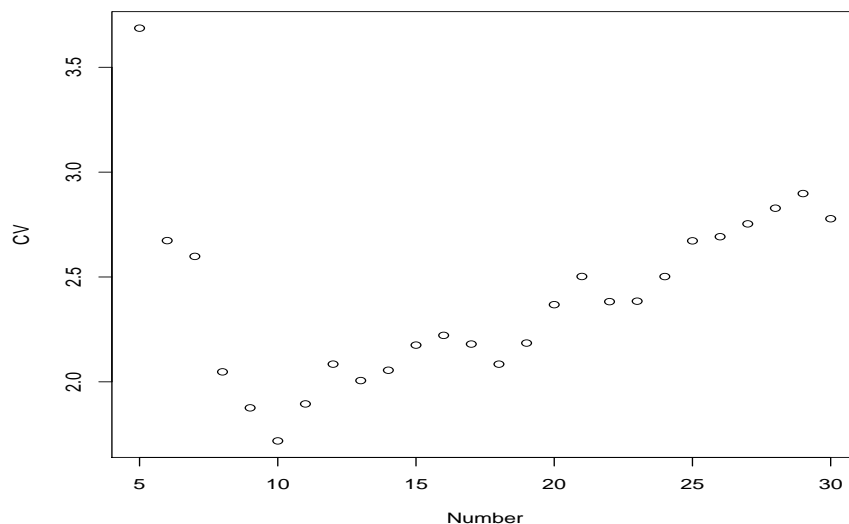


Figure 4: The average squared criterion for the sample autocovariance function  $\mathcal{C}_{e^i}(M)$  ( $p = 4$ ) at lag one for the Gaussian autoregressive time series  $X_t = 1.5X_{t-1} - 0.75X_{t-2} + \varepsilon_t$  where  $T = 200$ .

## 5 Concluding Remarks

In this paper we have introduced the method of orthogonal samples for estimating nuisance parameters for a broad class of statistics in time series. The objective of this scheme is not to “compete” with the many excellent methods that exist in the literature, but to consider the problem of nuisance parameter estimation from an alternative perspective, that can be used to develop a method which is extremely fast to implement and still gives reliable results. In future work our aim is combine orthogonal samples with the block bootstrap to construct a pivotal statistic. The hope is that by bootstrapping the pivotal statistic the bootstrap distribution can better approximate the pivotal statistic than the original statistic.

### Dedication

This paper is dedicated to the memory of Professor Emanuel Parzen who was an exceptional scientist and a wonderfully kind colleague. Without his fundamental contributions to time series, in particular his work on spectral analysis, this paper would not have been possible.

### Acknowledgements

This work has been partially supported by the National Science Foundation, DMS-1513647. SSR would like to thank Xianyang Zhang for several useful discussions. SSR would also like to thank the guest Editor Dimitris Politis and an anonymous referee for making several useful suggestions that improved the paper.

## References

- M.S. Bartlett. Periodogram analysis and continuous spectra. *Biometrika*, 37:1–16, 1950.
- A. K. Bera and M. L. Higgins. ARCH models: Properties, estimation and testing. *Journal of Economic Surveys*, 7:305–362, 1993.
- D.R. Brillinger. *Time Series, Data Analysis and Theory*. SIAM, San Francisco, 1981.
- P. J. Brockwell and R. A. Davis. *Time Series: Theory and Methods*. Springer, New York, 1991.
- W.W. Chen and R. S. Deo. Power transformations to induce normality and their applications. *Journal of the Royal Statistical Society (B)*, 66:117–130, 2004.
- R. Dahlhaus and D. Janas. A frequency domain bootstrap for ratio statistics in time series analysis. *Annals of Statistics*, 24:1934–1963, 1996.
- H. Dette and E. Paparoditis. Bootstrapping frequency domain tests in multivariate time series with an application to comparing spectral densities. *J. Royal Statistical Society (B)*, 71:831–857, 2009.
- F. Diebold. Testing for serial correlation in the presence of heteroskedasticity. *Proceedings of the American Statistical Association, Business and Economics Statistics Section*, pages 323–328, 1986.
- M. Eichler. Testing nonparametric and semi-parametric hypothesis in vector stationary processes. *J. Multivariate Statistics*, 99:968–1009, 2008.
- J. C. Escanciano and I. N Lobato. An automatic Portmanteau test for serial correlation. *Journal of Econometrics*, 151:140–149, 2009.
- J. Franke and W. Härdle. On bootstrapping kernel spectral estimates. *Ann. Statist.*, 20:121–145, 1992.
- Y. Hong. Consistent testing for serial correlation on unknown form. *Econometrica*, 64:837–864, 1996.
- C. Hurvich and S. Zeger. Frequency domain bootstrap methods for time series. *New York University, Graduate School of Business Administration*, 1987.
- C. Jentsch and D.N. Politis. Valid resampling of higher order statistics using linear process bootstrap and autoregressive bootstrap. *Communication in Statistics: Theory and Methods*, 42:1277–1293, 2013.

