Proxy samples for time series

Statistical inference for time series usually involves nuisance parameters which are difficult to estimate. In such situations, the general approach is to use a bootstrap scheme designed for time series to estimate the nuisance and/or the finite sample distribution function of the statistics of interest. Bootstrap schemes designed for time series (such as the block bootstrap) are simple to implement and have interesting theoretical properties, However, in practice, many of these bootstrap schemes can be computationally quite cumbersome and involve a user-chosen tuning parameter which can heavily influence the results.

In this paper we propose the method of proxy samples, which can be used to estimate nuisance parameters and the sampling distribution of certain test statistics. For a broad class of statistics, using the original data, a proxy sample is constructed by making a slight modification of the original statistic. The proxy sample is such that it is almost uncorrelated, furthermore it is almost uncorrelated to the original statistic. The most important feature is that it shares similar distributional properties to the centralised statistic of interest. These properties mean that it can be used to estimate the variance of the statistic. Furthermore, by using asymptotic normality of both the estimator and the corresponding proxy sample, we show that a confidence interval for the parameter of interest can easily be constructed using the standard t-distribution. Since the centralized statistic and the proxy sample share similar distribution properties we use the proxy sample to estimate the finite sample distribution of L_2-type test statistics. We show consistency of our methodology and illustrate the method with some simulations.

Proxy samples for time series

Suhasini Subba Rao
Department of Statistics, Texas A&M University
College Station, TX, U.S.A.
suhasini@stat.tamu.edu

May 29, 2015

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 proxy samples, which can be used to address some of these issues. For a broad class of statistics, using the original data, a proxy sample is constructed by making a slight modification of the original statistic, such that it shares similar distributional properties as the centralised statistic of interest. We use the proxy sample to estimate nuisance parameters and the finite sample distribution of the test statistics, focussing on Portmanteau and Goodness of Fit tests. The proposed method is simple and computationally fast to implement.

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

1 Introduction

A well known source of irritation for a time series analysist is that inference often depends on nuisance parameters which are not straightforward to estimate. The simplest example is the variance of the sample mean of a stationary time series, which is the sum of its autocovariance function. The problems mount if more sophisticated statistics are considered, such as the commonly used sample autocovariance function, spectral density estimators and quasi-Gaussian likelihood/Whittle likelihood estimators. All these estimators come under the canopy of integrated periodogram statistics which have the form

$$A_T(\phi) = \frac{1}{T}\sum_{k=1}^{T}\phi(\omega_k)|J_T(\omega_k)|^2, \quad \text{where} \quad J_T(\omega_k) = \frac{1}{\sqrt{2\pi T}}\sum_{t=1}^{T}X_t\exp(it\omega_k) \qquad (1)$$

with $\omega_k = \frac{2\pi k}{T}$. Under stationarity and some additional mixing-type and regularity conditions it is known that $T\mathrm{var}[A_T(\phi)] = V + O(T^{-1})$, where

$$V = \frac{1}{2\pi}\int_0^{2\pi} f(\omega)^2| \left(|\phi(\omega)|^2 + \phi(\omega)\overline{\phi(-\omega)}\right) d\omega + \frac{1}{(2\pi)^2}\int_0^{2\pi}\int_0^{2\pi} \phi(\omega_1)\overline{\phi(\omega_2)} f_4(\omega_1, -\omega_1, \omega_2) d\omega_1 d\omega_2. \tag{2}$$

Clearly this term is unwieldy and difficult to estimate. Some simplifications can be made under linearity of the time series. In particular, the fourth order cumulant term can be simplified (and in the special case of the Whittle likelihood this term is zero). However, for general nonlinear time series not even these simplifications are possible.

Several methods have been proposed to circumvent the need to estimate V. These include the method of self-normalisation proposed in Lobato [2001], Shao [2009] and Shao [2010a], where the limiting distribution is non-standard but free of nuisance parameters. The Haar-Fisz transform is a different type of self-normalisation, proposed in Fryzlewicz and Nason [2004] and Fryzlewicz and Nason [2006] which standardises the data in such a way that it is close to pivotal in its application. Alternatively, Dahlhaus and Janas [1996] propose transforming $A_T(\phi)$ into a 'ratio statistic'. The asymptotic variance of the ratio-statistic does not contain the fourth order cumulant f_4, which allows the spectral bootstrap (see Hurvich and Zeger [1987] and Franke and Härdle [1992]) to be employed to estimate the finite sample distribution of the statistic. The spectral bootstrap has the advantage that it can exploit the Fast Fourier Transform and is simple to use. Unfortunately this method only holds for univariate linear time series and even in the case of linearity, a multivariate extension is not possible. Possibly the most popular and widely used method for estimating finite sample distributions and nuisance parameters in time series are the block-based bootstrap, resampling methods, developed, for example, in Künsch [1989], Politis and Romano [1994], Romano and Thombs [1996], Politis et al. [1999], Lahiri [2003], Kirch and Politis [2011], Shao [2010b] and Kreiss and Lahiri [2012]. These methods avoid many of the issues mentioned above, but are computationally quite intensive, and usually require the selection of a user-chosen block length, where differing block lengths could lead to potentially different conclusions. The purpose of this paper is to propose an alternative method to the estimation of nuisance parameters, such as V, and the sampling distribution of a test statistic. The proposed method is simple and computationally fast, and will hopefully, complement and add to the existing arsenal of methods.

To motivate our approach let us consider two toy examples. In the first example, we suppose $\{X_t\}$ are independent, identically distributed (iid) random variables with mean μ and variance σ^2. The classical estimator of the variance of the sample mean, $\sqrt{T}\bar{X}$, where $\bar{X}_T = T^{-1}\sum_{t=1}^{T} X_t$ is the sample variance, which we now derive from an alternative perspective. To do this, we define the T dimensional vector $\underline{X} = (X_1, \ldots, X_T)$ and the T-

dimensional orthonormal vectors $\{\underline{e}_{j,T}; 1 \leq j \leq M\}$, whose elements sum to zero and squared elements sum to one (examples include the Discrete Fourier transform, the Discrete Wavelet transform and the Walsh transform). By construction, the mean of the transformed data $\underline{e}'_{j,T}\underline{X}$ is zero. However, straightforward calculations show that its variance is σ^2, furthermore $\{\underline{e}'_{j,T}\underline{X}\}$ is an uncorrelated sequence. Therefore, the sequence $\{\underline{e}'_{j,T}\underline{X}\}$ can be considered as a *centralised version* of the sample mean $\sqrt{T}\bar{X}$, which allows us to measure the uncertainty in $\sqrt{T}\bar{X}$. Based on these observations, an estimator of the variance of $\sqrt{T}\bar{X}$ is $\widetilde{\sigma}_T^2 = \frac{1}{M}\sum_{j=1}^{M}|\underline{e}'_{j,T}\underline{X}|^2$. Setting $M = (T-1)$, $\widetilde{\sigma}_T^2$ is the same as the sample variance. In future, we call $\{\underline{e}'_{j,T}\underline{X}; j = 1, \ldots, M\}$ a proxy sample for the sample mean, because it shares similar properties as the centralise sample mean.

