University of Cyprus

DEPARTMENT OF MATHEMATICS AND STATISTICS

# TESTING SERIAL DEPENDENCE BY THE DISTANCE COVARIANCE FUNCTION

DOCTOR OF PHILOSOPHY DISSERTATION

MARIA PITSILLOU

2016

DEPARTMENT OF MATHEMATICS AND STATISTICS

# TESTING SERIAL DEPENDENCE BY THE DISTANCE COVARIANCE FUNCTION

MARIA PITSILLOU

A Dissertation Submitted to the University of Cyprus in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy

NOVEMBER 2016

# VALIDATION PAGE

**Doctoral Candidate:** Maria Pitsillou

**Doctoral Thesis Title:** Testing Serial Dependence by the Distance Covariance Function

*The present Doctoral Dissertation was submitted in partial fulfillment of the requirements for the Degree of Doctor of Philosophy in the* ***Department of Mathematics and Statistics*** *and was approved on November 4th, 2016 by the members of the* ***Examination Committee.***

**Examination Committee:**

**Research Supervisor:** Konstantinos Fokianos, Professor,
Department of Mathematics and Statistics, University of Cyprus ________________

**Committee Member:** Tasos Christofides, Professor,
Department of Mathematics and Statistics, University of Cyprus ________________

**Committee Member:** Thomas Mikosch, Professor,
Department of Mathematics, University of Copenhagen ________________

**Committee Member:** Michael Neumann, Professor,
Department of Mathematics, Friedrich Schiller University Jena ________________

**Committee Member:** Efstathios Paparoditis, Professor,
Department of Mathematics and Statistics, University of Cyprus ________________

# DECLARATION OF DOCTORAL CANDIDATE

*The present doctoral dissertation was submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy of the University of Cyprus. It is a product of original work of my own, unless otherwise mentioned through references, notes, or any other statements.*

Maria Pitsillou

*To my Fay*

*and*

*George*

# Abstract

There has been a considerable recent interest in measuring dependence by employing the concept of *distance covariance function*, a new and appealing measure of dependence for random variables, introduced by Székely et al. (2007). This tool has been recently extended to time series analysis by Zhou (2012), but since then a limited number of works are discussing its properties. In this thesis, we develop a testing methodology based on distance covariance in the context of dependent data, and especially in time series. This is an important research topic because distance covariance - and its normalized form, the so-called distance correlation - can identify interesting links among the data, whereas the traditional correlation coefficient cannot unless the data are Gaussian and/or linearly related. Considering the univariate case, we construct a Box-Pierce type test statistic based on distance covariance for examining independence. Compared to the usual Box-Pierce test statistic - and its modified version, the Ljung-Box test statistic - the number of lags used for the construction of the proposed test statistic is not constant but grows with the sample size. Moreover, we extend the notion of distance covariance to multivariate time series by defining its matrix version. The information contained in this matrix is useful for identifying any possible relationships within and between the time series components. Based on this new concept, we introduce a multivariate Ljung-Box type test statistic with an increasing number of lags, suitable for testing independence.

The contributed R package **dCovTS** is also introduced. There is no available package in the literature regarding the distance covariance theory in time series. The proposed package provides functions that compute and plot distance covariance and correlation functions for both univariate and multivariate time series. Additionally, it includes functions for testing serial independence based on the proposed methodology presented in this

thesis. In the last part of the thesis, we discuss in detail the implementation of the package with several real data examples.

# Περίληψη

Τα τελευταία χρόνια έχει μελετηθεί εκτενώς η εξάρτηση τυχαίων μεταβλητών υιοθετώντας την έννοια της συνάρτησης *distance covariance* (συνδιακύμανση αποστάσεων). Πρόκειται για ένα νέο μέγεθος εξάρτησης, το οποίο έχει εισαχθεί από τους Székely et al. (2007). Πρόσφατα ο Zhou (2012) έχει επεκτείνει την έννοια του distance covariance στις χρονοσειρές. Ωστόσο, έκτοτε ένας περιορισμένος αριθμός άρθρων έχει μελετήσει τις ιδιότητες του νέου αυτού μεγέθους εξάρτησης. Στην παρούσα διατριβή αναπτύσσουμε μία μεθοδολογία για έλεγχο ανεξαρτησίας στις χρονοσειρές βασισμένη στη συνάρτηση distance covariance. Η σημαντικότητα ενός τέτοιου ερευνητικού θέματος προκύπτει από το γεγονός ότι η συνάρτηση distance covariance - και η κανονικοποιημένη της μορφή, η λεγόμενη distance correlation - μπορούν να προσδιορίσουν ενδιαφέρουσες σχέσεις μεταξύ των δεδομένων, τις οποίες η κλασσική συνάρτηση αυτοσυσχέτισης δεν ανιχνεύει, εκτός και αν τα δεδομένα ακολουθούν την κανονική κατανομή ή/και είναι γραμμικώς συσχετισμένα. Εξετάζοντας αρχικά την μονοδιάστατη περίπτωση, κατασκευάζουμε ένα έλεγχο ανεξαρτησίας με στατιστικό τύπου Box-Pierce. Σε αντίθεση με τα κριτήρια Box-Pierce και Ljung-Box, ο αριθμός των χρονικών υστερήσεων που χρησιμοποιείται στην κατασκευή του προτεινόμενου στατιστικού δεν είναι σταθερός αλλά αυξάνεται με το δειγματικό μέγεθος της χρονοσειράς. Επιπλέον, επεκτείνουμε την μεθοδολογία αυτή σε πολυδιάστατες χρονοσειρές ορίζοντας αρχικά την συνάρτηση distance covariance υπό μορφή πίνακα. Ο πίνακας distance covariance παρέχει πληροφορίες σχετικά με τις σχέσεις που δυνατόν να υπάρχουν μεταξύ των διάφορων συνιστωσών της χρονοσειράς. Με βάση αυτό το νέο μέγεθος εξάρτησης κατασκευάζουμε έλεγχο ανεξαρτησίας για πολυδιάστατες χρονοσειρές. Το στατιστικό που προκύπτει έχει την μορφή του πολυδιάστατου στατιστικού Ljung-Box, βασισμένο στους πίνακες distance convariance παρά στους κλασσικούς πίνακες αυτοσυσχέτησης.

Στο τελευταίο μέρος της εργασίας αυτής παρουσιάζουμε το R πακέτο **dCovTS**. Μέχρι τώρα, δεν υπάρχει κάποιο αντίστοιχο πακέτο στην R για υπολογισμό της συνάρτησης distance covariance στις χρονοσειρές. Έτσι, το προτεινόμενο πακέτο παρέχει συναρτήσεις που υπολογίζουν και σχεδιάζουν τις συναρτήσεις distance covariance και distance correlation τόσο για μονοδιάστες όσο και για πολυδιάστατες χρονοσειρές. Επιπρόσθετα, περιλαμβάνει συναρτήσεις που αφορούν στους ελέγχους ανεξαρτησίας όπως εξηγήθηκαν πιο πάνω και παρουσιάζονται σε αυτή την εργασία. Η παρουσίαση του πακέτου και των λειτουργιών του γίνεται μέσω διάφορων πραγματικών δεδομένων.

# Acknowledgments

This thesis would not have been possible without the encouragement and support of many people to which I am indebted and really grateful.

First of all, I would like to express my deep appreciation and sincere gratitude to my advisor, Professor Konstantinos Fokianos, for his excellent supervision. Without his knowledge, extremely helpful guidance, positive outlook and enthusiasm this work would not have been possible. He has been motivating, supportive and constantly available to discuss whenever I get stuck with my research providing the right direction to help me solve problems.

I would also like to thank the other members of my Ph.D. committee, Professor Tasos Christofides, Professor Thomas Mikosch, Professor Michael Neumann and Professor Efstathios Paparoditis, for their time and useful comments improving this work.

Moreover, I would like to acknowledge computational support by the European Grid Infrastructure and in particular Professor Marios Dikaiakos and his research team (Cy-Grid).

Finally, I would like to thank my parents and my sister for their unconditional love and constant support at all times. Certainly, without them, this thesis would not have been possible. I am also deeply indebted to my husband, George, for his personal support in often difficult times of my PhD. I couldn't have asked for a better partner; his faithful support all these years are really appreciated. Above all, I am really grateful to my beloved daughter, Fay, for bringing me so much happiness with her birth almost two years ago.

# Contents

# List of Figures

# List of Tables

# Chapter 1

# Introduction



The problem of measuring and detecting generic serial dependence is often encountered in time series analysis. The classical Pearson autocorrelation function (ACF) is a traditional tool for measuring dependence and constructing tests of independence. Several authors have also considered the use of the spectral density function, see Hong (1996, 1999) among others. Therefore, the methodologies for constructing test statistics for checking serial dependence can be divided into two main categories: time and frequency domain-based methodologies.

It is well known that time domain analysis, includes correlation-based tests such as those proposed by Box and Pierce (1970) and Ljung and Box (1978). The corresponding test statistics employ the ACF to test serial dependence. However, these tests are inconsistent for processes which are dependent but uncorrelated (Romano and Thombs, 1996; Shao, 2011). Another limitation of these tests is that the number of lags included in the construction of a test statistic is held constant in the asymptotic theory (Xiao and Wu, 2014). The latter may be a severe limitation in practice, since the actual dependence may be of higher order (Hong, 2000). Moreover, the ACF is suitable for detecting serial dependence in Gaussian models. Thus, the ACF fails to detect dependence for nonlinear and non-Gaussian models and alternative dependence measures are required. In Section 2.4.2 we provide a brief review of these alternative dependence tools and the corresponding tests of independence.

A different measure of dependence, which is termed *distance covariance function*, has been proposed recently by Székely et al. (2007) (but see also Feuerverger (1993) for an early treatment). It is defined as the weighted $L_2$-norm between the joint characteristic function of two random vectors of arbitrary, but not necessarily of equal dimensions, and their marginal characteristic functions. The sample version of distance covariance function can be viewed as a degenerate $V$-statistic. The limit distribution of degenerate $U$- and $V$-statistics for stationary and ergodic random variables, as well as for weakly dependent random variables, is examined thoroughly in the works by Dehling and Mikosch (1994) and Leucht and Neumann (2013a,b) among others. Since Székely et al.'s (2007) work, there has been a wide range of studies extending the distance covariance definition and methodology in various topics; a detailed literature review is presented in Sections 2.2 and 2.3.

Székely et al.'s (2007) distance covariance methodology is based on the assumption that the underlying data are independent and identically distributed (i.i.d). However, this assumption is often violated in many practical problems. Remillard (2009) proposed to extend the distance covariance methodology to a time series context in order to measure serial dependence. There have been few works on how to explore and measure serial dependence in time series based on distance covariance. Motivated by the work of Székely et al. (2007), Zhou (2012) recently defined the so-called *auto-distance covariance function* (ADCV) - and its rescaled version, the so-called *auto-distance correlation function* (ADCF), for a strictly stationary multivariate time series. Although Zhou (2012) developed a distance covariance methodology for multivariate time series, he did not explore the interrelationships between the various time series components. In Chapter 4, we investigate this by defining the matrix version of pairwise auto-distance covariance and correlation functions.

The key feature of the distance covariance function is that it identifies nonlinear dependence structures which are not detected by the ACF (for instance, stock return data). Indeed, compared to the ACF which measures the strength of linear dependencies and can be equal to zero even when the variables are related, ADCF vanishes only in the case where the observations are independent. This is better understood by examining Figure 1.1. The plot shows the ACF and ADCF of a second order bivariate nonlinear moving average process (NMA(2)) defined by $\mathbf{X}_t = (X_{t;1}, X_{t;2})$ with

$$X_{t;i} \;\;=\;\; \epsilon_t^{(i)}\epsilon_{t-1}^{(i)}\epsilon_{t-2}^{(i)}, \qquad i = 1, 2, \tag{1.1}$$

where $\{\epsilon_t^{(i)}, i = 1, 2\}$ is an i.i.d. sequence of standard normal random variables. It is easy to see that $X_{t;1}$ is independent of $X_{t;2}$ and $\{X_{t;i}, i = 1, 2\}$ are uncorrelated sequences. This fact is clearly discovered by the sample ACF (see Figure 1.1a). It is obvious though that $X_{t;i}$, $i = 1, 2$, is a sequence of dependent random variables; in fact $X_{t;i}$ depends on $X_{t-1;i}$ and $X_{t-2;i}$. This dependence structure is not discovered by the ACF but by the ADCF (see Figure 1.1b). The upper left and lower plots show that there is a lag-2 dependence in each of $X_{t;i}$, $i = 1, 2$. The horizontal bars shown in the plot are computed by the methodology outlined in Section 4.5.2. The rest of plots in Figure 1.1b show the independence of $X_{t;1}$ and $X_{t;2}$.

Zhou (2012) proved the weak consistency of distance covariance under conditions related to the so-called physical dependence measures. In this thesis, we prove the strong consistency of distance covariance under strong mixing conditions. Moreover, Zhou (2012) studied the asymptotic behavior of ADCV at a fixed lag order, while in our work we relax this assumption by considering an increasing number of lags. This is achieved by employing spectral domain methods, which allows us to incorporate a higher number of lags. From a frequency domain point of view, if a stationary time series is serially uncorrelated, then its standardized spectral density is uniformly distributed, i.e. it takes a constant value over the interval $(-\pi, \pi)$. Thus, any deviation of the normalized spectral density from uniformity provides strong evidence of correlation. However, standard spectral density approaches work sufficiently well for Gaussian processes. They become inappropriate for non-Gaussian models since they miss nonlinear processes with zero autocorrelation (for instance, autoregressive conditional heteroscedastic (ARCH), generalized ARCH (GARCH), bilinear, nonlinear or nonlinear moving average (NMA) models; see Priestley (1981); Hong (1999)). Motivated by this fact, Hong (1999) introduced a generalized spectral density approach that captures all form of dependencies, using the empirical characteristic function (ECF) and its derivatives. Some applications of the ECF include the work by Feuerverger (1993) who developed a consistent rank test for bivariate dependence, and Knight and

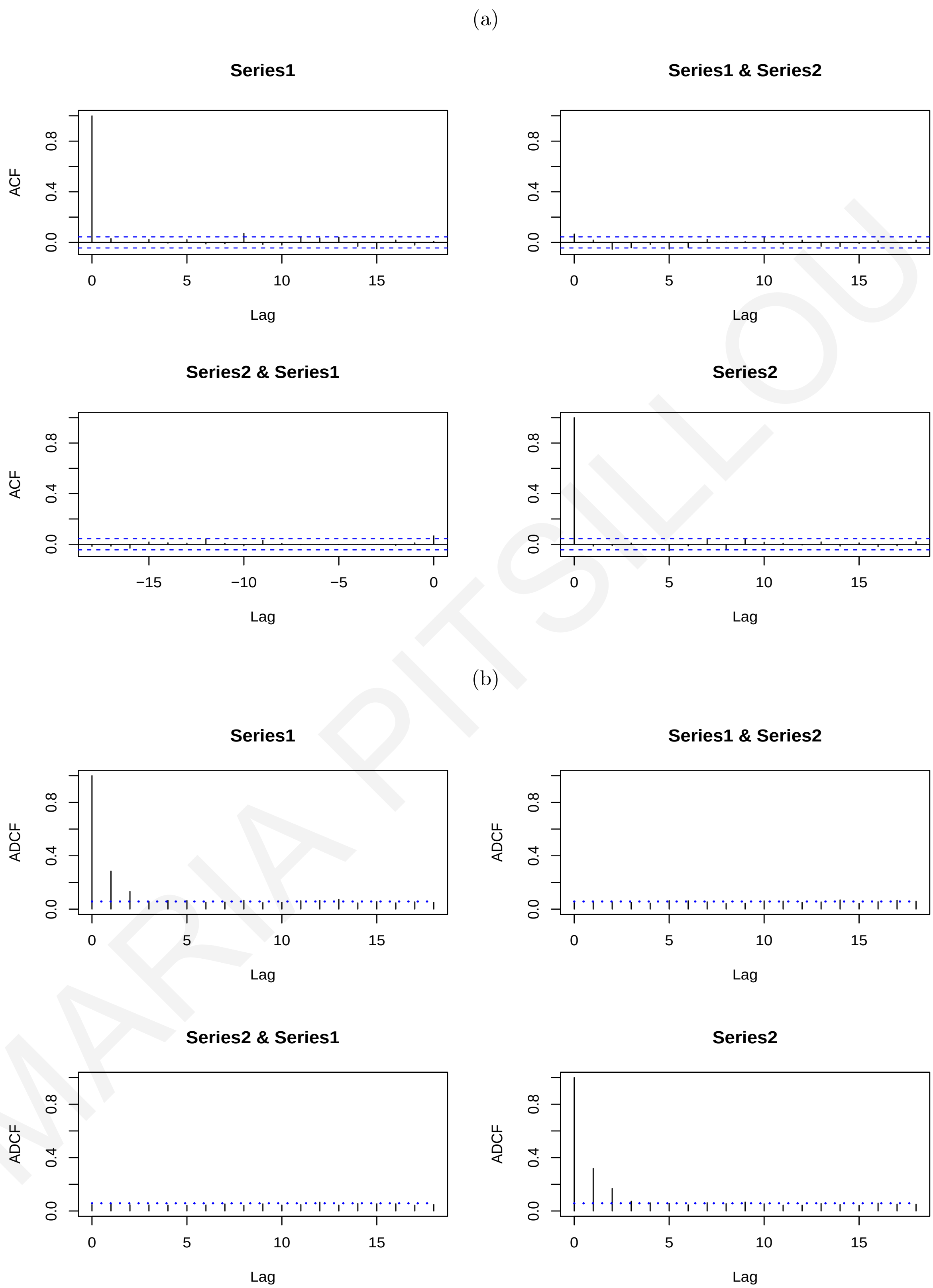



Figure 1.1: (a) Sample ACF of the bivariate NMA(2) series. (b) Sample ADCF of the bivariate NMA(2) series. Results are based on 2000 observations.

Yu (2002) who proposed an estimation method, based on the ECF for strictly stationary processes, leading to consistent and asymptotically normal estimators.

Based on Hong's (1999) approach, in Chapter 3 we employ the ADCV to propose a new univariate test for serial independence resulting to a Box-Pierce type test statistic. The main contribution of this work is that the number of lags included in the construction of test statistic grows with the sample size of the process. Moreover, because of the generalized spectral density approach, the proposed test statistic captures all pairwise dependencies. In addition, it is faster to compute than Hong's (1999) statistic, since it essentially avoids a two-dimensional integration. This contribution builds a bridge between the theory of distance covariance functions proposed by Székely et al. (2007) and the work of Hong (1999).

Subsequently, Chapter 4 deals with the aforementioned distance covariance testing methodology for multivariate processes. In several applications from various scientific fields, such as economics (e.g. Lutkepohl (2005); Verbeek (2012); Kirchgassner et al. (2013)), medicine (e.g. McLachlan et al. (2004); Parmigiani et al. (2003)) or environmetrics (e.g. Hipel and McLeod (1994); Manly (2008)) we usually observe several time series evolving simultaneously. For instance, in Section 4.5.4 we look into a two-dimensional time series of monthly log-returns of IBM stock and S&P 500 index. In Section 4.5.5 we investigate the dependence structure of a twelve-dimensional EEG time series data. Several other examples can be discussed but generally speaking, analyzing each time series separately, without taking into account the rest, might result to wrong conclusions because any interrelationships between series will not be discovered (Priestley, 1981). Thus, we introduce the ADCV matrix (and so ADCF matrix - see again Figure 1.1b) which identifies relationships among components of a vector series. We show that the sample version of ADCV matrix is a consistent estimator of the population ADCV. The sample ADCV matrix serves as a tool to construct tests for pairwise independence for multivariate time series. This is accomplished by following the work of Hong (1999). In particular, we introduce the generalized cross spectral density and the corresponding generalized spectral density matrix. Hence, extending the proposed univariate testing methodology, we construct multivariate test of independence in order to identify whether there is some inherent nonlinear interdependence between the component series. We note that our work in Chapter 4 can be seen as an extension to the work of Székely et al. (2007) since some of the results reported can be transferred to the independent data case. Indeed, using the ADCV matrix for identification of possible dependencies among the components of a random vector could potentially yield to dimension reduction problem.

Rizzo and Szekely (2014) introduced the **energy** package for R (R Core Team, 2014), which is a package that includes functions for the existing distance covariance methodology for random variables. Apart from this package, there is no available package for the corresponding methodology for dependent data. We fill this gap by publishing the **dCovTS** package, where in its first version provides functions for the aforementioned distance covariance methodology in time series. In Chapter 5, we discuss the implementation of these functions by providing several real data examples.

The concluding Chapter 6 summarizes the work and addresses potential extensions for further research. Below, we outline the main contributions of this thesis:

- We construct a univariate test of independence based on ADCV, by providing a Box-Pierce type test statistic which is consistent against all pairwise dependencies.
- We extend the notion of ADCV by considering its matrix version for strictly stationary time series and show how it is interpreted for real data analysis.
- We provide a consistent estimator of the ADCV matrix and obtain simultaneous confidence intervals (the horizontal bars shown in Figure 1.1b) to check visually independence. This method is based on simulation as we explain the theoretical challenges occurred when considering the ADCV matrix.
- We propose a test for testing independence for multivariate time series. The test statistic is quite analog to the multivariate Ljung-Box test statistic (Hosking, 1980; Li and McLeod, 1981) but it is formed with the ADCV matrices instead of the usual autocorrelation matrices.
- The number of lags included in the construction of the proposed test statistics is not constant, but increases with the sample size of the process.
- The suggested testing methodology is available in the package dCovTS for the R language.

## 1.1 Basic Definitions

We conclude this introductory chapter by giving some basic definitions required for the sequel. Throughout the remaining, we denote by bold $\{\mathbf{X}_t, t \in \mathbb{Z}\}$, with index set $\mathbb{Z} = \{0, \pm 1, \pm 2, \cdots\}$, a $d$-dimensional time series process, with components $\{X_{t;i}\}_{i=1}^{d}$.

**Definition 1.1.1** *(Strict Stationarity),* Brockwell and Davis (1991, Definition 1.3.3)
A time series $\{X_t, t \in \mathbb{Z}\}$ is said to be strictly stationary if the joint distributions of $(X_{t_1}, \cdots, X_{t_k})'$ and $(X_{t_1+j}, \cdots, X_{t_k+j})'$ are the same for all positive integers $k$ and for all $t_1, \cdots, t_k$, $j \in \mathbb{Z}$.

**Definition 1.1.2** *(Stationarity - Univariate case),* Brockwell and Davis (1991, Definition 1.3.2)
The time series $\{X_t, t \in \mathbb{Z}\}$ with mean $EX_t = \mu_t$ and covariance function $\gamma_X(r, s) := \text{Cov}(X_r, X_s) = E[(X_r - EX_r)(X_s - EX_s)]$, is said to be stationary if

(i) $E|X_t|^2 < \infty$, for all $t \in \mathbb{Z}$,

(ii) $EX_t = m$, $m$ is constant, for all $t \in \mathbb{Z}$,

(iii) $\gamma_X(r, s) = \gamma_X(r + t, s + t)$, for all $r, s, t \in \mathbb{Z}$.

**Definition 1.1.3** *(Stationarity - Multivariate case),* Brockwell and Davis (1991, Definition 11.1.1)
The time series $\{\mathbf{X}_t\}$ with mean vector $E\mathbf{X}_t := \boldsymbol{\mu}_t$ and covariance matrices $\Gamma(t + j, t) := E[(\mathbf{X}_{t+j} - \boldsymbol{\mu}_{t+j})(\mathbf{X}_t - \boldsymbol{\mu}_t)']$, $j = 0, \pm 1, \pm 2, \cdots$, is said to be stationary if $\boldsymbol{\mu}_t$ and $\Gamma(t + j, t)$ are independent of $t$.

A strictly stationary process with finite second moments is stationary, whereas the converse implication does not hold. We note that our asymptotic theory developed in this thesis is based on strictly stationary time series.

**Definition 1.1.4** *(Autocovariance and autocorrelation functions),* Brockwell and Davis (1991, Section 1.5)
The autocovariance function (ACV) of a univariate stationary time series $\{X_t\}$ is defined

by

$$\gamma(j) = \mathrm{Cov}(X_{t+j}, X_t) = E[(X_{t+j} - EX_{t+j})(X_t - EX_t)], \quad t, j \in \mathbb{Z},$$

with $\gamma(j) = \gamma(-j)$. The autocorrelation function (ACF) of $\{X_t\}$ is defined analogously as the function

$$\rho(j) = \mathrm{Cor}(X_{t+j}, X_t) = \gamma(j)/\gamma(0), \quad t, j \in \mathbb{Z}.$$

**Definition 1.1.5** *(Sample ACV and ACF),* Brockwell and Davis (1991, Definition 1.5.2) The sample ACV of a univariate stationary time series $\{X_t\}$ with a sample of size $n$, is defined by

$$\hat{\gamma}(j) = \frac{1}{n}\sum_{t=1}^{n-1}(X_{t+j} - \overline{X})(X_t - \overline{X}), \quad 0 \leq j < n,$$

and $\hat{\gamma}(-j) = \hat{\gamma}(j)$, $-n < j \leq 0$, where $\overline{X}$ denotes the sample mean $\overline{X} = n^{-1}\sum_{t=1}^{n} X_t$. The corresponding sample ACF is defined in terms of the sample autocovariance function as follows

$$\hat{\rho}(j) = \hat{\gamma}(j)/\hat{\gamma}(0), \quad |j| < n.$$

**Definition 1.1.6** *(Covariance and correlation matrices),* Brockwell and Davis (1991, Section 11.1)

The covariance matrix of a $d$-dimensional stationary time series $\{\mathbf{X}_t\}$ with mean vector $E(\mathbf{X}_t) = \boldsymbol{\mu}$, is given by

$$\Gamma(j) = E\Big[(\mathbf{X}_{t+j} - \boldsymbol{\mu})(\mathbf{X}_t - \boldsymbol{\mu})'\Big] = \Big[\gamma_{ab}(j)\Big]_{a,b=1}^{d}, \quad t, j, \in \mathbb{Z}.$$

The correlation matrix $r(\cdot)$ of $\{\mathbf{X}_t\}$ is defined by

$$r(j) = D^{-1}\Gamma(j)D^{-1} = \Big[\gamma_{ab}(j)/\big(\gamma_{aa}(0)\gamma_{bb}(0)\big)^{1/2}\Big]_{a,b=1}^{d} = \Big[r_{ab}(j)\Big]_{a,b=1}^{d}, \quad j \in \mathbb{Z},$$

where $D$ is a $d \times d$ diagonal matrix $D = diag\{\gamma_{11}(0)^{1/2}, \ldots, \gamma_{dd}(0)^{1/2}\}$.

**Definition 1.1.7** *(Sample covariance and correlation matrices),* Brockwell and Davis (1991, Section 11.2)

The sample covariance matrix of a $d$-dimensional stationary time series $\{\mathbf{X}_t\}$ with sample of size $n$, is given by

$$\widehat{\Gamma}(j) = \begin{cases} \dfrac{1}{n}\displaystyle\sum_{t=1}^{n-j}(\mathbf{X}_{t+j}-\overline{\mathbf{X}}_n)(\mathbf{X}_t-\overline{\mathbf{X}}_n)', & 0 \le j \le n-1; \\ \dfrac{1}{n}\displaystyle\sum_{t=-j+1}^{n}(\mathbf{X}_{t+j}-\overline{\mathbf{X}}_n)(\mathbf{X}_t-\overline{\mathbf{X}}_n)', & -n+1 \le j < 0, \end{cases}$$

where $\overline{\mathbf{X}}_n$ denotes the vector of sample means. Observe that the mean of the $i^{th}$ time series, $\mu_i$, is estimated by $n^{-1}\sum_{t=1}^{n} X_{t;i}$. Analogous to Definition 1.1.6, the sample correlation matrix is then

$$\hat{r}(j) = \left[\hat{\gamma}_{ab}(j)/(\hat{\gamma}_{aa}(0)\hat{\gamma}_{bb}(0))^{1/2}\right]_{a,b=1}^{d},$$

where $\hat{\gamma}_{ab}(j)$ denotes the $(a,b)$-component of $\Gamma(j)$.

**Definition 1.1.8** *(Strong mixing processes),* Doukhan (1994, Section 1.3.3)
A process $\{\mathbf{X}_t\}$ is said to be $\alpha$-mixing (strong mixing) if

$$\alpha(j) = \sup_{A\in\mathcal{F}_{-\infty}^{0}, B\in\mathcal{F}_{j}^{\infty}} |P(A)P(B) - P(A\cap B)| \to 0,$$

as $j \to \infty$, where $\mathcal{F}_j^s$ denotes the $\sigma$-algebra generated by $\{\mathbf{X}_t, j \le t \le s\}$.

Mixing condition implies ergodicity, whereas the converse does not hold. For more details on strong mixing properties and the definitions of other types of mixing processes, the reader is referred to Rosenblatt (1956), Doukhan (1994) and Bradley (1983, 2007).

# Chapter 2

# Measuring Dependence by Means of Distances: A Review



## 2.1 Introduction

Almost a decade ago, Székely et al. (2007) introduced the distance covariance and distance correlation as measures of dependence between two random variables of arbitrary dimensions. Since then, there have been a huge number of works extending the concept of distance covariance in various scientific fields for i.i.d. data. We review several of them in the first two sections of this chapter.

Compared to the case of i.i.d. data, only few papers have appeared in the literature where the distance covariance concept is employed as a measure of dependence for dependent data. In the last part of this chapter, we give a brief survey of these works and establish the basis in order to accomplish the distance covariance testing methodology for dependent data in the subsequent chapters of the thesis. We conclude the chapter by reviewing other dependence measures used for constructing tests of independence in time series. The intension is to examine the performance of our new developed tests to these related tests of independence available in the time series literature. More on this comparison can be found in Chapters 3 and 4 of this thesis.

## 2.2 On Distance Covariance Function

For the next two sections, $X$ denotes a univariate random variable, but $\mathbf{X}$ denotes a multivariate random vector.

Székely et al. (2007) introduced the distance covariance function as a new measure of dependence between random vectors, $\mathbf{X}$ and $\mathbf{Y}$, of arbitrary, not necessarily equal dimensions, say $p$ and $q$ respectively. The definition relies on the joint characteristic function of $\mathbf{X}$ and $\mathbf{Y}$,

$$\phi(\mathbf{t}, \mathbf{s}) = E\Big[\exp\Big(i(\mathbf{t}'\mathbf{X} + \mathbf{s}'\mathbf{Y})\Big)\Big]$$

and the marginal characteristic functions of $\mathbf{X}$ and $\mathbf{Y}$. For instance, in the case of $\mathbf{X}$ its marginal characteristic function is

$$\phi(\mathbf{t}) = E\Big[\exp\Big(i\mathbf{t}'\mathbf{X}\Big)\Big],$$

where $(\mathbf{t}, \mathbf{s}) \in \mathbb{R}^{p+q}$ and $i^2 = -1$. The marginal characteristic function of $\mathbf{Y}$ is defined similarly. The distance covariance function is defined as the nonnegative square root of a weighted $L_2$-distance between the joint and the product of the marginal characteristic functions of the random vectors, namely

$$V^2(\mathbf{X}, \mathbf{Y}) \;=\; \int_{\mathbb{R}^{p+q}} |\phi(\mathbf{t}, \mathbf{s}) - \phi(\mathbf{t})\phi(\mathbf{s})|^2 \, d\mathcal{W}(\mathbf{t}, \mathbf{s}) \tag{2.1}$$

where $\mathcal{W}(\cdot, \cdot) : \mathbb{R}^{p+q} \to \mathbb{R}$ is a weight function for which the above integral exists and whose choice is discussed later on. Rescaling (2.1) leads to the definition of the distance correlation function between $\mathbf{X}$ and $\mathbf{Y}$, which is the positive square root of

$$R^2(\mathbf{X}, \mathbf{Y}) = \begin{cases} \dfrac{V^2(\mathbf{X}, \mathbf{Y})}{\sqrt{V^2(\mathbf{X}, \mathbf{X})V^2(\mathbf{Y}, \mathbf{Y})}}, & V^2(\mathbf{X}, \mathbf{X})V^2(\mathbf{Y}, \mathbf{Y}) > 0; \\ 0, & \text{otherwise.} \end{cases} \tag{2.2}$$

The previous display shows that the distance correlation function is a coefficient analogous to Pearson's correlation coefficient. But, unlike the classical coefficient which measures the linear relationship between $\mathbf{X}$ and $\mathbf{Y}$ and could be zero even when the variables are dependent, the distance correlation vanishes only in the case where $\mathbf{X}$ and $\mathbf{Y}$ are independent. More properties of (2.1) and (2.2) are established later. The choice of the weight function $\mathcal{W}(\cdot,\cdot)$ is crucial. The following lemma gives a solution to this choice. We state it here because of its importance for the sequel.

**Lemma 2.2.1** *Székely et al. (2007, Lemma 1)*
If $0 < \alpha < 2$, then for all $\mathbf{x}$ in $\mathbb{R}^d$

$$\int_{\mathbb{R}^d} \frac{1-\cos(\mathbf{t}'\mathbf{x})}{|\mathbf{t}|_d^{d+\alpha}} d\mathbf{t} = C(d,\alpha)\, |\mathbf{x}|_d^{\alpha},$$

where $|\mathbf{x}|_d$ denotes the Euclidean norm of $\mathbf{x}$ in $\mathbb{R}^d$ and

$$C(d,\alpha) = \frac{2\pi^{d/2}\Gamma(1-\alpha/2)}{\alpha 2^{\alpha}\Gamma((d+\alpha)/2)}, \tag{2.3}$$

with $\Gamma(\cdot)$ denoting the complete gamma function. The integrals at 0 and $\infty$ are meant in the principal value sense: $\lim_{\epsilon\to 0}\int_{\mathbb{R}^d\setminus\{\epsilon B+\epsilon^{-1}B^c\}}$, where $B$ is the unit ball (centered at 0) in $\mathbb{R}^d$ and $B^c$ is the complement of $B$.

A proof of Lemma 2.2.1 can be found in Székely and Rizzo (2005). In view of Lemma 2.2.1 and for $\alpha = 1$, we result to a nonintegrable weight function of the form

$$\mathcal{W}(\mathbf{t},\mathbf{s}) \;=\; (c_p c_q\, |\mathbf{t}|_p^{1+p}\, |\mathbf{s}|_q^{1+q})^{-1}, \tag{2.4}$$

where $c_p$ and $c_q$ are given by

$$c_d = C(d,1) = \frac{\pi^{(1+d)/2}}{\Gamma((1+d)/2)}.$$

The weight function (2.4) results to a scale and rotation invariant distance correlation function. Note that Bakirov et al. (2006) proposed a more complicated weight function. As we will see, the choice of (2.4) yields to computational advantages as opposed to the choice proposed by Bakirov et al. (2006). The main properties of distance correlation function are listed below:

- If $E\Big(|\mathbf{X}|_p + |\mathbf{Y}|_q\Big) < \infty$, then the distance correlation, $R(\mathbf{X},\mathbf{Y})$, satisfies $0 \le R(\mathbf{X},\mathbf{Y}) \le 1$ and $R(\mathbf{X},\mathbf{Y}) = 0$ if and only if $\mathbf{X}$ and $\mathbf{Y}$ are independent.

- $R(\mathbf{X}, \mathbf{Y})$ is invariant under orthogonal transformations of the form $(\mathbf{X}, \mathbf{Y}) \to \Big(\alpha_1 + b_1 C_1 \mathbf{X}, \alpha_2 + b_2 C_2 \mathbf{Y}\Big)$, where $\alpha_1$, $\alpha_2$ are arbitrary vectors, $b_1$, $b_2$ are arbitrary nonzero numbers and $C_1$, $C_2$ are arbitrary orthogonal matrices.

- If $p = q = 1$ and $X$ and $Y$ have standard normal distributions with $r = \mathrm{Cov}(X, Y)$, then

  (i) $R(X, Y) \leq |r|$ and

  (ii) $R^2(X, Y) = \dfrac{r\arcsin r + \sqrt{1 - r^2} - r\arcsin r/2 - \sqrt{4 - r^2} + 1}{1 + \pi/3 - \sqrt{3}}$.

The idea of employing (2.1) for detecting independence was previously discussed by Feuerverger (1993) who considered measures of the form of (2.1). To develop nonparametric test, Feuerverger (1993) first suggested to replace the univariate sample points $X_i$ and $Y_i$, $i = 1, \ldots, n$, by approximate normal score quantities $X_i'$ and $Y_i'$. That is, with

$$X_i' \quad = \quad \Phi^{-1}\left(\frac{\mathrm{rank}(X_i) - 3/8}{n + 1/4}\right), \tag{2.5}$$

with $\Phi$ denoting the $N(0, 1)$ distribution function and similarly for $Y_i'$, Feuerverger (1993) proposed the statistic

$$\int_{\mathbb{R}^2} \left|\tilde{\phi}(t, s) - \tilde{\phi}(t)\tilde{\phi}(s)\right|^2 |t|^{-2} \, |s|^{-2} \, dtds,$$

where $\tilde{\phi}(t, s) = n^{-1} \sum_{j=1}^{n} e^{i(tX_j' + sY_j')}$ is the empirical joint characteristic function and $\tilde{\phi}(t) := \tilde{\phi}(t, 0)$ and $\tilde{\phi}(s) := \tilde{\phi}(0, s)$ are the empirical marginal characteristic functions. Clearly, this statistic is similar to the statistic proposed by Székely et al. (2007) (the empirical version of (2.1) given in (2.6)), where the main differences are the use of formula (2.5) and the restriction to the univariate case. Considering the unscored data, $X_i$ and $Y_i$, Feuerverger (1993) also considered a second statistic which is identical to that of Székely et al. (2007) (equation (2.1)), providing interesting choices for the integrable weight function $\mathcal{W}(\cdot, \cdot)$. More on the comparison between the distance covariance and the statistics proposed by Feuerverger (1993) can be found in Gretton et al. (2009).

## 2.3 Estimation and Testing

### 2.3.1 Estimation

In this section, we consider the empirical counterparts of (2.1) and (2.2), whose definition is based on the weighting function given by (2.4). Suppose that $\{\mathbf{X}_t\}$, $\{\mathbf{Y}_t\}$ is a random sample of size $n$ from the joint distribution of the random vectors $\mathbf{X}$ and $\mathbf{Y}$. Substituting the empirical joint characteristic function

$$\hat{\phi}(\mathbf{t}, \mathbf{s}) = n^{-1} \sum_{j=1}^{n} \exp\Big(i\mathbf{t}'\mathbf{X}_j + i\mathbf{s}'\mathbf{Y}_j\Big)$$

and the corresponding marginal characteristic functions $\hat{\phi}(\mathbf{t}) := \hat{\phi}(\mathbf{t}, 0)$ and $\hat{\phi}(\mathbf{s}) := \hat{\phi}(0, \mathbf{s})$ in (2.1), an intuitive sample version of distance covariance is given by the square root of

$$\widehat{V}^2(\mathbf{X}, \mathbf{Y}) = \frac{1}{c_p c_q} \int_{\mathbb{R}^{p+q}} \frac{\left|\hat{\phi}(\mathbf{t}, \mathbf{s}) - \hat{\phi}(\mathbf{t})\hat{\phi}(\mathbf{s})\right|^2}{|\mathbf{t}|_p^{1+p} \, |\mathbf{s}|_q^{1+q}} d\mathbf{t} d\mathbf{s}. \tag{2.6}$$

The empirical distance covariance and correlation measures can also be viewed as functions of the double centered distance matrices of the samples. In particular, we consider $n \times n$ Euclidean pairwise distance matrices with elements $(a_{ij}) = \Big(|\mathbf{X}_i - \mathbf{X}_j|_p\Big)$ and $(b_{ij}) = \Big(|\mathbf{Y}_i - \mathbf{Y}_j|_q\Big)$. These matrices are double centered so that their row and column sums are equal to zero. In other words, let

$$\begin{aligned} A_{ij} &= a_{ij} - \bar{a}_{i.} - \bar{a}_{.j} + \bar{a}_{..}, \\ B_{ij} &= b_{ij} - \bar{b}_{i.} - \bar{b}_{.j} + \bar{b}_{..}, \end{aligned}$$

where $\bar{a}_{i.} = \Big(\sum_{j=1}^{n} a_{ij}\Big)/n$, $\bar{a}_{.j} = \Big(\sum_{i=1}^{n} a_{ij}\Big)/n$, $\bar{a}_{..} = \Big(\sum_{i,j=1}^{n} a_{ij}\Big)/n^2$. Similarly for $\bar{b}_{i.}$,$\bar{b}_{.j}$ and $\bar{b}_{..}$. The sample distance covariance is defined by the square root of the statistic

$$\widehat{V}^2(\mathbf{X}, \mathbf{Y}) = \frac{1}{n^2} \sum_{i,j=1}^{n} A_{ij} B_{ij}. \tag{2.7}$$

The simplicity of the sample distance covariance version (2.7) is proved in Székely et al. (2007, Theorem 1) and it is essentially based on Lemma 2.2.1. It is obvious that we can directly compute the sample distance correlation in terms of (2.7) by (2.2) as

$$\widehat{R}^2(\mathbf{X},\mathbf{Y}) = \begin{cases} \dfrac{\widehat{V}^2(\mathbf{X},\mathbf{Y})}{\sqrt{\widehat{V}^2(\mathbf{X},\mathbf{X})\widehat{V}^2(\mathbf{Y},\mathbf{Y})}}, & \widehat{V}^2(\mathbf{X},\mathbf{X})\widehat{V}^2(\mathbf{Y},\mathbf{Y}) > 0; \\ 0, & \text{otherwise.} \end{cases}$$

Some of the main properties of sample distance covariance and correlation functions are outlined as follows:

- The estimators of distance covariance and distance correlation are both consistent, that is

$$\lim_{n\to\infty}\widehat{V}^2(\mathbf{X},\mathbf{Y}) = V^2(\mathbf{X},\mathbf{Y}),$$
$$\lim_{n\to\infty}\widehat{R}^2(\mathbf{X},\mathbf{Y}) = R^2(\mathbf{X},\mathbf{Y}),$$

  almost surely.

- $\widehat{V}^2(\mathbf{X},\mathbf{Y}) \geq 0$, where the equality holds when $\mathbf{X}$ and $\mathbf{Y}$ are independent.
- $0 \leq \widehat{R}(\mathbf{X},\mathbf{Y}) \leq 1$.
- $\widehat{R}(\mathbf{X}, \alpha + b\mathbf{X}C) = 1$, where $\alpha$ is a vector, $b$ is a nonzero real number and $C$ is an orthogonal matrix.

An unbiased estimator of distance covariance which was proposed by Székely and Rizzo (2014), is given by

$$\widetilde{V}^2(\mathbf{X},\mathbf{Y}) = \frac{1}{n(n-3)} \sum_{i\neq j} \widetilde{A}_{ij}\widetilde{B}_{ij} \tag{2.8}$$

for $n > 3$, where $\widetilde{A}_{ij}$ denotes the $(i,j)$th entry of the new centered matrix, or the so-called $\mathcal{U}$-centered matrix $\widetilde{A}$, defined by

$$\widetilde{A}_{ij} = \begin{cases} a_{ij} - \frac{1}{n-2}\sum_{l=1}^{n} a_{il} - \frac{1}{n-2}\sum_{k=1}^{n} a_{kj} + \frac{1}{(n-1)(n-2)}\sum_{k,l=1}^{n} a_{kl}, & i \neq j; \\ 0, & i = j. \end{cases}$$

Note that $\widetilde{A}_{ij}$ have the additional property that $E[\widetilde{A}_{ij}] = 0$ for all $i$,$j$. Although this centering turns out to be more complicated process than the process of obtaining (2.7), there are advantages of using the new centering. Following Székely and Rizzo (2014) define the Hilbert space $\mathcal{H}_n = \{\widetilde{A} : A = (a_{ij}) \in \mathcal{S}_n\}$, where $\mathcal{S}_n$ is the linear span of all $n \times n$ distance matrices $(a_{ij})$. The inner product of any pair of elements $C = (C_{ij})$ and $D = (D_{ij})$ in the linear span of $\mathcal{H}_n$ is defined by

$$\left(C \cdot D\right) = \frac{1}{n(n-3)} \sum_{i \neq j} C_{ij} D_{ij}.$$

The main advantages of using $\mathcal{U}$-centered matrices are listed below (Székely and Rizzo, 2014, Lemma 1):

1. $\widetilde{(\widetilde{A})} = \widetilde{A}$. That is, if $B$ is a matrix obtained by $\mathcal{U}$-centering the matrix $\widetilde{A} \in \mathcal{H}_n$, then $B = \widetilde{A}$.
2. $\widetilde{A}$ is invariant to double centering. More precisely, if $B$ is a matrix obtained by double centering the matrix $\widetilde{A}$, then it holds that $B = \widetilde{A}$.
3. Considering $B$ a matrix obtained by adding a constant $c$ to the off-diagonal elements of $\widetilde{A}$, then $\widetilde{B} = \widetilde{A}$. This invariance property with respect to $c$ holds only for the $\mathcal{U}$-centered matrices but it does not hold for the double centered matrices.

### 2.3.2 Asymptotic Tests

Following Székely et al. (2007, Theorem 5), it can be proved that under the null hypothesis of independence between the random vectors $\mathbf{X}$ and $\mathbf{Y}$, the statistic

$$n\widehat{V}^2(\mathbf{X}, \mathbf{Y}) \to Q \equiv \sum_{j=-\infty}^{\infty} \lambda_j Z_j^2,$$

in distribution, where $Z_j$ are independent standard normal variables, $\{\lambda_j\}$ are eigenvalues which depend on the joint distribution of the random vectors $(\mathbf{X}, \mathbf{Y})$ and $E[Q] = 1$. Large values of the proposed statistic support the alternative hypothesis that $\mathbf{X}$ and $\mathbf{Y}$ are not independent.

Székely and Rizzo (2014) defined the partial distance covariance and partial distance correlation to measure the dependence of two random vectors $\mathbf{X}$, $\mathbf{Y}$ given a third random vector $\mathbf{Z}$, where $\mathbf{X}$, $\mathbf{Y}$ and $\mathbf{Z}$ are in arbitrary dimensions. Based on these measures, the authors proposed a test for testing conditional independence. Testing for conditional independence based on distance covariance has also been studied by Poczos and Schneider (2012) and Wang et al. (2015).

### 2.3.3 Extensions of Distance Covariance

The $\alpha$-distance covariance function is a generalization of the distance covariance function introduced by Székely et al. (2007). For $\alpha \in (0, 2)$ the $\alpha$-distance covariance function is defined as the positive square root of

$$V^{2(\alpha)}(\mathbf{X}, \mathbf{Y}) = \frac{1}{C(p, \alpha)C(q, \alpha)} \int_{\mathbb{R}^{p+q}} \frac{|\phi(\mathbf{t}, \mathbf{s}) - \phi(\mathbf{t})\phi(\mathbf{s})|^2}{|\mathbf{t}|_p^{\alpha+p} \, |\mathbf{s}|_q^{\alpha+q}} d\mathbf{t} d\mathbf{s}, \tag{2.9}$$

where $C(p, \alpha)$ and $C(q, \alpha)$ are given by (2.3). Following similar strategy of obtaining (2.7), (2.9) can be computed by defining $a_{ij} = |\mathbf{X}_i - \mathbf{X}_j|_p^{\alpha}$ and $b_{ij} = |\mathbf{Y}_i - \mathbf{Y}_j|_q^{\alpha}$. This modified distance covariance measure unifies and extends the theory of distance covariance since (2.1) is a special case of (2.9) for $\alpha = 1$. Dueck et al. (2014) introduced an affinely invariant version of the distance correlation, whereas in Dueck et al. (2015), the authors considered the problem of computing distance correlation when the underlying joint distribution of the random vectors belongs to the class of Lancaster distributions.

Another extension was proposed by Székely and Rizzo (2009) who considered the notion of Brownian distance covariance which is based on Brownian motion/Wiener process for random vectors $\mathbf{X} \in \mathbb{R}^p$ and $\mathbf{Y} \in \mathbb{R}^q$. First, recall that a Wiener process $\{W(\mathbf{t}) : \mathbf{t} \in \mathbb{R}^d\}$ with expectation zero has covariance

$$\mathrm{Cov}\big(W(\mathbf{t}), W(\mathbf{s})\big) = |\mathbf{t}|_d + |\mathbf{s}|_d - |\mathbf{t} - \mathbf{s}|_d \,. \tag{2.10}$$

The authors considered two independent Brownian motions $\{W(\mathbf{t}) : \mathbf{t} \in \mathbb{R}^p\}$ and $\{W'(\mathbf{s}) : \mathbf{s} \in \mathbb{R}^q\}$ with covariance function (2.10) and defined the Brownian covariance of $\mathbf{X}$ and $\mathbf{Y}$

as the positive number whose square is

$$W^2(\mathbf{X},\mathbf{Y}) \;=\; E[\mathbf{X}_W \mathbf{X}'_W \mathbf{Y}_{W'} \mathbf{Y}'_{W'}], \tag{2.11}$$

where $\mathbf{X}_W$ is the $W$-centered version of $\mathbf{X}$ with respect to $W$ given by

$$\mathbf{X}_W \;=\; W(\mathbf{X}) - E[W(\mathbf{X})|W],$$

whenever the conditional expectation exists and $(W, W')$ are independent of $(\mathbf{X},\mathbf{Y},\mathbf{X}',\mathbf{Y}')$. The authors proved the surprising result that the Brownian covariance coincides with the population distance covariance (Székely and Rizzo, 2009, Theorem 8), that is

$$\begin{aligned} W^2(\mathbf{X},\mathbf{Y}) &= E\left|\mathbf{X}-\mathbf{X}'\right|_p \left|\mathbf{Y}-\mathbf{Y}'\right|_q + E\left|\mathbf{X}-\mathbf{X}'\right|_p E\left|\mathbf{Y}-\mathbf{Y}'\right|_q \\ &- E\left|\mathbf{X}-\mathbf{X}'\right|_p \left|\mathbf{Y}-\mathbf{Y}''\right|_q - E\left|\mathbf{X}-\mathbf{X}''\right|_p \left|\mathbf{Y}-\mathbf{Y}'\right|_q \\ &= V^2(\mathbf{X},\mathbf{Y}), \end{aligned}$$

with $(\mathbf{X}',\mathbf{Y}')$ and $(\mathbf{X}'',\mathbf{Y}'')$ being independent copies of $(\mathbf{X},\mathbf{Y})$. Székely and Rizzo (2009, 2013) have also extended this idea to other stochastic processes, such as fractional Brownian motions and to more general Gaussian processes, like the Laplace-Gaussian process. Moreover, a familiar special case is obtained when a bivariate random vector $(X, Y)$ is considered, that is when $p = q = 1$. In particular, if in (2.11) the one-dimensional stochastic processes $W$ and $W'$ are replaced by the identity function $id$ such that $X_{id} = X - E[X]$ and $Y_{id} = Y - E[Y]$, then we obtain the square of the classical Pearson covariance.

