Application of Distance Covariance to Extremes and Time Series and Inference for Linear Preferential Attachment Networks

Phyllis Wan

Submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy
in the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2018

ABSTRACT

Application of Distance Covariance to Extremes and Time Series and Inference for Linear Preferential Attachment Networks

Phyllis Wan

This thesis covers four topics: i) Measuring dependence in time series through distance covariance; ii) Testing goodness-of-fit of time series models; iii) Threshold selection for multivariate heavy-tailed data; and iv) Inference for linear preferential attachment networks.

Topic i) studies a dependence measure based on characteristic functions, called distance covariance, in time series settings. Distance covariance recently gathered popularity for its ability to detect nonlinear dependence. In particular, we characterize a general family of such dependence measures and use them to measure lagged serial and cross dependence in stationary time series. Assuming strong mixing, we establish the relevant asymptotic theory for the sample auto- and cross- distance correlation functions.

Topic ii) proposes a goodness-of-fit test for general classes of time series model by applying the auto-distance covariance function (ADCV) to the fitted residuals. Under the correct model assumption, the limit distribution for the ADCV of the residuals differs from that of an i.i.d. sequence by a correction term. This adjustment has essentially the same form regardless of the model specification.

Topic iii) considers data in the multivariate regular varying setting where the radial part R is asymptotically independent of the angular part Θ as R goes to infinity. The goal is to estimate the limiting distribution of Θ given $R \to \infty$, which characterizes the tail dependence of the data. A typical strategy is to look at the angular components of the data for which the radial parts exceed some threshold. We propose an algorithm to select the threshold based on distance covariance statistics and a subsampling scheme.

Topic iv) investigates inference questions related to the linear preferential attachment model for network data. Preferential attachment is an appealing mechanism based on the intuition "the rich get richer" and produces the well-observed power-law behavior in networks. We provide methods for fitting such a model under two data scenarios, when the network formation is given, and when only a single-time snapshot of the network is observed.

Contents

List of Figures

List of Tables

Acknowledgements

First and foremost, I would like to thank my advisor, Richard Davis. Without his direction and support this dissertation would not have been possible. I thank him for motivating me, educating me, challenging me and caring for me during the past five years, and for being my role model for the years to come.

My sincere gratitude goes to Sid Resnick and Thomas Mikosch, for their mentorship and support since the beginning of my study. I thank Muneya Matsui, Tiandong Wang and Jon Auerbach, for being great collaborators. And I am grateful to Bodhi Sen, Serena Ng and Peter Orbanz, for being my dissertation examiners.

Warmest thanks to the many professors in the Department of Statistics who have supported me during my study. Special thanks to Dood Kalicharan and Anthony Cruz who took care of us on administrative matters and more. A big thank you to all my dear colleagues for their company and cheers. Despite the scarcity of office windows, every work day was bright and sunny.

This journey was made colorful by dear friends and loved ones. My thanks go to Yuting M, Lisha Q, Jiasi S, Jessie H, Kristen J, Elaine L, Cristina P, Xufei W, You W, Yu H, Richard N, Yongbum C, Shihao Y, Adrien H, Michal W, Matthias K and all friends in New York and around the world. I thank Jun for always supporting me, and Jasper for his friendship.

Finally, I owe my most profound gratitude to my parents. Their love and encouragement have meant the world to me.

To my family.

Overview

This thesis is a compilation of four papers concerning problems in time series, extreme value theory and network analysis. Three of the four papers explore the usage of distance covariance, a dependence measure that recently rose to popularity for its ability to detect nonlinear dependence. The fourth paper considers inference methods for a network model. In this introduction, we provide an overview of the problems and our contributions.

0.1 Distance correlation in time series setting

In time series analysis, the autocorrelation function (ACF) is perhaps the most used dependence measure to assess serial dependence. It provides a measure of linear dependence and is closely linked with the class of ARMA models. On the other hand, the ACF gives only a partial description of dependence. As seen with financial time series, when the data are uncorrelated but dependent, the ACF is often non-informative. In this case, the dependence only becomes visible by examining the ACF applied to the absolute values or squares of the time series. In Chapter 1, we consider the application of *distance correlation*, in place of linear correlation, to measure dependences in time series.

The intuition of distance covariance is based on the property that two random vectors $X \in \mathbb{R}^p$ and $Y \in \mathbb{R}^q$ are independent if and only if $\varphi_{X,Y}(s,t) = \varphi_X(s)\,\varphi_Y(t)$, 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

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

where μ is a suitable measure. It is easy to see that if μ has a positive Lebesgue density on $\mathbb{R}^{p+q}$, X and Y are independent if and only if $T(X,Y;\mu) = 0$. Given observations

$\{(X_i, Y_i), i = 1, \ldots, n\}$, the sample version of the distance covariance is given by

$$\hat{T}(X, Y; \mu) = \int_{\mathbb{R}^{p+q}} \left| \hat{\varphi}_{X,Y}(s,t) - \hat{\varphi}_X(s)\, \hat{\varphi}_Y(t) \right|^2 \mu(ds, dt)\,, \quad (s,t) \in \mathbb{R}^{p+q},$$

where $\hat{\varphi}_{X,Y}, \hat{\varphi}_X, \hat{\varphi}_Y$ are the corresponding joint and marginal empirical characteristic functions. When $\mu = \mu_1 \times \mu_2$ and is symmetric about the origin, it can be shown that $\hat{T}(X, Y; \mu)$ has a V-statistic form and can be obtained in $O(n^2)$ computation.

The concept of distance covariance was first proposed by Feuerverger (1993) for univariate variables X and Y. It was later christened with its current name and brought to popularity in a series of papers by Székely and co-authors (see, for example, Székely et al. (2007)). It was first applied to time series setting when Zhou (2012) introduced the auto-distance covariance function. Most literature on distance covariance focus on the specific weight measure $\mu(s,t) \propto |s|^{-p-1}|t|^{-q-1}$, which has the advantage of being scale and rotation invariant.

In Chapter 1, we consider the general form of distance covariance and apply it to stationary univariate and multivariate time series. For time series $\{X_t\}$, serial dependence is measured auto-distance covariance functions. For bivariate time series $\{(X_t, Y_t)\}$, cross dependence is measured using cross-distance covariance functions. We establish the asymptotic results for these statistics under strong mixing.

The work in this chapter was published in Davis et al. (2018):

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

0.2 Goodness-of-fit testing for time series models

In many statistical modeling frameworks, goodness-of-fit tests are often administered to the residuals. In Chapter 2, we apply the auto-distance covariance function (ADCV) to the fitted residuals to assess goodness-of-fit for general classes of time series models.

It is known that the sequence of fitted residuals generally admits a different serial dependence than the sequence of iid innovations. Let $\hat{T}_h(Z; \mu)$ be the sample ADCV of an iid sequence $\{Z_t, 1 \leq t \leq n\}$ at lag h. From the results in Chapter 1,

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

where G_h is a centered Gaussian process. Let $\hat{T}_h(\hat{Z};\mu)$ be the sample ADCV in which the iid sequence is replaced by the residuals $\{\hat{Z}_t, 1 \leq t \leq n\}$. We show that

$$n\hat{T}_h(\hat{Z};\mu) \xrightarrow{d} \int |G_h(s,t) + \xi(s,t)|^2 \mu(ds,dt).$$

We demonstrate through simulations that the impact of the correction term $\xi(s,t)$ is non-trivial. This implies that adjustments are necessary when using this statistic to evaluate the goodness-of-fit of the model. Otherwise, an improper model may be accepted based on an incorrect threshold for the test statistics. Given a sequence of observations from the time series, the limit can be approximated through a parametric bootstrap.

A manuscript containing the results in this chapter is currently under development.

0.3 Threshold selection for multivariate heavy-tailed data

Regular variation is a typical assumption for modeling multivariate heavy-tailed data. A random vector $\mathbf{X} \in \mathbb{R}^d$ is *multivariate regularly varying* if the polar coordinates $(R, \boldsymbol{\Theta}) = (\|\mathbf{X}\|, \mathbf{X}/\|\mathbf{X}\|)$, where $\|\cdot\|$ is some norm, satisfy the conditions:

(a) R has a univariate Pareto-like tail;

(b) $\mathbb{P}(\boldsymbol{\Theta} \in \cdot | R > r)$ converges weakly to a probability measure $S(\cdot)$ as $r \to \infty$.

Here the limit S characterizes the tail dependence and is often the quantity of interest.

To estimate S, a common strategy is to look at the angular components $\boldsymbol{\Theta}$ of the data for which the radial parts R exceed some threshold. A large class of methods has been proposed in the literature to model these exceedances. The choice of threshold, however, has scarcely been discussed.

In order to select the threshold, the dependence between R and $\mathbf{\Theta}$ needs to be characterized. Linear correlation proves to be inadequate for this task for two reasons. First, R is heavy-tailed and often does not possess a second moment, thus violating the assumption for linear correlation. Second, $\mathbf{\Theta}$ can be of multiple dimensions. In Chapter 3, we propose to use distance covariance for this purpose. Given a sequence of thresholds $\{r_n\}$, we formally test the independence between R and $\mathbf{\Theta}$ conditional on $R > r_n$, using distance covariance. Our approach to this problem is based on the following two steps.

First, as $n \to \infty$, $(R_i, \mathbf{\Theta}_i)\mathbf{1}_{\{R_i > r_n\}}$, $i = 1, \ldots, n$, forms a triangular array. We generalize the limit theory of distance covariance in Chapter 1 to a triangular array setting. The results are given for both iid and weakly dependent data.

Second, the test of independence statistics are summarized in the form of p-values for different thresholds. To select an optimal threshold, we propose an algorithm which determines the change point from which the mean of the p-value distribution deviates from 0.5, the mean of its distribution under the null. This is done by subsampling the data and using a wild binary segmentation change point detection procedure. The subsampling scheme allows the method to be applicable to a wide range of weakly dependent data and also avoids the heavy computation in the calculation of distance covariance, a typical limitation for this measure.

The research in chapter will appear in a forthcoming paper, Wan and Davis (2018):

P. Wan and R.A. Davis. Threshold selection for multivariate heavy-tailed data. Extremes. *2018.*

0.4 Inference for preferential attachment model

Lastly, we turn our attention to another data type – networks. We are interested in the power-law behavior of the degree distributions observed in many networks, most notably in social networks. A discrete distribution D is said to possess a *power law* if

$$\mathbb{P}(D = i) \sim c \cdot i^{-\alpha}, \quad c > 0.$$

In other words, D is heavy-tailed. The study of power laws has always been of interest. In a network, the nodes with large degrees represent the individuals with large number of connections and hence are likely to be influential. If a network exhibits power laws in its degree distributions, the occurrence of nodes with large degrees is non-negligible.

Preferential attachment is a natural and appealing mechanism that models such behavior. It is based on the intuition of *the rich get richer*, that a connection is more likely to be made to an individual with many existing connections than one with less. Such models produce networks with the empirically observed power-law property and have been implemented empirically for many networks. However, until recently, few studies have focused on its mathematical properties and no rigorous estimation procedure has been proposed.

In Chapter 4, we bridge this gap by considering fitting a 5-parameter linear preferential model to directed networks. We proposed inference methods under two data scenarios. In the case where full history of the network formation is given, we derive the maximum likelihood estimator of the parameters and show strong consistency and asymptotical normality. In the case where only a single-time snapshot of the network is available, we propose an estimation method which combines method of moments with an approximation to the likelihood. The resulting estimator is also strongly consistent and performs quite well compared to the MLE estimator. We illustrate both estimation procedures through simulated data and explore the usage of this model in a real data example.

This work was published in Wan et al. (2017):

P. Wan, T. Wang, R.A. Davis, and S.I. Resnick. Fitting the linear preferential attachment model. Electron. J. Statist.*, 11:3738–3780, 2017.*

Chapter 1

Applications of distance correlation to time series

1.1 Introduction

In time series analysis, modeling serial dependence is typically the overriding objective. In order to achieve this goal, it is necessary to formulate a measure of dependence and this may depend on the features in the data that one is trying to capture. The autocorrelation function (ACF), which provides a measure of linear dependence, is perhaps the most used dependence measure in time series. It is closely linked with the class of ARMA models and provides guidance in both model selection and model confirmation. On the other hand, the ACF gives only a partial description of serial dependence. As seen with financial time series, data are typically uncorrelated but dependent so that the ACF is non-informative. In this case, the dependence becomes visible by examining the ACF applied to the absolute values or squares of the time series. In this chapter we consider the application of distance correlation in a time series setting, which can overcome some of the limitations of other dependence measures.

In recent years, the notions of distance covariance and correlation have become rather popular in applied statistics. Given vectors X and Y with values in $\mathbb{R}^p$ and $\mathbb{R}^q$, the *distance covariance* between X and Y with respect to a suitable measure μ on $\mathbb{R}^{p+q}$ is given by

$$T(X,Y;\mu) = \int_{\mathbb{R}^{p+q}} \left|\varphi_{X,Y}(s,t) - \varphi_X(s)\,\varphi_Y(t)\right|^2 \mu(ds,dt)\,, \tag{1.1}$$

where the characteristic function of any random vector $Z \in \mathbb{R}^d$ is denoted by $\varphi_Z(t) = \mathbb{E}[e^{\,i\langle t,Z\rangle}]$, $t \in \mathbb{R}^d$. The *distance correlation* is the corresponding version of T standardized to values in $[0,1]$. The quantity $T(X,Y;\mu)$ is zero if and only if $\varphi_{X,Y} = \varphi_X\,\varphi_Y$, μ-a.e. In many situations, for example when μ has a positive Lebesgue density on $\mathbb{R}^{p+q}$, we may conclude that X and Y are independent if and only if $T(X,Y;\mu) = 0$. An empirical version $T_n(X,Y;\mu)$ of $T(X,Y;\mu)$ is obtained if the characteristic functions in (1.1) are replaced by their corresponding empirical versions. Then one can build a test for independence between X and Y based on the distribution of T_n under the null hypothesis that X and Y are independent.

The use of empirical characteristic functions for univariate and multivariate sequences for inference purposes has a long history. In the 1970s and 1980s, Feuerverger and Mureika (1977), Csörgő (1981a,b,c) and many others proved fundamental asymptotic results for iid sequences, including Donsker-type theory for the empirical characteristic function. Statisticians have applied these methods for goodness-of-fit tests, changepoint detection, testing for independence, etc.; see for example Meintanis and coworkers (Meintanis and Iliopoulos (2008), Hlávka et al. (2011), Meintanis et al. (2015)), and the references therein. The latter authors employed the empirical distance covariance for finite measures μ. Feuerverger (1993) was the first to apply statistics of the form (1.1) for general measures. In particular, he advocated the infinite measure

$$\mu(ds,dt) = |s|^{-2}|t|^{-2} ds\, dt$$

for testing independence of univariate data. A series of papers by Székely et al.[1] (Székely et al. (2007), Székely and Rizzo (2009, 2014), see also the references therein) developed asymptotic techniques for the empirical distance covariance and correlation of iid sequences for the infinite measure μ given by

$$\mu(ds,dt) = c_{p,q}|s|^{-\alpha-p}|t|^{-\alpha-q} ds\, dt, \tag{1.2}$$

[1]They appeared to have coined the terms distance covariance and correlation.

where $c_{p,q}$ is a constant (see (1.15)) and $\alpha \in (0,2)$. With this choice of μ, the distance correlation, $T(X,Y;\mu)/(T(X,X;\mu)T(Y,Y;\mu))^{1/2}$ is invariant relative to scale and orthogonal transformations, two desirable properties for measures of dependence. As a consequence this choice of measure is perhaps the most common. However, there are other choices of measures for μ that are also useful depending on the context.

Dueck et al. (2014) studied the affinely invariant distance covariance given by $\tilde{T}(X,Y;\mu) = T(\Sigma_X^{-1}X, \Sigma_Y^{-1}Y)$, where Σ_X, Σ_Y are the respective covariance matrices of X and Y and μ is given by (1.2). They showed that the empirical version of $\tilde{T}(X,Y;\mu)/\sqrt{\tilde{T}(X,X;\mu)\tilde{T}(Y,Y;\mu)}$, where Σ_X and Σ_Y are estimated by their empirical counterparts, is strongly consistent. In addition, they provide explicit expressions in terms of special functions of the limit in the case when X,Y are multivariate normal. Further progress on this topic has been achieved in Sejdinovic et al. (2013) and Lyons (2013), who generalized distance correlation to a metric space.

In this chapter we are interested in the empirical distance covariance and correlation applied to a stationary sequence $((X_t,Y_t))$ to study serial dependence, where X_t and Y_t assume values in $\mathbb{R}^p$ and $\mathbb{R}^q$, respectively. We aim at an analog to the autocorrelation and autocovariance functions of classical time series analysis in terms of lagged distance correlation and distance covariance. Specifically we consider the lagged-distance covariance function $T(X_0,Y_h;\mu)$, $h \in \mathbb{Z}$, and its standardized version that takes values in $[0,1]$. We refer to these quantities as the *auto- and cross-distance covariance and correlation functions*. We provide asymptotic theory for the empirical auto- and cross-distance covariance and correlation functions under mild conditions. Under ergodicity we prove consistency and under α-mixing, we derive the weak limits of the empirical auto- and cross-distance covariance functions for both cases when X_0 and Y_h are independent and dependent.

From a modeling perspective, distance correlation has limited value in providing a clear description of the nature of the dependence in the time series. To this end, it may be difficult to find a time series model that produces a desired distance correlation. In contrast, one could always find an autoregressive (or more generally ARMA) process that matches the ACF for an arbitrary number of lags. The theme in this chapter will be to view the distance correlation more as a tool for testing independence rather than actually measuring dependence.

The literature on distance correlation for dependent sequences is sparse. To the best of our knowledge, Zhou (2012) was the first to study the auto-distance covariance and its empirical analog for stationary sequences. In particular, he proved limit theory for $T_n(X_0, X_h; \mu)$ under so-called physical dependence measure conditions on (X_t) and independence of X_0 and X_h. Fokianos and Pitsillou (2017) developed limit theory for a Ljung-Box-type statistic based on pairwise distance covariance $T_n(X_i, X_j; \mu)$ of a sample from a stationary sequence. In both papers, the measure μ is given by (1.2). The latter paper uses ideas from Hong (1999). He applied the empirical characteristic function of a strongly mixing time series for testing various hypotheses on the dependence structure of a time series; he called it a generalized spectral approach. His test statistic bears some resemblance with the distance covariance: it is an integral of the weighted squared difference between the Fourier transform of the sequence $\text{cov}(e^{iuX_0}, e^{ivX_h})$ and an empirical analog weighted by the density of a finite measure μ.

Typically, a crucial and final step in checking the quality of a fitted time series model is to examine the residuals for lack of serial dependence. The distance correlation can be used in this regard. However, as first pointed out in his discussion, Rémillard (2009) indicated that the behavior of the distance correlation when applied to the residuals of a fitted AR(1) process need not have the same limit distribution as that of the distance correlation based on the corresponding iid noise. We provide a rigorous proof of this result for a general AR(p) process with finite variance under certain conditions on the measure μ. Interestingly, the conditions preclude the use of the standard weight function (1.2) used in Székely et al. (2007). In contrast, if the noise sequence is heavy-tailed and belongs to the domain of attraction of a stable distribution with index $\beta \in (0, 2)$, the distance correlation functions for both the residuals from the fitted model and the iid noise sequence coincide.

The chapter is organized as follows. In Section 1.2 we commence with some basic results for distance covariance. We give conditions on the moments of X and Y and the measure μ, which ensure that the integrals $T(X,Y;\mu)$ in (1.1) are well-defined. We provide alternative representations of $T(X,Y;\mu)$ and consider various examples of finite and infinite measures μ. Section 1.3 is devoted to the empirical auto- and cross-distance covariance and correlation functions. Our main results on the asymptotic theory of these functions are provided in Section 1.3.1. Among them are an a.s. consistency result (Theorem 1.3.1) under the assumption of ergodicity and asymptotic normality under a strong mixing condition (Theorem 1.3.2). Another main result (Theorem 1.4.2) is concerned with the asymptotic behavior of the empirical auto-distance covariance function of the residuals of an autoregressive process for both the finite and infinite variance cases. In Section 1.5, we provide a small study of the empirical auto-distance correlation functions derived from simulated and real-life dependent data of moderate sample size. The proofs of Lemma 1.4.1 and Theorem 1.4.2, which are significant but very technical, are relegated to Section **??**.

1.2 Distance covariance for stationary time series

1.2.1 Conditions for existence

From (1.1), the distance covariance between two vectors X and Y is the squared L^2-distance between the joint characteristic function of (X,Y) and the product of the marginal characteristic functions of X and Y with respect to a measure μ on $\mathbb{R}^{p+q}$. Throughout we assume that μ is finite on sets bounded away from the origin, i.e., on sets of the form

$$D_\delta^c = \{(s,t) : |s| \wedge |t| > \delta\}\,, \qquad \delta > 0\,. \tag{1.3}$$

In what follows, we interpret (s,t) as a concatenated vector in $\mathbb{R}^{p+q}$ equipped with the natural norm $|(s,t)|_{\mathbb{R}^p\times\mathbb{R}^q} = \sqrt{|s|^2+|t|^2}$. We suppress the dependence of the norm $|\cdot|$ on the dimension. The symbol c stands for any positive constant, whose value may change from line to line, but is not of particular interest. Clearly if X and Y are independent, $T(X,Y;\mu)=0$. On the other hand, if μ is an infinite measure, and X and Y are dependent, extra conditions are needed to ensure that $T(X, Y; \mu)$ is finite. This is the content of the following lemma.

Lemma 1.2.1. *Let X and Y be two possibly dependent random vectors and one of the following conditions is satisfied:*

1. *μ is a finite measure on $\mathbb{R}^{p+q}$.*
2. *μ is an infinite measure on $\mathbb{R}^{p+q}$, finite on the sets D_δ^c, $\delta > 0$, such that*

$$\int_{\mathbb{R}^{p+q}} (1 \wedge |s|^\alpha)\,(1 \wedge |t|^\alpha)\,\mu(ds, dt) < \infty \tag{1.4}$$

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

3. *μ is infinite in a neighborhood of the origin and for some $\alpha \in (0, 2]$, $\mathbb{E}[|X|^\alpha] + \mathbb{E}[|Y|^\alpha] < \infty$ and*

$$\int_{\mathbb{R}^{p+q}} 1 \wedge |(s, t)|^\alpha\,\mu(ds, dt) < \infty\,. \tag{1.5}$$

Then $T(X, Y; \mu)$ is finite.

Remark 1.2.2. If $\mu = \mu_1 \times \mu_2$ for some measures μ_1 and μ_2 on $\mathbb{R}^p$ and $\mathbb{R}^q$, respectively, and if μ is finite on the sets D_δ^c then it suffices for (1.4) to verify that

$$\int_{|s| \le 1} |s|^\alpha\,\mu_1(ds) + \int_{|t| \le 1} |t|^\alpha\,\mu_2(dt) < \infty\,.$$

Proof. (1) Since the integrand in $T(X, Y; \mu)$ is uniformly bounded the statement is trivial. (2) By (1.3), $\mu(D_\delta^c) < \infty$ for any $\delta > 0$. Therefore it remains to verify the integrability of $|\varphi_{X,Y}(s,t) - \varphi_X(s)\,\varphi_Y(t)|^2$ on one of the sets D_δ. We consider only the case $|s| \vee |t| \le 1$; the cases when $|s| \le 1$, $|t| > 1$ and $|s| > 1$, $|t| \le 1$ are similar. An application of the Cauchy-Schwarz inequality yields

$$|\varphi_{X,Y}(s,t) - \varphi_X(s)\varphi_Y(t)|^2 \le (1 - |\varphi_X(s)|^2)\,(1 - |\varphi_Y(t)|^2)\,. \tag{1.6}$$

Since $1 - |\varphi_X(s)|^2 = \int_{\mathbb{R}^p} \big(1 - \cos\langle s, x\rangle\big)\,\mathbb{P}(X - X' \in dx)$ for an independent copy X' of X, a Taylor expansion and the fact that X, X' have finite αth moments yield for $\alpha \in (0, 2]$ and some constant $c > 0$,

$$\begin{aligned} 1 - |\varphi_X(s)|^2 &\leq \int_{\mathbb{R}^p} \big(2 \wedge |\langle s, x\rangle|^2\big)\, \mathbb{P}(X - X' \in dx) \\ &\leq 2 \int_{|\langle s,x\rangle| \leq \sqrt{2}} |\langle s, x\rangle/\sqrt{2}|^\alpha\, \mathbb{P}(X - X' \in dx) + 2\,\mathbb{P}(|\langle s, X - X'\rangle| > \sqrt{2}) \\ &\leq c\,|s|^\alpha\, \mathbb{E}[|X - X'|^\alpha] < \infty\,. \end{aligned} \tag{1.7}$$

In the last step we used Markov's inequality and the fact that $|\langle s, x\rangle| \leq |s|\,|x|$. A corresponding bound holds for $1 - |\varphi_Y(t)|^2$. Now, $T(X, Y; \mu) < \infty$ follows from (1.4) and (1.6).

(3) By (1.5), $\mu(\{(s, t) : |(s, t)| > 1\})$ is finite. Therefore we need to show integrability of $|\varphi_{X,Y}(s, t) - \varphi_X(s)\varphi_Y(t)|^2$ only for $|(s, t)| \leq 1$. Using the arguments from part (2) and the finiteness of the αth moments, we have

$$|\varphi_{X,Y}(s, t) - \varphi_X(s)\varphi_Y(t)|^2 \leq c\,(|s|^\alpha + |t|^\alpha) \leq c\,|(s, t)|^\alpha\,.$$

Now integrability of the left-hand side at the origin with respect to μ is ensured by (1.5). □

1.2.2 Alternative representations and examples

If $\mu = \mu_1 \times \mu_2$ for measures μ_1 and μ_2 on $\mathbb{R}^p$ and $\mathbb{R}^q$ we write for $x \in \mathbb{R}^p$ and $y \in \mathbb{R}^q$,

$$\begin{aligned} \hat{\mu}(x, y) &= \int_{\mathbb{R}^{p+q}} \cos(\langle s, x\rangle + \langle t, y\rangle)\, \mu(ds, dt)\,, \\ \hat{\mu}_1(x) &= \int_{\mathbb{R}^p} \cos\langle s, x\rangle\, \mu_1(ds)\,, \quad \hat{\mu}_2(y) = \int_{\mathbb{R}^q} \cos\langle t, y\rangle\, \mu_2(dt)\,, \end{aligned}$$

for the real parts of the Fourier transforms with respect to μ, μ_1, μ_2, respectively. We assume that these transforms are well-defined. Let (X', Y') be an independent copy of (X, Y), and let Y'' and Y''' be independent copies of Y which are also independent of $(X, Y), (X', Y')$. We have

$$\begin{aligned} T(X, Y; \mu) &= \int_{\mathbb{R}^{p+q}} \mathbb{E}\Big[e^{\,i\langle s, X - X'\rangle + i\langle t, Y - Y'\rangle} + e^{\,i\langle s, X - X'\rangle}\, e^{\,i\langle t, Y'' - Y'''\rangle} \\ &\qquad - e^{\,i\langle s, X - X'\rangle + i\langle t, Y - Y''\rangle} - e^{\,-i\langle s, X - X'\rangle - i\langle t, Y - Y''\rangle}\Big]\mu(ds, dt)\,. \end{aligned} \tag{1.8}$$

Notice that the complex-valued trigonometric functions under the expected value may be replaced by their real parts. We intend to interchange the integral with respect to μ and the expectation.

Finite measure μ

For a finite measure on $\mathbb{R}^{p+q}$, we may apply Fubini's theorem directly and interchange integration with expectation to obtain

$$\begin{aligned} T(X,Y;\mu) &= \mathbb{E}\big[\hat{\mu}(X-X',Y-Y')] + \mathbb{E}[\hat{\mu}(X-X',Y''-Y''')] \\ &\quad -2\,\mathbb{E}[\hat{\mu}(X-X',Y-Y'')\big]\,. \end{aligned} \tag{1.9}$$

If $\mu = \mu_1 \times \mu_2$ we also have

$$\begin{aligned} T(X,Y;\mu) &= \mathbb{E}[\hat{\mu}_1(X-X')\,\hat{\mu}_2(Y-Y')] + \mathbb{E}[\hat{\mu}_1(X-X')]\mathbb{E}[\hat{\mu}_2(Y-Y')] \\ &\quad -2\,\mathbb{E}[\hat{\mu}_1(X-X')\,\hat{\mu}_2(Y-Y'')]\,. \end{aligned}$$

Infinite measure μ

We consider an infinite measure μ on $\mathbb{R}^{p+q}$ which is finite on D_δ^c for any $\delta > 0$. We assume that $T(X,Y;\mu)$ is finite and $\mu = \mu_1 \times \mu_2$. In this case, we cannot pass from (1.8) to (1.9) because the Fourier transform $\hat{\mu}$ is not defined as a Lebesgue integral. We have

$$T(X,Y;\mu) = \int_{\mathbb{R}^{p+q}} \big(\mathbb{E}[\mathrm{COS}(s,t)] + \mathbb{E}[\mathrm{S}IN(s,t)]\big)\,\mu(ds,dt)\,, \tag{1.10}$$

where

$$\begin{aligned} \mathrm{COS}(s,t) &= \cos(\langle s,X-X'\rangle)\,\cos(\langle t,Y-Y'\rangle) + \cos(\langle s,X-X'\rangle)\,\cos(\langle t,Y''-Y'''\rangle) \\ &\quad -2\,\cos(\langle t,X-X'\rangle)\,\cos(\langle s,Y-Y''\rangle)\,, \\ \mathrm{SIN}(s,t) &= -\sin(\langle s,X-X'\rangle)\,\sin(\langle t,Y-Y'\rangle) - \sin(\langle s,X-X'\rangle)\,\sin(\langle t,Y''-Y'''\rangle) \\ &\quad +2\,\sin(\langle t,X-X'\rangle)\,\sin(\langle s,Y-Y''\rangle)\,. \end{aligned}$$

Using the fact that

$$\cos u\,\cos v = 1-(1-\cos u)-(1-\cos v)+(1-\cos u)(1-\cos v)\,,$$

calculation shows that

$$\mathbb{E}\big[\mathrm{COS}(s,t)\big] = \mathbb{E}\big[(1-\cos(\langle s,X-X'\rangle))\,(1-\cos(\langle t,Y-Y'\rangle))$$

$$+(1-\cos(\langle s, X-X'\rangle))\,(1-\cos(\langle t, Y''-Y'''\rangle))$$
$$-2\,(1-\cos(\langle t, X-X'\rangle))\,(1-\cos(\langle s, Y-Y''\rangle))\big].$$

A Taylor series argument shows that for $\alpha \in (0,2]$,

$$\begin{aligned}\mathbb{E}[|\mathrm{COS}(s,t)|] \;\le\; c\Big(&\mathbb{E}\big[(1\wedge|\langle s, X-X'\rangle/\sqrt{2}|^\alpha)\,(1\wedge|\langle t, Y-Y'\rangle/\sqrt{2}|^\alpha)\big]\\ &+\mathbb{E}\big[1\wedge|\langle s, X-X'\rangle/\sqrt{2}|^\alpha\big]\,\mathbb{E}\big[1\wedge|\langle t, Y-Y'\rangle/\sqrt{2}|^\alpha\big]\\ &+\mathbb{E}\big[(1\wedge|\langle t, X-X'\rangle/\sqrt{2}|^\alpha)\,(1\wedge|\langle s, Y-Y''\rangle/\sqrt{2}|^\alpha|\big]\Big).\end{aligned}$$

Under condition (1.4) the right-hand side is integrable with respect to μ if

$$\mathbb{E}[|X|^\alpha + |Y|^\alpha + |X|^\alpha\,|Y|^\alpha] < \infty\,. \tag{1.11}$$

An application of Fubini's theorem yields

$$\begin{aligned}&\int_{\mathbb{R}^{p+q}} \mathbb{E}[\mathrm{COS}(s,t)]\,\mu(ds,dt)\\ &= \mathbb{E}\Big[\int_{\mathbb{R}^{p+q}}\Big((1-\cos(\langle s, X-X'\rangle))\,(1-\cos(\langle t, Y-Y'\rangle))\\ &\quad+(1-\cos(\langle s, X-X'\rangle))\,(1-\cos(\langle t, Y''-Y'''\rangle))\\ &\quad-2\,(1-\cos(\langle t, X-X'\rangle))\,(1-\cos(\langle s, Y-Y''\rangle))\Big)\,\mu(ds,dt)\Big].\end{aligned}$$

If we assume that the restrictions μ_1, μ_2 of μ to $\mathbb{R}^p$ and $\mathbb{R}^q$ are symmetric about the origin then we have $\mathbb{E}\big[\mathrm{SIN}(s,t)\big] = -\mathbb{E}\big[\mathrm{SIN}(-s,t)\big] = -\mathbb{E}\big[\mathrm{SIN}(s,-t)\big]$. Together with the symmetry property of μ this implies that $\int_{\mathbb{R}^{p+q}} \mathbb{E}\big[\mathrm{SIN}(s,t)\big]\,\mu(ds,dt) = 0$.

We summarize these arguments. For any measure ν on $\mathbb{R}^d$ we write

$$\tilde{\nu}(s) = \int_{\mathbb{R}^d} (1-\cos\langle s, x\rangle)\,\nu(dx)\,, \qquad s \in \mathbb{R}^d\,.$$

Lemma 1.2.3. *Assume* (1.4) *and* (1.11) *for some* $\alpha \in (0,2]$. *If* μ_1, μ_2 *are symmetric about the origin and* $\mu = \mu_1 \times \mu_2$ *then*

$$\begin{aligned}T(X,Y;\mu) \;=\; &\mathbb{E}[\tilde{\mu}_1(X-X')\,\tilde{\mu}_2(Y-Y')] + \mathbb{E}[\tilde{\mu}_1(X-X')]\,\mathbb{E}[\tilde{\mu}_2(Y-Y')]\\ &-2\,\mathbb{E}[\tilde{\mu}_1(X-X')\tilde{\mu}_2(Y-Y'')]\,.\end{aligned} \tag{1.12}$$

Remark 1.2.4. For further use, we mention the alternative representation of (1.12):

$$\begin{aligned} T(X,Y;\mu) &= \operatorname{cov}\big(\tilde\mu_1(X-X'),\tilde\mu_2(Y-Y')\big) \\ &\qquad -2\operatorname{cov}\Big(\mathbb{E}[\tilde\mu_1(X-X')\mid X]\,,\mathbb{E}[\tilde\mu_2(Y-Y')\mid Y]\Big)\,. \end{aligned} \tag{1.13}$$

Examples

Example 1.2.5. *Assume that μ has density w on $\mathbb{R}^{p+q}$ given by*

$$w(s,t)=c_{p,q}\,|s|^{-\alpha-p}\,|t|^{-\alpha-q}\,,\qquad s\in\mathbb{R}^p\,,t\in\mathbb{R}^q\,, \tag{1.14}$$

for some positive constant $c_{p,q}=c_pc_q$. For any $d\ge 1$ and $\alpha\in(0,2)$, one can choose c_d such that

$$\int_{\mathbb{R}^d}(1-\cos\langle s,x\rangle)\,c_d\,|s|^{-\alpha-d}\,ds=|x|^\alpha\,. \tag{1.15}$$

Under the additional moment assumption (1.11) *we obtain from* (1.12)

$$\begin{aligned} T(X,Y;\mu) &= \mathbb{E}[|X-X'|^\alpha\,|Y-Y'|^\alpha]+\mathbb{E}[|X-X'|^\alpha]\,\mathbb{E}[Y-Y'|^\alpha] \\ &- \quad 2\,\mathbb{E}[|X-X'|^\alpha\,|Y-Y''|^\alpha]\,. \end{aligned} \tag{1.16}$$

This is the distance covariance *introduced by Székely et al. (2007).*

The distance covariance $T(X,Y;\mu)$ introduced in (1.16) has several good properties. It is homogeneous under positive scaling and is also invariant under orthonormal transformations of X and Y. Some of these properties are shared with other distance covariances when μ is infinite. We illustrate this for a Lévy measure μ on $\mathbb{R}^{p+q}$, i.e., it satisfies (1.5) for $\alpha=2$. In particular, μ is finite on sets bounded away from zero. Via the Lévy-Khintchine formula, a Lévy measure μ corresponds to an $\mathbb{R}^{p+q}$-valued infinitely divisible random vector (Z_1,Z_2) (with Z_1 assuming values in $\mathbb{R}^p$ and Z_2 in $\mathbb{R}^q$) and characteristic function

$$\begin{aligned} \varphi_{Z_1,Z_2}(x,y) &= \exp\Big\{-\int_{\mathbb{R}^{p+q}}\Big(e^{\,i\langle s,x\rangle+i\langle t,y\rangle}-1 \\ &\qquad -(i\langle x,s\rangle+i\langle y,t\rangle)\mathbf{1}(|(s,t)|\le 1)\Big)\,\mu(ds,dt)\Big\}\,. \end{aligned} \tag{1.17}$$

Lemma 1.2.6. *Assume that there exists an* $\alpha \in (0,2]$ *such that* $\mathbb{E}[|X|^\alpha] + \mathbb{E}[|Y|^\alpha] < \infty$ *and* μ *is a symmetric Lévy measure corresponding to* (1.17) *such that* (1.5) *holds. Then*

$$\begin{aligned} T(X,Y;\mu) &= Re\,\mathbb{E}\Big[-\log\varphi_{Z_1,Z_2}(X-X',Y-Y') - \log\varphi_{Z_1,Z_2}(X-X',Y''-Y''') \\ &\quad +2\log\varphi_{Z_1,Z_2}(X-X',Y-Y'')\Big]\,. \end{aligned} \tag{1.18}$$

Remark 1.2.7. We observe that (1.18) always vanishes if Z_1 and Z_2 are independent.

Proof. By the symmetry of the random vectors in (1.8) and the measure μ, we have

$$\begin{aligned} &\mathrm{Re}\int_{\mathbb{R}^{p+q}} \mathbb{E}\Big[e^{\,i\langle s,X-X'\rangle+i\langle t,Y-Y'\rangle}-1\Big]\,\mu(ds,dt) \\ &= \mathrm{Re}\int_{\mathbb{R}^{p+q}} \mathbb{E}\Big[e^{\,i\langle s,X-X'\rangle+i\langle t,Y-Y'\rangle}-1 \\ &\qquad -(i\langle s,X-X'\rangle+i\langle t,Y-Y'\rangle)\,\mathbf{1}\big(|(s,t)|\le 1\big)\Big]\,\mu(ds,dt) \\ &= \mathrm{Re}\,\mathbb{E}\big[-\log\varphi_{Z_1,Z_2}(X-X',Y-Y')\big]\,. \end{aligned}$$

The last step is justified if we can interchange the integral and the expected value. Therefore we have to verify that the following integral is finite:

$$\int_{\mathbb{R}^{p+q}} \mathbb{E}\Big[\Big|e^{\,i\langle s,X-X'\rangle+i\langle t,Y-Y'\rangle}-1-(i\langle s,X-X'\rangle+i\langle t,Y-Y'\rangle)\,\mathbf{1}\big(|(s,t)|\le 1\big)\Big|\Big]\,\mu(ds,dt)\,.$$

The integrals over the disjoint sets $\{(s,t): |(s,t)|\le 1\}$ and $\{(s,t):|(s,t)|>1\}$ are denoted by I_1 and I_2, respectively. The quantity I_2 is bounded since the integrand is bounded and μ is finite on sets bounded away from zero. A Taylor expansion shows for $\alpha\in(0,2]$,

$$\begin{aligned} I_1 &\le c\int_{|(s,t)|\le 1}\mathbb{E}\big[2\wedge(|\langle s,X-X'\rangle|+|\langle t,Y-Y'\rangle|)^2\big]\,\mu(ds,dt) \\ &\le c\,(\mathbb{E}|X|^\alpha]+\mathbb{E}|Y|^\alpha])\int_{|(s,t)|\le 1} 1\wedge|(s,t)|^\alpha\,\mu(ds,dt) \end{aligned}$$

and the right-hand side is finite by assumption.

Proceeding in the same way as above for the remaining expressions in (1.8), the lemma is proved. □

Example 1.2.8. *Assume that* μ *is a probability measure of a random vector* (Z_1,Z_2) *in* $\mathbb{R}^{p+q}$ *and that* Z_1 *and* Z_2 *are independent. Then*

$$T(X,Y;\mu) = \mathbb{E}[\varphi_{Z_1}(X-X')\,\varphi_{Z_2}(Y-Y')] + \mathbb{E}[\varphi_{Z_1}(X-X')]\,\mathbb{E}[\varphi_{Z_2}(Y''-Y''')]$$

$$-2\,\mathbb{E}[\varphi_{Z_1}(X-X')\,\varphi_{Z_2}(Y-Y'')]\,.$$

For example, consider independent symmetric Z_1 and Z_2 with multivariate β-stable distributions in $\mathbb{R}^p$ and $\mathbb{R}^q$, respectively, for some $\beta \in (0,2]$. They have joint characteristic function given by $\varphi_{Z_1,Z_2}(x,y) = e^{-(|x|^\beta+|y|^\beta)}$. Therefore

$$\begin{aligned} T(X,Y;\mu) &= \mathbb{E}[e^{-(|X-X'|^\beta+|Y-Y'|^\beta)}] + \mathbb{E}[e^{-|X-X'|^\beta}]\,\mathbb{E}[e^{-|Y-Y'|^\beta}] \\ &\quad -2\,\mathbb{E}[e^{-(|X-X'|^\beta+|Y-Y''|^\beta)}]\,. \end{aligned} \tag{1.19}$$

Example 1.2.9. *Assume that X and Y are integer-valued. Consider the spectral densities w_1 and w_2 on $[-\pi,\pi]$ of two real-valued second-order stationary processes and assume $\mu(s,t) = w_1(s)w_2(t)$. Denote the covariance functions on the integers corresponding to w_1 and w_2 by γ_1 and γ_2, respectively. We have the well-known relation*

$$\int_{-\pi}^{\pi} e^{itk}\,w_i(t)\,dt = \int_{-\pi}^{\pi}\cos(tk)\;w_i(t)\,dt = \gamma_i(k)\,, \qquad k\in\mathbb{Z}\,,$$

where we also exploit the symmetry of the functions w_i. If we restrict integration in (1.8) to $[-\pi,\pi]^2$ we obtain, abusing notation,

$$\begin{aligned} T(X,Y;\mu) &= \mathbb{E}[\gamma_1(X-X')\,\gamma_2(Y-Y')] + \mathbb{E}[\gamma_1(X-X')]\mathbb{E}[\gamma_2(Y-Y')] \\ &\quad -2\,\mathbb{E}[\gamma_1(X-X')\,\gamma_2(Y-Y'')]\,. \end{aligned}$$

The spectral density of a stationary process may have singularities (e.g. for fractional ARMA processes) but this density is integrable on $[-\pi,\pi]$. If w_1, w_2 are positive Lebesgue a.e. on $[0,\pi]$ then $T(X,Y;\mu) = 0$ if and only if X,Y are independent. Indeed, the characteristic function of an integer-valued random variable is periodic with period 2π.

Example 1.2.10. *To illustrate (1.18) we consider a symmetric α-stable vector (Z_1,Z_2) for $\alpha\in(0,2)$ with log-characteristic function*

$$-\log\varphi_{Z_1,Z_2}(x,y) = \int_{\mathbb{S}^{p+q-1}} |\langle s,x\rangle + \langle t,y\rangle|^\alpha\, m(ds,dt)$$

and m is a finite symmetric measure on the unit sphere $\mathbb{S}^{p+q-1}$ of $\mathbb{R}^{p+q}$. Then we have

$$T(X,Y;\mu) = \int_{\mathbb{S}^{p+q-1}} \mathbb{E}\big[|\langle s,X-X'\rangle + \langle t,Y-Y'\rangle|^\alpha + |\langle s,X-X'\rangle + \langle t,Y''-Y'''\rangle|^\alpha$$

$$-2\,|\langle s, X-X'\rangle + \langle t, Y'-Y''\rangle|^{\alpha}\big]\, m(ds,dt)\,.$$

A special case is the sub-Gaussian $\alpha/2$-stable random vectors with characteristic function $-\log\varphi_{Z_1,Z_2}(x,y) = |(x,y)'\Sigma(x,y)|^{\alpha/2}$, where Σ is the covariance matrix of an $\mathbb{R}^{p+q}$-valued random vector and we write (x,y) for the concatanation of any $x\in\mathbb{R}^p$ and $y\in\mathbb{R}^q$. Then

$$\begin{aligned} T(X,Y;\mu) &= \mathbb{E}\big[|(X-X',Y-Y')'\Sigma\,(X-X',Y-Y')|^{\alpha/2} \\ &\quad +[|(X-X',Y''-Y''')'\Sigma\,(X-X',Y''-Y''')|^{\alpha/2} \\ &\quad -2[|(X-X',Y-Y'')'\Sigma\,(X-X',Y-Y'')|^{\alpha/2}\big]\,. \end{aligned}$$

In particular, if Σ is block-diagonal with Σ_1 a $p\times p$ covariance matrix and Σ_2 a $q\times q$ covariance matrix, we have

$$\begin{aligned} T(X,Y;\mu) &= \mathbb{E}\big[|(X-X')'\Sigma_1\,(X-X') + (Y-Y')'\Sigma_2(Y-Y')|^{\alpha/2} \\ &\quad +|(X-X')'\Sigma_1\,(X-X') + (Y''-Y''')'\Sigma_2(Y''-Y''')|^{\alpha/2} \\ &\quad -2|(X-X')'\Sigma_1\,(X-X') + (Y-Y'')'\Sigma_2(Y-Y'')|^{\alpha/2}\big]\,, \end{aligned}$$

and if Σ is the identity matrix,

$$\begin{aligned} T(X,Y;\mu) &= \mathbb{E}\big[\big|\,|X-X'|^2+|Y-Y'|^2\big|^{\alpha/2} + \big|\,|X-X'|^2+|Y''-Y'''|^2\big|^{\alpha/2} \\ &\quad -2\big|\,|X-X'|^2+|Y-Y''|^2\big|^{\alpha/2}\big]\,. \end{aligned} \qquad (1.20)$$

We notice that for these examples, $T(X,Y;\mu)$ is scale homogeneous, i.e., $T(cX,cY;\mu) = |c|^{\alpha}T(X,Y;\mu)$, and (1.20) *is invariant under orthonormal transformations, i.e., $T(RX,SY;\mu) = T(X,Y;\mu)$ for orthonormal matrices R and S, properties also enjoyed by the weight function in Example 1.2.5.*

1.3 The empirical distance covariance function of a stationary sequence

In this section we consider the empirical distance covariance for a stationary time series $((X_t,Y_t))$ with generic element (X,Y) where X and Y assume values in $\mathbb{R}^p$ and $\mathbb{R}^q$, respectively. The empirical distance covariance is given by

$$T_n(X,Y;\mu) \;=\; \int_{\mathbb{R}^{p+q}} \left|\varphi^n_{X,Y}(s,t)-\varphi^n_X(s)\,\varphi^n_Y(t)\right|^2 \mu(ds,dt)\,,$$

where the empirical characteristic function is given by $\varphi^n_{X,Y}(s,t) = \dfrac{1}{n}\sum_{j=1}^n e^{\,i\,\langle s,X_j\rangle + i\,\langle t,Y_j\rangle}$, $n\ge 1$, and $\varphi^n_X(s)=\varphi^n_{X,Y}(s,0)$ and $\varphi^n_Y(s)=\varphi^n_{X,Y}(0,t)$.

1.3.1 Asymptotic results for the empirical distance correlation

Under the conditions of Lemma 1.2.1 that ensure the finiteness of $T(X,Y;\mu)$, we show that T_n is consistent for stationary ergodic time series; see (Samorodnitsky, 2016, Chapter 2) for a definition of ergodicity.

Theorem 1.3.1. *Consider a stationary ergodic time series $((X_j,Y_j))_{j=1,2,\ldots}$ with values in $\mathbb{R}^{p+q}$ and assume one of the three conditions in Lemma 1.2.1 are satisfied. Then*

$$T_n(X,Y;\mu) \stackrel{a.s.}{\to} T(X,Y;\mu), \quad \text{as } n\to\infty.$$

Proof. For $(s,t)\in\mathbb{R}^{p+q}$ the difference between the joint characteristic function with the product characteristic function and the empirical analog are given by

$$C(s,t)=\varphi_{X,Y}(s,t)-\varphi_X(s)\varphi_Y(t) \quad \text{and} \quad C_n(s,t)=\varphi^n_{X,Y}(s,t)-\varphi^n_X(s)\varphi^n_Y(t)\,.$$

Each of the processes $\varphi^n_{X,Y}$, φ^n_X, φ^n_Y is a sample mean of iid bounded continuous processes defined on $\mathbb{R}^{p+q}$. Consider the compact set

$$K_\delta=\{(s,t)\in\mathbb{R}^{p+q}: \delta\le |s|\wedge|t|\,, |s|\vee|t|\le 1/\delta\} \tag{1.21}$$

for small $\delta>0$. By the ergodic theorem on $\mathcal{C}(K_\delta)$, the space of continuous functions on K_δ, $\varphi^n_{X,Y}\stackrel{a.s.}{\to}\varphi_{X,Y}$ as $n\to\infty$; see Krengel (1985). Hence

$$\int_{K_\delta}|C_n(s,t)|^2\,\mu(ds,dt) \stackrel{a.s.}{\to} \int_{K_\delta}|C(s,t)|^2\,\mu(ds,dt)\,, \qquad n\to\infty\,.$$

It remains to show that

$$\lim_{\delta\downarrow 0}\limsup_{n\to\infty}\int_{K^c_\delta}|C_n(s,t)|^2\,\mu(ds,dt)=0 \quad \text{a.s.}$$

If μ is a finite measure we have

$$\lim_{\delta\downarrow 0}\limsup_{n\to\infty}\int_{K_\delta^c}|C_n(s,t)|^2\mu(ds,dt)\le c\,\lim_{\delta\downarrow 0}\mu(K_\delta^c)=0\,.$$

Now assume that μ is infinite on the axes or at zero and (1.4) holds. We apply inequality (1.6) under the assumption that (X,Y) has the empirical probability measure of the sample (X_j,Y_j), $j=1,\dots,n$. Since the empirical measure has all moments finite we obtain from (1.7) that for $\alpha\in(0,2]$,

$$1-|\varphi_X^n(s)|^2\le c\,|s|^\alpha\,\mathbb{E}_{n,X}[|X-X'|^\alpha]=c\,|s|^\alpha\,n^{-2}\sum_{1\le k,l\le n}|X_k-X_l|^\alpha\,,$$

where X,X' are independent and each of them has the empirical distribution of the X-sample. The right-hand side is a U-statistic which converges a.s. to $\mathbb{E}[|X-X'|^\alpha]$ as $n\to\infty$ provided this moment is finite. This follows from the ergodic theorem for U-statistics; see Aaronson et al. (1996). The same argument as for part (2) of Lemma 1.2.1 implies that on K_δ^c,

$$|C_n(s,t)|^2\le c\,\mathbb{E}_{n,X}[|X-X'|^\alpha]\,\mathbb{E}_{n,Y}[|Y-Y'|^\alpha]\,(1\wedge|s|^\alpha)\,(1\wedge|t|^\alpha)\,.$$

By the ergodic theorem,

$$\limsup_{n\to\infty}\int_{K_\delta^c}|C_n(s,t)|^2\mu(ds,dt)\le c\,\mathbb{E}[|X-X'|^\alpha]\,\mathbb{E}[|Y-Y'|^\alpha]\int_{K_\delta^c}(1\wedge|s|^\alpha)(1\wedge|t|^\alpha)\mu(ds,dt)$$

almost surely, and the latter integral converges to zero as $\delta\downarrow 0$ by assumption.

If the measure μ is infinite at zero and (1.5) holds the proof is analogous. □

In order to prove weak convergence of T_n we assume that the sequence $((X_i,Y_i))$ with values in $\mathbb{R}^{p+q}$ is α-mixing with rate function (α_h); see (Doukhan, 1994, p. 18) and (Ibragimov and Linnik, 1971, p. 305) for the definition. We have the following result.

Theorem 1.3.2. *Assume that $((X_j,Y_j))$ is a strictly stationary sequence with values in $\mathbb{R}^{p+q}$ such that $\sum_h \alpha_h^{1/r}<\infty$ for some $r>1$. Set $u=2r/(r-1)$ and write $X=(X^{(1)},\dots,X^{(p)})$ and $Y=(Y^{(1)},\dots,Y^{(q)})$.*

1. *Assume that X_0 and Y_0 are independent and for some $\alpha \in (u/2, u]$, $\epsilon \in [0, 1/2)$ and $\alpha' \leq \min(2, \alpha)$, the following hold:*

$$\mathbb{E}[|X|^\alpha + |Y|^\alpha] < \infty, \qquad \mathbb{E}\Big[\prod_{l=1}^{p} |X^{(l)}|^\alpha\Big] < \infty\,, \quad \mathbb{E}\Big[\prod_{l=1}^{q} |Y^{(l)}|^\alpha\Big] < \infty\,, \tag{1.22}$$

and

$$\int_{\mathbb{R}^{p+q}} (1 \wedge |s|^{\alpha'(1+\epsilon)/u})(1 \wedge |t|^{\alpha'(1+\epsilon)/u})\, \mu(ds, dt) < \infty\,. \tag{1.23}$$

Then

$$n\, T_n(X, Y; \mu) \xrightarrow{d} \|G\|_\mu^2 = \int_{\mathbb{R}^{p+q}} |G(s,t)|^2\, \mu(ds, dt)\,, \tag{1.24}$$

where G is a complex-valued mean-zero Gaussian process whose covariance structure is given in (1.29) with $h = 0$ and depends on the dependence structure of $((X_t, Y_t))$.

2. *Assume that X_0 and Y_0 are dependent and for some $\alpha \in (u/2, u]$, $\epsilon \in [0, 1/2)$ and for $\alpha' \leq \min(2, \alpha)$ the following hold:*

$$\mathbb{E}[|X|^{2\alpha} + |Y|^{2\alpha}] < \infty, \quad \mathbb{E}\Big[(1 \vee \prod_{l=1}^{p} |X^{(l)}|^\alpha)(1 \vee \prod_{k=1}^{q} |Y^{(k)}|^\alpha)\Big] < \infty\,, \tag{1.25}$$

and

$$\int_{\mathbb{R}^{p+q}} (1 \wedge |s|^{\alpha'(1+\epsilon)/u})(1 \wedge |t|^{\alpha'(1+\epsilon)/u})\, \mu(ds, dt) < \infty\,. \tag{1.26}$$

Then

$$\sqrt{n}\,(T_n(X, Y; \mu) - T(X, Y; \mu)) \xrightarrow{d} G'_\mu = \int_{\mathbb{R}^{p+q}} G'(s,t)\, \mu(ds, dt)\,, \tag{1.27}$$

where $G'(s,t) = 2Re\{G(s,t)C(s,t)\}$ is a mean-zero Gaussian process.

