



Goodness-of-fit testing for time series models via distance covariance

Phyllis Wan ^{a,*}, Richard A. Davis ^b

^a *Econometric Institute, Erasmus University Rotterdam, P.O. Box 1738, 3000DR Rotterdam, The Netherlands*

^b *Department of Statistics, Columbia University, 1255 Amsterdam Avenue, MC 4690, New York, NY 10027, USA*



ARTICLE INFO

Article history:

Received 28 November 2018

Received in revised form 21 April 2020

Accepted 2 May 2020

Available online 10 August 2020

JEL classification:

C01

C13

C22

Keywords:

Distance covariance

Time series models

Estimated residuals

Goodness-of-fit testing

Serial dependence

ABSTRACT

In many statistical modeling frameworks, goodness-of-fit tests are typically administered to the estimated residuals. In the time series setting, whiteness of the residuals is assessed using the sample autocorrelation function. For many time series models, especially those used for financial time series, the key assumption on the residuals is that they are in fact independent and not just uncorrelated. In this paper, we apply the auto-distance covariance function (ADCV) to evaluate the serial dependence of the estimated residuals. Distance covariance can discriminate between dependence and independence of two random vectors. The limit behavior of the test statistic based on the ADCV is derived for a general class of time series models. One of the key aspects in this theory is adjusting for the dependence that arises due to parameter estimation. This adjustment has essentially the same form regardless of the model specification. We illustrate the results in simulated examples.

© 2020 Elsevier B.V. All rights reserved.

1. Introduction

Let $\{X_j, j \in \mathbb{Z}\}$ be a stationary time series of random variables with finite mean and variance. Given consecutive observations of this time series X_1, \dots, X_n , we consider testing the plausibility that the data were generated from a parametric model. We consider causal models of the form

$$X_j = g(Z_{-\infty:j}; \beta), \quad (1.1)$$

where the Z_j 's are independent and identically distributed (iid) with mean zero and finite variance, $Z_{n_1:n_2}$ denotes the sequence $\{Z_j, n_1 \leq j \leq n_2\}$, and $\beta \in \mathbb{R}^d$ is the parameter vector. Assume further that the model (1.1) has the invertible representation

$$Z_j = h(X_{-\infty:j}; \beta). \quad (1.2)$$

The objective of this paper is to provide a validity check of the model (1.1) by testing the estimated residuals for independence.

Given observations $X_{1:n}$ and $\hat{\beta}$, an estimator for β , the innovations $\{Z_j\}$ can be approximated by the residuals based on the infinite sequence $X_{-\infty:j}$, defined as

$$\tilde{Z}_j := Z_j(\hat{\beta}) = h(X_{-\infty:j}; \hat{\beta}). \quad (1.3)$$

* Corresponding author.

E-mail addresses: wan@ese.eur.nl (P. Wan), rdavis@stat.columbia.edu (R.A. Davis).

Since we do not observe X_j for $j \leq 0$, we instead use the estimated residuals

$$\hat{Z}_j := h(Y_{-\infty:j}; \hat{\beta}), \quad j = 1, \dots, n, \tag{1.4}$$

where $\{Y_j\}$ is the infinite sequence with $Y_j = X_j, j \geq 1$ and $Y_j = 0$ for $j \leq 0$. If the time series $\{X_j\}$ is stationary and ergodic, the influence of $X_{-\infty:0}$ in (1.3) becomes negligible for large j and \hat{Z}_j and \tilde{Z}_j become close.

It is general practice to inspect $\{\hat{Z}_j\}$ for goodness-of-fit of the time series model. If (1.1) correctly describes the generating mechanism of $\{X_j\}$, one would expect $\{\hat{Z}_j\}$ to behave similarly as $\{Z_j\}$. However, the sequence $\{\hat{Z}_j, 1 \leq j \leq n\}$ is not iid since they are functions of $\hat{\beta}$, hence certain properties of $\{\hat{Z}_j\}$ can differ from that of $\{Z_j\}$, which in turn may impact sample statistics such as the sample autocorrelation of the residuals. This has been noted for specific time series models in the literature. For example, for the ARMA model, corrections have been made for statistics based on the residuals, see Section 9.4 of Brockwell and Davis (1991). For heteroscedastic GARCH models, the moment sum process of the residuals is notably different from that of iid innovations, see Kulperger and Yu (2005). Though $\{\tilde{Z}_j\}$ should be nearly independent under the true model assumption, the discrepancy between $\{\hat{Z}_j\}$ and $\{Z_j\}$ should be taken into account when designing a goodness-of-fit test.

In this paper, we characterize the serial dependence of the estimated residuals $\{\hat{Z}_j\}$ using distance covariance (Feuerverger, 1993; Székely et al., 2007). Distance covariance is a useful dependence measure with the ability to detect both linear and nonlinear dependence. It is zero if and only if independence occurs. Similar to the autocovariance function, Zhou (2012) proposed the auto-distance covariance function (ADCV) as a measure of serial dependence for time series. In this paper, we study the ADCV of the residuals and derive its limit when the model is correctly specified. We show that the limiting distribution of the ADCV of $\{\hat{Z}_j\}$ differs from that of its iid counterpart $\{Z_j\}$ and quantify the difference. While this limit is not distribution free, the parametric bootstrap can be used to approximate limit quantities. This is an extension of Section 4 of Davis et al. (2018) which considered this problem for AR processes.

Based on our theory, we propose to assess the goodness-of-fit test of time series models by applying a Box–Pierce type of statistic based on ADCV of the residuals. A similar test statistic has been proposed in the literature for identifying iid sequences (Hong, 1999; Fokianos and Pitsillou, 2017); however, the difference in limiting behavior of ADCV needs to be taken into consideration when applying to estimated residuals. The aim of this paper is to provide the theoretical background in order to apply this test to time series residuals.

The remainder of the paper is structured as follows. An introduction to distance correlation and ADCV along with some historical remarks are given in Section 2. In Section 3, we provide the limit result for the ADCV of the residuals for a general class of time series models. To implement the limiting results, we apply the parametric bootstrap, the methodology and theoretical justification of which is given in Section 4. We then apply the result to ARMA and GARCH models in Sections 5 and 6 and illustrate with simulation studies. A simulated example where the data does not conform with the model is demonstrated in Section 7. In Section 8, we apply the goodness-of-fit test to differentiate time series generated from an AR(1) model with iid noise and one with GARCH(1,1) noise where we demonstrate the power of our test. Lastly, in Section 9, the goodness-of-fit is applied to assess the validity of an AR model in a real example.

2. Distance covariance

Let $X \in \mathbb{R}^p$ and $Y \in \mathbb{R}^q$ be two random vectors, potentially of different dimensions. Let $\varphi_{X,Y}(s, t), \varphi_X(s), \varphi_Y(t)$ denote the joint and marginal characteristic functions of (X, Y) . We know that

$$X \perp\!\!\!\perp Y \iff \varphi_{X,Y}(s, t) = \varphi_X(s)\varphi_Y(t).$$

The distance covariance between X and Y is defined as

$$T(X, Y; \mu) = \int_{\mathbb{R}^{p+q}} |\varphi_{X,Y}(s, t) - \varphi_X(s)\varphi_Y(t)|^2 \mu(ds, dt), \quad (s, t) \in \mathbb{R}^{p+q},$$

where μ is a suitable measure on \mathbb{R}^{p+q} . In order to ensure that $T(X, Y; \mu)$ is well-defined, one of the following conditions is assumed to be satisfied (Davis et al., 2018):

1. μ is a finite measure;
2. μ is an infinite measure such that

$$\int_{\mathbb{R}^{p+q}} (1 \wedge |s|^\alpha)(1 \wedge |t|^\alpha)\mu(ds, dt) < \infty \quad \text{and} \quad \mathbb{E}[|XY|^\alpha + |X|^\alpha + |Y|^\alpha] < \infty, \quad \text{for some } \alpha \in (0, 2].$$

If μ has a positive Lebesgue density on \mathbb{R}^{p+q} , then X and Y are independent if and only if $T(X, Y; \mu) = 0$.

For a stationary series $\{X_j\}$, the auto-distance covariance (ADCV) is given by

$$T_h(X; \mu) := T(X_0, X_h; \mu) = \int_{\mathbb{R}^2} |\varphi_{X_0, X_h}(s, t) - \varphi_X(s)\varphi_X(t)|^2 \mu(ds, dt), \quad (s, t) \in \mathbb{R}^2.$$

Given observations $\{X_j, 1 \leq j \leq n\}$, the ADCV can be estimated by its sample version

$$\hat{T}_h(X; \mu) := \int_{\mathbb{R}^2} |C_n^X(s, t)|^2 \mu(ds, dt), \quad (s, t) \in \mathbb{R}^2,$$

where

$$C_n^X(s, t) := \frac{1}{n} \sum_{j=1}^{n-h} e^{isX_j + itX_{j+h}} - \frac{1}{n} \sum_{j=1}^{n-h} e^{isX_j} \frac{1}{n} \sum_{j=1}^{n-h} e^{itX_{j+h}}.$$

If we assume that $\mu = \mu_1 \times \mu_2$ and is symmetric about the origin, then under the conditions where $T_h(X; \mu)$ exists, $\hat{T}_h(X; \mu)$ is computable in an alternative expression similar to a V -statistic, see Section 2.2 of [Davis et al. \(2018\)](#) for details. It can be shown that if the X_j 's are iid, the process $\sqrt{n}C_n^X(s, t)$ converges weakly,

$$\sqrt{n}C_n^X \xrightarrow{d} G_h \quad \text{on } \mathcal{C}(K), \quad (2.1)$$

for any compact set $K \subset \mathbb{R}^2$, and

$$n\hat{T}_h(X; \mu) \xrightarrow{d} \int |G_h|^2 \mu(ds, dt),$$

where G_h is a zero-mean Gaussian process with covariance structure

$$\begin{aligned} \Gamma((s, t), (s', t')) &= \text{cov}(G_h(s, t), G_h(s', t')) \\ &= \mathbb{E}[(e^{i(s, X_0)} - \varphi_X(s))(e^{i(t, X_h)} - \varphi_X(t)) \\ &\quad \times (e^{-i(s', X_0)} - \varphi_X(-s'))(e^{-i(t', X_h)} - \varphi_X(-t'))]. \end{aligned}$$

Much literature on distance covariance focus on the specific weight measure $\mu(s, t)$ with density proportional to $|s|^{-p-1}|t|^{-q-1}$. This distance covariance has the advantage of being scale and rotational invariant, but imposes moment constraints on the variables under consideration. In our case, as will be shown in Section 3, this measure may not work when applied to the residuals (see also Section 4 of [Davis et al. \(2018\)](#) for a counterexample). To avoid this difficulty, we assume a finite measure for μ . In this case $\hat{T}_h(X; \mu)$ has the computable form

$$\begin{aligned} \hat{T}_h(X; \mu) &= \frac{1}{(n-h)^2} \sum_{i,j=1}^{n-h} \hat{\mu}(X_i - X_j, X_{i+h} - X_{j+h}) \\ &\quad + \frac{1}{(n-h)^4} \sum_{i,j,k,l=1}^{n-h} \hat{\mu}(X_i - X_j, X_{k+h} - X_{l+h}) \\ &\quad - 2 \frac{1}{(n-h)^3} \sum_{i,j,k=1}^{n-h} \hat{\mu}(X_i - X_j, X_{i+h} - X_{k+h}), \end{aligned}$$

where $\hat{\mu}(x, y) = \int \exp(isx + ity)\mu(ds, dt)$ is the Fourier transform with respect to μ .

The concept of distance covariance was first proposed by [Feuerverger \(1993\)](#) in the bivariate case and later popularized by [Székely et al. \(2007\)](#). The idea of ADCV was first introduced by [Zhou \(2012\)](#). For distance covariance in the time series context, we refer to [Davis et al. \(2018\)](#) for theory in a general framework.

For testing serial dependence in time series, [Fokianos and Pitsillou \(2017\)](#) suggested a Box-Ljung-type test statistic using the ADCV and pointed out that this corresponds to the serial dependence measure using the generalized spectral density proposed in [Hong \(1999\)](#). We will elaborate more on this in Section 8. The concept of distance covariance is closely related to the Hilbert-Schmidt Independence Criterion (HSIC), see [Gretton et al. \(2005\)](#). For example, the distance covariance with Gaussian measure coincides with the HSIC with a Gaussian kernel. In recent work, [Wang et al. \(2018\)](#) use HSIC for testing the cross dependence between two time series. The context behind these papers is assessing the serial and cross dependence between time series. In this paper, our focus is to apply ADCV in the context of model testing such that these techniques can be applied to the estimated residuals.