- C. Jentsch and D.N. Politis. Covariance matrix estimation and linear process bootstrap for multivariate time series of possibly increasing dimension. *Annals of Statistics*, 43: 1117–1140, 2015.
- N. M. Kiefer and T.J. Vogelsang. A new asymptotic theory for heteroskedasticity-autocorrelation robust tests. *Econometric Theory*, 21:1130–1164, 2005.
- C. Kirch and D. N. Politis. Tft-bootstrap: Resampling time series in the frequency domain to obtain replicates in the time domain. *Annals of Statistics*, 39:1427–1440, 2011.
- J. Krampe, J.-P. Kreiss, and E. Paparoditis. Estimated wold representation and spectral density driven bootstrap for time series. *Preprint*, 2016.
- J.-P. Kreiss. *Bootstrapping and Related Techniques*, chapter Bootstrap procedures for AR( $\infty$ )-processes, pages 107–113. Springer, 1992.
- J-P. Kreiss and S. N. Lahiri. *Handbook of Statistics, volume 30*, chapter Bootstrap methods for time series, pages 3–26. Elsevier, 2012.
- J.-P. Kreiss, E. Paparoditis, and D. N. Politis. On the range and validity of the autoregressive sieve bootstrap. *Annals of Statistics*, 39:2103–2130, 2010.
- H. R. Künsch. The jackknife and the bootstrap for general stationary observations. *Annals of Statistics*, 17:1217–1241, 1989.
- S. N. Lahiri. *Resampling Methods for Dependent Data*. Springer, New York, 2003.
- J. Lee and S. Subba Rao. A note on quadratic forms of nonstationary processes. *Under revision*, 2016.
- I. N. Lobato. Testing that a dependent process is uncorrelated. *Journal of the American Statistical Association*, 96:1066–1076, 2001.
- I.N. Lobato, J.C. Nankervis, and N.E. Savin. Testing for zero autocorrelation in the presence of statistical dependence. *Econometric Theory*, 18:730–743, 2002.
- T.L. McMurray and D.N. Politis. Banded and tapered estimates for autocovariance matrices and the linear process bootstrap. *Journal of Time Series Analysis*, 31:471–482, 2010.
- A. Milhoj. A test of fit in time series. *Biometrika*, 68:177–188, 1981.
- U. Müller. A theory of robust long-run variance estimation. *Journal of Econometrics*, 141: 1331–1352, 2007.