In the second example, we relax the assumptions on $\{X_t\}$, and suppose it is a stationary, short memory, time series with autocovariance function $\{c(j)\}$ and spectral density function $f(\omega)$. In this case, our objective is to estimate the long run variance, $\mathrm{var}[\sqrt{T}\bar{X}_T] = \sum_j c(j) + o(T^{-1})$, using orthogonal transformations of the data which share the same variance. A suitable transformation should preserve the variance and be close to uncorrelated. For stationary time series, an ideal candidate is the Discrete Fourier transform $\{J_T(\omega_k)\}$. We observe that $J_T(0) = \sqrt{T/(2\pi)}\bar{X}_T$ and $\mathrm{var}[J_T(0)] = f(0) + O(T^{-1})$, whereas for $k \geq 1$, $\mathrm{E}[J_T(\omega_k)] = 0$ and $\mathrm{var}[J_T(\omega_k)] \approx f(0) + O(kT^{-1})$. Further, it is well known that under stationarity the DFT is almost an orthogonal transformation of the time series, in the sense that if $k_1 \neq k_2$ then $\mathrm{cov}[J_T(\omega_{k_1}), J_T(\omega_{k_2})] = O(T^{-1})$ and $\mathrm{var}[J_T(\omega_k)] \approx f(0) + O(kT^{-1})$. Focusing on the frequencies close to the origin, we observe $\{J_T(\omega_k); 1 \leq k \leq M\}$ is a near uncorrelated sample whose variable is close to the variance of $\mathrm{var}[J_T(0)]$ when M is small. Thus $\{J_T(\omega_k); 1 \leq k \leq M\}$ can be considered as proxy sample of the sample mean. These arguments naturally lead us to use $M^{-1}\sum_{k=1}^{M}|J_T(\omega_k)|^2$ as an estimator of $\mathrm{var}[J_T(0)]$. We observe that this estimator can also be viewed as an estimator of the spectral density function at zero, noting that the typical estimator of the long run variance is the spectral density estimator at zero.

In this paper we generalise this notion. That is, for a given statistic we define a 'near independent proxy' sample, which, by construction, has properties that mimic the centralised statistic of interest. We work within the frequency domain as it allows us to exploit the near uncorrelated property of the discrete Fourier transform to define the proxy sample. Currently, it is unclear whether other transformations have similar properties. Nevertheless, defining estimators within the frequency domain is not a restrictive assumption as most estimators defined within the time domain can also be defined within the frequency domain. In Section 2 we define a proxy sample associated with $A_T(\phi)$ which shares similar properties to the centralised $A_T(\phi)$, namely the same variance and higher order cumulants. These properties allow us to use the proxy sample to estimate the nuisance parameter V, defined in (2). In

Section 3 we address the issue of testing. Since the proxy sample shares similar sampling properties with the centralised version of the statistic, it can be used to estimate the finite sample distribution, and the critical values, of the statistic under the null that the mean of the statistic is zero. This type of setting arises in several situations in time series and in Sections 3.1 and 3.2 we use the proxy sample to estimate the distribution of the test statistic for testing for uncorrelatedness and the Goodness of Fit test.

The methods discussed in Sections 2 and 3 depend on number of elements, M, in the proxy sample. The number of elements plays an analogous role to the bandwidth in nonparametric regression; too few elements in the proxy sample will make the estimator too variable, while too many terms will induce a bias. Therefore, in Section 4 we propose a cross validation method to select the number of terms in the proxy sample. We mention that evaluation of the proxy sample and cross validation criterion method requires $O(T \log T)$ and $O(|\mathcal{S}|T)$ operations, respectively, where $\mathcal{S}$ denotes the set over which the cross-validation criterion is minimised and $|\mathcal{S}|$ denotes its cardinality. The methods are illustrated in Section 5 with simulations. The proofs can be found in the supplementary material.

2 Estimation of nuisance parameters

In this section we consider statistics which have the form $A_T(\phi)$. In particular, we define a proxy sample associated with $A_T(\phi)$ which we use to estimate the asymptotic variance V (defined in (2)). Throughout this paper we will assume the time series is s-order stationary - that is all of moments of X_t up to the s-moment is invariant to shift (for example, a strictly stationary time series with finite s-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 $\kappa_s(j_1, \ldots, j_{s-1}) = \text{cum}(X_t, X_{t+j_1}, \ldots, X_{t+j_{s-1}})$. Furthermore, we define spectral density and s-order spectral density functions as

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

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 $\{X_t\}$ constant, since the DFT zeros a constant mean at most frequencies.

It is clear that $A_T(\phi)$ is an estimator of $A(\phi)$, where

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

and as mentioned in the introduction, several interesting statistics can be written in the form $A_T(\phi)$. Below, we give some well known examples.

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

$$\widehat{c}_T(j) = \frac{1}{T}\sum_{k=1}^{T} \exp(ij\omega_k)|J_T(\omega_k)|^2 = \widetilde{c}_T(j) + \widetilde{c}_T(T-j) = \widetilde{c}_T(j) + O\left(\frac{|j|}{T}\right), \tag{3}$$

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

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

(c) In order to test for goodness of fit of a model with spectral density function $g(\omega;\theta)$, Milhoj [1981] proposed estimating the jth 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 $\phi(\omega) = e^{ij\omega}g(\omega_k;\theta)^{-1}$.

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

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

If for some $\theta \in \Theta$, $f(\omega) = f(\omega;\theta)$ (where $f(\omega)$ is the spectral density of the observed time series) then the sampling properties of $\widehat{\theta}_T$ are based on the sampling properties of the derivative of the Whittle likelihood at θ, where

$$\nabla_\theta \mathcal{L}_T(\theta) = A_T(\phi) + \frac{1}{T}\sum_{k=1}^{T}\frac{1}{f(\omega_k;\theta)}\nabla_\theta f(\omega_k;\theta) \tag{4}$$

and $\phi(\omega) = \nabla_\theta f(\omega_k;\theta)^{-1}$. More precisely, under sufficient mixing conditions we can show $V = \lim_{T\to\infty}\mathrm{var}[\nabla_\theta\mathcal{L}_T(\theta)] = \lim_{T\to\infty} T\mathrm{var}[A_T(\phi)]$. Thus, by using (4) and the Taylor series expansion it is well known that

$$\sqrt{T}\left(\widehat{\theta}_T - \theta\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, W^{-1}VW^{-1}\right),$$

where $W = \frac{1}{2\pi}\int_0^{2\pi} f(\omega)^{-2}\nabla f_\theta(\omega;\theta)\nabla f_\theta(\omega;\theta)'d\omega$.

To construct the proxy sample associated with $A_T(\phi)$ we recall some of the pertinent features of the proxy sample associated with the sample mean; that is $\{J_T(\omega_k); k = 1,\ldots,M\}$ 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 classical results it is clear that despite $|J_T(\omega_k)|^2$ and $J_T(\omega_k)\overline{J_T(\omega_{k+r})}$ estimating very different quantities (the spectral density and zero respectively), in the case that r is small and $k > 0$, the variances are close. This suggests that in order to construct the proxy 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

$$A_T(\phi; r) = \frac{1}{T}\sum_{k=1}^{T} \phi(\omega_k) J_T(\omega_k)\overline{J_T(\omega_{k+r})},$$

noting that $A_T(\phi; 0) = A_T(\phi)$. In the following lemmas we show that $\{\sqrt{2}\Re A_T(\phi; r), \sqrt{2}\Im A_T(\phi; r)\}_r$ (where $\Re$ and $\Im$ denote the real and imaginary parts of a random variable) is a suitable proxy sample. We first show that in general $A_T(\phi; 0)$ and $A_T(\phi; r)$ $(r > 0)$ have differing means.