Remillard (2009) suggested to extend the notion of distance covariance by replacing the samples $\{\mathbf{X}_i\}$, $\{\mathbf{Y}_i\}$ by their normalized ranks, that is $\{R_{\mathbf{X},i}/n\}$ and $\{R_{\mathbf{Y},i}/n\}$ respectively, where $R_{X,ij}$ is the rank of $X_{ij}$. He further suggested to generalize the distance covariance to measure dependence between more than two random vectors, say $\mathbf{X}_1, \ldots, \mathbf{X}_d$, taking values in $\mathbb{R}^{p_1}, \ldots, \mathbb{R}^{p_d}$ respectively. Belu (2012) achieved this by defining the distance covariance as the difference between the $d$-dimensional joint characteristic function,

$$\phi(\mathbf{t}_1, \ldots, \mathbf{t}_d) = E\Big[\exp\Big(i(\mathbf{t}'_1\mathbf{X}_1 + \cdots + \mathbf{t}'_d\mathbf{X}_d)\Big)\Big],$$

and the product of the individual marginals,

$$\phi(\mathbf{t}_j) = E\Big[\exp\Big(i(\mathbf{t}_j'\mathbf{X}_j)\Big)\Big],$$

for $j = 1, \ldots, d$, $(\mathbf{t}_1, \ldots, \mathbf{t}_d) \in \mathbb{R}^{p_1 + \ldots p_d}$ and $i^2 = -1$, weighted similarly as in Székely et al. (2007). In particular, this extended version of distance covariance is defined as the positive square root of

$$V^2(\mathbf{X}_1, \ldots, \mathbf{X}_d) = \frac{1}{c_{p_1} \ldots c_{p_d}} \int_{\mathbb{R}^{p_1 + \cdots + p_d}} \frac{|\phi(\mathbf{t}_1, \ldots, \mathbf{t}_d) - \phi(\mathbf{t}_1) \ldots \phi(\mathbf{t}_d)|^2}{|\mathbf{t}_1|_{p_1}^{\alpha + p_1} \ldots |\mathbf{t}_d|_{p_d}^{\alpha + p_d}} d\mathbf{t}_1 \ldots d\mathbf{t}_d,$$

where $c_{p_j}$, $j = 1, \ldots, d$, is given by (2.3) and $\alpha$ is a positive constant that lies in the interval (0,2).

### 2.3.4 Distance Covariance in the Machine Learning Literature

Distance covariance and all its modifications discussed in Section 2.3.3, are mainly based on Euclidean distance spaces. Kosorok (2009) proposed a generalization of the Brownian distance covariance by considering alternative norms to Euclidean norms in order to increase power of the tests based on distance covariance. In addition, Lyons (2013) generalized the theory of distance covariance from Euclidean spaces to metric spaces of negative type. Moving far away from dependence measures based on Euclidean distances, one may also consider kernel-based measures of dependence into reproducing kernel Hilbert spaces (RKHS), as established in the machine learning community. Gretton et al. (2005) introduced the Hilbert-Schmidt independence criterion (HSIC) as the associated test statistic of testing independence. Extending Lyons's (2013) work, Sejdinovic et al. (2013) considered the link between the distance covariance function and the HSIC. For more details on recent methods for measuring dependence based on distance covariance one can referred to Josse and Holmes (2014) and the references therein.

## 2.4 Dependence Measures Based on Distances in Time Series and Related Tests of Independence

### 2.4.1 On Auto-Distance Covariance Function

Székely et al. (2007) distance covariance methodology discussed in Sections 2.2 and 2.3 is based on the assumption that the observations are i.i.d. However, in many applied problems this assumption is violated, so Remillard (2009) proposed an extension of the distance covariance methodology to the case of non-i.i.d. observations, especially for time series data, for measuring serial dependence. A few researchers since then have developed a distance covariance methodology in the context of time series (Zhou, 2012; Dueck et al., 2014; Davis et al., 2016). Zhou (2012) defined the ADCV for strictly stationary multivariate processes and studied its asymptotic distribution at a fixed lag. He further showed that the limiting distribution is identical to the limiting distribution obtained by Székely et al. (2007) in the case of independent data (see Section 2.3.2). Moreover, although Zhou (2012) defined the distance correlation coefficient to explore temporal nonlinear dependence structure in multivariate time series, his approach does not identify possible interrelationships between various time series components. Our work is different in various ways. Fokianos and Pitsillou (2016a,b) relaxed the assumption of fixed lag and constructed both univariate (Chapter 3) and multivariate (Chapter 4) tests of independence based on ADCV by considering an increasing number of lags. The proposed multivariate test of independence is based on the novel notion of distance covariance matrix which is calculated by considering all pairs of random variables and it identifies possible dependencies among and between different components of a vector time series.

Dueck et al. (2014) extended the notion of distance correlation to the affinely invariant cross-distance correlation to multivariate time series, by considering a four-dimensional time series of wind observations at and near the Stateline wind energy center in the Pacific Northwest of United States. However, they emphasized that this part of their study is purely exploratory and provided for illustration purposes in order to develop parametric and nonparametric bootstrap tests for Gaussianity in future work. Davis et al. (2016) also applied distance covariance methodology to stationary univariate and multivariate time series to study serial dependence under various choices of the weight function $\mathcal{W}(\cdot,\cdot)$. Moreover, one of their main results was concerned with the asymptotic distribution of the empirical ADCV when applied to the residuals of a fitted autoregressive process with finite or infinite variance.

Extending the work by Hong (1999), Chen and Hong (2012) developed a nonparametric test for the Markov property of a multivariate time series based on the conditional characteristic function. Based on the fact that the correlation coefficient is suitable for Gaussian data but it fails for nonlinear cases, many authors defined the correlation function in a local sense, including Tjøstheim and Hufthammer (2013), Berentsen and Tjøstheim (2014), Støve et al. (2014) and Støve and Tjøstheim (2014) among others. In the context of time series, Lacal and Tjøstheim (2016) defined a new measure of dependence, the so-called local Gaussian autocorrelation that works well for nonlinear models. The authors compared the proposed test statistic to distance covariance function and found that they both work in a similar way.

### 2.4.2 Other Dependence Distance Measures for Time Series Data

There are many other tests for examining serial independence in time series literature using both time domain and frequency domain methodologies, see Tjøstheim (1996) for a review. Recall that $\{\mathbf{X}_t, t \in \mathbb{Z}\}$ denotes a $d$-dimensional time series process, with components $\{X_{t;i}\}_{i=1}^{d}$. The most well known testing procedures for both univariate and multivariate time series are mainly based on the ACF, which usually serves as a measure of dependence. Two widely used univariate correlation-based tests are those proposed by Box and Pierce (1970)

$$\text{BP} = n \sum_{j=1}^{p} \hat{\rho}^2(j),$$

and Ljung and Box (1978)

$$\text{LB} = n(n+2) \sum_{j=1}^{p} (n-j)^{-1} \hat{\rho}^2(j),$$

where $p$ is chosen arbitrarily and denotes the maximum lag order employed for the tests. Most of these procedures were extended to the multivariate case; see Mahdi (2011) for an overview and a new proposed correlation-based test. The multivariate Ljung-Box test statistic (Hosking, 1980; Li and McLeod, 1981; Li, 2004)

$$mLB = n^2 \sum_{j=1}^{p} (n-j)^{-1} \text{trace}\{\widehat{\Gamma}'(j)\widehat{\Gamma}^{-1}(0)\widehat{\Gamma}(j)\widehat{\Gamma}^{-1}(0)\} \tag{2.12}$$

is widely used for testing $\Gamma(1) = \cdots = \Gamma(p) = 0$. Although these tests perform well under linear and Gaussian processes, their performance is poor against general types of nonlinear dependencies including those with zero autocorrelation (ARCH, bilinear, NMA processes). As Tjøstheim (1996) pointed out, although a standard procedure for increasing the power of the correlation-based tests in such nonlinear and non-Gaussian cases is to compute the correlation of the squared observations, generally such a procedure leads to a loss of power compared to the ordinary correlation based tests. Robinson (1991) presented a variety of test statistics for testing serial correlation under the presence of conditional heteroskedasticity, whereas Escanciano and Lobato (2009) further proposed an automatic Portmanteau test that allows for nonlinear dependencies and the lag order $p$ is selected automatically from the data.

Many studies in the literature have considered the problem of measuring dependence between $X_t$ and $Y_t \equiv X_{t-|j|}$ for $j = 0, \pm 1, \cdots$, in terms of the distance between the bivariate distribution of $(X_t, Y_t)$, $F_{X;Y}(x,y) = P(X_t \leq x, Y_t \leq y)$, and the product of their marginal distribution functions, $F_X(x) = P(X_t \leq x)$ and $F_Y(y) = P(Y_t \leq y)$. Well known distance measures for distribution functions are:

- Kolmogorov-Smirnov distance

$$D_1(j) = \sup_{(x,y)\in\mathbb{R}^2} |F_{X;Y}(x,y) - F_X(x)F_Y(y)| \tag{2.13}$$

- Cramer-von Mises type distance

$$D_2(j) = \int_{\mathbb{R}^2} \left\{F_{X;Y}(x,y) - F_X(x)F_Y(y)\right\}^2 dF_{X;Y}(x,y) \tag{2.14}$$

Replacing the theoretical distribution functions with their empirical analogues

$$\begin{aligned}\widehat{F}_{X;Y}(x,y) &= \frac{1}{(n-|j|)}\sum_{t=|j|+1}^{n}\mathbb{I}(X_t \le x)\mathbb{I}(Y_t \le y)\\ \widehat{F}_{X}(x) &= \frac{1}{(n-|j|)}\sum_{t=|j|+1}^{n}\mathbb{I}(X_t \le x)\end{aligned} \tag{2.15}$$

where $\mathbb{I}(\cdot)$ is the indicator function, we can calculate the corresponding estimated measures of dependence $\widehat{D}_1(\cdot)$ and $\widehat{D}_2(\cdot)$. Alternatively, similar distance measures are obtained by employing density functions (see Skaug and Tjøstheim (1996) and Bagnato et al. (2014) for an overview) or characteristic functions (Pinkse, 1998; Hong, 1999). Hong (1999, p. 1206) explained how tests based on the empirical characteristic function may have omnibus power against tests based on the empirical distribution function.

Skaug and Tjøstheim (1993) extended the work by Blum et al. (1961) and considered the asymptotic behavior of the Cramer-von Mises type statistic (2.14) at lag $j$, under ergodicity of $\{X_t\}$. Moreover, they constructed a test for pairwise independence among pairs $(X_t, X_{t-1})$, $(X_t, X_{t-2})$, ..., $(X_t, X_{t-p})$ using the statistic

$$\widehat{D}_{2p} = \sum_{i=1}^{p}\widehat{D}_2(i)$$

where $p$ is a fixed constant denoting the maximum lag order employed for the test. Under the null hypothesis that $\{X_t\}$ consists of i.i.d. random variables, $n\widehat{D}_{2p}$ converges in distribution (as $n \to \infty$) to $\sum_{i,j=1}^{\infty}\lambda_{ij}W_{ij}(p)$ where $\lambda_{ij}$ are nonzero eigenvalues and $\{W_{ij}(p), i, j \ge 1\}$ are i.i.d. Chi-squared variables with $p$ degrees of freedom. However, the latter test statistic is consistent against serial dependencies up to a finite order $p$. From a practical point of view, this may be restrictive since the actual serial dependence may be detected at lags larger than $p$. Motivated by the construction proposed by Skaug and Tjøstheim (1993), Hong (1998) introduced a statistic where the number of lags increases with the sample size and different weights are given to different lags; in other words let

$$V_n = \sum_{i=1}^{n-1}k^2(i/p)(n-i)\widehat{D}_2^2(i)$$

where $k(\cdot)$ is an appropriate chosen kernel function. He further proved that after proper standardization and for large $p$, the statistic is asymptotically standard normally distributed.

In general, incorporating a large number of lags in the asymptotic theory is an issue nicely addressed by the frequency domain framework. Paparoditis (2000, 2001) developed a new goodness-of-fit test with a test statistic based on the distance between the estimator of the ratio between the true and the hypothesized spectral density and the expected value of the estimator under the null. The asymptotic normality along with the properties of the test are discussed. Hong (1996) proposed three test statistics for testing serial independence in univariate time series by comparing a kernel-based standardized spectral density

$$\hat{f}(\omega) = \frac{1}{2\pi} \sum_{j=-(n-1)}^{(n-1)} k(j/p)\hat{\rho}(j)\cos(j\omega), \qquad \omega \in [-\pi, \pi] \tag{2.16}$$

where $p \to \infty$, $p/n \to 0$ and the kernel function $k(\cdot)$ satisfies some standard properties and the null spectral, $f_0(\omega) = 1/2\pi$ via divergence measures. In particular, he used a quadratic norm, the Hellinger metric and the Kullback-Leibler information criterion

$$\begin{aligned}
T_{1n} &= Q(\hat{f}, f_0) = \left[2\pi \int_{-\pi}^{\pi} \left\{\hat{f}(\omega) - f_0(\omega)\right\}^2 d\omega\right]^{1/2}, \\
T_{2n} &= H(\hat{f}, f_0) = \left[\int_{-\pi}^{\pi} \left\{\hat{f}^{1/2}(\omega) - f_0^{1/2}(\omega)\right\}^2 d\omega\right]^{1/2}, \\
T_{3n} &= I(\hat{f}, f_0) = -\int_{\omega:\hat{f}(\omega)>0} \ln\left(\hat{f}(\omega)/f_0(\omega)\right) f_0(\omega) d\omega
\end{aligned}$$

respectively. He further suggested a closed form expression for $T_{1n}$:

$$T_{1n} = n \sum_{j=1}^{n-1} k^2(j/p)\hat{\rho}^2(j).$$

Compared to the classical Portmanteau statistics, like BP and LB, all these suggested statistics, properly standardized, are asymptotically standard normally distributed and are derived without having to impose a specific alternative model. One of the main contributions of Xiao and Wu (2014) was that $T_{1n}$ in its standardized version remains true under the presence of serial correlation, whereas Shao (2011) proved its robustness to conditional heteroscedasticity. Although Hong's (1996) tests incorporate an increasing number of lags, they are based on the classical correlation coefficient and are not consistent against all pairwise dependencies of unknown form. To achieve this, Hong (2000) generalized both Cramer-von Mises and Kolmogorov-Smirnov type test statistics using a generalized spectral theory, in combination with the empirical distribution function. In fact, he first defined the following dependence measure

$$\rho_j^*(x, y) = F_{X;Y}(x, y) - F_X(x)F_Y(y). \tag{2.17}$$

Clearly, the distance dependence measure $\rho_j^*(\cdot, \cdot)$ defined in (2.17) vanishes only in the case where $X_t$ and $Y_t \equiv X_{t-|j|}$ are independent, leading to the observation that $\rho_j^*(\cdot, \cdot)$ can capture all pairwise dependencies including those with zero autocorrelation. Replacing the theoretical distribution functions with the empirical ones given in (2.15), one can find the empirical analogue of (2.17), $\hat{\rho}_j^*(\cdot, \cdot)$. One can observe that Skaug and Tjøstheim (1993) approach was merely based on (2.17). Hong (2000) introduced the generalized spectral density function as the Fourier transform of $\rho_j^*(\cdot, \cdot)$

$$h(\omega, x, y) \;\; = \;\; \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} \rho_j^*(x, y) e^{-ij\omega}, \quad \omega \in [-\pi, \pi], \tag{2.18}$$

and the corresponding generalized spectral distribution function

$$H(\lambda, x, y) \;\; := \;\; 2 \int_0^{\lambda\pi} h(\omega, x, y) d\omega, \quad \lambda \in [0, 1].$$

A kernel-based estimator of (2.18) can be similarly defined as in (2.16), with the sample autocorrelation function $\hat{\rho}(j)$ being replaced by the distance measure $\hat{\rho}_j^*(\cdot, \cdot)$ and so a corresponding estimator $\widehat{H}(\cdot, \cdot, \cdot)$ is derived. The suggested generalized Kolmogorov-Smirnov and Cramer-von Mises type test statistics are then derived by replacing $\widehat{H}(\lambda, x, y)$ and the null $\widehat{H}_0(\lambda, x, y) = \widehat{\rho}_0^*(x, y)\lambda$ in the analogous quantities given in (2.13) and (2.14) respectively. The author also proved the asymptotic normality of statistics under the null hypothesis of independence. Compared to related test statistics, the latter tests do not involve the choice of a lag order, require no moment conditions and have accurate sizes in finite samples.

In a latest work, Dette et al. (2015) introduced a "new" spectrum as the Fourier transform of the so-called copula cross-covariance kernel

$$\rho_j^U(\tau_1, \tau_2) := \mathrm{Cov}(\mathbb{I}(U_t \leq \tau_1), \mathbb{I}(U_{t-j} \leq \tau_2)), \tag{2.19}$$

where $(\tau_1, \tau_2) \in (0, 1)^2$ and $U_t := F_X(X_t)$. They proposed to estimate the corresponding spectral densities associated with measures (2.17) and (2.19) via Laplace periodograms. In addition, they highlighted that in the case of using (2.19), replacing the original observations with their ranks we may achieve an invariance property with respect to transformations of the marginal distributions.

# Chapter 3

# Testing for Pairwise Dependence in Univariate Time Series

## 3.1 Introduction

After having introduced a general framework of distance covariance function as it appears in the statistics literature, we introduce the notion of distance covariance function in time series, the so-called *auto-distance covariance function* (ADCV), and its rescaled version; the *so-called auto-distance correlation function* (ADCF) (Section 3.2). We show that the distance covariance methodology can be motivated by a generalized spectral domain point of view as explained in Section 3.3. In Section 3.4, we suggest the use of a new Box-Ljung type test statistic defined in terms of the ADCV, suitable for detecting pairwise dependencies in univariate stationary time series. We show that, under the null hypothesis of independence and under mild regularity conditions, the test statistic converges to a normal random variable. We illustrate several examples in Section 3.5, that complement our results. We finally note that, although the proposed methodology is for univariate processes, it can be extended for multivariate processes as we will see in the next chapter.

## 3.2 On Auto-Distance Covariance Function

Assume that $\{X_t, t \in \mathbb{Z}\}$ is a univariate strictly stationary time series (see Definition 1.1.1) and suppose that we have available a sample of size $n$. In what follows, we will make the following assumptions for developing the theory:

**Assumption 1** $\{X_t\}$ is a strictly stationary $\alpha$-mixing process with mixing coefficients $\alpha(j)$, $j \geq 1$.

**Assumption 2** $E\,|X_t| < \infty$.

**Assumption 3** The mixing coefficients of $\{X_t\}$, $\alpha(j)$, satisfy $(i)$ $\sum_{j=-\infty}^{\infty} \alpha(j) < \infty$, $(ii)$ $\alpha(j) = O(1/j^2)$.

Assumption 1 is useful for developing theoretical results about the ADCV. It is a rather natural assumption as a first step towards studying the estimation of (3.4) in the context of time series. Assumption 2 guarantees the finiteness of (3.4). Assumption 3$(i)$ implies the existence of the generalized spectral density (3.9), as we will see in the next section. Assumption 3$(ii)$ is the minimal condition needed to obtain a Marcinkiewicz-Zygmund type inequality (Doukhan and Louhichi, 1999, Lemma 6) for the proofs of Lemmas 3.5.1 and 3.5.2 given in the Appendix. Assumption 3$(ii)$ implies 3$(i)$. For a formal definition of a strong mixing process and its mixing coefficients, the reader is referred to Definition 1.1.8 of Chapter 1.

We will define the distance covariance function by resorting to the joint and marginal characteristic functions of the pair $\big(X_t, X_{t-|j|}\big)$. Denote the joint characteristic function of $X_t$ and $X_{t-|j|}$ by $\phi_{|j|}(u,v)$; that is

$$\phi_{|j|}(u,v) \;=\; E\Big[\exp\Big(i(uX_t + vX_{t-|j|})\Big)\Big], \quad j = 0, \pm 1, \pm 2, \ldots,$$

where $(u,v) \in \mathbb{R}^2$, and $i^2 = -1$. Furthermore, let

$$\phi(u) \;=\; E\Big[\exp\Big(iuX_t\Big)\Big],$$

be the marginal characteristic function of $X_t$. Following Hong (1999), and because of the assumed stationarity, we define

$$\sigma_j(u,v) \;=\; \mathrm{Cov}\left(e^{iuX_t}, e^{ivX_{t-|j|}}\right) = \phi_{|j|}(u,v) - \phi(u)\phi(v), \tag{3.1}$$

i.e. (3.1) denotes the covariance function between the two series $e^{iuX_t}$ and $e^{ivX_{t-|j|}}$. From (3.1) we note that $\sigma_j(u,v)$ is simply the difference between the joint characteristic function of $(X_t, X_{t-|j|})$ and the product of their marginals. Hence $\sigma_j(u,v) = 0, \forall\ (u,v) \in \mathbb{R}^2$ implies that the random variables $X_t$ and $X_{t-|j|}$ are independent. Define the $\|.\|_{\mathcal{W}}$-norm of $\sigma_j(u,v)$ by

$$\|\sigma_j(u,v)\|^2_{\mathcal{W}} \;=\; \int_{\mathbb{R}^2} |\sigma_j(u,v)|^2 \, d\mathcal{W}(u,v), \quad j = 0, \pm 1, \pm 2, \ldots, \tag{3.2}$$

where $\mathcal{W}(u,v)$ is an arbitrary positive weight function for which the above integral exists. In particular, the weight function $\mathcal{W}_0(u) = 1/(\pi\,|u|^2)$ yields

$$\mathcal{W}(u,v) \;=\; \mathcal{W}_0(u)\mathcal{W}_0(v) = \frac{1}{\pi\,|u|^2}\frac{1}{\pi\,|v|^2}, \quad (u,v) \in \mathbb{R}^2, \tag{3.3}$$

which results to the ADCV defined by the positive square root of

$$V^2_X(j) \;=\; \frac{1}{\pi^2}\int_{\mathbb{R}^2} \frac{|\sigma_j(u,v)|^2}{|u|^2\,|v|^2} du dv, \quad j = 0, \pm 1, \pm 2, \ldots. \tag{3.4}$$

It is clear from the above definition that $V^2_X(j) \geq 0, \forall\ \ j$, and that $X_t$ and $X_{t-|j|}$ are independent if and only if $V^2_X(j) = 0$. Although Hong (1999) defined implicitly (3.2) by using an integrable weight function $\mathcal{W}(\cdot,\cdot)$, it turns out that utilizing a nonintegrable weight function, like (3.3), yields a closed form expression of the estimate of the auto-distance covariance function. In addition, the calculation of the estimator based on (3.3) is faster than using Hong's (1999) approach. The auto-distance correlation function (ADCF) is the square root of

$$R^2_X(j) = \begin{cases} \dfrac{V^2_X(j)}{V^2_X(0)}, & V^2_X(0) \neq 0; \\ 0, & V^2_X(0) = 0. \end{cases} \tag{3.5}$$

Davis et al. (2016) provided alternative representations of ADCV and ADCF by considering various choices of finite and infinite measures $\mathcal{W}(\cdot,\cdot)$. However, choosing (3.3) leads to the fact that (3.5) is scale invariant and nonzero when $X_t$ and $X_{t-|j|}$ are dependent at lag $j$. In addition, the most important feature of (3.4) is that if it is calculated by using an integrable function then it might miss the potential dependence among observations (Székely et al., 2007, p. 2771). To develop an estimator for (3.4), define

$$\hat{\sigma}_j(u,v) \quad = \quad \hat{\phi}_j(u,v) - \hat{\phi}_j(u,0)\hat{\phi}_j(0,v), \quad j = 0, \pm 1, \pm 2, \dots \tag{3.6}$$

with

$$\hat{\phi}_j(u,v) \quad \equiv \quad \frac{1}{n-|j|} \sum_{t=|j|+1}^{n} e^{i(uX_t + vX_{t-|j|})}.$$

Then, the sample auto-distance covariance function is defined by

$$\widehat{V}_X^2(j) \quad = \quad \frac{1}{\pi^2} \int_{\mathbb{R}^2} \frac{|\hat{\sigma}_j(u,v)|^2}{|u|^2 \, |v|^2} du dv, \quad j = 0, \pm 1, \pm 2, \dots \tag{3.7}$$

Estimator (3.7) can be computed as follows: Let $Y_t = X_{t-|j|}$. Then, based on the sample $\{(X_t, Y_t) : t = 1 + |j|, \dots, n\}$, we can calculate the $(n-|j|) \times (n-|j|)$ Euclidean distance matrices $A = (A_{rl})$ and $B = (B_{rl})$ with elements

$$A_{rl} \quad = \quad a_{rl} - \bar{a}_{r.} - \bar{a}_{.l} + \bar{a}_{..},$$

with $a_{rl} = |X_r - X_l|$, $\bar{a}_{r.} = \left(\sum_{l=1+|j|}^{n} a_{rl}\right)/(n-|j|)$, $\bar{a}_{.l} = \left(\sum_{r=1+|j|}^{n} a_{rl}\right)/(n-|j|)$, $\bar{a}_{..} = \left(\sum_{r,l=1+|j|}^{n} a_{rl}\right)/(n-|j|)^2$, and quite analogously for $B_{rl}$. Then,

$$\widehat{V}_X^2(j) \quad = \quad \frac{1}{(n-|j|)^2} \sum_{r,l=1+|j|}^{n} A_{rl} B_{rl}. \tag{3.8}$$

Clearly (3.8) can be easily implemented for any given time series data, because it is computed by simple summation and multiplication. It is expected to perform better than the usual autocovariance function especially for nonlinear time series models. In addition, it is an appealing measure of dependence since its computation is based on linear combinations of distances among observations. Note that (3.4) (and its empirical analogue) have been studied by Zhou (2012) under the setup of multivariate time series. However, although in this chapter we will be focusing exclusively on univariate responses, our results can be extended to the case of multivariate time series, as we will see in the next chapter.

It is interesting to observe that in the special case of a Gaussian process, the ADCF can be expressed as a function of the usual autocorrelation function. In particular, we have the following proposition.

**Proposition 3.2.1** If $\{X_t\}$ is a Gaussian stationary time series such that $E(X_t) = 0$, $\text{Var}(X_t) = 1$ and $\rho(j) = \text{Cov}(X_t, X_{t-|j|})$, then

(i) $R_X(j) \leq |\rho(j)|, \quad j = 0, \pm 1, \pm 2, \ldots$

(ii) $R_X^2(j) = \dfrac{\rho(j)\text{arcsin}\rho(j) + \sqrt{1-\rho^2(j)} - \rho(j)\text{arcsin}\rho(j)/2 - \sqrt{4-\rho^2(j)} + 1}{1 + \pi/3 - \sqrt{3}}, \quad j = 0, \pm 1, \pm 2, \ldots,$

where $R_X(\cdot)$ is given by (3.5).

Hence, for a Gaussian process $\rho(j) = 0$ implies $R_X^2(j) = 0, \ \forall \ j \neq 0$. A detailed proof can be found in the Appendix of this chapter.

The following proposition shows the consistency of the estimator $\widehat{V}_X^2(\cdot)$.

**Proposition 3.2.2** Suppose that Assumptions 1 and 2 hold true. Then for all $j = 0, \pm 1, \pm 2, \ldots$

$$\widehat{V}_X^2(j) \to V_X^2(j),$$

almost surely, as $n \to \infty$.

The proof of the above proposition is based on similar arguments given in the proof of Székely et al. (2007, Theorem 2). The $\alpha$-mixing condition enables application of the ergodic theorem to the case of time series data. Under mild conditions, Zhou (2012) obtained the weak consistency of $\widehat{V}_X(\cdot)$. Note that in the approach taken by Zhou (2012) it is required that $E|X_t|^{1+\delta} < \infty$ for some $\delta > 0$. In our approach we require $E|X_t| < \infty$. However, Zhou (2012) proved this result using the physical dependence measure suggested by Wu (2005), whereas we employ the notion of $\alpha$-mixing (Rosenblatt, 1956; Doukhan, 1994).

## 3.3 Generalized Spectral Density Approach

We now discuss the connection of ADCV with the work by Hong (1999).

Recall (3.1) and suppose that $\sup_{(u,v)\in\mathbb{R}^2}\sum_j|\sigma_j(u,v)|<\infty$, which holds under Assumption 3($i$). Then, the Fourier transform of $\sigma_j(u,v)$ exists and is given by

$$f(\omega,u,v) \;=\; \frac{1}{2\pi}\sum_{j=-\infty}^{\infty}\sigma_j(u,v)e^{-ij\omega}, \quad \omega\in[-\pi,\pi]. \tag{3.9}$$

When $\sigma_j(u,v)=0$, $\forall\ \ j\neq 0$, then (3.9) reduces to the constant

$$f_0(\omega,u,v) \;=\; \frac{1}{2\pi}\sigma_0(u,v), \quad \omega\in[-\pi,\pi].$$

Therefore testing whether $f$ is constant with respect to $\omega$ implies that all $\sigma_j(u,v)=0$, i.e. $X_t$ and $X_{t-|j|}$ are independent for all $j\neq 0$. Hong (1999) studies a kernel-density estimator of $f(\omega,u,v)$ and calculates its $L_2$-distance from $f_0(\omega,u,v)$ to test for $f$ being constant. In addition to Assumptions 1 - 3 consider also the following:

**Assumption 4** Suppose that $k(\cdot)$ is a kernel function such that $k:\mathbb{R}\to[-1,1]$, is symmetric and is continuous at 0 and all except a finite number of points, with $k(0)=1$, $\int_{-\infty}^{\infty}k^2(z)dz<\infty$ and $|k(z)|\leq C|z|^{-b}$ for large $z$ and $b>1/2$.

Assumption 4 is mild and allows for kernels with bounded or unbounded support. Set the following nonparametric estimator of $f$,

$$\hat{f}_n(\omega,u,v) \;=\; \frac{1}{2\pi}\sum_{j=-(n-1)}^{(n-1)}(1-|j|/n)^{1/2}k(j/p)\hat{\sigma}_j(u,v)e^{-ij\omega}, \tag{3.10}$$

where $k(\cdot)$ is a kernel function satisfying Assumption 4, and $p$ is a bandwidth. Similarly, put

$$\hat{f}_0(\omega,u,v) \;=\; \frac{1}{2\pi}\hat{\sigma}_0(u,v), \quad \omega\in[-\pi,\pi]$$

where $\hat{\sigma}_0(\cdot,\cdot)$ is given by (3.6). Then, we consider the squared weighted norm of $\hat{f}_n$ minus $\hat{f}_0$; that is

$$\begin{aligned}
L_2^2\Big(\hat{f}_n(\omega,u,v),\hat{f}_0(\omega,u,v)\Big) &= \int_{-\pi}^{\pi}\|\hat{f}_n(\omega,u,v)-\hat{f}_0(\omega,u,v)\|^2_{\mathcal{W}}d\omega \\
&= \int_{\mathbb{R}^2}\int_{-\pi}^{\pi}|\hat{f}_n(\omega,u,v)-\hat{f}_0(\omega,u,v)|^2 d\omega d\mathcal{W} \\
&= \int_{\mathbb{R}^2}\int_{-\pi}^{\pi}\Big|\frac{1}{2\pi}\sum_{j=-(n-1)}^{n-1}(1-|j|/n)^{1/2}k(j/p)\hat{\sigma}_j(u,v)e^{-ij\omega} \\
&\quad -\frac{1}{2\pi}\hat{\sigma}_0(u,v)\Big|^2 d\omega d\mathcal{W} \\
&= \int_{\mathbb{R}^2}\int_{-\pi}^{\pi}\Big|\frac{1}{2\pi}\sum_{j=-(n-1)}^{-1}(1-|j|/n)^{1/2}k(j/p)\hat{\sigma}_j(u,v)e^{-ij\omega} \\
&\quad +\sum_{j=1}^{n-1}(1-j/n)^{1/2}k(j/p)\hat{\sigma}_j(u,v)e^{-ij\omega}+\frac{1}{2\pi}\hat{\sigma}_0(u,v) \\
&\quad -\frac{1}{2\pi}\hat{\sigma}_0(u,v)\Big|^2 d\omega d\mathcal{W} \\
&= \int_{\mathbb{R}^2}\int_{-\pi}^{\pi}\left|2\times\frac{1}{2\pi}\sum_{j=1}^{n-1}(1-j/n)^{1/2}k(j/p)\hat{\sigma}_j(u,v)e^{-ij\omega}\right|^2 d\omega d\mathcal{W} \\
&= \frac{2}{\pi}\int_{\mathbb{R}^2}\frac{1}{2\pi}\int_{-\pi}^{\pi}\left|\sum_{j=1}^{n-1}(1-j/n)^{1/2}k(j/p)\hat{\sigma}_j(u,v)e^{-ij\omega}\right|^2 d\omega d\mathcal{W},
\end{aligned}$$

where $d\mathcal{W}\equiv d\mathcal{W}(u,v)$. At this point, making use of the Parseval's identity we observe that

$$\frac{1}{2\pi}\int_{-\pi}^{\pi}|f(\omega)|^2\,d\omega=\sum_{j=1}^{n-1}|z_j|^2,$$

for $f(\omega)=\sum_{j=1}^{n-1}z_je^{-ij\omega}$ and $z_j=(1-j/n)^{1/2}k(j/p)\hat{\sigma}_j(u,v)$. Thus, we have that,

$$\begin{aligned}
L_2^2\Big(\hat{f}_n(\omega,u,v),\hat{f}_0(\omega,u,v)\Big) &= \frac{2}{\pi}\int_{\mathbb{R}^2}\sum_{j=1}^{n-1}\left|(1-j/n)^{1/2}k(j/p)\hat{\sigma}_j(u,v)\right|^2 d\mathcal{W} \\
&= \frac{2}{\pi}\int_{\mathbb{R}^2}\sum_{j=1}^{n-1}(1-j/n)k^2(j/p)\,|\hat{\sigma}_j(u,v)|^2\,d\mathcal{W}, \qquad (3.11)
\end{aligned}$$

for any suitable weighting function such that the above integral exists. In particular, for the choice of $\mathcal{W}(\cdot,\cdot)$ given by (3.3) we obtain that

$$L_2^2\Big(\hat{f}_n(\omega,u,v),\hat{f}_0(\omega,u,v)\Big) \;=\; \frac{2}{\pi}\sum_{j=1}^{n-1}(1-j/n)k^2(j/p)\widehat{V}_X^2(j).$$

This fact motivates our study of Box-Pierce type statistics based on auto-distance covariance function. Indeed, if $k(z)=1$ if $|z|\leq 1$ and 0 otherwise (i.e. in the case of uniform weighting), then the last expression becomes

$$L_2^2\Big(\hat{f}_n(\omega,u,v),\hat{f}_0(\omega,u,v)\Big) \;=\; \frac{2}{\pi}\sum_{j=1}^{p}(1-j/n)\widehat{V}_X^2(j). \tag{3.12}$$

Equation (3.12) can be viewed as a Box-Pierce type statistic for testing the hypotheses $V_X^2(j)=0$, $j=1,\ldots,p$, since the factor $(1-j/n)$ can be replaced by unity. It is interesting to observe that, by recalling (3.1) and letting

$$\rho_j(u,v) \;=\; \frac{\sigma_j(u,v)}{V_X(0)},$$

then working analogously we can obtain a test statistic for testing independence in terms of ADCF. Indeed, recall that $\sup_{(u,v)\in\mathbb{R}^2}\sum_j|\sigma_j(u,v)|<\infty$ under Assumption 3($i$), and define the Fourier transform of $\rho_j(u,v)$ by

$$g(\omega,u,v) \;=\; \frac{1}{2\pi}\sum_{j=-\infty}^{\infty}\rho_j(u,v)e^{-ij\omega}, \quad \omega\in[-\pi,\pi],$$

where under independence, it reduces to the constant

$$g_0(\omega,u,v) \;=\; \frac{1}{2\pi}\rho_0(u,v), \quad \omega\in[-\pi,\pi].$$

Analogous to (3.10), a kernel-density estimator of $g$ is given by

$$\hat{g}_n(\omega,u,v) \;=\; \frac{1}{2\pi}\sum_{j=-(n-1)}^{(n-1)}(1-|j|/n)^{1/2}k(j/p)\hat{\rho}_j(u,v)e^{-ij\omega},$$

where $\hat{\rho}_j(u,v)=\hat{\sigma}_j(u,v)/\widehat{V}_X(0)$ with $\hat{\sigma}_j(\cdot,\cdot)$ and $\widehat{V}_X(\cdot)$ defined by (3.6) and (3.8) respectively. Considering now the squared weighted norm between $\hat{g}_n(\omega,u,v)$ and $\hat{g}_0(\omega,u,v)=$

$\hat{\rho}_0(u,v)/2\pi$ , we get

$$
\begin{aligned}
L_2^2\Big(\hat{g}_n(\omega,u,v),\hat{g}_0(\omega,u,v)\Big) &= \int_{-\pi}^{\pi} \|\hat{g}_n(\omega,u,v)-\hat{g}_0(\omega,u,v)\|_{\mathcal{W}}^2 d\omega \\
&= \int_{\mathbb{R}^2}\int_{-\pi}^{\pi} |\hat{g}_n(\omega,u,v)-\hat{g}_0(\omega,u,v)|^2 d\omega d\mathcal{W} \\
&= \int_{\mathbb{R}^2}\int_{-\pi}^{\pi} \Big|\frac{1}{2\pi}\sum_{j=-(n-1)}^{n-1}(1-|j|/n)^{1/2}k(j/p)\hat{\rho}_j(u,v)e^{-ij\omega} \\
&\quad -\frac{1}{2\pi}\hat{\rho}_0(u,v)\Big|^2 d\omega d\mathcal{W} \\
&= \int_{\mathbb{R}^2}\int_{-\pi}^{\pi} \Big|\frac{1}{2\pi}\sum_{j=-(n-1)}^{-1}(1-|j|/n)^{1/2}k(j/p)\hat{\rho}_j(u,v)e^{-ij\omega} \\
&\quad +\sum_{j=1}^{n-1}(1-j/n)^{1/2}k(j/p)\hat{\rho}_j(u,v)e^{-ij\omega}\Big|^2 d\omega d\mathcal{W} \\
&= \int_{\mathbb{R}^2}\int_{-\pi}^{\pi} \left|2\times\frac{1}{2\pi}\sum_{j=1}^{n-1}(1-j/n)^{1/2}k(j/p)\hat{\rho}_j(u,v)e^{-ij\omega}\right|^2 d\omega d\mathcal{W} \\
&= \frac{2}{\pi}\int_{\mathbb{R}^2}\frac{1}{2\pi}\int_{-\pi}^{\pi}\left|\sum_{j=1}^{n-1}(1-j/n)^{1/2}k(j/p)\hat{\rho}_j(u,v)e^{-ij\omega}\right|^2 d\omega d\mathcal{W}.
\end{aligned}
$$

Using at this point the Parseval's identity as before and choosing the weighting function defined in (3.3), we obtain that

$$
\begin{aligned}
L_2^2\Big(\hat{g}_n(\omega,u,v),\hat{g}_0(\omega,u,v)\Big) &= \frac{2}{\pi}\int_{\mathbb{R}^2}\sum_{j=1}^{n-1}\left|(1-j/n)^{1/2}k(j/p)\hat{\rho}_j(u,v)\right|^2 d\mathcal{W} \\
&= \frac{2}{\pi}\int_{\mathbb{R}^2}\sum_{j=1}^{n-1}(1-j/n)k^2(j/p)\left|\hat{\rho}_j(u,v)\right|^2 d\mathcal{W} \\
&= \frac{2}{\pi}\sum_{j=1}^{n-1}(1-j/n)k^2(j/p)\frac{\int_{\mathbb{R}^2}|\hat{\sigma}_j(u,v)|^2 d\mathcal{W}}{\widehat{V}_X^2(0)} \\
&= \frac{2}{\pi}\sum_{j=1}^{n-1}(1-j/n)k^2(j/p)\frac{\widehat{V}_X^2(j)}{\widehat{V}_X^2(0)} \\
&= \frac{2}{\pi}\sum_{j=1}^{n-1}(1-j/n)k^2(j/p)\widehat{R}_X^2(j),
\end{aligned}
$$

where $\widehat{R}_X^2(\cdot)$ is given by (3.5). Therefore, in the case of uniform weighting function, the last expression becomes

$$\sum_{j=1}^{p}(1-j/n)\widehat{R}_X^2(j).$$

## 3.4 Main Results

In this section, we develop a test statistic for testing the hypotheses that the sequence $\{X_t, t=1,\ldots,n\}$ forms an i.i.d. sequence. The test statistic is motivated by (3.11) and is based on

$$T_n = \sum_{j=1}^{n-1}(n-j)k^2(j/p)\widehat{V}_X^2(j), \tag{3.13}$$

following Hong (1999). However, there is an important difference between the test statistic obtained by Hong (1999) and the one given by (3.13). The weight function chosen previously in (3.11) to form test statistics like (3.13) is assumed to be integrable. However, in our case we propose (3.13) by allowing nonintegrable weight functions $\mathcal{W}(\cdot,\cdot)$. We have the following results.

**Theorem 3.4.1** Suppose that Assumptions 2 and 4 hold and let $p = cn^\lambda$, where $c > 0$, $\lambda \in (0,1)$. Then, under the null hypothesis that $\{X_t\}$ is an i.i.d. sequence, we have that

$$M_n = \frac{T_n - \widehat{C}_0\sum_{j=1}^{n-1}k^2(j/p)}{\left[\widehat{D}_0\sum_{j=1}^{n-2}k^4(j/p)\right]^{1/2}} \to N(0,1),$$

as $n\to\infty$, in distribution, where

$$C_0 = \int_{\mathbb{R}^2}\sigma_0(u,-u)\sigma_0(v,-v)d\mathcal{W} = \left(\int_{\mathbb{R}}\frac{1-|\phi(u)|^2}{\pi|u|^2}du\right)^2 = \left[E\left|X_t - X_t'\right|\right]^2,$$

$$D_0 = 2\left(\int_{\mathbb{R}^2}|\sigma_0(u,u')|^2 d\mathcal{W}_0(u)d\mathcal{W}_0(u')\right)^2 = 2\left(\int_{\mathbb{R}^2}\frac{|\phi_0(u,u')-\phi(u)\phi(u')|^2}{\pi^2|u|^2|u'|^2}dudu'\right)^2 = 2V_X^4(0),$$

and $\widehat{C}_0$, $\widehat{D}_0$ are their sample counterparts and the expectation is taken with respect to the distribution of $X_t$ with $X_t'$ an independent copy of $X_t$.

The following proposition gives the asymptotic distribution of the statistic $T_n$ in the special case of a standard Gaussian process.

**Proposition 3.4.1** If $\{X_t\}$ is a Gaussian stationary time series such that $EX_t = 0$, $\text{Var}(X_t) = 1$ and $\rho(j) = \text{Cov}(X_t, X_{t-|j|})$ then under Assumption 4 and $p = cn^\lambda$ for $c > 0$ and $\lambda \in (0, 1)$,

$$M_n = \frac{T_n - 4/\pi \sum_{j=1}^{n-1} k^2(j/p)}{4\sqrt{2}/\pi(1 + \pi/3 - \sqrt{3})\left[\sum_{j=1}^{n-2} k^4(j/p)\right]^{1/2}} \to N(0, 1),$$

as $n \to \infty$, in distribution.

**Theorem 3.4.2** Suppose that Assumptions 1, 3($i$) and 4 hold and $p = cn^\lambda$ for $c > 0$ and $\lambda \in (0, 1)$. Then,

$$\frac{\sqrt{p}}{n} M_n \to \frac{\frac{\pi}{2} \int_{\mathbb{R}^2} \int_{-\pi}^{\pi} |f(\omega, u, v) - f_0(\omega, u, v)|^2 \, d\omega d\mathcal{W}(u, v)}{\left[D_0 \int_0^\infty k^4(z) dz\right]^{1/2}},$$

as $n \to \infty$, in probability.

The above theorem assures that under any alternative hypothesis, $M_n$ has asymptotic power 1 whenever the weighted squared norm of $f(\omega, u, v)$ minus $f_0(\omega, u, v)$ is positive. This is a consequence of the fact that

$$\int_{\mathbb{R}^2} \int_{-\pi}^{\pi} |f(\omega, u, v) - f_0(\omega, u, v)|^2 \, d\omega d\mathcal{W}(u, v) = \frac{2}{\pi} \sum_{j=1}^{\infty} V_X^2(j). \tag{3.14}$$

Clearly, (3.14) is equal to 0 if and only if $X_t$ and $X_{t-j}$ are independent for all $j \geq 1$. Therefore, the statistic $M_n$ is consistent against the hypothesis of pairwise dependence.

## 3.5 Applications

### 3.5.1 Investigating the Size of $T_n$

We first report some empirical results concerning the behavior of the test statistic $T_n$ given by (3.13). The simulations correspond to different sample sizes and we use standard nonparametric bootstrap (number of replications $b = 499$) to obtain critical values for studying the size and the power of the proposed statistic. The calculation of the test statistic is based on the use of R package **dCovTS** implemented in Chapter 5.

To examine the effects of using different kernel functions for constructing the test statistic $T_n$, we choose Lipschitz continuous functions, i.e. functions $k(\cdot)$ such that for any $z_1, z_2 \in \mathbb{R}$

$$|k(z_1) - k(z_2)| \leq C\,|z_1 - z_2|$$

for some constant $C$. In particular, we use the following

- The Daniell kernel (DAN), given by

$$k(z) = \frac{\sin(\pi z)}{\pi z}, \quad z \in \mathbb{R} - \{0\},$$

- The Parzen kernel (PAR), given by

$$k(z) = \begin{cases} 1 - 6(\pi z/6)^2 + 6\,|\pi z/6|^3, & |z| \leq 3/\pi, \\ 2(1 - |\pi z/6|)^3, & 3/\pi \leq |z| \leq 6/\pi, \\ 0, & \text{otherwise}, \end{cases}$$

- The Bartlett kernel (BAR), given by

$$k(z) = \begin{cases} 1 - |z|\,, & |z| \leq 1, \\ 0, & \text{otherwise}. \end{cases}$$

We also compare the test statistic $T_n$ with other test statistics, available in time series literature and reviewed in Section 2.4.2, to examine its relative performance. In particular, we consider the Box-Pierce (BP) test statistic

$$\mathrm{BP} = n \sum_{j=1}^{p} \hat{\rho}^2(j),$$

the Ljung-Box (LB) test statistic

$$\mathrm{LB} = n(n+2) \sum_{j=1}^{p} (n-j)^{-1} \hat{\rho}^2(j),$$

the test statistic proposed by Hong (1996)

$$T_n^{(1)} = n \sum_{j=1}^{n-1} k^2(j/p) \hat{\rho}^2(j).$$

Furthermore, we consider the test statistic obtained by Hong (1999)

$$T_n^{(2)} = \int_{\mathbb{R}^2} \sum_{j=1}^{n-1} (1 - j/n) k^2(j/p) \left| \hat{\sigma}_j(u,v) \right|^2 d\mathcal{W}(u,v),$$

where $\mathcal{W}(\cdot,\cdot)$ is an arbitrary integrable weight function. Note that, a convenient way to calculate $T_n^{(2)}$, is to employ the bivariate cumulative distribution function of a standard normal random variable (Chen and Hong, 2012); i.e. $\mathcal{W}(u,v) = \mathcal{W}_0(u)\mathcal{W}_0(v) = \Phi(u)\Phi(v)$. This allows us to consider a countable number $N$ of grid points $(u,v)$ for which the integral in equation (3.11), is replaced by its empirical mean. The number $N$ is chosen to be 500, because a larger choice of $N$ would not alter the results significantly. Table 3.1 shows the computational time taken to obtain $T_n$ and $T_n^{(2)}$, for various choices of sample size $n$ and bandwidth $p$ on a standard laptop with Intel Core i5 system and CPU 2.30 GHz. Clearly, $T_n^{(2)}$ is computationally more expensive than $T_n$, especially when $n$ and $p$ are large. The computational time of calculating $T_n$ is still quite high because $T_n$ is based on $\widehat{V}_X^2(j)$. The sequence $\{\widehat{V}_X^2(j), j = 1, \cdots, n\}$ is computed by employing a distance matrix among observations for each lag $j$. The computation therefore is quite demanding; some recent progress towards this issue was reported by Huo and Székely (2015).

Table 3.1: Computational time (in seconds) for computing the test statistics $T_n$ and $T_n^{(2)}$.

| $n$: | | 100 | | | 200 | | | 500 | | |
|---|---|---|---|---|---|---|---|---|---|---|
| $p$: | | 3 | 7 | 16 | 3 | 9 | 25 | 4 | 13 | 42 |
| $T_n$ | **BAR** | 0.13 | 0.14 | 0.18 | 1.20 | 1.12 | 1.12 | 18.39 | 18.56 | 17.94 |
| | **PAR** | 0.13 | 0.12 | 0.11 | 0.83 | 0.79 | 0.78 | 18.19 | 17.97 | 18.69 |
| | **DAN** | 0.11 | 0.12 | 0.11 | 0.80 | 0.80 | 0.78 | 18.93 | 18.48 | 18.47 |
| $T_n^{(2)}$ | **BAR** | 1.20 | 1.21 | 1.20 | 4.79 | 4.74 | 4.72 | 29.61 | 29.58 | 29.58 |
| | **PAR** | 1.21 | 1.22 | 1.22 | 4.82 | 4.79 | 4.77 | 29.62 | 29.58 | 29.69 |
| | **DAN** | 1.21 | 1.24 | 1.24 | 4.82 | 4.77 | 4.81 | 29.61 | 29.60 | 29.61 |

We now investigate the size of the test. Suppose that $\{X_t\}$ is an i.i.d. sequence of standard normal random variables. To examine the sensitivity of the test statistic $T_n$ on the values of bandwidth $p$, we use $p = n^{\lambda}$ with $\lambda = 1/5, 2/5, 3/5$. If $n = 100$, then $p$ takes approximately the values 3, 7 and 16. Similarly for other sample sizes. Table 3.2 contains achieved type I error rates at 5% and 10% nominal levels. We note that the proposed test statistic keeps its size closer to its nominal level. In fact the Bartlett kernel yields better approximations. Further support for the asymptotic normality of the proposed test statistic is given in Table 3.3. These results show that the asserted asymptotic normality is adequate especially for large sample sizes. Moreover, Figure 3.1 shows box plots and qq-plots for the sampling distribution of the standardized statistic $M_n$, verifying the adequacy of the normal distribution.