- U. Müller and M. Watson. Low frequency econometrics. *Preprint*, 2016.
- M. Neumann. Spectral density estimation via nonlinear wavelet methods for stationary nongaussian time series. *J. Time Series Analysis*, 1996.
- E. Parzen. On consistent estimates of the spectrum of a stationary time series. *Annals of Mathematical Statistics*, 28:329–348, 1957.
- D. N. Politis, J. P. Romano, and M. Wold. Subsampling for heteroskedastic time series. *Journal of Econometrics*, 81:281–318, 1997.
- D. N. Politis, J. P. Romano, and M. Wolf. *Subsampling*. Springer, New York, 1999.
- D.N. Politis and J. P. Romano. The stationary bootstrap. *Journal of the American Statistical Association*, 89:1303–1313, 1994.
- R. M. Robinson. Testing for strong serial correlation and dynamic conditional heteroskedasticity in multiple regression. *Journal of Econometrics*, 47:67–84, 1991.
- J. P. Romano and L. A. Thombs. Inference for autocorrelations under weak assumptions. *Journal of the American Statistical Association*, 91:590–600, 1996.
- X. Shao. Confidence intervals for spectral mean and ratio statistics. *Biometrika*, 96:107–117, 2009.
- X. Shao and D.N. Politis. Fixed b-subsampling and block bootstrap: improved confidence intervals based on p-value calibration. *Journal of the Royal Statistical Society (B)*, 75:323–343, 2013.
- V. Statulevicius and Jakimavicius. Estimates of semiinvariant and centered moments of stochastic processes with mixing: I. *Lithuanian Math. J.*, 28:226–238, 1988.
- Y. Sun. Robust trend influence with series variance estimator and testing-optimal smoothing parameter. *Journal of Econometrics*, 164:345–366, 2011.
- Y. Sun. A heteroskedasticity and autocorrelation robust F test using orthonormal series variance estimator. *Econometric Journal*, 1:1–26, 2013.
- Y. Sun. Fixed smoothing asymptotics in a two-set generalized method of moments framework. *Econometrica*, 82:2327–2370, 2014.
- Y. Sun, P.C.B. Phillips, and S. Jin. Optimal bandwidth selection in heteroskedasticity-autocorrelation robust testing. *Econometrica*, 76:175–194, 2008.



G. R. Terrell. The Wilson-Hiferty transform is locally saddlepoint. *Biometrika*, 57:445–453, 2003.

A. Weiss. ARCH and Bilinear time series models. *Journal of Business and Economic Statistics*, 4:69–70, 1986.

## Supplementary material

To prove the results we make heavy use of the following well known result. Suppose that the time series  $\{X_t\}$  satisfies Assumption 2.1 with  $p = s$ . Then we have

$$\text{cum}[J_T(\omega_{k_1}), \dots, J_T(\omega_{k_s})] = \frac{(2\pi)^{s/2-1}}{T^{s/2-1}} I_{k_1+\dots+k_s \in T\mathbb{Z}} f_s(\omega_{k_1}, \dots, \omega_{k_{s-1}}) + O(T^{-s/2}), \quad (39)$$

see Brillinger [1981], Theorem 4.3.2, for the details.

### A Proofs for Section 2

**PROOF of Lemma 2.1** This immediately follows from (39) and the Lipschitz continuity of  $f$  and  $\phi$  which allows the sum to be replaced by an integral.  $\square$

**PROOF of Theorem 2.1(i)** We first derive expressions for  $\text{cov}[A_T(\phi; r_1), A_T(\phi; r_2)]$  and  $\text{cov}[A_T(\phi; r_1), \overline{A_T(\phi; r_2)}]$  from which we can deduce the covariance of  $\Re A_T(\phi; r)$  and  $\Im A_T(\phi; r)$ .

To simplify notation we denote  $J_k = J_T(\omega_k)$ . By using indecomposable partitions and (39) we have