Lemma 2.1 *Let us suppose that $\{X_t\}$ is second order stationary time series where $\sum_j |jc(j)| < \infty$ and $\phi(\cdot)$ is a Lipschitz continuous function. Then we have*

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

Despite these terms having different expectations in the following lemma and corollary we show that they share similar second order properties.

Theorem 2.1 *Suppose $\{X_t\}$ is a fourth order stationary time series where $\sum_{j=-\infty}^{\infty} |jc(j)| < \infty$ and for $1 \leq i \leq 3$ $\sum_{j_1,j_2,j_3=-\infty}^{\infty} |j_i \kappa_4(j_1, j_2, j_3)| < \infty$ and the function $\phi : [0, 2\pi] \to \mathbb{R}$ is Lipschitz continuous. Then we have*

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

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

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

and

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

where

$$\begin{aligned} V(\omega_r) &= \frac{1}{2\pi}\int_0^{2\pi} f(\omega) f(\omega + \omega_r)| \left(|\phi(\omega)|^2 + \phi(\omega)\overline{\phi(-\omega - \omega_r)}\right) d\omega + \\ &\quad \frac{1}{(2\pi)^2}\int_0^{2\pi}\int_0^{2\pi} \phi(\omega_1)\overline{\phi(\omega_2)} f_4(\omega_1, -\omega_1 - \omega_r, \omega_2) d\omega_1 d\omega_2. \end{aligned} \tag{5}$$

We observe that the two assumptions stated in the above lemma, $\sum_j |jc(j)| < \infty$ and $\sum_{j_1,j_2,j_3=-\infty}^{\infty} |j_i\kappa_4(j_1,j_2,j_3)| < \infty$ (for $1 \leq i \leq 3$) imply 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 the assumptions in Theorem 2.1 hold. Let $V(\cdot)$ be defined as in (5). 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 and Lemma 2.1 imply for $M << T$, that the sequence $\{\sqrt{2}\Re A_T(\phi;r), \sqrt{2}\Im A_T(\phi;r); r = 1,\ldots,M\}$ are 'near uncorrelated' random variables with approximately the same variance. Based on these observations we propose the following estimator of $V(0)$

$$\widehat{V}(0) = \frac{T}{2M}\sum_{r=1}^{M}\left(2|\Re A_T(\phi;r)|^2 + 2|\Im A_T(\phi;r)|^2\right) = \frac{T}{M}\sum_{r=1}^{M}|A_T(\phi;r)|^2. \tag{6}$$

In the following we show that $\widehat{V}(0)$ is a mean square consistent estimator of V.

Lemma 2.2 *Suppose that $\{X_t\}$ is an eight order stationary time series where for $2 \leq s \leq 8$, $\sum_{j_1,\ldots,j_{s-1}} |j_i\kappa_s(j_1,\ldots,j_{s-1})| < \infty$ (for $1 \leq i \leq s-1$). Then we have*

$$\mathrm{cov}[|\sqrt{T}A_T(\phi;r_1)|^2, |\sqrt{T}A_T(\phi;r_2)|^2] = \begin{cases} V(\omega_r)^2 + O(T^{-1}) & 0 < r_1 = r_2 (= r) \\ O(T^{-1}) & 0 < r_1 < r_2 < T/2 \end{cases} \tag{7}$$

and

$$\mathrm{E}\left(\widehat{V}(0) - V(0)\right)^2 = O\left(\frac{M^2}{T^2} + \frac{1}{M}\right). \tag{8}$$

It is interesting to note that the estimator of $\widehat{V}(0)$ is analogous to kernel estimators in nonparametric regression, where M plays the role of window width (bandwidth multiplied by the length of time series). We note that we can directly apply (6) to estimate the statistics defined in Example 2.1(a,b,c). Missing from these examples is the estimator of the variance of the derivative of the Whittle likelihood $A_T(\nabla_\theta f(\cdot;\theta)^{-1})$ given in Example 2.1(d). This is because θ is unknown, and we only have an estimator $\widehat{\theta}_T$. Instead, we suggest using

$$\widehat{V_{\widehat{\theta}}}(0) = \frac{1}{M}\sum_{r=1}^{M}\left|\frac{1}{\sqrt{T}}\sum_{k=1}^{T}\nabla_\theta f(\omega_k;\widehat{\theta})^{-1}J_T(\omega_k)\overline{J_T(\omega_{k+r})}\right|^2 \tag{9}$$

as an estimator of $V_\theta = \lim_{T\to\infty} T\text{var}[A_T(\nabla_\theta f(\cdot;\theta)^{-1})]$. Similarly, in the general case that we want to estimate the variance of

$$A_T(\phi_\theta) = \frac{1}{T}\sum_{k=1}^{T}\phi(\omega_k;\theta)|J_T(\omega_k)|^2,$$

but only observe an estimator $\widehat{\theta}_T$ of θ, we can use

$$\widehat{V}_{\widehat{\theta}}(0) = \frac{T}{M}\sum_{r=1}^{M}|A_T(\phi_{\widehat{\theta}};r)|^2, \text{ where } \quad A_T(\phi_{\widehat{\theta}};r) = \frac{1}{T}\sum_{k=1}^{T}\phi(\omega_k;\widehat{\theta})J_T(\omega_k)\overline{J_T(\omega_{k+r})} \qquad (10)$$

as an estimator of $V_\theta = \lim_{T\to\infty} T\text{var}[A_T(\phi_\theta)]$. We now show that $\widehat{V}_{\widehat{\theta}}(0)$ consistently estimates V_θ.

Lemma 2.3 *Let us suppose that* $|\widehat{\theta}_T - \theta| = O_p(T^{-1/2})$, $\sup_{\theta,\omega}|\frac{\partial\phi(\omega;\theta)}{\partial\omega}| < \infty$ *and* $\sup_{\theta,\omega}|\frac{\partial^2\phi(\omega;\theta)}{\partial\theta^2}| < \infty$. *Let* $\widehat{V}_{\widehat{\theta}}(0)$ *be defined as in (10). Then we have*

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

where $V_\theta = \lim_{T\to\infty} T\text{var}[A_T(\phi_\theta)]$.

It immediately follows from the above lemma that if $f(\omega;\theta)$ is uniformly bounded away from zero and uniformly bounded from above for all θ and ω, and its first and second derivatives with respect to θ and ω are uniformly bounded, then (9) is a consistent estimator of V_θ if $M/T \to 0$ as $M \to \infty$ and $T \to \infty$.

We now state an asymptotic result, which will be useful in the construction of confidence intervals.

Theorem 2.2 *Let us suppose that* $\{X_t\}$ *is a stationary* α*-mixing time series, where the* α*-mixing coefficient* $\alpha(t)$ *is such that* $\alpha(t) \leq K|t|^{-s}$ *(for* $|t| \neq 0$*) and* $K < \infty$. *In addition we will assume that* $\phi : [0, 2\pi] \to \mathbb{R}$ *has a bounded second derivative,* $A_T(\phi)$ *is a real-valued random variable and for some* $r > 4s/(s-6)$ *we have* $\mathrm{E}|X_t|^r < \infty$. *Let* $A_{M,T} = (A_T(\phi;1), \ldots, A_T(\phi;M))'$. *Then for* M *fixed we have*

$$\sqrt{\frac{T}{V(0)}}\begin{pmatrix} A_T(\phi) - A(\phi) \\ \Re A_{M,T} \\ \Im A_{M,T}\end{pmatrix} \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \text{diag}\big(1, \frac{1}{\sqrt{2}}I_M, \frac{1}{\sqrt{2}}I_M\big)\right),$$

where I_M *denotes the identity matrix of dimension* M.