Table 3.2: Achieved type I error rates of the test statistics for testing the hypothesis that the data are i.i.d. The data are generated by the standard normal distribution. Achieved significance levels are given in percentages. The results are based on $b = 499$ bootstrap replications and 100 simulations.

| $n$: | | 100 | | | | | | 200 | | | | | | 500 | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $p$: | | 3 | | 7 | | 16 | | 3 | | 9 | | 25 | | 4 | | 13 | | 42 | |
| | | 10% | 5% | 10% | 5% | 10% | 5% | 10% | 5% | 10% | 5% | 10% | 5% | 10% | 5% | 10% | 5% | 10% | 5% |
| $T_n$ | **BAR** | 6 | 2 | 10 | 6 | 11 | 5 | 11 | 7 | 6 | 3 | 9 | 3 | 13 | 4 | 7 | 2 | 7 | 2 |
| | **PAR** | 12 | 6 | 16 | 6 | 4 | 1 | 10 | 5 | 6 | 3 | 7 | 2 | 10 | 5 | 14 | 9 | 11 | 5 |
| | **DAN** | 9 | 3 | 10 | 4 | 7 | 4 | 9 | 5 | 8 | 6 | 10 | 4 | 15 | 9 | 13 | 5 | 8 | 4 |
| BP | | 9 | 5 | 10 | 6 | 11 | 9 | 12 | 5 | 6 | 2 | 6 | 4 | 4 | 4 | 8 | 2 | 8 | 6 |
| LB | | 9 | 3 | 13 | 5 | 10 | 8 | 10 | 6 | 6 | 2 | 6 | 4 | 4 | 4 | 8 | 2 | 7 | 6 |
| $T_n^{(1)}$ | **BAR** | 6 | 2 | 4 | 1 | 11 | 5 | 12 | 8 | 10 | 3 | 9 | 6 | 10 | 4 | 13 | 5 | 9 | 5 |
| | **PAR** | 6 | 2 | 9 | 2 | 13 | 6 | 12 | 7 | 14 | 11 | 6 | 2 | 10 | 5 | 4 | 1 | 4 | 1 |
| | **DAN** | 8 | 4 | 15 | 6 | 19 | 8 | 8 | 5 | 15 | 8 | 16 | 6 | 13 | 5 | 9 | 5 | 10 | 4 |
| $T_n^{(2)}$ | **BAR** | 12 | 6 | 8 | 2 | 12 | 5 | 11 | 4 | 11 | 4 | 10 | 6 | 9 | 3 | 6 | 4 | 6 | 4 |
| | **PAR** | 12 | 3 | 13 | 6 | 8 | 5 | 10 | 3 | 7 | 4 | 9 | 2 | 9 | 5 | 10 | 6 | 8 | 2 |
| | **DAN** | 14 | 6 | 8 | 3 | 7 | 1 | 13 | 5 | 12 | 5 | 10 | 5 | 15 | 9 | 5 | 2 | 7 | 1 |

Table 3.3: Skewness, kurtosis and $p$-values obtained by performing a one-sample Kolmogorov-Smirnov test, for testing normality of the normalized test statistic $M_n$ given by Theorem 3.4.1. The results are based on $b = 499$ bootstrap replications and 100 simulations.

| | Skewness | | | Kurtosis | | | $p$-value | | |
|---|---|---|---|---|---|---|---|---|---|
| | BAR | PAR | DAN | BAR | PAR | DAN | BAR | PAR | DAN |
| $n = 100$ | | | | | | | | | |
| $p = 3$ | 0.741 | -0.032 | 0.2638 | 4.701 | 3.150 | 2.5546 | 0.627 | 0.740 | 0.9286 |
| $p = 7$ | 0.203 | 0.479 | 0.0855 | 2.478 | 3.832 | 2.8492 | 0.739 | 0.756 | 0.8669 |
| $p = 16$ | 0.178 | 0.404 | 0.5142 | 2.951 | 2.825 | 3.2691 | 0.982 | 0.949 | 0.3336 |
| $n = 200$ | | | | | | | | | |
| $p = 3$ | -0.169 | -0.171 | 0.2313 | 2.827 | 2.826 | 3.2557 | 0.977 | 0.989 | 0.9517 |
| $p = 9$ | 0.198 | 0.187 | 0.6011 | 2.320 | 2.301 | 3.2171 | 0.509 | 0.508 | 0.3376 |
| $p = 25$ | 0.338 | 0.335 | 0.4768 | 3.002 | 2.989 | 3.4548 | 0.905 | 0.882 | 0.4091 |
| $n = 500$ | | | | | | | | | |
| $p = 4$ | 0.490 | -0.097 | 0.113 | 2.764 | 2.422 | 2.768 | 0.338 | 0.814 | 0.9508 |
| $p = 13$ | 0.004 | 0.046 | 0.190 | 2.727 | 2.404 | 2.637 | 0.686 | 0.564 | 0.9344 |
| $p = 42$ | -0.016 | 0.312 | -0.003 | 2.680 | 2.710 | 2.660 | 0.460 | 0.980 | 0.9367 |

(a) (b)

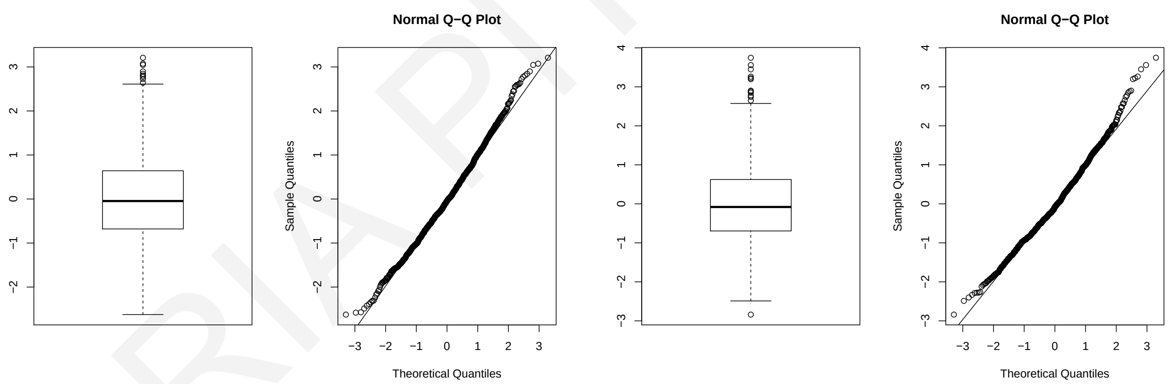



Figure 3.1: Box plots and qq-plots for the sampling distribution of the normalized statistic $M_n$. The results are based on (a) 500 and (b) 1000 observations, for $p = [n^{3/5}]$ and 1000 simulations. The statistic $M_n$ was calculated based on the Bartlett kernel.



### 3.5.2 Comparison Between ADCF and ACF

As mentioned at the beginning of this chapter, compared to the ACF where it measures the strength of linear dependencies and can be equal to zero even when the variables are related, ADCF vanishes only in the case where the observations are independent. We compare the performance of ADCF and ACF by considering the following alternative models:

- ARCH(2)-model

$$X_t = \sigma_t \epsilon_t, \quad \sigma_t^2 = 0.5 + 0.8X_{t-1}^2 + 0.1X_{t-2}^2 \tag{3.15}$$

- TAR(1)-model

$$X_t = \begin{cases} -1.5X_{t-1} + \epsilon_t, & X_{t-1} < 0; \\ 0.5X_{t-1} + \epsilon_t, & X_{t-1} \geq 0 \end{cases} \tag{3.16}$$

- NMA(2)-model

$$X_t = \epsilon_t \epsilon_{t-1} \epsilon_{t-2}, \tag{3.17}$$

where $\{\epsilon_t\}$ is a sequence of i.i.d. standard normal random variables. Note that (3.15) corresponds to an autoregressive conditional heteroscedastic model of order two (see Engle (1982)) and (3.16) corresponds to a threshold autoregressive model (TAR) of order one (see Tsay (2005, Section 4.1.2) for instance). The TAR model generates data with nonlinear dependence structure (Tong, 1990). Model (3.17) is an example of nonlinear moving average of order two. It is well known that the process $\{X_t\}$ generated by (3.17) consists of a sequence of 2-dependent but uncorrelated random variables.

Figure (3.2) compares the sample ACF and sample ADCF of models (3.15) – (3.17) with sample size $n = 2000$. In all these three cases, ADCF performs better than ACF, and correctly reflects the underlying nonlinear dependence structure of the aforementioned models. In particular, the sample ADCF for model (3.15) (Figure 3.2a) exhibits nice exponential decay strongly suggesting dependence among the data. Moreover, the ADCF for models (3.16) and (3.17) (Figures 3.2b and 3.2c) cuts off at lag 1 and lag 2 respectively, whereas the corresponding ACF plots fail to show any serial correlation among the observations. The shown critical values (blue dotted horizontal line) of the ADCF plots (second column) are the pairwise 95% critical values that correspond to the pairwise independence test. They are computed via the subsampling approach suggested by Zhou (2012, Section

5.1), where the choice of the block size is based on the minimum volatility method proposed by Politis et al. (1999, Section 9.4.2). Additionally, the 95% simultaneous critical values are also shown in the sample ADCF plots (last column). They are computed via the independent wild bootstrap approach (Dehling and Mikosch, 1994; Shao, 2010; Leucht and Neumann, 2013b) explained in detail in the appropriate section of Chapter 4.

### 3.5.3 Investigating the Power of $T_n$

For investigating the power of the test statistic $T_n$ we consider the data generating processes given by (3.15), (3.16) and (3.17). Figure 3.3 (respectively Figure 3.4) shows the power of all test statistics considered for various sample sizes and bandwidth parameters when the data are generated by (3.15) (respectively (3.16)). We note that in both cases $T_n$ and $T_n^{(2)}$ perform better than all the other test statistics in the sense that they achieve the maximum power. For bandwidth values of the form $n^{1/5}$ and $n^{2/5}$ the power of both test statistics increases to one, especially for large sample sizes. When $p = n^{3/5}$ then we note that, for the case of model (3.15), the power of $T_n^{(2)}$ is superior to the power of $T_n$. However, the simulation suggests that as the sample size tends to larger values, the power of $T_n$ approaches the power of $T_n^{(2)}$. The situation is reversed in Figure 3.4, but the fact that both tests give similar results is clearly depicted, especially for large values of sample size.

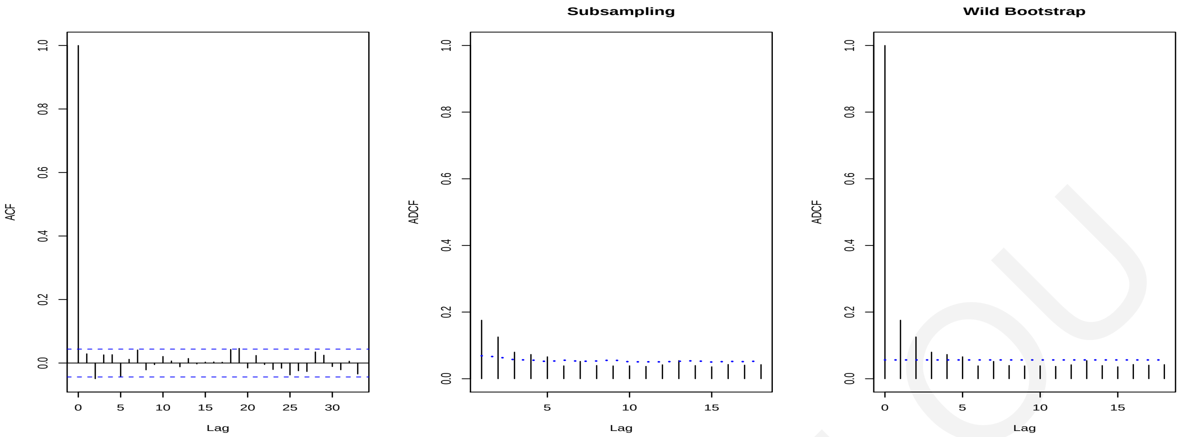
Figure 3.5 shows bootstrap $p$-values for various values of the bandwidth parameter and sample sizes when the data are generated by (3.17). We note again that the performance of both test statistics $T_n$ and $T_n^{(2)}$ is superior to the performance of the rest test statistics.



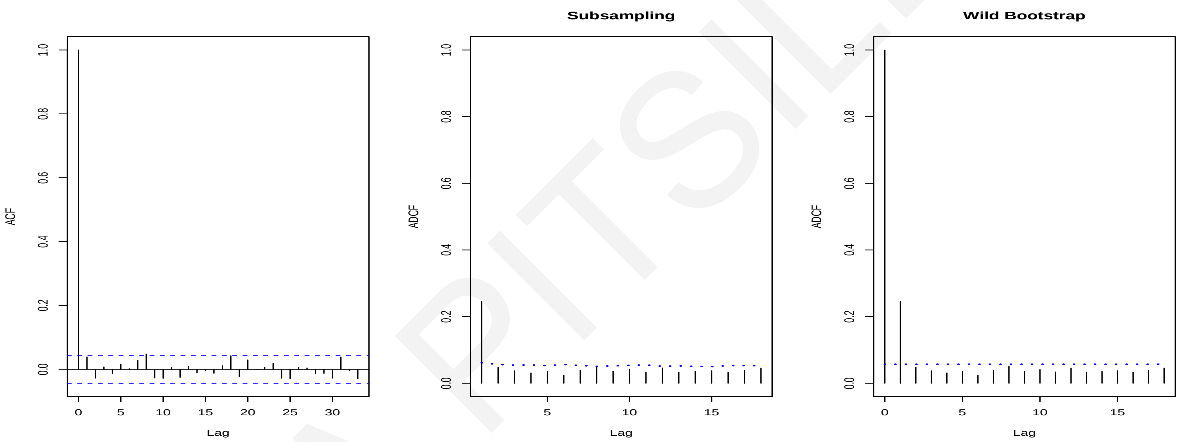
### 3.5.4 S&P 500 Stock Return Data

Typically, the stock return data do not have a constant variance, but highly volatile periods tend to be clustered. That is, there is a change in volatility over time. So, models that study this change in volatility need to be considered. It is well known that ARMA models assume constant variance. However, ARCH - and the generalized ARCH - models,

(a)




(b)




(c)

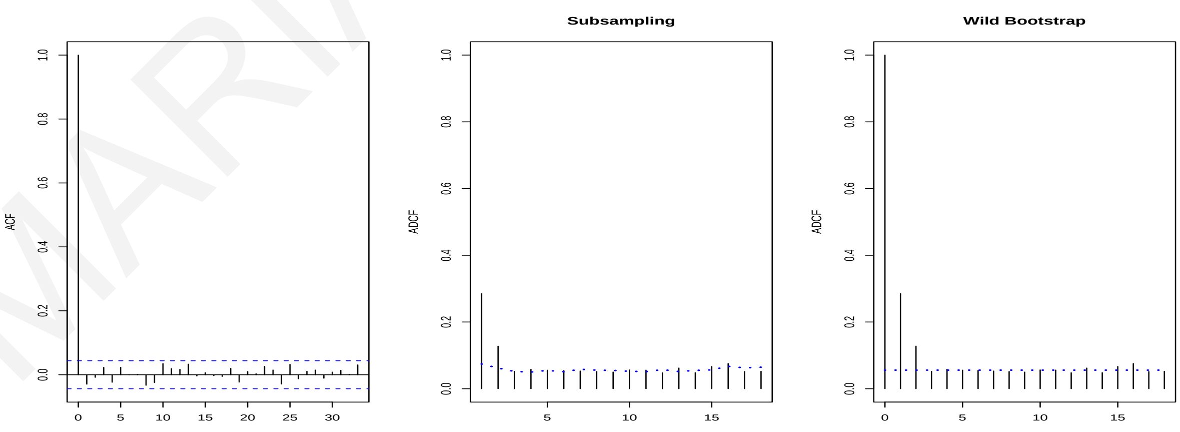



Figure 3.2: Comparison of the sample ADCF and sample ACF. Results are based on sample size $n = 2000$. (a) Data are generated by the ARCH(2) model given by (3.15). (b) Data are generated by the TAR(1) model given by (3.16). (c) Data are generated by the NMA(2) model given by (3.17).

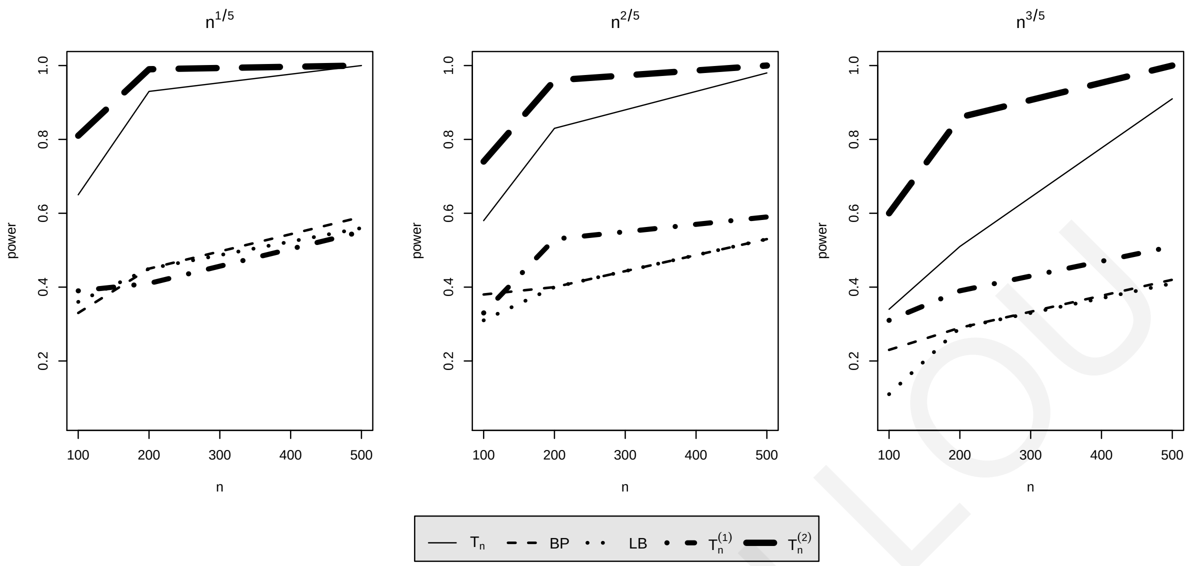



Figure 3.3: Achieved power of all test statistics. The data are generated by the ARCH(2) model given by (3.15). The results are based on $b = 499$ bootstrap replications and 100 simulations. The test statistics $T_n$, $T_n^{(1)}$, $T_n^{(2)}$ are calculated by employing the Daniell kernel.

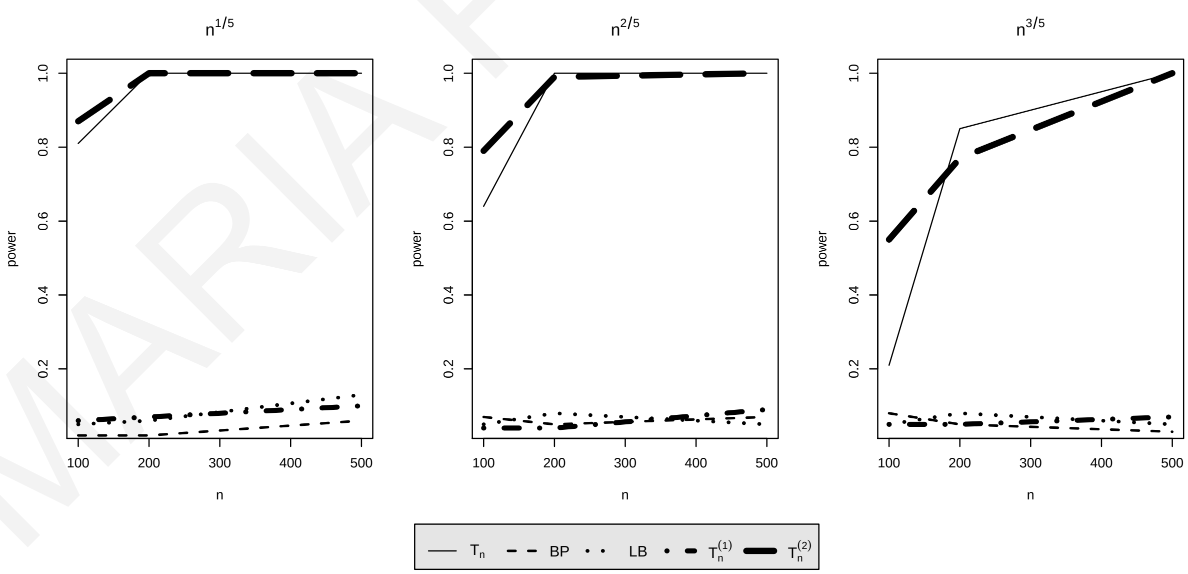



Figure 3.4: Achieved power of all test statistics. The data are generated by the TAR(1) model given by (3.16). The results are based on $b = 499$ bootstrap replications and 100 simulations. The test statistics $T_n$, $T_n^{(1)}$, $T_n^{(2)}$ are calculated by employing the Daniell kernel.

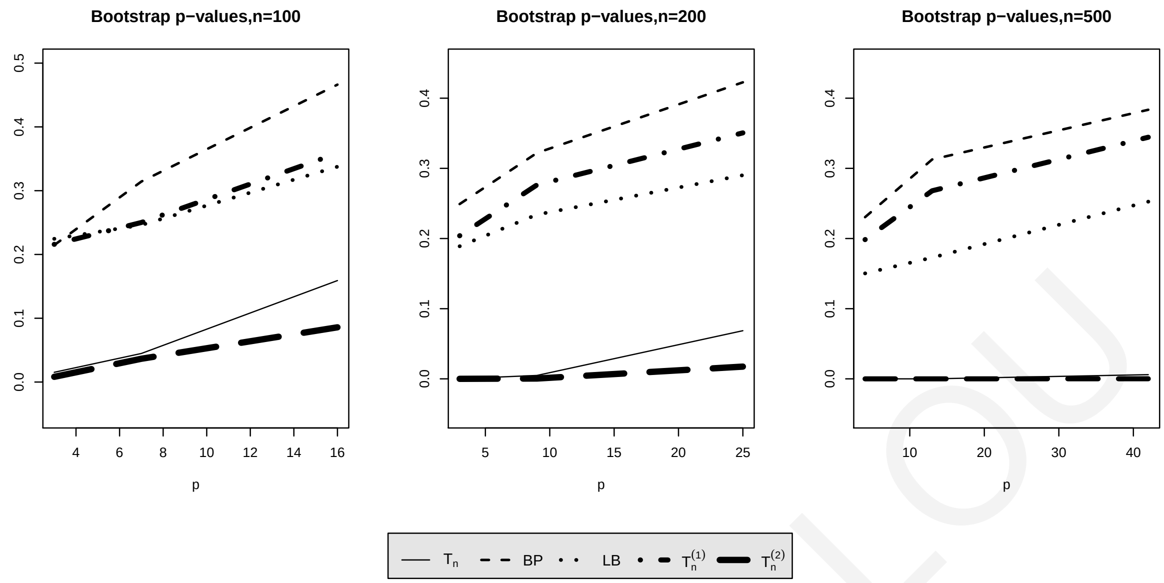



Figure 3.5: Bootstrap $p$-values of all test statistics as a function of the bandwidth $p$. The data are generated by the NMA(2) model given by (3.17). The results are based on $b = 499$ bootstrap replications and 100 simulations. The test statistics $T_n$, $T_n^{(1)}$, $T_n^{(2)}$ are calculated by employing the Daniell kernel.

are developed to model changes in volatility. We analyze monthly excess returns of the S&P 500 index starting from 1926. This series consists of 792 observations (Tsay, 2005, Example 3.3). Figure 3.6 shows the ACF (upper plot) and the ADCF (lower plots) of the original series and the squared original series. The ACF plot of the original series suggests a moderate serial correlation at lags 1 and 3, while the ACF plot of the squared series shows strong linear dependence. This is a common feature in financial returns. However, the ADCF plots strongly suggest dependence, especially when considering the shown critical values that correspond to the independence test. Tsay (2005) suggested an AR(3)-GARCH(1,1) model for the series. However, it is further observed that all autoregressive parameters are insignificant at the 5% significance level. Hence, a GARCH(1,1) model is fitted to these data. After data fitting it is of interest to study the behavior of the standardized residuals. The upper panel of Figure 3.7 show the ACF plots of the standardized residuals and the squared standardized residuals of the fitted model. These plots, fail to show any signal of serial correlation. On the other hand, their ADCF plots (lower plots of Figure 3.7) indicate that there is dependence among the residuals. Table 3.4 contains the $p$-values obtained by constructing tests of independence among the residuals for various choices of the bandwidth $p$. The BP, LB and $T_n^{(1)}$ test statistics yield large $p$-values suggesting no serial correlation between the residuals, whereas $T_n^{(2)}$ and $T_n$ give low $p$-values suggesting dependence among the observations. All results are calculated by using $b = 199$ and 499. The test statistics $T_n$, $T_n^{(1)}$ and $T_n^{(2)}$ are calculated by employing the Bartlett kernel. Any other choice of the kernel function, $k(\cdot)$, results to the same conclusions.

Table 3.4: P-values obtained by constructing tests of independence among the standardized residuals after fitting the GARCH(1,1) model to $S\&P$ 500 returns with Gaussian and standard normal inverse Gaussian innovations. All test statistics were calculated for both $b = 199$ and $b = 499$ bootstrap replications and by employing the Bartlett kernel.

| Bandwidth | Replications | Gaussian Model | | | | | NIG Model | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $p$ | $b$ | $T_n$ | $T_n^{(2)}$ | BP | LB | $T_n^{(1)}$ | $T_n$ | $T_n^{(2)}$ | BP | LB | $T_n^{(1)}$ |
| 4 | 199 | 0.020 | 0.015 | 0.640 | 0.695 | 0.670 | 0.650 | 0.380 | 0.890 | 0.880 | 0.875 |
| | 499 | 0.036 | 0.004 | 0.672 | 0.686 | 0.638 | 0.660 | 0.350 | 0.910 | 0.864 | 0.890 |
| 15 | 199 | 0.030 | 0.005 | 0.260 | 0.255 | 0.370 | 0.420 | 0.065 | 0.725 | 0.500 | 0.525 |
| | 499 | 0.016 | 0.004 | 0.242 | 0.280 | 0.388 | 0.404 | 0.174 | 0.730 | 0.498 | 0.538 |
| 55 | 199 | 0.030 | 0.005 | 0.580 | 0.665 | 0.305 | 0.410 | 0.210 | 0.625 | 0.815 | 0.795 |
| | 499 | 0.028 | 0.026 | 0.652 | 0.632 | 0.336 | 0.412 | 0.196 | 0.626 | 0.786 | 0.766 |

Therefore, ignoring dependence will lead to unreliable inference and forecasts. Analyzing in depth these data is outside the scope of this chapter. However, a model with heavy tailed innovations might be a better approximation for these data. Indeed, a GARCH(1,1) model with innovations following a standard normal-inverse gaussian distribution (NIG) seems to be more appropriate choice for the data. Figure 3.8 shows the ACF plots (upper panel) and the ADCF plots (lower panel) of the standardized residuals and the squared standardized residuals of the fitted model with NIG innovations. Both ACF and ADCF plots do not indicate any serial dependence among the residuals. Furthermore, all test statistics yield large $p$-values indicating the absence of dependence among observations–see Table 3.4.

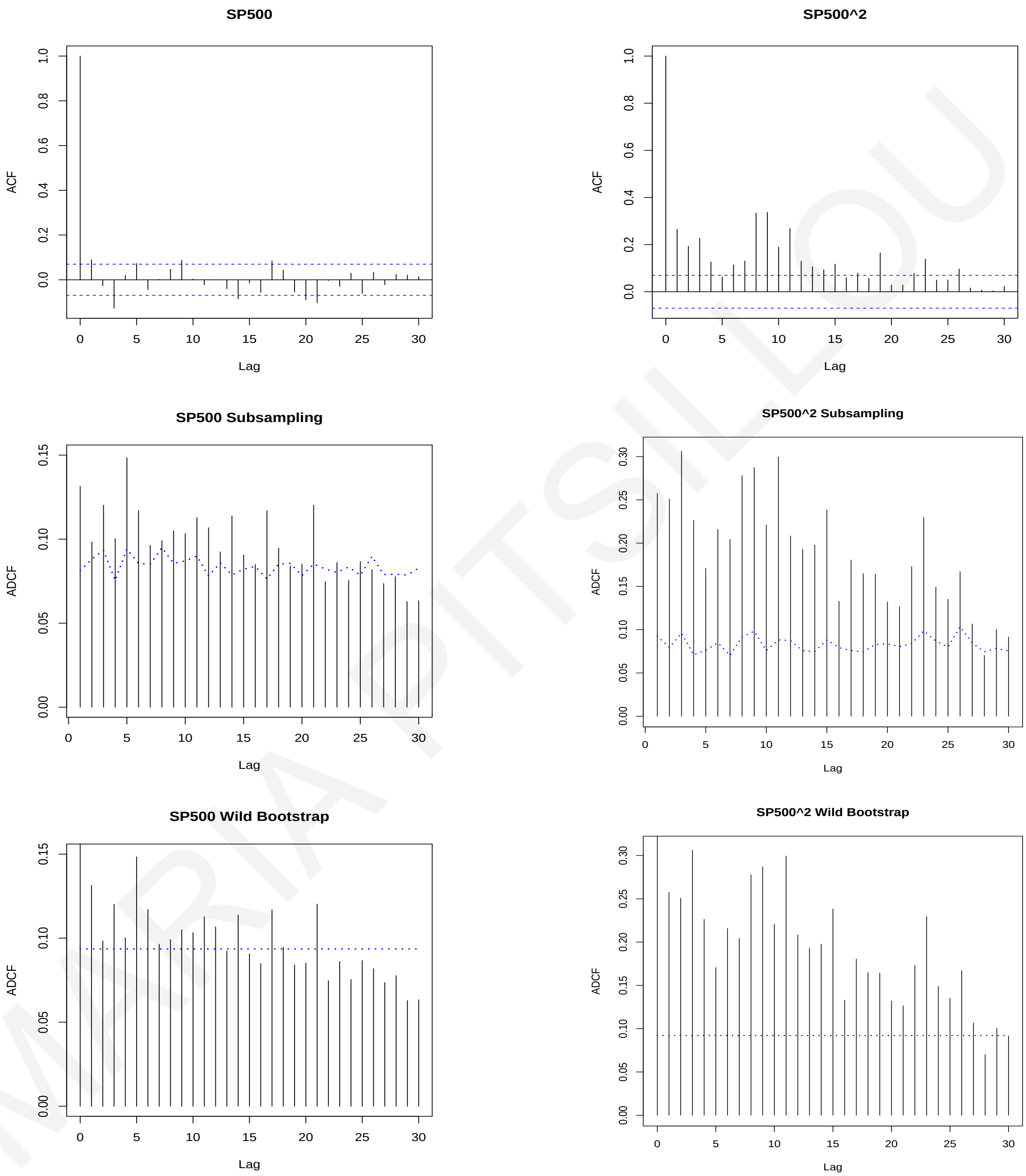



Figure 3.6: Upper plots: The ACF of the original series and the squared original series. Lower plots: The ADCF of the original and the squared original series.

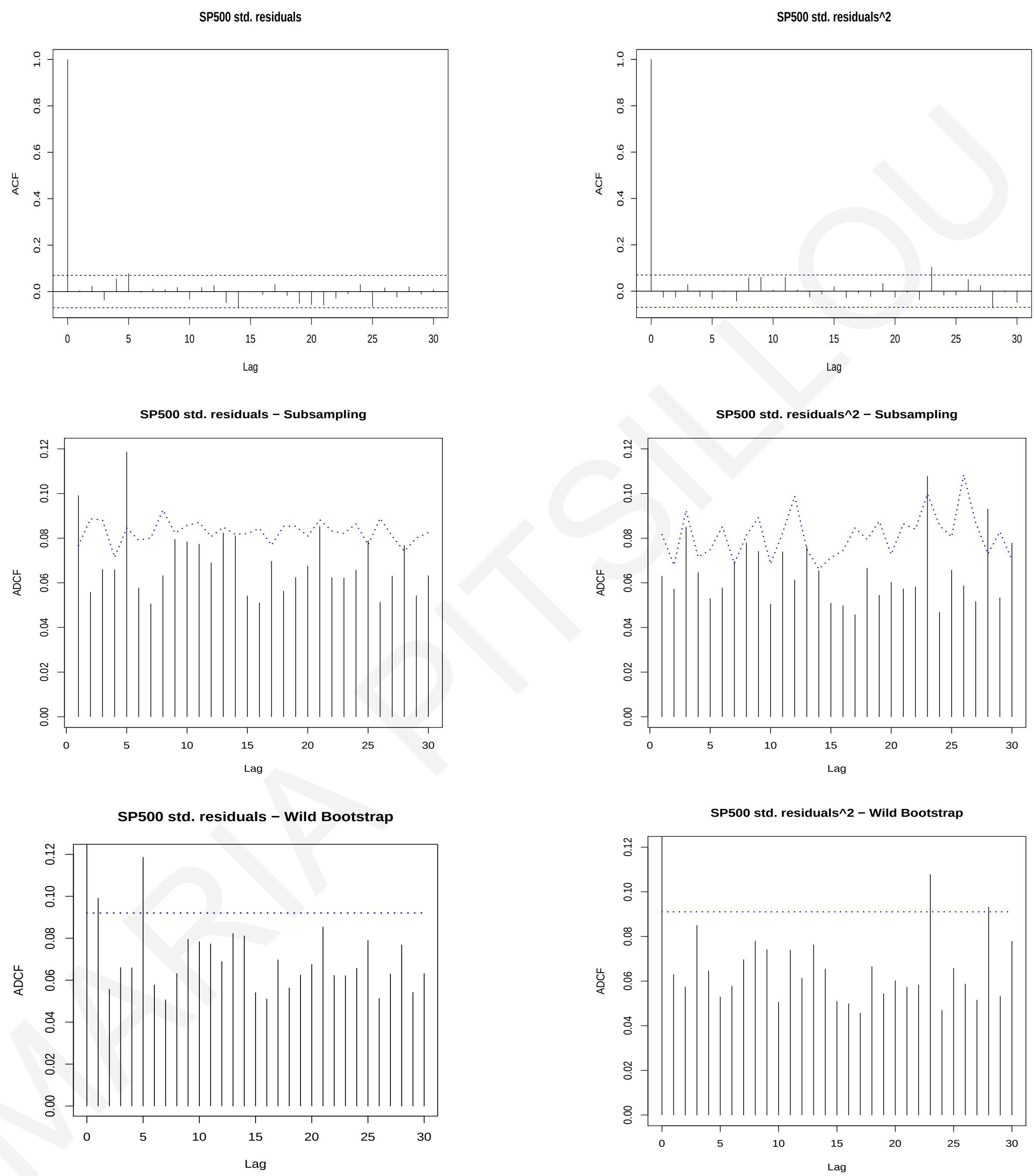



Figure 3.7: Upper plots: The ACF of the standardized residuals and the squared standardized residuals. Lower plots: The ADCF of the standardized residuals and squared standardized residuals. The results are obtained after fitting the GARCH(1,1) model with Gaussian innovations.

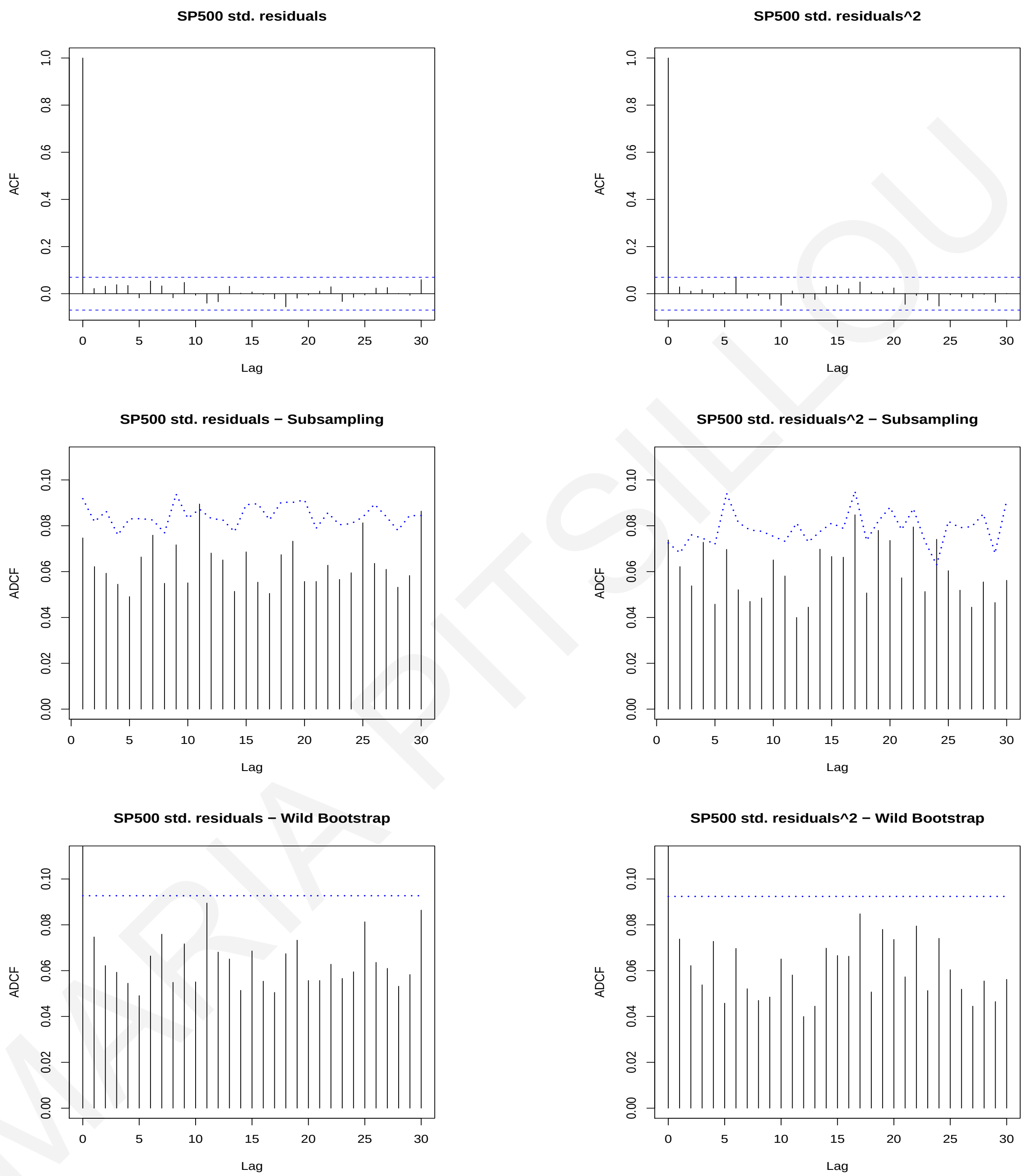



Figure 3.8: Upper plots: The ACF of the standardized residuals and the squared standardized residuals. Lower plots: The ADCF of the standardized residuals and squared standardized residuals. The results are obtained after fitting the GARCH(1,1) model with innovations following the standard NIG distribution.

# Appendix – Proofs

In this section, we give proofs of the main results discussed in this chapter. We first prove Proposition 3.2.1.

**Proof of Proposition 3.2.1** (i) It is well known that if $\{X_t\}$ follows a standard normal distribution with autocorrelation function $\rho(\cdot)$, then the joint characteristic function of the pair $(X_t, X_{t-|j|})$ and the corresponding marginal characteristic functions are given by

$$\phi_{|j|}(u,v) = \exp\Big(-(u^2+v^2)/2 - \rho(j)uv\Big),$$

$$\phi(u) = \exp\Big(-u^2/2\Big),$$

and

$$\phi(v) = \exp\Big(-v^2/2\Big)$$

respectively. Define the function $F(\rho(j))$ such that $V_X^2(j) = F(\rho(j))/\pi^2$, that is

$$\begin{aligned} F(\rho(j)) &= \int_{\mathbb{R}^2} \left|\phi_{|j|}(u,v) - \phi(u)\phi(v)\right|^2 \frac{du}{u^2}\frac{dv}{v^2} \\ &= \int_{\mathbb{R}^2} \left|e^{-(u^2+v^2)/2-\rho(j)uv} - e^{-u^2/2}e^{-v^2/2}\right|^2 \frac{du}{u^2}\frac{dv}{v^2} \\ &= \int_{\mathbb{R}^2} \left|e^{-(u^2+v^2)/2-\rho(j)uv} - e^{-(u^2+v^2)/2}\right|^2 \frac{du}{u^2}\frac{dv}{v^2} \\ &= \int_{\mathbb{R}^2} e^{-(u^2+v^2)}\big(e^{-\rho(j)uv} - 1\big)^2 \frac{du}{u^2}\frac{dv}{v^2} \\ &= \int_{\mathbb{R}^2} e^{-(u^2+v^2)}\big(1 - 2e^{-\rho(j)uv} + e^{-2\rho(j)uv}\big) \frac{du}{u^2}\frac{dv}{v^2}. \end{aligned} \tag{3.18}$$

Considering now the Taylor series of the exponential function we have the following

$$e^{-\rho(j)uv} = 1 - \rho(j)uv + \sum_{k=2}^{\infty} \frac{(-\rho(j)uv)^k}{k!},$$

and

$$e^{-2\rho(j)uv} = 1 - 2\rho(j)uv + \sum_{k=2}^{\infty} 2^k \frac{(-\rho(j)uv)^k}{k!}.$$

So,

$$
\begin{aligned}
(3.18) \quad &= \int_{\mathbb{R}^2} e^{-(u^2+v^2)} \sum_{k=2}^{\infty} \frac{2^k-2}{k!}(-\rho(j)uv)^k \frac{du}{u^2}\frac{dv}{v^2} \\
&= \int_{\mathbb{R}^2} e^{-(u^2+v^2)} \sum_{n=1}^{\infty} \frac{2^{2n}-2}{(2n)!}(-\rho(j)uv)^{2n} \frac{du}{u^2}\frac{dv}{v^2} \\
&= \rho^2(j)\left[\int_{\mathbb{R}^2} e^{-(u^2+v^2)} \sum_{n=1}^{\infty} \frac{2^{2n}-2}{(2n)!}(\rho(j)uv)^{2(n-1)} dudv\right] \\
&= \rho^2(j)\left[\sum_{n=1}^{\infty} \frac{2^{2n}-2}{(2n)!}\rho(j)^{2(n-1)} \int_{\mathbb{R}^2} e^{-(u^2+v^2)}(uv)^{2(n-1)} dudv\right] \\
&= \rho^2(j)G(\rho(j)).
\end{aligned}
$$

Clearly, $G(\rho(j))$ is a function consisting of positive terms, is nondecreasing in $\rho(j)$ and $G(\rho(j)) \leq G(1)$. Thus,

$$
R_X^2(j) \;=\; \frac{V_X^2(j)}{V_X^2(0)} = \frac{F(\rho(j))/\pi^2}{F(1)/\pi^2} = \rho^2(j)\frac{G(\rho(j))}{G(1)} \leq \rho^2(j).
$$

So, $R_X(j) \leq |\rho(j)|$.

(ii) We first note that $F(0) = F'(0) = 0$, thus $F(\rho(j)) = \int_0^{\rho(j)} \int_0^x F''(z)dzdx$, where the second derivative of $F(z)$ is

$$
\begin{aligned}
F''(z) \;&=\; \frac{d^2}{dz^2}\int_{\mathbb{R}^2} e^{-(u^2+v^2)}\left(1 - 2e^{-zuv} + e^{-2zuv}\right)\frac{du}{u^2}\frac{dv}{v^2} \\
&= \int_{\mathbb{R}^2} e^{-(u^2+v^2)}\left(4e^{-2zuv} - 2e^{-zuv}\right)dudv \\
&= 4V(z) - 2V\left(\frac{z}{2}\right),
\end{aligned}
$$

where

$$
V(z) = \int_{\mathbb{R}^2} e^{-u^2-v^2-2zuv} dudv.
$$

Using the results $\int_{\mathbb{R}} e^{-at^2} e^{-2bt} dt = e^{b^2/a}\sqrt{\pi/a}$ and $\int_{\mathbb{R}} e^{-at^2} dt = \sqrt{\pi/a}$, we observe that

$$\begin{aligned} V(z) &= \int_{\mathbb{R}} e^{-v^2} \int_{\mathbb{R}} e^{-u^2} e^{-2zvu} du dv \\ &= \int_{\mathbb{R}} e^{-v^2} e^{z^2 v^2} \sqrt{\pi} dv \\ &= \sqrt{\pi} \int_{\mathbb{R}} e^{-v^2(1-z^2)} dv \\ &= \sqrt{\pi}\sqrt{\frac{\pi}{1-z^2}} = \frac{\pi}{\sqrt{1-z^2}}. \end{aligned}$$

Therefore, we have the following result for the function $F(\rho(j))$:

$$\begin{aligned} F(\rho(j)) &= \int_0^{\rho(j)} \int_0^x \left( \frac{4\pi}{\sqrt{1-z^2}} - \frac{2\pi}{\sqrt{1-z^2/4}} \right) dz dx \\ &= 4\pi \int_0^{\rho(j)} \big(\arcsin(x) - \arcsin(x/2)\big) dx \\ &= 4\pi \Big[ x\arcsin(x) + \sqrt{1-x^2} - x\arcsin(x/2) - \sqrt{4-x^2} \Big]_0^{\rho(j)} \\ &= 4\pi \Big( \rho(j)\arcsin\rho(j) + \sqrt{1-\rho^2(j)} - \rho(j)\arcsin(\rho(j)/2) - \sqrt{4-\rho^2(j)} + 1 \Big). \end{aligned}$$

Finally, using the above result we get the required relation

$$R_X^2(j) = \frac{F(\rho(j))/\pi^2}{F(1)/\pi^2} = \frac{\rho(j)\arcsin\rho(j) + \sqrt{1-\rho^2(j)} - \rho(j)\arcsin(\rho(j)/2) - \sqrt{4-\rho^2(j)} + 1}{1 + \pi/3 - \sqrt{3}},$$

for $j = 0, \pm 1, \pm 2, \ldots$. □

We next prove two lemmas which will be employed in the proof of the remaining results of the chapter. In what follows $C$, possibly with a subscript, denotes a generic constant. We first define the pseudoestimator:

$$\bar{f}_n(\omega, u, v) = \frac{1}{2\pi} \sum_{j=-(n-1)}^{(n-1)} k(j/p)(1 - |j|/n)^{1/2} \tilde{\sigma}_j(u, v) e^{-ij\omega},$$

where

$$\tilde{\sigma}_j(u,v) = \frac{1}{n-|j|}\sum_{t=|j|+1}^{n}\psi_t(u)\psi_{t-|j|}(v) \tag{3.19}$$

and

$$\psi_t(u) \equiv e^{iuX_t} - \phi(u). \tag{3.20}$$

**Lemma 3.5.1** Suppose that $\{X_t, t \geq 1\}$ satisfies Assumptions 1 and 3$(ii)$. Then,

$$(n-|j|)^2 E\left|\hat{\sigma}_j(u,v) - \tilde{\sigma}_j(u,v)\right|^2 \leq C,$$

and

$$(n-|j|)E\left|\tilde{\sigma}_j(u,v)\right|^2 \leq C,$$

uniformly in $(u,v) \in \mathbb{R}^2$. The result of the Lemma is also true under independence.

**Proof of Lemma 3.5.1** Note that

$$\begin{aligned}\tilde{\sigma}_j(u,v) &= \frac{1}{n-|j|}\sum_{t=|j|+1}^{n} e^{i(uX_t+vX_{t-|j|})} - \frac{\phi(v)}{n-|j|}\sum_{t=|j|+1}^{n} e^{iuX_t} \\ &\quad - \frac{\phi(u)}{n-|j|}\sum_{t=|j|+1}^{n} e^{ivX_{t-|j|}} + \phi(u)\phi(v)\end{aligned}$$

Therefore, by subtracting $\tilde{\sigma}_j(u,v)$ from $\hat{\sigma}_j(u,v)$ we get:

$$
\begin{aligned}
\hat{\sigma}_j(u,v)-\tilde{\sigma}_j(u,v) &= -\frac{1}{n-|j|}\sum_{t=|j|+1}^{n} e^{i(uX_t)}\sum_{t=|j|+1}^{n} e^{i(vX_{t-|j|})}+\frac{\phi(v)}{n-|j|}\sum_{t=|j|+1}^{n} e^{iuX_t} \\
&\quad +\frac{\phi(u)}{n-|j|}\sum_{t=|j|+1}^{n} e^{ivX_{t-|j|}}-\phi(u)\phi(v) \\
&= -\frac{1}{n-|j|}\sum_{t=|j|+1}^{n} e^{iuX_t}\left[\frac{\sum_{t=|j|+1}^{n} e^{ivX_{t-|j|}}}{n-|j|}-\phi(v)\right] \\
&\quad +\phi(u)\left[\frac{\sum_{t=|j|+1}^{n} e^{ivX_{t-|j|}}}{n-|j|}-\phi(v)\right] \\
&= -\left\{\frac{\sum_{t=|j|+1}^{n} e^{iuX_t}}{n-|j|}-\phi(u)\right\}\left\{\frac{\sum_{t=|j|+1}^{n} e^{ivX_{t-|j|}}}{n-|j|}-\phi(v)\right\}. \\
&= -\frac{1}{(n-|j|)^2}\sum_{t=|j|+1}^{n}\psi_t(u)\sum_{t=|j|+1}^{n}\psi_{t-|j|}(v).
\end{aligned}
$$

The Cauchy – Schwarz inequality implies that

$$
E\left|\hat{\sigma}_j(u,v)-\tilde{\sigma}_j(u,v)\right|^2 \le \frac{1}{(n-|j|)^4}\left\{E\left|\sum_{t=|j|+1}^{n}\psi_t(u)\right|^4 E\left|\sum_{t=|j|+1}^{n}\psi_{t-|j|}(v)\right|^4\right\}^{1/2} \le \frac{C}{(n-|j|)^2}
$$

uniformly, because $E\left|\sum_{t=|j|+1}^{n}\psi_t(u)\right|^4 \le C(n-|j|)^2$ given Assumption 3($ii$) (Doukhan and Louhichi, 1999, Lemma 6).