$$\begin{aligned} & T \text{cov}[A_T(\phi; r_1), A_T(\phi; r_2)] \\ &= \frac{1}{T} \sum_{k_1, k_2=1}^T \phi(\omega_{k_1}) \overline{\phi(\omega_{k_2})} \text{cov}[J_{k_1} \overline{J_{k_1+r_1}}, J_{k_2} \overline{J_{k_2+r_2}}] \\ &= \frac{1}{T} \sum_{k_1, k_2=1}^T \phi(\omega_{k_1}) \overline{\phi(\omega_{k_2})} \left( \text{cov}[J_{k_1}, J_{k_2}] \text{cov}[\overline{J_{k_1+r_1}}, \overline{J_{k_2+r_2}}] \right. \\ &\quad \left. + \text{cov}[J_{k_1}, \overline{J_{k_2+r_2}}] \text{cov}[\overline{J_{k_1+r_1}}, J_{k_2}] + \text{cum}[J_{k_1}, \overline{J_{k_1+r_1}}, \overline{J_{k_2}}, J_{k_2+r_2}] \right) \\ &= \frac{1}{T} \sum_{k=1}^T \phi(\omega_k) \overline{\phi(\omega_k)} f(\omega_k) f(\omega_{k+r_1}) \delta_{r_1=r_2} + \frac{1}{T} \sum_{k=1}^T \phi(\omega_k) \overline{\phi(-\omega_{k+r})} f(\omega_k) f(\omega_{k+r_1}) \delta_{r_1=r_2} \\ &\quad + \frac{2\pi}{T^2} \sum_{k_1, k_2=1}^T \phi(\omega_{k_1}) \overline{\phi(\omega_{k_2})} f_4(\omega_{k_1}, -\omega_{k_1+r}, -\omega_{k_2}) \delta_{r_1, r_2} \Big) + O(T^{-1}). \end{aligned}$$

Thus we see that if  $r_1 \neq r_2$ , then  $|T \text{cov}[A_T(\phi; r_1), A_T(\phi; r_2)]| = O(T^{-1})$ . On the other hand if  $r_1 = r_2$  we replace the sum above with an integral to give  $T \text{var}[A_T(\phi; r_1)] = V(\omega_r) + O(T^{-1})$ .

We apply the same arguments to  $T \text{cov}[A_T(\phi; r_1), \overline{A_T(\phi; r_2)}]$  to give

$$\begin{aligned}
& T \text{cov}[A_T(\phi; r_1), \overline{A_T(\phi; r_2)}] \\
&= \frac{1}{T} \sum_{k_1, k_2=1}^T \phi(\omega_{k_1}) \phi(\omega_{k_2}) \text{cov}[J_{k_1} \overline{J_{k_1+r_1}}, \overline{J_{k_2} J_{k_2+r_1}}] \\
&= \frac{1}{T} \sum_{k_1, k_2=1}^T \phi(\omega_{k_1}) \phi(\omega_{k_2}) \left( \text{cov}[J_{k_1}, \overline{J_{k_2}}] \text{cov}[\overline{J_{k_1+r_1}}, J_{k_2+r_2}] \right. \\
&\quad \left. + \text{cov}[J_{k_1}, \overline{J_{k_2+r_2}}] \text{cov}[\overline{J_{k_1+r_1}}, J_{k_2+r_2}] + \text{cum}[J_{k_1}, \overline{J_{k_1+r_1}}, \overline{J_{k_2}}, J_{k_2+r_2}] \right) \\
&= \frac{1}{T} \sum_{k=1}^T \phi(\omega_k) \phi(-\omega_k) f(\omega_k) f(\omega_{k+r_1}) \delta_{r_1=-r_2 \text{ or } T-r_2} + \frac{1}{T} \sum_{k=1}^T \phi(\omega_k) \phi(-\omega_{k+r_1}) f(\omega_k) f(\omega_{k+r_1}) \delta_{r_1=-r_2 \text{ or } T-r_2} \\
&\quad + \frac{2\pi}{T^2} \sum_{k_1, k_2=1}^T \phi(\omega_{k_1}) \phi(\omega_{k_2}) f_4(\omega_{k_1}, -\omega_{k_1+r_1}, -\omega_{k_2}) \delta_{r_1=-r_2 \text{ or } T-r_2} \Big) + O(T^{-1}).
\end{aligned}$$

Since  $0 < r_1, r_2 < T/2$ , the above implies that  $T \text{cov}[A_T(\phi; r_1), \overline{A_T(\phi; r_2)}] = O(T^{-1})$ .

Finally, by using the identities  $\Re A_T(\phi; r) = \frac{1}{2}(A_T(\phi; r) + \overline{A_T(\phi; r)})$  and  $\Im A_T(\phi; r) = \frac{-i}{2}(A_T(\phi; r) - \overline{A_T(\phi; r)})$  and the above the result immediately follows.  $\square$

In order to prove Theorem 2.1(ii) we require the following result.