The proof of Theorem 1.3.2 is given in Section 1.6.

Remark 1.3.3. We notice that (1.23) and (1.26) are always satisfied if μ is a finite measure.

Remark 1.3.4. If (X_i) and (Y_i) are two independent iid sequences then the statement of Theorem 1.3.2(1) remains valid if for some $\alpha \in (0, 2]$, $\mathbb{E}[|X|^\alpha] + \mathbb{E}[|Y|^\alpha] < \infty$ and

$$\int_{\mathbb{R}^{p+q}} (1 \wedge |s|^\alpha)(1 \wedge |t|^\alpha)\, \mu(ds, dt) < \infty\,. \tag{1.28}$$

Remark 1.3.5. The distribution of the limit variable in (1.24) is generally not tractable. Therefore one must use numerical or resampling methods for determining the quantiles of $nT_n(X,Y;\mu)$. On the other hand, the limit distribution in (1.27) is normally distributed with mean 0 and variance σ_μ^2 that can be easily calculated from the covariance function of $G(s,t)$ and $C(s,t)$. Notice that if $C(s,t)=0$, the limit random variable in (1.27) is 0 and part (1) of the theorem applies. Again resampling or subsampling methods must be employed to determine quantiles of nT_n.

1.3.2 Testing serial dependence for multivariate time series

Define the *cross-distance covariance function* (CDCVF) of a strictly stationary sequence $((X_t,Y_t))$ by

$$T_\mu^{X,Y}(h)=T(X_0,Y_h;\mu)\,,\qquad h\in\mathbb{Z}\,,$$

and the *auto-distance covariance function* (ADCVF) of a stationary sequence (X_t) by

$$T_\mu^{X}(h)=T_\mu^{X,X}(h)\,,\qquad h\in\mathbb{Z}\,.$$

Here and in what follows, we assume that $\mu=\mu_1\times\mu_2$ for suitable measures μ_1 on $\mathbb{R}^p$ and μ_2 on $\mathbb{R}^q$. In the case of an ADCVF we also assume $\mu_1=\mu_2$. The empirical versions $T_{n,\mu}^X$ and $T_{n,\mu}^{X,Y}$ are defined correspondingly. For example, for integer $h\ge 0$, one needs to replace $\varphi_{X,Y}^n(s,t)$ in the definition of $T_n(X,Y;\mu)$ by

$$\varphi_{X_0,Y_h}^n(s,t)=\frac{1}{n}\sum_{j=1}^{n-h}e^{\,i\langle s,X_j\rangle+i\langle t,Y_{j+h}\rangle},\qquad s\in\mathbb{R}^p\,,t\in\mathbb{R}^q\,,\quad n\ge h+1\,,$$

with the corresponding modifications for the marginal empirical characteristic functions. For finite h, the change from the upper summation limit n to $n-h$ has no influence on the asymptotic theory.

We also introduce the corresponding *cross-distance correlation function* (CDCF) and *auto-distance correlation function* (ADCF) respectively:

$$R_\mu^{X,Y}(h)=\frac{T_\mu^{X,Y}(h)}{\sqrt{T_\mu^X(0)\,T_\mu^Y(0)}}\qquad \textit{and}\qquad R_\mu^X(h)=\frac{T_\mu^X(h)}{T_\mu^X(0)}\,,\qquad h\in\mathbb{Z}\,.$$

The quantities $R_\mu^{X,Y}(h)$ assume values in $[0,1]$, with the two endpoints representing independence and complete dependence. The empirical CDCF $R_{n,\mu}^{X,Y}$ and ADCF $R_{n,\mu}^{X}$ are defined by replacing the distance covariances $T_\mu^{X,Y}(h)$ by the corresponding empirical versions $T_{n,\mu}^{X,Y}(h)$.

The empirical ADCV was examined in Zhou (2012) and Fokianos and Pitsillou (2017) as an alternative tool for testing serial dependence, in the way that it also captures non-linear dependence. They always choose the measure $\mu = \mu_1 \times \mu_1$ with density (1.14).

In contrast to the autocorrelation and cross-correlation functions of standard stationary time series models (such as ARMA, GARCH) it is in general complicated (or impossible) to provide explicit (and tractable) expressions for $T_\mu^X(h)$ and $T_\mu^{X,Y}(h)$ or even to say anything about the rate of decay of these quantities when $h \to \infty$. However, in view of (1.13) we observe that

$$\begin{aligned} T_\mu^X(h) &= \operatorname{cov}\big(\tilde\mu_1(X_0 - X_0'), \tilde\mu_1(X_h - X_h')\big) \\ &\quad -2\operatorname{cov}\big(\mathbb{E}[\tilde\mu_1(X_0 - X_0') \mid X_0]\,, \mathbb{E}[\tilde\mu_1(X_h - X_0') \mid X_h]\big)\,. \end{aligned}$$

While this is not the autocovariance function of a stationary process, it is possible to bound each of the terms in case (X_t) is α-mixing with rate function (α_h). In this case, one may use bounds for the autocovariance functions of the stationary series $(\tilde\mu_1(X_t - X_t'))$ and $(\mathbb{E}[\tilde\mu_1(X_t - X_0') \mid X_t])$ which inherit α-mixing from (X_t) with the same rate function. For example, a standard inequality (Doukhan (1994), Section 1.2.2, Theorem 3(a)) yields that $T_\mu^X(h) \le c\,\alpha_h^{1/r}\,\big(\mathbb{E}[(\tilde\mu_1(X_0 - X_0'))^u]\big)^{2/u}$ for positive c and $r > 0$ such that $r^{-1} + 2u^{-1} = 1$. If $\tilde\mu_1$ is bounded we also have $T_\mu^X(h) \le c\,\alpha_h$ for some positive constant. Similar bounds can be found for $T_\mu^{X,Y}(h)$ provided $((X_t, Y_t))$ is α-mixing.

Next we give an example where the ADCVF can be calculated explicitly.

Example 1.3.6. *Consider a univariate strictly stationary Gaussian time series* (X_t) *with mean zero, variance* σ^2 *and autocovariance function* γ_X. *We choose a Gaussian probability measure* μ *which leads to the relation* (1.19). *Choose* N_1, N_2, N_3 *iid* $N(0,2)$*-distributed independent of the independent quantities* $(X_0, X_h), (X_0', X_h'), X_h''$. *Then for* $h \ge 0$,

$$T_\mu^X(h) = \mathbb{E}\big[e^{\,i N_1(X_0 - X_0') + iN_2(X_h - X_h')}\big] + \big(\mathbb{E}\big[e^{\,i N_1(X_0 - X_0')}\big]\big)^2 - 2\,\mathbb{E}\big[e^{\,i N_1(X_0 - X_0') + iN_2(X_h - X_h'')}\big]$$

$$
\begin{aligned}
&= \mathbb{E}\big[e^{\,i\,(N_1X_0+N_2X_h)-i(N_1X_0'+N_2X_h')}\big] + \big(\mathbb{E}\big[e^{\,i\,N_1(X_0-X_0')}\big]\big)^2 \\
&\quad -2\,\mathbb{E}\big[e^{\,i\,(N_1X_0+N_2X_h)-i(N_1X_0'+N_2X_h'')}\big] \\
&= \mathbb{E}\big[e^{\,i\,N_3\left(N_1^2\sigma^2+N_2^2\sigma^2+2\gamma_X(h)N_1N_2\right)^{1/2}}\big] + \big(\mathbb{E}\big[e^{\,i\,N_3(N_1^2\sigma^2)^{1/2}}\big]\big)^2 \\
&\quad -2\,\mathbb{E}\big[e^{\,i\,N_3\,\left(N_1^2\sigma^2+N_2^2\sigma^2+\gamma_X(h)N_1N_2\right)^{1/2}}\big] \\
&= \mathbb{E}\big[e^{-\left(N_1^2\sigma^2+N_2^2\sigma^2+2\gamma_X(h)N_1N_2\right)}\big] + \big(\mathbb{E}\big[e^{-N_1^2\sigma^2}\big]\big)^2 - 2\,\mathbb{E}\big[e^{-\left(N_1^2\sigma^2+N_2^2\sigma^2+\gamma_X(h)N_1N_2\right)}\big]\,.
\end{aligned}
$$

For the evaluation of this expression we focus on the first term, the other cases being similar. Observing that $\sigma^2 \pm \gamma_X(h)$ are the eigenvalues of the covariance matrix

$$
\begin{pmatrix} \sigma^2 & \gamma_X(h) \\ \gamma_X(h) & \sigma^2 \end{pmatrix},
$$

calculation shows that

$$
N_1^2\sigma^2 + N_2^2\sigma^2 + 2\gamma_X(h)N_1N_2 \stackrel{d}{=} N_1^2(\sigma^2-\gamma_X(h)) + N_2^2(\sigma^2+\gamma_X(h))\,.
$$

Now the moment generating function of a χ^2-distributed random variable yields

$$
\mathbb{E}\big[e^{-\left(N_1^2\sigma^2+N_2^2\sigma^2+2\gamma_X(h)N_1N_2\right)}\big] = \big(1+4(\sigma^2-\gamma_X(h))\big)^{-1/2}\,\big(1+4(\sigma^2+\gamma_X(h))\big)^{-1/2}\,.
$$

Proceeding in a similar fashion, we obtain

$$
\begin{aligned}
T_\mu^X(h) \;&=\; \big(1+4(\sigma^2-\gamma_X(h))\big)^{-1/2}\,\big(1+4(\sigma^2+\gamma_X(h))\big)^{-1/2} + (1+4\sigma^2)^{-1} \\
&\quad -2\,\big(1+4(\sigma^2-\gamma_X(h)/2)\big)^{-1/2}\,\big(1+4(\sigma^2+\gamma_X(h)/2)\big)^{-1/2}\,.
\end{aligned}
$$

If $\gamma_X(h) \to 0$ as $h \to \infty$ Taylor expansions yield $T_\mu^X(h) \sim 4\gamma_X^2(h)/(1+4\sigma^2)^3$. A similar result was given in Fokianos and Pitsillou (2017), where they derived an explicit expression for $T_\mu^X(h)$ for a stationary Gaussian process (X_t) with weight function (1.2).

If $((X_t, Y_t))$ is strictly stationary and ergodic then $((X_t, Y_{t+h}))$ is a strictly stationary ergodic sequence for every integer h. Then Theorem 1.3.1 applies.

Corollary 1.3.7. *Under the conditions of Theorem 1.3.1, for $h \geq 0$,*

$$
T_{n,\mu}^{X,Y}(h) \stackrel{a.s.}{\to} T_\mu^{X,Y}(h) \qquad \textit{and} \qquad T_{n,\mu}^{X}(h) \stackrel{a.s.}{\to} T_\mu^{X}(h)\,,
$$

and

$$R_{n,\mu}^{X,Y}(h) \stackrel{a.s.}{\rightarrow} R_{\mu}^{X,Y}(h) \qquad \textit{and} \qquad R_{n,\mu}^{X}(h) \stackrel{a.s.}{\rightarrow} R_{\mu}^{X}(h)\,.$$

Applying Theorem 1.3.2 and Theorem 1.3.1, we also have the following weak dependence result under α-mixing. Zhou (2012) proved the corresponding result under conditions on the so-called *physical dependence measure.*

Corollary 1.3.8. *Assume that X_0 and Y_h are independent for some $h \geq 0$ and the sequence $((X_t, Y_t))$ satisfies the conditions of Theorem 1.3.2. Then*

$$n\,T_{n,\mu}^{X,Y}(h) \stackrel{d}{\rightarrow} \|G_h\|_\mu^2 \qquad \textit{and} \qquad n\,R_{n,\mu}^{X,Y}(h) \stackrel{d}{\rightarrow} \frac{\|G_h\|_\mu^2}{\sqrt{T_\mu^X(0)\,T_\mu^Y(0)}}\,,$$

where G_h is a centered Gaussian process on $\mathbb{R}^{p+q}$.

Remark 1.3.9. From the proof of Theorem 1.3.2 (the central limit theorem for the multivariate empirical characteristic function) it follows that G_h has covariance function

$$\begin{aligned}\Gamma((s,t),(s',t')) &= \operatorname{cov}(G_h(s,t), G_h(s',t')) \\ &= \sum_{j\in\mathbb{Z}} \mathbb{E}\big[\big(e^{\,i\langle s, X_0\rangle} - \varphi_X(s)\big)\big(e^{\,i\langle t, Y_h\rangle} - \varphi_Y(t)\big) \\ &\qquad \times\big(e^{\,-i\langle s', X_j\rangle} - \varphi_X(-s')\big)\big(e^{\,-i\langle t', Y_{j+h}\rangle} - \varphi_Y(-t')\big)\big]\,. \qquad (1.29)\end{aligned}$$

In the special case when (X_t) and (Y_t) are independent sequences G_h is the same across all h with covariance function

$$\Gamma((s,t),(s',t')) = \big(\varphi_X(s-s') - \varphi_X(s)\varphi_X(s')\big)\big(\varphi_Y(t-t') - \varphi_Y(t)\varphi_Y(t')\big)\,.$$

Since G_h is centered Gaussian its squared L^2-norm $\|G_h\|_\mu^2$ has a weighted χ^2-distribution; see Kuo (1975), Chapter 1. The distribution of $\|G_h\|_\mu^2$ is not tractable and therefore one needs resampling methods for determining its quantiles.

Remark 1.3.10. Corollary 1.3.8 can be extended to the joint convergence of the function $n\,T_{n,\mu}^{X,Y}(h)$ at finitely many lags h, provided X_0 and Y_h are independent for these lags.

Remark 1.3.11. Corollary 1.3.8 does not apply when X_0 and Y_h are dependent. Then $n\,T_{n,\mu}^{X,Y}(h) \rightarrow \infty$ a.s. and $n\,R_{n,\mu}^{X,Y}(h) \rightarrow \infty$ a.s.

1.4 Auto-distance covariance of fitted residuals from AR(p) process

An often important problem in time series is to assess the goodness-of-fit of a particular model. As an illustration, consider a causal autoregressive process of order p (AR(p)) given by the difference equations,

$$X_t = \sum_{k=1}^{p} \phi_k X_{t-k} + Z_t\,, \quad t = 0, \pm 1, \ldots,$$

where (Z_t) is an iid sequence with a finite moment $\mathbb{E}|Z|^\kappa < \infty$ for some $\kappa > 0$. It is further assumed Z_t has mean 0 if $\kappa \geq 1$. It is often convenient to write the AR(p) process in the form, $Z_t = X_t - \boldsymbol{\phi}^T \mathbf{X}_{t-1}$, where $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_p)^T$, $p \geq 1$ and $\mathbf{X}_t = (X_t, \ldots, X_{t-p+1})^T$. Since the process is assumed causal, we can write $X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}$ for absolutely summable constants (ψ_j); see Brockwell and Davis (1991), p. 85. For convenience, we also write $\psi_j = 0$ for $j < 0$ and $\psi_0 = 1$.

The least-squares estimator $\widehat{\boldsymbol{\phi}}$ of $\boldsymbol{\phi}$ satisfies the relation

$$\widehat{\boldsymbol{\phi}} - \boldsymbol{\phi} = \Gamma_{n,p}^{-1} \frac{1}{n} \sum_{t=p+1}^{n} \mathbf{X}_{t-1} Z_t\,, \quad \text{where} \quad \Gamma_{n,p} = \frac{1}{n} \sum_{t=p+1}^{n} \mathbf{X}_{t-1}^T \mathbf{X}_{t-1}\,.$$

If $\sigma^2 = \mathrm{var}(Z_t) < \infty$, we have by the ergodic theorem,

$$\Gamma_{n,p} \stackrel{a.s.}{\to} \Gamma_p = \big(\gamma_X(j-k)\big)_{1 \leq j,k \leq p}\,, \qquad \text{where} \quad \gamma_X(h) = \mathrm{cov}(X_0, X_h)\,, h \in \mathbb{Z}\,. \tag{1.30}$$

Causality of the process implies that the partial sum $\sum_{t=p+1}^{n} \mathbf{X}_{t-1} Z_t$ is a martingale and applying the martingale central limit theorem yields

$$\sqrt{n}\,(\widehat{\boldsymbol{\phi}} - \boldsymbol{\phi}) \stackrel{d}{\to} \mathbf{Q}\,, \tag{1.31}$$

where $\mathbf{Q}$ is $N(\mathbf{0}, \sigma^2 \Gamma_p^{-1})$ distributed.

The residuals of the fitted model are given by

$$\widehat{Z}_t = X_t - \widehat{\boldsymbol{\phi}}^T \mathbf{X}_{t-1} = \big(\boldsymbol{\phi} - \widehat{\boldsymbol{\phi}}\big)^T \mathbf{X}_{t-1} + Z_t\,, \qquad t = p+1, \ldots, n\,. \tag{1.32}$$

For convenience, we set $\widehat{Z}_t = 0$, $t = 1, \ldots, p$ since this choice does not influence the asymptotic theory. Each of the residuals $\widehat{Z}_t$ depends on the estimated parameters and hence the residual process exhibits serial dependence. Nevertheless, we might expect the test statistic based on the distance covariance function of the residuals, given by

$$T_{n,\mu}^{\widehat{Z}}(h) = \int_{\mathbb{R}} |C_n^{\widehat{Z}}(s,t)|^2 \, \mu(ds, dt),$$

to behave in a similar fashion for the true noise sequence (Z_t). If the model is a good fit, then we would not expect $T_{n,\mu}^{\widehat{Z}}(h)$ to be extraordinarily large. As observed by Rémillard (2009), the limit distributions for $T_{n,\mu}^{\widehat{Z}}(h)$ and $T_{n,\mu}^{Z}(h)$ are not the same. As might be expected, the residuals, which are fitted to the actual data, tend to have smaller distance covariance than the true noise terms for lags less than p, if the model is correct. As a result, one can fashion a goodness-of-fit test based on applying the distance covariance statistics to the residuals. In the following theorem, we show that the distance covariance based on the residuals has a different limit than the distance covariance based on the actual noise, if the process has a finite variance. So in applying a goodness-of-fit test, one must make an adjustment to the limit distribution. Interestingly, if the noise has heavy-tails, the limits based on the residuals and the noise terms are the same and no adjustment is necessary.

For the formulation of the next result we need some auxiliary limit theory; the proofs are given in Section 1.7.

Lemma 1.4.1. *Consider an iid sequence (Z_t) with finite variance. Let*

$$C_n^Z(s,t) = \varphi_{Z_0,Z_h}^n(s,t) - \varphi_Z^n(s)\varphi_Z^n(t)\,.$$

1. *For every $h \geq 0$,*

$$\sqrt{n}\left(C_n^Z, \widehat{\boldsymbol{\phi}} - \boldsymbol{\phi}\right) \xrightarrow{d} (G_h, \mathbf{Q})\,,$$

where the convergence is in $\mathcal{C}(K) \times \mathbb{R}^p$, $K \subset \mathbb{R}^2$ is a compact set, G_h is the limit process of C_n^Z with covariance structure specified in Remark 1.3.9 for the sequence $((Z_t, Z_{t+h}))$, $\mathbf{Q}$ is the limit in (4.2), $(G_h, \mathbf{Q})$ are mean-zero and jointly Gaussian with covariance *matrix*

$$cov(G_h(s,t),\mathbf{Q}) = -\varphi_Z'(s)\,\varphi_Z'(t)\,\Gamma_p^{-1}\Psi_h\,, \qquad s,t\in\mathbb{R}\,, \tag{1.33}$$

where $\Psi_h = (\psi_{h-j})_{j=1,\dots,p}$ *and* φ_Z' *is the first derivative of* φ_Z.

2. *For every* $h \geq 0$,

$$\sqrt{n}\left(C_n^Z, C_n^{\widehat{Z}} - C_n^Z\right) \xrightarrow{d} (G_h,\xi_h)\,,$$

where $(G_h,\mathbf{Q})$ *are specified in* (4.4) *and*

$$\xi_h(s,t) = t\varphi_Z(t)\,\varphi_Z'(s)\Psi_h^T\mathbf{Q}, \qquad (s,t)\in K\,, \tag{1.34}$$

the convergence is in $\mathcal{C}(K,\mathbb{R}^2)$, $K\subset\mathbb{R}^2$ *is a compact set. In particular, we have*

$$\sqrt{n}\,C_n^{\widehat{Z}} \xrightarrow{d} G_h + \xi_h\,, \tag{1.35}$$

in $\mathcal{C}(K)$ *for* $K\subset\mathbb{R}^2$ *compact.*

Now we can formulate the following result; the proof is given in the Section **??**.

Theorem 1.4.2. *Consider a causal AR(p) process with iid noise* (Z_t). *Assume* μ *satisfies*

$$\int_{\mathbb{R}^2}\left[(1\wedge|s|^2)\,(1\wedge|t|^2)\,\mu(ds,dt) + (s^2+t^2)\,\mathbf{1}(|s|\wedge|t|>1)\mu(ds,dt) < \infty. \tag{1.36}$$

1. *If* $\sigma^2 = Var(Z) < \infty$, *then*

$$n\,T_{n,\mu}^{\widehat{Z}}(h) \xrightarrow{d} \|G_h+\xi_h\|_\mu^2 \qquad \text{and} \qquad n\,R_{n,\mu}^{\widehat{Z}}(h) \xrightarrow{d} \frac{\|G_h+\xi_h\|_\mu^2}{T_\mu^Z(0)}\,, \tag{1.37}$$

where (G_h,ξ_h) *are jointly Gaussian limit random fields on* $\mathbb{R}^2$. *The covariance structure of* G_h *is specified in Remark 1.3.9 for the sequence* $((Z_t,Z_{t+h}))$, ξ_h *and the joint limit structure of* (G_h,ξ_h) *are given in Lemma 1.4.1.*

2. *Assume that* Z *is in the domain of attraction of a stable law of index* $\alpha\in(0,2)$, *i.e.,* $\mathbb{P}(|Z|>x) = x^{-\alpha}L(x)$ *for* $x>0$, $L(\cdot)$ *is a slowly varying function at* ∞, *and*

$$\frac{\mathbb{P}(Z>x)}{\mathbb{P}(|Z|>x)} \to p \quad \text{and} \quad \frac{\mathbb{P}(Z<-x)}{\mathbb{P}(|Z|>x)} \to 1-p$$

as $x \to \infty$ for some $p \in [0,1]$ (Feller (1971), p. 313). Then we have

$$n\, T_{n,\mu}^{\widehat{Z}}(h) \xrightarrow{d} \|G_h\|_\mu^2 \qquad \text{and} \qquad n\, R_{n,\mu}^{\widehat{Z}}(h) \xrightarrow{d} \frac{\|G_h\|_\mu^2}{T_\mu^Z(0)}\,, \tag{1.38}$$

where G_h is a Gaussian limit random field on $\mathbb{R}^2$. The covariance structure of G_h is specified in Remark 1.3.9 for the sequence $((Z_t, Z_{t+h}))$.

Remark 1.4.3. Rémillard (2009) mentioned that $T_{n,\mu}^{Z}(h)$ and $T_{n,\mu}^{\widehat{Z}}(h)$ for an AR(1) process have distinct limit processes and he also suggested the limiting structure in (1.37).

Remark 1.4.4. The limit in (1.37) can be extended to cover ARMA processes and some non-linear processes that are invertible. This is the subject of Chapter 2.

The structure of the limit process in (1.37) is rather implicit. In applications, one needs to rely on resampling methods. Relation (1.37) can be extended to a joint convergence result for finitely many lags h but the dependence structure of the limiting vectors is even more involved. Condition (1.36) holds for probability measures $\mu = \mu_1 \times \mu_1$ on $\mathbb{R}^2$ with finite second moment but it does not hold for the benchmark measure $\mu = \mu_1 \times \mu_1$ described in (1.14). A reason for this is that $\|\xi_h\|_\mu^2$ is in general not well defined in this case. If Z_t has characteristic function φ_Z then by virtue of (4.5), $\|\xi_h\|_\mu^2$ is finite a.s. if and only if

$$\int_{-\infty}^{\infty} |t\varphi_Z(t)|^2 \mu_1(dt) \int_{-\infty}^{\infty} |\varphi_Z'(s)|^2 \mu_1(ds) < \infty\,.$$

Now assume that Z_t has a density function f and choose $\mu_1(dt) = c_1 t^{-2} dt$. Then by Plancherel's identity, the first integral becomes

$$\int_{-\infty}^{\infty} |\varphi_Z(t)|^2\, dt = c \int_{-\infty}^{\infty} f^2(t)\, dt\,.$$

If one chooses f to be a symmetric gamma distribution with shape parameter $\delta \in (0, 1/2)$, i.e., $f(z) = .5\beta^\delta |z|^{\delta-1} e^{-|z|\beta}/\Gamma(\delta)$, then the integral $\int_{-\infty}^{\infty} f^2(t)dt$ is infinity and hence the limit random variable in (1.37) cannot be finite.

AR simulation. We illustrate the results of Theorem 1.4.2. First, we generate independent replications of a time series $(X_t)_{t=1,\ldots,1000}$ from a causal AR(10) model with $Z_t \sim N(0,1)$

and

$$\boldsymbol{\phi} = (-0.140, 0.038, 0.304, 0.078, 0.069, 0.013, 0.019, 0.039, 0.148, -0.062).$$

In this and the following examples, we choose the weight measure $\mu = \mu_1 \times \mu_2$, where μ_i is the $N(0, 0.5)$-distribution and hence (1.36) is satisfied. From the independent replications of the simulated residuals we approximate the limit distribution $\|G_h + \xi_h\|_\mu^2 \,/\, T_\mu^Z(0)$ of $n\, R_{n,\mu}^{\widehat{Z}}(h)$ by the corresponding empirical distribution.

The left graph in Figure 1.1 shows the box-plots for $n\, R_{n,\mu}^{\widehat{Z}}(h)$ based on 1000 replications from the AR(10) model, each with sample size $n = 1000$. As seen from the plots, the distribution at each lag is heavily skewed. In the right panel of Figure 1.1, we compare the empirical 5%, 50%, 95% quantiles of $n\, R_{n,\mu}^{\widehat{Z}}(h)$ to those of $n\, R_{n,\mu}^{Z}(h)$, the scaled ADCF of iid noise, all of which have the same limit, $\|G_h\|_\mu^2 \,/\, T_\mu^Z(0)$. The asymptotic variance of the ADCF of the residuals is smaller than that of iid noise at initial lags, and gradually increases at larger lags to the values in the iid case. This behavior is similar to that of the ACF of the residuals of an AR process; see for example Chapter 9.4 of Brockwell and Davis (1991).

Theorem 1.4.2 provides a visual tool for testing the goodness-of-fit of an AR(p) model, by examining the serial dependence of the residuals after model fitting. Under the null hypothesis, we expect $n\, R_{n,\mu}^{\widehat{Z}}(h)$ to be well bounded by the 95% quantiles of the limit distribution $\|G_h + \xi_h\|_\mu^2 \,/\, T_\mu^Z(0)$. For a single time series, this quantity can be approximated using a parametric bootstrap (generating an AR(10) process from the estimated parameters and residuals); see for example Politis et al. (1999). In the right graph of Figure 1.1 we overlay the empirical 5%, 50%, 95% quantiles of $n\, R_{n,\mu}^{\widehat{Z}}(h)$ estimated from one particular realization of the time series. As can be seen in the graph, the parametric bootstrap provides a good approximation to the actual quantiles found via simulation. On the other hand, the quantiles found by simply bootstrapping the residuals provides a rather poor approximation, at least for the first 10 lags.

We now consider the same AR(10) model as before, but with noise having a t-distribution with 1.5 degrees of freedom. (Here the noise is in the domain of attraction of a stable distri-

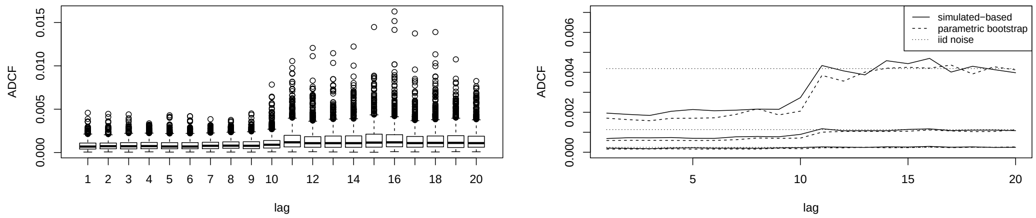


Figure 1.1: Distribution of $n\,R^{\widehat{Z}}_{n,\mu}(h)$, $n = 1000$ for the residuals of an AR(10) process with $N(0,1)$ innovations. Left: Box-plots from 1000 independent replications. Right: 5%, 50%, 95% empirical quantiles of $n\,R^{\widehat{Z}}_{n,\mu}(h)$ based on simulated residuals, on resampled residuals and on iid noise, respectively. The weight measure is $\mu = \mu_1 \times \mu_2$, with each $\mu_i \sim N(0, 0.5)$.

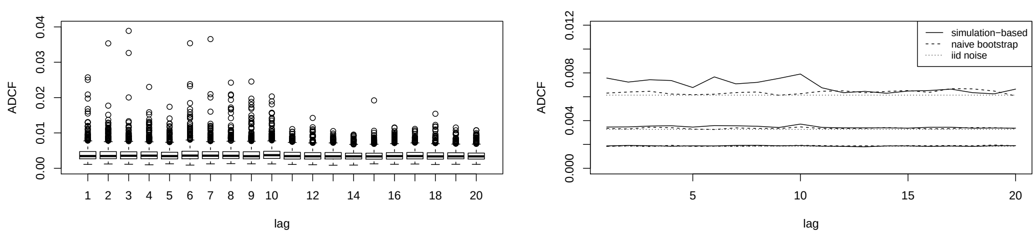


Figure 1.2: Distribution of $n\,R^{\widehat{Z}}_{n,\mu}(h)$ for residuals of AR process with $t_{1.5}$ innovations. Left: lag-wise box-plots. Right panel: empirical 5%, 50%, 95% quantiles from simulated residuals, empirical quantiles from resampled residuals, and empirical quantiles from iid noise. The weight measure is $\mu = \mu_1 \times \mu_2$, with each $\mu_i \sim N(0, 0.5)$.

bution with index 1.5.) The left graph of Figure 1.2 shows the box-plots of $n\ R^{\widehat{Z}}_{n,\mu}(h)$ based on 1000 replications, and the right graph shows the 5%, 50%, 95% quantiles of $n\,R^{\widehat{Z}}_{n,\mu}(h)$ and $n\,R^{Z}_{n,\mu}(h)$, both of which have the same limit distribution $\|G_h\|^2_\mu \,/\, T^Z_\mu(0)$. In this case, the quantiles of $\|G_h\|^2_\mu \,/\, T^Z_\mu(0)$ can be approximated naively by bootstrapping the fitted residuals $(\widehat{Z}_t)$ of the AR model. The left graph of Figure 1.2 overlays the 5%, 50%, 95% quantiles from bootstrapping with those from the simulations. The agreement is reasonably good.

We next provide an empirical example illustrating the limitation of using the measure in (1.14). Again, we use the same AR(10) model as before, but with noise now generated from the symmetric gamma distribution with $\delta = .2, \beta = .5$. The corresponding pair of graphs with boxplots and quantiles for $n\,R^{\widehat{Z}}_{n,\mu}(h)$ is displayed in Figure 1.3. Notice now that the box plots for the sampling distribution of the distance correlation for the first 10 lags are rather spread out compared to those at lags greater than 10. In particular, the sampling behavior of these distance correlations is directly opposite of what we observed in Figure 1.1 where a finite measure was used. To further illustrate this disparity, the plot on the right in Figure 1.3 displays the $95\%, 50\%, 5\%$ quantiles for the companion box plots (the dotted lines are the corresponding quantiles for iid noise with the Gamma(0.2,0.4) distribution). Now, compared to quantiles of distance correlation based on the iid noise, we see a stark difference. The median for the estimates based on the residuals using the weight function in (1.14) is nearly the same as the 95% quantile for the noise at lags 1-10. This illustrates the problem with using (1.14) as a weight function applied to the residuals.

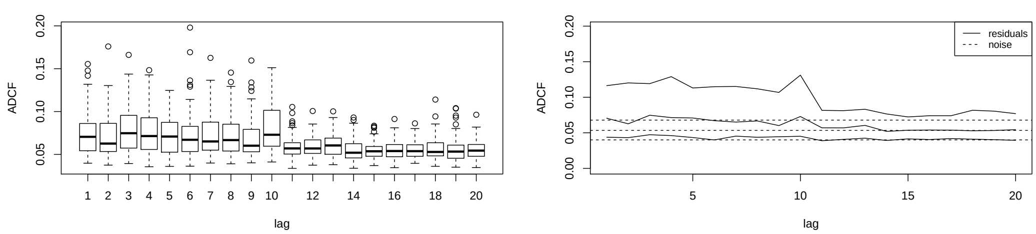


Figure 1.3: Distribution of $n\, R^{\widehat{Z}}_{n,\mu}(h)$, $n = 1000$ for residuals of AR process with a symmetric Gamma(0.2,0.5) noise. Left: box-plots from 500 independent replications. Right panel: empirical 5%, 50%, 95% quantiles from simulated residuals and from iid noise. The measure μ is given by (1.14).

1.5 Data Examples

1.5.1 Amazon daily returns

In this example, we consider the daily stock returns of Amazon from 05/16/1997 to 06/16/2004. Denoting the series by (X_t), Figure 1.4 shows the ACF of (X_t), (X_t^2), $(|X_t|)$ and ADCF of (X_t) with weight measure $\mu(ds, dt) = s^{-2}t^{-2}dsdt$. In the right panel, we compare the ADCF with the 5%, 50%, 95% confidence bounds of the ADCF for iid data, approximated by the corresponding empirical quantiles from 1000 random permutations. With most financial time series, which are typically uncorrelated, serial dependence can be detected by examining the ACF of the absolute values and squares. Interestingly for the Amazon data, the ACF of the squared data also fails to pick up any signal. On the other hand, the ADCF has no trouble detecting serial dependence without having to resort to applying any transformation.

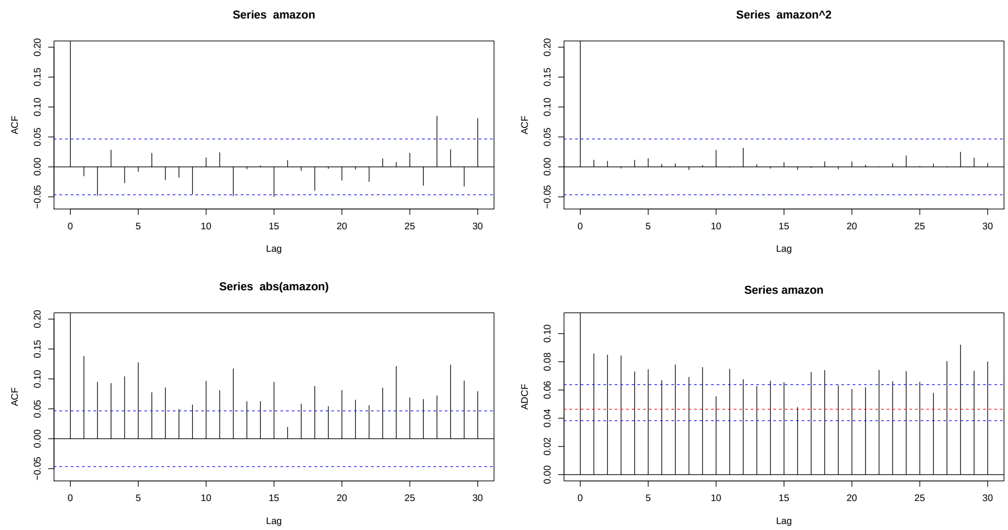


Figure 1.4: ACF and ADCF of daily stock returns of Amazon (X_t) from 05/16/1997 to 06/16/2004. Upper left: ACF of (X_t); Upper right: ACF of (X_t^2); Lower left: ACF of $(|X_t|)$; Lower right: ADCF of (X_t), the 5%, 50%, 95% confidence bounds of ADCF from randomly permuting the data.

1.5.2 Wind speed data

For the next example we consider the daily averages of wind speeds at Kilkenny's synoptic meteorological station in Ireland. The time series consists of 6226 observations from 1/1/1961 to 1/17/1978, after which a square root transformation has been applied to stabilize the variance. This transformation has also been suggested in previous studies (see, for example, Haslett and Raftery (1989)). The ACF of the data, displayed in Figure 1.5, suggests a possible AR model for the data. An AR(9) model was found to provide the best fit (in terms of minimizing AICC among all AR models) to the data. The ACF of the residuals (see upper right panel in Figure 1.5) shows that the serial correlation has been successfully removed. The ACF of the squared residuals and ADCF of the residuals are also plotted in the bottom panels Figure 1.5. For computation of the ADCF, we used the N(0,.5) distribution for the weight measure, which satisfies the condition (1.36). The ADCF of the residuals is well bounded by the confidence bounds for the ADCF of iid noise, shown by the dotted line in the plot. Without adjusting these bounds for the residuals, one would be tempted to conclude that the AR model is a good fit. However, the adjusted bounds for the ADCF of residuals, represented by the solid line in the plot and computed using a parametric bootstrap, suggest that some ADCF values among the first 8 lags are in fact larger than expected. Hence this sheds some doubt on the validity of an AR(9) model with iid noise for this data. A similar conclusion can be reached by inspecting the ACF of the squares of the residuals (see lower left panel in Figure 1.5).

One potential remedy for the lack of fit of the AR(9) model, is to consider a GARCH(1,1) model applied to the residuals. The GARCH model performs well in devolatilizing the AR-fitted residuals and no trace of a signal could be detected through the ACF of the GARCH-residuals applied to the squares and absolute values. The ADCF of the devolatilized residuals, seen in Figure 1.6, still presents some evidence of dependence. Here the confidence bounds plotted are for iid observations, obtained from 1000 random permutations of the GARCH-residuals and as such do not include an adjustment factor. Ultimately, a periodic AR model, which allows for periodicity in both the AR parameters and white noise variance might be a more desirable model.

1.6 Proof of Theorem 1.3.2

The proof follows from the following lemma.

Lemma 1.6.1. *Assume that* $\sum_h \alpha_h^{1/r} < \infty$ *for some* $r > 1$ *and set* $u = 2r/(r-1)$. *We also assume the moment conditions* (1.22) *(or* (1.25)*) for some* $\alpha > 0$ *if* X_0 *and* Y_0 *are*

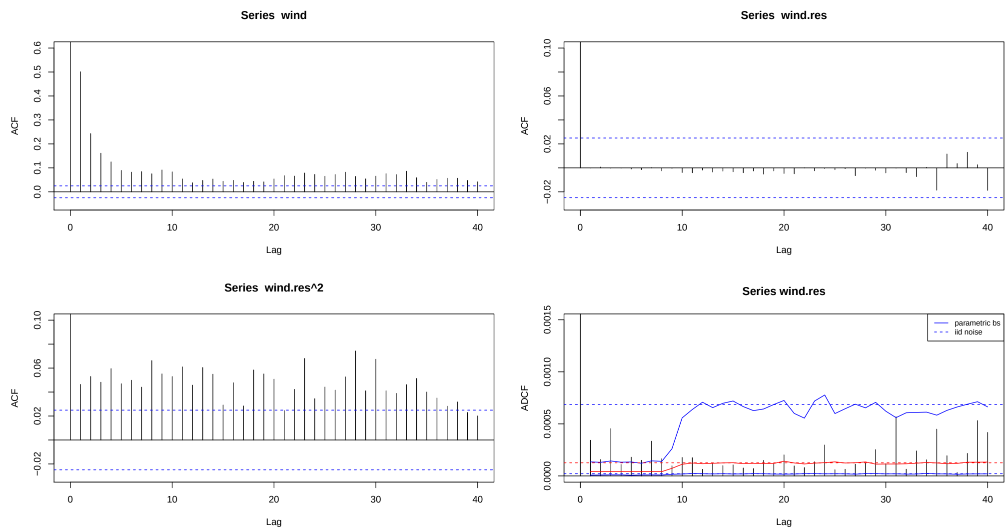


Figure 1.5: ACF and ADCF of Kilkenny wind speed time series and AR(9) fitted residuals. Upper left: ACF of the series. Upper right: ACF of the residuals. Lower left: ACF of the residual squares. Lower right: ADCF of the residuals, the 5%, 50%, 95% confidence bounds of ADCF for fitted residuals from 1000 parametric bootstraps, and that for iid noise from 1000 random permutations.

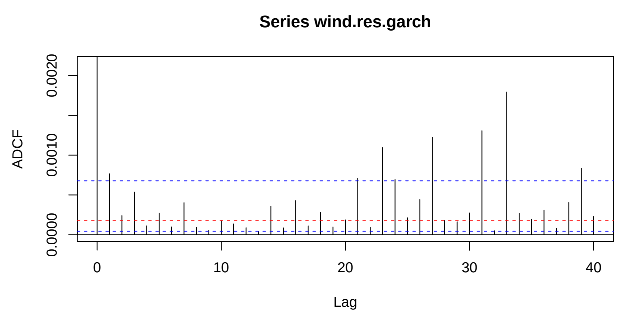


Figure 1.6: ADCF of the residuals of Kilkenny wind speed time series from AR(9)-GARCH fitting and the 5%, 50%, 95% confidence bounds of ADCF for iid noise from 1000 random permutations.

independent (dependent).

1. *For $\alpha \le 2$ there exists a constant $c > 0$ such that for $\epsilon \in [0, 1/2)$,*

$$n\,\mathbb{E}[|C_n(s,t) - C(s,t)|^2] \le c\left(1 \wedge |s|^{\alpha(1+\epsilon)/u}\right)\left(1 \wedge |t|^{\alpha(1+\epsilon)/u}\right), \qquad n \ge 1\,. \tag{1.39}$$

2. *If $\alpha \in (u/2, u]$ then $\sqrt{n}(\varphi^n_{X,Y} - \varphi_{X,Y}) \xrightarrow{d} G$ on compact sets $K \subset \mathbb{R}^{p+q}$ for some complex-valued mean-zero Gaussian field G.*

Remark 1.6.2. Notice that $C(s,t)=0$ when X_0 and Y_0 are independent.

Proof. (1) We focus on the proof under the assumption of independence. At the end, we indicate the changes necessary when X_0 and Y_0 are dependent.

We write

$$U_k = e^{\,i\langle s,X_k\rangle} - \varphi_X(s)\,, \qquad V_k = e^{\,i\langle t,Y_k\rangle} - \varphi_Y(t)\,, \qquad k\ge 1\,,$$

where we suppress the dependence of U_k and V_k on s and t, respectively. Then

$$\begin{aligned} n\,\mathbb{E}[|C_n(s,t)|^2] &= n\,\mathbb{E}\Big|\frac{1}{n}\sum_{k=1}^n U_k\,V_k - \frac{1}{n}\sum_{k=1}^n U_k\frac{1}{n}\sum_{l=1}^n V_l\Big|^2 \\ &\le 2\,n\,\mathbb{E}\Big[\Big|\frac{1}{n}\sum_{k=1}^n U_k\,V_k\Big|^2\Big] + 2\,n\,\mathbb{E}\Big[\Big|\frac{1}{n}\sum_{k=1}^n U_k\frac{1}{n}\sum_{l=1}^n V_l\Big|^2\Big] =: 2\,(I_1+I_2)\,. \end{aligned}$$

We have by stationarity

$$I_1 = \mathbb{E}[|U_0V_0|^2] + 2\sum_{h=1}^{n-1}(1-h/n)\,\mathrm{Re}\,\mathbb{E}[U_0V_0\,\overline{U_hV_h}]\,.$$

Since U_0 and V_0 are independent $\mathbb{E}[U_0V_0]=0$. In view of the α-mixing condition (see Doukhan (1994), Section 1.2.2, Theorem 3(a)) we have

$$\begin{aligned} \big|\mathrm{Re}\mathbb{E}[U_0V_0\,\overline{U_hV_h}]\big| &\le c\,\alpha_h^{1/r}\,(\mathbb{E}[|U_0V_0|^u])^{2/u} \qquad (1.40)\\ &= c\,\alpha_h^{1/r}\,(\mathbb{E}[|U_0|^u])^{2/u}(\mathbb{E}[|V_0|^u])^{2/u} \\ &\le c\,\alpha_h^{1/r}\,(\mathbb{E}[|U_0|^2])^{2/u}(\mathbb{E}[|V_0|^2])^{2/u}\,. \end{aligned}$$

In the last step we used that $u=2r/(r-1)>2$ and that $\max(|U_0|,|V_0|)\le 2$. We have for $\alpha\in(0,2]$

$$\mathbb{E}[|U_0|^2] = 1-|\varphi_X(s)|^2 \le \mathbb{E}[1\wedge|\langle s,X-X'\rangle|^\alpha] \le c\,\big(1\wedge|s|^\alpha\big)\,.$$

Therefore and since $\sum_h \alpha_h^{1/r}<\infty$ we have $I_1 \le c\,\big(1\wedge|s|^\alpha\big)^{2/u}\,\big(1\wedge|t|^\alpha\big)^{2/u}$.

Now we turn to I_2. By the Cauchy-Schwarz inequality and since $|\frac{1}{n}\sum_{k=1}^n U_k|$ and $|\frac{1}{n}\sum_{k=1}^n V_k|$ are bounded by 2 we have

$$I_2 \le 2\,n\,\Big(\mathbb{E}\Big|\frac{1}{n}\sum_{k=1}^n U_k\Big|^4\Big)^{1/2}\,\Big(\mathbb{E}\Big|\frac{1}{n}\sum_{k=1}^n V_k\Big|^4\Big)^{1/2}$$

$$\leq \ c\Big(n\,\mathbb{E}\Big|\frac{1}{n}\sum_{k=1}^{n}U_k\Big|^{2+\delta}\Big)^{1/2}\Big(n\,\mathbb{E}\Big|\frac{1}{n}\sum_{k=1}^{n}V_k\Big|^{2+\delta}\Big)^{1/2},$$

for any $\delta \in [0,2]$. In view of Lemma 18.5.1 in Ibragimov and Linnik (1971) we have for $\delta \in [0,1)$,

$$I_2 \ \leq \ c\Big(n\,\mathbb{E}\Big|\frac{1}{n}\sum_{k=1}^{n}U_k\Big|^{2}\Big)^{(2+\delta)/4}\Big(n\,\mathbb{E}\Big|\frac{1}{n}\sum_{k=1}^{n}V_k\Big|^{2}\Big)^{(2+\delta)/4},$$

Similar arguments as for I_1 show that

$$I_2 \ \leq \ c\,\big(1\wedge|s|^{\alpha(2+\delta)/4}\big)^{2/u}\,\big(1\wedge|t|^{\alpha(2+\delta)/4}\big)^{2/u}.$$

Combining the bounds for I_1 and I_2, we arrive at (1.39).

Now we indicate the changes necessary when X_0 and Y_0 are dependent. We use the notation above and, additionally, write $\widetilde{W}_k = U_kV_k - C(s,t)$. We have

$$C_n(s,t) - C(s,t) = \frac{1}{n}\sum_{k=1}^{n}\widetilde{W}_k - \frac{1}{n}\sum_{k=1}^{n}U_k\frac{1}{n}\sum_{l=1}^{n}V_l\,.$$

Then

$$n\,\mathbb{E}[|C_n(s,t)-C(s,t)|^2] \leq 2\,n\,\mathbb{E}\Big[\Big|\frac{1}{n}\sum_{k=1}^{n}\widetilde{W}_k\Big|^2\Big] + 2\,n\,\mathbb{E}\Big[\Big|\frac{1}{n}\sum_{k=1}^{n}U_k\frac{1}{n}\sum_{l=1}^{n}V_l\Big|^2\Big] = 2\,(I_1'+I_2)\,.$$

Since $\mathbb{E}[\widetilde{W}_0]=0$, we have by stationarity

$$I_1' \ = \ \mathbb{E}[|\widetilde{W}_0|^2] + 2\sum_{h=1}^{n-1}(1-h/n)\,\mathrm{Re}\,\mathbb{E}[\widetilde{W}_0\,\overline{\widetilde{W}_h}]\,.$$

Observe that $\mathbb{E}[|\widetilde{W}_0|^2] \leq 2(\mathbb{E}|U_0|^4\,\mathbb{E}|V_0|^4)^{1/2} + 2|C(s,t)|^2$ and

$$\begin{aligned} |U_0|^2 \ &\leq \ (|e^{\,i\langle s,X_0\rangle}-1| + \mathbb{E}[|1-e^{\,i\langle s,X_0\rangle}|])^2 \\ &\leq \ c\,(1\wedge(|s|\,|X_0|)^{\alpha/2})^2 + c\,(1\wedge(|s|^{\alpha/2}\,\mathbb{E}|X_0|^{\alpha/2}))^2\,. \end{aligned}$$

Since $\mathbb{E}[|X_0|^{2\alpha}] < \infty$ we have $\mathbb{E}[|U_0|^4] \leq c\,(1\wedge|s|^{2\alpha})$ and in a similar manner, $\mathbb{E}|V_0|^4 \leq c\,(1\wedge|t|^{2\alpha})$. We also have $|C(s,t)|^2 \leq c\,(1\wedge|s|^{\alpha})\,(1\wedge|t|^{\alpha})$. Finally, we conclude that

$$\mathbb{E}[|\widetilde{W}_0|^2] \leq c\,(1\wedge|s|^{\alpha})\,(1\wedge|t|^{\alpha})\,.$$

With the α-mixing condition we obtain

$$\big|\mathrm{Re}\,\mathbb{E}[\widetilde{W}_0\,\overline{\widetilde{W}_h}]\big| \;\le\; c\,\alpha_h^{1/r}\,(\mathbb{E}[|\widetilde{W}_0|^u])^{2/u} \le c\,\alpha_h^{1/r}\,(\mathbb{E}[|\widetilde{W}_0|^2])^{2/u}.$$

This together with $\sum_h \alpha_h^{1/r} < \infty$ yields $I_1' \le c\,\big(1 \wedge |s|^\alpha\big)^{2/u}\,\big(1 \wedge |t|^\alpha\big)^{2/u}$. The remaining term I_2 can be treated in the same way as in the independent case. Combining the bounds for I_1' and I_2, we arrive at (1.39).