PROOF. The proof immediately follows from Lemma 3.3, Lee and Subba Rao [2012]. Note the same result will hold under different type mixing conditions, including physical dependence (see Wu [2005] and, for quadratic forms, Wu and Shao [2007]). $\square$

The above lemma implies that for M fixed, the ratio

$$\frac{\sqrt{T}[A_T(\phi) - A(\phi)]}{\sqrt{\widehat{V}(0)}} \xrightarrow{\mathcal{D}} \frac{Z_0}{\sqrt{\frac{1}{M}\sum_{j=1}^{M} Z_j^2}} \sim t_M,$$

as $T \to \infty$, where $\{Z_j; j = 0, \ldots, M\}$ are iid standard normal random variables and t_M denotes the t-distribution with M degrees of freedom. Using this result a $(100 - \alpha)\%$ confidence interval for $A(\phi)$ is

$$\left[A_T(\phi) \pm t_M(\alpha/2)\sqrt{\frac{\widehat{V}(0)}{T}}\right].$$

We have seen that by construction the proxy sample has approximately the same variance as the statistic of interest, and both the original statistic and the proxy sample are asymptotically Gaussian. Thus the proxy sample can be considered as a centralised version of the statistic of interest. We use these properties in the following section.

3 Testing in Time Series

In this section we turn to the problem of hypothesis testing in time series. Many test statistics in time series can be formulated in terms of $\{A_T(\phi_j)\}_j$, where under the null hypothesis $\mathrm{E}[A_T(\phi_j)] = 0$ and under the alternative $\mathrm{E}[A_T(\phi_j)] \neq 0$. This motivates the popular ℓ_2 test statistic

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

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

By using the results in Section 2 that under the null hypothesis $\{\sqrt{2}\Re A_T(\phi_j; r), \sqrt{2}\Im A_T(\phi_j; r)\}$ and $A_T(\phi_j; r)$ have asymptotically the same marginal distribution. Furthermore, the covariances between $\{A_T(\phi_j)\}_j$ are asymptotically equivalent to the covariances between $\{\sqrt{2}\Re A_T(\phi_j; r), \sqrt{2}\Im A_T(\phi_j; r)\}_j$ (see Theorem 3.1). Based on these observations we define the proxy sample associated with S_T as

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

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

Theorem 3.1 *Suppose that $\{X_t\}$ is an 16-order stationary time series where for $2 \leq s \leq 16$, $\sum_{j_1,\ldots,j_{s-1}} |j_i \kappa_s(j_1,\ldots,j_{s-1})| < \infty$ (for $1 \leq i \leq s-1$). Furthermore, we assume $\{\phi_j\}$ are Lipschitz continuous functions and $\Re A_T(\phi_j) = A_T(\phi_j)$ (this final assumption is used to simplify analysis). Let V_{j_1,j_2} be defined as*

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

Then we have

(i) <u>*The mean*</u>

(a) Under the null hypothesis that $\mathrm{E}[A_T(\phi_j)] = 0$ for $1 \leq j \leq L$ we have

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

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

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

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

(ii) <u>*The covariance*</u>

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

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

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

(iii) <u>*Higher order cumulants*</u> *Suppose that $\{X_t\}$ is a $2p$-order stationary time series with $2 \leq s \leq 2p$ and $\sum_{j_1,\ldots,j_{s-1}} |j_i \kappa_s(j_1,\ldots,j_{s-1})| < \infty$. Let cum_p denote the pth order cumulant of a random variable. Then under the null hypothesis*

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

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

$$\widehat{F}_{M,T}(x) = \frac{1}{2M}\left(\sum_{r=1}^{M}\left[I(S_{T,R}(r) \leq x) + I(S_{T,I}(r) \leq x)\right]\right). \tag{11}$$

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 that at least one $\mathrm{E}[A_T(\phi_j)]$ for $j = 1, \ldots, L\}$ is non-zero, then $S_T = O(T)$. By Theorem 3.1(ii) the variance of $S_{T,R}(r)$ and $S_{T,I}(r)$ is finite and uniformly bounded for all r and T. This implies that $1 - \widehat{F}_{M,T}(S_T) \stackrel{\mathcal{P}}{\to} 0$ as M and $T \to \infty$, thus giving the procedure power. At this point we mention that under sufficient Brillinger-mixing conditions and the null hypothesis, it can be shown that the kth sample moment associated with the distribution of $\widehat{F}_{M,T}(x)$ consistently estimates the kth moment of S_T. However, to show a Glivenko-Cantelli type result of the form $\sup_{x\in\mathbb{R}} |\widehat{F}_{M,T}(x) - F(x)| \stackrel{\text{a.s.}}{\to} 0$ as $M \to \infty$ and $T \to \infty$, where F denotes the limiting distribution of Q_T under the null hypothesis (F is a generalised chi-squared) is beyond the scope of the current paper. Nevertheless, we believe that this result holds for a general class of time series (including those with only a finite number of moments). Our conjecture is supported by the results of the simulation in Section 5, where we apply the proposed methodology to a wide class of problems.

We now apply 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 independent, identically distributed (iid) random variables. In this case the classical Box-Pierce statistic is defined as

$$\widetilde{Q}_T = \frac{T}{\widetilde{c}_T(0)^2}\sum_{j=1}^{L}|\widetilde{c}_T(j)|^2, \tag{12}$$

where $\widetilde{c}_T(j)$ is defined in Example 2.1(a). If the null holds, then $\widetilde{Q}_T$ is asymptotically a chi-square distribution with L degrees of freedom. However, if the intention is to test for uncorrelatedness (ie. $H_0 : c(j) = 0$ for all $j \neq 0$ against the alternative $H_A : c(j) \neq 0$ for some j), without the additional constraint of independence, then it can be shown that

$$T\mathrm{cov}[\widetilde{c}_T(j_1), \widetilde{c}_n(j_2)] = c(0)^2\delta_{j_1,j_2} + c(0)^2\delta_{j_1,j_2}\delta_{j_1,0} + \sum_{k=-\infty}^{\infty}\kappa_4(j_1, k, k+j_2), \tag{13}$$

where δ_{j_1,j_2} is the dirac-delta function (see Brockwell and Davis [1987], Chapter 7, for the derivation in the case of a linear time series and Romano and Thombs [1996] in the general case). Consequently, under the null of uncorrelatedness, the distribution of $\widetilde{Q}_T$ is not a standard chi-square.

Diebold [1986], Weiss [1986], Robinson [1991], Bera and Higgins [1993] and Escanciano and Lobato [2009] avoid some of these issues by placing stronger conditions on the time series and assume that under the null hypothesis the time series are martingales differences (thus uncorrelated). This implies that the fourth order cumulant term in (13) is zero in the case that $j_1 \neq j_2$, which induces asymptotic uncorrelatedness between the sample covariances. Based on this observation they propose the robust Portmanteau test

$$Q_T^* = T \sum_{j=1}^{L} \frac{|\widetilde{c}_T(j)|^2}{\widehat{\tau}_j}, \tag{14}$$

where $\widehat{\tau}_j = \frac{1}{n-j} \sum_{t=j+1}^{n} (X_t - \bar{X})^2 (X_{t-j} - \bar{X})^2$. Under the null of martingale differences Q_T^* is asymptotically χ^2-distributed with L-degrees of freedom. However, if the intention is to test for uncorrelatedness, *without* additional assumptions on the structure, then, even under the null, Q_T^* will have a generalised chi-squared distribution, whose parameters are difficult to estimate. This problem motivated Romano and Thombs [1996] (using the block bootstrap) and Lobato [2001] (who developed the method of self-normalisation) to test for uncorrelatedness under these weaker conditions. Our intention in this section is to use proxy samples to test for serial correlation.