From the definition of $\tilde{\sigma}_j(u,v)$ in (3.19), it follows that:

$$
E\left|\tilde{\sigma}_j(u,v)\right|^2 = E\left|\frac{1}{n-|j|}\sum_{t=|j|+1}^{n}\psi_t(u)\psi_{t-|j|}(v)\right|^2 \le \frac{C}{n-|j|}
$$

uniformly. □

**Lemma 3.5.2** Suppose that $\{X_t, t\ge 1\}$ satisfies Assumptions 1, 3($ii$) and 4. For each $\gamma>0$, denote by $D(\gamma)$ the region $D(\gamma)=\{(u,v):\gamma\le|u|\le 1/\gamma, \gamma\le|v|\le 1/\gamma\}$. Then, under Assumption 4, for any fixed $\gamma > 0$,

$$\int_{D(\gamma)} \sum_{j=1}^{n-1} k^2(j/p)(n-j)\Big\{|\hat{\sigma}_j(u,v)|^2 - |\tilde{\sigma}_j(u,v)|^2\Big\} d\mathcal{W} \;=\; O_P(p/\sqrt{n}) = o_P(\sqrt{p})$$

as $p/n \to 0$. The result of the Lemma is also true under independence.

**Proof of Lemma 3.5.2** From the properties of kernels (Hong, 1999, p. 1213) we obtain that

$$\sum_{|j|<n} \frac{1}{n-|j|} k^2(j/p) \;=\; O(p/n)$$

for bandwidth $p = cn^\lambda$, $\lambda \in (0,1)$. In addition, we observe that

$$\begin{aligned}
\sum_{|j|<n} \frac{1}{n-|j|} k^2(j/p) &= \sum_{j=-(n-1)}^{n-1} \frac{1}{n-|j|} k^2(j/p) \\
&= \sum_{j=-(n-1)}^{-1} \frac{1}{n-|j|} k^2(j/p) + \frac{1}{n} k^2(0) + \sum_{j=1}^{n-1} \frac{1}{n-j} k^2(j/p) \\
&= 2\sum_{j=1}^{n-1} \frac{1}{n-j} k^2(j/p) + \frac{1}{n}
\end{aligned}$$

and so

$$\sum_{j=1}^{n-1} \frac{1}{n-j} k^2(j/p) \;=\; O(p/n). \tag{3.21}$$

Now, using that $\sqrt{n-j} \leq \sqrt{n}$, we further observe that

$$\sum_{j=1}^{n-1} \frac{1}{\sqrt{n-j}} k^2(j/p) \;\leq\; \sum_{j=1}^{n-1} \frac{\sqrt{n}}{n-j} k^2(j/p) = O(p/\sqrt{n}). \tag{3.22}$$

Now, define

$$Z_{n;p}^{\gamma} \;=\; \int_{D(\gamma)} \sum_{j=1}^{n-1} k^2(j/p)(n-j)\Big\{|\hat{\sigma}_j(u,v)|^2 - |\tilde{\sigma}_j(u,v)|^2\Big\} d\mathcal{W},$$

and observe that, for any fixed $\gamma > 0$, the chosen weight function $\mathcal{W}(u,v)$ in equation

(3.3), is bounded on $D(\gamma)$. Thus, $\mathcal{W}(u,v)$ is an integrable function in this region. To prove the first result of this Lemma, we need to show that for any $\epsilon > 0$, there exists a finite $M = \delta(\epsilon) > 0$ such that

$$P\left(\frac{\sqrt{n}\left|Z_{n;p}^{\gamma}\right|}{p} \geq M\right) \;\leq\; \epsilon.$$

Using,

$$|\hat{\sigma}_j(u,v)|^2 - |\tilde{\sigma}_j(u,v)|^2 \;=\; |\hat{\sigma}_j(u,v) - \tilde{\sigma}_j(u,v)|^2 + 2\mathrm{Re}\{|\hat{\sigma}_j(u,v) - \tilde{\sigma}_j(u,v)|\, \tilde{\sigma}_j(u,v)^*\},$$

where $*$ denotes complex conjugate, we get the following:

$$\begin{aligned}
E\left|Z_{n;p}^{\gamma}\right| \;&\leq\; \int_{D(\gamma)} \left\{\sum_{j=1}^{n-1} k^2(j/p)(n-j)E\Big(|\hat{\sigma}_j(u,v)|^2 - |\tilde{\sigma}_j(u,v)|^2\Big)\right\} d\mathcal{W} \\
&=\; \int_{D(\gamma)} \left\{\sum_{j=1}^{n-1} \frac{k^2(j/p)}{(n-j)}(n-j)^2 E\Big(|\hat{\sigma}_j(u,v) - \tilde{\sigma}_j(u,v)|^2\Big)\right\} d\mathcal{W} \\
&\quad +2\int_{D(\gamma)} \left\{\sum_{j=1}^{n-1} k^2(j/p)(n-j)\mathrm{Re}\{E\,|\hat{\sigma}_j(u,v) - \tilde{\sigma}_j(u,v)|\, \tilde{\sigma}_j(u,v)^*\}\right\} d\mathcal{W}.
\end{aligned}$$

Now Lemma 3.5.1, the Cauchy – Schwarz inequality, relations (3.21) and (3.22) and the fact that $\int_{D(\gamma)} d\mathcal{W} < \infty$ show that,

$$\begin{aligned}
E\left|Z_{n;p}^{\gamma}\right| \;&\leq\; \int_{D(\gamma)} \left\{\sum_{j=1}^{n-1} \frac{k^2(j/p)}{(n-j)}(n-j)^2 E\Big(|\hat{\sigma}_j(u,v) - \tilde{\sigma}_j(u,v)|^2\Big)\right\} d\mathcal{W} \\
&\quad +\; 2\int_{D(\gamma)} \sum_{j=1}^{n-1} \frac{k^2(j/p)}{\sqrt{n-j}} \sqrt{(n-j)^2 E\,|\hat{\sigma}_j(u,v) - \tilde{\sigma}_j(u,v)|^2}\sqrt{(n-j)E\,|\tilde{\sigma}_j(u,v)|^2}\, d\mathcal{W} \\
&\leq\; C_1 \int_{D(\gamma)} d\mathcal{W} \sum_{j=1}^{n-1} \frac{k^2(j/p)}{n-j} + C_2 \int_{D(\gamma)} d\mathcal{W} \sum_{j=1}^{n-1} \frac{k^2(j/p)}{\sqrt{n-j}} \\
&=\; O(p/\sqrt{n}) \qquad\qquad (3.23)
\end{aligned}$$

as $p/n \to 0$. By Markov's inequality, (3.23) implies the first result.

To prove the second result of the Lemma, we actually need to show that

$$P\left(\frac{\left|Z^{\gamma}_{n;p}\right|}{\sqrt{p}} \geq \epsilon\right) \to 0,$$

as $n, p \to \infty$ and $p/n \to 0$, $\forall\ \ \epsilon > 0$. By Markov's inequality and (3.23), the second result follows immediately. $\square$

**Proof of Proposition 3.2.2** Recall the region $D(\gamma)$ defined in Lemma 3.5.2, and for each $\gamma > 0$, define the random variables

$$\widehat{V}^2_{X;\gamma}(j) = \int_{D(\gamma)} \left|\hat{\sigma}_j(u,v)\right|^2 d\mathcal{W}. \tag{3.24}$$

Because of the SLLN for $\alpha$-mixing random variables and the fact that $\int_{D(\gamma)} d\mathcal{W} < \infty$, we obtain

$$\lim_{n\to\infty} \widehat{V}^2_{X;\gamma}(j) = V^2_{X;\gamma}(j) = \int_{D(\gamma)} \left|\sigma_j(u,v)\right|^2 d\mathcal{W},$$

almost surely. Clearly, $V^2_{X;\gamma}(j) \to V^2_X(j)$ as $\gamma$ tends to zero. So, it remains to prove that almost surely

$$\limsup_{\gamma\to 0} \limsup_{n\to\infty} \left|\widehat{V}^2_{X;\gamma}(j) - \widehat{V}^2_X(j)\right| = 0. \tag{3.25}$$

For each $\gamma > 0$, we obtain that

$$\begin{aligned}\left|\widehat{V}^2_{X;\gamma}(j) - \widehat{V}^2_X(j)\right| \ \leq\ & \int_{|u|<\gamma} \left|\hat{\sigma}_j(u,v)\right|^2 d\mathcal{W} + \int_{|u|>1/\gamma} \left|\hat{\sigma}_j(u,v)\right|^2 d\mathcal{W} \\ & + \int_{|v|<\gamma} \left|\hat{\sigma}_j(u,v)\right|^2 d\mathcal{W} + \int_{|v|>1/\gamma} \left|\hat{\sigma}_j(u,v)\right|^2 d\mathcal{W}. \end{aligned} \tag{3.26}$$

Now, for $z \in \mathbb{R}$, define $H(y) = \int_{|z|<y} (1 - \cos z)/\left|z\right|^2 dz$. This is bounded by $\pi$, and

$\lim_{y\to 0} H(y) = 0$. Because of $|x+y|^2 \leq 2|x|^2 + 2|y|^2$ and the Cauchy – Schwarz inequality

$$
\begin{aligned}
|\hat{\sigma}_j(u,v)|^2 &= |\tilde{\sigma}_j(u,v) + \hat{\sigma}_j(u,v) - \tilde{\sigma}_j(u,v)|^2 \\
&\leq 2\left|\frac{1}{n-|j|}\sum_{t=|j|+1}^{n}\psi_t(u)\psi_{t-|j|}(v)\right|^2 + 2\left|\frac{1}{n-|j|}\sum_{t=|j|+1}^{n}\psi_t(u)\frac{1}{n-|j|}\sum_{t=|j|+1}^{n}\psi_{t-|j|}(v)\right|^2 \\
&\leq 4\left\{\frac{1}{(n-|j|)}\sum_{t=|j|+1}^{n}|\psi_t(u)|^2\right\}\left\{\frac{1}{(n-|j|)}\sum_{t=|j|+1}^{n}\left|\psi_{t-|j|}(v)\right|^2\right\}. \qquad (3.27)
\end{aligned}
$$

But, the first summand in (3.26) satisfies:

$$
\begin{aligned}
\int_{|u|<\gamma}|\hat{\sigma}_j(u,v)|^2\, d\mathcal{W} \;\leq\; & \left\{\frac{4}{(n-|j|)}\sum_{t=|j|+1}^{n}\int_{|u|<\gamma}\frac{|\psi_t(u)|^2}{\pi|u|^2}du\right\} \\
& \times\left\{\frac{1}{(n-|j|)}\sum_{t=|j|+1}^{n}\int_{\mathbb{R}}\frac{\left|\psi_{t-|j|}(v)\right|^2}{\pi|v|^2}dv\right\}. \qquad (3.28)
\end{aligned}
$$

However, (3.20) yields

$$
\begin{aligned}
\left|\psi_{t-|j|}(v)\right|^2 &= \left(e^{ivX_{t-|j|}} - \phi(v)\right)\left(e^{ivX_{t-|j|}} - \phi(v)\right)^* \\
&= e^{ivX_{t-|j|}}\left(e^{ivX_{t-|j|}}\right)^* - \phi(v)^* e^{ivX_{t-|j|}} - \phi(v)\left(e^{ivX_{t-|j|}}\right)^* + \phi(v)\phi(v)^* \\
&= 1 + |\phi(v)|^2 - \phi(v)^* e^{ivX_{t-|j|}} - \phi(v)\left(e^{ivX_{t-|j|}}\right)^*
\end{aligned}
$$

and similarly for $|\psi_t(u)|^2$. Now, letting $X_{t-|j|} \equiv Y_t$ and using Lemma 2.2.1 we get

$$
\begin{aligned}
\int_{\mathbb{R}} \frac{\left|\psi_{t-|j|}(v)\right|^2}{\pi |v|^2} dv &= \int_{\mathbb{R}} \frac{1 + |\phi(v)|^2 - \phi(v)^* e^{ivX_{t-|j|}} - \phi(v)\left(e^{ivX_{t-|j|}}\right)^*}{\pi |v|^2} dv \\
&= E \int_{\mathbb{R}} \frac{1 + e^{iv(Y-Y')} - e^{iv(Y_t - Y)} - e^{iv(Y-Y_t)}}{\pi |v|^2} dv \\
&= E \int_{\mathbb{R}} \frac{1 + \cos(v|Y-Y'|) - \cos(v|Y_t - Y|) - \cos(v|Y_t - Y|)}{\pi |v|^2} dv \\
&= E \int_{\mathbb{R}} \frac{2 - 1 + \cos(v|Y-Y'|) - 2\cos(v|Y_t - Y|)}{\pi |v|^2} dv \\
&= E \int_{\mathbb{R}} \frac{-(1 - \cos(v|Y-Y'|)) + 2(1 - \cos(v|Y_t - Y|))}{\pi |v|^2} dv \\
&= 2E_Y |Y_t - Y| - E|Y - Y'| \leq 2(|Y_t| + E|Y|) = 2(\left|X_{t-|j|}\right| + E|X_1|),
\end{aligned}
$$

where the expectation $E_Y$ is taken with respect to $Y$ and we denote by $Y'$ the random variable which is a copy of $Y$ and independent of $Y_t$. Similarly

$$
\begin{aligned}
\int_{|u|<\gamma} \frac{|\psi_t(u)|^2}{\pi |u|^2} &= \int_{|u|<\gamma} \frac{1 + |\phi(u)|^2 - e^{iuX_t}\phi(u)^* - \phi(u)\left(e^{iuX_t}\right)^*}{\pi |u|^2} du \\
&= E \int_{|u|<\gamma} \frac{1 + \cos(u|X - X'|) - 2\cos(u|X_t - X|)}{\pi |u|^2} du \\
&= E \int_{|u|<\gamma} \frac{-(1 - \cos(u|X-X'|)) + 2(1 - \cos(u|X_t - X|))}{\pi |u|^2} du \\
&= 2E_X |X_t - X| H(\gamma |X_t - X|) - E|X - X'| H(\gamma |X - X'|) \\
&\leq 2E_X |X_t - X| H(|X_t - X| \gamma)
\end{aligned}
$$

where the expectation $E_X$ is taken with respect to $X$. Therefore, from (3.28)

$$
\begin{aligned}
\int_{|u|<\gamma} |\hat{\sigma}_j(u,v)|^2 d\mathcal{W} \leq\ & 16 \frac{1}{(n-|j|)} \sum_{t=|j|+1}^{n} \left(\left|X_{t-|j|}\right| + E|X_1|\right) \\
& \times \frac{1}{(n-|j|)} \sum_{t=|j|+1}^{n} E_X[|X_t - X| H(|X_t - X| \gamma)].
\end{aligned}
$$

Because of the Assumptions 1 and 2 and the ergodic theorem for $\alpha$-mixing processes we obtain

$$\frac{1}{n-|j|}\sum_{t=|j|+1}^{n}\Big(\left|X_{t-|j|}\right|+E\left|X_1\right|\Big) \;\;\rightarrow\;\; E\left|X_1\right|+E\left|X_1\right|=2E\left|X_1\right|,$$

$$\frac{1}{n-|j|}\sum_{t=|j|+1}^{n}E_X\left|X_t-X\right|H(\left|X_t-X\right|\gamma) \;\;\rightarrow\;\; E\left|X_0-X_1\right|H(\left|X_0-X_1\right|\gamma),$$

as $n\rightarrow\infty$, almost surely. Therefore,

$$\limsup_{n\rightarrow\infty}\int_{|u|<\gamma}\left|\hat{\sigma}_j(u,v)\right|^2 d\mathcal{W} \;\;\leq\;\; 32E\left|X_1\right|E\left|X_0-X_1\right|H(\left|X_0-X_1\right|\gamma)$$

and by Lebesgue's dominated convergence theorem,

$$\limsup_{\gamma\rightarrow 0}\limsup_{n\rightarrow\infty}\int_{|u|<\gamma}\left|\hat{\sigma}_j(u,v)\right|^2 d\mathcal{W} \;\;=\;\; 0.$$

For the second term of (3.26), (3.27) implies that $\left|\psi_t(u)\right|^2\leq 4$ and $1/(n-|j|)\sum_{t=|j|+1}^{n}\left|\psi_t(u)\right|^2\leq 4$. Therefore

$$\begin{aligned}\int_{|u|>1/\gamma}\left|\hat{\sigma}_j(u,v)\right|^2 d\mathcal{W} \;\;&\leq\;\; 16\int_{|u|>1/\gamma}\frac{du}{\pi\left|u\right|^2}\int_{\mathbb{R}}\frac{1}{n-|j|}\sum_{t=|j|+1}^{n}\frac{\left|\psi_{t-|j|}(v)\right|^2}{\pi\left|v\right|^2}dv\\ &\leq\;\; 16\gamma\frac{2}{n-|j|}\sum_{t=|j|+1}^{n}(\left|X_{t-|j|}\right|+E\left|X_1\right|).\end{aligned}$$

Then, almost surely for suitable chosen $\gamma>0$

$$\limsup_{\gamma\rightarrow 0}\limsup_{n\rightarrow\infty}\int_{|u|>1/\gamma}\left|\hat{\sigma}_j(u,v)\right|^2 d\mathcal{W} \;\;=\;\; 0.$$

The other two summands of (3.26) can be dealt in a similar way to obtain (3.25). □

**Proof of Theorem 3.4.1** Arguing as in Hong (1999), we first define

$$V_{tsj}(u,v) \;\;=\;\; C_{tsj}(u,v)+C_{stj}(u,v)^*,$$

where

$$C_{tsj}(u,v) \;=\; \psi_t(u)\psi_s(u)^*\psi_{t-j}(v)\psi_{s-j}(v)^*.$$

Since $C_{tsj}(u,v) = C_{stj}(u,v)^*$, $V_{tsj}(u,v)$ is real-valued and symmetric in $t$ and $s$; that is, $V_{tsj}(u,v) = V_{stj}(u,v)$. We then get the following result:

$$\begin{aligned}
\sum_{j=1}^{n-1} k^2(j/p)(n-j)\left|\tilde{\sigma}_j(u,v)\right|^2 &= \sum_{j=1}^{n-1} \frac{k^2(j/p)}{n-j}\left|\sum_{t=j+1}^{n} \psi_t(u)\psi_{t-j}(v)\right|^2 \\
&= \sum_{j=1}^{n-1} \frac{k^2(j/p)}{n-j}\left|\psi_{j+1}(u)\psi_1(v) + \cdots + \psi_n(u)\psi_{n-j}(v)\right|^2 \\
&= \sum_{j=1}^{n-1} \frac{k^2(j/p)}{n-j}\Bigg[\sum_{t=j+1}^{n} \left|\psi_t(u)\right|^2\left|\psi_{t-j}(v)\right|^2 \\
&\quad +2\sum_{t=j+2}^{n}\sum_{s=j+1}^{t-1} \left|\psi_t(u)\right|\left|\psi_s(u)\right|\left|\psi_{t-j}(v)\right|\left|\psi_{s-j}(v)\right|\Bigg] \\
&= \sum_{j=1}^{n-1} \frac{k^2(j/p)}{n-j}\Bigg[\sum_{t=j+1}^{n} \left|\psi_t(u)\right|^2\left|\psi_{t-j}(v)\right|^2 \\
&\quad +2\sum_{t=j+2}^{n}\sum_{s=j+1}^{t-1} \left|\psi_t(u)\right|\left|\psi_s(u)^*\right|\left|\psi_{t-j}(v)\right|\left|\psi_{s-j}(v)^*\right|\Bigg] \\
&= \sum_{j=1}^{n-1} \frac{k^2(j/p)}{n-j}\left[\sum_{t=j+1}^{n} C_{ttj}(u,v) + \sum_{t=j+2}^{n}\sum_{s=j+1}^{t-1} V_{tsj}(u,v)\right] \\
&= \widehat{C}(u,v) + \widehat{V}(u,v) \qquad (3.29)
\end{aligned}$$

where

$$\widehat{C}(u,v) \;=\; \sum_{j=1}^{n-1} \frac{k^2(j/p)}{n-j}\left[\sum_{t=j+1}^{n} C_{ttj}(u,v)\right],$$

$$\widehat{V}(u,v) \;=\; \sum_{j=1}^{n-1} \frac{k^2(j/p)}{n-j}\left[\sum_{t=j+2}^{n}\sum_{s=j+1}^{t-1} V_{tsj}(u,v)\right].$$

Consider the first term of (3.29). We observe that $\int_{D(\gamma)} C_{ttj}(u,v)d\mathcal{W}$ and $\int_{D(\gamma)} C_{ssj}(u,v)d\mathcal{W}$ are two independent integrals unless $t=s$ or $s\pm j$. In addition

$$
\begin{aligned}
E\int_{D(\gamma)} C_{ttj}(u,v)d\mathcal{W} &= E\int_{D(\gamma)} |\psi_t(u)|^2\,|\psi_{t-j}(v)|^2\,d\mathcal{W} \\
&= E\int_{D(\gamma)} |\psi_t(u)|^2\,d\mathcal{W}_0(u) E\int_{D(\gamma)} |\psi_{t-j}(v)|^2\,d\mathcal{W}_0(v) \\
&= \int_{D(\gamma)} 1-|\phi(u)|^2\,d\mathcal{W}_0(u)\int_{D(\gamma)} 1-|\phi(v)|^2\,d\mathcal{W}_0(v) \\
&= \int_{D(\gamma)} \sigma_0(u,-u)d\mathcal{W}_0(u)\int_{D(\gamma)} \sigma_0(v,-v)d\mathcal{W}_0(v) \\
&= \int_{D(\gamma)} \sigma_0(u,-u)\sigma_0(v,-v)d\mathcal{W} =: C_0^\gamma < \infty.
\end{aligned}
$$

It follows that $E\left\{\sum_{t=j+1}^n\left[\int_{D(\gamma)} C_{ttj}(u,v)d\mathcal{W} - C_0^\gamma\right]\right\}^2 \le C(n-j)$, under independence. We now prove that

$$
\int_{D(\gamma)} \widehat{C}(u,v)d\mathcal{W} - C_0^\gamma\sum_{j=1}^{n-1} k^2(j/p) \quad = \quad O_P(p/\sqrt{n}). \tag{3.30}
$$

To prove (3.30), we actually need to show that for any $\epsilon > 0$, there exists a finite $M = \delta(\epsilon) > 0$ such that

$$
P\left(\frac{\sqrt{n}}{p}\left|\int_{D(\gamma)} \widehat{C}(u,v)d\mathcal{W} - C_0^\gamma\sum_{j=1}^{n-1} k^2(j/p)\right| \ge M\right) \le \epsilon.
$$

At first place, we observe that $E\int_{D(\gamma)} \widehat{C}(u,v)d\mathcal{W} = C_0^\gamma\sum_{j=1}^{n-1} k^2(j/p)$. Applying the

Cauchy – Schwarz inequality and (3.21) we obtain that

$$
\begin{aligned}
E\left(\int_{D(\gamma)} \widehat{C}(u,v)d\mathcal{W} - C_0^\gamma \sum_{j=1}^{n-1} k^2(j/p)\right)^2 &= E\left(\sum_{j=1}^{n-1} \frac{k^2(j/p)}{n-j}\left[\sum_{t=j+1}^{n}\left\{\int_{D(\gamma)} C_{ttj}(u,v)d\mathcal{W} - C_0^\gamma\right\}\right]\right)^2 \\
&= E\left(\sum_{j=1}^{n-1} \frac{k(j/p)}{\sqrt{n-j}}\left[\sum_{t=j+1}^{n} \frac{k(j/p)}{\sqrt{n-j}}\right.\right. \\
&\qquad \left.\left. \times\left\{\int_{D(\gamma)} C_{ttj}(u,v)d\mathcal{W} - C_0^\gamma\right\}\right]\right)^2 \\
&\leq E\left(\sum_{j=1}^{n-1} \frac{k^2(j/p)}{n-j} \sum_{j=1}^{n-1}\left[\sum_{t=j+1}^{n} \frac{k(j/p)}{\sqrt{n-j}}\right.\right. \\
&\qquad \left.\left. \times\left\{\int_{D(\gamma)} C_{ttj}(u,v)d\mathcal{W} - C_0^\gamma\right\}\right]^2\right) \\
&= \sum_{j=1}^{n-1} \frac{k^2(j/p)}{n-j} \sum_{j=1}^{n-1} \frac{k^2(j/p)}{n-j} \\
&\qquad \times E\left(\sum_{t=j+1}^{n}\left\{\int_{D(\gamma)} C_{ttj}(u,v)d\mathcal{W} - C_0^\gamma\right\}\right)^2 \\
&\leq O(p/\sqrt{n}).
\end{aligned}
$$

By Markov's inequality this implies the relation (3.30). Lemma 3.5.2 and Equations (3.29) and (3.30) yield that

$$
\begin{aligned}
\int_{D(\gamma)}\left\{\sum_{j=1}^{n-1} k^2(j/p)(n-j)\left|\hat{\sigma}_j(u,v)\right|^2\right\}d\mathcal{W} &= \int_{D(\gamma)}\left\{\sum_{j=1}^{n-1} k^2(j/p)(n-j)\left|\tilde{\sigma}_j(u,v)\right|^2\right\}d\mathcal{W} \\
&\quad + O_P(p/\sqrt{n}) \\
&= \int_{D(\gamma)} \widehat{C}(u,v)d\mathcal{W} + \int_{D(\gamma)} \widehat{V}(u,v)d\mathcal{W} \\
&\quad + O_P(p/\sqrt{n}) \\
&= C_0^\gamma \sum_{j=1}^{n-1} k^2(j/p) + \widehat{V}_n^\gamma + O_P(p/\sqrt{n}), \quad (3.31)
\end{aligned}
$$

where $\widehat{V}_n^\gamma \equiv \int_{D(\gamma)} \widehat{V}(u,v)d\mathcal{W}$.

Because of Assumption 4, we obtain by applying Hong (1999, Theorem A3) on $D(\gamma)$ that

$$\widehat{V}_n^\gamma = \widehat{V}_{ng}^\gamma + o_P(\sqrt{p}) \tag{3.32}$$

where

$$\widehat{V}_{ng}^\gamma = \sum_{t=g+2}^{n} \sum_{s=1}^{t-g-1} \sum_{j=1}^{g} \frac{k^2(j/p)}{n-j} \int_{D(\gamma)} V_{tsj}(u,v) d\mathcal{W}$$

and $g \equiv g(n)$ such that $g/p \to 0, g/n \to 0$. Now, by applying Hong (1999, Theorem A4) on $D(\gamma)$ we get the following:

$$\left[ pD_0^\gamma \int_0^\infty k^4(z)dz \right]^{-1/2} \widehat{V}_{ng}^\gamma \to N(0,1) \tag{3.33}$$

as $n \to \infty$ in distribution, where

$$D_0^\gamma = 2\left[ \int_{D(\gamma)} |\sigma_0(u,u')|^2 \, d\mathcal{W}_0(u) d\mathcal{W}_0(u') \right]^2.$$

Then, using (3.31), (3.32) and (3.33) we have the following:

$$\frac{\int_{D(\gamma)} \left\{ \sum_{j=1}^{n-1} k^2(j/p)(n-j) \left|\hat{\sigma}_j(u,v)\right|^2 \right\} d\mathcal{W} - C_0^\gamma \sum_{j=1}^{n-1} k^2(j/p)}{\left[ pD_0^\gamma \int_0^\infty k^4(z)dz \right]^{1/2}} \to N(0,1). \tag{3.34}$$

as $n \to \infty$ in distribution.

Observe that $\widehat{C}_0^\gamma - C_0^\gamma = O_P(1/\sqrt{n})$ and that $\sum_{j=1}^{n-1} k^2(j/p) = O(p)$. Hence $C_0^\gamma$ can be replaced by $\widehat{C}_0^\gamma$ asymptotically given $p/n \to 0$. Furthermore, $p^{-1} \sum_{j=1}^{n-2} k^4(j/p) \to \int_0^\infty k^4(z)dz$ and $\widehat{D}_0^\gamma \to D_0^\gamma$, in probability. We conclude that the factor $pD_0^\gamma \int_0^\infty k^4(z)dz$

can be replaced by $\widehat{D}_0^\gamma \sum_{j=1}^{n-2} k^4(j/p)$, by Slutsky's theorem. Thus, (3.34) becomes

$$\frac{\int_{D(\gamma)} \left\{ \sum_{j=1}^{n-1} k^2(j/p)(n-j) \left|\hat{\sigma}_j(u,v)\right|^2 \right\} d\mathcal{W} - \widehat{C}_0^\gamma \sum_{j=1}^{n-1} k^2(j/p)}{\left[ \widehat{D}_0^\gamma \sum_{j=1}^{n-2} k^4(j/p) \right]^{1/2}} \quad \to \quad N(0,1). \qquad (3.35)$$

Write $T_{n;\gamma}$ as the test-statistic $T_n$ defined on $D(\gamma)$ rather than on $\mathbb{R}^2$, i.e.

$$T_{n;\gamma} \;=\; \int_{D(\gamma)} \left\{ \sum_{j=1}^{n-1} k^2(j/p)(n-j) \left|\hat{\sigma}_j(u,v)\right|^2 \right\} d\mathcal{W}.$$

and note that

$$T_n - T \;=\; T_n - T_{n;\gamma} + T_{n;\gamma} - T_{.\gamma} + T_{.\gamma} - T,$$

where $T_{.\gamma}$ is defined as an asymptotically distributed normal random variable such that

$$\frac{T_{.\gamma} - \widehat{C}_0^\gamma \sum_{j=1}^{n-1} k^2(j/p)}{\left[ \widehat{D}_0^\gamma \sum_{j=1}^{n-2} k^4(j/p) \right]^{1/2}} \quad \to \quad N(0,1),$$

in distribution, as $n \to \infty$ such that $p/n \to 0$. In addition, $T$ is an asymptotically normally distributed random variable with

$$\frac{T - \widehat{C}_0 \sum_{j=1}^{n-1} k^2(j/p)}{\left[ \widehat{D}_0 \sum_{j=1}^{n-2} k^4(j/p) \right]^{1/2}} \quad \to \quad N(0,1)$$

as $\gamma \to 0$. Obviously, (3.35) shows that

$$T_{n;\gamma} - T_{.\gamma} \;=\; o_P(1),$$

as $n \to \infty$. But, it is also true that

$$T_{.\gamma} - T = o_P(1),$$

as $\gamma \to 0$, because $\widehat{C}_0^\gamma \to \widehat{C}_0$ and $\widehat{D}_0^\gamma \to \widehat{D}_0$. If we show that

$$\limsup_{\gamma\to 0} \limsup_{n\to\infty} |T_n - T_{n;\gamma}| = 0, \tag{3.36}$$

almost surely, then the proof of Theorem 3.4.1 will be completed.

For each $\gamma > 0$:

$$\begin{aligned} |T_n - T_{n;\gamma}| &= \left| \int_{\mathbb{R}^2} \left\{ \sum_{j=1}^{n-1} k^2(j/p)(n-j) \left|\hat{\sigma}_j(u,v)\right|^2 \right\} d\mathcal{W} \right. \\ &\quad \left. - \int_{D(\gamma)} \left\{ \sum_{j=1}^{n-1} k^2(j/p)(n-j) \left|\hat{\sigma}_j(u,v)\right|^2 \right\} d\mathcal{W} \right| \\ &= \sum_{j=1}^{n-1} k^2(j/p)(n-j) \left| \int_{\mathbb{R}^2} \left|\hat{\sigma}_j(u,v)\right|^2 d\mathcal{W} - \int_{D(\gamma)} \left|\hat{\sigma}_j(u,v)\right|^2 d\mathcal{W} \right| \\ &= \sum_{j=1}^{n-1} k^2(j/p)(n-j) \left| \widehat{V}^2_{X;\gamma}(j) - \widehat{V}^2_X(j) \right| \end{aligned} \tag{3.37}$$

where $\widehat{V}^2_{X;\gamma}(j)$ is defined as in (3.24). Now, recall result (3.25). Combine (3.37) and (3.25) to finally get the required relation (3.36). □

**Proof of Proposition 3.4.1** Note that if $X_t$ follows a standard normal distribution and $X'_t$ is an i.i.d. copy of $X_t$, then $X_t - X'_t$ follows the normal distribution with mean 0 and variance 2, i.e $X_t - X'_t \sim N(0,2)$. It is well known that the random variable $|X_t - X'_t|$ follows a half-normal distribution with mean given by

$$E\left|X_t - X'_t\right| = \frac{2}{\sqrt{\pi}},$$

and so

$$C_0 = \left(E\left|X_t - X'_t\right|\right)^2 = 4/\pi. \tag{3.38}$$

Moreover, recall that $D_0 = 2V_X^4(0)$ and that $V_X^2(0) = F(1)/\pi^2$. By a careful check in the calculations derived in the proof of Proposition 3.2.1 we obtain that

$$D_0 = 2\left[\frac{4}{\pi}(1 + \pi/3 - \sqrt{3})\right]^2. \tag{3.39}$$

From the asymptotic normality of $T_n$ derived in Theorem 3.4.1 and combining (3.38) and (3.39) we finally get the required result. □

**Proof of Theorem 3.4.2** We need to show the following: (i) $1/p\sum_{j=1}^{n} k^4(j/p) \to \int_0^\infty k^4(j/p)$. This follows from Assumption 4, $p \to \infty$ and $p/n \to 0$. In addition, we need that (ii) $EL^2_{2:\gamma}(\hat{f}_n, f) \to 0$ where

$$L^2_{2;\gamma}(\hat{f}_n, f) \equiv \int_{D(\gamma)} \int_{-\pi}^{\pi} \left|\hat{f}_n(\omega, u, v) - f(\omega, u, v)\right|^2 d\omega d\mathcal{W}(u, v)$$

and

$$\hat{f}_n(\omega, u, v) \;=\; \frac{1}{2\pi} \sum_{j=-(n-1)}^{(n-1)} (1 - |j|/n)^{1/2} k(j/p)\hat{\sigma}_j(u, v)e^{-ij\omega}.$$

This condition is established along the lines of Hong (1999, Theorem 2) on $D(\gamma)$ given Assumptions 1, 3($i$) and 4. Furthermore, by applying Markov's inequality we get (iii) $\widehat{C}_0^\gamma = O_P(1)$ and (iv) $\widehat{D}_0^\gamma \to D_0^\gamma$ in probability.

Combining (i) and (iv) and by using Slutsky's theorem we get

$$\frac{1}{\sqrt{p}}\left[\widehat{D}_0^\gamma \sum_{j=1}^{n-2} k^4(j/p)\right]^{1/2} \;\to\; \left[D_0^\gamma \int_0^\infty k^4(z)dz\right]^{1/2}, \tag{3.40}$$

in probability.

Now, recall (3.11) and (3.13) and define

$$L^2_{2;\gamma}(\hat{f}_n, \hat{f}_0) \equiv \int_{D(\gamma)} \int_{-\pi}^{\pi} \left|\hat{f}_n(\omega, u, v) - \hat{f}_0(\omega, u, v)\right|^2 d\omega d\mathcal{W} = \frac{2}{\pi} T_{n;\gamma}.$$

Using the inequality $|x+y|^2 \leq 2|x|^2 + 2|y|^2$ and after some calculations we observe that

$$\frac{\pi}{2}\left[\frac{1}{n}L^2_{2;\gamma}(\hat{f}_n,\hat{f}_0) - L^2_{2;\gamma}(f,f_0)\right] \;\leq\; L^2_{2;\gamma}(\hat{f}_n,f).$$

Based on the last expression, we obtain that

$$E\left|\frac{\pi}{2}\left[\frac{1}{n}L^2_{2;\gamma}(\hat{f}_n,\hat{f}_0) - L^2_{2;\gamma}(f,f_0)\right] - \frac{1}{n}\sum_{j=1}^{n-1}k^2(j/p)\widehat{C}_0^{\gamma}\right| \leq EL^2_{2;\gamma}(\hat{f}_n,f) + \frac{1}{n}\sum_{j=1}^{n-1}k^2(j/p)E\left|\widehat{C}_0^{\gamma}\right|.$$

By applying Markov's inequality and (ii) and (iii), the last result yields

$$\frac{\pi}{2}\frac{1}{n}L^2_{2;\gamma}(\hat{f}_n,\hat{f}_0) - \frac{1}{n}\sum_{j=1}^{n-1}k^2(j/p)\widehat{C}_0^{\gamma} \;\to\; \frac{\pi}{2}L^2_{2;\gamma}(f,f_0),$$

as $n\to\infty$, in probability, i.e.

$$\frac{1}{n}\left[T_{n;\gamma} - \widehat{C}_0^{\gamma}\sum_{j=1}^{n-1}k^2(j/p)\right] \;\to\; \frac{\pi}{2}\int_{D(\gamma)}\int_{-\pi}^{\pi}|f(\omega,u,v)-f_0(\omega,u,v)|^2\,d\omega d\mathcal{W}, \quad (3.41)$$

as $n\to\infty$, in probability.

Therefore, combining (3.40) and (3.41) we get the required result on $D(\gamma)$. However, considering $\widehat{C}_0^{\gamma} \to \widehat{C}_0$ and $\widehat{D}_0^{\gamma} \to \widehat{D}_0$ as $\gamma \to 0$ and (3.36), the required result is now proved on $\mathbb{R}^2$. $\square$

# Chapter 4

# Testing Pairwise Independence for Multivariate Time Series by the Auto-Distance Correlation Matrix

## 4.1 Introduction

In this chapter, we introduce the notions of multivariate auto-distance covariance and auto-distance correlation functions. In fact, we extend them in a different direction by putting forward their matrix version. By doing so, possible interrelationships among the components of a multivariate time series are identified. Section 4.2 is devoted to the definition and interpretation of the distance covariance matrix, whereas in Section 4.3 we show that it can be consistently estimated. The empirical distance covariance matrix is used as a tool for developing a testing methodology for testing pairwise dependence in multivariate time series. The resulting test statistic is analogous to the multivariate Ljung-Box test statistic, and its asymptotic properties are derived in Section 4.4. To visually check independence for a multivariate time series, we construct a plot based on the distance correlation matrix, where the shown critical values are obtained simultaneously via a bootstrap methodology explained in Section 4.5. We conclude this chapter with several simulated and real data examples.

## 4.2 On Auto-Distance Covariance Matrix

### 4.2.1 Definitions

Consider a $d$-variate strictly stationary time series (see Definition 1.1.1), denoted by $\{\mathbf{X}_t, t \in \mathbb{Z}\}$ with components $\{X_{t;r}\}_{r=1}^{d}$ and suppose that its cumulative distribution function (c.d.f) is denoted by $F(x_1, x_2, \ldots, x_d)$. Let further $F_r(x)$ denote the marginal distribution of $\{X_{t;r}\}$ with $r = 1, 2, \ldots, d$. Suppose we have available a sample of size $n$, that is $\{\mathbf{X}_t, t = 1, 2, \ldots, n\}$. For the rest of the chapter we impose the following assumptions.

**Assumption 1** $\{\boldsymbol{X}_t\}$ is a strictly stationary and ergodic process.

**Assumption 2** $(i)$ $E\left|X_{t;r}\right| < \infty$, $\forall\ t$ and $\forall\ r = 1, 2, \ldots, d$. $(ii)$ $E\left|X_{t;r}\right|^2 < \infty$, $\forall\ t$ and $\forall\ r = 1, 2, \ldots, d$.

Assumption 1 is used for developing an asymptotic theory for the ADCV matrix defined in (4.4). Assumption 2($i$) guarantees the finiteness of the elements of (4.4). While Assumption 2($ii$) is used for proving the weak consistency of the ADCV matrix, its strong consistency is established by imposing - among others - Assumption 2($i$); details are discussed in the corresponding Section 4.3.

Extending the work of Székely et al. (2007), Zhou (2012) defined the distance covariance function for multivariate time series processes, but without taking into account possible cross-dependence relationships between all possible pairs of the component series of $\{\mathbf{X}_t\}$. Aiming to fill this gap, we define the pairwise distance covariance function as the distance between the joint characteristic function and the marginal characteristic functions of the pair $(X_{t;r}, X_{t-|j|;m})$, for $r, m = 1, 2, \ldots, d$. Denote the joint characteristic function of $X_{t;r}$ and $X_{t-|j|;m}$ by $\phi_{|j|}^{(r,m)}(u, v)$; that is

$$\phi_{|j|}^{(r,m)}(u, v) = E\left[\exp\Big(i(uX_{t;r} + vX_{t-|j|;m})\Big)\right], \qquad j = 0, \pm 1, \pm 2, \ldots,$$

where $(u,v) \in \mathbb{R}^2$, $r,m = 1,2,\ldots,d$ and $i^2=-1$. Furthermore, let

$$\phi^{(r)}(u) = E\left[\exp\Big(i(uX_{t;r})\Big)\right],$$

be the marginal characteristic function of $X_{t;r}$ for $r=1,2,\ldots,d$. Denote by

$$\Sigma_{|j|}(u,v) = \left[\sigma_j^{(r,m)}(u,v)\right]_{d\times d}, \qquad j=0,\pm1,\pm2,\ldots,$$

the $d\times d$ matrix whose $(r,m)$ element is $\sigma_j^{(r,m)}(u,v)$ which is simply the covariance function between $e^{iuX_{t;r}}$ and $e^{ivX_{t-|j|;m}}$; that is

$$\begin{aligned}\sigma_j^{(r,m)}(u,v) &= \operatorname{Cov}\Big(e^{iuX_{t;r}}, e^{ivX_{t-|j|;m}}\Big) \\ &= \phi_{|j|}^{(r,m)}(u,v) - \phi^{(r)}(u)\phi^{(m)}(v). \end{aligned} \tag{4.1}$$

It is easily seen from (4.1) that if $\sigma_j^{(r,m)}(u,v)=0\ \forall\ (u,v)\in\mathbb{R}^2$, then the random variables $X_{t;r}$ and $X_{t-|j|;m}$ are independent, for all $j\neq 0$. Let the $\|\cdot\|_{\mathcal{W}}$-norm of $\sigma_j^{(r,m)}(u,v)$ be defined by

$$\|\sigma_j^{(r,m)}(u,v)\|_{\mathcal{W}}^2 = \int_{\mathbb{R}^2}\left|\sigma_j^{(r,m)}(u,v)\right|^2 d\mathcal{W}(u,v), \qquad j=0,\pm1,\pm2,\ldots,$$

where $\mathcal{W}(\cdot,\cdot)$ is an arbitrary positive weight function for which the above integral exists. For instance, Székely et al. (2007) employed a nonintegrable weight function, of the form

$$\mathcal{W}(u,v) = \mathcal{W}_0(u)\mathcal{W}_0(v) = \frac{1}{\pi|u|^2}\frac{1}{\pi|v|^2}, \qquad (u,v)\in\mathbb{R}^2. \tag{4.2}$$

This choice of $\mathcal{W}(\cdot,\cdot)$ would be of central focus in this work. Obviously, (4.2) is a nonintegrable function in $\mathbb{R}^2$. However, other choices of $\mathcal{W}(\cdot,\cdot)$ with $\int d\mathcal{W}<\infty$ are possible. Davis et al. (2016) provided various choices of finite and infinite measures $\mathcal{W}(\cdot,\cdot)$ leading to alternative definitions of the distance covariance and correlation functions in a time series context. Moreover, following Hong (1999) and Chen and Hong (2012) we can assume that $\mathcal{W}(\cdot,\cdot):\mathbb{R}^2\to\mathbb{R}^+$ is nondecreasing with bounded total variation. Such an assumption obviously holds for the choice of $\mathcal{W}(u,v)=\Phi(u)\Phi(v)$, where $\Phi(\cdot)$ is the c.d.f of a standard normal random variable. In what follows we use (4.2) though. This choice is supported by the fact that integrable weight functions might miss potential dependence among observations (see Székely et al. (2007, p. 2771)).

**Definition 4.2.1** The *pairwise auto-distance covariance function* between $X_{t;r}$ and $X_{t-|j|;m}$ is denoted by $V_{rm}(j)$ and it is defined by the positive square root of

$$V^2_{rm}(j) \;=\; \|\sigma^{(r,m)}_j(u,v)\|^2_{\mathcal{W}}, \quad r,m=1,\ldots,d, \quad j=0,\pm1,\pm2,\ldots \tag{4.3}$$

with $\mathcal{W}(\cdot,\cdot)$ given by (4.2). The *auto-distance covariance matrix* of $\{\mathbf{X}_t\}$ at lag $j$ will be denoted by $V(j)$ and it is the $d\times d$ matrix

$$V(j) = \Big[V_{rm}(j)\Big]^d_{r,m=1}, \quad j=0,\pm1,\pm2,\ldots. \tag{4.4}$$

Clearly, $V^2_{rm}(j)\geq 0$, $\forall\ j$ and $X_{t;r}$ and $X_{t-|j|;m}$ are independent if and only if $V^2_{rm}(j)=0$. Furthermore, we define $d\times d$ matrices of the form

$$V^{(2)}(j) \;=\; \Big[V^2_{rm}(j)\Big]^d_{r,m=1}, \quad j=0,\pm1,\pm2,\ldots. \tag{4.5}$$

Based on the above definition, it is natural to define the auto-distance correlation matrix. We have the following definition.

**Definition 4.2.2** The *pairwise auto-distance correlation function* between $X_{t;r}$ and $X_{t-|j|;m}$ is denoted by $R_{rm}(j)$ and it is defined as the positive square root of

$$R^2_{rm}(j) \;=\; \frac{V^2_{rm}(j)}{\sqrt{V^2_{rr}(0)}\sqrt{V^2_{mm}(0)}}, \quad r,m=1,\ldots,d, \quad j=0,\pm1,\pm2,\ldots$$

provided that $V_{rr}(0)V_{mm}(0)\neq 0$. The *auto-distance correlation matrix* of $\{\mathbf{X}_t\}$ at lag $j$ will be denoted by $R(j)$ and it is the $d\times d$ matrix

$$R(j) \;=\; \Big[R_{rm}(j)\Big]^d_{r,m=1}, \quad j=0,\pm1,\pm2,\ldots.$$

Similarly, define the $d \times d$ matrices, say $R^{(2)}(j)$, by

$$R^{(2)}(j) = \left[ R^2_{rm}(j) \right]_{r,m=1}^{d}, \quad j = 0, \pm 1, \pm 2, \ldots .$$

Then (4.5) shows that

$$R^{(2)}(j) = D^{-1} V^{(2)}(j) D^{-1}$$

where $D = \mathrm{diag}\{\sqrt{V^2_{rr}(0)}, r = 1, 2, \ldots, d\}$. This is the usual formula we obtain for the relationship between the standard correlation and covariance matrix. All above population quantities exist and are well defined because of standard properties of the characteristic function.

$V_{rm}(j)$ measures the dependence of $X_{t;r}$ on $X_{t-|j|;m}$. In addition, if $V_{rm}(j) > 0$, we say that the series $X_{t;m}$ leads the series $X_{t;r}$ at lag $j$. There are two main properties of the pairwise distance covariance function when $j \neq 0$. Firstly, in general, $V_{rm}(j) \neq V_{mr}(j)$ for $r \neq m$, since they measure different dependence structure between the series $\{X_{t;r}\}$ and $\{X_{t;m}\}$ for all $r, m = 1, 2, \ldots, d$.

Based on the above discussion, the elements of distance covariance matrices $\{V(j), j = 0, \pm 1, \pm 2, \ldots\}$ can be interpreted as follows:

1. For all $j \in \mathbb{Z}$, the diagonal elements $\left(V_{rr}(j)\right)_{r=1}^{d}$ are the auto-distance covariance function of $\{X_{t;r}\}$ and they express the dependence among the pairs of observations of the series $\{X_{t;r}\}$.

2. The off-diagonal elements $\left(V_{rm}(0)\right)_{r=1}^{d}$ measure the concurrent dependence between $\{X_{t;r}\}$ and $\{X_{t;m}\}$. If $V_{rm}(0) > 0$, $\{X_{t;r}\}$ and $\{X_{t;m}\}$ are concurrently dependent.

3. For $j \in \mathbb{Z} - \{0\}$, $\left(V_{rm}(j)\right)_{r,m=1}^{d}$ measures the dependence of $\{X_{t;r}\}$ on the past values of $\{X_{t-|j|;m}\}$. Thus, if $V_{rm}(j) = 0$ for all $j \in \mathbb{Z} - \{0\}$, then $\{X_{t;r}\}$ does not depend on any past values of $\{X_{t;m}\}$.

4. For all $j \in \mathbb{Z}$, $V_{rm}(j) = V_{mr}(j) = 0$ implies that $\{X_{t;r}\}$ and $\{X_{t;m}\}$ are independent. Moreover, for all $j \in \mathbb{Z} - \{0\}$, if $V_{rm}(j) = 0$ and $V_{mr}(j) = 0$ then $\{X_{t;r}\}$ and $\{X_{t;m}\}$

have no lead-lag relationship.

5. If for all $j \in \mathbb{Z} - \{0\}$ $V_{rm}(j) = 0$ but there exists some $j \in \mathbb{Z} - \{0\}$ such that $V_{mr}(j) > 0$, then $\{X_{t;r}\}$ does not depend on the past values of $\{X_{t;m}\}$, but $\{X_{t;m}\}$ depends on some past values of $\{X_{t;r}\}$.

We now turn to the estimation problem.

