5-2012

New Results in Multivariate Time Series with Applications

Nan Su
Clemson University, nsu@clemson.edu

New Results in Multivariate Time Series with Applications

A Dissertation
Presented to
the Graduate School of
Clemson University

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy
Mathematical Science

---

by
Nan Su
May 2012

Accepted by:
Dr. Robert Lund, Committee Chair
Dr. Colin Gallagher
Dr. Chanseok Park
Dr. K.B. Kulasekera

Abstract

This dissertation presents some new results in stationary multivariate time series.

The asymptotic properties of the sample autocovariance are established, that is, we derive a multivariate version of Bartlett's Classic Formula. The estimation of the autocovariance function plays a crucial role in time series analysis, in particular for the identification problem. Explicit formula for vector autoregressive (p) and vector moving average (q) processes are presented as examples. We also address linear processes driven by non-independent errors, a feature that permits consideration of multivariate GARCH processes.

We next compare several techniques to discriminate two multivariate stationary signals. The compared methods include Gaussian likelihood ratio variance/covariance matrix tests and spectral-based tests gauging equality of the autocovariance function of the two signals. A simulation study is presented that illuminates the various properties of the methods. An analysis of experimentally collected gearbox data is also presented.

Acknowledgments

During the past few years at Clemson, many people have guided, encouraged, and supported me. Without them, my life at Clemson University could not have been stimulating and enriching as it was. I would like to take this opportunity to thank them.

First and foremost, I am indebted to my advisor Dr. Robert Lund for his guidance and support. I thank him for giving me the confidence to tackle difficult problems in my dissertation. His knowledge and insight in the area of time series modeling and analysis has significantly improved the quality of this work.

I would like to deeply thank Dr. Colin Gallagher, Dr. K.B. Kulasekera, Dr. Chanseok Park and all other mathematics faculty who have taught me during the past several years. I enjoyed the learning from all the professors. It is an experience I will treasure forever.

I would also like to thank the Department of Mathematical Sciences for providing me financial support me during my stay at Clemson. I appreciate all the help from the alumni and current graduate students in our department and around U.S.

Finally, I thank my parents. Without their love and support I could not have done this. I dedicate this thesis to them.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Time Series Overview

Time series analysis is a part of statistics; i.e., the study of the reduction of data. Time series are processes that are recorded in a temporal order. The subject is mathematically elaborate, yet realistic. In its modern form, time series analysis dates from the early 1950s and the advent of high speed computing. The first reasonably connected account in this sense is probably [1], which is still perhaps worth studying.

Definition 1.1 A time series is a set of time-ordered observations $\{X_t\}$, X_t being at the observation at time t.

A time series model entails specification of a suitable probability model for the observed data. A complete time series model for $\{X_t\}$ is a specification of the joint distribution of $\{X_t\}$, of which $\{x_t\}$ is postulated to be a realization. In practice, often only means and variance/covariance are modeled.

Definition 1.2 Let $\{X_t\}$ be a time series with $E(X_t^2) < \infty$. The mean function of $\{X_t\}$ is

$$\mu_X(t) = E(X_t).$$

The covariance function of $\{X_t\}$ is

$$\gamma_X(r, s) = \mathrm{Cov}(X_r, X_s) = E[(X_r - \mu_X(r))(X_s - \mu_X(s))]$$

for all integers r and s.

Definition 1.3 $\{X_t\}$ is (weakly) stationary if

(i) $\mu_X(t)$ does not depend on t,

and

(ii) $\gamma_X(t+h, t)$ does not depend on t for each h.

We point out that a strictly stationary time series $\{X_t, t = 0, \pm 1, \ldots\}$ means that $(X_1, \ldots, X_n)$ and $(X_{1+h}, \ldots, X_{n+h})$ have the joint distributions for all integers h and $n > 0$. Henceforth, we will use the term stationary to mean weak stationarity; if a process is stationary in a strict sense, we will use the term strictly stationary (should this be relevant).

Because the mean $E(X_t) = \mu_X(t)$ of a stationary time series does not dependent on time t, we write $\mu_X(t) = \mu$. Also, because $\gamma_X(r, s)$ depends on r and s only through $|r - s|$, We write

$$\gamma_X(h) := \gamma_X(h, 0) = \gamma_X(t+h, t).$$

Definition 1.4 Let $\{X_t\}$ be a stationary time series. The autocovariance function (ACVF) of $\{X_t\}$ at lag h is

$$\gamma_X(h) = \mathrm{Cov}(X_{t+h}, X_t).$$

The autocorrelation function (ACF) of $\{X_t\}$ at lag h is

$$\rho_X(h) = \frac{\gamma_X(h)}{\gamma_X(0)} = \mathrm{Cor}(X_{t+h}, X_t).$$

The impact of time series analysis on scientific applications is partially appreciated by producing an abbreviated list of the diverse fields in which important time series problems arise. For example, many familiar time series occur in the field of economics, where we are continually exposed to daily stock market quotations or monthly unemployment figures. Social scientists follow populations series, such as birthrates or school enrollments. An epidemiologist might be interested in the number of influenza cases observed over time.

Some of the most intensive and sophisticated applications of time series methods have been to problems in the physical and environmental sciences. This fact accounts for the basic engineering flavor permeating the topic's language of the topics. One of the earliest recorded series is the monthly sunspot numbers studied by Schuster [32]. More modern investigations center on whether warming is present in global temperature measurements or whether levels of pollution may influence daily mortality in Los Angeles. The modeling of speech series is an important problem related to the efficient transmission of voice recordings. Here, a time series characteristic known as the power spectrum is used to help computers recognize and translate speech. Geophysical time series, such as those produced by yearly depositions of various kinds, can provide long-range proxies for temperature and rainfall. Seismic recordings can aid in mapping fault lines or in distinguishing between earthquakes and nuclear explosions.

Methods for time series analysis may be divided into two classes: time-domain methods and frequency-domain methods.

The time domain approach is generally motivated by the presumption that correlation between adjacent points in time is best explained in terms of dependence between the current value on past values. The time domain approach focuses on modeling future value of a series as a parametric function of the current and past values. In this scenario, linear regression models are often used to describe the present value of a series in terms of past values and on past values of other series.

Conversely, the frequency domain approach assumes the primary characteristics of interest relate to periodic or systematic sinusoidal variations found naturally in most data. These periodic variations are often caused by the biological, physical, or environmental phenomena of interest. In spectral analysis, the partition of the variation in a time series is accomplished by evaluating separately the variance associated with each periodicity of interest. This variance profile over frequency is called the power spectrum.

Many time series arising in practice are best considered as components of some vector-valued (multivariate) time series $\{X_t\}$ having not only serial dependence within each component series $\{X_{ti}\}$, but also interdependence between the different component series $\{X_{ti}\}$ and $\{X_{tj}\}$, $i \neq j$. For instance, in a system consisting of investment, income, and consumption, one may want to understand the likely impact of a change in income. Alternatively, given a particular theory, is it consistent with the relations implied by a multivariate time series model which is developed with the help of statistical tools? Questions regarding the structure of the relationships between the variables involved are occasionally investigated in the context of multivariate time series analysis. Obtaining insight into the dynamic structure of a system is one objective of multivariate time series analysis. Much of the theory of univariate time series extends in a natural way to the multivariate case; however, new problems arise.

As in the univariate case, a particularly important role is played by the class of multivariate stationary time series, defined as follows.

Definition 1.5 The m-variate series $\{X_t\}$ is (weakly) stationary if

(i) $\mu_X(t) = E[X_t]$ does not depend on t,

and

(ii) $\Gamma_X(t+h,t) = cov(X_{t+h} - E[X_{t+h}], X_t - E[X_t])$ does not depend on t for each h.

For a stationary time series, we shall use the notation

$$\mu := E[X_t] = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_m \end{pmatrix} \tag{1.1.1}$$

and

$$\Gamma(h) = E[(X_{t+h} - \mu)(X_t - \mu)'] = \begin{bmatrix} \gamma_{11}(h) & \dots & \gamma_{1m}(h) \\ \vdots & \ddots & \vdots \\ \gamma_{m1}(h) & \dots & \gamma_{mm}(h) \end{bmatrix}.$$

We shall refer to μ as the mean of the series and to $\Gamma(h)$ as the covariance matrix at lag h. Notice that if $\{X_t\}$ is stationary with covariance matrix function $\Gamma(\cdot)$, then for each i, $\{X_{ti}\}$ is univariate stationary with covariance function $\gamma_{ii}(\cdot)$. The function $\gamma_{ij}(\cdot)$, $i \neq j$, is called the cross-covariance function of the two series $\{X_{ti}\}$ and $\{X_{tj}\}$. It should be noted that $\gamma_{ij}(\cdot)$ is not in general the same as $\gamma_{ji}(\cdot)$. The correlation matrix function $R(\cdot)$ is defined by

$$R(h) := \begin{bmatrix} \rho_{11}(h) & \dots & \rho_{1m}(h) \\ \vdots & \ddots & \vdots \\ \rho_{m1}(h) & \dots & \rho_{mm}(h) \end{bmatrix},$$

where $\rho_{ij}(h) = \gamma_{ij}(h)/[\gamma_{ii}(0)\gamma_{jj}(0)]^{1/2}$. The basic properties of $\Gamma(\cdot)$ are

1. $\Gamma(h) = \Gamma'(-h)$,
2. $|\gamma_{ij}(h)| \leq [\gamma_{ii}(0)\gamma_{jj}(0)]^{1/2}, i, j = 1, \dots, m$,
3. $\gamma_{ii}(\cdot)$ is an autocovariance function, $i = 1, \dots, m$, and
4. $\sum_{j,k=1}^{n} a_j' \Gamma(j-k) a_k \geq 0$ for all $n \in \{1, 2, \dots\}$ and $a_1, \dots a_n \in \mathrm{R}^m$.

1.2 Univariate Versions of Bartlett's Formula

If $\{X_t\}$ is a real-valued stationary process, then from a second-order point of view it is characterized by its mean μ and its autocovariance function $\gamma(\cdot)$. The estimation of μ, $\gamma(\cdot)$, and the autocorrelation function $\rho(\cdot) = \gamma(\cdot)/\gamma(0)$ from observations of $X_1, \dots, X_n$, therefore plays a crucial role in problems of inference, and in particular, in the problem of constructing an appropriate model for the data.

A univariate Bartlett's result states that the first L sample autocovariances are asymptotically normal as $n \to \infty$:

$$\begin{pmatrix} \hat{\gamma}(0) \\ \hat{\gamma}(1) \\ \hat{\gamma}(2) \\ \vdots \\ \hat{\gamma}(L) \end{pmatrix} \sim AN \left(\begin{pmatrix} \gamma(0) \\ \gamma(1) \\ \gamma(2) \\ \vdots \\ \gamma(L) \end{pmatrix}, \frac{W}{n} \right). \tag{1.2.2}$$

The matrix W is clarified further below. The assumptions needed here are that $\{X_t\}$ has the linear process representation $X_t = \sum_{k=-\infty}^{\infty} \psi_k Z_{t-k}$, where $\{Z_t\}$ is i.i.d. with a finite fourth moment. In particular, we will need that $\{X_t\}$ is fourth-order stationary, which implies that $E[X_t X_{t+i} X_{t+r} X_{t+r+j}]$ does not depend on t. In (1.2.2), W is an $(L+1) \times (L+1)$ matrix with entries $W_{i,j} = \lim_{n\to\infty} n\mathrm{Cov}(\hat{\gamma}(i), \hat{\gamma}(j))$. It is known that

$$W_{i,j} = \sum_{r=-\infty}^{\infty} [E[X_t X_{t+i} X_{t+r} X_{t+r+j}] - \gamma(i)\gamma(j)], 0 \le i, j \le L.$$

The term $E[X_t X_{t+i} X_{t+r} X_{t+r+j}]$ does not depend on time t since $\{X_t\}$ is strictly stationary. Proposition 7.3.1 in Brockwell and Davis [3] provides the equivalent form

$$W_{i,j} = (\eta - 3)\gamma(i)\gamma(j) + \sum_{k=-\infty}^{\infty} [\gamma(k)\gamma(k-i+j) + \gamma(k+j)\gamma(k-i)], \tag{1.2.3}$$

for $0 \le i, j \le L$. Here, $E[Z_t^4] = \eta\sigma^4$. When $\{Z_t\}$ is Gaussian, $\eta = 3$ and the first term in (1.2.3) is zero.

The corresponding asymptotic property of the sample autocorrelation function is stated as follows:

$$\begin{pmatrix} \hat{\rho}(1) \\ \hat{\rho}(2) \\ \vdots \\ \hat{\rho}(L) \end{pmatrix} \sim AN \left(\begin{pmatrix} \rho(1) \\ \rho(2) \\ \vdots \\ \rho(L) \end{pmatrix}, \frac{V}{n} \right), \tag{1.2.4}$$

where V is the covariance matrix whose (i, j)-element is given by classical Bartlett's formula:

$$\begin{aligned} v_{ij} &= \{\rho(k+i)\rho(k+j) + \rho(k-i)\rho(k+j) + 2\rho(i)\rho(j)\rho^2(k) \qquad (1.2.5) \\ &- 2\rho(i)\rho(k)\rho(k+j) - 2\rho(j)\rho(k)\rho(k+i)\} \end{aligned}$$

1.3 Research Goals

Our objective is to derive forms of the above results in multivariate setting; that is, establish asymptotic normality of the random matrices $\hat{\Gamma}(0)$, $\hat{\Gamma}(1), \ldots$, $\hat{\Gamma}(L)$ as a function of the sample size n and identify the limiting information matrix. As noted above, this issue has been classically settled in univariate treatments on time series, for example, Bartlett [1], Hannan [8], Brockwell and Davis [3], and Shumway and Stoffer [14]. However, no one has yet tackled the multivariate case. In the next section, a compact multivariate version formula analogue to (1.2.3) is presented. One will appreciate the difficulties encountered in its derivation there.

We also show a parallel result to (1.2.6) for the asymptotic sample autocorrelation structure by applying a multivariate delta method. This work verifies the bivariate Gaussian formula in Theorem 11.2.3 of Brockwell and Davis [3], which was simply stated but not derived.

Proofs essentially rely on joint asymptotic normality of sample autocovariances. Under Gaussian assumptions, our formulae greatly simplify. However, the matrix calculations are still intense. For example, we will often need to reshape the vector or matrix. Here permutation matrices and properties of Kronecker products help.

Examples similar to those common in univariate case are also presented. We specifically, consider the first order causal autoregression satisfying

$$X_t = \Phi X_{t-1} + Z_t, \tag{1.3.6}$$

where $\{Z_t\}$ is zero mean d-variate white noise with covariance matrix Σ. We assume that Φ is invertible. We identify each piece in Equation (2.3.10), which reduces to that in Example 7.2.3 in Brockwell and Davis [3] when $d = 1$.

Next, we move to a qth order moving-average $\{X_t\}$ satisfying

$$X_t = Z_t + \Theta_1 Z_{t-1} + \cdots + \Theta_q Z_{t-q},$$

where $\{Z_t\}$ is zero mean white noise with covariance matrix Σ. We again consider the Gaussian case and concentrate on computation of the sum appearing below in (2.3.10).

1.4 Application

Our applications compares several techniques that are used to discriminate two multivariate stationary signals. The compared methods include Gaussian likelihood ratio variance/covariance matrix tests and spectral-based tests gauging equality of the autocovariance function (over all lags) of the two signals. We show how one can make inappropriate conclusions with PCA tests, even when dimension augmentation techniques are used to incorporate non-zero lag autocovariances into the analysis. The various discrimination methods are first discussed. A simulation study is then presented that illuminates the various properties of the methods. An analysis of experimentally collected gearbox data is also presented.

Elaborating, given two d-dimensional series $\{\mathbf{X}_t\}$ and $\{\mathbf{Y}_t\}$ that are preprocessed to a zero-mean stationary setting, we considers how to assess whether (or not) the two signals have the same time series dynamics. This is useful in discrimination and classification pursuits. For example, if a test signal $\{\mathbf{Y}_t\}$ is deemed to have different dynamics than a reference signal $\{\mathbf{X}_t\}$ that is known to be "healthy", the test signal could be deemed unhealthy. Signal discrimination problems are fundamental ([15] [16]) and are well-developed when discriminating series via means or first moments; here, Hotelling T^2 or Q statistics are frequently relied upon ([17], [18]). In 1986, Coates [19] considered discrimination of two univariate constant-mean series based on their sample autocovariances. Speech signals, for example, are typically of constant mean, regardless of what words are being spoken. Here, word-to-word changes are best identified through autocovariances shifts and monitoring of the mean is insufficient to identify dynamic changes. Kakizawa [20] seeks to discriminate an earthquake from a covert underground nuclear test; again, the crux issue lies with constant-mean data.