(2) We need an analog of S. Csörgő's central limit theorem (Csörgő, 1981a,b,c) for the empirical characteristic function of an iid multivariate sequence with Gaussian limit. For ease of notation we focus on the X-sequence; the proof for the (X,Y)-sequence is analogous and therefore omitted. The convergence of the finite-dimensional distributions of $\sqrt{n}(\varphi_X^n - \varphi_X)$ follows from Theorem 18.5.2 in Ibragimov and Linnik (1971) combined with the Cramér-Wold device. We need to show tightness of the normalized empirical characteristic function on compact sets. We use the sufficient condition of Theorem 3 in Bickel and Wichura (1971) for multiparameter processes. We evaluate the process on cubes $(s,t] = \prod_{k=1}^p (s_k, t_k]$, where $s = (s_1 \ldots, s_p)$ and $t = (t_1, \ldots, t_p)$ and $s_i < t_i$, $i = 1, \ldots, p$. The increment of the normalized empirical characteristic function on $(s,t]$ is given by

$$\begin{aligned} I_n(s,t] &= \sqrt{n}(\varphi_X^n(s,t] - \varphi_X(s,t]) \\ &= \frac{\sqrt{n}}{n}\sum_{r=1}^n \Big\{ \sum_{k_1=0,1} \cdots \sum_{k_p=0,1} (-1)^{p-\sum_j k_j}\big(\prod_{l=1}^p e^{\,i(s_l+k_l(t_l-s_l))X_r^{(l)}} \\ &\qquad -\mathbb{E}\big[\prod_{l=1}^p e^{\,i(s_l+k_l(t_l-s_l))X_r^{(l)}}\big]\big)\Big\} =: \frac{1}{\sqrt{n}}\sum_{r=1}^n W_r\,, \end{aligned} \tag{1.41}$$

where $X_r = (X_r^{(1)}, \ldots, X_r^{(p)})$ and

$$W_r = \prod_{l=1}^p (e^{\,it_l X_r^{(l)}} - e^{\,is_l X_r^{(l)}}) - \mathbb{E}\big[\prod_{l=1}^p (e^{\,it_l X_r^{(l)}} - e^{\,is_l X_r^{(l)}})\big]\,.$$

We apply the sums $\sum_{k_j=0,1}$ inductively to derive (1.41). Observe that

$$\mathbb{E}\big[|I_n(s,t]|^2\big] = \mathbb{E}[|W_0|^2] + 2\sum_{h=1}^{n-1}(1-h/n)\mathrm{Re}\,\mathbb{E}[W_0\,\overline{W}_h]\,.$$

By the Lipschitz property of trigonometric functions we have for some constant $c > 0$ and $\alpha \in (0, 2]$,

$$|e^{\,is_l X_r^{(l)}} - e^{\,it_l X_r^{(l)}}|^2 \le c\,(1 \wedge |t_l - s_l|^2 (X_r^{(l)})^2/4) \le c\,(1 \wedge |s_l - t_l|^\alpha |X_r^{(l)}|^\alpha/4^\alpha)\,.$$

Proceeding as for (1.40) and noticing that $\alpha \le 2 \le u$, we have

$$\begin{aligned}|\mathbb{E}[W_0\,\overline{W}_h]| &\le c\,\alpha_h^{1/r}\,(\mathbb{E}[|W_0|^u])^{2/u}\\ &\le \alpha_h^{1/r} \prod_{l=1}^p |s_l - t_l|^{2\alpha/u} \big(\mathbb{E}\big[\prod_{l=1}^p |X_0^{(l)}|^\alpha\big]\big)^{2/u}.\end{aligned}$$

Using the summability of $(\alpha_h^{1/r})$ and the moment condition on X_0, we may conclude that

$$\mathbb{E}\big[|I_n(s,t)|^2\big] \le c \prod_{l=1}^p |s_l - t_l|^{2\alpha/u}.$$

If $2\alpha/u > 1$ the condition of Theorem 3 in Bickel and Wichura (1971) yields that the processes $(\sqrt{n}(\varphi_X^n - \varphi_X))$ are tight on compact sets. $\square$

Proof of Theorem 1.3.2(1). Recall the definition of K_δ from (1.21) and that X_0 and Y_0 are independent. From Lemma 1.6.1 and the continuous mapping theorem we have

$$\int_{K_\delta} |\sqrt{n} C_n(s,t)|^2\,\mu(ds,dt) \xrightarrow{d} \int_{K_\delta} |G(s,t)|^2\,\mu(ds,dt)\,, \qquad n \to \infty\,.$$

From (1.23), (1.39) and the dominated convergence theorem, for any $\varepsilon > 0$, some $\epsilon \in (0, 1/2]$ and $\alpha' \le \min(2, \alpha)$,

$$\begin{aligned}&\lim_{\delta \downarrow 0} \limsup_{n\to\infty} \mathbb{P}\left(\int_{K_\delta^c} |\sqrt{n}\,C_n(s,t)|^2\,\mu(ds,dt) > \varepsilon\right)\\ \le\ &\varepsilon^{-1} \lim_{\delta\downarrow 0}\limsup_{n\to\infty} \int_{K_\delta^c} \mathbb{E}[|\sqrt{n} C_n(s,t)|^2]\,\mu(ds,dt)\\ \le\ &\lim_{\delta\downarrow 0} \int_{K_\delta^c} c\,\big(1 \wedge |s|^{\alpha'(1+\epsilon)/u}\big)\,\big(1 \wedge |t|^{\alpha'(1+\epsilon)/u}\big)\mu(ds,dt) = 0\,.\end{aligned}$$

$\square$

Proof of Theorem 1.3.2(2). Now we assume that X_0 and Y_0 are dependent. We observe that

$$\sqrt{n}\,(T_n(s,t;\mu) - T(s,t;\mu)) = \int_{\mathbb{R}^{p+q}} \sqrt{n}(|C_n(s,t)|^2 - |C(s,t)|^2)\,\mu(ds,dt).$$

In view of Lemma 1.6.1(2) and the a.s. convergence of C_n on compact sets the continuous mapping theorem implies that for some Gaussian mean-zero process G',

$$\int_{K_\delta} \sqrt{n}\{(C_n(s,t) - C(s,t))\overline{C}_n(s,t) + C(s,t)(\overline{C}_n(s,t) - \overline{C}(s,t))\}\,\mu(ds,dt)$$
$$\stackrel{d}{\to} \int_{K_\delta} G'(s,t)\,\mu(ds,dt)\,, \qquad n\to\infty\,,$$

where $G'X(s,t) = 2\mathrm{Re}\{G(s,t)C(s,t)\}$. We have

$$\big||C_n|^2 - |C|^2\big| = \big||C_n - C|^2 + 2\,\mathrm{Re}\,(\overline{C}\,(C_n - C))\big| \le c\,|C_n - C|\,.$$

By Markov's inequality, (1.39) and (1.23),

$$\begin{aligned}
&\lim_{\delta\downarrow 0}\limsup_{n\to\infty} \mathbb{P}\left(\int_{K_\delta^c} \sqrt{n}\big||C_n(s,t)|^2 - |C(s,t)|^2\big|\,\mu(ds,dt) > \varepsilon\right)\\
&\le\ c\lim_{\delta\downarrow 0}\limsup_{n\to\infty}\int_{K_\delta^c}\big(n\,\mathbb{E}[|C_n - C|^2]\big)^{1/2}\,\mu(ds,dt)\\
&\le\ \lim_{\delta\downarrow 0}\int_{K_\delta^c} c\left(1\wedge|s|^{\alpha'(1+\epsilon)/u}\right)\left(1\wedge|t|^{\alpha'(1+\epsilon)/u}\right)\mu(ds,dt) = 0\,.
\end{aligned}$$

□

1.7 Proof of Theorem 1.4.2

We prove the result for the residuals calculated from least square estimates (LSEs). One may show that the same result holds for maximum likelihood and Yule-Walker estimates. The least squares estimator $\widehat{\boldsymbol{\phi}}$ of $\boldsymbol{\phi}$ satisfies the relation

$$\widehat{\boldsymbol{\phi}} - \boldsymbol{\phi} \ = \ \Gamma_{n,p}^{-1}\frac{1}{n}\sum_{t=p+1}^{n}\mathbf{X}_{t-1}\,Z_t\,,$$

where

$$\Gamma_{n,p} \ = \ \frac{1}{n}\sum_{t=p+1}^{n}\mathbf{X}_{t-1}^T\mathbf{X}_{t-1}\,.$$

If $\sigma^2 = \mathrm{var}(Z_t) < \infty$, we have by the ergodic theorem,

$$\Gamma_{n,p} \stackrel{a.s.}{\to} \Gamma_p = \big(\gamma_X(j-k)\big)_{1\le j,k\le p}\,, \qquad \text{where } \gamma_X(h) = \mathrm{cov}(X_0, X_h)\,, h\in\mathbb{Z}\,. \tag{4.1}$$

Causality of the process implies that the partial sum $\sum_{t=p+1}^{n} \mathbf{X}_{t-1} Z_t$ is a martingale and applying the martingale central limit theorem yields

$$\sqrt{n}\,(\widehat{\boldsymbol{\phi}} - \boldsymbol{\phi}) \xrightarrow{d} \mathbf{Q}\,, \tag{4.2}$$

where $\mathbf{Q}$ is $N(\mathbf{0}, \sigma^2 \Gamma_p^{-1})$ distributed.

Keeping this in mind, we start with a joint central limit theorem for C_n^Z and $\widehat{\boldsymbol{\phi}}$.

Lemma 1.7.1. *Consider an iid sequence (Z_t) with finite variance.*

1. *For every $h \geq 0$,*

$$\sqrt{n}\left(C_n^Z, \widehat{\boldsymbol{\phi}} - \boldsymbol{\phi}\right) \xrightarrow{d} (G_h, \mathbf{Q})\,,$$

where the convergence is in $\mathcal{C}(K) \times \mathbb{R}^p$, $K \subset \mathbb{R}^2$ is a compact set, G_h is the limit process of C_n^Z with covariance structure specified in Remark 3.9 for the sequence $((Z_t, Z_{t+h}))$, $\mathbf{Q}$ is the limit in (4.2), $(G_h, \mathbf{Q})$ are mean-zero and jointly Gaussian with covariance matrix

$$cov(G_h(s,t), \mathbf{Q}) = -\varphi_Z'(s)\,\varphi_Z'(t)\,\Gamma_p^{-1}\Psi_h\,, \qquad s,t \in \mathbb{R}\,, \tag{4.4}$$

where $\Psi_h = (\psi_{h-j})_{j=1,\ldots,p}$ and φ_Z' is the first derivative of φ_Z.

2. *For every $h \geq 0$,*

$$\sqrt{n}\left(C_n^Z, C_n^{\widehat{Z}} - C_n^Z\right) \xrightarrow{d} (G_h, \xi_h)\,,$$

where $(G_h, \mathbf{Q})$ are specified in (4.4) and

$$\xi_h(s,t) = t\varphi_Z(t)\,\varphi_Z'(s)\Psi_h^T\mathbf{Q}, \qquad (s,t) \in K\,, \tag{4.5}$$

the convergence is in $\mathcal{C}(K, \mathbb{R}^2)$, $K \subset \mathbb{R}^2$ is a compact set. In particular, we have

$$\sqrt{n}\,C_n^{\widehat{Z}} \xrightarrow{d} G_h + \xi_h\,. \tag{4.6}$$

Proof of part (1). We observe that, uniformly for $(s,t) \in K$,

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

$$= \frac{1}{n}\sum_{j=1}^{n}\big(e^{\,isZ_j}-\varphi_Z(s)\big)\big(e^{\,itZ_{j+h}}-\varphi_Z(t)\big)$$
$$-\frac{1}{n}\sum_{j=1}^{n}\big(e^{\,isZ_j}-\varphi_Z(s)\big)\frac{1}{n}\sum_{j=1}^{n}\big(e^{\,itZ_j}-\varphi_Z(t)\big)+O_{\mathbb{P}}(n^{-1})\,.$$

In view of the functional central limit theorem for the empirical characteristic function of an iid sequence (see Csörgő (1981a,1981b)) we have uniformly for $(s,t)\in K$,

$$\sqrt{n}\,C_n^Z(s,t) = \frac{1}{\sqrt{n}}\sum_{j=1}^{n}\big(e^{\,isZ_j}-\varphi_Z(s)\big)\big(e^{\,itZ_{j+h}}-\varphi_Z(t)\big)+O_{\mathbb{P}}(n^{-1/2})$$
$$= I_n(s,t)+O_{\mathbb{P}}(n^{-1/2})\,.$$

Therefore it suffices to study the convergence of the finite-dimensional distributions of $\big(I_n,\sqrt{n}\,(\widehat{\boldsymbol\phi}-\boldsymbol\phi)\big)$. In view of (4.1) it suffices to show the convergence of the finite-dimensional distributions of $\big(I_n,(1/\sqrt{n})\sum_{j=1}^n \mathbf{X}_{j-1}Z_j\big)$. This convergence follows by an application of the martingale central limit theorem and the Cramér-Wold device. It remains to determine the limiting covariance structure, taking into account the causality of the process (X_t). We have

$$\operatorname{cov}\Big(I_n,\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\mathbf{X}_{j-1}Z_j\Big) = \frac{1}{n}\mathbb{E}\Big[\sum_{j=1}^{n}\sum_{k=1}^{n}\big(e^{\,isZ_j}-\varphi_Z(s)\big)\big(e^{\,itZ_{j+h}}-\varphi_Z(t)\big)\mathbf{X}_{k-1}Z_k\Big]\,.$$

By causality, X_k and Z_j are independent for $k<j$. Hence $\mathbb{E}[(e^{\,isZ_j}-\varphi_Z(s))(e^{\,itZ_{j+h}}-\varphi_Z(t))X_{l-k}Z_l]$ is non-zero if and only if $l=j+h$ and $k\le h$, resulting in

$$\mathbb{E}\big[\big(e^{\,isZ_j}-\varphi_Z(s)\big)\big(e^{\,itZ_{j+h}}-\varphi_Z(t)\big)\,X_{l-k}\,Z_l\big]$$
$$= \mathbb{E}\big[X_{j+h-k}\big(e^{\,isZ_j}-\varphi_Z(s)\big)\big]\,\mathbb{E}\big[Z_{j+h}\big(e^{\,itZ_{j+h}}-\varphi_Z(t)\big)\big]$$
$$= \psi_{h-k}\,\mathbb{E}\big[Z\big(e^{\,isZ}-\varphi_Z(s)\big)\big]\,\mathbb{E}\big[Z\big(e^{\,itZ}-\varphi_Z(t)\big)\big]$$
$$= -\psi_{h-k}\,i\mathbb{E}\big[Ze^{\,isZ}\big]\,i\mathbb{E}\big[Ze^{\,itZ}\big]$$
$$= -\psi_{h-k}\,\varphi_Z'(s)\,\varphi_Z'(t)\,.$$

This implies (4.4). □

Proof of part (2). We observe that, uniformly for $(s,t)\in K$,

$$C_n^{\widehat Z}(s,t)-C_n^Z(s,t) \tag{1.42}$$

$$
\begin{aligned}
&= \frac{1}{n}\sum_{j=1}^{n} e^{\,isZ_j+itZ_{j+h}}\,\big(e^{\,i\,(\boldsymbol{\phi}-\widehat{\boldsymbol{\phi}})^T(s\mathbf{X}_{j-1}+t\mathbf{X}_{j+h-1})}-1\big)\\
&\quad +\frac{1}{n}\sum_{j=1}^{n}\big(1-e^{\,i(\boldsymbol{\phi}-\widehat{\boldsymbol{\phi}})^T s\mathbf{X}_{j-1}}\big)e^{\,isZ_j}\frac{1}{n}\sum_{j=1}^{n}e^{\,itZ_{j+h}}\\
&\quad +\frac{1}{n}\sum_{j=1}^{n}e^{\,i(\boldsymbol{\phi}-\widehat{\boldsymbol{\phi}})^T s\mathbf{X}_{j-1}+isZ_j}\frac{1}{n}\sum_{j=1}^{n}\big(1-e^{\,i(\boldsymbol{\phi}-\widehat{\boldsymbol{\phi}})^T t\mathbf{X}_{j+h-1}}\big)e^{\,itZ_{j+h}}+O_{\mathbb{P}}(n^{-1})\\
&= E_{n1}(s,t)+E_{n2}(s,t)+E_{n3}(s,t)+O_{\mathbb{P}}(n^{-1})\,. \qquad (1.43)
\end{aligned}
$$

Write

$$
\widetilde{E}_{n1}(s,t) \;=\; i\,(\boldsymbol{\phi}-\widehat{\boldsymbol{\phi}})^T\,\frac{1}{n}\sum_{j=1}^{n}(s\mathbf{X}_{j-1}+t\,\mathbf{X}_{j+h-1})\,e^{\,isZ_j+itZ_{j+h}}\,.
$$

In view of the uniform ergodic theorem, (4.2) and the causality of (X_t) we have

$$
\begin{aligned}
\sqrt{n}\widetilde{E}_{n1}(s,t) &\stackrel{d}{\to} -i\mathbf{Q}^T\mathbb{E}\big[(s\mathbf{X}_0+t\mathbf{X}_h)\,e^{\,i(sZ_1+tZ_{h+1})}\big] \qquad (1.44)\\
&= -t\varphi_Z(t)\varphi_Z'(s)\Psi_h^T\mathbf{Q}=\xi_h(s,t)\,,
\end{aligned}
$$

where the convergence is in $\mathcal{C}(K)$. By virtue of part (1) and the mapping theorem we have the joint convergence $\sqrt{n}(C_n^Z,\widetilde{E}_{n1})\stackrel{d}{\to}(G_h,\xi_h)$ in $\mathcal{C}(K,\mathbb{R}^2)$. Denoting the sup-norm in $\mathcal{C}(K)$ by $\|\cdot\|$, it remains to show that $\sqrt{n}\big(\|E_{n2}\|+\|E_{n3}\|+\|E_{n1}-\widetilde{E}_{n1}\|\big)\stackrel{\mathbb{P}}{\to}0$. The proof for E_{n2} and E_{n3} is analogous to (1.44) by observing that the limiting expectation is zero. We have by a Taylor expansion for some positive constant c,

$$
\sqrt{n}\|E_{n1}(s,t)-\widetilde{E}_{n1}(s,t)\| \;\le\; c\,\big|\sqrt{n}(\boldsymbol{\phi}-\widehat{\boldsymbol{\phi}})\big|^2\sup_{(s,t)\in K}\frac{1}{n^{3/2}}\sum_{j=1}^{n}\big|s\mathbf{X}_{j-1}+t\mathbf{X}_{j+h-1}\big|^2\stackrel{\mathbb{P}}{\to}0\,.
$$

In the last step we used the uniform ergodic theorem and (4.2). □

Proof of Theorem 1.4.2(1). We proceed as in the proof of Theorem 1.3.2. By virtue of (4.6) and the continuous mapping theorem we have

$$
\int_{K_\delta}|\sqrt{n}\,C_n^{\widehat{Z}}(s,t)|^2\,\mu(ds,dt)\stackrel{d}{\to}\int_{K_\delta}|G(s,t)+\xi_h(s,t)|^2\,\mu(ds,dt)\,,\qquad n\to\infty\,.
$$

Thus it remains to show that

$$
\lim_{\delta\downarrow 0}\limsup_{n\to\infty}\mathbb{P}\Big(\int_{K_\delta^c}|\sqrt{n}C_n^{\hat{Z}}(s,t)|^2\mu(ds,dt)>\varepsilon\Big)=0\,,\qquad \varepsilon>0\,. \qquad (1.45)
$$

Following the lines of the proof of Theorem 1.3.2, we have

$$\lim_{\delta\downarrow 0}\limsup_{n\to\infty}\int_{K_\delta^c}\mathbb{E}[|\sqrt{n}C_n^Z(s,t)|^2]\,\mu(ds,dt)=0\,;$$

see also Remark 3.4. Thus it suffices to show

$$\lim_{\delta\downarrow 0}\limsup_{n\to\infty}\mathbb{P}\Big(\int_{K_\delta^c}|\sqrt{n}(C_n^{\hat Z}(s,t)-C_n^Z(s,t))|^2\mu(ds,dt)>\varepsilon\Big)=0\,,\qquad \varepsilon>0\,.$$

For convenience we redefine

$$C_n^Z=\frac{1}{n}\sum_{j=p+1}^{n-h}e^{\,isZ_j+itZ_{j+h}}-\frac{1}{n}\sum_{j=p+1}^{n-h}e^{\,isZ_j}\,\frac{1}{n}\sum_{j=p+1}^{n-h}e^{\,itZ_{j+h}}\,.$$

This version does not change previous results for C_n^Z.

Using telescoping sums, we have for $\bar n=n-p-h$,

$$\begin{aligned}
&\frac{\bar n}{n}(C_n^{\hat Z}(s,t)-C_n^Z(s,t))\\
&=\frac{1}{\bar n}\sum_{j=p+1}^{n-h}A_jB_j-\frac{1}{\bar n}\sum_{j=p+1}^{n-h}A_j\frac{1}{\bar n}\sum_{j=p+1}^{n-h}B_j-\frac{1}{\bar n}\sum_{j=p+1}^{n-h}U_j\sum_{j=p+1}^{n-h}B_j-\frac{1}{\bar n}\sum_{j=p+1}^{n-h}V_j\sum_{j=p+1}^{n-h}A_j\\
&\quad+\frac{1}{\bar n}\sum_{j=p+1}^{n-h}U_jB_j+\frac{1}{\bar n}\sum_{j=p+1}^{n-h}V_jA_j=:\sum_{j=1}^{6}I_{nj}(s,t),
\end{aligned}$$

where, suppressing the dependence on s,t in the notation,

$$\begin{aligned}
U_j&=e^{\,isZ_j}-\varphi_Z(s)\,,\qquad V_j=e^{\,itZ_{j+h}}-\varphi_Z(t)\,,\\
A_j&=e^{\,isZ_j}\,(e^{\,is(\boldsymbol{\phi}-\widehat{\boldsymbol{\phi}})'X_{j-1}}-1)\,,\qquad B_j=e^{\,itZ_{j+h}}\,(e^{\,is(\boldsymbol{\phi}-\widehat{\boldsymbol{\phi}})'X_{j+h-1}}-1).
\end{aligned}$$

Write $K_n=|\sqrt{n}(\boldsymbol{\phi}-\widehat{\boldsymbol{\phi}})|$ and $c>0$ for any positive constant which may differ from line to line. By Taylor expansions we have

$$\begin{aligned}
&n\,|I_{n1}(s,t)|^2\le\Big(\frac{\sqrt{n}}{\bar n}\sum_{j=p+1}^{n-h}|A_jB_j|\Big)^2\\
\le\;&c\Big(\frac{\sqrt{n}}{\bar n}\sum_{j=p+1}^{n-h}(1\wedge|s|\,|\boldsymbol{\phi}-\widehat{\boldsymbol{\phi}}|\,|X_{j-1}|)\,(1\wedge|t|\,|\boldsymbol{\phi}-\widehat{\boldsymbol{\phi}}||X_{j+h-1}|)\Big)^2\\
\le\;&c\Big(\,\min\big(|s\,t|\,K_n^2\,\frac{1}{\bar n^{3/2}}\sum_{j=p+1}^{n-h}|X_{j-1}\,X_{j+h-1}|\,,\;|s|\,K_n\,\frac{1}{\bar n}\sum_{j=p+1}^{n-h}|X_{j-1}|\,,
\end{aligned}$$

$$|t|\, K_n\, \frac{1}{\bar n} \sum_{j=p+1}^{n-h} |X_{j+h-1}|\Big)\Big)^2 .$$

The quantities K_n are stochastically bounded. From ergodic theory (see Example 2.19 in Samorodnitsky (2016)), $n^{-1}\sum_{j=1}^n |X_j| = O_{\mathbb P}(1)$ and $n^{-3/2}\sum_{j=1}^n |X_j\, X_{j+h}| = o_{\mathbb P}(1)$. Hence

$$n\,|I_{n1}(s,t)|^2 \;\le\; \min(s^2,t^2,(st)^2)\, O_{\mathbb P}(1) \;\le\; \big((1\wedge s^2)\,(1\wedge t^2) + (s^2+t^2)\mathbf{1}(|s|\wedge |t|\ge 1)\big)\; O_{\mathbb P}(1),$$

where the term $O_{\mathbb P}(1)$ does not depend on s and t. Thus we conclude for $k=1$ that

$$\lim_{\delta\downarrow 0}\limsup_{n\to\infty} \mathbb P\Big(n\int_{K_\delta^c} |I_{nk}(s,t)|^2\,\mu(ds,dt) > \varepsilon\Big) = 0\,, \qquad \varepsilon>0\,. \tag{1.46}$$

A similar argument yields

$$\begin{aligned}
n\,|I_{n2}(s,t)|^2 \;&\le\; \Big(\frac{\sqrt n}{\bar n^2}\sum_{j,k=p+1}^{n-h} |A_j|\,|B_k|\Big)^2\\
&\le\; \Big(\frac{\sqrt n}{\bar n^2}\sum_{j,k=p+1}^{n-h} (1\wedge |s|\,|\boldsymbol\phi-\widehat{\boldsymbol\phi}|\,|X_{j-1}|)(1\wedge |t|\,|\boldsymbol\phi-\widehat{\boldsymbol\phi}|\,|X_{k+h-1}|)\Big)^2\\
&\le\; c\Bigg(\min\Big(|st|\,K_n^2\,\frac{1}{\bar n^{5/2}}\sum_{j,k=p+1}^{n-h} |X_{j-1}\,X_{k+h-1}|,\\
&\qquad\qquad |s|\,K_n\,\frac{1}{\bar n}\sum_{j=p+1}^{n-h}|X_{j-1}|\,,\,|t|\,K_n\,\frac{1}{\bar n}\sum_{k=p+1}^{n-h}|X_{k+h-1}|\Big)\Bigg)^2\\
&\le\; \min(s^2,t^2,(st)^2)\,O_{\mathbb P}(1).
\end{aligned}$$

Then (1.46) holds for $k=2$. Taylor expansions also yield

$$\begin{aligned}
n\,|I_{n3}(s,t)|^2 \;&\le\; \Big(\frac{\sqrt n}{\bar n^2}\sum_{j,k=p+1}^{n-h} |U_j|\,|B_k|\Big)^2\\
&\le\; c\,\Big(\frac{\sqrt n}{\bar n^2}\sum_{j,k=p+1}^{n-h} (1\wedge \frac12 |s|\,(|Z_j|+\mathbb E|Z|))(1\wedge |t|\,|\boldsymbol\phi-\widehat{\boldsymbol\phi}|\,|X_{k+h-1}|)\Big)^2\\
&\le\; \min(t^2,(st)^2)\,O_{\mathbb P}(1).
\end{aligned}$$

This proves (1.46) for $k=3$. By a symmetry argument but with the corresponding bound $\min(s^2,(st)^2)\,O_{\mathbb P}(1)$, (1.46) for $k=4$ follows as well. By Taylor expansion, we also have

$$n\,|I_{n5}(s,t)|^2 \;\le\; \Big(\frac{\sqrt n}{\bar n}\sum_{j=p+1}^{n-h} |U_j|\,|B_j|\Big)^2$$

$$\leq \; c \left(\frac{\sqrt{n}}{\bar{n}} \sum_{j=p+1}^{n-h} (1 \wedge \frac{1}{2}|s| \, (|Z_j| + \mathbb{E}|Z|))(1 \wedge |t| \, |\boldsymbol{\phi} - \widehat{\boldsymbol{\phi}}| \, |X_{j+h-1}|) \right)^2$$
$$\leq \; \min(t^2, (st)^2) \, O_{\mathbb{P}}(1).$$

We may conclude that (1.46) holds for $k = 5$. The case $k = 6$ follows in a similar way with the corresponding bound $\min(s^2, (st)^2) \, O_{\mathbb{P}}(1)$. □

Proof of Theorem 1.4.2(2). We follow the proof of Theorem 4.2(1) by first showing that

$$\sqrt{n} \, C_n^{\widehat{Z}} \stackrel{d}{\to} G_h \tag{1.47}$$

in $\mathcal{C}(K)$ for $K \subset \mathbb{R}^2$ compact, and then (1.45). The convergence $\sqrt{n} \, C_n^Z \stackrel{d}{\to} G_h$ in $\mathcal{C}(K)$ continues to hold as in the proof of Theorem 4.2(1) since the conditions in Csörgő (1981a,1981b) are satisfied if some moment of Z is finite. For (1.47) it suffices to show that

$$\sqrt{n} \, (C_n^{\widehat{Z}} - C_n^Z) \stackrel{p}{\to} 0 \tag{1.48}$$

in $\mathcal{C}(K)$. Recalling the decomposition (1.43), we now can show directly that $\sup_{|s|,|t| \leq M} \sqrt{n} |E_{ni}(s,t)| \stackrel{p}{\to} 0$ for any $M > 0$ and $i = 1, 2, 3$, which implies (1.48). We focus only on the case $i = 1$ to illustrate the method; the cases $i = 2, 3$ are analogous. We observe that for $\delta > 0$,

$$\sup_{|s|,|t| \leq M} \sqrt{n} |E_{n1}(s,t)| \; \leq \; \sup_{|s|,|t| \leq M} \sqrt{n} |\boldsymbol{\phi} - \widehat{\boldsymbol{\phi}}| \, \frac{1}{n} \sum_{j=p+1}^{n-h} |s \mathbf{X}_{j-1} + t \, \mathbf{X}_{j+h-1}|$$
$$\leq \; M \, n^{\frac{1}{\delta}} |\boldsymbol{\phi} - \widehat{\boldsymbol{\phi}}| \; n^{-\frac{1}{\delta} - \frac{1}{2}} \sum_{j=1}^{n} |\mathbf{X}_j| \, . \tag{1.49}$$

On the other hand, under the conditions of Theorem 4.2(2) Hannan and Kanter (1977) showed for $\delta > \alpha$,

$$n^{1/\delta} \, (\boldsymbol{\phi} - \widehat{\boldsymbol{\phi}}) \stackrel{a.s.}{\to} 0.$$

For $\alpha \in (1, 2)$, $\mathbb{E}[|\mathbf{X}|] < \infty$ and since we can choose $\delta = 2$ such that $1/\delta + 1/2 = 1$. The ergodic theorem finally yields that the right-hand side in (1.49) converges to zero a.s. As regards the case $\alpha \in (0, 1]$, we have $\mathbb{E}[|\mathbf{X}|^{\alpha-\gamma}] < \infty$ for any small γ and

$$\mathbb{E}\Big[\big| n^{-1/\delta - 1/2} \sum_{j=1}^{n} |\mathbf{X}_j| \big|^{\alpha-\gamma} \Big] \leq n^{-(\alpha-\gamma)(1/\delta + 1/2) + 1} \, \mathbb{E}[|\mathbf{X}|^{\alpha-\gamma}] \to 0.$$

If we choose δ close to α and γ close to zero the right-hand side in (1.49) converges to zero in probability.

Using the same bounds as in part (1), but writing this time $K_n = n^{1/\delta}|\boldsymbol{\phi} - \widehat{\boldsymbol{\phi}}|$, we have

$$\begin{aligned} n\,|I_{n1}(s,t)|^2 &\le c\Big(\min\big(|s\,t|\,K_n^2\,n^{-1/2-2/\delta}\sum_{j=1}^n |X_{j-1}X_{j+h-1}|\,,|s|\,K_n\,n^{-1/\delta-1/2}\sum_{j=0}^n |X_j|,\\ &\qquad |t|\,K_n\,n^{-1/\delta-1/2}\sum_{j=0}^n |X_j|\big)\Big)^2\\ &\le c\,\min(\,|s\,t|^2,\,|s|^2,\,|t|^2\,)\,\max\Big(K_n^2\,n^{-1/2-2/\delta}\,\sum_{j=1}^n |X_{j-1}X_{j+h-1}|,\\ &\qquad K_n\,n^{-1/\delta-1/2}\,\sum_{j=0}^n |X_j|\Big)^2\,. \end{aligned}$$

The same argument as above shows that $n^{-1/\delta-1/2}\sum_{j=0}^n |X_j| = O_{\mathbb{P}}(1)$ for δ close to α. Since $2|X_{j-1}X_{j+h-1}| \le X_{j-1}^2 + X_{j+h-1}^2$ a similar argument shows that $n^{-1/2-2/\delta}\sum_{j=1}^n |X_{j-1}X_{j+h-1}| = O_{\mathbb{P}}(1)$. These facts establish (1.46) for $k=1$. The same arguments show that bounds analogous to part (1) can be derived for $n\,|I_{nk}(s,t)|^2$ for $k = 2,\dots,6$. We omit further details. □

Acknowledgements

We would like to thank the two anonymous referees for careful reading and constructive comments. This research by Davis and Wan was supported in part by Army MURI grant W911NF-12-1-0385. Muneya Matsui's research is partly supported by JSPS Grant-in-Aid for Young Scientists B (16K16023) and Nanzan University Pache Research Subsidy I-A-2 for the 2017 academic year. Thomas Mikosch's research is partly supported by the Danish Research Council Grant DFF-4002-00435.

Chapter 2

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

2.1 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 $X_1, \ldots, X_n$, 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

$$X_j := f(X_{-\infty:j}, Z_j; \boldsymbol{\beta}), \tag{2.1}$$

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 chapter is to provide a validity check of the model (2.1) by inspecting the residuals.

A typical assumption for time series models is that the recursion (2.1) is casual and invertible, that is,

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

and

$$Z_j = Z_j(\boldsymbol{\beta}) = h(X_{-\infty:j}; \boldsymbol{\beta}) \tag{2.2}$$

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

$$\tilde{Z}_j := Z_j(\hat{\boldsymbol{\beta}}) = h(X_{-\infty:j}; \hat{\boldsymbol{\beta}}), \tag{2.3}$$

the residuals based on the infinite sequence $\{X_j, j \le n\}$. If the recursion (2.1) 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

$$\hat{Z}_j := h(Y_{-\infty:j}; \hat{\boldsymbol{\beta}}), \quad j = 1, \ldots, n, \tag{2.4}$$

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 (2.3) becomes negligible for large j and $\hat{Z}_j$ and $\tilde{Z}_j$ become indistinguishable.

While $\hat{Z}_1, \ldots, \hat{Z}_n$ 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 chapter, 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) (i.e., Section 1.4 of this thesis) which considered this problem for AR processes.

The remainder of the chapter is structured as follows. An introduction to distance correlation and ADCV along with some historical remarks are given in Section 2.2. In Section 2.3, we provide the limit result for the ADCV of the residuals for a general class of time series models. Two points regarding implementing theory are discussed in Section 2.4. We then apply the result to ARMA and GARCH models in Section 2.5 and 2.6 and illustrate with simulation studies. A simulated example where the data does not conform with the model is also demonstrated in Section 2.7.

2.2 Distance covariance

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

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

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

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

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

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

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

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

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,

$$\sqrt{n}C_n^X \xrightarrow{d} G_h \quad \text{on } \mathcal{C}(K), \tag{2.5}$$

for 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

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

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) (i.e., Chapter 1 of this thesis) 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 2.3, 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

$$\hat{T}_h(X;\mu) = \frac{1}{(n-h)^2}\sum_{i,j=1}^{n-h} \hat{\mu}(X_i - X_j, X_{i+h} - X_{j+h})$$

$$+\frac{1}{(n-h)^4}\sum_{i,j,k,l=1}^{n-h}\hat{\mu}(X_i-X_j,X_{k+h}-X_{l+h})$$
$$-2\frac{1}{(n-h)^3}\sum_{i,j,k=1}^{n-h}\hat{\mu}(X_i-X_j,X_{i+h}-X_{k+h}),$$

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.

2.3 General result

Let $X_1,\dots,X_n$ be the observed sequence from a stationary time series $\{X_j\}$ generated from (2.1), and let $\hat{Z}_1,\dots,\hat{Z}_n$ be the estimated residual calculated through (2.4). 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 $\{X_k,k\le j\}$. 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 $\mathbf{Q}$.

(M2) Assume that the function h in the invertible representation (2.2) is continuously differentiable, and writing

$$\mathbf{L}_j(\boldsymbol{\beta}) := \frac{\partial}{\partial \boldsymbol{\beta}} h(X_{-\infty:j}; \boldsymbol{\beta}), \tag{2.6}$$

we have

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

(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

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

Theorem 2.3.1. *Let $X_1, \ldots, X_n$ be a sequence of observations generated from a causal and invertible time series model* (2.1)*. 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* (2.4) *satisfying conditions (M1)–(M3). Further assume that the weight measure μ satisfies*

$$\int_{\mathbb{R}^2} \left[(1 \wedge |s|^2)\,(1 \wedge |t|^2) + (s^2 + t^2)\,\mathbf{1}(|s| \wedge |t| > 1) \right] \mu(ds, dt) < \infty. \tag{2.7}$$

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

$$\xi_h(s, t) := it\mathbf{Q}^T \mathbb{E}\left[\left(e^{isZ_0} - \varphi_Z(s)\right) e^{itZ_h} \mathbf{L}_h(\boldsymbol{\beta}) \right], \tag{2.8}$$

with $\mathbf{Q}$ being the limit distribution of $\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ and $\mathbf{L}_h$ as defined in (2.6)*.*

The proof of the theorem is provided in Section 2.8.

2.4 Two notes on implementation

2.4.1 Auto-distance correlation function (ADCF)

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):

$$n\hat{R}_h(\hat{Z}; \mu) \xrightarrow{d} \frac{\|G_h + \xi_h\|_\mu^2}{T_0(Z; \mu)}, \tag{2.9}$$

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

2.4.2 Parametric bootstrap

The limit in (2.9) 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 $X_1, \ldots, X_n$, let $\hat{\boldsymbol{\beta}}$ be the parameter estimate and $\hat{Z}_1, \ldots, \hat{Z}_n$ be the estimated residuals. A set of bootstrapped residuals can be obtained as follows:

1. Sample iid $Z_1^*, \ldots, Z_n^*$ from the empirical distribution of $\{\hat{Z}_j\}$, i.e., with replacement from $\hat{Z}_1, \ldots, \hat{Z}_n$.
2. Generate $X_1^*, \ldots, X_n^*$ from the time series model with parameter value $\hat{\boldsymbol{\beta}}$ and residual sequence $Z_1^*, \ldots, Z_n^*$.

3. Re-fit the time series model. Obtained the parameter estimate $\hat{\boldsymbol{\beta}}^*$ and the estimated residuals $\hat{Z}_1^*, \ldots, \hat{Z}_n^*$.

Let $n\hat{R}_h(\hat{Z}^*, \mu)$ be the ADCF calculated from the bootstrapped residuals $\hat{Z}_1^*, \ldots, \hat{Z}_n^*$. This procedure is repeated B times to obtain $n\hat{R}_h^{(1)}(\hat{Z}^*, \mu), \ldots, n\hat{R}_h^{(B)}(\hat{Z}^*, \mu)$. 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.

2.5 Example: ARMA(p,q)

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

$$X_t = \sum_{i=1}^{p} \phi_i X_{t-i} + Z_t + \sum_{j=1}^{q} \theta_j Z_{t-j}, \tag{2.10}$$

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 (2.10) 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

$$\hat{Z}_t = \sum_{j=0}^{t-1} \pi_j(\hat{\boldsymbol{\beta}}) X_{t-j}. \tag{2.11}$$

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.

Corollary 2.5.1. *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* (2.11)*. Assume that μ satisfies* (2.7)*, 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 (2.5) *and ξ_h in* (2.8)*.*

The proof of Corollary 2.5.1 is given in Section 2.9.

Remark 2.5.2. 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).

2.5.1 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 2.1, 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 (2.7), the ADCFs are evaluated using the Gaussian weight measure $N(0, 0.5^2)$. Confirming the results in Theorem 2.3.1 and Corollary 2.5.1, 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.

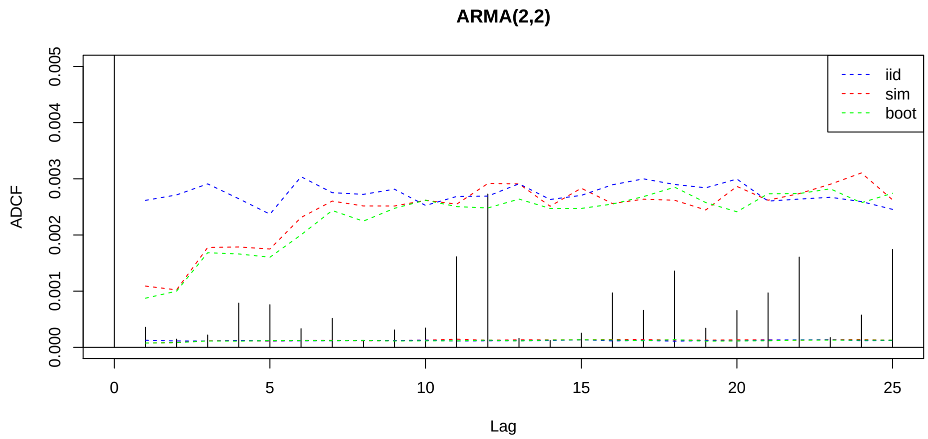


Figure 2.1: Empirical 5% and 95% quantiles of the ADCF for a) iid innovations; b) estimated residuals; c) bootstrapped residuals; from a ARMA(2,2) model.

2.6 Example: GARCH(p,q)

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

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i X_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2, \quad \alpha_0 > 0,\ \alpha_i \geq 0,\ \beta_j \geq 0. \tag{2.12}$$

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 (2.12) 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

$$\hat{Z}_t = X_t/\hat{\sigma}_t. \tag{2.13}$$

Define the parameter space by

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

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 $\boldsymbol{\theta}$ lies in the interior of $\boldsymbol{\Theta}$.

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

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

(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 $B(z) = 1 - \sum_{j=1}^p \beta_j z^j$ 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

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

where

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

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.

Corollary 2.6.1. *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* (2.13)*. Assume that (Q1)–(Q4) holds and that* μ *satisfies* (2.7)*, 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* (2.5) *and* ξ_h *in* (2.8)*.*

The proof of Corollary 2.6.1 is given in Section 2.10.

2.6.1 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 2.2, 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.

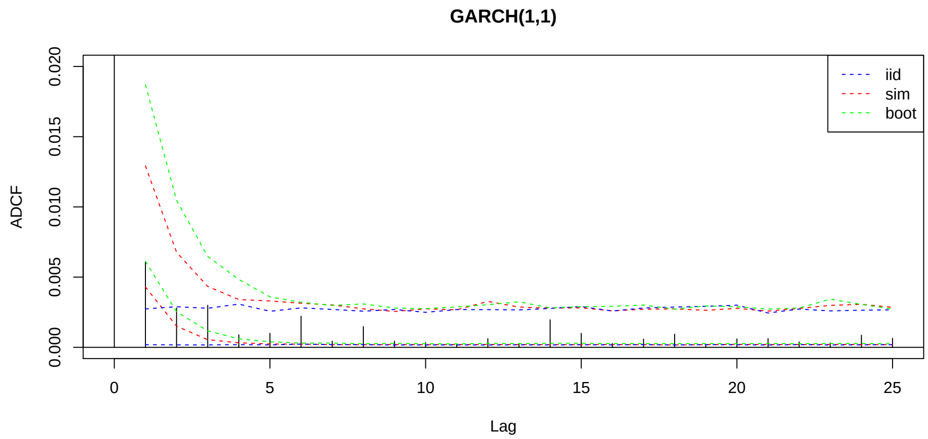


Figure 2.2: Empirical 5% and 95% quantiles of the ADCF for a) iid innovations; b) estimated residuals; c) bootstrapped residuals; from a GARCH(1,1) model.

2.7 Example: Non-causal AR(1)

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 2.3, 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.

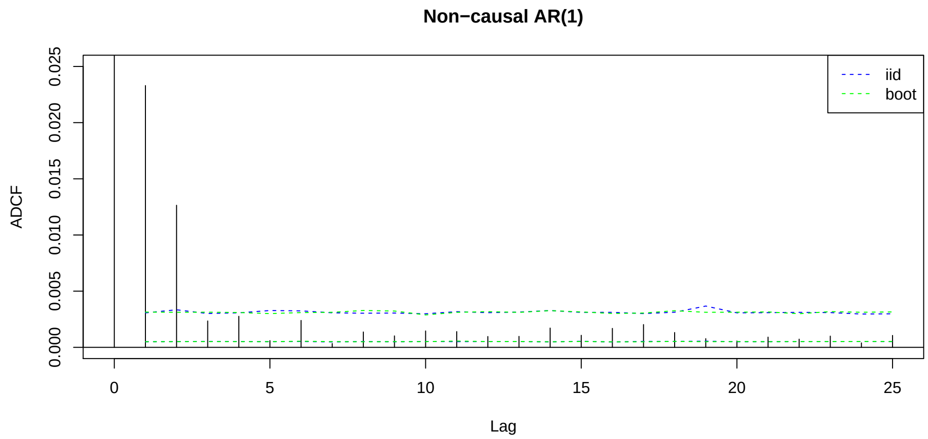


Figure 2.3: Empirical 5% and 95% quantiles of the ADCF for a) iid innovations; b) bootstrapped residuals; from non-causal AR(1) data fitted with a causal AR(1) model.

In the following appendices, we provide proofs to Theorem 2.3.1 and Corollaries 2.5.1 and 2.6.1. Throughout the proofs, c denotes a general constant whose value may change from line to line.

2.8 Proof of Theorem 2.3.1

Proof. The proof proceeds in the following steps with the aids of Propositions 2.8.1, 2.8.2 and 2.8.3. 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 2.8.1 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 $\mathbb{R}^2$. 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

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

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

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

and

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

These are shown in Propositions 2.8.2 and 2.8.3, respectively.

□

Proposition 2.8.1. *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

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

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

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

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

By assumption (M3),

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

It follows from a Taylor expansion that

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

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)) that

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

Hence, for fixed (s,t),

$$
E_n(s,t)\xrightarrow{d}\mathbf{Q}^T\mathbf{C}_h(s,t).
$$

Note that

$$\frac{1}{n}\sum_{j=1}^{n-h} e^{\,itZ_{j+h}} \xrightarrow{p} \varphi_Z(t),$$

and

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

We have, for fixed (s,t),

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

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

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

and

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

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

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

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

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.

□

Proposition 2.8.2. *Under the conditions of Theorem 3.1,*

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

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

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

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,

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

$$|st|^2\left(\frac{1}{\sqrt{n}}\sum_{j=1}^{n-h}|\hat{Z}_j-Z_j|^2\frac{1}{\sqrt{n}}\sum_{j=1}^{n-h}|\hat{Z}_{j+h}-Z_{j+h}|^2\right)\Bigg)$$

For $k=1,2$,

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

Therefore

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

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

$$\lim_{\delta\to0}\limsup_{n\to\infty}\mathbb{P}\left(\int_{K_\delta^c}n|I_{n1}(s,t)|^2\mu(ds,dt)>\varepsilon\right)=0.$$

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. □

Proposition 2.8.3. *Under the conditions of Theorem 3.1,*

$$\lim_{\delta\to0}\mathbb{P}\left(\int_{K_\delta^c}|G_h+\xi_h|^2\mu(ds,dt)>\varepsilon\right)=0.$$

Proof. Note that

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

This implies

$$\lim_{\delta\to0}\mathbb{P}\left(\int_{K_\delta^c}|\xi_h|^2\mu(ds,dt)>\varepsilon\right)=0.$$

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

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

and the proposition is proved. □

2.9 Proof of Corollary 2.5.1

Proof. In the following we verify conditions (M1), (M2), (M3) in Theorem 2.3.1.

(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

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

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

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

(M3): Note that

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

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 $m < n$,

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left| \tilde{Z}_t - \hat{Z}_t \right|^k \leq \sum_{j=0}^{m} |\pi_j(\hat{\boldsymbol{\beta}})|^k \frac{1}{\sqrt{n}} \sum_{t=1}^{m} |X_{t-j}|^k + \sum_{j=m+1}^{\infty} |\pi_j(\hat{\boldsymbol{\beta}})|^k \frac{1}{\sqrt{n}} \sum_{t=1}^{n} |X_{t-j}|^k =: I_1 + I_2 \tag{2.14}$$

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

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

□

2.10 Proof of Corollary 2.6.1

Proof. In the following we verify conditions (M1), (M2), (M3) in Theorem 2.3.1.

(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

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

(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

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

Hence

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

(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

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

□

2.11 Conclusion

In this chapter, 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.

Chapter 3

Threshold selection for multivariate heavy-tailed data

3.1 Introduction

For multivariate heavy-tailed data, the principal objective is often to study dependence in the 'tail' of the distribution. To achieve this goal, the assumption of multivariate regular variation is typically used as a starting point. A random vector $\mathbf{X} \in \mathbb{R}^d$ is said to be *multivariate regularly varying* if the polar coordinates $(R, \mathbf{\Theta}) = (\|\mathbf{X}\|, \mathbf{X}/\|\mathbf{X}\|)$, where $\|\cdot\|$ is some norm, satisfy the conditions

(a) R is univariate regularly varying, i.e., $\mathbb{P}(R > r) = L(r)r^{-\alpha}$, where $L(\cdot)$ is a slowly varying function at infinity;

(b) $\mathbb{P}(\mathbf{\Theta} \in \cdot | R > r)$ converges weakly to a measure $S(\cdot)$ as $r \to \infty$.

The α is referred to as the index of the regular variation, while the S is called the angular distribution and characterizes the limiting tail dependence. There are other equivalent definitions of regular variation (Resnick, 2002), but this one is the most convenient for our purposes.

Given observations $\{\mathbf{X}_i\}_{i=1}^n$ and their corresponding polar coordinates $\{(R_i, \mathbf{\Theta}_i)\}_{i=1}^n$, a straightforward procedure for estimating S is to look at angular components of the data for which the radii are greater than a large threshold r_0, that is, $\mathbf{\Theta}_i$ for which $R_i > r_0$. In most studies, one takes r_0 to be a large empirical quantile of R. While there has been extensive research on choosing a threshold for which the distribution of R is regularly varying (i.e., limit condition (a)), little research has been devoted to ensuring the threshold is large enough for the independence of Θ and R to be reasonable (i.e., limit condition (b)). To this end, de Haan and de Ronde (1998) fit a parametric extreme value distribution model to each marginal and examined the parameter stability plot of each coordinate. The Stărică plot (Stărică, 1999) looked at the joint tail empirical measure, but was, in some way, equivalent to only examining the extremal behavior of R. Resnick (2007) suggested an automatic threshold selection from the Stărică plot but observed that the thresholds were sometimes systematically underestimated. In their study of the threshold based inference for parametric max-stable processes, Jeon and Smith (2014) suggested choosing the threshold by minimizing the MSE of the estimated parameters.

In this chapter, we propose an algorithm which selects the threshold for modeling S. Our motivation is the implied property that $(R, \boldsymbol{\Theta})$ given $R > r$ become independent as $r \to \infty$. Given a sequence of candidate threshold levels, we test the degree of dependence between R and $\boldsymbol{\Theta}$ for the truncated data above each level. The dependence measure we use is the distance covariance introduced by Székely et al. (2007). This measure has the ability to account for various types of dependence and to be applicable to data in higher dimensions. The resulting test statistics are given in the form of p-values and are compared across all levels through a subsampling scheme. This enables us to extract more information from the test statistics while not overloading the computational burden.

The remainder of the chapter is organized as follows. We first provide some theoretical background on multivariate regular variation in Section 3.2. The distance covariance and its theoretical properties are introduced in Section 3.3. Applying this dependence measure in our conditioning setting, we propose a test statistic and prove relevant theoretical results in Section 3.4. Our proposed algorithm for threshold selection is presented in Section 3.5, and illustrated on simulated and real examples in Section 3.6. The chapter concludes with a discussion.

3.2 Multivariate regular variation and problem set-up

One way to approach multivariate heavy-tailed data is through the notion of multivariate regular variation. For a detailed review, see, for example, Chapter 6 of Resnick (2007). Let $\mathbf{X} = (X_1, \ldots, X_d)$ be a d-dimensional random variable defined on the cone $\mathbb{R}^d_+ = [\mathbf{0}, \infty)\backslash\{\mathbf{0}\}$. Define the polar coordinate transformation

$$T(\mathbf{X}) = (\|\mathbf{X}\|, \mathbf{X}/\|\mathbf{X}\|) =: (R, \boldsymbol{\Theta}), \tag{3.1}$$

where $\|\cdot\|$ denotes some norm. Then $\mathbf{X}$ is regularly varying if and only if there exists a probability measure $S(\cdot)$ on $\mathbb{S}^{d-1}$, the unit sphere in $\mathbb{R}^d$, and a function $b(t) \to \infty$, such that

$$t\mathbb{P}\left[(R/b(t), \boldsymbol{\Theta}) \in \cdot\right] \xrightarrow{v} \nu_\alpha \times S, \quad t \to \infty, \qquad \text{on } (0,\infty) \times \mathbb{S}^{d-1}, \tag{3.2}$$

where $\xrightarrow{v}$ denotes vague convergence, and ν_α is a measure defined on $(0, \infty]$ such that

$$\nu_\alpha(x, \infty] = x^{-\alpha}, \quad x > 0.$$

Here $b(t)$ can be chosen as the $1 - t^{-1}$-quantile, i.e.,

$$b(t) = \inf\{s | \mathbb{P}(R \leq s) \geq 1 - t^{-1}\}.$$

The convergence (3.2) implies that

$$\mathbb{P}\left[\left(\frac{R}{r}, \boldsymbol{\Theta}\right) \in \cdot \,\middle|\, R > r\right] \xrightarrow{w} \nu_\alpha \times S, \quad r \to \infty, \qquad \text{on } [1,\infty) \times \mathbb{S}^{d-1}, \tag{3.3}$$

where $\xrightarrow{w}$ denotes weak convergence. In other words, given that $R > r$ for r large, the conditional distribution of R/r and $\boldsymbol{\Theta}$ are independent in the limit. In view of (3.3), we restrict the measure ν_α to $[1, \infty)$ throughout the remainder of the chapter. The angular measure S characterizes the tail dependence structure of $\mathbf{X}$. If S is concentrated on $\{e_i, i = 1, \ldots d\}$, where $e_i = (0, \ldots, 0, 1, 0, \ldots, 0)$, then the components of $\mathbf{X}$ are asymptotically independent in the tail, a case known as asymptotic independence. If S has mass lying in the subspace $\{(t_1, \ldots, t_d) \in \mathbb{S}^{d-1} | t_i > 0, t_j > 0, i \neq j\}$, then an extreme observation in the X_i direction implicates a positive probability of an extreme observation in the X_j direction, a case known as asymptotic dependence. Hence the estimation of S from observations is an important problem, and often the primary goal, in multivariate heavy-tailed modeling.

The following convergence is implied from (3.3):

$$\mathbb{P}(\boldsymbol{\Theta} \in \cdot | R > r) \stackrel{w}{\to} S(\cdot), \quad r \to \infty. \tag{3.4}$$

This suggests estimating S using the angular data $(\boldsymbol{\Theta}_i)$ whose radial parts satisfy $R_i > r_0$ for r_0 large. The motivation behind our method is to seek r_0 such that when $R > r_0$, R and $\boldsymbol{\Theta}$ are virtually independent. Given a candidate threshold sequence $\{r_k\}$, we formally test the independence between $(R_i, \boldsymbol{\Theta}_i)$ among the observations satisfying $R_i > r_k$. The use of Pearson's correlation as the dependence measure is unsuitable in this case, for two reasons. First, correlation is only applicable to univariate random variables, whereas $\boldsymbol{\Theta}$ lies on the sphere of dimension $d-1$. Second, correlation only describes the linear relationship between two random variables, thus having zero correlation is not a sufficient condition for independence. Instead, we use a more powerful dependence measure, the *distance covariance*, which is introduced in the next section.

3.3 Distance covariance

In this section, we briefly review the definition and some properties of the distance covariance. More detailed descriptions and proofs can be found in Székely et al. (2007) and Davis et al. (2018).

Let $X \in \mathbb{R}^p$ and $Y \in \mathbb{R}^q$ be two random vectors, then the distance covariance between X and Y is defined as

$$T(X, Y; \mu) = \int_{\mathbb{R}^{p+q}} \left|\varphi_{X,Y}(s,t) - \varphi_X(s)\,\varphi_Y(t)\right|^2 \mu(ds, dt)\,, \quad (s,t) \in \mathbb{R}^{p+q}, \tag{3.5}$$

where $\varphi_{X,Y}(s,t), \varphi_X(s), \varphi_Y(t)$ denote the joint and marginal characteristic functions of (X, Y) and 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 throughout the paper (Davis et al., 2018):

1. μ is a finite measure on $\mathbb{R}^{p+q}$;
2. μ is an infinite measure on $\mathbb{R}^{p+q}$ such that

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

and

$$\mathbb{E}[|XY|^\alpha + |X|^\alpha + |Y|^\alpha] < \infty$$

for some $\alpha \in (0, 2]$.

One advantage of distance covariance over, say, Pearson's covariance, is that, if μ has a positive Lebesgue density on $\mathbb{R}^{p+q}$, then X and Y are independent if and only if $T(X, Y; \mu) = 0$. Another attractive property of this dependence measure is that it readily applies to random vectors of different dimensions.

To estimate $T(X, Y; \mu)$ from observations $(X_1, Y_1), \ldots, (X_n, Y_n)$, define the empirical distance covariance

$$T_n(X, Y; \mu) = \int_{\mathbb{R}^{p+q}} \left|\hat{\varphi}_{X,Y}(s, t) - \hat{\varphi}_X(s)\, \hat{\varphi}_Y(t)\right|^2 \mu(ds, dt)\,,$$

where $\hat{\varphi}_{X,Y}(s, t) = \frac{1}{n} \sum_{j=1}^n e^{i\,\langle s, X_j\rangle + i\,\langle t, Y_j\rangle}$ and $\hat{\varphi}_X(s) = \hat{\varphi}_{X,Y}(s, 0), \hat{\varphi}_Y(t) = \hat{\varphi}_{X,Y}(0, t)$ are the respective empirical characteristic functions. If we assume that $\mu = \mu_1 \times \mu_2$ and is symmetric about the origin, then under the conditions where $T(X, Y; \mu)$ exists, $T_n(X, Y; \mu)$ also has the computable form

$$\begin{aligned} T_n(X, Y; \mu) &= \frac{1}{n^2} \sum_{i,j=1}^n \tilde{\mu}_1(X_i - X_j)\, \tilde{\mu}_2(Y_i - Y_j) \\ &+ \frac{1}{n^4} \sum_{i,j,k,l=1}^n \tilde{\mu}_1(X_i - X_j)\tilde{\mu}_2(Y_k - Y_l) - \frac{2}{n^3} \sum_{i,j,k=1}^n \tilde{\mu}_1(X_i - X_j)\tilde{\mu}_2(Y_i - Y_k), \end{aligned}$$

where $\tilde{\mu}(x) = \int (1 - \cos\langle s, x\rangle)\, \mu(ds)$ (Davis et al., 2018).