We will use proxy samples to estimate the distribution of Portmanteau statistic under the null that $H_0 : c(j) = 0$ for all $1 \leq j \leq L$. We recall from Example 2.1(a) that $A_T(e^{ij\cdot})$ is an estimator of the autocovariance $\widehat{c}_T(j)$. Therefore, to test for uncorrelatedness at lag $j = 1, \ldots, L$ we define the test statistic

$$Q_T = T \sum_{j=1}^{L} |A_T(e^{ij\cdot})|^2. \tag{15}$$

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

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

and the empirical distribution

$$\widehat{F}_{M,T}(x) = \frac{1}{2M} \left(\sum_{r=1}^{M} [I(Q_{T,R}(r) \leq x) + I(Q_{T,I}(r) \leq x)] \right). \tag{16}$$

We reject the null at the $\alpha\%$-level if $1 - \widehat{F}_{M,T}(Q_T) < \alpha\%$. Results of the corresponding simulation study is given in in Section 5.1, 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 proxy samples can be used in a goodness of fit statistic. 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 (see, for example, Milhoj [1981], Hong [1996] and Chen and Deo [2004]). We mentioned in Example 2.1(c) that $A_T(e^{ij\cdot}g(\cdot;\theta)^{-1})$ is an estimator of the covariance of the residual process at lag j. If the null were true, then the residual covariance, $A_T(e^{ij\cdot}g(\cdot;\theta)^{-1})$, is estimating zero. Motivated by Milhoj [1981] we define the test statistic as

$$G_T = T \sum_{j=1}^{L} \left| A_T(e^{ij\cdot}g(\cdot;\theta)^{-1}) \right|^2. \tag{17}$$

Furthermore, motivated by (16), we estimate the finite sample distribution of G_T with

$$\widehat{F}_{M,T}(x) = \frac{1}{2M} \left(\sum_{r=1}^{M} \left[I(G_{T,R}(r) \leq x) + I(G_{T,I}(r) \leq x) \right] \right), \tag{18}$$

where $\{G_{T,R}(r), G_{R,I}(r)\}$ is the corresponding proxy sample defined as

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

A similar result stated in Theorem 3.1 also applies G_T, $G_{T,R}(r)$ and $G_{T,I}(r)$. Thus we use $\widehat{F}_{M,T}(x)$ as an approximation of the distribution of G_T under the null of no serial correlation. We reject the null at the $\alpha\%$-level if $1 - \widehat{F}_{M,T}(G_T) < \alpha\%$.

In Section 5.2 we illustrate this method with some simulations.

4 Selection of M

It is clear that both the estimator $\widehat{V}$ and empirical distribution function $\widehat{F}_{M,T}$ discussed in Sections 2 and 3 rely on the choice of M. In this section we propose a cross validation criterion for selecting M. Our proposed method is based on the results derived in Section 2. Using Theorem 2.1 we note that $\{A_T(\phi; r); 1 \leq r < T/2\}$ is an almost uncorrelated, zero mean sequence with variance $\mathrm{var}[A_T(\phi; r)] \approx V(\omega_r)$ (where $V(\omega_r)$ is defined in (5)). These observations together with (7) imply that

$$\left(\frac{|\sqrt{T} A_T(\phi; r)|^2}{V(\omega_r)} - 1 \right) \qquad 1 \leq r < T/2, \tag{19}$$

is an almost uncorrelated sequence with mean zero and variance one. We use this sequence as the building blocks of the cross validation criterion. In order to select M we extend the estimator defined in (6) to all frequencies $V(\omega_r)$. More precisely, we use $\widehat{V}(\omega_r)$ as an estimator of $V(\omega_r)$ where

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

noting that for $r = 0$ we have the estimator defined in (6). Therefore, replacing $\widehat{V}(\omega_r)$ in (19) with $\widehat{V}_M(\omega_r)$ we see that if M is too large, it will induce a bias, whereas if M is too small it's variance will be inflated. Using these observations we propose to select the M which minimises the average squared error for frequencies in the neighbourhood of zero. We define the average squared error as

$$\mathcal{C}_\phi(M) = \frac{4}{T} \sum_{r=1}^{T/p} \left(\frac{\left|\sqrt{T} A_T(\phi; r)\right|^2}{\widehat{V}_M(\omega_r)} - 1 \right)^2,$$

noting that we have only used the T/p frequencies which are closest to 0. To select M we use the cross validation criterion $\widehat{M} = \arg\min_{M \in \mathcal{S}} \mathcal{C}_\phi(M)$, where S is the set over which we do the selection. To illustrate how the criterion behaves, in Figure 1 we give a plot of $\mathcal{C}_\phi(M)$ over M for $\phi(\omega) = e^{i\omega}$ (which corresponds to the statistic which estimates the autocovariance at lag one).

5 Simulations

In this section we illustrate the proxy sampling methodology by using it to test for uncorrelatedness and for goodness of fit (described in Section 3). All tests are done at the $\alpha = 5\%$ and $\alpha = 10\%$ nominal levels. The methods are compared to the block bootstrap method, where to obtain the bootstrap critical values 1000 bootstrap samples were taken. Throughout this section we let $\{Z_t\}$ and $\{\varepsilon_t\}$ denote independent, identically distributed standard normal random variables and chi-square with one degree of freedom random variables respectively.

5.1 Testing for uncorrelatedness

In this section we illustrate the test for uncorrelatedness using the proxy sample method described in Section 3.1. We use the test statistic Q_T (defined in (15)), using $L = 5$, and obtain the critical values using the estimated distribution function, $\widehat{F}_{M,T}$ defined in (16). We compare our method with the regular Box-Pierce statistic (defined in (12)) and the robust Portmanteau test statistic defined in (14), for both these methods we obtained the critical values using the χ^2 distribution with five degrees of freedom. In addition, we compare our method to the results of the bootstrap test where the critical values are obtained using the block bootstrap procedure. Namely, the critical values for Q_T are obtained using the centralised empirical distribution constructed with samples from the block bootstrap procedure, with block bootstrap length $B = 5, 10$ and 20.

To select M in the proxy sample method we use the cross-validation 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

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

For the case that $n = 100$ and $n = 200$ we used $\mathcal{S} = \{10, \ldots, 25\}$ and for $n = 500$ we used $\mathcal{S} = \{10, \ldots, 40\}$.

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 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

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

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, moreover unlike the stationary case, there are significant correlations between the DFT separated by $T/12$ frequencies). We used the sample sizes $T = 100$ and $T = 500$.

The results are given in Tables 1 and 2. We observe that overall the proxy sampling method keeps to the nominal level, with a mild inflation of the type I error for independent data (normal and t-distribution), which is probably because the 5th and 10th percentile is estimated using a maximum of 40 points (depending on the set $\mathcal{S}$). 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. Suprisingly, 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.

Model	Proxy Q_T		Regular $\widetilde{Q}_T$		Robust Q_T^*		Block Bootstrap					
							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 1: Test for uncorrelatedness, under the null hypothesis, $T = 100$ over 5000 replications.