The classical way of discriminating $\{\mathbf{X}_t\}$ and $\{\mathbf{Y}_t\}$ through second order characteristics is via a Gaussian likelihood ratio. Such a test compares the sample variance matrix of the two series. Elaborating, conclusions are based on how different the two sample variance matrices

$$N^{-1}\sum_{t=1}^{N}\mathbf{X}_t\mathbf{X}_t', \qquad N^{-1}\sum_{t=1}^{N}\mathbf{Y}_t\mathbf{Y}_t'$$

are from each other. Here, N is the sample length of the two series, which are assumed equal for convenience. When the dimension d is large, this comparison is typically made after a dimension reduction transformation, usually some type of principal component analysis (PCA), is done. Without dimension reduction aspects, covariance comparisons are not truly PCA techniques; however, they share the commonality in that conclusions are made only from sample variances.

Basing signal equality conclusions exclusively on sample variances can produce erroneous conclusions when the two series are not multivariate white noise. A more comprehensive test would compare the sample autocovariances

$$\hat{\mathbf{\Gamma}}_X(h) = N^{-1} \sum_{t=1}^{N-h} \mathbf{X}_{t+h}\mathbf{X}_t'$$

and

$$\hat{\mathbf{\Gamma}}_Y(h) = N^{-1} \sum_{t=1}^{N-h} \mathbf{Y}_{t+h}\mathbf{Y}_t'$$

over all suitable lags $h \geq 0$. Such tests for multivariate series were discussed in [20], [21], [22], and the references within.

Bassily [21] and Lund [22] attack the problem with frequency domain techniques. Specifically, two multivariate covariance functions are equal if and only if their spectral densities are equal at all frequencies (the spectrum is assumed to have no point masses). From this, signal equality tests that compare the periodograms of both series were devised (Section 3.2 elaborates). Chapter 3 rehashes these methods and shows how one can fool variance-based tests for signal equality, even when the dimension is augmented to account for non-zero autocovariances at higher lags. The pros and cons of the various methods are demonstrated by simulating multivariate stationary signals with various properties and then applying the tests. An application to a series of gearbox vibrations is included.

Chapter 2

Multivariate Versions of Bartlett's Formula

This section quantifies the form of the asymptotic covariance matrix of the sample autocovariances in a multivariate stationary time series — the classic Bartlett formula. Such quantification is useful in many statistical inferences involving autocovariances. While joint asymptotic normality of the sample autocovariances is well-known in univariate settings, explicit forms of the asymptotic covariances have not been investigated in the general multivariate non-Gaussian case. We fill this gap by providing such an analysis, bookkeeping all skewness terms. Additionally, following a recent univariate paper by Francq and Zakoian, we consider linear processes driven by non-independent errors, a feature that permits consideration of multivariate GARCH processes.

2.1 Introduction

Considers a d-dimensional stationary time series $\{X_t\}$ satisfying the linear process representation

$$X_t = \sum_{k=-\infty}^{\infty} \Psi_k Z_{t-k}, \tag{2.1.1}$$

where $\sum_{k=-\infty}^{\infty} |\Psi_k| < \infty$ in a component by component sense. Throughout, we assume that $\{Z_t\}$ is d-dimensional white noise with finite fourth moments and covariance matrix $\Sigma = E[Z_t Z_t']$; stronger assumptions on $\{Z_t\}$ will occasionally be imposed. Let $\mu = E[X_t]$ be the series mean and

$$\Gamma(h) = E[(X_{t+h} - \mu)(X_t - \mu)'] = \{\gamma_{i,j}(h)\}_{i,j=1}^{d}$$

be the theoretical lag h autocovariance. Our notation for the sample autocovariances is

$$\hat{\Gamma}(h) = \frac{1}{n} \sum_{t=1}^{n-h} (X_{t+h} - \bar{X})(X_t - \bar{X})' = \{\hat{\gamma}_{i,j}(h)\}_{i,j=1}^{d},$$

where the data are $X_1, \ldots, X_n$ and $\bar{X} = n^{-1} \sum_{t=1}^{n} X_t$.

When Σ is invertible, it is possible to reduce consideration to $\Sigma = I_d$, where I_d denotes the d-dimensional identity matrix. This is seen by noting that

$$X_t = \sum_{k=-\infty}^{\infty} \Psi_k^* Z_{t-k}^*,$$

where $\Psi_k^* = \Psi_k \Sigma^{1/2}$ and $\{Z_t^*\}$ defined pointwise by $Z_t^* = \Sigma^{-1/2} Z_t$ is zero mean d-variate white noise with covariance matrix I_d. This reduction, however, does not overly simplify our future computations; hence, we work with the model as written in (2.1.1).

Our objective is to establish joint asymptotic normality of the random matrices $\hat{\Gamma}(0)$, $\hat{\Gamma}(1), \ldots, \hat{\Gamma}(L)$ as a function of the sample size n and identify the limiting information matrix. This issue has been classically settled in univariate treatments on time series, for example, Bartlett [1], Hannan [8], Brockwell and Davis [3], and Shumway and Stoffer [14]. There, the noise process $\{Z_t\}$ is commonly assumed to be independent and identically distributed (IID) with a finite fourth moment. Recently, Francq and Zakoian [5] considered univariate extensions of Bartlett's formula when $\{Z_t\}$ is not IID, but rather satisfied a fourth-order symmetry condition. This permits inferences in GARCH and other processes satisfying (2.1.1) where independence of the Z_t's does not hold. Later, we investigate multivariate results in this setting.

General multivariate versions of Bartlett's result do not exist. However, many authors have trodden adjacent to the problem. Hannan [8], Romano and Thombs [12], and Berlinet and Francq [2] give the formula

$$\lim_{n\to\infty} n\mathrm{Cov}(\hat{\gamma}_{i,j}(h), \hat{\gamma}_{i',j'}(k)) = \sum_{\ell=-\infty}^{\infty} \mathrm{Cov}(X_t(i)X_{t-h}(j), X_{t+\ell}(i')X_{t+\ell-k}(j'))$$

for fourth-order stationary series but do not attempt to derive an asymptotic covariance matrix in terms of the moments of $\{Z_t\}$ or second-order properties of $\{X_t\}$. Brockwell and Davis [3] state Bartlett's multivariate formula for sample autocorrelations, but do not provide proof or consider autocovariances (asymptotic results for autocovariances and autocorrelations structurally differ when the higher order cumulants of $\{Z_t\}$ are non-zero, which is the non-Gaussian case). Shumway and Stoffer [14] handle the multivariate case by citing Brockwell and Davis [3]. While Fuller [6] (Theorem 6.4.1) does consider multivariate autocovariances and autocorrelations, his arguments only apply to Gaussian processes, where skewness terms are zero. Lütkepohl [11] and Reinsel [13], two other prominent multivariate time series references, do not pursue the issue. Given this, it seems worthwhile to derive a multivariate version of Bartlett's result in as much generality as possible. And while our arguments are largely bookkeeping, the bookkeeping is sometimes cumbersome.

Arguments justifying normality in the limiting distribution of sample covariances and correlations follow the same line of reasoning as Brockwell and Davis [3] (Chapter 7) when $\{Z_t\}$ is IID with a finite fourth moment; we will not repeat this logic here. Instead, we focus on identifying an explicit form for the limiting covariance, which is

$$\lim_{n\to\infty} nE[\hat{\Gamma}(p) \otimes \hat{\Gamma}(q)],$$

where $\otimes$ denotes the usual Kronecker product (see Appendix A). In this pursuit, we define the covariance between two random matrices X and Y as

$$\mathrm{Cov}(X, Y) = E[X \otimes Y] - E[X] \otimes E[Y].$$

There is a caveat here: a non-singular limit distribution for $\sqrt{n}\hat{\Gamma}(0)$ does not exist as a $d \times d$ multivariate normal random matrix. This is simply because $\hat{\gamma}_{i,j}(0) = \hat{\gamma}_{j,i}(0)$ for $i, j \in \{1, \ldots, d\}$, so that some of the components of $\hat{\Gamma}(0)$ are redundant. Nonetheless, it is convenient to allow singular covariance matrices in the limit and we do not mention this issue further.

2.2 Preliminaries

Clarifying notation, suppose that $c_i \in \mathbf{R}^n$ for $i = 1, \ldots, m$ and set $C = [c_1, \ldots, c_m]$. Then $\mathrm{vec}(C)$ is defined as the mn-dimensional vector formed by stacking the columns of C on top of one another:

$$\mathrm{vec}(C) = \begin{pmatrix} c_1 \\ \vdots \\ c_m \end{pmatrix} \in \mathbf{R}^{mn}.$$

The Kronecker product of the $m \times n$ matrix A and $r \times s$ matrix B is defined as the $mr \times ns$ matrix of form

$$A \otimes B = \begin{bmatrix} a_{1,1}B & \ldots & a_{1,n}B \\ \vdots & \ddots & \vdots \\ a_{m,1}B & \ldots & a_{m,n}B \end{bmatrix}.$$

Several identities that will be used repeatedly are worth collecting here. If C is an $n \times p$ matrix and D is an $s \times t$ matrix, then

$$(A \otimes B)' = (A' \otimes B') \tag{2.2.2}$$

and

$$(A \otimes B)(C \otimes D) = (AC) \otimes (BD), \tag{2.2.3}$$

both sides of (2.2.3) being $mr \times pt$ matrices. Also,

$$\mathrm{vec}(ABC) = (C' \otimes A)\mathrm{vec}(B). \tag{2.2.4}$$

Another useful identity is

$$\mathrm{vec}(A \otimes B) = (I_d \otimes K \otimes I_d)[\mathrm{vec}(A) \otimes \mathrm{vec}(B)], \tag{2.2.5}$$

where K is the $d^2 \times d^2$ matrix such that $\mathrm{vec}(A') = K\mathrm{vec}(A)$. The form of K is discussed on page 466 of Lütkepohl [11]; however, we note that all entries in K are either zero or one.

While Kronecker products are not commutative, they are permutation equivalent. This means that there exist permutation matrices P and Q such that

$$(A \otimes B) = P(B \otimes A) \tag{2.2.6}$$

and

$$(A \otimes B) = (B \otimes A)Q \tag{2.2.7}$$

Here, P and $Q = P'$ are $d^2 \times d^2$ orthogonal permutation matrices whose entries are either zero or unity. In fact, some analysis will show that the unit entries of P are generated by $P_{d(\nu-1)+j+1,\nu+jd} = 1$, $1 \le \nu \le d$ and $0 \le j \le d-1$. One has $P^2 = Q^2 = I_{d^2}$.

Chain rule derivative relations for matrices of appropriate dimensions are

$$\frac{\partial \mathrm{vec}(BAC)}{\partial \mathrm{vec}(A)} = C' \otimes B, \quad \frac{\partial \mathrm{vec}(BC)}{\partial \mathrm{vec}(A)} = (C' \otimes I)\frac{\partial \mathrm{vec}(B)}{\partial \mathrm{vec}(A)} + (I \otimes B)\frac{\partial \mathrm{vec}(C)}{\partial \mathrm{vec}(A)}, \tag{2.2.8}$$

where I denotes an identity matrix of appropriate dimension. Moreover, if A is an $n \times n$ matrix and B an $m \times m$ matrix, the $nm \times nm$ Kronecker sum is defined as

$$A \oplus B = A \otimes I_m + I_n \otimes B, \tag{2.2.9}$$

where I_d denotes the d-dimensional identity matrix.

2.3 Results

Our first result considers asymptotic normality in the simplest case: that where $\{Z_t\}$ is IID.

Theorem 2.1 Consider $\{X_t\}$ in (1.1) and suppose that $\{Z_t\}$ is zero mean IID noise with the $d \times d$ variance matrix $E[Z_t Z_t'] = \Sigma$ and the $d^2 \times d^2$ skewness matrix $\eta = E[Z_t Z_t' \otimes Z_t Z_t'] < \infty$. Then the asymptotic normality

$$\begin{pmatrix} \hat{\Gamma}(p) \\ \hat{\Gamma}(q) \end{pmatrix} \sim AN\left(\begin{pmatrix} \Gamma(p) \\ \Gamma(q) \end{pmatrix}, n^{-1} \begin{pmatrix} V_{p,p} & V_{p,q} \\ V_{q,p} & V_{q,q} \end{pmatrix} \right)$$

holds, where $V_{i,j}$ is a $d^2 \times d^2$ dimensional matrix with the structure

$$V_{i,j} = S_{i,j} + \sum_{k=-\infty}^{\infty} [\mathrm{vec}(\Gamma(k))\mathrm{vec}(\Gamma(k-i+j))' + P(\Gamma(k-i) \otimes \Gamma(k+j))], \tag{2.3.10}$$

where $S_{i,j}$ is the skewness that has the form

$$\mathrm{vec}(S_{i,j}) = [\widetilde{\Gamma}(i) \otimes \widetilde{\Gamma}(j)]\mathrm{vec}(M)$$

with $M = \eta - \mathrm{vec}(\Sigma)\mathrm{vec}(\Sigma)' - P(\Sigma \otimes \Sigma) - \Sigma \otimes \Sigma$ and

$$\widetilde{\Gamma}(i) = \sum_{\ell=-\infty}^{\infty} \Psi_{\ell+i} \otimes \Psi_{\ell}. \tag{2.3.11}$$

Here, P is a $d^2 \times d^2$ orthogonal permutation matrix whose entries are either zero or unity. The unit entries of P are generated as $P_{d(\nu-1)+j+1,\nu+jd} = 1$ for $1 \le \nu \le d$ and $0 \le j \le d-1$. The quantity $\widetilde{\Gamma}(i)$ is a second moment quantity and satisfies $E[X_{t+i} \otimes X_t] = \widetilde{\Gamma}(i)\mathrm{vec}(\Sigma)$.

The result in Theorem 2.1 reduces to the classical result when $d = 1$. The component $S_{i,j}$ is viewed as the contribution due to skewness of Z_t. In the case where $\{X_t\}$ is Gaussian, $S_{i,j} = 0$. It is interesting to note the differences between (2.3.10) and the univariate version of (2.3.10), which is

$$\lim_{n\to\infty} n\mathrm{Cov}(\hat{\gamma}(p), \hat{\gamma}(q)) = (\eta - 3)\gamma(p)\gamma(q) + \sum_{k=-\infty}^{\infty} [\gamma(k)\gamma(k-p+q) + \gamma(k+q)\gamma(k-p)],$$

with $\eta = E[Z_t^4]$. In the univariate case, one does not need a permutation matrix P to scramble the orders of the components. Also, there is no need to stack components with vec operations, nor does the Kronecker product arise. The form of the skewness is also more unwieldy.

The next result is a component by component result of Theorem 2.1 and can be obtained by extracting sub-blocks of the information matrix in Theorem 2.1, or by arguing from scratch with four-fold summations as in the Proof of Theorem 2.1. The details are left to the reader.

Theorem 2.2 Under the assumptions of Theorem 2.1,

$$\begin{pmatrix} \hat{\gamma}_{a,b}(p) \\ \hat{\gamma}_{c,d}(q) \end{pmatrix} \sim AN\left(\begin{pmatrix} \gamma_{a,b}(p) \\ \gamma_{c,d}(q) \end{pmatrix}, n^{-1} \begin{pmatrix} m_{p,p} & m_{p,q} \\ m_{q,p} & m_{q,q} \end{pmatrix} \right),$$

where

$$m_{i,j} = s_{i,j} + \sum_{h=-\infty}^{\infty} [\gamma_{a,c}(h)\gamma_{b,d}(h-i+j) + \gamma_{a,d}(h+j)\gamma_{b,c}(h-i)]. \qquad (2.3.12)$$

and $s_{i,j}$ is the (c,d)th entry in the (a,b)th $d^2 \times d^2$ subblock of $S_{i,j}$ in Theorem 2.1. ♣

We now move to settings where $\{Z_t\}$ is not IID. Akin to Francq and Zakoian [5], we make the symmetry assumption

$$E[Z_{t_1} Z'_{t_2} \otimes Z_{t_3} Z'_{t_4}] = 0 \tag{2.3.13}$$

when $t_1 \neq t_2$, $t_1 \neq t_3$ and $t_1 \neq t_4$. To proceed, we need notation for the stationary series $\{Z_t Z'_t\}$ and $\{Z_t \otimes Z_t\}$. For this, we make the definitions

$$\Gamma_{Z_t Z'_t}(h) = \mathrm{Cov}((Z_{t+h} Z'_{t+h}), (Z_t Z'_t)) = E[Z_{t+h} Z'_{t+h} \otimes Z_t Z'_t] - \Sigma \otimes \Sigma \tag{2.3.14}$$

and

$$\begin{aligned} \Gamma_{Z_t \otimes Z_t}(h) &= \mathrm{Cov}((Z_{t+h} \otimes Z_{t+h}), (Z_t \otimes Z_t)) \\ &= E[(Z_{t+h} \otimes Z_{t+h})(Z_t \otimes Z_t)'] - E[Z_{t+h} \otimes Z_{t+h}] E[Z_t \otimes Z_t]'. \end{aligned} \tag{2.3.15}$$

Observe that $\Gamma_{Z_t Z'_t}(h)$ and $\Gamma_{Z_t \otimes Z_t}(h)$ are $d^2 \times d^2$ matrices. Define the memory $\kappa = \sum_{h=-\infty}^{\infty} |\Gamma_{Z_t Z'_t}(h)|$ and when all components of κ are finite, set

$$\kappa^* = \left[\sum_{h=-\infty}^{\infty} \Gamma_{Z_t Z'_t}(h) \right] - \Gamma_{Z_t Z'_t}(0).$$

Our next result establishes the form of the limiting information matrix of the sample autocovariances. For this, additional assumptions on $\{Z_t\}$ are needed to ensure asymptotic normality. In fact, counterexamples exist where sample autocovariances are asymptotically non-Gaussian when $\{Z_t\}$ does not mix rapidly enough (even in one dimension). Mixing conditions that are sufficient to induce asymptotic normality of the sample autocovariances are presented in Hannan [9], Chanda [4], Romano and Thombs [12], Berlinet and Francq

[2], and Giraitis [7]; we refer the reader to these references for more.

