June 8, 2018

Goodness-of-Fit Testing for Time Series Models via Distance Covariance

Phyllis Wan and Richard A. Davis

Department of Statistics
Columbia University
1255 Amsterdam Avenue, MC 4690
New York, NY 10027
e-mail: phyllis@stat.columbia.edu; rdavis@stat.columbia.edu

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 correlation function (ADCF) to evaluate the serial dependence of the estimated residuals. Distance correlation can discriminate between dependence and independence and is applicable for multivariate data. The limit behavior of the test statistic based on the ADCF 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.

Keywords and phrases: distance covariance, time series models, estimated residuals, goodness-of-fit testing, serial dependence.

1. Introduction

?⟨sec:intro⟩? 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 $X_1, \ldots, X_n$, we are interested in whether the sequence can plausibly be viewed as generated from a parametric model, more precisely, whether $\{X_j\}$ is generated from the recursion

$$X_j := f(X_{-\infty:j}, Z_j; \boldsymbol{\beta}), \tag{1.1}$$ eq:model

where $X_{n_1:n_2}$ denotes the sequence $\{X_j, n_1 \le j \le n_2\}$, the Z_j's are iid with finite second moments, and $\boldsymbol{\beta} \in \mathbb{R}^d$ is the parameter vector. The objective of this paper is to provide a validity check of the model (1.1) by inspecting the residuals.

A typical assumption for time series models is that the recursion (1.1) is casual and invertible, that is,

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

and

$$Z_j = Z_j(\boldsymbol{\beta}) = h(X_{-\infty:j}; \boldsymbol{\beta}) \tag{1.2}$$ eq:invert

for some functions g and h. Here we write $Z_j(\boldsymbol{\beta})$ to indicate its dependency on $\boldsymbol{\beta}$. Given the observations $X_{1:n}$, let $\hat{\boldsymbol{\beta}}$ be an estimator of $\boldsymbol{\beta}$. Then the innovations $\{Z_j\}$ can be approximated by

$$\tilde{Z}_j := Z_j(\hat{\boldsymbol{\beta}}) = h(X_{-\infty:j}; \hat{\boldsymbol{\beta}}), \tag{1.3}$$ eq:tildez

the residuals based on the infinite sequence $\{X_j, j \le n\}$. If the recursion (1.1) describes the generating mechanism of $\{X_j\}$, one would expect $\{\tilde{Z}_j\}$ to inherit the properties of $\{Z_j\}$. In reality, we do not observe X_j for $j \le 0$ and instead rely on the estimated residuals

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

where $\{Y_j\}$ is the infinite sequence with $Y_j = X_j$, $1 \le j \le n$ and $Y_j = 0$ for $j \le 0$. 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 indistinguishable.

While $\hat{Z}_1, \dots, \hat{Z}_n$ are derived to approximate the iid innovation $\{Z_j\}$, the sequence itself is not iid since they are functions of $\hat{\boldsymbol{\beta}}$. This has been noted for specific time series models in the literature. For example, for ARMA model, corrections have been proposed for statistics based on the residuals, see Section 9.4 of Brockwell and Davis (1991). For the heteroscedastic GARCH models, the moment sum process of the residuals were studied in Kulperger and Yu (2005). Still, if the model assumption is true, $\{\hat{Z}_j\}$ should possess a serial dependence structure consistent with the model.

In this paper, we evaluate the serial dependence of residuals using distance covariance. Distance covariance is a usefull dependence measure with the ability to detect both linear and nonlinear dependence. It is zero if and only if independence occurs. We study the auto-distance covariance function (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. This is an extension of Section 4 of Davis et al. (2018) which considered this problem for AR processes.

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 theoretical results, we justify the use of parametric bootstrap in Section 4. We then apply the result to ARMA and GARCH models in Section 5 and 6 and illustrate with simulation studies. A simulated example where the data does not conform with the model is also demonstrated in Section 7.

2. Distance covariance

⟨sec:dcor⟩ Let $X \in \mathbb{R}^p$ and $Y \in \mathbb{R}^q$ be two random vectors, potentially of different dimensions. Then

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

where $\varphi_{X,Y}(s,t), \varphi_X(s), \varphi_Y(t)$ denote the joint and marginal characteristic functions of (X,Y). The *distance covariance* between X and Y is defined as

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

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

$$\iint_{\mathbb{R}^{p+q}} (1 \wedge |s|^\alpha)(1 \wedge |t|^\alpha)\mu(ds,dt) < \infty$$

and

$$\mathbb{E}[|XY|^\alpha + |X|^\alpha + |Y|^\alpha] < \infty, \quad \text{for some } \alpha \in (0,2].$$

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

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

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

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

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 a alternative V-statistic like form, 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,

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

eq:gh

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

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

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

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

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

Most 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 variable sevaluated. In our case, as will be shown in Section 3, we require a finite measure for μ and shall use a Gaussian measure. 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.

It should be noted that the concept of distance covariance is closely related to Hilbert-Schmidt Independence Criterion (HSIC), see Gretton et al. (2005). For example, the distance covariance with Gaussian measure coincides with the HSIC with Gaussian kernel. In a recent (unpublished) work, Zhu and Li use HSIC for testing the cross dependence between two time series.