Model	Proxy Q_T		Regular $\widetilde{Q}_T$		Robust Q_T^*		Block Bootstrap					
							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 2: Test for uncorrelatedness, under the null hypothesis, $T = 500$ over 5000 replications.

Models under the alternative of correlation

To access 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 $\{U_{t,2}\}$ are independent of each other, $\{Y_{1,t}\}$ is defined above and $\{U_{t,2}\}$ 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, $\{U_{t,2}\}$ is the ARCH model defined in the previous section and $\{U_{3,t}\}$ is the Gaussian autoregressive process $U_{3,t} = 0.5U_{4,t-1} + Z_t$. We used the sample sizes $T = 100$, $T = 200$ and $T = 500$.

The result are given in Tables 3, 4 and 5. The power for most of the methods are relatively close (the regular Box-Pierce statistic having the largest power). Overall, in terms of power, the proxy 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	Proxy Q_T		Regular $\widetilde{Q}_T$		Robust Q_T^*		Block Bootstrap					
							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 3: Test for uncorrelatedness, under the alternative hypothesis, $T = 100$ over 5000 replications.

5.2 Goodness of fit test

In this section we illustrate the goodness of fit test using the proxy sample method described in Section 3.2 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 $\omega \in [0, 2\pi]$. We use the test statistic G_T (defined in (17)), with $L = 5$, and obtain the critical values using the estimated distribution function, $\widehat{F}_{M,T}$ defined in (18). We compare our method to the results of the bootstrap test where the critical values are obtained using the block bootstrap procedure. We use the block bootstrap length $B = 5, 10, 20, 30$ and 40.

Model	Proxy Q_T		Regular $\widetilde{Q}_T$		Robust Q_T^*		Block Bootstrap					
							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 4: Test for uncorrelatedness under the alternative hypothesis, $T = 200$ over 5000 replications.

Model	Proxy Q_T		Regular $\widetilde{Q}_T$		Robust Q_T^*		Block Bootstrap					
							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 5: Test for uncorrelatedness under the alternative hypothesis, $T = 500$ over 5000 replications.

To select M in the proxy sample method, we focus on $j = 1$ and use $M = \arg\min_{M\in\mathcal{S}} C_\phi(M)$, where

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

For $n = 100$ and $n = 200$ we used $\mathcal{S} = \{10,\ldots,25\}$ and for $n = 500$ we used $\mathcal{S} = \{10,\ldots,40\}$.

Models under the null hypothesis

The first model is the Gaussian autoregressive process

$$X^G_{0.6,t} = 0.6X^G_{0.6,t-1} + Z_t,$$

with spectral density $f(\omega) = g(\omega;\theta) = (2\pi)^{-1}|1 - 0.6e^{i\omega}|^2$. The second model is the non-Gaussian autoregressive process

$$X^\chi_{0.6,t} = 0.6X^\chi_{0.6,t-1} + \varepsilon_t,$$

with spectral density $f(\omega) = g(\omega;\theta) = 2(2\pi)^{-1}|1 - 0.6e^{i\omega}|^2$. The third model is the non-Gaussian autoregressive process

$$X^\chi_{0.9,t} = 0.9X^\chi_{0.9,t-1} + \varepsilon_t,$$

with spectral density $f(\omega) = g(\omega;\theta) = 2(2\pi)^{-1}|1 - 0.9e^{i\omega}|^2$. We used the sample sizes $T = 100$ and $T = 500$.

The result are given in Tables 6 and 7. We observe that the proxy sampling method tends to keeps to the nominal level for both $T = 100$ and $T = 500$, with an underestimation of the type I error for model $\{X^\chi_{0,9,t}\}$ (which is also seen for the Block bootstrap). On the other hand for $T = 100$, the block bootstrap underestimates the nominal level when the block is too small ($B = 5$ and 10) and over inflates the type I error when the block is too large ($B = 40$ and sometimes $B = 30$). The ideal block length seems to be somewhere between $B = 20 - 40$, though it appears to depend on the model. In the case that $T = 500$, the nominal level is underestimated for all the block lengths considered.

Model	Proxy Q_T		Block Bootstrap									
			B=5		B=10		B=20		B=30		B=40	
	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
$X_{0.6,t}^{G}$	2.32	4.94	0.00	0.00	0.18	0.96	2.18	7.06	4.66	11.24	8.64	17.84
$X_{0.6,t}^{\chi}$	2.24	4.4	0	0.02	0.04	0.26	0.94	3.48	2.24	6.82	4.94	13.00
$X_{0.9,t}^{\chi}$	0.96	1.74	0	0	0	0.02	0.04	0.26	0.2	0.96	0.56	2.64

Table 6: Goodness of fit test, under the null, T=100 over 5000 replications

Model	Proxy Q_T		Block Bootstrap									
			B=5		B=10		B=20		B=30		B=40	
	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
$X_{0.6,t}^{G}$	5.24	10.28	0	0	0.08	0.52	1.54	3.92	2.48	6	3.32	7.62
$X_{0.6,t}^{\chi}$	5.2	9.78	0	0	0.02	0.16	0.52	1.84	0.04	0.14	2.12	6.68
$X_{0.9,t}^{\chi}$	2.24	4.84	0	0	0	0	0	0	0.04	0.62	0.04	0.62

Table 7: Goodness of fit test, under the null, T=500 over 5000 replications

Models under the alternative hypothesis

To access the empirical power of the test we use realisations from the same models considered in the null, namely $X_{0.6,t}^{0.9}, X_{0.6,t}^{\chi}$ and $X_{0.9,t}^{\chi}$ (their corresponding spectral density functions are given in the previous section). To each of these models we fit the spectral density function $g(\omega;\phi,\sigma) = (2\pi)^{-1}\sigma^2|1-\phi\exp(i\omega)|^{-2}$ for different values of ϕ and σ (though σ will always be correctly specified). We used the sample sizes $T = 100$, $T = 200$ and $T = 500$.

The result are given in Tables 8, 9 and 10. For all the sample sizes considered, the power of the Block bootstrap test increases with the block length, though we recall that the largest block size the type I errors were highly inflated. Overall, the power of the proxy sample test is comparable (and often larger) than the power of the block bootstrap tests which use the larger blocks.

Model	Null	Proxy Q_T		Block Bootstrap									
				B=5		B=10		B=20		B=30		B=40	
		5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
$X_{0.6,t}^{G}$	$\sigma = 1, \phi = 0.3$	60.4	71.24	1.98	11.32	35.42	55.12	53.24	70.78	61.2	76.5	68.62	80.34
$X_{0.6,t}^{\chi}$	$\sigma = \sqrt{2}, \phi = 0.3$	64.32	76.12	1.3	8.68	33.2	56.12	54.24	74.58	62.4	80.42	71.06	84.14
$X_{0.6,t}^{G}$	$\sigma = 1, \phi = 0.45$	15.1	22.28	0	0	1.94	7.24	10.94	22.06	15.9	28.62	23.14	36.36
$X_{0.6,t}^{\chi}$	$\sigma = \sqrt{2}, \phi = 0.45$	13.98	22.38	0	0	0.82	3.96	6.42	16.78	10.72	23.6	17.9	32.74
$X_{0.9,t}^{\chi}$	$\sigma = \sqrt{2}, \phi = 0.7$	41.14	51.82	0	0	3.2	10.28	17.54	36	25.54	45.58	36.24	56.44

Table 8: Goodness of fit test, under the alternative, T=100 over 5000 replications