### 4.2.2 Estimation

To develop an estimator of (4.3) and consequently an estimator of (4.4), define first

$$\hat{\sigma}_j^{(r,m)}(u,v) = \hat{\phi}_{|j|}^{(r,m)}(u,v) - \hat{\phi}^{(r)}(u)\hat{\phi}^{(m)}(v), \qquad j = 0, \pm 1, \pm 2, \ldots,$$

with

$$\hat{\phi}_{|j|}^{(r,m)}(u,v) = \frac{1}{n-|j|} \sum_{t=|j|+1}^{n} e^{i(uX_{t;r} + vX_{t-|j|;m})}.$$

Thus, the sample pairwise ADCV is defined by the positive square root of

$$\widehat{V}_{rm}^2(j) \;=\; \frac{1}{\pi^2} \int_{\mathbb{R}^2} \frac{\left|\hat{\sigma}_j^{(r,m)}(u,v)\right|^2}{|u|^2 |v|^2} du dv, \qquad j = 0, \pm 1, \pm 2, \ldots \tag{4.6}$$

To calculate (4.6), let $Y_{t;m} = X_{t-|j|;m}$. Then, based on the sample $\{(X_{t;r}, Y_{t;m}) : t = 1+|j|, \ldots, n\}$, we calculate the $(n-|j|)\times(n-|j|)$ Euclidean distance matrices $A^r = (A^r_{ts})$ and $B^m = (B^m_{ts})$ with elements

$$A^r_{ts} = a^r_{ts} - \bar{a}^r_{t.} - \bar{a}^r_{.s} + \bar{a}^r_{..},$$

with $\alpha^r_{ts} = |X_{t;r} - X_{s;r}|$, $b^m_{ts} = |Y_{t;m} - Y_{s;m}|$ and

$$\bar{\alpha}^r_{t.} \;=\; \frac{\sum_{s=1+|j|}^{n} a^r_{ts}}{(n-|j|)}, \quad \bar{\alpha}^r_{.s} = \frac{\sum_{t=1+|j|}^{n} a^r_{ts}}{(n-|j|)}, \quad \bar{\alpha}^r_{..} = \frac{\sum_{t=1+|j|}^{n}\sum_{s=1+|j|}^{n} a^r_{ts}}{(n-|j|)^2}.$$

Similarly, we define the quantities $\bar{b}^m_{t.}$, $\bar{b}^m_{.s}$, $\bar{b}^m_{..}$ and $B^m_{ts}$. Then, following the arguments of Chapter 3, we obtain that

$$\widehat{V}^2_{rm}(j) \quad = \quad \frac{1}{(n-|j|)^2} \sum_{t,s} A^r_{ts} B^m_{ts}. \tag{4.7}$$

By (4.5) define the $d \times d$ matrices

$$\widehat{V}^{(2)}(j) \quad = \quad \left[ \widehat{V}^2_{rm}(j) \right], \qquad j = 0, \pm 1, \pm 2, \ldots.$$

The sample ADCV matrix is given by

$$\widehat{V}(j) \quad = \quad \left[ \widehat{V}_{rm}(j) \right], \qquad j = 0, \pm 1, \pm 2, \ldots.$$

## 4.3 Asymptotic Properties of the Sample Distance Covariance Matrix

We first show that $\widehat{V}^2_{rm}(j)$ can be expressed as a $V$-statistic of order two. Recall that a $V$-statistic of order $p$ on the basis of a sample $\{X_t, t = 1, 2, \ldots, n\}$ of $\mathbb{R}^d$-valued random variables, where $d \geq 1$ and $n \geq p$, is defined by (see, for instance, Kowalski and Tu (2008, Chapters 3 and 5) and Serfling (1980, Section 5.1.2))

$$V_n \quad = \quad \frac{1}{n^p} \sum_{i_1=1}^{n} \cdots \sum_{i_p=1}^{n} h(X_{i_1}, \ldots, X_{i_p})$$

for any real valued measurable kernel function $h : \mathbb{R}^d \times \cdots \times \mathbb{R}^d \to \mathbb{R}$.

Let $u, u' \in \mathbb{R}$ and suppose, in general, that $X$ is a real valued random variable with c.d.f $F_X(\cdot)$. Then, define

$$m_X(u) := E\,|X-u| = \int_{\mathbb{R}} |x-u|\, dF_X(x),$$

$$\bar{m}_X := E[m_X(X)] = \int_{\mathbb{R}} m_X(u) dF_X(u) = \int_{\mathbb{R}} \left\{ \int_{\mathbb{R}} |x-u|\, dF_X(x) \right\} dF_X(u),$$

$$d_X(u,u') = |u-u'| - m_X(u) - m_X(u') + \bar{m}_X.$$

With some abuse of notation and setting $X \equiv X_{t;r}$ and $Y \equiv X_{t-|j|;m}$ we obtain that

$$V^2_{rm}(j) = E\Big[d_X(X_1,X_2)d_Y(Y_1,Y_2)\Big]$$

where $X$, $X_1$ and $X_2$ and $Y$, $Y_1$ and $Y_2$ are i.i.d. copies (Székely and Rizzo, 2013, p. 1262). Hence, we have defined a kernel $h : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}$ by

$$h(x,y;x',y') = d_X(x,x')d_Y(y,y'), \tag{4.8}$$

such that

$$V^2_{rm}(j) \;=\; \int_{\mathbb{R}^2}\int_{\mathbb{R}^2} h(x,y;x',y')dF(x,y)dF(x',y').$$

Thus, the empirical analogue of the parameter $V^2_{rm}(j)$ is a $V$-statistic of order two with kernel (4.8). Indeed,

$$\begin{aligned}
\widehat{V}^2_{rm}(j) &= \frac{1}{(n-|j|)^2}\sum_{t=1+|j|}^{n}\sum_{s=1+|j|}^{n} h\Big(X_{t;r},Y_{t;m};X_{s;r},Y_{s;m}\Big)\\
&= \frac{1}{(n-|j|)^2}\sum_{t=1+|j|}^{n}\sum_{s=1+|j|}^{n} d_X(X_{t;r},X_{s;r})d_Y(Y_{t;m},Y_{s;m})\\
&= \frac{1}{(n-|j|)^2}\sum_{t=1+|j|}^{n}\sum_{s=1+|j|}^{n} \Big\{|X_{t;r}-X_{s;r}| - m_X(X_{t;r}) - m_X(X_{s;r}) + \bar{m}_X\Big\}\\
&\quad \times\Big\{|Y_{t;m}-Y_{s;m}| - m_Y(Y_{t;m}) - m_Y(Y_{s;m}) + \bar{m}_Y\Big\}\\
&= \frac{1}{(n-|j|)^2}\sum_{t=1+|j|}^{n}\sum_{s=1+|j|}^{n}\\
&\quad \times\left\{|X_{t;r}-X_{s;r}| - \frac{\sum_s |X_{t;r}-X_{s;r}|}{n-|j|} - \frac{\sum_t |X_{t;r}-X_{s;r}|}{n-|j|} + \frac{\sum_t\sum_s |X_{t;r}-X_{s;r}|}{(n-|j|)^2}\right\}\\
&\quad \times\left\{|Y_{t;m}-Y_{s;m}| - \frac{\sum_s |Y_{t;m}-Y_{s;m}|}{n-|j|} - \frac{\sum_t |Y_{t;m}-Y_{s;m}|}{n-|j|} + \frac{\sum_t\sum_s |Y_{t;m}-Y_{s;m}|}{(n-|j|)^2}\right\}\\
&= \frac{1}{(n-|j|)^2}\sum_t\sum_s\Big\{\alpha^r_{ts} - \bar{\alpha}^r_{t.} - \bar{\alpha}^r_{.s} + \bar{\alpha}^r_{..}\Big\} \times \Big\{b^m_{ts} - \bar{b}^m_{t.} - \bar{b}^m_{.l} + \bar{b}^m_{..}\Big\}\\
&= \frac{1}{(n-|j|)^2}\sum_t\sum_s A^r_{ts}B^m_{ts}.
\end{aligned}$$

Note that the kernel function is a symmetric, continuous and positive semidefinite function. We now show that under pairwise independence, the $V$-statistic $\widehat{V}^2_{rm}(\cdot)$ is degenerate. First note that, under Assumption $2(i)$, Lemma 2.2.1 and Fubini's theorem, we obtain that (see also Székely and Rizzo (2013, p. 1261))

$$\begin{aligned} V^2_{rm}(j) &= E[h(X,Y;X',Y')] = E[d_X(X,X')d_Y(Y,Y')] \\ &= E\left|X-X'\right|\left|Y-Y'\right| + E\left|X-X'\right| E\left|Y-Y''\right| - 2E\left|X-X'\right|\left|Y-Y''\right|. \end{aligned} \quad (4.9)$$

where $X'$ is an i.i.d. copy of X and $Y'$ and $Y''$ are i.i.d. copies of $Y$. Indeed, recall that $V^2_{rm}(j)$ is a weighted integral defined by (4.3). Simple algebra gives that the numerator of the integral includes terms of the form

$$E\Big[\cos\left(u(X-X')\right)\cos\left(v(Y-Y')\right)\Big].$$

Applying the identity

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

employing the Fubini's theorem and Lemma 2.2.1 we get terms of the form

$$E\int_{\mathbb{R}^2} \frac{\Big[1-\cos\left(u(X-X')\right)\Big]\Big[1-\cos\left(v(Y-Y')\right)\Big]}{\pi^2\left|u\right|^2\left|v\right|^2} du dv = E\left|X-X'\right| E\left|Y-Y'\right|.$$

Employing similar steps for all terms obtained in the integral (4.3), we finally get (4.9). Using the result (4.9) and assuming that the data are pairwise independent, then

$$E[h(x,y;X,Y)] \;\; = \;\; 0,$$

which shows that $\widehat{V}^2_{rm}(j)$ is a degenerate $V$-statistic of order two.

Based on the above observation, the following proposition shows the strong consistency of the estimator $\widehat{V}(j)$. Proofs of all results are listed in the Appendix.

**Proposition 4.3.1** Let $\{\mathbf{X}_t\}$ be a $d$-variate process satisfying Assumptions 1 and $2(i)$ with distribution function $F(x_1, x_2, \ldots, x_d)$ and marginal distribution function $F_r(x)$ for

$r = 1, \ldots, d$. Then, for all $j = 0, \pm 1, \pm 2, \ldots$

$$\widehat{V}(j) \to V(j)$$

almost surely as $n \to \infty$.

**Remark 4.3.1** The above result follows by the proof of Székely et al. (2007) and Proposition 3.2.2 of Chapter 3. In particular, under strict stationarity, ergodicity and existence of first moments the strong consistency of the sample distance covariance can be established by considering individually the elements of $\widehat{V}^{(2)}(j)$ and then the elements of $\widehat{V}(j)$. We mention that our assumptions are minimal for proving this result. Related work by Borovcova et al. (1999) requires stationarity, ergodicity, existence of second moments, almost surely $(F_r \times F_m)$ continuity of $h(\cdot)$ and uniform integrability. Under these assumptions, it can be shown that $\widehat{V}(j)$ is a weakly consistent estimator of $V(j)$ (see Borovcova et al. (1999, Theorem 1) in connection to Aaronson et al. (1996, Proposition 2.8)). Furthermore, by dropping the continuity assumption and replacing it with $\beta$-mixing we obtain again the weak consistency of $\widehat{V}(j)$ (see Borovcova et al. (1999, Theorem 2) in connection to Aaronson et al. (1996, Proposition 2.8)).

The following theorem reveals the limiting distribution of the sample pairwise ADCV, $\widehat{V}^2_{rm}(j)$, when $\{\boldsymbol{X}_t\}$ is a pairwise independent sequence.

**Theorem 4.3.1** Suppose that Assumptions 1 and $2(i)$ hold. Then under pairwise independence and for fixed $j$

$$(n - |j|)\widehat{V}^2_{rm}(j) \quad \to \quad Z := \sum_k \lambda_k Z_k^2, \tag{4.10}$$

where $(Z_k)_k$ is an i.i.d. sequence of $N(0,1)$ random variables and $(\lambda_k)_k$ is a sequence of nonzero eigenvalues which satisfy the Hilbert - Schmidt equation

$$E\Big[h(x, y; X, Y)\Phi(X, Y)\Big] = \lambda\Phi(x, y),$$

where $h(\cdot)$ is a kernel defined by (4.8) and is represented as

$$h(x, y; x', y') = \sum_{k=1}^{\infty} \lambda_k \Phi_k(x, y)\Phi_k(x', y').$$

Here, $(\Phi_k)_k$ is the sequence of the corresponding orthonormal eigenfunctions (for more details see Leucht and Neumann (2013a,b)).

The above result follows from Leucht and Neumann (2013a, Theorem 1). It shows that the asymptotic distribution of $\widehat{V}^2_{rm}(j)$ is non standard when the hypothesis of interest is that of independence. In fact, we note the following associated problems:

1. Generally, it is of interest to approximate the asymptotic distribution of the matrix variate $U$-statistic $\{V^{(2)}(j), j = 0, \pm 1, \pm 2, \dots\}$, regardless whether the assumption of independence holds true. This is a problem which has not been addressed to the literature, to the best of our knowledge (see Chen (2016) for some recent work on this topic).

2. Under independence, it is of interest to form simultaneous confidence intervals for $V(j)$ as it is done in the case of the standard autocorrelation function. But (4.10) cannot be employed in applications and therefore some simulation based method should be applied. We discuss further this point in Section 4.5.2.

## 4.4 Testing Independence

To derive a test statistic for testing the null hypothesis of multivariate independence and investigate its asymptotic null distribution, we further impose the following conditions.

**Assumption 3** Suppose that $\{\mathbf{X}_t\}$ is an $\alpha$-mixing process with mixing coefficients $\{\alpha_j\}$ satisfying (i) $\sum_j \alpha_j < \infty$ or (ii) $\alpha(j) = O(j^{-2})$.

**Assumption 4** $K : \mathbb{R} \to [-1, 1]$ is symmetric and is continuous at 0 and all except a finite number of points, with $K(0) = 1$, $\int_{-\infty}^{\infty} K^2(z)dz < \infty$ and $|K(z)| \leq C\,|z|^{-b}$ for large $z$ and $b > 1/2$.

Assumption 3(i) ensures the existence of the generalized spectral density (4.12), whereas its second part, 3(ii), is the minimal condition for proving Lemmas 4.5.1 and 4.5.2 in the Appendix. Assumption 4 is standard for univariate kernel functions $K(\cdot)$, also stated in Chapter 3. It is used for obtaining kernel-based estimators for (4.12). A wide range of well known kernel functions with bounded (e.g. QS, Daniell) or unbounded support (e.g. Bartlett, Parzen) satisfy Assumption 4.

For establishing a multivariate distance covariance testing methodology, we first need to define the following matrix norm. In particular, for an $m \times n$ matrix $A$, recall that its Frobenius norm is defined as

$$\|A\|_F = \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n} |\alpha_{ij}|^2} = \sqrt{\mathrm{tr}\{A^* A\}},$$

where $A^*$ denotes the conjugate transpose of $A$, and $\mathrm{tr}\{A\}$ denotes the trace of the matrix $A$.

### 4.4.1 The Generalized Spectral Density Approach

Recall (4.1) and suppose that

$$\sup_{(u,v)\in\mathbb{R}^2} \sum_{j=-\infty}^{\infty} \left|\sigma_j^{(r,m)}(u,v)\right| < \infty, \tag{4.11}$$

which holds under Assumption 3. Thus, the sequence of covariance matrices $\{\Sigma_{|j|}(u,v), j = 0, \pm 1, \pm 2, \dots\}$ has absolutely summable components for all $(u,v) \in \mathbb{R}^2$. We can then define the Fourier transform of $\sigma_j^{(r,m)}(\cdot,\cdot)$ as

$$f^{(r,m)}(\omega,u,v) = \frac{1}{2\pi}\sum_{j=-\infty}^{\infty} \sigma_j^{(r,m)}(u,v) e^{-ij\omega}, \qquad \omega \in [-\pi,\pi]. \tag{4.12}$$

Under (4.11), $f^{(r,m)}(\cdot,\cdot,\cdot)$ is bounded and uniformly continuous. If $r = m$, then $f^{(r,m)}(\omega,u,v)$ is called the generalized spectrum or generalized spectral density of $X_{t;r}$ at frequency $\omega$ for all $(u,v) \in \mathbb{R}^2$. If $r \neq m$, then $f^{(r,m)}(\omega,u,v)$ is called the generalized cross-spectrum or generalized cross spectral density of $X_{t;r}$ and $X_{t;m}$ at frequency $\omega$ for all $(u,v) \in \mathbb{R}^2$

(Brillinger (1981, Section 7.1), Priestley (1981, Section 9.1) and Chen and Hong (2012) in our context). Collecting all elements defined by (4.12) in a $d \times d$ matrix, we obtain

$$\begin{aligned} F(\omega, u, v) &= \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} \Sigma_{|j|}(u, v) e^{-ij\omega} \\ &= \left[ f^{(r,m)}(\omega, u, v) \right]_{r,m=1}^{d} \end{aligned}$$

which is called the generalized spectral density matrix. Under the null hypothesis of independence, $F(\cdot, \cdot, \cdot)$ reduces to the matrix

$$F_0(\omega, u, v) = \frac{1}{2\pi} \left[ \sigma_0^{(r,m)}(u, v) \right]_{r,m=1}^{d}.$$

In general $F_0(\cdot, \cdot, \cdot)$ is not a diagonal matrix, but when $X_{t;r}$ and $X_{t;m}$ are independent for all $r, m = 1, 2, \ldots, d$ then $F_0(\cdot, \cdot, \cdot)$ reduces to a diagonal matrix. We consider the following class of kernel-density estimators,

$$\hat{f}^{(r,m)}(\omega, u, v) = \frac{1}{2\pi} \sum_{j=-(n-1)}^{(n-1)} (1 - |j|/n)^{1/2} K(j/p) \hat{\sigma}_j^{(r,m)}(u, v) e^{-ij\omega}, \quad \omega \in [-\pi, \pi],$$

where $p$ is a bandwidth parameter and $K(\cdot)$ is a univariate kernel function with the properties stated in Assumption 4.

Then, we can form the matrices

$$\widehat{F}(\omega, u, v) = \left[ \widehat{f}^{(r,m)}(\omega, u, v) \right]_{r,m=1}^{d}$$

and

$$\widehat{F}_0(\omega, u, v) = \frac{1}{2\pi} \left[ \hat{\sigma}_0^{(r,m)}(u, v) \right]_{r,m=1}^{d}$$

respectively. We then consider the squared $L_2$-distance between $\widehat{F}(\cdot, \cdot, \cdot)$ and $\widehat{F}_0(\cdot, \cdot, \cdot)$ by the following

$$\begin{aligned} L_2^2\Big(\widehat{F}(\omega,u,v),\widehat{F}_0(\omega,u,v)\Big) &= \int_{\mathbb{R}^2}\int_{-\pi}^{\pi}\|\widehat{F}(\omega,u,v)-\widehat{F}_0(\omega,u,v)\|_F^2 d\omega d\mathcal{W}(u,v) \\ &= \int_{\mathbb{R}^2}\int_{-\pi}^{\pi}\mathrm{tr}\Big\{\Big(\widehat{F}(\omega,u,v)-\widehat{F}_0(\omega,u,v)\Big)^* \\ &\qquad \times\Big(\widehat{F}(\omega,u,v)-\widehat{F}_0(\omega,u,v)\Big)\Big\}d\omega d\mathcal{W}(u,v). \end{aligned}$$

Working analogously as in the univariate case for obtaining the similar result (3.11), we get the following

$$L_2^2\Big(\widehat{F}(\omega,u,v),\widehat{F}_0(\omega,u,v)\Big) = \frac{2}{\pi}\sum_{r,m}\sum_{j=1}^{n-1}(1-j/n)K^2(j/p)\int_{\mathbb{R}^2}\Big|\hat{\sigma}_j^{(r,m)}(u,v)\Big|^2 d\mathcal{W}(u,v),$$

for any suitably weighting function $\mathcal{W}(\cdot,\cdot)$ - see the discussion after equation (4.2). In particular, employing (4.2) yields to

$$\begin{aligned} L_2^2\Big(\widehat{F}(\omega,u,v),\widehat{F}_0(\omega,u,v)\Big) &= \frac{2}{\pi}\sum_{r,m}\sum_{j=1}^{n-1}(1-j/n)K^2(j/p)\widehat{V}_{rm}^2(j) \\ &= \frac{2}{\pi}\sum_{j=1}^{n-1}(1-j/n)K^2(j/p)\mathrm{tr}\{\widehat{V}^*(j)\widehat{V}(j)\}. \end{aligned} \tag{4.13}$$

Equation (4.13) can be formed in terms of the distance correlation matrix by working analogously. Indeed, recall $D = \mathrm{diag}\{\sqrt{V_{rr}^2(0)}, r = 1,2,\dots,d\}$ and define the $d \times d$ matrix

$$R_{|j|}(u,v) = D^{-1/2}\Sigma_{|j|}(u,v)D^{-1/2}$$

with elements

$$\rho_j^{(r,m)}(u,v) = \frac{\sigma_j^{(r,m)}(u,v)}{\sqrt{V_{rr}(0)}\sqrt{V_{mm}(0)}}.$$

By recalling (4.11), we can define the Fourier transform of $\rho_j^{(r,m)}(\cdot,\cdot)$ by

$$g^{(r,m)}(\omega,u,v) = \frac{1}{2\pi}\sum_{j=-\infty}^{\infty}\rho_j^{(r,m)}(u,v)e^{-ij\omega}, \qquad \omega \in [-\pi,\pi],$$

which form the $d \times d$ matrix

$$\begin{aligned} G(\omega, u, v) &= \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} R_{|j|}(u, v) e^{-ij\omega} \\ &= \Big[ g^{(r,m)}(\omega, u, v) \Big]_{r,m=1}^{d}. \end{aligned}$$

Under independence, $G(\cdot, \cdot, \cdot)$ reduces to

$$G_0(\omega, u, v) = \frac{1}{2\pi} \Big[ \rho_0^{(r,m)}(u, v) \Big]_{r,m=1}^{d}.$$

An analogous to (6.3) kernel-density estimator of $g^{(r,m)}(\cdot, \cdot)$ is given by

$$\hat{g}^{(r,m)}(\omega, u, v) = \frac{1}{2\pi} \sum_{j=-(n-1)}^{(n-1)} (1 - |j| / n)^{1/2} K(j/p) \hat{\rho}_j^{(r,m)}(u, v) e^{-ij\omega}, \quad \omega \in [-\pi, \pi].$$

We can then define the estimators of $G(\cdot, \cdot, \cdot)$ and $G_0(\cdot, \cdot, \cdot)$ by

$$\widehat{G}(\omega, u, v) = \Big[ \hat{g}^{(r,m)}(\omega, u, v) \Big]_{r,m=1}^{d}$$

and

$$\widehat{G}_0(\omega, u, v) = \frac{1}{2\pi} \Big[ \hat{\rho}_0^{(r,m)}(u, v) \Big]_{r,m=1}^{d}$$

respectively. Considering now the squared $L_2$-distance between $G(\cdot, \cdot, \cdot)$ and $G_0(\cdot, \cdot, \cdot)$ we get

$$\begin{aligned} L_2^2\Big(\widehat{G}(\omega, u, v), \widehat{G}_0(\omega, u, v)\Big) &= \int_{\mathbb{R}^2} \int_{-\pi}^{\pi} \|\widehat{G}(\omega, u, v) - \widehat{G}_0(\omega, u, v)\|_F^2 d\omega d\mathcal{W}(u, v) \\ &= \int_{\mathbb{R}^2} \int_{-\pi}^{\pi} \operatorname{tr}\Bigg\{ \Big(\widehat{G}(\omega, u, v) - \widehat{G}_0(\omega, u, v)\Big)^* \\ &\qquad \times \Big(\widehat{G}(\omega, u, v) - \widehat{G}_0(\omega, u, v)\Big) \Bigg\} d\omega d\mathcal{W}(u, v). \end{aligned}$$

After some calculations and choosing the weighting function defined in (4.2) we find that (compare to the multivariate Ljung-Box statistic given in (2.12))

$$\begin{aligned}
L_2^2\Big(\widehat{G}(\omega,u,v),\widehat{G}_0(\omega,u,v)\Big) &= \frac{2}{\pi}\sum_{r,m}\sum_{j=1}^{n-1}(1-j/n)K^2(j/p)\int_{\mathbb{R}^2}\Big|\widehat{\rho}_j^{(r,m)}(u,v)\Big|^2 d\mathcal{W}(u,v) \\
&= \frac{2}{\pi}\sum_{r,m}\sum_{j=1}^{n-1}(1-j/n)K^2(j/p)\frac{\widehat{V}_{rm}^2(j)}{\sqrt{\widehat{V}_{rr}^2(0)}\sqrt{\widehat{V}_{mm}^2(0)}} \\
&= \frac{2}{\pi}\sum_{r,m}\sum_{j=1}^{n-1}(1-j/n)K^2(j/p)\widehat{R}_{rm}^2(j) \\
&= \frac{2}{\pi}\sum_{j=1}^{n-1}(1-j/n)K^2(j/p)\mathrm{tr}\Big\{\widehat{R}^*(j)\widehat{R}(j)\Big\} \\
&= \frac{2}{\pi}\sum_{j=1}^{n-1}(1-j/n)K^2(j/p) \\
&\quad\times\mathrm{tr}\Big\{[\widehat{D}^{-1/2}\widehat{V}(j)\widehat{D}^{-1/2}]^*\widehat{D}^{-1/2}\widehat{V}(j)\widehat{D}^{-1/2}\Big\} \\
&= \frac{2}{\pi}\sum_{j=1}^{n-1}(1-j/n)K^2(j/p)\mathrm{tr}\Big\{\widehat{V}^*(j)\widehat{D}^{-1}\widehat{V}(j)\widehat{D}^{-1}\Big\}. \quad (4.14)
\end{aligned}$$

Equations (4.13) and (4.14) motivate our study of multivariate tests of independence. In particular, it is of interest to test the null hypothesis that the vector series $\{\mathbf{X}_t\}$ is i.i.d. regardless of the possible dependence between time series components $\{X_{t;r}\}$ for $r = 1, 2, \ldots, d$. Equation (4.14) can be viewed as a multivariate Ljung-Box type statistic based on the distance covariance matrix rather than on Pearson covariance matrix. Indeed, choosing the truncated periodogram window, that is $K(z) = 1$ for $|z| \leq 1$ and 0 otherwise, equation (4.14) becomes

$$L_2^2\Big(\widehat{G}(\omega,u,v),\widehat{G}_0(\omega,u,v)\Big) = \frac{2}{\pi}\sum_{j=1}^{p}(1-j/n)\mathrm{tr}\{\widehat{V}^*(j)\widehat{D}^{-1}\widehat{V}(j)\widehat{D}^{-1}\}$$

which can be viewed as a multivariate Ljung-Box type statistic for testing that $V(j) = 0$, $j = 1, 2, \ldots, p$, because $(1 - j/n)$ can be replaced by unity.

Define

$$T_n^{(r,m)} = \sum_{j=1}^{n-1}(n-j)K^2(j/p)\widehat{V}_{rm}^2(j).$$

The test statistic motivated by (4.13) is based on

$$\widetilde{T}_n = \sum_{r,m} T_n^{(r,m)} = \sum_{j=1}^{n-1} (n-j)K^2(j/p)\mathrm{tr}\{\widehat{V}^*(j)\widehat{V}(j)\}.$$

Similarly, we can consider

$$\overline{T}_n = \sum_{r,m} \frac{T_n^{(r,m)}}{\sqrt{\widehat{V}_{rr}^2(0)}\sqrt{\widehat{V}_{mm}^2(0)}} = \sum_{j=1}^{n-1} (n-j)K^2(j/p)\mathrm{tr}\{\widehat{V}^*(j)\widehat{D}^{-1}\widehat{V}(j)\widehat{D}^{-1}\}.$$

using (4.14). We have the following results:

**Theorem 4.4.1** Suppose that Assumption 4 is true and let $p = cn^\lambda$, where $c > 0$, $\lambda \in (0,1)$. Then, if $\{\boldsymbol{X}_t\}$ is an i.i.d. sequence, we have that

$$M_n^{(r,m)} = \frac{T_n^{(r,m)} - \widehat{C}_0^{(r,m)} \sum_j K^2(j/p)}{\left[\widehat{D}_0^{(r,m)} \sum_j K^4(j/p)\right]^{1/2}} \rightarrow N(0,1)$$

as $n \to \infty$ in distribution, where

$$C_0^{(r,m)} = \int_{\mathbb{R}^2} \sigma_0^{(r,r)}(u,-u)\sigma_0^{(m,m)}(v,-v)d\mathcal{W}(u,v),$$

$$D_0^{(r,m)} = 2\int_{\mathbb{R}^4} \left|\sigma_0^{(r,r)}(u,u')\sigma_0^{(m,m)}(u,u')\right|^2 d\mathcal{W}(u,v)d\mathcal{W}(u',v') = 2V_{rr}^2(0)V_{mm}^2(0),$$

and $\widehat{C}_0^{(r,m)}$, $\widehat{D}_0^{(r,m)}$ are their sample counterparts.

The previous theorem implies the following results.

**Corollary 4.4.1** Suppose that Assumption 4 is true and let $p = cn^\lambda$, where $c > 0$, $\lambda \in (0,1)$. Then, under the null hypothesis that $\{\boldsymbol{X}_t\}$ is an i.i.d. sequence, we have that

$$\widetilde{M}_n \equiv \frac{\widetilde{T}_n - \sum_{r,m} \widehat{C}_0^{(r,m)} \sum_{j=1}^{n-1} K^2(j/p)}{\left[\sum_{r,m} \widehat{D}_0^{(r,m)} \sum_{j=1}^{n-2} K^4(j/p)\right]^{1/2}} \rightarrow N(0,1)$$

as $n \to \infty$ in distribution.

**Corollary 4.4.2** Suppose that Assumption 4 is true and let $p = cn^\lambda$, where $c > 0$,

$\lambda \in (0,1)$. Then, under the null hypothesis that $\{\boldsymbol{X}_t\}$ is an i.i.d. sequence, we have that

$$\overline{M}_n \equiv \frac{\overline{T}_n - \sum_{r,m} \widehat{\underline{C}}_0^{(r,m)} \sum_{j=1}^{n-1} K^2(j/p)}{d\Big[2\sum_{j=1}^{n-2} K^4(j/p)\Big]^{1/2}} \quad \to \quad N(0,1)$$

as $n \to \infty$ in distribution, where

$$\underline{C}_0^{(r,m)} = \frac{C_0^{(r,m)}}{V_{rr}(0)V_{mm}(0)},$$

and $\widehat{\underline{C}}_0^{(r,m)}$ is the corresponding empirical analogue.

The following result illustrates the consistency of the test statistics.

**Theorem 4.4.2** Suppose that Assumptions 1, 3(i) and 4 hold and $p = cn^\lambda$ for $c > 0$ and $\lambda \in (0,1)$. Then,

$$\frac{\sqrt{p}}{n}\widetilde{M}_n \quad \to \quad \frac{\frac{\pi}{2}L_2^2\Big(F(\omega,u,v), F_0(\omega,u,v)\Big)}{\Big[\sum_{r,m} D_0^{(r,m)} \int_0^\infty K^4(z)dz\Big]^{1/2}}$$

and

$$\frac{\sqrt{p}}{n}\overline{M}_n \quad \to \quad \frac{\frac{\pi}{2}L_2^2\Big(G(\omega,u,v), G_0(\omega,u,v)\Big)}{d\Big[2\int_0^\infty K^4(z)dz\Big]^{1/2}},$$

as $n \to \infty$, in probability.

The above results depend on the value of the bandwidth parameter $p$ and the sample size $n$. We will not discuss the issue of choosing the bandwidth parameter but our limited experience shows that choosing roughly $p \geq 15$ for a sample size of $n = 500$ yields a better asymptotic approximation. Because we deal with a testing problem, it is preferable in applications to vary the value of $p$ and then examine closely the sensitivity of the results. For small $n$, we suggest the use of simulation based techniques to approximate the distribution of $\widetilde{T}_n$ (or $\overline{T}_n$). This topic is discussed next.

## 4.5 Computation of Test Statistic With Applications

### 4.5.1 Bootstrap Methodology

To examine the empirical behavior of the proposed test statistic we present a limited simulation study. We suggest a resampling method to approximate the asymptotic distribution of $\widetilde{T}_n$ (equivalently $\overline{T}_n$). Recalling Section 4.3, the test statistics $\widetilde{T}_n$ (or $\overline{T}_n$) are functions of degenerate $V$-statistic of order two. Dehling and Mikosch (1994) proposed wild bootstrap techniques to approximate the distribution of degenerate $U$-statistics for the case of i.i.d. data. In a recent contribution Leucht and Neumann (2013a,b) proposed the use of a novel scheme of dependent wild bootstrap (Shao, 2010) to approximate the distribution of degenerate $U$- and $V$-statistics calculated for dependent data. More specifically, the method relies on generating auxiliary random variables $(W^*_{tn})_{t=1}^{n-j}$ and compute the bootstrap realizations of $\widehat{V}^2_{rm}(j)$ as

$$\widehat{V}^{2*}_{rm}(j) = \frac{1}{(n-j)^2}\sum_{t,s=1}^{n-j} W^*_{tn} h(X_{t;r}, Y_{t;m}; X_{s;r}, Y_{s;m}) W^*_{sn}$$

where $h(\cdot,\cdot)$ is defined by (4.8), for $r, m = 1, 2, \ldots, d$ and $j = 1, 2, \ldots, n-1$. Then, the bootstrap realization of $\widetilde{T}_n$ is computed by

$$\widetilde{T}^*_n = \sum_{j=1}^{n-1}(n-j)K^2(j/p)\sum_{r,m}\widehat{V}^{2*}_{rm}(j).$$

To test for independence, we repeat the above steps $b$ times to obtain $\widetilde{T}^*_{n,1}, \widetilde{T}^*_{n,2}, \ldots, \widetilde{T}^*_{n,b}$ and then approximate the $p$-value of the test statistic by

$$p - \text{value} := \frac{1}{b+1}\Big(\sum_{i=1}^{b}\mathbb{I}\{\widetilde{T}^*_{n,i} \geq \widetilde{T}_n\}\Big),$$

where $\mathbb{I}(\cdot)$ denotes the indicator function. We work analogously when employing $\overline{T}_n$; details are omitted. Shao (2010) highlighted that the methodology of wild bootstrap for time series extends that of Wu (1986) by allowing the auxiliary random variables $W^*_{tn}$ to be dependent. In fact, Leucht and Neumann (2013b) proposed to generate the sequence $W^*_{tn}$ by a first order autoregressive model. In the case of independent data, Dehling and Mikosch (1994) studied the limit distribution of degenerate $U$-statistic based on independent auxiliary variable $W^*_{tn}$. Because we test independence, we generate $W^*_{tn}$ as i.i.d. standard normal variables.

### 4.5.2 Obtaining Simultaneous Critical Values for the ADCF Plots

It is customary in time series analysis to plot the ordinary autocorrelation function together with simultaneous confidence intervals for checking the white noise assumption. The critical values employed for obtaining confidence intervals are deduced by the asymptotic normality of the vector which consists of the first $q$ sample autocorrelations under the white noise assumption (Brockwell and Davis, 1991, Theorem 7.2.1). Our aim is to illustrate a similar plot but in terms of the distance correlation function. This task is quite complicated though since if we form the vector of the first $q$ sample distance correlation function then this consists of $V$-statistics (which under the hypothesis of pairwise independence are degenerate). Theorem 4.3.1 makes this point precise. To overcome this difficulty we resort to Monte Carlo simulation. We explain the steps in what follows.

We first note that critical values chosen by the wild bootstrap maintain asymptotically the nominal size of a given test statistic for testing a hypothesis of interest. Given $b$ bootstrap realizations of $\widehat{R}_{rm}(j)$ say $\{\widehat{R}^*_{rm,i}(j), i = 1, 2, \ldots, b\}$ we compute the $p$-value

$$p_{rm}(j) \quad = \quad \frac{1}{b+1}\Big(\sum_{i=1}^{b} \mathbb{I}\{\widehat{R}^*_{rm,i}(j) \geq \widehat{R}_{rm}(j)\}\Big).$$

Then considering $\{p_{rm}(j), j = 1, 2, \ldots, q\}$, we note that these correspond to the $p$-values obtained by testing hypotheses $R_{rm}(j) = 0$, $j = 1, 2, \ldots, q$. Because this is a multiple testing situation, we adjust the $p$-values to obtain a new set $\{\tilde{p}_{rm}(j), j = 1, 2, \ldots, q\}$ using the False Discovery Rate method suggested by Benjamini and Hochberg (1995), at some prespecified level $\alpha$. Based on the adjusted $p$-values we obtain critical points $\{c_{rm}(j), j = 1, 2, \ldots, q\}$ which satisfy

$$\tilde{p}_{rm}(j) \quad = \quad \frac{\#\{\widehat{R}^*_{rm,i}(j) \geq c_{rm}(j)\}}{b}, \qquad j = 1, 2, \ldots, q,$$

Table 4.1: Simultaneous empirical critical values at a significance level $\alpha = 0.05$, for various sample sizes and dimensions. Results are based on $b = 499$ bootstrap replications and 100 simulations.

| | N(0,1) | Pois(4) | Gamma(1,1) | Beta(2,3) | U(1,1) | $X_4^2$ | Exp(1) | $\rho = 0.9$ | $\rho = 0.4$ |
|---|---|---|---|---|---|---|---|---|---|
| $n = 500$ | | | | | | | | | |
| $d = 2$ | 0.116 | 0.117 | 0.118 | 0.113 | 0.109 | 0.117 | 0.114 | 0.117 | 0.116 |
| $d = 3$ | 0.117 | 0.116 | 0.130 | 0.115 | 0.111 | 0.119 | 0.119 | 0.119 | 0.117 |
| $d = 4$ | 0.118 | 0.117 | 0.132 | 0.118 | 0.112 | 0.121 | 0.118 | 0.119 | 0.119 |
| $d = 5$ | 0.121 | 0.116 | 0.129 | 0.114 | 0.113 | 0.122 | 0.122 | 0.116 | 0.117 |
| $n = 600$ | | | | | | | | | |
| $d = 2$ | 0.106 | 0.105 | 0.102 | 0.103 | 0.100 | 0.105 | 0.106 | 0.106 | 0.106 |
| $d = 3$ | 0.107 | 0.105 | 0.106 | 0.106 | 0.102 | 0.112 | 0.107 | 0.108 | 0.106 |
| $d = 4$ | 0.108 | 0.106 | 0.104 | 0.104 | 0.104 | 0.108 | 0.106 | 0.105 | 0.107 |
| $d = 5$ | 0.109 | 0.107 | 0.106 | 0.102 | 0.106 | 0.110 | 0.107 | 0.110 | 0.107 |
| $n = 700$ | | | | | | | | | |
| $d = 2$ | 0.098 | 0.095 | 0.099 | 0.096 | 0.092 | 0.098 | 0.099 | 0.097 | 0.098 |
| $d = 3$ | 0.098 | 0.097 | 0.098 | 0.097 | 0.092 | 0.103 | 0.099 | 0.099 | 0.099 |
| $d = 4$ | 0.100 | 0.099 | 0.096 | 0.097 | 0.095 | 0.101 | 0.099 | 0.099 | 0.099 |
| $d = 5$ | 0.099 | 0.097 | 0.098 | 0.098 | 0.093 | 0.100 | 0.103 | 0.099 | 0.098 |
| $n = 800$ | | | | | | | | | |
| $d = 2$ | 0.091 | 0.090 | 0.092 | 0.091 | 0.088 | 0.092 | 0.090 | 0.095 | 0.092 |
| $d = 3$ | 0.091 | 0.093 | 0.093 | 0.090 | 0.089 | 0.095 | 0.092 | 0.094 | 0.094 |
| $d = 4$ | 0.093 | 0.091 | 0.095 | 0.090 | 0.087 | 0.093 | 0.093 | 0.092 | 0.094 |
| $d = 5$ | 0.092 | 0.092 | 0.095 | 0.090 | 0.089 | 0.097 | 0.094 | 0.092 | 0.092 |
| $n = 900$ | | | | | | | | | |
| $d = 2$ | 0.087 | 0.084 | 0.088 | 0.085 | 0.080 | 0.086 | 0.085 | 0.086 | 0.084 |
| $d = 3$ | 0.086 | 0.085 | 0.089 | 0.084 | 0.081 | 0.088 | 0.086 | 0.087 | 0.088 |
| $d = 4$ | 0.087 | 0.086 | 0.090 | 0.085 | 0.083 | 0.089 | 0.086 | 0.088 | 0.088 |
| $d = 5$ | 0.087 | 0.086 | 0.093 | 0.085 | 0.084 | 0.087 | 0.086 | 0.087 | 0.087 |
| $n = 1000$ | | | | | | | | | |
| $d = 2$ | 0.081 | 0.080 | 0.082 | 0.081 | 0.079 | 0.081 | 0.081 | 0.083 | 0.082 |
| $d = 3$ | 0.081 | 0.081 | 0.084 | 0.082 | 0.080 | 0.083 | 0.083 | 0.083 | 0.084 |
| $d = 4$ | 0.083 | 0.081 | 0.084 | 0.081 | 0.081 | 0.082 | 0.083 | 0.082 | 0.082 |
| $d = 5$ | 0.083 | 0.081 | 0.085 | 0.083 | 0.079 | 0.084 | 0.083 | 0.084 | 0.083 |

where $\#\{A\}$ denotes the number of times the event $A$ occurs. The horizontal line in the plots (see for instance Figure 1.1) corresponds to $c = \max_{r,m,j} c_{rm}(j)$. This is somehow a conservative approach but guarantees that all simultaneous confidence intervals are at a given level $\alpha$.

Table 4.1 illustrates that critical values obtained under independence are not sensitive to the choice of response distribution or dimension but depend on the sample size, as it should be expected. The first seven columns of Table 4.1 have been obtained by considering univariate data. The last two columns correspond to independent samples drawn from a $d$-dimensional normal distribution with mean zero and equicorrelation matrix $\boldsymbol{\rho}$ with $\rho_{ij} = \rho$, $i \neq j$. We set $\rho = 0.9$ and 0.4 respectively.

### 4.5.3 Monte Carlo Simulation Results

The empirical results presented in this section are based on the test statistic $\overline{T}_n$; see Corollary 4.4.2. Simulations were run for $n = 500, 1000$ and the bootstrap procedure described in Section 4.5.1 was applied to calculate the power and the nominal level of the test.

The test statistic $\overline{T}_n$ is calculated by using univariate kernel functions $K(\cdot)$ which are Lipschitz continuous; that is for any $z_1, z_2 \in \mathbb{R}$

$$|K(z_1) - K(z_2)| \leq C\,|z_1 - z_2|\,,$$

for some constant $C$. We use the Daniell kernel (DAN), the Parzen kernel (PAR) and the Bartlett kernel (BAR) whose definition is given in Section 3.5 and satisfy the above condition. Although Lipschitz condition rules out the truncated kernel (TRUNC)

$$K(z) = \begin{cases} 1, & |z| \leq 1, \\ 0, & \text{otherwise}, \end{cases}$$

results based on this kernel are also provided. We compare the performance of $\overline{T}_n$ to the performance of multivariate Ljung-Box statistic (Hosking, 1980; Li and McLeod, 1981) given in (2.12). We re-state it here for the reader's convenience:

$$mLB = n^2 \sum_{j=1}^{p} (n-j)^{-1} \mathrm{trace}\{\widehat{\Gamma}'(j)\widehat{\Gamma}^{-1}(0)\widehat{\Gamma}(j)\widehat{\Gamma}^{-1}(0)\}.$$

We first investigate the size of the proposed test under $H_0$. We consider a sample of bivariate standard normal time series. Table 4.2 reports the achieved level of all test statistics at 5% and 10% nominal levels. The results indicate that the proposed test statistics approximate the nominal levels quite adequately.

Table 4.2: Achieved type I error of the test statistics for testing the hypothesis that the data are i.i.d. The data are generated by the bivariate standard normal distribution. Achieved significance levels are given in percentages. The value of bandwidth $p$ is chosen by $p = [3n^{\lambda}]$, $\lambda = 0.1$, 0.2 and 0.3. The results are based on $b = 499$ bootstrap replications and 100 simulations.

| $n$ : | | 500 | | | | | | 1000 | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $p$ : | | 6 | | 11 | | 20 | | 6 | | 12 | | 24 | |
| | | 10% | 5% | 10% | 5% | 10% | 5% | 10% | 5% | 10% | 5% | 10% | 5% |
| $\overline{T}_n$ | BAR | 10 | 4 | 11 | 4 | 13 | 5 | 6 | 3 | 4 | 1 | 12 | 6 |
| | TRUNC | 9 | 3 | 14 | 7 | 8 | 7 | 8 | 3 | 8 | 5 | 5 | 3 |
| | PAR | 9 | 6 | 7 | 3 | 11 | 5 | 8 | 3 | 11 | 6 | 12 | 6 |
| | DAN | 7 | 2 | 5 | 2 | 3 | 0 | 7 | 2 | 12 | 7 | 11 | 5 |
| $mLB$ | | 10 | 6 | 8 | 4 | 11 | 7 | 7 | 3 | 10 | 7 | 9 | 6 |

Furthermore Table 4.3 presents some empirical evidence of the asymptotic normality of $\overline{T}_n$.

Table 4.3: Skewness, kurtosis and $p$-values obtained by performing a one-sample Kolmogorov-Smirnov test, for testing normality of the normalized test statistic $\overline{M}_n$ given by Corollary 4.4.2. The results are based on $b = 499$ bootstrap replications and 100 simulations.

| | Skewness | | | | Kurtosis | | | | $p$-value | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | BAR | PAR | DAN | TRUNC | BAR | PAR | DAN | TRUNC | BAR | PAR | DAN | TRUNC |
| $n = 500$ | | | | | | | | | | | | |
| $p = 6$ | -0.043 | -0.119 | -0.026 | -0.058 | 2.010 | 2.619 | 3.263 | 2.569 | 0.590 | 0.997 | 0.998 | 0.957 |
| $p = 11$ | 0.506 | 0.383 | 0.116 | 0.005 | 3.472 | 2.668 | 3.110 | 3.392 | 0.489 | 0.851 | 0.999 | 0.994 |
| $p = 20$ | -0.127 | 0.377 | -0.066 | 0.047 | 2.550 | 2.915 | 2.588 | 2.546 | 0.950 | 0.502 | 0.943 | 0.983 |
| $n = 1000$ | | | | | | | | | | | | |
| $p = 6$ | -0.234 | -0.045 | 0.054 | -0.092 | 2.313 | 2.338 | 2.442 | 2.692 | 0.703 | 0.930 | 0.877 | 0.985 |
| $p = 12$ | 0.110 | -0.233 | -0.285 | 0.357 | 3.223 | 2.836 | 2.768 | 3.181 | 0.469 | 0.972 | 0.868 | 0.917 |
| $p = 24$ | -0.233 | -0.239 | 0.691 | 0.425 | 2.398 | 3.101 | 3.744 | 3.164 | 0.665 | 0.955 | 0.546 | 0.535 |

To investigate the power of the proposed test, we consider the bivariate NMA(2) process discussed in the Introduction of the thesis and given by (1.1), and the following data generating processes:

- Bivariate ARMA(1,1)-model

$$\begin{bmatrix} X_{t;1} \\ X_{t;2} \end{bmatrix} - \begin{bmatrix} 0.2 & -0.6 \\ 0.3 & 1 \end{bmatrix} \begin{bmatrix} X_{t-1;1} \\ X_{t-1;2} \end{bmatrix} = \begin{bmatrix} \epsilon_{t;1} \\ \epsilon_{t;2} \end{bmatrix} - \begin{bmatrix} -0.5 & 0 \\ 0 & -0.6 \end{bmatrix} \begin{bmatrix} \epsilon_{t-1;1} \\ \epsilon_{t-1;2} \end{bmatrix} \quad (4.15)$$

where $\{\epsilon_t\}$ is a bivariate sequence of serially uncorrelated random vectors with mean zero and covariance matrix $\Sigma = \begin{bmatrix} 4 & 0.8 \\ 0.8 & 1 \end{bmatrix}$.

- Bivariate GARCH(1,1)-model

$$X_{t;i} = h_{t;i}^{1/2} \epsilon_{t;i}, \qquad i = 1, 2 \tag{4.16}$$

where

$$\begin{bmatrix} h_{t;1} \\ h_{t;2} \end{bmatrix} = \begin{bmatrix} 0.003 \\ 0.005 \end{bmatrix} + \begin{bmatrix} 0.2 & 0.1 \\ 0.1 & 0.3 \end{bmatrix} \begin{bmatrix} X_{t-1;1}^2 \\ X_{t-1;2}^2 \end{bmatrix} - \begin{bmatrix} 0.4 & 0.05 \\ 0.05 & 0.5 \end{bmatrix} \begin{bmatrix} h_{t-1;1} \\ h_{t-1;2} \end{bmatrix}$$

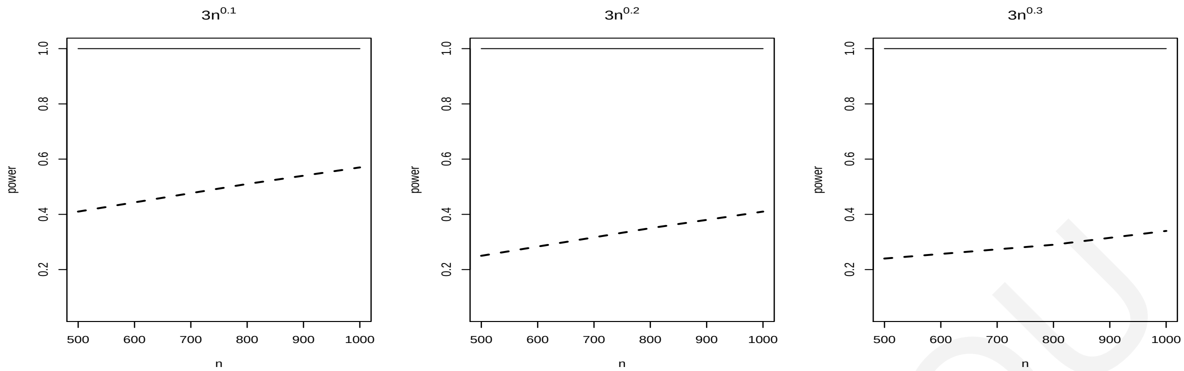
and $\{\epsilon_t\}$ is a bivariate sequence of uncorrelated random vectors with mean zero and unconditional correlation matrix $R = \begin{bmatrix} 1 & 0.4 \\ 0.4 & 1 \end{bmatrix}$.