Theorem 2.3 Consider $\{X_t\}$ in (2.1.1) where $\{Z_t\}$ satisfies (2.3.13) and suppose that $\kappa < \infty$ and that $\{Z_t\}$ mixes rapidly enough to guarantee asymptotic normality of the sample autocovariances (for example, satisfies Theorem 2.1 in Chanda [4]). Then for any non-negative integer p and q,

$$\begin{pmatrix} \hat{\Gamma}(p) \\ \hat{\Gamma}(q) \end{pmatrix} \sim AN\left(\begin{pmatrix} \Gamma(p) \\ \Gamma(q) \end{pmatrix}, n^{-1} \begin{pmatrix} W_{p,p} & W_{p,q} \\ W_{q,p} & W_{q,q} \end{pmatrix} \right).$$

Here, $W_{i,j}$ is a $d^2 \times d^2$ dimensional matrix with the structure

$$W_{i,j} = V_{i,j} + V^*_{i,j},$$

where

$$\begin{aligned} \mathrm{vec}(V^*_{i,j}) &= [\widetilde{\Gamma}(i) \otimes \widetilde{\Gamma}(j)]\mathrm{vec}(\kappa^*) \\ &+ \sum_{\ell \neq i} (\widetilde{\Gamma}(\ell) \otimes \widetilde{\Gamma}(\ell - i + j))\mathrm{vec}(\Gamma_{Z_t \otimes Z_t}(\ell - i)) \\ &+ \sum_{\ell \neq i} (\widetilde{\Gamma}(\ell) \otimes \widetilde{\Gamma}(-\ell + i + j))\mathrm{vec}(P\Gamma_{Z_t Z_t'}(\ell - i)). \end{aligned}$$

Here, $V_{i,j}$ and $\widetilde{\Gamma}(\cdot)$ are defined by (2.3.10) and (2.3.11) respectively.

For pairwise autocovariances, we obtain the following. The result is obtained in a similar manner to which Theorem 2.2 follows from Theorem 2.1.

Theorem 2.4 Under the assumptions of Theorem 2.3,

$$\begin{pmatrix} \hat{\gamma}_{a,b}(p) \\ \hat{\gamma}_{c,d}(q) \end{pmatrix} \sim AN\left(\begin{pmatrix} \gamma_{a,b}(p) \\ \gamma_{c,d}(q) \end{pmatrix}, n^{-1} \begin{pmatrix} w_{p,p} & w_{p,q} \\ w_{q,p} & w_{q,q} \end{pmatrix} \right).$$

Here, $w_{i,j} = m_{i,j} + m^*_{i,j}$, where $m_{i,j}$ and $m^*_{i,j}$ are the (c, d)th entry in the (a, b)th $d^2 \times d^2$

subblock of $V_{i,j}$ in Theorem 2.1 and $V^*_{i,j}$ in Theorem 2.3, respectively.

2.4 Discussion

Remark 2.5 Asymptotic properties of the sample autocorrelation function can also be quantified. The lag h autocorrelation $\rho(h)$ is

$$\rho(h) = D^{-1/2}\Gamma(h)D^{-1/2} = \left[\frac{\gamma_{i,j}(h)}{\sqrt{\gamma_{i,i}(0)\gamma_{j,j}(0)}}\right]_{i,j=1}^{d},$$

where $D = \text{diag}(\Gamma(0))$. The lag h sample autocorrelation is

$$\hat{\rho}(h) = \hat{D}^{-1/2}\hat{\Gamma}(h)\hat{D}^{-1/2}.$$

Observe that $(\hat{\rho}(p), \hat{\rho}(q))$ depends only on $\hat{\Gamma}(0)$, $\hat{\Gamma}(p)$, and $\hat{\Gamma}(q)$. The partial derivative matrix of this transformation is

$$J = \begin{pmatrix} \frac{\partial\rho(p)}{\partial\Gamma(0)} & \frac{\partial\rho(p)}{\partial\Gamma(p)} & \frac{\partial\rho(p)}{\partial\Gamma(q)} \\ \frac{\partial\rho(q)}{\partial\Gamma(0)} & \frac{\partial\rho(q)}{\partial\Gamma(p)} & \frac{\partial\rho(q)}{\partial\Gamma(q)} \end{pmatrix}.$$

Notice that J has a 2×3 block structure where each block is a $d^2 \times d^2$ matrix. Matrix derivatives are defined by stacking elements of the matrix in the usual vec fashion; e.g.,

$$\frac{\partial\rho(p)}{\partial\Gamma(0)} = \frac{\partial\text{vec}(\rho(p))}{\partial\text{vec}(\Gamma(0))}.$$

Applying (2.2.8) several times gives

$$\frac{\partial\text{vec}(\rho(p))}{\partial\Gamma(0)} = (D^{-1/2}\Gamma(p)' \otimes I_d)\Lambda + (I_d \otimes D^{-1/2}\Gamma(p))\Lambda,$$

where $\Lambda = \partial\text{vec}(D^{-1/2})/\partial\text{vec}(\Gamma(0))$ is the $d^2 \times d^2$ diagonal matrix whose only non-zero entries are

$$\Lambda_{i,i} = -\frac{1}{2}\gamma_{i,i}^{-3/2}(0)$$

when $i = k(d+1)+1$ for some k in $\{0, 1, \ldots, d-1\}$.

The remaining blocks in J are similarly computed:

$$J = \begin{pmatrix} [D^{-1/2}\Gamma'(p) \oplus D^{-1/2}\Gamma(p)]\Lambda & D^{-1/2} \otimes D^{-1/2} & 0 \\ [D^{-1/2}\Gamma'(q) \oplus D^{-1/2}\Gamma(q)]\Lambda & 0 & D^{-1/2} \otimes D^{-1/2} \end{pmatrix}. \tag{2.4.16}$$

Now let

$$\Upsilon = \begin{pmatrix} V_{0,0} & V_{0,p} & V_{0,q} \\ V_{p,0} & V_{p,p} & V_{p,q} \\ V_{q,0} & V_{q,p} & V_{q,q} \end{pmatrix}$$

denote the asymptotic covariance structure of $(\hat{\Gamma}(0), \hat{\Gamma}(p), \hat{\Gamma}(q))$ specified in Theorem 2.1. Applying a multivariate delta method (see Proposition 6.4.3 in Brockwell and Davis [3], for example) gives

$$\begin{pmatrix} \hat{\rho}(p) \\ \hat{\rho}(q) \end{pmatrix} \sim AN\left(\begin{pmatrix} \rho(p) \\ \rho(q) \end{pmatrix}, n^{-1}J\Upsilon J'\right).$$

An implication here is that

$$\lim_{n\to\infty} n\mathrm{Cov}(\hat{\rho}(p), \hat{\rho}(q)) = [J\Upsilon J']_{1,2},$$

where $[\cdot]_{1,2}$ denotes the row 1, column 2 $d \times d$ subblock of the matrix in brackets. Applying (2.4.16) reveals

$$\begin{aligned}\lim_{n\to\infty} n\mathrm{Cov}(\hat{\rho}(p),\hat{\rho}(q)) &= [D^{-1/2}\Gamma'(p)\oplus D^{-1/2}\Gamma(p)]V_{0,0}[D^{-1/2}\Gamma'(q)\oplus D^{-1/2}\Gamma(q)]\Lambda^2\\ &\quad +(D^{-1/2}\otimes D^{-1/2})V_{p,0}[D^{-1/2}\Gamma'(q)\oplus D^{-1/2}\Gamma(q)]\Lambda\\ &\quad +[D^{-1/2}\Gamma'(p)\oplus D^{-1/2}\Gamma(p)]V_{0,q}(D^{-1/2}\otimes D^{-1/2})\Lambda\\ &\quad +(D^{-1/2}\otimes D^{-1/2})V_{p,q}(D^{-1/2}\otimes D^{-1/2}). \qquad (2.4.17)\end{aligned}$$

♣

Remark 2.6 Suppose that $\{X_t\}$ is Gaussian so that skewness contributions are zero. We seek to explicitly identify

$$\lim_{n\to\infty} n\mathrm{Cov}(\hat{\rho}_{a,b}(p),\hat{\rho}_{c,d}(q)).$$

Observe that $(\hat{\rho}(p),\hat{\rho}(q))$ depends only on $(\hat{\gamma}_{a,a}(0),\hat{\gamma}_{a,b}(p),\hat{\gamma}_{b,b}(0),\hat{\gamma}_{c,c}(0),\hat{\gamma}_{c,d}(q),\hat{\gamma}_{d,d}(0))$. The partial derivative matrix of this transformation has the form

$$J^* = \begin{pmatrix} A_1 & A_2 & A_3 & 0 & 0 & 0\\ 0 & 0 & 0 & B_1 & B_2 & B_3\end{pmatrix},$$

where

$$A_1 = -\frac{1}{2}\gamma_{a,a}^{-\frac{3}{2}}(0)\gamma_{a,b}(p)\gamma_{b,b}^{-\frac{1}{2}}(0), \quad A_2 = \gamma_{a,a}^{-\frac{1}{2}}(0)\gamma_{b,b}^{-\frac{1}{2}}(0), \quad A_3 = -\frac{1}{2}\gamma_{a,a}^{-\frac{1}{2}}(0)\gamma_{a,b}(p)\gamma_{b,b}^{-\frac{3}{2}}(0)$$

and

$$B_1 = -\frac{1}{2}\gamma_{c,c}^{-\frac{3}{2}}(0)\gamma_{c,d}(q)\gamma_{d,d}^{-\frac{1}{2}}(0), \quad B_2 = \gamma_{c,c}^{-\frac{1}{2}}(0)\gamma_{d,d}^{-\frac{1}{2}}(0) \quad B_3 = -\frac{1}{2}\gamma_{c,c}^{-\frac{1}{2}}(0)\gamma_{c,d}(q)\gamma_{d,d}^{-\frac{3}{2}}(0).$$

Let Ξ denote the limiting covariance structure of

$$(\hat{\gamma}_{a,a}(0), \hat{\gamma}_{a,b}(p), \hat{\gamma}_{b,b}(0), \hat{\gamma}_{c,c}(0), \hat{\gamma}_{c,d}(q), \hat{\gamma}_{d,d}(0))'$$

specified in Theorem 2.2. Arguing as in the last remark with a delta method shows that

$$\begin{aligned}
\lim_{n\to\infty} n\mathrm{Cov}(\hat{\rho}_{a,b}(p), \hat{\rho}_{c,d}(q)) &= \left[J^*\Xi(J^*)'\right]_{1,2} \\
&= A_1\Xi_{1,4}B_1 + A_2\Xi_{2,4}B_1 + A_3\Xi_{3,4}B_1 + A_1\Xi_{1,5}B_2 + A_2\Xi_{2,5}B_2 \\
&+ A_3\Xi_{3,5}B_2 + A_1\Xi_{1,6}B_3 + A_2\Xi_{2,6}B_3 + A_3\Xi_{3,6}B_3.
\end{aligned}$$

Expanding and simplifying gives

$$\begin{aligned}
\lim_{n\to\infty} n\mathrm{Cov}(\hat{\rho}_{a,b}(p), \hat{\rho}_{c,d}(q)) &= \sum_{h=-\infty}^{\infty} \Big[\rho_{a,c}(h)\rho_{b,d}(h-p+q) + \rho_{a,d}(h+q)\rho_{b,c}(h-p) \\
&- \rho_{a,b}(p)\{\rho_{a,c}(h)\rho_{a,d}(h+q) + \rho_{b,c}(h)\rho_{b,d}(h+q)\} \\
&- \rho_{c,d}(q)\{\rho_{a,c}(h)\rho_{b,c}(h-p) + \rho_{a,d}(h)\rho_{b,d}(h-p)\} \\
&+ \frac{1}{2}\rho_{a,b}(p)\rho_{c,d}(q)\{\rho^2_{a,c}(h) + \rho^2_{b,c}(h) + \rho^2_{a,d}(h) + \rho^2_{b,d}(h)\}\Big].
\end{aligned}$$

This verifies the bivariate Gaussian formula in Theorem 11.2.3 of Brockwell and Davis [3]. We have been unable to verify the formula in Corollary 6.4.1.1 of Fuller [6] (which should be the same) and suspect a typographical error.

2.5 Examples

This section presents multivariate extensions of two classical time series derivations: identifying $V_{p,p}$ in first order autoregressions and general moving-averages. The reader will gain feel for the complexity of the computations in multivariate settings.

Example 3.1 Consider the first order causal autoregression satisfying

$$X_t = \Phi X_{t-1} + Z_t, \tag{2.5.18}$$

where $\{Z_t\}$ is zero mean d-variate white noise with covariance matrix Σ. We assume that Φ is invertible. Causality implies that all eigenvalues of Φ are less than unity in absolute value. For simplicity, we work with a Gaussian series so that skewness terms are zero. Our goal here is to identify $V_{p,p}$. This is useful since $\hat{\Gamma}(p) \sim AN(\Gamma(p), V_{p,p}/n)$. Equation (2.3.10) gives

$$V_{p,p} = \sum_{k=-\infty}^{\infty} \left[\mathrm{vec}(\Gamma(k))\mathrm{vec}(\Gamma(k))' + P(\Gamma(k-p) \otimes \Gamma(k+p)) \right]. \tag{2.5.19}$$

To compute $V_{p,p}$, we need the autocovariances of the model. Taking variances on both sides of (2.5.18) produces

$$\Gamma(0) - \Phi\Gamma(0)\Phi' = \Sigma. \tag{2.5.20}$$

While this equation cannot be solved explicitly for $\Gamma(0)$, it is possible to obtain $\mathrm{vec}(\Gamma(0))$ in an explicit manner. The components of $\Gamma(0)$ can be recovered from $\mathrm{vec}(\Gamma(0))$ as follows: for $1 \leq \ell \leq d^2$, the ℓth component of $\mathrm{vec}(\Gamma(0))$ is $\gamma_{i,j}(0)$ with $i = \lfloor (\ell-1)/d \rfloor + 1$ and $j = \ell - (i-1)d$. To obtain $\mathrm{vec}(\Gamma(0))$, take vecs of both sides of (2.5.20), apply (2.2.4), and solve the resulting equation to get

$$\mathrm{vec}(\Gamma(0)) = (I_{d^2} - \Phi \otimes \Phi)^{-1}\mathrm{vec}(\Sigma). \tag{2.5.21}$$

The causality assumption guarantees that $I_{d^2} - \Phi \otimes \Phi$ is invertible. Covariances at higher lags are obtained from (2.5.18) and are $\Gamma(h) = \Phi^h\Gamma(0)$ and $\Gamma(-h) = \Gamma(0)(\Phi')^h$ for $h \geq 1$. For $h \geq 1$, (2.2.4) and induction give $\mathrm{vec}(\Gamma(h)) = (I_d \otimes \Phi)^h \mathrm{vec}(\Gamma(0))$. Combining this with

(2.5.21) produces

$$\mathrm{vec}(\Gamma(k)) = (I_d \otimes \Phi)^k (I_{d^2} - \Phi \otimes \Phi)^{-1} \mathrm{vec}(\Sigma), \quad k \geq 0. \tag{2.5.22}$$

Similarly,

$$\mathrm{vec}(\Gamma(-k)) = (\Phi' \otimes I_d)^k (I_{d^2} - \Phi \otimes \Phi)^{-1} \mathrm{vec}(\Sigma), \quad k > 0. \tag{2.5.23}$$

The quantity $\mathrm{vec}(V_{p,p})$, a d^4-dimensional vector, can now be explicitly calculated as follows. We begin by working on the summation in (2.5.19) involving the vec terms. Applying (2.5.22) gives

$$\sum_{k=0}^{\infty} \mathrm{vec}(\Gamma(k))\mathrm{vec}(\Gamma(k))' = \sum_{k=0}^{\infty} (I_d \otimes \Phi)^k BB' (I_d \otimes \Phi')^k = \sum_{k=0}^{\infty} U_1^k G (U_1')^k := S_1,$$

where $U_1 = (I_d \otimes \Phi)$, $B = (I_{d^2} - \Phi \otimes \Phi)^{-1} \mathrm{vec}(\Sigma)$, and $G = BB'$. The infinite geometric sum S_1 can be shown to satisfy the relationship

$$S_1 - U_1 S_1 U_1' = G.$$

Now take vecs on both sides of this equation and argue as in the lines that produced (2.5.21) from (2.5.20) to get

$$\mathrm{vec}(S_1) = (I_{d^4} - U_1 \otimes U_1)^{-1} \mathrm{vec}(G).$$

Similar arguments produce

$$\sum_{k=-\infty}^{-1} \mathrm{vec}(\Gamma(k))\mathrm{vec}(\Gamma(k))' = \sum_{k=1}^{\infty} U_2^k G (U_2')^k := S_2,$$

where $U_2 = (\Phi' \otimes I_d)$ and

$$\mathrm{vec}(S_2) = (I_{d^4} - U_2 \otimes U_2)^{-1}(U_2 \otimes U_2)\mathrm{vec}(G).$$