**Lemma A.1** *Suppose that the time series  $\{X_t\}$  satisfies Assumption 2.1 with  $p = 2n$ . Then we have*

$$\text{cum}[\sqrt{T} A_T(\phi; r_1), \dots, \sqrt{T} A_T(\phi; r_n)] = O\left(\frac{1}{T^{n/2-1}}\right).$$

**PROOF** Bu using indecomposable partitions and (39) the result immediately follows.  $\square$

We make use of Lemma A.1 below.

**PROOF of Theorem 2.1(ii)** To prove (8) we note that

$$\begin{aligned}
& \text{cov}[|\sqrt{T} A_T(\phi; r_1)|^2, |\sqrt{T} A_T(\phi; r_2)|^2] = \left| \text{cov}[\sqrt{T} A_T(\phi; r_1), \sqrt{T} A_T(\phi; r_2)] \right|^2 \\
& + \left| \text{cov}[\sqrt{T} A_T(\phi; r_1), \sqrt{T} \overline{A_T(\phi; r_2)}] \right|^2 + T^2 \text{cum}_4 \left( A_T(\phi; r_1), \overline{A_T(\phi; r_1)}, A_T(\phi; r_2), \overline{A_T(\phi; r_2)} \right).
\end{aligned}$$

Thus we see that (8) follows immediately from the above, Theorem 2.1 and Lemma 2.2.  $\square$

**PROOF of Lemma 2.2** By using Corollary 2.1 we can show that

$$|\mathbb{E}[\widehat{V}_M(0)] - V(0)| = O(M/T). \quad (40)$$

To prove (17) we use the classical variance bias decomposition

$$\mathbb{E} \left( \widehat{V}_M(0) - V(0) \right)^2 = \text{var}[\widehat{V}_M(0)] + \left[ \mathbb{E}[\widehat{V}_M(0)] - V(0) \right]^2.$$

To bound  $\text{var}[\widehat{V}_M(0)]$  we note that

$$\begin{aligned}\text{var}[\widehat{V}_M(0)] &= \frac{1}{M^2} \sum_{r_1, r_2=1}^M \left\{ \left| \text{cov}[\sqrt{T}A_T(\phi; r_1), \sqrt{T}A_T(\phi; r_2)] \right|^2 + \left| \text{cov}[\sqrt{T}A_T(\phi; r_1), \sqrt{T}A_T(\phi; r_2)] \right|^2 \right. \\ &\quad \left. + \text{cum} \left( \sqrt{T}A_T(\phi; r_1), \sqrt{T}A_T(\phi; r_1), \sqrt{T}A_T(\phi; r_2), \sqrt{T}A_T(\phi; r_2) \right) \right\} \\ &= O(M^{-1} + T^{-1}),\end{aligned}$$

where the last line follows immediately from (8). Altogether by using the above and (40), we obtain the desired result.  $\square$

**PROOF of Lemma 2.3** By making a Taylor expansion of  $A_T(\phi_{\widehat{\theta}}; r)$  about  $\phi_{\theta}$  we have

$$A_T(\phi_{\widehat{\theta}}; r) = A_T(\phi_{\theta}; r) + (\widehat{\theta}_T - \theta) A_T(\nabla_{\theta} \phi_{\theta}; r) + (\widehat{\theta}_T - \theta)^2 \frac{1}{T} \sum_{k=1}^T \nabla_{\theta}^2 \phi(\omega_k; \theta) \Big|_{\theta=\bar{\theta}_k} J_T(\omega_k) \overline{J_T(\omega_{k+r})},$$

where  $\bar{\theta}_k$  lies between  $\theta$  and  $\widehat{\theta}$ . Thus

$$\left| A_T(\phi_{\widehat{\theta}}; r) - A_T(\phi_{\theta}; r) - (\widehat{\theta}_T - \theta) A_T(\nabla_{\theta} \phi_{\theta}; r) \right| = O_p(T^{-2}).$$