Figure 4.1 shows the power of both $\overline{T}_n$ and $mLB$ statistics considered for various sample sizes and bandwidth parameters when the data are generated by models (1.1), (4.15) and (4.16) respectively. Clearly, when the data are generated by the nonlinear models NMA(2) and GARCH(1,1) (Figures 4.1a and 4.1c) $\overline{T}_n$ performs better, whereas in the case of the bivariate ARMA(1,1) (Figure 4.1b) both test statistics achieve similar power.

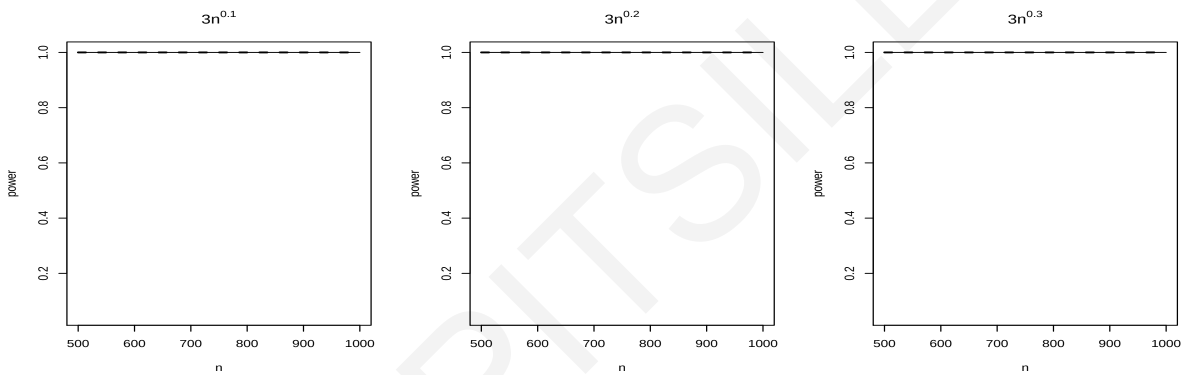
### 4.5.4 Application to a Bivariate Series

We study the relation between monthly log returns of the stocks of IBM and the S&P 500 composite index during the period January 1937 to December 2011 (900 observations in total). Figure 4.2 shows the ACF and ADCF plots of the original series, whereas Figure 4.3 shows the ACF and ADCF of the squared series. The first ACF plot (upper panel of Figure 4.2) indicates that there is no correlation among observations, whereas the ACF plot of the squared series (upper panel of Figure 4.3) confirms the conditional heteroscedasticity in monthly log-returns. However, both ADCF plots (lower panels of Figures 4.2 and 4.3) suggest strongly dependence among observations. The horizontal line in the ADCF plots has been drawn following the methodology outlined in Section 4.5.2. Applying the proposed testing methodology to the bivariate log returns directly, we note

(a)




(b)




(c)

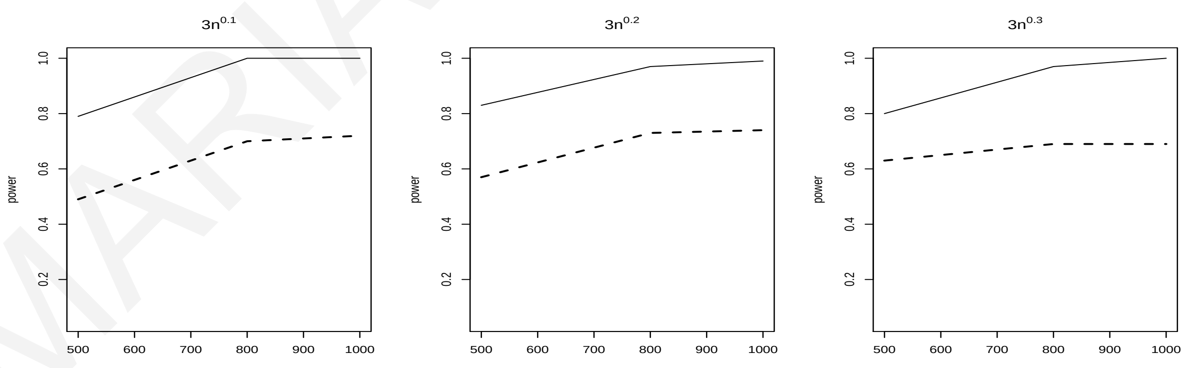





Figure 4.1: Achieved power of all test statistics. The results are based on $b = 499$ bootstrap replications and 100 simulations. The test statistic $\overline{T}_n$ is calculated by employing the Bartlett kernel. Solid line corresponds to $\overline{T}_n$, whereas dashed line corresponds to $mLB$. (a) The data are generated by the bivariate NMA(2) model given by (1.1). (b) The data are generated by the bivariate ARMA(1,1) model given by (4.15). (c) The data are generated by the bivariate GARCH(1,1) model given by (4.16).

Table 4.4: P-values of tests of independence among the residuals after fitting both VAR(3) and VAR(4) model to the bivariate log returns $\big(IBM_t, SP_t\big)$. P-values obtained after employing univariate EGARCH(1,1) models to the bivariate series are also presented. All results are computed based on $b = 499$ wild bootstrap replications. $\overline{T}_n$ is calculated based on the Bartlett kernel.

| Model fitted | $p$ | $mLB$ | $\overline{T}_n$ |
|---|---|---|---|
| VAR(3) | 6 | 0.714 | 0.032 |
| | 12 | 0.954 | 0.022 |
| | 24 | 0.798 | 0.002 |
| VAR(4) | 6 | 0.998 | 0.030 |
| | 12 | 0.998 | 0.012 |
| | 24 | 0.972 | 0.008 |
| EGARCH(1,1) | 6 | 0.155 | 0.164 |
| | 12 | 0.718 | 0.148 |
| | 24 | 0.845 | 0.144 |

that the test statistic $\overline{T}_n$ yields low $p$-values (0.004, 0.002 and 0) when $p = 6$, 12 and 24 respectively. The multivariate Ljung-Box statistic ($mLB$) yields large $p$-values (0.122, 0.488 and 0.348 respectively) for the same choices of $p$. Assuming that the bivariate series $\big(IBM_t, SP_t\big)$ follows a VAR model and employing the AIC to choose its order we obtain that a fourth order VAR fits well the data. Figure 4.4 shows the ACF (upper panel) and ADCF plots (lower panel) of the residuals of the fitted model. Note that the ADCF plot still indicates that there is dependence among the residuals. Table 4.4 shows the $p$-values of constructing tests of independence among the residuals after fitting both VAR(3) and VAR(4) model to the data. In both cases, $mLB$ statistic gives large $p$-values, whereas $\overline{T}_n$ yields low $p$-values suggesting dependence among the residuals. However, entertaining univariate exponential GARCH(1,1) (EGARCH) models seems to be a more appropriate choice for the data. Indeed, Figures 4.5 and 4.6 present the ACF and ADCF plots of the standardized residuals and their squared series of the two univariate models respectively, indicating no serial dependence among the observations. Moreover, both test statistics, $\overline{T}_n$ and $mLB$, yield large $p$-values, confirming the model adequacy (Table 4.4).

### 4.5.5 Application to a 12-Dimensional Time Series

We analyze the electroencephalograph (EEG) signals recorded extracranially and intracranially from twelve segments of both healthy and epileptic subjects. The dataset used here

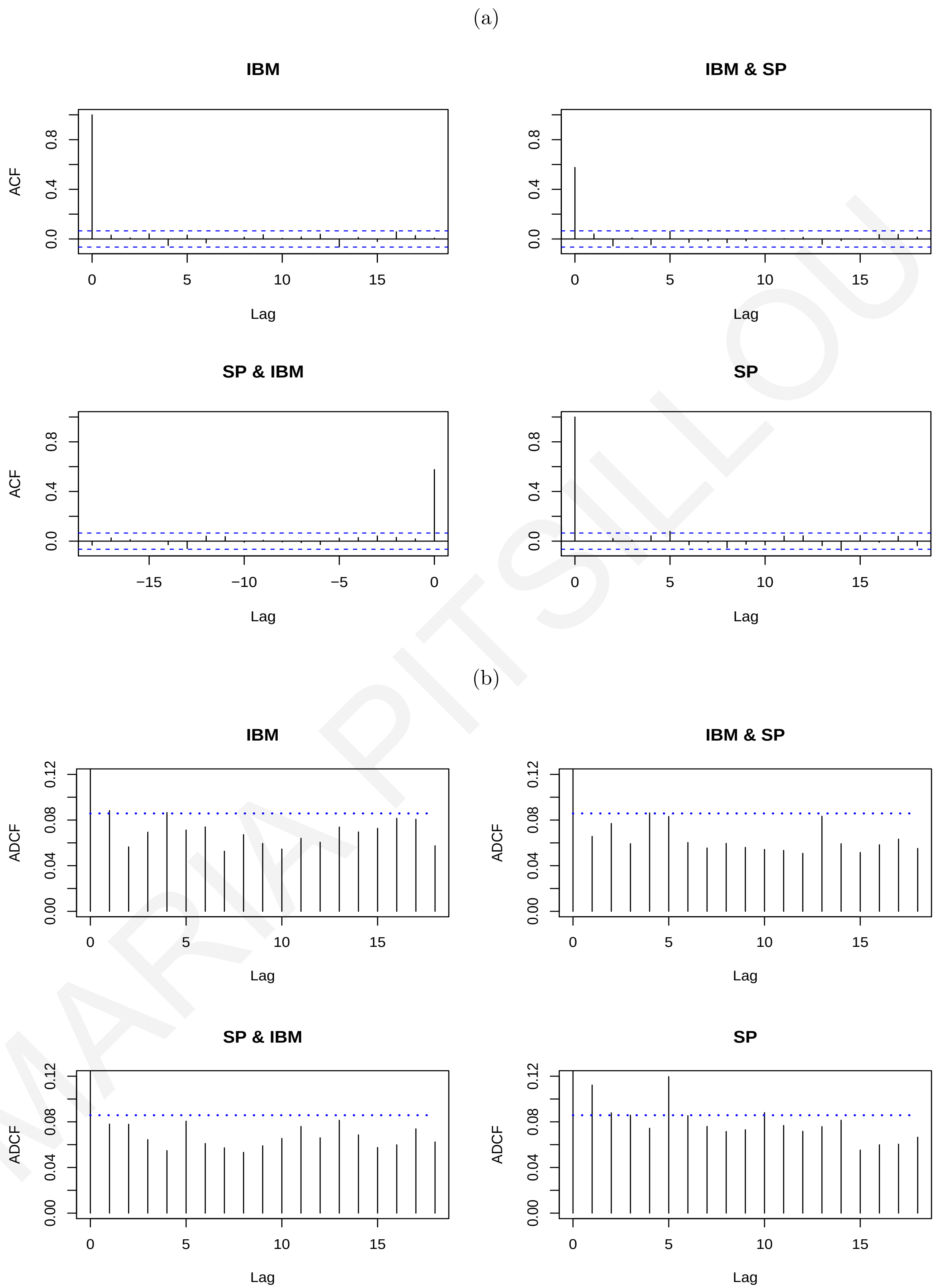
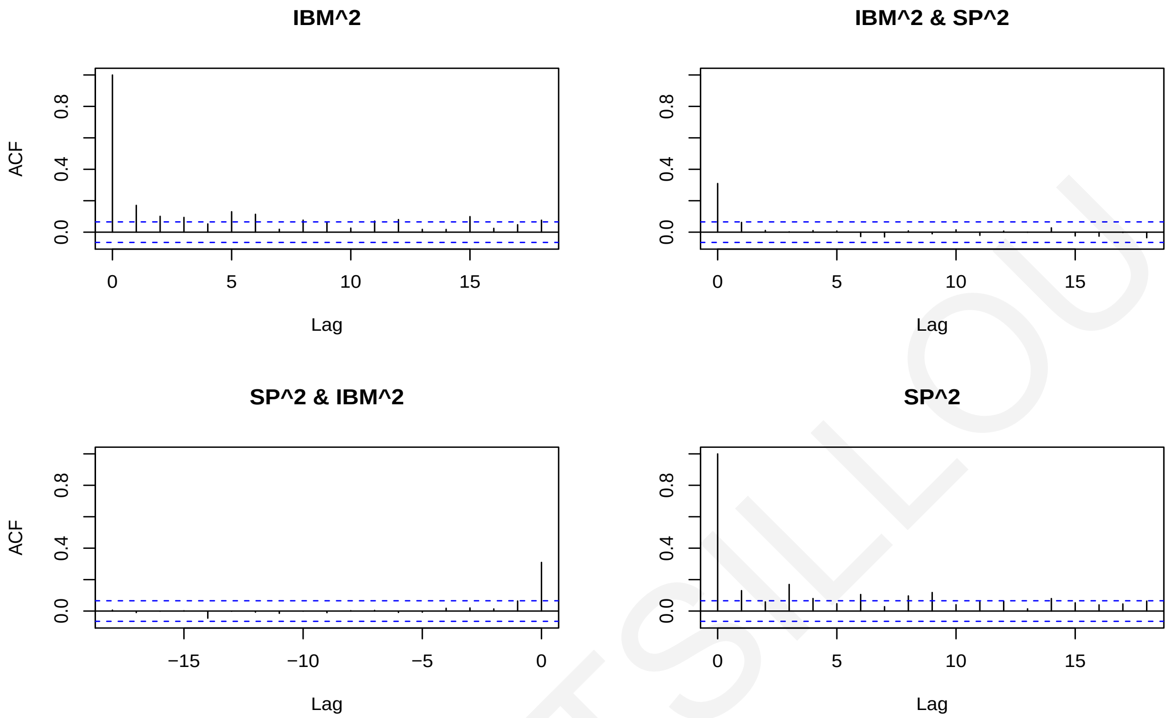



Figure 4.2: (a) The sample ACF of the original series. (b) The sample ADCF of the original series.

(a)




(b)

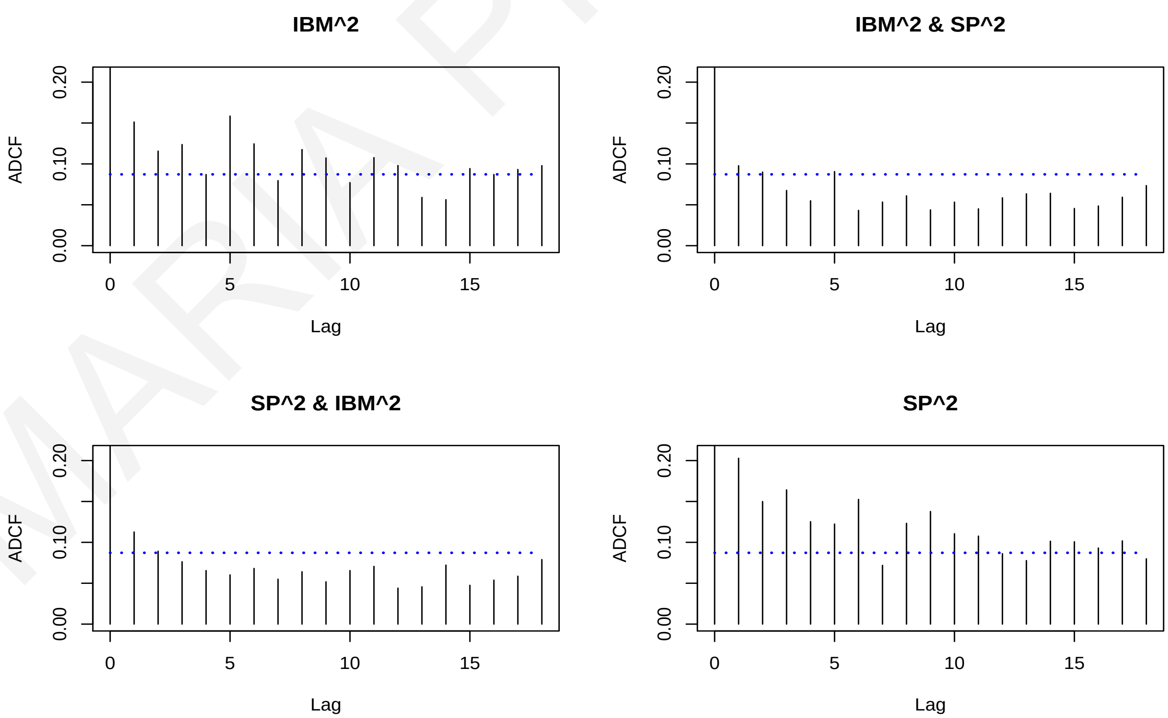
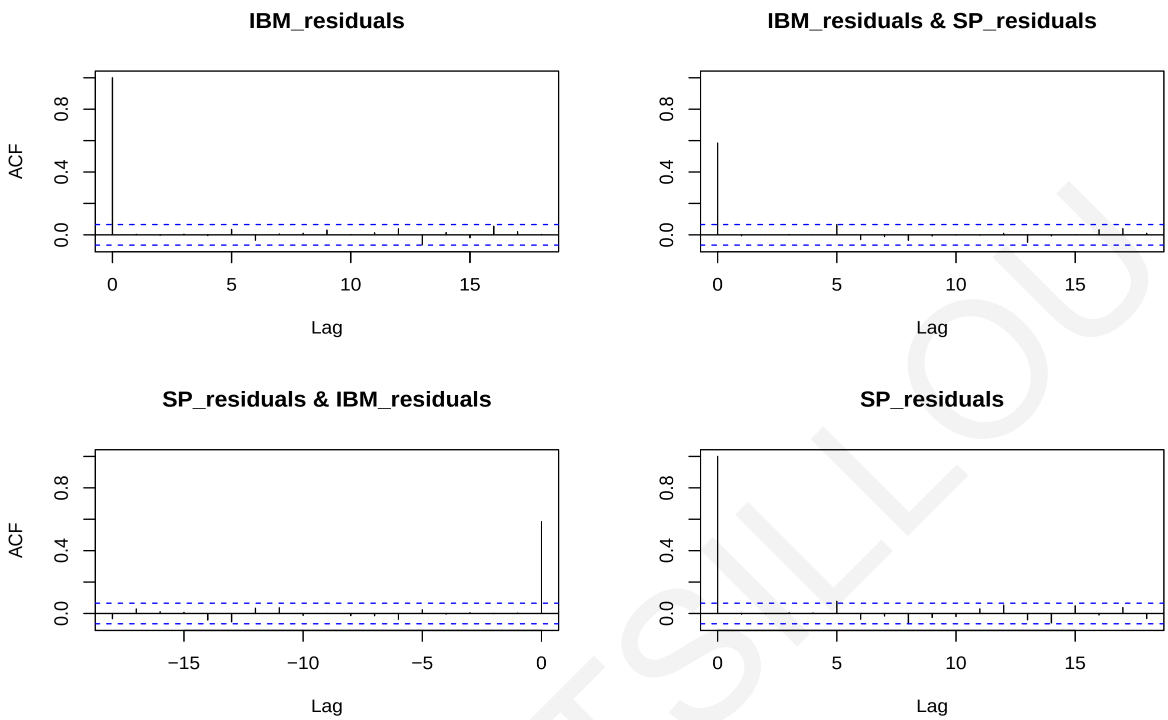



Figure 4.3: (a) The sample ACF of the squared series. (b) The sample ADCF of the squared series.

(a)




(b)

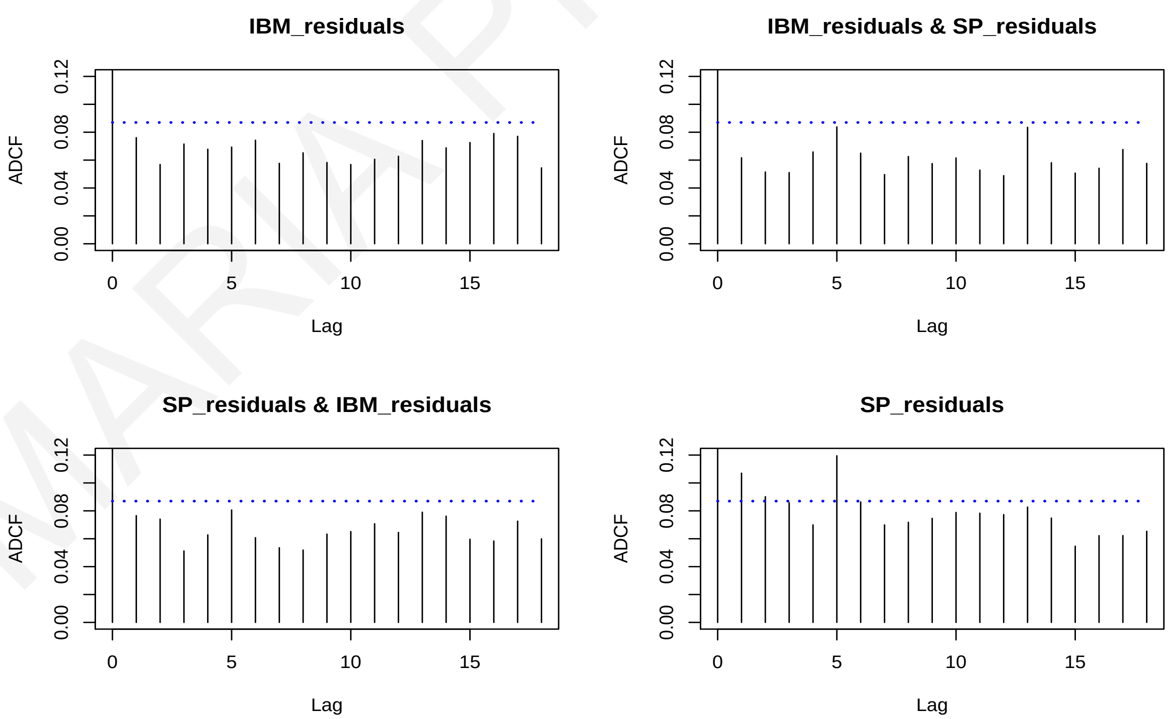
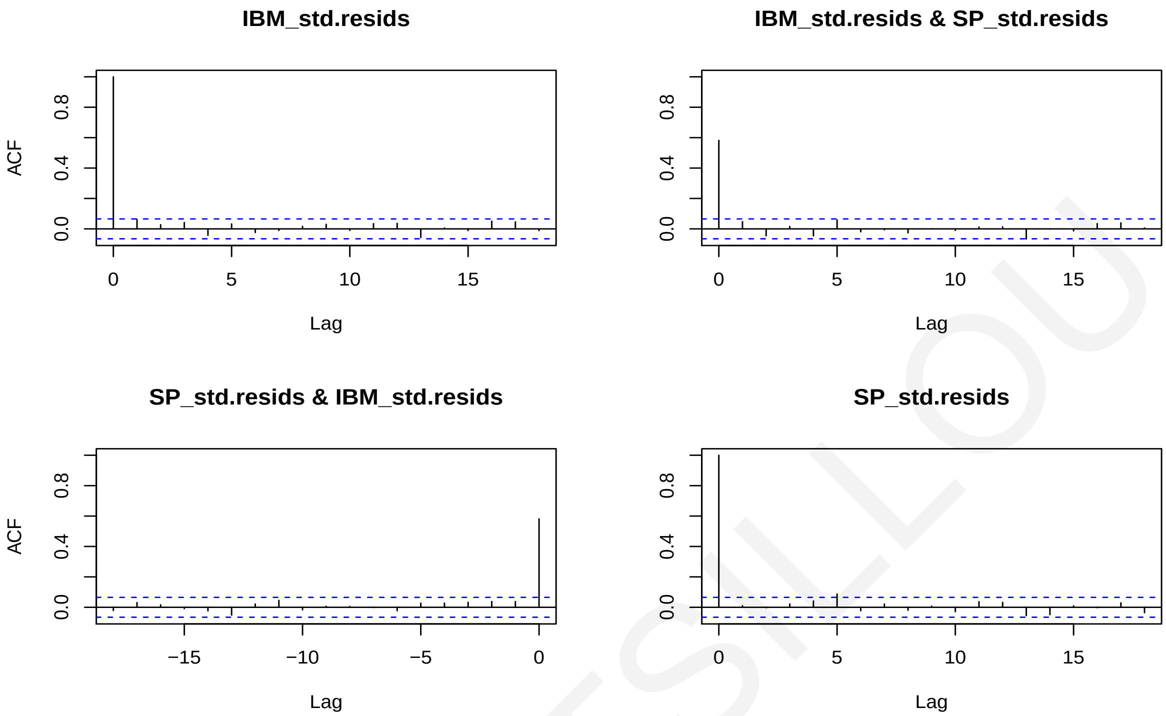



Figure 4.4: (a) The sample ACF of the residuals from VAR(4) model. (b) The sample ADCF of the residuals from VAR(4) model.

(a)




(b)

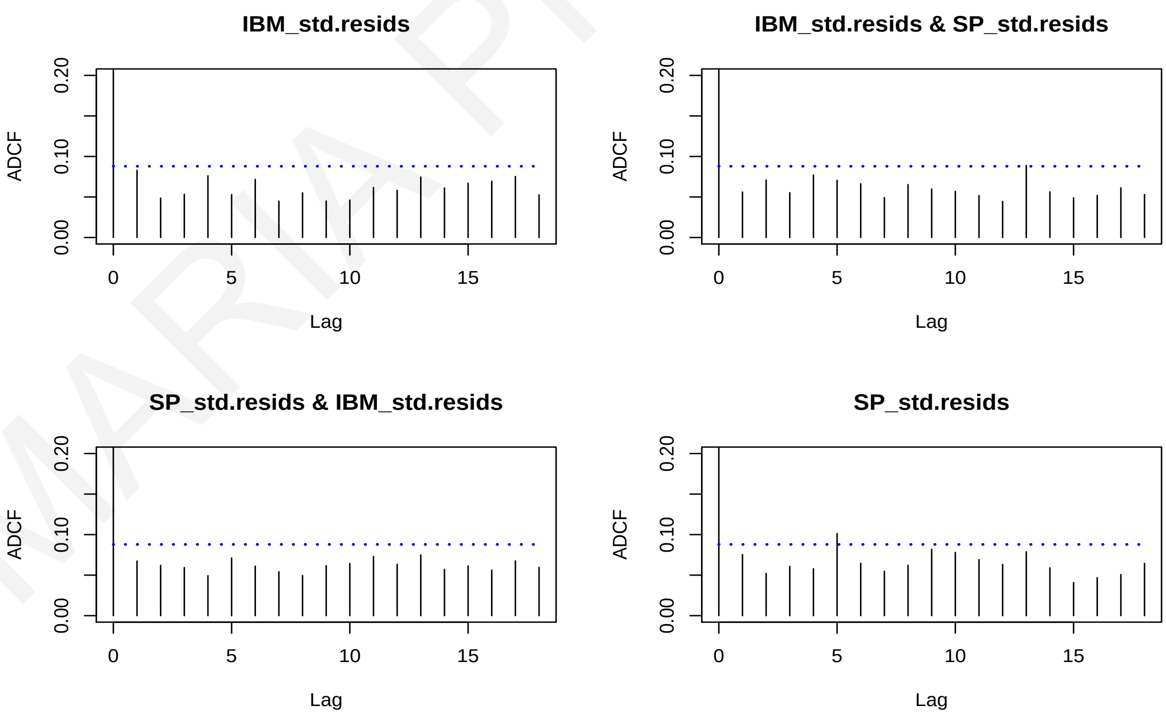



Figure 4.5: (a) The sample ACF of the standardized residuals after fitting univariate EGARCH(1,1) models. (b) The sample ADCF of the standardized residuals after fitting univariate EGARCH(1,1) models.

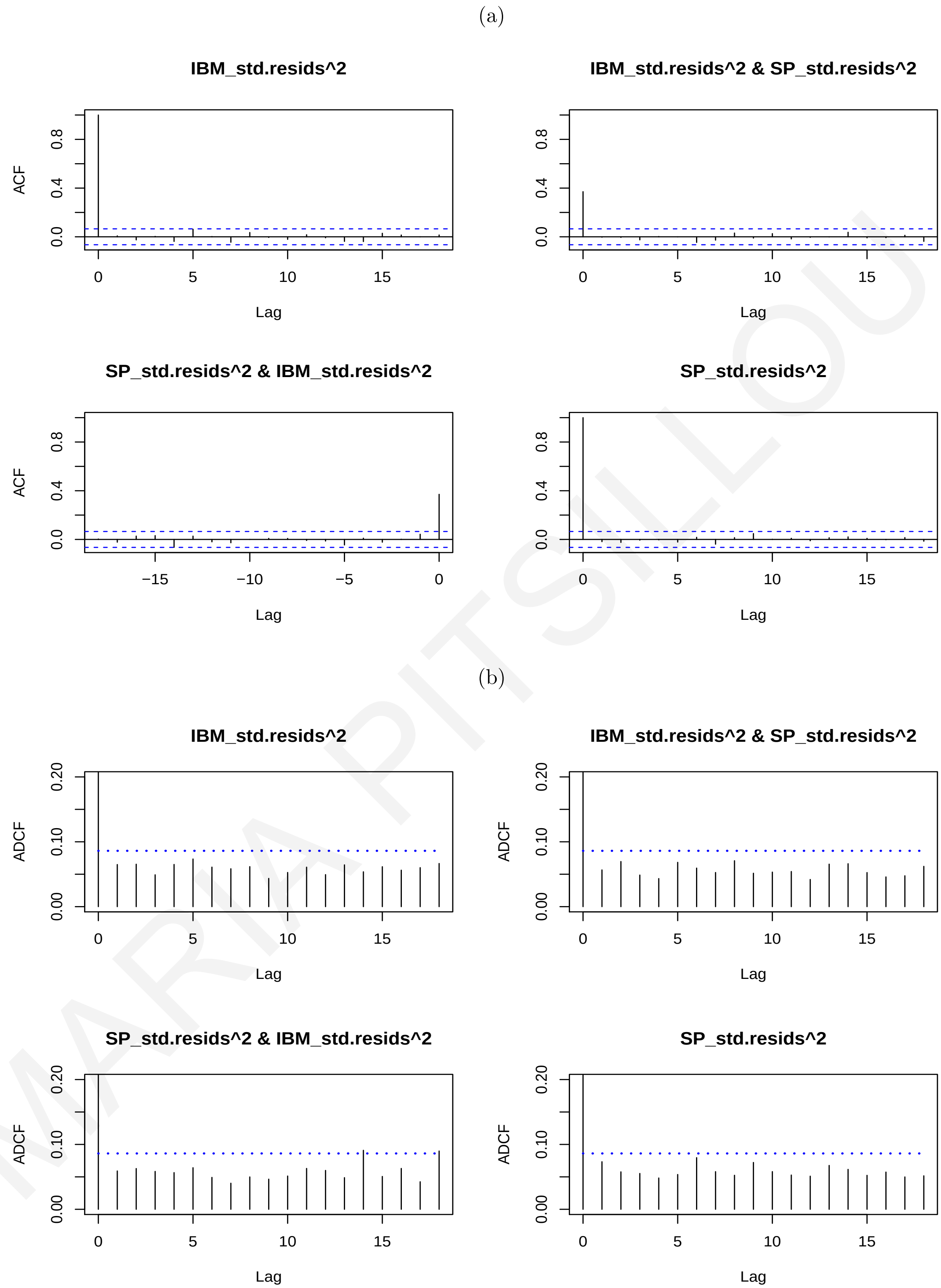



Figure 4.6: (a) The sample ACF of the squared standardized residuals after fitting univariate EGARCH(1,1) models. (b) The sample ADCF of the squared standardized residuals after fitting univariate EGARCH(1,1) models.

is a subset of an original data set publicly available by Andrzejaki et al. (2001). The selected data set contains three EEG time series recorded from healthy volunteers during the relaxed state with eyes open and closed. It also consists of three EEG signals obtained from patients during seizure interval. Six time series recorded from patients during seizure-free interval were also included in the data set; three of them were recorded from the epileptic zone and three from the hippocampal formation of the opposite hemisphere of the brain. Each of these EEG signals consists of 400 observations. We assume that this 12-dimensional time series follows a VAR model. Applying the AIC we obtain that a fifth order VAR model fits well the data. To check the adequacy of the model fit, we look at the behavior of the residuals. Constructing tests of independence among the residuals, both $mLB$ and $\overline{T}_n$ yields large $p$-values. In particular, $mLB$ gives $p$-values equal to 1 and $\overline{T}_n$ gives $p$-values equal to 0.882, 0.852 and 0.782 when $p = 6$, 10 and 19 respectively. All $p$-values are calculated for $b = 499$ wild bootstrap replications. The test statistic $\overline{T}_n$ is calculated by employing the Bartlett kernel. However, other choices of bandwidth values, $p$, and kernel functions, $K(\cdot)$, yield similar results.

# Appendix – Proofs

In this section we prove the main theorems of the chapter. Note that some of the arguments used for the proofs are similar to those given in the Appendix of Chapter 3, and thus they are omitted.

**Proof of Theorem 4.3.1** In Section 4.3 we showed that under Assumption 2$(i)$ and pairwise independence, $\widehat{V}^2_{rm}(\cdot)$ can be expressed as a degenerate $V$-statistic of order 2 with a measurable, symmetric, continuous and semidefinite kernel function given by (4.8). By Assumption 1, we may apply Theorem 1 of Leucht and Neumann (2013a) and get:

$$(n-|j|)V^2_{rm}(j) \to Z := \sum_k \lambda_k Z_k^2,$$

as $n \to \infty$ in distribution. □

For the rest of the proofs, we first define

$$\bar{f}^{(r,m)}(\omega, u, v) = \frac{1}{2\pi} \sum_{j=-(n-1)}^{(n-1)} K(j/p)(1-|j|/n)^{1/2} \tilde{\sigma}_j^{(r,m)}(u,v) e^{-ij\omega},$$

where

$$\tilde{\sigma}_j^{(r,m)}(u,v) = \frac{1}{n-|j|} \sum_{t=|j|+1}^{n} \psi_{t;r}(u)\psi_{t-|j|;m}(v) \tag{4.17}$$

and

$$\psi_{t;r}(u) \equiv e^{iuX_{t;r}} - \phi^{(r)}(u).$$

The corresponding pseudoestimator of the generalized spectral density matrix is defined as

$$\overline{F}(\omega, u, v) = \frac{1}{2\pi} \sum_{j=-(n-1)}^{(n-1)} K(j/p)(1-|j|/n)^{1/2} \widetilde{\Sigma}_{|j|}(u,v) e^{-ij\omega},$$

where $\widetilde{\Sigma}_{|j|}(\cdot,\cdot)$ is the covariance matrix of $e^{iu\mathbf{X}_t}$ with elements given by (4.17). For the proof of Theorem 4.4.1, we will need the following two lemmas whose proof is omitted as it follows closely the arguments of the corresponding proofs given in the Appendix of Chapter 3.

**Lemma 4.5.1** Suppose that $\{\boldsymbol{X}_t\}$ satisfies Assumptions 1 and 3(ii). Then we have that $(n-|j|)^2 E\left|\widehat{\sigma}_j^{(r,m)}(u,v) - \tilde{\sigma}_j^{(r,m)}(u,v)\right|^2 \leq C$ and $(n-|j|)E\left|\tilde{\sigma}_j^{(r,m)}(u,v)\right|^2 \leq C$ uniformly in $(u,v) \in \mathbb{R}^2$ for $r, m = 1, 2, \ldots, p$.

**Lemma 4.5.2** Suppose that $\{\boldsymbol{X}_t\}$ satisfies Assumptions 1, 3(ii) and 4. For each $\gamma > 0$, let $D(\gamma) = \{(u,v) : \gamma \leq |u| \leq 1/\gamma, \gamma \leq |v| \leq 1/\gamma\}$. Then

$$\int_{D(\gamma)} \sum_{j=1}^{n-1} K^2(j/p)(n-j)\left\{\left|\hat{\sigma}_j^{(r,m)}(u,v)\right|^2 - \left|\tilde{\sigma}_j^{(r,m)}(u,v)\right|^2\right\} d\mathcal{W}(u,v) \quad = \quad O_P(p/\sqrt{n}) = o_P(\sqrt{p})$$

for $r, m = 1, 2, \ldots, p$ as $p/n \to 0$.

**Proof of Theorem 4.4.1** Following similar arguments as in the corresponding proof given in Chapter 3, it can be shown that (Hong, 1999)

$$\sum_{j=1}^{n-1} K^2(j/p)(n-j)\left|\tilde{\sigma}_j^{(r,m)}(u,v)\right|^2 \quad = \quad \widehat{C}^{rm}(u,v) + \widehat{V}^{rm}(u,v) \tag{4.18}$$

where

$$\widehat{C}^{rm}(u,v) \quad = \quad \sum_{j=1}^{n-1} \frac{K^2(j/p)}{n-j}\left[\sum_{t=j+1}^{n} C_{ttj}^{rm}(u,v)\right],$$

$$\widehat{V}^{rm}(u,v) \quad = \quad \sum_{j=1}^{n-1} \frac{K^2(j/p)}{n-j}\left[\sum_{t=j+2}^{n}\sum_{s=j+1}^{t-1} V_{tsj}^{rm}(u,v)\right],$$

with

$$V_{tsj}^{rm}(u,v) \quad = \quad C_{tsj}^{rm}(u,v) + C_{stj}^{rm}(u,v)^*$$

and

$$C_{tsj}^{rm}(u,v) \;=\; \psi_{t;r}(u)\psi_{s;r}(u)^*\psi_{t-j;m}(v)\psi_{s-j;m}(v)^*$$

where $*$ denotes complex conjugate.

For the first summand of (4.18), it holds that $\int_{D(\gamma)} C_{ttj}^{rm}(u,v)d\mathcal{W}$ and $\int_{D(\gamma)} C_{ssj}^{rm}(u,v)d\mathcal{W}$ are independent integrals unless $t=s$ or $s\pm j$. In addition,

$$E\int_{D(\gamma)} C_{ttj}^{rm}(u,v)d\mathcal{W}(u,v) \;=\; C_0^{rm\gamma} \equiv \int_{D(\gamma)} \sigma_0^{(r,r)}(u,-u)\sigma_0^{(m,m)}(v,-v)d\mathcal{W}(u,v) < \infty,$$

shows that $E\Big\{\sum_{t=j+1}^{n}\Big[\int_{D(\gamma)} C_{ttj}^{rm}(u,v)d\mathcal{W} - C_0^{rm\gamma}\Big]\Big\}^2 \le C(n-j)$. Hence, by Markov's inequality, Cauchy-Schwarz inequality and the properties of the kernel function, we obtain that

$$\int_{D(\gamma)} \widehat{C}^{rm}(u,v)d\mathcal{W} \;=\; O_P(p/\sqrt{n}) + C_0^{rm\gamma}\sum_{j=1}^{n-1}K^2(j/p). \tag{4.19}$$

So, using Lemma 4.5.2, equations (4.18) and (4.19) we have the following:

$$\begin{aligned}
\int_{D(\gamma)}\Big\{\sum_{j=1}^{n-1}K^2(j/p)(n-j)\left|\widehat{\sigma}_j^{(r,m)}(u,v)\right|^2\Big\}d\mathcal{W} \;&=\; \int_{D(\gamma)}\Big\{\sum_{j=1}^{n-1}K^2(j/p)(n-j)\left|\tilde{\sigma}_j^{(r,m)}(u,v)\right|^2\Big\}d\mathcal{W}\\
&\quad + O_P(p/\sqrt{n})\\
&= \int_{D(\gamma)} \widehat{C}^{rm}(u,v)d\mathcal{W} + \int_{D(\gamma)} \widehat{V}^{rm}(u,v)d\mathcal{W}\\
&\quad + O_P(p/\sqrt{n})\\
&= C_0^{rm\gamma}\sum_{j=1}^{n-1}K^2(j/p) + \widehat{V}_n^{rm\gamma} + O_P(p/\sqrt{n}),
\end{aligned}$$

where $\widehat{V}_n^{rm\gamma} \equiv \int_{D(\gamma)} \widehat{V}^{rm}(u,v)d\mathcal{W}$. Therefore, the test statistic, $T_{n;\gamma}^{(r,m)}$, takes the form:

$$\begin{aligned}
T_{n;\gamma}^{(r,m)} \;&=\; \int_{D(\gamma)}\Big\{\sum_{j=1}^{n-1}K^2(j/p)(n-j)\left|\widehat{\sigma}_j^{(r,m)}(u,v)\right|^2\Big\}d\mathcal{W}\\
&= C_0^{rm\gamma}\sum_{j=1}^{n-1}K^2(j/p) + \widehat{V}_n^{rm\gamma} + O_P(p/\sqrt{n}).
\end{aligned} \tag{4.20}$$

Given Assumption 4 and by applying Hong (1999, Theorem A.3) on $D(\gamma)$, we obtain

$$\widehat{V}_n^{rm\gamma} = \widehat{V}_{ng}^{rm\gamma} + o_P(\sqrt{p}) \tag{4.21}$$

where

$$\widehat{V}_{ng}^{rm\gamma} = \sum_{t=g+2}^{n} \sum_{s=1}^{t-g-1} \sum_{j=1}^{g} \frac{K^2(j/p)}{n-j} \int_{D(\gamma)} V_{tsj}^{rm}(u,v) d\mathcal{W}$$

and $g \equiv g(n)$ such that $g/p \to 0, g/n \to 0$. Now, by applying Hong (1999, Theorem A.4) on $D(\gamma)$ we get the following:

$$\left[ pD_0^{rm\gamma} \int_0^\infty K^4(z)dz \right]^{-1/2} \widehat{V}_{ng}^{rm\gamma} \to N(0,1) \tag{4.22}$$

as $n \to \infty$ in distribution, where

$$D_0^{rm\gamma} = 2 \int_{D(\gamma)} \left| \sigma_0^{(r,r)}(u,u') \sigma_0^{(m,m)}(v,v') \right|^2 d\mathcal{W}(u,v) d\mathcal{W}(u',v').$$

Equations (4.20), (4.21) and (4.22) yield to

$$\frac{\int_{D(\gamma)} \left\{ \sum_{j=1}^{n-1} K^2(j/p)(n-j) \left| \hat{\sigma}_j^{(r,m)}(u,v) \right|^2 \right\} d\mathcal{W} - C_0^{rm\gamma} \sum_{j=1}^{n-1} K^2(j/p)}{\left[ pD_0^{rm\gamma} \int_0^\infty K^4(z)dz \right]^{1/2}} \to N(0,1). \tag{4.23}$$

as $n \to \infty$, in distribution.

Observe that $\widehat{C}_0^{rm\gamma} - C_0^{rm\gamma} = O_P(1/\sqrt{n})$ and that $\sum_{j=1}^{n-1} K^2(j/p) = O(p)$. Combining both results, we can replace $C_0^{rm\gamma}$ by $\widehat{C}_0^{rm\gamma}$ when $p/n \to 0$. In addition, $p^{-1} \sum_{j=1}^{n-2} K^4(j/p) \to \int_0^\infty K^4(z)dz$ and that $\widehat{D}_0^{rm\gamma} \to D_0^{rm\gamma}$ in probability. By Slutsky's theorem, $pD_0^{rm\gamma} \int_0^\infty K^4(z)dz$ can be replaced by $\widehat{D}_0^{rm\gamma} \sum_{j=1}^{n-2} K^4(j/p)$.

Summarizing, (4.23) becomes

$$\frac{T_{n;\gamma}^{(r,m)} - \widehat{C}_0^{rm\gamma} \sum_{j=1}^{n-1} K^2(j/p)}{\left[ \widehat{D}_0^{rm\gamma} \sum_{j=1}^{n-2} K^4(j/p) \right]^{1/2}} \to N(0,1),$$

as $n \to \infty$ in distribution. The rest of the proof follows by similar arguments given in

Chapter 3, by showing that

$$\limsup_{\gamma \to 0} \limsup_{n \to \infty} \left| T_n^{(r,m)} - T_{n;\gamma}^{(r,m)} \right| \quad = \quad 0. \tag{4.24}$$

□

**Proof of Corollary 4.4.1** From Theorem 4.4.1 and under the null hypothesis of independence, the random variables $T_n^{(r,m)}$ satisfy

$$\frac{T_n^{(r,m)} - \widehat{C}_0^{(r,m)} \sum_{j=1}^{n-1} K^2(j/p)}{\left[ \widehat{D}_0^{(r,m)} \sum_{j=1}^{n-2} K^4(j/p) \right]^{1/2}} \quad \to \quad N(0,1) \tag{4.25}$$

in distribution, as $n \to \infty$, for $r, m = 1, \ldots, d$. Then, it can be shown by arguments quite analogous to Hong (1999, Proof of Theorem 3) that

$$\widetilde{M}_n = \frac{\sum_{r,m} T_n^{(r,m)} - \sum_{r,m} \widehat{C}_0^{(r,m)} \sum_{j=1}^{n-1} K^2(j/p)}{\left[ \sum_{r,m} \widehat{D}_0^{(r,m)} \sum_{j=1}^{n-2} K^4(j/p) \right]^{1/2}} \quad \to \quad N(0,1),$$

in distribution, as $n \to \infty$. □

**Proof of Corollary 4.4.2** Recall that $\overline{T}_n^{(r,m)}$ may be written as

$$\overline{T}_n^{(r,m)} \quad = \quad \frac{1}{\sqrt{\widehat{V}_{rr}^2(0)}\sqrt{\widehat{V}_{mm}^2(0)}} T_n^{(r,m)}.$$

By recalling result (4.25), we get

$$\frac{\overline{T}_n^{(r,m)} - \underline{\widehat{C}}_0^{(r,m)} \sum_{j=1}^{n-1} K^2(j/p)}{\left[ \underline{\widehat{D}}_0^{(r,m)} \sum_{j=1}^{n-2} K^4(j/p) \right]^{1/2}} \to N(0,1), \tag{4.26}$$

in distribution, as $n \to \infty$, for $r, m = 1, 2, \ldots, d$, where

$$\underline{\widehat{C}}_0^{(r,m)} = \frac{\widehat{C}_0^{(r,m)}}{\widehat{V}_{rr}(0)\widehat{V}_{mm}(0)}, \quad \underline{\widehat{D}}_0^{(r,m)} = \frac{\widehat{D}_0^{(r,m)}}{\widehat{V}_{rr}^2(0)\widehat{V}_{mm}^2(0)} = 2.$$

The sum of the normal random variables defined in (4.26) is also normally distributed, namely

$$\overline{M}_n = \frac{\overline{T}_n - \sum_{r,m} \underline{\widehat{C}}_0^{(r,m)} \sum_{j=1}^{n-1} K^2(j/p)}{d\left[2\sum_{j=1}^{n-2} K^4(j/p)\right]^{1/2}} \quad \to \quad N(0,1),$$

in distribution, as $n \to \infty$ and the proof is now completed. □

**Proof of Theorem 4.4.2** We prove the first result of the theorem. Recall $D(\gamma)$ defined in Lemma 4.5.2. For the proof we show the following: (i) $1/p \sum_{j=1}^{n} K^4(j/p) \to \int_0^\infty K^4(j/p)$ given Assumption 4 and $p \to \infty$ and $p/n \to 0$,
(ii) $E \int_{D(\gamma)} \int_{-\pi}^{\pi} \left| \hat{f}^{(r,m)}(\omega, u, v) - f^{(r,m)}(\omega, u, v) \right|^2 d\omega d\mathcal{W}(u, v) \to 0$ which is proved similarly to the proof of Hong (1999, Proof of Theorem 2, p. 1213) on $D(\gamma)$ for all $r, m = 1, \ldots, d$ given Assumptions 1, 3(i) and 4. Additionally, by applying Markov's inequality we get (iii) $\widehat{C}_0^{rm\gamma} = O_P(1)$ and (iv) $\widehat{D}_0^{rm\gamma} \to D_0^{rm\gamma}$ in probability.

Combining remarks (i) and (iv) and by Slutsky's theorem we get

$$\frac{1}{p}\widehat{D}_0^{rm\gamma} \sum_{j=1}^{n-2} K^4(j/p) \quad \to \quad D_0^{rm\gamma} \int_0^\infty K^4(z)dz,$$

in probability. Then, from the properties of convergence in probability,

$$\frac{1}{\sqrt{p}}\left[\sum_{r,m} \widehat{D}_0^{rm\gamma} \sum_{j=1}^{n-2} K^4(j/p)\right]^{1/2} \quad \to \quad \left[\sum_{r,m} D_0^{rm\gamma} \int_0^\infty K^4(z)dz\right]^{1/2}, \tag{4.27}$$

in probability.