It follows that

$$\begin{aligned}\mathrm{vec}\left(\sum_{k=-\infty}^{\infty} \mathrm{vec}(\Gamma(k))\mathrm{vec}(\Gamma(k))'\right) &= (I_{d^4} - U_1 \otimes U_1)^{-1}\mathrm{vec}(G) \\ &+ (I_{d^4} - U_2 \otimes U_2)^{-1}(U_2 \otimes U_2)\mathrm{vec}(G). \quad (2.5.24)\end{aligned}$$

To evaluate the vec of the second summation in (2.5.19), we partition the infinite summation into three pieces. First, use the linearity of vec and (2.2.4) to get

$$\mathrm{vec}\left(\sum_{k=p}^{\infty} P[\Gamma(k-p) \otimes \Gamma(k+p)]\right) = \sum_{k=p}^{\infty}(I_{d^2} \otimes P)\mathrm{vec}\left(\Gamma(k-p) \otimes \Gamma(k+p)\right). \quad (2.5.25)$$

Using (2.2.5) in (2.5.25) and applying (2.5.22) and (2.2.3) gives

$$\mathrm{vec}\left(\sum_{k=p}^{\infty} P[\Gamma(k-p) \otimes \Gamma(k+p)]\right) = CS_3(B \otimes B),$$

where $C = (I_{d^2} \otimes P)(I_d \otimes K \otimes I_d)$ and S_3 is the geometric sum

$$S_3 = I_{d^2} \otimes U_1^{2p} + U_1 \otimes U_1^{2p+1} + \cdots .$$

One can sum this geometric series explicitly to get $S_3 = (I_{d^4} - U_1 \otimes U_1)^{-1}(I_{d^2} \otimes U_1^{2p})$ and hence

$$\mathrm{vec}\left(\sum_{k=p}^{\infty} P[\Gamma(k-p) \otimes \Gamma(k+p)]\right) = C(I_{d^4} - U_1 \otimes U_1)^{-1}(I_{d^2} \otimes U_1^{2p})(B \otimes B). \quad (2.5.26)$$

Similar arguments provide

$$\operatorname{vec}\left(\sum_{k=-\infty}^{-p} P[\Gamma(k-p)\otimes\Gamma(k+p)]\right) = C(I_{d^4} - U_2\otimes U_2)^{-1}(U_2^{2p}\otimes I_{d^2})(B\otimes B). \quad (2.5.27)$$

Causality guarantees invertibility of $(I_{d^4} - U_1\otimes U_1)$ and $(I_{d^4} - U_2\otimes U_2)$. The finite sum is similarly handled. In the end, one gets

$$\operatorname{vec}\left(\sum_{k=-p+1}^{p-1} P[\Gamma(k-p)\otimes\Gamma(k+p)]\right) = C(I_{d^4} - U_2^{-1}\otimes U_1)^{-1}(U_2^{2p-1}\otimes U_1 - I_{d^2}\otimes U_1^{2p})(B\otimes B). \quad (2.5.28)$$

Now combine (2.5.24), (2.5.26), (2.5.27), and (2.5.28) and simplify to get

$$\begin{aligned}
\operatorname{vec}(V_{p,p}) &= \big[(I_{d^4} - U_1\otimes U_1)^{-1} + (I_{d^4} - U_2\otimes U_2)^{-1}(U_2\otimes U_2)\big]\operatorname{vec}(G) \\
&+ C\big[(I_{d^4} - U_1\otimes U_1)^{-1}(I_{d^2}\otimes U_1^{2p}) + (I_{d^4} - U_2\otimes U_2)^{-1}(U_2^{2p}\otimes I_{d^2}) \\
&+ (I_{d^4} - U_2^{-1}\otimes U_1)^{-1}(U_2^{2p-1}\otimes U_1 - I_{d^2}\otimes U_1^{2p})\big](B\otimes B).
\end{aligned}$$

This expression reduces to that in Example 7.2.3 in Brockwell and Davis [3] when $d = 1$.

Example 3.2 Consider a qth order moving-average $\{X_t\}$ satisfying

$$X_t = Z_t + \Theta_1 Z_{t-1} + \cdots + \Theta_q Z_{t-q},$$

where $\{Z_t\}$ is zero mean white noise with covariance matrix Σ. We again consider the Gaussian case and concentrate on computation of the sum in (2.3.10). Tests for moving-averages of order q are sometimes constructed by assessing whether or not $\hat{\Gamma}(q+1)$ is significantly different from zero. This is quantified via

$$\hat{\Gamma}(q+1) \sim AN\left(0, \frac{V_{q+1,q+1}}{n}\right)$$

and the task is to identify $V_{q+1,q+1}$.

The covariance function of $\{X_t\}$ is

$$\Gamma(h) = \sum_{i=0}^{q-h} \Theta_{i+h}\Sigma\Theta_i',$$

for $h = 0, \ldots, q$ with the convention that $\Theta_0 = I_d$. Also, $\Gamma(-h) = \Gamma(h)'$. Observe that $\Gamma(h) = 0$ when $|h| > q$. Thus, (2.3.10) gives

$$V_{q+1,q+1} = \sum_{k=-q}^{q} \text{vec}(\Gamma(k))\text{vec}(\Gamma(k))'.$$

This expression reduces to the classical $\sigma^4[\gamma(0) + 2\sum_{k=1}^{q}\gamma(k)^2]$ in the univariate case (see Example 7.2.2 in Brockwell and Davis [3]). We also note that $V_{k,k} = V_{q+1,q+1}$ for $k \geq 2$.

2.6 Proofs

The following Lemma is the basis of our computations.

Lemma 1 Under the assumptions of Theorem 2.1,

$$\begin{aligned} E[X_tX_{t+p}' \otimes X_{t+h+p}X_{t+h+p+q}'] &= R_{p,q}(h+p) + \Gamma(p)\otimes\Gamma(q) + \text{vec}(\Gamma(h+p))\text{vec}(\Gamma(h+q))' \\ &+ P\Gamma(h)\otimes\Gamma(h+p+q), \end{aligned} \tag{2.6.29}$$

where

$$R_{p,q}(h+p) = \sum_{i=-\infty}^{\infty} (\Psi_i \otimes \Psi_{h+p+i})[\eta - \text{vec}(\Sigma)\text{vec}(\Sigma)' - P(\Sigma\otimes\Sigma) - \Sigma\otimes\Sigma](\Psi_{p+i}\otimes\Psi_{h+p+q+i})'.$$

Proof of Lemma 1 Expanding X_t with (2.1.1) and taking expectations provides

$$\begin{aligned}
& E[X_t X'_{t+p} \otimes X_{t+h+p} X'_{t+h+p+q}] \\
= & \sum_{i=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \sum_{\ell=-\infty}^{\infty} E[\Psi_i Z_{t-i} Z'_{t+p-j} \Psi'_j \otimes \Psi_k Z_{t+h+p-k} Z'_{t+h+p+q-\ell} \Psi'_\ell] \\
= & \sum_{T_1} E[\Psi_i Z_{t-i} Z'_{t+p-j} \Psi'_j \otimes \Psi_k Z_{t+h+p-k} Z'_{t+h+p+q-\ell} \Psi'_\ell] \\
+ & \sum_{T_2} E[\Psi_i Z_{t-i} Z'_{t+p-j} \Psi'_j \otimes \Psi_k Z_{t+h+p-k} Z'_{t+h+p+q-\ell} \Psi'_\ell] \\
+ & \sum_{T_3} E[\Psi_i Z_{t-i} Z'_{t+p-j} \Psi'_j \otimes \Psi_k Z_{t+h+p-k} Z'_{t+h+p+q-\ell} \Psi'_\ell] \\
+ & \sum_{T_4} E[\Psi_i Z_{t-i} Z'_{t+p-j} \Psi'_j \otimes \Psi_k Z_{t+h+p-k} Z'_{t+h+p+q-\ell} \Psi'_\ell] \\
=: & \text{ I+II+III+IV},
\end{aligned}$$

when the zero-mean IID structure of $\{Z_t\}$ is used. Here, T_i describes the following indices that must be summed over for a fixed t:

$$\begin{aligned}
T_1 &= \{(i,j,k,\ell) : t-i = t+p-j = t+h+p-k = t+h+p+q-\ell\} \\
&= \{-\infty < i < \infty; j = p+i; k = h+p+i; \ell = h+p+q+i\}. \qquad (2.6.30)
\end{aligned}$$

Similar reasoning gives

$$\begin{aligned}
T_2 &= \{(i,j,k,\ell) : t-i = t+p-j, t+h+p-k = t+h+p+q-\ell, t-i \neq t+h+p-k\} \\
&= \{-\infty < i < \infty; -\infty < k < \infty; j = p+i; \ell = h+p+q+i; k \neq h+p+i\}. \qquad (2.6.31)
\end{aligned}$$

$$
\begin{aligned}
T_3 &= \{(i,j,k,\ell) : t-i = t+h+p-k, t+p-j = t+h+p+q-\ell, t-i \neq t+p-j\} \\
&= \{-\infty < i < \infty; -\infty < j < \infty; k = h+p+i; \ell = h+q+j; j \neq p+i\}. \qquad (2.6.32)
\end{aligned}
$$

$$
\begin{aligned}
T_4 &= \{(i,j,k,\ell) : t-i = t+h+p+q-\ell, t+p-j = t+h+p-k, t-i \neq t+p-j\} \\
&= \{-\infty < i < \infty; -\infty < j < \infty; k = h+p+j; \ell = h+p+q+i; j \neq p+i\}. \; (2.6.33)
\end{aligned}
$$

Note that T_1 requires a single summation whereas T_2, T_3, and T_4 require double summations. We examine each of these terms case by case. For notation, let $M_t = Z_t Z_t'$ and observe that $\eta = E[M_t \otimes M_t]$.

The first term is easy to evaluate with application of (2.2.3):

$$
\begin{aligned}
\mathrm{I} &= \sum_{i=-\infty}^{\infty} E[(\Psi_i M_t \otimes \Psi_{h+p+i} M_t)(\Psi_{p+i}' \otimes \Psi_{h+p+q+i}') \\
&= \sum_{i=-\infty}^{\infty} (\Psi_i \otimes \Psi_{h+p+i}) E[M_t \otimes M_t] (\Psi_{p+i} \otimes \Psi_{h+p+q+i})' \\
&= \sum_{i=-\infty}^{\infty} (\Psi_i \otimes \Psi_{h+p+i}) \eta (\Psi_{p+i}' \otimes \Psi_{h+p+q+i}').
\end{aligned}
$$

For the second term, suppose that $t \neq s$ so that Z_t and Z_s are uncorrelated. Now simplify II into

$$\begin{aligned}
\text{II} &= \sum_{i=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} E[\Psi_i Z_t Z_t' \Psi_{p+i}' \otimes \Psi_k Z_s Z_s' \Psi_{q+k}'] \\
&- \sum_{i=-\infty}^{\infty} E[\Psi_i Z_t Z_t' \Psi_{p+i}' \otimes \Psi_{h+p+i} Z_s Z_s' \Psi_{h+p+q+i}'].
\end{aligned}$$

Using (2.1.1), one can verify that

$$\begin{aligned}
\Gamma(p) \otimes \Gamma(q) &= \sum_{i=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} E[\Psi_i Z_t Z_t' \Psi_{p+i}' \otimes \Psi_j Z_s Z_s' \Psi_{j+q}'] \\
&= \sum_{i=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} (\Psi_i \otimes \Psi_j)(\Sigma \otimes \Sigma)(\Psi_{i+p}' \otimes \Psi_{j+q}').
\end{aligned}$$

Combining the last two relations and applying (2.2.4) identifies II as

$$\begin{aligned}
\text{II} &= \Gamma(p) \otimes \Gamma(q) - \sum_{i=-\infty}^{\infty} E[\Psi_i Z_t Z_t' \Psi_{h+p+i}' \otimes \Psi_{p+i} Z_s Z_s' \Psi_{h+p+q+i}'] \\
&= \Gamma(p) \otimes \Gamma(q) - \sum_{i=-\infty}^{\infty} (\Psi_i \otimes \Psi_{p+i})(\Sigma \otimes \Sigma)(\Psi_{h+p+i}' \otimes \Psi_{h+p+q+i}'). \qquad (2.6.34)
\end{aligned}$$

For the third term,

$$
\begin{aligned}
\mathrm{III} &= \sum_{i=-\infty}^{\infty}\sum_{j=-\infty}^{\infty} E[\Psi_i Z_t Z_s' \Psi_j' \otimes \Psi_{h+p+i} Z_t Z_s' \Psi_{h+q+j}'] \\
&- \sum_{i=-\infty}^{\infty} E[\Psi_i Z_t Z_s' \Psi_{p+i}' \otimes \Psi_{h+p+i} Z_t Z_s' \Psi_{h+p+q+i}'] \\
&= \sum_{i=-\infty}^{\infty}\sum_{j=-\infty}^{\infty} E[(\Psi_i Z_t Z_s' \otimes \Psi_{h+p+i} Z_t Z_s')(\Psi_j' \otimes \Psi_{h+q+j}')] \\
&- \sum_{i=-\infty}^{\infty} E[(\Psi_i Z_t Z_s' \otimes \Psi_{h+p+i} Z_t Z_s')(\Psi_{p+i}' \otimes \Psi_{h+p+q+i}')] \\
&= \sum_{i=-\infty}^{\infty}\sum_{j=-\infty}^{\infty} (\Psi_i \otimes \Psi_{h+p+i}) E[Z_t Z_s' \otimes Z_t Z_s'](\Psi_j' \otimes \Psi_{h+q+j}')
\end{aligned}
$$

$$
- \sum_{i=-\infty}^{\infty} (\Psi_i \otimes \Psi_{h+p+i}) E[Z_t Z_s' \otimes Z_t Z_s'](\Psi_{p+i}' \otimes \Psi_{h+p+q+i}').
$$

But when $t \neq s$, $E[Z_t Z_s' \otimes Z_t Z_s'] = E[(Z_t \otimes Z_t)(Z_s' \otimes Z_s')] = \mathrm{vec}(\Sigma)\mathrm{vec}(\Sigma)'$, and hence,

$$
\begin{aligned}
\mathrm{III} &= \sum_{i=-\infty}^{\infty}\sum_{j=-\infty}^{\infty} (\Psi_i \otimes \Psi_{h+p+i})\mathrm{vec}(\Sigma)\mathrm{vec}(\Sigma)'(\Psi_j' \otimes \Psi_{h+q+j}') \\
&- \sum_{i=-\infty}^{\infty} (\Psi_i \otimes \Psi_{h+p+i})\mathrm{vec}(\Sigma)\mathrm{vec}(\Sigma)'(\Psi_{p+i}' \otimes \Psi_{h+p+q+i}').
\end{aligned}
$$

Now apply (2.2.4) to get

$$
\begin{aligned}
\mathrm{III} &= \sum_{i=-\infty}^{\infty}\sum_{j=-\infty}^{\infty} \mathrm{vec}(\Psi_{h+p+i}\Sigma\Psi_i')\mathrm{vec}(\Psi_{h+q+j}\Sigma\Psi_j')' \\
&- \sum_{i=-\infty}^{\infty} \mathrm{vec}(\Psi_{h+p+i}\Sigma\Psi_i')\mathrm{vec}(\Psi_{h+p+q+i}\Sigma\Psi_{p+i}')' \\
&= \mathrm{vec}(\Gamma(h+p))\mathrm{vec}(\Gamma(h+q))' - \sum_{i=-\infty}^{\infty} (\Psi_i \otimes \Psi_{h+p+i})\mathrm{vec}(\Sigma)\mathrm{vec}(\Sigma)'(\Psi_{p+i}' \otimes \Psi_{h+p+q+i}').
\end{aligned}
$$

The last term is the hardest. For this term with $t \neq s$, start with

$$
\begin{aligned}
\mathrm{IV} &= \sum_{i=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} E[\Psi_i Z_{t-i} Z'_{t+p-j} \Psi'_j \otimes \Psi_{h+j} Z_{t+p-j} Z'_{t-i} \Psi'_{h+p+q+i}] \\
&- \sum_{i=-\infty}^{\infty} (\Psi_i \otimes \Psi_{h+p+i}) E[Z_t Z'_s \otimes Z_s Z'_t](\Psi'_{p+i} \otimes \Psi'_{h+p+q+i}) \\
&= \sum_{i=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} E[(\Psi_i Z_{t-i} Z'_{t+p-j} \otimes \Psi_{h+j} Z_{t+p-j} Z'_{t-i})(\Psi'_j \otimes \Psi'_{h+p+q+i})] \\
&- \sum_{i=-\infty}^{\infty} (\Psi_i \otimes \Psi_{h+p+i}) E[Z_t Z'_s \otimes Z_s Z'_t](\Psi'_{p+i} \otimes \Psi'_{h+p+q+i}) \\
&= \sum_{i=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} (\Psi_i \otimes \Psi_{h+j}) E[Z_{t-i} Z'_{t+p-j} \otimes Z_{t+p-j} Z'_{t-i}](\Psi'_j \otimes \Psi'_{h+p+q+i}) \\
&- \sum_{i=-\infty}^{\infty} (\Psi_i \otimes \Psi_{h+p+i}) E[Z_t Z'_s \otimes Z_s Z'_t](\Psi'_{p+i} \otimes \Psi'_{h+p+q+i}).
\end{aligned}
$$