Model	Null	Proxy Q_T		Block Bootstrap									
				B=5		B=10		B=20		B=30		B=40	
		5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
$X_{0.6,t}^{G}$	$\sigma = 1, \phi = 0.3$	92.88	96.5	39.62	63.78	85.76	93.20	91.22	95.72	93.36	97.04	93.48	97.40
$X_{0.6,t}^{\chi}$	$\sigma = \sqrt{2}, \phi = 0.3$	96.04	98.5	33.5	62.1	88.86	96.02	93.9	98.4	95.14	98.80	95.66	98.92
$X_{0.6,t}^{G}$	$\sigma = 1, \phi = 0.45$	38.38	49.5	0	0.1	8.88	20.38	27.16	42.2	33.52	48.28	39.08	53.82
$X_{0.6,t}^{\chi}$	$\sigma = \sqrt{2}, \phi = 0.3$	40.46	53.08	0	0	6.04	15.86	21.28	38.98	27.34	46.52	34.24	53.28
$X_{0.9,t}^{\chi}$	$\sigma = \sqrt{2}, \phi = 0.7$	91.44	94.78	0	0	29.9	49.5	70.4	84.94	80.98	91	85.78	93.9

Table 9: Goodness of fit test, under the alternative, T=200 over 5000 replications

Model	Null	Proxy Q_T		Block Bootstrap									
				B=5		B=10		B=20		B=30		B=40	
		5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
$X_{0.6,t}^{G}$	$\sigma = 1, \phi = 0.3$	99.98	100	98.88	99.72	100	100	100	100	100	100	100	100
$X_{0.6,t}^{\chi}$	$\sigma = \sqrt{2}, \phi = 0.3$	99.98	100	99.2	99.92	100	100	100	100	99.92	100	99.92	100
$X_{0.6,t}^{G}$	$\sigma = 1, \phi = 0.45$	86.90	92.52	0.18	1.54	48.88	65.62	77.36	86.7	81	89.93	82.58	90.72
$X_{0.6,t}^{\chi}$	$\sigma = \sqrt{2}, \phi = 0.45$	89.92	94.22	0.06	0.56	42.16	61.88	75.60	87.44	81.94	90.8	84.24	92.26
$X_{0.9,t}^{G}$	$\sigma = \sqrt{2}, \phi = 0.7$	99.96	99.98	0	0	94.46	98.12	99.86	99.96	99.9	99.98	99.92	99.98

Table 10: Goodness of fit test, under the alternative, T=500 over 5000 replications

Acknowledgements

The author thanks Soutir Bandyopadhyay and Carsten Jentsch, whose work inspired me to look into this problem. Furthermore the author also thanks Piotr Fryzlewicz and Hernando Omabo for making several insightful comments.

A Appendix

To prove the results we make heavy use of the following well known result. Suppose that $\{X_t\}_t$ is an s-order stationary time series where $\sum_{j_1,\ldots,j_{s-1}} |j_i \kappa_s(j_1,\ldots,j_{s-1})| < \infty$ (for $1 \leq i \leq s-1$). Then we have

$$\mathrm{cum}\left[J_T(\omega_{k_1}),\ldots,J_T(\omega_{k_s})\right] = \frac{1}{T^{s/2-1}} I_{k_1+\ldots+k_s \in T\mathbb{Z}} f_s(\omega_{k_1},\ldots,\omega_{k_{s-1}}) + O(\frac{1}{T^{s/2}}), \tag{20}$$

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

A.1 Proofs for Section 2

PROOF of Lemma 2.1 This immediately follows from (20) and the Lipschitz continuity of f and ϕ which allows the sum to be replaced by an integral. □

PROOF of Theorem 2.1 We first derive expressions for $\mathrm{cov}[A_T(\phi;r_1), A_T(\phi;r_2)]$ and $\mathrm{cov}[A_T(\phi;r_1), \overline{A_T(\phi;r_2)}]$ 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

(20) we have

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

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

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

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

Finally, by using the identities $\Re A_T(\phi; r) = \frac{1}{2}(A_T(\phi; r_1) + \overline{A_T(\phi; r_1)})$ and $\Im A_T(\phi; r) = \frac{-i}{2}(A_T(\phi; r_1) - \overline{A_T(\phi; r_1)})$ and the above the result immediately follows. $\square$

Next we state a result which will be useful in the subsequent proofs.

Lemma A.1 *Suppose that $\{X_t\}$ is an $2n$-order stationary time series where for $2 \leq s \leq 2n$, $\sum_{j_1,\ldots,j_{s-1}} |j_i \kappa_s(j_1,\ldots,j_{s-1})| < \infty$ (for $1 \leq i \leq s-1$). Then we have*

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

PROOF The proof immediately follows from using indecomposible partitions and (20). $\Box$

We make use of Lemma A.1 below.

PROOF of Lemma 2.2 To prove (7) we note that

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

Thus we see that (7) follows immediately from the above, Theorem 2.1 and Lemma 2.2. To prove (8) we use the classical variance bias decomposition

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

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

$$\widehat{V}(0) \quad = \quad \frac{1}{M^2}\sum_{r_1,r_2=1}^{M}\left(\left|\mathrm{cov}[\sqrt{T}A_T(\phi;r_1), \sqrt{T}A_T(\phi;r_2)]\right|^2 = O(M^{-1} + T^{-1}),\right.$$

where the last line follows immediately from (7). By using Corollary 2.1 we can show that $|\mathrm{E}[\widehat{V}(0)] - V(0)| = O(M/T)$. Altogether, this gives the desired result. $\Box$

PROOF of Lemma 2.3 By using a Taylor expansion we have

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

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

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

Therefore

$$\begin{aligned}\widehat{V}_{\widehat{\theta}} \quad &= \quad \frac{T}{M}\sum_{r=1}^{M}|A_T(\phi_\theta;r)|^2 + \underbrace{|\widehat{\theta}_T - \theta|^2}_{O_p(T^{-1})}\underbrace{\frac{T}{M}\sum_{r=1}^{M}|A_T(\nabla_\theta\phi_\theta;r)|^2}_{=O_p(1),\text{ by Lemma 2.2}} + O_p(T^{-1}) \\ &= \quad \frac{T}{M}\sum_{r=1}^{M}|A_T(\phi_\theta;r)|^2 + O_p(T^{-1}). \end{aligned} \tag{21}$$

By using Lemma 2.2 we have $\mathrm{E}(\frac{T}{M}\sum_{r=1}^{M}|A_T(\phi_\theta;r)|^2 - V_\theta)^2 = O(M^{-1} + M/T)$, thus by using this and (21) we obtain the result. $\Box$

A.2 Proofs for Section 3.1

To prove Theorem 3.1 we use the following definition

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

The following lemma facilitates the proof of Theorem 3.1.