3. General result

Let X_1, \dots, X_n be observations from a stationary time series $\{X_j\}$ generated from (1.1) with $\beta = \beta_0$. Let $\hat{Z}_1, \dots, \hat{Z}_n$ be the estimated residual calculated through (1.4). In this section, we examine the ADCV of the residuals

$$\hat{T}_h(\hat{Z}; \mu) := \|C_n^{\hat{Z}}\|_{\mu}^2 = \int |C_n^{\hat{Z}}|^2 \mu(ds, dt),$$

where

$$C_n^{\hat{Z}}(s, t) := \frac{1}{n} \sum_{j=1}^{n-h} e^{is\hat{Z}_j + it\hat{Z}_{j+h}} - \frac{1}{n} \sum_{j=1}^{n-h} e^{is\hat{Z}_j} \frac{1}{n} \sum_{j=1}^{n-h} e^{it\hat{Z}_{j+h}}.$$

To provide the limiting result for $\hat{T}_h(\hat{Z}; \mu)$, we require the following assumptions.

(M1) Let \mathcal{F}_j be the σ -algebra generated by $\{X_k, k \leq j\}$. We assume that the parameter estimate $\hat{\beta}$ is of the form

$$\sqrt{n}(\hat{\beta} - \beta_0) = \frac{1}{\sqrt{n}} \sum_{j=1}^n \mathbf{m}(Z_{-\infty:j}; \beta_0) + o_p(1), \tag{3.1}$$

where \mathbf{m} is a vector-valued function of the infinite sequence $X_{-\infty:j}$ such that

$$\mathbb{E}[\mathbf{m}(Z_{-\infty:j}; \beta_0) | \mathcal{F}_{j-1}] = \mathbf{0}, \quad \mathbb{E}|\mathbf{m}(Z_{-\infty:0}; \beta_0)|^2 < \infty. \tag{3.2}$$

This representation can be readily found in most likelihood-based estimators, for example, the Yule–Walker estimator for AR processes, quasi-MLE for GARCH processes, etc. In these cases \mathbf{m} can be taken as the likelihood score function. By the martingale central limit theorem, (3.1) and (3.2) imply that

$$\sqrt{n}(\hat{\beta} - \beta_0) \xrightarrow{d} \mathbf{Q},$$

for a random Gaussian vector \mathbf{Q} .

(M2) Assume that the function h in the invertible representation (1.2) is continuously differentiable, and writing

$$\mathbf{L}_j(\beta) := \frac{\partial}{\partial \beta} h(X_{-\infty:j}; \beta), \tag{3.3}$$

we assume

$$\mathbb{E}\|\mathbf{L}_0(\beta_0)\|^2 < \infty.$$

(M3) Assume that $\{\hat{Z}_j\}$, the estimated residuals based on the *finite* sequence of observations, is close to $\{\tilde{Z}_j\}$, the fitted residuals based on the *infinite* sequence, such that

$$\frac{1}{\sqrt{n}} \sum_{j=1}^n |\hat{Z}_j - \tilde{Z}_j|^k = o_p(1), \quad k = 1, 2.$$

Theorem 3.1. Let X_1, \dots, X_n be a sequence of observations generated from the causal and invertible time series model (1.1) and (1.2) with $\beta = \beta_0$. Let $\hat{\beta}$ be an estimator of β and let $\hat{Z}_1, \dots, \hat{Z}_n$ be the estimated residuals calculated through (1.4) satisfying conditions (M1)–(M3). Furthermore assume that the weight measure μ satisfies

$$\int_{\mathbb{R}^2} [(1 \wedge |s|^2)(1 \wedge |t|^2) + (s^2 + t^2) \mathbf{1}(|s| \wedge |t| > 1)] \mu(ds, dt) < \infty. \tag{3.4}$$

Then

$$n\hat{T}_h(\hat{Z}; \mu) \xrightarrow{d} \|G_h + \xi_h\|_{\mu}^2,$$

where G_h is the limiting distribution for $n\hat{T}_h(Z; \mu)$, the ADCV based on the iid innovations Z_1, \dots, Z_n , and the correction term ξ_h is given by

$$\xi_h(s, t) := it\mathbf{Q}^T \mathbb{E} \left[(e^{isZ_0} - \varphi_Z(s)) e^{itZ_h} \mathbf{L}_h(\beta_0) \right], \tag{3.5}$$

with \mathbf{Q} being the limit distribution of $\sqrt{n}(\hat{\beta} - \beta_0)$ and \mathbf{L}_h as defined in (3.3).

The proof of the theorem is provided in [Appendix A](#).

Remark 3.2. Distance correlation, analogous to linear correlation, is the normalized version of distance covariance, defined as

$$R(X, Y; \mu) := \frac{T(X, Y; \mu)}{\sqrt{T(X, X; \mu)T(Y, Y; \mu)}} \in [0, 1].$$

The auto-distance correlation function (ADCF) of a stationary series $\{X_j\}$ at lag h is given by

$$R_h(X; \mu) := R(X_0, X_h; \mu),$$

and its sample version $\hat{R}_h(X; \mu)$ can be defined similarly. It can be shown that the ADCF for the residuals from an AR(p) model has the limiting distribution (Davis et al., 2018):

$$n\hat{R}_h(\hat{Z}; \mu) \xrightarrow{d} \frac{\|G_h + \xi_h\|_\mu^2}{T_0(Z; \mu)}, \quad (3.6)$$

and the result can be easily generalized to other models. In the examples in Sections 5 and 6, we shall use ADCF in place of ADCV.

4. Parametric bootstrap

The limit in (3.6) is not distribution-free and is generally intractable. In order to use the result, we propose to approximate the limit through the parametric bootstrap described below.

Given observations X_1, \dots, X_n , let $\hat{\beta}$ be the parameter estimate and $\hat{Z}_1, \dots, \hat{Z}_n$ be the estimated residuals. A set of bootstrapped residuals can be obtained as follows:

1. Let \hat{F}_n be the mean-corrected empirical distribution of $\{\hat{Z}_j, 1 \leq j \leq n\}$;
2. Generate X_1^*, \dots, X_n^* from the time series model with parameter value $\hat{\beta}$ and innovation sequence $\{Z_j^*\}$ generated from \hat{F}_n ;
3. Re-fit the time series model. Obtained the parameter estimate $\hat{\beta}^*$ and the estimated residuals $\hat{Z}_1^*, \dots, \hat{Z}_n^*$.

Let $n\hat{T}_h(\hat{Z}^*, \mu)$ be the ADCV calculated from the bootstrapped residuals $\hat{Z}_1^*, \dots, \hat{Z}_n^*$. In Theorem 4.2, we show that when the sample size n is large, the empirical distribution of $\{n\hat{T}_h(\hat{Z}^*, \mu)\}$ forms a good representation of the limiting distribution of $n\hat{T}_h(\hat{Z}, \mu)$, the ADCV of the actual fitted residuals. Before stating the theorem, we first state the relevant conditions. We denote by \mathbb{P}_n and \mathbb{E}_n the probability and expectation conditional on the observations X_1, \dots, X_n .

(M1') Let $\mathcal{F}_j, \mathcal{F}_j^*$ be the σ -algebra generated by $\{Z_k, k \leq j\}$ and $\{Z_k^*, k \leq j\}$, respectively. We assume that condition (M1) holds, i.e., (3.1) and (3.2) hold. In addition, as $n \rightarrow \infty$, for any $\epsilon > 0$,

$$\mathbb{P}_n \left(\left| \frac{1}{n} \sum_{j=1}^n \mathbb{E}_n[\mathbf{m}^T(Z_{-\infty;j}^*; \hat{\beta}) \mathbf{m}(Z_{-\infty;j}^*; \hat{\beta}) | \mathcal{F}_{j-1}^*] - \tau^2 \right| > \epsilon \right) \xrightarrow{p} 0$$

for some $\tau > 0$, and

$$\mathbb{P}_n \left(\frac{1}{n} \sum_{j=1}^n \mathbb{E}_n[\mathbf{m}^T(Z_{-\infty;j}^*; \hat{\beta}) \mathbf{m}(Z_{-\infty;j}^*; \hat{\beta}) \mathbf{1}_{\{|\mathbf{m}(Z_{-\infty;j}^*; \hat{\beta})| > \sqrt{n}\epsilon\}} | \mathcal{F}_{j-1}^*] > \epsilon \right) \xrightarrow{p} 0.$$

(M2') Assume that the function h in the invertible representation (1.2) is continuously differentiable and

$$\mathbb{P} \left[\sup_n \mathbb{E}_n \|\mathbf{L}_0^*(\hat{\beta})\|^2 < \infty \right] = 1,$$

where

$$\mathbf{L}_j^*(\beta) := \frac{\partial}{\partial \beta} h(X_{-\infty;j}^*; \beta). \quad (4.1)$$

(M3') Assume that the estimated residuals based on the finite sequence of observations, \hat{Z}_j^* , are close to the fitted residuals based on the infinite sequence, \tilde{Z}_j^* , such that for any $\epsilon > 0$,

$$\mathbb{P}_n \left(\frac{1}{\sqrt{n}} \sum_{j=1}^n |\hat{Z}_j^* - \tilde{Z}_j^*|^k > \epsilon \right) \rightarrow 0, \quad k = 1, 2.$$

Remark 4.1. Condition (M1') ensures that $\sqrt{n}(\hat{\beta}^* - \hat{\beta})$ provides a good approximation to \mathbf{Q} , the limit of $\sqrt{n}(\hat{\beta} - \beta_0)$. These conditions are standard for the martingale central limit theorem, see, for example, Scott (1973). Conditions (M2') and (M3') are parallel arguments to conditions (M2) and (M3).

Theorem 4.2. Assuming conditions (M1'), (M2') and (M3') hold, the ADCV of the bootstrapped residuals $\{\hat{Z}_{1:n}^*\}$ satisfies

$$\sup_t \left| \mathbb{P}_n \left(n\hat{T}_h(\hat{Z}^*, \mu) \leq t \right) - \mathbb{P} \left(\|G_h + \xi_h\|_\mu^2 \leq t \right) \right| \xrightarrow{p} 0.$$

5. Example: ARMA(p, q)

Consider the causal, invertible ARMA(p, q) process that follows the recursion,

$$X_t = \sum_{i=1}^p \phi_i X_{t-i} + Z_t + \sum_{j=1}^q \theta_j Z_{t-j}, \tag{5.1}$$

where $\beta = (\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q)^T$ is the vector of parameters and $\{Z_t\}$ is iid with mean 0 and variance σ^2 . Denote the AR and MA polynomials by $\phi(z) = 1 - \sum_{i=1}^p \phi_i z^i$ and $\theta(z) = 1 + \sum_{j=1}^q \theta_j z^j$, and let B be the backward operator such that

$$BX_t = X_{t-1}.$$

Then the recursion (5.1) can be represented by

$$\phi(B)X_t = \theta(B)Z_t.$$

It follows from invertibility that $\phi(z)/\theta(z)$ has the power series expansion

$$\frac{\phi(z)}{\theta(z)} = \sum_{j=0}^{\infty} \pi_j(\beta)z^j,$$

where $\sum_{j=0}^{\infty} |\pi_j(\beta)| < \infty$, and

$$Z_t = Z_t(\beta) = \sum_{j=0}^{\infty} \pi_j(\beta)X_{t-j}.$$

Given an estimate of the parameters $\hat{\beta}$, the residuals based on the infinite sequence $\{X_{-\infty:n}\}$ are given by

$$\tilde{Z}_t := Z_t(\hat{\beta}) = \sum_{j=0}^{\infty} \pi_j(\hat{\beta})X_{t-j}.$$

Based on the observed data X_1, \dots, X_n , the estimated residuals are

$$\hat{Z}_t = \sum_{j=0}^{t-1} \pi_j(\hat{\beta})X_{t-j}. \tag{5.2}$$

One choice for $\hat{\beta}$ is the pseudo-MLE based on Gaussian likelihood

$$L(\beta, \sigma^2) \propto \sigma^{-n} |\Sigma|^{-1/2} \exp\left\{-\frac{1}{2\sigma^2} \mathbf{X}_n^T \Sigma^{-1} \mathbf{X}_n\right\},$$

where $\mathbf{X}_n = (X_1, \dots, X_n)^T$ and the covariance $\Sigma = \Sigma(\beta) := \text{Var}(\mathbf{X}_n)/\sigma^2$ is independent of σ^2 . The pseudo-MLE $\hat{\beta}$ and $\hat{\sigma}^2$ are taken to be the values that maximize $L(\beta, \sigma^2)$. It can be shown that $\hat{\beta}$ is consistent and asymptotically normal even for non-Gaussian Z_t (Brockwell and Davis, 1991).