Taking conditional expectations identifies the inner bracketed term:

$$
\begin{aligned}
E[Z_{t-i} Z'_{t+p-j} \otimes Z_{t+p-j} Z'_{t-i}] &= E[E[Z_{t-i} Z'_{t+p-j} \otimes Z_{t+p-j} Z'_{t-i}] \mid Z_{t+p-j}]) \\
&= E[E[(I_d \otimes Z_{t+p-j})(Z_{t-i} \otimes Z'_{t-i})(Z'_{t+p-j} \otimes I_d) \mid Z_{t+p-j}]) \\
&= E[(I_d \otimes Z_{t+p-j}) \Sigma (Z'_{t+p-j} \otimes I_d)].
\end{aligned}
$$

Hence,

$$\begin{aligned}
& \sum_{i=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} (\Psi_i \otimes \Psi_{h+j}) E[Z_{t-i} Z'_{t+p-j} \otimes Z_{t+p-j} Z'_{t-i}](\Psi'_j \otimes \Psi'_{h+p+q+i}) \\
= & \sum_{i=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} E[(\Psi_i \otimes \Psi_{h+j})(I_n \otimes Z_{t+p-j})\Sigma(Z'_{t+p-j} \otimes I_n)(\Psi'_j \otimes \Psi'_{h+p+q+i})] \\
= & \sum_{i=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} E[(\Psi_i \otimes \Psi_{h+j} Z_{t+p-j})\Sigma(Z'_{t+p-j} \otimes I_n)(\Psi'_j \otimes \Psi'_{h+p+q+i})].
\end{aligned}$$

Applying (2.2.6) gives

$$\begin{aligned}
& \sum_{i=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} (\Psi_i \otimes \Psi_{h+j}) E[Z_{t-i} Z'_{t+p-j} \otimes Z_{t+p-j} Z'_{t-i}](\Psi'_j \otimes \Psi'_{h+p+q+i}) \\
= & \sum_{i=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} E[P(\Psi_{h+j} Z_{t+p-j} \otimes \Psi_i)\Sigma(Z'_{t+p-j} \otimes I_n)(\Psi'_j \otimes \Psi'_{h+p+q+i})] \\
= & \; P \sum_{i=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} E[(\Psi_{h+j} Z_{t+p-j} \otimes \Psi_i \Sigma)(Z'_{t+p-j} \otimes I_n)(\Psi'_j \otimes \Psi'_{h+p+q+i})] \\
= & \; P \sum_{i=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} E[(\Psi_{h+j} Z_{t+p-j} Z'_{t+p-j} \Psi'_j) \otimes (\Psi_i \Sigma \Psi'_{h+p+q+i})] \\
= & \; P\Gamma(h) \otimes \Gamma(h+p+q).
\end{aligned}$$

Similar reasoning gives

$$\begin{aligned}
& \sum_{i=-\infty}^{\infty} (\Psi_i \otimes \Psi_{h+p+i}) E[Z_t Z'_s \otimes Z_s Z'_t](\Psi'_{p+i} \otimes \Psi'_{h+p+q+i}) \\
= & \sum_{i=-\infty}^{\infty} (\Psi_i \otimes \Psi_{h+p+i})(P(\Sigma \otimes \Sigma))(\Psi'_{p+i} \otimes \Psi'_{h+p+q+i}).
\end{aligned}$$

Therefore,

$$\mathrm{IV} = P\Gamma(h) \otimes \Gamma(h+p+q) - \sum_{i=-\infty}^{\infty} (\Psi_i \otimes \Psi_{h+p+i})(P(\Sigma \otimes \Sigma))(\Psi'_{p+i} \otimes \Psi'_{h+p+q+i}).$$

Putting the above computations together establishes the Lemma. ♣

Proof of Theorem 2.1: The argument is essentially the same as that on page 227 of Brockwell and Davis [3]. We provide the main points for the sake of completeness.

Observe that

$$E[\hat{\Gamma}^*(p) \otimes \hat{\Gamma}^*(q)] = n^{-2} \sum_{s=1}^{n} \sum_{t=1}^{n} E[X_t X'_{t+p} \otimes X_s X'_{s+q}],$$

where $\hat{\Gamma}^*(h) = n^{-1} \sum_{t=1}^{n} X_t X'_{t+h}$ is an unbiased estimator of $\Gamma(h)$ that can be shown to have the same asymptotic properties as $\hat{\Gamma}(h)$. Applying Lemma 1 gives

$$\begin{aligned} E[\hat{\Gamma}^*(p) \otimes \hat{\Gamma}^*(q)] &= n^{-2} \sum_{s=1}^{n} \sum_{t=1}^{n} \Big[R_{p,q}(t-s) + \Gamma(p) \otimes \Gamma(q) + \mathrm{vec}(\Gamma(s-t))\mathrm{vec}(\Gamma(s-t-p+q))' \\ &+ P\Gamma(s-t-p) \otimes \Gamma(s-t+q) \Big]. \end{aligned}$$

where

$$R_{p,q}(k) = \sum_{i=-\infty}^{\infty} (\Psi_i \otimes \Psi_{i+k})[\eta - \mathrm{vec}(\Sigma)\mathrm{vec}(\Sigma)' - P(\Sigma \otimes \Sigma) - \Sigma \otimes \Sigma](\Psi'_{i+p} \otimes \Psi'_{i+k+q}).$$

Subtracting $\Gamma(p) \otimes \Gamma(q)$ and regrouping terms by diagonals gives

$$\mathrm{Cov}(\hat{\Gamma}^*(p), \hat{\Gamma}^*(q)) = n^{-1} \sum_{|k|<n} (1 - |k|/n)(T_k + R_{p,q}(k)),$$

where

$$T_k = \text{vec}(\Gamma(k))\text{vec}(\Gamma(k-p+q))' + P\Gamma(k-p) \otimes \Gamma(k+q), \qquad -n < k < n.$$

Dominated convergence now gives

$$\lim_{n\to\infty} n\text{Cov}(\hat{\Gamma}^*(p), \hat{\Gamma}^*(q)) = S_{p,q} + \sum_{k=-\infty}^{\infty} [\text{vec}(\Gamma(k))\text{vec}(\Gamma(k-p+q))' + P(\Gamma(k-p) \otimes \Gamma(k+q))],$$

where $S_{p,q}$ is the skewness

$$S_{p,q} = \sum_{\ell_1=-\infty}^{\infty} \sum_{\ell_2=-\infty}^{\infty} (\Psi_{\ell_1} \otimes \Psi_{\ell_2}) M (\Psi'_{\ell_1+p} \otimes \Psi'_{\ell_2+q}) \tag{2.6.35}$$

and $M = \eta - \text{vec}(\Sigma)\text{vec}(\Sigma)' - P(\Sigma \otimes \Sigma) - \Sigma \otimes \Sigma$. Since the limiting properties of starred and unstarred versions of $\hat{\Gamma}(h)$ are the same, this proves the result except for the skewness statements.

Taking vecs of both sides (2.6.35) and applying (2.2.4) gives

$$\text{vec}(S_{i,j}) = \left[\sum_{\ell_1=-\infty}^{\infty} \sum_{\ell_2=-\infty}^{\infty} (\Psi_{\ell_1+i} \otimes \Psi_{\ell_2+j})(\Psi_{\ell_1} \otimes \Psi_{\ell_2}) \right] \text{vec}(M).$$

Now use (2.2.7), associativity of Kronecker products, and the definition of $\widetilde{\Gamma}(j)$ to get

$$\text{vec}(S_{i,j}) = \left[\sum_{\ell_1=-\infty}^{\infty} \Psi_{\ell_1+i} \otimes \widetilde{\Gamma}(j) \otimes \Psi_{\ell_1} \right] Q\text{vec}(M).$$

Use the same commutative tactics and the fact that $Q^2 = I_{d^2}$ to get

$$\text{vec}(S_{i,j}) = [\widetilde{\Gamma}(i) \otimes \widetilde{\Gamma}(j)]\text{vec}(M).$$

Finally, to get $E[X_{t+i} \otimes X_t] = \widetilde{\Gamma}(i)\text{vec}(\Sigma)$, expand with (2.1.1) to get

$$E[X_{t+i} \otimes X_t] = \sum_{\ell_1=-\infty}^{\infty} \sum_{\ell_2=-\infty}^{\infty} (\Psi_{\ell_1} \otimes \Psi_{\ell_2}) E[Z_{t+i-\ell_1} \otimes Z_{t-\ell_2}].$$

But since $E[Z_{t+i-\ell_1} \otimes Z_{t-\ell_2}]$ is zero unless $\ell_2 = \ell_1 - i$ and is vec(Σ) when $\ell_2 = \ell_1 - i$, the identity follows.

Proof of Theorem 2.3: We need a modification of Lemma 1. For this, expand as before to get

$$\begin{aligned} & E[X_t X'_{t+p} \otimes X_{t+h} X'_{t+h+q}] \\ = & \sum_{\ell_1=-\infty}^{\infty} \sum_{\ell_2=-\infty}^{\infty} \sum_{\ell_3=-\infty}^{\infty} \sum_{\ell_4=-\infty}^{\infty} [(\Psi_{\ell_1} Z_{t-\ell_1})(\Psi_{\ell_2} Z_{t+p-\ell_2})' \otimes (\Psi_{\ell_3} Z_{t+h-\ell_3})(\Psi_{\ell_4} Z_{t+h+q-\ell_4})']. \end{aligned}$$

By the symmetry condition (2.3.13), one encounters non-zero summands only when all four indices in four-fold sum agree, or if there are two pairs of indices that agree (see also Francq and Zakoian [5]). Proceeding as in the proof of Lemma 1 gives

$$\begin{aligned} & E[X_t X'_{t+p} \otimes X_{t+h} X'_{t+h+q}] \\ = & \sum_{\ell_1=-\infty}^{\infty} \sum_{\ell_2=-\infty}^{\infty} (\Psi_{\ell_1} \otimes \Psi_{\ell_2}) E[Z_{t-\ell_1} Z'_{t-\ell_1} \otimes Z_{t+p-\ell_2} Z'_{t+p-\ell_2}] (\Psi_{\ell_1+p} \otimes \Psi_{\ell_2+q})' \\ + & \sum_{\ell_1=-\infty}^{\infty} \sum_{\ell_2=-\infty}^{\infty} (\Psi_{\ell_1} \otimes \Psi_{\ell_1+h}) E[Z_{t-\ell_1} Z'_{t+p-\ell_2} \otimes Z_{t-\ell_1} Z'_{t+p-\ell_2}] (\Psi_{\ell_2} \otimes \Psi_{\ell_2+h-p+q})' \\ + & \sum_{\ell_1=-\infty}^{\infty} \sum_{\ell_2=-\infty}^{\infty} (\Psi_{\ell_1} \otimes \Psi_{\ell_2+h-p}) E[Z_{t-\ell_1} Z'_{t+p-\ell_2} \otimes Z_{t+p-\ell_2} Z'_{t-\ell_1}] (\Psi_{\ell_2} \otimes \Psi_{\ell_1+h+q})' \\ - & 2 \sum_{\ell=-\infty}^{\infty} (\Psi_{\ell} \otimes \Psi_{\ell+h}) \eta (\Psi_{\ell+p} \otimes \Psi_{\ell+h+q})'. \end{aligned}$$

Applying (2.3.14) now gives

$$
\begin{aligned}
& E[X_t X'_{t+p} \otimes X_{t+h} X'_{t+h+q}] \\
= & \sum_{\ell_1=-\infty}^{\infty} \sum_{\ell_2=-\infty}^{\infty} (\Psi_{\ell_1} \otimes \Psi_{\ell_2})[\Gamma_{Z_t Z'_t}(\ell_2 - \ell_1 - p) + \Sigma \otimes \Sigma](\Psi_{\ell_1+p} \otimes \Psi_{\ell_2+q})' \\
+ & \sum_{\ell_1=-\infty}^{\infty} \sum_{\ell_2=-\infty}^{\infty} (\Psi_{\ell_1} \otimes \Psi_{\ell_1+h})[\Gamma_{Z_t \otimes Z_t}(\ell_2 - \ell_1 - p) + \mathrm{vec}(\Sigma)\mathrm{vec}(\Sigma)'](\Psi_{\ell_2} \otimes \Psi_{\ell_2+h-p+q})' \\
+ & \sum_{\ell_1=-\infty}^{\infty} \sum_{\ell_2=-\infty}^{\infty} (\Psi_{\ell_1} \otimes \Psi_{\ell_2+h-p}) P[\Gamma_{Z_t Z'_t}(\ell_2 - \ell_1 - p) + \Sigma \otimes \Sigma](\Psi_{\ell_2} \otimes \Psi_{\ell_1+h+q})' \\
- & 2 \sum_{\ell=-\infty}^{\infty} (\Psi_\ell \otimes \Psi_{\ell+h}) \eta (\Psi_{\ell+p} \otimes \Psi_{\ell+h+q})'.
\end{aligned}
$$

The argument finishing the proof of Theorem 2.1 provides

$$
W_{p,q} = \lim_{n\to\infty} n\mathrm{Cov}(\hat{\Gamma}(p), \hat{\Gamma}(q)) = \sum_{h=-\infty}^{\infty} \mathrm{Cov}(X_t X'_{t+p}, X_{t+h} X'_{t+h+q}).
$$

Therefore,

$$
\begin{aligned}
W_{p,q} = & \sum_{\ell_1=-\infty}^{\infty} \sum_{\ell_2=-\infty}^{\infty} (\Psi_{\ell_1} \otimes \Psi_{\ell_2}) \sum_{h=-\infty}^{\infty} \Gamma_{Z_t Z'_t}(h) (\Psi_{\ell_1+p} \otimes \Psi_{\ell_2+q})' \\
+ & \sum_{h=-\infty}^{\infty} \sum_{\ell_1=-\infty}^{\infty} \sum_{\ell_2=-\infty}^{\infty} (\Psi_{\ell_1} \otimes \Psi_{\ell_1+h})[\Gamma_{Z_t \otimes Z_t}(\ell_2 - \ell_1 - p) + \mathrm{vec}(\Sigma)\mathrm{vec}(\Sigma)'](\Psi_{\ell_2} \otimes \Psi_{\ell_2+h-p+q})' \\
+ & \sum_{h=-\infty}^{\infty} \sum_{\ell_1=-\infty}^{\infty} \sum_{\ell_2=-\infty}^{\infty} (\Psi_{\ell_1} \otimes \Psi_{\ell_2+h-p}) P[\Gamma_{Z_t Z'_t}(\ell_2 - \ell_1 - p) + \Sigma \otimes \Sigma](\Psi_{\ell_2} \otimes \Psi_{\ell_1+h+q})' \\
- & 2 \sum_{h=-\infty}^{\infty} \sum_{\ell=-\infty}^{\infty} (\Psi_\ell \otimes \Psi_{\ell+h}) \eta (\Psi_{\ell+p} \otimes \Psi_{\ell+h+q})'.
\end{aligned}
$$

Combining the first and last two terms in the above equation, and at the same time separating out the $\mathrm{vec}(\Sigma)\mathrm{vec}(\Sigma)'$ and $\Sigma \otimes \Sigma$ terms, it follows that

$$
\begin{aligned}
W_{p,q} &= \sum_{\ell_1=-\infty}^{\infty}\sum_{\ell_2=-\infty}^{\infty}(\Psi_{\ell_1}\otimes\Psi_{\ell_2})(\kappa-2\eta)(\Psi_{\ell_1+p}\otimes\Psi_{\ell_2+q})' \\
&+ \sum_{h=-\infty}^{\infty}\sum_{\ell_1=-\infty}^{\infty}\sum_{\ell_2=-\infty}^{\infty}(\Psi_{\ell_1}\otimes\Psi_{\ell_1+h})\Gamma_{Z_t\otimes Z_t}(\ell_2-\ell_1-p)(\Psi_{\ell_2}\otimes\Psi_{\ell_2+h-p+q})' \\
&+ \sum_{h=-\infty}^{\infty}\sum_{\ell_1=-\infty}^{\infty}\sum_{\ell_2=-\infty}^{\infty}(\Psi_{\ell_1}\otimes\Psi_{\ell_2+h-p})P\Gamma_{Z_tZ_t'}(\ell_2-\ell_1-p)(\Psi_{\ell_2}\otimes\Psi_{\ell_1+h+q})' \\
&+ \sum_{h=-\infty}^{\infty}\mathrm{vec}(\Gamma(h))\mathrm{vec}(\Gamma(h-p+q)')+P\Gamma(h-p)\otimes\Gamma(h+q).
\end{aligned}
$$