3. General result

⟨sec:meta⟩ Let $X_1,\ldots,X_n$ be the observed sequence from a stationary time series $\{X_j\}$ generated from (1.1), and let $\hat{Z}_1,\ldots,\hat{Z}_n$ be the estimated residual calculated through (1.4). In this section, we examine the ADCV of the residuals

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

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.

?⟨cond:m1⟩? (M1) Let $\mathcal{F}_j$ be the σ-algebra generated by $\{X_k, k \le j\}$. We assume that the parameter estimate $\hat{\boldsymbol{\beta}}$ is of the form

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

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

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

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. By the martingale central limit theorem, this implies that

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

for a random Gaussian vector $\mathbf{Q}$.

?⟨cond:m2⟩? (M2) Assume that the function h in the invertible representation (1.2) is continuously differentiable, and writing

$$\mathbf{L}_j(\boldsymbol{\beta}) := \frac{\partial}{\partial\boldsymbol{\beta}} h(X_{-\infty:j};\boldsymbol{\beta}), \tag{3.1}$$ eq:bigL

we have

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

?⟨cond:m3⟩? (M3) Assume the estimated residuals based on the finite sequence of observations, $\hat{Z}_j$, is close to the fitted residuals based on the infinite sequence, $\tilde{Z}_j$, such that

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

⟨thm:meta⟩ **Theorem 3.1.** *Let $X_1, \ldots, X_n$ be a sequence of observations generated from a causal and invertible time series model* (1.1). *Let $\hat{\boldsymbol{\beta}}$ be an estimator of $\boldsymbol{\beta}$ and let $\hat{Z}_1, \ldots, \hat{Z}_n$ be the estimated residuals calculated through* (1.4) *satisfying conditions (M1)–(M3). Further assume that the weight measure μ satisfies*

$$\iint_{\mathbb{R}^2} \left[(1 \wedge |s|^2)(1 \wedge |t|^2) + (s^2 + t^2)\,\mathbf{1}(|s| \wedge |t| > 1)\right]\mu(ds, dt) < \infty. \tag{3.2}$$ eq:7

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 $Z_1, \ldots, Z_n$, and the correction term is given by

$$\xi_h(s,t) := it\mathbf{Q}^T\mathbb{E}\left[\left(e^{isZ_0} - \varphi_Z(s)\right)e^{itZ_h}\mathbf{L}_h(\boldsymbol{\beta})\right], \tag{3.3}$$ eq:xi

with $\mathbf{Q}$ being the limit distribution of $\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ and $\mathbf{L}_h$ as defined in (B.1).

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 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):

$$n\hat{R}_h(\hat{Z};\mu) \xrightarrow{d} \frac{\|G_h+\xi_h\|_\mu^2}{T_0(Z;\mu)}, \tag{3.4}$$ eq:adcf:limit

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

4. Parametric bootstrap

⟨sec:boot⟩ **Messy notation!**

The limit in (3.4) is not distribution-free and generally intractable. In order to use the result, we propose to approximate the limit through parametric bootstrap, described in the following.

Given observations $X_1,\dots,X_n$, let $\hat{\boldsymbol\beta}$ 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. Sample iid $Z_1^*,\dots,Z_n^*$ from the empirical distribution of $\{\hat{Z}_j\}$, i.e., with replacement from $\hat{Z}_1,\dots,\hat{Z}_n$.
2. Generate $X_1^*,\dots,X_n^*$ from the time series model with parameter value $\hat{\boldsymbol\beta}$ and residual sequence $Z_1^*,\dots,Z_n^*$.
3. Re-fit the time series model. Obtained the parameter estimate $\hat{\boldsymbol\beta}^*$ and the estimated residuals $\hat{Z}_1^*,\dots,\hat{Z}_n^*$.

Let $n\hat{R}_h(\hat{Z}^*,\mu)$ be the ADCF calculated from the bootstrapped residuals $\hat{Z}_1^*,\dots,\hat{Z}_n^*$. This procedure is repeated B times to obtain $n\hat{R}_h^{(1)}(\hat{Z}^*,\mu),\dots,n\hat{R}_h^{(B)}(\hat{Z}^*,\mu)$. When the sample size n is large, the empirical distribution of $\{n\hat{R}_h^{(b)}(\hat{Z}^*,\mu)\}$ provides an approximation for the limiting distribution of $n\hat{R}_h(\hat{Z};\mu)$. The theoretical convergence of the bootstrapped ADCF is currently under investigation.

Given the following condition, we show in Theorem 4.1 that these statistics form a good representation of the limit distribution of $n\hat{T}_h(\hat{Z},\mu)$, the ADCV of the actual fitted residuals.