We have the following result for the ADCV of ARMA residuals.

Corollary 5.1. *Let $\{X_t, 1 \leq j \leq n\}$ be observations from a causal and invertible ARMA(p, q) time series and $\{\hat{Z}_t, 1 \leq t \leq n\}$ be the estimated residuals defined in (5.2) using the pseudo-MLE $\hat{\beta}$. Assume that μ satisfies (3.4), then*

$$n\hat{T}_h(\hat{Z}; \mu) \xrightarrow{d} \|G_h + \xi_h\|_{\mu}^2,$$

where (G_h, ξ_h) is a joint Gaussian process defined on \mathbb{R}^2 with G_h as specified in (2.1) and ξ_h in (3.5).

The proof of Corollary 5.1 is given in Appendix C.

Remark 5.2. In the case where the distribution of Z_t is in the domain of attraction of an α -stable law with $\alpha \in (0, 2)$, and the parameter estimator $\hat{\beta}$ has convergence rate faster than $n^{-1/2}$, i.e.,

$$a_n(\hat{\beta} - \beta) = O_p(1), \quad \text{for some } a_n = o(n^{-1/2}),$$

(Davis, 1996), the ADCV of the residuals has limit

$$n\hat{T}_h(\hat{Z}; \mu) \xrightarrow{d} \|G_h\|_{\mu}^2,$$

where the correction term ξ_h disappears. For a proof in the AR(p) case, see Theorem 4.2 of Davis et al. (2018).

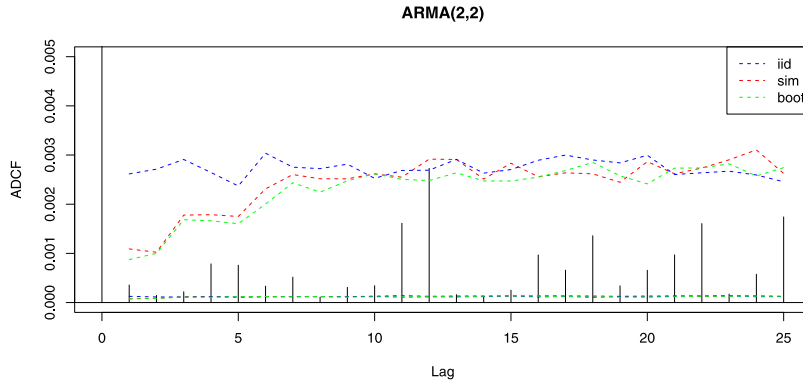


Fig. 1. Empirical 5% and 95% quantiles of the ADF for (a) iid innovations; (b) estimated residuals; (c) bootstrapped residuals; from a ARMA(2,2) model.

5.1. Simulation

We generate time series of length $n = 2000$ from an ARMA(2,2) model with standard normal innovations and parameter values

$$\beta = (\phi_1, \phi_2, \theta_1, \theta_2) = (1.2, -0.32, -0.2, -0.48).$$

For each simulation, an ARMA(2,2) model is fitted to the data. In Fig. 1, we compare the empirical 5% and 95% quantiles for the ADF of

- (a) iid innovations from 1000 independent simulations;
- (b) estimated residuals from 1000 independent simulations of $\{X_t\}$;
- (c) estimated residuals through 1000 independent parametric bootstrap samples from one realization of $\{X_t\}$.

In order to satisfy condition (3.4), the ADFs are evaluated using the Gaussian weight measure $N(0, 0.5^2)$. Confirming the results in Theorem 3.1 and Corollary 5.1, the simulated quantiles of $\hat{R}_h(\tilde{Z}; \mu)$ differ significantly from that of $\hat{R}_h(Z; \mu)$, especially when h is small. Given one realization of the time series, the quantiles estimated by the parametric bootstrap correctly capture this effect.

6. Example: GARCH(p, q)

In this section, we consider the GARCH(p, q) model,

$$X_t = \sigma_t Z_t,$$

where the Z_t 's are iid innovations with mean 0 and variance 1 and

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^p \alpha_i X_{t-i}^2 + \sum_{j=1}^q \beta_j \sigma_{t-j}^2, \quad \alpha_0 > 0, \alpha_i \geq 0, \beta_j \geq 0. \tag{6.1}$$

Let $\theta = (\alpha_0, \alpha_1, \dots, \alpha_p, \beta_1, \dots, \beta_q)$ denote the parameter vector. We write the conditional variance $\sigma_t^2 = \sigma_t^2(\theta)$ to denote it as a function of θ .

Iterating the recursion in (6.1) gives

$$\sigma_t^2(\theta) = c_0(\theta) + \sum_{i=1}^{\infty} c_i(\theta) X_{t-i}^2,$$

for suitably defined functions c_i 's, see Berkes et al. (2003). Given an estimator $\hat{\theta}$, an estimator for $\sigma_t^2(\theta)$ based on the infinite sequence $\{X_j, j \leq t\}$ can be written as

$$\tilde{\sigma}_t^2 := \sigma_t^2(\hat{\theta}_n) = c_0(\hat{\theta}_n) + \sum_{i=1}^{\infty} c_i(\hat{\theta}_n) X_{t-i}^2,$$

and the unobserved residuals are given by

$$\tilde{Z}_t = X_t / \tilde{\sigma}_t.$$

In practice, $\hat{\sigma}_t^2$ can be approximated by the truncated version

$$\hat{\sigma}_t^2(\hat{\theta}_n) := c_0(\hat{\theta}_n) + \sum_{i=1}^{t-1} c_i(\hat{\theta}_n)X_{t-i}^2,$$

and the estimated residual \hat{Z}_t is given by

$$\hat{Z}_t = X_t / \hat{\sigma}_t. \tag{6.2}$$

Define the parameter space by

$$\Theta = \{\mathbf{u} = (s_0, s_1, \dots, s_p, t_1, \dots, t_q) : t_1 + \dots + t_q \leq \rho_0, \underline{u} \leq \min(\mathbf{u}) \leq \max(\mathbf{u}) \leq \bar{u}\},$$

for some $0 < \underline{u} < \bar{u}$, $0 < \rho_0 < 1$ and $q\underline{u} < \rho_0$, and assume the following conditions:

(Q1) The true value θ lies in the interior of Θ .

(Q2) For some $\zeta > 0$,

$$\lim_{x \rightarrow 0} x^{-\zeta} \mathbb{P}\{|Z_0| \leq x\} = 0.$$

(Q3) For some $\delta > 0$,

$$\mathbb{E}|Z_0|^{4+\delta} < \infty.$$

(Q4) The GARCH(p, q) representation is minimal, i.e., the polynomials $A(z) = \sum_{i=1}^p \alpha_i z^i$ and $B(z) = 1 - \sum_{j=1}^q \beta_j z^j$ do not have common roots.

Given observations $\{X_t, 1 \leq t \leq n\}$, [Berkes et al. \(2003\)](#) proposed a quasi-maximum likelihood estimator for θ given by

$$\hat{\theta}_n := \arg \max_{\mathbf{u} \in \Theta} \sum_{t=1}^n l_t(\mathbf{u}),$$

where

$$l_t(\mathbf{u}) := -\frac{1}{2} \log \hat{\sigma}_t^2(\mathbf{u}) - \frac{X_t^2}{2\hat{\sigma}_t^2(\mathbf{u})}.$$

Provided that (Q1)–(Q4) are satisfied, the quasi-MLE $\hat{\theta}_n$ is consistent and asymptotically normal.

Consider the estimated residuals for the GARCH(p, q) model based on $\hat{\theta}_n$. We have the following result.

Corollary 6.1. *Let $\{X_t, 1 \leq j \leq n\}$ be observations from a GARCH(p, q) time series and $\{\hat{Z}_t, 1 \leq t \leq n\}$ be the estimated residuals defined in (6.2) based on the quasi-MLE $\hat{\theta}_n$. Assume that (Q1)–(Q4) hold and that μ satisfies (3.4), we have*

$$n\hat{T}_h(\hat{Z}; \mu) \xrightarrow{d} \|G_h + \xi_h\|_\mu^2,$$

where (G_h, ξ_h) is a joint Gaussian process defined on \mathbb{R}^2 with G_h as specified in (2.1) and ξ_h in (3.5).

The proof of [Corollary 6.1](#) is given in [Appendix D](#).

6.1. Simulation

We generate time series of length $n = 2000$ from a GARCH(1,1) model with parameter values

$$\theta = (\alpha_0, \alpha_1, \beta_1) = (0.5, 0.1, 0.8).$$

For each simulation, a GARCH(1,1) model is fitted to the data. In [Fig. 2](#), we compare the empirical 5% and 95% quantiles for the ADCF of

- (a) iid innovations from 1000 independent simulations;
- (b) estimated residuals from 1000 independent simulations of $\{X_t\}$;
- (c) estimated residuals through 1000 independent parametric bootstrap samples from one realization of $\{X_t\}$.

Again the ADCFs are based on the Gaussian weight measure $N(0, 0.5^2)$. The difference between the quantiles of $\hat{R}_h(\hat{Z}; \mu)$ and $\hat{R}_h(Z; \mu)$ can be observed. For this GARCH model, the correction has the opposite effect than in the previous ARMA example – the ADCF for residuals is larger than that for iid variables, especially for small lags.

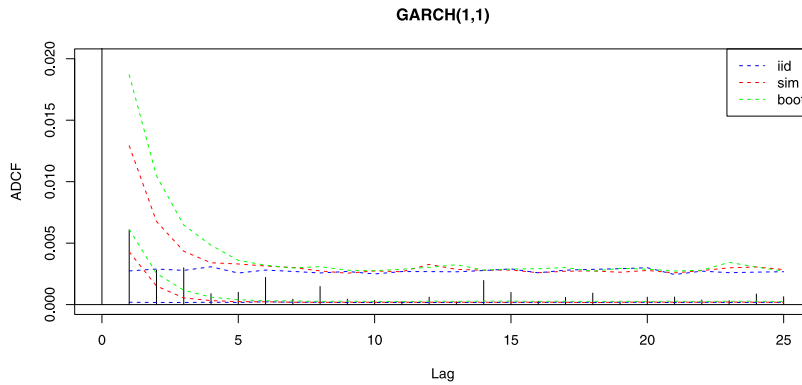


Fig. 2. Empirical 5% and 95% quantiles of the ADFCF for (a) iid innovations; (b) estimated residuals; (c) bootstrapped residuals; from a GARCH(1,1) model.

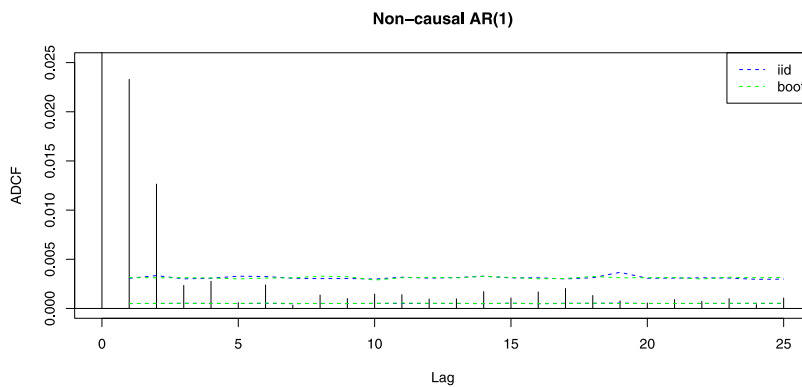


Fig. 3. Empirical 5% and 95% quantiles of the ADFCF for (a) iid innovations; (b) bootstrapped residuals; from non-causal AR(1) data fitted with a causal AR(1) model.