Now use $\kappa=\eta-\Sigma\otimes\Sigma+\kappa^*$, $\Gamma_{Z_t\otimes Z_t}(0)=\eta-\mathrm{vec}(\Sigma)\mathrm{vec}(\Sigma)'$, $\Gamma_{Z_tZ_t'}(0)=\eta-\Sigma\otimes\Sigma$, and separate out the lag zero summands to get

$$
\begin{aligned}
W_{p,q} &= \sum_{\ell_1=-\infty}^{\infty}\sum_{\ell_2=-\infty}^{\infty}(\Psi_{\ell_1}\otimes\Psi_{\ell_2})(\kappa^*+\Gamma_{Z_tZ_t'}(0)-2\eta)(\Psi_{\ell_1+p}\otimes\Psi_{\ell_2+q})' \\
&+ \sum_{h=-\infty}^{\infty}\mathrm{vec}(\Gamma(h))\mathrm{vec}(\Gamma(h-p+q)')+P\Gamma(h-p)\otimes\Gamma(h+q) \\
&+ \sum_{h=-\infty}^{\infty}\sum_{\ell_1=-\infty}^{\infty}\sum_{\ell_2=-\infty}^{\infty}(\Psi_{\ell_1}\otimes\Psi_{\ell_1+h})\Gamma_{Z_t\otimes Z_t}(\ell_2-\ell_1-p)(\Psi_{\ell_2}\otimes\Psi_{\ell_2+h-p+q})' \\
&+ \sum_{h=-\infty}^{\infty}\sum_{\ell_1=-\infty}^{\infty}\sum_{\ell_2=-\infty}^{\infty}(\Psi_{\ell_1}\otimes\Psi_{\ell_2+h-p})P\Gamma_{Z_tZ_t'}(\ell_2-\ell_1-p)(\Psi_{\ell_2}\otimes\Psi_{\ell_1+h+q})' \\
&=: V_{p,q}+V^*_{p,q},
\end{aligned}
$$

where $V_{p,q}$ shares the form as (2.3.10) and

$$
\begin{aligned}
V^*_{p,q} &= \sum_{\ell_1=-\infty}^{\infty}\sum_{\ell_2=-\infty}^{\infty}(\Psi_{\ell_1}\otimes\Psi_{\ell_2})\kappa^*(\Psi_{\ell_1+p}\otimes\Psi_{\ell_2+q})' \\
&+ \sum_{T_p}(\Psi_{\ell_1}\otimes\Psi_{\ell_1+h})\Gamma_{Z_t\otimes Z_t}(\ell_2-\ell_1-p)(\Psi_{\ell_2}\otimes\Psi_{\ell_2+h-p+q})' \\
&+ \sum_{T_p}(\Psi_{\ell_1}\otimes\Psi_{\ell_2+h-p})P\Gamma_{Z_tZ_t'}(\ell_2-\ell_1-p)(\Psi_{\ell_2}\otimes\Psi_{\ell_1+h+q})'. \qquad (2.6.36)
\end{aligned}
$$

Arguing as in the proof of Theorem 2.1 gives

$$\begin{aligned}
\mathrm{vec}(V^*_{p,q}) &= (\widetilde{\Gamma}(p) \otimes \widetilde{\Gamma}(q))\mathrm{vec}(\kappa^*) \\
&+ \sum_{T_p} (\widetilde{\Gamma}(\ell_2 - \ell_1) \otimes \widetilde{\Gamma}(\ell_2 - \ell_1 - p + q))\mathrm{vec}(\Gamma_{Z_t \otimes Z_t}(\ell_2 - \ell_1 - p)) \\
&+ \sum_{T_p} (\widetilde{\Gamma}(\ell_2 - \ell_1) \otimes \widetilde{\Gamma}(\ell_1 - \ell_2 + p + q))\mathrm{vec}(P\Gamma_{Z_t Z_t'}(\ell_2 - \ell_1 - p)),
\end{aligned}$$

where the index set is $T_p = \{(h, \ell_1, \ell_2) : -\infty < h, \ell_1, \ell_2 < \infty, \ell_2 - \ell_1 - p \neq 0\}$. Setting $\ell = \ell_2 - \ell_1$, we obtain

$$\begin{aligned}
\mathrm{vec}(V^*_{p,q}) &= (\widetilde{\Gamma}(p) \otimes \widetilde{\Gamma}(q))\mathrm{vec}(\kappa^*) \\
&+ \sum_{\ell \neq p} (\widetilde{\Gamma}(\ell) \otimes \widetilde{\Gamma}(\ell - p + q))\mathrm{vec}(\Gamma_{Z_t \otimes Z_t}(\ell - p)) \\
&+ \sum_{\ell \neq p} (\widetilde{\Gamma}(\ell) \otimes \widetilde{\Gamma}(-\ell + p + q))\mathrm{vec}(P\Gamma_{Z_t Z_t'}(\ell - p)).
\end{aligned}$$

2.7 Concluding Remarks

Some issues with the above work are enumerated here.

First, it would be desirable to have a central limit theorem for general d-variate processes satisfying (2.1.1), where $\{Z_t\}$ obeys an “easily checkable” set of mixing conditions. It is not enough for Z_t to have polynomial moments of all orders — counterexamples exist where $\{Z_t\}$ is white noise whose polynomial moments are all finite but where $\hat{\Gamma}(h)$ is not asymptotically normal. The most relevant result in the literature guaranteeing asymptotic normality of sample autocovariance appears to be Theorem 2.1 of Chanda [4].

Second, it is not clear to us whether (2.4.17) can be further simplified, nor is it clear that the asymptotic covariance depends only on autocorrelations and not on autocovariances as in the univariate case.

Finally, we have not pursued a multivariate GARCH example (of course, even univariate GARCH computations are difficult). The logical choice here would work in a multivariate ARCH(1) setting.

Chapter 3

A Comparison of Multivariate Signal Discrimination Techniques

We compare several techniques to discriminate two multivariate stationary signals. The compared methods include Gaussian likelihood ratio variance/covariance matrix tests — perhaps best viewed as principal component analyses (PCA) without dimension reduction aspects — and spectral-based tests gauging equality of the autocovariance function (over all lags) of the two signals. We show how one can make inappropriate conclusions with PCA tests, even when dimension augmentation techniques are used to incorporate non-zero lag autocovariances into the analysis. The various discrimination methods are first discussed. A simulation study is then presented that illuminates the various properties of the methods. An analysis of experimentally collected gearbox data is also presented.

3.1 Introduction

Given two d-dimensional series $\{\mathbf{X}_t\}$ and $\{\mathbf{Y}_t\}$ that are preprocessed to a zero-mean stationary setting, we considers how to assess whether (or not) the two signals have the same time series dynamics. This is useful in discrimination and classification pursuits. For example, if a test signal $\{\mathbf{Y}_t\}$ is deemed to have different dynamics than a reference signal $\{\mathbf{X}_t\}$ that is known to be "healthy", the test signal could be deemed unhealthy. Signal discrimination problems are fundamental ([15] [16]) and are well-developed when discriminating series via means or first moments; here, Hotelling T^2 or Q statistics are frequently relied upon ([17], [18]). In 1986, [19] considered discrimination of two univariate constant-mean series based on their sample autocovariances. Speech signals, for example, are typically of constant mean, regardless of what words are being spoken. Here, word-to-word changes are best identified through autocovariances shifts and monitoring of the mean is insufficient to identify dynamic changes. [20] seeks to discriminate an earthquake from a covert underground nuclear test; again, the crux issue lies with constant-mean data.

The classical way of discriminating $\{\mathbf{X}_t\}$ and $\{\mathbf{Y}_t\}$ through second order characteristics is via a Gaussian likelihood ratio. Such a test compares the sample variance matrix of the two series. Elaborating, conclusions are based on how different the two sample variance matrices

$$N^{-1}\sum_{t=1}^{N}\mathbf{X}_t\mathbf{X}_t', \quad N^{-1}\sum_{t=1}^{N}\mathbf{Y}_t\mathbf{Y}_t'$$

are from each other. Section 3.2 shows how to do this. Here, N is the sample length of the two series, which are assumed equal for convenience. When the dimension d is large, this comparison is typically made after a dimension reduction transformation, usually some type of principal component analysis (PCA), is done. Without dimension reduction aspects, covariance comparisons are not truly PCA techniques; however, they share the commonality in that conclusions are made only from sample autocovariances.

Basing signal equality conclusions exclusively on sample variances can produce erroneous conclusions when the two series are not multivariate white noise. A more comprehensive test would compare the sample autocovariances

$$\hat{\mathbf{\Gamma}}_X(h) = N^{-1}\sum_{t=1}^{N-h}\mathbf{X}_{t+h}\mathbf{X}_t'$$

and

$$\hat{\mathbf{\Gamma}}_Y(h) = N^{-1} \sum_{t=1}^{N-h} \mathbf{Y}_{t+h}\mathbf{Y}_t'$$

over all suitable lags $h \geq 0$. Such tests for multivariate series were discussed in [20], [21], [22], and the references within.

PCA methods have been extended to handle cases where correlation at non-zero series lags is present. This is typically done through a dimension augmentation scheme. For example, if $\mathbf{\Gamma}_X(1)$ and/or $\mathbf{\Gamma}_Y(1)$ are believed to be non-zero, one could compare the sample covariance matrices of the $2d$-dimensional vectors $\{\mathbf{X}_t^*\}$ and $\{\mathbf{Y}_t^*\}$, where

$$\mathbf{X}_t^* = (X_{2t-1,1}, \ldots, X_{2t-1,d}, X_{2t,1}, \ldots, X_{2t,d})'$$

and

$$\mathbf{Y}_t^* = (Y_{2t-1,1}, \ldots, Y_{2t-1,d}, Y_{2t,1}, \ldots, Y_{2t,d})'.$$

If the sample variance of $\{\mathbf{X}_t^*\}$ and $\{\mathbf{Y}_t^*\}$ agree, then one concludes that $\mathbf{\Gamma}_X(0) = \mathbf{\Gamma}_Y(0)$ and $\mathbf{\Gamma}_X(1) = \mathbf{\Gamma}_Y(1)$. Higher order comparisons are constructed from analogous reasoning. Of course, such dimension augmentation tactics shorten the observed series length; also, there is no clear maximum lag to augment by when autocovariances at all lags are non-zero, the typical case in practice.

Bassily [21] and Lund [22] attack the problem with different techniques. Specifically, two multivariate covariance functions are equal if and only if their spectral densities are equal at all frequencies (the spectrum is assumed to have no point masses). From this, signal equality tests that compare the periodograms of both series were devised (Section 3.2 elaborates). This paper rehashes these methods and shows how one can fool variance-based tests for signal equality, even when the dimension is augmented to account for non-zero autocovariances at higher lags. The pros and cons of the various methods are demonstrated by simulating multivariate stationary signals with various properties and then applying the tests. An application to a series of gearbox vibrations is included.

3.2 Background

We work with two zero-mean d-dimensional covariance stationary signals $\{\mathbf{X}_t\}$ and $\{\mathbf{Y}_t\}$ observed at times $t = 1, \ldots, N$. The covariance matrices at lag $h \geq 0$ are

$$\mathbf{\Gamma}_X(h) = E[\mathbf{X}_{t+h}\mathbf{X}_t'], \quad \mathbf{\Gamma}_Y(h) = E[\mathbf{Y}_{t+h}\mathbf{Y}_t'].$$

3.2.1 Testing Equality of Variances

The classical test for signal equality of zero-mean stationary series merely compares the sample variance matrices of the two observed series. The null hypothesis is that $\mathbf{\Gamma}_X(0) = \mathbf{\Gamma}_Y(0)$. A Gaussian likelihood ratio statistic for testing this hypothesis is

$$\lambda = \left(2^d \frac{\det\left[\hat{\mathbf{\Gamma}}_X(0)\hat{\mathbf{\Gamma}}_Y(0)\right]^{1/2}}{\det\left[\hat{\mathbf{\Gamma}}_X(0) + \hat{\mathbf{\Gamma}}_Y(0)\right]} \right)^N, \tag{3.2.1}$$

where det indicates matrix determinant. This statistic is derived in [24], pg. 404. Values of λ are in $[0, 1]$ and the null hypothesis is rejected when λ is too small to be explained by random chance. Authors have used this test when the series are non-Gaussian white noise without drastic performance degradations. Here, the usual central limit caveat applies: the test works well for large N provided marginal distributions of the series are not heavy-tailed. Applying the test when the data are autocorrelated (i.e, not white noise) is more problematic. This aspect will be demonstrated in the next Section.

In great generality, $-2\ln(\lambda)$ has an asymptotic (as $N \to \infty$) chi-squared distribution ([25], [26]). The degrees of freedom is equal to the number of parameters that are saved when the two signals have the same covariance matrix. Since covariance matrices of a d-dimensional signal are $d \times d$ symmetric matrices, $d(d+1)/2$ free parameters are saved; that is, $d(d+1)/2$ is the appropriate degrees of freedom. Phrased another way, λ asymptotically behaves as $e^{-L/2}$, where L is a chi-squared random variate with $d(d+1)/2$ degrees of freedom. From this, it follows that λ has the asymptotic density

Table 3.1: Ninety-Fifth Percentiles for λ

Dimension	Ninety-Fifth Percentile
1	0.1478
2	0.02025
3	0.001839
4	1.035e-4
5	3.580e-6

$$f_\lambda(x) = \frac{[-\ln(x)]^{d(d+1)/4-1}}{\Gamma\left(\frac{d(d+1)}{4}\right)}, \qquad 0 \le x \le 1.$$

Here, $\Gamma(\alpha)$ represents the usual Gamma function at argument $\alpha > 0$ (the use of Γ as both a covariance and a function should cause no confusion). This density can be used to extract percentiles; however, exact formulas cannot be given since the antiderivative of $\ln(x)^\beta$ for $\beta > 0$ has no explicit formula. Table I lists how small λ must be to warrant rejection of equal variances with 95% statistical confidence for several valued of d. A plot of the asymptotic density of λ for $d = 2$ is shown in Fig. 3.1.

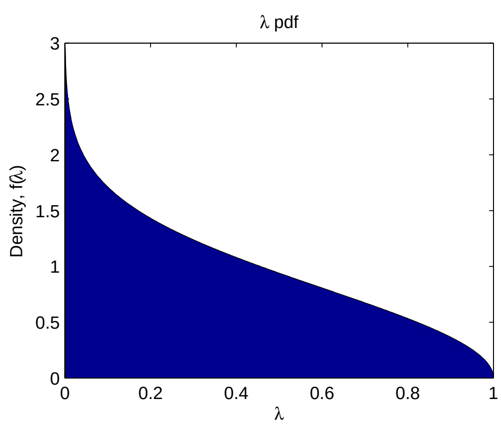

Figure 3.1: Plot of the PDF for λ with d=2

3.2.2 Testing Equality of the Autocovariance Functions

A spectral approach to testing equality of multivariate autocovariance functions was developed in [21]. Since $\mathbf{\Gamma}_X(h) = \mathbf{\Gamma}_Y(h)$ for all lags $h \geq 0$ if and only if $\mathbf{f}_X(\lambda) = \mathbf{f}_Y(\lambda)$ for all frequencies $\lambda \in [0, 2\pi)$ (with respect to the Lebesgue measure), where

$$\mathbf{f}_X(\lambda) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \mathbf{\Gamma}_X(h) e^{-i\lambda h}$$

and

$$\mathbf{f}_Y(\lambda) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \mathbf{\Gamma}_Y(h) e^{-i\lambda h}$$

are the theoretical spectral densities of $\{\mathbf{X}_t\}$ and $\{\mathbf{Y}_t\}$ at frequency λ, respectively.

Bassily [21] estimates the spectral densities of the two series and statistically compares their ratios. Specifically, the discrete Fourier transforms (DFTs) of the series are first computed via

$$\mathbf{J}_X(\lambda_j) = N^{-1/2} \sum_{t=1}^{N} \mathbf{X}_t e^{-it\lambda_j}$$

and

$$\mathbf{J}_Y(\lambda_j) = N^{-1/2} \sum_{t=1}^{N} \mathbf{Y}_t e^{-it\lambda_j}$$

at all Fourier frequencies $\lambda_j = 2\pi j/N$, for $j = 0, \ldots, N-1$ (see [27] and [28] for Fourier transform basics). The raw (unsmoothed) spectral densities are estimated via

$$\hat{\mathbf{f}}_X(\lambda_j) = \frac{\mathbf{J}_X(\lambda_j)\mathbf{J}_X^*(\lambda_j)}{2\pi}, \qquad \hat{\mathbf{f}}_Y(\lambda_j) = \frac{\mathbf{J}_Y(\lambda_j)\mathbf{J}_Y^*(\lambda_j)}{2\pi}.$$

Here, the asterisk denotes complex conjugation. The raw spectral estimates are then smoothed in a uniform manner over $2M+1$ Fourier frequencies closest to the Fourier frequency being considered:

$$\hat{\mathbf{f}}_X^s(\lambda_j) = \frac{\sum_{k=-M}^{M} \hat{\mathbf{f}}_X(\lambda_{j+k})}{2M+1}, \qquad \hat{\mathbf{f}}_Y^s(\lambda_j) = \frac{\sum_{k=-M}^{M} \hat{\mathbf{f}}_Y(\lambda_{j+k})}{2M+1}.$$

Here, M is a positive integer, representing a smoothing bandwidth, that satisfies $2M+1 \geq d$ (this is needed for technical reasons rooted in the finiteness of variances). The choice of M does not usually influence practical conclusions about signal equality. In smoothing the raw spectral estimates, frequencies outside of $[0, 2\pi)$ are rounded modulo 2π to mimic the periodic nature of the DFT; for example, $\hat{\mathbf{f}}_X(\lambda_{j+N}) = \hat{\mathbf{f}}_X(\lambda_j)$.