?⟨cond:m1p⟩? (M1') Let $\mathcal{F}_j,\mathcal{F}_j^*$ be the σ-algebra generated by $\{X_k, k\le j\}$ and $\{X_k^*, k\le j\}$, respectively. We assume that the parameter estimate $\hat{\boldsymbol\beta}$ is of the form

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

where $\mathbf{m}$ satisfies

$$\mathbb{E}[\mathbf{m}(X^*_{-\infty:j};\hat{\boldsymbol\beta})|\mathcal{F}^*_{j-1}] = \mathbb{E}[\mathbf{m}(X_{-\infty:j};\boldsymbol\beta)|\mathcal{F}_{j-1}] = \mathbf{0}, \quad \mathbb{E}\sup_n|\mathbf{m}(X^*_{-\infty:0};\hat{\boldsymbol\beta})|^2<\infty.$$

?⟨cond:m2p⟩? (M2') Assume that the function h in the invertible representation (1.2) is continuously differentiable, and writing

$$\mathbf{L}_j^*(\boldsymbol\beta) := \frac{\partial}{\partial\boldsymbol\beta}h(X^*_{-\infty:j};\boldsymbol\beta), \tag{4.1}$$ eq:bigL

we have

$$\mathbb{E}\sup_n\|\mathbf{L}_0^*(\hat{\boldsymbol\beta})\|^2<\infty.$$

?⟨cond:m3p⟩? (M3') Assume the estimated residuals based on the finite sequence of observations, $\hat{Z}_j^*$, is close to the fitted residuals based on the infinite sequence, $\tilde{Z}_j^*$, such that

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

⟨thm:boot⟩ **Theorem 4.1.** *Assume that the conditions (M1'), (M2') and (M3'). Then the ADCV of the bootstrapped residuals* $\{\hat{Z}_{1:n}^*\}$ *satisfies*

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

conditioning on the observed data $X_1, \ldots, X_n$.

5. Example: ARMA(p,q)

⟨sec:arma⟩ Consider the causal, invertible ARMA(p, q) process that follows the recursion

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

where $\boldsymbol{\beta} = (\phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q)^T$ is the vector of parameters and $\{Z_t\}$ is the sequence of mean 0 and uncorrelated innovation. 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, i.e.,

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

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

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

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

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

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

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

Based on the observed data $X_1, \ldots, X_n$, the estimate residuals are

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

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

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

where $\mathbf{X}_n = (X_1, \ldots, X_n)^T$ and the covariance $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\boldsymbol{\beta}) := \mathrm{Var}(\mathbf{X}_n)/\sigma^2$ is independent of σ^2. The pseudo-MLE $\hat{\boldsymbol{\beta}}$ and $\hat{\sigma}^2$ are taken to be the values that maximize $L(\boldsymbol{\beta}, \sigma^2)$. It can be shown that $\hat{\boldsymbol{\beta}}$ 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.

⟨thm:arma⟩ **Corollary 5.1.** *Let $\{X_t, 1 \le j \le n\}$ be observations from a causal and invertible ARMA(p,q) time series and $\{\hat{Z}_t, 1 \le t \le n\}$ be the estimated residuals defined in (5.2). Assume that μ satisfies (3.2), 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 in $\mathbb{R}^2$ with G_h as specified in (2.1) and ξ_h in (3.3).

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 a α-stable law with $\alpha \in (0,2)$, and the parameter estimator $\hat{\boldsymbol{\beta}}$ has convergence rate faster than $n^{-1/2}$, i.e.,

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

(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, see Theorem 4.2 of Davis et al. (2018).

5.1. Simulation

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

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

For each simulation, an ARMA(2,2) model is fitted to the data. In Figure 1, 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;
c) estimated residuals from 1000 independent parametric bootstrap samples from one realization of $\{X_t\}$.

In order to satisfy the requirement (3.2), the ADCFs 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 $\hat{R}_h(\hat{Z};\mu)$ 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 parametric boostrap correctly capture this effect.

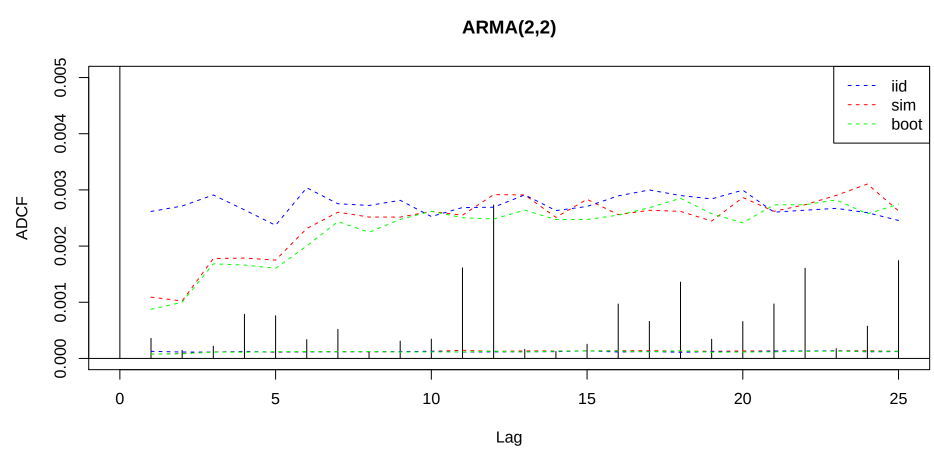

FIG 1. *Empirical 5% and 95% quantiles of the ADCF for a) iid innovations; b) estimated residuals; c) bootstrapped residuals; from a ARMA(2,2) model.*

⟨fig:arma⟩

6. Example: GARCH(p,q)

⟨sec:garch⟩ In this section, we consider a 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

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

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

Iterating the recursion in (6.1) gives

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

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

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

and the unobserved residuals are given by

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

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

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

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

$$\hat{Z}_t = X_t / \hat{\sigma}_t. \tag{6.2}$$ eq:garch:residua

Define the parameter space by

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

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

?⟨cond:q1⟩? (Q1) The true value $\boldsymbol{\theta}$ lies in the interior of $\boldsymbol{\Theta}$.