The most popular choice of μ, first mentioned by Feuerverger (1993) and then more extensively studied by Székely et al. (2007), is

$$\mu(ds, dt) = c_{p,q}|s|^{-\kappa-p}|t|^{-\kappa-q} ds\, dt\,. \tag{3.6}$$

where $c_{p,q}$ is as defined in Lemma 1 of Székely et al. (2007). This choice of μ gives $\tilde{\mu}(x)\tilde{\mu}(y) = |x|^\kappa |y|^\kappa$. Moreover, this is the only choice of μ for which the distance covariance is invariant relative to scale and orthogonal transformations. Note that in order for the integral (3.5) to exist, it is required that

$$E[|X|^\kappa |Y|^\kappa + |X|^\kappa + |Y|^\kappa] < \infty.$$

We will utilize the described weight measure (3.6) with $\kappa = 1$ in our simulations and data analyses in Section 3.6, but applied to the log transformation on R to ensure that the moment condition is satisfied.

As detailed in Davis et al. (2018), if the sequence $\{(X_i, Y_i)\}$ is stationary and ergodic, then

$$T_n(X, Y; \mu) \xrightarrow{a.s.} T(X, Y; \mu). \tag{3.7}$$

Further, if X and Y are independent, then under an α-mixing condition,

$$n\, T_n(X, Y; \mu) \xrightarrow{d} \int_{\mathbb{R}^{p+q}} |G(s,t)|^2 \mu(s,t) \tag{3.8}$$

for some centered Gaussian field G. On the other hand, if X and Y are dependent, then

$$\sqrt{n}(T_n(X, Y; \mu) - T(X, Y; \mu)) \xrightarrow{d} G'_\mu$$

for some non-trivial limit G'_μ, implying that $n\, T_n(X, Y; \mu)$ diverges as $n \to \infty$. Naturally one can devise a test of independence between X and Y using the statistic $n\, T_n(X, Y; \mu)$: the null hypothesis of independence is rejected at level χ if $n\, T_n(X, Y; \mu) > c$, where c is the upper χ-quantile of $\int_{\mathbb{R}^{p+q}} |G(s,t)|^2 \mu(s,t)$.

In practice, the distribution $\int_{\mathbb{R}^{p+q}} |G(s,t)|^2 \mu(s,t)$ is intractable and is typically approximated through bootstrap. Hence the main drawback of using distance covariance is the computation burden it brings for large sample size: the computation of a single distance covariance statistic requires $O(n^2)$ operations, while finding the cut-off values via resampling requires much more additional computation. Our method, however, overcomes this problem through subsampling the data, as will be described in Section 3.5.

3.4 Theoretical results

Let $\{\mathbf{X}_i\}_{i=1}^n$ be iid observations in $\mathbb{R}^d$ from a multivariate regularly varying distribution $\mathbf{X}$ satisfying (3.1) and (3.3), and $\{(R_i, \mathbf{\Theta}_i)\}_{i=1}^n$ be their polar coordinate transformations. Given a threshold r_n, we measure the dependence between R/r_n and $\mathbf{\Theta}$ conditional on $R > r_n$ by the empirical distance covariance of $(R/r_n, \mathbf{\Theta})$ on the set $\{R > r_n\}$:

$$T_n := \int_{\mathbb{R}^{d+1}} |C_n(s,t)|^2 \mu(ds, dt), \tag{3.9}$$

with

$$C_n(s,t) := \hat{\varphi}_{\frac{R}{r_n},\Theta|r_n}(s,t) - \hat{\varphi}_{\frac{R}{r_n}|r_n}(s)\hat{\varphi}_{\Theta|r_n}(t),$$

where $\hat{\varphi}_{\frac{R}{r_n},\Theta|r_n}$ is the conditional empirical characteristic function of $(R/r_n, \mathbf{\Theta})$,

$$\hat{\varphi}_{\frac{R}{r_n},\Theta|r_n}(s,t) = \frac{1}{\sum_{j=1}^n \mathbf{1}_{\{R_j > r_n\}}} \sum_{j=1}^n e^{isR_j/r_n + it^T \mathbf{\Theta}_j} \mathbf{1}_{\{R_j > r_n\}}, \ s \in \mathbb{R}, \ \ t = (t_1, \ldots, t_d)^T \in \mathbb{R}^d,$$

and $\hat{\varphi}_{\frac{R}{r_n}|r_n}, \hat{\varphi}_{\Theta|r_n}$ are the corresponding empirical conditional marginal characteristic functions,

$$\hat{\varphi}_{\frac{R}{r_n}|r_n}(s) = \hat{\varphi}_{\frac{R}{r_n},\Theta|r_n}(s,0), \qquad \hat{\varphi}_{\Theta|r_n}(t) = \hat{\varphi}_{\frac{R}{r_n},\Theta|r_n}(0,t).$$

In this section, we establish the limiting results (3.7) and (3.8) adapted to the conditional distance covariance. For ease of notation, let

$$p_n := \mathbb{P}(R > r_n) \ , \qquad \hat{p}_n := \frac{1}{n} \sum_{j=1}^n \mathbf{1}_{\{R_j > r_n\}}$$

be the theoretical and empirical probability of exceedance, and let

$$\varphi_{\frac{R}{r_n},\Theta|r_n}(s,t) := \mathbb{E}\left[e^{isR/r_n + it^T \mathbf{\Theta}} | R > r_n\right] = \frac{\mathbb{E}\left[e^{isR/r_n + it^T \mathbf{\Theta}} \mathbf{1}_{R > r_n}\right]}{p_n},$$

and

$$\varphi_{\frac{R}{r_n}|r_n}(s) := \varphi_{\frac{R}{r_n},\Theta|r_n}(s,0), \qquad \varphi_{\Theta|r_n}(t) := \varphi_{\frac{R}{r_n},\Theta|r_n}(0,t),$$

be the theoretical conditional joint and marginal characteristic functions.

Recall from (3.3) that as $n \to \infty$, R/r_n and $\mathbf{\Theta}$ become asymptotically independent and converge to ν_α and S respectively. Denote the characteristic functions of the corresponding limit distributions by

$$\begin{aligned} \varphi_R(s) &:= \int_1^\infty \exp(isr)\alpha r^{-\alpha-1}dr = \lim_{n\to\infty} \varphi_{\frac{R}{r_n}|r_n}(s), & (3.10) \\ \varphi_\Theta(t) &:= \int_{\mathbb{S}^{d-1}} \exp(it\theta)S(d\theta) = \lim_{n\to\infty} \varphi_{\Theta|r_n}(t). & (3.11) \end{aligned}$$

We have the following results.

Theorem 3.4.1. *1. Let $\mathbf{X}_1, \ldots, \mathbf{X}_n$ be iid observations generated from $\mathbf{X}$, where $\mathbf{X}$ is multivariate regularly varying with index $\alpha > 1$. Let T_n be the conditional empirical distance covariance between the angular and radial component defined in* (3.9)*. Further assume that $np_n \to \infty$ and the weight measure μ satisfies*

$$\int_{\mathbb{R}^{d+1}} (1 \wedge |s|^\beta)(1 \wedge |t|^2)\mu(ds, dt) < \infty, \tag{3.12}$$

for some $1 < \beta < 2 \wedge \alpha$. Then

$$T_n \xrightarrow{p} 0.$$

2. In addition, if $\{r_n\}$ satisfies

$$np_n \int_{\mathbb{R}^{d+1}} |\varphi_{\frac{R}{r_n},\Theta|r_n}(s,t) - \varphi_{\frac{R}{r_n}|r_n}(s)\varphi_{\Theta|r_n}(t)|^2 \mu(ds, dt) \to 0, \tag{3.13}$$

then

$$n\hat{p}_n T_n \xrightarrow{d} \int_{\mathbb{R}^{d+1}} |Q(s,t)|^2 \mu(ds, dt), \tag{3.14}$$

where Q is a centered Gaussian process with covariance function

$$cov(Q(s,t), Q(s',t')) = (\varphi_R(s-s') - \varphi_R(s)\varphi_R(-s'))(\varphi_\Theta(t-t') - \varphi_\Theta(t)\varphi_\Theta(-t')) \tag{3.15}$$

with $\varphi_R, \varphi_\Theta$ as defined in (3.10) *and* (3.11)*.*

Remark 3.4.2. *In the case where $\mathbf{X}$ is regularly varying with index $\alpha \leq 1$, similar results hold if we replace R/r_n with $\log(R/r_n)$ for which all moments exist.*

The proof of the theorem is delayed to Section 3.8. In the following remark, we discuss certain sufficient conditions for assumption (3.13).

Remark 3.4.3. *Assume that $\mu = \mu_1 \times \mu_2$, where μ_1, μ_2 are measures on $\mathbb{R}$ and $\mathbb{R}^d$, respectively, and symmetric about the origin. From Section 2.2 of Davis et al. (2018), condition* (3.13) *is equivalent to*

$$\begin{aligned} np_n \Bigg(& \mathbb{E}[\tilde{\mu}_1(\frac{R}{r_n} - \frac{R'}{r_n})\, \tilde{\mu}_2(\mathbf{\Theta} - \mathbf{\Theta}')|R, R' > r_n] \\ & + \mathbb{E}[\tilde{\mu}_1(\frac{R}{r_n} - \frac{R'}{r_n})|R, R' > r_n]\, \mathbb{E}[\tilde{\mu}_2(\mathbf{\Theta} - \mathbf{\Theta}')|R, R' > r_n] \\ & - 2\, \mathbb{E}[\tilde{\mu}_1(\frac{R}{r_n} - \frac{R'}{r_n})\tilde{\mu}_2(\mathbf{\Theta} - \mathbf{\Theta}'')|R, R' > r_n] \Bigg) \to 0, \end{aligned} \tag{3.16}$$

where

$$\tilde{\mu}_i(x) = \int (1 - \cos(x^T s))\, \mu_i(ds), \qquad i = 1, 2.$$

Let $P_{\frac{R}{r_n},\Theta|r_n}$ denote the conditional joint distribution of $(R/r_n, \Theta)$ given $R > r_n$ and $P_{\frac{R}{r_n}|r_n}$, $P_{\Theta|r_n}$ be the respective conditional marginals. Then (3.16) *can be expressed as*

$$\begin{aligned} & np_n \int_{(1,\infty)\times\mathbb{S}^{d-1}} \int_{(1,\infty)\times\mathbb{S}^{d-1}} \tilde{\mu}_1(T - T')\, \tilde{\mu}_2(\mathbf{\Theta} - \mathbf{\Theta}') \\ & \qquad \Big(P_{\frac{R}{r_n},\Theta|r_n}(dT, d\mathbf{\Theta}) P_{\frac{R}{r_n},\Theta|r_n}(dT', d\mathbf{\Theta}') \\ & \qquad + P_{\frac{R}{r_n}|r_n}(dT) P_{\Theta|r_n}(d\mathbf{\Theta}) P_{\frac{R}{r_n}|r_n}(dT') P_{\Theta|r_n}(d\mathbf{\Theta}') \\ & \qquad - 2 P_{\frac{R}{r_n},\Theta|r_n}(dT, d\mathbf{\Theta}) P_{\frac{R}{r_n}|r_n}(dT') P_{\Theta|r_n}(d\mathbf{\Theta}') \Big) \\ = & \int_{(1,\infty)\times\mathbb{S}^{d-1}} \int_{(1,\infty)\times\mathbb{S}^{d-1}} \tilde{\mu}_1(T - T')\, \tilde{\mu}_2(\mathbf{\Theta} - \mathbf{\Theta}') \\ & \qquad \sqrt{np_n} \Big(P_{\frac{R}{r_n},\Theta|r_n}(dT, d\mathbf{\Theta}) - P_{\frac{R}{r_n}|r_n}(dT) P_{\Theta|r_n}(d\mathbf{\Theta}) \Big) \\ & \qquad \sqrt{np_n} \Big(P_{\frac{R}{r_n},\Theta|r_n}(dT', d\mathbf{\Theta}') - P_{\frac{R}{r_n}|r_n}(dT') P_{\Theta|r_n}(d\mathbf{\Theta}') \Big) \\ \to & \ 0, \end{aligned} \tag{3.17}$$

where $(R', \mathbf{\Theta}')$, $(R'', \mathbf{\Theta}'')$ are iid copies of $(R, \mathbf{\Theta})$. One way to verify (3.17) *is to assume a second-order like condition on the distribution of $(R, \mathbf{\Theta})$. For example, assume that*

$$\frac{P_{\frac{R}{r_n},\Theta|r_n} - \nu_\alpha \times S}{A(r_n)} \xrightarrow{w} \chi, \qquad on\ [1, \infty) \times \mathbb{S}^{d-1},$$

where χ is a signed measure such that $\chi([r,\infty] \times B)$ is finite for all $r \geq 1$ and B Borel set in $\mathbb{S}^{d-1}$, the unit sphere in $\mathbb{R}^d$, and the scalar function $A(t) \to 0$ as $t \to \infty$. When the components of $\mathbf{X}$ are asymptotically independent, this is equivalent to the second order condition for multivariate regular variation (Resnick, 2002). If we choose the sequence r_n such that $\sqrt{np_n} \to \infty$ and $\sqrt{np_n}A(r_n) \to 0$, then

$$\begin{aligned}
&\sqrt{np_n}A(r_n)\,\frac{P_{\frac{R}{r_n},\Theta|r_n}((\cdot,\cdot)) - P_{\frac{R}{r_n}|r_n}(\cdot)\times P_{\Theta|r_n}(\cdot)}{A(r_n)} \\
=\;& \sqrt{np_n}A(r_n)\left(\frac{P_{\frac{R}{r_n},\Theta|r_n}((\cdot,\cdot)) - \nu_\alpha \times S((\cdot,\cdot))}{A(r_n)} - \frac{\left(P_{\frac{R}{r_n}|r_n}(\cdot) - \nu_\alpha(\cdot)\right)\times P_{\Theta|r_n}(\cdot)}{A(r_n)}\right. \\
&\left. - \frac{\nu_\alpha(\cdot)\times(P_{\Theta|r_n}(\cdot) - S(\cdot))}{A(r_n)}\right) \xrightarrow{w} 0
\end{aligned}$$

on $[1,\infty] \times \mathbb{S}^{d-1}$. In the case where μ_1, μ_2 are finite measures, $\tilde{\mu}_1, \tilde{\mu}_2$ are bounded and (3.17) *is satisfied since the integrand can be written as*

$$\begin{aligned}
np_nA^2(r_n)\int_{(1,\infty)\times\mathbb{S}^{d-1}} \Bigg[\int_{(1,\infty)\times\mathbb{S}^{d-1}} &\tilde{\mu}_1(T-T')\,\tilde{\mu}_2(\mathbf{\Theta}-\mathbf{\Theta}') \\
&\frac{P_{\frac{R}{r_n},\Theta|r_n}(dT,d\mathbf{\Theta}) - P_{\frac{R}{r_n}|r_n}(dT)P_{\Theta|r_n}(d\mathbf{\Theta})}{A(r_n)}\Bigg] \\
&\frac{P_{\frac{R}{r_n},\Theta|r_n}(dT',d\mathbf{\Theta}') - P_{\frac{R}{r_n}|r_n}(dT')P_{\Theta|r_n}(d\mathbf{\Theta}')}{A(r_n)} \;\to\; 0.
\end{aligned}$$

In the special case that $|A| \in RV_\rho$ for $\rho < 0$, (3.17) *is met provided r_n is chosen such that*

$$O(n^{\frac{1}{\alpha+2|\rho|}+\epsilon}) \leq r_n \leq o(n^{\frac{1}{\alpha}}), \quad \textit{for some } \epsilon > 0.$$

When the measures μ_1, μ_2 are infinite, (3.13) can be verified in specific cases. This is illustrated in the following example.

Example 3.4.4. *Let $\mathbf{X}$ follow a bivariate logistic distribution, i.e., $\mathbf{X}$ has cdf*

$$\mathbb{P}(X_1 < x_1, X_2 < x_2) = \exp(-(x_1^{-1/\gamma} + x_2^{-1/\gamma})^\gamma), \quad \gamma \in (0,1). \tag{3.18}$$

Then $\mathbf{X}$ has asymptotically independent components if and only if $\gamma = 1$. It can be shown that $\mathbf{X}$ is regularly varying with index $\alpha = 1$, i.e., $p_n = \mathbb{P}(R > r_n) \sim r_n^{-1}$ as $r_n \to \infty$.

Using the L_1-norm, $\|(x_1,x_2)\| = |x_1|+|x_2|$, the polar coordinate transform is $(R,\Theta) = (X_1+X_2, X_1/(X_1+X_2)) \in (0,\infty)\times[0,1]$ and the pdf of (R,Θ) is

$$\begin{aligned} f_{R,\Theta}(r,\theta) &= r^{-2}\left(\theta(1-\theta)\right)^{-\frac{\gamma+1}{\gamma}}\left(\theta^{-\frac{1}{\gamma}}+(1-\theta)^{-\frac{1}{\gamma}}\right)^{\gamma-2} e^{-r^{-1}\left(\theta^{-\frac{1}{\gamma}}+(1-\theta)^{-\frac{1}{\gamma}}\right)^{\gamma}} \\ &\quad \left(r^{-1}\left(\theta^{-\frac{1}{\gamma}}+(1-\theta)^{-\frac{1}{\gamma}}\right)^{\gamma} - \frac{\gamma-1}{\gamma}\right). \end{aligned}$$

We now consider the case of the infinite weight measure μ given in (3.6) with $\kappa = 1$ and derive the condition on the sequence $\{r_k\}$ for which the conditions of Theorem 3.4.1 hold. First observe that

$$\begin{aligned} f_{\frac{R}{r_n},\Theta|r_n}(t,\theta) &= t^{-2}\left(\theta(1-\theta)\right)^{-\frac{\gamma+1}{\gamma}}\left(\theta^{-\frac{1}{\gamma}}+(1-\theta)^{-\frac{1}{\gamma}}\right)^{\gamma-2} e^{-r_n^{-1}t^{-1}\left(\theta^{-\frac{1}{\gamma}}+(1-\theta)^{-\frac{1}{\gamma}}\right)^{\gamma}} \\ &\quad \left(r_n^{-1}t^{-1}\left(\theta^{-\frac{1}{\gamma}}+(1-\theta)^{-\frac{1}{\gamma}}\right)^{\gamma} - \frac{\gamma-1}{\gamma}\right) \\ &\to t^{-2}\frac{1-\gamma}{\gamma}\left(\theta(1-\theta)\right)^{-\frac{\gamma+1}{\gamma}}\left(\theta^{-\frac{1}{\gamma}}+(1-\theta)^{-\frac{1}{\gamma}}\right)^{\gamma-2}, \quad \text{as } n\to\infty, \qquad (3.19) \\ &=: f_T(t)f_\Theta(\theta), \end{aligned}$$

and

$$\begin{aligned} & r_n\left|f_{\frac{R}{r_n},\Theta|r_n}(t,\theta) - f_T(t)f_\Theta(\theta)\right| \\ \le\ & f_T(t)f_\Theta(\theta)\left(r_n\left|e^{-r_n^{-1}t^{-1}\left(\theta^{-\frac{1}{\gamma}}+(1-\theta)^{-\frac{1}{\gamma}}\right)^{\gamma}}-1\right|\right. \\ & \left. +e^{-r_n^{-1}t^{-1}\left(\theta^{-\frac{1}{\gamma}}+(1-\theta)^{-\frac{1}{\gamma}}\right)^{\gamma}} t^{-1}\left(\theta^{-\frac{1}{\gamma}}+(1-\theta)^{-\frac{1}{\gamma}}\right)^{\gamma}\frac{\gamma}{1-\gamma}\right) \\ \le\ & f_T(t)f_\Theta(\theta)\left(t^{-1}\left(\theta^{-\frac{1}{\gamma}}+(1-\theta)^{-\frac{1}{\gamma}}\right)^{\gamma} + t^{-1}\left(\theta^{-\frac{1}{\gamma}}+(1-\theta)^{-\frac{1}{\gamma}}\right)^{\gamma}\frac{\gamma}{1-\gamma}\right) \\ \le\ & t^{-3}\left(\theta^{-\frac{1}{\gamma}}+(1-\theta)^{-\frac{1}{\gamma}}\right)^{2\gamma-2}\frac{1}{1-\gamma} \\ \le\ & ct^{-3}, \quad \text{for } t\ge 1 \text{ and } \theta\in[0,1], \end{aligned}$$

where c denotes a generic a constant whose value may change from line to line throughout the proof, and the last inequality comes from the facts that

$$\theta(1-\theta) \le \frac{1}{4} \quad \text{and} \quad \left(\theta^{-\frac{1}{\gamma}}+(1-\theta)^{-\frac{1}{\gamma}}\right)^{2\gamma-2} \le \left(\frac{1}{2}\right)^{\frac{2-2\gamma}{\gamma}} < \infty.$$

Letting

$$h_n(t,\theta) \;:=\; \frac{f_{\frac{R}{r_n},\Theta|r_n}(t,\theta) - f_{\frac{R}{r_n}|r_n}(t)f_{\Theta|r_n}(\theta)}{r_n^{-1}},$$

we have

$$\begin{aligned}
&\max\left(\int_0^1\int_1^\infty |h_n(t,\theta)|dtd\theta, \int_0^1\int_1^\infty |\log(t)h_n(t,\theta)|dtd\theta\right)\\
&\le \int_0^1\int_1^\infty |th_n(t,\theta)|dtd\theta\\
&\le \int_0^1\int_1^\infty \left|\frac{f_{\frac{R}{r_n},\Theta|r_n}(t,\theta) - f_T(t)f_\Theta(\theta)}{t^{-1}r_n^{-1}}\right| dtd\theta\\
&\quad + \int_0^1\int_1^\infty \left|f_T(t)\frac{f_{\Theta|r_n}(\theta)-f_\Theta(\theta)}{t^{-1}r_n^{-1}}\right| dtd\theta\\
&\quad + \int_0^1\int_1^\infty \left|f_{\Theta|r_n}(\theta)\frac{f_{\frac{R}{r_n}|r_n}(t)-f_T(t)}{t^{-1}r_n^{-1}}\right| dtd\theta,
\end{aligned}$$

where the first term can be bounded by

$$\int_0^1\int_1^\infty ct^{-2}dtd\theta < \infty,$$

and the other terms can be bounded in the same way. Since R has infinite first moment, we apply the distance correlation to $\log R$ and $\mathbf{\Theta}$. The integral in (3.17) is bounded by

$$\begin{aligned}
&\frac{np_n}{r_n^2}\int_0^1\int_1^\infty\int_0^1\int_1^\infty |\log t - \log t'||\theta-\theta'||h_n(t,\theta)||h_n(t',\theta')|dtd\theta dt'd\theta'\\
\le\;& c\frac{n}{r_n^3}\int_0^1\int_1^\infty\int_0^1\int_1^\infty (|\log t| + |\log t'|)|h_n(t,\theta)||h_n(t',\theta')|dtd\theta dt'd\theta'\\
\le\;& c\frac{n}{r_n^3}\left(\int_0^1\int_1^\infty |\log(t)h_n(t,\theta)|dtd\theta\right)\left(\int_0^1\int_1^\infty |h_n(t,\theta)|dtd\theta\right) \le c\frac{n}{r_n^3},
\end{aligned}$$

which converges to zero if $n = o(r_n^3)$. Therefore if $\{r_n\}$ is chosen such that $r_n = o(n)$ and $n = o(r_n^3)$, then Theorem 3.4.1 holds.

The result in Theorem 3.4.1 can be generalized from iid to a regularly varying time series setting, which we present in the next theorem. For a multivariate stationary time series $\{\mathbf{X}_t\}$ and $h \ge 1$, set $\mathbf{Y}_h = (\mathbf{X}_0, \ldots, \mathbf{X}_h)$. Then $\{\mathbf{X}_t\}$ is regularly varying if

$$\frac{\mathbb{P}(x^{-1}\mathbf{Y}_h \in \cdot)}{\mathbb{P}(x^{-1}\|\mathbf{X}_0\| > 1)} \stackrel{v}{\to} \mu_h^*(\cdot), \quad x\to\infty,$$

for some non-null measure μ_h^* on $\overline{\mathbb{R}}_0^{(h+1)d} = \overline{\mathbb{R}}^{(h+1)d} \backslash \{\mathbf{0}\}$, $\overline{\mathbb{R}} = \mathbb{R} \cup \{\pm\infty\}$, with the property that $\mu_h^*(tC) = t^{-\alpha}\mu_h^*(C)$ for any $t > 0$ and Borel set $C \subset \overline{\mathbb{R}}_0^{(h+1)d}$. See, for example, page 979 of Davis and Mikosch (2009). It follows easily that

$$\frac{\mathbb{P}(x^{-1}(\mathbf{X}_0, \mathbf{X}_h) \in \cdot)}{\mathbb{P}(\|\mathbf{X}_0\| > x)} \xrightarrow{v} \mu_h(\cdot), \tag{3.20}$$

where

$$\mu_h(D) = C \cdot \mu_h^*(\{\mathbf{s} \in \overline{\mathbb{R}}^{(h+1)d} : (\mathbf{s}_1, \mathbf{s}_h) \in D\}).$$

Assume that $\{\mathbf{X}_t\}$ is α-mixing. We assume the following conditions between $\{\mathbf{X}_t\}$ and the sequence of threshold $\{r_n\}$, which can be verified for various time series models (Davis and Mikosch, 2009).

(**M**) Assume $p_n^{-1} = \mathbb{P}^{-1}(\|\mathbf{X}_1\| > r_n) = o(n^{1/3})$ and that there exists a sequence $\{l_n\}$ such that $l_n \to \infty$, $l_n p_n \to 0$, and

i)

$$\left(\frac{1}{p_n}\right)^{\delta} \sum_{h=l_n}^{\infty} \alpha_h^{\delta} \to 0 \text{ for some } \delta \in (0,1); \tag{3.21}$$

ii)

$$\lim_{h\to\infty} \limsup_{n\to\infty} \frac{1}{p_n} \sum_{j=h}^{l_n} \mathbb{P}(\|\mathbf{X}_0\| > r_n, \|\mathbf{X}_j\| > r_n) = 0; \tag{3.22}$$

iii)

$$n p_n \alpha_{l_n} \to 0. \tag{3.23}$$

Theorem 3.4.5. *Let $\{\mathbf{X}_t\}$ be a multivariate regularly varying time series with tail index $\alpha > 1$ and α-mixing with coefficients $\{\alpha_h\}_{h\geq 0}$. Assume the same conditions for the weight measure μ and the sequence of thresholds $\{r_n\}$ in Theorem 3.4.1, i.e.,* (3.12), (3.13) *hold, and that condition (**M**) holds. Then*

$$n\hat{p}_n T_n \xrightarrow{d} \int_{\mathbb{R}^{d+1}} |Q'(s,t)|^2 \mu(ds, dt),$$

where Q' is a centered Gaussian process. In particular,

$$T_n \xrightarrow{p} 0.$$

The proof of Theorem 3.4.5 is given in Section 3.9.

Note that the limiting distributions Q in Theorem 3.4.1 and Q' in Theorem 3.4.5 are both intractable. In practice, quantiles of the distributions are calculated using resampling methods. While in the iid case this can be done straightforwardly, in the weakly dependent case one needs to apply the block bootstrap or stationary bootstrap to obtain the desired result (see Davis et al. (2012)). In the following section, we present a threshold selection framework with a subsampling scheme that does not require independence between the observations.

3.5 Threshold selection

In this section, we propose a procedure to select the threshold for estimating the spectral measure S from observations $\mathbf{X}_1, \cdots, \mathbf{X}_n$. Let us first consider the case where a specific threshold r_n is given. Then (3.9) specifies the empirical distance covariance between R/r_n and $\mathbf{\Theta}$ conditional on $R > r_n$. Under the assumption (3.13), we have from Theorem 3.4.1,

$$n\hat{p}_n T_n \to \int_{\mathbb{R}^{d+1}} |Q(s,t)|^2 \mu(ds, dt),$$

where $n\hat{p}_n$ is the number of observations such that $R_i > r_n$. In practice, the limit distribution $\int |Q|^2 \mu(s,t)$ is intractable, but one can resort to bootstrapping. Consider the hypothesis testing framework:

$$\begin{aligned} H_0 &: \quad R/r_n \text{ and } \mathbf{\Theta} \text{ are independent with respect to } \mathbb{P}[\cdot|R > r_n]; \\ H_1 &: \quad R/r_n \text{ and } \mathbf{\Theta} \text{ are not independent with respect to } \mathbb{P}[\cdot|R > r_n]. \end{aligned}$$

Define the p-value for testing H_0 versus H_1 to be

$$pv = \mathbb{P}\left(\int_{\mathbb{R}^{d+1}} |Q(s,t)|^2 \mu(ds, dt) > u \right)\Bigg|_{u = n\hat{p}_n T_n}. \tag{3.24}$$

Under H_0, pv follows $U(0,1)$. Under H_1, $n\hat{p}_n T_n$ diverges and pv should be sufficiently small.

Now consider a decreasing sequence of candidate thresholds $\{r_k\}$. From (3.24), a sequence of p-values $\{pv_k\}$, each corresponding to a threshold r_k, can be obtained. Our goal is to find the smallest threshold r^* such that conditional on $R > r^*$, Θ can reasonably be considered independent of R. Note that the pv_k's are not independent for each k since they are computed from the same set of data. Conventional multiple testing procedures, such as Bonferroni correction, are problematic to implement for dependent p-values. To counter these limitations, we propose an intuitive and direct method based on subsampling.

The idea is outlined as follows: For a fixed level r_k, we choose a subsample of size n_k from the conditional empirical cdf $\hat{F}_{\frac{R}{r_n},\Theta|r_k}$ of $(R_i/r_k, \Theta_i)$ with $R_i > r_k$, $i = 1, \ldots, n$. For this subsample, we compute the distance covariance $T_{n,k}$. To compute a p-value of $T_{n,k}$ under the assumption that the conditional empirical distribution is a product of the conditional marginals, we take a large number (L) of subsamples of size n_k from

$$\tilde{F}_{\frac{R}{r_n},\Theta|r_k}(d\theta, dr) = \hat{F}_{\Theta|r_k}(d\theta)\hat{F}_{\frac{R}{r_n}|r_k}(dr),$$

and calculate the value $\tilde{T}^{(l)}_{n,k}, l = 1, \ldots, L$ for each subsample. The p-value of $T_{n,k}$, pv_k, is then the empirical p-value of $T_{n,k}$ relative the $\{\tilde{T}^{(l)}_{n,k}\}_{l=1,\ldots,L}$. This process, starting with an initial subsample of n_k from $\hat{F}_{\frac{R}{r_n},\Theta|r_k}$ is repeated m times, which produces m estimates $\{pv^{(j)}_k\}_{j=1,\ldots,m}$ of the pv_k, which are independent conditional on the original sample. These are then averaged

$$\overline{pv}_k = \frac{1}{m}\sum_{j=1}^{m} pv^{(j)}_k.$$

So for the sequence of levels $\{r_k\}$, we produce a sequence of independent p-values $\{\overline{pv}_k\}$.

Our choice of threshold r at which $(\Theta, R)|R > r$ are independent (and dependent otherwise) will be based on an examination of the path of the mean p-values, $\{\overline{pv}_k\}$. Note the following two observations:

- If R and $\boldsymbol{\Theta}$ are independent given $R > r_k$, then the $pv^{(1)}_k, \ldots, pv^{(m)}_k$ will be iid and approximately $U(0, 1)$-distributed, so that $\overline{pv}_k$ should center around 0.5.
- If R and $\boldsymbol{\Theta}$ are dependent given $R > r_k$, then the $pv^{(j)}_k$'s will be well below 0.5 (closer to 0), and so will $\overline{pv}_k$.

By studying the sequence $\{\overline{pv}_k\}$, which we call the mean p-value path, we choose the threshold to be the smallest r_k such that $\overline{pv}_l$ is around 0.5 for $l < k$. A well-suited change-point method for our situation is the CUSUM algorithm, by Page (1954), which detects the changes in mean in a sequence by looking at mean-corrected partial sums. In our algorithm, we use a spline fitting method that is based on the CUSUM approach called wild binary segmentation (WBS), proposed by Fryzlewicz (2014). The WBS procedure uses the CUSUM statistics of subsamples and fits a piecewise constant spline to $\{\overline{pv}_k\}$. In our setting, we may choose r_k to be the knot of the spline after which the fitted value is comfortably below 0.5.

There are several advantages to using the subsampling scheme. First, recall that the p-value path $\{pv_k\}$, which is obtained from the whole data set, has complicated serial structure and varies greatly from each realization. In contrast, the mean p-values $\overline{pv}_k$ from subsampling are conditionally independent and will center around 0.5 with small variance when the total sample size n and the number of subsample m is large. This, in turns, helps to present a justifiable estimation for the threshold. Second, the calculation of distance covariance can be extremely slow for moderate sample size. Using smaller sample sizes for the subsamples, our computational burden is greatly reduced. In addition, this procedure is amenable to parallel computing, reducing the computation time even further. Third, the subsampling makes it possible to accommodate stationary but dependent data, waiving the stringent independent assumption.

The idea of looking at the mean p-value path is inspired by Mallik et al. (2011), which used the mean of p-values from multiple independent tests to detect change points in population means.

3.6 Data Illustration

In this section, we demonstrate our threshold selection method through simulated and real data examples.

In practice, we set the sequence of thresholds $\{r_k\}$ to be the upper quantiles of R corresponding to $\{q_k\}$, a pre-specified sequence. The subsample size n_k at each threshold r_k is set as $n_k = n_0 \cdot q_k$ for some $n_0 << n$. This is designed such that for any r_k, each subsample is a n_0/n fraction of all the eligible data points with $R > r_k$. Then the choice of $\{n_k\}$ boils down to the choice of n_0, which should reflect the following considerations: i) n_0 should be large enough to ensure good resolution of p-values at all levels; ii) n_0/n should be sufficient small such that the subsamples do not contain too much overlap in observations; iii) larger n_0 requires heavier computation for the distance correlation. In our examples, where the total sample size n ranges from 3000 to 20000, we find n_0 between 500 and 1000 to be a suitable choice. The number of subsamples m can be set as large as computation capacity allows. In our examples, we take $m = 60$.

For all the examples, we choose the weight function μ for distance covariance to be (3.6) with $\kappa = 1$, and the number of replications used to calculate each p-value is $L = 200$. To ensure that the moment conditions are met, the distance correlation is applied to the log of the radial part R in all examples.

3.6.1 Simulated data with known threshold

To illustrate our methodology, we simulate observations from a distribution with a known threshold for which R and Θ become independent.

Let R be the absolute value of a t-distribution with 2 degrees of freedom and Θ_1, Θ_2 be independent random variables such that $\Theta_1 \sim U(0,1)$, $\Theta_2 \sim Beta(3,3)$. Set

$$\Theta | R = \begin{cases} \Theta_1, & \text{if } R > r_{0.2}, \\ \Theta_2, & \text{if } R \le r_{0.2}, \end{cases}$$

where $r_{0.2}$ is the upper 20%-quantile of R. Then R and Θ are independent given $R > r$ if and only if $r \ge r_{0.2}$. Let $(X_{i1}, X_{i2}) = (R_i\Theta_i, R_i(1-\Theta_i))$, $i = 1, \ldots, n$, be the simulated observations. We generate $n = 10000$ iid observations from this distribution. Figures 3.1a, 3.1b and 3.1c show the data in Cartesian and polar coordinates. Our goal is to recover the tail angular distribution by choosing the appropriate threshold.

A sequence of candidate thresholds $\{r_k\}$ is selected to be the empirical upper quantiles of R corresponding to $\{q_k\}$, 150 equidistant points between 0.01 and 0.4. We apply the procedure described in Section 3.5 to the data. For each r_k, the mean p-value $\overline{pv}_k$ is calculated using $m = 60$ random subsamples, each of size $n_k = 500 \cdot q_k$, from the observations with $R_i > r_k$. Figure 3.1d shows the mean p-value path. For the WBS algorithm, we set the threshold to be the largest r_k such that for all thresholds r (quantile level q) such that $r < r_k$ ($q > q_k$), the fitted spline of the p-value stays below 0.45[1]. The threshold levels chosen is 20.4%, which are in good agreement with the true independence level 0.2. The empirical cdfs of the truncated Θ_i's corresponding to the chosen thresholds is shown in Figure 3.1e. We can see that the true tail angular cdf (i.e., $U(0, 1)$) is accurately recovered.

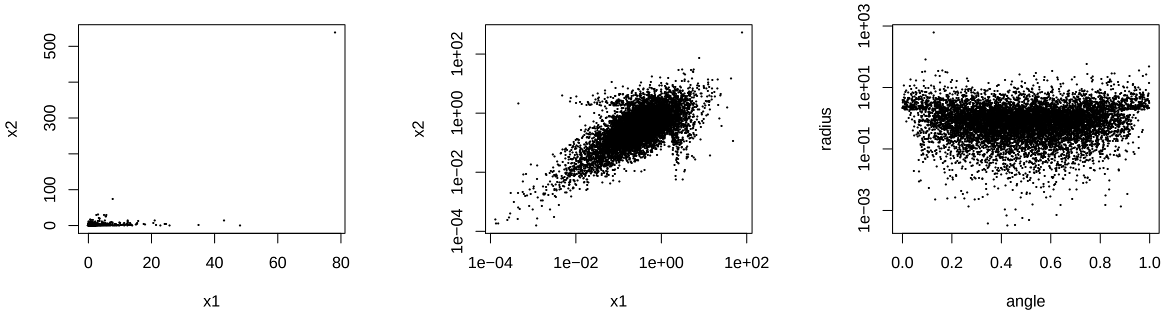
3.6.2 Simulated logistic data

We simulate data from a bivariate logistic distribution, which is bivariate regularly varying. Recall from Example 3.4.4 that (X_1, X_2) follows a bivariate logistic distribution if it has cdf (3.18). In this example, we set $\gamma = 0.8$ and generate $n = 10000$ iid observations from this distribution. Similar to the previous example, for each threshold r_k corresponding to the upper q_k quantile, where $\{q_k\}$ is chosen to be the 150 equidistant points between 0.01 and 0.3. The mean p-value $\overline{pv}_k$ is calculated using $m = 60$ random subsamples of size $n_k = 500 \cdot q_k$ from the observations with $R_i > r_k$.

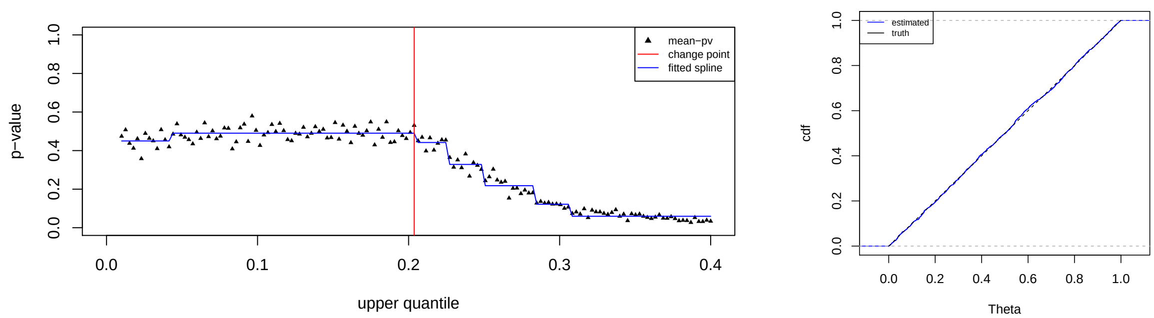
Figures 3.2a , 3.2b and 3.2c show the scatterplots of the data. Here the L_1-norm is used to transform the data into polar coordinates. Our algorithms suggests using 7.4% of the data to estimate the angular distribution. The estimated cdf of the angular distribution is shown with the theoretical limiting cdf, derived from (3.19), in Figure 3.2e. So even though R and Θ are not independent for any threshold r_k, our procedures produce good estimates of the limiting distribution of Θ.

---

[1]Of course, other selection rules can be used. For example, a more conservative approach would be choosing the threshold as the largest r_k such that for $r > r_k$, the fitted spline of the p-value stays above 0.45.



(a) (b) (c)



(d) (e)

Figure 3.1: (a) scatterplot of (X_{i1}, X_{i2}); (b) scatterplot of (X_{i1}, X_{i2}) in log-log scale; (c) scatterplot of (R_i, Θ_i); (d) mean p-value path (black triangles), fitted WBS spline (blue line), and the chosen threshold quantile (red vertical line); (e) estimated cdf of Θ using the threshold chosen, compared with the truth (black dotted).

3.6.3 Real data

In this example, we look at the following exchange rate returns relative to the US dollar: Deutsche mark (DEM), British pound (GBP), Canadian dollar (CAD), and Swiss franc (CHF). The time spans for the data are 1990-01-01 to 1998-12-31 with a total of 3287 days of observations. We examine the pairs GBP/CHF, CAD/CHF, DEM/CHF and estimate the angular density in the tail for each pair. Figures 3.3a–3.3c present the scatter plots of the data. The marginals of the observations are standardized using the rank transformation proposed in Joe et al. (1992):

$$Z_i = 1/\log\{n/(Rank(X_i) - .5)\}, \quad i = 1, \ldots, n.$$

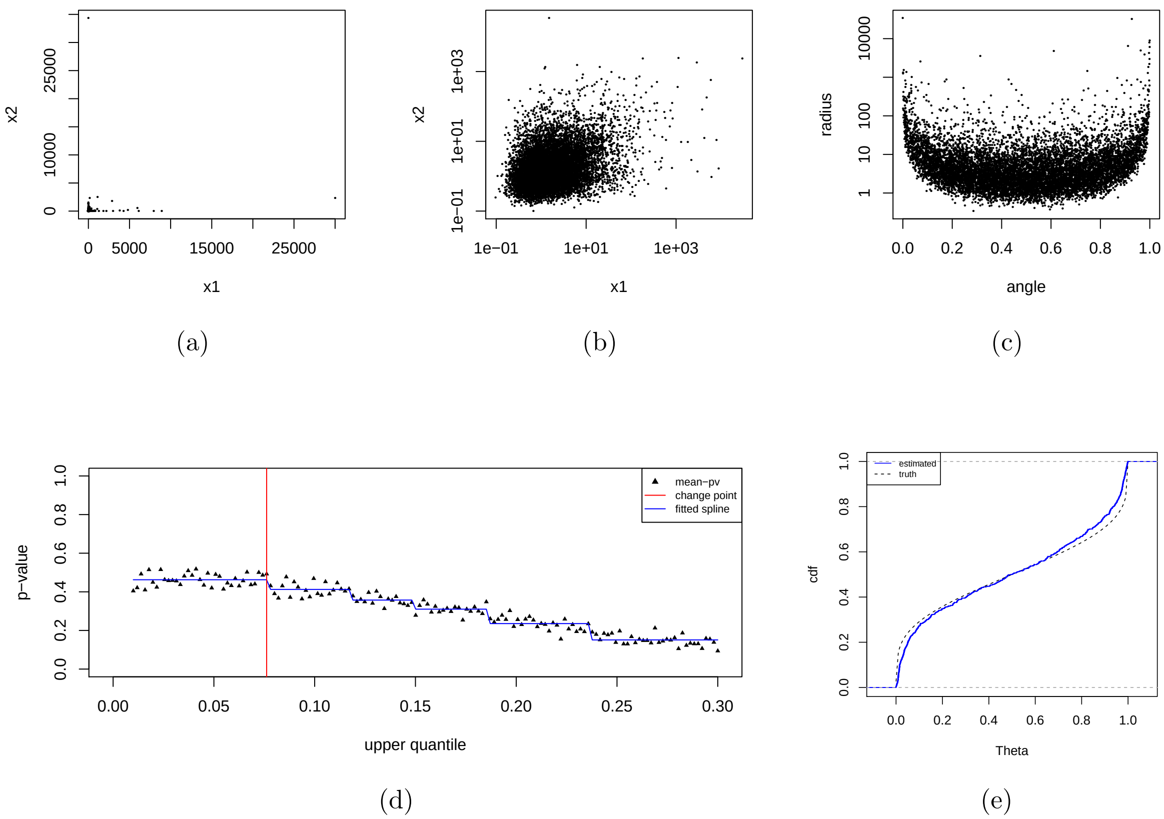


Figure 3.2: (a) scatterplot of (X_{i1}, X_{i2}); (b) scatterplot of (X_{i1}, X_{i2}) in log-log scale; (c) scatterplot of (R_i, Θ_i); (d) mean p-value path (black triangles), fitted WBS spline (blue line), and the chosen threshold quantile (red vertical line); (e) estimated cdf of Θ using the threshold chosen, compared with the theoretical limiting cdf (black dotted).

Again $\{q_k\}$ is chosen to be the 150 equidistant points between 0.01 and 0.3, and the mean p-value $\overline{pv}_k$ is calculated using $m = 60$ random subsamples of size $n_k = 500 \cdot q_k$ from the observations with $R_i > r_k$. Note that while it may not be reasonable to view the observations as iid, the subsampling scheme can still be applied to choose the threshold of independence between R and Θ.

The mean p-value paths are shown in Figures 3.4a–3.4c. The threshold levels selected for the three pairs are 9.6%, 7.4%, 16%, respectively. Figures 3.3d–3.3f show the shape of the estimated angular densities for each pairs. As expected, the tails of the two central European exchange rates, DEM and CHF, are highly dependent. In contrast, that of CAD

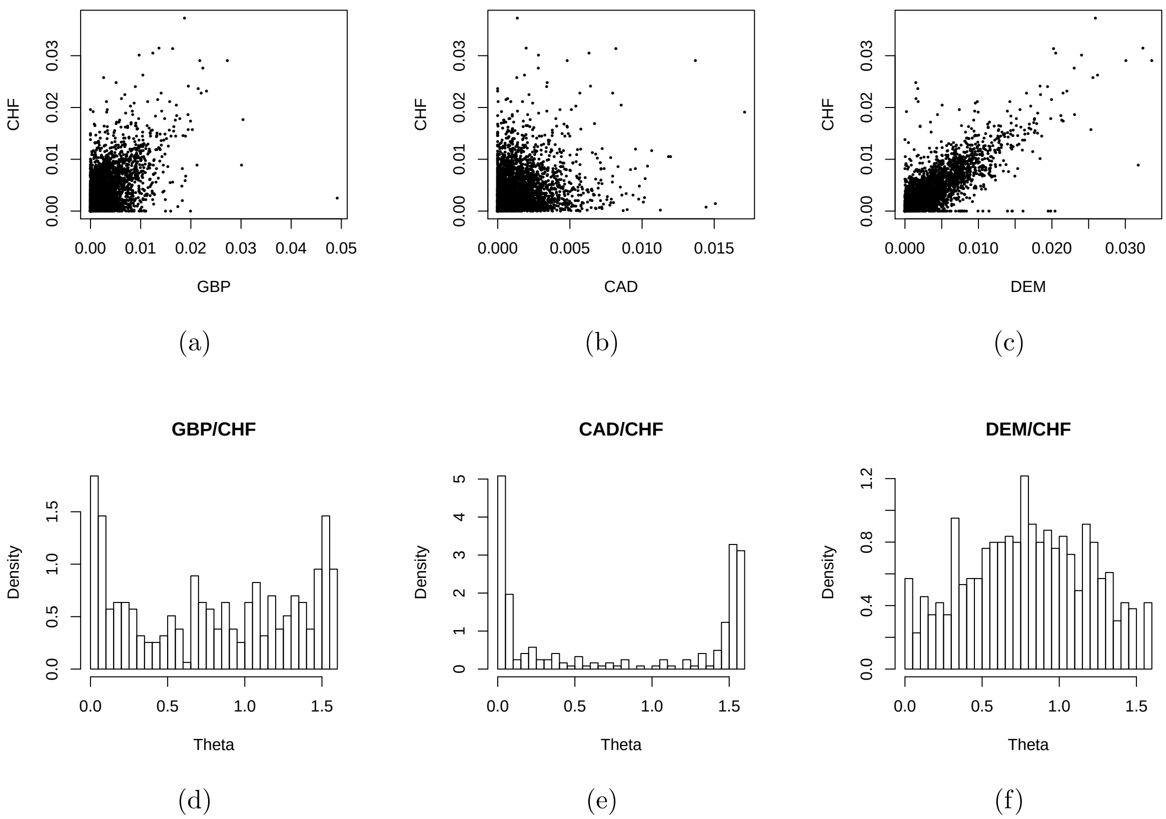


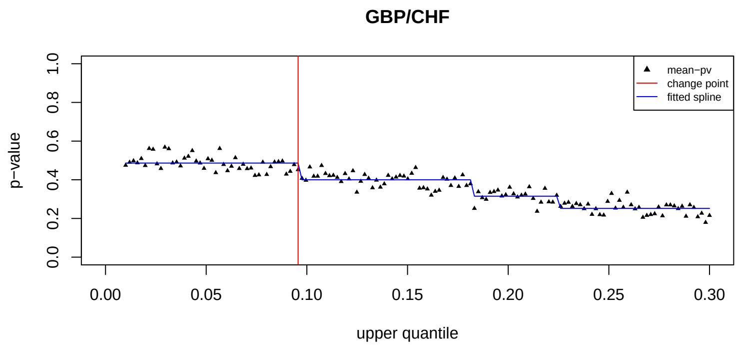
Figure 3.3: Analysis of the paired exchange rate returns: CHF/DEM, CHF/GBP, CHF/CAD with respect to USD between 1990-01-01 to 1998-12-31. (a)–(c): Scatter plots of the standardized paired exchange rate returns; (d)–(f): Estimated angular densities using the estimated thresholds chosen.

and CHF are almost independent.

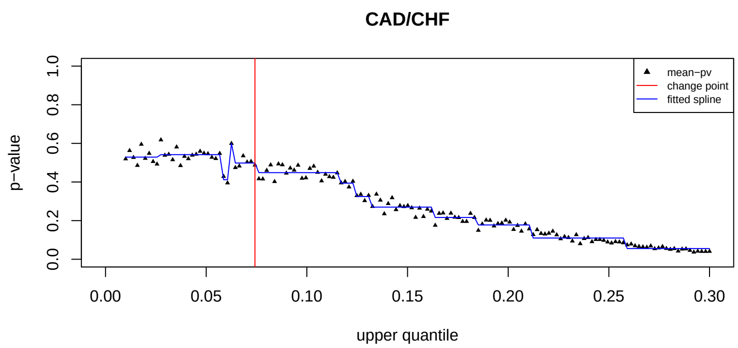
3.6.4 Simulated non-regularly varying data

In this example, we generate data from a model which is not regularly varying. Let R be a random variable from the standard Pareto distribution:

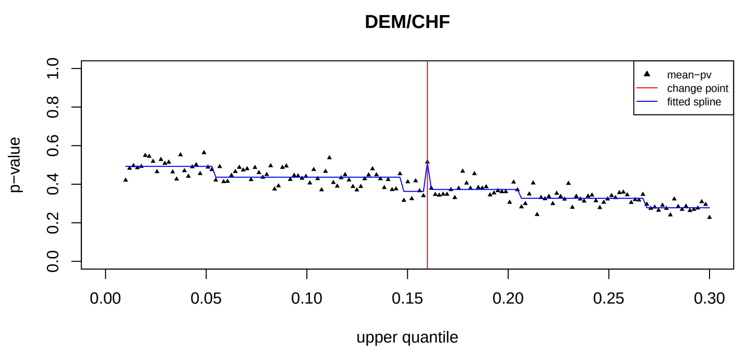
$$\mathbb{P}(R > r) = r^{-1}, \quad r \geq 1.$$



(a)



(b)



(c)

Figure 3.4: Analysis of the paired exchange rate returns: CHF/DEM, CHF/GBP, CHF/CAD with respect to USD between 1990-01-01 to 1998-12-31. (a)–(c): mean p-value paths (black triangles), fitted WBS splines (blue lines) and the chosen threshold quantiles (red vertical line).

Let Θ_1, Θ_2 be independent random variables such that $\Theta_1 \sim U(0, 0.5)$, $\Theta_2 \sim U(0.5, 1)$. Set

$$\Theta|R \sim \begin{cases} \Theta_1, & \text{if } \log R \in (2k, 2k+1] \text{ for some integer } k, \\ \Theta_2, & \text{if } \log R \in (2k+1, 2k+2] \text{ for some integer } k. \end{cases}$$

For any positive integer k, it can be verify that

$$\mathbb{P}(\Theta \in (0, 0.5)|R > e^{2k}) = \frac{1 - e^{-1}}{1 - e^{-2}},$$

while

$$\mathbb{P}(\Theta \in (0, 0.5)|R > e^{2k+1}) = \frac{e^{-1} - e^{-2}}{1 - e^{-2}}.$$

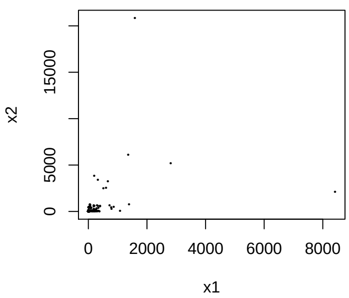
Hence $\mathbb{P}(\Theta \in \cdot|R > r)$ does not convergence as $r \to \infty$ and $\mathbf{X} = (R\Theta, R(1-\Theta))$ is not regularly varying.

Let $(X_{i1}, X_{i2}) = (R_i\Theta_i, R_i(1-\Theta_i))$, $i = 1, \ldots, n$, be iid observations from this distribution, where $n = 20000$. Figures 3.5a, 3.5b and 3.5c show the data in Cartesian and polar coordinates. We apply our threshold selection algorithm to the data, with the threshold upper quantile levels q_k chosen as the 150 equidistant points between 0.01 and 0.2. The mean p-value $\overline{pv}_k$ is calculated using $m = 60$ random subsamples of size $n_k = 500 \cdot q_k$ from the observations with $R_i > r_k$. This is shown in Figure 3.5d.

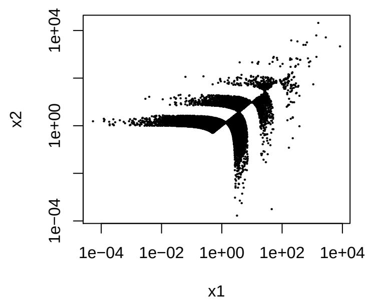
In this model, the radial part R is regularly varying, but Θ and R are dependent given $R > r$ for any r. We expect the mean p-values to be well below 0.5, as are observed. No threshold is selected by the algorithm. This suggests that our technique can potentially be used to detect misspecified models from the regular variation assumption, especially in the scenario where the heavy-tailedness of R is observed but dependence between R and Θ is suspected.