By recalling (4.13) we define

$$\begin{aligned} L^2_{2;\gamma}\Big(\widehat{F}(\omega,u,v),\widehat{F}_0(\omega,u,v)\Big) &\equiv \int_{D(\gamma)}\int_{-\pi}^{\pi} \mathrm{tr}\Big\{\Big(\widehat{F}(\omega,u,v)-\widehat{F}_0(\omega,u,v)\Big)^* \\ &\qquad \times\Big(\widehat{F}(\omega,u,v)-\widehat{F}_0(\omega,u,v)\Big)\Big\} d\omega d\mathcal{W}(u,v) \\ &= \frac{2}{\pi}\sum_{r,m} T^{(r,m)}_{n;\gamma} = \frac{2}{\pi}\widetilde{T}_{n;\gamma}. \end{aligned}$$

Using the inequality $|x+y| \leq 2|x|^2 + 2|y|^2$ and after some calculations, we observe that

$$\begin{aligned} &\frac{\pi}{2}\left[\frac{1}{n}L^2_{2;\gamma}\Big(\widehat{F}(\omega,u,v),\widehat{F}_0(\omega,u,v)\Big) - L^2_{2;\gamma}\Big(F(\omega,u,v),F_0(\omega,u,v)\Big)\right] \\ &\qquad \leq \sum_{r,m}\int_{D(\gamma)}\int_{-\pi}^{\pi}\left|\widehat{f}^{(r,m)}(\omega,u,v) - f^{(r,m)}(\omega,u,v)\right|^2 d\omega d\mathcal{W}. \end{aligned}$$

Thus,

$$\begin{aligned} &E\left|\frac{\pi}{2}\left[\frac{1}{n}L^2_{2;\gamma}\Big(\widehat{F}(\omega,u,v),\widehat{F}_0(\omega,u,v)\Big) - L^2_{2;\gamma}\Big(F(\omega,u,v),F_0(\omega,u,v)\Big)\right] - \frac{1}{n}\sum_{r,m}\widehat{C}_0^{rm\gamma}\sum_{j=1}^{n-1}K^2(j/p)\right| \\ &\leq \sum_{r,m}E\int_{D(\gamma)}\int_{-\pi}^{\pi}\left|\widehat{f}^{(r,m)}(\omega,u,v) - f^{(r,m)}(\omega,u,v)\right|^2 d\omega d\mathcal{W} + \frac{1}{n}\sum_{j=1}^{n-1}K^2(j/p)\sum_{r,m}E\left|\widehat{C}_0^{rm\gamma}\right|. \end{aligned}$$

By applying Markov's inequality and using remarks (ii) and (iii), the last expression yields

$$\begin{aligned} &\frac{\pi}{2}\frac{1}{n}L^2_{2;\gamma}\Big(\widehat{F}(\omega,u,v),\widehat{F}_0(\omega,u,v)\Big) - \frac{1}{n}\sum_{r,m}\widehat{C}_0^{rm\gamma}\sum_{j=1}^{n-1}K^2(j/p) \\ &\qquad \rightarrow \frac{\pi}{2}L^2_{2;\gamma}(F(\omega,u,v),F(\omega,u,v)), \end{aligned}$$

in probability, i.e.,

$$\frac{1}{n}\Big[\widetilde{T}_{n;\gamma} - \sum_{r,m}\widehat{C}_0^{rm\gamma}\sum_{j=1}^{n-1}K^2(j/p)\Big] \;\rightarrow\; \frac{\pi}{2}L^2_{2;\gamma}\Big(F(\omega,u,v),F_0(\omega,u,v)\Big), \tag{4.28}$$

in probability. Combining now (4.27) and (4.28) we get the required result on $D(\gamma)$. However, considering $\widehat{C}_0^{rm\gamma} \rightarrow \widehat{C}_0^{(r,m)}$ and $\widehat{D}_0^{rm\gamma} \rightarrow \widehat{D}_0^{(r,m)}$ as $\gamma \rightarrow 0$ and (4.24), the first result is now proved on $\mathbb{R}^2$.

The proof of the second result is derived by following similar arguments. In particular, the first remark (i) remains the same. Following (Hong, 1999, Proof of Theorem 2), remark (ii) becomes
$E\int_{D(\gamma)}\int_{-\pi}^{\pi}\left|\hat{g}^{(r,m)}(\omega,u,v)-g^{(r,m)}(\omega,u,v)\right|^2 d\omega d\mathcal{W}(u,v)\to 0$, where

$$\hat{g}^{(r,m)}(\omega,u,v)=\frac{1}{2\pi}\sum_{j=-(n-1)}^{(n-1)}(1-|j|/n)^{1/2}K(j/p)\frac{\sigma_j^{(r,m)}(u,v)}{\sqrt{V_{rr}(0)V_{mm}(0)}}e^{-ij\omega},$$

for all $r,m=1,\ldots,d$ given Assumptions 1, 3(i) and 4. Remark (iii) becomes $\widehat{\underline{C}}_0^{rm\gamma}=O_P(1)$. The quantity $\underline{C}_0^{rm\gamma}$ is defined as the corresponding quantity $C_0^{rm\gamma}$ being divided by the term $\{V_{rr;\gamma}(0)V_{mm;\gamma}(0)\}$ (or its empirical analogue where appropriate), with $V_{rm;\gamma}(0)$ given by

$$V_{rm;\gamma}^2(0)=\int_{D(\gamma)}\left|\sigma_0^{(r,m)}(u,v)\right|^2 d\mathcal{W}(u,v),$$

for all $r,m=1,2,\ldots,d$. Thus, applying the same steps followed in the proof of the first result of the theorem, equations (4.27) and (4.28) become now

$$\frac{d}{\sqrt{p}}\left[2\sum_{j=1}^{n-2}K^4(j/p)\right]^{1/2} \quad\to\quad d\left[2\int_0^\infty K^4(z)dz\right]^{1/2}, \tag{4.29}$$

in probability, and

$$\frac{1}{n}\Big[\overline{T}_{n;\gamma}-\sum_{r,m}\widehat{\underline{C}}_0^{rm\gamma}\sum_{j=1}^{n-1}K^2(j/p)\Big] \quad\to\quad \frac{\pi}{2}L^2_{2;\gamma}\Big(G(\omega,u,v),G_0(\omega,u,v)\Big), \tag{4.30}$$

in probability, where

$$\overline{T}_{n;\gamma} \quad=\quad \sum_{r,m}\overline{T}_{n;\gamma}^{(r,m)},$$

and

$$\overline{T}_{n;\gamma}^{(r,m)}=\frac{T_{n;\gamma}^{(r,m)}}{V_{rr;\gamma}(0)V_{mm;\gamma}(0)}.$$

Equations (4.29) and (4.30) prove the second result on $D(\gamma)$. We further obtain that following similar arguments as in Chapter 3, one can show that

$$\limsup_{\gamma\to 0}\limsup_{n\to\infty}\left|\overline{T}_n^{(r,m)}-\overline{T}_{n;\gamma}^{(r,m)}\right| \;=\; 0. \tag{4.31}$$

Considering $\widehat{\underline{C}}_0^{rm\gamma} \to \widehat{\underline{C}}_0^{(r,m)}$ as $\gamma \to 0$ and (4.31), the proof of the second result is now proved on $\mathbb{R}^2$. □

# Chapter 5

# dCovTS: An R Package that Implements Distance Covariance and Correlation Theory in Time Series

## 5.1 Introduction

As we have already highlighted in the previous chapters, although Székely et al. (2007) initially introduced the concept of distance covariance under the assumption that the data are i.i.d., Zhou (2012) extended this notation to a time series framework and developed an asymptotic theory based at a fixed lag order. In Chapters 3 and 4, we relaxed this assumption and constructed a distance covariance testing methodology for both univariate and multivariate time series by considering an increasing number of lags. The **energy** (Rizzo and Szekely, 2014) package for R, is a package that involves a wide range of functions for the existing distance covariance methodology. However, there is no package for the aforementioned distance covariance methodology in time series. Thus, we aim at filling this gap by publishing an R package named **dCovTS**.

The current version of **dCovTS** package (version number 1.1) is available from CRAN and can be downloaded via `https://cran.r-project.org/web/packages/dCovTS/`. The aim of the **dCovTS** package is to provide a set of functions that compute and plot dis-

Table 5.1: Functions in **dCovTS**

| Function | Description |
|---|---|
| ADCF, mADCF | Estimates distance correlation for a univariate and multivariate time series respectively |
| ADCV, mADCV | Estimates distance covariance for a univariate and multivariate time series respectively |
| ADCFplot, mADCFplot | Plots sample distance correlation in a univariate and multivariate time series framework respectively |
| kernelFun | Computes univariate kernel function, $k(\cdot)$ |
| UnivTest | Performs a univariate test of independence based on $T_n$ |
| mADCFtest, mADCVtest | Perform multivariate tests of independence based on $\overline{T}_n$ and $\widetilde{T}_n$ respectively |

Table 5.2: Datasets in **dCovTS**

| Data | Description |
|---|---|
| ibmSp500 | Monthly returns of IBM and S&P 500 composite index from January 1926 to December 2011 |
| MortTempPart | Mortality, temperature and pollution data measured daily in Los Angeles County over the period 1970-1979 |

tance covariance and correlation functions in both univariate and multivariate time series. Moreover, it offers functions that perform univariate and multivariate tests of independence based on distance covariance function as explained in Chapters 3 and 4 (see Table 5.1). The package also provides two real datasets listed in Table 5.2. In Section 5.2, we recall the main features of this testing methodology and we demonstrate how they can be implemented with **dCovTS**. In the last Section 5.3, we apply **dCovTS** to several real data examples. A more detailed description of the functions and datasets can be found in the help files of the package.

## 5.2 Implementation of dCovTS Package

In this section, we recall the main results provided in Chapters 3 and 4 and demonstrate how they can be developed in R from our package **dCovTS**.

### 5.2.1 Distance Covariance Function via dCovTS

Consider first the univariate case. Recall that the ADCV, $V_X(j)$, between $X_t$ and $X_{t-|j|}$ is defined as the positive square root of

$$V_X^2(j) \;=\; \frac{1}{\pi^2}\int_{\mathbb{R}^2} \frac{\left|\phi_{|j|}(u,v)-\phi(u)\phi(v)\right|^2}{|u|^2\,|v|^2}dudv, \qquad j=0,\pm1,\pm2,\ldots \tag{5.1}$$

where $\phi_{|j|}(u,v) = E\Big[\exp\Big(i(uX_t + vX_{t-|j|})\Big)\Big]$ is the joint characteristic function of $X_t$ and $X_{t-|j|}$ and $\phi(u) := \phi_{|j|}(u,0)$ , $\phi(v) := \phi_{|j|}(0,v)$ are their corresponding marginal characteristic functions, for $(u,v)\in\mathbb{R}^2$.

Rescaling (5.1), one can define the ADCF as the positive square root of

$$R_X^2(j) \;=\; \frac{V_X^2(j)}{V_X^2(0)}, \qquad j=0,\pm1,\pm2,\ldots$$

for $V_X^2(0)\neq 0$ and zero otherwise.

Székely and Rizzo (2014) proposed an unbiased version of the sample distance covariance. Considering a sample of size $n$, in the context of time series data this is given by

$$\widetilde{V}_X^2(j) = \frac{1}{(n-|j|)(n-|j|-3)}\sum_{r\neq l}\widetilde{A}_{rl}\widetilde{B}_{rl}, \tag{5.2}$$

for $n>3$, where $\widetilde{A}_{rl}$ is the $(r,l)$ element of the so-called $\mathcal{U}$-centered matrix $\widetilde{A}$, defined by

$$\widetilde{A}_{rl} = \begin{cases} a_{rl} - \dfrac{1}{n-|j|-2}\displaystyle\sum_{t=1+|j|}^{n} a_{rt} - \dfrac{1}{n-|j|-2}\displaystyle\sum_{s=1+|j|}^{n} a_{sl} + \dfrac{1}{(n-|j|-1)(n-|j|-2)}\displaystyle\sum_{t,s=1+|j|}^{n} a_{ts}, & r\neq l; \\ 0, & r=l, \end{cases}$$

with $a_{rl} = |X_r - X_l|$, $r,l = 1+|j|,\ldots,n$. $\widetilde{B}_{rl}$ is defined analogously with $b_{rl} = |Y_r - Y_l|$, where $Y_t \equiv X_{t-|j|}$.

The functions ADCV() and ADCF() in **dCovTS** return the empirical quantities $\widehat{V}_X(\cdot)$ and $\widehat{R}_X(\cdot)$ respectively, discussed in Chapter 3. Moreover, based on the definitions of $V_X(\cdot)$ and $R_X(\cdot)$, we observe that $\widehat{V}_X^2(j) = \widehat{V}_X^2(-j)$ and $\widehat{R}_X^2(j) = \widehat{R}_X^2(-j)$, and thus results based on negative lags are omitted from the package. Using the same functions with argument unbiased=TRUE, the results correspond to the unbiased squared quantities $\widetilde{V}_X^2(\cdot)$ and $\widetilde{R}_X^2(\cdot)$. Note that the default option has been set to unbiased=FALSE (corresponding to empirical version discussed in Chapter 3).

We now turn on to the multivariate case. Recall that the pairwise ADCV between $X_{t;r}$ and $X_{t-|j|;m}$, is denoted by $V_{rm}(j)$ and it is defined as the nonnegative square root of

$$V_{rm}^2(j) = \frac{1}{\pi^2} \int_{\mathbb{R}^2} \frac{\left| \phi_{|j|}^{(r,m)}(u,v) - \phi^{(r)}(u)\phi^{(m)}(v) \right|^2}{|u|^2 \, |v|^2} dudv, \quad j = 0, \pm 1, \pm 2, \ldots$$

where $\phi_{|j|}^{(r,m)}(u,v) = E\Big[\exp\Big(i(uX_{t;r} + vX_{t-|j|;m})\Big)\Big]$ denotes the joint characteristic function of $X_{t;r}$ and $X_{t-|j|;m}$ and $\phi^{(r)}(u) := \phi_{|j|}^{(r,m)}(u,0)$, $\phi^{(m)}(v) := \phi_{|j|}^{(r,m)}(0,v)$ are their corresponding marginal characteristic functions. The ADCV matrix, $V(j)$, is then defined by

$$V(j) = \Big[V_{rm}(j)\Big]_{r,m=1}^{d}, \quad j = 0, \pm 1, \pm 2, \ldots$$

The pairwise ADCF between $X_{t;r}$ and $X_{t-|j|;m}$, $R_{rm}(j)$, is a coefficient that lies in the interval $[0,1]$ and also measures dependence and is defined as the positive square root of

$$R_{rm}^2(j) = \frac{V_{rm}^2(j)}{\sqrt{V_{rr}^2(0)}\sqrt{V_{mm}^2(0)}},$$

for $V_{rr}(0)V_{mm}(0) \neq 0$ and zero otherwise. The ADCF matrix of $\mathbf{X}_t$, is then defined as

$$R(j) = \Big[R_{rm}(j)\Big]_{r,m=1}^{d}, \quad j = 0, \pm 1, \pm 2, \ldots$$

Analogously to (5.2), an unbiased estimator of $\widehat{V}_{rm}^2(\cdot)$ is given by

$$\widetilde{V}_{rm}^2(j) = \frac{1}{(n-|j|)(n-|j|-3)} \sum_{t \neq s} \widetilde{A}_{ts}^r \widetilde{B}_{ts}^m,$$

where $\widetilde{A}_{ts}^r$ are computed appropriately, with $a_{ts}^r = |X_{t;r} - X_{s;r}|$, $t, s = 1 + |j|, \ldots, n$. $\widetilde{B}_{ts}^m$ is computed analogously with $b_{ts}^m = |Y_{t;m} - Y_{s;m}|$, where $Y_{t;m} \equiv X_{t-|j|;m}$.

In Chapter 4, we showed that under the null hypothesis of pairwise independence, the empirical counterpart of the squared pairwise ADCV, $\widehat{V}_{rm}^2(\cdot)$, is a degenerate $V$-statistic of order two with a measurable kernel function that is symmetric, continuous and semidefinite. Following the same arguments, one can show that under the null hypothesis of pairwise independence, $\widehat{V}_X^2(\cdot)$ is a degenerate $V$-statistic of order two. For more properties of these functions the reader is referred to Chapter 4.

The sample ADCV matrix, $\widehat{V}(\cdot)$, and ADCF matrix, $\widehat{R}(\cdot)$, whose definitions are discussed in Chapter 4, are calculated from **dCovTS** using the functions mADCV() and mADCF() respectively. The unbiased estimators of ADCV and ADCF matrices are obtained from **dCovTS** using the argument unbiased=TRUE. We note that the functions of **dCovTS** that calculate ADCV and ADCF are mainly based on the functions dcov() and dcor() respectively from **energy** (Rizzo and Szekely, 2014) package.

The distance correlation plots for both univariate and multivariate time series are obtained by the ADCFplot() and mADCFplot() functions respectively, where the shown critical values (blue dotted horizontal line) are the simultaneous 95% empirical critical values computed by employing the independent wild bootstrap described in Section 4.5.2 (argument bootMethod="Wild Bootstrap"). In the case of a univariate time series, we also use the subsampling approach suggested by Zhou (2012, Section 5.1) in order to compute the pairwise 95% critical values (argument bootMethod="Subsampling"). In addition, the package provides the ordinary independent bootstrap methodology to derive simultaneous 95% critical values for the ADCF plots (argument bootMethod="Independent Bootstrap") for both univariate and multivariate time series. Recall that these are computed by using the biased definition of distance covariance and correlation.

### 5.2.2 Univariate and Multivariate Tests of Independence Based on Distance Covariance via dCovTS

In Chapters 3 and 4, we developed a distance covariance testing methodology considering an increasing number of lags by employing Hong's (1999) generalized spectral domain methodology. We recall the resulting test statistics below. We proposed the following

Portmanteau type statistic based on ADCV

$$T_n = \sum_{j=1}^{n-1} (n-j) k^2(j/p) \widehat{V}_X^2(j), \tag{5.3}$$

where $p$ is a bandwidth of the form $p = cn^\lambda$ for $c > 0$ $\lambda \in (0,1)$, $k(\cdot)$ is a univariate kernel function satisfying Assumption 4 (stated in Chapters 3 and 4). kernelFun() in **dCovTS** computes a number of such kernel functions including the truncated (the default), Bartlett, Daniell, QS and Parzen kernels.

We also considered a similar test statistic based on ADCF

$$\sum_{j=1}^{n-1} (n-j) k^2(j/p) \widehat{R}_X^2(j). \tag{5.4}$$

The function UnivTest() from **dCovTS** package performs univariate tests of independence based on (5.3) and its rescaled version (5.4), using the arguments testType="covariance" (the default) and testType="correlation" respectively.

The proposed test statistic for testing pairwise independence in a multivariate time series framework is based on the ADCV matrix, $V(\cdot)$, and it is given by

$$\widetilde{T}_n = \sum_{j=1}^{n-1} (n-j) k^2(j/p) \text{tr}\{\widehat{V}^*(j)\widehat{V}(j)\}. \tag{5.5}$$

where $k(\cdot)$ is a univariate kernel function and $p$ is a bandwidth as both described before and $\widehat{V}^*(\cdot)$ denotes the complex conjugate matrix of $\widehat{V}(\cdot)$. In Chapter 4, we formed the statistic (5.5) in terms of the ADCF matrix as follows

$$\overline{T}_n = \sum_{j=1}^{n-1} (n-j) k^2(j/p) \text{tr}\{\widehat{V}^*(j)\widehat{D}^{-1}\widehat{V}(j)\widehat{D}^{-1}\}. \tag{5.6}$$

where $D = diag\{V_{rr}(0), r = 1, 2, \ldots, d\}$. The multivariate tests of independence based on $\widetilde{T}_n$ and $\overline{T}_n$ are performed via mADCVtest() and mADCFtest() respectively in **dCovTS** package.

All test statistics $T_n$, $\widetilde{T}_n$ and $\overline{T}_n$ of equations (5.3), (5.5) and (5.6) respectively, are functions of degenerate $V$-statistics of order two. In the case of independent data, Dehling and Mikosch (1994) studied the wild bootstrap distribution for degenerate $U$-statistics. Based on Leucht and Neumann (2013a,b) we suggest the use of a new variant of wild bootstrap (Shao, 2010), the so-called independent wild bootstrap suitable for dependent data. Thus, the empirical $p$-values of the tests are derived based on this methodology. In Chapter 4, all the steps followed are described in detail. The package also provides the ordinary independent bootstrap methodology to derive empirical $p$-values of the tests.

The computation of the bootstrap replications, and thus the empirical $p$-values and the critical values, can be distributed to multiple cores simultaneously (argument parallel=TRUE). To do this, the **doParallel** (Analytics and Weston, 2015) package needs to be installed first, in order to register a computing cluster.

## 5.3 Applications

In this section, we apply **dCovTS** on both univariate and multivariate real data. Note that for the reader's convenience we include the main R commands used with the help of **dCovTS** for the analysis of these examples.

### 5.3.1 Univariate Time Series

#### Pollution, Temperature and Mortality Data

We first consider the pollution, temperature and mortality data measured daily in Los Angeles County over the 10 year period 1970-1979 (Shumway et al., 1988). The data are available in our package by the argument MortTempPart and contain 508 observations and 3 variables representing the mortality (cmort), temperature (tempr) and pollutant particulates (part) data.

```
> library(dCovTS) #loading the package
> data(MortTempPart)
> MortTempPart[1:10,] # the first ten observations
```

```
     cmort tempr  part
1    97.85 72.38 72.72
2   104.64 67.19 49.60
3    94.36 62.94 55.68
4    98.05 72.49 55.16
5    95.85 74.25 66.02
6    95.98 67.88 44.01
7    88.63 74.20 47.83
8    90.85 74.88 43.60
9    92.06 64.17 24.99
10   88.75 67.09 40.41
> attach(MortTempPart)
```

Following the analysis of Shumway and Stoffer (2011), the possible effects of temperature ($T_t$) and pollutant particulates ($P_t$) on daily cardiovascular mortality ($M_t$) are examined via regression. In particular, once the temperature is adjusted for its mean ($T_. = 74.3$), we fit the following regression model using the function lm() already existing in R

$$\begin{aligned} \widehat{M_t} &= 2831.49 - 1.396_{(0.101)} t - 0.472_{(0.032)}(T_t - T.) \\ &\quad + 0.023_{(0.003)}(T_t - T.)^2 + 0.255_{(0.019)} P_t, \end{aligned} \tag{5.7}$$

where the standard errors of the estimators are given in parentheses. Figure 5.1 provides the ACF, partial correlation (PACF) and ADCF plots of the residuals of model (5.7). The plots shown in Figure 5.1 suggest an AR(2) process for the residuals. The new fit is

$$\begin{aligned} \widehat{M_t} &= 3075.15 - 1.517_{(0.423)} t - 0.019_{(0.050)}(T_t - T.) \\ &\quad + 0.015_{(0.002)}(T_t - T.)^2 + 0.155_{(0.027)} P_t, \end{aligned} \tag{5.8}$$

where the standard errors of the estimators are given in parentheses. The above model fit is derived using arima() function of R. The correlation plots for the residuals from the new model (5.8) are shown in Figure 5.2 indicating that there is no serial dependence. The calls for both model fits and their diagnostic plots are given below. ADCF plots (lower

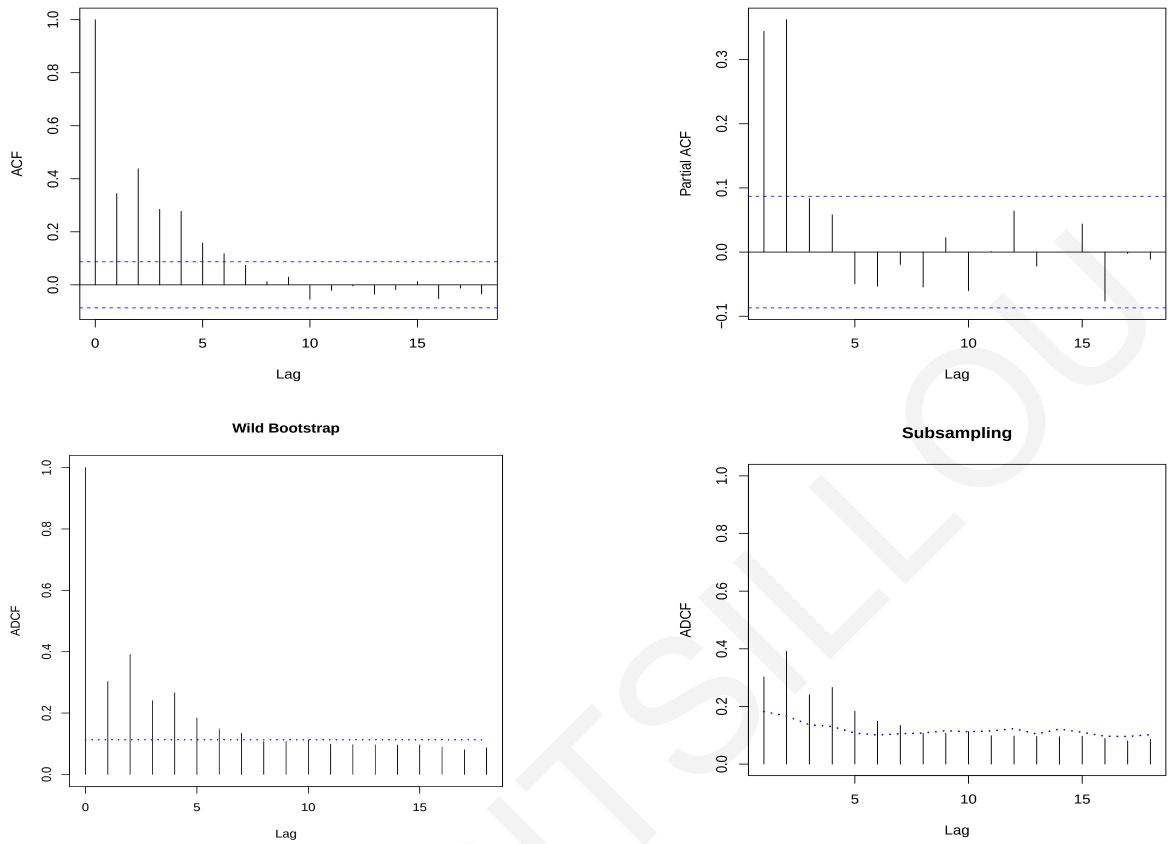



Figure 5.1: Sample ACF, PACF and ADCF plots of the mortality residuals of model (5.7).

plots of Figures 5.1 and 5.2) are constructed using both resampling schemes explained in the previous chapter: independent wild bootstrap (with $b = 499$ replications) and Subsampling.

```
> temp  = tempr-mean(tempr)  # center temperature
> temp2 = temp^2
> trend = time(cmort)
> fit = lm(cmort~ trend + temp + temp2 + part, na.action=NULL)
> Residuals <- as.numeric(resid(fit))
> ##Correlation plots
> acf(Residuals,lag.max=18,main="")
> pacf(Residuals,lag.max=18,main="")
> ADCFplot(Residuals,MaxLag=18,main="Wild Bootstrap",bootMethod="Wild")
> ADCFplot(Residuals,MaxLag=18,main="Subsampling",bootMethod="Subsampling")
```

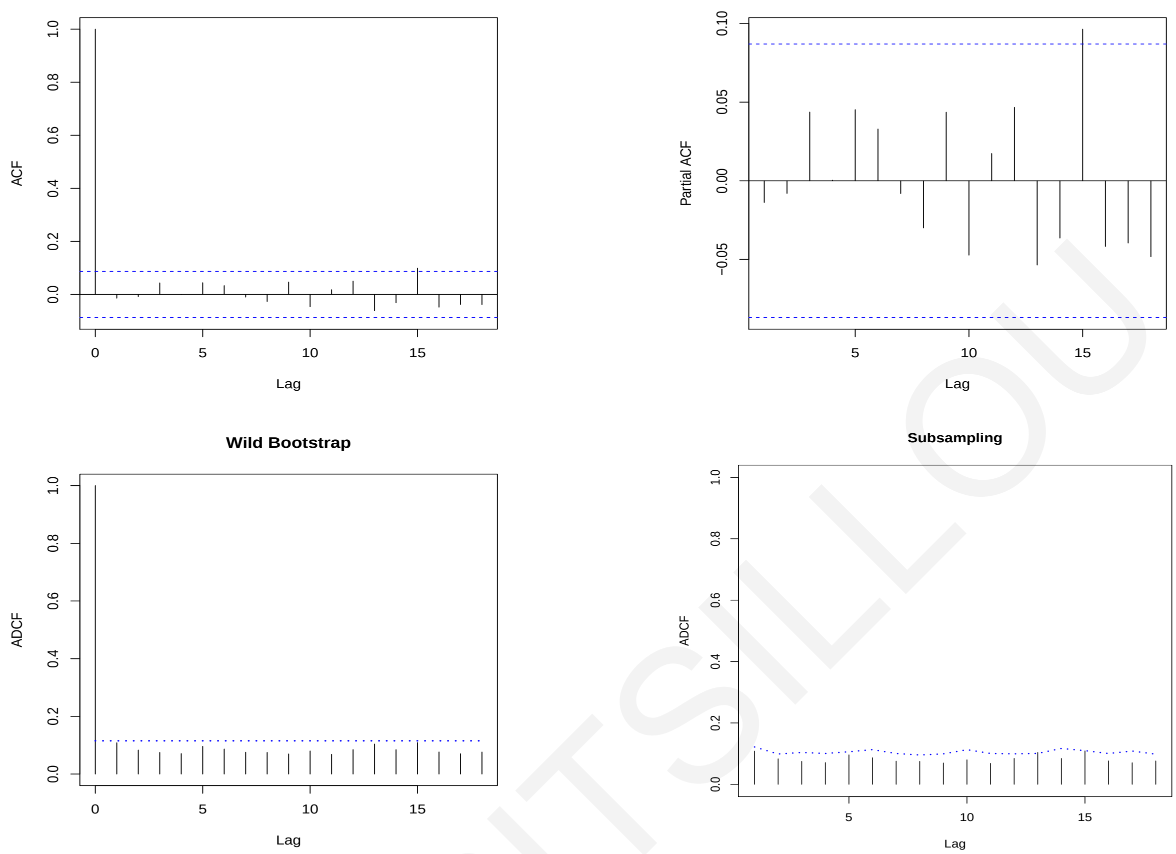



Figure 5.2: Sample ACF, PACF and ADCF plots of the mortality residuals of model (5.8) indicating that the new residuals can be taken as white noise.

```
> fit2 <- arima(cmort, order=c(2,0,0), xreg=cbind(trend,temp,temp2,part))
> Residuals2 <- as.numeric(residuals(fit2))
> ##Correlation plots
> acf(Residuals2,lag.max=18,main="")
> pacf(Residuals2,lag.max=18,main="")
> ADCFplot(Residuals2,MaxLag=18,main="Wild Bootstrap",bootMethod="Wild")
> ADCFplot(Residuals2,MaxLag=18,main="Subsampling",bootMethod="Subsampling")
```

To formally confirm the absence of any serial dependence among the new residuals of model (5.8), as shown in Figure 5.2, we perform univariate tests of independence based on the test statistic $T_n$ given in (5.3) using the UnivTest() function from our package with argument testType="covariance", which is the default. In order to examine the effect of using different bandwidths, we choose $p = [3n^{\lambda}]$ for $\lambda$=0.1, 0.2 and 0.3, that is $p = 6$, 11, and 20 and we apply Bartlett kernel. The resulting $p$-values are 0.118, 0.170 and 0.208 respectively suggesting acceptance of independence. P-values are calculated for $b = 499$ independent wild bootstrap replications. Bootstrap procedure can be computed on multiple cores simultaneously (argument parallel=TRUE) in order to be computationally less expensive (they take about 10, 14 and 23 seconds respectively on a standard laptop with Intel Core i5 system and CPU 2.30 GHz):

```
> UnivTest(Residuals2, type="bartlett", p=6, b=499, parallel=TRUE)

        Univariate test of independence based on distance covariance

data: Residuals2, kernel type: bartlett, bandwidth=6, replicates 499
Tn = 67.7344, p-value = 0.118

> UnivTest(Residuals2, type="bartlett", p=11, b=499, parallel=TRUE)

        Univariate test of independence based on distance covariance

data:  Residuals2, kernel type: bartlett, bandwidth=11, replicates 499
Tn = 125.6674, p-value = 0.170

> UnivTest(Residuals2, type="bartlett", p=20, b=499, parallel=TRUE)

        Univariate test of independence based on distance covariance

data:  Residuals2, kernel type: bartlett, bandwidth=20, replicates 499
Tn = 225.9266, p-value = 0.208
```

We compare the proposed test statistic with other test statistics, given in Chapter 3, in order to check its performance. We recall these statistics below. We consider the Box-

Pierce (Box and Pierce, 1970) test statistic,

$$\text{BP} = n \sum_{j=1}^{p} \hat{\rho}^2(j),$$

the Ljung-Box (Ljung and Box, 1978) test statistic,

$$\text{LB} = n(n+2) \sum_{j=1}^{p} (n-j)^{-1} \hat{\rho}^2(j),$$

the test statistic proposed by Hong (1996),

$$T_n^{(1)} = n \sum_{j=1}^{n-1} k^2(j/p) \hat{\rho}^2(j)$$

and the test statistic suggested by Hong (1999),

$$T_n^{(2)} = \int_{\mathbb{R}^2} \sum_{j=1}^{n-1} (n-j) k^2(j/p) \left| \hat{\sigma}_j(u,v) \right| d\mathcal{W}(u,v), \tag{5.9}$$

where $\mathcal{W}(\cdot,\cdot) : \mathbb{R}^2 \to \mathbb{R}$ is an arbitrary nondecreasing function with bounded total variation. For the aforestated bandwidth values, all these alternative test statistic give large $p$-values indicating the absence of any serial dependence among the new residuals. More precisely, BP and LB give 0.848, 0.906, 0.170 and 0.844, 0.901, 0.142 respectively. BP and LB based tests are performed in R by the function Box.test() as follows:

```
> box1 <-  Box.test(Residuals2,lag=6)
> box2 <-  Box.test(Residuals2,lag=11)
> box3 <-  Box.test(Residuals2,lag=20)
> ljung1 <-  Box.test(Residuals2,lag=6,type="Ljung")
> ljung2 <-  Box.test(Residuals2,lag=11,type="Ljung")
> ljung3 <-  Box.test(Residuals2,lag=20,type="Ljung")
```

The $p$-values obtained by $T_n^{(1)}$ are 0.896, 0.930 and 0.870 respectively. $T_n^{(2)}$ gives the following $p$-values: 0.854, 0.752 and 0.504 respectively. $T_n^{(1)}$ and $T_n^{(2)}$ are calculated by employing the Bartlett kernel. These $p$-values are calculated for $b = 499$ ordinary bootstrap replications. The R functions for constructing these test statistics are beyond the scope of this

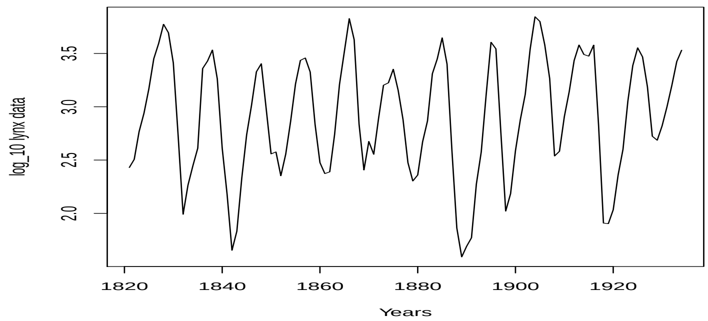



Figure 5.3: The time-series plot of the Canadian lynx data.

chapter and are available upon request.

### The Canadian Lynx Data

We now study the numbers of the Canadian lynx trapped in the Mackenzie River district of northwest Canada, recorded annually from 1821 to 1934. The data are available from the R package **datasets** by the name lynx. The logarithms of the lynx data (at a log 10 scale) are plotted in Figure 5.3, exhibiting a periodic fluctuation with asymmetric population cycles.

```
> ## Reading and plotting the lynx data
> log.lynx <- log10(lynx)
> plot.ts(log.lynx,ylab="log_10 lynx data",xlab="Years")
```

Following the detailed analysis of Fan and Yao (2003, p. 136), we fit the following two-regime TAR model

$$X_t = \begin{cases} 0.546 + 1.032X_{t-1} - 0.173X_{t-2} + 0.171X_{t-3} - 0.431X_{t-4} & \\ +0.332X_{t-5} - 0.284X_{t-6} + 0.210X_{t-7} + \epsilon_t^{(1)}, & X_{t-2} \leq 3.116; \\ 2.632 + 1.492X_{t-1} - 1.324X_{t-2} + \epsilon_t^{(2)}, & X_{t-2} > 3.116, \end{cases} \tag{5.10}$$

where $\{\epsilon_t^{(i)}\}$ are i.i.d. sequences of N(0,1) random variable, for $i = 1, 2$. The above model is derived in R, using the function tar() from the **TSA** (Chan and Ripley, 2012) package. Like all statistical fitting, we look at the behavior of the residuals as the most popular diagnostic tool. Figure 5.4 shows the sample ADCF plots of the residuals of model (5.10), suggesting no serial dependence among the residuals.

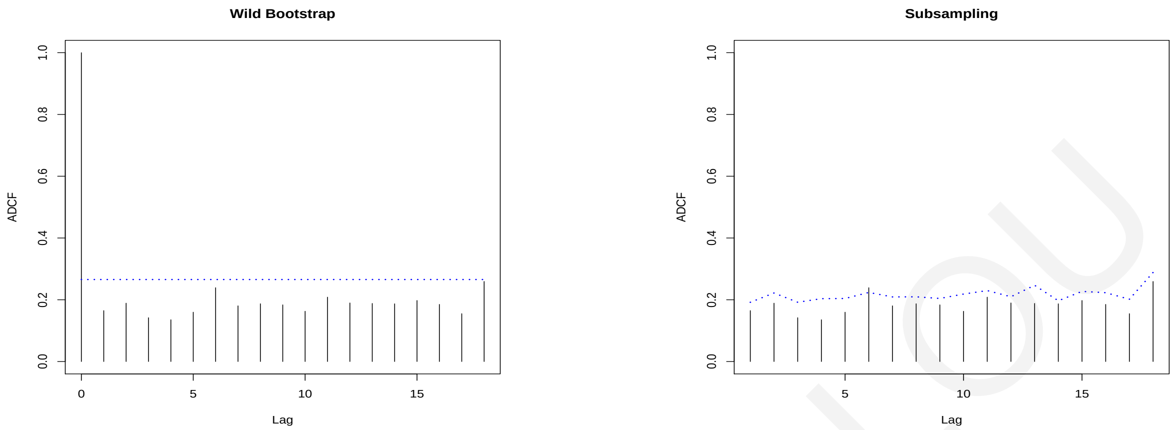



Figure 5.4: Sample ADCF plots of the residuals of the TAR model (5.10) fitted to the Canadian lynx data.

The R-code for the above TAR model fit and the corresponding diagnostic plots is given below:

```
> ## TAR model fit
> library(TSA)
> fit <- tar(log.lynx,p1=10,p2=10,d=2,threshold=3.116, estimate.thd = F)
> res <- residuals(fit)
> res <- res[!is.na(res)]     #removing the NA's
> ADCFplot(res,MaxLag=18,bootMethod="Wild",main="Wild Bootstrap")
> ADCFplot(res,MaxLag=18,bootMethod="Subs",main="Subsampling")
```

To formally confirm the adequacy of model (5.10), we construct tests of independence among the residuals. The resulting $p$-values are provided in Table 5.3, for various choices of the bandwidth parameter $p$. All test statistics yield large $p$-values indicating the absence of serial dependence among the residuals. We note that the kernel-based test statistics $T_n$, $T_n^{(1)}$ and $T_n^{(2)}$ are computed based on the Bartlett kernel. However, any other choices of the kernel function and the bandwidth parameter $p$, yield the same conclusions. Indeed, Figure 5.5 presents the $p$-values obtained by constructing tests of independence among the residuals based on the proposed statistic $T_n$, for various choices of the kernel function and the bandwidth parameter, $p$. The main message from these plots is that the statistic

$T_n$ is not very sensitive to the choice of the kernel function and the bandwidth parameter $p$. P-values are calculated for $b = 499$ independent wild bootstrap replications. The R commands for the tests of independence based on $T_n$, BP and LB are given below:

```
> #Tn tests
> UnivTest(res,type="bar",p=3,b=499,parallel=TRUE)

        Univariate test of independence based on distance covariance

data:  res, kernel type: bartlett, bandwidth=3, replicates 499
Tn = 0.0243, p-value = 0.552

> UnivTest(res,type="bar",p=7,b=499,parallel=TRUE)

        Univariate test of independence based on distance covariance

data:  res, kernel type: bartlett, bandwidth=7, replicates 499
Tn = 0.0774, p-value = 0.632

> UnivTest(res,type="bar",p=17,b=499,parallel=TRUE)

        Univariate test of independence based on distance covariance

data:  res, kernel type: bartlett, bandwidth=17, replicates 499
Tn = 0.2343, p-value = 0.586

> #BP and LB tests
> box1 <-  Box.test(res,lag=3)
> box2 <-  Box.test(res,lag=7)
> box3 <-  Box.test(res,lag=17)
> ljung1 <-  Box.test(res,lag=3,type="Ljung")
> ljung2 <-  Box.test(res,lag=7,type="Ljung")
```

Table 5.3: P-values obtained by constructing tests of independence among the residuals after fitting the TAR model (5.10) to Canadian lynx data. All test statistics are calculated for $b = 499$ independent wild bootstrap replications. The statistics $T_n$, $T_n^{(1)}$ and $T_n^{(2)}$ are computed by employing the Bartlett kernel.

| $p$ | $T_n$ | $T_n^{(2)}$ | BP | LB | $T_n^{(1)}$ |
|---|---|---|---|---|---|
| 3 | 0.552 | 0.804 | 0.978 | 0.976 | 0.958 |
| 7 | 0.632 | 0.916 | 0.904 | 0.886 | 0.934 |
| 17 | 0.586 | 0.770 | 0.699 | 0.564 | 0.738 |

```
> ljung3 <-  Box.test(res,lag=17,type="Ljung")
```

### Monthly Log Returns of Intel Stock

In this final example, we analyze the monthly log stock returns of Intel Corporation from January 1973 to December 2003, for 372 observations. The univariate time series is available from **FinTS** (Graves, 2014) package under the name m.intc7303:

```
> library(FinTS)
> data(m.intc7303)
> data <- zoo(m.intc7303,as.yearmon(index(m.intc7303)))
> data <- ts(data)
> ldata <- log(data+1)   #convert into log returns

> #ACF and ADCF plots of the series
> acf(ldata,main="log_intel",lag.max=18)
> acf(ldata^2,main="log_intel^2",lag.max=18)
> ADCFplot(ldata,main="Wild Bootstrap",ylim=c(0,0.2))
> ADCFplot(ldata,bootMethod="Subs",main="Subsampling",ylim=c(0,0.2))
```

Figure 5.6 shows the sample ACF of the log returns and their squared values (upper panel), as well as the sample ADCF plots of the log series (lower panel). Clearly, the sample ACF of the squared returns and the sample ADCF plots of the log returns indicate the existence of conditional heteroscedasticity. Thus, we fit the following GARCH(1,1) model

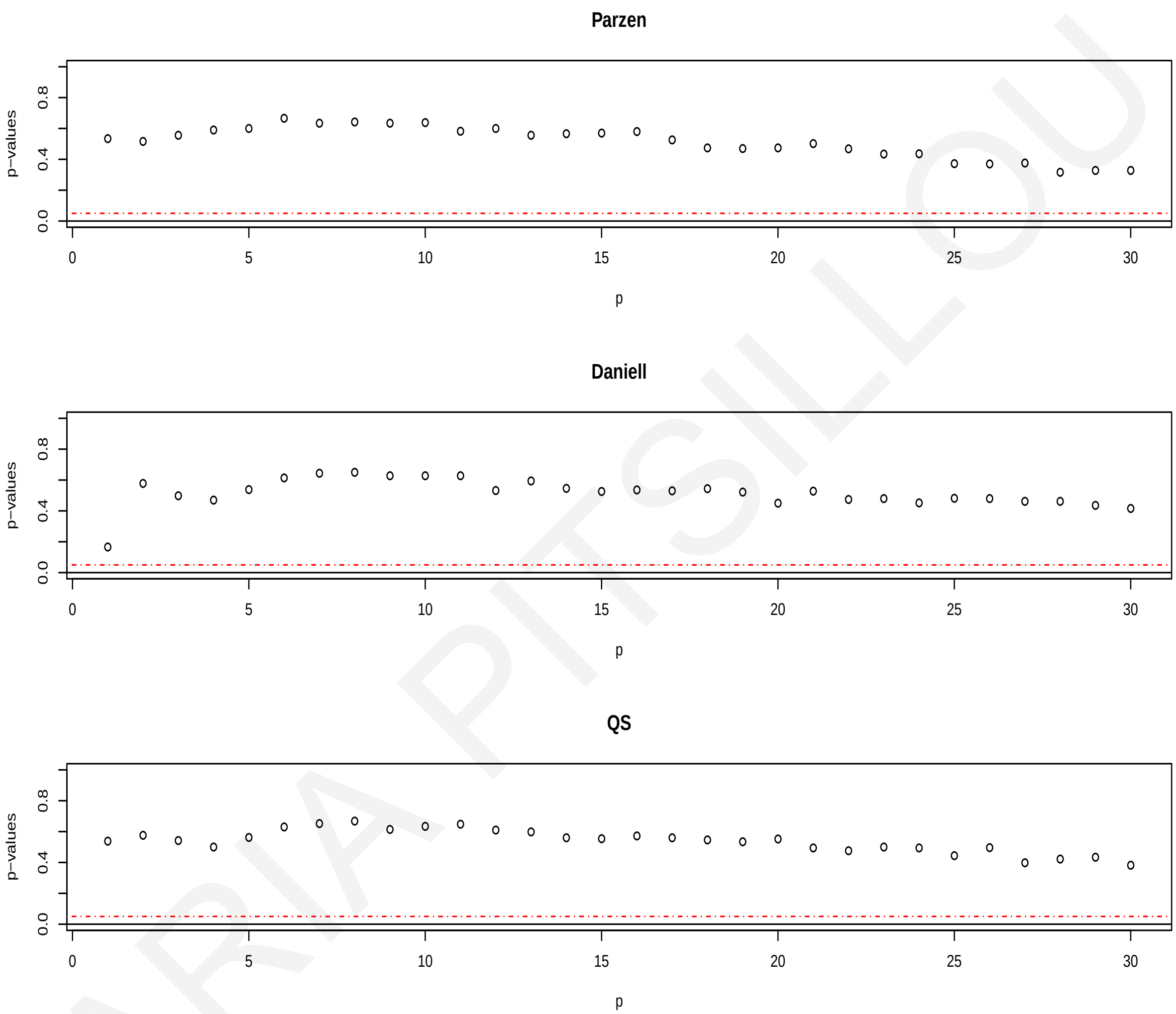



Figure 5.5: P-values obtained by performing tests of independence based on $T_n$, among the residuals of the TAR model (5.10) for the Canadian lynx data. Results are based on $b = 499$ independent wild bootstrap replications for various choices of the bandwidth parameter, $p$, and for three different choices of the kernel function, the Parzen, the Daniell and the QS. The red dotted line is marked at the significance level 0.05.