?⟨cond:q2⟩? (Q2) For some $\zeta > 0$,

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

?⟨cond:q3⟩? (Q3) For some $\delta > 0$,

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

?⟨cond:q4⟩? (Q4) The GARCH(p,q) representation is minimal, i.e., the polynomials $A(z) = \sum_{i=1}^{p} \alpha_i z^i$ and $B(z) = 1 - \sum_{j=1}^{p} \beta_j z^j$ do not have common roots.

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

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

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 $\hat{\boldsymbol{\theta}}_n$ is consistent and asymptotically normal.

For the ADCV of the residuals based on $\hat{\boldsymbol{\theta}}_n$, we have the following result.

⟨thm:garch⟩ **Corollary 6.1.** *Let $\{X_t, 1 \le j \le n\}$ be observations from a GARCH(p,q) time series and $\{\hat{Z}_t, 1 \le t \le n\}$ be the estimated residuals defined in (6.2). Assume that (Q1)–(Q4) holds and that μ satisfies (3.2), 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 in $\mathbb{R}^2$ with G_h as specified in (2.1) and ξ_h in (3.3).

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

$$\boldsymbol{\beta} = (\alpha_0, \alpha_1, \beta_1) = (0.5, 0.1, 0.8).$$

For each simulation, a GARCH(1,1) model is fitted to the data. In Figure 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;
c) estimated residuals from 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 the GARCH model, the correction has the opposite effect than in the ARMA model – the ADCF for residuals are larger than that for iid variables, especially for small lags.

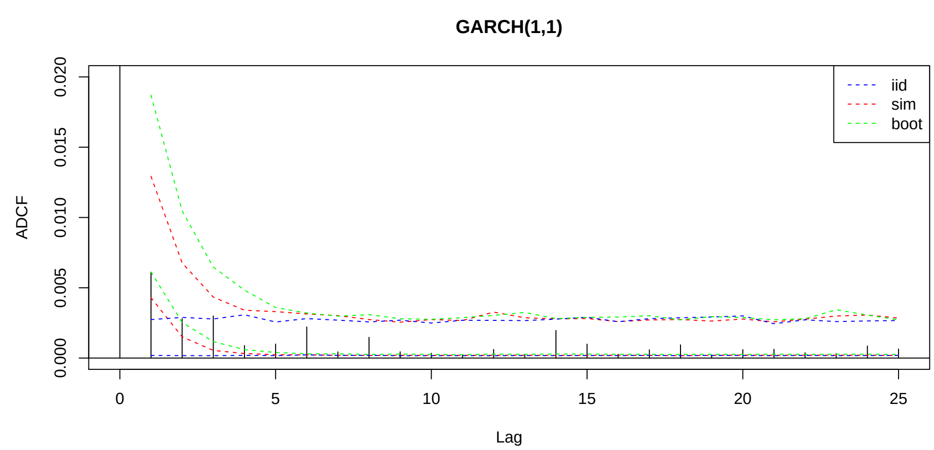

Fig 2. *Empirical 5% and 95% quantiles of the ADCF for a) iid innovations; b) estimated residuals; c) bootstrapped residuals; from a GARCH(1,1) model.*

⟨fig:garch⟩

7. Example: Non-causal AR(1)

⟨sec:noncausal⟩ In this section, we consider an example where the model is wrongly specified. We generate time series of length $n = 2000$ from a non-causal AR(1) model with $\phi = 1.67$ and t-distributed noise with degree of freedom 2.5. Then we fit a causal AR(1) model, where $|\phi| < 1$, to the data and obtain the corresponding residuals. Again the ADCF is evaluated using the Gaussian weight measure $N(0, 0.5^2)$ and in Figure 3, we plot the 5% and 95% ADCF quantiles of:

a) estimated residuals from 1000 independent simulations;
b) estimated residuals from 1000 independent parametric bootstrap samples from one realization of $\{X_t\}$.

The ADCFs of the bootstrapped residuals provide an approximation for the limiting distribution of the ADCF of the residuals given the model is correctly specified. In this case, the ADCFs 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.