3.7 Discussion

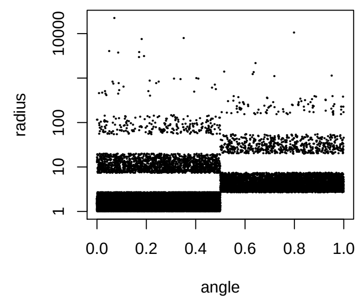
In this chapter, we propose a threshold selection procedure for multivariate regular variation, for which R and Θ are approximately independent for R beyond the threshold. While our



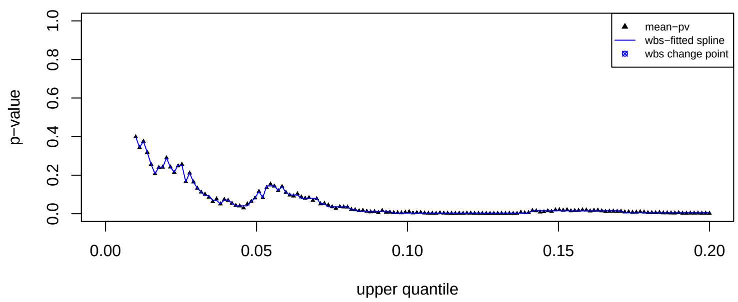
(a)



(b)



(c)



(d)

Figure 3.5: (a) scatterplot of (X_{i1}, X_{i2}); (b) scatterplot of (X_{i1}, X_{i2}) in log-log scale; (c) scatterplot of (R_i, Θ_i); (d) mean p-value path (black triangles), fitted WBS spline (blue line), and the chosen threshold quantile (red vertical line).

problem is set in the multivariate heavy-tailed setting and we utilize distance covariance as our measure of dependence, our algorithm is essentially a change point detection method based on p-values generated through subsampling schemes. Hence this may be generalized to other problem settings and potentially incorporates other dependence measures. Though we have proposed an automatic selection for the threshold based on the fitted mean p-value path, we would like to emphasize that, like the Hill plot, this should be viewed as a visual tool rather than an optimal selection criterion. The final threshold should be based on the proposed procedure in conjunction with visual inspection of the p-value path.

We note that the choice of norm in the polar coordinate transformation (3.1) may result in significant differences in the choice of thresholds, which indicates the rate of convergence to the limit spectral density. This is especially evident in the near 'asymptotic independence' case, where the mass of the angular distribution concentrates on the axes.

As an illustration, we simulated iid observations $\{(X_{i1}, X_{i2})\}_{i=1,\dots,n}$ from the bivariate logistic distribution, where the cdf is given in (3.18), with $\gamma = 0.95$ and $n = 10000$. We apply the polar coordinate transformation with respect to the L_p-norm for $p = 0.2, 1, 5$. Note that in the case of $p = 0.2$, L_p is only a quasi-norm as it does not satisfy the triangular inequality. However, it can be shown that (3.4) holds and the limiting angular distribution exists for bivariate logistic distribution. We compare the threshold selection results in Figure 3.6. Note that in the cases of the L_1 and L_5-norms, the threshold levels are chosen to be upper 5% and 12%, respectively, while in the case of the $L_{0.2}$-norm, it is not possible to select the threshold as the dependence between R and Θ at all levels were shown to be significant. Indeed, this can be seen in Figure 3.7, where we compare the histogram of $X_1^p/(X_1^p + X_2^p)$ given $\|X\|_p$ is large across three levels of truncations, 2%, 5% and 12%, together with the theoretical limiting density curve. For the $L_{0.2}$-norm, the limiting angular density is poorly approximated by the truncated data for all levels. For the other two norms, the truncated observations according to the selected threshold provide decent approximations to the true limiting density of the angular component. One possible explanation for this is that under the $L_{0.2}$-norm, the threshold is concave and hence observations on the diagonal are much easier to be classified as "extremes" than those near the axis. As a result, the estimator of the angular density uses more observations near the diagonal, which may not be, in fact, close enough to the limit. This choice of norm is an interesting topic and is the subject of ongoing research.

3.8 Proof of Theorem 3.4.1

Note from the definition of the empirical distance covariance in (3.9), the integrand can be

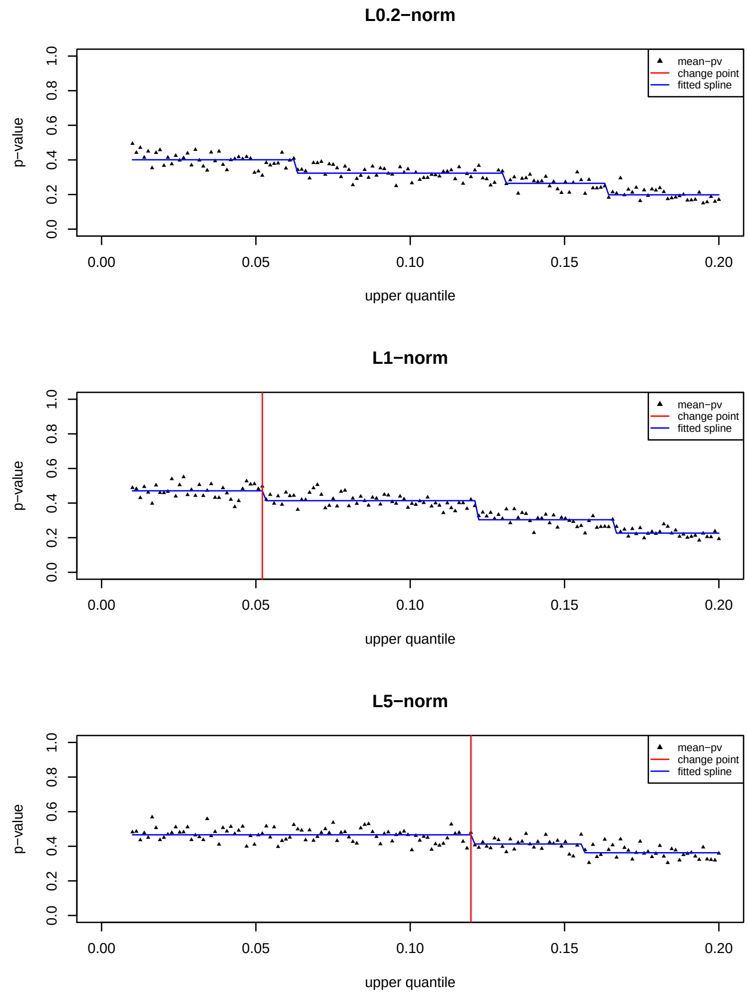


Figure 3.6: Simulated logistic data of sample size $n = 10000$ with $\gamma = 0.95$. Threshold selection algorithm applied under the $L_{0.2}$-, L_1- and L_5-norms: mean p-value paths (black triangles), fitted WBS splines (blue lines) and the chosen threshold quantiles (red vertical line).

expressed as

$$
\begin{aligned}
C_n(s,t) &= \frac{1}{n\hat{p}_n}\sum_{j=1}^{n} e^{isR_j/r_n + it^T\boldsymbol{\Theta}_j}\mathbf{1}_{\{R_j>r_n\}} \\
&\quad - \frac{1}{n\hat{p}_n}\sum_{j=1}^{n} e^{isR_j/r_n}\mathbf{1}_{\{R_j>r_n\}}\frac{1}{n\hat{p}_n}\sum_{k=1}^{n} e^{it^T\boldsymbol{\Theta}_k}\mathbf{1}_{\{R_k>r_n\}} \\
&= \frac{1}{n\hat{p}_n}\sum_{j=1}^{n}\left(e^{isR_j/r_n} - \varphi_{\frac{R}{r_n}|r_n}(s)\right)\left(e^{it^T\boldsymbol{\Theta}_j} - \varphi_{\Theta|r_n}(t)\right)\mathbf{1}_{\{R_j>r_n\}} \\
&\quad - \frac{1}{n\hat{p}_n}\sum_{j=1}^{n}\left(e^{isR_j/r_n} - \varphi_{\frac{R}{r_n}|r_n}(s)\right)\mathbf{1}_{\{R_j>r_n\}}\ \frac{1}{n\hat{p}_n}\sum_{k=1}^{n}\left(e^{it^T\boldsymbol{\Theta}_k} - \varphi_{\Theta|r_n}(t)\right)\mathbf{1}_{\{R_k>r_n\}}.
\end{aligned}
$$

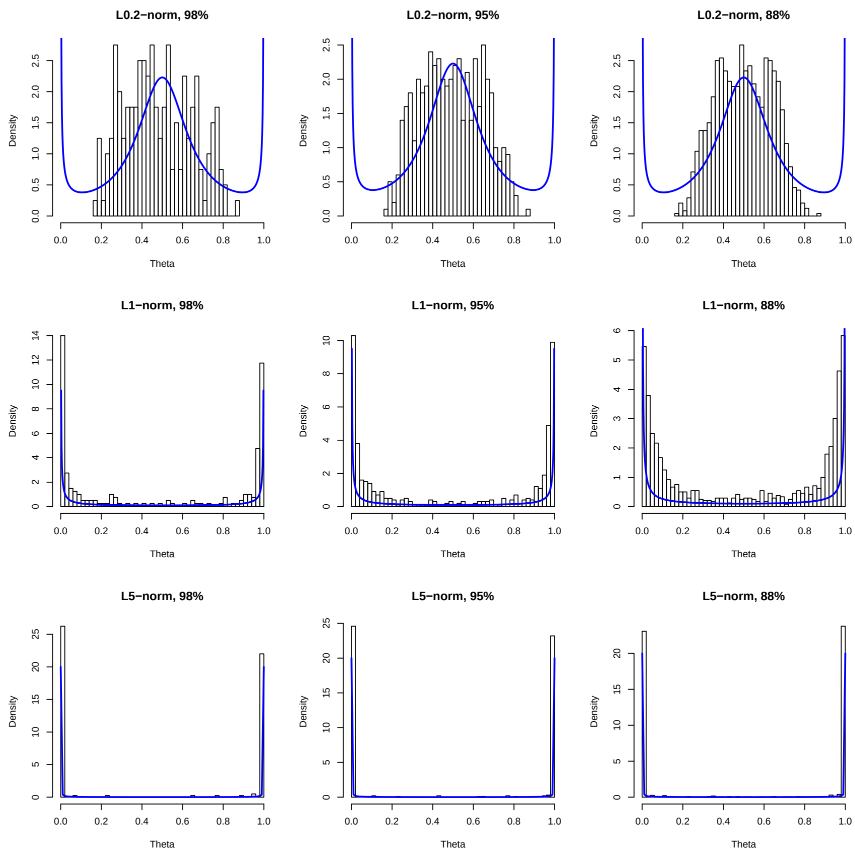


Figure 3.7: Simulated logistic data of sample size $n = 10000$ with $\gamma = 0.95$. Histogram of $X_1^p/(X_1^p + X_2^p)$ for truncated levels 2%, 5% and 12% for $p = 0.2, 1, 5$.

Writing $U_{jn} = \left(e^{isR_j/r_n} - \varphi_{\frac{R}{r_n}|r_n}(s)\right)\mathbf{1}_{\{R_j>r_n\}}$, $V_{jn} = \left(e^{it^T\mathbf{\Theta}_j} - \varphi_{\Theta|r_n}(t)\right)\mathbf{1}_{\{R_j>r_n\}}$, we have

$$C_n(s,t) = \frac{p_n}{\hat{p}_n}\frac{1}{n}\sum_{j=1}^n \frac{U_{jn}V_{jn}}{p_n} - \left(\frac{p_n}{\hat{p}_n}\right)^2 \frac{1}{n}\sum_{j=1}^n \frac{U_{jn}}{p_n}\frac{1}{n}\sum_{k=1}^n \frac{V_{kn}}{p_n}.$$

Since $\mathbb{E}U_{jn} = \mathbb{E}V_{jn} = 0$ and $\mathbb{E}U_{jn}V_{jn}/p_n = \varphi_{\frac{R}{r_n},\Theta|r_n}(s,t) - \varphi_{\frac{R}{r_n}|r_n}(s)\varphi_{\Theta|r_n}(t)$, it is convenient to mean correct the summands and obtain

$$\begin{aligned} C_n(s,t) &= \frac{p_n}{\hat{p}_n}\frac{1}{n}\sum_{j=1}^n\left(\frac{U_{jn}V_{jn}}{p_n} - \left(\varphi_{\frac{R}{r_n},\Theta|r_n}(s,t) - \varphi_{\frac{R}{r_n}|r_n}(s)\varphi_{\Theta|r_n}(t)\right)\right) \\ &\quad - \left(\frac{p_n}{\hat{p}_n}\right)^2 \frac{1}{n}\sum_{j=1}^n \frac{U_{jn}}{p_n}\frac{1}{n}\sum_{k=1}^n \frac{V_{kn}}{p_n} + \frac{p_n}{\hat{p}_n}\left(\varphi_{\frac{R}{r_n},\Theta|r_n}(s,t) - \varphi_{\frac{R}{r_n}|r_n}(s)\varphi_{\Theta|r_n}(t)\right) \\ &=: \left(\frac{p_n}{\hat{p}_n}\right)\tilde{E}_1 - \left(\frac{p_n}{\hat{p}_n}\right)^2\tilde{E}_{21}\tilde{E}_{22} + \left(\frac{p_n}{\hat{p}_n}\right)\tilde{E}_3 \\ &=: \left(\frac{p_n}{\hat{p}_n}\right)\tilde{E}_1 - \left(\frac{p_n}{\hat{p}_n}\right)^2\tilde{E}_2 + \left(\frac{p_n}{\hat{p}_n}\right)\tilde{E}_3 \end{aligned}$$

Note that $\tilde{E}_1, \tilde{E}_{21}, \tilde{E}_{22}$ are averages of iid zero-mean random variables and $\tilde{E}_3$ is non-random. We first prove the second part of Theorem 3.4.1. The first part of Theorem 3.4.1 follows easily in a similar fashion.

Proof of Theorem 3.4.1(2). In order to show (3.14), it suffices to establish that

$$n\hat{p}_n \int_{\mathbb{R}^{d+1}} \left(\frac{p_n}{\hat{p}_n}\right)^2 |\tilde{E}_1|^2 \mu(ds,dt) \xrightarrow{d} \int_{\mathbb{R}^{d+1}} |Q(s,t)|^2 \mu(ds,dt), \tag{3.25}$$

and

$$\left| n\hat{p}_n T_n - n\hat{p}_n \int_{\mathbb{R}^{d+1}} \left(\frac{p_n}{\hat{p}_n}\right)^2 |\tilde{E}_1|^2 \mu(ds,dt) \right| \xrightarrow{p} 0, \tag{3.26}$$

where (3.26) can be implied by

$$n\hat{p}_n \int_{\mathbb{R}^{d+1}} \left(\frac{p_n}{\hat{p}_n}\right)^2 |\tilde{E}_2|^2 \mu(ds,dt) + n\hat{p}_n \int_{\mathbb{R}^{d+1}} \left(\frac{p_n}{\hat{p}_n}\right)^2 |\tilde{E}_3|^2 \mu(ds,dt) \xrightarrow{p} 0. \tag{3.27}$$

Notice that

$$\mathbb{E}\left|\frac{\hat{p}_n}{p_n} - 1\right|^2 = \mathbb{E}\left|\frac{1}{n}\sum_{j=1}^n \left(\frac{\mathbf{1}_{\{R_j > r_n\}}}{p_n} - 1\right)\right|^2 = \frac{1}{n}\mathbb{E}\left|\frac{\mathbf{1}_{\{R_1 > r_n\}}}{p_n} - 1\right|^2 \le \frac{1}{np_n}O(1) + \frac{1}{n}O(1) \to 0.$$

Hence $\hat{p}_n/p_n \xrightarrow{p} 1$ and for (3.25) and (3.27), it is equivalent to prove that

$$np_n \int_{\mathbb{R}^{d+1}} |\tilde{E}_1|^2 \mu(ds,dt) \xrightarrow{d} \int_{\mathbb{R}^{d+1}} |Q(s,t)|^2 \mu(ds,dt) \tag{3.28}$$

and

$$np_n \int_{\mathbb{R}^{d+1}} |\tilde{E}_2|^2 \mu(ds,dt) + np_n \int_{\mathbb{R}^{d+1}} |\tilde{E}_3|^2 \mu(ds,dt) \xrightarrow{p} 0. \tag{3.29}$$

We will show the convergence (3.28) in Proposition 3.8.1. By (3.13),

$$np_n \int_{\mathbb{R}^{d+1}} |\tilde{E}_3|^2 \mu(ds,dt) \to 0.$$

So that (3.29) holds provided

$$np_n \int_{\mathbb{R}^{d+1}} |\tilde{E}_2|^2 \mu(ds,dt) \xrightarrow{p} 0, \tag{3.30}$$

which follows in a similar fashion as Proposition 3.8.1.

□

Proposition 3.8.1. *Assume μ satisfies*

$$\int_{\mathbb{R}^{d+1}} (1 \wedge |s|^\beta)(1 \wedge |t|^2)\mu(ds, dt) < \infty,$$

and that $np_n \to \infty$ as $n \to \infty$, then

$$np_n \int_{\mathbb{R}^{d+1}} |\tilde{E}_1|^2 \mu(ds, dt) \xrightarrow{d} \int_{\mathbb{R}^{d+1}} |Q(s,t)|^2 \mu(ds, dt),$$

where Q is a centered Gaussian process with covariance function (3.15).

Proof of Proposition 3.8.1. We first show that

$$\sqrt{np_n}\tilde{E}_1 \xrightarrow{d} Q(s,t), \quad \text{on } \mathcal{C}(\mathbb{R}^{d+1}) \tag{3.31}$$

which can be implied by the finite distributional convergence of $\sqrt{np_n}\tilde{E}_1(s,t)$ and its tightness on $\mathcal{C}(\mathbb{R}^{d+1})$.

Write

$$\sqrt{np_n}\tilde{E}_1 = \frac{1}{\sqrt{n}} \sum_{j=1}^n \left(\frac{U_{jn} V_{jn}}{\sqrt{p_n}} - \sqrt{p_n}(\varphi_{\frac{R}{r_n},\Theta|r_n}(s,t) - \varphi_{\frac{R}{r_n}|r_n}(s)\varphi_{\Theta|r_n}(t)) \right) =: \frac{1}{\sqrt{n}} \sum_{j=1}^n Y_{jn},$$

where Y_{jn}'s are iid random variables with mean 0. For fixed (s,t), note that

$$\mathrm{Var}(Y_{1n}) = \mathbb{E}|Y_{1n}|^2 = \frac{\mathbb{E}|U_{1n}V_{1n}|^2}{p_n}(1+o(1)) = \frac{\mathbb{E}\mathbf{1}_{\{R_1 > r_n\}}}{p_n} O(1) < \infty.$$

On the other hand, any $\delta > 0$,

$$\mathbb{E}|Y_{1n}|^{2+\delta} = \frac{\mathbb{E}|U_{1n}V_{1n}|^{2+\delta}}{p_n^{1+\delta/2}}(1+o(1)) \le c\frac{\mathbb{E}\mathbf{1}_{\{R_1 > r_n\}}}{p_n^{1+\delta/2}}(1+o(1)) = O(p_n^{-\delta/2})$$

Then we can apply the central limit theorem for triangular arrays by checking the Lyapounov condition (see, e.g., Billingsley (1995)) for the Y_{jn}'s:

$$\frac{\sum_{j=1}^n \mathbb{E}|Y_{jn}|^{2+\delta}}{\left(\mathrm{Var}\left(\sum_{j=1}^n Y_{jn}\right)\right)^{\frac{2+\delta}{2}}} = \frac{O(np_n^{-\frac{\delta}{2}})}{n^{1+\frac{\delta}{2}}\mathrm{Var}(Y_{1n})^{1+\frac{\delta}{2}}} = O((np_n)^{-\frac{\delta}{2}}) \to 0.$$

It follows easily that for fixed (s,t),

$$\sqrt{np_n}\tilde{E}_1 \xrightarrow{d} Q(s,t).$$

The finite-dimensional distribution can be obtained using the Cramér-Wold device and the covariance function can be verified through calculations.

We now show the tightness of $\sqrt{np_n}\tilde{E}_1$. Note that

$$\begin{aligned}
\tilde{E}_1(s,t) &= \frac{1}{n}\sum_{j=1}^{n}\frac{\left(e^{isR_j/r_n}-\varphi_{\frac{R}{r_n}|r_n}(s)\right)\left(e^{it^T\boldsymbol{\Theta}_j}-\varphi_{\Theta|r_n}(t)\right)\mathbf{1}_{\{R_j>r_n\}}}{p_n} \\
&\quad -\left(\varphi_{\frac{R}{r_n},\Theta|r_n}(s,t)-\varphi_{\frac{R}{r_n}|r_n}(s)\varphi_{\Theta|r_n}(t)\right) \\
&= \left(\frac{1}{n}\sum_{j=1}^{n}\frac{e^{isR_j/r_n+it^T\boldsymbol{\Theta}_j}\mathbf{1}_{\{R_j>r_n\}}}{p_n}-\varphi_{\frac{R}{r_n},\Theta|r_n}(s,t)\right) \\
&\quad -\left(\frac{1}{n}\sum_{j=1}^{n}\frac{e^{isR_j/r_n}\mathbf{1}_{\{R_j>r_n\}}}{p_n}-\varphi_{\frac{R}{r_n}|r_n}(s)\right)\varphi_{\Theta|r_n}(t) \\
&\quad -\left(\frac{1}{n}\sum_{j=1}^{n}\frac{e^{it^T\boldsymbol{\Theta}_j}\mathbf{1}_{\{R_j>r_n\}}}{p_n}-\varphi_{\Theta|r_n}(t)\frac{\hat{p}_n}{p_n}\right)\varphi_{\frac{R}{r_n}|r_n}(s) \\
&=: \tilde{E}_{11}+\tilde{E}_{12}+\tilde{E}_{13}.
\end{aligned}$$

Without loss of generality, we show the tightness for $\sqrt{np_n}\tilde{E}_{11}$ and that of $\sqrt{np_n}\tilde{E}_{12}$ and $\sqrt{np_n}\tilde{E}_{13}$ follows from the same argument.

First we introduce some notation following that from Bickel and Wichura (1971). Fix $(s,t),(s',t')\in\mathbb{R}^{d+1}$ where $s<s'$ and $t<t'$. Let B be the subset of $\mathbb{R}^{d+1}$ of the form

$$B := ((s,t),(s',t')] = (s,s']\times\prod_{k=1}^{d}(t_k,t'_k]\subset\mathbb{R}^{d+1}.$$

For ease of notation, we suppress the dependence of B on $(s,t),(s',t')$. Define the increment of the stochastic process $\tilde{E}_{11}$ on B to be

$$\begin{aligned}
\tilde{E}_{11}(B) &:= \frac{1}{n}\sum_{j=1}^{n}\sum_{z_0=0,1}\sum_{z_1=0,1}\cdots\sum_{z_d=0,1}(-1)^{d+1-\sum_j z_j} \\
&\qquad \tilde{E}_{11}\left(s+z_0(s'-s),t_1+z_1(t'_1-t_1),\ldots,t_d+z_d(t'_d-t_d)\right).
\end{aligned}$$

From a sufficient condition of Theorem 3 of Bickel and Wichura (1971), the tightness of $\sqrt{np_n}\tilde{E}_1$ is implied if the following statement holds for any $(s,t),(s',t')$ and corresponding B,

$$\mathbb{E}|\sqrt{np_n}\tilde{E}_{11}(B)|^2\le c|s-s'|^{\beta}\prod_{k=1}^{d}|t_k-t'_k|^{\beta},\quad\text{for some }\beta>1.$$

It follows that

$$
\begin{aligned}
&\mathbb{E}\left|\sqrt{np_n}\left(\tilde{E}_{11}(B)\right)\right|^2 \\
= \;& np_n \mathbb{E}\left| \sum_{z_0=0,1} \cdots \sum_{z_d=0,1} (-1)^{d+1-\sum_j z_j} \frac{1}{n} \sum_{j=1}^{n} e^{i(s+z_0(s'-s))R_j/r} \prod_{k=1}^{d} e^{i(t_k+z_k(t'_k-t_k))\Theta_{jk}} \frac{\mathbf{1}_{\{R_j>r_n\}}}{p_n} \right. \\
& \left. - \sum_{z_0=0,1} \cdots \sum_{z_d=0,1} (-1)^{d+1-\sum_j z_j} \mathbb{E}\left[\left(e^{i(s+z_0(s'-s))R/r}\right) \prod_{k=1}^{d} e^{i(t_k+z_k(t'_k-t_k))\Theta_k} \frac{\mathbf{1}_{\{R>r_n\}}}{p_n}\right] \right|^2 \\
= \;& np_n \mathbb{E}\left| \frac{1}{n} \sum_{j=1}^{n} (e^{isR_j/r_j} - e^{is'R_j/r_j}) \prod_{k=1}^{d} (e^{it_k\Theta_{jk}} - e^{it'_k\Theta_{jk}}) \frac{\mathbf{1}_{\{R_j>r_n\}}}{p_n} \right. \\
& \left. - \mathbb{E}\left[(e^{isR/r} - e^{is'R/r}) \prod_{k=1}^{d} (e^{it_k\Theta_k} - e^{it'_k\Theta_k}) \frac{\mathbf{1}_{\{R>r_n\}}}{p_n}\right] \right|^2 \\
= \;& p_n \mathrm{Var}\left((e^{isR/r_j} - e^{is'R/r_j}) \prod_{k=1}^{d} (e^{it_k\Theta_k} - e^{it'_k\Theta_k}) \frac{\mathbf{1}_{\{R>r_n\}}}{p_n} \right) \qquad (3.32) \\
\le \;& \mathbb{E}\left[\left| (e^{isR/r} - e^{is'R/r}) \prod_{k=1}^{d} (e^{it_k\Theta_k} - e^{it'_k\Theta_k}) \right|^2 \middle| R > r_n \right].
\end{aligned}
$$

From a Taylor series argument,

$$
|e^{ix} - e^{ix'}|^2 \le c\left(1 \wedge |x-x'|^2\right) \le c\left(1 \wedge |x-x'|^\beta\right) \le c|x-x'|^\beta, \quad \text{for any } \beta \in (0,2].
$$

Hence for any $\beta \in (1, 2 \wedge \alpha)$,

$$
\begin{aligned}
\mathbb{E}\left|\sqrt{np_n}\tilde{E}_{11}(B)\right|^2 \;&\le\; c|s-s'|^\beta \prod_{k=1}^{d} |t_i - t'_i|^\beta \mathbb{E}\left[(R/r_n)^\beta \prod_{k=1}^{d} |\Theta_k|^\beta | R > r_n\right] \\
&<\; c|s-s'|^\beta \prod_{k=1}^{d} |t_i - t'_i|^\beta,
\end{aligned}
$$

since $|\Theta_k|^\beta$'s are bounded and $\sup_n \mathbb{E}[(R/r_n)^\beta | R > r_n] < \infty$ by the regular variation assumption. This proves the tightness.

Define the bounded set

$$
K_\delta = \{(s,t) |\; \delta < |s| < 1/\delta, \delta < |t| < 1/\delta\}, \quad \text{for } \delta < .5.
$$

Then, using (3.31), we have from the continuous mapping theorem,

$$
np_n \int_{K_\delta} |\tilde{E}_1|^2 \mu(ds,dt) \xrightarrow{d} \int_{K_\delta} |Q(s,t)|^2 \mu(ds,dt). \qquad (3.33)
$$

On the other hand, for any $\beta < 2 \wedge \alpha$, we have

$$
\begin{aligned}
& \mathbb{E}|\sqrt{np_n}\tilde{E}_1|^2 \\
= \;& np_n \mathbb{E}\left|\frac{1}{n}\sum_{j=1}^{n}\left(\frac{U_{jn}V_{jn}}{p_n} - \mathbb{E}\left[\frac{U_{jn}V_{jn}}{p_n}\right]\right)\right|^2 \\
\leq \;& \frac{\mathbb{E}|U_{jn}V_{jn} - \mathbb{E}U_{jn}V_{jn}|^2}{p_n} \\
\leq \;& \frac{c\,\mathbb{E}|U_{jn}V_{jn}|^2}{p_n} \qquad\qquad (3.34) \\
= \;& \frac{c\,\mathbb{E}\left[\left|e^{isR_j/r_n} - \varphi_{\frac{R}{r_n}|r_n}(s)\right|^2 \left|e^{it^T\mathbf{\Theta}_j} - \varphi_{\Theta|r_n}(t)\right|^2 \mathbf{1}_{\{R_j > r_n\}}\right]}{p_n} \\
\leq \;& \frac{c\,\mathbb{E}\left[\left(\left|e^{isR_j/r_n} - 1\right|^2 + \left|\varphi_{\frac{R}{r_n}|r_n}(s) - 1\right|^2\right)\left(\left|e^{it^T\mathbf{\Theta}_j} - 1\right|^2 + \left|\varphi_{\Theta|r_n}(t) - 1\right|^2\right)\mathbf{1}_{\{R_j > r_n\}}\right]}{p_n} \\
\leq \;& \frac{c\,\mathbb{E}\left[\left(1 \wedge \left|\frac{sR_j}{r_n}\right|^2 + \mathbb{E}\left[1 \wedge \left|\frac{sR_j}{r_n}\right|^2 |\frac{R}{r_n} > 1\right]\right)\left(1 \wedge |t\mathbf{\Theta}_j|^2 + \mathbb{E}\left[1 \wedge |t\mathbf{\Theta}_j|^2 |\frac{R}{r_n} > 1\right]\right)\mathbf{1}_{\{R_j > r_n\}}\right]}{p_n} \\
\leq \;& \frac{c\,\mathbb{E}\left[\left(1 \wedge \left|\frac{sR_j}{r_n}\right|^\beta + \mathbb{E}\left[1 \wedge \left|\frac{sR_j}{r_n}\right|^\beta |\frac{R}{r_n} > 1\right]\right)\left(1 \wedge |t\mathbf{\Theta}_j|^2 + \mathbb{E}\left[1 \wedge |t\mathbf{\Theta}_j|^2 |\frac{R}{r_n} > 1\right]\right)\mathbf{1}_{\{R_j > r_n\}}\right]}{p_n} \\
\leq \;& \frac{c\,\mathbb{E}\left[\left(1 \wedge |s|^\beta\right)\left(\left|\frac{R_j}{r_n}\right|^\beta + \mathbb{E}\left[\left|\frac{R}{r_n}\right|^\beta |\frac{R}{r_n} > 1\right]\right)(1 \wedge |t|^2)\,\mathbf{1}_{\{R_j > r_n\}}\right]}{p_n} \\
\leq \;& c\,\mathbb{E}\left[\left(1 \wedge |s|^\beta(|R_j/r_n|^\beta + \mathbb{E}[|R/r_n|^\beta | R > r_n])\right)\left(1 \wedge |t|^2\right)|R > r_n\right] \\
\leq \;& c(1 \wedge |s|^\beta)(1 \wedge |t|^2).
\end{aligned}
$$

Therefore for any $\epsilon > 0$,

$$
\begin{aligned}
\lim_{\delta\to 0}\limsup_{n\to\infty}\mathbb{P}\left[np_n\int_{K_\delta^c}|\tilde{E}_1|^2\mu(ds,dt) > \epsilon\right] \;&\leq\; \frac{1}{\epsilon}\lim_{\delta\to 0}\limsup_{n\to\infty}\int_{K_\delta^c}\mathbb{E}|\sqrt{np_n}\tilde{E}_1|^2\mu(ds,dt) \\
&\leq\; \frac{1}{\epsilon}\lim_{\delta\to 0}\limsup_{n\to\infty}\int_{K_\delta^c}c(1 \wedge |s|^\beta)(1 \wedge |t|^2)\mu(ds,dt) \\
&\to\; 0
\end{aligned}
$$

by the dominated convergence theorem. This combined with (3.33) shows the convergence of $np_n\int|\tilde{E}_1|^2\mu(ds,dt)$ to $\int|Q(s,t)|^2\mu(ds,dt)$, and hence completes the proof of the proposition.

□

Proof of Theorem 3.4.1(2) (cont.) Now it remains to show (3.30). Similar to the proof of Proposition 3.8.1, we can show that

$$\sqrt{np_n}\tilde{E}_{21} \xrightarrow{d} Q'$$

for a centered Gaussian process Q', and

$$\tilde{E}_{22} \xrightarrow{p} 0.$$

Hence

$$\sqrt{np_n}\tilde{E}_2 = \sqrt{np_n}\tilde{E}_{21}\tilde{E}_{22} \xrightarrow{p} 0.$$

The argument then follows similarly from the continuous mapping theorem and bounding the tail integrals. □

Proof of Theorem 3.4.1(1). Similar to the proof of Theorem 3.4.1(2), it suffices to show that

$$\int |\tilde{E}_i|^2 \mu(ds, dt) \xrightarrow{p} 0, \quad i = 1, 2, 3. \tag{3.35}$$

The convergence (3.35) for $i = 1, 2$ follows trivially from the more general results (3.28) and (3.30) in the proof of Theorem 3.4.1(2). Hence it suffices to show

$$\int |\tilde{E}_3|^2 \mu(ds, dt) \to 0, \tag{3.36}$$

where we recall that $\tilde{E}_3 := \varphi_{\frac{R}{r_n},\Theta|r_n}(s,t) - \varphi_{\frac{R}{r_n}|r_n}(s)\varphi_{\Theta|r_n}(t)$ is non-random.

Let $P_{\frac{R}{r_n},\Theta|r_n}(\cdot) = P\left[\left(\frac{R}{r_n}, \boldsymbol{\Theta}\right) \in \cdot | \frac{R}{r_n} > 1\right]$ and $P_{\frac{R}{r_n}|r_n}, P_{\Theta|r_n}$ be the corresponding marginal measures. Then from (3.3),

$$P_{\frac{R}{r_n},\Theta|r_n} - P_{\frac{R}{r_n}|r_n}P_{\Theta|r_n} \xrightarrow{v} \nu_\alpha \times S - \nu_\alpha \times S = 0,$$

and hence for fixed (s, t),

$$\tilde{E}_3(s,t) = \int e^{isT + it^T\boldsymbol{\Theta}} \left(P_{\frac{R}{r_n},\Theta|r_n} - P_{\frac{R}{r_n}|r_n}P_{\Theta|r_n}\right)(dT, d\boldsymbol{\Theta}) \to 0.$$

For any $\beta < 2 \wedge \alpha$, using the same argument in (3.34),

$$|\tilde{E}_3|^2 = \left(\frac{\mathbb{E}|U_{jn}V_{jn}|}{p_n}\right)^2 \leq c(1 \wedge |s|^\beta)(1 \wedge |t|^2).$$

Then (3.36) follows from (3.12) and dominated convergence. This concludes the proof.

□

3.9 Proof of Theorem 3.4.5

Following the same notation and steps as the proof of Theorem 3.4.1 in Section 3.8, it suffices to prove the following convergences for the mixing case:

$$\frac{\hat{p}_n}{p_n} \xrightarrow{p} 1, \tag{3.37}$$

$$np_n \int_{\mathbb{R}^{d+1}} |\tilde{E}_1|^2 \mu(ds, dt) \xrightarrow{d} \int_{\mathbb{R}^{d+1}} |Q'(s,t)|^2 \mu(ds, dt) \tag{3.38}$$

and

$$np_n \int_{\mathbb{R}^{d+1}} |\tilde{E}_2|^2 \mu(ds, dt) \xrightarrow{p} 0. \tag{3.39}$$

We prove (3.37) and (3.38) in Propositions 3.9.2 and 3.9.3, respectively. The proof of (3.39) follows in a similar fashion. The proofs of both propositions rely on the following lemma.

Throughout this proof we make use of the results that if $\{Z_t\}$ is stationary and α-mixing with coefficient $\{\alpha_h\}$, then

$$|\mathrm{cov}(Z_0, Z_h)| \leq c\alpha_h^{\delta} \left(\mathbb{E}|Z_0|^{2/(1-\delta)}\right)^{1-\delta}, \quad \text{for any } \delta \in (0,1), \tag{3.40}$$

see Section 1.2.2, Theorem 3(a) of Doukhan (1994).

Lemma 3.9.1. *Let $\{\mathbf{X}_t\}$ be a multivariate stationary time series that is regularly varying and α-mixing with mixing coefficient $\{\alpha_h\}$. For a sequence $r_n \to \infty$, set $p_n = \mathbb{P}(\|\mathbf{X}_0\| > r_n)$. Let f_1, f_2 be bounded functions which vanish outside $\overline{\mathbb{R}}^d \backslash B_1(\mathbf{0})$, where $B_1(\mathbf{0})$ is the unit open ball $\{\mathbf{x} | \|\mathbf{x}\| < 1\}$, with sets of discontinuity of measure zero. Set,*

$$S_n^{(i)} = \sum_{t=1}^{n} \left(f_i\left(\frac{\mathbf{X}_t}{r_n}\right) - \mathbb{E} f_i\left(\frac{\mathbf{X}_0}{r_n}\right)\right), \quad i = 1, 2.$$

*Assume that condition (**M**) holds for $\{\alpha_h\}$ and $\{r_n\}$. Then*

$$\frac{1}{\sqrt{np_n}} (S_n^{(1)}, S_n^{(2)})^T \xrightarrow{d} N(\mathbf{0}, \Sigma), \tag{3.41}$$

where the covariance matrix $[\Sigma_{ij}]_{i,j=1,2} = [\sigma^2(f_i, f_j)]_{i,j=1,2}$ with

$$\sigma^2(f_1, f_2) := \sigma_0^2(f_1, f_2) + 2\sum_{h=1}^{\infty} \sigma_h^2(f_1, f_2) \tag{3.42}$$

and

$$\sigma_h^2(f_1, f_2) = \int f_1 f_2 d\mu_h, \quad h \geq 0. \tag{3.43}$$

In particular,

$$\frac{1}{np_n}(S_n^{(1)}, S_n^{(2)})^T \xrightarrow{p} \mathbf{0}.$$

The proof of Lemma 3.9.1 is provided after the proofs of the propositions.

Proposition 3.9.2. *Assume that condition (**M**) holds, then*

$$\frac{\hat{p}_n}{p_n} \xrightarrow{p} 1,$$

Proof. We have

$$\frac{\hat{p}_n}{p_n} - 1 = \frac{1}{n}\sum_{j=1}^n \left(\frac{\mathbf{1}_{\{R_j > r_n\}}}{p_n} - 1\right) = \frac{1}{np_n}\sum_{j=1}^n (\mathbf{1}_{\{R_j > r_n\}} - p_n).$$

Apply Lemma 3.9.1 to $f(\mathbf{x}) = \mathbf{1}_{\{\|\mathbf{x}\|>1\}}$ and the result follows. □

Proposition 3.9.3. *Assume that condition (**M**) holds, and that μ and $\{r_n\}$ satisfies* (3.12) *and* (3.13), *respectively, then*

$$np_n \int_{\mathbb{R}^{d+1}} |\tilde{E}_1|^2 \mu(ds, dt) \xrightarrow{d} \int_{\mathbb{R}^{d+1}} |Q'(s,t)|^2 \mu(ds, dt),$$

where Q' is a centered Gaussian process.

Proof. Let us first establish the convergence of $\sqrt{np_n}\tilde{E}_1(s,t)$ for fixed (s,t). Take

$$\begin{aligned} f_1(\mathbf{x}) &= \mathrm{Re}\left\{\left(e^{is\|\mathbf{x}\|} - \mathbb{E}[e^{is\|\mathbf{x}\|} \,|\, \|\mathbf{x}\| > 1]\right)\left(e^{it\mathbf{x}/\|\mathbf{x}\|} - \mathbb{E}[e^{it\mathbf{x}/\|\mathbf{x}\|} \,|\, \|\mathbf{x}\| > 1]\right)\mathbf{1}_{\|\mathbf{x}\|>1}\right\} \\ f_2(\mathbf{x}) &= \mathrm{Im}\left\{\left(e^{is\|\mathbf{x}\|} - \mathbb{E}[e^{is\|\mathbf{x}\|} \,|\, \|\mathbf{x}\| > 1]\right)\left(e^{it\mathbf{x}/\|\mathbf{x}\|} - \mathbb{E}[e^{it\mathbf{x}/\|\mathbf{x}\|} \,|\, \|\mathbf{x}\| > 1]\right)\mathbf{1}_{\|\mathbf{x}\|>1}\right\}. \end{aligned}$$

Then from Lemma 3.9.1,

$$\frac{1}{\sqrt{np_n}}(S_n^{(1)}, S_n^{(2)})^T = \sqrt{np_n}(\mathrm{Re}\{\tilde{E}_1(s,t)\}, \mathrm{Im}\{\tilde{E}_1(s,t)\}) \xrightarrow{d} N(\mathbf{0}, \Sigma),$$

where the covariance structure Σ can be derived from (3.42) and (3.43). This implies that

$$\sqrt{np_n}\tilde{E}_1(s,t) \xrightarrow{d} Q'(s,t),$$

where $Q'(s,t)$ is a zero-mean complex normal process with covariance matrix $\Sigma_{11}+\Sigma_{22}$ and relation matrix $\Sigma_{11}-\Sigma_{22}+i(\Sigma_{12}+\Sigma_{21})$.

The finite-dimensional distributional convergence of $\sqrt{n\hat{p}_n}\tilde{E}_1$ to a $Q'(s,t)$ can be generalized using the Cramér-Wold device and we omit the calculation of the covariance structure. The tightness condition for the functional convergence follows the same arguments in the proof of Proposition 3.8.1 from Bickel and Wichura (1971), with equality (3.32) replaced by a variance calculation of the sum of α-mixing components using the inequality (3.40) and condition (3.33) is verified through the same argument. This completes the proof of Proposition 3.9.3.

□

of Lemma 3.9.1. The proof follows from that of Theorem 3.2 in Davis and Mikosch (2009). Here we outline the sketch of the proof and detail only the parts that differ from their proof.

By the vague convergence in (3.20), we have

i)

$$\frac{1}{p_n}\mathbb{E}\left[f_i\left(\frac{\mathbf{X}_0}{r_n}\right)\right]\to\int f_i d\mu_0 \quad\text{and}\quad \frac{1}{p_n}\mathbb{E}\left[f_i^2\left(\frac{\mathbf{X}_0}{r_n}\right)\right]\to\int f_i^2 d\mu_0;$$

ii)

$$\frac{1}{p_n}\mathrm{Var}\left[f_i\left(\frac{\mathbf{X}_0}{r_n}\right)\right]=\frac{1}{p_n}\mathbb{E}\left[f_i^2\left(\frac{\mathbf{X}_0}{r_n}\right)\right]-p_n\left(\frac{1}{p_n}\mathbb{E}\left[f_i\left(\frac{\mathbf{X}_0}{r_n}\right)\right]\right)^2\to\int f_i^2 d\mu_0=\sigma_0^2(f_i,f_i);$$

iii)

$$\frac{1}{p_n}\mathrm{cov}\left[f_i\left(\frac{\mathbf{X}_0}{r_n}\right),f_j\left(\frac{\mathbf{X}_h}{r_n}\right)\right]\ \to\ \int f_i f_j d\mu_h\ =\ \sigma_h^2(f_i,f_j).$$

Let us first consider the marginal convergence of $\frac{1}{\sqrt{np_n}}S_n^{(i)}$ for $i=1,2$. Without loss of generality, we suppress the dependency on i and set

$$Y_{tn}:=f\left(\frac{\mathbf{X}_t}{r_n}\right)-\mathbb{E}f\left(\frac{\mathbf{X}_0}{r_n}\right).$$

Then

$$\frac{1}{p_n}\mathrm{Var}\left[Y_{1n}\right]\to\sigma_0^2(f,f)\quad\text{and}\quad\frac{1}{p_n}\mathrm{cov}\left(Y_{1n},Y_{(h+1)n}\right)\to\sigma_h^2(f,f).$$

We also have the following two results for $|\text{cov}(Y_{1n}, Y_{(h+1)n})|$:

$$
\begin{aligned}
&\lim_{h\to\infty}\limsup_{n\to\infty}\sum_{j=h}^{l_n}\frac{1}{p_n}|\text{cov}(Y_{1n}, Y_{(j+1)n})| \\
\le\;& \lim_{h\to\infty}\limsup_{n\to\infty}\sum_{j=h}^{l_n}\frac{1}{p_n}\mathbb{E}\left|f\left(\frac{\mathbf{X}_0}{r_n}\right)f\left(\frac{\mathbf{X}_j}{r_n}\right)\right| + \sum_{j=h}^{l_n}\frac{1}{p_n}\mathbb{E}\left(\left|f\left(\frac{\mathbf{X}_0}{r_n}\right)\right|\right)^2 \\
\le\;& \lim_{h\to\infty}\limsup_{n\to\infty}\sum_{j=h}^{l_n}\frac{c}{p_n}\mathbb{E}\left(\mathbf{1}_{\{\|\mathbf{X}_0\|>r_n\}}\right)\left(\mathbf{1}_{\{\|\mathbf{X}_j\|>r_n\}}\right) + \sum_{j=h}^{l_n}\frac{c}{p_n}\left(\mathbb{E}\mathbf{1}_{\{\|\mathbf{X}_0\|>r_n\}}\right)^2 \\
\le\;& \lim_{h\to\infty}\limsup_{n\to\infty}\sum_{j=h}^{l_n}\frac{c}{p_n}\mathbb{P}\left(\|\mathbf{X}_0\|>r_n, \|\mathbf{X}_j\|>r_n\right) + cl_np_n \\
=\;& 0 \qquad (3.44)
\end{aligned}
$$

from condition (3.22), and

$$
\begin{aligned}
\lim_{n\to\infty}\sum_{j=l_n}^{\infty}\frac{1}{p_n}|\text{cov}(Y_{1n}, Y_{(j+1)n})| \;\le\;& \lim_{n\to\infty}\sum_{j=l_n}^{\infty}\frac{1}{p_n}\left|\text{cov}\left[f\left(\frac{\mathbf{X}_0}{r_n}\right), f\left(\frac{\mathbf{X}_j}{r_n}\right)\right]\right| \\
\le\;& \lim_{n\to\infty}\sum_{j=l_n}^{\infty}\frac{1}{p_n}\alpha_j^\delta\left(\mathbb{E}\left|f\left(\frac{\mathbf{X}_0}{r_n}\right)\right|^{2/(1-\delta)}\right)^{1-\delta} \\
\le\;& \lim_{n\to\infty}\sum_{j=l_n}^{\infty}\frac{c}{p_n}\alpha_j^\delta\left(\mathbb{E}\left(\mathbf{1}_{\{\|\mathbf{X}_0\|>r_n\}}\right)^{2/(1-\delta)}\right)^{1-\delta} \\
\le\;& \lim_{n\to\infty}\sum_{j=l_n}^{\infty}c\alpha_j^\delta p_n^{-\delta} \\
=\;& 0 \qquad (3.45)
\end{aligned}
$$

from condition (3.21).

We apply the same technique of small/large blocks as used in Davis and Mikosch (2009). Let m_n and l_n be the sizes of big and small blocks, respectively, where $l_n \ll m_n \ll n$. Let $I_{kn} = \{(k-1)m_n+1, \ldots, km_n\}$ and $J_{kn} = \{(k-1)m_n+1, \ldots, (k-1)m_n+l_n\}$, $k = 1, \ldots, n/m_n$, be the index sets of big and small blocks respectively. Set $\tilde{I}_{kn} = I_{kn}\backslash J_{kn}$, i.e., $\tilde{I}_{kn}$ are the big blocks with the first l_n observations removed. For simplicity, we set $m_n := 1/p_n$ and assume that the number of big blocks $n/m_n = np_n$ is integer-valued. The non-integer case can be generalized without additional difficulties.

Denote

$$S_n(B) := \sum_{t \in B} Y_{tn},$$

then

$$\sum_{t=1}^{n} Y_{tn} = S_n(1:n) = \sum_{k=1}^{np_n} S_n(I_{kn}) = \sum_{k=1}^{np_n} S_n(\tilde{I}_{kn}) + \sum_{k=1}^{np_n} S_n(J_{kn}).$$

Let $\{\tilde{S}_n(\tilde{I}_{kn})\}_{k=1,\ldots,np_n}$ be iid copies of $\tilde{S}_n(\tilde{I}_{1n})$. To prove the convergence of $\frac{1}{\sqrt{np_n}} S_n(1:n)$, it suffices to show the following:

$$\frac{1}{\sqrt{np_n}} \sum_{k=1}^{np_n} \tilde{S}_n(\tilde{I}_{kn}) \text{ and } \frac{1}{\sqrt{np_n}} \sum_{k=1}^{np_n} S_n(\tilde{I}_{kn}) \text{ has the same limiting distribution,} \tag{3.46}$$

$$\frac{1}{\sqrt{np_n}} \sum_{k=1}^{np_n} S_n(J_{kn}) \xrightarrow{p} 0, \tag{3.47}$$

and

$$\frac{1}{\sqrt{np_n}} \sum_{k=1}^{np_n} \tilde{S}_n(\tilde{I}_{kn}) \xrightarrow{d} N(0, \sigma^2(f,f)). \tag{3.48}$$

The statement (3.46) holds if

$$np_n \alpha_{l_n} \to 0, \quad \text{as } n \to \infty. \tag{3.49}$$

This follows from the same argument in equation (6.2) in Davis and Mikosch (2009).

For condition (3.47), it suffices to show that

$$\frac{1}{np_n} \mathrm{Var}\left(\sum_{k=1}^{np_n} S_n(J_{kn}) \right) \to 0.$$

Note that

$$\frac{1}{np_n} \mathrm{Var}\left(\sum_{k=1}^{np_n} S_n(J_{kn}) \right) \leq \mathrm{Var}(S_n(J_{1n})) + 2 \sum_{h=1}^{np_n-1} (1 - \frac{h}{np_n}) |\mathrm{cov}(S_n(J_{1n}), S_n(J_{(h+1)n}))| =: P_1 + P_2.$$

We have

$$\begin{aligned} \limsup_{n\to\infty} P_1 &= \limsup_{n\to\infty} \mathrm{Var}\left(\sum_{j=1}^{l_n} Y_{jn} \right) \\ &\leq \limsup_{n\to\infty} l_n p_n \left(\frac{\mathrm{Var}(Y_{1n})}{p_n} + 2 \sum_{j=1}^{l_n-1} (1 - \frac{j}{l_n}) \frac{|\mathrm{cov}(Y_{1n}, Y_{(j+1)n})|}{p_n} \right) \end{aligned}$$

$$
\begin{aligned}
&\le \limsup_{n\to\infty} l_n p_n \frac{\mathrm{Var}(Y_{1n})}{p_n} + \lim_{h\to\infty}\limsup_{n\to\infty} 2l_n p_n \sum_{j=1}^{h-1} \frac{|\mathrm{cov}(Y_{1n}, Y_{(j+1)n})|}{p_n} \\
&\quad + \lim_{h\to\infty}\limsup_{n\to\infty} 2l_n p_n \sum_{j=h}^{l_n-1} \frac{|\mathrm{cov}(Y_{1n}, Y_{(j+1)n})|}{p_n} \\
&= 0
\end{aligned}
$$

where the last step follows from dominated convergence and (3.44). And for the other term,

$$
\begin{aligned}
P_2 &\le 2\sum_{h=1}^{np_n-1} \sum_{s\in J_{1n}} \sum_{t\in J_{(h+1)n}} |\mathrm{cov}(Y_{sn}, Y_{tn})| \\
&\le 2\sum_{h=1}^{np_n-1} l_n \sum_{k=h/p_n-l_n+1}^{h/p_n} \left|\mathrm{cov}(Y_{1n}, Y_{(k+1)n})\right| \\
&\le 2l_n p_n \sum_{k=1/p_n-l_n+1}^{\infty} \frac{\left|\mathrm{cov}(Y_{1n}, Y_{(k+1)n})\right|}{p_n} \\
&\le 2l_n p_n \sum_{k=l_n+1}^{\infty} \frac{\left|\mathrm{cov}(Y_{1n}, Y_{(k+1)n})\right|}{p_n} \to 0.
\end{aligned}
$$

Note that $1/p_n = m_n$ is the size of big blocks I_{kn}'s and $1/p_n - l_n + 1 = m_n - l_n + 1$ is the distance between consecutive small blocks $(J_{kn}, J_{(k+1)n})$'s. The last limit follows from (3.45).