Bassily [21] base signal equality conclusions on the statistic

$$\bar{\Delta} = \frac{1}{\frac{N}{2}-1} \sum_{j=1}^{\frac{N}{2}-1} |\Delta(\lambda_j)|. \tag{3.2.2}$$

Here, the $\Delta(\lambda_j)$'s are the log determinant of the ratios of the smoothed spectral density estimates:

$$\Delta(\lambda_j) = \log\left(\det\left(\hat{\mathbf{f}}_X^s(\lambda_j)\right)\right) - \log\left(\det\left(\hat{\mathbf{f}}_Y^s(\lambda_j)\right)\right). \tag{3.2.3}$$

Under the null hypothesis of equal autocovariance functions, $\Delta(\omega_j)$ should be statistically close to zero for every non-zero Fourier frequency λ_j. Bassily [21] show that $\Delta(\lambda_j)$ has an asymptotic distribution that does not depend on j for $j = 1, 2, \ldots, N/2-1$ or the common spectral density of $\{\mathbf{X}_t\}$ and $\{\mathbf{Y}_t\}$. From this structure, a test for signal equality based on $\overline{\Delta}$ is easily constructed based on the central limit theorem (the $\Delta(\lambda_j)$'s for varying j are approximately independent). Such a test rejects equality of autocovariance functions when

$$\bar{\Delta} > \mu_M + z_\alpha \frac{\sigma_M}{\sqrt{\frac{N}{2}-1}}. \tag{3.2.4}$$

Here, z_α denotes a quantile that cuts off an upper tail area of α in the standard normal distribution ($z_\alpha = 1.645$ when $\alpha = 0.05$) and μ_M and σ_M are the theoretical mean and variance of $|\Delta(\lambda_j)|$. Note that this is a one sided test.

The constants μ_M and σ_M^2 depend on both M and d and are difficult to derive. Lund [22] derives explicit expressions when $d = 1$, but the computations for the multidimensional case are intense. However, simulations with Gaussian white noise readily provide good estimates of them. These estimates are given in tables in [21].

The detection power of the $\bar{\Delta}$ statistic can be increased if the signals are known to be band-limited. Specifically, in the spectrums of $\{\mathbf{X}_t\}$ and $\{\mathbf{Y}_t\}$ are known to limited to the interval $[\lambda_L, \lambda_U]$, then (3.2.2) is modified to

$$\bar{\Delta} = C^{-1} \sum_{\{j:\lambda_j \in [\lambda_L, \lambda_U]\}} |\Delta(\lambda_j)|,$$

where C is the number of distinct Fourier frequencies in the interval $[\lambda_L, \lambda_U]$. The rejection region is the same as in (3.2.4), except that $N/2 - 1$ is replaced by C. One should take C large enough to induce asymptotic normality of averages (a typical rule of thumb is to have $C \geq 30$.) Detection power increases because many frequencies where no differences occur are excluded in the analysis, accentuating the importance of differences in the considered frequency increments.

3.3 Method Comparison

This section studies the properties of the λ and $\bar{\Delta}$ statistics through specifically designed simulations to illustrate various points. In all cases, the issues are apparent in dimension $d = 2$ and at 95% statistical confidence. The smoothing parameter $M = 5$ and series length $N = 1024$ are common to all runs. In all cases, one hundred thousand simulations were conducted.

First, the λ and $\bar{\Delta}$ statistics were computed for each simulated realization of $\{\mathbf{X}_t\}$ and $\{\mathbf{Y}_t\}$, each realization containing zero-mean Gaussian white noise. In this case, the covariance matrix of $\mathbf{X}_t$ and $\mathbf{Y}_t$ was taken as the two-dimensional identity matrix. Hence, this case, which we refer to as Case I, is a scenario where the two signals have the same dynamics. Table 3.2 shows empirically aggregated proportions of runs where the λ and $\bar{\Delta}$

reject the null hypothesis of signal equality at level 95%. As both proportions are close to 5%, both methods have worked well in this case.

Our second case is one where $\{\mathbf{X}_t\}$ and $\{\mathbf{Y}_t\}$ do not have the same variance (lag-zero covariance matrix). Here, $\{\mathbf{X}_t\}$ and $\{\mathbf{Y}_t\}$ are zero-mean Gaussian white noise with the covariance matrices

$$\mathbf{\Gamma}_X(0) = \begin{bmatrix} 1.0 & 0.0 \\ 0.0 & 1.0 \end{bmatrix}, \qquad \mathbf{\Gamma}_Y(0) = \begin{bmatrix} 1.1 & 0.1 \\ 0.1 & 1.0 \end{bmatrix},$$

respectively. Table 3.2 displays the proportions of times the λ and $\bar{\Delta}$ tests reject signal equality at confidence 95%. In this case, the likelihood ratio statistic λ has worked best, drastically so, as seen by its larger empirical rejection proportion. This is not unexpected: while both methods should ideally reject signal equality, the two signals differ only in their variances; covariances at all higher lags are zero. While the λ statistic focuses solely on variance differences, the $\bar{\Delta}$ statistics must consider all covariance lags. This essentially degrades the detection power of the $\bar{\Delta}$ test in this case.

Case III considers a situation where $\{\mathbf{X}_t\}$ and $\{\mathbf{Y}_t\}$ have the same variances, but where there is non-zero autocorrelation at non-zero lags; that is, the series under consideration are not multivariate white noise. We do this by examining solutions to the vector autoregressive moving-average (VARMA) model of autoregressive order 2 and moving-average order 1. Specifically, both $\{\mathbf{X}_t\}$ and $\{\mathbf{Y}_t\}$ obey the VARMA difference equation

$$\mathbf{X}_t = \mathbf{\Phi}_1 \mathbf{X}_{t-1} + \mathbf{\Phi}_2 \mathbf{X}_{t-2} + \mathbf{Z}_t + \mathbf{\Theta}_1 \mathbf{Z}_{t-1}.$$

Here, the autoregressive matrix coefficients were chosen as

$$\mathbf{\Phi}_1 = \begin{bmatrix} 0.40 & 0.05 \\ 0.05 & 0.30 \end{bmatrix}, \quad \mathbf{\Phi}_2 = \begin{bmatrix} -0.48 & 0.10 \\ 0.10 & -0.06 \end{bmatrix},$$

and the moving-average coefficient matrix was selected as

Table 3.2: Method Detection and False Alarm Probabilities

	λ	$\bar{\Delta}$
Case I	5.14%	5.39%
Case II	57.69%	7.05%
Case III	19.04%	5.54%
Case IV	73.19%	5.62%
Case V	8.12%	100.00%
Case VI	100.00%	100.00%
Case VII	7.61%	15.64%

$$\mathbf{\Theta}_1 = \begin{bmatrix} 0.30 & 0.10 \\ 0.10 & 0.50 \end{bmatrix}.$$

Here, $\{\mathbf{Z}_t\}$ is chosen as white noise with an identity covariance matrix. The Case III performance characteristics reverse from Case II with the λ statistic erroneously rejecting signal equality about 19% of the time. Most statisticians view this false alarm rate as unacceptable in a 95% test. The $\bar{\Delta}$ statistic, however, rejects signal equality at approximately the intended 5% rate.

Case IV represents an exacerbated version of Case III. Here, the two series are taken as vector autoregressions of order one. Specifically, both series follow the VAR(1) dynamics

$$\mathbf{X}_t = \mathbf{\Phi}\mathbf{X}_{t-1} + \mathbf{Z}_t,$$

where $\{\mathbf{Z}_t\}$ is taken as Gaussian white noise with the identity covariance matrix and

$$\mathbf{\Phi} = \begin{bmatrix} 0.90 & 0.10 \\ -0.10 & 0.90 \end{bmatrix}.$$

The dynamics of this model lie near the boundary of the multivariate causality region of a VAR(1) model, as is seen by the near unit diagonal entries in $\mathbf{\Phi}$. In this case, the λ statistic erroneously rejects signal equality at a whopping 75% rate. The false alarm (Type I error) of the $\bar{\Delta}$ test is also getting a bit larger than the specified 5%, but not drastically so. Taken together, the last two cases show that likelihood ratio tests to detect variances changes perform suboptimally unless the signals are known to be white noise. At this point, one can also question the detection power of the $\bar{\Delta}$ statistic as it performed poorly in in the one case where the signals were truly different (Case II). The next three cases will perhaps remedy this concern.

Case V moves to a situation designed to fool the λ statistic. Specifically, our $\{\mathbf{X}_t\}$ is taken as the first-order moving-average satisfying

$$\mathbf{X}_t = \mathbf{Z}_t + \mathbf{\Theta}\mathbf{Z}_{t-1}.$$

and $\{\mathbf{Y}_t\}$ is taken as white noise

$$\mathbf{Y}_t = \eta_t.$$

The caveat here is that we select the parameters $\mathbf{\Theta}$, $\mathrm{Var}(\mathbf{Z}_t) = \mathbf{\Sigma}_Z$, and $\mathrm{Var}(\eta_t) = \mathbf{\Sigma}_\eta$ so that $\mathbf{\Gamma}_X(0) = \mathbf{\Gamma}_Y(0)$. To do this, we take

$$\mathbf{\Theta} = \begin{bmatrix} 0.70 & 0.30 \\ 0.30 & 0.50 \end{bmatrix}, \quad \mathbf{\Sigma}_Z = \begin{bmatrix} 1.00 & 0.00 \\ 0.00 & 1.00 \end{bmatrix},$$

and

$$\mathbf{\Sigma}_\eta = \begin{bmatrix} 1.58 & 0.36 \\ 0.36 & 1.34 \end{bmatrix},$$

In this case, the two series have different dynamics, but have the same lag-zero variance matrix. The empirical probabilities in Table 3.2 reflect this property: the λ statistic opts for equivalent signal dynamics only slightly more than the 5% nominal false alarm rate; however, the $\bar{\Delta}$ statistic makes the correct conclusion of signal inequality in all of the one hundred thousand runs.

Summarizing to this point, the λ test degrades under correlation but is more powerful at detecting variance changes when only variance changes are truly present.

One can reduce equality of autocovariance problems to variance comparisons through dimension augmentation techniques. For example, suppose that the signal's autocovariances are only non-zero at lags $0, 1, \ldots, \kappa$ and set

$$\mathbf{X}_n^* = (\mathbf{X}'_{(n-1)(\kappa+1)+1}, \ldots, \mathbf{X}'_{n(\kappa+1)})'.$$

Then $\{\mathbf{X}_t\}$ and $\{\mathbf{Y}_t\}$ have the same autocovariances at lags $h = 0, \ldots, \kappa$ if $\{\mathbf{X}_n^*\}$ and $\{\mathbf{Y}_n^*\}$ have equal variances. For example, in Case V, $\mathbf{X}_t^* = (X_{2t-1,1}, X_{2t-1,2}, X_{2t,1}, X_{2t,2})'$ and $\mathbf{Y}_t^* = (Y_{2t-1,1}, Y_{2t-1,2}, Y_{2t,1}, Y_{2t,2})'$. Of course, such tactics may not represent an efficient way of proceeding when κ is large as series sample sizes are reduced.

Case VI shows empirical probabilities of signal equality rejection when 4-dimensional vectors are made to analyze the signals generated in Case IV. We will not rerun the $\bar{\Delta}$ analyses, preferring to emphasize that the $\bar{\Delta}$ method naturally handles autocorrelation and that there is no need to do any sort of dimension augmentation. The rejection probability of the λ statistic in Case V increases to 100% when the dimension is augmented to four dimensions. Since moving averages are completely characterized by their lag-zero and lag-one autocovariances, dimension augmentation works very well here.

Selection of the dimension to augment by is problematic. If one selects the augmentation dimension too small, higher order covariances will not be considered (which is suboptimal if these autocovariances are non-zero). On the other hand, if the selected dimension is too large, then the sample size becomes significantly smaller and discrimination power is lost.

Our last case is intended to show that there are no easy ways of selecting augmentation dimensions. We do this by constructing two series where the signals have different dynamics, but where both the lag-zero and lag-one autocovariances agree. That is, we want $\{\mathbf{X}_t\}$ and $\{\mathbf{Y}_t\}$ to have different dynamics, but $\mathbf{\Gamma}_X(0) = \mathbf{\Gamma}_Y(0)$ and $\mathbf{\Gamma}_X(1) = \mathbf{\Gamma}_Y(1)$. Case VII shows signal equality rejection probabilities in such a case. This was done by mixing two univariate signals with equal lag-zero and lag-one autocovariances. Specifically, suppose that $\{X_{t,1}\}$, and $\{X_{t,2}\}$, the components of $\{\mathbf{X}_t\}$, both follow the AR(1) dynamics

$$X_{t,1} = \phi X_{t-1,1} + Z_{t,1}, \qquad X_{t,2} = \phi X_{t-1,2} + Z_{t,2},$$

where $\{Z_{t,1}\}$ and $\{Z_{t,2}\}$ are independent zero-mean unit-variance Gaussian white noise series. Hence, the two components of $\{\mathbf{X}_t\}$ are independent AR(1) series having the same univariate covariances at all lags. Now suppose that both components of $\{\mathbf{Y}_t\}$ obey MA(1) dynamics:

$$Y_{t,1} = \eta_{t,1} + \theta_1 \eta_{t-1,1}, \qquad Y_{t,2} = \eta_{t,2} + \theta_2 \eta_{t-1,2},$$

where $\{\eta_{t,1}\}$ and $\{\eta_{t,2}\}$ are independent zero-mean variance σ_η^2 Gaussian white noise series. A simple computation shows that $\{X_{t,1}\}$ and $\{Y_{t,1}\}$ have the same lag-zero and lag-one autocovariances when

$$\phi = \frac{\theta}{1+\theta^2}, \qquad \sigma_\eta^2 = \frac{1+\theta^2}{1+\theta^2+\theta^4}.$$

To mix the two components (so that $\{X_{t,1}\}$ and $\{X_{t,2}\}$ are not independent), set

$$\mathbf{X}_t = \mathbf{L}\begin{pmatrix} X_{t,1} \\ X_{t,2} \end{pmatrix}, \qquad \mathbf{Y}_t = \mathbf{L}\begin{pmatrix} Y_{t,1} \\ Y_{t,2} \end{pmatrix},$$

where

$$\mathbf{L} = \begin{bmatrix} 1/2 & 1/3 \\ -1/3 & 1/2 \end{bmatrix}.$$

Then $\{\mathbf{X}_t\}$ and $\{\mathbf{Y}_t\}$ have different signal dynamics, yet, by construction, $\mathbf{\Gamma}_X(0) = \mathbf{\Gamma}_Y(0)$ and $\mathbf{\Gamma}_X(1) = \mathbf{\Gamma}_Y(1)$.

The Case VII probabilites use $\phi = 1/4$. The values $\theta_1 = 2 + \sqrt{3}$, $\theta_2 = 2 - \sqrt{3}$, and $\sigma_\eta^2 = 0.9952$ were then chosen to satisfy the above constraints. The Table 3.2 rejection proportions show that while the $\bar{\Delta}$ statistic does not detect signal inequality well, the λ statistic is almost completely fooled. Because of this, we do not consider comparing signals whose autocovariances match to a higher number of lags as the pattern is clear: the λ statistic will have more difficulty correctly discriminating such signals.

Overall, the $\bar{\Delta}$ tests seems to perform well without the need for dimension augmentation. Performance of the classical λ test can degrade should autocorrelations in the series be present (i.e., this test perform well for white noise discrimination only). We suggest that the $\bar{\Delta}$ statistic be considered should conclusions on signal equality have importance.

3.4 Gearbox Analysis

To demonstrate discrimination capabilities on actual data, the λ and $\bar{\Delta}$ statistics will be computed for three experimentally collected gearbox vibration signals of dimension $d = 2$. Our goal here lies with fault diagnosis. In fault diagnosis schemes, a known healthy signal is compared to a test signal, which may be healthy or unhealthy. An unhealthy signal is indicative of faults. Such an approach has been used to diagnose faults in wind turbine gearboxes ([29] [30]), gas turbines ([31] [33]), electric motors ([34] [35] [36] [37]), and general rotating components ([38]). See [39] [40] [41] [42] [43] [44] [45] [46] [47] and [48] for other fault detection research.