Therefore

$$\begin{aligned}\widehat{V}_{\widehat{\theta}, M}(0) &= \frac{T}{M} \sum_{r=1}^M |A_T(\phi_{\theta}; r)|^2 + \underbrace{|\widehat{\theta}_T - \theta|^2}_{O_p(T^{-1})} \underbrace{\frac{T}{M} \sum_{r=1}^M |A_T(\nabla_{\theta} \phi_{\theta}; r)|^2}_{=O_p(1), \text{ by Lemma 2.2}} + O_p(T^{-1}) \\ &= \frac{T}{M} \sum_{r=1}^M |A_T(\phi_{\theta}; r)|^2 + O_p(T^{-1}).\end{aligned}\tag{41}$$

By using Lemma 2.2 we have

$$\mathbb{E} \left\{ \frac{T}{M} \sum_{r=1}^M |A_T(\phi_{\theta}; r)|^2 - V_{\theta}(0) \right\}^2 = O(M^{-1} + M/T).$$

Therefore, by using the above and (41) we obtain the result.  $\square$

## B Proofs for Section 3.1

To prove Theorem 3.1 we use the following definition

$$\begin{aligned}V_{j_1, j_2}(\omega_r) &= \frac{1}{2\pi} \int_0^{2\pi} f(\omega) f(\omega + \omega_r) \left( \phi_{j_1}(\omega) \overline{\phi_{j_2}(\omega)} + \phi_{j_1}(\omega) \overline{\phi_{j_2}(-\omega - \omega_r)} \right) d\omega + \\ &\quad \frac{1}{2\pi} \int_0^{2\pi} \int_0^{2\pi} \phi_{j_1}(\omega_1) \overline{\phi_{j_2}(\omega_2)} f_4(\omega_1, -\omega_1 - \omega_r, -\omega_2) d\omega_1 d\omega_2 + O(T^{-1}).\end{aligned}$$

The following lemma facilitates the proof of Theorem 3.1.

**Lemma B.1** *Suppose that the time series  $\{X_t\}$  satisfies Assumption 2.1 with  $p = 16$ . Then we have*

$$\begin{aligned} T\text{cov}[A_T(\phi_{j_1}; 0), A_T(\phi_{j_2}; 0)] &= V_{j_1, j_2} + O(T^{-1}) \\ T\text{cov}\left[A_T(\phi_{j_1}; 0), \overline{A_T(\phi_{j_2}; 0)}\right] &= V_{j_1, j_2} + O(T^{-1}). \end{aligned} \quad (42)$$

For all  $0 < r_1, r_2 < T/2$  we have

$$\begin{aligned} T\text{cov}[A_T(\phi_{j_1}; r_1), A_T(\phi_{j_2}; r_2)] &= \begin{cases} V_{j_1, j_2}(\omega_r) + O(T^{-1}) & r_1 = r_2 \\ O(T^{-1}) & r_1 \neq r_2 \end{cases} \\ T\text{cov}\left[A_T(\phi_{j_1}; r_1), \overline{A_T(\phi_{j_2}; r_2)}\right] &= O(T^{-1}). \end{aligned} \quad (43)$$

For all  $0 \leq r_1, r_2, r_3, r_4 < T/2$  we have

$$T^2\text{cum}[A_T(\phi_{j_1}; r_1), A_T(\phi_{j_2}; r_2), A_T(\phi_{j_3}; r_3), A_T(\phi_{j_4}; r_4)] = O(T^{-1}). \quad (44)$$