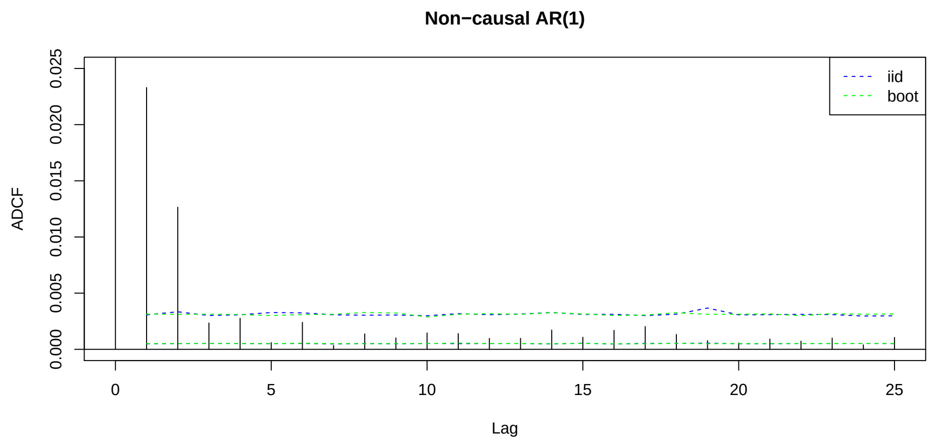

Fig 3. *Empirical* 5% *and* 95% *quantiles of the ADCF for a) iid innovations; b) bootstrapped residuals; from non-causal AR(1) data fitted with a causal AR(1) model.*

⟨fig:ncar⟩

8. Conclusion

In this paper, we examined the serial dependence of estimated residuals for time series models via the auto-distance covariance function (ADCV) and derived the asymptotic result for general classes of time series models. We showed theoretically that the limiting behavior differs from the ADCV for iid innovations by a correction term. This indicated that adjustments should be made when testing the goodness-of-fit of the model by inspecting the serial dependence of residuals. We illustrated 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 is incorrectly fitted with a causal model and showed that ADCV correctly detected model misspecification when applied to the residuals.

References

kokoszka:2003 [1] I. Berkes, L. Horváth, and P. Kokoszka. GARCH processes: structure and estimation. *Bernoulli*, 9(2): 201–227, 2003.

ingsley:1999 [2] P. Billingsley. *Convergence of Probability Measures.* Wiley, New York., 2nd edition, 1999.

l:davis:1991 [3] P.J. Brockwell and R.A. Davis. *Time Series: Theory and Methods.* Springer, New York., 1991.

davis:1996 [4] R.A. Davis. Gauss-Newton and M-estimation for ARMA proces with infinite variance. *Stoch. Process. Appl.*, 63:75–95, 1996.

osch:wan:2018 [5] R.A. Davis, M. Matsui, T. Mikosch, and P. Wan. Applications of distance covariance to time series. *Bernoulli*, 24(4A):3087–3116, 2018.

erverger:1993 [6] A. Feuerverger. A consistent test for bivariate dependence. *Internat. Statis. Rev.*, 61(3):419–433, 1993.

cholkopf:2005 [7] A. Gretton, O. Bousquet, A. Smola, and Schölkopf B. Measuring statistical dependence with Hilbert-Schmidt norms. In Sanjay Jain, Hans Ulrich Simon, and Etsuji Tomita, editors, *Algorithmic Learning Theory*, pages 63–77, Berlin, Heidelberg, 2005. Springer Berlin Heidelberg.

erger:yu:2005 [8] R. Kulperger and H. Yu. High moment partial sum processes of residuals in garch models and their applications. *Ann. Statist.*, 33(5):2395–2422, 2005.

neumann:2009 [9] A. Leucht and M.H. Neumann. Consistency of general bootstrap methods for degenerate U-type and V-type statistics. *J. Multiv. Anal.*, 100:1622–1633, 2009.

dnitsky:2016 [10] G. Samorodnitsky. *Stochastic Processes and Long Range Dependence.* Springer, New York., 2016.

akirov:2007 [11] G.J. Székely, M.L. Rizzo, and N.K. Bakirov. Measuring and testing dependence by correlation of distances. *Ann. Statist.*, 35:2769–2794, 2007.

zhou:2012 [12] Z. Zhou. Measuring nonlinear dependence in time-series, a distance correlation approach. *J. Time Ser. Anal.*, 33:438–457, 2012.

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

⟨app:meta⟩ *Proof.* The proof proceeds in the following steps with the aids 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

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

and

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

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

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

It follows from the continuous mapping theorem that

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

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

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

and

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

These are shown in Propositions A.2 and A.3, respectively.

□

⟨prop:joint:conv⟩ **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

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

then

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

eq:decomp:en

We now derive the limit of $E_n(s,t)$. For fixed s and t,

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

By assumption (M3),

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

It follows from a Taylor expansion that

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

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, it follows from the ergodic theorem (see, for example, Corollary 2.1.8 of Samorodnitsky (2016)) **Need uniform ergodic theorem** that