Lemma A.2 *Suppose that $\{X_t\}$ is an 16-order stationary time series where for $2 \leq s \leq 16$, $\sum_{j_1,\ldots,j_{s-1}} |j_i \kappa_s(j_1,\ldots,j_{s-1})| < \infty$ (for $1 \leq i \leq s-1$). Then we have*

$$\begin{aligned} T\mathrm{cov}\left[A_T(\phi_{j_1};0), A_T(\phi_{j_2};0)\right] &= V_{j_1,j_2} + O(T^{-1}) \\ T\mathrm{cov}\left[A_T(\phi_{j_1};0), \overline{A_T(\phi_{j_2};0)}\right] &= V_{j_1,j_2} + O(T^{-1}). \end{aligned} \tag{22}$$

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

$$\begin{aligned} T\mathrm{cov}\left[A_T(\phi_{j_1};r_1), A_T(\phi_{j_2};r_2)\right] &= \begin{cases} V_{j_1,j_2}(\omega_r) + O(T^{-1}) & r_1 = r_2 \\ O(T^{-1}) & r_1 \neq r_2 \end{cases} \\ T\mathrm{cov}\left[A_T(\phi_{j_1};r_1), \overline{A_T(\phi_{j_2};r_2)}\right] &= O(T^{-1}). \end{aligned} \tag{23}$$

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

$$T^2\mathrm{cum}\left[A_T(\phi_{j_1};r_1), A_T(\phi_{j_2};r_2), A_T(\phi_{j_3};r_3), A_T(\phi_{j_4};r_4)\right] = O(T^{-1}). \tag{24}$$

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

$$T^2\mathrm{cov}[A_T(\phi_{j_1};r_1)^2, A_T(\phi_{j_2};r_2)^2] = \begin{cases} 2V_{j_1,j_2}(\omega_r)^2 + O(T^{-1}) & r_1 = r_2 (= r) \\ O(T^{-1}) & r_1 \neq r_2 \end{cases} \tag{25}$$

$$T^2\mathrm{cov}[A_T(\phi_{j_1};r_1)^2, A_T(\phi_{j_2};r_2)\overline{A_T(\phi_2;r_2)}] = O(T^{-1}) \tag{26}$$

$$T^2\mathrm{cov}[|A_T(\phi_{j_1};r_1)|^2, |A_T(\phi_{j_2};r_2)|^2] = \begin{cases} V_{j_1,j_2}(\omega_r)^2 + O(T^{-1}) \\ O(T^{-1}) \end{cases} \tag{27}$$

Finally

$$|V_{j_1,j_2}(\omega_r) - V_{j_1,j_2}| \leq K|r|T^{-1}, \tag{28}$$

where K is a finite constant.

PROOF The proof of (22) and (23) is identical to the proof of Theorem 2.1, thus we omit the details. The proof of (24) follows from Lemma A.1, thus we omit the details. To prove

(25) we use indecomposable partitions to decompose the term in the product of covariances and a fourth order cumulant term. Specifically

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

By using (23) we obtain (25). A similar proof applies to (26)-(27).

Finally, to prove (28) we simply use the Lipschitz continuity of f, f_4 and ϕ_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

$$\begin{aligned}
\mathrm{E}[S_T] &= T\sum_{j=1}^{L}\mathrm{E}\left|(A_T(\phi_j) - \mathrm{E}[A_T(\phi_j)) + \mathrm{E}[A_T(\phi_j)]\right|^2 \\
&= T\sum_{j=1}^{L}\mathrm{E}\left|(A_T(\phi_j) - \mathrm{E}[A_T(\phi_j)) + \mathrm{E}[A_T(\phi_j)]\right|^2 \\
&= T\sum_{j=1}^{L}\mathrm{var}[A_T(\phi_j)] + \sum_{j=1}^{L}\left|\mathrm{E}[A_T(\phi_j)]\right|^2. \qquad (29)
\end{aligned}$$

Using that under the null $\mathrm{E}[A_T(\phi_j)] = 0$, and substituting this into the above we have

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

where the last line follows from (23). This gives (ia). To prove (ib) we note that

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

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

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

thus proving (ib).

To prove (iia) we note that since $A_T(\phi_j)$ is real and under the null $\mathrm{E}[A_T(\phi_j)]$, then expanding $\mathrm{var}[S_T]$ gives

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

where the last line follows from (23) and (24). Thus proving (iia).

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

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

Thus

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

Thus by using (25)-(27) we have

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

Now we recall that $S_{T,R}(r) = 2T\sum_{j=1}^{L}|\Re A_T(\phi_j;r)|^2$, thus by using the above and (28) we have

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

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

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

References

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, 1987.

W. Chen and R. Deo. A generalized Portmanteau goodness of fit test for time series. *Econometric Theory*, 20:382–416, 2004.

R. Dahlhaus and D. Janas. A frequency domain bootstrap for ratio statistics in time series analysis. *Annals of Statistics*, 24:1934–1963, 1996.

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.

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.

P. Fryzlewicz and G. Nason. A Haar-Fisz algorithm for poisson intensity estimation. *J. Comp. Graph. Stat.*, 13:621–638, 2004.

P. Fryzlewicz and G. Nason. Haar-Fisz estimation of evolutionary wavelet spectra. *Journal of the Royal Statistical Society Series B*, 68:611–634, 2006.

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 Adminstration*, 1987.

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-P. Kreiss and S. N. Lahiri. *Handbook of Statistics, volume 30*, chapter Bootstrap methods for time series, pages 3–26. Elsevier, 2012.

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*, 2012.

I. N. Lobato. Testing that a dependent process is uncorrelated. *Journal of the American Statistical Association*, 96:1066–1076, 2001.

A. Milhoj. A test of fit in time series. *Biometrika*, 68:177–188, 1981.

D. N. Politis, J. P. Romano, and M. Wold. Subsampling for heteroskadastic 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. A self-normalized approach to confidence interval construction in time series. *Journal of the Royal Statistical Society (B)*, 72:343–366, 2010a.

X. Shao. The dependent wild bootstrap. *Journal of the American Statistical Association*, 105:218–235, 2010b.

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

W. B. Wu. Nonlinear system theory: Another look at dependence. *Proceedings of the National Academy of Sciences*, 102:14150–14154, 2005.

W. B. Wu and X. Shao. A limit theorem for quadratic forms and its applications. *Econometric Theory*, 23:930–951, 2007.

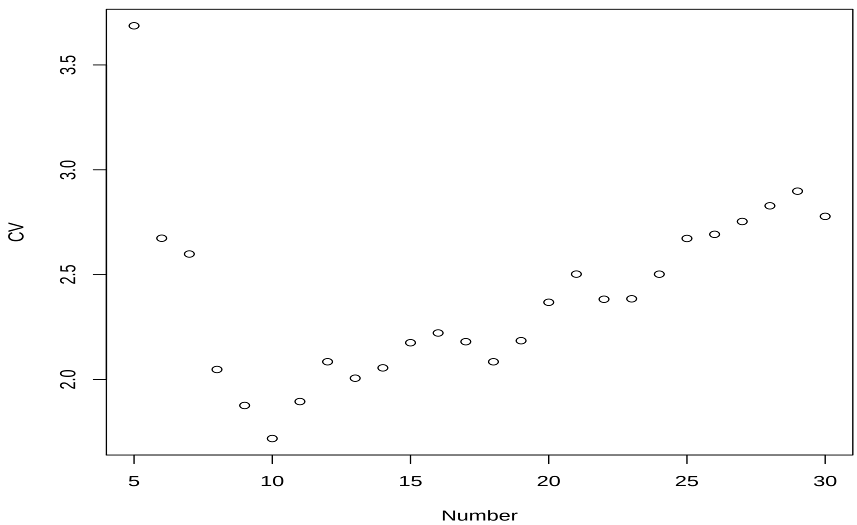

Figure 1: Cross-validation criterion for the sample autocovariance function $\mathcal{C}_{e^{i\cdot}}(M)$ at lag one for the Gaussian autoregressive time series $X_t = 1.5X_{t-1} - 0.75X_{t-2} + \varepsilon_t$ where $T = 200$.