For  $0 < r < T/2$  and  $0 < r_1, r_2 < T/2$  we have

$$T^2\text{cov}[A_T(\phi_{j_1}; r_1)^2, A_T(\phi_{j_2}; r_2)^2] = \begin{cases} 2V_{j_1, j_2}(\omega_r)^2 + O(T^{-1}) & r_1 = r_2 (= r) \\ O(T^{-1}) & r_1 \neq r_2 \end{cases} \quad (45)$$

$$T^2\text{cov}[A_T(\phi_{j_1}; r_1)^2, A_T(\phi_{j_2}; r_2)\overline{A_T(\phi_{j_2}; r_2)}] = O(T^{-1}) \quad (46)$$

$$T^2\text{cov}[|A_T(\phi_{j_1}; r_1)|^2, |A_T(\phi_{j_2}; r_2)|^2] = \begin{cases} V_{j_1, j_2}(\omega_r)^2 + O(T^{-1}) \\ O(T^{-1}) \end{cases} \quad (47)$$

Finally

$$|V_{j_1, j_2}(\omega_r) - V_{j_1, j_2}| \leq K|r|T^{-1}, \quad (48)$$

where  $K$  is a finite constant.

**PROOF** The proof of (42) and (43) is identical to the proof of Theorem 2.1, thus we omit the details. The proof of (44) follows from Lemma A.1, thus we omit the details. To prove (45) we use indecomposable partitions to decompose it in terms of the product of covariances and a fourth order cumulant term. Specifically

$$\begin{aligned} &T^2\text{cov}[A_T(\phi_{j_1}; r)^2, A_T(\phi_{j_2}; r_2)^2] \\ &= 2T^2|\text{cov}[A_T(\phi_{j_1}; r_1), A_T(\phi_{j_1}; r_2)]|^2 + T^2\text{cum}[A_T(\phi_{j_1}; r_1), A_T(\phi_{j_1}; r_1), \overline{A_T(\phi_{j_1}; r_2)}, \overline{A_T(\phi_{j_1}; r_2)}]. \end{aligned}$$

By using (43) we obtain (45). A similar proof applies to (46) and (47).

Finally, to prove (48) we simply use the Lipschitz continuity of  $f$ ,  $f_4$  and  $\phi_j$ . Thus we have proved the result.  $\square$

**PROOF of Theorem 3.1** To prove (i), we note that under both the null and alternative the following expansion is valid

$$\begin{aligned} \mathbb{E}[S_T] &= T \sum_{j=1}^L \mathbb{E} |(A_T(\phi_j) - \mathbb{E}[A_T(\phi_j)] + \mathbb{E}[A_T(\phi_j)])|^2 \\ &= T \sum_{j=1}^L \text{var}[A_T(\phi_j)] + T \sum_{j=1}^L |\mathbb{E}[A_T(\phi_j)]|^2. \end{aligned} \quad (49)$$

Using that under the null  $\mathbb{E}[A_T(\phi_j)] = O(T^{-1})$ , and substituting this into the above we have

$$\mathbb{E}[S_T] = \sum_{j=1}^L \text{var}[\sqrt{T}A_T(\phi_j)] + O(T^{-1}) = \sum_{j=1}^L V_{j,j} + O(T^{-1}),$$

where the last line follows from (43). This gives (ia). To prove (ib) we note that

$$\mathbb{E}[S_{T,R}(r)] = \frac{1}{2} \sum_{j=1}^L \text{var} \left[ \sqrt{T} \left( A_T(\phi_j; r) + \overline{A_T(\phi_j; r)} \right) \right] + \frac{1}{2} \sum_{j=1}^L \left| \sqrt{T} \mathbb{E} \left( A_T(\phi_j; r) + \overline{A_T(\phi_j; r)} \right) \right|^2.$$

Under both the null and alternative  $\mathbb{E}[A_T(\phi_j; r)] = O(T^{-1})$  for  $0 < r < T/2$ . Thus

$$\begin{aligned} \mathbb{E}[S_{T,R}(r)] &= \frac{1}{2} \sum_{j=1}^L \left( 2\text{var}[\sqrt{T}A_T(\phi_j; r)] + 2\Re\text{cov}[\sqrt{T}A_T(\phi_j; r), \overline{\sqrt{T}A_T(\phi_j; r)}] \right) + O(T^{-1}) \\ &= \sum_{j=1}^L V_{j,j} + O(T^{-1}), \end{aligned}$$

thus proving (ib).