The vibration data used here comes from The Prognostics and Health Management Society (PHM Society) as part of their *2009 PHM Challenge Competition Data Set*. Similar data sets are found at NASA's Prognostics Center of Excellence's prognostic-data-repository (`http://ti.arc.nasa.gov/tech/dash/pcoe/prognostic-data-repository/`). The data were collected from a generic, three-axis gearbox with accelerometers mounted on the input side and output side (see fig. 3.2). The input pinion had 32 teeth, the input-side idler gear 96 teeth, the output-side idler gear 48 teeth, and the output gear 80 teeth, resulting in the 5 to 1 reduction ratio.

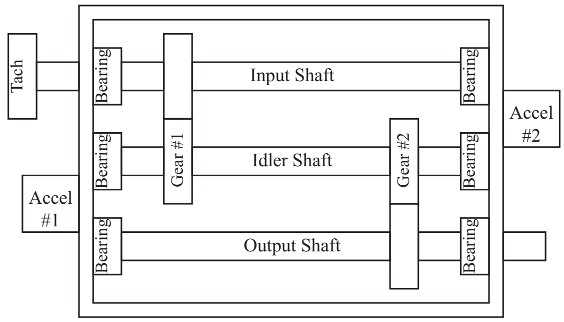

Figure 3.2: System diagram of the generic industrial gearbox used in the 2009 PHM Society competition showing the location of the accelerometers and the physical relation of the components.

The vibration data set, as a whole, contains over 560 two dimensional series. These series correspond to gearbox runs at 30, 35, 40, 45, and 50 Hz under high and low loadings, all repeated twice. This frequency sequence was run again for numerous fault cases, including chipped teeth, broken teeth, eccentric gears, bent shafts, imbalanced shafts, and inner and outer bearing defects. This battery of was repeated for helical and spur gears. The series were collected at 66.6kHz and are of length $N = 266000$.

Our investigation focuses on three series. Series A and B were collected from the gearbox when no faults are present (healthy data). Series C was collected after various faults have been introduced (faulty data). The faults present in series C include an eccentric gear, a gear with a broken tooth, and a bearing with a fouled ball.

Figure 3.3 plot segments of the component series. Notice that the data appear to have a constant mean (roughly) and were sampled at a very high frequency. In fact, the entire data length corresponds to only x.xx seconds of runtime. In truth, non-stationarity is likely present in these series. Plausibly, there are many deterministic sinusoids embedded in the series, a prominent one residing at 30Hz. We will combat local variance change aspects by making sliding subsegments of length 1024.

Smoothed periodograms of the components of the healthy series A and faulty series C are plotted in fig. 3.4. The smoothing uniformly weights eleven adjacent periodogram ordinates. The periodograms of the healthy and faulty data appear pretty similar. Observe

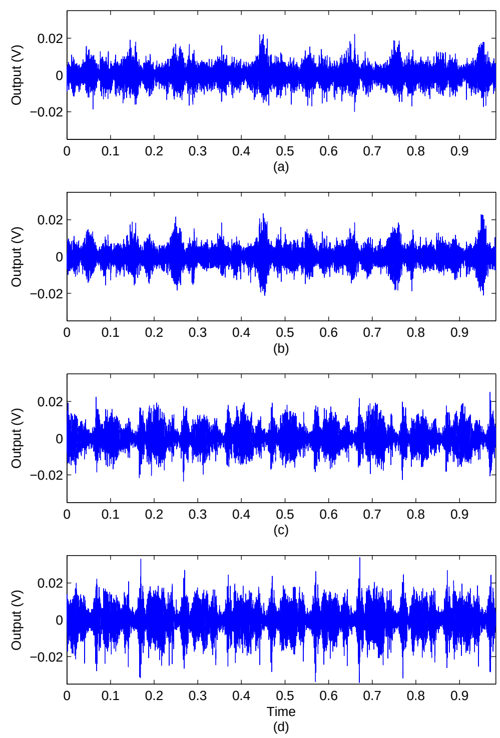

Figure 3.3: Sample of data to be analyzed. (a) Gear 1 Signal 1. (b) Gear 1 Signal 2. (c) Gear 2 Signal 1. (d) Gear 2 Signal 2.

that all significant spectral content is located below 12,000Hz and that the more significant spectral content is found below 5,000Hz. This is to be expected. The input shaft for the data being analyzed is rotating at 30Hz and with a gear reduction ratio of 5:1, the output shaft will be rotating at 6Hz. The spectral contributions from the rotating shafts and gears as well as the tooth interactions are expected to be at lower frequencies, particularly below 1,000Hz. Because of this, we band-limit all $\bar{\Delta}$ statistics to $[0, 1000]$Hz. That is not to say the higher frequencies are totally negligible. A broken tooth, for example, creates a short-duration disturbance once-per-gear revolution. This once per cycle, short-duration disturbance may be similar to a impulse-train type disturbance and may affect the system accordingly. Impact Technologies identified such behavior and exploited it in their ImpactEnergy detection algorithms ([49] [50] [51] [52] [53] [54]).

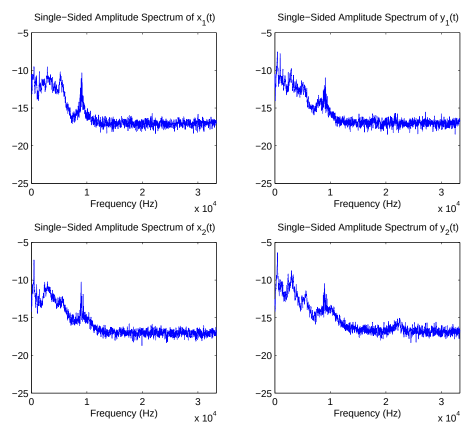

Figure 3.4: Periodograms of healthy and faulty signals. (top left) Healthy signal, input accelerometer. (bottom left) Healthy signal, output accelerometer. (top right) Faulty signal, input accelerometer. (bottom right) Faulty signal, output accelerometer. Notice the change in the periodograms from healthy signal to faulty signal.

To compare signals, each series will be segmented into non-overlapping segments of length 1024, resulting in roughly 250 subsegments. Each subsegment will be compared to the corresponding subsegment in the other series and referred to as a trial. Each trial calculates a λ and a band-limited $\bar{\Delta}$. Once all 250 comparisons are made, the percentage of trials that exceed the 95^{th} percentile for each corresponding statistic will be reported.

Table 3.3: Detection Powers

Test/Comparison	λ	$\bar{\Delta}$
Healthy (A)-Healthy (B)	96.4%	50.0%
Healthy (A)-Faulty (C)	100.0%	87.2%
Healthy (B)-Faulty (C)	96.4%	88.4%

Table 3.3 summarizes the outcomes. For the case where the comparison is between two like signals, the λ statistic declares them different 96.4% of the time (all conclusions are made at level 5%) while the $\bar{\Delta}$ statistic declares them different only 50.0% of the time. In truth, there are likely some subtle differences between the two healthy case runs. However, as there is significant non-zero correlation at many lags in this data, one believes the $\bar{\Delta}$ results to be more realistic.

When comparing signal A to signal C, the λ statistic declares them different 100.0% of the time while the $\bar{\Delta}$ statistic declares them different 87.2% of the time. When comparing signal B to signal C, the λ statistic declares them different 96.4 the $\bar{\Delta}$ statistic declares them different xx.x% of the time. Overall, it appears that both statistics capably identified that the signals were born of two different processes.

3.5 Conclusion

This paper compared two multivariate signal discrimination techniques under various scenarios. The likelihood ratio statistic λ rejects signal equality in a reliable manner only when the series considered are white noise. However, when the series are white noise, it has a larger detection probability than the $\bar{\Delta}$ statistic. In cases where autocovariances at higher lags are non-zero, the $\bar{\Delta}$ statistic is more reliable. In fact, a simple VAR(1) case was constructed where the false alarm rate of the λ statistic was approximately 15 times higher than advertised. Applications to an experimental set of gearbox vibrations showed similar structure. Overall, it is wise to base signal equality conclusions on the $\bar{\Delta}$ statistic when signals are not multivariate white noises.

Bibliography

[1] M.S. Bartlett, An Introduction to Stochastic Processes, Cambridge University Press, London, 1955.

[2] A. Berlinet, C. Francq, On Bartlett's formula for non-linear processes, Journal of Time Series Analysis 18 (1997) 535-552.

[3] P.J. Brockwell, R.A. Davis, Time Series: Theory and Methods, second ed., Springer-Verlag, New York City, 1991.

[4] K.C. Chanda, Asymptotic properties of serial covariances for nonlinear stationary processes, Journal of Multivariate Analysis 47 (1993) 163-171.

[5] C. Francq, Q. Zakoïan, Bartlett's formula for a general class of nonlinear processes, Journal of Time Series Analysis 30 (2009) 449-465.

[6] W.A. Fuller, Introduction to Statistical Time Series, Second Edition, Wiley, New York City, 1996.

[7] L. Giraitis, P. Kokoszka, R. Leipus, Stationary ARCH models: Dependence structure and central limit theorem, Econometric Theory 16 (2000) 3-22.

[8] E.J. Hannan, Multiple Time Series, John Wiley & Sons Inc., 1970.

[9] E.J. Hannan, The asymptotic distribution of serial autocovariances, Annals of Statistics 4 (1976) 396-399.

[10] R.B. Lund, H. Bassily, B. Vidakovic, Testing equality of autocovariance functions, Journal of Time Series Analysis 30 (2009) 332-348.

[11] H. Lütkepohl, New Introduction to Multiple Time Series Analysis, second ed., Springer, New York City, 2005.

[12] J.P. Romano, L.A. Thombs, Inference for autocorrelations under weak assumptions, Journal of the American Statistical Association 91 (1996) 590-600.

[13] G.C. Reinsel, Elements of Multivariate Time Series Analysis, Springer-Verlag, New York City, 1997.

[14] R.H. Shumway, D.S. Stoffer, Time Series Analysis and its Applications: with R Examples, second ed., Springer, New York City, 2006.

[15] C. E. Shannon and W. Weaver, *The Mathematical Theory of Communication*, The University of Illinois Press: Urbana, Champaign, Illinois, 1949.

[16] S. Kullback and R. A. Leibler, “On information and sufficiency,The Annals of Mathematical Statistics 22 (1951) 79-86.

[17] H. Hotelling, Multivariate quality control, in: C. Eisenhart, M. W. Hastay, and W. A. Wallis, (Eds.) Techniques of Statistical Analysis, 1st ed., McGraw-Hill Book Company, Inc., New York, New York, first edition, 1947.

[18] The generalization of student’s ratio, Ann. Math. Statist.2 (1931) 360-378.

[19] D. S. Coates and P. J. Diggle, Tests for comparing two estimated spectral densities,Journal of Time Series Analysis 7 (1986) 7-20.

[20] Y. Kakizawa, R. H. Shumway, and M. Taniguchi, Discrimination and clustering for multivariate time series, Journal of the American Statistical Association 93 (1998) 328-340.

[21] H. Bassily, R. Lund, and J. Wagner, Fault detection in multivariate signals with applications to gas turbines, IEEE Transactions on Signal Processing 57 (2009) 835-42.

[22] R. Lund, H. Bassily, and B. Vidakovic, Testing equality of stationary autocovariances, Journal of Time Series Analysis 30 (2009) 332-348.

[23] L. Kilian, BOOK REVIEWS - New Introduction to Multiple Time Series Analysis by Helmut Ltkepohl Springer, 2005, Econometric Theory 22 (2006) 961-967.

[24] T. W. Anderson, An Introduction to Multivariate Statistical Analysis, New York, Second edition, 1984.

[25] A. C. Davison, Statistical Models, Cambridge University Press, Cambridge, 2003.

[26] G. Casella and R. Berger, Statistical Inference, Duxbury Press, Pacific Grove, CA, 2nd edition, 2001.

[27] P. Bloomfield, Fourier Analysis of Time Series, Wiley, New York, 1976.

[28] G. M. Jenkins and D. G. Watts, Spectral Analysis and Its Applications, Holden-Day, San Francisco, CA, 1968.

[29] Y. Amirat, M. Benbouzid, B. Bensaker, and R. Wamkeue,condition monitoring and fault diagnosis in wind energy conversion systems: A review, in: Proceedings of IEEE International Electric Machines and Drives Conference, IEMDC 2007, volumn 2, Antalya, Turkey, 1434-1439.

[30] M. C. Garcia, M. A. Sanz-Bobi, and J. del Pico, SIMAP: Intelligent system for predictive maintenance. application to the health condition monitoring of a windturbine gearbox, Computers in Industry 57 (2006), 552-568.

[31] S. Gupta, A. Ray, S. Sarkar, and M. Yasar, Fault detection and isolation in aircraft gas turbine engines. part 1: underlying concept, Proceedings of the Institution of Mechanical Engineers, Part G: Journal of Aerospace Engineering 222 (2008), 307-318.

[32] A. Schuster, The periodicity of sun-spots, The Astrophysical Journal, 23(2) (1906), 101C109.

[33] R. Sekhon, H. Bassily, and J. Wagner, A comparison of two trending strategies for gas turbine performance prediction, Journal of Engineering for Gas Turbines and Power 130 (2008), 041601-1-041601-10.

[34] J. Stack, T. Habetler, and R. Harley, Bearing fault detection via autoregressive stator current modeling,IEEE Transactions on Industry Applications 40 (2004) 740-747.

[35] M. Benbouzid, Bibliography on induction motors faults detection and diagnosis, IEEE Transactions on Energy Conversion 14 (1999) 1065-1074.

[36] M. E. H. Benbouzid, A review of induction motors signature analysis as a medium for faults detection,Industrial Electronics, IEEE Transactions on 47 (2000) 984-993.

[37] P. Vecer, M. Kreidl, and R. Smid, Condition indicators for gearbox condition monitoring systems Acta Polytechnica, Czech Technical University in Prague 45 (2005) 35-43.

[38] M. Artioli, G. Capolino, F. Filippetti, and A. Yazidi, A general purpose software for distance monitoring and diagnosis of electrical machines (2003) 272-276.

[39] J. Gertler, Survey of model-based failure detection and isolation in complex plants, IEEE Control Systems Magazine 8 (1988) 3-11.

[40] J. J. Gertler, Fault Detection and Diagnosis in Engineering Systems, Marcel Dekker, Inc., New York, NY, 1998.

[41] R. Isermann, Fault-Diagnosis Systems: An Introduction from Fault Detection to Fault Tolerance Springer, Germany, 2006.

[42] R. Patton, P. Frank, and R. Clark, Eds., Fault Diagnosis in Dynamic Systems: Theory and Application, Prentice-Hall, Inc., Upper Saddle River, NJ, 1989.

[43] R. J. Patton, P. M. Frank, and R. N. Clark, Eds., Issues of Fault Diagnosis for Dynamic Systems, Springer, London, 2000.

[44] J. Dron, L. Rasolofondraibe, C. Couet, and A. Pavan, Fault detection and monitoring of a ball bearing benchtest and a production machine via autoregressive spectrum analysis, Journal of Sound and Vibration 218 (1998) 501-525.

[45] S. E. Oraby, A. F. Al-Modhuf, and D. R. Hayhurst, A diagnostic approach for turning tool based on the dynamic force signals, Journal of Manufacturing Science and Engineering 127 (2005) 463-475.

[46] J. Caiado, N. Crato, and D. Peña, An interpolated periodogram-based metric for comparison of time series with unequal lengths, 2006. Published in the Munich Personal RePEc Archive. MPRA Paper No. 2075, posted 07. November 2007 / 02:12.

[47] J. Caiado, N. Crato, and D. Pena, A periodogram-based metric for time series classification, Computational Statistics & Data Analysis 50 (2006), 2668-2684.

[48] R. K. Mobley, An Introduction to Predictive Maintenance, Butterworth-Heinemann, New York, second edition, 2002.

[49] C. Byington, M. Watson, D. Edwards, and P. Stoelting, A model-based approach to prognostics and health management for flight control actuators, in 2004 IEEE Aerospace Conference Proceedings, volume 6, Big Sky, MT 3551–3562.

[50] M. Watson, C. Byington, D. Edwards, and S. Amin, Dynamic modeling and wear-based remaining useful life prediction of high power clutch systems, STLE Tribology Transactions 48 (2005) 208-217.

[51] M. J. Watson, C. S. Byington, and A. Behbahani, Very high frequency monitoring system for engine gearbox and generator health management,in Proceedings of the SAE AeroTech Congress & Exhibition, 2007-01-3878.

[52] G. Vachtsevanos, F. L. Lewis, M. Roemer, A. Hess, and B. Wu, Intelligent Fault Diagnosis and Prognosis for Engineering Systems, John Wiley & Sons, Inc., Hoboken, New Jersey, 2006.

[53] M. J. Roemer, C. S. Byington, and G. J. Kacprzynski, An overview of selected prognostic technologies with application to engine health management, in Proceedings of GT2006 ASME Turbo Expo 2006: Power for Land, Sea, and Air, no. GT2004-54135, Barcelona, Spain.

[54] M. J. Roemer and C. S. Byington, "Prognostics and health management software for gas turbine engine bearings," in Proceedings of GT2007 ASME Turbo Expo 2007: Power for Land, Sea, and Air, 4790X, ASME, Montreal, Canada, 2007, 795-802.