7. Example: Non-causal AR(1)

In this section, we consider an example where the model is misspecified. We generate time series of length $n = 2000$ from a non-causal AR(1) model

$$X_t = \phi X_{t-1} + Z_t$$

with $\phi = 1.67$ and Z_t 's from a t -distribution with 2.5 degrees of freedom. Then we fit a causal AR(1) model, where $|\phi| < 1$, to the data and obtain the corresponding residuals. Again we use the Gaussian weight measure $N(0, 0.5^2)$ when evaluating the ADFCF of the residuals. In Fig. 3, the 5% and 95% ADFCF quantiles are plotted for:

- estimated residuals from 1000 independent simulations of $\{X_t\}$;
- estimated residuals through 1000 independent parametric bootstrap samples from one realization of $\{X_t\}$.

The ADFCFs of the bootstrapped residuals provide an approximation for the limiting distribution of the ADFCF of the residuals given the model is correctly specified. In this case, the ADFCFs of the estimated residuals significantly differ from the quantiles of that of the bootstrapped residuals. This indicates the time series does not come from the assumed causal AR model.

8. Example: Assessing the goodness-of-fit of AR model

In this section, we illustrate the use of our methodology for testing the compatibility of the estimated residuals to the independent noise assumption. In this example, the null model we use is the AR(1) model with independent noise; that is, the time series $\{X_j\}$ is generated from

$$X_j = \phi X_{j-1} + Z_j, \quad (8.1)$$

where $\{Z_j\}$ is a sequence of iid noise with mean 0 and finite variance. Given observations of a time series, we are interested in testing the hypotheses

H_0 : the data generating model for the observations is (8.1);

H_a : the data generating model for the observations is not (8.1).

The test will be performed based on the estimated residuals from fitting the AR(1) model to the observed data. Let $\hat{R}_h(\hat{Z}; \mu)$ denote the ADCF of the estimated residuals at lag h . Under the null hypothesis, $\hat{R}_h(\hat{Z}; \mu)$ should follow the convergence in (3.6). We use a Box–Pierce type statistic

$$B_n = \sum_{h=1}^p (1 - h/n) \hat{R}_h(\hat{Z}; \mu), \tag{8.2}$$

whose convergence under H_0 can be justified from (3.6). Note that the statistic B_n belongs to the family of measures

$$\sum_{j=1}^{n-1} (n-j) k^2(j/p) \hat{R}_h(\hat{Z}; \mu),$$

where $k(\cdot)$ is a suitable kernel, which was proposed by Hong (1999) for detecting serial dependence in time series, and later summarized in Fokianos and Pitsillou (2017) in the expression of ADCF.

In the following simulation study, we generate time series from AR(1) model with GARCH(1,1) noise, where the noise sequence $\{Z_j\}$ is generated from

$$Z_j = \sigma_j \epsilon_j,$$

where $\epsilon_j \stackrel{iid}{\sim} N(0, 1)$,

$$\sigma_j^2 = 1 + \alpha Z_{j-1}^2 + \beta \sigma_{j-1}^2,$$

and $\alpha, \beta \geq 0, \alpha + \beta < 1$. This modeling approach is commonly used in the context of financial time series. The case $(\alpha, \beta) = (0, 0)$ corresponds to the null model whereas otherwise corresponds to the alternative model. We simulate time series from various parameter sets and test for the goodness-of-fit of the null model. For each time series, the test statistic (8.2) and its distribution is approximated using the parametric bootstrap. As with the previous sections, the ADCF is calculated using $N(0, 0.5^2)$ as the weight measure. We reject the null hypothesis if the test statistic does not lie in its central 95% confidence interval approximated from the bootstrapped samples. The rejection rates are recorded in Table 1 (in %). Each rejection rate is based on 500 replications of the same test. The columns with $(\alpha, \beta) = (0, 0)$ correspond to the scenarios where the model is correctly specified and hence give the size of the test, whereas the other columns give the power of the test relative to a GARCH alternative for the noise. We see that our test produces stable sizes and satisfactory power as the alternative model deviates from the null assumption.

We compare our test with the standard approach that uses the Ljung–Box statistic on the residuals (Ljung and Box, 1978) by replicating the same experiment. The Ljung–Box statistic is defined as

$$Q = n(n+2) \sum_{h=1}^p \frac{1}{n-h} \hat{\rho}_h(\hat{Z}),$$

where $\hat{\rho}_h(\hat{Z})$ is the empirical auto-correlation function (ACF) of the fitted residual at lag h . This statistic captures serial correlation but might not be powerful in the presence of non-linear dependence. Table 2 shows the rejection rates of the test using the Ljung–Box statistic for the same sets of models. The distribution of Q is approximated using parametric bootstrap samples. In comparison with ADCF, the Ljung–Box statistic exhibits poor power in the case of model misspecification.

9. Example: Kilkenny wind speed data

In this section, we revisit the example of the daily averages of wind speeds at Kilkenny’s synoptic meteorological station in Ireland, which was previously considered in Davis et al. (2018), Section 5.2. The time series consists of 6226 observations from 1/1/1961 to 1/17/1978. A square root transformation has been applied to stabilize the variance. In Davis et al. (2018), an AR(9) model was fitted to the time series based on the suggestion of prior analyses. The fitted residuals exhibited no obvious auto-correlation and its ADCFs lie within the confidence band for that of iid observations. However, parametric bootstrap showed that the ADCF of the residuals exceeded its bootstrapped confidence bands at initial lags, posing questions to the goodness-of-fit of the AR model.

In Table 3 we formally test the goodness-of-fit of the AR(9) model using the statistic in (8.2) and compare its performance with the Ljung–Box statistic. For both statistics, the test distributions are approximated by the parametric bootstrap. As we can see, while the Ljung–Box statistics are not significant, our test rejects the AR(9) model and indicates that the fitted residuals are not compatible with an AR(9) model driven by iid noise.

Table 1

Rejection rate (in %) for the goodness-of-fit of AR (1) model on data generated from AR(1) model with GARCH(1,1) noise using the statistic (8.2). The rejection rates are calculated for varying number of lags p , sample size n , AR coefficient ϕ and GARCH coefficient (α, β) . The columns with $(\alpha, \beta) = (0, 0)$ correspond to the null model and hence represent the size of the test. The columns with $(\alpha, \beta) \neq (0, 0)$ correspond to the power on misspecified model.

$n = 200$									
	$\phi = 0.2$			$\phi = 0.5$			$\phi = 0.8$		
	(α, β)			(α, β)			(α, β)		
	(0,0)	(.3,.3)	(.3,.6)	(0,0)	(.3,.3)	(.3,.6)	(0,0)	(.3,.3)	(.3,.6)
$p = 2$	5.2	26.0	88.2	4.8	21.6	88.0	2.8	24.4	88.4
$p = 5$	4.8	19.4	80.0	4.8	17.0	80.8	3.8	19.2	84.2
$p = 8$	5.6	16.2	75.2	5.0	13.0	74.8	5.4	13.4	76.6
$n = 500$									
	$\phi = 0.2$			$\phi = 0.5$			$\phi = 0.8$		
	(α, β)			(α, β)			(α, β)		
	(0,0)	(.3,.3)	(.3,.6)	(0,0)	(.3,.3)	(.3,.6)	(0,0)	(.3,.3)	(.3,.6)
$p = 2$	4.0	66.2	100.0	4.0	64.2	100.0	4.6	63.4	100.0
$p = 5$	5.2	46.4	99.8	4.6	51.0	100.0	3.8	42.4	99.4
$p = 8$	5.6	37.8	99.4	4.6	41.2	98.8	4.4	33.0	99.6

Table 2

Rejection rate (in %) for the goodness-of-fit of AR (1) model on data generated from AR(1) model with GARCH(1,1) noise using the Ljung–Box statistics based on ACF. The rejection rates are calculated for varying number of lags p , sample size n , AR coefficient ϕ and GARCH coefficient (α, β) . The columns with $(\alpha, \beta) = (0, 0)$ correspond to the null model and hence represent the size of the test. The columns with $(\alpha, \beta) \neq (0, 0)$ correspond to the power on misspecified model.

$n = 200$									
	$\phi = 0.2$			$\phi = 0.5$			$\phi = 0.8$		
	(α, β)			(α, β)			(α, β)		
	(0,0)	(.3,.3)	(.3,.6)	(0,0)	(.3,.3)	(.3,.6)	(0,0)	(.3,.3)	(.3,.6)
$p = 2$	4.0	6.6	15.2	5.2	11.0	20.0	6.4	9.4	16.0
$p = 5$	7.4	6.6	10.6	5.6	6.4	12.4	4.2	5.4	13.8
$p = 8$	6.6	6.0	10.2	4.6	5.0	9.0	5.8	8.6	10.0
$n = 500$									
	$\phi = 0.2$			$\phi = 0.5$			$\phi = 0.8$		
	(α, β)			(α, β)			(α, β)		
	(0,0)	(.3,.3)	(.3,.6)	(0,0)	(.3,.3)	(.3,.6)	(0,0)	(.3,.3)	(.3,.6)
$p = 2$	6.2	9.0	16.8	6.0	8.8	20.2	7.6	9.2	19.6
$p = 5$	5.6	7.0	12.0	6.2	6.8	15.0	5.0	6.2	16.2
$p = 8$	5.4	5.8	10.8	6.2	6.8	12.4	5.2	5.0	14.0

Table 3

The p -values for testing the goodness-of-fit of the AR(9) model on the Kilkenny wind speed time series using the Ljung–Box statistic and Box–Pierce statistic based on ADCF, both for lag options $p = 5, 10, 20$.

p -value	Ljung–Box			Box–Pierce based on ADCF		
	$p = 5$	$p = 10$	$p = 20$	$p = 5$	$p = 10$	$p = 20$
	0.840	0.656	0.260	0.000	0.004	0.014

10. Conclusion

In this paper, we propose a goodness-of-fit procedure for time series models by examining the serial dependence of estimated residuals. The dependence is measured using the auto-distance covariance function (ADCV) and its limiting behavior is derived for general classes of time series models. We show that the limiting law often differs from that of the ADCV based on iid innovations by a correction term. This indicates that adjustments should be made when testing the goodness-of-fit of the model. We illustrate the result on simulated examples of ARMA and GARCH processes and discover that the adjustments could be in either direction – the quantiles of ADCV for residuals could be larger or smaller than that for iid innovations. We also studied an example when a non-causal AR process was incorrectly fitted with a causal model and showed that ADCV correctly detected model misspecification when applied to the residuals.

Appendices

In the following [Appendices](#), we provide proofs to [Theorem 3.1](#) and [Corollaries 5.1](#) and [6.1](#). Throughout the proofs, c denotes a general constant whose value may change from line to line.