To finish the proof, we need to establish the central limit theorem in (3.48). Note the $\tilde{S}_n(\tilde{I}_{ln})$'s are iid with $\mathbb{E}\tilde{S}_n(\tilde{I}_{ln}) = 0$. We now calculate its variance. Recall that $1/p_n - l_n$ is the size of $\tilde{I}_{1n}$, the big block with small block removed. Then

$$
\begin{aligned}
&\mathrm{Var}\left(\tilde{S}_n(\tilde{I}_{1n})\right) \\
&= \mathrm{Var}\left(\sum_{j=1}^{1/p_n-l_n} Y_{jn}\right) \\
&= (\frac{1}{p_n} - l_n)\mathrm{Var}(Y_{jn}) + 2\sum_{k=1}^{1/p_n-l_n-1} (1/p_n - l_n - k)\mathrm{cov}(Y_{1n}, Y_{(k+1)n}) \\
&= (\frac{1}{p_n} - l_n)\mathrm{Var}(Y_{jn}) + 2\left(\sum_{k=1}^{h} + \sum_{k=h+1}^{l_n} + \sum_{k=l_n+1}^{1/p_n-l_n-1}\right)\left(1 - \frac{l_n+k}{1/p_n}\right)\frac{1}{p_n}\mathrm{cov}(Y_{1n}, Y_{(k+1)n}) \\
&:= I_0 + I_1 + I_2 + I_3.
\end{aligned}
$$

Here

$$\lim_{n\to\infty} I_0 = \lim_{n\to\infty} (1 - l_n p_n)\frac{1}{p_n}\mathrm{Var}(Y_{jn}) = \sigma_0^2(f,f),$$

and

$$\lim_{n\to\infty} I_1 = \lim_{n\to\infty} 2\sum_{k=1}^{h} (1 - p_n(l_n + k))\,\frac{\mathrm{cov}(Y_{1n}, Y_{(k+1)n})}{p_n} = 2\sum_{k=1}^{h} \sigma_k^2(f,f).$$

We also have

$$\lim_{h\to\infty} \limsup_{n\to\infty} |I_2| \leq \lim_{h\to\infty} \limsup_{n\to\infty} \sum_{k=h+1}^{l_n} \frac{|\mathrm{cov}(Y_{1n}, Y_{(k+1)n})|}{p_n} = 0$$

from (3.44), and

$$\lim_{n\to\infty} |I_3| \leq \lim_{n\to\infty} \sum_{k=l_n}^{\infty} \frac{|\mathrm{cov}(Y_{1n}, Y_{(k+1)n})|}{p_n} = 0$$

from (3.45). Therefore

$$\lim_{n\to\infty} \mathrm{Var}\left(\tilde{S}_n(\tilde{I}_{1n})\right) = \lim_{n\to\infty} I_0 + \lim_{h\to\infty}\lim_{n\to\infty} I_1 = \sigma_0^2(f,f) + 2\sum_{k=1}^{\infty} \sigma_k^2(f,f) =: \sigma^2(f,f)$$

as defined. To show that this infinite sum converges, it suffices to show that

$$\sum_{h=1}^{\infty} \mu_h(\{(\mathbf{x}, \mathbf{x}')| \|\mathbf{x}\| > 1, \|\mathbf{x}'\| > 1\}) < \infty.$$

This follows from (3.22) in condition (**M**), for if

$$\sum_{h=1}^{\infty} \mu_h(\{(\mathbf{x}, \mathbf{x}')| \|\mathbf{x}\| > 1, \|\mathbf{x}'\| > 1\}) = \infty,$$

then

$$\begin{aligned}\limsup_{n\to\infty} \frac{1}{p_n}\sum_{j=h}^{l_n} \mathbb{P}(\|\mathbf{X}_0\| > r_n, \|\mathbf{X}_j\| > r_n) &\geq \liminf_{n\to\infty} \sum_{j=h}^{l_n} \mathbb{P}(\|\mathbf{X}_0\| > r_n, \|\mathbf{X}_j\| > r_n | \|\mathbf{X}_0\| > r_n) \\ &\geq \sum_{j=h}^{\infty} \mu_j(\{(\mathbf{x}, \mathbf{x}')| \|\mathbf{x}\| > 1, \|\mathbf{x}'\| > 1\}) = \infty,\end{aligned}$$

which leads to a contradiction.

To apply the central limit theorem, we verify the Lindeberg's condition,

$$\mathbb{E}\left[(\tilde{S}_n(\tilde{I}_{1n}))^2 \mathbf{1}_{\{|\tilde{S}_n(\tilde{I}_{1n})| > \epsilon\sqrt{np_n}\}}\right] \leq \mathbb{E}\left[\left(\sum_{j=1}^{1/p_n - l_n} Y_{jn}\right)^2 \mathbf{1}_{\{|\tilde{S}_n(\tilde{I}_{1n})| > \epsilon\sqrt{np_n}\}}\right]$$

$$
\begin{aligned}
&\leq \mathbb{E}\left[c(1/p_n - l_n)^2 \mathbf{1}_{\{|\tilde{S}_n(\tilde{I}_{1n})|>\epsilon\sqrt{np_n}\}}\right] \\
&\leq c\frac{1}{p_n^2}\mathbb{P}\left[|\tilde{S}_n(\tilde{I}_{1n})| > \epsilon\sqrt{np_n}\right] \\
&\leq c\frac{1}{p_n^2}\frac{\mathrm{Var}\left[\tilde{S}_n(\tilde{I}_{1n})\right]}{\epsilon^2 np_n} = O(\frac{1}{np_n^3}) \to 0.
\end{aligned}
$$

This completes the proof for the convergence of $\frac{1}{\sqrt{np_n}}S_n(1:n)$.

The joint convergence of $\frac{1}{\sqrt{np_n}}(S_n^{(1)}, S_n^{(2)})^T$ follows from the same line of argument together with the Crámer-Wold device. In particular,

$$
\frac{1}{np_n}\mathrm{cov}\left(S_n^{(i)}, S_n^{(j)}\right) = \sigma^2(f_i, f_j), \quad i, j = 1, 2.
$$

This completes the proof of the lemma. □

Remark 3.9.4. *Lemma 3.9.1 itself is a more general result of independent interest. The result can be generalized for functions f_i defined on $\overline{\mathbb{R}}^d\backslash\{\mathbf{0}\}$ with compact support. In this case, condition (3.22) should be modified to*

$$
\lim_{h\to\infty}\limsup_{n\to\infty}\frac{1}{p_n}\sum_{j=h}^{l_n}\mathbb{P}(\|\mathbf{X}_0\| > \epsilon r_n, \|\mathbf{X}_j\| > \epsilon r_n) = 0
$$

for some $\epsilon > 0$, where support$(f) \subseteq \overline{\mathbb{R}}^d\backslash B_\epsilon(\mathbf{0})$. Also, as seen during the proof of the lemma, the conditions on p_n, l_n, and α_t can be further relaxed.

Acknowledgement

The foreign exchange rate data are obtained from OANDA through the R-package 'qrmtools'. We would like to thank Bodhisattva Sen for helpful discussions. We would also like to thank the editor and referees for their many constructive and insightful comments.

Chapter 4

Fitting the linear preferential attachment model

4.1 Introduction

The preferential attachment mechanism, in which edges and nodes are added to the network based on probabilistic rules, provides an appealing description for the evolution of a network. The rule for how edges connect nodes depends on node degree; large degree nodes attract more edges. The idea is applicable to both directed and undirected graphs and is often the basis for studying social networks, collaborator and citation networks, and recommender networks. Elementary descriptions of the preferential attachment model can be found in Easley and Kleinberg (2010) while more mathematical treatments are available in Durrett (2010), van der Hofstad (2017), Bhamidi (2007). Also see Kolaczyk and Csárdi (2014) for a statistical survey of methods for network data, Rinaldo et al. (2013) for consideration of statistics of an undirected network and Yan et al. (2016) for asymptotics of a directed exponential random graph models. Limit theory for estimates of an undirected preferential attachment model was considered in Gao and van der Vaart (2017).

For many networks, empirical evidence supports the hypothesis that in- and out-degree distributions follow a power law. This property has been shown to hold in linear preferential attachment models, which makes preferential attachment an attractive choice for network modeling Durrett (2010), van der Hofstad (2017), Krapivsky et al. (2001), Krapivsky and

Redner (2001), Bollobás et al. (2003). While the marginal degree power laws in a simple linear preferential attachment model were established in Krapivsky et al. (2001), Krapivsky and Redner (2001), Bollobás et al. (2003), the joint regular variation (see Resnick (2008, 2007)) which is akin to a *joint power law*, was only recently established (Samorodnitsky et al., 2016, Resnick and Samorodnitsky, 2015). In addition, it was shown in Wang and Resnick (2016) that the joint probability mass function of the in- and out-degrees is multivariate regularly varying. This is a key result as the degrees of a network are integer-valued.

In this chapter, we discuss methods of fitting a simple linear preferential attachment model, which is parametrized by $\boldsymbol{\theta} = (\alpha, \beta, \gamma, \delta_{\text{in}}, \delta_{\text{out}})$. The first three parameters, α, β, γ, correspond to probabilities of the 3 scenarios for adding an edge and hence sum to 1, i.e., $\alpha + \beta + \gamma = 1$. The other two, δ_{in} and δ_{out}, are tuning parameters related to growth rates. The tail indices of the marginal power laws for the in- and out-degrees can be expressed as explicit functions of $\boldsymbol{\theta}$ (see (4.5) and (4.6) below). The graph $G(n) = (V(n), E(n))$, where $V(n)$ is the set of nodes and $E(n)$ is the set of edges at the nth iteration, evolves based on postulates that describe how new edges and nodes are formed. This construction of the network is Markov in the sense that the probabilistic rules for obtaining $G(n+1)$ once $G(n)$ is known do not require prior knowledge of earlier stages of the construction.

The Markov structure of the model allows us to construct a likelihood function based on observing $G(n_0), G(n_0+1), \dots, G(n_0+n)$. After deriving the likelihood function, we show that it has a unique maximum at $\hat{\boldsymbol{\theta}} = (\hat{\alpha}, \hat{\beta}, \hat{\gamma}, \hat{\delta}_{\text{in}}, \hat{\delta}_{\text{out}})$ and that the resulting maximum likelihood estimator is strongly consistent and asymptotically normal. The normality is proved using a martingale central limit theorem applied to the score function. The limiting distribution also reveals that $(\hat{\alpha}, \hat{\beta}, \hat{\gamma})$, $\hat{\delta}_{\text{in}}$, and $\hat{\delta}_{\text{out}}$ are asymptotically independent. From these results, asymptotic properties of the MLE for the power law indices can be derived.

For some network data, only a snapshot of the nodes and edges is available at a single point in time, that is, only $G(n)$ is available for some n. In such cases, we propose an estimation procedure for the parameters of the network using an approximation to the likelihood and method of moments. This also produces strongly consistent estimators. These estimators perform reasonably well compared to the MLE where the entire evolution of the network is known but predictably there is some loss of efficiency.

We illustrate the estimation procedure for both scenarios using simulated data. Simulation plays an important role in the process of modeling networks since it provides a way to assess the performance of model fitting procedures in the idealized setting of knowing the true model. Also, after fitting a model to real data, simulation provides a check on the quality of fit. Departures from model assumptions can often be detected via simulation of multiple realizations from the fitted network. Hence it is important to have efficient simulation algorithms for producing realizations of the preferential attachment network for a given set of parameter values. We adopt a simulation method, learned from Joyjit Roy, that was inspired by Atwood et al. (2015) and is similar to that of Tonelli et al. (2010).

Our fitting methods are implemented in a real data setting using the Dutch Wiki talk network (Kunegis, 2013). While one should not expect the simple 5-parameter (later extended to 7-parameter) linear preferential attachment model to fully explain a network with millions of edges, it does provide a reasonable fit to the tail behavior of the degree distributions. We are also able to detect important structural features in the network through fitting the model over separate time intervals.

Often it is difficult to believe in the existence of a true model, especially one whose parameters remain constant over time. Allowing, as we do, a preferential attachment model with only a few parameters and no possibility for node removal may seem simplistic and unrealistic for social network data. Of course, preferential attachment is only one mechanism for network formation and evidence for its use in fields outside data networks is mixed (Jones and Handcock, 2003a,b) and we restrict attention to linear preferential attachment. Even imperfect models have the potential to capture salient properties in the data, such as heavy-tailedness of the in-degree and out-degree distributions, and to identify departures from model assumptions. While maximum likelihood estimation is essentially the gold standard for cases when the underlying model is a good representation of the data, it may perform poorly in case the model is far from being appropriate. In Wan et al. (2018), we consider a semi-parametric estimation approach for network models that exhibit heavy-tailed degree distributions. This alternative estimation methodology borrows ideas from extreme value theory.

The rest of the chapter is structured as follows. In Section 4.2, we formulate the linear preferential attachment network model and present an efficient simulation method for the network. Section 4.3 gives parameter estimators when either the full history is known or when only a single snapshot in time is available. We test these estimators against simulated data in Section 4.5 and then explore the Wiki talk network in Section 4.6.

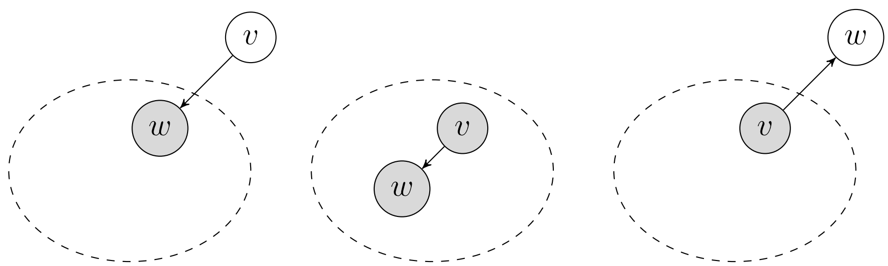
4.2 Model specification and simulation

In this section, we present the linear preferential attachment model in detail and provide a fast simulation algorithm for the network.

4.2.1 The linear preferential attachment model

The directed edge preferential attachment model (Bollobás et al., 2003, Krapivsky and Redner, 2001) constructs a growing directed random graph $G(n) = (V(n), E(n))$ whose dynamics depend on five non-negative real numbers α, β, γ, δ_{in} and δ_{out}, where $\alpha + \beta + \gamma = 1$ and $\delta_{\text{in}}, \delta_{\text{out}} > 0$. To avoid degenerate situations, assume that each of the numbers α, β, γ is strictly smaller than 1. We obtain a new graph $G(n)$ by adding one edge to the existing graph $G(n-1)$ and index the constructed graphs by the number n of edges in $E(n)$. We start with an arbitrary initial finite directed graph $G(n_0)$ with at least one node and n_0 edges. For $n > n_0$, $G(n) = (V(n), E(n))$ is a graph with $|E(n)| = n$ edges and a random number $|V(n)| = N(n)$ of nodes. If $u \in V(n)$, $D_{\text{in}}^{(n)}(u)$ and $D_{\text{in}}^{(n)}(u)$ denote the in- and out-degree of u respectively in $G(n)$. There are three scenarios that we call the α, β and γ-schemes, which are activated by flipping a 3-sided coin whose outcomes are $1, 2, 3$ with probabilities α, β, γ. More formally, we have an iid sequence of multinomial random variables $\{J_n, n > n_0\}$ with cells labelled $1, 2, 3$ and cell probabilities α, β, γ. Then the graph $G(n)$ is obtained from

$G(n-1)$ as follows.



α-scheme β-scheme γ-scheme

- If $J_n = 1$ (with probability α), append to $G(n-1)$ a new node $v \in V(n) \setminus V(n-1)$ and an edge (v, w) leading from v to an existing node $w \in V(n-1)$. Choose the existing node $w \in V(n-1)$ with probability depending on its in-degree in $G(n-1)$:

$$\mathbb{P}[\text{choose } w \in V(n-1)] = \frac{D_{\text{in}}^{(n-1)}(w) + \delta_{\text{in}}}{n - 1 + \delta_{\text{in}} N(n-1)}\,. \tag{4.1}$$

- If $J_n = 2$ (with probability β), add a directed edge (v, w) to $E(n-1)$ with $v \in V(n-1) = V(n)$ and $w \in V(n-1) = V(n)$ and the existing nodes v, w are chosen independently from the nodes of $G(n-1)$ with probabilities

$$\mathbb{P}[\text{choose } (v, w)] = \Big(\frac{D_{\text{out}}^{(n-1)}(v) + \delta_{\text{out}}}{n - 1 + \delta_{\text{out}} N(n-1)}\Big)\Big(\frac{D_{\text{in}}^{(n-1)}(w) + \delta_{\text{in}}}{n - 1 + \delta_{\text{in}} N(n-1)}\Big).$$

- If $J_n = 3$ (with probability γ), append to $G(n-1)$ a new node $w \in V(n) \setminus V(n-1)$ and an edge (v, w) leading from the existing node $v \in V(n-1)$ to the new node w. Choose the existing node $v \in V(n-1)$ with probability

$$\mathbb{P}[\text{choose } v \in V(n-1)] = \frac{D_{\text{out}}^{(n-1)}(v) + \delta_{\text{out}}}{n - 1 + \delta_{\text{out}} N(n-1)}\,. \tag{4.2}$$

Note that this construction allows the possibility of having self loops in the case where $J_n = 2$, but the proportion of edges that are self loops goes to 0 as $n \to \infty$. Also, multiple edges are allowed between two nodes.

4.2.2 Power law of degree distributions

Given an observed network with n edges, let $N_{ij}(n)$ denote the number of nodes with in-degree i and out-degree j. If the network is generated from the linear preferential attachment model described above, then from Bollobás et al. (2003), there exists a proper probability distribution $\{f_{ij}\}$ such that almost surely

$$\frac{N_{ij}(n)}{N(n)} \to f_{ij} =: \frac{p_{ij}}{1-\beta}, \quad n \to \infty. \tag{4.3}$$

Consider the limiting marginal in-degree distribution $p_i^{\text{in}} := \sum_j p_{ij}$. It is calculated from (Bollobás et al., 2003, Equation (3.10)) that

$$p_0^{\text{in}} = \frac{\alpha}{1 + a_1(\delta_{\text{in}})\delta_{\text{in}}},$$

and for $i \geq 1$,

$$p_i^{\text{in}} = \frac{\Gamma(i+\delta_{\text{in}})\Gamma(1+\delta_{\text{in}}+a_1(\delta_{\text{in}})^{-1})}{\Gamma(i+1+\delta_{\text{in}}+a_1(\delta_{\text{in}})^{-1})\Gamma(1+\delta_{\text{in}})} \left(\frac{\alpha\delta_{\text{in}}}{1+a_1(\delta_{\text{in}})\delta_{\text{in}}} + \frac{\gamma}{a_1(\delta_{\text{in}})} \right),$$

where

$$a_1(\lambda) := \frac{\alpha+\beta}{1+\lambda(1-\beta)}, \quad \lambda > 0.$$

Moreover, p_i^{in} satisfies

$$p_i^{\text{in}} := \sum_{j=0}^{\infty} p_{ij} \sim C_{\text{in}} i^{-\iota_{\text{in}}} \text{ as } i \to \infty, \quad \text{as long as } \alpha\delta_{\text{in}} + \gamma > 0, \tag{4.4}$$

for some finite positive constant C_{in}, where the power index

$$\iota_{\text{in}} = 1 + \frac{1+\delta_{\text{in}}(\alpha+\gamma)}{\alpha+\beta} \tag{4.5}$$

Similarly, the limiting marginal out-degree distribution has the same property:

$$p_j^{\text{out}} := \sum_{i=0}^{\infty} p_{ij} \sim C_{\text{out}} i^{-\iota_{\text{out}}} \text{ as } j \to \infty, \quad \text{as long as } \gamma\delta_{\text{out}} + \alpha > 0,$$

for C_{out} positive and

$$\iota_{\text{out}} = 1 + \frac{1+\delta_{\text{out}}(\alpha+\gamma)}{\beta+\gamma}. \tag{4.6}$$

Algorithm 1: Simulating a directed edge preferential attachment network

Algorithm

Input: $\alpha, \beta, \delta_{\text{in}}, \delta_{\text{out}}$, the parameter values; $G(n_0) = (V(n_0), E(n_0))$, the initialization graph; n, the targeted number edges

Output: $G(n) = (V(n), E(n))$, the resulted graph

```
t ← n_0
while t < n do
    N(t) ← |V(t)|
    Generate U ~ Uniform(0, 1)
    if U < α then
        v^(1) ← N(t) + 1
        v^(2) ← Node_Sample(E(t), 2, δ_in)
        V(t) ← Append(V(t), N(t) + 1)
    else if α ≤ U < α + β then
        v^(1) ← Node_Sample(E(t), 1, δ_out)
        v^(2) ← Node_Sample(E(t), 2, δ_in)
    else if U ≥ α + β then
        v^(1) ← Node_Sample(E(t), 1, δ_out)
        v^(2) ← N(t) + 1
        V(t) ← Append(V(t), N(t) + 1)
    E(t + 1) ← Append(E(t), (v^(1), v^(2)))
    t ← t + 1
end
return G(n) = (V(n), E(n))
```

Function *Node_Sample*

Input: $E(t)$, the edge list up to time t; $j = 1, 2$, the node to be sample, representing outgoing and incoming nodes, respectively; $\delta \in \{\delta_{\text{in}}, \delta_{\text{out}}\}$, the offset parameter

Output: the sampled node, v

```
Generate W ~ Uniform(0, t + N(t)δ)
if W ≤ t then
    v ← v^(j)_⌈W⌉
else if W > t then
    v ← ⌈(W − t)/δ⌉
return v
```

4.2.3 Simulation algorithm

We describe an efficient simulation procedure for the preferential attachment network given the parameter values $(\alpha, \beta, \gamma, \delta_{\text{in}}, \delta_{out})$, where $\alpha + \beta + \gamma = 1$. The simulation cost of the algorithm is linear in time. This algorithm, which was provided by Joyjit Roy during his graduate work at Cornell University, is presented below for completeness. Note that this simulation algorithm is specifically designed for the case where the preferential attachment probabilities (4.1)–(4.2) are linear in the degrees. A similar idea for the simulation of the Yule-Simon process appeared in Tonelli et al. (2010). Efficient simulation methods for the case where the preferential attachment probabilities are non-linear are studied in Atwood et al. (2015), where their algorithm trades some efficiency for the flexibility to model non-linear preferential attachment.

Using the notation from the introduction, at time $t = 0$, we initiate with an arbitrary graph $G(n_0) = (V(n_0), E(n_0))$ of n_0 edges, where the elements of $E(n_0)$ are represented in form of $(v_i^{(1)}, v_i^{(2)}) \in V(n_0) \times V(n_0)$, $i = 1, \dots, n_0$, with $v_i^{(1)}, v_i^{(2)}$ denoting the outgoing and incoming vertices of the edge, respectively. To grow the network, we update the network at each stage from $G(n-1)$ to $G(n)$ by adding a new edge $(v_n^{(1)}, v_n^{(2)})$. Assume that the nodes are labeled using positive integers starting from 1 according to the time order in which they are created, and let the random number $N(n) = |V(n)|$ denote the total number of nodes in $G(n)$.

Let us consider the situation where an existing node is to be chosen from $V(n)$ as the vertex of the new edge. Naively sampling from the multinomial distribution requires $O(N(n))$ evaluations, where $N(n)$ increases linearly with n. Therefore the total cost to simulate a network of n edges is $O(n^2)$. This is significantly burdensome when n is large, which is usually the case for observed networks. Algorithm 1 describes a simulation algorithm which uses the alias method (Kronmal and Jr., 1979) for node sampling. Here sampling an existing node from $V(n)$ requires only constant execution time, regardless of n. Hence the cost to simulate $G(n)$ is only $O(n)$. This method allows generation of a graph with 10^7 nodes on a personal laptop in less than 5 seconds.

To see that the algorithm indeed produces the intended network, it suffices to consider the case of sampling an existing node from $V(n-1)$ as the incoming vertex of the new edge. In the function Node_Sample in Algorithm 1, we generate $W \sim \text{Uniform}(0, n-1+N(n-1)\delta_{\text{in}})$ and set

$$v \leftarrow v^{(j)}_{\lceil W \rceil} \mathbf{1}_{\{W \le n-1\}} + \left\lceil \frac{W-(n-1)}{\delta_{\text{in}}} \right\rceil \mathbf{1}_{\{W > n-1\}}.$$

Then

$$\begin{aligned} \mathbb{P}(v=w) &= \mathbb{P}\left(v^{(j)}_{\lceil W \rceil} = w\right) \mathbb{P}(W \le n-1) \\ &\quad + \mathbb{P}\left(\left\lceil \frac{W-(n-1)}{\delta_{\text{in}}} \right\rceil = w\right) \mathbb{P}(W > n-1) \\ &= \frac{D^{(n-1)}_{\text{in}}(w)}{n-1} \frac{n-1}{n-1+N(n-1)\delta_{\text{in}}} \\ &\quad + \frac{1}{N(n-1)} \frac{N(n-1)\delta_{\text{in}}}{n-1+N(n-1)\delta_{\text{in}}} \\ &= \frac{D^{(n-1)}_{\text{in}}(w) + \delta_{\text{in}}}{n-1+N(n-1)\delta_{\text{in}}}, \end{aligned}$$

which corresponds to the desired selection probability (4.1).

4.3 Parameter estimation: MLE based on the full network history

In this section, we estimate the preferential attachment parameter vector $\boldsymbol{\theta} = (\alpha, \beta, \delta_{\text{in}}, \delta_{\text{out}})$ under two assumptions about what data is available. In the first scenario, the full evolution of the network is observed, from which the likelihood function can be computed. The resulting MLE is strongly consistent and asymptotically normal. For the second scenario, the data only consist of one snapshot of the network with n edges, without the knowledge of the network history that produced these edges. For this scenario we give an estimation approach through approximating the score function and moment matching, which produces parameter estimators that are also strongly consistent but less efficient than those based on the full evolution of the network. In both cases, the estimators are uniquely determined.

4.3.1 Likelihood calculation

Assume the network begins with the graph $G(n_0)$ (consisting of n_0 edges) and then evolves according to the description in Section 4.2.1 with parameters $(\alpha, \beta, \delta_{\text{in}}, \delta_{\text{out}})$, where $\delta_{\text{in}}, \delta_{\text{out}} > 0$ and α, β are non-negative probabilities. The γ is implicitly defined by $\gamma = 1 - \alpha - \beta$. To avoid trivial cases, we will also assume $\alpha, \beta, \gamma < 1$ for the rest of the chapter. For MLE estimation we restrict the parameter space for $\delta_{\text{in}}, \delta_{\text{out}}$ to be $[\epsilon, K]$, for some sufficiently small $\epsilon > 0$ and large K. In particular, the true value of $\delta_{\text{in}}, \delta_{\text{out}}$ is assumed to be contained in (ϵ, K). Let $e_t = (v_t^{(1)}, v_t^{(2)})$ be the newly created edge when the random graph evolves from $G(t-1)$ to $G(t)$. We sometimes refer to t as the time rather than the number of edges.

Assume we observe the initial graph $G(n_0)$ and the edges $\{e_t\}_{t=n_0+1}^n$ in the order of their formation. For $t = n_0+1, \dots, n$, the values of the following variables are known:

- $N(t)$, the number of nodes in graph $G(t)$;
- $D_{\text{in}}^{(t-1)}(v)$, $D_{\text{out}}^{(t-1)}(v)$, the in- and out-degree of node v in $G(t-1)$, for all $v \in V(t-1)$;
- J_t, the scenario under which e_t is created.

Then the likelihood function is

$$\begin{aligned} &L(\alpha, \beta, \delta_{\text{in}}, \delta_{\text{out}} |\, G(n_0), (e_t)_{t=n_0+1}^n) \\ &\qquad = \prod_{t=n_0+1}^n \left(\alpha \frac{D_{\text{in}}^{(t-1)}(v_t^{(2)}) + \delta_{\text{in}}}{t-1+\delta_{\text{in}} N(t-1)} \right)^{\mathbf{1}_{\{J_t=1\}}} \\ &\qquad \times \prod_{t=n_0+1}^n \left(\beta \Big(\frac{D_{\text{in}}^{(t-1)}(v_t^{(2)}) + \delta_{\text{in}}}{t-1+\delta_{\text{in}} N(t-1)} \Big) \Big(\frac{D_{\text{out}}^{(t-1)}(v_t^{(1)}) + \delta_{\text{out}}}{t-1+\delta_{\text{out}} N(t-1)} \Big) \right)^{\mathbf{1}_{\{J_t=2\}}} \\ &\qquad \times \prod_{t=n_0+1}^n \left((1-\alpha-\beta) \frac{D_{\text{out}}^{(t-1)}(v_t^{(1)}) + \delta_{\text{out}}}{t-1+\delta_{\text{out}} N(t-1)} \right)^{\mathbf{1}_{\{J_t=3\}}} \end{aligned} \tag{4.7}$$

and the log likelihood function is

$$\begin{aligned} &\log L(\alpha, \beta, \delta_{\text{in}}, \delta_{\text{out}} |\, G(n_0), (e_t)_{t=n_0+1}^n) \qquad (4.8) \\ &= \log \alpha \sum_{t=n_0+1}^n \mathbf{1}_{\{J_t=1\}} + \log \beta \sum_{t=n_0+1}^n \mathbf{1}_{\{J_t=2\}} + \log(1-\alpha-\beta) \sum_{t=n_0+1}^n \mathbf{1}_{\{J_t=3\}} \\ &\quad + \sum_{t=n_0+1}^n \log \left(D_{\text{in}}^{(t-1)}(v_t^{(2)}) + \delta_{\text{in}} \right) \mathbf{1}_{\{J_t \in \{1,2\}\}} \end{aligned}$$

$$
\begin{aligned}
&+\sum_{t=n_0+1}^{n} \log\left(D_{\text{out}}^{(t-1)}(v_t^{(1)})+\delta_{\text{out}}\right)\mathbf{1}_{\{J_t\in\{2,3\}\}}\\
&\quad-\sum_{t=n_0+1}^{n} \log(t-1+\delta_{\text{in}}N(t-1))\mathbf{1}_{\{J_t\in\{1,2\}\}}\\
&\quad\quad-\sum_{t=n_0+1}^{n} \log(t-1+\delta_{\text{out}}N(t-1))\mathbf{1}_{\{J_t\in\{2,3\}\}}.
\end{aligned}
$$

The score functions for $\alpha, \beta, \delta_{\text{in}}, \delta_{\text{out}}$ are calculated as follows:

$$
\begin{aligned}
&\frac{\partial}{\partial\alpha}\log L(\alpha,\beta,\delta_{\text{in}},\delta_{\text{out}}|\ G(n_0),(e_t)_{t=n_0+1}^{n})\\
&\qquad = \frac{1}{\alpha}\sum_{t=n_0+1}^{n}\mathbf{1}_{\{J_t=1\}} - \frac{1}{1-\alpha-\beta}\sum_{t=n_0+1}^{n}\mathbf{1}_{\{J_t=3\}}, \qquad (4.9)\\
&\frac{\partial}{\partial\beta}\log L(\alpha,\beta,\delta_{\text{in}},\delta_{\text{out}}|\ G(n_0),(e_t)_{t=n_0+1}^{n})\\
&\qquad = \frac{1}{\beta}\sum_{t=n_0+1}^{n}\mathbf{1}_{\{J_t=2\}} - \frac{1}{1-\alpha-\beta}\sum_{t=n_0+1}^{n}\mathbf{1}_{\{J_t=3\}}, \qquad (4.10)\\
&\frac{\partial}{\partial\delta_{\text{in}}}\log L(\alpha,\beta,\delta_{\text{in}},\delta_{\text{out}}|\ G(n_0),(e_t)_{t=n_0+1}^{n})\\
&\qquad = \sum_{t=n_0+1}^{n}\frac{1}{D_{\text{in}}^{(t-1)}(v_t^{(2)})+\delta_{\text{in}}}\mathbf{1}_{\{J_t\in\{1,2\}\}}\\
&\qquad\quad - \sum_{t=n_0+1}^{n}\frac{N(t-1)}{t-1+\delta_{\text{in}}N(t-1)}\mathbf{1}_{\{J_t\in\{1,2\}\}}, \qquad (4.11)\\
&\frac{\partial}{\partial\delta_{\text{out}}}\log L(\alpha,\beta,\delta_{\text{in}},\delta_{\text{out}}|\ G(n_0),(e_t)_{t=n_0+1}^{n})\\
&\qquad = \sum_{t=n_0+1}^{n}\frac{1}{D_{\text{out}}^{(t-1)}(v_t^{(1)})+\delta_{\text{out}}}\mathbf{1}_{\{J_t\in\{2,3\}\}}\\
&\qquad\quad - \sum_{t=n_0+1}^{n}\frac{N(t-1)}{t-1+\delta_{\text{out}}N(t-1)}\mathbf{1}_{\{J_t\in\{2,3\}\}}.
\end{aligned}
$$

Note that the score functions (4.9), (4.10) for α and β do not depend on δ_{in} and δ_{out}. One can show that the Hessian matrix of the log-likelihood for (α, β) is positive definite. Setting (4.9) and (4.10) to zero gives the unique MLE estimates for α and β,

$$
\hat{\alpha}^{MLE} = \frac{1}{n-n_0}\sum_{t=n_0+1}^{n}\mathbf{1}_{\{J_t=1\}}, \qquad (4.12)
$$

$$
\hat{\beta}^{MLE} = \frac{1}{n-n_0}\sum_{t=n_0+1}^{n}\mathbf{1}_{\{J_t=2\}}. \qquad (4.13)
$$

These estimates are strongly consistent by applying the strong law of large numbers for the $\{J_t\}$ sequence.

Next, consider the first term of the score function for δ_{in} in (4.11), and we have

$$\begin{aligned}\sum_{t=n_0+1}^{n} \frac{1}{D_{\text{in}}^{(t-1)}(v_t^{(2)}) + \delta_{\text{in}}} \mathbf{1}_{\{J_t \in \{1,2\}\}} \\ = \sum_{i=0}^{\infty} \frac{1}{i+\delta_{\text{in}}} \sum_{t=n_0+1}^{n} \mathbf{1}_{\left\{D_{\text{in}}^{(t-1)}(v_t^{(2)})=i, J_t \in \{1,2\}\right\}}.\end{aligned}$$

Observe that $\left\{D_{\text{in}}^{(t-1)}(v_t^{(2)}) = i, J_t \in \{1,2\}\right\}$ describes the event that the in-degree of node $v_t^{(2)} \in V(t-1)$ is i at time $t-1$ and is augmented to $i+1$ at time t. For each $i \geq 1$, such an event happens at some stage $t \in \{n_0+1, n_0+2, \ldots, n\}$ only for those nodes with in-degree $\leq i$ at time n_0 and in-degree $> i$ at time n. Let $N_{ij}(n)$ denote the number of nodes with in-degree i and out-degree j at time n, and $N_i^{\text{in}}(n)$ and $N_{>i}^{\text{in}}(n)$ to be the number of nodes with in-degree equal to i and greater than i, respectively, i.e.,

$$N_i^{\text{in}}(n) = \sum_{j=0}^{\infty} N_{ij}(n), \quad N_{>i}^{\text{in}}(n) = \sum_{k>i} N_k^{\text{in}}(n).$$

Then

$$\sum_{t=n_0+1}^{n} \mathbf{1}_{\left\{D_{\text{in}}^{(t-1)}(v_t^{(2)})=i, J_t \in \{1,2\}\right\}} = N_{>i}^{\text{in}}(n) - N_{>i}^{\text{in}}(n_0), \quad i \geq 1.$$

On the other hand, when $i=0$, $\left\{D_{\text{in}}^{(t-1)}(v_t^{(2)}) = 0, J_t \in \{1,2\}\right\}$ occurs for some t if and only if all of the following three events happen:

(i) $v_t^{(2)}$ has in-degree > 0 at time n;

(ii) $v_t^{(2)}$ does not have in-degree > 0 at time n_0;

(iii) $v_t^{(2)}$ was not created under the γ-scheme (otherwise it would have been born with in-degree 1).

This implies:

$$\sum_{t=n_0+1}^{n} \mathbf{1}_{\left\{D_{\text{in}}^{(t-1)}(v_t^{(2)})=0, J_t \in \{1,2\}\right\}} = N_{>0}^{\text{in}}(n) - N_{>0}^{\text{in}}(n_0) - \sum_{t=n_0+1}^{n} \mathbf{1}_{\{J_t=3\}},$$

since there are, in total, $\sum_{t=n_0+1}^n \mathbf{1}_{\{J_t=3\}}$ nodes created under the γ-scheme. Therefore,

$$\begin{aligned}\sum_{t=n_0+1}^{n} \frac{1}{D_{\text{in}}^{(t-1)}(v_t^{(2)}) + \delta_{\text{in}}} \mathbf{1}_{\{J_t\in\{1,2\}\}} &= \sum_{i=0}^{\infty} \frac{1}{i+\delta_{\text{in}}} \sum_{t=n_0+1}^{n} \mathbf{1}_{\left\{D_{\text{in}}^{(t-1)}(v_t^{(2)})=i, J_t\in\{1,2\}\right\}} \\ &= \sum_{i=0}^{\infty} \frac{N_{>i}^{\text{in}}(n) - N_{>i}^{\text{in}}(n_0)}{i+\delta_{\text{in}}} - \frac{\sum_{t=n_0+1}^n \mathbf{1}_{\{J_t=3\}}}{\delta_{\text{in}}}. \quad (4.14)\end{aligned}$$

Setting the score function (4.11) for δ_{in} to 0 and dividing both sides by $n-n_0$ leads to

$$\begin{aligned}\frac{1}{n-n_0} \sum_{i=0}^{\infty} \frac{N_{>i}^{\text{in}}(n) - N_{>i}^{\text{in}}(n_0)}{i+\delta_{\text{in}}} - \frac{1}{\delta_{\text{in}}(n-n_0)} \sum_{t=n_0+1}^{n} \mathbf{1}_{\{J_t=3\}} \\ - \frac{1}{n-n_0} \sum_{t=n_0+1}^{n} \frac{N(t-1)}{t-1+\delta_{\text{in}} N(t-1)} \mathbf{1}_{\{J_t\in\{1,2\}\}} = 0, \quad (4.15)\end{aligned}$$

where the only unknown parameter is δ_{in}. In Section 4.3.2, we show that the solution to (4.15) actually maximizes the likelihood function in δ_{in}. Similarly, the MLE for δ_{out} can be solved from

$$\begin{aligned}\frac{1}{n-n_0} \sum_{j=0}^{\infty} \frac{N_{>j}^{\text{out}}(n) - N_{>j}^{\text{out}}(n_0)}{j+\delta_{\text{out}}} - \frac{\frac{1}{n-n_0}\sum_{t=n_0+1}^n \mathbf{1}_{\{J_t=1\}}}{\delta_{\text{out}}} \\ - \frac{1}{n-n_0} \sum_{t=n_0+1}^{n} \frac{N(t-1)}{t-1+\delta_{\text{out}} N(t-1)} \mathbf{1}_{\{J_t\in\{2,3\}\}} = 0,\end{aligned}$$

where $N_{>j}^{\text{out}}(n)$ is defined in the same fashion as $N_{>i}^{\text{in}}(n)$.

Remark 4.3.1. *The arguments leading to* (4.14) *allow us to rewrite the likelihood function* (4.7)*:*

$$\begin{aligned}& L(\alpha, \beta, \delta_{in}, \delta_{out} |\, G(n_0), (e_t)_{t=n_0+1}^n) \\ &= \alpha^{\sum_{t=n_0+1}^n \mathbf{1}_{\{J_t=1\}}} \beta^{\sum_{t=n_0+1}^n \mathbf{1}_{\{J_t=2\}}} (1-\alpha-\beta)^{\sum_{t=n_0+1}^n \mathbf{1}_{\{J_t=3\}}} \\ &\quad \times \prod_{t=n_0+1}^{n} (t-1+\delta_{in}N(t-1))^{-\mathbf{1}_{\{J_t\in\{1,2\}\}}} (t-1+\delta_{out}N(t-1))^{-\mathbf{1}_{\{J_t\in\{2,3\}\}}} \\ &\quad \times \prod_{t=n_0+1}^{n} \left[\prod_{i=0}^{\infty} (i+\delta_{in})^{\mathbf{1}_{\left\{D_{in}^{(t-1)}(v_t^{(2)})=i, J_t\in\{1,2\}\right\}}} \prod_{j=0}^{\infty} (j+\delta_{out})^{\mathbf{1}_{\left\{D_{out}^{(t-1)}(v_t^{(1)})=j, J_t\in\{2,3\}\right\}}} \right] \\ &= \alpha^{\sum_{t=n_0+1}^n \mathbf{1}_{\{J_t=1\}}} \beta^{\sum_{t=n_0+1}^n \mathbf{1}_{\{J_t=2\}}} (1-\alpha-\beta)^{\sum_{t=n_0+1}^n \mathbf{1}_{\{J_t=3\}}} \\ &\quad \times \prod_{t=n_0+1}^{n} \Big[(t-1+\delta_{in}N(t-1))^{-\mathbf{1}_{\{J_t\in\{1,2\}\}}} (t-1+\delta_{out}N(t-1))^{-\mathbf{1}_{\{J_t\in\{2,3\}\}}}\end{aligned}$$

$$\delta_{in}^{-\mathbf{1}_{\{J_t=3\}}}\ \delta_{out}^{-\mathbf{1}_{\{J_t=1\}}}\Big] \times \prod_{i=0}^{\infty} (i+\delta_{in})^{N^{in}_{>i}(n)-N^{in}_{>i}(n_0)} \prod_{j=0}^{\infty}(j+\delta_{out})^{N^{out}_{>j}(n)-N^{out}_{>j}(n_0)}.$$

Hence by the factorization theorem, $N(n_0)$, $(J_t)_{t=n_0+1}^n$, $(N^{in}_{>i}(n)-N^{in}_{>i}(n_0))_{i\geq 0}$, $(N^{out}_{>j}(n)-N^{out}_{>j}(n_0))_{j\geq 0}$ *are sufficient statistics for* $(\alpha,\beta,\delta_{in},\delta_{out})$.

4.3.2 Consistency of MLE

We remarked after (4.12) and (4.13) that $\hat{\alpha}^{MLE}$ and $\hat{\beta}^{MLE}$ converge almost surely to α and β. We now prove that the MLE of $(\delta_{\text{in}},\delta_{\text{out}})$ is also strongly consistent. Note that if we initiate the network with $G(n_0)$ (for both n_0 and $N(n_0)$ finite), then almost surely for all $i,j\geq 0$,

$$\frac{N^{\text{in}}_{>i}(n_0)}{n} \leq \frac{N(n_0)}{n} \to 0, \quad \frac{N^{\text{out}}_{>j}(n_0)}{n} \leq \frac{N(n_0)}{n} \to 0, \quad \text{as } n\to\infty,$$

and $(n-n_0)/n\to 1$. In other words, n_0, $N^{\text{in}}_{>i}(n_0)$, $N^{\text{out}}_{>j}(n_0)$ are all $o(n)$. So for simplicity, we assume that the graph is initiated with finitely many nodes and no edges, that is, $n_0=0$ and $N(0)\geq 1$. In particular, these assumptions imply the sum of the in-degrees at time n is equal to n.

Let $\Psi_n(\cdot),\Phi_n(\cdot)$ be the functional forms of the terms in the log-likelihood function (4.8) involving δ_{in} and δ_{out} respectively, normalized by $1/n$, i.e.,

$$\begin{aligned}\Psi_n(\lambda) &:= \sum_{i=0}^{\infty}\frac{N^{\text{in}}_{>i}(n)}{n}\log(i+\lambda) - \frac{\log\lambda}{n}\sum_{t=1}^{n}\mathbf{1}_{\{J_t=3\}}\\ &\quad - \frac{1}{n}\sum_{t=1}^{n}\log\left(t-1+\lambda N(t-1)\right)\mathbf{1}_{\{J_t\in\{1,2\}\}},\\ \Phi_n(\mu) &:= \sum_{j=0}^{\infty}\frac{N^{\text{out}}_{>j}(n)}{n}\log(j+\mu) - \frac{\log\mu}{n}\sum_{t=1}^{n}\mathbf{1}_{\{J_t=1\}}\\ &\quad - \frac{1}{n}\sum_{t=1}^{n}\log\left(t-1+\mu N(t-1)\right)\mathbf{1}_{\{J_t\in\{2,3\}\}}.\end{aligned}$$

The following theorem gives the consistency of the MLE of δ_{in} and δ_{out}.

Theorem 4.3.2. *Suppose* $\delta_{in},\delta_{out}\in(\epsilon,K)\subset(0,\infty)$. *Define*

$$\hat{\delta}_{in}^{MLE} = \hat{\delta}_{in}^{MLE}(n) := \underset{\epsilon\leq\lambda\leq K}{\operatorname{argmax}}\,\Psi_n(\lambda), \quad \hat{\delta}_{out}^{MLE} = \hat{\delta}_{out}^{MLE}(n) := \underset{\epsilon\leq\mu\leq K}{\operatorname{argmax}}\,\Phi_n(\mu).$$

Then these are the MLE estimators of $\delta_{in}, \delta_{out}$ and they are strongly consistent; that is,

$$\hat{\delta}_{in}^{MLE} \xrightarrow{a.s.} \delta_{in}, \qquad \hat{\delta}_{out}^{MLE} \xrightarrow{a.s.} \delta_{out}, \qquad n \to \infty.$$

Proof of Theorem 4.3.2. We only verify the consistency of $\hat{\delta}_{\text{in}}^{MLE}$ since similar arguments apply to $\hat{\delta}_{\text{out}}^{MLE}$. Define

$$\psi_n(\lambda) \;:=\; \Psi_n'(\lambda) \;=\; \sum_{i=0}^{\infty} \frac{N_{>i}^{\text{in}}(n)/n}{i+\lambda} - \frac{\frac{1}{n}\sum_{t=1}^{n} \mathbf{1}_{\{J_t=3\}}}{\lambda} - \frac{1}{n}\sum_{t=1}^{n} \frac{N(t-1)}{t-1+\lambda N(t-1)} \mathbf{1}_{\{J_t \in \{1,2\}\}}.$$

Let us consider a limit version of ψ_n:

$$\psi(\lambda) \;:=\; \sum_{i=0}^{\infty} \frac{p_{>i}^{\text{in}}(\delta_{\text{in}})}{i+\lambda} - \frac{\gamma}{\lambda} - (1-\beta)a_1(\lambda), \tag{4.16}$$

where $p_{>i}^{\text{in}}(\delta_{\text{in}}) := \sum_{k>i} p_k^{\text{in}}(\delta_{\text{in}})$ with $p_k^{\text{in}}(\delta_{\text{in}}) := p_k^{\text{in}}$ as defined in (4.4), and

$$a_1(\lambda) \;:=\; \frac{\alpha+\beta}{1+\lambda(1-\beta)}, \qquad \lambda > 0.$$

Here we write $p_i^{\text{in}}(\delta_{\text{in}})$ to emphasize the dependence on δ_{in}. In Lemmas 4.7.1 and 4.7.2, provided in Section 4.7, it is shown that $\psi(\cdot)$ has a unique zero at δ_{in}, where $\psi(\lambda) > 0$ when $\lambda < \delta_{\text{in}}$ and $\psi(\lambda) < 0$ when $\lambda > \delta_{\text{in}}$, and

$$\sup_{\lambda \geq \epsilon} |\psi_n(\lambda) - \psi(\lambda)| \;\to\; 0. \tag{4.17}$$

Since ψ is continuous, for any $\kappa > 0$ arbitrarily small, there exists $\varepsilon_\kappa > 0$ such that $\psi(\lambda) > \varepsilon_\kappa$ for $\lambda \in [\epsilon, \delta_{\text{in}} - \kappa]$ and $\psi(\lambda) < -\varepsilon_\kappa$ for $\lambda \in [\delta_{\text{in}} + \kappa, K]$. From (4.17),

$$\mathbb{P}\left(\exists N_\kappa \; s.t. \sup_{n > N_\kappa} \sup_{\lambda \in [\epsilon, K]} |\psi_n(\lambda) - \psi(\lambda)| < \varepsilon_\kappa/2 \right) \;=\; 1. \tag{4.18}$$

Note $\sup_{\lambda \in [\epsilon, K]} |\psi_n(\lambda) - \psi(\lambda)| < \varepsilon_\kappa/2$ implies

$$\psi_n(\lambda) \;\geq\; \psi(\lambda) - \sup_{\lambda \in [\epsilon, K]} |\psi_n(\lambda) - \psi(\lambda)| \;\geq\; \varepsilon_\kappa - \varepsilon_\kappa/2 \;>\; 0, \quad \lambda \in [\epsilon, \delta_{\text{in}} - \kappa),$$

and

$$\psi_n(\lambda) \;\leq\; \psi(\lambda) + \sup_{\lambda \in [\epsilon, K]} |\psi_n(\lambda) - \psi(\lambda)| \;\leq\; -\varepsilon_\kappa + \varepsilon_\kappa/2 \;<\; 0, \quad \lambda \in (\delta_{\text{in}} + \kappa, K].$$

These jointly indicate that $\delta_{\text{in}} - \kappa \leq \hat{\delta}_{\text{in}}^{MLE} \leq \delta_{\text{in}} + \kappa$. Hence (4.18) implies

$$\mathbb{P}\left(\lim_{n\to\infty} |\hat{\delta}_{\text{in}}^{MLE} - \delta_{\text{in}}| \leq \kappa \right) \;=\; 1,$$

for arbitrary $\kappa > 0$. That is, $\hat{\delta}_{\text{in}}^{MLE} \xrightarrow{\text{a.s.}} \delta_{\text{in}}$. □

4.3.3 Asymptotic normality of MLE

In the following theorem, we establish the asymptotic normality for the MLE estimator

$$\hat{\boldsymbol{\theta}}_n^{MLE} = (\hat{\alpha}^{MLE}, \hat{\beta}^{MLE}, \hat{\delta}_{\text{in}}^{MLE}, \hat{\delta}_{\text{out}}^{MLE}).$$

Theorem 4.3.3. *Let $\hat{\boldsymbol{\theta}}_n^{MLE}$ be the MLE estimator for $\boldsymbol{\theta}$, the parameter vector of the preferential attachment model. Then*

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_n^{MLE} - \boldsymbol{\theta}) \xrightarrow{d} N\left(\mathbf{0}, \Sigma(\boldsymbol{\theta})\right),$$

where

$$\Sigma^{-1}(\boldsymbol{\theta}) = I(\boldsymbol{\theta}) := \begin{bmatrix} \frac{1-\beta}{\alpha(1-\alpha-\beta)} & \frac{1}{1-\alpha-\beta} & 0 & 0 \\ \frac{1}{1-\alpha-\beta} & \frac{1-\alpha}{\beta(1-\alpha-\beta)} & 0 & 0 \\ 0 & 0 & I_{in} & 0 \\ 0 & 0 & 0 & I_{out} \end{bmatrix}, \tag{4.19}$$

with

$$\begin{aligned} I_{in} &:= \sum_{i=0}^{\infty} \frac{p_{>i}^{in}}{(i+\delta_{in})^2} - \frac{\gamma}{\delta_{in}^2} - \frac{(\alpha+\beta)(1-\beta)^2}{\left(1+\delta_{in}(1-\beta)\right)^2}, \\ I_{out} &:= \sum_{j=0}^{\infty} \frac{p_{>j}^{out}}{(j+\delta_{out})^2} - \frac{\alpha}{\delta_{out}^2} - \frac{(\gamma+\beta)(1-\beta)^2}{\left(1+\delta_{out}(1-\beta)\right)^2}. \end{aligned} \tag{4.20}$$

In particular, $I(\boldsymbol{\theta})$ is the asymptotic Fisher information matrix for the parameters, and hence the MLE estimator is efficient.

Remark 4.3.4. *From Theorem 4.3.3, the estimators $(\hat{\alpha}^{MLE}, \hat{\beta}^{MLE})$, $\hat{\delta}_{in}^{MLE}$, and $\hat{\delta}_{out}^{MLE}$ are asymptotically independent.*

Proof of Theorem 4.3.3. We first show the limiting distributions for the MLE's, i.e. $(\hat{\alpha}^{MLE}, \hat{\beta}^{MLE})$, $\hat{\delta}_{\text{in}}^{MLE}$ and $\hat{\delta}_{\text{out}}^{MLE}$. From (4.12) and (4.13),

$$(\hat{\alpha}^{MLE}, \hat{\beta}^{MLE}) = \frac{1}{n}\sum_{t=1}^{n}\left(\mathbf{1}_{\{J_t=1\}}, \mathbf{1}_{\{J_t=2\}}\right),$$

where $\{J_t\}$ is a sequence of iid random variables. Hence the limiting distribution of the pair $\left(\hat{\alpha}^{MLE}, \hat{\beta}^{MLE}\right)$ follows directly from standard central limit theorem for sums of independent random variables.

Next we show the asymptotic normality for $\hat{\delta}_{\text{in}}^{MLE}$; the argument for $\hat{\delta}_{\text{out}}^{MLE}$ is similar. Recall from (4.11) that the score function for δ_{in} can be written as

$$\frac{\partial}{\partial \delta_{\text{in}}} \log L(\alpha, \beta, \delta_{\text{in}}, \delta_{\text{out}})\Big|_{\delta} =: \sum_{t=1}^{n} u_t(\delta),$$

where u_t is defined by

$$u_t(\delta) := \frac{1}{D_{\text{in}}^{(t-1)}(v_t^{(2)}) + \delta} \mathbf{1}_{\{J_t \in \{1,2\}\}} - \frac{N(t-1)}{t - 1 + \delta N(t-1)} \mathbf{1}_{\{J_t \in \{1,2\}\}}. \tag{4.21}$$

The MLE estimator $\hat{\delta}_{\text{in}}^{MLE}$ can be obtained by solving $\sum_{t=1}^{n} u_t(\delta) = 0$. By a Taylor expansion of $\sum_{t=1}^{n} u_t(\delta)$,

$$0 = \sum_{t=1}^{n} u_t(\hat{\delta}_{\text{in}}^{MLE}) = \sum_{t=1}^{n} u_t(\delta_{\text{in}}) + (\hat{\delta}_{\text{in}}^{MLE} - \delta_{\text{in}}) \sum_{t=1}^{n} \dot{u}_t(\hat{\delta}_{\text{in}}^{*}), \tag{4.22}$$

where $\dot{u}_t$ denotes the derivative of u_t and $\hat{\delta}_{\text{in}}^{*} = \delta_{\text{in}} + \xi(\hat{\delta}_{\text{in}}^{MLE} - \delta_{\text{in}})$ for some $\xi \in [0, 1]$. An elementary transformation of (4.22) gives

$$n^{1/2}(\hat{\delta}_{\text{in}}^{MLE} - \delta_{\text{in}}) = \left(-\frac{1}{n^{-1} \sum_{t=1}^{n} \dot{u}_t(\hat{\delta}_{\text{in}}^{*})} \right) \left(n^{-1/2} \sum_{t=1}^{n} u_t(\delta_{\text{in}}) \right).$$

To establish

$$n^{1/2}(\hat{\delta}_{\text{in}}^{MLE} - \delta_{\text{in}}) \xrightarrow{d} N(0, I_{\text{in}}^{-1}),$$

where I_{in} is as defined in (4.19), it suffices to show the following two results:

(i) $n^{-1/2} \sum_{t=1}^{n} u_t(\delta_{\text{in}}) \xrightarrow{d} N(0, I_{\text{in}})$,

(ii) $n^{-1} \sum_{t=1}^{n} \dot{u}_t(\hat{\delta}_{\text{in}}^{*}) \xrightarrow{p} -I_{\text{in}}$.

These are proved in Lemmas 4.7.3 and 4.7.4 in the Section 4.7.1, respectively.