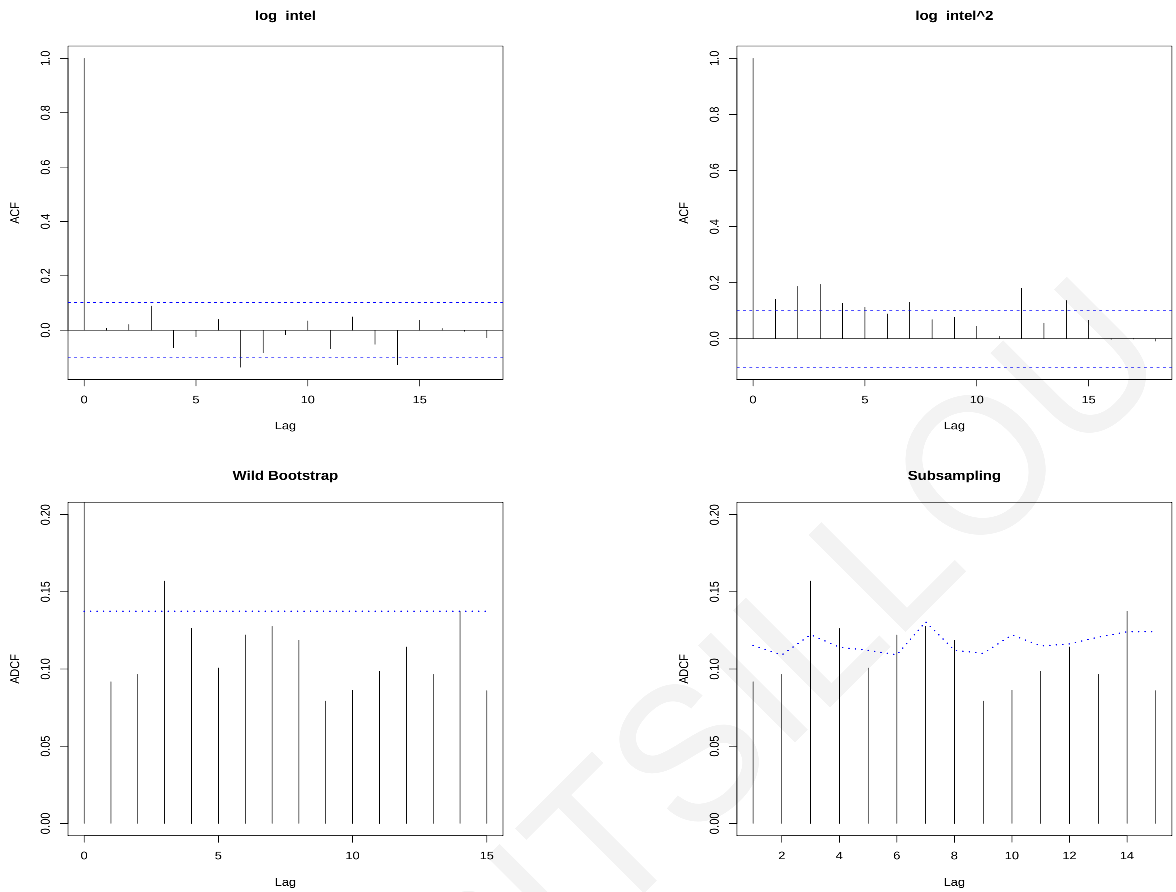



Figure 5.6: Sample ACF and ADCF plots of the monthly log returns of intel stock.

with normal innovations

$$X_t = \sigma_t \epsilon_t, \quad \sigma_t^2 = 0.001 + 0.087X_{t-1}^2 + 0.846\sigma_{t-1}^2. \tag{5.11}$$

The above model fit is derived in R using garch() function from the package **tseries** (Trapletti and Hornik, 2015). After model fitting, we look at the behavior of the standardized residuals by performing tests of independence among the residuals. All test statistics give large $p$-values (Table 5.4) suggesting that the model used is suitable for the data. The R commands for the model fit and the corresponding tests of independence based on $T_n$, BP and LB are given below:

```
> # model fit
> library(tseries)
> model <- garch(ldata, order = c(1,1))
> res <- residuals(model)
```

Table 5.4: P-values obtained by constructing tests of independence among the residuals after fitting the GARCH(1,1) model given in (5.11) to Intel stock data. All test statistics are calculated for $b = 499$ bootstrap replications. The statistics $T_n$, $T_n^{(1)}$ and $T_n^{(2)}$ are computed by employing the Bartlett kernel.

| $p$ | $T_n$ | $T_n^{(2)}$ | BP | LB | $T_n^{(1)}$ |
|---|---|---|---|---|---|
| 4 | 0.906 | 0.950 | 0.496 | 0.488 | 0.756 |
| 11 | 0.332 | 0.498 | 0.542 | 0.523 | 0.552 |
| 35 | 0.386 | 0.550 | 0.734 | 0.663 | 0.464 |

```
> res <- res[!is.na(res)]
> std.res <- res/sd(res)

> # Tests of independence
> UnivTest(std.res,"bar",p=4,b=499, parallel=TRUE)

        Univariate test of independence based on distance covariance

data:  std.res, kernel type: bartlett, bandwidth=4, replicates 499
Tn = 0.66593, p-value = 0.906

> UnivTest(std.res,"bar",p=11,b=499, parallel=TRUE)

        Univariate test of independence based on distance covariance

data:  std.res, kernel type: bartlett, bandwidth=11, replicates 499
Tn = 3.9101, p-value = 0.332

> UnivTest(std.res,"bar",p=35,b=499, parallel=TRUE)

        Univariate test of independence based on distance covariance

data:  std.res, kernel type: bartlett, bandwidth=35, replicates 499
Tn = 13.431, p-value = 0.386
```

```
> box1 <- Box.test(std.res,lag=4)
> box2 <- Box.test(std.res,lag=11)
> box3 <- Box.test(std.res,lag=35)
> ljung1 <- Box.test(std.res,lag=4,type="Ljung")
> ljung2 <- Box.test(std.res,lag=11,type="Ljung")
> ljung3 <- Box.test(std.res,lag=35,type="Ljung")
```

### 5.3.2 Multivariate Time Series

#### IBM and S&P 500 Data

We now analyze the monthly log returns of the stocks of International Business Machines (IBM) and the S&P 500 composite index starting from 29 May 1936 to 28 November 1975 for 474 observations. A larger dataset is available in our package by the object ibmSp500 starting from January 1926 for 1032 observations. It is actually a combination of two smaller datasets: the first one was first reported by Tsay (2010) and the second one was first reported by Tsay (2014). Below, we give the R commands for reading this smaller dataset from the package:

```
> data(ibmSp500)
> new_data <- ibmSp500[224:588,2:3]
> lseries <- log(new_data+1)
> at=scale(lseries,center=T,scale=F)
```

We first construct test of independence among the series based on the multivariate Ljung-Box statistic (Hosking, 1980; Li and McLeod, 1981) defined by

$$mLB = n^2 \sum_{j=1}^{p} (n-j)^{-1} \text{trace}\{\widehat{\Gamma}'(j)\widehat{\Gamma}^{-1}(0)\widehat{\Gamma}(j)\widehat{\Gamma}^{-1}(0)\}.$$

Choosing $p = [3n^{\lambda}]$ for $\lambda$=0.1, 0.2 and 0.3, that is $p = 6$, 11 and 20, $mLB$ gives large $p$-values (0.090, 0.159 and 0.235 respectively) suggesting no serial correlation among observations. However, constructing test of independence among the squared series, $mLB$

yields low $p$-values close to zero implying strong linear dependence. This confirms the conditional heteroscedasticity in the monthly log returns. These tests are preformed in R using LjungBox() function from **portes** (Mahdi and McLeod, 2012) package. As expected, the proposed test statistic $\overline{T}_n$ detects the dependence among the original series giving low $p$-values (0.030, 0.014 and 0.016 respectively). The statistic $\overline{T}_n$ is computed by employing the Bartlett kernel. As in the univariate examples described above, to speed up the computation of the empirical $p$-values for our proposed test, the bootstrap procedure can be computed on multiple cores simultaneously (argument parallel=TRUE). The calls for all the above multivariate tests of independence can be found below ($\overline{T}_n$ tests take about 1, 2 and 3 minutes respectively for $b = 499$ bootstrap replications on a standard laptop with Intel Core i5 system and CPU 2.30 GHz):

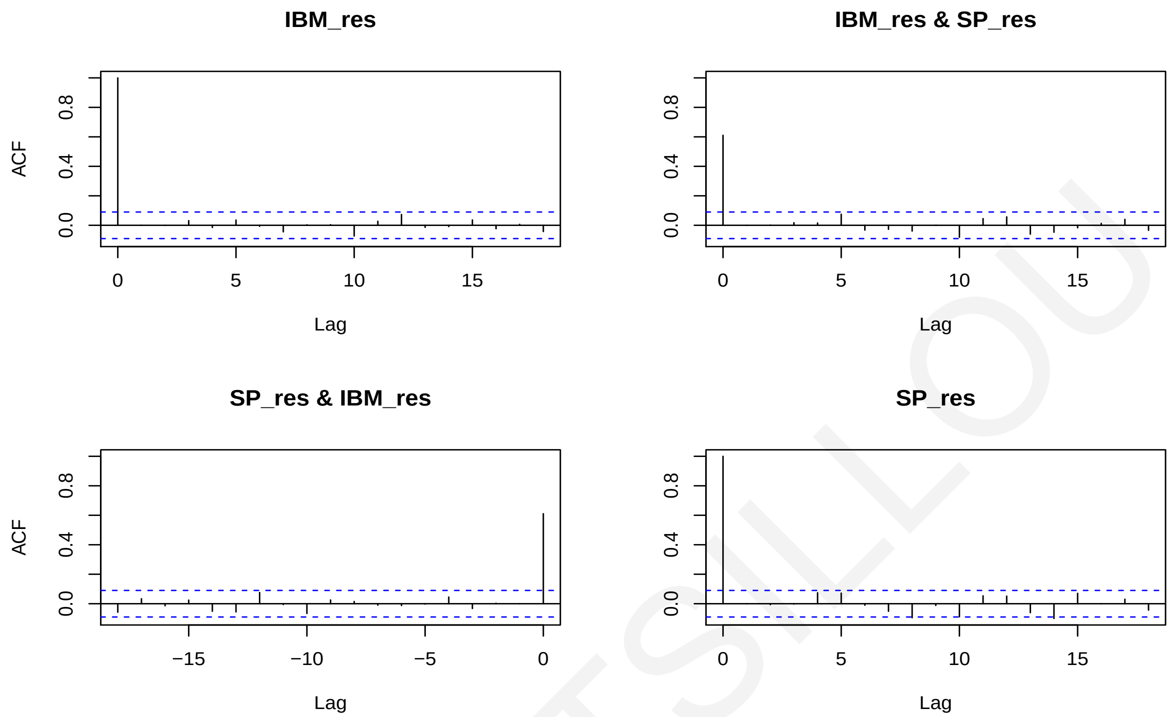
```
> ##mLB tests
> test1 <- LjungBox(at,c(6,11,20))
> test2 <- LjungBox(at^2,c(6,11,20))

> ## \bar{T}_n tests
> mADCFtest(at,p=6,b=499,type="bar",parallel=TRUE)

        Multivariate test of independence based on distance correlation

data:  at, kernel type: bartlett, bandwidth=6, replicates 499
Tnbar = 32.4818, p-value = 0.03

> mADCFtest(at,"bartlett",p=11,type="bar",b=499,parallel=TRUE)

        Multivariate test of independence based on distance correlation

data:  at, kernel type: bartlett, bandwidth=11, replicates 499
Tnbar = 62.2055, p-value = 0.014

> mADCFtest(at,"bartlett",p=20,type="bar",b=499,parallel=TRUE)
```

```
        Multivariate test of independence based on distance correlation


data:  at, kernel type: bartlett, bandwidth=20, replicates 499
Tnbar = 113.405, p-value = 0.016
```

Assuming that the bivariate log returns follows a VAR model and employing the AIC to choose its best order, we obtain that a VAR(2) model fits well the data. Figure 5.7 shows the ACF plots (upper panel) and ADCF plots (lower panel) of the residuals after fitting a VAR(2) model to the original bivariate log return series using the function VAR() from the **MTS** (Tsay, 2015) package. Based on these plots, the residuals of VAR(2) model do not have any strong dependence. Constructing tests of independence based on $\overline{T}_n$ and $mLB$ for the same choices of bandwidth, $p$ =6, 11, 20, we confirm this visual result. The resulting $p$-values are given in the following R demonstration:

```
> ##ACF and ADCF plots


> library(MTS)
> model <- VAR(at,2)
> resids <- residuals(model)
> colnames(resids) <- c("IBM_res","SP_res")
> windows(9,6)
> acf(resids,lag.max=18)
> mADCFplot(resids,MaxLag=18,ylim=c(0,0.25))


> ## Tests of independence based on \bar{T}_n
> mADCFtest(resids,"bartlett",p=6,b=499,type="bar",parallel=TRUE)


        Multivariate test of independence based on distance correlation


data:  resids, kernel type: bartlett, bandwidth=6, replicates 499
Tnbar = 25.4731, p-value = 0.254
```

(a)




(b)

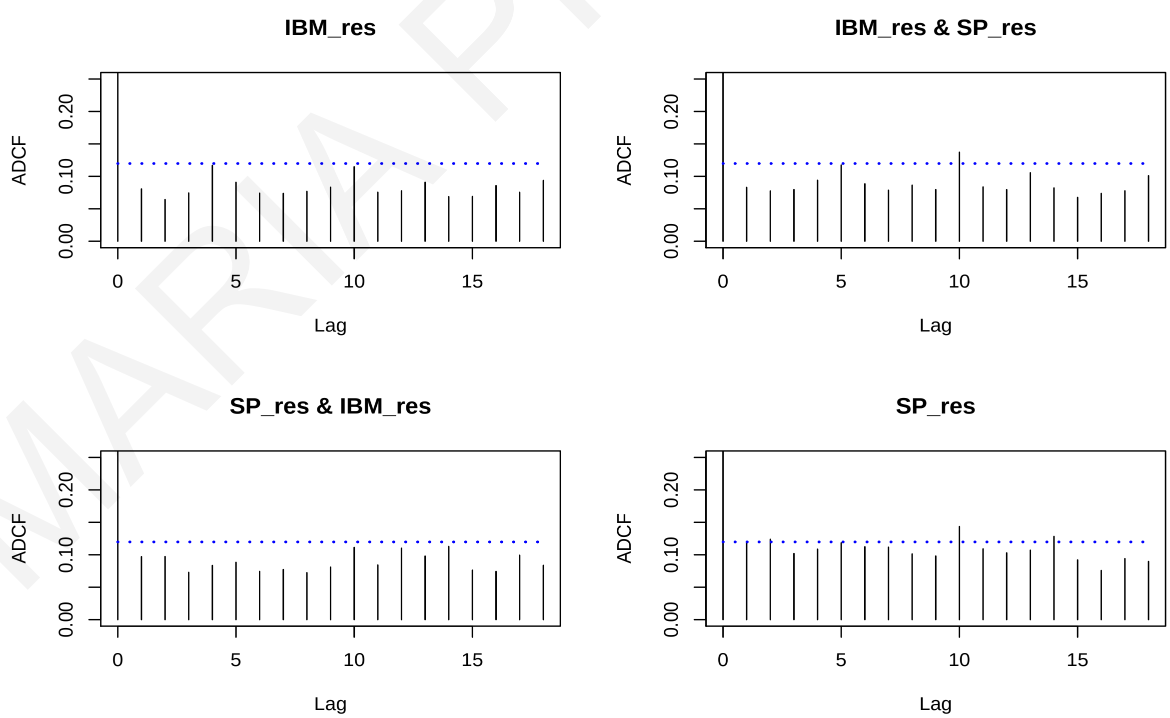



Figure 5.7: The sample ACF (upper panel) and sample ADCF (lower panel) of the residuals after fitting VAR(2) model to the bivariate series IBM and S&P500.

```
> mADCFtest(resids,"bartlett",p=11,type="bar",b=499,parallel=TRUE)

        Multivariate test of independence based on distance correlation

data:  resids, kernel type: bartlett, bandwidth=11, replicates 499
Tnbar = 52.9218, p-value = 0.19

> mADCFtest(resids,"bartlett",p=20,type="bar",b=499,parallel=TRUE)

        Multivariate test of independence based on distance correlation

data:  resids, kernel type: bartlett, bandwidth=20, replicates 499
Tnbar = 103.2249, p-value = 0.098

> ## Tests of independence based on mLB
> LjungBox(resids,lags=c(6,11,20))
 Lags Statistic df    pvalue
    6  13.44822 24 0.9581479
   11  35.69476 44 0.8094291
   20  68.74206 80 0.8111627
```

**Growth Rates of Real Gross Domestic Product of UK, Canada and US**

We now consider the quarterly percentage growth rates of real gross domestic product (GDP) of UK, Canada and US from the first quarter of 1980 to the second quarter of 2011. The three dimensional time series corresponding to the data is available from **MTS** (Tsay, 2015) package under the name qgdp:

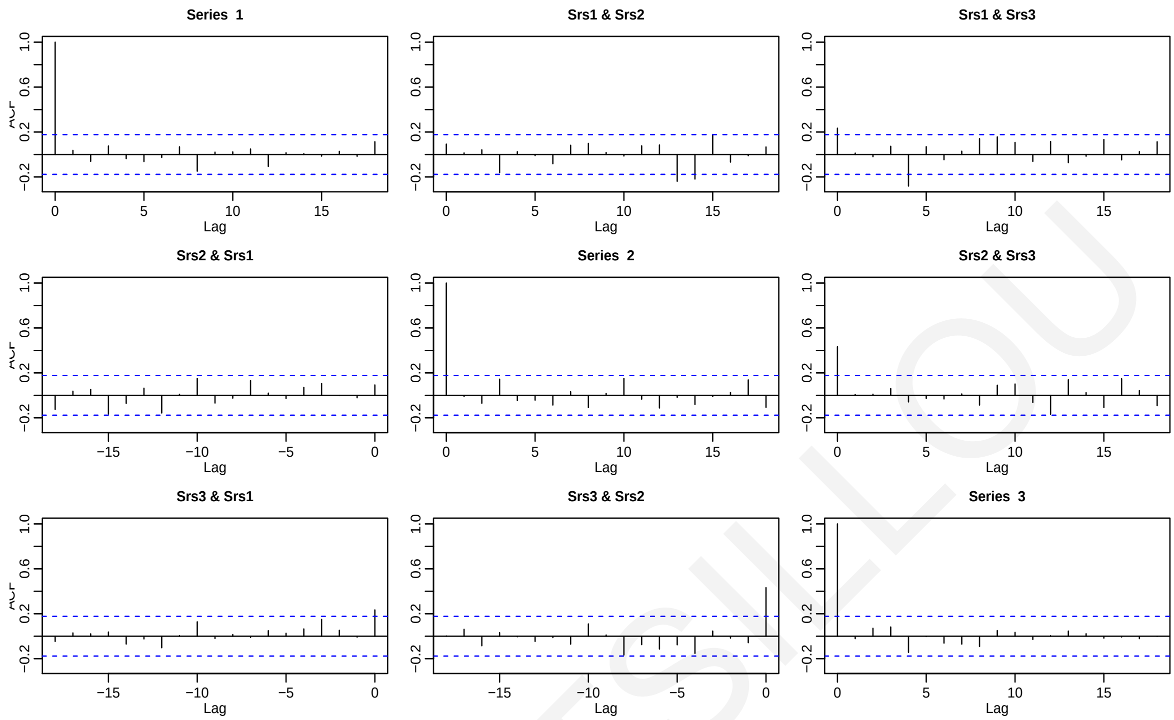
```
> library(MTS)
> data("mts-examples",package="MTS")
> Y <- qgdp[,3:5]
```

```
> gdp <- log(Y)
> z=gdp[2:126,]-gdp[1:125,] ## growth rates
> z=z*100 # percentage growth rates
```

Following the analysis of Tsay (2014, p. 51), we employ a VAR(2) model to the data, where the VAR order was determined by the AIC criterion. After model fitting we perform a residual analysis, by looking at the sample ACF plots (upper panel) and ADCF plots (lower panel) of Figure 5.8 of the three residual series. These plots fail to show any strong dependence among the residuals. Indeed, constructing tests of independence based on $\overline{T}_n$ and $mLB$, we get large $p$-values for bandwidth parameters $p$=6, 10 and 18. In particular, the corresponding $p$-values given by $\overline{T}_n$ and $mLB$ are 0.308, 0.226, 0.102 and 0.885, 0.809, 0.769 respectively. Note that the statistic $\overline{T}_n$ was computed by employing a Parzen kernel. The R commands for the VAR(2) model fit, the corresponding residual plots of Figure 5.8 and the tests of independence mentioned above are given below:

```
> model <- VAR(z,2)
> resi <- model$residuals
> acf(resi,lag.max=18)
> mADCFplot(resi,MaxLag=18)

### Tests of independence based on \bar{T}_n

> mADCFtest(resi,type="parzen",p=6,b=499,parallel=TRUE)

        Multivariate test of independence based on distance correlation

data:  resi, kernel type: parzen, bandwidth=6, replicates 499
Tnbar = 98.838, p-value = 0.308

> mADCFtest(resi,type="parzen",p=10,b=499,parallel=TRUE)

        Multivariate test of independence based on distance correlation
```

(a)




(b)

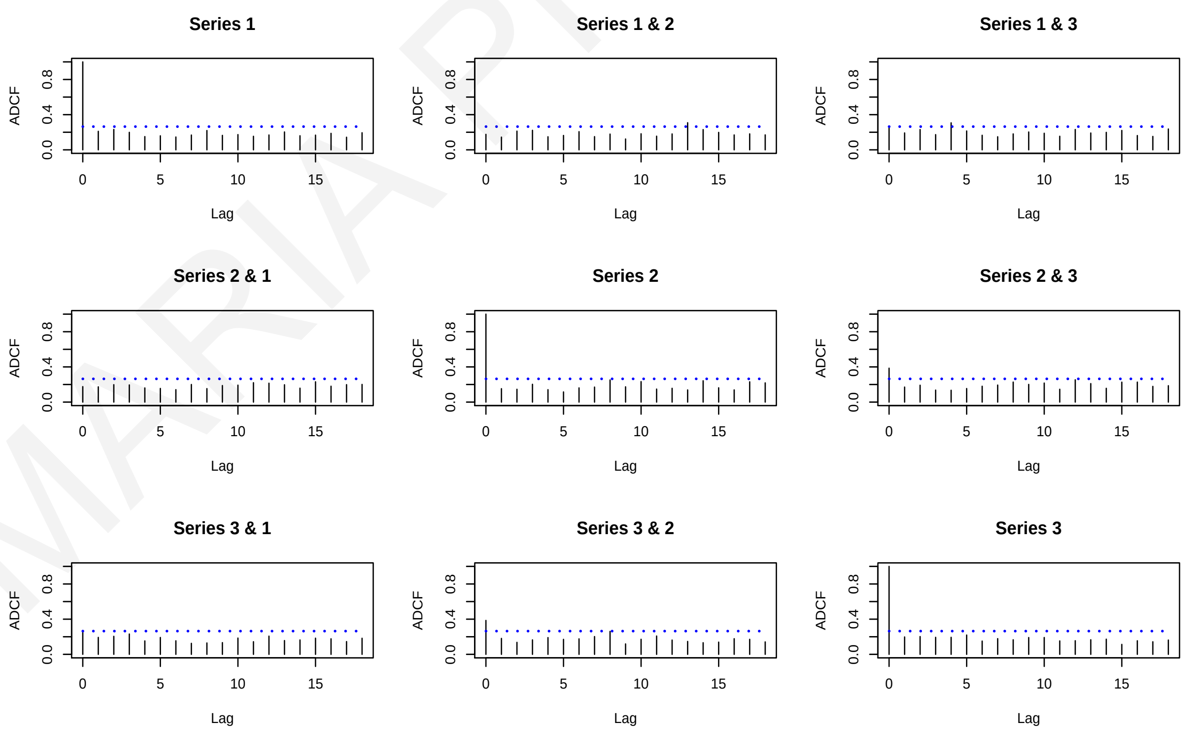



Figure 5.8: The sample ACF (upper panel) and sample ADCF (lower panel) of the residuals of the VAR(2) model for the percentage quarterly growth rates of real gross domestic products of UK, Canada and US.

```
data:  resi, kernel type: parzen, bandwidth=10, replicates 499
Tnbar = 170.75, p-value = 0.226


> mADCFtest(resi,type="parzen",p=18,b=499,parallel=TRUE)


        Multivariate test of independence based on distance correlation


data:  resi, kernel type: parzen, bandwidth=18, replicates 499
Tnbar = 311.56, p-value = 0.102


### Tests of independence based on mLB
> LjungBox(resi, lags=c(6,10,18))
```

**Monthly Unemployment Rates of the 50 States in the United States**

Lastly, we study the monthly unemployment rates of the US 50 states (Tsay, 2014, p. 4) from January 1976 to September 2011 for 429 observations. Note that the data were seasonally adjusted and are available from Tsay's (2014) book site `http://faculty.chicagobooth.edu/ruey.tsay/teaching/mtsbk/`. In this example, we consider the differenced monthly rates of the first five states corresponding to Alabama (AL), Alaska (AK), Arizona (AZ), Arkansas (AR) and California (CA) for the last 300 observations. The first order differencing of the data is performed via diffM() function from **MTS** (Tsay, 2015) package:

```
> data <- read.table("m-unemp-states.txt",header=T)
> dim(data)
[1] 429  50
> rates <- diffM(data)
>  dim(rates)
[1] 428  50
> zt <- tail(rates[,1:5],300)
> head(zt)
```

```
      AL   AK   AZ   AR   CA
[1,] -0.1 -0.1 -0.2 -0.1 -0.2
[2,] -0.2  0.0 -0.3 -0.1 -0.1
[3,] -0.1  0.0 -0.2 -0.1 -0.1
[4,]  0.0  0.0 -0.2 -0.1  0.0
[5,]  0.0  0.1 -0.1 -0.1  0.0
[6,]  0.2  0.3  0.0 -0.1  0.1
```

Assuming that the 5-dimensional series $(AL_t, AK_t, AZ_t, AR_t, CA_t)$ follows a VAR model and employing the AIC to choose its order, we obtain that a fifth order VAR model fits well the data. After a VAR(5) data fitting, we construct multivariate tests of independence among the 5-variate residual series for bandwidth parameters $p$=6, 10 and 17, to check the adequacy of the model fit. Constructing tests of independence among the residuals based on $mLB$ and $\overline{T}_n$, both test statistics give large $p$-values close to 1, indicating that a VAR(5) model is adequate for the data. We note that the kernel-based statistic $\overline{T}_n$ is computed based on the truncated kernel (the default) for $b = 499$ bootstrap replications. The R commands for the above model fit and the corresponding tests of independence among the residuals are as follows:

```
> var5 <- VAR(zt,5)
> residuals <- var5$residuals

> mADCFtest(residuals,p=6,b=499, parallel=TRUE)

       Multivariate test of independence based on distance correlation

data:  residuals, kernel type: truncated, bandwidth=6, replicates 499
Tnbar = 438.9153, p-value = 0.99

> mADCFtest(residuals,p=10,b=499,parallel=TRUE)

        Multivariate test of independence based on distance correlation
```

```
data:  residuals, kernel type: truncated, bandwidth=10, replicates 499
Tnbar = 763.8492, p-value = 0.916

> mADCFtest(residuals,p=17,b=499,parallel=TRUE)

        Multivariate test of independence based on distance correlation

data:  residuals, kernel type: truncated, bandwidth=17, replicates 499
Tnbar = 1296.841, p-value = 0.932

> LjungBox(residuals,lags=c(6,10,17))
 Lags Statistic  df    pvalue
    6  56.04981 150 1.0000000
   10 159.56430 250 0.9999983
   17 316.24291 425 0.9999778
```

# Chapter 6

# Conclusions and Future Work

## 6.1 Conclusions

In this thesis, we have developed a novel distance covariance testing methodology for testing pairwise dependence in both univariate and multivariate time series. The main contribution of this work is that the resulting test statistics are calculated for an increasing number of lags, whereas the results based on ADCV available in the literature (reviewed in Section 2.4.1) have been obtained under the assumption of a fixed lag order.

The first part of the thesis (Chapter 3) considers the univariate case suggesting a test statistic that is based on the ADCV and is motivated by a spectral domain point of view. In fact, we compared a kernel-based generalized estimated spectral density to the null spectral density, by a weighted quadratic norm. It turns out that the resulting test statistic is defined in terms of the ADCV and its asymptotic distribution is standard normal, suitably normalized. As Hong (1999) pointed out, the use of the generalized spectral density allows us to detect pairwise dependence in both linear and nonlinear time series structures. Our approach differs from that of Hong (1999), since our test statistic is calculated by means of a nonintegrable weighting function. The nonintegrability of the weight function yields more interesting results. In addition, we allow the number of lags tested in $H_0$ to increase with the sample size $n$. Empirical results suggest that our new test of independence has better power than the Portmanteau tests of Box and Pierce (1970) and Ljung and Box (1978) and the test proposed by Hong (1996), against a nonlinear structure. The proposed test is quite close in terms of power to the test proposed by Hong (1999), denoted by $T_n^{(2)}$, and in some cases it outperforms $T_n^{(2)}$.

In the second part of this work (Chapter 4), we extend the ADCV function in the context of multivariate time series by defining its matrix version. The information contained in this matrix is useful for examining any possible relationships within and between time series. In fact, we introduced the generalized spectral density matrix and compared it to the corresponding null matrix obtained under serial independence, by a squared weighted Frobenius norm. The resulting test statistic is a multivariate Ljung-Box ($mLB$) type test statistic defined in terms of the ADCV matrices, where its standardized version is asymptotically normally distributed. Simulations and real data examples propose that our new developed multivariate test of independence performs better than the one based on the $mLB$, especially against non-Gaussian and nonlinear data structures. Several extensions of this work are discussed in the next section.

The test statistics derived from the univariate and multivariate methodology demonstrated in Chapters 3 and 4 respectively, depend on a bandwidth parameter $p$. A cross-validation method might be suitable to choose the bandwidth parameter but we can also vary $p$ to examine the sensitivity of the results obtained. However, in our data examples we did not discover any notable relation between $p$ and the outcome of all test statistics. Moreover, we can obtain the optimal kernel function $k(.)$ that maximizes the asymptotic power of the test statistics under some conditions. In this sense, the Daniell kernel is the optimal kernel which maximizes the power of the test statistic proposed by Hong (1999, Theorem 6).

The R package **energy** (Rizzo and Szekely, 2014) provides functions that covers distance methodology for random variables. Dropping the assumption of i.i.d. data, there is no published package that includes functions about distance covariance for time series data. Many R users can now use the package **dCovTS** which fills this gap by providing functions that compute distance covariance and correlation functions for both univariate and multivariate time series. We also include functions that develop univariate and multivariate tests of serial dependence based on distance covariance and correlation functions as described in Chapters 3 and 4.

## 6.2 Future Work

The results presented in this thesis can be seen as a starting point for further research. Below, we briefly name few prospective investigations related to the theory of distance covariance in time series.

### 6.2.1 Other Types of Data

A possible extension of the methodology outlined in this thesis is to explore the use of ADCV (univariate version and matrix version) in areas where the assumption of strict stationarity is relaxed. For this goal, the framework introduced by Dahlhaus (1996) concerning locally stationary processes can be quite useful to define a local ADCV. Such a measure will detect local dependencies in non-stationary time series since the framework of locally stationary processes allows segmented approximation to the process by a stationary process. Moreover, experiencing with other types of dependent data, like space/time data, spatial data and data observed in an irregular lattice is another challenging topic which offers special attention for research since such data is observed quite frequently in applications.

### 6.2.2 ADCV Matrix in High Dimensions

Another possible direction for further research is to study the behavior of ADCV matrix in high dimensions. High-dimensionality occurs in cases where the time series dimension $d$ is of much larger order than the observed sample $n$, that is $d \gg n$ (Bühlmann and van der Geer, 2011). Many important examples of high dimensional data, in particular those studied in economics and finance, environmetrics or medical imaging have the feature that the observations are dependent over time; this characteristic adds up to the expected dimensional dependence. Classical models for time series analysis assume a stationary correlation structure and employ spectral domain based methods (equivalently methods which are based on the sample autocovariance matrix) to carry out inference and prediction. In spite of this, to the best of our knowledge, no work exists that analyzes the behavior of ADCV matrix, especially for high-dimensions. We envision to approach this problem by the theory of $U$- and $V$-statistics. It can be shown that the ADCV matrix is matrix variate $U$-statistic. Therefore the study of matrix variate $U$-statistics for dependent data will yield asymptotic results for the ADCV (Chen, 2016). Here, we need to point out that this problem is demanding but its solution will give insights for several other problems. In particular, from the discussion of Chapter 4, we expect that the ADCV matrix will be quite useful on determining possible dependencies among and between time series. This is an important step for understating the dependence structure of the observed process and the ways in which the dimensionality affects the inference.

### 6.2.3 Graphical Modeling Based on ADCV Matrix

Closely related to the research topic of high dimensionality is the study of graphical models for multivariate time series (Brillinger, 1996; Dahlhaus, 2000; Eichler, 2008, 2012). In general, graphical models describe conditional independence relationships among the components of a $d$-dimensional time series $\{\mathbf{X}_t\}$ by means of a graph. Particularly, a graph $G = (I, E)$ consists of a set of vertices $I = \{1, 2, \ldots, d\}$ representing the components of the series, and a set of edges $E \subset \{(a, b) \in I \times I\}$ indicating conditional dependence. Considering multivariate stationary time series, Dahlhaus (2000) proposed the use of partial frequency methods and especially the use of partial spectral density and partial spectral coherence. His approach leads to the definition of the so-called partial correlation graph, where an edge $a - b$ is missing from the graph if and only if the corresponding series $\{X_{t;a}\}$ and $\{X_{t;b}\}$ are uncorrelated after removing the linear effects of all the other components $\{X_{t;I\setminus\{a,b\}}\}$ for all $t \in \mathbb{Z}$.

Motivated by this, an extension for further research is to define the so-called *partial distance correlation graph* which is constructed based on the ADCV matrix, where a missing edge $a - b$ indicates conditional independence for the corresponding components. In fact, we first propose instead of using the original series $\{\mathbf{X}_t\}$, to use the series $\{\mathbf{W}_t\}$ with components

$$W_{t;r} = |X - X'| - m_X(X) - m_X(X') + \bar{m}_X,$$

where $m_X(u) := E\,|X-u|$ and $\bar{m}_X := E\,|m_X(X)|$ for $u \in \mathbb{R}$, with $X \equiv X_{t;r}$ and $X'$ be an i.i.d. copy of $X$, for $r = 1, 2, \ldots, d$. We note that if $W_{t;a}$ and $W_{t;b}$ are independent given $W_{t;Iab}$, then $X_{t;a}$ and $X_{t;b}$ are independent given $X_{t;I_{ab}}$ (and vice versa), where we have set $I_{ab} = I \setminus \{a, b\}$ for ease of notation. By recalling Section 4.3, we obtain that the transformed series $\{\mathbf{W}_t\}$ is a zero-mean process with covariance matrix

$$V^{(2)}(j) = \left[V^2_{rm}(j)\right]^d_{r,m=1}, \qquad j = 0, \pm1, \pm2, \ldots.$$

Motivated by Brillinger (1981, Chapter 5), we may consider that two fixed components $\{W_{t;a}\}$ and $\{W_{t;b}\}$ are *partially independent* if the corresponding partial cross-spectrum

$$f_{ab/I_{ab}}(\omega) = f_{ab}(\omega) - f_{aI_{ab}}(\omega)\left[f_{I_{ab}I_{ab}}(\omega)\right]^{-1} f_{I_{ab}b}(\omega), \tag{6.1}$$

or its rescaled version, the so-called partial spectral coherence

$$R_{ab/I_{ab}}(\omega) \;=\; \frac{f_{ab/I_{ab}}(\omega)}{\sqrt{f_{aa/I_{ab}}(\omega) f_{bb/I_{ab}}(\omega)}}, \tag{6.2}$$

is zero for all frequencies $\omega \in \Pi$, where

$$\begin{aligned} f_{ab}(\omega) &= \frac{1}{2\pi}\sum_{j=-\infty}^{\infty} \mathrm{Cov}(W_{t;a}, W_{t-|j|;b}) e^{-ij\omega} \\ &= \frac{1}{2\pi}\sum_{j=-\infty}^{\infty} V^2_{ab}(j) e^{-ij\omega}. \end{aligned}$$

Considering a sample of size $n$, a kernel-based estimator for $f_{ab}(\cdot)$ is given by

$$\hat{f}_{ab}(\omega) \;=\; \frac{1}{2\pi}\sum_{j=-(n-1)}^{(n-1)} (1 - |j|/n)^{1/2} k(j/p) \widehat{V}^2_{ab}(j) e^{-ij\omega}, \quad \omega \;\in\; \Pi, \tag{6.3}$$

where $k(\cdot)$ is a univariate kernel function, $p$ is a bandwidth parameter and $\widehat{V}^2_{ab}(\cdot)$ is given by (4.7). Therefore, substituting (6.3) in the definitions of (6.1) and (6.2), we obtain $\hat{f}_{ab/I_{ab}}(\cdot)$ and $\widehat{R}_{ab/I_{ab}}(\cdot)$ respectively. Studying the properties of $R_{ab/I_{ab}}(\cdot)$ and its sample version is an important problem on its own. In addition, developing a test statistic based on $R_{ab/I_{ab}}(\cdot)$ for testing conditional independence among the components of multivariate time series is another challenging problem. The testing problem is of the following form

$$H_0 : R_{ab/I_{ab}}(\omega) = 0 \quad \text{against} \quad H_1 : R_{ab/I_{ab}}(\omega) \neq 0,$$

for all $\omega \in \Pi$. Extending Eichler's (2008) testing methodology which is based on the classical partial spectral coherence, we may consider the following test statistic

$$S_n = \int_{\Pi} \|\widehat{R}_{ab/I_{ab}}(\omega)\|^2 d\omega,$$

where $\|\cdot\|$ denotes the Euclidean norm. The above test statistic can be approximated by the following

$$S_n = \frac{2\pi}{n} \sum_{j=1}^{n} \left\| \widehat{R}_{ab/I_{ab}}\Big(\frac{2\pi j}{n}\Big) \right\|^2.$$

Considering several real data problems from economics, finance, biology, environmetrics and other scientific fields, we may compare these results to the existing methodology based on classical approaches. Clearly, studying the asymptotic properties of $S_n$ gives rise to develop a new statistical testing theory based on the notion of ADCV matrix.

### 6.2.4 Enhance dCovTS Package

A number of possible extensions of the first version of **dCovTS** package can be seen as further research. In particular, we can enhance the package with R functions that cover the proposed theory for the different types of dependent data presented in Section 6.2.1 and for high-dimensional data discussed in Section 6.2.2. Moreover, a related R theory for the suggested graphical modeling theory briefly explained in Section 6.2.3 can be developed for a future version of this package.

# Bibliography



Aaronson, J., R. Burton, H. Dehling, D. Gilat, T. Hill, and B. Weiss (1996). Strong laws for L- and U-statistics. *Transactions of the American Mathematical Society* **348**, 2845–2866.

Analytics, R. and S. Weston (2015). *doParallel: Foreach Parallel Adaptor for the 'parallel' Package*. R package version 1.0.10.

Andrzejaki, R. G., K. Lehnertz, F. Mormann, C. Rieke, P. David, and C. E. Elger (2001). Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state. *Physical Review E* **64**, 8 pages.

Bagnato, L., L. De Capitani, and A. Punzo (2014). Testing serial independence via density-based measures of divergence. *Methodology and Computing in Applied Probability* **16**, 627–641.

Bakirov, N. K., M. L. Rizzo, and G. J. Székely (2006). A multivariate nonparametric test of independence. *Journal of Multivariate Analysis* **97**, 1742–1756.

Belu, A. C. (2012). *Multivariate measures of dependence for random variables and Lévy processes*. Ph. D. thesis, Department of Statistics. Case Western Reserve University.

Benjamini, Y. and Y. Hochberg (1995). Controlling the false discovery rate: A practical and powerful approach. *Journal of the Royal Statistical Society* **57**, 289–300.

Berentsen, G. D. and D. Tjøstheim (2014). Recognizing and visualizing departures from independence in bivariate data using local Gaussian correlation. *Statistics and Computing* **24**, 785–801.

Blum, J. R., J. Kiefer, and M. Rosenblatt (1961). Distribution free tests of independence based on the sample distribution function. *Annals of Mathematical Statistics* **32**, 485–498.

Borovcova, S., R. Burton, and H. Dehling (1999). Consistency of the Takens estimator for the correlation dimension. *The Annals of Applied Probability* **9**, 376–390.

Box, G. E. P. and D. A. Pierce (1970). Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. *Journal of the American Statistical Association* **65**, 1509–1526.

Bradley, R. C. (1983). Approximation theorems for strongly mixing random variables. *Michigan Mathematical Journal.* **30**, 69–81.

Bradley, R. C. (2007). *Introduction to Strong Mixing Conditions*, Volume 1-3. Heber City (Utah): Kendrick Press.

Brillinger, D. R. (1981). *Time Series : Data Analysis and Theory.* New York: McGraw-Hill.

Brillinger, D. R. (1996). Remarks concerning graphical models for time series and point pocesses. *Revista de Econometria* **16**, 1–23.

Brockwell, P. J. and R. A. Davis (1991). *Time series : theory and methods.* New York: Springer-Verlag. Second Edition.

Bühlmann, P. and S. van der Geer (2011). *Statistics for high-dimenisonal data.* Berlin: Springer-Verlag.

Chan, K.-S. and B. Ripley (2012). *TSA: Time Series Analysis.* R package version 1.01.

Chen, B. and Y. Hong (2012). Testing for the Markov property in time series. *Econometric Theory* **28**, 130–178.

Chen, X. (2016). Sup-norm asymptotics of high-dimensional matrix-variate U-statistics and its applications. http://arxiv.org/abs/1602.00199.

Dahlhaus, R. (1996). On the Kullback-Leibler information divergence of locally stationary processes. *Stochastic Processes Applications* **62**, 139–168.

Dahlhaus, R. (2000). Graphical interaction models for multivariate time series. *Metrika* **51**, 157–172.

Davis, R. A., M. Matsui, T. Mikosch, and P. Wan (2016). Applications of distance correlation to time series. http://arxiv.org/abs/1606.05481.

Dehling, H. and T. Mikosch (1994). Random quadratic forms and the boostrap for U-statistics. *Journal of Multivariate Analysis* **51**, 392–413.

Dette, H., M. Hallin, T. Kley, and S. Volgushev (2015). Of copulas, quantiles, ranks and spectra: An $L_1$-approach to spectral analysis. *Bernoulli* **21**, 781–831.

Doukhan, P. (1994). *Mixing: Properties and Examples.* New York: Springer-Verlag.

Doukhan, P. and S. Louhichi (1999). A new weak dependence condition and applications to moment inequalities. *Stochastic Processes and their Applications* **84**, 313–342.

Dueck, J., D. Edelmann, T. Gneiting, and D. Richards (2014). The affinely invariant distance correlation. *Bernoulli* **20**, 2305–2330.

Dueck, J., D. Edelmann, and D. Richards (2015). Distance correlation coefficients for Lancaster distributions. http://arxiv.org/pdf/1502.01413.pdf.

Eichler, M. (2008). Testing nonparametric and semiparametric hypotheses in vector stationary processes. *Journal of Multivariate Analysis* **99**, 968–1009.

Eichler, M. (2012). Graphical modelling of multivariate time series. *Probability Theory and Related Fileds* **153**, 233–268.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflations. *Econometrica* **50**, 987–1007.

Escanciano, J. C. and I. N. Lobato (2009). An automatic Portmanteau test for serial correlation. *Journal of Econometrics* **151**, 140–149.

Fan, J. and Q. Yao (2003). *Nonlinear time series: Nonparametric and Parametric Methods.* New York: Springer-Verlag.

Feuerverger, A. (1993). A consistent test for bivariate dependence. *International Statistical Review* **61**, 419–433.

Fokianos, K. and M. Pitsillou (2016a). Consistent testing for pairwise dependence in time series. *Technometrics*. http://dx.doi.org/10.1080/00401706.2016.1156024.

Fokianos, K. and M. Pitsillou (2016b). Testing pairwise independence for multivariate time series by the auto-distance correlation matrix. Submitted for publication.

Graves, S. (2014). *FinTS: Companion to Tsay (2005) Analysis of Financial Time Series.* R package version 0.4-5.

Gretton, A., K. Fukumizu, and B. K. Sriperumbudur (2009). Discussion of: Brownian distance covariance. *The Annals of Applied Statistics* **3**, 1285–1294.

Gretton, A., R. Herbrich, A. Smola, O. Bousque, and B. Schö lkop (2005). Kernel methods for measuring independence. *Journal of Machine Learning Research* **6**, 2075–2129.

Hipel, K. W. and A. I. McLeod (1994). *Time Series Modelling of Water Resources and Environmental Systems.* Amsterdam: Elsevier Science Pub Co.

Hong, Y. (1996). Consistent testing for serial correlation of unknown form. *Econometrica* **64**, 837–864.

Hong, Y. (1998). Testing for pairwise serial independence via the empirical distribution function. *Journal of the Royal Statistical Society, Series B* **60**, 429–453.

Hong, Y. (1999). Hypothesis testing in time series via the emprical characteristic function: A generalized spectral density approach. *Journal of the American Statistical Association* **94**, 1201–1220.

Hong, Y. (2000). Generalized spectral tests for serial dependence. *Journal of the Royal Statistical Society, Series B* **62**, 557–574.

Hosking, J. R. M. (1980). Multivariate Portmanteau statistic. *Journal of the American Statistical Association* **75**, 349–386.

Huo, X. and G. J. Székely (2015). Fast computing for distance covariance. *Technometrics*. DOI: 10.1080/00401706.2015.105443.

Josse, J. and S. Holmes (2014). Tests of independence and beyond. http://arxiv.org/pdf/1307.7383.pdf.

Kirchgassner, G., J. Walters, and U. Halser (2013). *Introduction to Modern Time Series Analysis*. New York: Springer.

Knight, J. L. and J. Yu (2002). Empirical characteristic function in time series estimation. *Econometric Theory* **18**, 691–721.

Kosorok, M. R. (2009). Discussion of: Brownian distance covariance. *The Annals of Applied Statistics* **3**, 1270–1278.

Kowalski, J. and X. M. Tu (2008). *Modern Applied U-Statistics.* Hoboken, New Jersey: Wiley Series in Probability and Statistics.

Lacal, V. and D. Tjøstheim (2016). Local Gaussian autocorrelation and tests of serial independence. *Journal of Time Series Analysis*. DOI: 10.1111/jtsa.12195.

Leucht, A. and M. H. Neumann (2013a). Degenerate U- and V-statistics under ergodicity: asymptotics, bootstrap and applications in statistics. *Annals of the Institute of Statistical Mathematics* **65**, 349–386.

Leucht, A. and M. H. Neumann (2013b). Dependent wild bootstrap for degenerate U- and V- statistics. *Journal of Multivariate Analysis* **117**, 257–280.

Li, W. (2004). *Diagnostic Checks in Time Series.* Boca Raton, FL: Chapman & Hall/CRC.

Li, W. K. and A. I. McLeod (1981). Distribution of the residuals autocorrelations in multivariate ARMA time series models. *Journal of the Royal Statistical Society Series B* **43**, 231–239.

Ljung, G. M. and G. E. P. Box (1978). On a measure of lack of fit in time series models. *Biometrika* **62**, 297–303.

Lutkepohl, H. (2005). *New Introduction to Multiple Time Series Analysis.* Berlin Heidelberg: Springer-Verlag.

Lyons, R. (2013). Distance covariance in metric spaces. *The Annals of Probability* **41**, 3284–3305.

Mahdi, E. (2011). *Diagnostic Checking, Time Series and Regression.* Ph. D. thesis, The University of Western Ontario.

Mahdi, E. and A. I. McLeod (2012). Improved multivariate Portmanteau diagnostic test. *Journal of Time Series Analysis* **33**, 211–222.

Manly, B. F. J. (2008). *Statistics for Environmental Science and Management.* Boca Raton: Chapman & Hall.

McLachlan, G. J., K. Do, and C. Ambroise (2004). *Analyzing Microarray Gene Expression Data.* Hoboken, New Jersey: Wiley.

Paparoditis, E. (2000). Spectral density based goodness-of-fit tests for time series models. *The Scandinavian Journal of Statistics* **27**, 143–176.

Paparoditis, E. (2001). On some power properties of goodness-of-fit tests in time series. In C. Charalambides, M. Koutras, and N. Balakrishnan (Eds.), *Probability and Statistical Models with Applications*, pp. 333–348. Boca Raton: Chapman & Hall/CRC.

Parmigiani, G., E. S. Garrett, R. A. Irixarry, and S. L. Zeger (2003). *The Analysis of Gene Expression Data: Methods and Software.* New Jersey: Springer-Verlag.

Pinkse, J. (1998). Consistent nonparametric testing for serial independence. *Journal of Econometrics* **84**, 205–231.

Poczos, B. and J. Schneider (2012). Conditional distance variance and correlation. http://www.cs.cmu.edu/ bapoczos/articles/poczos12distancecorr.pdf.

Politis, N. P., J. P. Romano, and M. Wolf (1999). *Subsampling.* New York: Springer-Verlag.

Priestley, M. B. (1981). *Spectral Analysis and Time Series.* London: Academic Press.

R Core Team (2014). *R: A Language and Environment for Statistical Computing.* Vienna, Austria: R Foundation for Statistical Computing.

Remillard, B. (2009). Discussion of: Brownian distance covariance. *The Annals of Applied Statistics* **3**, 1295–1298.

Rizzo, M. L. and G. J. Szekely (2014). *energy: E-statistics (energy statistics)*. R package version 1.6.2.

Robinson, P. M. (1991). Testing for strong serial correlation and dynamic conditional heteroskedasticity in multiple regression. *Journal of Econometrics* **47**, 67–84.

Romano, P. and L. A. Thombs (1996). Inference for autocorrelations under weak assumptions. *Journal of the Americal Statistical Association* **91**, 590–600.

Rosenblatt, M. (1956). A central limit theorem and a strong mixing condition. *Proceedings of the National Academy of Sciences USA* **42**, 43–47.

Sejdinovic, D., B. Sriperumbudur, A. Gretton, and K. Fukumizu (2013). Equivalence of distance-based and RKHS-based statistics in hypothesis testing. *The Annals of Statistics* **41**, 2263–2291.

Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics*. Wiley Series in Probability and Statistics.

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

Shao, X. (2011). Testing for white noise under unknown dependence and its applications to diagnostic checking for time series models. *Econometric Theory* **27**, 312–343.

Shumway, R. H., R. S. Azari, and Y. Pawitan (1988). Modeling mortality fluctuations in Los Angeles as functions of pollution and weather effects. *Environmetnal research* **45**, 224–241.

Shumway, R. H. and D. S. Stoffer (2011). *Time Series Analysis and Its Applications With R Examples*. New York: Springer-Verlag. Third Edition.

Skaug, H. J. and D. Tjøstheim (1993). A nonparametric test of serial independence based on the empirical distribution function. *Biometrika* **80**, 591–602.

Skaug, H. J. and D. Tjøstheim (1996). Measures of distance between densities with application to testing for serial independence. *Time Series Analysis in Memory of E. J. Hannan (eds. P. Robinson and M. Rosenblatt, New York)*.

Støve, B. and D. Tjøstheim (2014). Asymmetric dependence patterns in financial returns: An empirical investigation using local Gaussian correlation. In N. Haldrup, M. Meitz, and P. Saikkonen (Eds.), *Essays in Nonlinear Time Series Econometrics*, pp. 307–329. Oxford: Oxford Univeristy Press.

Støve, B., D. Tjøstheim, and K. O. Hufthammer (2014). Using local Gaussian correlation in a nonlinear re-examination of financial contagion. *Journal of Empirical Finance* **25**, 62–82.

Székely, G. J. and M. L. Rizzo (2005). Hierarchical clustering via joint between - within distances: Extending Ward's minimum variance method. *Journal of Classification* **22**, 151–183.

Székely, G. J. and M. L. Rizzo (2009). Brownian distance covariance. *The Annals of Statistics* **3**, 1236–1265.

Székely, G. J. and M. L. Rizzo (2013). Energy statistics: A class of statistics based on distances. *Journal of Statistical Planning and Inference* **143**, 1249–1272.

Székely, G. J. and M. L. Rizzo (2014). Partial distance correlation with methods for dissimilarities. *The Annals of Statistics* **42**, 2382–2412.

Székely, G. J., M. L. Rizzo, and N. K. Bakirov (2007). Measuring and testing dependence by correlation of distances. *The Annals of Statistics* **35**, 2769–2794.

Tjøstheim, D. (1996). Measures of dependence and tests of independence. *Statistics* **28**, 249–284.

Tjøstheim, D. and K. O. Hufthammer (2013). Local Gaussian correlation: A new measure of dependence. *Journal of Econometrics* **172**, 33–48.

Tong, H. (1990). *Nonlinear Time Series:A Dynamical System Approach.* New York: Oxford University Press.

Trapletti, A. and K. Hornik (2015). *tseries: Time Series Analysis and Computational Finance.* R package version 0.10-34.

Tsay, R. S. (2005). *Analysis of Financial Time Series.* Hoboken, NJ: Wiley Series in Probability and Statistics. Second Edition.

Tsay, R. S. (2010). *Analysis of Financial Time Series.* Hoboken, NJ: Wiley Series in Probability and Statistics. Third Edition.

Tsay, R. S. (2014). *Multivariate Time Series Analysis With R and Financial Applications.* Hoboken, NJ: Wiley Series in Probability and Statistics.

Tsay, R. S. (2015). *MTS: All-Purpose Toolkit for Analyzing Multivariate Time Series (MTS) and Estimating Multivariate Volatility Models.* R package version 0.33.

Verbeek, M. (2012). *A Guide to Modern Econometrics.* United Kingdom: Wiley Series in Probability and Statistics.

Wang, X., W. Pan, W. Hu, Y. Tian, and H. Zhang (2015). Conditional distance correlation. *Journal of the American Statistical Association.* DOI: 10.1080/01621459.2014.993081.

Wu, C. F. J. (1986). Jackknife, bootstrap and other resampling methods in regression analysis. *The Annals of Statistics* **14**, 1261–1295.

Wu, W. B. (2005). Nonlinear system theory: Another look at dependence. *Proceedings of the National Academy of Sciences USA* **102**(40), 14150–14154.

Xiao, H. and W. B. Wu (2014). Portmanteau test and simultaneous inference for serial covariances. *Statistica Sinica* **24**, 577–600.

Zhou, Z. (2012). Measuring nonlinear dependence in time-series, a distance correlation approach. *Journal of Time Series Analysis* **33**, 438–457.