Appendix A. Proof of Theorem 3.1

Proof. The proof proceeds in the following steps with the aid of [Propositions A.1, A.2](#) and [A.3](#). Write

$$n\hat{T}_h(\hat{Z}; \mu) = \|\sqrt{n}C_n^{\hat{Z}}\|_{\mu}^2 = \|\sqrt{n}C_n^{\hat{Z}} - \sqrt{n}C_n^Z + \sqrt{n}C_n^Z\|_{\mu}^2,$$

where

$$C_n^{\hat{Z}}(s, t) := \frac{1}{n} \sum_{j=1}^{n-h} e^{is\hat{Z}_j + it\hat{Z}_{j+h}} - \frac{1}{n} \sum_{j=1}^{n-h} e^{is\hat{Z}_j} \frac{1}{n} \sum_{j=1}^{n-h} e^{it\hat{Z}_{j+h}}$$

and

$$C_n^Z(s, t) := \frac{1}{n} \sum_{j=1}^{n-h} e^{isZ_j + itZ_{j+h}} - \frac{1}{n} \sum_{j=1}^{n-h} e^{isZ_j} \frac{1}{n} \sum_{j=1}^{n-h} e^{itZ_{j+h}}.$$

We first show in [Proposition A.1](#) that

$$(\sqrt{n}(C_n^{\hat{Z}} - C_n^Z), \sqrt{n}C_n^Z) \xrightarrow{d} (\xi_h, G_h), \quad \text{on } \mathcal{C}(K),$$

where K is any compact set in \mathbb{R}^2 . This implies

$$\sqrt{n}C_n^{\hat{Z}} \xrightarrow{d} \xi_h + G_h, \quad \text{on } \mathcal{C}(K).$$

For $\delta \in (0, 1)$, define the compact set

$$K_{\delta} = \{(s, t) | \delta \leq s \leq 1/\delta, \delta \leq t \leq 1/\delta\}.$$

It follows from the continuous mapping theorem that

$$n \int_{K_{\delta}} |C_n^{\hat{Z}}|^2 \mu(ds, dt) \xrightarrow{d} \int_{K_{\delta}} |G_h + \xi_h|^2 \mu(ds, dt).$$

To complete the proof, it remains to justify that we can take $\delta \downarrow 0$. For this it suffices to show that for any $\varepsilon > 0$,

$$\lim_{\delta \rightarrow 0} \limsup_{n \rightarrow \infty} \mathbb{P} \left(\int_{K_{\delta}^c} |\sqrt{n}C_n^{\hat{Z}}|^2 \mu(ds, dt) > \varepsilon \right) = 0,$$

and

$$\lim_{\delta \rightarrow 0} \mathbb{P} \left(\int_{K_{\delta}^c} |G_h + \xi_h|^2 \mu(ds, dt) > \varepsilon \right) = 0.$$

These are shown in [Propositions A.2](#) and [A.3](#), respectively. \square

Proposition A.1. Given the conditions [\(M1\)–\(M3\)](#),

$$(\sqrt{n}(C_n^{\hat{Z}} - C_n^Z), \sqrt{n}C_n^Z) \xrightarrow{d} (\xi_h, G_h), \quad \text{on } \mathcal{C}(K),$$

for any compact $K \subset \mathbb{R}^2$.

Proof. We first consider the marginal convergence of $\sqrt{n}(C_n^{\hat{Z}} - C_n^Z)$. Denote

$$E_n(s, t) := \frac{1}{\sqrt{n}} \sum_{j=1}^{n-h} \left(e^{is\hat{Z}_j + it\hat{Z}_{j+h}} - e^{isZ_j + itZ_{j+h}} \right),$$

then

$$\begin{aligned}
 \sqrt{n}(C_n^{\hat{Z}}(s, t) - C_n^Z(s, t)) &= \frac{1}{\sqrt{n}} \sum_{j=1}^{n-h} \left(e^{is\hat{Z}_j + it\hat{Z}_{j+h}} - e^{isZ_j + itZ_{j+h}} \right) \\
 &\quad - \frac{1}{\sqrt{n}} \sum_{j=1}^{n-h} \left(e^{is\hat{Z}_j} - e^{isZ_j} \right) \frac{1}{n} \sum_{j=1}^{n-h} e^{itZ_{j+h}} \\
 &\quad - \frac{1}{n} \sum_{j=1}^{n-h} e^{is\hat{Z}_j} \frac{1}{\sqrt{n}} \sum_{j=1}^{n-h} \left(e^{it\hat{Z}_{j+h}} - e^{itZ_{j+h}} \right) \\
 &= E_n(s, t) - E_n(s, 0) \frac{1}{n} \sum_{j=1}^{n-h} e^{itZ_{j+h}} - E_n(0, t) \frac{1}{n} \sum_{j=1}^{n-h} e^{is\hat{Z}_j}.
 \end{aligned} \tag{A.1}$$

We now derive the limit of $E_n(s, t)$. Observe that uniformly for $(s, t) \in K$,

$$\begin{aligned}
 E_n(s, t) &= \frac{1}{\sqrt{n}} \sum_{j=1}^{n-h} e^{isZ_j + itZ_{j+h}} \left(e^{is(\hat{Z}_j - Z_j) + it(\hat{Z}_{j+h} - Z_{j+h})} - 1 \right) \\
 &= \frac{1}{n} \sum_{j=1}^{n-h} e^{isZ_j + itZ_{j+h}} (is\sqrt{n}(\hat{Z}_j - Z_j) + it\sqrt{n}(\hat{Z}_{j+h} - Z_{j+h})) + o_p(1), \\
 &= \frac{1}{n} \sum_{j=1}^{n-h} e^{isZ_j + itZ_{j+h}} (is\sqrt{n}(\hat{Z}_j - \tilde{Z}_j) + it\sqrt{n}(\hat{Z}_{j+h} - \tilde{Z}_{j+h})) \\
 &\quad + \frac{1}{n} \sum_{j=1}^{n-h} e^{isZ_j + itZ_{j+h}} (is\sqrt{n}(\tilde{Z}_j - Z_j) + it\sqrt{n}(\tilde{Z}_{j+h} - Z_{j+h})) + o_p(1) \\
 &=: E_{n1}(s, t) + E_{n2}(s, t) + o_p(1).
 \end{aligned}$$

By assumption (M3),

$$|E_{n1}(s, t)| \leq |s| \frac{1}{\sqrt{n}} \sum_{j=1}^{n-h} |\hat{Z}_j - \tilde{Z}_j| + |t| \frac{1}{\sqrt{n}} \sum_{j=1}^{n-h} |\hat{Z}_{j+h} - \tilde{Z}_{j+h}| \xrightarrow{p} 0, \quad \text{in } \mathcal{C}(K).$$

It follows from a Taylor expansion that

$$E_{n2}(s, t) = \sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})^T \frac{1}{n} \sum_{j=1}^{n-h} e^{isZ_j + itZ_{j+h}} (is\mathbf{L}_j(\boldsymbol{\beta}^*) + it\mathbf{L}_{j+h}(\boldsymbol{\beta}^*)),$$

where $\boldsymbol{\beta}^* = \boldsymbol{\beta} + \epsilon(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ for some $\epsilon \in [0, 1]$. Since $\mathbf{L}_j(\boldsymbol{\beta})$ is stationary and ergodic, in view of the uniform ergodic theorem,

$$\frac{1}{n} \sum_{j=1}^{n-h} e^{isZ_j + itZ_{j+h}} (is\mathbf{L}_j(\boldsymbol{\beta}) + it\mathbf{L}_{j+h}(\boldsymbol{\beta})) \xrightarrow{p} \mathbb{E} [e^{isZ_j + itZ_{j+h}} (is\mathbf{L}_j(\boldsymbol{\beta}) + it\mathbf{L}_{j+h}(\boldsymbol{\beta}))] =: \mathbf{C}_h(s, t), \quad \text{in } \mathcal{C}(K).$$

Hence,

$$E_n(s, t) \xrightarrow{d} \mathbf{Q}^T \mathbf{C}_h(s, t), \quad \text{in } \mathcal{C}(K).$$

Note that

$$\frac{1}{n} \sum_{j=1}^{n-h} e^{itZ_{j+h}} \xrightarrow{p} \varphi_Z(t), \quad \text{in } \mathcal{C}(K),$$

and

$$\frac{1}{n} \sum_{j=1}^{n-h} e^{is\hat{Z}_j} = \frac{1}{n} \sum_{j=1}^{n-h} e^{isZ_j} + \frac{1}{\sqrt{n}} E_n(s, 0) \xrightarrow{p} \varphi_Z(s), \quad \text{in } \mathcal{C}(K). \tag{A.2}$$

We have

$$\sqrt{n}(C_n^{\hat{Z}} - C_n^Z) \xrightarrow{d} \mathbf{Q}^T (\mathbf{C}_h(s, t) - \mathbf{C}_h(s, 0)\varphi_Z(t) - \mathbf{C}_h(0, t)\varphi_Z(s)), \quad \text{in } \mathcal{C}(K).$$

To further simplify the above expression, notice that $\mathbf{L}_j(\boldsymbol{\beta})$ is a function of $X_{-\infty:j}$ and independent of Z_{j+h} by causality. Hence

$$\begin{aligned} \mathbf{C}_h(s, t) &= \mathbb{E} \left[e^{isZ_j} i \mathbf{L}_j(\boldsymbol{\beta}) \right] \mathbb{E} \left[e^{itZ_{j+h}} \right] + \mathbb{E} \left[e^{isZ_j + itZ_{j+h}} it \mathbf{L}_{j+h}(\boldsymbol{\beta}) \right] \\ &= \mathbf{C}_h(s, 0) \varphi_Z(t) + \mathbb{E} \left[e^{isZ_j + itZ_{j+h}} it \mathbf{L}_{j+h}(\boldsymbol{\beta}) \right], \end{aligned}$$

and

$$\begin{aligned} \mathbf{Q}^T (\mathbf{C}_h(s, t) - \mathbf{C}_h(s, 0) \varphi_Z(t) - \mathbf{C}_h(0, t) \varphi_Z(s)) \\ = \mathbf{Q}^T (\mathbb{E} \left[e^{isZ_j + itZ_{j+h}} it \mathbf{L}_{j+h}(\boldsymbol{\beta}) \right] - \mathbb{E} \left[e^{itZ_{j+h}} it \mathbf{L}_{j+h}(\boldsymbol{\beta}) \right] \varphi_Z(s)) = \xi_h(s, t). \end{aligned} \tag{A.3}$$

This justifies the marginal convergence of $\sqrt{n}(C_n^Z - C_n^Z)$.

For the joint convergence of $\sqrt{n}(C_n^Z - C_n^Z)$ and $\sqrt{n}C_n^Z$, we recall assumption (M1)

$$\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) = \frac{1}{\sqrt{n}} \sum_{j=1}^n \mathbf{m}(X_{-\infty:j}; \boldsymbol{\beta}) + o_p(1)$$

and also note from the proof of Theorem 1 in Davis et al. (2018) that

$$\sqrt{n}C_n^Z = \frac{1}{\sqrt{n}} \sum_{j=1}^n (e^{isZ_j} - \varphi_Z(s))(e^{itZ_{j+h}} - \varphi_Z(t)) + o_p(1) \xrightarrow{d} G_h, \quad \text{in } C(K).$$

By martingale central limit theorem,

$$\left(\frac{1}{\sqrt{n}} \sum_{j=1}^n \mathbf{m}(X_{-\infty:j}; \boldsymbol{\beta}), \quad \frac{1}{\sqrt{n}} \sum_{j=1}^{n-h} (e^{isZ_j} - \varphi_Z(s))(e^{itZ_{j+h}} - \varphi_Z(t)) \right)$$

converges jointly to (\mathbf{Q}, G_h) . This implies the joint convergence of $\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ and $\sqrt{n}C_n^Z$. Since ξ_h is continuous and its randomness only depends on \mathbf{Q} , the joint convergence $\sqrt{n}C_n^Z$ and $\sqrt{n}C_n^Z - \sqrt{n}C_n^Z$ also follows. \square

Proposition A.2. Under the conditions of Theorem 3.1,

$$\lim_{\delta \rightarrow 0} \limsup_{n \rightarrow \infty} \mathbb{P} \left(\int_{K_\delta^c} |\sqrt{n}C_n^Z|^2 \mu(ds, dt) > \varepsilon \right) = 0.$$

Proof. Using telescoping sums, $C_n^Z - C_n^Z$ has the following decomposition,

$$\begin{aligned} C_n^Z - C_n^Z &= \frac{1}{n} \sum_{j=1}^{n-h} A_j B_j - \frac{1}{n} \sum_{j=1}^{n-h} A_j \frac{1}{n} \sum_{j=1}^{n-h} B_j - \frac{1}{n} \sum_{j=1}^{n-h} U_j \frac{1}{n} \sum_{j=1}^{n-h} B_j - \frac{1}{n} \sum_{j=1}^{n-h} V_j \frac{1}{n} \sum_{j=1}^{n-h} A_j \\ &\quad + \frac{1}{n} \sum_{j=1}^{n-h} U_j B_j + \frac{1}{n} \sum_{j=1}^{n-h} V_j A_j =: \sum_{k=1}^6 I_{nk}(s, t), \end{aligned}$$

where

$$U_j = e^{isZ_j} - \varphi_Z(s), \quad V_j = e^{itZ_{j+h}} - \varphi_Z(t), \quad A_j = e^{is\hat{Z}_j} - e^{isZ_j}, \quad B_j = e^{it\hat{Z}_{j+h}} - e^{itZ_{j+h}}.$$