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

Hence,

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

Note that

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

and

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

We have

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

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

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

and

$$\begin{aligned}
&\mathbf{Q}^T \left(\mathbf{C}_h(s,t) - \mathbf{C}_h(s,0)\varphi_Z(t) - \mathbf{C}_h(0,t)\varphi_Z(s)\right) \\
&= \mathbf{Q}^T \left(\mathbb{E}\left[\mathrm{e}^{isZ_j+itZ_{j+h}} it\mathbf{L}_{j+h}(\boldsymbol{\beta})\right] - \mathbb{E}\left[\mathrm{e}^{itZ_{j+h}} it\mathbf{L}_{j+h}(\boldsymbol{\beta})\right]\varphi_Z(s)\right) = \xi_h(s,t). \qquad \text{(A.2)}
\end{aligned}$$

eq:xi:cal

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

For the joint convergence of $\sqrt{n}(C_n^{\hat{Z}} - C_n^Z)$ 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

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

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 non-random and continuous, the joint convergence $\sqrt{n}C_n^Z$ and $\sqrt{n}C_n^{\hat{Z}} - \sqrt{n}C_n^Z$ also follows.

□

⟨prop:a2⟩ **Proposition A.2.** *Under the conditions of Theorem 3.1,*

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

Proof. Using telescoping sums, $C_n^{\hat{Z}} - C_n^Z$ has the following decomposition,

$$C_n^{\hat{Z}} - C_n^Z = \frac{1}{n}\sum_{j=1}^{n-h} A_jB_j - \frac{1}{n}\sum_{j=1}^{n-h} A_j \frac{1}{n}\sum_{j=1}^{n-h} B_j - \frac{1}{n}\sum_{j=1}^{n-h} U_j \frac{1}{n}\sum_{j=1}^{n-h} B_j - \frac{1}{n}\sum_{j=1}^{n-h} V_j \frac{1}{n}\sum_{j=1}^{n-h} A_j$$

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

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,

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

For $k = 1, 2$,

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

Therefore

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

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

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

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. □

⟨prop:a3⟩ **Proposition A.3.** *Under the conditions of Theorem 3.1,*

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

Proof. Note that

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

This implies

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

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

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

and the proposition is proved. □

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. The result for bootstrap estimator in Theorem 4.1 follows as a special case.

Let $\{Z_{1:n,n}\}$ be a triangular array such that

$$Z_{jn}\overset{iid}{\sim}F_n,\quad \forall j=1,\ldots,n,$$

where the distribution F_n converges to F

$$F_n\xrightarrow{d}F.$$

Let $\{\boldsymbol{\beta}_n\}$ be a sequence of parameter vectors such that

$$\boldsymbol{\beta}_n\to\boldsymbol{\beta}.$$

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

$$X_{j+1,n}=f(X_{-\infty:j,n},Z_{tn};\boldsymbol{\beta}_n).$$

Let $\hat{\boldsymbol{\beta}}_n$ be the parameter estimate from $\{X_{1:n,n}\}$, $\{\hat{Z}_{1:n,n}\}$ be the fitted residuals calculated through (1.4), and $T_n^*(h)$ be the ADCV of $\{\hat{Z}_{1:n,n}\}$ at lag h. We require the following conditions.

?⟨cond:n1⟩? (N1) Let $\mathcal{F}_j$ and $\mathcal{F}_{jn}$ be the σ-algebra generated by $\{X_k,k\le j\}$ and $\{X_{kn},k\le j\}$, respectively. We assume that the parameter estimate $\hat{\boldsymbol{\beta}}$ is of the form

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

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

where $\mathbf{m}$ satisfies

$$\mathbb{E}[\mathbf{m}(X_{-\infty:j,n};\boldsymbol{\beta})|\mathcal{F}_{j-1,n}]=\mathbb{E}[\mathbf{m}(X_{-\infty:j,n};\boldsymbol{\beta})|\mathcal{F}_{j-1}]=\mathbf{0},\quad \mathbb{E}\sup_n|\mathbf{m}(X_{-\infty:0};\boldsymbol{\beta}_n)|^2<\infty.$$

?⟨cond:n2⟩? (N2) Assume that the function h in the invertible representation (1.2) is continuously differentiable, and writing

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

we have

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

?⟨cond:n3⟩? (N3) For fixed j, let $\tilde{Z}_{jn}$ be the fitted residual based on the unobserved infinite sequence $\{X_{-\infty:j,n}\}$ obtained from (1.3), and $\hat{Z}_{jn}$ be the estimated residuals based on the finite sequence $\{X_{0:j,n}\}$ obtained from (1.4). Assume that $\tilde{Z}_{jn}$ is close to $\hat{Z}_{jn}$ such that

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

⟨thm:ta⟩ **Theorem B.1.** *Assume that (N1), (N2), (N3) and* (3.2) *holds, then*

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

Remark B.2. To prove Theorem 4.1, take $\boldsymbol{\beta}_n = \hat{\boldsymbol{\beta}}$ and $Z_{tn} = Z_t^*$. 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.

Proof of Theorem B.1. Note that $T_n^*(h)$ can be written as

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

where

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

and

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

The result is proved in two propositions. In Proposition B.3, 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

$$\lim_{\delta \to 0} \limsup_{n \to \infty} \mathbb{P} \quad \iint\Big(_{K_\delta^c} n|C_n^{\hat{Z}_n}|^2 \mu(ds,dt) > \varepsilon \Big) = 0.$$

This is done in Proposition B.4. □

⟨prop:b1⟩ **Proposition B.3.** *Given (N1), (N2) and (N3) are satisfied we have*

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

Proof. The proof is divided into the following steps.