To establish the joint asymptotic normality of the MLE estimator $\hat{\boldsymbol{\theta}}_n^{MLE}$, denote the joint score function vector for $\boldsymbol{\theta}$ by

$$\frac{\partial}{\partial \boldsymbol{\theta}} \log L(\boldsymbol{\theta}) =: \mathbf{S}_n(\boldsymbol{\theta}) = \left(S_n(\alpha), S_n(\beta), S_n(\delta_{\text{in}}), S_n(\delta_{\text{out}})\right)^T,$$

where $S_n(\alpha), S_n(\beta), S_n(\delta_{\text{in}}), S_n(\delta_{\text{out}})$ are the score functions for $\alpha, \beta, \delta_{\text{in}}, \delta_{\text{out}}$, respectively. A multivariate Taylor expansion gives

$$\mathbf{0} = \mathbf{S}_n\left(\hat{\boldsymbol{\theta}}_n^{MLE}\right) = \mathbf{S}_n(\boldsymbol{\theta}) + \dot{\mathbf{S}}_n\left(\hat{\boldsymbol{\theta}}_n^{*}\right)\left(\hat{\boldsymbol{\theta}}_n^{MLE} - \boldsymbol{\theta}\right), \tag{4.23}$$

where $\dot{\mathbf{S}}_n$ denotes the Hessian matrix of the log-likelihood function $\log L(\boldsymbol{\theta})$, and $\hat{\boldsymbol{\theta}}_n^* = \boldsymbol{\theta} + \boldsymbol{\xi} \circ \left(\hat{\boldsymbol{\theta}}_n^{MLE} - \boldsymbol{\theta}\right)$ for some vector $\boldsymbol{\xi} \in [0,1]^4$, where "$\circ$" denotes the Hadamard product. From Remark 4.3.1, the likelihood function $L(\boldsymbol{\theta})$ can be factored into

$$L(\boldsymbol{\theta}) \;=\; f_1(\alpha,\beta) f_2(\delta_{\text{in}}) f_3(\delta_{\text{out}}).$$

Hence

$$\frac{1}{n}\dot{\mathbf{S}}_n(\hat{\boldsymbol{\theta}}_n^*) \;=\; \begin{bmatrix} \frac{\partial^2 \log L_n(\hat{\boldsymbol{\theta}}_n^*)}{\partial \alpha^2} & \frac{\partial^2 \log L_n(\hat{\boldsymbol{\theta}}_n^*)}{\partial \alpha \partial \beta} & 0 & 0 \\ \frac{\partial^2 \log L_n(\hat{\boldsymbol{\theta}}_n^*)}{\partial \beta \partial \alpha} & \frac{\partial^2 \log L_n(\hat{\boldsymbol{\theta}}_n^*)}{\partial \beta^2} & 0 & 0 \\ 0 & 0 & \frac{\partial^2 \log L_n(\hat{\boldsymbol{\theta}}_n^*)}{\partial \delta_{\text{in}}^2} & 0 \\ 0 & 0 & 0 & \frac{\partial^2 \log L_n(\hat{\boldsymbol{\theta}}_n^*)}{\partial \delta_{\text{out}}^2} \end{bmatrix} \stackrel{p}{\to} I(\boldsymbol{\theta}) \qquad (4.24)$$

as implied in the previous part of the proof, where $I(\boldsymbol{\theta})$ (defined in (4.19)) is positive semi-definite.

Note that $(S_n(\alpha), S_n(\beta)), S_n(\delta_{\text{in}}), S_n(\delta_{\text{out}})$ are pairwise uncorrelated. As an example, observe that

$$\begin{aligned} \mathbb{E}[S_n(\alpha)S_n(\delta_{\text{in}})] \;&=\; \int \frac{\partial \log L(\boldsymbol{\theta})}{\partial \alpha}\frac{\partial \log L(\boldsymbol{\theta})}{\partial \delta_{\text{in}}} L(\boldsymbol{\theta}) d\mathbf{x} \\ &=\; \int \frac{\partial \log f_1(\alpha,\beta)}{\partial \alpha}\frac{\partial \log f_2(\delta_{\text{in}})}{\partial \delta_{\text{in}}} f_1(\alpha,\beta) f_2(\delta_{\text{in}}) f_3(\delta_{\text{out}}) d\mathbf{x} \\ &=\; \int \frac{\partial f_1(\alpha,\beta)}{\partial \alpha}\frac{\partial f_2(\delta_{\text{in}})}{\partial \delta_{\text{in}}} f_3(\delta_{\text{out}})\, d\mathbf{x} \\ &=\; \frac{\partial^2}{\partial \alpha \partial \delta_{\text{in}}} \int L(\boldsymbol{\theta}) d\mathbf{x} \\ &=\; 0 \;=\; \mathbb{E}[S_n(\alpha)]\mathbb{E}[S_n(\delta_{\text{in}})]. \end{aligned}$$

Using the Cramér-Wold device, the joint convergence of $\mathbf{S}_n(\boldsymbol{\theta})$ follows easily, i.e.,

$$n^{-1/2}\mathbf{S}_n(\boldsymbol{\theta}) \stackrel{d}{\to} N(\mathbf{0}, I(\boldsymbol{\theta})).$$

From here, the result of the theorem follows from (4.23) and (4.24). □

4.4 Parameter estimation based on one snapshot

Based only on the single snapshot $G(n)$, we propose a parameter estimation procedure. We assume that the choice of the snapshot does not depend on any endogenous information related to the network. The snapshot merely represents a point in time where the data is available. Since no information on the initial graph $G(n_0)$ is available, we merely assume n_0 and $N(n_0)$ are fixed and $n \to \infty$.

Among the sufficient statistics for $(\alpha, \beta, \delta_{\text{in}}, \delta_{\text{out}})$ derived in Remark 4.3.1, $\big(N^{\text{in}}_{>i}(n)\big)_{i\geq 0}$, $\big(N^{\text{out}}_{>j}(n)\big)_{j\geq 0}$ are computable from $G(n)$, but the $(J_t)_{t=1}^n$ are not. However, when n is large, we can use the following approximations according to the proof of Lemma 4.7.2:

$$\frac{1}{n}\sum_{t=n_0+1}^{n} \mathbf{1}_{\{J_t=3\}} \approx 1-\alpha-\beta,$$

and

$$\frac{1}{n}\sum_{t=n_0+1}^{n} \frac{N(t)}{t+\delta_{\text{in}}N(t)}\mathbf{1}_{\{J_t\in\{1,2\}\}} \approx (\alpha+\beta)\frac{1-\beta}{1+\delta_{\text{in}}(1-\beta)}.$$

Substituting in (4.15), we estimate δ_{in} in terms of α and β by solving

$$\sum_{i=0}^{\infty} \frac{N^{\text{in}}_{>i}(n)/n}{i+\delta_{\text{in}}} - \frac{1-\alpha-\beta}{\delta_{\text{in}}} - \frac{(\alpha+\beta)(1-\beta)}{1+(1-\beta)\delta_{\text{in}}} = 0. \tag{4.25}$$

Note that a strongly consistent estimator of β can be obtained directly from $G(n)$:

$$\tilde{\beta} = 1 - \frac{N(n)}{n} \xrightarrow{\text{a.s.}} \beta.$$

To obtain an estimate for α, we make use of the recursive formula for $\{p^{\text{in}}_i\}$ in (4.36a):

$$\left(1+\frac{(\alpha+\beta)\delta_{\text{in}}}{1+(1-\beta)\delta_{\text{in}}}\right)p^{\text{in}}_0 = \alpha, \tag{4.26}$$

and replace p^{in}_0 by $N^{\text{in}}_0(n)/n$ for large n,

$$\left(1+\frac{(\alpha+\beta)\delta_{\text{in}}}{1+(1-\beta)\delta_{\text{in}}}\right)\frac{N^{\text{in}}_0(n)}{n} = \alpha. \tag{4.27}$$

Plug the strongly consistent estimator $\tilde{\beta}$ into (4.25) and (4.27), and we claim that solving the system of equations:

$$\sum_{i=0}^{\infty} \frac{N^{\text{in}}_{>i}(n)/n}{i+\delta_{\text{in}}} - \frac{1-\alpha-\tilde{\beta}}{\delta_{\text{in}}} - \frac{(\alpha+\tilde{\beta})(1-\tilde{\beta})}{1+(1-\tilde{\beta})\delta_{\text{in}}} = 0, \tag{4.28a}$$

$$\left(1+\frac{(\alpha+\tilde{\beta})\delta_{\text{in}}}{1+(1-\tilde{\beta})\delta_{\text{in}}}\right)\frac{N^{\text{in}}_0(n)}{n} = \alpha, \tag{4.28b}$$

gives the unique solution $(\tilde{\alpha}, \tilde{\delta}_{\text{in}})$ which is strongly consistent for $(\alpha, \delta_{\text{in}})$.

Theorem 4.4.1. *The solution* $(\tilde{\alpha}, \tilde{\delta}_{in})$ *to the system of equations in* (4.28) *is unique and strongly consistent for* (α, δ_{in})*, i.e.*

$$\tilde{\alpha} \xrightarrow{a.s.} \alpha, \quad \tilde{\delta}_{in} \xrightarrow{a.s.} \delta_{in}.$$

The proof of Theorem 4.4.1 is given in Section 4.8.

The parameters $\tilde{\delta}_{\text{out}}$ and $\tilde{\gamma}$ can be estimated by a mirror argument. We summarize the estimation procedure for $(\alpha, \beta, \gamma, \delta_{\text{in}}, \delta_{\text{out}})$ from the snapshot $G(n)$ as follows:

1. Estimate β by $\tilde{\beta} = 1 - N(n)/n$.
2. Obtain $\tilde{\delta}^0_{\text{in}}$ by solving (i.e., matching (4.28a) and (4.28b))
$$\sum_{i=1}^{\infty} \frac{N^{\text{in}}_{>i}(n)}{n} \frac{i}{i+\delta_{\text{in}}}(1+\delta_{\text{in}}(1-\tilde{\beta})) \;=\; \frac{\frac{N^{\text{in}}_0(n)}{n} + \tilde{\beta}}{1 - \frac{N^{\text{in}}_0(n)}{n}\frac{\delta_{\text{in}}}{1+(1-\tilde{\beta})\delta_{\text{in}}}}.$$
3. Estimate α by
$$\tilde{\alpha}^0 \;=\; \frac{\frac{N^{\text{in}}_0(n)}{n} + \tilde{\beta}}{1 - \frac{N^{\text{in}}_0(n)}{n}\frac{\tilde{\delta}^0_{\text{in}}}{1+(1-\tilde{\beta})\tilde{\delta}^0_{\text{in}}}} - \tilde{\beta}.$$
4. Obtain $\tilde{\delta}^0_{\text{out}}$ by solving
$$\sum_{j=1}^{\infty} \frac{N^{\text{out}}_{>j}(n)}{n} \frac{j}{j+\delta_{\text{out}}}(1+\delta_{\text{out}}(1-\tilde{\beta})) \;=\; \frac{\frac{N^{\text{out}}_0(n)}{n} + \tilde{\beta}}{1 - \frac{N^{\text{out}}_0(n)}{n}\frac{\delta_{\text{out}}}{1+(1-\tilde{\beta})\delta_{\text{out}}}}.$$
5. Estimate γ by
$$\tilde{\gamma}^0 \;=\; \frac{\frac{N^{\text{out}}_0(n)}{n} + \tilde{\beta}}{1 - \frac{N^{\text{out}}_0(n)}{n}\frac{\tilde{\delta}^0_{\text{out}}}{1+(1-\tilde{\beta})\tilde{\delta}^0_{\text{out}}}} - \tilde{\beta}.$$

Note that even though all three estimators $\tilde{\alpha}^0, \tilde{\beta}, \tilde{\gamma}^0$ are strongly consistent and hence $\tilde{\alpha}^0 + \tilde{\beta} + \tilde{\gamma}^0 \xrightarrow{\text{a.s.}} 1$, Step 1–5 do not necessarily imply the strict equality

$$\tilde{\alpha}^0 + \tilde{\beta} + \tilde{\gamma}^0 \;=\; 1.$$

We recommend adding the following two steps for a re-normalization to overcome this defect.

6. Re-normalize the probabilities
$$(\tilde{\alpha}, \tilde{\beta}, \tilde{\gamma}) \leftarrow \left(\frac{\tilde{\alpha}^0(1-\tilde{\beta})}{\tilde{\alpha}^0 + \tilde{\gamma}^0}, \tilde{\beta}, \frac{\tilde{\gamma}^0(1-\tilde{\beta})}{\tilde{\alpha}^0 + \tilde{\gamma}^0} \right).$$

7. Plug $\tilde{\alpha}$ into (4.28a) to update the estimate of δ_{in}, i.e., solve for $\tilde{\delta}_{\text{in}}$ from

$$\sum_{i=0}^{\infty} \frac{N^{\text{in}}_{>i}(n)/n}{i+\tilde{\delta}_{\text{in}}} - \frac{1-\tilde{\alpha}-\tilde{\beta}}{\tilde{\delta}_{\text{in}}} - \frac{(\tilde{\alpha}+\tilde{\beta})(1-\tilde{\beta})}{1+(1-\tilde{\beta})\tilde{\delta}_{\text{in}}} = 0.$$

Similarly, solve for $\tilde{\delta}_{\text{out}}$ from

$$\sum_{j=0}^{\infty} \frac{N^{\text{out}}_{>j}(n)/n}{j+\tilde{\delta}_{\text{out}}} - \frac{1-\tilde{\gamma}-\tilde{\beta}}{\tilde{\delta}_{\text{out}}} - \frac{(\tilde{\gamma}+\tilde{\beta})(1-\tilde{\beta})}{1+(1-\tilde{\beta})\tilde{\delta}_{\text{out}}} = 0.$$

4.5 Simulation study

We now apply the estimation procedures described in Sections 4.3 and 4.4 to simulated data, which allows us to compare the estimation results using the full history of the network with that using just one snapshot. Algorithm 1 is used to simulate realizations of the preferential attachment network.

4.5.1 MLE

For the scenario of observing the full history of the network, we simulated 5000 independent replications of the preferential attachment network with 10^5 edges under the true parameter values

$$\boldsymbol{\theta} = (\alpha, \beta, \delta_{\text{in}}, \delta_{\text{out}}) = (0.3,\ 0.5,\ 2,\ 1). \tag{4.29}$$

For each realization, the MLE estimate of the parameters was computed and standardized as

$$\frac{\sqrt{n}\left((\hat{\boldsymbol{\theta}}_n^{MLE})_i - (\boldsymbol{\theta})_i\right)}{\hat{\sigma}_{ii}}, \tag{4.30}$$

where $(\hat{\boldsymbol{\theta}}_n)_i$ and $(\boldsymbol{\theta})_i$ denote the i-th components of $\hat{\boldsymbol{\theta}}_n^{MLE}$ and $\boldsymbol{\theta}$ respectively, and $\hat{\sigma}_{ii}^2$ is the i-th diagonal component of the matrix $\hat{\Sigma} := \Sigma(\hat{\boldsymbol{\theta}}_n^{MLE})$. The explicit formula for the entries

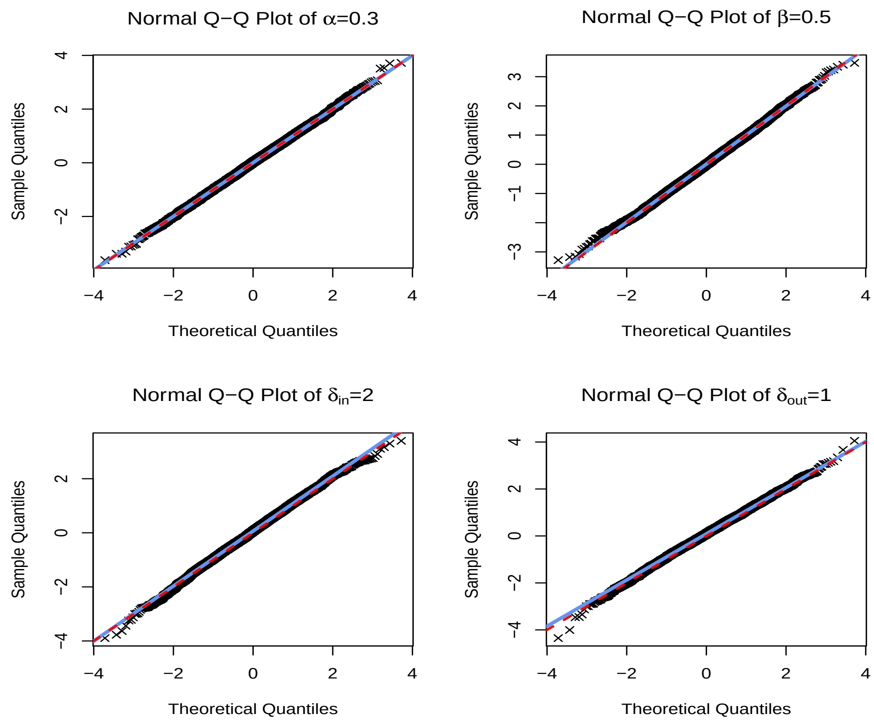


Figure 4.1: Normal QQ-plots in black for normalized estimates in (4.30) under 5000 replications of a preferential attachment network with 10^5 edges and $\boldsymbol{\theta} = (0.3, 0.5, 2, 1)$. The fitted lines in blue are the traditional qq-lines (given by R) used to check normality of the estimates. The red dashed line represents the $y = x$ line in all plots.

of $\hat{\Sigma}$ is

$$\hat{\Sigma} = \begin{bmatrix} \hat{\alpha}^{MLE}\left(1-\hat{\alpha}^{MLE}\right) & -\hat{\alpha}^{MLE}\hat{\beta}^{MLE} & 0 & 0 \\ -\hat{\alpha}^{MLE}\hat{\beta}^{MLE} & \hat{\beta}^{MLE}\left(1-\hat{\beta}^{MLE}\right) & 0 & 0 \\ 0 & 0 & \hat{I}_{\text{in}}^{-1} & 0 \\ 0 & 0 & 0 & \hat{I}_{\text{out}}^{-1} \end{bmatrix},$$

where, see (4.19) and (4.20),

$$\hat{I}_{\text{in}} = \sum_{i=0}^{\infty} \frac{N_{>i}^{\text{in}}(n)/n}{\left(i+\hat{\delta}_{\text{in}}^{MLE}\right)^2} - \frac{1-\hat{\alpha}^{MLE}-\hat{\beta}^{MLE}}{\left(\hat{\delta}_{\text{in}}^{MLE}\right)^2} - \frac{\left(\hat{\alpha}^{MLE}+\hat{\beta}^{MLE}\right)\left(1-\hat{\beta}^{MLE}\right)^2}{\left(1+\hat{\delta}_{\text{in}}^{MLE}(1-\hat{\beta}^{MLE})\right)^2},$$

$$\hat{I}_{\text{out}} = \sum_{j=0}^{\infty} \frac{N_{>j}^{\text{out}}(n)/n}{\left(j+\hat{\delta}_{\text{out}}^{MLE}\right)^2} - \frac{\hat{\alpha}^{MLE}}{\left(\hat{\delta}_{\text{out}}^{MLE}\right)^2} - \frac{\left(1-\hat{\alpha}^{MLE}\right)\left(1-\hat{\beta}^{MLE}\right)^2}{\left(1+\hat{\delta}_{\text{out}}^{MLE}(1-\hat{\beta}^{MLE})\right)^2}.$$

By the strong consistency of the MLEs combined with Lemma 4.7.2, we have that $\hat{\Sigma} \xrightarrow{\text{a.s.}} \Sigma$.

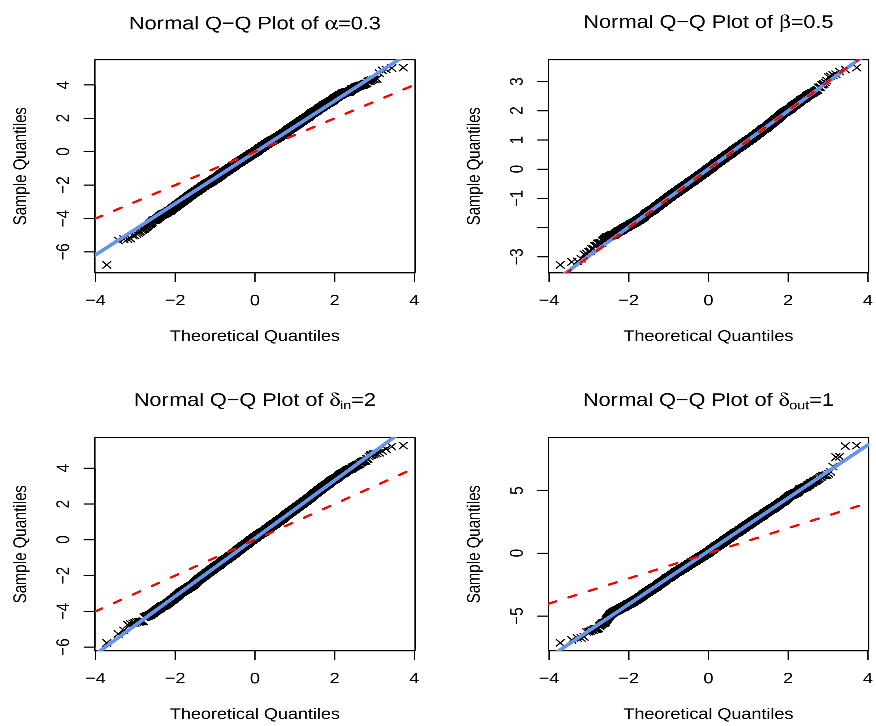


Figure 4.2: Normal QQ-plots for the normalized estimates in (4.31) under 5000 replications of a preferential attachment network with 10^5 edges and $\boldsymbol{\theta} = (0.3, 0.5, 2, 1)$. The fitted lines in blue are the traditional qq-lines used to check normality of the estimates. The red dashed line represents the $y = x$ line in all plots.

The QQ-plots of the normalized MLEs are shown in Figure 4.1, all of which line up quite well with the $y = x$ line (the red dashed line). This is consistent with the asymptotic theory described in Theorem 4.3.3. Confidence intervals for $\boldsymbol{\theta}$ can be obtained using this theorem. Given a single realization, an approximate $(1 - \varepsilon)$-confidence interval for $(\boldsymbol{\theta})_i$ is

$$(\hat{\boldsymbol{\theta}}_n^{MLE})_i \pm z_{\varepsilon/2}\sqrt{\frac{\hat{\sigma}_{ii}^2}{n}} \quad \text{for } i = 1, \ldots, 4,$$

where $z_{\varepsilon/2}$ is the upper $\varepsilon/2$ quantile of $N(0, 1)$.

4.5.2 One snapshot

We used the same simulated data as in Section 4.5.1 to obtain parameter estimates $\tilde{\boldsymbol{\theta}}_n := (\tilde{\alpha}, \tilde{\beta}, \tilde{\delta}_{\text{in}}, \tilde{\delta}_{\text{out}})$ through only the final snapshot, i.e., the set of directed edges without timestamps, following the procedure described at the end of Section 4.4. For the purpose of comparison with MLE, Figure 4.2 gives the QQ-plots for the normalized estimates from the snapshots using the same standardizations for the MLEs, i.e.,

$$\frac{\sqrt{n}\left((\tilde{\boldsymbol{\theta}}_n)_i - (\boldsymbol{\theta})_i\right)}{\hat{\sigma}_{ii}}, \quad i = 1, \ldots, 4, \tag{4.31}$$

where $(\tilde{\boldsymbol{\theta}}_n)_i$ denotes the i-th components of $\tilde{\boldsymbol{\theta}}_n$. Again, the fitted lines in blue are the traditional QQ-lines and the red dashed lines are the $y = x$ line. The QQ-plot for $\tilde{\beta}$ exhibits the same shape as for $\hat{\beta}^{MLE}$, since the two estimates are identical.

From Figure 4.2, we see that the snapshot estimates of all four parameters are consistent and approximately normal, i.e., the QQ-plots are linear. However, the slopes of the QQ-lines for $\tilde{\alpha}, \tilde{\delta}_{\text{in}}, \tilde{\delta}_{\text{out}}$ are much steeper than the diagonal line, indicating a loss of efficiency for $\tilde{\boldsymbol{\theta}}_n$ compared with $\hat{\boldsymbol{\theta}}_n$. Indeed the estimator variance is inflated for all parameters except for β, where $\tilde{\beta}$ coincides with the true MLE. This is as expected since knowing only the final snapshot provides far less information than the whole network history.

Recall that for a consistent estimator T_n of a one-dimensional parameter θ constructed from a random sample of size n, the asymptotic relative efficiencies (ARE) of T_n is defined by

$$ARE(T_n) := \lim_{n\to\infty} \frac{\text{Var}(\sqrt{n}T_n^*)}{\text{Var}(\sqrt{n}T_n)},$$

where T_n^* denotes the asymptotically efficient estimator. We may compute the ARE's for the snapshot parameter estimates

$$\begin{aligned}
ARE(\tilde{\alpha}) &= \lim_{n\to\infty} \frac{n\text{Var}(\hat{\alpha}^{MLE})}{n\text{Var}(\tilde{\alpha})} \approx \frac{\widehat{\text{Var}}(\hat{\alpha}^{MLE})}{\widehat{\text{Var}}(\tilde{\alpha})} \approx 0.398,\\
ARE(\tilde{\delta}_{\text{in}}) &= \lim_{n\to\infty} \frac{n\text{Var}(\hat{\delta}_{\text{in}}^{MLE})}{n\text{Var}(\tilde{\delta}_{\text{in}})} \approx \frac{\widehat{\text{Var}}(\hat{\delta}_{\text{in}}^{MLE})}{\widehat{\text{Var}}(\tilde{\delta}_{\text{in}})} \approx 0.392,\\
ARE(\tilde{\delta}_{\text{out}}) &= \lim_{n\to\infty} \frac{n\text{Var}(\hat{\delta}_{\text{out}}^{MLE})}{n\text{Var}(\tilde{\delta}_{\text{out}})} \approx \frac{\widehat{\text{Var}}(\hat{\delta}_{\text{out}}^{MLE})}{\widehat{\text{Var}}(\tilde{\delta}_{\text{out}})} \approx 0.226,
\end{aligned}$$

where $\widehat{\text{Var}}$ denotes the sample variance of the parameter estimate based on the 5000 replications. Note that $ARE(\tilde{\beta}) = 1$ since $\tilde{\beta} = \hat{\beta}^{MLE}$.

Given a single realization, the variances of the snapshot estimates can be estimated through resampling as follows. Using the estimated parameter $\tilde{\boldsymbol{\theta}}_n$, simulate 10^4 independent bootstrap replicates of the network with $n = 10^5$ edges. For each simulated network, the snapshot estimate, $\tilde{\boldsymbol{\theta}}_n^* := \left(\tilde{\alpha}^*, \tilde{\beta}^*, \tilde{\delta}_{\text{in}}^*, \tilde{\delta}_{\text{out}}^*\right)$, is computed. The sample variance of these 10^4 snapshot estimates can then be used as an approximation for the variance of $\tilde{\boldsymbol{\theta}}_n$ so that assuming asymptotic normality, a $(1-\varepsilon)$-confidence interval for $\boldsymbol{\theta}$ can be approximated by

$$(\tilde{\boldsymbol{\theta}}_n)_i \pm z_{\varepsilon/2}\sqrt{\widehat{\text{Var}}\left((\tilde{\boldsymbol{\theta}}_n^*)_i\right)} \quad \text{for } i = 1, \ldots, 4,$$

where $z_{\varepsilon/2}$ is the upper $\varepsilon/2$ quantile of $N(0,1)$.

4.5.3 Sensitivity test

Now we investigate the sensitivity of our estimates while values of the parameters $(n, \alpha, \beta, \delta_{\text{in}}, \delta_{\text{out}})$ are allowed to vary. First consider the impact of n, the number of edges in the network. To do so we held the parameters fixed with values given by (4.29): $(\alpha, \beta, \delta_{\text{in}}, \delta_{\text{out}}) = (0.3,\ 0.5,\ 2,\ 1)$ and varied the value of n. The QQ-plots (not presented) for standardized estimates using both full MLE and one-snapshot methods were produced to check the asymptotic normality. When $n = 500, 1000$, diagnostics revealed departures from normality for both the MLE and the snapshot estimates. However, after increasing n to 10000, estimates obtained from both approaches appeared normally distributed as expected.

For each value of n in Table 4.1, 5000 replicates of the network with n edges and parameters $\boldsymbol{\theta} = (0.3, 0.5, 2, 1)$ were generated. For each realization, the MLE's $\hat{\boldsymbol{\theta}}_n^{MLE}$ were computed using the full history of the network and the one-snapshot estimates $\tilde{\boldsymbol{\theta}}_n$ were obtained using the 7-step snapshot method proposed in Section 4.4, pretending that only the last snapshot $G(n)$ was available. The mean for these two estimators were recorded in Table 4.1. There is little bias for both estimates of α and β, even for small values of n. On the other hand, there is some bias for estimated δ_{in} and δ_{out} for $n \leq 5000$. The magnitude of the biases for both types of estimates decrease as n increases. Also the ARE's of the snapshot estimator stay within a narrow band as n increases.

Next we held $(n, \delta_{\text{in}}, \delta_{\text{out}}) = (10^5, 2, 1)$ fixed and experimented with various values of (α, β) in Table 4.2. For each choice of (α, β), 5000 independent realizations of the network

n	$Mean(\hat{\boldsymbol{\theta}}_n^{MLE})$	$Mean(\tilde{\boldsymbol{\theta}}_n)$	$ARE(\tilde{\boldsymbol{\theta}}_n)$
1000	(0.300, 0.500, 2.076, 1.054)	(0.301, 0.500, 2.128, 1.066)	(0.408, 1.000, 0.397, 0.228)
5000	(0.300, 0.500, 2.022, 1.013)	(0.301, 0.500, 2.036, 1.010)	(0.414, 1.000, 0.386, 0.236)
10000	(0.300, 0.500, 2.011, 1.006)	(0.301, 0.500, 2.019, 1.006)	(0.408, 1.000, 0.388, 0.232)
50000	(0.300, 0.500, 2.003, 1.002)	(0.300, 0.500, 2.005, 1.002)	(0.399, 1.000, 0.393, 0.230)
100000	(0.300, 0.500, 2.001, 1.001)	(0.300, 0.500, 2.003, 1.000)	(0.392, 1.000, 0.382, 0.223)

Table 4.1: Mean of $\hat{\boldsymbol{\theta}}_n^{MLE}$ and $\tilde{\boldsymbol{\theta}}_n$ with ARE's of $\tilde{\boldsymbol{\theta}}_n$ relative to $\hat{\boldsymbol{\theta}}_n^{MLE}$ for $\boldsymbol{\theta} = (0.3, 0.5, 2, 1)$ under different choices of n.

were generated and the means of the MLE $\hat{\boldsymbol{\theta}}_n^{MLE}$ and the one-snapshot estimates $\tilde{\boldsymbol{\theta}}_n$ were recorded. Overall, the biases for $\hat{\boldsymbol{\theta}}_n^{MLE}$ are remarkably small for virtually all combinations of parameter values, except for those parameter choices where one of (α, β) is extremely small. The biases for the snapshot estimates $\tilde{\boldsymbol{\theta}}_n$ exhibit a similar property, but the magnitudes of the biases are consistently larger than those in the MLE case.

In general, the snapshot estimators are able to achieve 20%–50% efficiency over the range of parameters considered. The loss of efficiency might be less than one would expect given the substantial reduction in the data available to produce the snapshot estimates. It is worth noting that in the case where $(\alpha, \beta) = (0.7, 0.2)$, the efficiencies of the snapshot estimators for α and δ_{in} are much larger (0.73 and 0.79, respectively). A heuristic explanation for this increase is that the parameter $\gamma = 1 - \alpha - \beta = 0.1$ is relatively small. By the implicit constraints used for the snapshot estimates, we have

$$\tilde{\alpha} + \tilde{\gamma} \;=\; 1 - \tilde{\beta} \;=\; 1 - \hat{\beta}^{MLE} \;=\; \hat{\alpha}^{MLE} + \hat{\gamma}^{MLE},$$

that is, the snapshot estimate of the sum $\alpha + \gamma$ is the same as the MLE for the sum. Now if γ is small, one would expect the resulting estimates to also be small so that $\tilde{\alpha}$ would be nearly the same as $\hat{\alpha}^{MLE}$. Hence the ARE would be close to 1. On the other hand, in the case of a larger γ, see the bottom row of Table 4.2 in which $\gamma = 0.6$, the ARE for α is not as large (0.42), but the ARE for $\tilde{\delta}_{\text{out}}$ is (0.63).

(α, β)	$Mean(\hat{\boldsymbol{\theta}}_n^{MLE})$	$Mean(\tilde{\boldsymbol{\theta}}_n)$	$ARE(\tilde{\boldsymbol{\theta}}_n)$
(0.001, 0.99)	(0.001, 0.990, 2.034, 1.016)	(0.001, 0.990, 2.071, 1.049)	(0.291, 1.000, 0.147, 0.316)
(0.01, 0.9)	(0.010, 0.900, 2.004, 1.001)	(0.010, 0.900, 2.008, 1.004)	(0.331, 1.000, 0.207, 0.381)
(0.1, 0.8)	(0.100, 0.800, 2.003, 1.001)	(0.100, 0.800, 2.004, 1.002)	(0.353, 1.000. 0.264, 0.216)
(0.2, 0.6)	(0.200, 0.600, 2.002, 1.001)	(0.200, 0.600, 2.003, 1.001)	(0.364, 1.000, 0.309, 0.236)
(0.5, 0.3)	(0.500, 0.300, 2.001, 1.001)	(0.500, 0.300, 2.002, 1.000)	(0.472, 1.000, 0.529, 0.202)
(0.7, 0.2)	(0.700, 0.200, 2.002, 1.000)	(0.700, 0.200, 2.002, 1.000)	(0.726, 1.000, 0.793, 0.217)
(0.1, 0.3)	(0.100, 0.300, 2.001, 1.001)	(0.100, 0.300, 2.002, 1.000)	(0.420, 1.000, 0.313, 0.629)

Table 4.2: Mean of $\hat{\boldsymbol{\theta}}_n^{MLE}$ and $\tilde{\boldsymbol{\theta}}_n$ with ARE's of $\tilde{\boldsymbol{\theta}}_n$ relative to $\hat{\boldsymbol{\theta}}_n^{MLE}$ for $(n, \delta_{\text{in}}, \delta_{\text{out}}) = (10^5, 2, 1)$ under different choices of (α, β).

4.6 Real network example

In this section, we explore fitting a preferential attachment model to a social network. As illustration, we chose the Dutch Wiki talk network dataset, available on KONECT (Kunegis, 2013) . The nodes represent users of Dutch Wikipedia, and an edge from node A to node B refers to user A writing a message on the talk page of user B at a certain time point. The network consists of 225,749 nodes (users) and 1,554,699 edges (messages). All edges are recorded with timestamps.

In order to accommodate all the edge formulation scenarios appeared in the dataset, we extend our model by appending the following two interaction schemes ($J_n = 4, 5$) in addition to the existing three ($J_n = 1, 2, 3$) described in Section 4.2.1.

- If $J_n = 4$ (with probability ξ), append to $G(n-1)$ two new nodes $v, w \in V(n)\setminus V(n-1)$ and an edge connecting them (v, w).
- If $J_n = 5$ (with probability ρ), append to $G(n-1)$ a new node $v \in V(n) \setminus V(n-1)$ with self loop (v, v).

These scenarios have been observed in other social network data, such as the network that models Facebook wall posts, again available on KONECT (Kunegis, 2013). They occur in small proportions and can be easily accommodated by a slight modification in the model fitting procedure. The new model has parameter vector $(\alpha, \beta, \gamma, \xi, \delta_{\text{in}}, \delta_{\text{out}})$, and ρ is implicitly defined through $\rho = 1 - (\alpha + \beta + \gamma + \xi)$. Similar to the derivations in Section 4.3, the MLE estimators for $\alpha, \beta, \gamma, \xi$ are

$$\hat{\alpha}^{MLE} = \frac{1}{n}\sum_{t=1}^{n} \mathbf{1}_{\{J_t=1\}}, \quad \hat{\beta}^{MLE} = \frac{1}{n}\sum_{t=1}^{n} \mathbf{1}_{\{J_t=2\}},$$
$$\hat{\gamma}^{MLE} = \frac{1}{n}\sum_{t=1}^{n} \mathbf{1}_{\{J_t=3\}}, \quad \hat{\xi}^{MLE} = \frac{1}{n}\sum_{t=1}^{n} \mathbf{1}_{\{J_t=4\}},$$

and $\delta_{\text{in}}, \delta_{\text{out}}$ can be obtained through solving

$$\sum_{i=0}^{\infty} \frac{N^{\text{in}}_{>i}(n)/n}{i+\delta_{\text{in}}} - \frac{\frac{1}{n}\sum_{t=1}^{n} \mathbf{1}_{\{J_t\in\{3,4,5\}\}}}{\delta_{\text{in}}} - \frac{1}{n}\sum_{t=1}^{n} \frac{N(t)}{t+\delta_{\text{in}}N(t)} \mathbf{1}_{\{J_t\in\{1,2\}\}} = 0,$$
$$\sum_{j=0}^{\infty} \frac{N^{\text{out}}_{>j}(n)/n}{j+\delta_{\text{out}}} - \frac{\frac{1}{n}\sum_{t=1}^{n} \mathbf{1}_{\{J_t\in\{1,4,5\}\}}}{\delta_{\text{out}}} - \frac{1}{n}\sum_{t=1}^{n} \frac{N(t)}{t+\delta_{\text{out}}N(t)} \mathbf{1}_{\{J_t\in\{2,3\}\}} = 0.$$

We first naively fit the linear preferential attachment model to the full network using MLE. The MLE estimators are

$$(\hat{\alpha}, \hat{\beta}, \hat{\gamma}, \hat{\xi}, \hat{\rho}, \hat{\delta}_{\text{in}}, \hat{\delta}_{\text{out}}) = (3.08\times 10^{-3}, 8.55\times 10^{-1}, 1.39\times 10^{-1}, 4.76\times 10^{-5}, 3.06\times 10^{-3}, 0.547, 0.134). \tag{4.33}$$

To evaluate the goodness-of-fit, 20 network realizations of the same size were simulated from the fitted model. We overlaid the empirical in- and out-degree frequencies of the original network with that of the simulations. If the model fits the data well, the degree frequencies of the data should lie within the range formed by that of the simulations, which gives an informal confidence region for the degree distributions. From Figure 4.3, we see that while the data roughly agrees with the simulations in the out-degree frequencies, the deviation in the in-degree frequencies is noticeable.

To better understand the discrepancy in the in-degree frequencies, we examined the link data and their timestamps and discovered bursts of messages originating from certain nodes over small time intervals. According to Wikipedia policy (Wikipedia, 2016), certain administrating accounts are allowed to send group messages to multiple users simultaneously. These bursts presumably represent broadcast announcements generated from these accounts. These administrative broadcasts can also be detected if we apply the linear preferential

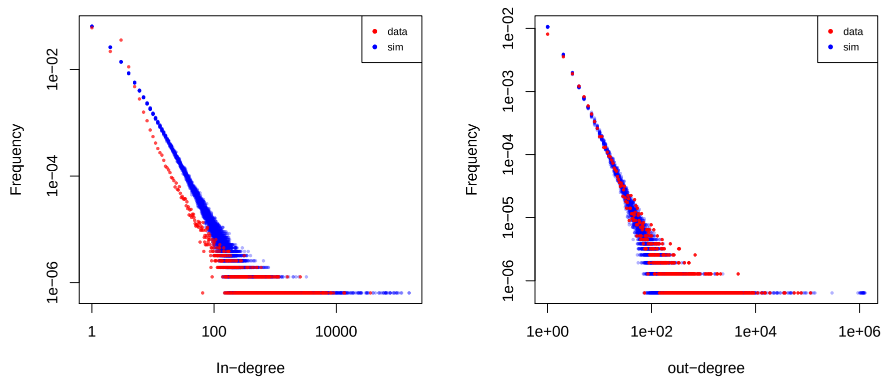


Figure 4.3: Empirical in- and out-degree frequencies of the full Wiki talk network (red) and that from 20 realizations of the linear preferential attachment network with fitted parameter values (4.33) from MLE (blue). The scatter plots for the degree frequencies from the 20 simulations are overlaid together to form an informal confidence region for the degree distribution of the fitted model

attachment model to the network in local time intervals. We divided the total time frame down to sub-intervals of varying length each containing the formation of 10^4 edges. The number 10^4 is chosen to ensure good asymptotics as shown in Table 4.1. This process generated 155 networks,

$$G(n_{k-1}), \ldots, G(n_k - 1), \quad k = 1, \ldots, 155.$$

For each of the 155 datasets, we fit a preferential attachment model using MLE. The resulting estimates $(\hat{\delta}_{\text{in}}, \hat{\delta}_{\text{out}})$ are plotted against the corresponding timeline on the upper left panel of Figure 4.4. Notice that $\hat{\delta}_{\text{in}}$ exhibits large spikes at various times. Recall from (4.1), a large value of δ_{in} indicates that the probability of an existing node v receiving a new message becomes less dependent on its in-degree, i.e., previous popularity. These spikes appear to be directly related to the occurrences of group messages. This plot is truncated after the day 2016/3/16, on which a massive group message of size 48,957 was sent and the model can no longer be fit.

We identified 37 users who have sent, at least once, 40 or more consecutive messages in the message history. This is evidence that group messages were sent by this user. We presume these nodes are administrative accounts; they are responsible for about 30% of the

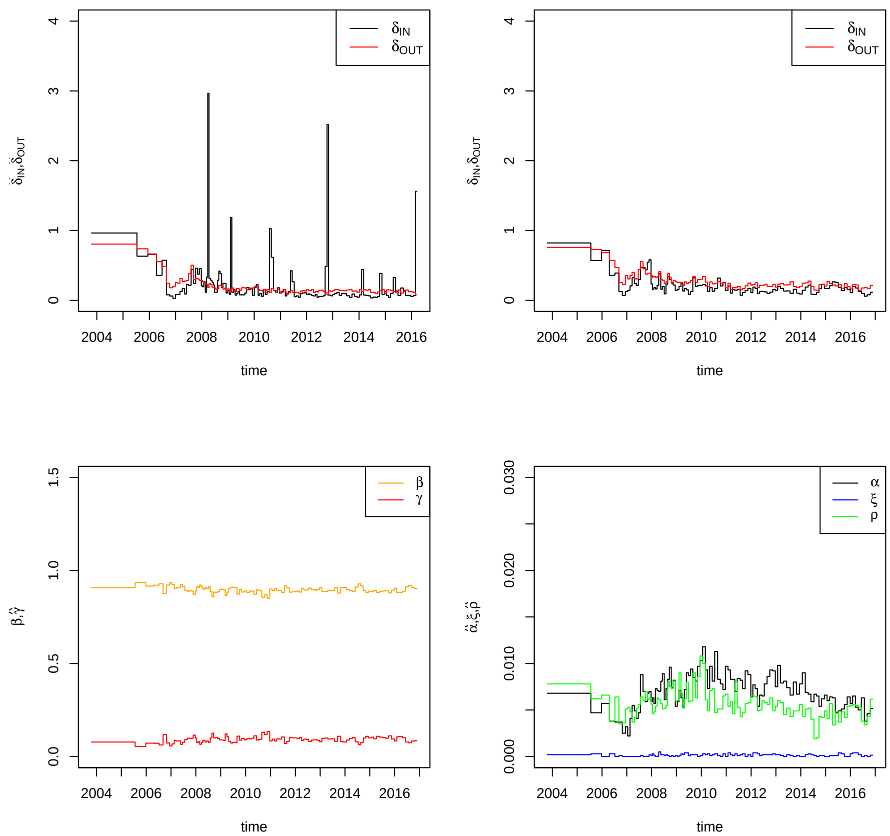


Figure 4.4: Local parameter estimates of the linear preferential attachment model for the full and reduced Wiki talk network. Upper left: $(\hat{\delta}_{\text{in}}, \hat{\delta}_{\text{out}})$ for the full network. Upper right, lower left, lower right: $(\hat{\delta}_{\text{in}}, \hat{\delta}_{\text{out}})$, $(\hat{\beta}, \hat{\gamma})$, $(\hat{\alpha}, \hat{\xi}, \hat{\rho})$ for the reduced network, respectively.

total messages sent. Since their behavior cannot be regarded as normal social interaction, we excluded messages from these accounts from the dataset in our analysis. We then also removed nodes with zero in- and out-degrees.

The re-estimated parameters after the data cleaning are displayed in the other three panels of Figure 4.4. Here all parameter estimates are quite stable through time.

The reduced network now contains 112,919 nodes and 1,086,982 edges, to which we fit the linear preferential attachment model. The fitted parameters based on MLE for our reduced dataset are

$$(\hat{\alpha}, \hat{\beta}, \hat{\gamma}, \hat{\xi}, \hat{\rho}, \hat{\delta}_{\text{in}}, \hat{\delta}_{\text{out}}) =$$

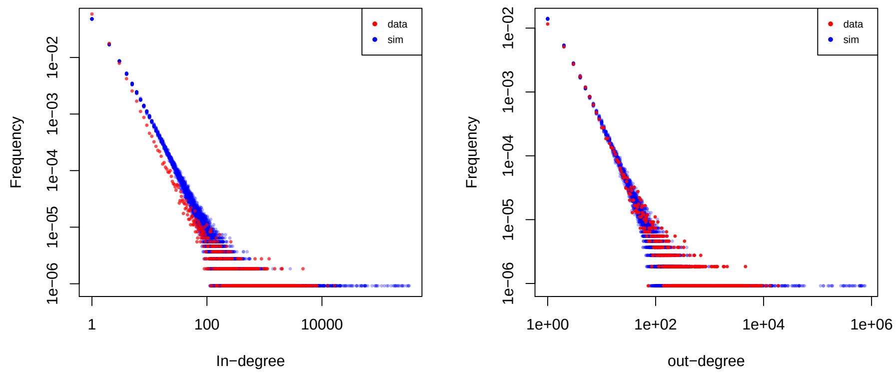


Figure 4.5: Empirical in- and out-degree frequencies of the reduced Wiki talk network (red) and that from 20 realizations of the linear preferential attachment network with fitted parameter values (4.34) from MLE (blue).

$$(6.95 \times 10^{-3}, 8.96 \times 10^{-1}, 9.10 \times 10^{-2}, 1.44 \times 10^{-4}, 5.61 \times 10^{-3}, 0.174, 0.257). \tag{4.34}$$

Again the degree distributions of the data and 20 simulations from the fitted model are displayed in Figure 4.5. The out-degree distribution of the data agrees reasonably well with the simulations. For the in-degree distribution, the fit is better than that for the entire dataset (Figure 4.3). However, for smaller in-degrees, the fitted model over-estimates the in-degree frequencies. We speculate that in many social networks, the out-degree is in line with that predicted by the preferential attachment model. An individual node would be more likely to reach out to others if having done so many times previously. For in-degrees, the situation is complicated and may depend on a multitude of factors. For instance, the choice of recipient may depend on the community that the sender is in, the topic being discussed in the message, etc. As an example a group leader might send messages to his/her team on a regular basis. Such examples violate the base assumptions of the preferential attachment model and could result in the deviation between the data and the simulations.

Next we consider the estimation method of Section 4.4 applied to a single snapshot of the data. In order to implement this procedure, we donned blinders and assumed that our dataset consists only of the information of the wiki data at the last timestamp. That is, information about administrative broadcasts, and other aspects of the data learned by

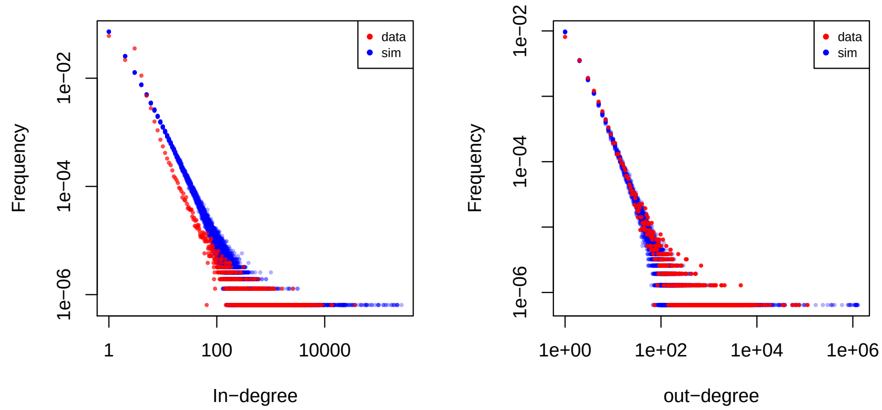


Figure 4.6: Empirical in- and out-degree frequencies of the full Wiki talk network (red) and that from 20 realizations of the linear preferential attachment network with fitted parameter values (4.35) from the snapshot estimator (blue).

looking at the previous history of the data are unavailable. In particular, we would have no knowledge of the existence of the two additional scenarios corresponding to $J_n = 4, 5$. With this in mind, we fit the three scenario model using the methods in Section 4.4. The fitted parameters are

$$(\tilde{\alpha}, \tilde{\beta}, \tilde{\gamma}, \tilde{\delta}_{\text{in}}, \tilde{\delta}_{\text{out}}) = (5.80 \times 10^{-4}, 8.55 \times 10^{-1}, 1.45 \times 10^{-1}, 0.199, 0.165). \tag{4.35}$$

The comparison of the degree distributions between the data and simulations from the fitted model is displayed in Figure 4.6 and is not too dissimilar to the plots in Figure 4.3 that are based on maximum likelihood estimation using the full network data. In particular, the out-degree distribution is matched reasonably well, but the fitted model does a poor job of capturing the in-degree distribution.

We see from this example that while the linear preferential attachment model is perhaps too simplistic for the Wiki talk network dataset, it has the ability to illuminate some gross features, such as the out-degrees, as well as to capture important structural changes such as the group message behavior. Consequently, despite its limitation, this model may be used as a building block for more flexible models. Modification to the existing model formulation and more careful analysis of change points in parameters is a direction for future research.

4.7 For the proof of Theorem 4.3.2: Lemmas 4.7.1 and 4.7.2

Lemma 4.7.1. *For $\lambda > 0$, the function $\psi(\lambda)$ in (4.16) has a unique zero at δ_{in} and, $\psi(\lambda) > 0$ when $\lambda < \delta_{in}$ and $\psi(\lambda) < 0$ when $\lambda > \delta_{in}$.*

Proof. The probabilities $\{p_i^{\text{in}}(\lambda)\}$ satisfy the recursions in i (cf. Bollobás et al. (2003)):

$$p_0^{\text{in}}(\lambda)\left(\lambda + \frac{1}{a_1(\lambda)}\right) = \frac{\alpha}{a_1(\lambda)}, \tag{4.36a}$$

$$p_1^{\text{in}}(\lambda)\left(1 + \lambda + \frac{1}{a_1(\lambda)}\right) = \lambda p_0^{\text{in}}(\lambda) + \frac{\gamma}{a_1(\lambda)},$$

$$p_2^{\text{in}}(\lambda)\left(2 + \lambda + \frac{1}{a_1(\lambda)}\right) = (1+\lambda)p_1^{\text{in}}(\lambda),$$

$$\vdots$$

$$p_i^{\text{in}}(\lambda)\left(i + \lambda + \frac{1}{a_1(\lambda)}\right) = (i-1+\lambda)p_{i-1}^{\text{in}}(\lambda), \quad (i \geq 2),$$

where $a_1(\lambda) := (\alpha+\beta)/(1+\lambda(1-\beta))$. Summing the recursions in (4.36) from 0 to i, we get (with the convention that $\sum_{i=0}^{-1} = 0$)

$$\sum_{k=0}^{i} p_k^{\text{in}}(\lambda)\left(k + \lambda + \frac{1}{a_1(\lambda)}\right) = \sum_{k=0}^{i-1}(k+\lambda)p_k^{\text{in}}(\lambda) + \frac{\alpha}{a_1(\lambda)} + \frac{\gamma}{a_1(\lambda)}\mathbf{1}_{\{i\geq 1\}}, \quad i \geq 0,$$

which can be simplified to

$$\frac{1}{a_1(\lambda)}\sum_{k=0}^{i} p_k^{\text{in}}(\lambda) + (i+\lambda)p_i^{\text{in}}(\lambda) = \frac{1-\beta}{a_1(\lambda)} - \frac{\gamma}{a_1(\lambda)}\mathbf{1}_{\{i=0\}}, \quad i \geq 0. \tag{4.37}$$

From (4.3),

$$\sum_{i=0}^{\infty} p_i^{\text{in}}(\lambda) = \sum_{i,j} p_{ij}(\lambda) = 1-\beta. \tag{4.38}$$

Hence by rearranging (4.37), we have

$$(i+\lambda)p_i^{\text{in}}(\lambda) + \frac{\gamma}{a_1(\lambda)}\mathbf{1}_{\{i=0\}} = \frac{1}{a_1(\lambda)}\left(1 - \beta - \sum_{k=0}^{i} p_k^{\text{in}}(\lambda)\right) = \frac{1}{a_1(\lambda)}p_{>i}^{\text{in}}(\lambda),$$

or equivalently,

$$p_{>i}^{\text{in}}(\lambda) = a_1(\lambda)(i+\lambda)p_i^{\text{in}}(\lambda) + \gamma\mathbf{1}_{\{i=0\}}. \tag{4.39}$$

Now with the help of (4.38) and (4.39), we can rewrite $\psi(\lambda)$ in the following way:

$$\begin{aligned}
\psi(\lambda) &= \sum_{i=0}^{\infty} \frac{p_{>i}^{\text{in}}(\delta_{\text{in}})}{i+\lambda} - \frac{\gamma}{\lambda} - (1-\beta)a_1(\lambda) \\
&= \sum_{i=0}^{\infty} \frac{p_{>i}^{\text{in}}(\delta_{\text{in}})}{i+\lambda} - \frac{\gamma}{\lambda} - \sum_{i=0}^{\infty} \frac{p_i^{\text{in}}(\delta_{\text{in}})a_1(\lambda)(i+\lambda)}{i+\lambda} \\
&= \sum_{i=0}^{\infty} \frac{a_1(\delta_{\text{in}})(i+\delta_{\text{in}})p_i^{\text{in}}(\delta_{\text{in}}) + \gamma \mathbf{1}_{\{i=0\}}}{i+\lambda} - \frac{\gamma}{\lambda} - \sum_{i=0}^{\infty} \frac{p_i^{\text{in}}(\delta_{\text{in}})a_1(\lambda)(i+\lambda)}{i+\lambda} \\
&= \sum_{i=0}^{\infty} \frac{p_i^{\text{in}}(\delta_{\text{in}})}{i+\lambda} \Big(a_1(\delta_{\text{in}})(i+\delta_{\text{in}}) - a_1(\lambda)(i+\lambda) \Big) \\
&= \sum_{i=0}^{\infty} \frac{p_i^{\text{in}}(\delta_{\text{in}})}{i+\lambda} \int_{\lambda}^{\delta_{\text{in}}} \frac{\partial}{\partial s} \Big(a_1(s)(i+s) \Big) ds \\
&= \sum_{i=0}^{\infty} \frac{p_i^{\text{in}}(\delta_{\text{in}})}{i+\lambda} \int_{\lambda}^{\delta_{\text{in}}} \frac{(\alpha+\beta)(1-i(1-\beta))}{(1+s(1-\beta))^2} ds \\
&= \left(\sum_{i=0}^{\infty} \frac{p_i^{\text{in}}(\delta_{\text{in}})}{i+\lambda} (1-i(1-\beta)) \right) \int_{\lambda}^{\delta_{\text{in}}} \frac{\alpha+\beta}{(1+s(1-\beta))^2} ds \\
&= C(\lambda) \int_{\lambda}^{\delta_{\text{in}}} \frac{\alpha+\beta}{(1+s(1-\beta))^2} ds. \qquad (4.40)
\end{aligned}$$

The series defining $C(\lambda)$ converges absolutely for any $\lambda > 0$ since

$$\begin{aligned}
\sum_{i=0}^{\infty} \left| \frac{p_i^{\text{in}}(\delta_{\text{in}})}{i+\lambda} (1-i(1-\beta)) \right| &< \sum_{i=0}^{\infty} p_i^{\text{in}}(\delta_{\text{in}}) \left| \frac{i(1-\beta)}{i+\lambda} + \frac{1}{i+\lambda} \right| \\
&< (1-\beta)(1-\beta+\frac{1}{\lambda}) < \infty.
\end{aligned}$$

Summing over i in (4.39), we get by monotone convergence

$$\sum_{i=0}^{\infty} p_{>i}^{\text{in}}(\lambda) = \sum_{i=0}^{\infty} i p_i^{\text{in}}(\lambda) = a_1(\lambda) \sum_{i=0}^{\infty} i p_i^{\text{in}}(\lambda) + a_1(\lambda)\lambda \sum_{i=0}^{\infty} p_i^{\text{in}}(\lambda) + \gamma.$$

The infinite series converge because $p_i^{\text{in}}(\lambda)$ is a power law with index greater than 2; see (4.4) and (4.5). Solving for the infinite series we get

$$\sum_{i=0}^{\infty} i p_i^{\text{in}}(\lambda) = \frac{a_1(\lambda)\lambda}{1-a_1(\lambda)} (1-\beta) + \frac{\gamma}{1-a_1(\lambda)} = 1. \qquad (4.41)$$

Hence we have

$$C(\lambda) = \sum_{i \le (1-\beta)^{-1}} \frac{p_i^{\text{in}}(\delta_{\text{in}})}{i+\lambda} (1-i(1-\beta)) - \sum_{i > (1-\beta)^{-1}} \frac{p_i^{\text{in}}(\delta_{\text{in}})}{i+\lambda} (i(1-\beta)-1)$$

$$
\begin{aligned}
&> \sum_{i=0}^{\infty} \frac{p_i^{\text{in}}(\delta_{\text{in}})}{(1-\beta)^{-1}+\lambda}(1-i(1-\beta)) \\
&= \frac{1}{(1-\beta)^{-1}+\lambda}\sum_{i=0}^{\infty} p_i^{\text{in}}(\delta_{\text{in}}) - \frac{1-\beta}{(1-\beta)^{-1}+\lambda}\sum_{i=0}^{\infty} i p_i^{\text{in}}(\delta_{\text{in}}) \\
&= \frac{1}{(1-\beta)^{-1}+\lambda}(1-\beta) - \frac{1-\beta}{(1-\beta)^{-1}+\lambda}1 \;=\; 0.
\end{aligned}
$$

Now recall from (4.40) that $\psi(\lambda)$ is of the form

$$
\psi(\lambda) \;=\; C(\lambda)\int_{\lambda}^{\delta_{\text{in}}} \frac{\alpha+\beta}{(1+s(1-\beta))^2}ds,
$$

where $C(\lambda) > 0$ for all $\lambda > 0$. Therefore $\psi(\cdot)$ has a unique zero at δ_{in} and $\psi(\lambda) > 0$ when $\lambda < \delta_{\text{in}}$ and $\psi(\lambda) < 0$ when $\lambda > \delta_{\text{in}}$. □

We show the uniform convergence of ψ_n to ψ in the next lemma.