From a Taylor expansion,

$$\begin{aligned} n|I_{n1}(s, t)|^2 &\leq \left(\frac{1}{\sqrt{n}} \sum_{j=1}^{n-h} |A_j B_j| \right)^2 \\ &\leq \left(\frac{1}{\sqrt{n}} \sum_{j=1}^{n-h} |e^{is(\hat{Z}_j - Z_j)} - 1| |e^{it(\hat{Z}_{j+h} - Z_{j+h})} - 1| \right)^2 \\ &\leq c \left(\frac{1}{\sqrt{n}} \sum_{j=1}^{n-h} (1 \wedge |s| |\hat{Z}_j - Z_j|) (1 \wedge |t| |\hat{Z}_{j+h} - Z_{j+h}|) \right)^2 \end{aligned}$$

$$\begin{aligned} &\leq c \min \left(|s|^2 \left(\frac{1}{\sqrt{n}} \sum_{j=1}^{n-h} |\hat{Z}_j - Z_j| \right)^2, |t|^2 \left(\frac{1}{\sqrt{n}} \sum_{j=1}^{n-h} |\hat{Z}_{j+h} - Z_{j+h}| \right)^2, \right. \\ &\quad \left. |st|^2 \left(\frac{1}{\sqrt{n}} \sum_{j=1}^{n-h} |\hat{Z}_j - Z_j| |\hat{Z}_{j+h} - Z_{j+h}| \right)^2 \right) \\ &\leq c \min \left(|s|^2 \left(\frac{1}{\sqrt{n}} \sum_{j=1}^{n-h} |\hat{Z}_j - Z_j| \right)^2, |t|^2 \left(\frac{1}{\sqrt{n}} \sum_{j=1}^{n-h} |\hat{Z}_{j+h} - Z_{j+h}| \right)^2, \right. \\ &\quad \left. |st|^2 \left(\frac{1}{\sqrt{n}} \sum_{j=1}^{n-h} |\hat{Z}_j - Z_j|^2 \frac{1}{\sqrt{n}} \sum_{j=1}^{n-h} |\hat{Z}_{j+h} - Z_{j+h}|^2 \right) \right) \end{aligned}$$

For $k = 1, 2$,

$$\begin{aligned} \frac{1}{\sqrt{n}} \sum_{j=1}^{n-h} |\hat{Z}_j - Z_j|^k &\leq c \left(\frac{1}{\sqrt{n}} \sum_{j=1}^{n-h} |\hat{Z}_j - \tilde{Z}_j|^k + \frac{1}{\sqrt{n}} \sum_{j=1}^{n-h} |\tilde{Z}_j - Z_j|^k \right) \\ &\leq o_p(1) + c \frac{1}{n^{(k-1)/2}} \|\sqrt{n}(\hat{\beta} - \beta)\|^k \frac{1}{n} \sum_{j=1}^{n-h} \|\mathbf{L}_j(\beta^*)\|^k \\ &= O_p(1). \end{aligned}$$

Therefore

$$n|I_{n1}(s, t)|^2 \leq \min(|s|^2, |t|^2, |st|^2)O_p(1) \leq ((1 \wedge |s|^2)(1 \wedge |t|^2) + (s^2 + t^2) \mathbf{1}(|s| \wedge |t| > 1))O_p(1),$$

where the $O_p(1)$ term does not depend on (s, t) . This implies that

$$\lim_{\delta \rightarrow 0} \limsup_{n \rightarrow \infty} \mathbb{P} \left(\int_{K_\delta^c} n|I_{n1}(s, t)|^2 \mu(ds, dt) > \varepsilon \right) = 0.$$

Similar arguments show that $n|I_{n2}(s, t)|^2$ is bounded by $\min(|s|^2, |t|^2, |st|^2)O_p(1)$, $n|I_{n3}(s, t)|^2$ and $n|I_{n5}(s, t)|^2$ are bounded by $\min(|t|^2, |st|^2)O_p(1)$, and $n|I_{n4}(s, t)|^2$ and $n|I_{n6}(s, t)|^2$ are bounded by $\min(|s|^2, |st|^2)O_p(1)$, and the result of the proposition follows. \square

Proposition A.3. Under the conditions of [Theorem 3.1](#),

$$\lim_{\delta \rightarrow 0} \mathbb{P} \left(\int_{K_\delta^c} |G_h + \xi_h|^2 \mu(ds, dt) > \varepsilon \right) = 0.$$

Proof. Note that

$$\begin{aligned} |\xi(s, t)|^2 &\leq c|t|^2 \|\mathbf{Q}\|^2 \mathbb{E} |e^{isZ_0} - \varphi_Z(s)|^2 \mathbb{E} |\mathbf{L}_h(\beta)|^2 \\ &\leq c|t|^2 \|\mathbf{Q}\|^2 \mathbb{E} [(1 \wedge |s|^2)(Z_0 + \mathbb{E}|Z|)^2] \mathbb{E} |\mathbf{L}_h(\beta)|^2 \\ &\leq |t|^2 (1 \wedge |s|^2) O_p(1). \end{aligned}$$

This implies

$$\lim_{\delta \rightarrow 0} \mathbb{P} \left(\int_{K_\delta^c} |\xi_h|^2 \mu(ds, dt) > \varepsilon \right) = 0.$$

On the other hand, it was shown in [Davis et al. \(2018\)](#) that $\int |G_h|^2 \mu(ds, dt)$ exists as the limit of $n\hat{T}_h(Z; \mu)$. Hence

$$\lim_{\delta \rightarrow 0} \mathbb{P} \left(\int_{K_\delta^c} |G_h|^2 \mu(ds, dt) > \varepsilon \right) = 0,$$

and the proposition is proved. \square

Appendix B. Proof of bootstrap consistency: A generalized theorem for triangular arrays

In this section, we generalize the convergence of ADCV for residuals for triangular arrays, from which the companion result for the bootstrap estimator in Theorem 4.2 can be derived.

Let $\{Z_{1:n}^{(n)}\}$ be a triangular array of random variables where for each n , $Z_j^{(n)}$'s are defined on the probability space $(\Omega, \mathcal{F}^{(n)}, \mathbb{P}_n)$ such that

$$Z_j^{(n)} \stackrel{iid}{\sim} F^{(n)}.$$

Assume that the distribution $F^{(n)}$ converges to F in distribution,

$$F^{(n)} \xrightarrow{d} F,$$

where F is the distribution of Z . Let $\{\beta^{(n)}\}$ be a sequence of parameter vectors such that

$$\beta^{(n)} \rightarrow \beta.$$

For each n , let $\{X_{1:n}^{(n)}\}$ be a time series generated from the model (1.1) with parameter vector $\beta^{(n)}$ and innovation sequence $\{Z_{-\infty:n}^{(n)}\}$,

$$X_j^{(n)} = g(Z_{-\infty:j}^{(n)}; \beta^{(n)}).$$

Let $\hat{\beta}^{(n)}$ and $\{\hat{Z}_{1:n}^{(n)}\}$ be the corresponding estimates and residuals based on $\{X_{1:n}^{(n)}\}$. We consider $T_n^{(n)}(h)$, the ADCV of $\{\hat{Z}_{1:n}^{(n)}\}$ at lag h . We require the following conditions.

(N1) Let $\mathcal{F}_j^{(n)}$ be the σ -algebra generated by $\{Z_k^{(n)}, k \leq j\}$, respectively. We assume that for any $\epsilon > 0$,

$$\mathbb{P}_n \left(\left| \sqrt{n}(\hat{\beta}^{(n)} - \beta^{(n)}) - \frac{1}{\sqrt{n}} \sum_{j=1}^n \mathbf{m}(Z_{-\infty:j}^{(n)}; \beta^{(n)}) \right| > \epsilon \right) \rightarrow 0.$$

where

$$\mathbb{E}_n[\mathbf{m}(Z_{-\infty:j}^{(n)}; \beta^{(n)}) | \mathcal{F}_{j-1}^{(n)}] = \mathbf{0}.$$

Further we assume that as $n \rightarrow \infty$,

$$\frac{1}{n} \sum_{j=1}^n \mathbb{E}_n[\mathbf{m}^T(Z_{-\infty:j}^{(n)}; \beta^{(n)}) \mathbf{m}(Z_{-\infty:j}^{(n)}; \beta^{(n)}) | \mathcal{F}_{j-1}^{(n)}] \xrightarrow{p} \tau^2,$$

for some $\tau > 0$, and

$$\mathbb{P}_n \left(\frac{1}{n} \sum_{j=1}^n \mathbb{E}_n[\mathbf{m}^T(Z_{-\infty:j}^{(n)}; \beta^{(n)}) \mathbf{m}(Z_{-\infty:j}^{(n)}; \beta^{(n)}) \mathbf{1}_{\{|\mathbf{m}(Z_{-\infty:j}^{(n)}; \beta^{(n)})| > \sqrt{n}\epsilon\}} | \mathcal{F}_{j-1}^{(n)}] > \epsilon \right) \rightarrow 0, \quad \forall \epsilon > 0.$$

(N2) Assume that the function h in the invertible representation (1.2) is continuously differentiable, and writing

$$\mathbf{L}_j^{(n)}(\beta) := \frac{\partial}{\partial \beta} h(X_{-\infty:j}^{(n)}; \beta), \tag{B.1}$$

we have

$$\sup_n \mathbb{E}_n \|\mathbf{L}_0^{(n)}(\beta^{(n)})\|^2 < \infty.$$

(N3) For fixed j , let $\tilde{Z}_j^{(n)}$ be the fitted residual based on the unobserved infinite sequence $\{X_{-\infty:j}^{(n)}\}$ obtained from (1.3), and $\hat{Z}_j^{(n)}$ be the estimated residuals based on the finite sequence $\{X_{1:j}^{(n)}\}$ obtained from (1.4). Assume that $\tilde{Z}_j^{(n)}$ is close to $\hat{Z}_j^{(n)}$ such that for any $\epsilon > 0$,

$$\mathbb{P}_n \left(\frac{1}{\sqrt{n}} \sum_{j=1}^n |\hat{Z}_j^{(n)} - \tilde{Z}_j^{(n)}|^k > \epsilon \right) \rightarrow 0, \quad k = 1, 2.$$

Theorem B.1. Assume that (N1), (N2), (N3) and (3.4) hold, then

$$\sup_t \left| \mathbb{P}_n(nT_n^{(n)}(h) \leq t) - \mathbb{P}(\|G_h + \xi_h\|_\mu^2 \leq t) \right| \rightarrow 0.$$

Proof of Theorem 4.2. Take $\beta^{(n)} = \hat{\beta}$ and $Z_t^{(n)} = Z_t^*$. Here, conditional on the data, Z_t^* 's are iid and follow the empirical distribution from $\{\hat{Z}_{1:n}\}$, which converges to the distribution of Z from (A.2). The result follows from Theorem B.1. \square

Proof of Theorem B.1. Let Z_1, Z_2, \dots be a sequence of random variable such that $Z_j \stackrel{iid}{\sim} F$. For each j , we have $Z_j^{(n)} \xrightarrow{d} Z_j$. By the Skorohod representation theorem, there exists a sufficiently rich probability space $(\tilde{\Omega}, \tilde{\mathcal{A}}, \tilde{\mathbb{P}})$ where $\tilde{\Omega} = \{(\omega_1, \omega_2, \dots) : \omega_j \in \Omega_0\}$ for some Ω_0 , and functions $z : \Omega_0 \rightarrow \mathbb{R}, z^{(n)} : \Omega_0 \rightarrow \mathbb{R}$, such that for each j ,

$$\tilde{Z}_j^{(n)} = z^{(n)}(\omega_j) \sim F^{(n)}, \quad \tilde{Z}_j = z(\omega_j) \sim F,$$

and

$$\tilde{Z}_j^{(n)} \xrightarrow{a.s.} \tilde{Z}_j.$$

This argument is similar to that in Leucht and Neumann (2009). Since we are only concerned about the distributional limit of $nT_n^{(n)}(h)$, we may assume without loss of generality that $Z_j^{(n)}$'s and Z_j 's are defined on the same probability space, and $Z_j^{(n)} \xrightarrow{a.s.} Z_j$ for each j . It suffices to prove that in this case,

$$nT_n^{(n)}(h) \xrightarrow{d} \|G_h + \xi_h\|_{\mu}^2.$$

Note that $T_n^{(n)}(h)$ can be written as

$$T_n^{(n)}(h) = \int |C_n^{\hat{Z}_n}(s, t)|^2 \mu(ds, dt) = \int |C_n^{\hat{Z}_n} - C_n^{Z_n} + C_n^{Z_n}|^2 \mu(ds, dt)$$

where

$$C_n^{\hat{Z}_n}(s, t) := \frac{1}{n} \sum_{j=1}^{n-h} e^{is\hat{Z}_j^{(n)} + it\hat{Z}_{j+h}^{(n)}} - \frac{1}{n} \sum_{j=1}^{n-h} e^{isZ_j^{(n)}} \frac{1}{n} \sum_{j=1}^{n-h} e^{itZ_{j+h}^{(n)}}$$

and

$$C_n^{Z_n}(s, t) := \frac{1}{n} \sum_{j=1}^{n-h} e^{isZ_j^{(n)} + itZ_{j+h}^{(n)}} - \frac{1}{n} \sum_{j=1}^{n-h} e^{isZ_j} \frac{1}{n} \sum_{j=1}^{n-h} e^{itZ_{j+h}}.$$