Dealing with the triangular array. Consider a sequence $Z_1, Z_2, \ldots$ where $Z_j \overset{iid}{\sim} F$. For each j, we have $Z_{jn} \xrightarrow{d} Z_j$. By Skorohod representation theorem, there exists a sufficiently rich probability space $(\tilde{\Omega}, \tilde{\mathcal{A}}, \tilde{P})$ where $\tilde{\Omega} = \{(\omega_1, \omega_2, \ldots) : \omega_j \in \Omega_0\}$ for some Ω_0, and functions $z : \Omega_0 \to \mathbb{R}$, $z_n : \Omega_0 \to \mathbb{R}$, such that for each j,

$$\tilde{Z}_{jn} = z_n(\omega_j) \sim F_n, \quad \tilde{Z}_j = z(\omega_j) \sim F,$$

and

$$\tilde{Z}_{jn} \xrightarrow{\text{a.s.}} \tilde{Z}_j.$$

This argument is similar to that in Leucht and Neumann (2009). Since we are only concerned about the distributional limit of $nT_n^*(h)$, we may assume without loss of generality that $Z_{jn} \xrightarrow{\text{a.s.}} Z_j$ for each j.

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 $\sqrt{n} C_n^Z \xrightarrow{d} G_h$, where

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

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

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

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

$$C_n^{Z_n}(s,t) := \frac{1}{n}\sum_{j=1}^{n-h} \left(U_{jn} V_{jn} - \frac{1}{n}\sum_{j=1}^{n-h} U_{jn} \frac{1}{n}\sum_{j=1}^{n-h} V_{jn} \right),$$

where $U_{jn}(s) := e^{isZ_{jn}} - \varphi_{Z_n}(s)$ and $V_{jn}(t) := e^{itZ_{j+h,n}} - \varphi_{Z_n}(t)$. Without loss of generality, here we only show

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

For fixed s, the convergence follows since

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

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

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

We have

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

Note that

$$
\mathbb{E}|e^{isZ_{jn}}-e^{is'Z_{jn}}|^2 \leq \mathbb{E}|e^{i(s-s')Z_{jn}}-1|^2 \leq 2\mathbb{E}|Z_{jn}|^2|s-s'|^2.
$$

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

Convergence of $\sqrt{n}(\hat{\boldsymbol\beta}_n-\boldsymbol\beta_n)$. In this part we show that

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

eq:conv:betaboot

Let $\{X_{1:n}\}$ be the time series generated from the time series model (1.1) with parameter vector $\boldsymbol\beta$ and innovation sequence $\{Z_{1:n}\}$. From the proof of Proposition A.1,

$$
\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\mathbf{m}(X_{-\infty:j};\boldsymbol\beta) \xrightarrow{d} \mathbf{Q}.
$$

It suffices to show that

$$
\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\mathbf{m}(X_{-\infty:j,n};\boldsymbol\beta_n) - \frac{1}{\sqrt{n}}\sum_{j=1}^{n}\mathbf{m}(X_{-\infty:j};\boldsymbol\beta) \xrightarrow{p} 0. \tag{B.3}
$$

eq:conv:mboot

We have

$$
\begin{aligned}
& \mathbb{E}\left|\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\left(\mathbf{m}(X_{-\infty:j,n};\boldsymbol\beta_n)-\mathbf{m}(X_{-\infty:j};\boldsymbol\beta)\right)\right|^2 \\
=\ & \frac{1}{n}\sum_{j=1}^{n}\mathbb{E}\left|\mathbf{m}(X_{-\infty:j,n};\boldsymbol\beta_n)-\mathbf{m}(X_{-\infty:j};\boldsymbol\beta)\right|^2 \\
& +\frac{2}{n}\sum_{1\leq i<j\leq n}\mathbb{E}\left(\mathbf{m}(X_{-\infty:i,n};\boldsymbol\beta_n)-\mathbf{m}(X_{-\infty:i};\boldsymbol\beta)\right)\left(\mathbf{m}(X_{-\infty:j,n};\boldsymbol\beta_n)-\mathbf{m}(X_{-\infty:j};\boldsymbol\beta)\right) \\
=\ & \mathbb{E}\left|\mathbf{m}(X_{-\infty:0,n};\boldsymbol\beta_n)-\mathbf{m}(X_{-\infty:0};\boldsymbol\beta)\right|^2 \\
& +\frac{2}{n}\sum_{1\leq i<j\leq n}\mathbb{E}\left[\mathbb{E}\left[\left(\mathbf{m}(X_{-\infty:i,n};\boldsymbol\beta_n)-\mathbf{m}(X_{-\infty:i};\boldsymbol\beta)\right)\left(\mathbf{m}(X_{-\infty:j,n};\boldsymbol\beta_n)-\mathbf{m}(X_{-\infty:j};\boldsymbol\beta)\right)\mid \mathcal{F}_i,\mathcal{F}_{in}\right]\right]
\end{aligned}
$$