Lemma 4.7.2. *As $n \to \infty$, for any $\epsilon > 0$,*

$$
\sup_{\lambda \geq \epsilon} |\psi_n(\lambda) - \psi(\lambda)| \;\xrightarrow{a.s.}\; 0.
$$

Proof. By the definition of ψ, $p_{>i}^{\text{in}}(\delta_{\text{in}})$ is a function of δ_{in} and is a constant with respect to λ. Hence we suppress the dependence on δ_{in} and simply write it as $p_{>i}^{\text{in}}$ when considering the difference $\psi_n - \psi$ as a function of λ:

$$
\begin{aligned}
\psi_n(\lambda) - \psi(\lambda) \;&=\; \sum_{i=0}^{\infty} \frac{N_{>i}^{\text{in}}(n)/n - p_{>i}^{\text{in}}}{i+\lambda} - \frac{1}{\lambda}\left(\frac{1}{n}\sum_{t=1}^{n} \mathbf{1}_{\{J_t=3\}} - (1-\alpha-\beta)\right) \\
&\quad -\frac{1}{n}\sum_{t=1}^{n}\left(\frac{N(t-1)}{t-1+\lambda N(t-1)}\mathbf{1}_{\{J_t\in\{1,2\}\}} - \frac{(1-\beta)(\alpha+\beta)}{1+\lambda(1-\beta)}\right).
\end{aligned}
$$

Thus,

$$
\begin{aligned}
\sup_{\lambda\geq\epsilon} |\psi_n(\lambda) - \psi(\lambda)| \;&\leq\; \sup_{\lambda\geq\epsilon}\sum_{i=0}^{\infty} \frac{\left|N_{>i}^{\text{in}}(n)/n - p_{>i}^{\text{in}}\right|}{i+\lambda} + \sup_{\lambda\geq\epsilon}\frac{1}{\lambda}\left|\frac{1}{n}\sum_{t=1}^{n}\mathbf{1}_{\{J_t=3\}} - (1-\alpha-\beta)\right| \\
&\quad + \sup_{\lambda\geq\epsilon}\left|\frac{1}{n}\sum_{t=1}^{n}\frac{N(t-1)}{t-1+\lambda N(t-1)}\mathbf{1}_{\{J_t\in\{1,2\}\}} - \frac{(1-\beta)(\alpha+\beta)}{1+\lambda(1-\beta)}\right|
\end{aligned}
\tag{4.42}
$$

For the first term, note that for all $i \geq 0$,

$$
iN_{>i}^{\text{in}}(n) \;=\; \sum_{k=i+1}^{\infty} N_k^{\text{in}}(n) i \;\leq\; \sum_{k=1}^{\infty} kN_k^{\text{in}}(n) \;=\; n,
$$

since the assumption on initial conditions implies the sum of in-degrees at n is n. Therefore $N^{\text{in}}_{>i}(n)/n \le i^{-1}$ for $i \ge 1$, and it then follows that

$$\sum_{i=0}^{\infty} \frac{\left|N^{\text{in}}_{>i}(n)/n - p^{\text{in}}_{>i}\right|}{i+\lambda} \le \sum_{i=0}^{M} \frac{\left|N^{\text{in}}_{>i}(n)/n - p^{\text{in}}_{>i}\right|}{i+\lambda} + \sum_{i=M+1}^{\infty} \frac{1/i}{i+\lambda} + \sum_{i=M+1}^{\infty} \frac{p^{\text{in}}_{>i}}{i+\lambda}.$$

Note that the last two terms on the right side can be made arbitrarily small uniformly on $[\epsilon, \infty)$ if we choose M sufficiently large. Recall the convergence of the degree distribution $\{N_{ij}(n)/N(n)\}$ to the probability distribution $\{f_{ij}\}$ in (4.3), we have

$$\frac{N^{\text{in}}_{>i}(n)}{n} = \frac{N(n)}{n}\,\frac{N^{\text{in}}_{>i}(n)}{N(n)} \xrightarrow{\text{a.s.}} (1-\beta) \sum_{l\ge 0, k>i} f_{kl} = p^{\text{in}}_{>i}, \quad \forall i \ge 0.$$

Hence, for any fixed M,

$$\sum_{i=0}^{M} \frac{\left|N^{\text{in}}_{>i}(n)/n - p^{\text{in}}_{>i}\right|}{i+\epsilon} \xrightarrow{\text{a.s.}} 0, \quad \text{as } n \to \infty.$$

which implies further that choosing M arbitrarily large gives

$$\sup_{\lambda\ge\epsilon} \sum_{i=0}^{\infty} \frac{\left|N^{\text{in}}_{>i}(n)/n - p^{\text{in}}_{>i}\right|}{i+\lambda} \le \sum_{i=0}^{M} \frac{\left|N^{\text{in}}_{>i}(n)/n - p^{\text{in}}_{>i}\right|}{i+\epsilon} + \sum_{i=M+1}^{\infty} \frac{1/i}{i+\epsilon} + \sum_{i=M+1}^{\infty} \frac{p^{\text{in}}_{>i}}{i+\epsilon} \xrightarrow{\text{a.s.}} 0.$$

The second term in (4.42) converges to 0 almost surely by strong law of large numbers, and the third term in (4.42) can be written as

$$\Bigg| \frac{1}{n} \sum_{t=1}^{n} \left(\frac{N(t-1)}{t-1+\lambda N(t-1)} - \frac{(1-\beta)}{1+\lambda(1-\beta)} \right) \mathbf{1}_{\{J_t \in \{1,2\}\}} + \frac{1-\beta}{1+\lambda(1-\beta)} \frac{1}{n} \sum_{t=1}^{n} \left(\mathbf{1}_{\{J_t\in\{1,2\}\}} - (\alpha+\beta) \right) \Bigg|,$$

which is bounded by

$$\left| \frac{1}{n} \sum_{t=1}^{n} \frac{N(t-1)}{t-1+\lambda N(t-1)} - \frac{(1-\beta)}{1+\lambda(1-\beta)} \right| + \frac{1-\beta}{1+\lambda(1-\beta)} \left| \frac{1}{n} \sum_{t=1}^{n} \mathbf{1}_{\{J_t\in\{1,2\}\}} - (\alpha+\beta) \right|.$$

We have

$$\begin{aligned} &\sup_{\lambda\ge\epsilon} \left| \frac{1}{n} \sum_{t=1}^{n} \frac{N(t-1)}{t-1+\lambda N(t-1)} - \frac{(1-\beta)}{1+\lambda(1-\beta)} \right| \\ &\quad = \sup_{\lambda\ge\epsilon} \left| \frac{1}{n} \sum_{t=1}^{n} \frac{N(t-1)/(t-1) - (1-\beta)}{(1+\lambda N(t-1)/(t-1))(1+\lambda(1-\beta))} \right| \end{aligned}$$

$$\leq \frac{1}{n}\sum_{t=1}^{n}\left|\frac{N(t-1)/(t-1)-(1-\beta)}{(1+\epsilon N(t-1)/(t-1))(1+\epsilon(1-\beta))}\right|,$$

which converges to 0 almost surely by Cesàro convergence of random variables, since

$$\left|\frac{N(n)/n-(1-\beta)}{(1+\epsilon N(n)/n)(1+\epsilon(1-\beta))}\right| \xrightarrow{\text{a.s.}} 0, \quad \text{as } n\to\infty.$$

Further, by the strong law of large numbers,

$$\begin{aligned}
&\sup_{\lambda\geq\epsilon}\frac{1-\beta}{1+\lambda(1-\beta)}\left|\frac{1}{n}\sum_{t=1}^{n}\mathbf{1}_{\{J_t\in\{1,2\}\}}-(\alpha+\beta)\right| \\
&\quad\leq \frac{1-\beta}{1+\epsilon(1-\beta)}\left|\frac{1}{n}\sum_{t=1}^{n}\mathbf{1}_{\{J_t\in\{1,2\}\}}-(\alpha+\beta)\right| \xrightarrow{\text{a.s.}} 0, \quad \text{as } n\to\infty.
\end{aligned}$$

Hence the third term of (4.42) also goes to 0 almost surely as $n\to\infty$. The result of the lemma follows. □

4.7.1 For the proof of Theorem 4.3.3: Lemmas 4.7.3 and 4.7.4

Lemma 4.7.3. *As $n\to\infty$,*

$$n^{-1/2}\sum_{t=1}^{n}u_t(\delta_{in}) \xrightarrow{d} N(0, I_{in}). \tag{4.43}$$

Proof. Let $\mathcal{F}_n=\sigma(G(0),\dots,G(n))$ be the σ-field generated by the information contained in the graphs. We first observe that $\{\sum_{t=1}^{n}u_t(\delta_{\text{in}}),\mathcal{F}_n,n\geq 1\}$ is a martingale. To see this, note from (4.21) that $|u_t(\delta)|\leq 2/\delta$ and

$$\begin{aligned}
&\mathbb{E}[u_t(\delta_{\text{in}})|\mathcal{F}_{t-1}] \\
&= \mathbb{E}\left[\frac{1}{D_{\text{in}}^{(t-1)}(v_t^{(2)})+\delta_{\text{in}}}\mathbf{1}_{\{J_t\in\{1,2\}\}}\middle|\mathcal{F}_{t-1}\right]-\frac{N(t-1)}{t-1+\delta_{\text{in}}N(t-1)}\mathbb{E}[\mathbf{1}_{\{J_t\in\{1,2\}\}}|\mathcal{F}_{t-1}] \\
&= \mathbb{E}\left[\frac{1}{D_{\text{in}}^{(t-1)}(v_t^{(2)})+\delta_{\text{in}}}\middle| J_t=1,\mathcal{F}_{t-1}\right]\mathbb{P}[J_t=1] \\
&\quad +\mathbb{E}\left[\frac{1}{D_{\text{in}}^{(t-1)}(v_t^{(2)})+\delta_{\text{in}}}\middle| J_t=2,\mathcal{F}_{t-1}\right]\mathbb{P}[J_t=2]-(\alpha+\beta)\frac{N(t-1)}{t-1+\delta_{\text{in}}N(t-1)} \\
&= (\alpha+\beta)\sum_{v\in V_{t-1}}\frac{1}{D_{\text{in}}^{(t-1)}(v)+\delta_{\text{in}}}\frac{D_{\text{in}}^{(t-1)}(v)+\delta_{\text{in}}}{t-1+\delta_{\text{in}}N(t-1)}-(\alpha+\beta)\frac{N(t-1)}{t-1+\delta_{\text{in}}N(t-1)}
\end{aligned}$$

$$= (\alpha+\beta)\left(\sum_{v\in V_{t-1}} \frac{1}{t-1+\delta_{\rm in}N(t-1)} - \frac{N(t-1)}{t-1+\delta_{\rm in}N(t-1)}\right) = 0,$$

which satisfies the definition of a martingale difference. Hence

$$\left\{n^{-1/2}\sum_{r=1}^{t} u_r(\delta_{\rm in})\right\}_{t=1,\dots,n}$$

is a zero-mean, square-integrable martingale array. The convergence (4.43) follows from the martingale central limit theory (cf. Theorem 3.2 of Hall and Heyde (1980)) if the following three conditions can be verified:

(a) $n^{-1/2}\max_t |u_t(\delta_{\rm in})| \xrightarrow{p} 0$,

(b) $n^{-1}\sum_t u_t^2(\delta_{\rm in}) \xrightarrow{p} I_{\rm in}$,

(c) $\mathbb{E}\left(n^{-1}\max_t u_t^2(\delta_{\rm in})\right)$ is bounded in n.

Since $|u_t(\delta_{\rm in})| \le 2/\delta_{\rm in}$, we have

$$n^{-1/2}\max_t |u_t(\delta_{\rm in})| \le \frac{2}{n^{1/2}\delta_{\rm in}} \to 0,$$

and

$$n^{-1}\max_t u_t^2 \le \frac{4}{n\delta_{\rm in}^2} \to 0.$$

Hence conditions (a) and (c) are straightforward.

To show (b), observe that

$$\begin{aligned}
\frac{1}{n}\sum_{t=1}^{n} u_t^2(\delta_{\rm in}) &= \frac{1}{n}\sum_{t=1}^{n} \mathbf{1}_{\{J_t\in\{1,2\}\}}\left(\frac{1}{D_{\rm in}^{(t-1)}(v_t^{(2)})+\delta_{\rm in}} - \frac{N(t-1)}{t-1+\delta_{\rm in}N(t-1)}\right)^2 \\
&= \frac{1}{n}\sum_{t=1}^{n} \frac{\mathbf{1}_{\{J_t\in\{1,2\}\}}}{\left(D_{\rm in}^{(t-1)}(v_t^{(2)})+\delta_{\rm in}\right)^2} \\
&\quad - \frac{2}{n}\sum_{t=1}^{n} \frac{\mathbf{1}_{\{J_t\in\{1,2\}\}}}{D_{\rm in}^{(t-1)}(v_t^{(2)})+\delta_{\rm in}}\frac{N(t-1)}{t-1+\delta_{\rm in}N(t-1)} \\
&\quad + \frac{1}{n}\sum_{t=1}^{n} \mathbf{1}_{\{J_t\in\{1,2\}\}}\left(\frac{N(t-1)}{t-1+\delta_{\rm in}N(t-1)}\right)^2 \\
&=: T_1 - 2T_2 + T_3.
\end{aligned}$$

Following the calculations in the proof of Lemma 4.7.2, we have for T_1,

$$T_1 = \sum_{i=0}^{\infty} \frac{N^{\text{in}}_{>i}(n)/n}{(i+\delta_{\text{in}})^2} - \frac{1}{\delta_{\text{in}}^2}\frac{1}{n}\sum_{t=1}^{n} \mathbf{1}_{\{J_t=3\}} \xrightarrow{p} \sum_{i=0}^{\infty} \frac{p^{\text{in}}_{>i}}{(i+\delta_{\text{in}})^2} - \frac{\gamma}{\delta_{\text{in}}^2}.$$

We then rewrite T_2 as

$$\begin{aligned} T_2 &= \frac{1}{n}\sum_{t=1}^{n} \frac{\mathbf{1}_{\{J_t\in\{1,2\}\}}}{D_{\text{in}}^{(t-1)}(v_t^{(2)})+\delta_{\text{in}}} \left(\frac{N(t-1)/(t-1)}{1+\delta_{\text{in}}N(t-1)/(t-1)} - \frac{1-\beta}{1+\delta_{\text{in}}(1-\beta)} \right) \\ &\quad + \frac{1}{n}\sum_{t=1}^{n} \frac{\mathbf{1}_{\{J_t\in\{1,2\}\}}}{D_{\text{in}}^{(t-1)}(v_t^{(2)})+\delta_{\text{in}}} \frac{1-\beta}{1+\delta_{\text{in}}(1-\beta)} \\ &=: T_{21}+T_{22}, \end{aligned}$$

where

$$|T_{21}| \leq \frac{1}{n}\sum_{t=1}^{n} \frac{1}{\delta_{\text{in}}} \left| \frac{N(t-1)/(t-1)}{1+\delta_{\text{in}}N(t-1)/(t-1)} - \frac{1-\beta}{1+\delta_{\text{in}}(1-\beta)} \right| \xrightarrow{p} 0$$

by Cesàro's convergence and

$$\begin{aligned} T_{22} &= \frac{1-\beta}{1+\delta_{\text{in}}(1-\beta)} \left(\sum_{i=0}^{\infty} \frac{N^{\text{in}}_{>i}(n)/n}{i+\delta_{\text{in}}} - \frac{1}{\delta_{\text{in}}}\frac{1}{n}\sum_{t=1}^{n} \mathbf{1}_{\{J_t=3\}} \right) \\ &\xrightarrow{p} \frac{1-\beta}{1+\delta_{\text{in}}(1-\beta)} \left(\sum_{i=0}^{\infty} \frac{p^{\text{in}}_{>i}}{i+\delta_{\text{in}}} - \frac{\gamma}{\delta_{\text{in}}} \right) = \frac{(\alpha+\beta)(1-\beta)^2}{(1+\delta_{\text{in}}(1-\beta))^2}, \end{aligned}$$

where the equality follows from (4.39). For T_3, similar to T_1, we have

$$\begin{aligned} T_3 &= \frac{1}{n}\sum_{t=1}^{n} \mathbf{1}_{\{J_t\in\{1,2\}\}} \left(\left(\frac{N(t-1)/(t-1)}{1+\delta_{\text{in}}N(t-1)/(t-1)} \right)^2 - \frac{(1-\beta)^2}{(1+\delta_{\text{in}}(1-\beta))^2} \right) \\ &\quad + \frac{(1-\beta)^2}{(1+\delta_{\text{in}}(1-\beta))^2} \frac{1}{n}\sum_{t=1}^{n} \mathbf{1}_{\{J_t\in\{1,2\}\}} \xrightarrow{p} \frac{(\alpha+\beta)(1-\beta)^2}{(1+\delta_{\text{in}}(1-\beta))^2}. \end{aligned}$$

Combining these results together,

$$\begin{aligned} \frac{1}{n}\sum_{t=1}^{n} u_t^2(\delta_{\text{in}}) &= T_1 - 2(T_{21}+T_{22}) + T_3 \\ &\xrightarrow{p} \sum_{i=0}^{\infty} \frac{p^{\text{in}}_{>i}}{(i+\delta_{\text{in}})^2} - \frac{\gamma}{\delta_{\text{in}}^2} - \frac{(\alpha+\beta)(1-\beta)^2}{(1+\delta_{\text{in}}(1-\beta))^2} = I_{\text{in}}. \end{aligned} \tag{4.44}$$

This completes the proof. □

Lemma 4.7.4. *As $n \to \infty$,*

$$\frac{1}{n}\sum_{t=1}^{n} \dot{u}_t(\hat{\delta}^*_{in}) \xrightarrow{p} -I_{in}.$$

Proof. The result of this lemma can be established by showing first

$$\frac{1}{n}\sum_{t=1}^{n} \dot{u}_t(\delta_{\text{in}}) \stackrel{p}{\to} -I_{\text{in}} \tag{4.45}$$

and then

$$\left|\frac{1}{n}\sum_{t=1}^{n} \dot{u}_t(\hat{\delta}^*_{\text{in}}) - \frac{1}{n}\sum_{t=1}^{n} \dot{u}_t(\delta_{\text{in}})\right| \stackrel{p}{\to} 0. \tag{4.46}$$

We first observe that

$$\begin{aligned} \dot{u}_t(\delta) &= -\left(\frac{1}{D_{\text{in}}^{(t-1)}(v_t^{(2)})+\delta}\right)^2 \mathbf{1}_{\{J_t\in\{1,2\}\}} + \left(\frac{N(t-1)}{t-1+\delta N(t-1)}\right)^2 \mathbf{1}_{\{J_t\in\{1,2\}\}} \\ &= -u_t^2(\delta) - 2u_t(\delta)\frac{N(t-1)}{t-1+\delta N(t-1)}. \end{aligned}$$

Recall the definition and convergence result for T_2 and T_3 in Lemma 4.7.3, we have

$$\frac{1}{n}\sum_{t=1}^{n} u_t(\delta_{\text{in}})\frac{N(t-1)}{t-1+\delta_{\text{in}}N(t-1)} = T_2 - T_3 \stackrel{p}{\to} 0.$$

Also from (4.44),

$$\frac{1}{n}\sum_{t=1}^{n} u_t^2(\delta_{\text{in}}) \stackrel{p}{\to} I_{\text{in}}.$$

Hence

$$\frac{1}{n}\sum_{t=1}^{n} \dot{u}_t(\delta_{\text{in}}) = -\frac{1}{n}\sum_{t=1}^{n} u_t^2(\delta_{\text{in}}) - \frac{2}{n}\sum_{t=1}^{n} u_t(\delta_{\text{in}})\frac{N(t-1)}{t-1+\delta_{\text{in}}N(t-1)} \stackrel{p}{\to} -I_{\text{in}}$$

and (4.45) is established.

By construction and definition, we have $\hat{\delta}_{\text{in}}, \hat{\delta}^*_{\text{in}}, \delta_{\text{in}} > 0$. To prove (4.46), note that

$$\begin{aligned} |u_t(\hat{\delta}^*_{\text{in}}) - u_t(\delta_{\text{in}})| &\le \mathbf{1}_{\{J_t\in\{1,2\}\}}\left|\frac{1}{D_{\text{in}}^{(t-1)}(v_t^{(2)})+\hat{\delta}^*_{\text{in}}} - \frac{1}{D_{\text{in}}^{(t-1)}(v_t^{(2)})+\delta_{\text{in}}}\right| \\ &\quad +\mathbf{1}_{\{J_t\in\{1,2\}\}}\left|\frac{N(t-1)}{t-1+\hat{\delta}^*_{\text{in}}N(t-1)} - \frac{N(t-1)}{t-1+\delta_{\text{in}}N(t-1)}\right| \\ &\le \mathbf{1}_{\{J_t\in\{1,2\}\}}\left|\frac{\delta_{\text{in}}-\hat{\delta}^*_{\text{in}}}{\left(D_{\text{in}}^{(t-1)}(v_t^{(2)})+\hat{\delta}^*_{\text{in}}\right)\left(D_{\text{in}}^{(t-1)}(v_t^{(2)})+\delta_{\text{in}}\right)}\right| \\ &\quad +\mathbf{1}_{\{J_t\in\{1,2\}\}}\left|\frac{(N(t-1))^2(\delta_{\text{in}}-\hat{\delta}^*_{\text{in}})}{\left(t-1+\hat{\delta}^*_{\text{in}}N(t-1)\right)(t-1+\delta_{\text{in}}N(t-1))}\right| \end{aligned}$$

$$\leq \quad \frac{2|\hat{\delta}^*_{\text{in}} - \delta_{\text{in}}|}{\hat{\delta}^*_{\text{in}}\delta_{\text{in}}}.$$

Then

$$|u_t^2(\hat{\delta}^*_{\text{in}}) - u_t^2(\delta_{\text{in}})| \;=\; \left|u_t(\hat{\delta}^*_{\text{in}}) - u_t(\delta_{\text{in}})\right| \left|u_t(\hat{\delta}^*_{\text{in}}) + u_t(\delta_{\text{in}})\right| \;\leq\; \frac{2\left|\hat{\delta}^*_{\text{in}} - \delta_{\text{in}}\right|}{\hat{\delta}^*_{\text{in}}\delta_{\text{in}}} \left(\frac{2}{\hat{\delta}^*_{\text{in}}} + \frac{2}{\delta_{\text{in}}}\right),$$

and

$$\begin{aligned}
&\left|u_t(\hat{\delta}^*_{\text{in}}\frac{N(t-1)}{t-1+\hat{\delta}^*_{\text{in}}N(t-1)} - u_t(\delta_{\text{in}})\frac{N(t-1)}{t-1+\delta_{\text{in}}N(t-1)}\right| \\
\leq\;& \left|u_t(\hat{\delta}^*_{\text{in}}) - u_t(\delta_{\text{in}})\right| \frac{\frac{N(t-1)}{t-1}}{1+\delta_{\text{in}}\frac{N(t-1)}{t-1}} + \left|u_t(\hat{\delta}^*_{\text{in}})\right| \left|\frac{\frac{N(t-1)}{t-1}}{1+\hat{\delta}^*_{\text{in}}\frac{N(t-1)}{t-1}} - \frac{\frac{N(t-1)}{t-1}}{1+\delta_{\text{in}}\frac{N(t-1)}{t-1}}\right| \\
\leq\;& \frac{2\left|\hat{\delta}^*_{\text{in}} - \delta_{\text{in}}\right|}{\hat{\delta}^*_{\text{in}}\delta_{\text{in}}}\frac{1}{\delta_{\text{in}}} + \frac{2}{\hat{\delta}^*_{\text{in}}}\frac{\left|\hat{\delta}^*_{\text{in}} - \delta_{\text{in}}\right|}{\hat{\delta}^*_{\text{in}}\delta_{\text{in}}}.
\end{aligned}$$

From Theorem 4.3.2, $\hat{\delta}^{MLE}_{\text{in}}$ is consistent for δ_{in}, hence

$$\left|\hat{\delta}^*_{\text{in}} - \delta_{\text{in}}\right| \;\leq\; \left|\hat{\delta}^{MLE}_{\text{in}} - \delta_{\text{in}}\right| \;\xrightarrow{p}\; 0.$$

We have

$$\begin{aligned}
&\left|\frac{1}{n}\sum_{t=1}^{n}\dot{u}_t(\hat{\delta}^*_{\text{in}}) - \frac{1}{n}\sum_{t=1}^{n}\dot{u}_t(\delta_{\text{in}})\right| \\
\leq\;& \frac{1}{n}\sum_{t=1}^{n}\left|\dot{u}_t(\hat{\delta}^*_{\text{in}}) - \dot{u}_t(\delta_{\text{in}})\right| \\
\leq\;& \frac{1}{n}\sum_{t=1}^{n}\left|u_t^2(\hat{\delta}^*_{\text{in}}) - u_t^2(\delta_{\text{in}})\right| + \frac{2}{n}\sum_{t=1}^{n}\left|u_t(\hat{\delta}^*_{\text{in}})\frac{N(t-1)}{t-1+\hat{\delta}^*_{\text{in}}N(t-1)} - u_t(\delta_{\text{in}})\frac{N(t-1)}{t-1+\delta_{\text{in}}N(t-1)}\right| \\
\leq\;& \frac{2\left|\hat{\delta}^*_{\text{in}} - \delta_{\text{in}}\right|}{\hat{\delta}^*_{\text{in}}\delta_{\text{in}}}\left(\frac{2}{\hat{\delta}^*_{\text{in}}} + \frac{2}{\delta_{\text{in}}}\right) + \frac{4\left|\hat{\delta}^*_{\text{in}} - \delta_{\text{in}}\right|}{\hat{\delta}^*_{\text{in}}\delta_{\text{in}}}\frac{1}{\delta_{\text{in}}} + \frac{4}{\hat{\delta}^*_{\text{in}}}\frac{\left|\hat{\delta}^*_{\text{in}} - \delta_{\text{in}}\right|}{\hat{\delta}^*_{\text{in}}\delta_{\text{in}}} \;\xrightarrow{p}\; 0.
\end{aligned}$$

This proves (4.46) and completes the proof of Lemma 4.7.4. □

4.8 Proof of Theorem 4.4.1

Proof. First observe that $\sum_i iN_i^{\text{in}}(n)$ sums up to the total number of edges n, so

$$\sum_{i=0}^{\infty}\frac{N^{\text{in}}_{>i}(n)}{n} \;=\; \sum_{i=0}^{\infty}\frac{iN_i^{\text{in}}(n)}{n} \;=\; 1.$$

We can re-write (4.28a) as

$$\begin{aligned}
\alpha + \tilde{\beta} &= \left(\frac{1}{\delta_{\text{in}}} - \sum_{i=0}^{\infty} \frac{N^{\text{in}}_{>i}(n)/n}{i+\delta_{\text{in}}} \right) \Big/ \left(\frac{1}{\delta_{\text{in}}} - \frac{1-\tilde{\beta}}{1+\delta_{\text{in}}(1-\tilde{\beta})} \right) \\
&= \left(\sum_{i=0}^{\infty} \frac{N^{\text{in}}_{>i}(n)/n}{\delta_{\text{in}}} - \sum_{i=0}^{\infty} \frac{N^{\text{in}}_{>i}(n)/n}{i+\delta_{\text{in}}} \right) \Big/ \left(\frac{1}{\delta_{\text{in}}(1+\delta_{\text{in}}(1-\tilde{\beta}))} \right) \\
&= \sum_{i=1}^{\infty} \frac{N^{\text{in}}_{>i}(n)}{n} \frac{i}{i+\delta_{\text{in}}} \left(1+\delta_{\text{in}}(1-\tilde{\beta})\right) =: f_n(\delta_{\text{in}}), \qquad (4.47)
\end{aligned}$$

and (4.28b) as

$$\alpha + \tilde{\beta} = \left(\frac{N^{\text{in}}_0(n)}{n} + \tilde{\beta} \right) \Big/ \left(1 - \frac{N^{\text{in}}_0(n)}{n} \frac{\delta_{\text{in}}}{1+(1-\tilde{\beta})\delta_{\text{in}}} \right) =: g_n(\delta_{\text{in}}).$$

Then $\tilde{\delta}_{\text{in}}$ can be obtained by solving

$$f_n(\delta) - g_n(\delta) = 0, \qquad \delta \in [\epsilon, K].$$

Similar to the proof of Theorem 4.3.2, we define the limit versions of f_n, and g_n as follows:

$$\begin{aligned}
f(\delta) &:= \sum_{i=1}^{\infty} p^{\text{in}}_{>i} \frac{i}{i+\delta}(1+\delta(1-\beta)), \\
g(\delta) &:= \left(p^{\text{in}}_0 + \beta\right) \Big/ \left(1 - p^{\text{in}}_0 \frac{\delta}{1+(1-\beta)\delta} \right), \qquad \delta \in [\epsilon, K].
\end{aligned}$$

Now we apply the re-parametrization

$$\eta := \frac{\delta}{1+\delta(1-\beta)} \in \left[\frac{1}{\epsilon^{-1}+1-\beta}, \frac{1}{K^{-1}+1-\beta} \right] =: \mathcal{I} \qquad (4.48)$$

to f and g, such that

$$\begin{aligned}
\tilde{f}(\eta) &:= f(\delta(\eta)) = \sum_{i=1}^{\infty} \frac{p^{\text{in}}_{>i}}{1+(i^{-1}-(1-\beta))\eta}, \\
\tilde{g}(\eta) &:= g(\delta(\eta)) = \frac{p^{\text{in}}_0 + \beta}{1-\eta p^{\text{in}}_0}.
\end{aligned}$$

Note that for all $\eta \in \mathcal{I}$:

- Set $b_i(\eta) := (i^{-1} - (1-\beta))\eta$, then $1 + b_i(\eta) > 0$ for all $i \geq 1$. So that $\tilde{f}(\eta) > 0$ on $\mathcal{I}$;
- $\tilde{f}(\eta) \leq \frac{1}{1-(1-\beta)\eta} \sum_{i=0}^{\infty} p^{\text{in}}_{>i} \leq 1 + (1-\beta)K < \infty$.

Meanwhile, $\tilde{g}$ is also well defined and strictly positive for $\eta \in \mathcal{I}$ because

$$1/p_0^{\text{in}} \, > \, 1/(1-\beta) \, > \, \eta. \tag{4.49}$$

The first inequality holds since:

$$\begin{aligned} 1/p_0^{\text{in}} \, > \, 1/(1-\beta) \;&\Leftrightarrow\; p_0^{\text{in}} \, < \, 1-\beta \\ &\Leftrightarrow\; \frac{\alpha}{1+\frac{(\alpha+\beta)\delta_{\text{in}}}{1+(1-\beta)\delta_{\text{in}}}} < 1-\beta \\ &\Leftrightarrow\; \alpha+\beta \, < \, 1+\frac{(1-\beta)(\alpha+\beta)\delta_{\text{in}}}{1+(1-\beta)\delta_{\text{in}}} \\ &\Leftrightarrow\; \alpha+\beta \, < \, 1+(1-\beta)\delta_{\text{in}}. \end{aligned}$$

We know $\alpha+\beta<1$ by our model assumption, thus verifying (4.49).

Define for $\eta \in \mathcal{I}$,

$$\tilde{h}(\eta) := \frac{1}{\tilde{f}(\eta)} - \frac{1}{\tilde{g}(\eta)} = \left(\sum_{i=1}^{\infty} \frac{p_{>i}^{\text{in}}}{1+(i^{-1}-(1-\beta))\eta}\right)^{-1} - \frac{1-\eta p_0^{\text{in}}}{p_0^{\text{in}}+\beta},$$

then it follows that

$$\tilde{h}(\eta)=0 \quad \Leftrightarrow \quad \tilde{f}(\eta)=\tilde{g}(\eta), \qquad \eta \in \mathcal{I}.$$

We now show that $\tilde{h}$ is concave and $\tilde{h}(\eta) \to 0$ as $\eta \to 0$, then the uniqueness of the solution follows.

First observe that

$$\begin{aligned} \frac{\partial^2}{\partial \eta^2}\tilde{h}(\eta) &= \frac{\partial^2}{\partial \eta^2}\left(\sum_{i=1}^{\infty} \frac{p_{>i}^{\text{in}}}{1+(i^{-1}-(1-\beta))\eta}\right)^{-1} \\ &= \frac{\partial^2}{\partial \eta^2}\left(\sum_{i=1}^{\infty} \frac{p_{>i}^{\text{in}}}{1+b_i(\eta)}\right)^{-1} \\ &= 2\left(\sum_{i=1}^{\infty} \frac{p_{>i}^{\text{in}}}{1+b_i(\eta)}\right)^{-3} \left[\frac{\partial}{\partial \eta}\left(\sum_{i=1}^{\infty} \frac{p_{>i}^{\text{in}}}{1+b_i(\eta)}\right)\right]^2 \\ &\qquad - \left(\sum_{i=1}^{\infty} \frac{p_{>i}^{\text{in}}}{1+b_i(\eta)}\right)^{-2} \frac{\partial^2}{\partial \eta^2}\left(\sum_{i=1}^{\infty} \frac{p_{>i}^{\text{in}}}{1+b_i(\eta)}\right). \end{aligned} \tag{4.50}$$

We now claim that

$$\frac{\partial}{\partial \eta}\left(\sum_{i=1}^{\infty} \frac{p_{>i}^{\text{in}}}{1+b_i(\eta)}\right) = \sum_{i=1}^{\infty} \frac{\partial}{\partial \eta}\left(\frac{p_{>i}^{\text{in}}}{1+b_i(\eta)}\right) = -\sum_{i=1}^{\infty} \frac{p_{>i}^{\text{in}}(i^{-1}-(1-\beta))}{(1+b_i(\eta))^2}, \tag{4.51}$$

$$\frac{\partial^2}{\partial \eta^2}\left(\sum_{i=1}^{\infty} \frac{p^{\text{in}}_{>i}}{1+b_i(\eta)}\right) = \sum_{i=1}^{\infty} \frac{\partial^2}{\partial \eta^2}\left(\frac{p^{\text{in}}_{>i}}{1+b_i(\eta)}\right) = 2\sum_{i=1}^{\infty} \frac{p^{\text{in}}_{>i}(i^{-1}-(1-\beta))^2}{(1+b_i(\eta))^3}. \tag{4.52}$$

It suffices to check:

$$\sum_{i=1}^{\infty} \sup_{\eta\in\mathcal{I}} \left|\frac{\partial}{\partial \eta}\left(\frac{p^{\text{in}}_{>i}}{1+b_i(\eta)}\right)\right| < \infty, \qquad \sum_{i=1}^{\infty} \sup_{\eta\in\mathcal{I}} \left|\frac{\partial^2}{\partial \eta^2}\left(\frac{p^{\text{in}}_{>i}}{1+b_i(\eta)}\right)\right| < \infty.$$

Note that for $i \geq 1$,

$$\begin{aligned}
\sup_{\eta\in\mathcal{I}} \left|\frac{\partial}{\partial \eta}\left(\frac{p^{\text{in}}_{>i}}{1+b_i(\eta)}\right)\right| &= \sup_{\eta\in\mathcal{I}} \frac{p^{\text{in}}_{>i}|i^{-1}-(1-\beta)|}{(1+b_i(\eta))^2} \\
&\leq (2-\beta)\sup_{\eta\in\mathcal{I}} \frac{p^{\text{in}}_{>i}}{(1+b_i(\eta))^2} \\
&\leq (2-\beta)(1+(1-\beta)K)^2 p^{\text{in}}_{>i}.
\end{aligned}$$

Recall (4.41), we then have

$$\sum_{i=0}^{\infty} p^{\text{in}}_{>i} = \sum_{i=0}^{\infty}\sum_{k>i} p^{\text{in}}_k = \sum_{k=0}^{\infty}\sum_{i=0}^{k-1} p^{\text{in}}_k = \sum_{k=0}^{\infty} k p^{\text{in}}_k = 1.$$

Hence,

$$\begin{aligned}
\sum_{i=1}^{\infty} \sup_{\eta\in\mathcal{I}} \left|\frac{\partial}{\partial \eta}\left(\frac{p^{\text{in}}_{>i}}{1+b_i(\eta)}\right)\right| &\leq (2-\beta)(1+(1-\beta)K)^2 \sum_{i=0}^{\infty} p^{\text{in}}_{>i} \\
&= (2-\beta)(1+(1-\beta)K)^2 < \infty,
\end{aligned}$$

which implies (4.51). Equation (4.52) then follows by a similar argument. Combining (4.50), (4.51) and (4.52) gives

$$\begin{aligned}
\frac{\partial^2}{\partial \eta^2}\tilde{h}(\eta) &= 2\left(\sum_{i=1}^{\infty} \frac{p^{\text{in}}_{>i}}{1+b_i(\eta)}\right)^{-3} \\
&\quad \times \left[\left(\sum_{i=1}^{\infty} \frac{p^{\text{in}}_{>i}(i^{-1}-(1-\beta))}{(1+b_i(\eta))^2}\right)^2 \left(\sum_{i=1}^{\infty} \frac{p^{\text{in}}_{>i}}{1+b_i(\eta)}\right)\left(\sum_{i=1}^{\infty} \frac{p^{\text{in}}_{>i}(i^{-1}-1+\beta)^2}{(1+b_i(\eta))^3}\right)\right] \\
&< 0,
\end{aligned}$$

by the Cauchy-Schwarz inequality. Hence $\tilde{h}$ is concave on $\mathcal{I}$.

From Lemma 4.7.1, $\psi(\delta_{\text{in}}) = 0$ where $\psi(\cdot)$ is as defined in (4.16). Hence we have $f(\delta_{\text{in}}) = \alpha+\beta$ in a similar derivation to that of (4.47). Also from (4.26), we have $g(\delta_{\text{in}}) = \alpha+\beta$. Hence, δ_{in} is a solution to $f(\delta) = g(\delta)$.

Under the $\delta \mapsto \eta$ reparametrization in (4.48), we have that $\tilde{f}(\eta_{\text{in}}) = \tilde{g}(\eta_{\text{in}})$ where $\eta_{\text{in}} := \delta_{\text{in}}/(1 + \delta_{\text{in}}(1-\beta))$, and also

$$\lim_{\eta \downarrow 0} \tilde{f}(\eta) = \sum_{i=1}^{\infty} p^{\text{in}}_{>i} = 1 - p^{\text{in}}_{>0} = \beta + p^{\text{in}}_0 = \lim_{\eta \downarrow 0} \tilde{g}(\eta).$$

This, along with the concavity of $\tilde{h}$, implies that η_{in} is the unique solution to $\tilde{h}(\eta) = 0$, or equivalently, to $\tilde{f}(\eta) = \tilde{g}(\eta)$ on $\mathcal{I}$.

Let $\tilde{f}_n(\eta) := f_n(\delta(\eta))$, $\tilde{g}_n(\eta) := g_n(\delta(\eta))$. We can show in a similar fashion that $\tilde{\eta} := \tilde{\delta}_{\text{in}}/(1 - \tilde{\delta}_{\text{in}}(1-\tilde{\beta}))$ is the unique solution to $\tilde{f}_n(\eta) = \tilde{g}_n(\eta)$. Using an analogue of the arguments in the proof of Theorem 4.7.2, we have

$$\sup_{\eta \in \mathcal{I}} |\tilde{f}_n(\eta) - \tilde{f}(\eta)| \xrightarrow{\text{a.s.}} 0, \quad \sup_{\eta \in \mathcal{I}} |\tilde{g}_n(\eta) - \tilde{g}(\eta)| \xrightarrow{\text{a.s.}} 0,$$

and therefore $\tilde{\eta} \xrightarrow{\text{a.s.}} \eta_{\text{in}}$. Since $\delta \mapsto \eta$ is a one-to-one transformation from $[\epsilon, K]$ to $\mathcal{I}$, we have that $\tilde{\delta}_{\text{in}}$ is the unique solution to $f_n(\delta) = g_n(\delta)$ and that $\tilde{\delta}_{\text{in}} \xrightarrow{\text{a.s.}} \delta_{\text{in}}$. On the other hand, $\tilde{\alpha}$ can be solved uniquely by plugging $\tilde{\delta}_{\text{in}}$ into (4.47) and is also strongly consistent, which completes the proof.

□

Acknowledgement

Research of the four authors was partially supported by Army MURI grant W911NF-12-1-0385. Don Towsley from University of Massachusetts introduced us to the model and within his group, James Atwood graciously supplied us with a simulation algorithm designed for a class of growth models broader than the one specified in Section 4.2.1; this later became Atwood et al. (2015). Joyjit Roy, formerly of Cornell, created an efficient algorithm designed to capitalize on the linear growth structure. Finally, we appreciate the many helpful and sensible comments from the referees and editors.

Bibliography

J. Aaronson, R.M. Burton, H. Dehling, D. Gilat, T. Hill, and B. Weiss. Strong laws of L- and U-statistics. *Trans. Amer. Math. Soc.*, 348:2845–2866, 1996.

J. Atwood, B. Ribeiro, and D. Towsley. Efficient network generation under general preferential attachment. *Computational Social Networks*, 2(1):7, 2015.

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

S. Bhamidi. Universal techniques to analyze preferential attachment trees: Global and local analysis. 2007. URL `http://www.unc.edu/~bhamidi/preferent.pdf`.

P.J. Bickel and M.J. Wichura. Convergence criteria for multiparameter stochastic processes and some applications. *Ann. Statist.*, 42:1656–1670, 1971.

P. Billingsley. *Probability and measure.* Wiley-Interscience, 3rd edition, 1995.

B. Bollobás, C. Borgs, J. Chayes, and O. Riordan. Directed scale-free graphs. In *Proceedings of the Fourteenth Annual ACM-SIAM Symposium on Discrete Algorithms (Baltimore, 2003)*, pages 132–139, New York, 2003. ACM.

P.J. Brockwell and R.A. Davis. *Time Series: Theory and Methods.* Springer, New York., 1991.

S. Csörgő. Limit behaviour of the empirical characteristic function. *Ann. Probab.*, 9:130–144, 1981a.

S. Csörgő. Multivariate characteristic functions and tail behaviour. *Z. Wahrsch. verw. Gebiete*, 55:197–202, 1981b.

S. Csörgő. Multivariate empirical characteristic functions. *Z. Wahrsch. verw. Geb.*, 55: 203–229, 1981c.

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

R.A. Davis and T. Mikosch. The extremogram: A correlogram for extreme events. *Bernoulli*, 15(4):977 – 1009, 2009.

R.A. Davis, T. Mikosch, and I. Cribben. Towards estimating extremal serial dependence via the bootstrapped extremogram. *J. Econometrics*, 170(1):142–152, 2012.

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

L. de Haan and J. de Ronde. Sea and wind: Multivariate extremes at work. *Extremes*, 1: 7–46, 1998.

P. Doukhan. *Mixing: Properties and Examples.* Springer-Verlag, New York, 1994.

J. Dueck, D. Edelmann, T. Gneiting, and D. Richards. The affinely invariant distance correlation. *Bernoulli*, 20:2305–2330, 2014.

R.T. Durrett. *Random Graph Dynamics.* Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge, 2010. ISBN 978-0-521-15016-3.

D. Easley and J. Kleinberg. *Networks, Crowds, and Markets.* Cambridge University Press, Cambridge, 2010.

W. Feller. *An Introduction to Probability Theory and its Applications.*, volume II. Wiley, New York, 2nd edition, 1971.

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

A. Feuerverger and R.A. Mureika. The empirical characteristic function and its applications. *Ann. Statist.*, 4:88–97, 1977.

K. Fokianos and M. Pitsillou. Consistent testing for pairwise dependence in time series. *Technometrics*, 59(2):262–270, 2017.

P. Fryzlewicz. Wild binary segmentation for multiple change-point detection. *Ann. Statist.*, 42(6):2243–2281, 2014.

F. Gao and A. van der Vaart. On the asymptotic normality of estimating the affine preferential attachment network models with random initial degrees. *Stochastic Processes and their Applications*, 2017.

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.

P. Hall and C. C. Heyde. *Martingale Limit Theory and its Application.* Academic Press Inc., New York, 1980. Probability and Mathematical Statistics.

E.J. Hannan and M. Kanter. Autoregressive process with infinite variance. *J. Appl. Probab.*, 14:411–415, 1977.

J. Haslett and A.E. Raftery. Space-time modelling with long-memory dependence: Assessing ireland's wind power. *J. R. Stat. Soc., Ser. C, Appl. Stat.*, 38:1–50, 1989.

Z. Hlávka, M. Hušková, and S.G. Meintanis. Tests for independence in non-parametric heteroscedastic regression models. *J. Multiv. Anal.*, 102:816–827, 2011.

Y. Hong. Hypothesis testing in time series via the empirical characteristic function: A generalized spectral density approach. *J. Amer. Statist. Assoc.*, 94:1201–1220, 1999.

I.A. Ibragimov and Yu.V. Linnik. *Independent and Stationary Sequences of Random Variables.* Wolters–Noordhoff, Groningen., 1971.

S. Jeon and R.L. Smith. Dependence structure of spatial extremes using threshold approach. arXiv:1209.6344v1, 2014.

H. Joe, R.L. Smith, and I. Weissman. Bivariate threshold methods for extremes. *JRSS. B.*, 54(1):171–183, 1992.

J. H. Jones and M. S. Handcock. An assessment of preferential attachment as a mechanism for human sexual network formation. *Proceedings of the Royal Society of London B: Biological Sciences*, 270(1520):1123–1128, 2003a.

J.H. Jones and M.S. Handcock. Social networks (communication arising): Sexual contacts and epidemic thresholds. *Nature*, 423(6940):605–606, 2003b.

E.D. Kolaczyk and G. Csárdi. *Statistical Analysis of Network Data with R.* Use R! Springer, New York, 2014.

P. Krapivsky, G. Rodgers, and S. Redner. Degree distributions of growing networks. *Phys. Rev. Lett*, 86, 2001.

P.L. Krapivsky and S. Redner. Organization of growing random networks. *Physical Review E*, 63(6):066123:1–14, 2001.

U. Krengel. *Ergodic Theorems. With a supplement by Antoine Brunel.* Walter de Gruyter, Berlin., 1985.

R.A. Kronmal and A.V. Peterson Jr. On the alias method for generating random variables from a discrete distribution. *The American Statistician*, 33(4):214–218, 1979.

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.

J. Kunegis. Konect: the Koblenz network collection. In *Proceedings of the 22nd International Conference on World Wide Web*, pages 1343–1350. ACM, 2013.

H.H. Kuo. Gaussian measures in Banach spaces. In *Lecture Notes in Math.*, volume 463. Springer, Berlin., 1975.

R. Lyons. Distance covariance in metric spaces. *Ann. Probab.*, 41:3284–3305, 2013.

A. Mallik, B. Sen, M. Banerjee, and G. Michailidis. Threshold estimation based on a p-value framework in dose-response and regression settings. *Biometrika*, 98(4):887–900, 2011.

S.G. Meintanis and G. Iliopoulos. Fourier methods for testing multivariate independence. *Comput. Statist. Data Anal.*, 52:1884–1895, 2008.

S.G. Meintanis, J. Ngatchou-Wandji, and E. Taufer. Goodness-of-fit tests for multivariate stable distributions based on the empirical characteristic function. *J. Multivariate Anal.*, 140:171–192, 2015.

E.S. Page. Continuous inspection schemes. *Biometrika*, 41(1):100–115, 1954.

D.N. Politis, J.P. Romano, and M. Wolf. *Subsampling.* Springer, New York, 1999.

B. Rémillard. Discussion of: Brownian distance covariance. *Ann. Statist.*, 3:1295–1298, 2009.

S.I. Resnick. Hidden regular variation, second order regular variation and asymptotic independence. *Extremes*, 5:303–336, 2002.

S.I. Resnick. *Heavy Tail Phenomena: Probabilistic and Statistical Modeling.* Springer Series in Operations Research and Financial Engineering. Springer-Verlag, New York, 2007. ISBN: 0-387-24272-4.

S.I. Resnick. *Extreme Values, Regular Variation and Point Processes.* Springer, New York, 2008. ISBN 978-0-387-75952-4. Reprint of the 1987 original.

S.I. Resnick and G. Samorodnitsky. Tauberian theory for multivariate regularly varying distributions with application to preferential attachment networks. *Extremes*, 18(3):349–367, 2015.

A. Rinaldo, S. Petrović, and S. E. Fienberg. Maximum likelihood estimation in the β-model. *Ann. Statist.*, 41(3):1085–1110, 2013.

G. Samorodnitsky. *Stochastic Processes and Long Range Dependence.* Springer, New York., 2016.

G. Samorodnitsky, S. Resnick, D. Towsley, R. Davis, A. Willis, and P. Wan. Nonstandard regular variation of in-degree and out-degree in the preferential attachment model. *Journal of Applied Probability*, 53(1):146–161, March 2016. doi: 10.1017/jpr.2015.15.

D. Sejdinovic, B. Sriperumbudur, A. Gretton, and K. Fukumizu. Equivalence of distance-based and rkhs-based statistics in hypothesis testing. *Ann. Statist.*, 41:2263–2291, 2013.

C. Stărică. Multivariate extremes for models with constant conditional correlations. *J. Empir. Finan.*, 6:515–553, 1999.

G.J. Székely and M.L. Rizzo. Brownian distance covariance. *Ann. Appl. Stat.*, 3:1236–1265, 2009.

G.J. Székely and M.L. Rizzo. Partial distance correlation with methods for dissimilarities. *Ann. Statist.*, 42:2382–2412, 2014.

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

R. Tonelli, G. Concas, and M. Locci. Three efficient algorithms for implementing the preferential attachment mechanism in Yule-Simon stochastic process. *WSEAS Trans. Inf. Sci. App*, 7, 2010.

R. van der Hofstad. *Random Graphs and Complex Networks. Vol. 1.* Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge, 2017.

P. Wan and R.A. Davis. Threshold selection for multivariate heavy-tailed data. *Extremes*, 2018.

P. Wan, T. Wang, R.A. Davis, and S.I. Resnick. Fitting the linear preferential attachment model. *Electron. J. Statist.*, 11:3738–3780, 2017.

P. Wan, T. Wang, R.A. Davis, and S.I. Resnick. Are extreme value estimation methods useful for network data? arXiv:1712.07166, 2018.

T. Wang and S.I. Resnick. Multivariate regular variation of discrete mass functions with applications to preferential attachment networks. *Methodology and Computing in Applied Probability*, pages 1–14, 2016.

Wikipedia, 2016. Wikipedia, Mass message senders – Wikipedia, the free encyclopedia. `https://en.wikipedia.org/wiki/Wikipedia:Mass_message_senders`, 2016. Accessed: 2017-01-27.

T. Yan, C. Leng, and J. Zhu. Asymptotics in directed exponential random graph models with an increasing bi-degree sequence. *Ann. Statist.*, 44(1):31–57, 2016.

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