The result is proved in two propositions. In Proposition B.2, we show the joint convergence

$$(\sqrt{n}C_n^{Z_n}, \sqrt{n}(C_n^{\hat{Z}_n} - C_n^{Z_n})) \xrightarrow{d} (G_h, \xi_h), \quad \text{in } \mathcal{C}(K),$$

where K is any compact set in \mathbb{R}^2 . This implies that

$$\sqrt{n}C_n^{\hat{Z}_n} \xrightarrow{d} G_h + \xi_h, \quad \text{in } \mathcal{C}(K).$$

Then we justify the convergence of the integral by showing that for any $\varepsilon > 0$,

$$\lim_{\delta \rightarrow 0} \limsup_{n \rightarrow \infty} \mathbb{P}_n \left(\int_{K_\delta^c} n|C_n^{\hat{Z}_n}|^2 \mu(ds, dt) > \varepsilon \right) = 0.$$

This is done in Proposition B.3. \square

Proposition B.2. Given that (N1), (N2) and (N3) are satisfied, we have

$$(\sqrt{n}C_n^{Z_n}, \sqrt{n}(C_n^{\hat{Z}_n} - C_n^{Z_n})) \xrightarrow{d} (G_h, \xi_h), \quad \text{in } \mathcal{C}(K).$$

Proof. The proof is divided into the following steps.

Convergence of $C_n^{Z_n}$. In this part we show that

$$C_n^{Z_n} \xrightarrow{d} G_h, \quad \text{in } \mathcal{C}(K).$$

From Proposition A.1, we have $\sqrt{n}C_n^{Z_n} \xrightarrow{d} G_h$, where

$$C_n^Z(s, t) := \frac{1}{n} \sum_{j=1}^{n-h} e^{isZ_j + itZ_{j+h}} - \frac{1}{n} \sum_{j=1}^{n-h} e^{isZ_j} \frac{1}{n} \sum_{j=1}^{n-h} e^{itZ_{j+h}}.$$

It suffices to show that

$$\sqrt{n}(C_n^{Z_n} - C_n^Z) \xrightarrow{P} 0, \quad \text{in } \mathcal{C}(K).$$

Note that

$$C_n^Z(s, t) := \frac{1}{n} \sum_{j=1}^{n-h} U_j V_j - \frac{1}{n} \sum_{j=1}^{n-h} U_j \frac{1}{n} \sum_{j=1}^{n-h} V_j,$$

where $U_j := e^{isZ_j} - \varphi_Z(s)$ and $V_j := e^{itZ_{j+h}} - \varphi_Z(t)$ with $\mathbb{E}U_j V_j = \mathbb{E}U_j = \mathbb{E}V_j = 0$. Similarly,

$$C_n^{Z^{(n)}}(s, t) := \frac{1}{n} \sum_{j=1}^{n-h} U_j^{(n)} V_j^{(n)} - \frac{1}{n} \sum_{j=1}^{n-h} U_j^{(n)} \frac{1}{n} \sum_{j=1}^{n-h} V_j^{(n)},$$

where $U_j^{(n)}(s) := e^{isZ_j^{(n)}} - \varphi_{Z^{(n)}}(s)$ and $V_j^{(n)}(t) := e^{itZ_{j+h}^{(n)}} - \varphi_{Z^{(n)}}(t)$. Without loss of generality, here we only show

$$\sqrt{n} \left(\frac{1}{n} \sum_{j=1}^{n-h} U_j^{(n)} - \frac{1}{n} \sum_{j=1}^{n-h} U_j \right) = \frac{1}{\sqrt{n}} \sum_{j=1}^{n-h} (U_j^{(n)} - U_j) \xrightarrow{p} 0, \quad \text{in } \mathcal{C}(K).$$

For fixed s , the convergence follows since

$$\mathbb{E} \left| \frac{1}{\sqrt{n}} \sum_{j=1}^{n-h} (U_j^{(n)} - U_j) \right|^2 \leq \mathbb{E} |U_j^{(n)} - U_j|^2 \rightarrow 0,$$

from bounded convergence. The finite dimensional convergence can be generalized using the Cramér–Wold device. It remains to prove the tightness of $\frac{1}{\sqrt{n}} \sum_{j=1}^{n-h} (U_{j_n} - U_j)$. By equation (7.12) of Billingsley (1999), the tightness of the process can be implied by

$$\mathbb{E} \left| \frac{1}{\sqrt{n}} \sum_{j=1}^{n-h} (U_j^{(n)}(s) - U_j(s)) - \frac{1}{\sqrt{n}} \sum_{j=1}^{n-h} (U_j^{(n)}(s') - U_j(s')) \right|^2 \leq |s - s'|^{\delta+1} O(1), \quad \text{for some } \delta > 0.$$

We have

$$\begin{aligned} & \mathbb{E} \left| \frac{1}{\sqrt{n}} \sum_{j=1}^{n-h} (U_j^{(n)}(s) - U_j(s)) - \frac{1}{\sqrt{n}} \sum_{j=1}^{n-h} (U_j^{(n)}(s') - U_j(s')) \right|^2 \\ & \leq \mathbb{E} \left| U_j^{(n)}(s) - U_j(s) - (U_j^{(n)}(s') - U_j(s')) \right|^2 \\ & \leq 2\mathbb{E} |e^{isZ_j^{(n)}} - e^{is'Z_j^{(n)}}|^2 + 2|\varphi_{Z^{(n)}}(s) - \varphi_{Z^{(n)}}(s')|^2 + 2\mathbb{E} |e^{isZ_j} - e^{is'Z_j}|^2 + 2|\varphi_Z(s) - \varphi_Z(s')|^2. \end{aligned}$$

Note that

$$\mathbb{E} |e^{isZ_j^{(n)}} - e^{is'Z_j^{(n)}}|^2 \leq \mathbb{E} |e^{i(s-s')Z_j^{(n)}} - 1|^2 \leq 2\mathbb{E} |Z_j^{(n)}|^2 |s - s'|^2.$$

The rest of the term can be bounded similarly. And the tightness is proved.

Convergence of $\sqrt{n}(C_n^{\hat{Z}^{(n)}}(s, t) - C_n^{Z^{(n)}}(s, t))$. In this part we show that

$$\sqrt{n}(C_n^{\hat{Z}^{(n)}}(s, t) - C_n^{Z^{(n)}}(s, t)) \xrightarrow{d} \xi_h, \quad \text{in } \mathcal{C}(K).$$

Similar to (A.1), we have

$$\sqrt{n}(C_n^{\hat{Z}^{(n)}}(s, t) - C_n^{Z^{(n)}}(s, t)) = E_n^{(n)}(s, t) - E_n^{(n)}(s, 0) \frac{1}{n} \sum_{j=1}^{n-h} e^{itZ_{j+h}^{(n)}} - E_n^{(n)}(0, t) \frac{1}{n} \sum_{j=1}^{n-h} e^{isZ_j^{(n)}}$$

where

$$E_n^{(n)}(s, t) := \frac{1}{\sqrt{n}} \sum_{j=1}^{n-h} \left(e^{is\hat{Z}_j^{(n)} + it\hat{Z}_{j+h}^{(n)}} - e^{isZ_j^{(n)} + itZ_{j+h}^{(n)}} \right).$$

From the decomposition of ξ_h in (A.3), it suffices to show that

$$E_n^{(n)}(s, t) \xrightarrow{d} \mathbf{Q}^T \mathbf{C}_h(s, t), \quad \text{in } \mathcal{C}(K).$$

Uniformly on $(s, t) \in K$, we have

$$\begin{aligned} E_n^{(n)}(s, t) &= \frac{1}{n} \sum_{j=1}^{n-h} e^{isZ_j^{(n)} + itZ_{j+h}^{(n)}} (is\sqrt{n}(\hat{Z}_j^{(n)} - \tilde{Z}_j^{(n)}) + it\sqrt{n}(\hat{Z}_{j+h}^{(n)} - \tilde{Z}_{j+h}^{(n)})) \\ &\quad + \frac{1}{n} \sum_{j=1}^{n-h} e^{isZ_j^{(n)} + itZ_{j+h}^{(n)}} (is\sqrt{n}(\tilde{Z}_j^{(n)} - Z_j^{(n)}) + it\sqrt{n}(\tilde{Z}_{j+h}^{(n)} - Z_{j+h}^{(n)})) + o_p(1) \\ &=: E_{n1}^{(n)}(s, t) + E_{n2}^{(n)}(s, t) + o_p(1). \end{aligned}$$

From condition (N3),

$$|E_{n1}^{(n)}(s, t)| \leq \frac{|s| + |t|}{\sqrt{n}} \sum_{j=1}^n |\hat{Z}_j^{(n)} - \tilde{Z}_j^{(n)}| \xrightarrow{p} 0, \quad \text{in } \mathcal{C}(K).$$

It suffices to show that $E_{n2}^{(n)}(s, t) \xrightarrow{d} \mathbf{Q}^T \mathbf{C}_h(s, t)$. By Taylor expansion,

$$E_{n2}^{(n)}(s, t) = \sqrt{n}(\hat{\boldsymbol{\beta}}^{(n)} - \boldsymbol{\beta}^{(n)})^T \frac{1}{n} \sum_{j=1}^{n-h} e^{isZ_j^{(n)} + itZ_{j+h}^{(n)}} (is\mathbf{L}_j^{(n)}(\boldsymbol{\beta}^{(n)*}) + it\mathbf{L}_{j+h}^{(n)}(\boldsymbol{\beta}^{(n)*})),$$

where $\boldsymbol{\beta}^{(n)*} = \epsilon\boldsymbol{\beta}^{(n)} + (1 - \epsilon)\hat{\boldsymbol{\beta}}^{(n)}$ for some $\epsilon \in [0, 1]$. From condition (N1),

$$\sqrt{n}(\hat{\boldsymbol{\beta}}^{(n)} - \boldsymbol{\beta}^{(n)}) \xrightarrow{d} \mathbf{Q}, \tag{B.2}$$

follows from the martingale central limit theorem, see Theorem 2 of Scott (1973). It remains to show that

$$\frac{1}{n} \sum_{j=1}^{n-h} e^{isZ_j^{(n)} + itZ_{j+h}^{(n)}} (is\mathbf{L}_j^{(n)}(\boldsymbol{\beta}^{(n)*}) + it\mathbf{L}_{j+h}^{(n)}(\boldsymbol{\beta}^{(n)*})) \xrightarrow{p} \mathbf{C}_h(s, t), \quad \text{in } \mathcal{C}(K).$$

The marginal convergence follows from (N2) and the weak law of large number for triangular arrays, see, for example, Theorem 2.2.6 of Durrett (2010). The convergence in $\mathcal{C}(K)$ can be extended similar to previous proofs.