To prove (iia) we note that since  $A_T(\phi_j)$  is real and under the null  $\mathbb{E}[A_T(\phi_j)] = O(T^{-1})$ , then expanding  $\text{var}[S_T]$  gives

$$\begin{aligned} \text{var}[S_T] &= \sum_{j_1, j_2=1}^L \text{cov} \left( |\sqrt{T}A_T(\phi_{j_1})|^2, |\sqrt{T}A_T(\phi_{j_2})|^2 \right) \\ &= \sum_{j_1, j_2=1}^L \left( 2\text{cov} \left[ \sqrt{T}A_T(\phi_{j_1}), \sqrt{T}A_T(\phi_{j_2}) \right]^2 + \right. \\ &\quad \left. \text{cum} \left[ \sqrt{T}A_T(\phi_{j_1}), \sqrt{T}A_T(\phi_{j_1}), \sqrt{T}A_T(\phi_{j_2}), \sqrt{T}A_T(\phi_{j_2}) \right] \right) \\ &= 2 \sum_{j_1, j_2=1}^L V_{j_1, j_2}^2 + O(T^{-1}), \end{aligned}$$

where the last line follows from (43) and (44). Thus proving (iia).

We now prove (iib), where we derive an expression for  $\text{cov}[|\sqrt{T}\Re A_T(\phi_{j_1}; r)|^2, |\sqrt{T}\Re A_T(\phi_{j_2}; r)|^2]$ . To simplify notation let  $A_T(r) = \sqrt{T}A_T(\phi_j; r)$ . Using this notation we write  $\sqrt{T}\Re A_T(\phi_j; r) = \frac{1}{2}(A_T(r) + \overline{A_T(r)})$  and

$$|\sqrt{T}\Re A_T(\phi_j; r)|^2 = \frac{1}{4} \left( A_T(r)^2 + A_T(r)\overline{A_T(r)} + \overline{A_T(r)}A_T(r) + \overline{A_T(r)}^2 \right).$$

Thus

$$\begin{aligned} & \text{cov}[|\sqrt{T}\Re A_T(\phi_j; r_1)|^2, |\sqrt{T}\Re A_T(\phi_j; r_2)|^2] \\ &= \frac{1}{16} \text{cov} \left[ A_T(r_1)^2 + A_T(r_1)\overline{A_T(r_1)} + \overline{A_T(r_1)}A_T(r_1) + \overline{A_T(r_1)}^2, \right. \\ & \quad \left. A_T(r_2)^2 + A_T(r_2)\overline{A_T(r_2)} + \overline{A_T(r_2)}A_T(r_2) + \overline{A_T(r_2)}^2 \right]. \end{aligned}$$

Thus by using (45)-(47) we have

$$\text{cov}[|\sqrt{T}\Re A_T(\phi_j; r_1)|^2, |\sqrt{T}\Re A_T(\phi_j; r_2)|^2] = \begin{cases} \frac{1}{2} \sum_{j_1, j_2=1}^L V_{j_1, j_2}(\omega_r) + O(T^{-1}) & r_1 = r_2 \\ O(T^{-1}) & r_1 \neq r_2 \end{cases}$$

Now we recall that  $S_{T,R}(r) = 2 \sum_{j=1}^L |\sqrt{T}\Re A_T(\phi_j; r)|^2$ , thus by using the above and (48) we have

$$\text{cov}[S_{T,R}(r_1), S_{T,R}(r_2)] = \begin{cases} 2 \sum_{j_1, j_2=1}^L V_{j_1, j_2} + O(T^{-1} + |r|T^{-1}) & r_1 = r_2 (= r) \\ O(T^{-1}) & r_1 \neq r_2. \end{cases}$$

The same arguments apply to  $\text{cov}[S_{T,R}(r_1), S_{T,R}(r_2)]$  and  $\text{cov}[S_{T,R}(r_1), S_{T,I}(r_2)]$ , which gives us (iib).

To prove (iii) we use the same method used to prove (ii) together with Lemma A.1.  $\square$