Since $\mathbb{E}[\mathbf{m}(X_{-\infty:j,n};\boldsymbol\beta_n)-\mathbf{m}(X_{-\infty:j};\boldsymbol\beta)\mid \mathcal{F}_i,\mathcal{F}_{in}]=0$, the second term disappears. By causality, $\mathbf{m}(X_{-\infty:0};\boldsymbol\beta)$ can be expressed as a function of $Z_{-\infty:0}$ and $\boldsymbol\beta$, and

$$
\mathbf{m}(X_{-\infty:0,n};\boldsymbol\beta_n)-\mathbf{m}(X_{-\infty:0};\boldsymbol\beta) = \tilde{\mathbf{m}}(Z_{-\infty:0,n};\boldsymbol\beta_n)-\tilde{\mathbf{m}}(Z_{-\infty:0};\boldsymbol\beta_n) \xrightarrow{\text{a.s.}} 0.
$$

Hence

$$
\mathbb{E}\left|\mathbf{m}(X_{-\infty:0,n};\boldsymbol\beta_n)-\mathbf{m}(X_{-\infty:0};\boldsymbol\beta)\right|^2 \to 0
$$

by condition (N1) and dominated convergence. This justifies (B.3) and hence (B.2).

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

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

Denote

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

Then similar to (A.1), we have

$$\sqrt{n}(C_n^{\hat{Z}_n}(s,t) - C_n^{Z_n}(s,t)) = E_n^n(s,t) - E_n^n(s,0)\frac{1}{n}\sum_{j=1}^{n-h} \mathrm{e}^{itZ_{j+h,n}} - E_n^n(0,t)\frac{1}{n}\sum_{j=1}^{n-h} \mathrm{e}^{is\hat{Z}_{jn}}.$$

From the decomposition of ξ_h in (A.2), it suffices to show that

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

We have

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

From condition (N3),

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

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

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

where $\boldsymbol{\beta}_n^* = \epsilon\boldsymbol{\beta}_n + (1-\epsilon)\hat{\boldsymbol{\beta}}_n$ for some $\epsilon \in [0,1]$. We have shown in the previous part that

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_n) \xrightarrow{d} \mathbf{Q}.$$

It remains to show that

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

This follows from

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

and

$$\begin{aligned} &\frac{1}{n}\sum_{j=1}^{n-h} \mathrm{e}^{isZ_{jn}+itZ_{j+h,n}} (is\mathbf{L}_j^n(\boldsymbol{\beta}_n^*) + it\mathbf{L}_{j+h}^n(\boldsymbol{\beta}_n^*)) - \frac{1}{n}\sum_{j=1}^{n-h} \mathrm{e}^{isZ_j+itZ_{j+h}} (is\mathbf{L}_j(\boldsymbol{\beta}) + it\mathbf{L}_{j+h}(\boldsymbol{\beta})) \\ \le\ & \frac{|s|+|t|}{n} \sum_{j=1}^{n} |\mathbf{L}_j^n(\boldsymbol{\beta}_n^*) - \mathbf{L}_j(\boldsymbol{\beta})| \\ \xrightarrow{p}\ & 0, \quad \text{in } \mathcal{C}(K), \end{aligned}$$

from dominated convergence.

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 implies 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

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

The join 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 $\sqrt{n}(C_n^{\hat{Z}}(s,t) - C_n^{Z}(s,t))$ in Proposition A.1. □

⟨prop:b2⟩ **Proposition B.4.**

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

Proof. This follows the same steps in the proof of Proposition A.2 by replacing all $\hat{Z}_j$ with $\hat{Z}_{jn}$ and Z_j with Z_{jn}. □

Appendix C: Proof of Corollary 5.1

⟨app:arma⟩ *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

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

while

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

Hence

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

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

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

(M3): Note that

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

For $k = 1, 2$,

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

For any $m < n$,

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left| \tilde{Z}_t - \hat{Z}_t \right|^k \leq \sum_{j=0}^{m} |\pi_j(\hat{\boldsymbol{\beta}})|^k \frac{1}{\sqrt{n}} \sum_{t=1}^{m} |X_{t-j}|^k + \sum_{j=m+1}^{\infty} |\pi_j(\hat{\boldsymbol{\beta}})|^k \frac{1}{\sqrt{n}} \sum_{t=1}^{n} |X_{t-j}|^k =: I_1 + I_2. \qquad \text{(C.1)}$$ eq:I1I2

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

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

and there exists $K > 0$ such that

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

It follows that

$$\sum_{j=0}^{\infty} |\pi_j(\hat{\boldsymbol{\beta}})|^k < \infty, \quad k = 1, 2.$$

Now for (C.1), I_1 converges to zero in probability for fixed m, while I_2 converges to zero uniformly as $m \to \infty$. This implies that

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

□

Appendix D: Proof of Corollary 6.1

⟨app:garch⟩ *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 $\hat{\boldsymbol{\theta}}_n$ has limiting distribution

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

where

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

(M2): We have

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

and

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

Lemma 3.1 of Kulperger and Yu (2005) showed that

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

Hence

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

(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

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

□