Joint convergence of $\sqrt{n}C_n^{Z_n}(s, t)$ and $\sqrt{n}(C_n^{\hat{Z}_n}(s, t) - C_n^{Z_n}(s, t))$. The above proofs imply that

$$\sqrt{n}C_n^{Z_n}(s, t) - \sqrt{n}C_n^Z(s, t) \xrightarrow{p} 0, \quad \text{in } \mathcal{C}(K),$$

and

$$\sqrt{n}(C_n^{\hat{Z}_n}(s, t) - C_n^{Z_n}(s, t)) - \sqrt{n}(C_n^{\hat{Z}_n}(s, t) - C_n^Z(s, t)) \xrightarrow{p} 0, \quad \text{in } \mathcal{C}(K).$$

The joint convergence of $\sqrt{n}C_n^{Z_n}(s, t)$ and $\sqrt{n}(C_n^{\hat{Z}_n}(s, t) - C_n^{Z_n}(s, t))$ follows from the joint convergence of $\sqrt{n}C_n^Z(s, t)$ and $\sqrt{n}(C_n^{\hat{Z}_n}(s, t) - C_n^Z(s, t))$ in Proposition A.1. \square

Proposition B.3. For any $\varepsilon > 0$,

$$\lim_{\delta \rightarrow 0} \limsup_{n \rightarrow \infty} \mathbb{P} \left(\int_{K_\delta^c} n|C_n^{\hat{Z}_n}|^2 \mu(ds, dt) > \varepsilon \right) = 0.$$

Proof. This follows the same steps in the proof of Proposition A.2 by replacing all \hat{Z}_j with $\hat{Z}_j^{(n)}$ and Z_j with $Z_j^{(n)}$. \square

Appendix C. Proof of Corollary 5.1

Proof. In the following we verify conditions (M1), (M2), (M3) in Theorem 3.1.

(M1): It can be shown that the pseudo-MLE for $\boldsymbol{\beta}$ satisfies the representation in (M1). We refer to Chapter 10.8 of Brockwell and Davis (1991) for details.

(M2): From

$$Z_t = \frac{\phi(B)}{\theta(B)} X_t =: h(X_{-\infty:t}, \boldsymbol{\beta}),$$

we have

$$\frac{\partial}{\partial \phi_i} h(X_{-\infty:t}, \boldsymbol{\beta}) = \frac{B^i}{\theta(B)} X_t = \frac{1}{\theta(B)} X_{t-i}, \quad i = 1, \dots, p,$$

while

$$\frac{\partial}{\partial \theta_i} h(X_{-\infty:t}, \boldsymbol{\beta}) = \frac{B^i \phi(B)}{(\theta(B))^2} X_t = \frac{B^i}{\theta(B)} Z_t = \frac{1}{\theta(B)} Z_{t-j}, \quad j = 1, \dots, q.$$

Hence

$$\mathbf{L}_0(\boldsymbol{\beta}) = \frac{\partial}{\partial \boldsymbol{\beta}} h(X_{-\infty:0}; \boldsymbol{\beta}) = \frac{1}{\theta(B)} (X_{-1}, \dots, X_{-p}, Z_{-1}, \dots, Z_{-q})^T.$$

By the definition of invertibility, there exists a power series for $1/\theta(z)$ such that

$$\frac{1}{\theta(z)} = \sum_{j=0}^{\infty} \xi_j(\boldsymbol{\beta}) z^j,$$

with $\sum_{j=0}^{\infty} |\xi_j(\boldsymbol{\beta})| < \infty$. Therefore

$$\mathbb{E} \|\mathbf{L}_0(\boldsymbol{\beta})\|^2 \leq p \sum_{j=0}^{\infty} |\xi_j(\boldsymbol{\beta})|^2 \mathbb{E}|X_0|^2 + q \sum_{k=0}^{\infty} |\xi_j(\boldsymbol{\beta})|^2 \mathbb{E}|Z_0|^2 < \infty.$$

(M3): Note that

$$\tilde{Z}_t - \hat{Z}_t = \sum_{j=t}^{\infty} \pi_j(\hat{\boldsymbol{\beta}}) X_{t-j}.$$

For $k = 1, 2$,

$$\frac{1}{\sqrt{n}} \sum_{t=1}^n |\tilde{Z}_t - \hat{Z}_t|^k \leq \frac{1}{\sqrt{n}} \sum_{t=1}^n \sum_{j=t}^{\infty} |\pi_j(\hat{\boldsymbol{\beta}}) X_{t-j}|^k = \sum_{j=0}^{\infty} |\pi_j(\hat{\boldsymbol{\beta}})|^k \frac{1}{\sqrt{n}} \sum_{t=1}^{j \wedge n} |X_{t-j}|^k.$$

For any $m < n$,

$$\frac{1}{\sqrt{n}} \sum_{t=1}^n |\tilde{Z}_t - \hat{Z}_t|^k \leq \sum_{j=0}^m |\pi_j(\hat{\boldsymbol{\beta}})|^k \frac{1}{\sqrt{n}} \sum_{t=1}^m |X_{t-j}|^k + \sum_{j=m+1}^{\infty} |\pi_j(\hat{\boldsymbol{\beta}})|^k \frac{1}{\sqrt{n}} \sum_{t=1}^n |X_{t-j}|^k =: I_1 + I_2. \tag{C.1}$$

Consider the coefficients $\pi_j(\hat{\boldsymbol{\beta}})$'s. By causality, the power series

$$\frac{\phi(z)}{\theta(z)} = \sum_{j=0}^{\infty} \pi_j(\boldsymbol{\beta}) z^j$$

converges for all $|z| < 1 + \epsilon$ for some $\epsilon > 0$. Then there exists a compact set \mathbf{C}_β containing $\boldsymbol{\beta}$ such that for any $\hat{\boldsymbol{\beta}} \in \mathbf{C}_\beta$, $\sum_{j=0}^{\infty} \pi_j(\hat{\boldsymbol{\beta}}) z^j$ converges for all $|z| < 1 + \epsilon/2$. In particular,

$$\pi_j(\hat{\boldsymbol{\beta}}) (1 + \epsilon/4)^j \rightarrow 0, \quad j \rightarrow \infty,$$

and there exists $K > 0$ such that

$$|\pi_j(\hat{\boldsymbol{\beta}})| \leq K(1 + \epsilon/4)^{-j}.$$

It follows that for $k = 1, 2$,

$$\sum_{j=0}^{\infty} |\pi_j(\hat{\boldsymbol{\beta}})|^k < \infty$$

and

$$\sum_{j=m}^{\infty} |\pi_j(\hat{\boldsymbol{\beta}})|^k < c(1 + \epsilon/4)^{-km}.$$

Now for (C.1), I_1 converges to zero in probability for fixed m , while I_2 converges to zero uniformly as $m \rightarrow \infty$ with order greater than $O(\log(n))$. This implies that

$$\frac{1}{\sqrt{n}} \sum_{t=1}^n |\tilde{Z}_t - \hat{Z}_t|^k \xrightarrow{P} 0, \quad k = 1, 2. \quad \square$$

Appendix D. Proof of Corollary 6.1

Proof. In the following we verify conditions (M1), (M2), (M3) in Theorem 3.1.

(M1): Given conditions (Q1)–(Q4), Berkes et al. (2003) showed that $\hat{\theta}_n$ has limiting distribution

$$\sqrt{n}(\hat{\theta}_n - \theta) = \frac{1}{\sqrt{n}} \sum_{t=1}^n \frac{1}{2}(1 - Z_t^2) \left\langle \frac{\partial \log \sigma_t^2(\theta)}{\partial \theta}, \mathbf{B}_0^{-1} \right\rangle + o_p(1) \xrightarrow{d} N(\mathbf{0}, \mathbf{B}_0^{-1} \mathbf{A}_0 \mathbf{B}_0^{-1}),$$

where

$$\mathbf{A}_0 = \text{cov} \left[\frac{\partial l_0(\theta)}{\partial \theta} \right], \quad \mathbf{B}_0 = \mathbb{E} \left[\frac{\partial^2 l_0(\theta)}{\partial \theta^2} \right].$$

(M2): We have

$$Z_t(\theta) = h(X_{-\infty:j}, \theta) = \frac{X_t}{\sigma_t(\theta)},$$

and

$$\mathbf{L}_0(\theta) = \frac{\partial}{\partial \theta} h(X_{-\infty:0}; \theta) = -\frac{X_0}{2\sigma_0^3(\theta)} \frac{\partial \sigma_0^2(\theta)}{\partial \theta} = -\frac{1}{2} Z_0 \frac{\partial \log \sigma_0^2(\theta)}{\partial \theta}.$$

Lemma 3.1 of Kulperger and Yu (2005) showed that

$$\mathbb{E} \left(\sup_{\mathbf{u} \in \Theta} \left| \frac{\partial \log \sigma_t^2(\mathbf{u})}{\partial \mathbf{u}} \right| \right)^k < \infty, \quad \text{for any } k > 0.$$

Hence

$$\mathbb{E} \|\mathbf{L}_0(\theta)\|^2 = \mathbb{E} \left| \frac{1}{2} Z_0 \frac{\partial \log \sigma_0^2(\theta)}{\partial \theta} \right|^2 \leq \frac{1}{4} \left(\mathbb{E} |Z_0|^4 \mathbb{E} \left| \frac{\partial \log \sigma_0^2(\theta)}{\partial \theta} \right|^4 \right)^{1/2} < \infty.$$

(M3): Theorem 1.3 and Lemma 3.5 of Kulperger and Yu (2005) show, respectively, that

$$\frac{1}{\sqrt{n}} \sum_{t=1}^n |\hat{Z}_t - \tilde{Z}_t| = o_p(1),$$

and

$$\sum_{t=1}^n |\hat{Z}_t - \tilde{Z}_t| = O_p(1).$$

Hence

$$\frac{1}{\sqrt{n}} \sum_{t=1}^n |\hat{Z}_t - \tilde{Z}_t|^2 \leq \frac{1}{\sqrt{n}} \sum_{t=1}^n |\hat{Z}_t - \tilde{Z}_t| \sum_{t=1}^n |\hat{Z}_t - \tilde{Z}_t| = o_p(1). \quad \square$$

References

- Berkes, I., Horváth, L., Kokoszka, P., 2003. GARCH processes: structure and estimation. *Bernoulli* 9 (2), 201–227.
- Billingsley, P., 1999. *Convergence of Probability Measures*, second ed. Wiley, New York.
- Brockwell, P.J., Davis, R.A., 1991. *Time Series: Theory and Methods*. Springer, New York.
- Davis, R.A., 1996. Gauss-Newton and M -estimation for ARMA proces with infinite variance. *Stoch. Process. Appl.* 63, 75–95.
- Davis, R.A., Matsui, M., Mikosch, T., Wan, P., 2018. Applications of distance covariance to time series. *Bernoulli* 24 (4A), 3087–3116.
- Durrett, R.T., 2010. *Probability: Theory and Examples*, fourth ed. Cambridge University Press.
- Feuerverger, A., 1993. A consistent test for bivariate dependence. *Internat. Statist. Rev.* 61 (3), 419–433.
- Fokianos, K., Pitsillou, M., 2017. Consistent testing for pairwise dependence in time series. *Technometrics* 59 (2), 262–270.
- Gretton, A., Bousquet, O., Smola, A., Schölkopf, B., 2005. Measuring statistical dependence with Hilbert–Schmidt norms. In: Jain, Sanjay, Simon, Hans Ulrich, Tomita, Etsuji (Eds.), *Algorithmic Learning Theory*. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 63–77.
- Hong, Y., 1999. Hypothesis testing in time series via the empirical characteristic function: A generalized spectral density approach. *J. Amer. Statist. Assoc.* 94, 1201–1220.
- Kulperger, R., Yu, H., 2005. High moment partial sum processes of residuals in garch models and their applications. *Ann. Statist.* 33 (5), 2395–2422.
- Leucht, A., Neumann, M.H., 2009. Consistency of general bootstrap methods for degenerate U -type and V -type statistics. *J. Multivariate Anal.* 100, 1622–1633.
- Ljung, G.M., Box, G.E.P., 1978. On a measure of a lack of fit in time series models. *Biometrika* 65 (2), 297–303.
- Scott, D.J., 1973. Central limit theorems for martingales and for processes with stationary increments using a skorokhod representation approach. *Adv. Appl. Probab.* 5 (1), 119–137.
- Székely, G.J., Rizzo, M.L., Bakirov, N.K., 2007. Measuring and testing dependence by correlation of distances. *Ann. Statist.* 35, 2769–2794.
- Wang, G., Li, W.K., Zhu, K., 2018. New hsc-based tests for independence between two stationary multivariate time series. arXiv:1804.09866.
- Zhou, Z., 2012. Measuring nonlinear dependence in time-series, a distance correlation approach. *J. Time Series Anal.* 33, 438–457.